THE COMPLETE
TECHNICAL PAPER PROCEEDINGS
FROM:



5G Fixed Wireless Access

A Technical Paper prepared for SCTE by

Mikko Jarvela
Solution Manager
Nokia
3201 Olympus Blvd, Dallas TX 75019
(469) 831-5973
mikko.jarvela@nokia.com

Harri Holma
Senior Advisor
Nokia
Karakaari 7, 02610 Espoo Finland
+358405232609
harri.holma@nokia.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. 5G Downlink Capacity Evolution ... 3
3. Uplink Capacity Evolution ... 6
   - 3.1. Uplink SU-MIMO baseline ... 6
   - 3.2. Uplink MU-MIMO ... 7
   - 3.3. UL-MIMO impact on user data rate ... 8
   - 3.4. 5G Uplink 3-layer CA ... 8
   - 3.5. Uplink Throughput Evolution Summary ... 9
4. Slicing for FWA ... 9
5. mmWave for FWA ... 13
   - 5.1. NR Dual Connectivity with mmWave ... 14
6. Conclusion ... 15

Abbreviations ... 16

Bibliography & References ... 16

List of Figures

Title **Page Number**

Figure 1 – 5G TDD cell capacity in live networks ... 3
Figure 2 – 5G Multi-user MIMO gain in a live network ... 4
Figure 3 – Capacity gain due to CPE optimization ... 5
Figure 4 – 5G TDD FWA cell capacity evolution steps ... 5
Figure 5 – 5G TDD FWA subscribers per cell ... 6
Figure 6 – Uplink Spectral Efficiencies in Networks Today (SU-MIMO) ... 7
Figure 7 – Capacity Gain of Uplink MU-MIMO ... 8
Figure 8 – 5G UL-MIMO gain in field tests ... 8
Figure 9 – 5G Uplink 3-layer CA Solution ... 9
Figure 10 – 5G Slicing architecture for RAN ... 10
Figure 11 – 4G/5G FWA slicing end-to-end deployment ... 10
Figure 12 – Balancing traffic between eMBB and FWA slices ... 11
Figure 13 – Average DL throughput (Mbps) before and after enabling RRP ... 11
Figure 14 – Average user throughput view for FWA slice ... 12
Figure 15 – Impact of low latency feature application on slot length statistics ... 12
Figure 16 – Low latency optimized slice example ... 12
Figure 17 – Percentage of households that can be offloaded to mmWave from existing macro sites ... 13
Figure 18 – Field trial results for 11 km extended range mmWave site ... 14

List of Tables

Title **Page Number**

Table 1 – Number of Subscribers Supported in Uplink vs MIMO technology ... 7
Table 2 – Estimated Throughput for 5G Uplink Capacity Evolution Steps ... 9

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

5G radio with the new frequency bands such as C-band provides significant capacity through increased bandwidth and higher spectral efficiency potential with mMIMO. This high capacity can be efficiently utilized by Fixed Wireless Access (FWA) customers. The high radio efficiency is important for FWA services because typical FWA customers consume more data than mobile customers. The need for a 5G FWA service is high in many markets because the fiber rollouts are expensive and take time, and because 4G LTE radio capacity is typically too limited for a FWA service.

This white paper analyses 5G radio capacity, the impact of device and end-to-end optimization, the benefits to FWA provided by network slicing, and the usage of millimeter wave (mmWave) spectrum in FWA applications. The field results show that 5G radio with 100 MHz spectrum can support more than 100 FWA subscribers per cell. The experience also shows that the device (Customer Premises Equipment, CPE) and end-to-end optimization can boost FWA performance. Further capacity and even higher data rates can be provided with 5G mmWave. Existing base station sites with mmWave can offload up to 50% of sub-6 GHz traffic. The default radio scheduler implementation does not differentiate between data services (eMBB vs. FWA) but with slicing, scheduling weights and radio resource partition can be separately optimized per service. With slicing, low-latency features can be also applied and optimized selectively for FWA service.

A 5G FWA service allows mobile operators to efficiently monetize their current excess 5G TDD radio capacity without waiting for a 5G core network and standalone architecture for other new use cases. 5G FWA is an untapped capability offered by 5G radio which can be monetized today.

2. 5G Downlink Capacity Evolution

A typical mobile subscriber uses 20 GB of data per month while a fixed network subscriber uses 20 times that amount: 400 GB per month or more. Therefore, FWA services requires very high radio capacity. 5G radio enables an attractive starting point for FWA because 5G is designed to provide high capacity and high data rates.

The main solutions enabling high capacity are the new mid-band TDD spectrum and massive MIMO beamforming antenna. Figure 1 illustrates the practical 5G TDD capacity in live networks today using 100 MHz bandwidth using either single user MIMO (SU-MIMO) or multi-user MIMO (MU-MIMO). The graphs illustrate the carried traffic in terms of gigabytes per hour cell as a function of cell Physical Resource Block (PRB) loading. Each dot represents data in a single cell. A linear trendline is added to the graphs to illustrate the achievable capacity with 100% loading.

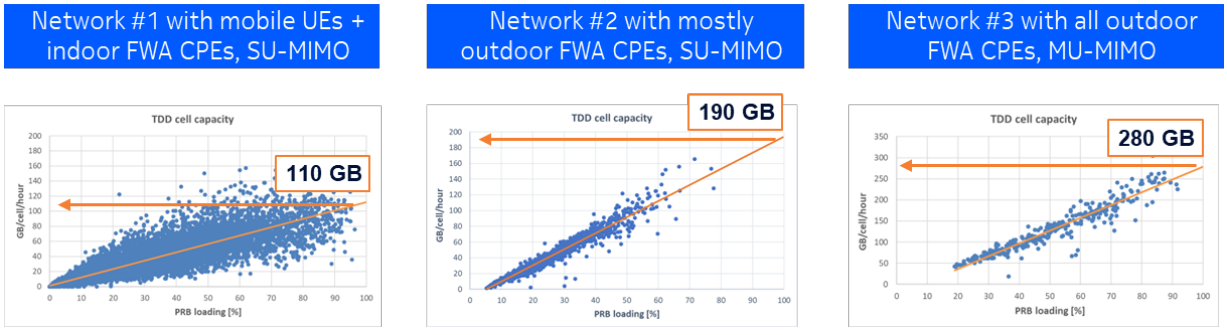


Figure 1 – 5G TDD cell capacity in live networks

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The results show that current 5G with SU-MIMO supports 110-190 GB/cell/hour. Network #1 shows 110 GB/hour with mostly mobile devices and indoor FWA units while Networks #2 has relatively more FWA traffic and some outdoor units. The different types of traffic and devices can explain part of the difference in the maximum capacities. Network #3 shows that the current 5G FWA capacity can be as high as 280 GB/hour in a cell with all FWA outdoor CPEs and when MU-MIMO is enabled.

Let's then look in more detail at the steps to boost the 5G radio capacity for the FWA use case starting from the 110 GB/cell/hour baseline we see in live networks today (Figure 3). The main technology components are Multiuser-MIMO (MU-MIMO), FWA specific optimization, FWA CPE RX enhancement, and a second phase of extreme MU-MIMO with interference cancellation.

1. MU-MIMO refers to the beamforming solution where multiple devices can reuse the same resources in the time and frequency domain in the different beams. The field results indicate that MU-MIMO provides typically 50-70% capacity gains during high load (see Figure 2).

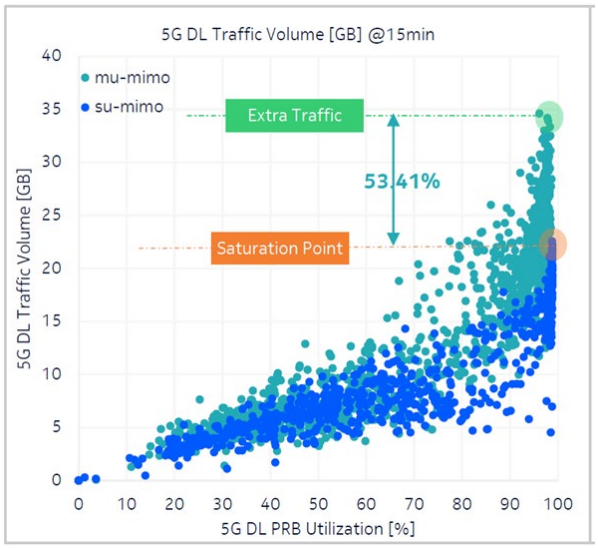


Figure 2 – 5G Multi-user MIMO gain in a live network

We assume a 50% MU-MIMO gain in the evolution steps model below.

2. FWA CPE performance can be optimized including CPE antenna type and CPE locations. This provides a major capacity enhancement potential. CPE outdoor antenna provides a 60% capacity gain, and its deployment can be selective based on location.

   CPEs in a weak signal area cause two challenges: 1) the data rate is low especially in the uplink, and 2) the CPE consumes more downlink resources because of the high power required. As shown in Figure 3, a 29% capacity gain can be expected by moving weak signal CPEs to a better position (7dB signal level improvement assumed). 62% capacity gain was seen by using outdoor antennas for weak signal CPEs (15 dB signal level improvement assumed).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

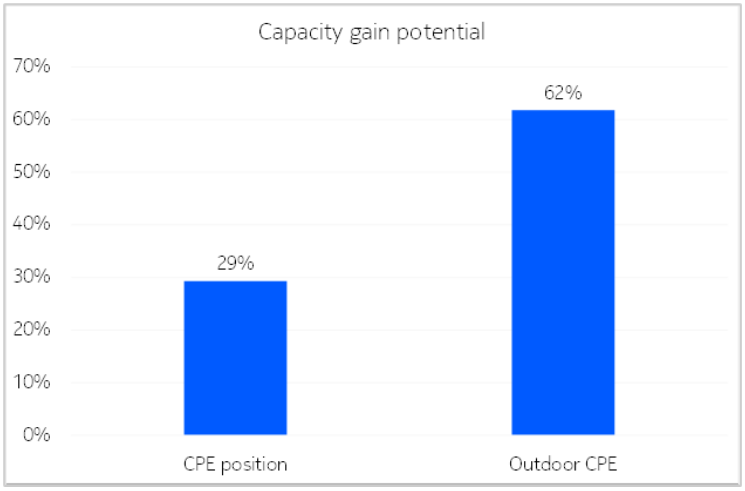


Figure 3 – Capacity gain due to CPE optimization

We assume in the following capacity evolution analysis a 40% gain due to outdoor CPE antenna based on the signal level statistics collected from an indoor FWA CPE.

3. FWA optimization in the radio network refers to multiple solutions for specifically optimizing the FWA service including beamforming, packet scheduler, traffic steering and slicing. We assume here 20% gain from FWA optimization.

4. CPEs with 8RX receiver can boost capacity by 10-15% according to measurements.

5. Extreme MU-MIMO refers to the further enhancement of MU-MIMO for minimizing interference between co-scheduled devices with Zero-Forcing algorithms. We assume here 25% gain.

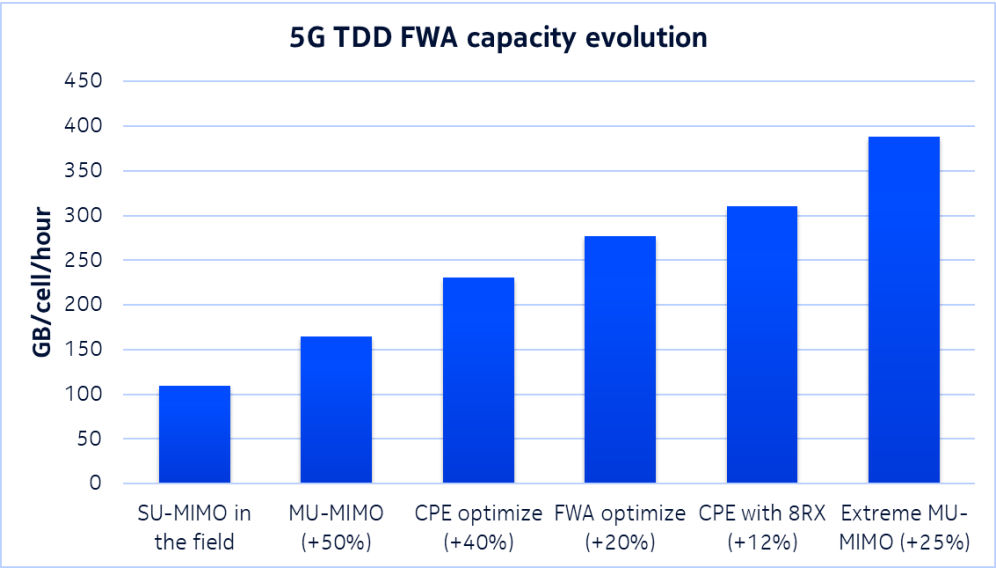


Figure 4 – 5G TDD FWA cell capacity evolution steps

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

When combining these steps together, the expected cell capacity will grow from the current lowest figure of 110 GB/cell/hour to approximately 390 GB/cell/hour.

Figure 5 illustrates the maximum number of FWA subscribers per cell assuming that the average usage is 500 GB/month, maximum 90% cell loading is allowed during peak hour, with the peak accounting for 10% of the daily traffic which are typical figures in the current FWA deployments. The typical FWA subscriber data usage is 300 – 500 GB/month currently. The results show that up to 150-200 FWA subscribers can be supported with a single 5G TDD sector using 100 MHz bandwidth.

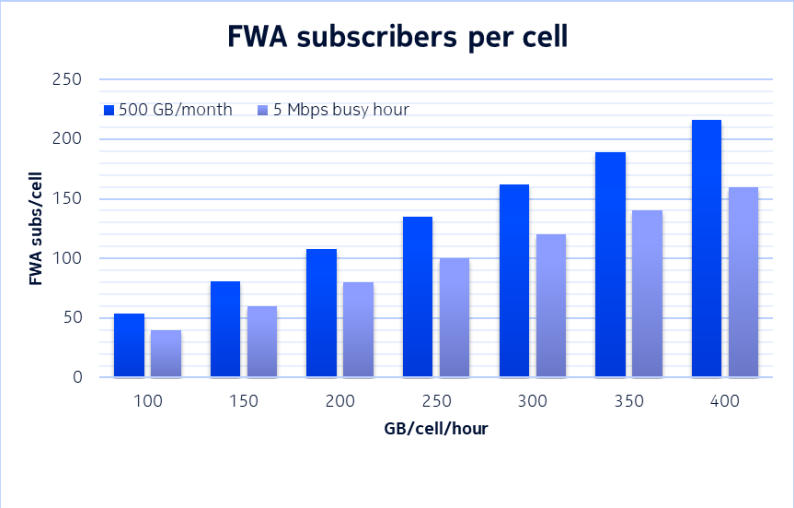


Figure 5 – 5G TDD FWA subscribers per cell

3. Uplink Capacity Evolution

This chapter provides insights into 5G uplink capacity illustrating typical spectral efficiencies that are currently seen, and showcasing advanced technologies that can further improve the uplink capacity in the future.

3.1. Uplink SU-MIMO baseline

The median uplink spectral efficiency among radio networks today is 2.2 bps/Hz when using SU-MIMO (Single User MIMO). This is illustrated in Figure 6 below.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

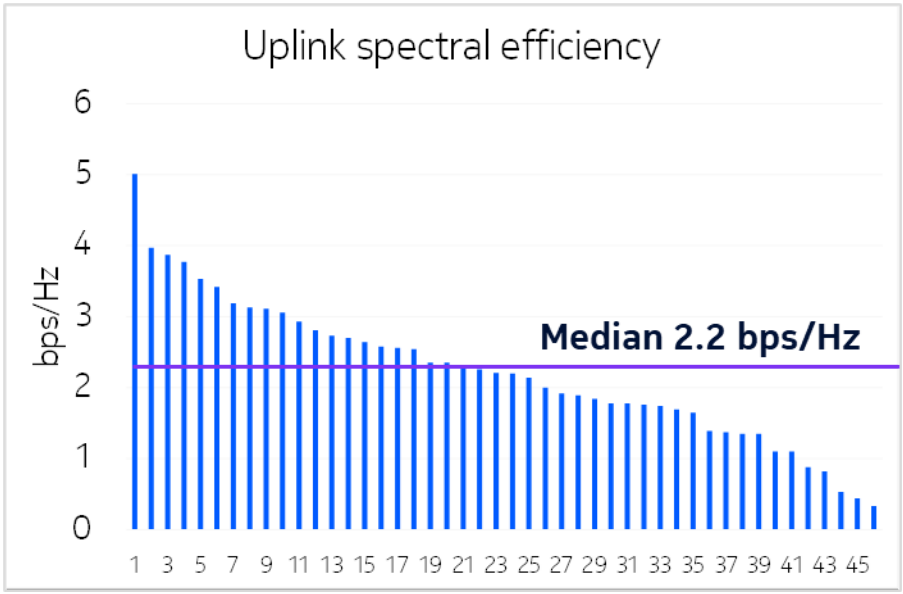


Figure 6 – Uplink Spectral Efficiencies in Networks Today (SU-MIMO)

3.2. Uplink MU-MIMO

In terms of number of subscribers supported, SU-MIMO median spectral efficiency shown above would translate to 21 users per cell with 10 Mbps uplink, with assumptions as shown in Table 1. With MU-MIMO, the spectral efficiency is estimated to increase to 5.4 bps/Hz which would mean that 50 users can be supported at 10 Mbps.

Table 1 – Number of Subscribers Supported in Uplink vs MIMO technology

MIMO Scheme	SU-MIMO	MU-MIMO
Spectral Efficiency	2.2 bps/Hz	5.4 bps/Hz
Cell Throughput (100 MHz TDD carrier)	52 Mbps	125 Mbps
User Activity	25%	25%
Average Throughput of 10 Mbps User	2.5 Mbps	2.5 Mbps
Number of Users Supported at 10 Mbps	21 Users	50 Users

Figure 7 shows an example comparison of uplink data volumes achieved with MU-MIMO and SU-MIMO. It illustrates that the uplink data volume increases by more than 70% with MU-MIMO activation even as the PRB loading goes down. The spectral efficiency is increased by a factor of 2.5.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

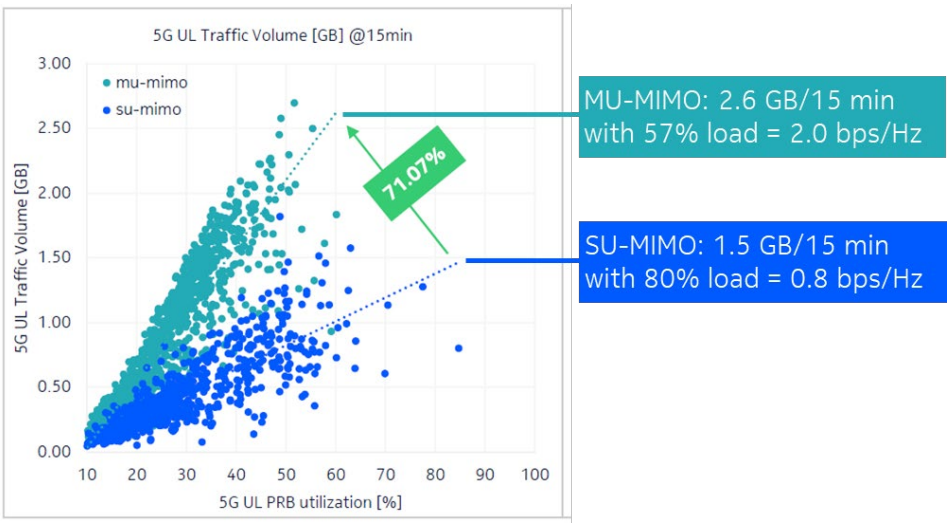


Figure 7 – Capacity Gain of Uplink MU-MIMO

3.3. UL-MIMO impact on user data rate

One benefit of 5G over LTE is that in uplink 5G devices support 2 transmit paths compared to only 1 transmit path in LTE. Therefore, 5G supports rank 2 UL-MIMO (Uplink MIMO) whereas LTE devices only support rank 1. Although standardized since Rel10, in practice there is no UL-MIMO in LTE, and it has always been limited to UL-CA which also has had little success.

In NR, NSA UEs always needed to have 2TX (one for the uplink LTE leg, one for the UL-NR leg). It means that for existing devices UL-MIMO has also become feasible in NR SA (Standalone) operation. Currently UL-MIMO is limited to NR TDD but it is forecasted to also be supported for NR FDD in the future.

Figure 8 shows UL-MIMO gains achieved with a regular UE in drive tests. With 64QAM, the UL-MIMO the user throughput gain was +61%, and with 256QAM the gain was +52% when compared with rank 1 throughput.

Field Performance (NR, cmW TDD, 64TRX, n77, 100MHz)	NSA (1Tx) Concurrent	SA (2Tx) Concurrent	
Drive Route 1, TCP Uplink Single UE Throughput 64QAM, Average, 1086 gain %	~3.4% (SSB: 82.7 Mbps, 1086: 85.4 Mbps)	~8.7% (SSB: 126.7 Mbps, 1086: 137.7 Mbps)	+61% 2TX gain
Drive Route 1, TCP Uplink Single UE Throughput 256QAM, Average, 1086 gain %	~11.1% (SSB: 86.3 Mbps, 1086: 95.9 Mbps)	~11.7% (SSB: 130.6 Mbps, 1086: 145.9 Mbps)	+52% 2TX gain

Figure 8 – 5G UL-MIMO gain in field tests

3.4. 5G Uplink 3-layer CA

3-layer uplink CA (Carrier Aggregation) combines UL-MIMO solution described above with Uplink CA and is specifically designed for FWA use cases, see Figure 9. This solution will initially be available for FDD+TDD CA scenario, however 3GPP Release 17 also defines the TDD+TDD solution.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The solution extends FR1 TDD + FDD uplink Carrier Aggregation so that UL-MIMO can be used in the TDD carrier for devices that are capable of 3TX transmission.

It is expected that this solution will provide the following improvement of peak throughput compared to 3GPP Release 15 uplink CA solution:

- +75% uplink throughput improvement (100 MHz TDD + 10 MHz FDD)
- +65% uplink throughput improvement (100 MHz TDD + 20 MHz FDD)

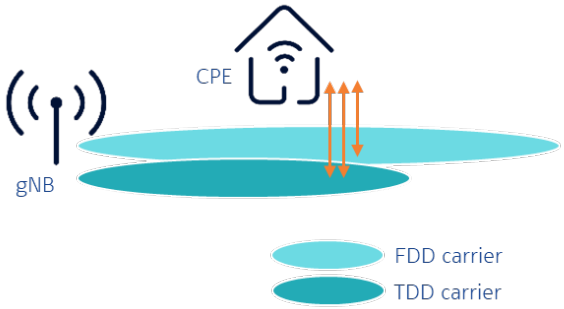


Figure 9 – 5G Uplink 3-layer CA Solution

Additional 5G uplink capacity evolution solutions are discussed in Chapter 5.

3.5. Uplink Throughput Evolution Summary

Table 2 summarizes the estimated 5G uplink capacity evolution steps in terms of cell throughput gain and user peak throughput gain.

Table 2 – Estimated Throughput for 5G Uplink Capacity Evolution Steps

	UL SU-MIMO	UL MU-MIMO	UL MIMO	UL 3-layer CA 20 MHz FDD
Cell Throughput (100 MHz TDD)	52 Mbps	125 Mbps		
User Peak Throughput (100 MHz TDD)	110 Mbps		220 Mbps	320 Mbps

4. Slicing for FWA

With 5G slicing, logical networks with varying characteristics and use cases can be created on top of common 5G RAN infrastructure.

The following characteristics can be defined by operators individually per slice:

- Slice-specific Performance Management
- Radio Scheduling Priority (scheduling weight)
- Radio Resource Partition
- Radio Latency Parameters

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Radio Layer Management
- Radio & Core Security Parameters

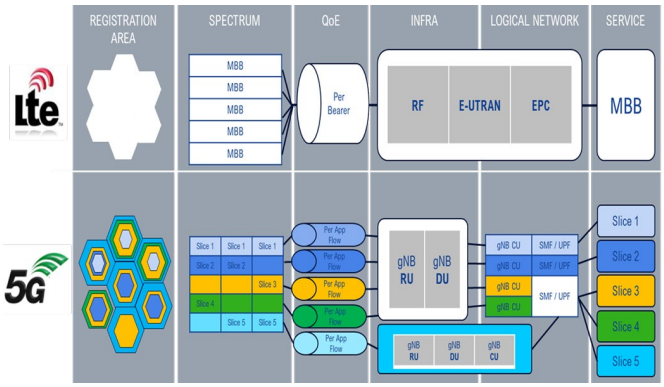


Figure 10 – 5G Slicing architecture for RAN

Slicing solution for FWA enables high performance sliced internet access services – see Figure 11 for high level architecture of the solution. Different FWA service tiers and subscriptions can be established with slicing. Slicing allows setting up various service and application category slices for FWA users, e.g. internet, voice, IPTV, streaming, gaming etc. The network load can be dynamically balanced between FWA users and mobile users, while ensuring a minimum capacity for the FWA users.

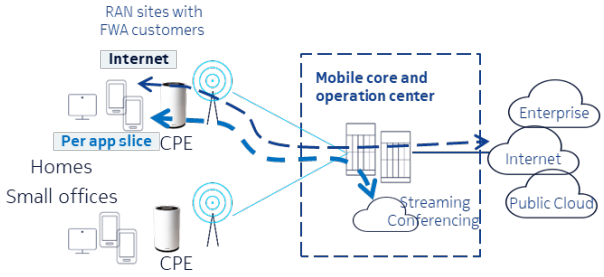


Figure 11 – 4G/5G FWA slicing end-to-end deployment

Let's look in detail the following key benefits of the FWA slicing solution inside 5G RAN:

- Enhanced radio scheduling provides dedicated scheduling weight and radio resource partitioning per slice
- Optimization of frequency layer management allowing separate mobility and handover frequency targets and thresholds
- Enhanced performance visibility with separate KPIs per slice
- Optimization of low latency services using scheduling request optimization per slice

Default radio scheduler implementation does not differentiate between data services (eMBB vs. FWA) With slicing, the scheduling weights and radio resource partition can be separately optimized per service.

Figure 12 below shows an example where unwanted FWA traffic peaks can saturate the cell capacity on a site. Slice based radio resource partitioning was applied to better balance the traffic between the eMBB and FWA services.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

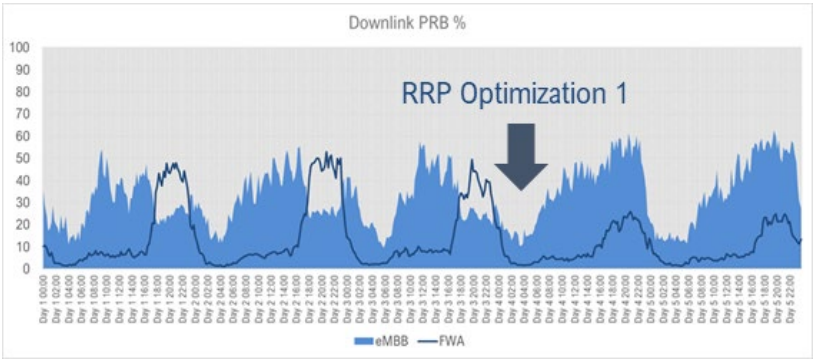


Figure 12 – Balancing traffic between eMBB and FWA slices

Another use case for slicing based radio resource partitioning is shown in Figure 13 below. Prior to applying RRP, the FWA end user experience was inconsistent resulting in highly varying downlink throughput. After activating RRP the biggest dips in the downlink throughput were eliminated.

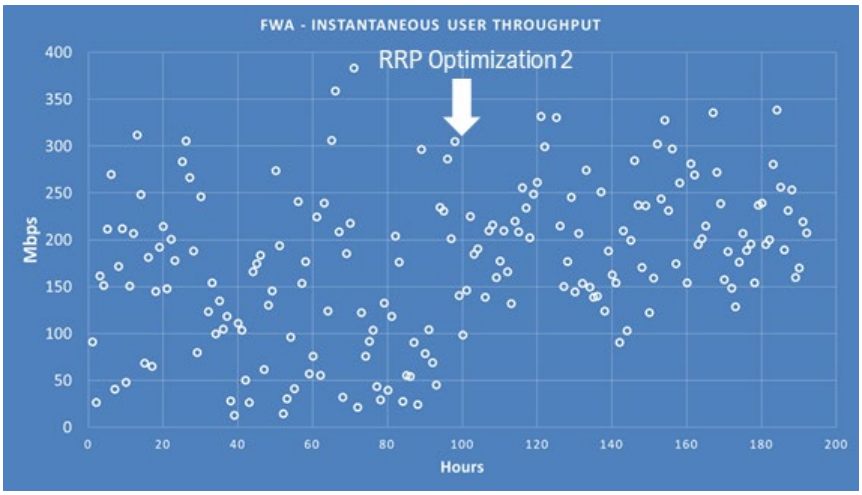


Figure 13 – Average DL throughput (Mbps) before and after enabling RRP

Regular network counters do not give sufficient visibility between handset and CPE performance. With slicing, the network and device performance can be differentiated, resulting in an enhanced view per service, as shown in the example of Figure 14.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

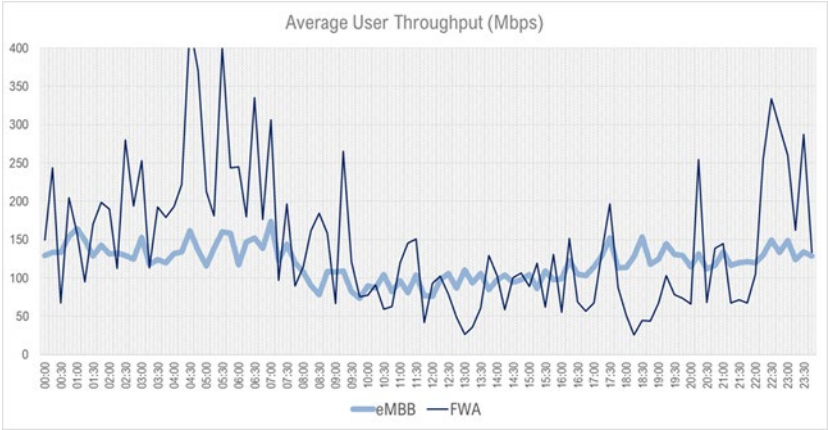


Figure 14 – Average user throughput view for FWA slice

Slicing enables application and optimization of low latency features selectively per service.

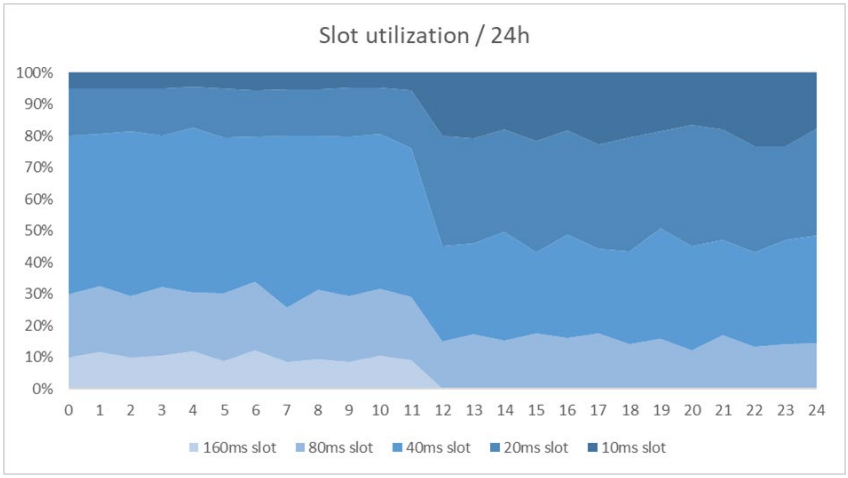


Figure 15 – Impact of low latency feature application on slot length statistics

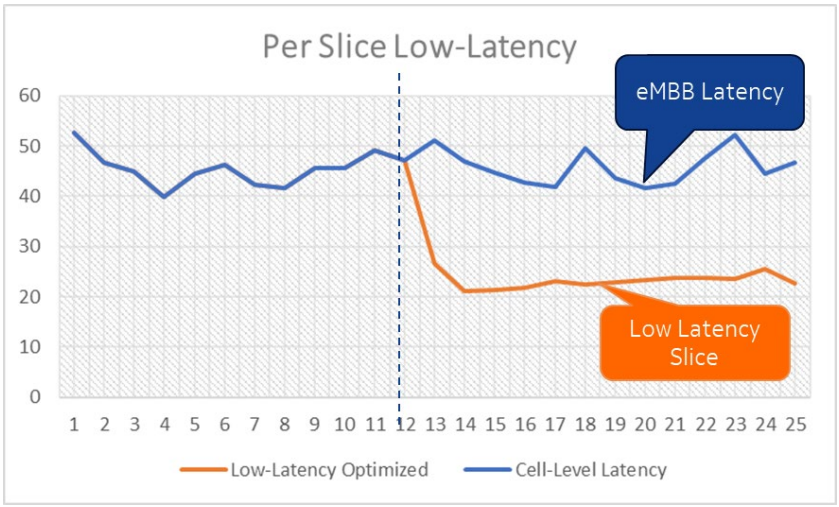


Figure 16 – Low latency optimized slice example

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

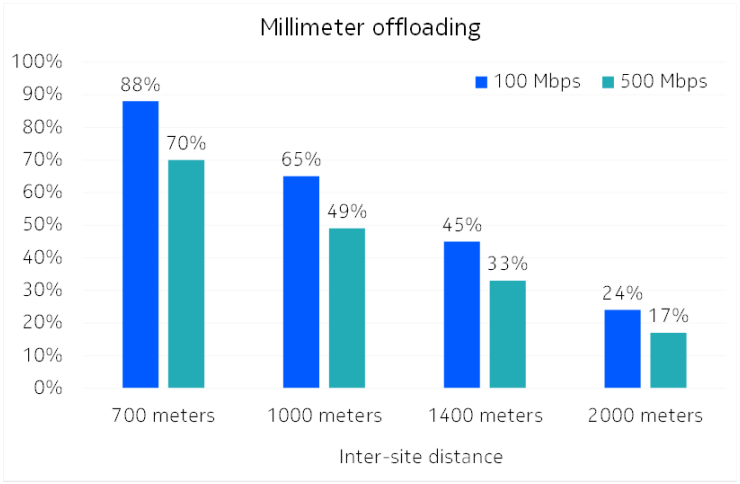
5. mmWave for FWA

Using mmWave bands for smartphones is challenging due to coverage limitations. The average share of mmWave availability in dense urban areas is 2-3% which is very low compared to mid-band TDD share of 60-70%.

However, mmWave can be attractive for sub-urban FWA even when using existing sites to deploy. Typically, an operator has access to more spectrum bandwidth on mmWave bands (400-800 MHz) vs midband TDD (up to 100 MHz). High gain CPE antenna technology can be used to optimize the mmWave link budget even for Non-line of sight (NLOS) propagation conditions.

Nokia has demonstrated throughput of 3.5 Gbps at 6 km distance. Therefore, in Line-of-sight (LOS) conditions, premium fiber-like service can be offered over distances of 6 km or more. Even in areas with high foliage, typically more than 100 users can be reached in LOS conditions from the existing base station sites.

A suburban case study was conducted to illustrate the potential to offload sub-6 GHz traffic to mmWave using existing macro base station sites and considering the realistic propagation conditions. 3-sector mmWave deployment was assumed with 30 m antenna height. The study assumed a 17 dBi outdoor rooftop antenna for the home CPE. The results are shown in Figure 17 for two different downlink service levels: 100 Mbps and 500 Mbps. In cases with a 1 km inter-site distance, mmWave connection is available in 50-65% of households. In cases with a 1.4 km inter-site site distance, mmWave offloading with 100 Mbps service still works for 45% of households.





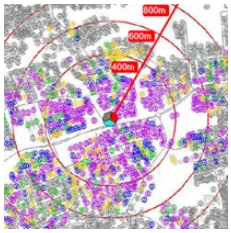


Figure 17 – Percentage of households that can be offloaded to mmWave from existing macro sites

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Similarly, another measurement campaign with high gain mmWave CPE antenna was executed in a suburban environment with limited foliage but mostly NLOS propagation conditions. The results showed that throughputs of 700-800 Mbps can be achieved consistently over distances of 500 meters from the base stations with extensions even up to 900 meters demonstrating efficient offload capability to mmWave.

A mmWave FWA field trial was conducted recently that demonstrates the potential of mmWave for providing extreme radio coverage. In the trial a sustained 2.1 Gbps downlink data rate connection was established using mmWave over a distance of 11 km. In the uplink, 57 Mbps throughput was achieved. 26 GHz mmWave band (n258) was utilized with 8 x 100 MHz component carriers and NSA architecture with LTE 2100 as the anchor band. A high gain 27 dBi antenna at 21 meters height was used with the CPE with an EIRP of 47 dBm. The gNB mmWave radio was radiating at 60 dBm power at the height of 46 meters.

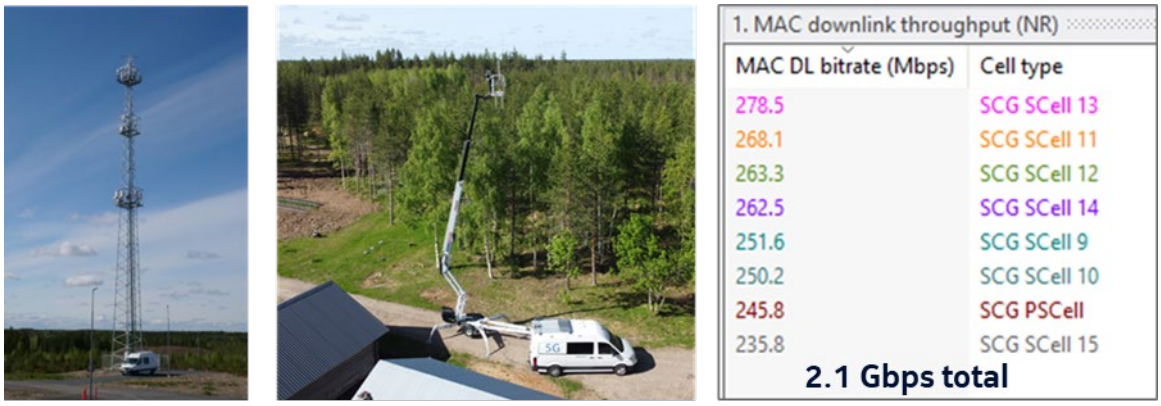


Figure 18 – Field trial results for 11 km extended range mmWave site

This achievement demonstrates the reach and connectivity speeds that 5G mmWave can deliver. It also lays the foundations for high-quality internet connectivity solutions delivered via FWA to areas where wired connections are not always possible.

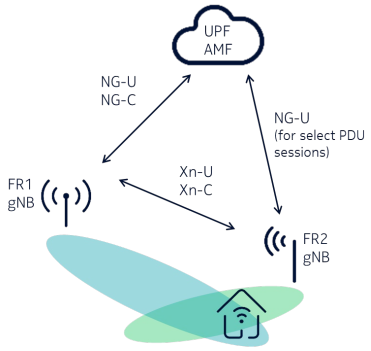
5.1. NR Dual Connectivity with mmWave

A key evolution step for mmWave FWA over the next few years is FR1-FR2 NR Dual Connectivity (NR-DC) functionality which provides access to the high throughput of FR2 while maintaining the connection stability of FR1.

Figure below shows the principle of FR1-FR2 NR-DC functionality. The Primary Cell (PCell) belongs to an FR1 band (FDD or TDD) and the Primary Secondary Cell (PSCell) uses an FR2 band. The total

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

throughput is estimated to reach up to 6-8.5 Gbps when a 100 MHz TDD carrier is combined with 8x100 MHz of FR2 spectrum in downlink.



The NR-DC functionality described above can be extended to include uplink CA as well to increase uplink throughput. Up to 4 uplink Component Carriers (CC) can be supported in FR2. In addition, FR1 Carrier Aggregation (CA) can be supported for FR1 to further increase the downlink throughput.

6. Conclusion

5G technology combined with mid-band TDD spectrum and massive MIMO antenna can provide a major boost to the radio capacity that it can support also Fixed Wireless Access services for home broadband connectivity. The analysis shows that up to 150-200 FWA subscribers per 100 MHz sector can be supported on the existing base station grid without adding new sites. High capacity on existing sites enables monetization of 5G for FWA services already today. Evolution steps were presented showing the future potential for further capacity enhancements for both downlink and uplink.

5G RAN slicing can provide several benefits to efficiently manage FWA user traffic load in the network. The radio resource can be partitioned per slice, guaranteeing the bandwidth for both mobile and FWA services. Frequency layer management allows separate mobility and handover frequency targets and thresholds for the different slices. The performance can also be monitored and visualized with separate KPIs per traffic type. Scheduling for low latency services can also be optimized per slice.

Incremental mmWave spectrum can triple FWA network capacity in NLOS conditions. It is possible to offload even 50% of households to mmWave while using existing base station sites. Combining mmWave with sub-6 GHz spectrum enables more aggressive mmWave radio planning, delivering more mmWave offloading while always providing a robust service with sub-6 GHz. In LOS conditions, mmWave enables fiber-like services over distances over 6 km, while providing enough capacity to connect hundreds of users per cell.

An extended range mmWave trial demonstrates that mmWave solutions can be an essential building block for operators to efficiently deliver widespread, multi-gigabit 5G broadband coverage to their customers in urban, suburban, and rural areas, complementing sub-6 GHz spectrum assets.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

bps	bits per second
CA	Carrier aggregation
CC	Component carrier
CPE	Customer premises equipment
DC	Dual connectivity
eMBB	Enhanced mobile broadband
FDD	Frequency division duplex
FWA	Fixed wireless access
GB	Gigabyte
LOS	Line of sight
MIMO	Multiple-input multiple-output
MU-MIMO	Multi-user MIMO
NLOS	Non-line of sight
NSA	Non-standalone
PCell	Primary cell
PRB	Physical resource block
PSCell	Primary secondary cell
RAN	Radio access network
RRP	Radio resource partitioning
RX	Receive
SA	Standalone
SCTE	Society of Cable Telecommunications Engineers
SU-MIMO	Single user MIMO
TDD	Time division duplex
TX	Transmit

Bibliography & References

499104_213179_FWA_White_Paper, Nokia

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23 CELEBRATING 40 YEARS

2023 Fall Technical Forum SCTE • CABLELABS • NCTA

A Pragmatic Approach to Deploying Broadband in LATAM with Standards-Based PON Technologies

A Technical Paper prepared for SCTE by

Marko Estacio
Sales Solution Architect
CommScope
AK 9 # 113-52 Of. 1606, Bogota Colombia
Marko.Estacio@CommScope.com

Brian Lane
Senior Product Manager
CommScope
3871 Lakefield Drive Suwanee, GA 30024
Brian.Lane@CommScope.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. The broadband opportunity ... 3
3. The broadband landscape in Latin America ... 4
4. Scenarios for adding broadband capacity ... 5
5. Technologies of choice for broadband deployments ... 6
   5.1. Hybrid Fiber Coax ... 6
   5.2. Passive Optical Networks ... 6
   5.3. Fixed Wireless Access ... 6
6. Selecting the right approach ... 6
   6.1. Success-based deployments ... 7
   6.2. Greenfield footprint expansion ... 8
   6.3. Greenfield deployments ... 8
   6.4. Brownfield and overbuild ... 8
7. The PON overlay approach ... 8
   7.1. Pre-connectorized technology for cost-effective, speedy deployments ... 10
   7.2. The advantage of R-OLT over C-OLT ... 10
8. Managing and scaling the PON ... 11
9. Conclusion ... 12

Abbreviations ... 12

Bibliography & References ... 12

List of Figures

Title **Page Number**

Figure 1 - Broadband users in key LATAM countries in Jan 2023 (in millions) [1] ... 4
Figure 2 - More speed does not always result in higher ARPU [2] ... 5
Figure 3 - Proportion of urban and rural areas in Latin America [3] ... 5
Figure 4 - Illustrative example of distribution of usage among subscribers [4] ... 7
Figure 5 - FTTH overlay network with pre-connectorized tap terminals ... 9
Figure 6 - R-OLT in existing node ... 9
Figure 7 - Comparing C-OLT to R-OLT (CommScope) ... 11

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Today, broadband is recognized as an essential utility, on par with electricity and water. Gone are the days when broadband was used only for web browsing and entertainment. Today, education, healthcare delivery, and many essential services are delivered over broadband. Furthermore, today, broadband is a fundamental part of business models in most industries, particularly as the world goes through a tsunami of digital transformation. From video streaming to digital healthcare, remote education, wireless backhaul, work from home, intelligent farming, the metaverse and so many others, the opportunities are endless.

This has caused broadband consumption to soar and has created new requirements for operators such as reliability, latency, security, to name a few. It has led operators worldwide to increase capacity in served areas, and to extend broadband deployments to previously unserved areas. In doing so, they must juggle numerous parameters that include technology solutions, capex availability, competition, ARPU trends, and regulation to name a few.

2. The broadband opportunity

These days, broadband is no longer merely about connectivity. It has become a lifeline, a basic infrastructure, powering businesses and consumer markets. Some of the current and emerging broadband opportunities for operators include:

- Residential
- Small/Medium Businesses
- 4/5G Mobile Backhaul
- WiFi Backhaul
- IoT, Digitization, M2M
- Wholesale

These transforming markets are creating new opportunities, leading to new entrants and a hyper-competitive business environment but also challenges for operators: Ubiquitous broadband access is becoming essential, and indeed mandated by governments in some countries (though not in Latin America for the most part). Furthermore, the definition of broadband is changing. Gone are the days when broadband meant 25Mbps downstream and 3 Mbps upstream. In the United States, the FCC is pushing a 100Mbps definition of broadband and the market is evolving towards 10G symmetric broadband.

Evolution of the full-service operator. Cable operators today offer fixed line services but also mobile service. Customers increasingly do not view the two media as separate but expect broadband access wherever they are. Fixed and mobile convergence is finally possible.

The lines between business and residential are blurring. More business grade services are now delivered in the home, and this trend will only grow. For example, medical grade services are increasingly delivered in the home, more businesses are being conducted in residential settings and so on. These trends will impose new requirements on broadband delivery in the home, such as low latency, guaranteed bandwidth, security and others.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Augmented reality is now a reality in many businesses, and the metaverse over time will be transformative not only in business environments but also in residential ones.

While Latin America is somewhat of a laggard in this respect, substantial public (and private) funding is going into broadband. In the United States, a staggering $100 billion is being invested in "broadband for all".

3. The broadband landscape in Latin America

The number of users in Latin America is growing rapidly. For example, Brazil, who had 182 million users in January 2023 is expected to have 190 million in 2027.

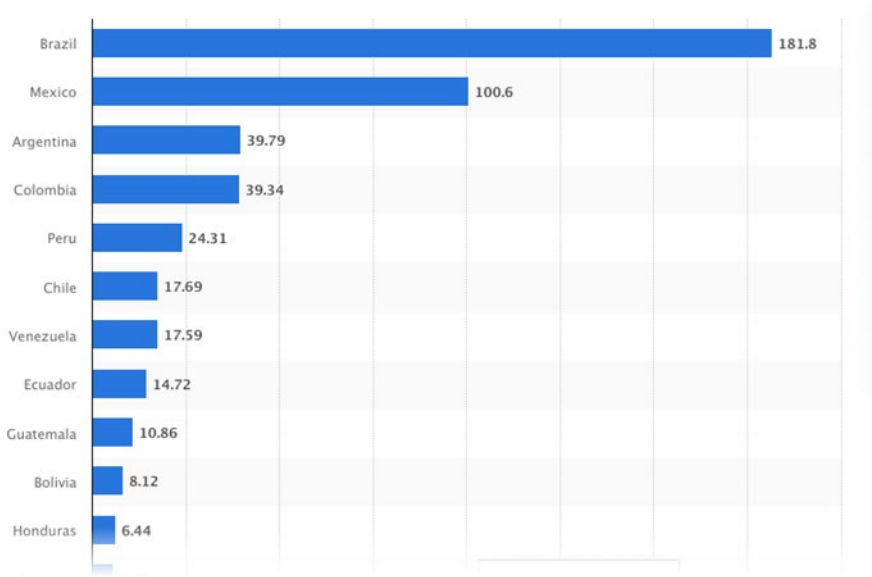


Figure 1 - Broadband users in key LATAM countries in Jan 2023 (in millions) [1]

However, growth in broadband use is not linearly correlated with ARPU growth. This is partly due to lack of affordability and causes challenges for operators who need to invest to add broadband capacity but cannot realize an attractive return on investment. This is compounded by the fact that unlike some developed countries where governments are investing massively in broadband (e.g. the BEAD program in the United States), in Latin America, governments are largely leaving it to the private sector to fund broadband expansion.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

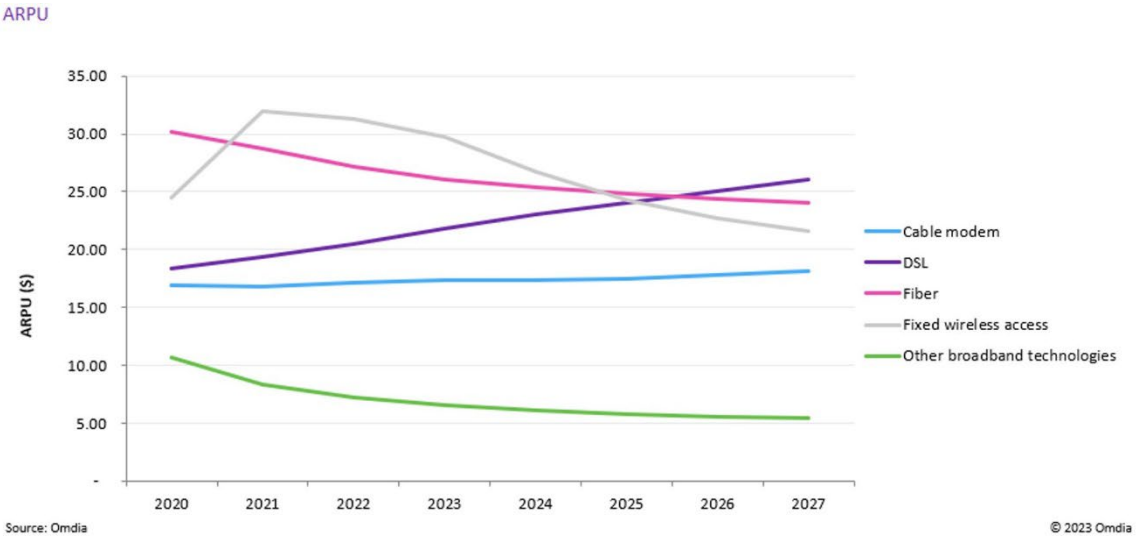


Figure 2 - More speed does not always result in higher ARPU [2]

Latin America and the Caribbean (LAC) is a large and diverse geographical region, encompassing 27 countries and more than 600 million people, and covering nearly 20 million square kilometers of forests, mountain ranges, glaciers, deserts, islands, and urban centers [3].

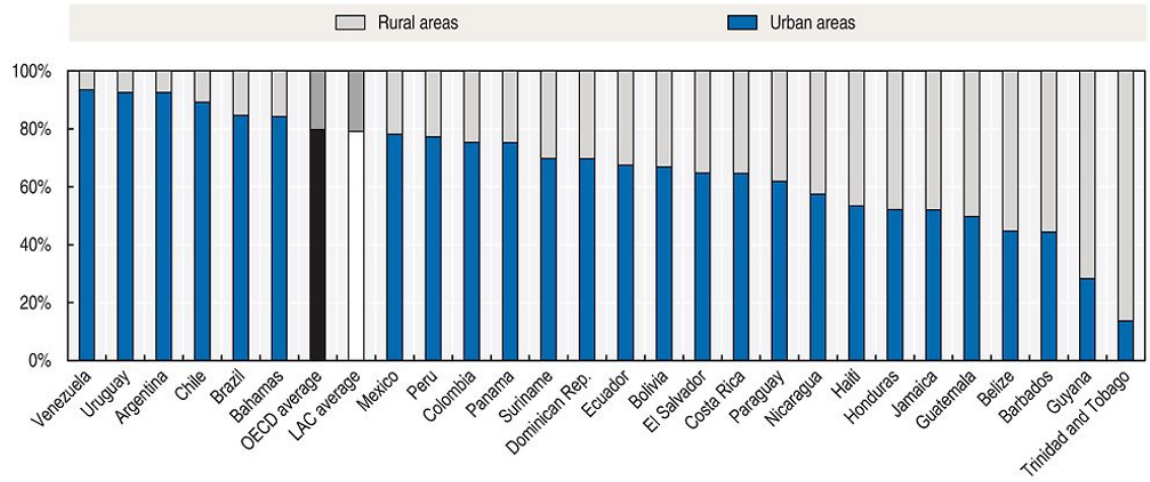


Figure 3 - Proportion of urban and rural areas in Latin America [3]

The regulatory framework in the region is also fragmented between countries, and sometimes within countries.

4. Scenarios for adding broadband capacity

Cable operators have primarily four scenarios for expanding their last mile, driven by market conditions.

- **Success based deployments** – for example developers demanding fiber builds, Small Office Home Office (SOHO) and other subscribers using gigabit symmetric services and businesses requiring increased Service Level Agreements (SLAs).
- **Greenfield** deployments, including multi-dwelling units (MDUs).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- **Greenfield footprint expansion** into adjacent areas that leverage the existing plant to enable more home passings.
- **Brownfield and overbuild** – capacity additions in existing footprint, particularly in the upstream for symmetric services, and for meeting competitive pressure.

5. Technologies of choice for broadband deployments

Cable operators have primarily two technology approaches:

5.1. Hybrid Fiber Coax

HFC is the traditional technology used by most cable operators. It largely uses DOCSIS® protocols and continues to evolve to deliver more capacity upstream and downstream. Cable operators can use DOCSIS® 3.1, and mid-split (or high split in some cases), to add capacity and to improve upstream throughput. Some operators are using the distributed access architecture (DAA) approach to add capacity, either in a Remote-PHY or Remote-MACPHY approach. On the horizon is the migration to DOCSIS® 4.0, already being tested by major MSOs, using either Full Duplex DOCSIS (FDX) or Extended Spectrum DOCSIS (ESD).

Operators who have not committed to DAA and are looking for a technology with high throughput and significant future growth capacity or are deploying in a greenfield area are increasingly turning to passive optical networks (PON).

5.2. Passive Optical Networks

PON uses a point-to-multipoint fiber distribution network where one strand of single-mode fiber serves many end users, by sending data upstream and downstream through unpowered splitters and fiber distribution equipment. These passive elements of the PON are supported by optical line terminals (OLT)s in the network, that connect through the passive network elements to optical network terminals (ONT) which are at the subscriber's location. The PON enables the operator to bring fiber all the way to the subscriber's premise, thus delivering gigabit bandwidth.

However, bringing fiber to the premises can be a costly value proposition. The cost can vary from a few hundred dollars to multi-thousand dollars, depending on the terrain topology, distance, and many other factors. This makes PON deployments particularly challenging in the CALA region where average revenue per user (ARPU) is relatively low compared to the cost of equipment. Therefore, overbuilding an existing HFC network with PON can be cost-prohibitive, and difficult to justify and sustain.

5.3. Fixed Wireless Access

Fixed Wireless Access (FWA) uses radio links between two fixed points to provide wireless broadband. It consists of a base station connected to a fixed network. While widely used by large mobile network operators who own spectrum (for example Verizon and T-Mobile in the United States), FWA is not widely used among cable operators, particularly in areas where newer 5G technologies have limited deployments like in Latin America.

6. Selecting the right approach

PON technologies are typically well suited to meet the capacity needs of high bandwidth users but are often associated with high deployment costs. Therefore, operators should take a measured approach and

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

consider multiple factors and evaluate many options as they set out to meet the needs of heavy users. Factors the operator should consider include:

- Brownfield or greenfield
- Spectrum availability
- Size of service group
- Trajectory of bandwidth usage and longevity of new infrastructure
- Availability of capacity in the HFC plant
- Topography
- Capex and Opex
- Available skillset
- Regulatory considerations
- Cost of overbuilding aerial fiber
- Subscriber mix and traffic consumption habits
- Size of the node and spare capacity

6.1. Success-based deployments

In areas of high bandwidth consumption, the operator can do an analysis to identify whether high bandwidth use is across most of the users in a service group, or a small subset. In some cases, a percentage of subscribers can be responsible for most of the heavy broadband use.

CommScope's internal research has identified the following trends in the CALA region:

- 25% of subscribers are responsible for 75% of bandwidth consumption in Central America.
- 30% of subscribers are responsible for 70% of bandwidth consumption in the Caribbean.

Measurement in service groups across several regions indicated that typically the 20/80 rule applies: 20% of users account for 80% of broadband use.

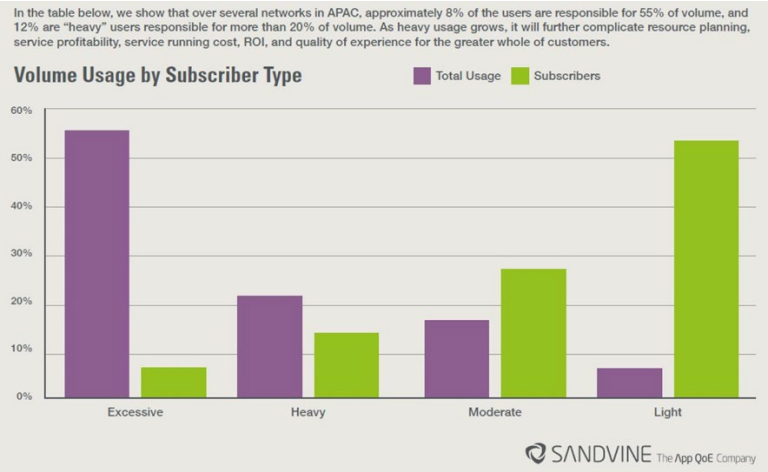


Figure 4 - Illustrative example of distribution of usage among subscribers [4]

For such deployments, operators can use a bandwidth shedding strategy where they use PON to selectively serve customers who pay for a higher tier of service.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6.2. Greenfield footprint expansion

In this case the operator can tap into the existing network to build capacity into the adjacent greenfield area. Having infrastructure, including fiber, hubs or an HFC node in the vicinity of the new deployment enables the operator to expand cost effectively. Effectively, an operator can expand coverage to an adjacent greenfield by deploying a Remote OLT on the outside plant and thus take advantage to build the PON network from a location closer to the user.

The PON can effectively be deployed as an overlay network tapping into existing assets.

6.3. Greenfield deployments

The approach in this case depends on the density of the area, topology, and other factors. For example, in rural areas, or areas of low density, it makes sense to use long haul optics and use a remote PON technology to bring capacity to the area. In dense areas, a large chassis may be more appropriate.

6.4. Brownfield and overbuild

PON can be used effectively to offload traffic, mostly from heavy users, improving the QoE for all subscribers in the overlay area.

7. The PON overlay approach

An operator that has HFC network assets can selectively add capacity using DOCSIS® technology. For example, the operator can upgrade using DOCSIS 3.1 technology, with mid-split or high-split configurations to deliver more upstream capacity. Another common tactic adopted by operators is to decrease the service group size by node segmentation, and in recent years, DAA has been adopted as the preferred HFC capacity expansion method due to benefits such as smaller footprint on hubs and better performance.

However, in areas of market demand from high consumption users, or competitive offerings of symmetric services from other ISPs, PON is the most cost-effective technology to deliver the desired capacity, particularly if deployed as an extension of the HFC plant. If the node has sufficient room, a remote OLT (R-OLT) module can be inserted in the node. This module can initially deliver GPON speeds and provide a more extended reach due to a higher optical budget. If equipped with XGS-PON or dual SFP a typical service group of 128 subscriber can be served per port.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

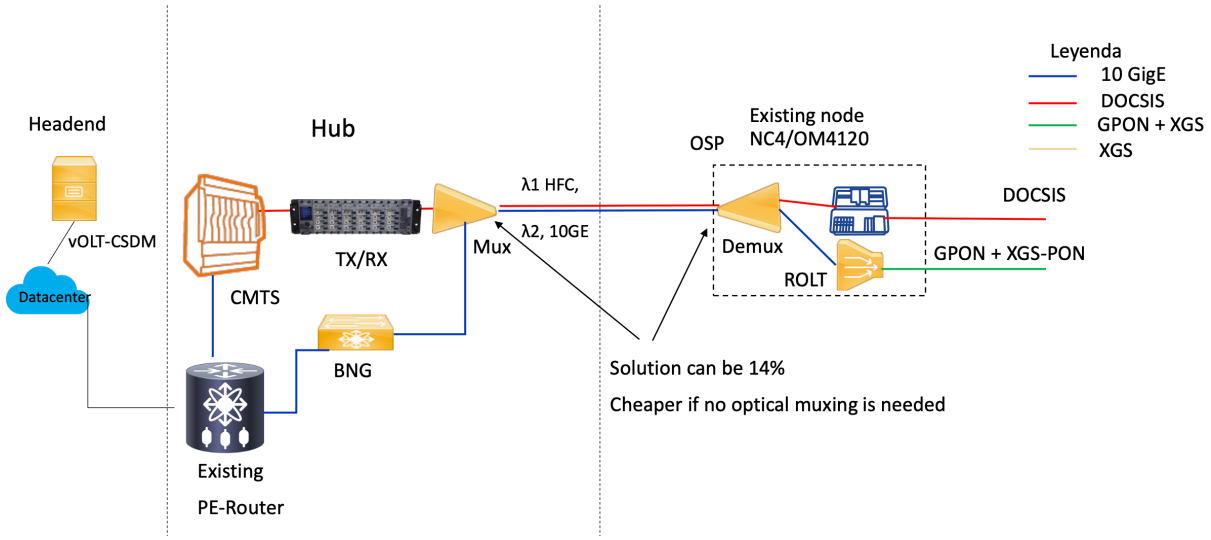


Figure 5 - FTTH overlay network with pre-connectorized tap terminals

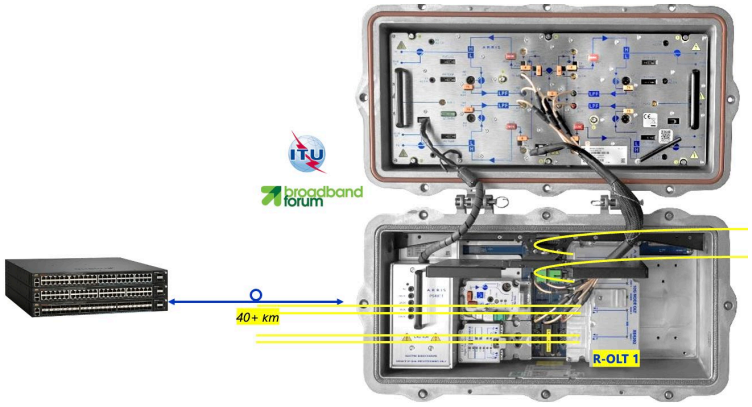


Figure 6 - R-OLT in existing node

There are many benefits for this approach:

- Using an existing cable path from the HFC network with aerial deployment means that no expensive underground feeder is required
- Minimizes future investment in HFC
- Meets the needs of heavy users, but at the same time frees capacity for remaining HFC subs, improving the quality of experience for all
- Solution fully compliant with ITU and Broadband Forum Architecture
- Requires minimal Day 1 investment.
- Enables a pay a cost-effective pay as you grow strategy.
- Possible to extend to 100% HP coverage whenever the decision is made to make full FTTH overbuild in the area.
- Is compatible with existing topology; can be housed in nodes and cabinets, which are common in the cable outside plant and leverage the power supply infrastructure already present in the cable network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Has a common management environment with existing HFC infrastructure, which is software based and cloud-native and that uses modern technologies such as YANG and NETCONF.
- Prepares the access network to offer high speeds and low latency services to support new use cases such as 5G X-haul, Edge Computing connectivity, etc.

7.1. Pre-connectorized technology for cost-effective, speedy deployments

PON deployments can be costly, particularly if they require building new facilities and infrastructure to host and connect the OLTs. Therefore, a cost-effective approach is to use a small-form universal hardened connector system for high-density FTTH environments. These connectors replace splices, reducing installed cost. They limit splicing to an absolute minimum, are easy to install, and therefore have a low turn-up time and require less skilled personnel. These hardened OLTs are embedded into the access network at existing points of presence in nodes, cabinets and hub sites leverages existing fiber runs to multiple PON systems, by using long reach dense wavelength-division multiplexing (DWDM). This allows service providers to expand their network capacity while avoiding the need to pull new fiber. Deploying these energy efficient PON actives in proximity to subscribers maximizes the utilization of the OLT, enabling it to serve more subscribers. It also provides a launch point for expansion to serve customers in areas adjacent to the plant, enabling a targeted, pay-as-you grow strategy while optimizing capex and opex.

7.2. The advantage of R-OLT over C-OLT

The R-OLT approach, wherever feasible, presents many advantages over a centralized OLT (C-OLT) including a more efficient port utilization and the ability to optimize the area design for a 128 split.

While there are many configurations and parameters to consider, here is a sample comparison, that makes multiple assumptions; for example, the feeder network is underground, which adds to the cost, but some operators can deploy aerial feeding network. This comparison is for illustrative purposes only.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

	Centralized OLT		Remote OLT
OLT Location	In headend location, collocated with CMTS (ISP, shelf or chassis based)		ROLT In existing Cable node (or using new cable node)
Feeder network	$$$ Requires new underground feeder		0 reuse existing fiber to the node, WDM + GE backhaul
Fiber Access network	$$$ **centralized** split with splitter near the OLT, more distribution fiber required	$ distributed split architecture allows more economical OSP build	$$ "Skinny fiber network" approach possible to minimize day 1 investment, build fast, and selectively target customers + extend later.
OSP Access Split Ratio	Max Split of 64 due to link budget	Max 64 Split, due to link budget distribute 1:4 – 1:16	Up to 128 split possible, optimize design for the area
OLT port usage	$ OLT usage can be optimized, pay as you grow approach possible.	$$$ Need to install OLT ports for 100% of homes passed day 1. Low port utilization.	$$ Hardened OLT deeper in the network allows to guarantee high OLT port utilization. Expand as you grow in 2 port increments.
Upgrade path to XGS-PON	$$	$$$ Lowest port utilization drives high number of ports to upgrade	$ Few ports to upgrade, combo capability present day 1
TCO	**$$$**	**$$$**	**$**

* Link budgets and distances, fiber design and OLT port usage will vary greatly depending on the target area

Figure 7 - Comparing C-OLT to R-OLT (CommScope)

8. Managing and scaling the PON

A main advantage of next-generation PON networks is that management and control can benefit from key innovations, such as software-defined networking (SDN). The separation of the OLT control and management functions from the physical OLT, opens the way to the utilization of cloud native control and management software and the adoption of open protocols such as NETCONF and Rest APIs for installation, provisioning, and management.

When management and control functions are separated, it becomes easier to install and manage numerous OLTs embedded throughout the outside plant, even in remote locations. When the OLT's control plane software resides in the cloud, tasks such as initial software updates, system checks, and other start-up tasks can be automated by the cloud-based domain management function which recognizes new OLTs as soon as they are powered up in the field. It then uses automation to bring these devices online quickly, which has the benefit of obviating the need for busy technicians to commission new OLTs. This process converts OLT deployment into a routine task, with hands-free onboarding and less operational impact from PON rollouts.

Another advantage of adherence to standards promoted by ITU and BBF among others that have defined network functions, interfaces, protocols, and abstractions models for all network functions in the PON access network, allowing different vendor interoperability. The result is reduced capital and operational costs, a faster time to market and flexibility to use best of breed technologies from multiple vendors, easy integration into existing network environments, and less long and costly OSS/BSS integration projects.

The adoption of cloud-based domain management provides an opportunity to evolve network management, by shifting old Network Management Paradigm from NMS polling NE´s to NE's

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

publishing in real time richer analytics to multiple consumer systems that may be subscribed to all or a subset of all network analytics.

9. Conclusion

Like all world regions, Latin America is seeing an unprecedented growth in bandwidth consumption, and an exploding need for broadband, higher speeds, symmetrical bandwidth, low latency to name a few. Like their peers across the world, Latin American operators are racing to meet the demand, but are judicious in deploying their capital and mindful of the limitations of their operating resources. Therefore, a pragmatic approach for them is to maximize the use of their existing deployed assets, primarily using DOCSIS®, and to use PON technologies to upgrade capacity where needed, for example in high consumption areas, areas adjacent to their HFC footprint, and greenfield areas. An R-OLT solution can be a cost effective, high throughput approach to satisfy market needs. Distributed R-OLTs can be managed by a standards-based disaggregated management and control environment, which reduces cost over the long run and minimizes vendor lock-in.

Abbreviations

ARPU	Average Revenue per Unit
BEAD	Broadband Equity, Access, and Deployment
BBF	Broadband Forum
BSS	Business Support System
DAA	Distributed Access Architecture
DOCSIS®	Data Over Cable Service Interface Specification
DWDM	Dense wavelength-division multiplexing
ESD	Extended Spectrum DOCSIS®
FCC	Federal Communications Commission
FDX	Full Duplex DOCSIS®
FWA	Fixed Wireless Access
GPON	Gigabit PON
HFC	Hybrid Fiber Coaxial
ITU	International Telecommunication Union
NE	Network Element
NETCONF	Network Configuration Protocol
NMS	Network Management System
OSS	Operational Support System
OSP	Outside Plant
PON	Passive Optical Network
REST	Representational State Transfer
R-OLT	Remote Optical Line Terminal
SDN	Software Defined Network
XGS-PON	10 Gigabit Symmetrical PON
X-haul	Backhaul-fronthaul
YANG	Yet Another Next Generation

Bibliography & References

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

[1] Statista, «Internet users in LatAm by country 2023,» [En línea]. Available: https://www.statista.com/statistics/186919/number-of-internet-users-in-latin-american-countries/. [Último acceso: 20 07 2023].

[2] OMDIA Research, «OMDIA, Total Fixed Broadband Subscription and Revenue Forecast, February 2023,» Informa Tech, London, 2023.

[3] OECD Library, «Home | OECD iLibrary,» OECD IDB, [En línea]. Available: https://www.oecd-ilibrary.org/sites/9789264251823-4-en/index.html?itemId=/content/component/9789264251823-4-en. [Último acceso: 20 07 2023].

[4] Sandvine, «Sandvine GIPR 2023.pdf,» Sandvine, [En línea]. Available: https://www.sandvine.com/hubfs/Sandvine_Redesign_2019/Downloads/2023/reports/Sandvine%20GIPR%202023.pdf?hsCtaTracking=b765258e-dbaf-4572-92b2-b877952153cf%7C798369b3-6aad-4db4-b4a6-bedb583bd41c. [Último acceso: 20 07 2023].

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

A Supply Chain of Weak Links

Open Source Versus Proprietary Software Threat Analysis

A Technical Paper prepared for SCTE by

Brian A. Scriber
Distinguished Technologist and VP of Security Technologies
CableLabs
858 Coal Creek Circle, Louisville, CO 80027
𝕏 @brianscriber
b.scriber@cablelabs.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

Abstract ... 3
Categories and Subject Descriptors ... 3
General Terms ... 3
Keywords ... 3
1. Introduction ... 3
2. Defining Open Source and Proprietary Software ... 3
3. Software Licensing ... 4
4. Defining Free and Open Source Software ... 4
5. Defining Proprietary Software ... 5
6. Security Assumptions ... 5

Analysis ... 5
7. Empirical vs. Intuitive ... 5
8. Count vs Severity vs Impact Breadth ... 6

Threat Comparison ... 6
9. Forking and Versioning ... 6
10. Patching ... 6
11. Malign Code Insertion and Insider Threats ... 7
12. Continuous Security Review ... 7
13. Software Use ... 7
14. Developer Expertise ... 8
15. Privacy Engineering ... 8
16. Liability and Regulatory Risk ... 8
17. Software Updates and Trust ... 9
18. Linking Libraries and SBOM ... 9
19. Software Monoculture ... 9
20. Publication of CVEs ... 12
21. Hybridization and Proprietization of Open Source ... 12
22. Economic Factors and Software Library Friction ... 12
23. Policy Implications ... 13
24. Mitigations ... 13

Outlook ... 15
Abbreviations ... 15
References ... 15

List of Tables

Title **Page Number**

Table 1 Activities Supported by Software License Model ... 4
Table 2 Threat Comparisons Where FOSS and Proprietary Software Differ ... 11
Table 3 Open-Source Security Foundation SLSA ... 14

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abstract

Research explores the assumptions, resourcing, and maintenance realities of software in both closed and open ecosystems. This work is an evaluation of the Software Lifecycle using a security lens to highlight advantages and disadvantages of each approach at different development stages. An aggregation of risks and threats is provided to build an overview of the myths and realities of ecosystem transparency, modifiability, and ownership while answering questions about forking, hybridization, and proprietization. With recent supply-chain attacks in the networking industry, and identification of malicious actors within the open-source ecosystems, these macro-threats are evaluated for applicability to each approach: monoculture vulnerability analysis, presumption of security review, motivation for feature additions, and software patching.

Categories and Subject Descriptors

D.2.8 [SOFTWARE ENGINEERING]: Metrics

K.6.5 [MANAGEMENT OF COMPUTING AND INFORMATION SYSTEMS]: Security and Protection

K.4.1 [COMPUTERS AND SOCIETY]: Public Policy Issues

General Terms

Security, Economics

Keywords

Security, Metrics, Open source software, Closed source software, Economics, Policy, Monoculture

1. Introduction

There exists a debate centered on the security of open source versus closed source software; this is predicated upon the false idea that there is one choice that the enterprise or implementation team must make, and that after selecting one of these two models, the implications of that decision can be ignored going forward. This paper will address the metrics that go into these considerations, defining the legal frameworks and contribution models for the different types of software, but it does so not to provide an answer to either one of these being more secure than the other, but with the objective of identifying the threats particularly inherent to adoption of either. Awareness of the threats lends itself to the considered mitigation steps necessary to better buttress a software service to be resistant or resilient to those threats.

2. Defining Open Source and Proprietary Software

Without clarification of definitions, it is possible to get tangled into emotional debates that could distract from the point of this paper which is to identify the inherent threats different across the approaches. The definitions are important, and for other contexts, increasingly so, but for purposes here the classification is based on a threat analysis perspective, not as heavily upon the licensing or legal frameworks.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Software Licensing

There are different licensing models that often play into definitions of different software models, for purposes of this paper, one key element for consideration is if the license publishes the source code, allows for modification and whether it allows for commercial hybridization or proprietization, but this paper will not tackle an otherwise contentious issue of whether software licensing nuance makes software open or not. From a perspective to be reviewed later, determining who contributed, or potentially who and how the contribution was reviewed, would be interesting concepts to consider, and only a very few licenses are starting to explore those caveats.

Table 1 Activities Supported by Software License Model

License	Allows for Commercial Hybridization or Proprietization	Requires contributions be attributed to author	Includes explicit notes on contribution review
CC BY	Yes	Yes	No
CC BY-SA	Yes	Yes	No
CC (other)	No	No	No
GPL v3	Yes	No	No
Apache v2.0	Yes	General	No
MIT	Yes	No	No
Apple Public License 2.0	Limited	No	No
BSD 3.0	Yes	No	No
EUPL 1.2	Yes*	No	No
Open SSL 3 (Now Apache 2.0)	Yes	Some[i]	No

Free and and open also get confused when looking at software like Adobe Acrobat Reader, which is closed source, but offered free of charge, or tools like the Oracle MySQL database which is open source, but subject to payment requirements. For purposes of this paper, the actual payment requirements are less important than the development methodology.

4. Defining Free and Open Source Software

Free and Open-Source Software (FOSS) is commonly used term inclusive of both the free software and open-source software models. For the purposes here, we can cover all those open models listed above under Software Licensing as FOSS because the threats from the software development lifecycle will be similar and the resultant code from non-hybridized or proprietized software will be available for review.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

This paper will look at any open contribution model software with subsequent open review of the source code as FOSS. Oracle's MySQL, given this definition, would be considered FOSS for purposes of this paper.

5. Defining Proprietary Software

Proprietary software shall be defined here as those solutions that are developed privately (either commercially or with no initial intention of public release of the final source code) and for which all contributors have a commercial agreement (employee, contractor, subcontractor or similar) for the final released solution. The development process and the resultant deliverables are controlled in a non-public manner. Adobe's Acrobat Reader would be considered proprietary for purposes of this paper.

6. Security Assumptions

While comparative security in FOSS vs proprietary code bases has been a topic of research and measurement[ii,iii], new threats have emerged during the decade since the some of the most recent novel work in this space and some salient questions remain in performing a comparison.

Outstanding questions for security considerations include the following: What metrics exist for comparative analysis? What do those metrics focus on? Is reliability/availability a reliable proxy for security? Who's writing the code? Who is reading the code? How frequently is the code reviewed and how efficient are those reviews at finding vulnerabilities before adversaries? What are the economic hurdles to getting access to the code or to modifying the code for malicious purposes? How do embedded software and vulnerabilities therein contribute to the overall security posture?

Analysis

7. Empirical vs. Intuitive

From a methodological perspective, even comparing published vulnerabilities in FOSS and proprietary solutions limits the sampling to only those known/public vulnerabilities in widespread software packages. Comparison between the mean time between vulnerability disclosure and fix has shown some advantages toward FOSS[iv].

While the population in that prior study was extremely limited, the impact of the time between discovery and fix is increasingly important because we see examples[v] of adversaries adopting attack behavior to leverage recent disclosures within hours of CVE publication (e.g., Citrix CVE-2022- 27518, Microsoft Exchange Server CVE-2022-41040, Confluence Data Center CVE-2022-26134). Metrics for security currently tend toward reliability and availability and less toward the frequency or severity of vulnerabilities in a software solution or toward the measurement of how widespread and broadly impacting a vulnerability in a particular software package might be (e.g., Log4Shell CVE-2021-44228). Without at least these three axes, comparisons are going to remain difficult.

Availability of metrics for those three dimensions (count, severity, and impact breadth) are difficult to obtain or even approximate, and the danger of comparing empirical data with intuitive assumptions is a significant risk to the industry.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

8. Count vs Severity vs Impact Breadth

Modern software is manifold, and while legacy systems may have had a monolithic approach to all the software running in a solution, the modern software leverages packages and libraries in compiled code, it uses callouts to multiple systems and microservices, the software also relies upon dynamic elements in configuration or deployment-optimization files. While legacy measurements looked at a system independently and heavily weighed vulnerability count and severity of vulnerability, the approach failed to account for the supply-chain implication of the modern technical solution space by including something beyond the severity concern of a compromise to this specific system. That vector beyond count and severity is the impact breadth, which addresses the concern with embedded libraries that are widely used[vi], hardware deficiencies that allow compromise (e.g., Spectre variants, CVE-2017-5753), or even compilers that have hidden vulnerabilities allowing for illicit code insertion or modification[vii].

For a true measure or comparison between systems or solutions, we would need to weight the count, severity, and impact breadth, and find the product of all three.

Threat Comparison

Threats across the supply chain, software lifecycle, and in the related processes abound in both FOSS and proprietary solutions, but over the last ten years the supply chain attacks have changed the risk calculus and bring consideration to different weights for existing and newer risks. Table 2, below, looks at threats where the implications differ between FOSS and proprietary software.

9. Forking and Versioning

In FOSS environments, forked codebases split reviewers and active participants across multiple projects. Fixes in one branch may not be adopted in base software or other branches and take time to migrate if they do. Fixes may not be appropriate for other branches or may introduce new vulnerabilities.

Support for older versions of software is the analogous behavior in the proprietary world. Budgets may prevent upgrades, and dependencies between versions may force organizations to stay on platforms or services that have known vulnerabilities.

10. Patching

Reliability engineering looks at mean time between failures, but security patching can look at mean time between vulnerability discovery and deployed patch. In this, FOSS seems to have a slight advantage over proprietary solutions[viii]. Not having a managed patching mechanism may be hurting FOSS though, perhaps more than speed to fix can make up for: the Log4Shell vulnerability (CVE-2021-44228), had a fix available within a day, but nearly a third of current downloads of that package are of an earlier, vulnerable, version[ix]. The true horror from research on downloaded packages is that 96% of the time that a software package is downloaded, a safer, more recent, version is available[x]. Updating build scripts, links that don't always point to HEAD, and time available for keeping dependencies fresh are all security considerations. Another concern with FOSS is that patching and mitigation actions are public and when the CVE and fix are published, threat actors can take this knowledge, along with sample exploit code and look for similar toolsets that may have the same code profile (even cut/paste identical code) to apply to create new zero-day vulnerabilities[xi]. Examples of this came after the Log4Shell vulnerability (CVE-2021-44228) and leveraging the privilege escalation vulnerability in Linux (CVE-2021-4034).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Proprietary solutions and hybridized solutions often have a maintenance contract related to updates and patching, the solution may or may not be managed, but the update process is frequently run by the solution provider. From an economic perspective software support is often based upon the expected revenue from newer versions of the software that can include patched updates from prior revisions. If the solution isn't widely adopted, if the version is a couple revisions behind the lead for that package, or if the company supporting it cannot invest in support, patching can take a back seat. With proprietary code, the users have no option to adopt the code and support it themselves and are beholden to the software provider.

11. Malign Code Insertion and Insider Threats

FOSS projects open contributor model is often paired with a review prior to adoption by maintainers[xii]. Those maintainers may not be experts in reviewing for security or privacy vulnerabilities allowing for insertion. Intentional contributions enabling nuanced vulnerabilities also occur, the field from which these are drawn include nation-state APTs, hactivism, organized crime and more. These vulnerabilities can be leveraged for data exfiltration, ransomware insertion, machine hijacking, fraud, or denial of service.

The closed list of contributors in corporate or proprietary solutions spaces is not proof against malign code insertion, there are examples of insider threat actors having actually engaged exfiltration or code insertion that include departing employees (e.g., Proofpoint[xiii]), through acquired companies' vulnerabilities (e.g., the Marriott-Starwood acquisition hack in 2014), accidental exposure of proprietary secrets by employees or contractors (e.g. Microsoft posts to GitHub in 2022[xiv]).

12. Continuous Security Review

The claim is made that "given enough eyeballs, all bugs are shallow"[xv] and that this fundamentally improves the quality and security of FOSS. Opposite questions also arise: "Sure the source code is available But is anyone reading it."[xvi] Questions about the reviewers include their number, qualifications, incentives/motivations, experience with the particular code base and cryptography, and their belief or trust in existing code.

Proprietary code differs in the reviews and motivations. Presumably, threat analysis is being performed (consciously or not) within corporations where vulnerabilities are being balanced against revenues or risk. For any system, commensurate processes and personnel would be assigned to write, review, test, and monitor. Those personnel face the same questions as FOSS, but the answers may differ for proprietary issues and for resource prioritization.

13. Software Use

In FOSS, the software is available for download, compilation, and use, often "as is" and without warranty. Adversaries have leveraged the trust in using open-source software during campaigns targeting senior developers, posing as recruiters and asking the developer to use modified versions of open source tools to demonstrate mastery[xvii]. As will be shown in section 17, Software Updates and Trust, checksums are not reliable proof against this type of attack.

Proprietary software is not proof against these types of threats, and with hybridization, the risks associated with composite software are present in a large majority of software packages[xviii]. When the included software is purchased as a binary library from other providers, particularly when that happens without an SBOM, the unknowns are effectively being granted full citizenship in the now proprietary

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

software or internal process with little understanding of what may happen with the process, the data, or the distributed final binaries.

14. Developer Expertise

FOSS volunteers may not be security experts or have access to training or mentorship from those with that expertise. Private employers look for developer experience in open-source projects and encourage that activity[xix]. This drives even the experienced developers to contribute and review open-source projects to gain employment opportunities. With some FOSS implementations (see Table 1) explicitly citing authors, external credit is available and can lead to being hired to positions with higher compensation[xx], but with that credit comes accountability if vulnerabilities are found.

Proprietary projects face the same hurdle as FOSS when developers leave projects, but hand-offs, training, documentation, backup developers with depth on important projects can be retained or recruited within private companies. For security concerns, it may be easier to find compensated help than it will be to find time from those already employed willing to volunteer time to FOSS. When the cybersecurity job market sees 35% annual growth[xxi] with 3.5 million open cybersecurity positions[xxii] open globally, available time from qualified security developers is challenging even for those who can compensate for that time.

15. Privacy Engineering

Privacy engineering is a relatively new discipline[xxiii]. While developers in FOSS are as educated as developers in proprietary solutions, qualifications for privacy work include strong policy support, some legal support, and technical understanding of data architecture and understanding of movement of data through and across systems. While open-source efforts have attracted legal support through groups like the Law Center[xxiv], it may still be early to expect policy support.

Privacy engineering in the proprietary realm is influenced by a few factors, a few of these include corporate public perception, regulation and the associated enforcement efforts, as well as the financial benefits of the data economy that may work against the first two. Looking at liability and risk concerns as part of the motivation in incentives for support of technologies like privacy protections could advantage proprietary software over FOSS, but data economy tradeoffs play into this calculus as well.

16. Liability and Regulatory Risk

The EU Product Liability Directive (PLD) update work in 2023 appears to be struggling with FOSS[xxv]. In one clause it states: “With the aim of not hampering innovation: (i)free and open-source software developed or supplied outside the course of commercial activity, as well as (ii) the source code of software, should be excluded from the definition of products covered under the proposal (Recital13).”[xxvi] In a later clause of the full proposal[xxvii], it states “However where software is supplied in exchange for a price or personal data is used other than exclusively for improving the security, compatibility or interoperability of the software, and is therefore supplied in the course of a commercial activity, the Directive should apply.” The implication being that if FOSS is ever incorporated into a product, it is subject to liability risk, and with respect to the EU Cyber Resilience Act and the PLD, major FOSS tool providers wrote in opposition “it will have a chilling effect on open source software development as a global endeavor, with the net effect of undermining the EU’s own expressed goals for innovation, digital sovereignty, and future prosperity.”[xxviii].

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Software liability for proprietary solutions certainly carries with it the legislative and regulatory risk of providing a service, but add to this the relatively recent additions to privacy regulation (GDPR in the EU and state-level privacy regulation in the US, PIPL in China, PIPEDA in Canada, along with several other sectoral or national privacy regulations in place now). Cybersecurity and Infrastructure Security Agency (CISA) in the US has subpoena powers related to cyber incidents and they are advocating for SBOM requirements, particularly for anything deemed critical infrastructure. Violations, delays, or inaccuracies in reporting may have consequential ramifications.

17. Software Updates and Trust

FOSS builds offer checksums on the binary distributions and allow users to compile these themselves as well. Checksums are not always present, but in over 88% of cases users may not notice checksums, understand checksums, know how to go about verifying checksums, or simply don't bother to verify packages. Over 33% of those asked specifically to verify software do so improperly and fail to catch a mismatch (partial pre-image attack)[xxix].

Signed binaries from PKI-backed digital certificates of trusted entities are intended to verify the source and the package contents match those of the trusted entity. While security policies, IDS/IPS, and software update mechanisms work to automate the verification of these packages, that doesn't mean supply chain issues can't occur with those credentials or the verification assumptions. The SolarWinds supply-chain attack inserted malicious code upstream of the signing[xxx], and the Aug 2023 Microsoft Azure attack using compromised Managed Service Account credentials to sign requests for new credentials which exposed user email and potentially virtual hosts[xxxi].

18. Linking Libraries and SBOM

FOSS linking is done publicly. There are tools to help organize these, and even some tools to help you find our who may be including other libraries in their builds. For a cybercriminal, knowing that a vulnerability exists in a particular software package, and perhaps having an exploit at the ready is more valuable when there is a clear list of other software that is now vulnerable because of the binary inclusion, however many levels.

Proprietary software still has the linked vulnerability problem, but the non-public nature of linking requires more trial-and-error work on the part of the cybercriminal looking for a way to apply an exploit. Software Bill of Material (SBOM) efforts of late[xxxii] are pushing for knowledge of all linked libraries by software owners. While that knowledge is valuable and can help with tracking and stopping vulnerabilities, the publication of this list of included libraries could be used as a roadmap for criminal adversaries.

19. Software Monoculture

In agriculture, plant propagation, breeding, and genetic similarities are used to help formulate herbicides and pesticides that can work without impact to target crops, it also helps in terms of standardization of equipment used for planting and harvesting and the efficiency of yield for a given quantity of land can be higher through monoculture farming[xxxiii]. The danger of these similarities, or identical genetic makeup in some cases, is that a threat to a single plant becomes a threat to all plants. One pest[xxxiv], one disease, one change in economics or distribution pricing can have a massive impact to the monoculture. This is entirely analogous to the work taking place in software.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Reliance upon existing public software is economically incentivized, and when underlying libraries are leveraged across multiple included platforms, services, or other software it drives further consolidation. In FOSS, the classic example of this is the reliance upon the OpenSSL library vulnerability (CVE-2014-0160) and the resultant Heartbleed impact which made the vulnerability in that package a concern across the ecosystem. Another more recent example has broader implications from the monoculture perspective, the Log4Shell vulnerability (CVE-2021-44228), which was pervasive with mitigation efforts that continue today.

Proprietary software is not immune to these considerations. When the operating system, browser, cloud offering, microservice, or tooling are all identical, a risk to one can become a global crisis. The other implication to using services outside of software run independently (e.g., an online email management service), is that an organization effectively makes each of the organizations running those services an insider to their data, their systems, and this opens the horizontal attack threat surface. When Microsoft Azure services were attacked in 2023[xxxv], Outlook 365 and Azure subscribers were all vulnerable.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 2 Threat Comparisons Where FOSS and Proprietary Software Differ

Threat	FOSS Implications	Proprietary Implications
Forking and Versioning	Update propagation across forked codebases, reviewers and participants split across projects, update applicability to forks, timeframe, new vulnerabilities.	Support for older versions of software: budgets may delay/prevent upgrades, dependencies between versions may force organizations on vulnerable platforms.
Patching	Mean time between vuln. ID and patch vs. 96% of patched software updates are available but ignored. Public patching and mitigation alerts adversaries.	Maintenance contracts, update process is frequently run by the solution provider, feature lag, security patch timeframes, code abandonment risk.
Malign Code Insertion & Insider Threats	Maintainer expertise in developing and reviewing for security, allowing for insertion of intentional malware contributions.	Active employee/contractor data or code exfiltration or code insertion. Departing employees, acquired companies, accidental exposure of proprietary secrets.
Continuous Security Review	Reviewer count, qualifications, incentives/motivations, experience with the project code, cryptography, and their belief or trust in existing code.	Liability pits resource prioritization against revenues and risk: development, review, testing, and monitoring; FOSS questions remain regarding experience and trust.
Software Use	FOSS is often trusted by technical users, it is offered "as is" and without warranty, and checksums are not proof of invulnerability.	Transitive corporate trust of aggregated software or packages (FOSS or proprietary) through business agreements and redistribution.
Developer Expertise	Existing security expertise, access to training or mentorship, pull toward commercial/paid development, contribution credit/accountability.	Developers leave projects, hand-offs, training, documentation, backup developers with depth can be retained or recruited, cybersecurity job market is tight.
Privacy Engineering	Technical understanding of data architecture and data movement across systems, strong policy and legal understanding and research tooling.	Balancing liability/risk of damaging public perception of corporate entities along with regulation and enforcement, against financial benefits of the data economy.
Liability and Regulatory Risk	EU Product Liability Directive update (2023) introduces significant legislative and regulatory risk to existing and future open-source projects and teams.	Global privacy legislation/regulation; security oversight (CISA/DHS Subpoena Powers), cyber incident reporting, SBOM, and critical infrastructure designations
Software Updates and Trust	Checksum presence, awareness, understanding, verification knowledge, consistency of practice, inability to catch partial pre-image attacks.	Signed binaries from PKI-backed certificates of trusted entities, IDS/IPS, and automation aren't proof vs. attacks upstream of signing or credential theft.
Linking Libraries and SBOM	Linking is done publicly, organization tools help cyber criminals map vulnerabilities to exploits to all software packages that include that vulnerability.	Linked vulnerability problem still exists but requires more trial-and-error work on the part of the cybercriminal. Publication of SBOM could remove this advantage.
Software Monoculture	Economic incentives for consolidation lead to wide-striking vulnerabilities such as Log4Shell and Heartbleed/OpenSSL; diversifying increases the attack surface.	Identical operating systems, browsers, and tooling have broad risk. Hyperscaling and microservices are business insiders that can yield wide and immediate compromise.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

20. Publication of CVEs

When the fixes to published CVEs are made in proprietary code, the mechanism by which the solution has been secured is not readily available to all observers, however in FOSS, the mitigation is published and can lead to providing a roadmap for other potential vulnerability exploits.

The implications of CVE publication, notification periods, and bug bounty programs are beyond the scope of this paper, but the public response, the measurable time between vulnerability listing and fix, are all information that the adversaries can use to understand which projects may be faster than others to update, and when fixes may be more challenging to adopt (e.g., such as when an API changes or parameter structure is modified). The proprietary response may make the fix visible to adversaries who are leveraging or testing that vulnerability, but for those not observing as actively, it may be difficult to discern which companies are faster than others at closing vulnerabilities.

21. Hybridization and Proprietization of Open Source

With the economic benefits of code reuse and libraries, FOSS currently exists in between 80-90% of all modern application code[xxxvi]. With aggregation, forking, and rates like this, can claims for a distinction between the two hold up? Proprietization may not be a word (yet), but the hybrid nature of modern software, the long supply chains, and the somewhat nebulous practices of different open-contributor projects mean significant threat analysis by cybersecurity professionals.

22. Economic Factors and Software Library Friction

Development costs involve not just the software being custom written for the solution, development costs now include the analysis, retrieval, documentation, testing and integration of other software to integrate into the final solution. If those are FOSS, the time spent to integrate those is still part of this equation. Adding to this are the maintenance costs, including staying up to date on patches, API changes, updates to tests, internal data management, deployment architecture, regulatory compliance checks (e.g., privacy regulations on data movement), security review and decisions on how to handle software at its end-of-life. For each new feature addition to a final software solution, the maintenance and development costs, referred to above, must be undertaken along with ensuring that the update process is managed, ensuring that new vulnerabilities aren't introduced, and population of release notes are made. For providers of library solutions or packages that get used by other software or included in other software, the list of activities (and related costs) increases: any other managed software must also be tested, version differences and mandatory upgrades must be considered, API updates, documentation changes, breaking changes, migration paths, support for prior versions, compatibility, interoperability and deployability must be part of the decision and release process. Do SBOMs need to be updated, are known vulnerabilities closed, are license requirements met, do contributions need to be submitted back to source projects, and are there any new regulatory requirements around reporting that need to be met.
All of this introduces friction into the process, in a proprietary solution necessary steps may be mandated by internal processes, checklists, and requirements from groups or clients that have their own list of requirements to check. In the FOSS environment, volunteers would need to do this job, or limited funds must be used to make sure to stay close to the verification and validation processes described above. As these may be considered overhead by some developers, it's possible that some of these steps could be overlooked.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

23. Policy Implications

As governments and policymakers look to regulations to help get a better hold on supply-chain insecurities, several actions are taking place. Governments limiting sources of origin for some technical solutions such as the FCC's actions to limit access to the U.S. market[xxxvii] but foreign actors, malicious actors, and organized crime can all make contributions to most open-source projects so that part of the chain is potentially still vulnerable. The European Union is exploring product liability extensions which could have an impact to FOSS, and we noted earlier in this paper the response to those new potential regulations. The US Government also recognizes the risk in this space as they actively seek feedback on open-source security for purposes of focus and prioritization through their 2023 RFI[xxxviii]. Whether government pressure changes the culture or development models of FOSS remains to be seen, but discussion has turned to considerations like SBOM, component risk, mandated mean time between disclosure and patching, product support lifetime expectations, minimum expectations for security and privacy, and, like what we see in the EU, the reconsideration of product liability in relation to modern concepts, tooling, and services provided by computational infrastructure. Government policy will play a part in the nuances between FOSS software security and that of proprietary or commercial endeavors; for those endeavors that utilize, aggregate, or include FOSS as part of their deliverables, the implications could be dramatic, and this could also have an impact on FOSS contribution.

24. Mitigations

This paper has addressed several threats that have a difference in implications, ramifications or nuance between FOSS and proprietary solutions. This can be overwhelming to see and can feel futile to fight, but there are some improving solutions and tools to help address supply-chain security that are available today and coming soon. Some considerations follow:

- Everyone should follow best practices for DevSecOps[xxxix] including securing the entire software lifecycle (development process and production environment), manage authentication and authorization (use least-privilege principles, encrypt sensitive data, role-based access control, two-factor authentication, access control lists), monitor (logging and monitoring, penetration tests, intrusion detection and prevention systems (IDS/IPS), audit regularly), train anyone engaged or authorized for any part of the process, use secrets management tools like Hardware Security Modules (HSMs), and have a disaster recovery plan (including regular table-top exercises and drills).
- For providers of open source software and tooling to handle versioning (e.g., GitHub, SourceForge, etc.), institute a practice of notification of applicable CVEs, generate and maintain the SBOM for each software package, complicate or prevent downloading of binaries or source versions that have been patched in subsequent versions, remove exemplar code or tools that exist for exclusively malicious purposes, and help to track, attribute, and verify contributors to FOSS.
- For providers of open source software and tooling to handle versioning (e.g., GitHub, SourceForge, etc.), institute a practice of notification of applicable CVEs, generate and maintain the SBOM for each software package, complicate or prevent downloading of binaries or source versions that have been patched in subsequent versions, remove exemplar code or tools that exist for exclusively malicious purposes, and help to track, attribute, and verify contributors to FOSS.
- For those providing builds or using builds, consider the Open-Source Security Foundation SLSA[xl] version 1.0 levels from 2023(see Table 3, below).
- For those in industries identified by CISA as being part of the Critical Infrastructure (Chemical Sector, Commercial Facilities Sector, Communications Sector, Critical Manufacturing Sector, Dams Sector, Defense Industrial Base Sector, Emergency Services Sector, Energy Sector, Financial Services Sector, Food and Agriculture Sector, Government Facilities Sector, Healthcare and Public Health Sector, Information Technology Sector, Nuclear Reactors, Materials and Waste Sector, Transportation Systems Sector, and the Water and Wastewater Systems), there are several resources for improving supply chain security[xli] including this report from the FCC and CSRIC VIII[xlii] which list the following suggestions for addressing open source security:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

  - Strict internal requirements should exist to protect the company and its customers.
  - Third-party suppliers should be held to the same company standards.
  - Policies should apply equally to open-source software as with proprietary software.
  - Third-party software should be sourced by a centralized configuration management team.
  - Centralized configuration management teams should ensure sources are reputable.
  - A gating subprocess should validate that patches are applied, and scans are completed.
  - A post-scan analysis should be used to reveal issue severity, priority, and applicability.
- For small and medium-sized businesses who are engaging in software in any manner are addressed by CISA who have highlighted the top six risks where those businesses should focus their attention and growth: Cyber Expertise, Executive Commitment to Cybersecurity, Supply Chain Risk Management, Single Source Suppliers, Supplier Disruption and Visibility into Supplier Cybersecurity Practices[xliii].

Table 3 Open-Source Security Foundation SLSA

SLSA Level	Description	Requirements	Example
0	No guarantees.	SLSA 0 represents the lack of any SLSA level.	Unknown provenance
1	Documentation of the build process	**The build process must be fully scripted/automated and generate provenance.** Provenance is metadata about how an artifact was built, including the build process, top-level source, and dependencies. Knowing the provenance allows software consumers to make risk-based security decisions. Provenance at SLSA 1 does not protect against tampering, but it offers a basic level of code source identification and can aid in vulnerability management.	Unsigned provenance
2	Tamper resistance of the build service	**Requires using version control and a hosted build service that generates authenticated provenance.** These additional requirements give the software consumer greater confidence in the origin of the software. At this level, the provenance prevents tampering to the extent that the build service is trusted. SLSA 2 also provides an easy upgrade path to SLSA 3.	Hosted source/build, signed provenance
3	Extra resistance to specific threats	**The source and build platforms meet specific standards to guarantee the auditability of the source and the integrity of the provenance respectively.** We envision an accreditation process whereby auditors certify that platforms meet the requirements, which consumers can then rely on. SLSA 3 provides much stronger protections against tampering than earlier levels by preventing specific classes of threats, such as cross-build contamination.	Security controls on host, non-falsifiable provenance
4	Highest levels of confidence and trust	**Requires two-person review of all changes and a hermetic, reproducible build process.** Two-person review is an industry best practice for catching mistakes and deterring bad behavior. Hermetic builds guarantee that the provenance's list of dependencies is complete. Reproducible builds, though not strictly required, provide many auditability and reliability benefits. Overall, SLSA 4 gives the consumer a high degree of confidence that the software has not been tampered with.	Two-party review + hermetic builds

The above are not the complete list of options available for mitigation, but these steps are intended to help organizations, FOSS tool providers, and software enterprises contribute to a safer software supply chain.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Outlook

If one assigns a point to each threat and count wins for FOSS versus proprietary solutions, one would be doing it wrong. Each piece of software is going to have a unique history, contribution model, algorithmic approach, versioning, lifecycle management, as well as unique teams to develop, support, test, configure and deploy the solution. For each one of these, a considered approach of the best model, open or proprietary, belongs in the architectural decision process. Are specialists necessary, are they available in the open-source community, are they qualified and incented to work on the project? These just scratch the surface of the questions that will need to be answered. It's hard to ask them all, which is why looking at varying threats and how they have evolved since the research literature changed is worth our effort.

Abbreviations

API	Application Programming Interface
CISA	Cybersecurity and Infrastructure Security Agency (part of Department of Homeland Security)
CVSS	NIST Common Vulnerability Scoring System
DHS	Department of Homeland Security
FOSS	Free or Open Source Software
HSM	Hardware Security Modules
IDS	Intrusion Detection System
IPS	Intrusion Prevention System
PLD	European Union Product Liability Directive
SBOM	Software Bill of Materials
SLSA	Supply-chain Levels for Software Artifacts

References

[i] https://wiki.openssl.org/index.php/Contributions
[ii] Schryen, Guido & Kadura, Rouven. (2009). Open source vs. closed source software: Towards measuring security. Proceedings of the ACM Symposium on Applied Computing. 2016-2023. 10.1145/1529282.1529731.
[iii] Schryen, Guido, "Security of Open Source and Closed Source Software: An Empirical Comparison of Published Vulnerabilities" (2009). *AMCIS 2009 Proceedings*. 387. https://aisel.aisnet.org/amcis2009/387
[iv] Schryen, Guido, "Security of Open Source and Closed Source Software: An Empirical Comparison of Published Vulnerabilities" (2009). *AMCIS 2009 Proceedings*. 387. https://aisel.aisnet.org/amcis2009/387
[v] CrowdStrike. (2023) CrowdStrike 2023 Global Threat Report., https://go.crowdstrike.com/rs/281-OBQ-266/images/CrowdStrike2023GlobalThreatReport.pdf
[vi] Ghafoor, Imran & Jattala, Imran & Durrani, Shakeel & Tahir, Ch. (2014). Analysis of OpenSSL Heartbleed vulnerability for embedded systems. 314-319. 10.1109/INMIC.2014.7097358.
[vii] Thompson, K. Reflections on Trusting Trust. *Comm. Of the ACM. 27*, 8 (1984), 761-763

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

[viii] Schryen, Guido, "Security of Open Source and Closed Source Software: An Empirical Comparison of Published Vulnerabilities" (2009). *AMCIS 2009 Proceedings*. 387.
[ix] Fox, B., *The EU's Product Liability Directive could kill open source,* Tech Radar, 10 July, 2023, https://www.techradar.com/pro/the-eus-product-liability-directive-could-kill-open-source
[x] ibid
[xi] CrowdStrike. (2023) CrowdStrike 2023 Global Threat Report., https://go.crowdstrike.com/rs/281-OBQ-266/images/CrowdStrike2023GlobalThreatReport.pdf
[xii] https://wiki.openssl.org/index.php/Contributions
[xiii] Proofpoint, Inc. v. Samuel Boone, Case# 1:2021cv00667, July 29, 2021, US District Court for the Western District of Texas, Judge Lee Yeakel
[xiv] https://www.vice.com/en/article/m7gb43/microsoft-employees-exposed-login-credentials-azure-github
[xv] Raymond, E. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary troff edition* (1999) https://lists.gnu.org/archive/html/groff/2021-11/pdfRt8tRop5yy.pdf
[xvi] Levy, E. *Wide open source*, http://www.securityfocus.com/news/19, 2000.
[xvii] Goodin, D. *Numerous orgs hacked after installing weaponized open source apps*, Ars Technica https://arstechnica.com/information-technology/2022/09/north-korean-threat-actors-are-weaponizing-all-kinds-of-open-source-apps/ , 29 Sept 2022
[xviii] Fox, B., *The EU's Product Liability Directive could kill open source,* Tech Radar, 10 July, 2023, https://www.techradar.com/pro/the-eus-product-liability-directive-could-kill-open-source
[xix] https://blogs.vmware.com/opensource/2020/08/25/boost-your-career-through-open-source-contribution/
[xx] Hann, Il-Horn; Roberts, Jeff; Slaughter, Sandra; and Fielding, Roy, "Economic Incentives for Participating in Open Source Software Projects" (2002). ICIS 2002 Proceedings. Paper 33.
[xxi] US Bureau of Labor Statistics: https://www.bls.gov/ooh/computer-and-information-technology/information-security-analysts.htm 8 Sep, 2022
[xxii] Cybercrime Magazine: https://cybersecurityventures.com/jobs/ 14 Apr, 2023.
[xxiii] Fennessy, C. International Association of Privacy Professionals: *Privacy engineering: The what, why and how* https://iapp.org/news/a/privacy-engineering-the-what-why-and-how/ 8 Aug, 2019
[xxiv] https://www.zdnet.com/article/lawyers-ride-shotgun-for-open-source/
[xxv] De Luca, S., European Parliament, "Briefing: EU Legislation in Progress: New Product Liability Directive" https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739341/EPRS_BRI(2023)739341_EN.pdf May, 2023
[xxvi] Ibid, pp 5.
[xxvii] Document 52022PC0495, Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on liability for defective products, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022PC0495, {SEC(2022) 343 final} - {SWD(2022) 315 final} - {SWD(2022) 316 final} - {SWD(2022) 317 final}, 28 Sep 2022
[xxviii] Harris, J., *Open Letter to the European Commission on the Cyber Resilience Act*, https://newsroom.eclipse.org/news/announcements/open-letter-european-commission-cyber-resilience-act , 17 April 2023
[xxix] Alexandre Meylan, Mauro Cherubini, Bertil Chapuis, Mathias Humbert, Igor Bilogrevic, and Kévin Huguenin. 2020. A Study on the Use of Checksums for Integrity Verification of Web Downloads. *ACM Trans. Priv. Secur.* 24, 1, Article 4 (September 2020), 36 pages. **https://doi.org/10.1145/3410154**
[xxx] Constantin, L., *SolarWinds attack explained: And why it was so hard to detect*, CSO, Dec 15, 2020, https://www.csoonline.com/article/570191/solarwinds-supply-chain-attack-explained-why-organizations-were-not-prepared.html
[xxxi] Schneier, B., *Microsoft Signing Key Stolen by Chinese*, Aug 2023, https://www.schneier.com/blog/archives/2023/08/microsoft-signing-key-stolen-by-chinese.html
[xxxii] https://www.cisa.gov/sbom
[xxxiii] McGuire, A., *Ecological Theories, Meta-Analysis, and the Benefits of Monocultures*. Washington State University
[xxxiv] JM Sirota, E Grafius, W Boylan-Pett, B Ferrari, P Kolarik, A Pyle, B Scriber, Colorado Potato Beetle Control... Insecticide and Acaricide Tests 18 (1), 153-155, https://academic.oup.com/amt/article-abstract/18/1/153/4573455, 1991
[xxxv] Schneier, B., *Microsoft Signing Key Stolen by Chinese*, Aug 2023, https://www.schneier.com/blog/archives/2023/08/microsoft-signing-key-stolen-by-chinese.html

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

[xxxvi] Fox, B., *The EU's Product Liability Directive could kill open source,* Tech Radar, 10 July, 2023, https://www.techradar.com/pro/the-eus-product-liability-directive-could-kill-open-source
[xxxvii] Federal Communications Commission FCC 22-84 *Protecting Against National Security Threats to the Communications Supply Chain through the Equipment Authorization Program* https://docs.fcc.gov/public/attachments/FCC-22-84A1.pdf 25 Nov 2022
[xxxviii] Office of the National Cyber Director, *Request for Information on Open-Source Software Security: Areas of Long-Term Focus and Prioritization*, https://www.federalregister.gov/documents/2023/08/10/2023-17239/request-for-information-on-open-source-software-security-areas-of-long-term-focus-and-prioritization , 10 Aug 2023
[xxxix] Baig, A. DevOps.com 15 DevSecOps Best Practices, https://devops.com/15-devsecops-best-practices/ 15 Apr 2022
[xl] SLSA specification version 1.0, Open Source Security Foundation, https://slsa.dev/spec/v1.0/ 2023
[xli] CISA *Resources and Tools* https://www.cisa.gov/resources-tools/all-resources-tools
[xlii] FCC, *CSRIC Report on Recommended Best Practices to Improve Supply Chain Security*, https://www.cisa.gov/resources-tools/resources/csric-report-recommended-best-practices-improve-supply-chain-security 21 Sep 2022
[xliii] CISA, Reducing ICT Supply Chain Risk in Samll and Medium-Sized Businesses https://www.cisa.gov/sites/default/files/2023-05/fs_reducing-ict-supply-chain-risk-smb_fact-sheet_508.pdf May 2023

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Accounting for Every MHz of Bandwidth: Data & Algorithms for Artifact Discovery and Close-Packing of QAMs in Support of Spectrum Activation

A Technical Paper prepared for SCTE by

Maher Harb
Distinguished Engineer, Data Science
Comcast
1800 Arch St, Philadelphia, PA 19103
+1 (215) 990-8376
maher_harb@comcast.com

Wenlong Shen
Senior Engineer, Machine Learning
Comcast
1800 Arch St, Philadelphia, PA 19103
+1 (267) 260-3798
wenlong_shen@comcast.com

Matt Stehman
Senior Engineer, Machine Learning
Comcast
1800 Arch St, Philadelphia, PA 19103
+1 (267) 260-3956
matthew_stehman@comcast.com

Sanket Walavalkar
Executive Director, Next Generation Access Network
Comcast
1800 Arch St, Philadelphia, PA 19103
+1 (917) 592-6789
sanket_walavalkar@comcast.com

Dan Rice
VP, Next Generation Access Network
Comcast
4100 E Dry Creek Rd, Centennial, CO 80122
+1 (720) 512-3730
daniel_rice4@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Phase 1 of Activations: Mid-split and OFDM at 810-1002 MHz 4
3. An Object-Oriented Model for the Full Band Capture 7
   3.1. MDU Insert within 810-1002 MHz 8
   3.2. Roll-off within 810-1002 MHz 9
4. The Pipeline for Artifact Detection 9
5. Phase 2 of Activations: 2nd OFDM 11
6. Phase 3 of Activations: Close-Packing of Video 13
7. Conclusion 15

Abbreviations 16

Bibliography & References 17

List of Figures

Title **Page Number**

Figure 1 - Top 25 configurations for RPDs that haven't yet undergone mid-split activation. ~46% of this RPD population is covered by the top 25. Aside from the upstream SC-QAM block, there exists high variability across configurations. 5

Figure 2 - Top 25 configurations for RPDs that are mid-split active. ~57% of this RPD population is covered by the top 25. Though upstream and OFDM blocks are aligned across permutations, there remains high variability across video and downstream SC-QAM. 6

Figure 3 - Example FBC that shows an undocumented local insert at 855 MHz. The overlaid colored shadings correspond to the RPD configuration according to the color map in legend of Figure 1. 9

Figure 4 - Example FBC that shows roll-off at the 999 MHz carrier. The overlaid colored shadings correspond to the RPD configuration according to the color map in legend of Figure 1. 9

Figure 5 - Sensitivity curve for insertion detection. The x-axis is the selection threshold in the number of cable modems that show roll-off. The y-axis is the corresponding number of impacted nodes. The "elbow method" would set the threshold close to ~10. 10

Figure 6 - A graph of the network showing all elements under a node. Cable modems with roll-off are highlighted in pink (top center area). The amplifier (red circle) responsible for the roll-off is marked with an asterisk. This amplifier was missed in the upgrade construction activity. 11

Figure 7 - Top panel: Spectrum configuration pre-2nd ofdm shows a "wasted" gap between video and downstream SC-QAM. Middle panel: an FBC from a cable modem reveals an undocumented MDU insert at 777 MHz. Bottom panel: The recommended configuration moves the SC-QAM block to the edge of the video, places a 2nd OFDM channel and creates exclusion bands per policy and for the detected insertion. 12

Figure 8 - Top 25 configurations for RPDs post 2nd OFDM placement. 13

Figure 9 - Top panel: packing of video QAMs with programs in which color mapping indicates encoding level. Bottom panel: same data color-coded to indicate the program category. 14

Figure 10 - Top panel: starting state in which video QAMs are neither completely filled nor closely packed. Middle panel: the view after invoking the greedy algorithm for improving the fill factor. Bottom panel: The view after closely packing the video QAMs. 15

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Tables

Title **Page Number**

Table 1 - 2×2 decision matrix showing how to interpret a channel's combination of energy & occupancy states. 7

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Comcast's network is undergoing an evolution towards the virtualized Cable Modem Termination System (vCMTS) and remote physical layer (PHY) architecture as part of the journey to 10G [1]. As of early 2023, millions of homes passed can enjoy the multi-Gbps downstream and a few 100 Mbps upstream speeds made possible by deploying a full-width orthogonal frequency division multiplexing (OFDM) channel at 810-1002 megahertz (MHz) and an orthogonal frequency division multiple access (OFDMA) channel in the mid-split (40-85 MHz) region of the spectrum across the vCMTS platform. The path to 10G involves incrementally deploying additional OFDM/OFDMA channels under the Data Over Cable Service Interface Specification (DOCSIS) 4.0 Full Duplex (FDX) technology [2]. However, reconfiguring the spectrum across markets and localities in support of the 10G roadmap is a daunting task because of the complex spectrum management involving the linear video and single carrier-quadrature amplitude modulation (SC-QAM) channels that need to be accommodated, moved, or phased out by converting linear QAM-based video to IP-based video. Adding to the challenge is the prevalence of local video insertions used by multi-dwelling units (MDUs), managed by the local regions, that may not be properly documented in a central location. In this technical paper, we introduce a methodology for discovering spectrum occupancy based on an object-oriented model of the cable modem's full band capture (FBC). Alongside the configuration data collected from the vCMTS, the FBC model allows the discovery of a host of artifacts that, if not addressed, would be problematic for turning up energy in new regions of the spectrum. The introduced methodology served as the basis for a machine learning pipeline established for automating the required spectrum enablement pre-checks and for generating detailed reporting on all discovered issues.

Furthermore, the deployed pipeline for artifact discovery was repurposed to support the effort for deploying a second OFDM channel on the vCMTS platform within the 618-810 MHz range. For this second phase of spectrum activation, the pipeline was also tasked with generating remote PHY device (RPD) configurations tailored for each node, with the goal of eliminating the vacant spectrum (gaps) between video and downstream SC-QAM and creating OFDM exclusion bands informed both by regional policies and by the discovery of insertions on the node.

Lastly, we also present our early explorations of the third phase of spectrum activation, which involves repacking video channels to free up the spectrum for FDX. Traditionally, video programming templates have been constructed locally at a market level and were not directly subject to centralized planning and optimization. These early explorations reveal opportunities to free up the spectrum by ensuring that each video QAM is fully packed with multiplexed programs (TV channels). Developing an artificial intelligence (AI) driven pipeline to automatically calculate the capacity requirements and optimal spectrum allocations given a variety of variables such a customer premise equipment (CPE) and network technology penetrations, traffic load, infrastructure and technology constraints, dynamic efficiencies from PHY and media access control (MAC) layer optimization micro-services is our vision. Fully automating the spectrum management per service group not only optimizes capacity and speed, but removes the inevitable human errors, time to configure and inefficiencies required when operating networks at the scale of 10s of millions of broadband customers.

2. Phase 1 of Activations: Mid-split and OFDM at 810-1002 MHz

In a previous NCTA technical paper, we shared our experience around enabling the mid-split spectrum by activating OFDMA on the vCMTS platform [3]. The focus of that contribution was on the profile management application (PMA) solution for constructing interval usage code (IUC) profiles, the capacity performance of OFDMA, and the upstream speeds that can be made accessible to our customers with OFDMA. Since then, thousands of nodes are now mid-split enabled. The deployment of mid-split

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

occurred in tandem with standardizing the OFDM channel placement to the 810-1002 MHz frequency range. The rationale behind syncing the 2 efforts (mid-split enablement and OFDM expansion) is to (1) offer higher upstream & downstream speeds simultaneously and (2) minimize the number of incremental configuration changes toward the end-state, in which FDX is fully deployed and symmetrical multi gigabit per second (Gbps) speeds are possible.

The 2 main challenges in the deployment journey were "cleaning up" the spectrum channel map to make it compatible with the newly opened regions of the spectrum and implementing automations around pre-enablement checks, and around the generation & transaction of the spectrum configurations that are tailored to each node on the network. To this end, one obvious exercise involved swapping legacy amplifiers for ones that are mid-split and 1 gigahertz (GHz) capable. However, issues surrounding deployment go beyond missing amplifier upgrades. To highlight those challenges, consider the high variability in spectrum configuration before mid-split deployment depicted in Figure 1. The figure shows that the top 25 configurations cover ~46% of the pre-mid-split RPD population. Note that an active carrier was deliberately placed at 999 MHz to aid in discovering potential roll-off. In all the permutations shown in Figure 1, the 810-1002 MHz spectrum is almost free for placement of the OFDM channel. Though there are edge cases not included in the top 25 in which SC-QAM channels need to be relocated to free up the 810-1002 MHz region. Furthermore, some permutations explicitly show the existence of a local video insertion while others may have an active insertion that is undocumented.

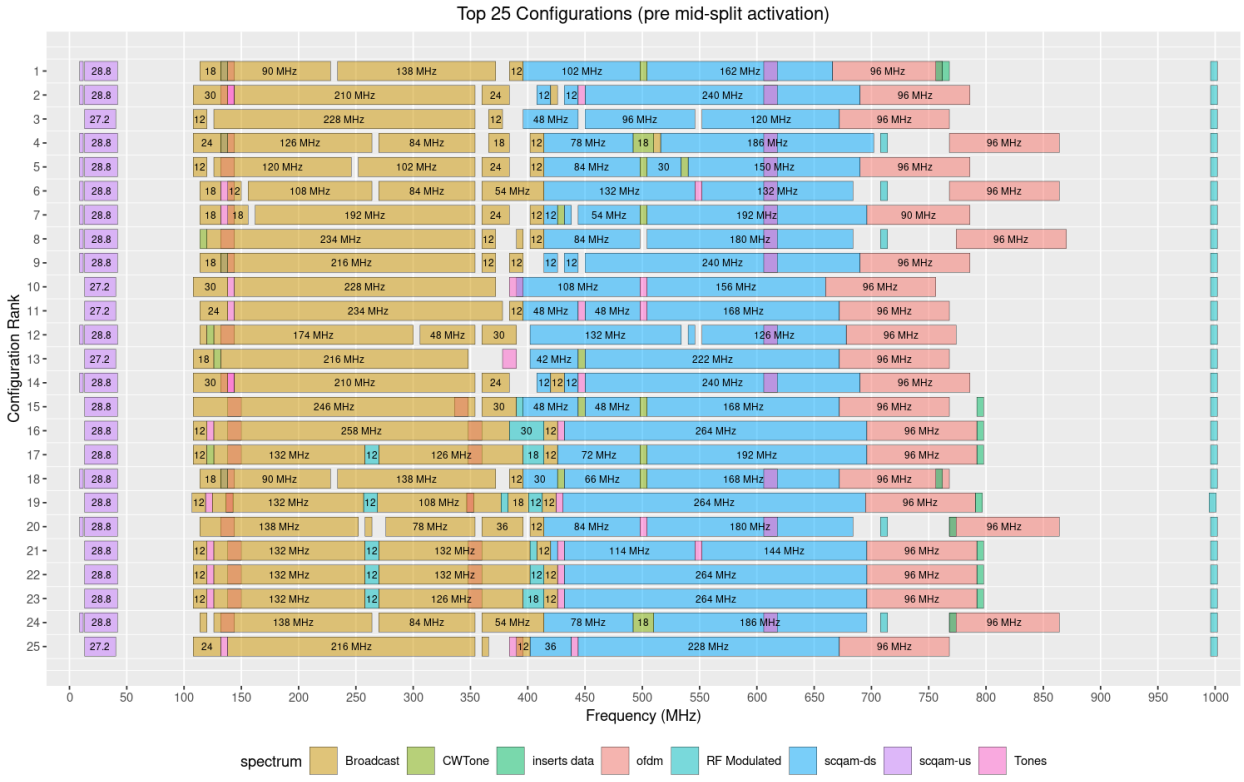


Figure 1 - Top 25 configurations for RPDs that haven't yet undergone mid-split activation. ~46% of this RPD population is covered by the top 25. Aside from the upstream SC-QAM block, there exists high variability across configurations.

In contrast, Figure 2 shows the top 25 configurations post mid-split activation, covering ~57% of this RPD population. The upstream and OFDM regions are now standardized across this population, while the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

high variability remains in the video and downstream SC-QAM regions (to be addressed over the next iterations of spectrum evolution).

Given the high variability in configurations, transforming the spectrum from the picture of Figure 1 to that of Figure 2 involved dealing with the following list of potential problems:

- Existence of undocumented local video insertions (usually serving an MDU) within the mid-split or the 810-1002 MHz region. In general, this type of problem extends to any new region of the spectrum to be used for activating DOCSIS channels.
- Existence of severe roll-off in the 810-1002 MHz region. This could be indicative of a missed legacy amplifier or localized to a tap serving a few dwellings.
- Existence of severe ingress in a currently vacant region of the spectrum, which will be used for relocating downstream SC-QAMs channels to free up the 810-1002 MHz for OFDM.

It was imperative to detect each of the scenarios above pre-deployment to avoid negatively impacting customers. Henceforth, we will refer to the existence of any of the above as an "artifact". In another previous technical paper, we shared a few concepts around artifact detection, based on an analysis of the cable modem's FBC [4]. These early explorations were ad-hoc in nature and involved developing a separate algorithm for the detection of each artifact. Next, we present a unified approach for artifact detection based on an object-oriented model of the FBC.

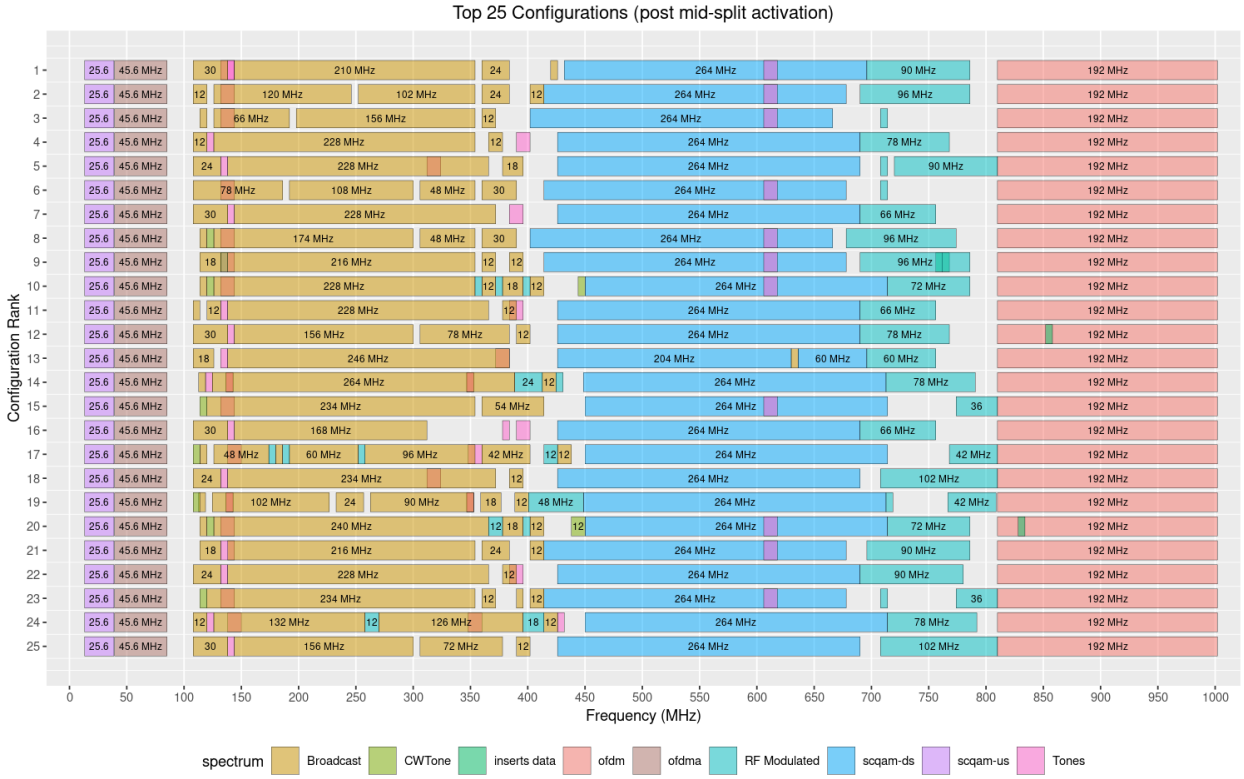


Figure 2 - Top 25 configurations for RPDs that are mid-split active. ~57% of this RPD population is covered by the top 25. Though upstream and OFDM blocks are aligned across permutations, there remains high variability across video and downstream SC-QAM.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. An Object-Oriented Model for the Full Band Capture

The idea of the new approach to pre-enablement checks is to conduct these systematically rather than what was previously a set of ad-hoc rules, each relating to a specific region of the spectrum and a specific type of artifact. The new approach recognizes that any of the checks performed falls broadly under a single task: matching the spectrum configuration against the cable modem's FBC and deciding on whether the two views agree or conflict. The latter is an indication of a host of problems which we are interested in detecting. The high-level concept is summarized in the 2x2 decision matrix shown in Table 1.

It follows that we benefit from a generic approach to processing the FBC, in which signal is aggregated across each 6 MHz channel and a Low/High label is assigned to each based on some mechanism for designating that the channel exhibits an energy level commensurate with signal vs. noise.

Table 1 - 2×2 decision matrix showing how to interpret a channel's combination of energy & occupancy states.

		Channel (QAM) Energy (based on FBC)	
		Low	**High**
Channel (QAM) Occupancy (based on RPD configuration)	**Vacant**	• No evidence of noise or local insertions • Clear for placing QAMs/OFDM	• Possible local insertion • Possible ingress, etc. (some other RF (Radio Frequency) pattern)
	Occupied	• For a 999 MHz carrier, indicates existence of roll-off • Possible ingress, etc. (some other RF pattern)	• Healthy spectrum in general • For a 999 MHz carrier, indicates no evidence for roll-off

An object-oriented model was built for this purpose. The FBC object is instantiated with 3 inputs:

- **FBC data**: an array (float) representing the raw measurement typically at ~117 kilohertz (KHz) resolution.
- **Configuration data**: a dictionary in which keys are the type of spectrum and values are arrays of the (start,end) frequencies of the contiguous spectrum blocks. Example:

```
{'broadcast': [(114.0, 120.0), (126.0, 222.0), (228.0, 240.0), (246.0, 414.0),
               (426.0, 438.0)],
'rf modulated': [(222.0, 228.0), (240.0, 246.0), (414.0, 426.0), (438.0, 450.0),
                 (714.0, 786.0)],
'tones': [(120.0, 126.0), (138.0, 150.0), (348.0, 360.0)],
'local inserts': [(792.0, 798.0)],
'ofdm': [(810.0, 1002.0)],
'ofdma': [(39.4, 85.0)],
'scqam-ds': [(450.0, 714.0)],
'scqam-us': [(13.2, 38.8)]}
```

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- **HRC flag**: a Boolean indicating whether the plant follows the standard (False) or the harmonic related carrier (HRC) plan in which QAM channels are offset by 1.25 MHz (True).

Once instantiated, the following host of pre-processing steps are invoked by the class:

- The spectrum is "cleaned" by removing guard regions. These have fixed locations depending on whether the plan is standard vs. HRC.
- Energy is aggregated per channel/QAM according to the standard EIA mapping.
- A clustering algorithm determines the threshold that best separates the channels across spectrum into 2 groups: noise and signal. Consequently, each channel is labeled as Low/High depending on whether it belongs to the noise/signal group. It is critical to create such a designation without knowledge of the spectrum configuration according to the RPD.
- An additional label is added to the channel based on knowledge of the RPD configuration. It contains all spectrum types assigned to the channels.

Here's an example of instantiating the FBC object and querying a single channel:

```
example_spectrum = FullBandCapture(raw_spectrum=x, config=config_dict, hrc=is_hrc)
example_spectrum()[89]
```

Executing the code above returns the attributes of electronic industries association (EIA) channel 89 (corresponds to a center frequency of 615 MHz according to the EIA plan):

```
{'frequency': 615, 'range': (5174, 5218), 'energy': -7.98, 'configuration': ['scqam-
ds', 'Tones'], 'label': 'High'}
```

We are now ready to "interrogate" the FBC object to check for the presence of artifacts of interest. Below are some examples.

3.1. MDU Insert within 810-1002 MHz

The example shown in Figure 3 represents detection of an MDU insert in the 810-1002 MHz region intended for OFDM expansion. With the instantiated FBC object, detecting inserts involves the following single line of (P*ython*) code:

```
{eia:attributes for eia,attributes in example_spectrum.channel_data.items()
if not any(attributes['configuration']) and
   attributes['label']=='High' and
   attributes['frequency']>810}
```

Executing the code above returns the suspect insert at 855 MHz:

```
{134: {'frequency': 855, 'range': (7222, 7266), 'energy': -19.98, 'configuration': [],
'label': 'High'}}
```

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

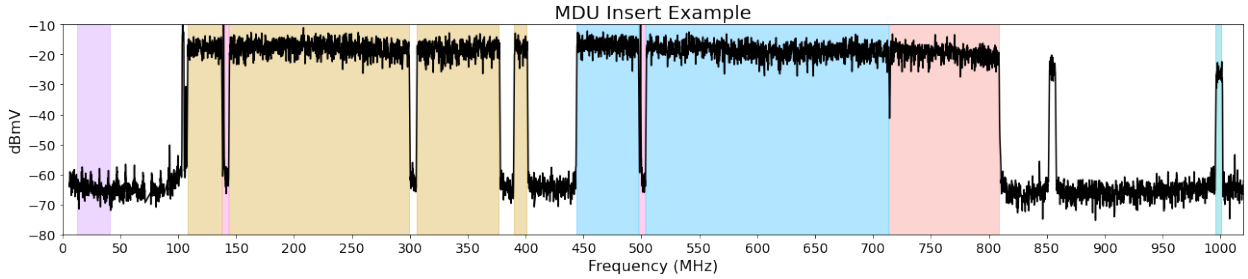


Figure 3 - Example FBC that shows an undocumented local insert at 855 MHz. The overlaid colored shadings correspond to the RPD configuration according to the color map in legend of Figure 1.

3.2. Roll-off within 810-1002 MHz

The example shown in Figure 4 represents detection of potential roll-off in the 810-1002 MHz region intended for OFDM expansion. The roll-off is inferred from the relatively low signal of the 999 MHz carrier (15 dBmV lower than the average signal level). Once again, with the instantiated FBC object, detecting roll-off involves executing the following single line of (P*ython*) code:

```
{eia:attributes for eia,attributes in example_spectrum_channel_data.items()
if attributes['configuration']==['RF Modulated'] and
   attributes['energy']< (example_spectrum_signal_level-15) and
   attributes['frequency']==999}
```

Executing the code above returns the low signal 999 MHz carrier:

```
[{158: {'frequency': 999, 'range': (8451, 8495), 'energy': -37.77, 'configuration':
['RF Modulated'], 'label': 'Low'}}]
```

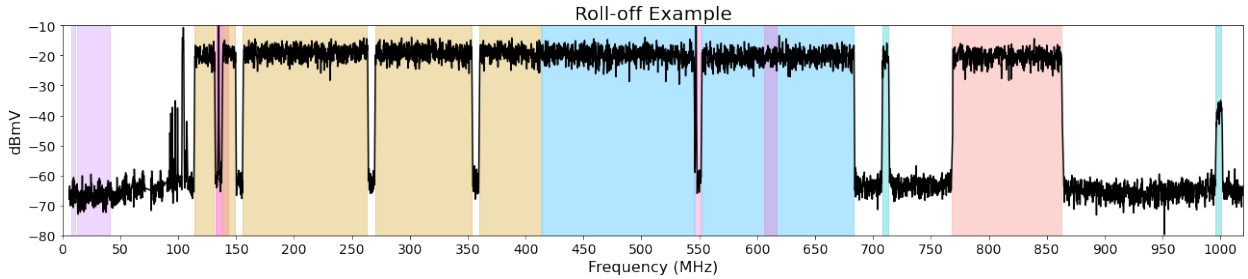


Figure 4 - Example FBC that shows roll-off at the 999 MHz carrier. The overlaid colored shadings correspond to the RPD configuration according to the color map in legend of Figure 1.

4. The Pipeline for Artifact Detection

The pipeline for artifact detection scans the network once a day, searching for the type of artifacts described in the previous section. Note that the "heavy lifting" is done by the FBC model, which, in effect, processes the raw data in the same way regardless of the intended artifact. Since the detection mechanism based on FBC discovers artifacts at a cable modem's level, this data is aggregated to the RPD level for decision-making on whether to move forward with spectrum activation or block the RPD until the discovered issues are resolved. This aggregation is rule-based, requiring a certain number of cable modems to exhibit the artifact for the RPD to be blocked from spectrum activation. A typical threshold

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

used is ~10 cable modems. In setting the threshold, we attempt to strike a delicate balance between False Positives and False Negatives. One approach for threshold selection is based on constructing a sensitivity curve such as the one shown in Figure 5. A rule-of-thumb for threshold selection follows the "elbow method"; i.e., setting the threshold close is the elbow within the arm-resembling sensitivity curve. This method does not consider the unequal impact of generating False Positive and False. For example, if the consequence of False Negatives is catastrophic then it makes sense to shift the threshold further to the left on the curve (i.e., threshold < 10).

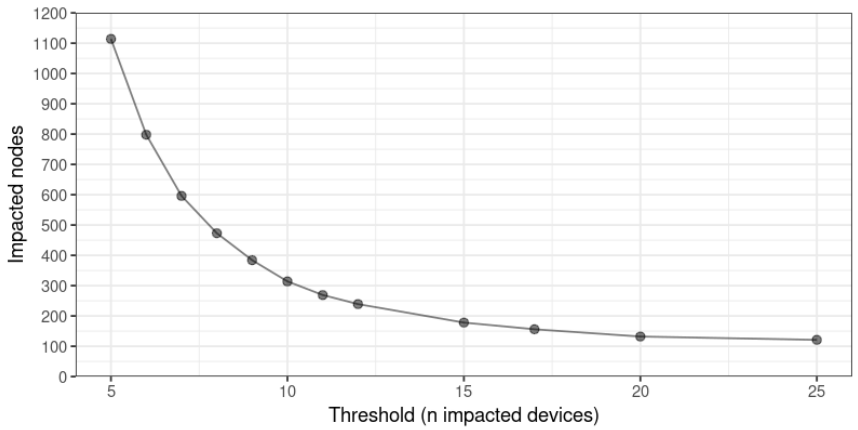


Figure 5 - Sensitivity curve for insertion detection. The *x*-axis is the selection threshold in the number of cable modems that show roll-off. The *y*-axis is the corresponding number of impacted nodes. The "elbow method" would set the threshold close to ~10.

Since the pipeline uses fresh telemetry, once the issue is fixed the blocked RPD will automatically be removed from the blocked list for spectrum activation on the following pipeline run. Therefore, manual management of a blocked/cleared RPD list is not required.

The pipeline also employs algorithms for narrowing down the root cause of the issue where possible. For example, when severe roll-off is observed impacting many devices on the node, we invoke graph algorithms powered by the graph representation of the network [5, 6] to determine if the issue is related to a missed amplifier vs. multiple separate and unrelated problems at each dwelling. This additional context helps the field techs to quickly narrow down the search for the root cause of the problem or work with the customers to resolve their home issues.

Figure 6 shows an example root cause analysis in which 23 cable modems exhibited severe roll-off at the 999 MHz carrier (depicted as pink-shaded rounded boxes on the graph). A least common ancestor algorithm determined that the source of the problem is the amplifier marked with an asterisk. This sort of finding is invaluable to the field techs to swap the missed amplifier in a timely manner.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

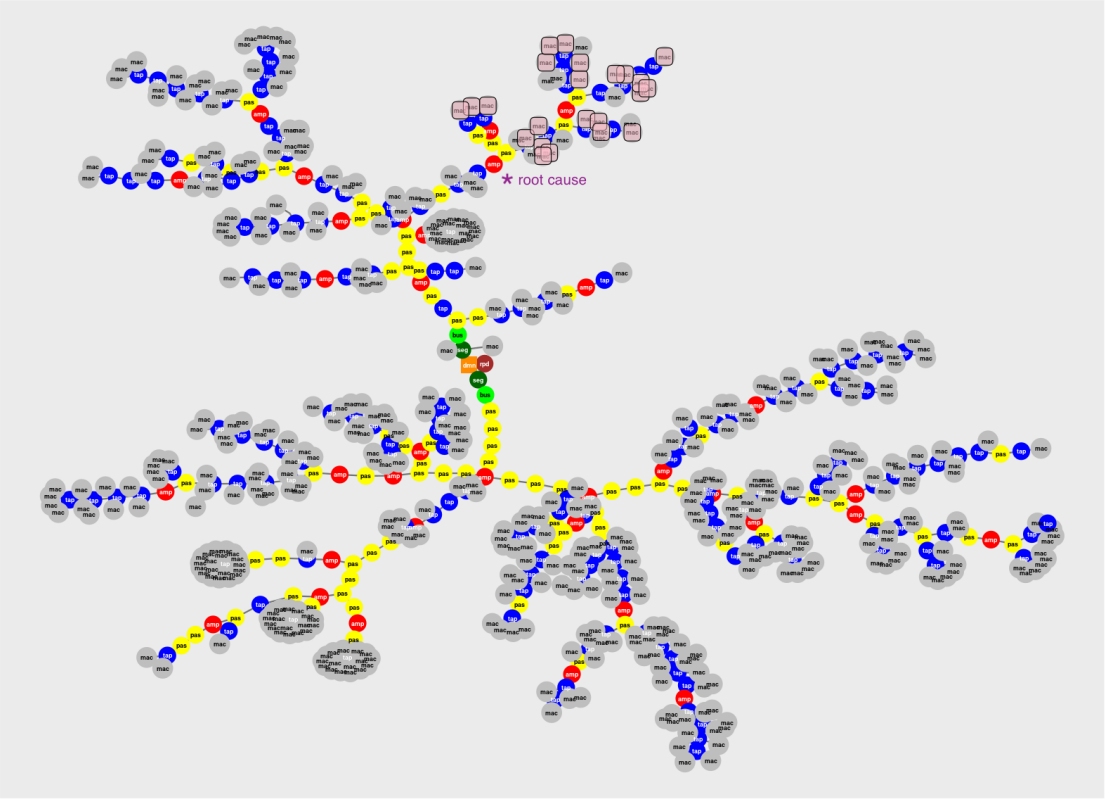


Figure 6 - A graph of the network showing all elements under a node. Cable modems with roll-off are highlighted in pink (top center area). The amplifier (red circle) responsible for the roll-off is marked with an asterisk. This amplifier was missed in the upgrade construction activity.

5. Phase 2 of Activations: 2nd OFDM

Phase 2 of spectrum activation is centered around adding a 2nd OFDM channel. This time however, the OFDM channel will not have a standardized location but will be constrained by the availability of spectrum from the edge of the video block (varies by node) to the edge of the first OFDM channel (810 MHz). The basic set of rules that govern the placement of the 2nd OFDM channel is the following:

- Move the downstream SC-QAM spectrum to the top edge of the video and remove any gaps within the downstream SC-QAM blocks, if available and permissible.
- Place a second OFDM channel with an upper edge at 810 MHz (lower edge of the first OFDM channel) and a lower edge at either the top of the downstream SC-QAM or 618 MHz (whichever is higher); this ensures that the OFDM channel width is no larger than 192 MHz.
- Create exclusion bands within the OFDM channel to accommodate detected MDU inserts as well as static exclusion frequencies dictated by policy. These static exclusions are intended to support several standard locations for video channels that will facilitate the video shift for FDX by simulcasting some video controller data and a standard location for QAM video in MDUs compatible with DOCSIS 4.0 spectrum.
- Check that created channel satisfies DOCSIS specifications; for example, in relation to the ratio of exclusion regions to usable OFDM spectrum, the proximity of exclusion bands to edges, etc.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Create a JavaScript Object Notation (JSON) formatted file to include the new configuration (to be transacted by an automated flow).
- The physical link channel (PLC) and OFDM location are further configured to support cellular band leakage measurement markers based on learning of legacy analog nodes configuration clusters so that leakage is detectable across all nodes in the region without more complicated leakage configuration geo-zones.

An example of the above logic in action is shown in Figure 7. The top panel shows the current configuration of an RPD. Notice the wide vacancy between the broadcast video and the downstream SC-QAM block. For this RPD, regional policy dictates establishing exclusion bands reserved for future local insertion usage at 747, 753, and 795 MHz. The middle panel, however, reveals an undocumented insert active today at 777 MHz. The bottom panel shows the recommended configuration, in which SC-QAM is moved to the edge of the video, a 2nd OFDM is placed between 648 and 810 MHz, and 3 DOCSIS standard compliant exclusion bands are created to cover the 4 insert locations (note, 747 and 753 are adjacent, requiring a single exclusion band).

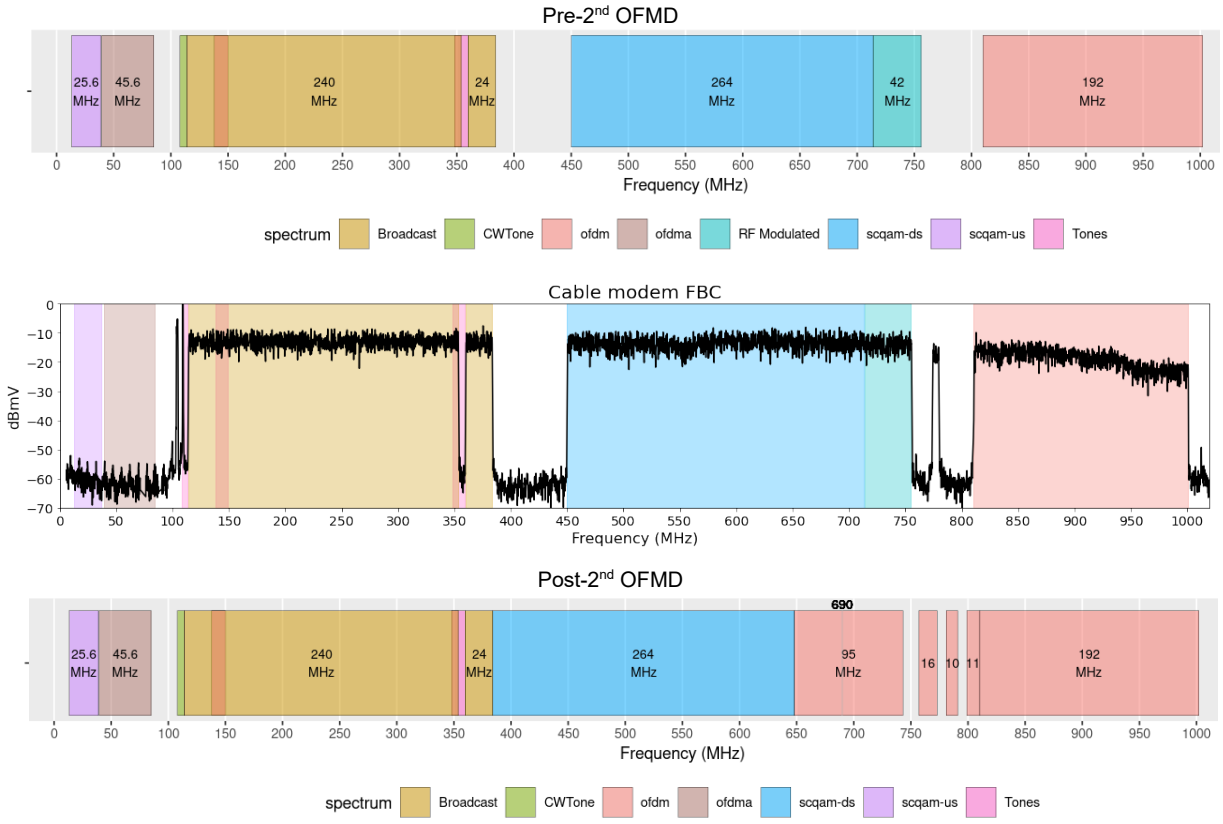


Figure 7 - Top panel: Spectrum configuration pre-2nd ofdm shows a "wasted" gap between video and downstream SC-QAM. Middle panel: an FBC from a cable modem reveals an undocumented MDU insert at 777 MHz. Bottom panel: The recommended configuration moves the SC-QAM block to the edge of the video, places a 2nd OFDM channel and creates exclusion bands per policy and for the detected insertion.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

With the established logic, Figure 8 offers a view of the top 25 configurations post-deployment of the 2nd OFDM channel. The spectrum is becoming more packed. Next, we'll shift our focus to the broadcast video regions.

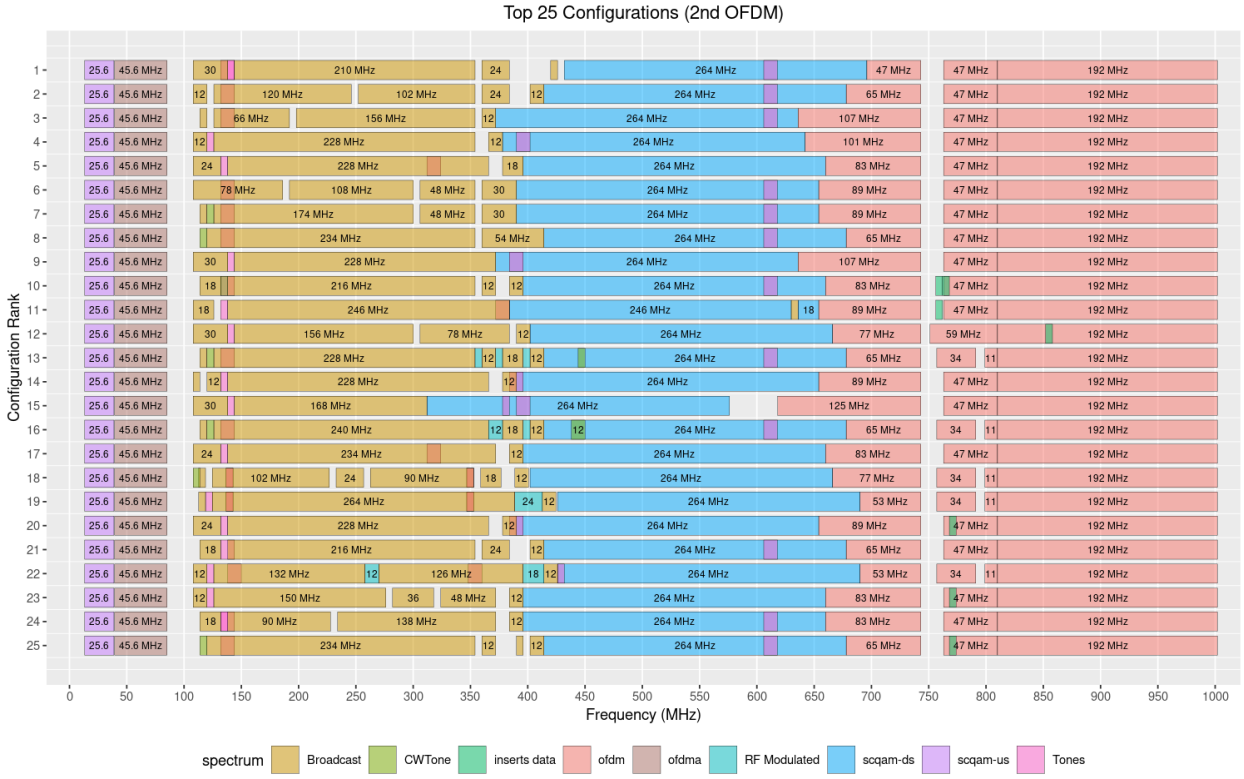


Figure 8 - Top 25 configurations for RPDs post 2nd OFDM placement.

6. Phase 3 of Activations: Close-Packing of Video

Phases 1& 2 of spectrum activation do not impact the video configuration. As seen in Figures 1-2 & 8 the video blocks are somewhat fragmented and can be made to occupy a narrower range of spectrum. But, making video use of spectrum more efficient goes beyond tighter channel packing. This is because within each QAM, several programming single program transport streams (SPTS) (TV channels) are multiplexed together in a multi-program transport stream (MPTS). Depending on the video encoding level, e.g., music, standard definition (SD), high definition (HD), of each program there may be room for moving content around such that each QAM is fully utilized.

For the ensuing discussion, to avoid confusion, we use the term "QAM" to refer to the 6 MHz frequency channel and "program" to refer to the TV channel. We present some findings based on exploratory data analysis performed on an example programming template. The first task was to examine how each video QAM fill factor varies by program encoding definition. The assumption is that a single QAM carries up to 15 SD programs, 9 HD programs, or 50 music choice programs. Figure 9 (top panel) shows the distribution of content by QAM for a given template highlighting the encoding level (fill color). The view confirms the assumed limits. However, we also see examples of QAMs having programs with a mix of different encoding levels. We also see a QAM that extends beyond the limit as it contains 50 music choice programs and 1 SD program (at ~340 MHz).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

It is also assumed that programs of similar categories (e.g., Sports, Family, Movies, Latin, etc.) are typically multiplexed together. To validate this hypothesis, we assigned each program a category obtained by performing text mining on the full program name. These categories shown in the bottom panel of Figure 9 represent our classification of the type of programming. We do indeed see some clustering of similar type programs on the same QAM.

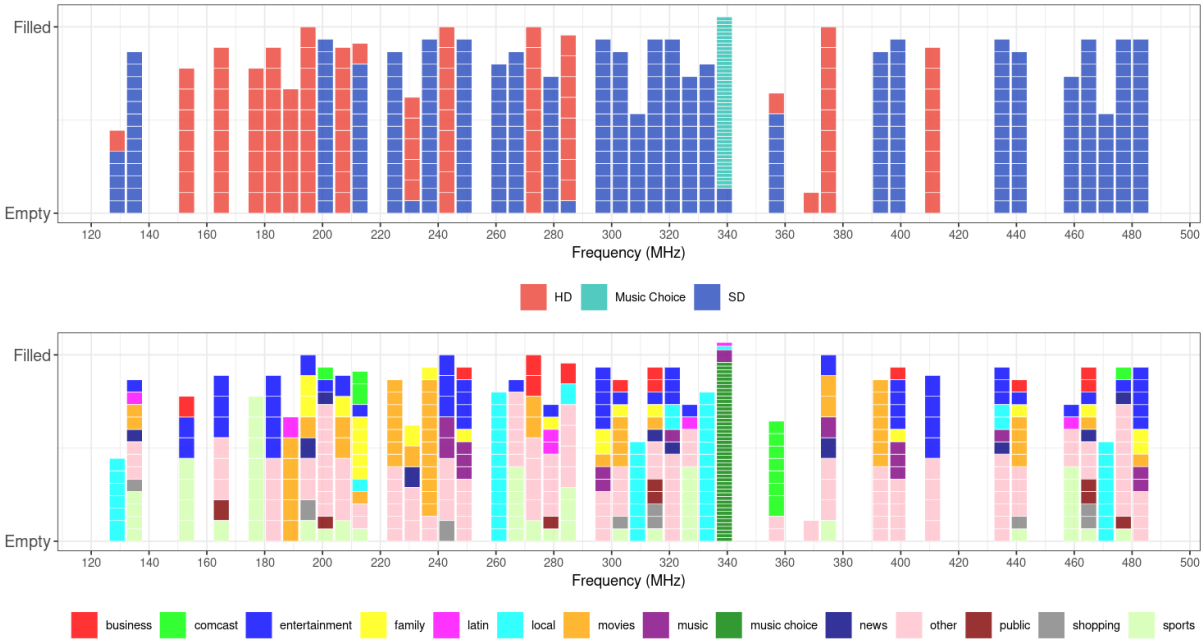


Figure 9 - Top panel: packing of video QAMs with programs in which color mapping indicates encoding level. Bottom panel: same data color-coded to indicate the program category.

We propose an algorithm for repacking video that is a "greedy" approach to relocate programs from scarcely populated QAMs to densely populated QAMs, with the preference for filling each QAM with programs of the same type of encoding, if possible. In more detail:

- Partially populated QAMs are ranked according to fill factor (least to most).
- Programs from the least populated QAMs get migrated to the most populated QAMs (preferably the same encoding level).
- The process repeats iteratively while tracking a performance metric on the overall fill factor of the template (entire video spectrum).
- Once no further gains are made, the process terminates.
- Lastly, QAM frequencies are reassigned to remove any gaps between video QAMs. This step (bottom panel of Figure 10) can be informed by the policy.

Figure 10 is a representative example of the above logic in action. The top panel represents the current state, in which the video occupies 42 QAMs. The middle panel shows the result of applying the greedy algorithm to improve the fill factor. For this template, 7 QAMs were freed, and the final lineup is almost filled, with few exceptions. The bottom panel shows the state after reassigning QAMs to frequencies such that the entire video is one contiguous block.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

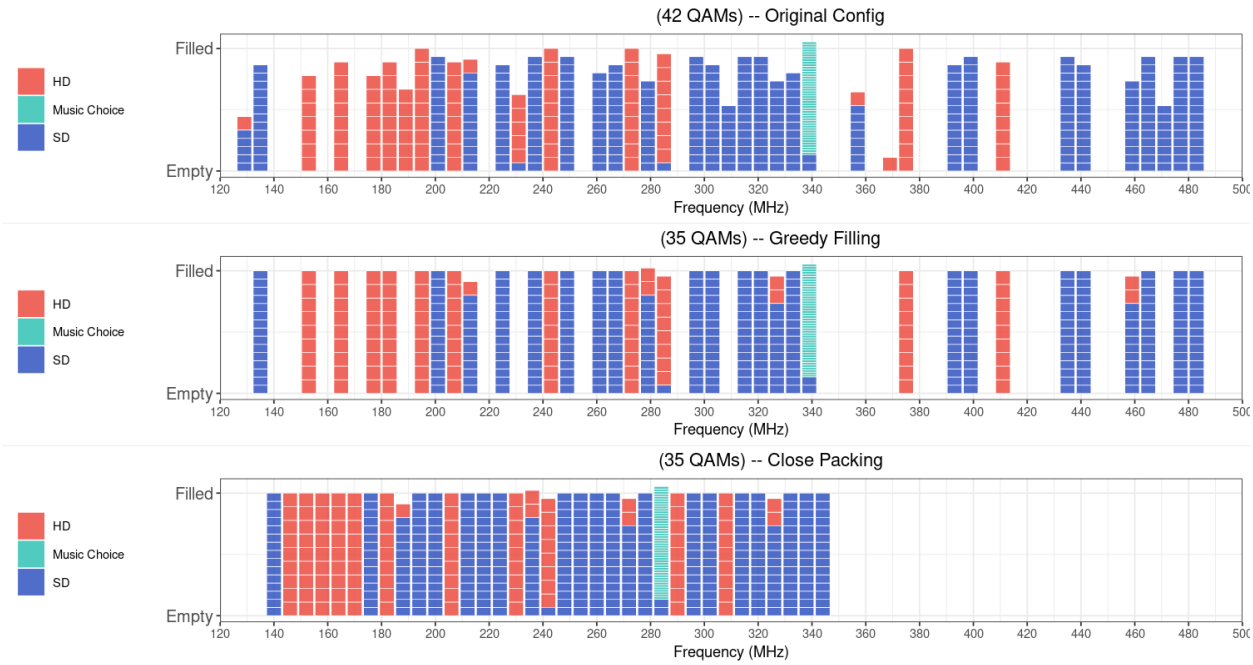


Figure 10 - Top panel: starting state in which video QAMs are neither completely filled nor closely packed. Middle panel: the view after invoking the greedy algorithm for improving the fill factor. Bottom panel: The view after closely packing the video QAMs.

This allocation and efficiency of video management is helpful in the following ways.

- Video can be consolidated to reduce the number of QAMs so that the spectrum can be used more efficiently, enabling increased DOCSIS IP (Internet Protocol) spectrum, for higher speed 10G broadband products.
- It allows us to manage the spectrum appropriately as we evolve to DOCSIS 4.0 based on the variety of broadband and set top box (STB) device capabilities.
- It allows us to easily evolve QAM Video to IP video efficiently over time as we expand the FDX spectrum.

This Machine learning pipeline is currently under development aligned with our launch of multi-Gbps symmetrical service around the time of the SCTE conference. We will report on the FDX version of this pipeline that is being prepared to calculate optimal allocations of spectrum for D3.0 modems, D3.1 modems, D4.0 modems , IP Video and QAM video spectrum collectively required at each stage of broadband speed increases that will be deployed in the next two years as we evolve to full FDX Band where the video channels will be reduced and shifted to the spectrum above the FDX band. This will result in a fully automated spectrum activation and management pipeline that we will describe in a future SCTE forum.

7. Conclusion

We presented a host of algorithms in support of activating the spectrum. In phase 1 of activations (mid-split and 810-1002 MHz), we focused on developing a pipeline for artifact detection based on an object-oriented model of the cable modem full-band capture. In phase 2, the pipeline was extended to generate the spectrum configuration for the second OFDM channel including exclusion bands informed by the discovery of local insertions. Finally, phase 3 of activation, which is currently in the development stage, is

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

concerned with freeing up the spectrum for FDX by reducing the footprint of broadcast video. This is achieved through re-multiplexing of programs such that each video QAM is fully utilized.

This base spectrum management pipeline for activation of new spectrum will be converged with the real time spectrum management introduced in last year's SCTE program [4] and in other Comcast papers in this year's SCTE program [7] to achieve the fully automated AI-supported pipeline for deploying, activating and ongoing management of our valuable spectrum assets.

Abbreviations

AI	artificial intelligence
CPE	customer premise equipment
DOCSIS	data over cable service interface specification
EIA	Electronic Industries Association
FBC	full band capture
FDX	full duplex
Gbps	gigabits per second
GHz	gigahertz
HD	high definition
HRC	harmonic related carrier
IP	internet protocol
IUC	interval usage code
JSON	JavaScript object notation
KHz	kilohertz
MAC	media access control
MDU	multi dwelling units
MHz	megahertz
MPTS	multi-program transport stream
NCTA	National Cable & Telecommunications Association
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
PHY	physical
PLC	physical link channel
PMA	profile management application
QAM	quadrature amplitude modulation
RF	radio frequency
RPD	remote PHY device
SC-QAM	single carrier quadrature amplitude modulation
SCTE	Society of Cable Telecommunications Engineers
SD	standard definition
SPTS	single program transport streams
STB	set top box
TV	television
vCMTS	virtual cable modem termination system

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

1. *Roaring into the '20s with 10G*, R. Howald, R. Thompson, S. Ozer, D. Rice & L. Wolcott, T. Cloonan, R. Cloonan, J. Ulm, and J. Ariesen; NCTA technical paper, 2020.

2. *Developing the DOCSIS 4.0 Playbook for the Season of 10G*, R. Howald, J. Williams, J. Cave, O. Ekundare, and M. Petersen; NCTA technical paper, 2021.

3. *Deploying PMA-Enabled OFDMA in Mid-Split and High-Split*, M. Harb, D. Rice, K. Dugan, J. Fereirra, and R. Narayanaswamy; NCTA technical paper, 2022.

4. *Time is Ripe for D3.0 Farming: Achieving Optimal Spectral Efficiency by Allocating D3.0 Spectrum to OFDM*, M. Harb, C. Humble, S. Ozer, and D. Rice; NCTA technical paper, 2022.

5. *How Network Topology Impacts Rf Performance: A Study Powered By Graph Representation Of The Access Network*, M. Harb, K. Subramanya, R. Narayanaswamy, S. Walavalkar, and D. Rice; NCTA technical paper, 2021.

6. *Graph Algorithms and Real Time Telemetry for Intelligent Plant Operations*, B. Lutz and M. Stehman; NCTA technical paper, 2023.

7. *Towards Fully Automated HFC Spectrum Management*, J. Zhu, M. Harb, D. Rice, J. Howe, and C. Humble Insert; NCTA technical paper, 2023.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Adapting D4.0 Scheduler to Achieve the Optimal Latency

A Technical Paper prepared for SCTE by

Hongbiao Zhang
Architect Wireline Solutions
Casa Systems
100 Old River Rd
Andover, MA01886
978 688 6706
hongbiao.zhang@casa-systems.com

Chain Lee
Principal Software Engineer, R&D SW Cable
Casa Systems
100 Old River Rd
Andover, MA01886
978 688 6706
clee@casa-systems.com

Vishnuvinod Sambasivan
Principal SQA Engineer, R&D SW Cable
Casa Systems
100 Old River Rd
Andover, MA01886
978 688 6706
vishnuvinod.sambasivan@casa-systems.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Proactive Granting ... 3
3. MGB Requirements and Implications ... 4
4. D4.0 Upstream Scheduling Enhancements ... 5
   - 4.1. Filling up Extra Space ... 6
   - 4.2. Selecting OFDMA Frame Size ... 6
   - 4.3. Scheduling Across Multiple Channels ... 6
   - 4.4. Aggregating Service Flows ... 6
   - 4.5. Skipping Small Bursts ... 7
   - 4.6. Allocating Proactive Grants ... 7
5. Experiments and Analysis ... 8
   - 5.1. The Test Cases ... 8
   - 5.2. Test A: Performance Degradation with MGB ... 9
   - 5.3. Test B: Performance Enhancements ... 9
   - 5.4. Further Analytical Results ... 10
     - 5.4.1. UHS-300 ... 11
     - 5.4.2. UHS-396 ... 12
     - 5.4.3. UHS-492 ... 13
6. Conclusion ... 13

Abbreviations ... 14

Bibliography & References ... 14

List of Figures

Title **Page Number**

Figure 1 – Effect of MGB ... 5
Figure 2 – Test A: Throughput and Latency vs N ... 9
Figure 3 – Test B: Throughput and Latency vs M with N = 70 ... 10
Figure 4 – Test A Theoretical vs Simulation Results ... 10
Figure 5 – Test B Theoretical vs Simulation Results ... 11

List of Tables

Title **Page Number**

Table 1 - FDD Minimum Grant Bandwidth ... 4
Table 2 – Throughput vs N and M (UHS-300) ... 12
Table 3 – Throughput vs N and M (UHS-396) ... 12
Table 4 – Throughput vs N and M (UHS-492) ... 13

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The internet today has seen a blast of cloud-based services that has an increasing demand for fast access and short response time, a.k.a. low latency requirements. These services include online gaming, cloud computing, online trading and monetary transactions, video conferencing, VR/AR, etc. To cable networks, they appear as Over-the-Top (OTT) applications that are not managed by operators. Therefore, they cannot be serviced with higher priority like the MSO-offered voice or video. This implies that the cable systems need to provide low latency not only to MSO-managed QoS services, but also to all best effort services just the same.

In the past, DOCSIS cable is known for large latency as compared with competition technologies such as FTTP, especially the media access latency in the upstream direction. In D3.1, Low Latency DOCSIS (LLD) was proposed to alleviate the problem. This includes Proactive Grant Service (PGS) that potentially reduces request-grant cycle by 2/3. Despite the advantages, to this day the authors have not seen large deployment of this mechanism, possibly and understandably due to the fact that the current sub-split and mid-split plants still have limited upstream capacity, and thus favor bandwidth utilization over low latency.

The introduction of high-split and ultra-split DOCSIS pushes the equilibrium to a new level: with the advent of additional upstream bandwidth that is several times larger than the current, it is affordable to pre-allocate some grants prior to any requests so as achieve a smaller access latency. Similarly, it is also affordable to allocate some more grants on top of the requested bytes. In other words, it does not require CMTS to use accurate grant counts to match the requests.

On the other hand, ultra-split DOCSIS as defined in D4.0 requires Minimum Grant Bandwidth (MGB) to be enforced for any transmitting CM at any time of its transmission across the entire extended spectrum, whenever the extended spectrum exceeds 192 MHz. This requirement imposes new challenges to upstream scheduling, including that of PGS.

The paper assumes D4.0 Frequency Division Duplex (FDD) is adopted, yet most of the discussions and solutions apply to Full Duplex (FDX) DOCSIS as well. In this paper we show that both latency and data speed of user applications' might be significantly impacted due to the requirements of MGB. We further discuss various methods in D4.0 upstream scheduling that tackle the MGB issue, and how PGS could be realized using these methods. Thus, it is expected to achieve an optimal latency as well as the desired data speed in operators' D4.0 networks.

2. Proactive Granting

In the past, DOCSIS cable is known for large latency as compared with competition technologies such as FTTP, especially the media access latency in the upstream direction. The lengthy media access time is due to the request-and-grant based mechanism, which requires a request being sent by a CM and a subsequent grant being assigned by the CMTS before any upstream data can be transmitted by the CM. One of the reasons for adopting this mechanism since day one is possibly because early stage DOCSIS systems have a relatively low data rate in the upstream direction. Meanwhile there are a great number of users sharing the same upstream channel, with varying distances from the CMTS. These attributes suggest that upstream scheduling design would favor bandwidth utilization over latency, and the CM and CMTS have to use accurate byte counts in calculation.

In D3.1, Low Latency DOCSIS (LLD) was proposed to address the latency issues. Among a set of LLD features, Proactive Grant Service (PGS) was introduced and targeted at reducing the request-grant cycle

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

([4]). In a nutshell, PGS allows certain amount of traffic to take advantage of proactive grants without going through the request-grant cycle, thus reducing media acquisition time. This improvement in latency comes at a cost of bandwidth utilization, as the pre-allocated grants will be wasted if the targeted service flow has nothing to transmit at the designated grant time. However as we've shown in our experiments ([3]), we can achieve great mean latency and jitter even when the instantaneous traffic rate goes several times over the proactive grant rate. In other words, in PGS operations, latency is insensitive to the fluctuation of the offered traffic rate, as long as the traffic rate doesn't exceed the Maximum Sustained Rate (MSR) limit. Therefore, it is possible to pre-allocate only a small amount of bandwidth regardless of the actual traffic rate, yet achieving an ideal balance between low latency and bandwidth utilization.

It should be noted that the usage of PGS is not limited to special applications, instead, it could be used for all best effort traffic. Even with service flows configured with Best Effort (BE) scheduling type, it is allowed for CMTS to treat them as PGS services and allocate proactive grants without CMs noticing it.

3. MGB Requirements and Implications

In D4.0, Minimum Grant Bandwidth refers to the smallest grant that a CMTS is allowed to assign to a CM in the FDX or FDD upstream channels ([1], [2]). It is needed to ensure fidelity of a D4.0 CM operating in FDX or FDD spectrum. For example, for a total of 2 FDD channels, or 192 MHz in spectrum, the MGB is 10.4 MHz, or 26 minislots across the two channels. For a total of 3 FDD channels, or 288MHz in spectrum, the MGB becomes 16.0 MHz, or 40 minislots across the 3 channels. For a total of 4 FDD channels, or 384 MHz in spectrum, the MGB is 21.2 MHz, or 53 minislots across the 4 channels. The MGB has to be enforced at any time when the corresponding CM is transmitting. The numerical values are shown in the table below.

Table 1 - FDD Minimum Grant Bandwidth

Split Name	FDD Upstream Spectrum (MHz)	Minimum Grant Bandwidth (MHz)	Equivalent Minislots
UHS-300	192	10.4	26
UHS-396	288	16.0	40
UHS-492	384	21.2	53

As compared with a non-FDD channel where a grant allocated for a transmitting CM could be as small as 1 minislot, enforcing a minimum of 26, 40, or 53 minislots for FDD channels makes the scheduling choices much more restrictive. We can see that with MGB imposed for each transmitting CM, there could only be a maximum of 18 transmitting CMs at any instance across the entire extended spectrum. With 2K FFT and a Cyclic Prefix (CP) of 2.5us, there could be between 2.5 - 7.4 OFDMA frames per millisecond depending on the frame size (K), which means there could be at most 44 - 133 CMs on average being scheduled in 1 millisecond time. This is barely enough to support proactive granting for just a couple of hundreds of users alone if the Guaranteed Grant Interval is chosen as 1ms.

We assume that a D4.0 ultra-split network offers more than enough upstream bandwidth to meet users' overall demands of traffic, had it not needed to fulfill MGB. Under this condition, we will demonstrate how enforcing MGB might impact the performance of user applications. These applications appear as OTT to the cable system and are treated with best effort and with the same priority. Therefore round robin is used among the service flows regardless of the offered rate. We will present a concrete example with numerical data in a later section, but here are some intuitions. First, we exam a scenario when there are

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

numerous SFs and traffic load is balanced among them, in which case the scheduler will ensure all of them get serviced in a timely fashion. Even if the total number of SFs exceed what can be scheduled in a scheduling cycle (due to MGB and smaller number of scheduling opportunities), the ones not serviced in the current scheduling cycle may be deferred to the next. As a result, their queues may accumulate quickly, but then get drained quickly as well when they get the scheduling opportunity (again due to MGB), in a round robin fashion. This may cause additional latency but only by a few scheduling cycles.

Now we exam a different scenario when traffic load among the various SFs is not balanced. For example, if there are a group of SFs with light but frequent traffic, and one or more additional SFs with heavy traffic. In that case the latter SFs may experience substantial traffic loss and unbounded latency, as the former SFs occupy much more bandwidth than necessary due to the effect of MGB, resulting in the latter being underserved. An example with MGB of 40 minislots is illustrated in the figure below.

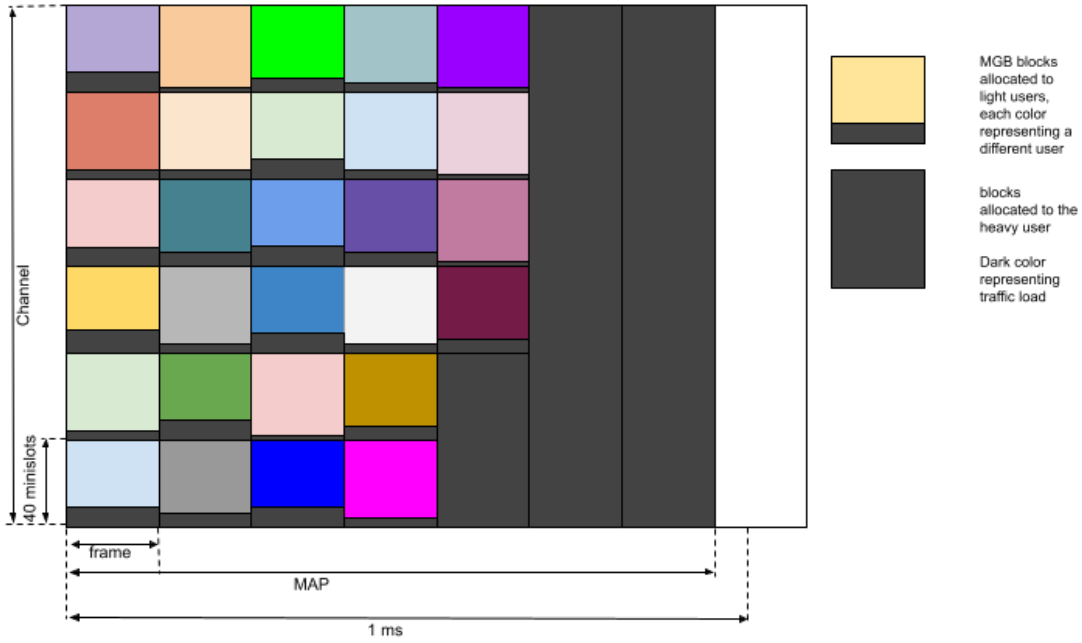


Figure 1 – Effect of MGB

The second scenario may occur when a group of upstream users are transmitting light but frequent traffic, whereas one additional user starts transmitting at a peak rate. This additional user will suffer both traffic loss and large latency, including any delay sensitive traffic therein. Even if the corresponding SF is configured with a Minimum Reserved Traffic Rate and so that portion can be treated with a higher priority, it wouldn't help. This is because the delay sensitive traffic is not distinguishable from others within the same SF, therefore there is no guarantee that it can be served with a priority.

For simplicity of illustration, we assumed in the above context that each cable modem contains one upstream service flow only, therefore the limit of total CMs in a scheduling cycle becomes the total SFs permitted. In the next section we will discuss the case with multiple service flows per cable modem.

4. D4.0 Upstream Scheduling Enhancements

In this section, we will present enhancement methods in D4.0 upstream scheduling algorithms that target at mitigating the effect of MGB. A CMTS may adopt one or more of these methods per its design choices.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.1. Filling up Extra Space

The D4.0 specification suggests that the "minimum grant bandwidth can be met through any combination of probe, ranging, OUDP testing SID, and data grant allocations across any of the Extended Upstream Channels in the CM's Extended Transmit Channel Set" ([1]). However probing signals may span the entire channel for only a subset of symbols. They can't be combined with other bursts to fulfil MGB. Besides, all of probing, ranging and OUDP testing are supposed to be allocated per CMTS's own schedules, and it serves no purpose if they're offered too often or at random. Instead, it makes sense to fill up the remaining space of a MGB block using data grant allocations. These extra grants help to reduce access latency for newly arrived packets, as the bandwidth would otherwise be wasted anyway. Per the specification, a CM is required to fill up the extra space of a grant with paddings ([1]).

4.2. Selecting OFDMA Frame Size

As mentioned above, there could only be a maximum of 18 simultaneous transmitting CMs at any time using the FDD channels. To accommodate more users, it is desirable to adopt a smaller OFDMA frame size, e.g., K = 6, which gives more granularity to grant allocations. With a 2K FFT, i.e., symbol time is 20us, with K = 6, and CP = 2.5 us, the resulted OFDMA frame time is 6 x (20 + 2.5) = 135 microseconds. Therefore, the maximum number of CMs that can be accommodated in a 1 millisecond scheduling cycle is 18 x 1000/135 = 133.2 CMs on average.

4.3. Scheduling Across Multiple Channels

Prior to D4.0, upstream SC-QAM channels and OFDMA channels are typically configured separately and operating independently. In D4.0 however, with the MGB requirements, it is desirable to concatenate all the FDD channels in order to make a more efficient usage of the extended spectrum. For example, if MGB is 40 minislots, a remaining 30 minislots on the first channel can be allocated to the next requesting CM, with an additional 10 or more minislots from a second channel being allocated to the same CM. In this way the MGB requirement is satisfied for the CM and the remaining 30 minislots on the first channel will not be wasted.

The implication is that all the FDD channels have to operate synchronously so that their symbols are aligned. Additionally, the scheduling cycle has to be identical for all of them, so that they can be scheduled as a unified resource pool. We require that these channels are configured with the same frame size K and the same Cyclic Prefix, and that the CMTS software forces the same OFDMA frame boundary across all the channels.

Note it is not required to synchronize between FDD channels and non-FDD channels.

4.4. Aggregating Service Flows

Prior to D4.0, each service flow is scheduled independently, with no regard to other SFs from the same CM. As MGB is enforced on a per CM basis, it makes sense to consider requests from all SFs of the same CM concurrently and allocate grants to multiple of them in the same frame, expecting the summation of all the grants satisfying MGB, or at least close to it. The aggregation results in a more efficient usage of channels' resources.

A question comes up as how we should handle the situation if multiple SFs of a CM have different scheduling priorities. When we aggregate these SFs, we may have ignored their priorities as compared with other SFs from different CMs. In other words, we may favor some lower priority SFs from one CM over a higher priority SF from another CM.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

One possible method is to extend the concept of Aggregate Service Flow (ASF) and hierarchical QoS so that there could be e.g. 1 – 4 ASFs per cable modem, and there are multiple individual service flows in each ASF, with each SF having its own QoS parameters. For example, if there are 2 ASFs for each CM, ASF 1 may contain all SFs used for management, signaling protocols and voice, whereas ASF 2 may contain all other SFs. In each scheduling cycle, strict priority may be used between ASF 1 and ASF 2 regardless of the CMs. The total requests of an ASF may be scheduled on either FDD channels or non-FDD channels (see the next subsection). In the former case, MGB will be satisfied for grants allocated to the current ASF, with the remaining space possibly filled up by another ASF with a lower priority and from the same CM. Among the total allocated grants for the ASF, the CMTS will satisfy individual SFs one by one using round robin, weighted round robin, or strict priority, per QoS parameters of the ASF and individual SFs.

In a way, an ASF is similar to the concept of "T-CONT" in GPON, whereas an individual SF is similar to a GEM port.

As an extreme case, each SF is an ASF, meaning that each SF is scheduled independently regardless of the CMs, thus reverting to the original upstream scheduling scenario. As another extreme case, all SFs from the same CM belong to one ASF, meaning all CMs of the same ASF type have the same relative priority, regardless of individual SFs therein. By carefully grouping SFs into ASFs, we are able to achieve a balance between prioritization and aggregation.

4.5. Skipping Small Bursts

In the past, a scheduler may prefer channels in a higher spectrum over those in a lower spectrum, i.e., it will schedule grants on a higher spectrum channel until the current map space is filled, at which point it will schedule grants on a lower spectrum channel. In D4.0 however, to further mitigate the issue with MGB, we may deliberately leave small data grants to non-FDD channels, and only schedule large data grants on FDD channels. That is, when scheduling a FDD channel, if the total bytes of all aggregated requests from a CM or an ASF are way smaller than what is required to satisfy MGB, we may skip such requests in the current channel, so they will be picked up by non-FDD channels. Similarly, when scheduling a non-FDD channel, if the total bytes of all aggregated requests from a CM or an ASF are close to or exceed MGB, we may skip such requests in the current channel, so they will be picked up by FDD channels. The threshold setting of what size to pick up or skip may depend on relative utilization levels between the FDD channels and the non-FDD channels.

Note in the context above and thereafter, a "request" may refer to either an explicit or piggyback request from a CM, or an artificial request that is inserted by the CMTS itself to satisfy the Guaranteed Grant Rate of a PGS SF.

4.6. Allocating Proactive Grants

With the mechanism introduced in subsection 4.5, we can now adopt PGS to further reduce access latency, especially that of initial bursts. As mentioned in section 2, it is sufficient and necessary to pre-allocate a small amount of bandwidth for a SF, therefore a PGS grant typically contains a small number of bytes as compared with MGB. When scheduling a FDD channel, these PGS bytes will be aggregated with any other requests from the same or different SFs belonging to the same CM or ASF, if there are any. If the total bytes are large enough, they will be scheduled on the current channel and will conform to MGB. If however there're no other requests to aggregate, or if the aggregated total bytes are too small as compared with MGB, they will be skipped on the current channel and left to non-FDD channels.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Experiments and Analysis

In this section, we use a concrete example to illustrate how MGB may impact performance of an SF. We further demonstrate by adopting some of the methods introduced in section 4, it is possible to mitigate the effect of MGB and achieve an optimal performance.

As a full implementation of D4.0 is not ready yet for either CMTS or CM vendors, we perform our experiments with simulation on a high-split system. We also design an analytical model and validate it with simulation. Using that, we may derive further results for scenarios that cannot be experimented.

Here again for simplicity, we assume that each SF belongs to a different CM or ASF, so each SF will be allocated with grants of at least MGB. Also we assume the traffic demands from all users not exceeding the overall channels' capacity if we ignore the effect of MGB.

5.1. The Test Cases

Consider a Casa CMTS system configured with 2 OFDMA channels and one ATDMA channel on an upstream port, with their specifications as follows:

- Channel 1: ATDMA, 5.8 – 12.2 MHz, 64-QAM
- Channel 2: OFDMA, 70 – 101 MHz, K = 6, 1024-QAM
- Channel 3: OFDMA, 108 – 204 MHz, K = 6, 1024-QAM
- MAP interval: 1ms

Channel 3 is used to simulate an FDD channel in a UHS-300 plant, so that the grants allocated to any SF in any frame will span at least 26 minislots in the frequency domain. Consequently, there can be at most 9 SFs allocated in an OFDMA frame, and on average, at most 66 of them can be accommodated in one scheduling cycle.

A high-split CM is attached to the upstream port, with channel bonding that spans all 3 channels. A test SF is created on this CM, with QoS parameters as follows:

- Scheduling type: PGS
- Guaranteed Grant Rate: 5 mbps
- Guaranteed Grant Interval: 1ms

This test SF has an offered traffic rate of 700 mbps and a packet size of 1024 bytes from a Spirent test equipment. We measure the throughput and latency of it using the same equipment.

Consider a group of N synthetical SFs with a PGS scheduling type, a Guaranteed Grant Rate (GGR) of 1.5 mbps, and a Guaranteed Grant Interval (GGI) of 1ms. These SFs are simulated at the scheduler by inserting data grants in each scheduling cycle that can accommodate the specified parameters. That is, if using Channel 3 each of them will be assigned with 26 minislots in each scheduling cycle, and if using Channel 2 each of them will be assigned with 4 minislots. Among the N synthetic SFs, we denote M as the number of SFs that are granted on Channel 2.

In the first test case (denoted as test case A), we increase N from 0 to 100, and allocate grants for all of the synthetic SFs on Channel 3, i.e., M = 0. We observe how performance of the test SF degrades. Then in the next test case (denoted as test case B), we fix N and gradually move the synthetic SFs from Channel 3 to Channel 2, i.e., increase M from 0 to N, and observe how performance of the test SF is recovered.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.2. Test A: Performance Degradation with MGB

In this test we increase N from 0 to 100 and allocate grants for all of the N synthetic SFs on Channel 3. As can be seen from the figure below, initially the test SF is the only one serviced by all channels, so it has 100% throughput and a minimum latency of 1.408 ms. When we increase N up to 20, some traffic of the test SF is pushed to Channel 1 and 2, but the total can still be satisfied by the remaining bandwidth, so the impact to its performance is negligible. When N continues to increase, traffic of the test SF can no longer be satisfied by the remaining bandwidth, and its performance degrades quickly. Buffer limit is reached and the latency increases as the throughput decreases. When N reaches 70 and above, traffic of the test SF is primarily serviced by Channel 1 and 2, therefore the performance metrics stay almost flat.

Note in reality, the lower channels are most likely occupied by legacy CMs, meaning the test SF will have to compete with them for resources, therefore its performance will be even worse.

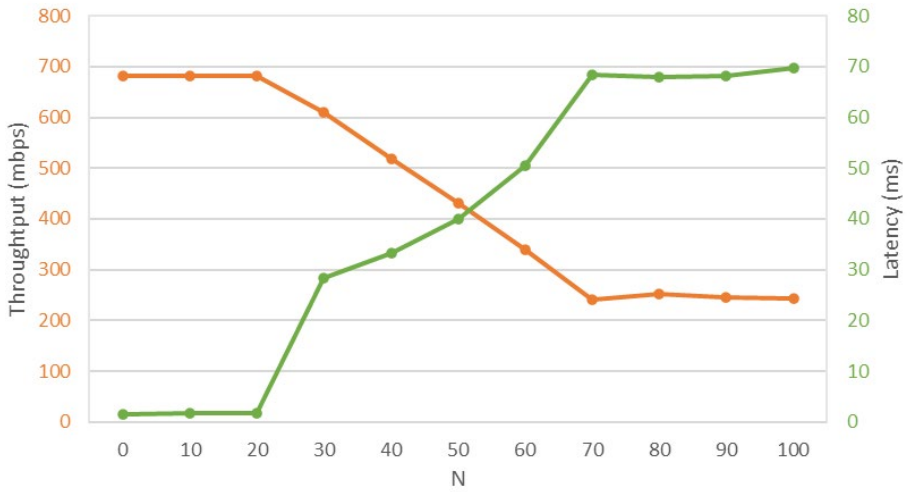


Figure 2 – Test A: Throughput and Latency vs N

5.3. Test B: Performance Enhancements

In this test we fix N as 70, meaning that the system has to accommodate 70 synthetic SFs. Now increase M from 0 to 70, meaning we gradually move these SFs from Channel 3 to Channel 2. This corresponds to the method introduced in subsection 4.5.

As we can see from the figure below, performance of the test SF starts to improve as M increases. When M reaches 60, its performance has recovered almost completely, with a full throughput and a minimum latency of 1.586 ms.

Note in reality, the lower channels are most likely occupied by the legacy CMs, meaning the M SFs will have to compete resources with them. However this is much better than the case where light users occupy large chucks of resources on FDD channels, and push heavy users to the lower channels.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

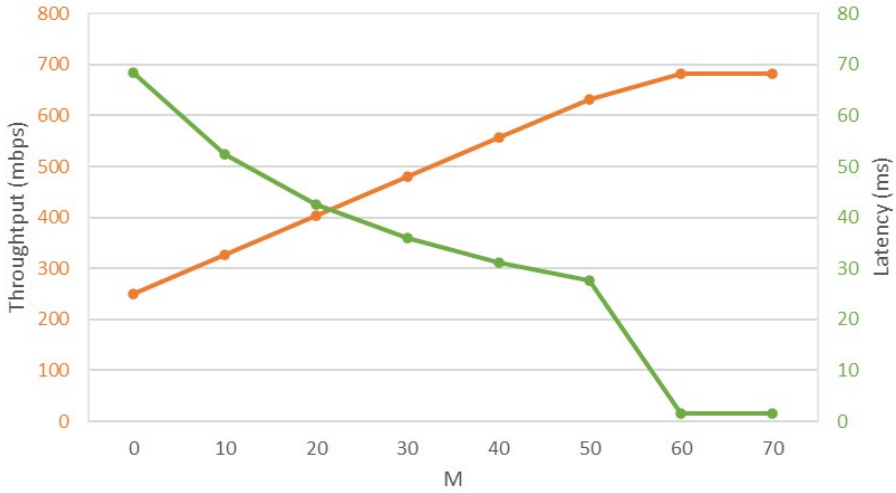


Figure 3 – Test B: Throughput and Latency vs M with N = 70

5.4. Further Analytical Results

We've designed an analytical model to calculate the performance metrics of various scenarios. Before using that, let's first validate the model with simulation results measured in Test A and Test B. Figure 4 below depicts a comparison of theoretical and simulation results for both throughput and latency of the test SF in Test A, and Figure 5 depicts the same in Test B. As we can see from the figures, the theoretical model matches well with the simulation.

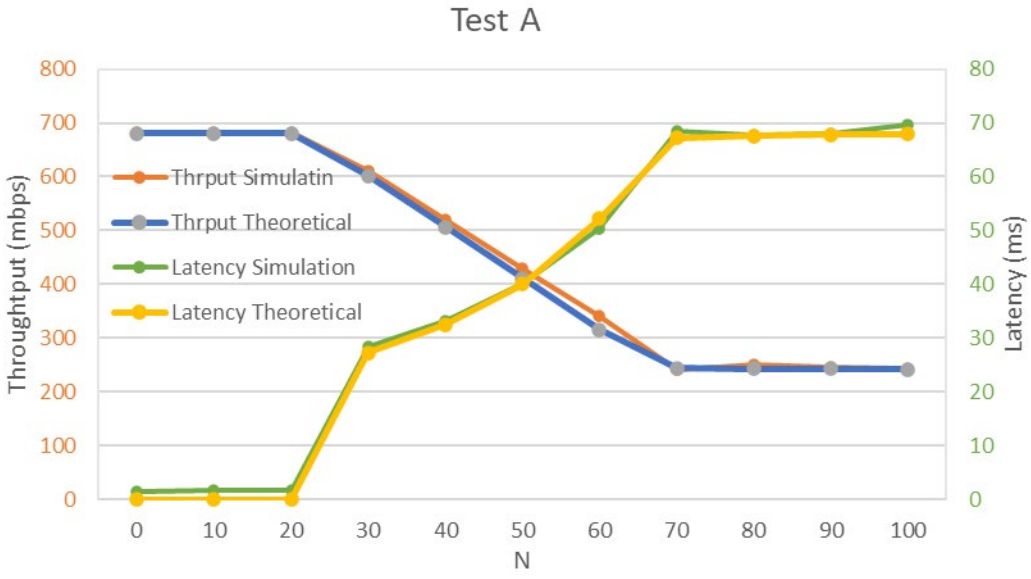


Figure 4 – Test A Theoretical vs Simulation Results

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

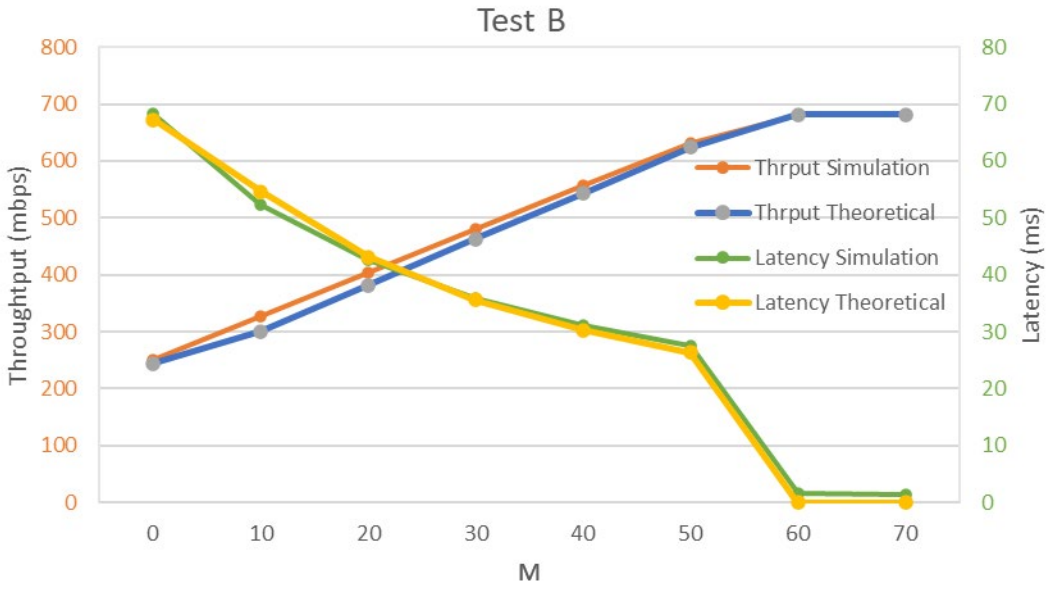


Figure 5 – Test B Theoretical vs Simulation Results

We can then use the analytical model to derive performance metrics under various conditions, as follows.

5.4.1. UHS-300

The system under study now is equipped with the following:

- Channel 1: Not present
- Channel 2: OFDMA, 8 – 101 MHz, K = 6, 1024-QAM
- Channel 3: OFDMA, 108 – 204 MHz, K = 6, 1024-QAM
- Channel 4: OFDMA, 204 – 300 MHz, K = 6, 1024-QAM

The system contains 2 FDD channels and an upstream upper edge of 300 MHz, i.e., UHS-300. N represents all light but frequent users whose traffic can be accommodated with PGS-like service with about 1.5 Mbps grant rate and 1ms grant interval. M represents how many of the N users are served using the non-FDD channel. Under this condition we evaluate the total bandwidth available to a heavy user under test. The table below enumerates the throughput of it with different N and M combinations.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 2 – Throughput vs N and M (UHS-300)

N / M	0	20	40	60	80	100	120	140	160	180	200	220	240
0	1933.9	1743.0	1552.0	1361.1	1170.1	979.2	788.3	637.6	636.4	635.5	634.8	634.2	633.7
20	0.0	1904.5	1713.6	1522.7	1331.7	1140.8	949.8	758.9	608.2	607.1	606.2	605.4	604.9
40	0.0	0.0	1875.2	1684.2	1493.3	1302.3	1111.4	920.4	729.5	578.8	577.7	576.8	576.1
60	0.0	0.0	0.0	1845.8	1654.8	1463.9	1273.0	1082.0	891.1	700.1	549.5	548.3	547.4
80	0.0	0.0	0.0	0.0	1816.4	1625.5	1434.5	1243.6	1052.6	861.7	670.8	520.1	518.9
100	0.0	0.0	0.0	0.0	0.0	1787.0	1596.1	1405.2	1214.2	1023.3	832.3	641.4	490.7
120	0.0	0.0	0.0	0.0	0.0	0.0	1757.7	1566.7	1375.8	1184.8	993.9	802.9	612.0
140	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1728.3	1537.3	1346.4	1155.5	964.5	773.6
160	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1698.9	1508.0	1317.0	1126.1	935.1
180	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1669.5	1478.6	1287.6	1096.7
200	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1640.2	1449.2	1258.3
220	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1610.8	1419.8
240	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1581.4

The MGB is 26 minislots. As shown in the table, the heavy user may achieve a peak rate of ~1.9 Gbps, but the rate quickly drops to ~630 Mbps when the number of light users N reaches 133. The lowest rate of ~630 Mbps is what the lower channel has to offer. When the light users are moved to the lower channel, the data rate is gradually recovered, up till ~1.58 Gbps.

5.4.2. UHS-396

Here we consider a system with 3 FDD channels and an upstream upper edge of 396 MHz, i.e., UHS-396, as follows:

- Channel 1: Not present
- Channel 2: OFDMA, 8 – 101 MHz, K = 6, 1024-QAM
- Channel 3: OFDMA, 108 – 204 MHz, K = 6, 1024-QAM
- Channel 4: OFDMA, 204 – 300 MHz, K = 6, 1024-QAM
- Channel 5: OFDMA, 300 – 396 MHz, K = 6, 1024-QAM

The MGB becomes 40 minislots in this case. The throughput with any combination of N and M is enumerated in the following table. Explanation is omitted here.

Table 3 – Throughput vs N and M (UHS-396)

N / M	0	20	40	60	80	100	120	140	160	180	200	220	240
0	2586.72	2292.96	1999.2	1705.44	1411.68	1117.92	824.16	642.2094	640.484	639.1399	638.0633	637.1815	636.4461
20	0	2557.344	2263.584	1969.824	1676.064	1382.304	1088.544	794.784	612.8334	611.108	609.7639	608.6873	607.8055
40	0	0	2527.968	2234.208	1940.448	1646.688	1352.928	1059.168	765.408	583.4574	581.732	580.3879	579.3113
60	0	0	0	2498.592	2204.832	1911.072	1617.312	1323.552	1029.792	736.032	554.0814	552.356	551.0119
80	0	0	0	0	2469.216	2175.456	1881.696	1587.936	1294.176	1000.416	706.656	524.7054	522.98
100	0	0	0	0	0	2439.84	2146.08	1852.32	1558.56	1264.8	971.04	677.28	495.3294
120	0	0	0	0	0	0	2410.464	2116.704	1822.944	1529.184	1235.424	941.664	647.904
140	0	0	0	0	0	0	0	2381.088	2087.328	1793.568	1499.808	1206.048	912.288
160	0	0	0	0	0	0	0	0	2351.712	2057.952	1764.192	1470.432	1176.672
180	0	0	0	0	0	0	0	0	0	2322.336	2028.576	1734.816	1441.056
200	0	0	0	0	0	0	0	0	0	0	2292.96	1999.2	1705.44
220	0	0	0	0	0	0	0	0	0	0	0	2263.584	1969.824
240	0	0	0	0	0	0	0	0	0	0	0	0	2234.208

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.4.3. UHS-492

We now consider a system with 4 FDD channels and an upstream upper edge of 492 MHz, i.e., UHS-492, as follows:

- Channel 1: Not present
- Channel 2: OFDMA, 8 – 101 MHz, K = 6, 1024-QAM
- Channel 3: OFDMA, 108 – 204 MHz, K = 6, 1024-QAM
- Channel 4: OFDMA, 204 – 300 MHz, K = 6, 1024-QAM
- Channel 5: OFDMA, 300 – 396 MHz, K = 6, 1024-QAM
- Channel 6: OFDMA, 396 – 492 MHz, K = 6, 1024-QAM

The MGB is 53 minislots in this case. The throughput with different combinations of N and M is listed as follows. Explanation is omitted.

Table 4 – Throughput vs N and M (UHS-492)

N	0	20	40	60	80	100	120	140	160	180	200	220	240
M													
0	3239.52	2850.288	2461.056	2071.824	1682.592	1293.36	904.128	646.8391	644.5386	642.7465	641.311	640.1354	639.1549
20	0	3210.144	2820.912	2431.68	2042.448	1653.216	1263.984	874.752	617.4631	615.1626	613.3705	611.935	610.7594
40	0	0	3180.768	2791.536	2402.304	2013.072	1623.84	1234.608	845.376	588.0871	585.7866	583.9945	582.559
60	0	0	0	3151.392	2762.16	2372.928	1983.696	1594.464	1205.232	816	558.7111	556.4106	554.6185
80	0	0	0	0	3122.016	2732.784	2343.552	1954.32	1565.088	1175.856	786.624	529.3351	527.0346
100	0	0	0	0	0	3092.64	2703.408	2314.176	1924.944	1535.712	1146.48	757.248	499.9591
120	0	0	0	0	0	0	3063.264	2674.032	2284.8	1895.568	1506.336	1117.104	727.872
140	0	0	0	0	0	0	0	3033.888	2644.656	2255.424	1866.192	1476.96	1087.728
160	0	0	0	0	0	0	0	0	3004.512	2615.28	2226.048	1836.816	1447.584
180	0	0	0	0	0	0	0	0	0	2975.136	2585.904	2196.672	1807.44
200	0	0	0	0	0	0	0	0	0	0	2945.76	2556.528	2167.296
220	0	0	0	0	0	0	0	0	0	0	0	2916.384	2527.152
240	0	0	0	0	0	0	0	0	0	0	0	0	2887.008

6. Conclusion

Proactive granting is an important tool that significantly reduces access latency for any best effort traffic, especially that of initial bursts. However, proactive granting in D4.0 may result in inefficient channel utilization on FDD channels due to the MGB restrictions. Our study demonstrates that by adopting various enhancement methods with upstream scheduling, it is possible to reduce the effect of MGB to the minimum, thus achieving the best utilization of channels' bandwidth. It is therefore possible to obtain an optimal latency for all traffic.

Our study also brings forth new questions, as there are other complexities in D4.0 that might require operators to reconsider their service offerings and SLAs. For example, the choice between channel utilization and prioritization becomes non-trial. Besides, an accurate realization of weighted round robin scheduling based on numerical values e.g., MSR may become difficult. These topics are left for future studies.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

ASF	Aggregate Service Flow
ATDMA	Advanced Time Division Time Access
BE	Best Effort
CM	Cable Modem
CMTS	Cable Modem Termination System
CP	Cyclic Prefix
FDD	Frequency Division Duplex
FDX	Full Duplex
GGI	Guaranteed Grant Interval
GGR	Guaranteed Grant Rate
LLD	Low Latency DOCSIS
MGB	Minimum Grant Bandwidth
MSR	Maximum Sustained traffic Rate
OFDMA	Orthogonal Frequency Division Multiple Access
OTT	Over-the-Top
PGS	Proactive Grant Service
SF	Service Flow
SLA	Service Level Agreement
UHS	Ultra-high Split
WRR	Weighted Round Robin

Bibliography & References

1. *Data-Over-Cable Service Interface Specifications DOCSIS® 4.0, MAC and Upper layer Protocols Interface Specification*, CM-SP-MULPIv4.0-I07-230503
2. *Data-Over-Cable Service Interface Specifications DOCSIS® 4.0, Physical layer Specification,* CM-SP-PHYv4.0-I06-221019
3. *H. Zhang P. Wolff, V. Sambasivan, Obtaining Low Latency in Upstream DOCSIS Transmissions,* SCTE Technical Journal, Vol 2, Num 1, Mar 2022.
4. *Data-Over-Cable Service Interface Specifications DOCSIS® 3.1, MAC and Upper layer Protocols Interface Specification*, CM-SP-MULPIv3.1-I21-201020

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23 CELEBRATING 40 YEARS

2023 Fall Technical Forum SCTE • CABLELABS • NCTA

An FPGA-based Coaxial Channel Emulator for Automated Testing and Validation

A Technical Paper prepared for SCTE by

Richard A Primerano
Principal Engineer
Comcast
richard_primerano@comcast.com

Rick Spanbauer
Executive Director, HFC Architecture
Comcast
richard_spanbauer@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 3
   1.1. RF Impairements in the Coaxial Plant 3
2. Modeling Channel Impairments 4
   2.1. Modeling Impairments in the Lab with Coax 4
   2.2. Digital Emulation of RF Impairments 5
      2.2.1. Emulating Channel Reflections 6
      2.2.2. Emulating Channel Tilt 8
3. Other Applications of the Channel Emulator Hardware 8
   3.1. Network Analysis Mode 8
   3.2. Wideband Signal Generation 9
4. Conclusion 10

Abbreviations 11

Bibliography & References 11

List of Figures

Title **Page Number**

Figure 1 – Some sources of reflections in the HFC plant 3
Figure 2 – Creating echoes in the lab with coaxial cable. 4
Figure 3 - Xilinx ZCU-111 evaluation board [7] 5
Figure 4 - Components of the DOCSIS channel emulator 6
Figure 5 - Channel emulator web-based user interface 6
Figure 6 - A single reflection in the channel emulator (top) and coaxial equivalent (bottom). 7
Figure 7 - VSA output in response to digital channel emulator 7
Figure 8 - Enabling tilt (left) and tilt-plus-reflection (right) in the emulator 8
Figure 9 - Network analyzer S_{21} sweep of channel emulator 9
Figure 10 - User interface for wideband signal generation 9
Figure 11 - 800 MHz IBW signal generation using the ZCU-111 10

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

RF signals propagating through the coaxial plant experience numerous impairments. These include natural attenuation, micro-reflections, ingress, and distortions. These impairments can be counteracted by the predistortion, equalization, and echo-cancellation components built into Data-Over-Cable Service Interface Specifications (DOCSIS®) devices. To assist in validating the performance of these DOCSIS devices, it is necessary to subject them to channels with varying levels of impairments. Today, impaired channels are created by building physical mockups of the hybrid fiber/coax (HFC) plant with cable, splitters, and attenuators. To date, changing micro-reflection levels, for example, requires physically swapping cables and attenuators. This physical approach to channel emulation can be impractical and time consuming when needing to test many scenarios.

In this paper, we present a *Field-Programmable-Gate-Array-based Coaxial Channel Emulator* that allows engineers to emulate reflections, channel tilt, ingress, and distortions in an automated testing framework. We show network analyzer sweeps of an actual coaxial channel side-by-side with the emulator, demonstrating close agreement between the two. This approach to testing has several benefits over traditional techniques. Since it is all-digital, tests are fully repeatable, and the channel can be reconfigured in software.

The emulator is capable of modeling channels with hundreds of meters of cable, tens of taps, and arbitrary tilt, ingress, and distortion. Replicating this with hardware would require a very large, expensive setup. The emulator provides 800 MHz of instantaneous bandwidth and has both a web-based user interface and a software application programming interface (API) allowing engineers to vary channel characteristics in real time. The system fits in a 2U rack enclosure, providing a compact alternative to cable-based emulators.

1.1. RF Impairements in the Coaxial Plant

Figure 1 shows a portion of the HFC plant between the remote PHY device (RPD) and the customer. As the signal travels, it becomes distorted by impairments in the channel. Some common channel impairments are discussed below.

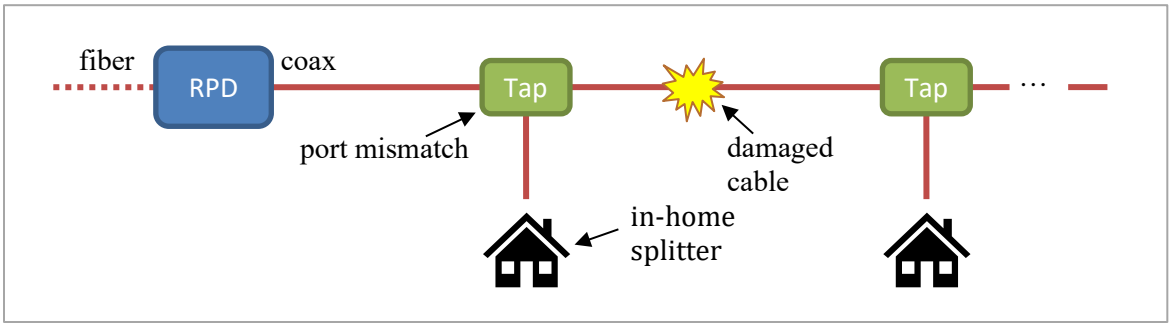


Figure 1 – Some sources of reflections in the HFC plant

Discrete impedance mismatch – Every interface between the coaxial cable and active/passive devices in the path causes an impedance mismatch. A cable properly terminated onto a tap may exhibit very low mismatch while a corroded ground block may exhibit a large mismatch. In each case, a reflection is generated, but in properly terminated connections, small reflections do not significantly impact system performance. Because these mismatches are electrically short (significantly less than one wavelength), they can be considered as discrete discontinuities along the cable. Sharp bends, cracks and nicks in cabling or

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

damaged connectors also create this type of localized discontinuity. Impedance mismatches can lead to strong frequency-selective behavior in the channel.

Distributed impedance mismatch – Cable damage and improperly installed connectors can allow water to enter the cable. This can cause an impedance mismatch in the cable spanning tens of feet, significantly longer than the wavelength of the RF signals traveling on the cable. This type of damage leads to distributed mismatch, which has different characteristics than the discrete mismatches discussed above. A detailed discussion of the causes and effects of water-soaked cable is found in [3].

Nonlinearities – Many devices in the network can contribute nonlinearity to RF signals. The most obvious source are distribution amplifiers. These nonlinearities have been studied extensively [5]. Additionally, corroded connectors creating Schottky diode-like contacts can also be a source of nonlinearity [6]. Unlike linear forms of distortion, nonlinearities lead to the generation of harmonics outside the band of the original signal, often causing interference on other frequencies.

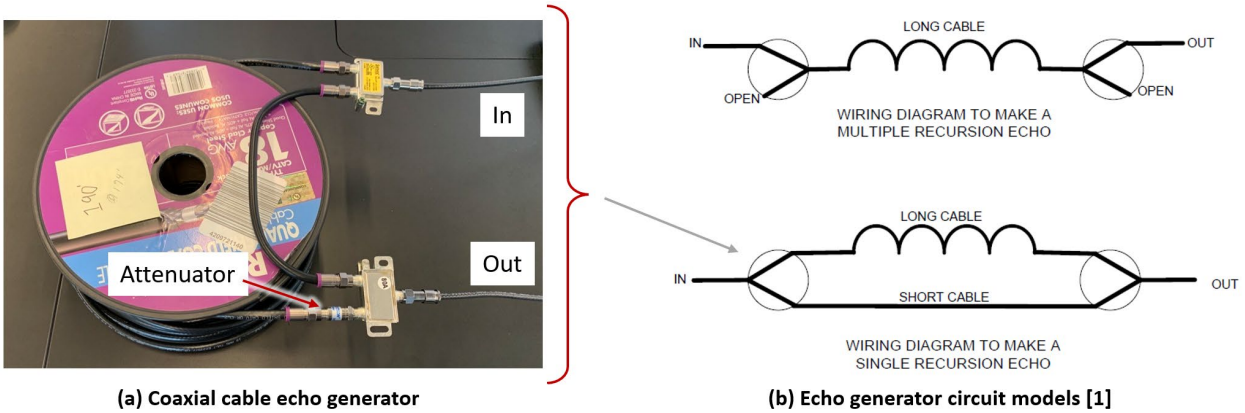
Ingress – The cable network's RF spectrum overlaps in frequency with numerous over-the-air (OTA) Federal Communications Commission (FCC) regulated bands. Cable damage and improperly installed connectors provide a path for these OTA signals to enter the network, degrading RF performance. Comcast has done extensive analysis of RF ingress power levels in different geographic areas with details presented in [4].

2. Modeling Channel Impairments

The RF transmit and receive electronics on each end of the DOCSIS channel have signal processing capabilities to correct for most channel impairments (this is also true for non-DOCSIS digital signals). For example, non-flat channel response due to reflections and tilt is counteracted by predistortion (on the transmit side) and equalization (on the receive side). Full duplex (FDX) DOCSIS nodes and smart amplifiers also have echo-cancellation and equalization to manage impairments due to reflections.

2.1. Modeling Impairments in the Lab with Coax

To test the performance of the DOCSIS channel over a range of operating conditions, channel impairments can be simulated in the lab. Several methods of simulating echoes in the lab are presented in [1]. These involve connecting lengths of cable with splitters, attenuators, and mismatches. Figure 2 shows a coaxial cable-based channel emulator that produces "single recursion echoes" and the circuit diagrams for two common coaxial emulators. The emulator in subfigure (a) is modeled by the bottom circuit in subfigure (b).



(a) Coaxial cable echo generator

(b) Echo generator circuit models [1]

Figure 2 – Creating echoes in the lab with coaxial cable.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In the lab, a coaxial channel emulator would be physically modified by changing the length of cables or the value of attenuators to vary null spacing and depth. This process allows engineers to stress channel equalization capabilities and ensure that the DOCSIS hardware operates reliably over the expected range of channel impairments. While this approach is simple and effective when modeling simple reflections, it suffers from several drawbacks. First, the engineer must physically change components for each impairment level they wish to test. This is a relatively slow process and makes automated testing impossible. Second, cable spools are large and heavy, so emulating a full node leg would require a large, complex setup. Finally, since the emulator is composed of physical components, it is difficult to ensure repeatable testing over time and across test sites. Similarly, non-linear impairments such as a corrosion-caused diode effect have been simulated with cable and circuit boards, or actual amplifiers with bias changes. This approach is awkward and it is often difficult to tune the hardware to simulate real network conditions.

2.2. Digital Emulation of RF Impairments

Many challenges faced with cable-based emulators can be addressed with digital techniques. With current field-programmable gate array (FPGA) technology, it is possible to build a wide bandwidth, low latency all-digital channel emulator to model complex DOCSIS networks. In this project, we have built a prototype digital emulator around the Xilinx ZCU-111 evaluation board. The board (shown in Figure 3) contains a Xilinx XCZU28DR FPGA. This ASIC has eight analog-to-digital (ADC) and eight digital-to-analog (DAC) channels, running in excess of 4 Gsps. The chip also contains quad-core and dual-core ARM processors.

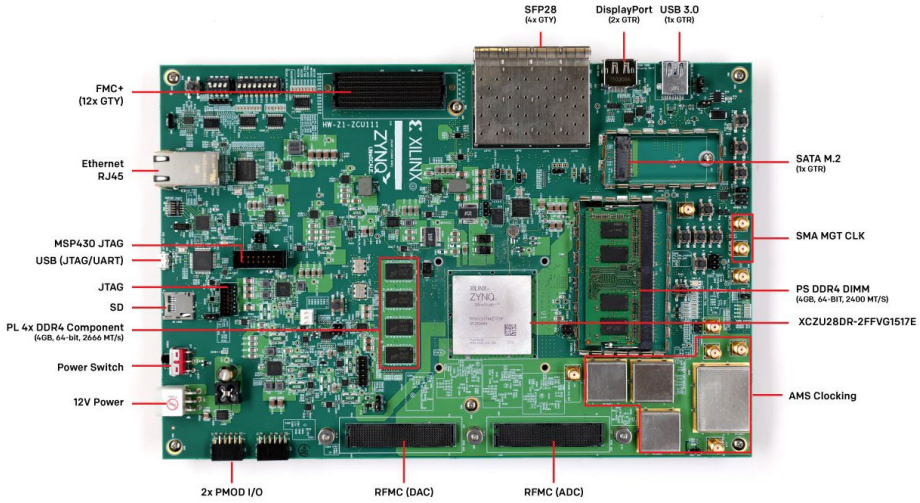


Figure 3 - Xilinx ZCU-111 evaluation board [7]

The digital channel emulator block diagram is shown in Figure 4. DOCSIS signals are digitized with the built-in ADC and converted back to RF with the DAC. The design allows the user to dynamically change the reflection characteristics of the channel and add tilt. A waveform memory allows the injection of arbitrary RF signals into model ingress. Finally, receive-side and transmit-side IQ capture modules provide real-time diagnostic capabilities.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

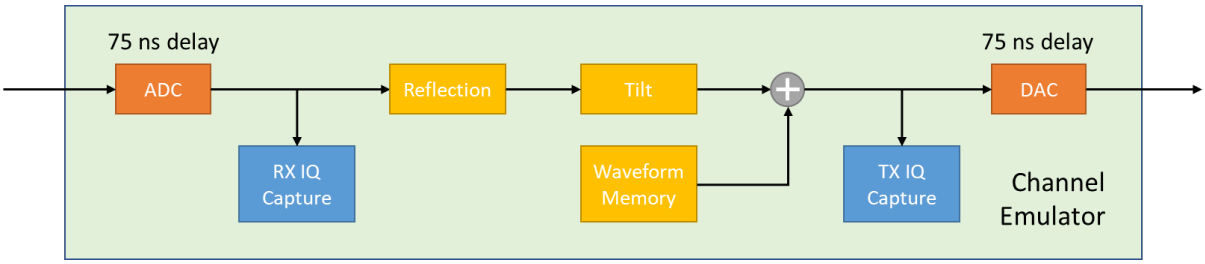


Figure 4 - Components of the DOCSIS channel emulator

The FPGA's quad-core ARM processor runs Linux and hosts a web-based user interface (UI) that gives operators full control over all hardware settings in real time. The UI, shown in Figure 5, allows the operator to set channel tilt in decibels per gigahertz (dB/GHz). In the current hardware build, there are six reflection generator modules. Each is controlled by setting the distance (i.e., the length of cable) between reflections and the mismatch caused by the reflection.

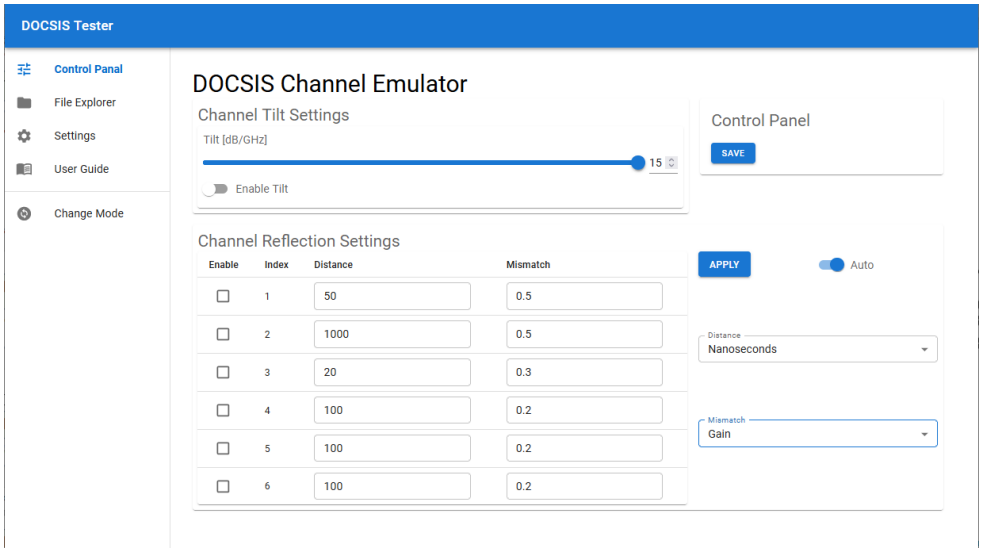


Figure 5 - Channel emulator web-based user interface

2.2.1. Emulating Channel Reflections

The six reflection modules are all identical in construction. To demonstrate their operation, we show the effect of a single digital reflection module. In the top of Figure 6, we show the hardware setup. A diagnostic modem is configured to transmit 16-QAM data continuously while a Fieldfox spectrum analyzer running Keysight 89600 vector signal analyzer (VSA) software is set to recover the data.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

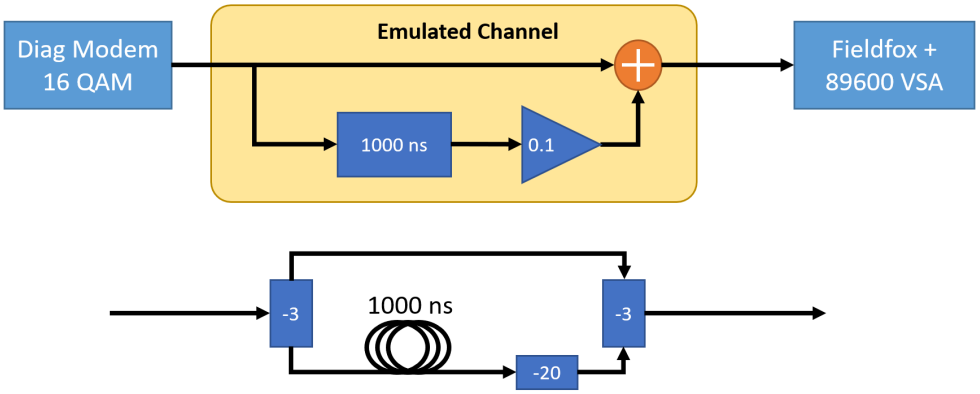


Figure 6 - A single reflection in the channel emulator (top) and coaxial equivalent (bottom).

Between the modem and VSA, we place the channel emulator with a single reflection enabled. The reflection works by splitting the incoming IQ signal, sending one path through a delay and attenuator, and summing the two paths. This is the digital equivalent of the *single recursion echo* in Figure 2. The bottom of Figure 6 shows how this echo would be built using coaxial cable. The delay used in this test was 1000 nanoseconds and the amplitude mismatch of the delayed path was -20 dB.

The Keysight 89600 VSA software has built-in channel equalization capability similar to that of the equalizer present in the DOCSIS 3.1 SC-QAM receiver channels. In Figure 7, we show the VSA output in response to the digital channel emulator. The top-left pane shows the 16-QAM constellation from the diagnostic modem and the bottom-left shows the spectrum of the signal. The top-right pane shows the VSA equalizer coefficients after convergence and the bottom-right shows the estimated channel frequency response.

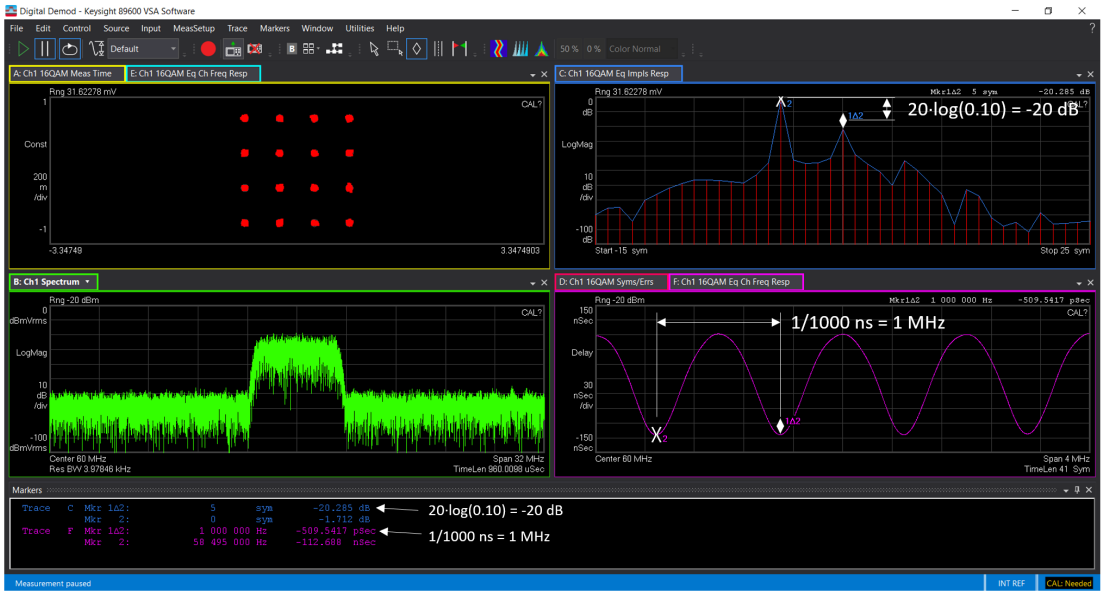


Figure 7 - VSA output in response to digital channel emulator

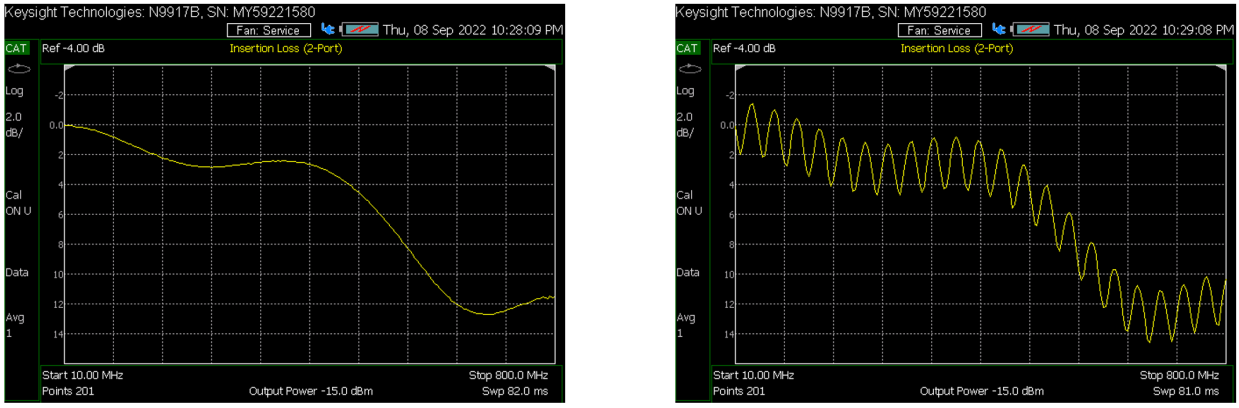
With the digital channel emulator echo path delay set to 1000 nanoseconds, we expect to see a 1 MHz ripple in the channel response. This can be verified in the bottom-right plot. Also, with a -20 dB echo level, we

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

expect to see a 20 dB delta between the equalizer's main tap and first delay. Again, this can be verified in the top-right plot.

2.2.2. Emulating Channel Tilt

In the coaxial plant, frequency-selective attenuation leads to tilt. The channel emulator models tilt using an 11-tap programmable finite impulse response (FIR) filter. By changing the filter coefficients in real time, emulated channel tilt can also be changed. Since tilt is controlled by FIR coefficients, we are free to specify arbitrary channel frequency response curves, not just the typical constant-slope response we associate with most tilt models. Figure 8 shows network analyzer sweeps of the channel emulator with a non-linear tilt profile enabled (left) and with tilt plus reflection (right).



Tilt Enabled **Tilt and Reflection Enabled**

Figure 8 - Enabling tilt (left) and tilt-plus-reflection (right) in the emulator

3. Other Applications of the Channel Emulator Hardware

The channel emulator is built on a general-purpose FPGA evaluation board. As such, the hardware can be designed to perform a variety of RF test functions. Two functions discussed here are *network analysis* and *wideband signal generation*.

3.1. Network Analysis Mode

To perform a network analyzer S_{21} measurement,[1] a tone is transmitted into the device-under-test (DUT) and the magnitude/phase of that tone reflected from DUT output port is measured. This process is repeated at many closely spaced frequencies to gather the complete measurement.

We have designed a bitstream for the FPGA that performs tone sweep and magnitude/phase measurement completely in fabric, allowing the emulator to operate as a network analyzer. Figure 9 shows a comparison of a Fieldfox spectrum analyzer sweep with the FPGA hardware. Each instrument was through-calibrated before testing. The right-hand figure shows agreement between the two devices of about 0.5 dB.

[1] S_{21} is what is known as a scattering parameter (S-parameter) for a two-port device or network, specifically forward voltage gain or forward transmission coefficient ($E_{Transmitted}/E_{Incident}$). Testing a two-port device or network from Port 1 to Port 2 gives the S-parameter S_{21}.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

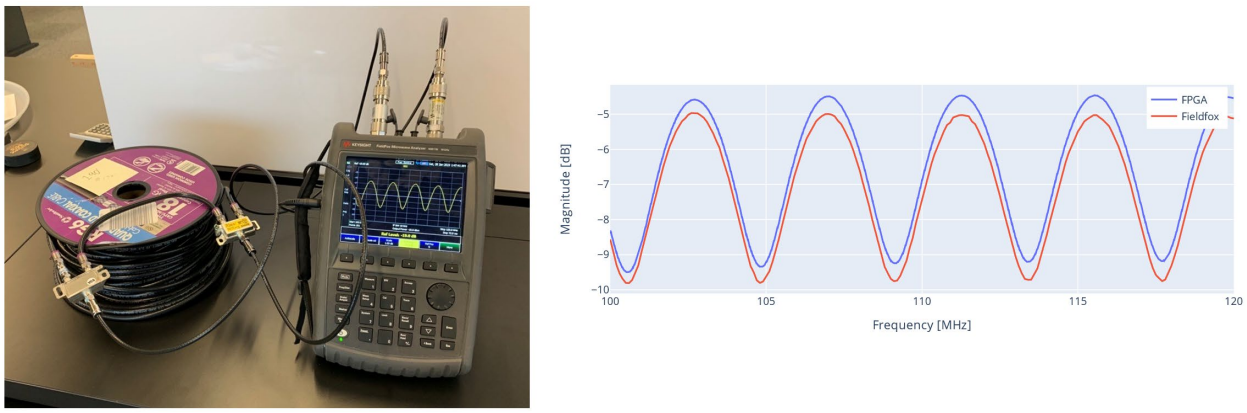


Figure 9 - Network analyzer S_{21} sweep of channel emulator

3.2. Wideband Signal Generation

In DOCSIS testing, it is often necessary to generate wideband signals simulating a fully-loaded spectrum. The channel emulator contains a waveform memory designed for ingress emulation, but this same capability can also be used for wideband channel loading. A mode has been added to the emulator's user interface that allows operators to specify wideband signals in terms of frequency ranges and period/duty cycle per range. The UI is shown in Figure 10.

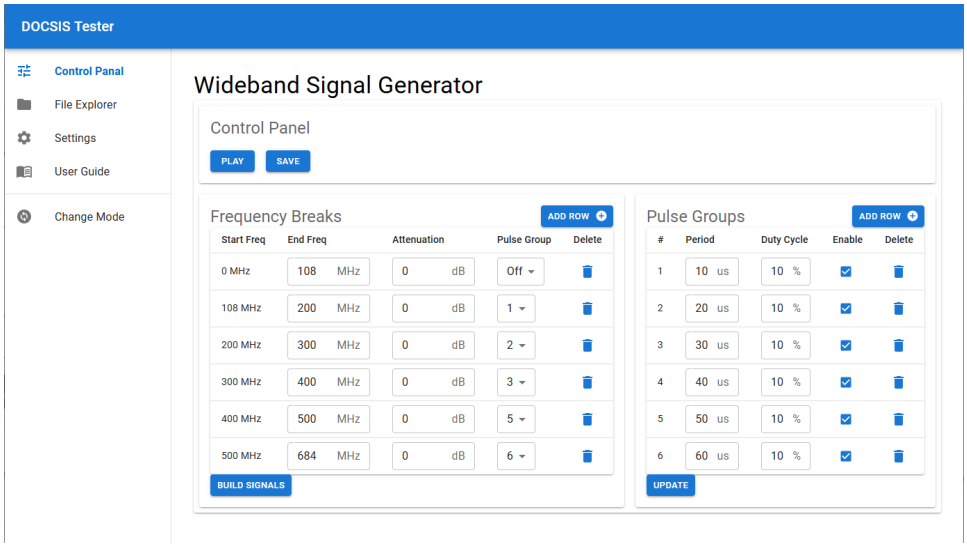


Figure 10 - User interface for wideband signal generation

Figure 11 shows a screen shot of the device producing a signal spanning 800 MHz.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

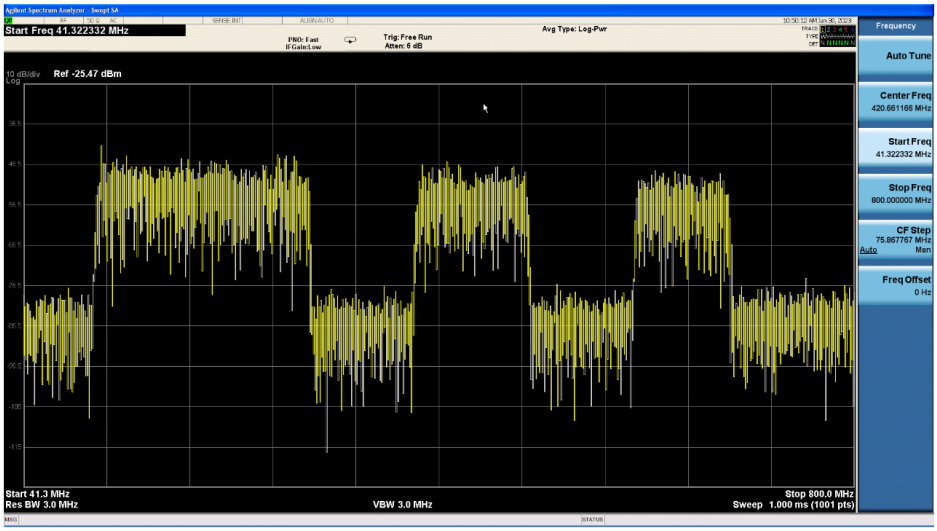


Figure 11 - 800 MHz IBW signal generation using the ZCU-111

4. Conclusion

In this paper, we have shown how commercially available FPGA hardware can be used for DOCSIS testing. The channel emulator allows users to dynamically create reflections, tilt, and ingress. Since this is an all-digital approach, repeatability can be achieved across test sites and testing can be fully automated. Additionally, the hardware footprint is significantly smaller than the commonly used coaxial emulators. Other common testing capabilities, including network analysis and signal generation can be incorporated into the same hardware platform, leading to a compact, low cost DOCSIS analysis tool.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

ADC	analog-to-digital converter
API	application programming interface
ASIC	application specific integrated circuit
DAC	digital-to-analog converter
dB	decibel
dB/GHz	decibels per gigahertz
DOCSIS	Data-Over-Cable Service Interface Specifications
DUT	device under test
FCC	Federal Communications Commission
FDX	full duplex [DOCSIS]
FIR	finite impulse response
FPGA	field programmable gate array
Gsps	gigasamples per second
HFC	hybrid fiber/coax
IQ	in-phase/quadrature
MHz	megahertz
ns	nanoseconds
OTA	over-the-air
QAM	quadrature amplitude modulation
RF	radio frequency
RPD	remote PHY device
SC-QAM	single carrier quadrature amplitude modulation
UI	user interface
VSA	vector signal analyzer

Bibliography & References

[1] PNM Best Practices: HFC Networks (DOCSIS 3.0), CM-GL-PNMP-V03-160725; Cable Television Laboratories, Inc. 2010-2016
[2] When Physical Layer Simulation Gets Real, Next-Gen Network Modeling; R. Narayanaswamy et al., SCTE Fall Technical Forum, October 2021
[3] Water Can Run, But It Can't Hide, PNM Finds Soaked Cables; K. Fox et al., SCTE Fall Technical Forum, October 2021
[4] *Deploying PMA-Enabled OFDMA in Mid-Split and High-Split*; M. Harb et al., SCTE Fall Technical Forum, October 2022
[5] Machine Learning Techniques for Equalizing Nonlinear Distortion; R. Thompson et al., SCTE Fall Technical Forum, October 2020
[6] Detection of Passive Intermodulation in Drop Wiring by Burst Transmission Analysis; T. Williams et al., SCTE Fall Technical Forum, October 2022
[7] *ZCU111 Evaluation Board User Guide*; Xilinx, UG1271 (v1.2) October 2, 2018

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

An Innovative Data Driven Approach to Optimize WiFi Performance

A Technical Paper prepared for SCTE by

Maryam Miri
Device Technology Specialist
Rogers Communications Inc
8200 Dixie Rd, Brampton, ON L6T0C1
647-203-2375
maryam.miri@rci.rogers.com

Kevin Esdale
Sr. Mgr, Data Prod & Automation
Rogers Communications Inc
8200 Dixie Rd, Brampton, ON L6T0C1
647-203-6060
kevin.esdale@rci.rogers.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Background ... 3
   - 2.1. Data Driven Proactive Schemes ... 3
   - 2.2. In-home WiFi Performance Optimization ... 4
3. System Model Architecture ... 6
   - 3.1. Video Quality Degradation Root Cause Analysis ... 6
   - 3.2. Decision Tree Development ... 9
   - 3.3. Clean Home Development ... 10
4. Results and discussion ... 11
5. Conclusion ... 12

Abbreviations ... 13

Bibliography & References ... 13

List of Figures

Title **Page Number**

Figure 1 - WiFi Network Topology ... 4
Figure 2 – Video Quality Degradation by Interference Injection ... 6
Figure 3 – Number of Hourly Fingerprints versus Achievable Bit Rate (2.4GHz of XB6s) ... 7
Figure 4 – Response Time versus Achievable Bit Rate (2.4GHz of XB6s) ... 7
Figure 5 – Achievable Bit Rate versus RSSI (2.4GHz of XB6s) ... 8
Figure 6 – Achievable Bit Rate versus RSSI (5GHz of XB7s) ... 8
Figure 7 – Achievable Bit Rate versus Channel Utilization (2.4GHz of XB7s) ... 9
Figure 8 – Example Branches of Developed Decision Tree (AP=1 depicts that the STB in question never changed its connection) ... 9
Figure 9 – Cumulative Improvement of Clean-Home Lists ... 11
Figure 10 – Distribution of Improvement Percentage ... 11
Figure 11 – Total Video Quality Degradation Events ... 12

List of Tables

Title **Page Number**

Table 1 – Video Quality Degradation Events ... 5
Table 2 – Specific Filters to Avoid External Effects ... 6
Table 3 – Clean-Home KPIs ... 10

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Applying a proactive communication approach with the focus of customers with technical issues is an effective business strategy and a proven method that enhances both customer satisfaction as well as brand reputation. In a proactive program, the organizations initiate communication with customers and by keeping customers informed of feasible technical solutions, they stay ahead of possible complaints.

Leveraging data driven approaches in a proactive communication program can significantly elevate the success rate. By utilizing data analytics tools, different insights, patterns, and correlations can be extracted. This helps in identifying the underlying issues and their general causations and hence, a more goal-oriented communication effort. In this paper, we describe the implementation of an end-to-end solution that by leveraging the analysis on millions of recorded data points, improves the wireless fidelity (WiFi) experience of users.

In other words, by identifying and addressing the underlying in-home network issues that cause video quality degradations, we offer solutions to enhance customers' experience. To achieve this goal, we analyzed reported data from various sources such as set-top boxes (STBs), gateways (GWs), point of deliveries (PODs), and user feedback to identify patterns or trends that indicate issues affecting video quality. Then, by correlating the video quality impairments to the WiFi telemetries such as throughput, response time, and latency we isolate the WiFi degradation root causes such as poor received signal strength index (RSSI), low signal to noise ratio (SNR), and high band utilizations. Finally, by defining the root causes' thresholds and developing a decision tree, we confidently propose educated solutions based on the root causes. By proactively reaching out to more than 5000 customers so far and employing the solutions, the results illustrate a notable improvement in the video quality. The observations of different cohorts depict that the pre-post video quality comparisons sustained greater than 70% improvement post implementations.

2. Background

WiFi has introduced a fundamental shift in human lifestyle by becoming the preferred medium of our everyday communications. Over the last 5 years, there has been a significant increase in the total number of connected devices in a household. Dominated in the internet of things (IoT) space, a huge growth in the total number of devices is expected over the coming years. This has increased the demand for connectivity and convenience in various aspects of our lives and the need for devices to be interconnected and able to communicate with each other. Therefore, having access to a reliable WiFi network with a robust infrastructure to support the growing number of devices is a crucial need today. It is worth noting that the reliability of WiFi depends on the performance of both access network and the in-home network.

2.1. Data Driven Proactive Schemes

Data mining has become one of the major methods in uncovering hidden insights and trends within large volumes of data in recent years. Identifying patterns, relationships, and insights from large volumes of data allows organizations to develop optimal strategic planning and marketing campaigns. Analyzing operational data enables organizations to identify potential bottlenecks and areas for improvement. This can lead to enhanced cost efficiency, customer satisfaction, and overall stronger customer relationships. Utilizing data to detect patterns and possible impairments in access network has been greatly discussed in [1]-[4-10]. In [1], using an unsupervised model, the authors analyze RF impairments observable by proactive network maintenance (PNM) and combine it with CableLab's spectral impairment detector (SID) observations [2]-[3] to improve SID's impairment classifiers. By transforming the state-of-the-art object detection algorithm into an anomaly detection model in [4], a flexible and high-performance anomaly detection solution is proposed for cable industry. The authors in [4] reach the mean average precision (mAP)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

of 97.82% in localizing and classifying network anomalies via developing a 1D-Convolutional Neural Network that includes 45 1D-Convolutional layers. In [5], a detection model is proposed that accurately differentiates among RF wave types and successfully detects the network impairment signatures. Also, by narrowing down the origin of the impairments through a network topology graph representation, the problem resolving mechanism has been further improved.

2.2. In-home WiFi Performance Optimization

Inspired by the above discussions and considering the major effect of WiFi performance on various aspects of modern life, in this paper, we analyze the reported data from controlled devices such as STB, POD, and GW to identify the factors that impact the in-home WiFi performance negatively and determine potential solutions. Figure 1 displays the topology of a WiFi network in a home environment with a streaming device that uses WiFi as the primary connection. As this figure depicts, depending on parameters such as the strength of the received signal from the APs, STB can connect to either GW or the POD. The received power by the receiver r from the transmitter t can be approximated by

$$P_r = P_t G_t G_r (\frac{\lambda}{4\pi d})^2,$$

in which a free space is assumed, and d, P_t, G_t, and G_r represent the distance between the transmitter and the receiver, the transmitted power, the transmitter antenna directivity gain, and the receiver antenna directivity gain, respectively. Also, according to the Shannon capacity theorem, the maximum achievable data transmission rate on a noisy channel is limited by the channel bandwidth and the signal to interference plus noise ratio (SINR) at the receiver [11]. This sheds light on the importance of a proper AP connection as well as frequency band selection.

In other words, the quality of the connection plays a crucial role in determining the selected modulation, error correction techniques and ultimately, the data transmission rates. It is worth noting that the total raw bit rate of a video that needs to be transmitted through IPTV depends on the video resolution as well as multiple other attributes. I.e., for a video stream with the resolution of $n_x \times n_y$pixels per frame, the frame rate of n_f, and the total n_a stereo channels of audio, the bit rate can be calculated as

$$R_{b_total} = 3n_b \times n_x \times n_y \times n_f \times r_{dc} \times r_{cc} + n_a \times R_a,$$

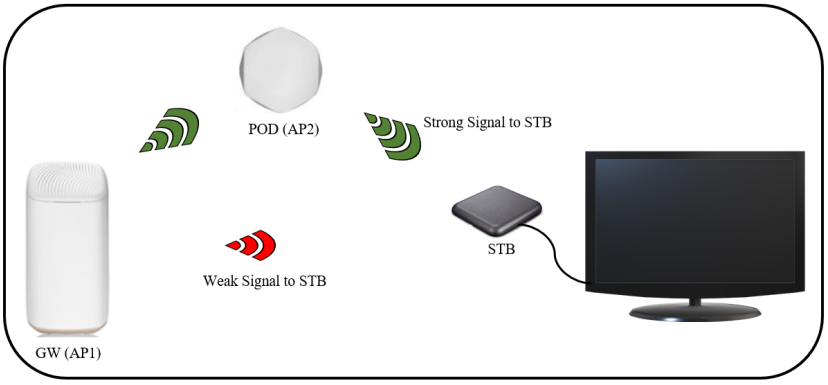


Figure 1 - WiFi Network Topology

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

where n_b is the number of quantization bits per pixel, r_{dc} and r_{cc} show the data compression ratio and colour compression ratio, and R_a represents the audio transmission rate (comprised of audio sampling rate and the number of quantization bits per sample).

In the case of a WiFi connection degradation where a modulation with low transmission rate is selected, the channel profile usually changes to a lower resolution, i.e., from 18.7Mbps to a lower data rate such as 6Mbps for a 4k video. Hence, if the WiFi connection to the STB is in the poor stages, it is impossible to stream a 4k resolution video with a bit rate of 18.7Mbps without observing defects such as automatic bit rate shifting (ABR), black screen, frame freezing and skipping, and pixelation. Furthermore, poor in-home coverage may result in other disturbances such as the video player aborting the manifest file that makes it impossible to recover the video content resulting in onscreen errors.

By tracking the reported profile changes as well as other video events and correlating them to the recorded in-home telemetries such as SNR and WiFi bandwidth, we can detect the video quality degradation events and offer solutions based on the root causes. Table 1 illustrates the events that we track in our data as the video quality degradation fingerprints and a brief description regarding each one of those events.

To verify the accuracy of these fingerprints, we conducted an SINR test experiment with 5 different scenarios by changing the STB AP as well as the location of the interference generator and monitored the reported events and telemetries by GW, STB, and POD. In each one of those scenarios, we ingested a partial interference with the bandwidth of 20MHz into an 80MHz WiFi channel, while streaming a 4k video on IPTV to observe the effects on the perceived quality of the video and the recorded errors. For further analysis and comparisons, two laptops, one streaming a YouTube channel and the other one connected to a Microsoft teams video call were also left in close range of the STB. As Figure 2 shows, once the experiment began the video frames on TV became frozen, the lowest profile alerted, and the manifest failed messages appeared in reported logs by the STB. Also, degradation was observed on both laptops; the YouTube channel streaming stopped, and the teams video call dropped or disrupted during all the scenarios. All in all, our SINR test scenarios not only verified the accuracy of aforementioned fingerprints reported by the STB, but also confirmed these markers can generally represent a degraded experience of other devices that have the same WiFi connection/proximity as the STB.

Table 1 – Video Quality Degradation Events

Channel Resolution	Tracked Event	Description
4k resolution video	Fog_Warn_Lowest_Profile_set	The bit rate shifts to 6.3Mbps
	Fog_Info_LowProfile	The bit rate shifts to a lower rate such as 12.8Mbps.
UHD resolution video	Fog_Warn_Lowest_Profile_set	The bit rate shifts to 3.8Mbps
	Fog_Info_LowProfile	The bit rate shifts to a lower rate such as 4.6Mbps.
4k and UHD resolution video	Fog_Warn_Mfstdwnld_Failed	Manifest file failed to download or Aborted
	Fog_ERR_NoUpdatePlaylist	No Update for the last playlist file
	Fog_Info_CDNTimeout	Timeout occurred reaching the URL for the next video file

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

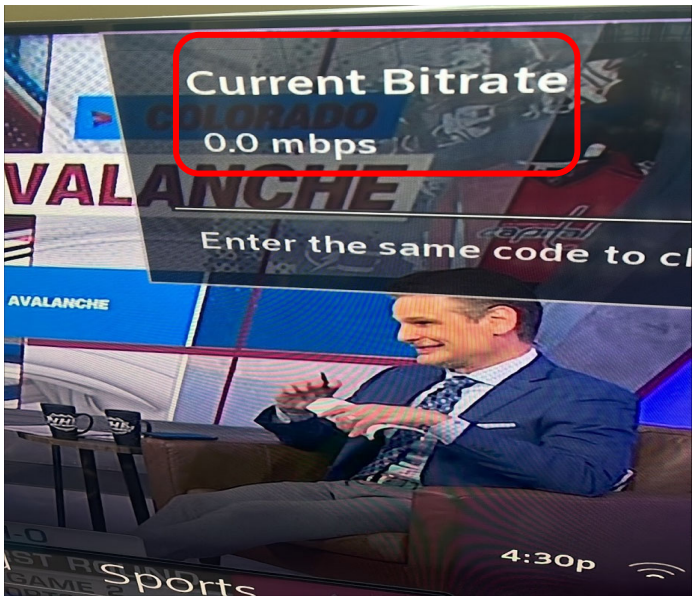


Figure 2 – Video Quality Degradation by Interference Injection

3. System Model Architecture

This section begins with a discussion on the root cause analysis conducted in our model, employing a data-driven approach. Next, we present the development of a decision tree that provides diverse solutions derived from the root cause analysis. Finally, we elaborate an illustrative use case of this study, demonstrating the practical implementation of an algorithm called clean-home.

3.1. Video Quality Degradation Root Cause Analysis

To perform the root cause analysis for customers encountering the fingerprint issues mentioned above, we employ the hourly collected values reported by GW, POD, and STB. In order to eliminate any video quality degradations resulting from factors outside the home environment, we exclude all degraded RF network telemetry data that could contribute to poor video experience. Put differently, we employ the filters outlined in Table 2 on hourly data of customers with degraded video quality to define appropriate thresholds for various WiFi parameters, such as RSSI and channel utilization.

Table 2 – Specific Filters to Avoid External Effects

Hourly Parameter	Value
# of GW/STB reboots	0
# of GW re-registrations	0
STB connection type	WiFi
# of WiFi crashes and WiFi disconnects	0
# of upStream/downStream docsis errors	0
ODM_score	<200
Node_score	<25
Area Congestion (upstream_score+downstream_score>1000)	No

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Also, considering the varying performances of different GW models (XB6, XB7, and XB8), we categorize the hourly data points that satisfy all the filters mentioned above based on the respective GW models. Due to the distinct channel bandwidth and supported standards of different operational WiFi bands (such as 2G, 5G, and 6G bands), it is necessary to further classify the data based on the STB connection band as well. Subsequently, to identify the root causes behind video quality degradations and establish degradation thresholds for parameters such as RSSI and channel utilization, we utilize diverse correlation graphs corresponding to each classified data group. Figure 3 shows the number of hourly fingerprints versus the reported achievable rate for STBs connected to 2.4GHz bandwidth of an XB6 GW. Figure 4 displays response time versus the achievable bit rate for the same STB connection. As depicted in Figure 3, for achievable bit rates of 150 Mb/s and higher, it is evident that not only are there no instances of more than 4 fingerprints, but the occurrences of fingerprints also become infrequent and scattered. Therefore, the 150 Mb/s bit rate can be regarded as one of the thresholds that determines the quality of IPTV signals for STBs connected to the 2.4 GHz band of an XB6. The noticeable deterioration of response time in Figure 4 when the bit rates are below 150 Mb/s further validates the reliability of this threshold.

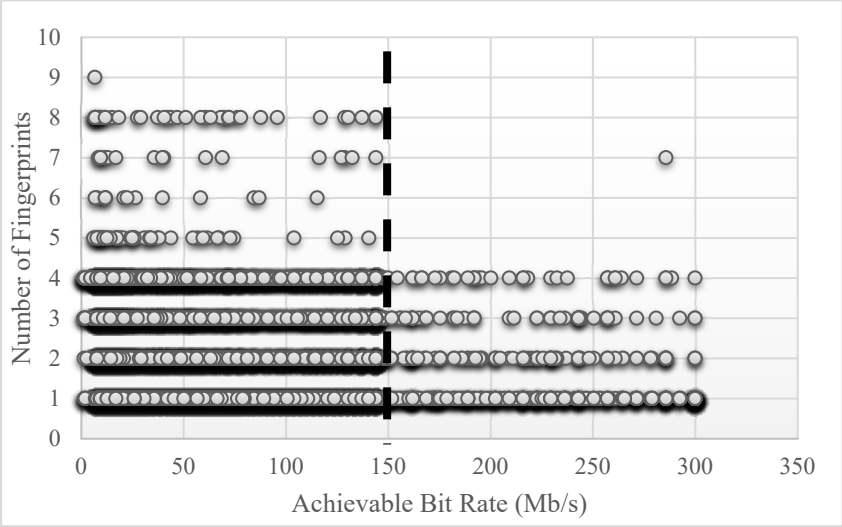


Figure 3 – Number of Hourly Fingerprints versus Achievable Bit Rate (2.4GHz of XB6s)

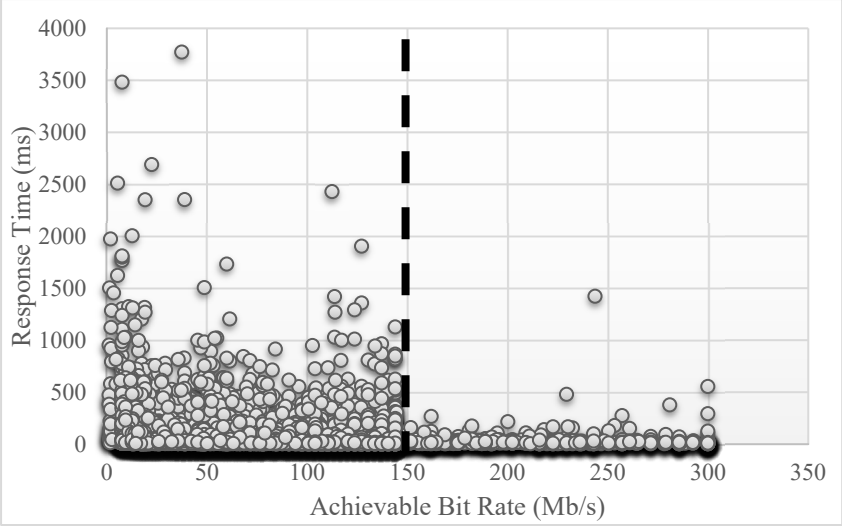


Figure 4 – Response Time versus Achievable Bit Rate (2.4GHz of XB6s)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In Figure 5, the relationship between the achievable bit rate and the STB RSSI is illustrated for the connection type mentioned above. As depicted in the figure, when the RSSIs are below -62 dBm, the achievable bit rate rarely exceeds 150 Mb/s. Therefore, we can designate -62 dBm as the threshold for the RSSI values that indicates service degradation. To summarize, it can be concluded that RSSIs below -62 dBm lead to achievable bit rates lower than 150 Mb/s, increased response times, and overall degradation in video quality. Consequently, a low RSSI below the threshold of -62 dBm is identified as one of the root causes incorporated into our decision tree specifically for 2.4GHz bandwidth connections on XB6 GWs. It's important to note that employing a similar analytical approach reveals that a bit rate of 866 Mb/s serves as the critical threshold for ensuring optimal service quality for STBs operating on the 5GHz band of an XB7 device. Figure 6 underscores the essential requirement that the RSSI must surpass the -68 dBm threshold to attain this desired bit rate. Hence, we can designate the RSSI of -68 dBm as one of the thresholds indicatives of service degradation for STBs connected to the 5GHz band of an XB7.

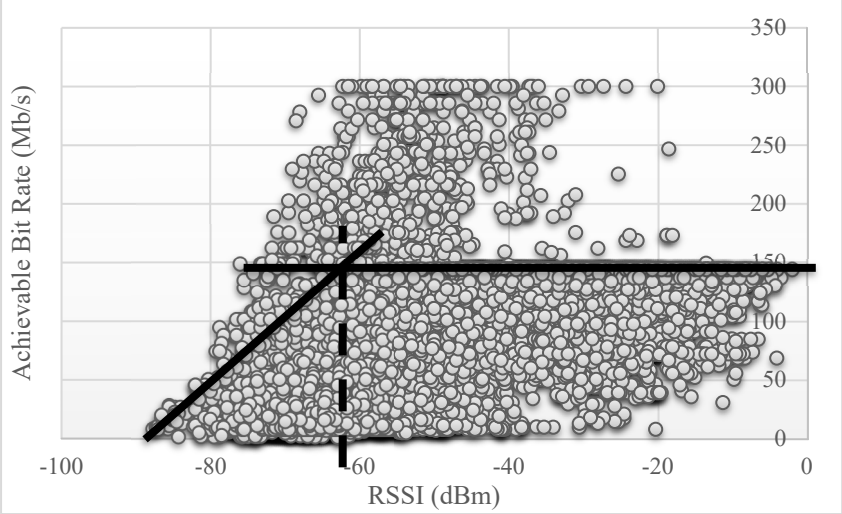


Figure 5 – Achievable Bit Rate versus RSSI (2.4GHz of XB6s)

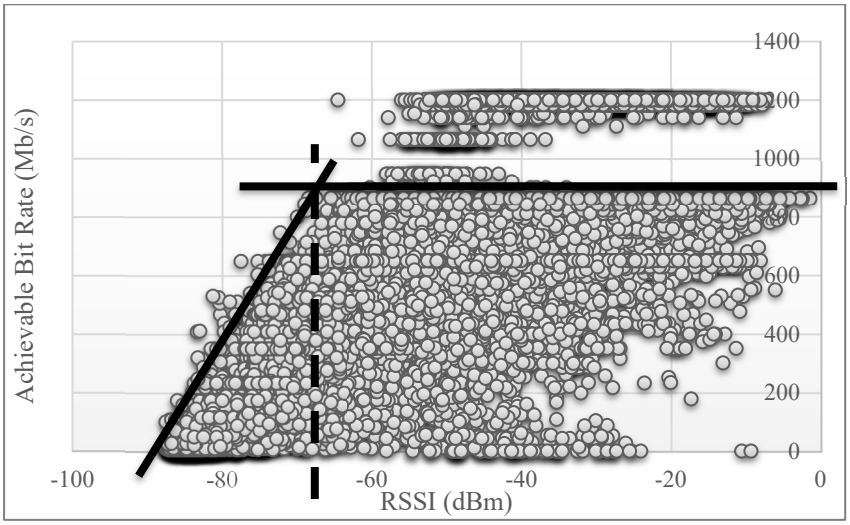


Figure 6 – Achievable Bit Rate versus RSSI (5GHz of XB7s)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

As mentioned earlier, channel utilization also has a major effect on the video quality and overall service performance. Figure 7 displays the achievable bit rate versus the channel utilization for STBs connected to 2.4GHz bandwidth of an XB7 GW. As depicted in this figure, for utilization below 70% there is a notable concentration of samples with achievable bit rates of more than 150 Mb/s as opposed to greater than 70%. Based on this analysis, 70% can be considered the threshold for the WiFi band congestion in the decision tree.

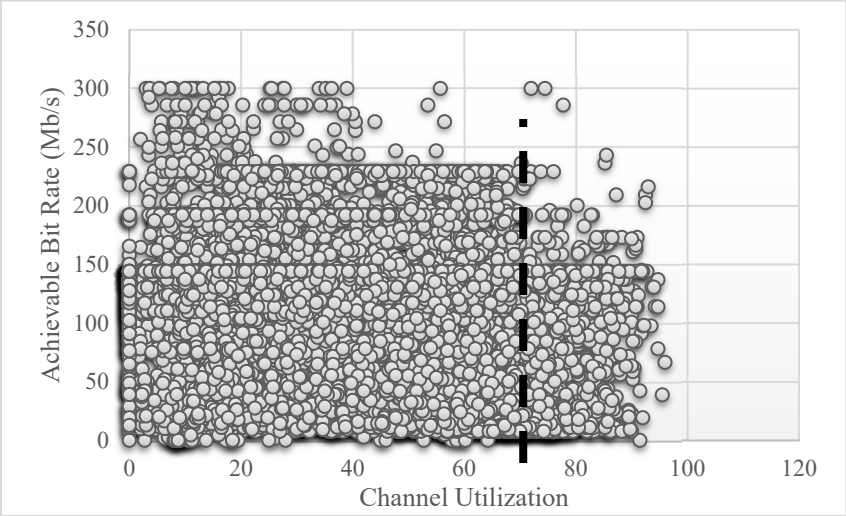


Figure 7 – Achievable Bit Rate versus Channel Utilization (2.4GHz of XB7s)

3.2. Decision Tree Development

Combining the individual filters, models, WiFi Band and WiFi parameters thresholds allows us to begin structuring the branches within our decision tree. Figure 8 elegantly navigates a specific branch (shown by the red box) for a particular model, traversing Band, RSSI and Utilization ultimately leading to a definitive resolution. This figure clearly demonstrates that the customer's needs can be met by incorporating a POD, effectively expanding the optimal WiFi coverage throughout the house. This, in turn, ensures that the STB in question is directed towards a suitable Access Point (AP) operating on a broader bandwidth and resulting in an enhanced RSSI.

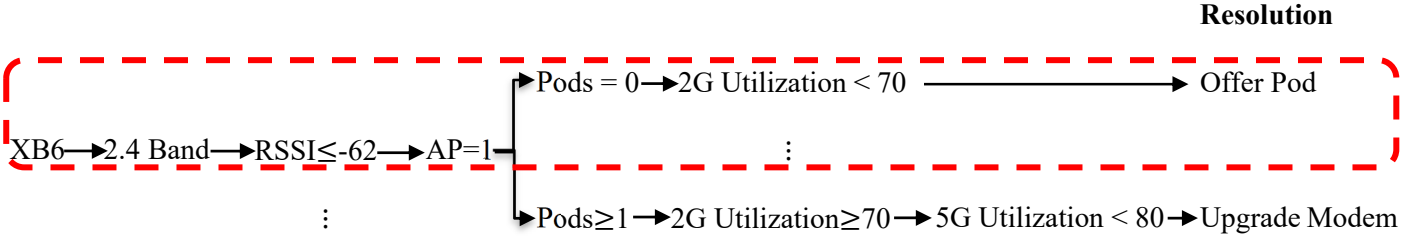


Figure 8 – Example Branches of Developed Decision Tree (AP=1 depicts that the STB in question never changed its connection)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.3. Clean Home Development

Ultimately, the output resolutions obtained from the decision tree discussed above, combined with additional KPIs, are fed into another algorithm known as clean-home. This process allows for an extensive analysis and optimization of various factors, contributing to an overall improvement in home network performance and video streaming quality. Table 3 presents a comprehensive list of all KPIs, accompanied by concise explanations, which constitute essential components of the clean-home algorithm. Implementing this algorithm helps us to enhance the efficiency of our proactive program. In simpler terms, each customer has a daily report from the clean-home algorithm, highlighting all detected KPIs and issues. Hence, as we contact the customer, we can address all the identified problems simultaneously, streamlining the resolution process and optimizing our interactions with customers.

It is essential to highlight that our actions concerning customers are prioritized according to the KPIs as well as the number of days the KPI is being degraded. For each customer that appears in the clean-home list, we look at a seven-day window and calculate a normalized score using the following formula

$$customer\ score = \frac{\sum_{i=KPI} w_i d_i}{7 \sum_{i=KPI} w_i},$$

where w_i represents the allocated weight for each KPI and d_i represents the number of days that the KPI has been identified as degraded during the last 7 days. By utilizing this formula, we can objectively prioritize actions and allocate resources more efficiently, ensuring an optimal proactive approach that resolves customer issues. The subsequent section delves into the comprehensive results obtained thus far from the clean-home algorithm, providing detailed insights on the observed improvements.

Table 3 – Clean-Home KPIs

KPI	Explanation
Video quality, upgrade GW to a higher model	Video quality degradation event observed for three consecutive days; solution: upgrade GW
Video quality, add pod for better coverage	Video quality degradation event observed for three consecutive days; solution: offer POD
Video quality, customer equipment	Video quality degradation event observed for three consecutive days, customer is using a third-party device; solution: offer pod
ODM Impaired	In-home RF issue, when ODM_score>200 and node_score<25
GW re-registration	If the modem goes offline (checking all offline events) at the same time the IPTV customer is using the service for three consecutive days
Daisy chained POD	Multiple PODs are interconnected in a serial structure
Poor POD backhaul	90% of the reported backhaul RSSI of the POD is less than -75 dBm for three consecutive days
Poor coverage, add pod	Low normalized home score calculated according to STB parameters such as RSSI, packet loss, number of errors, phyrate,…; solution: offer POD

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. Results and discussion

In this section, we will discuss the improvement results obtained from the clean-home algorithm. Through the utilization of the clean-home algorithm, we have proactively engaged with over 5000 customers, implementing various tailored solutions to enhance their experience. These solutions include providing PODs to improve coverage, recommending POD relocation for better backhaul signal, upgrading customer GWs or PODs to newer versions with advanced technologies, especially when multiple PODs are interconnected in a serial structure.

Figure 9 illustrates the cumulative improvements for each clean-home list to date. To calculate the improvement percentages, we compare the WiFi performance of customers one week before contacting them with their WiFi performance three weeks after the solution is provided. Also, to ensure a fair comparison, we carefully select a control group that we don't do any actions on; By comparing the progress of customers who received specific actions with those who didn't receive any interventions, we effectively assess the impact of our actions. As shown in this figure, for all the clean-home lists, the improvements achieved for customers who received specific actions are nearly double when compared to the control group. This observation highlights the significant impact of the interventions in enhancing the WiFi performance and indicates the effectiveness of the actions taken.

Figure 10 illustrates the distribution of improvement percentages. As Figure 10-a depicts, for the majority of the customers all the issues are being resolved. Figure 10-b emphasizes that more than 72% of the customers that we did an action on are experiencing 80-100% improvement. Finally, in Figure 11, we present a specific comparison of the total number of video quality degradation events that occurred one

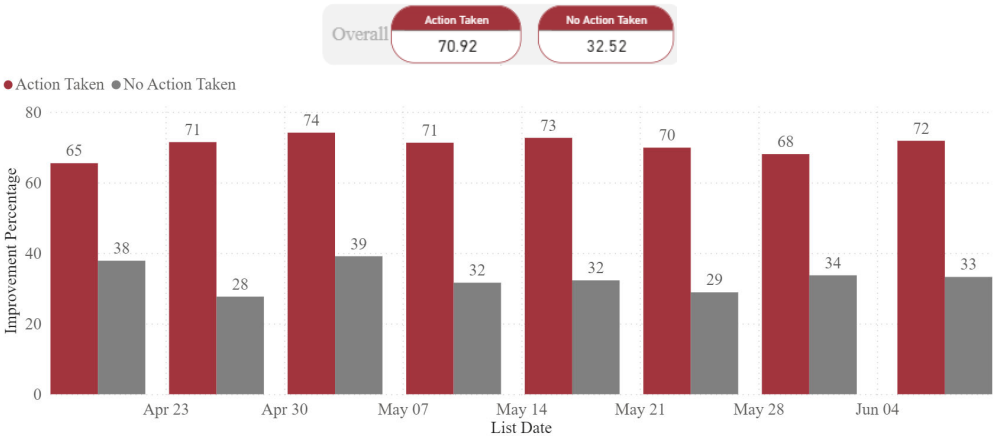
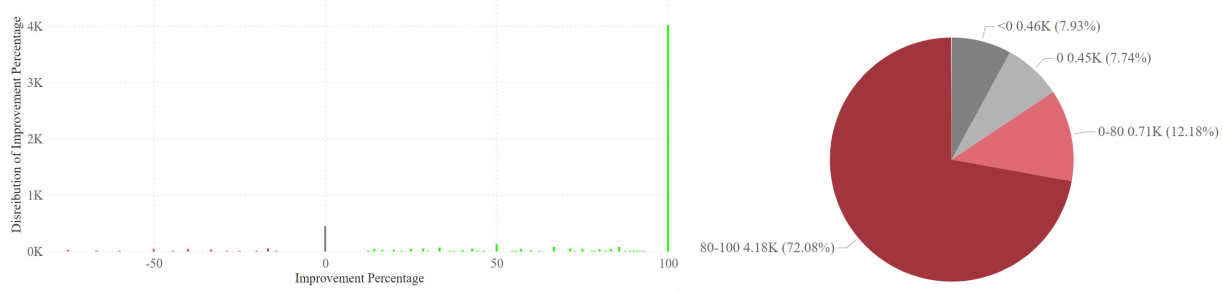


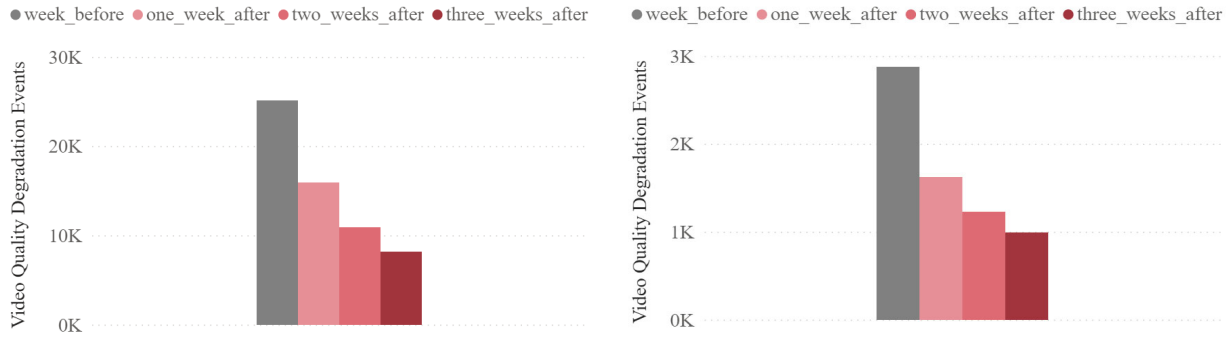
Figure 9 – Cumulative Improvement of Clean-Home Lists



a) All Improvement Percentages b) Bucketized Improvement Percentages

Figure 10 – Distribution of Improvement Percentage

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



a) Offer POD to Customer b) Upgrade Customer's GW

Figure 11 – Total Video Quality Degradation Events

week before the resolution was provided, with those happening one, two, and three weeks after. We explore this comparison across various solutions, including offering POD and upgrading the customer's GW. When comparing the number of video quality degradation events from the week before to the three weeks after, we observe significant improvements at approximately 70% for both solutions. In summary, all the results conclusively demonstrate the effectiveness of our proactive approach and the reliability and success of our strategies in addressing the WiFi in-home challenges.

5. Conclusion

In this paper, we have presented a proactive approach that by leveraging data-driven techniques, effectively identifies and addresses in-home network issues. Through extensive experimentation and engagement with over 5000 customers so far, we have demonstrated the remarkable effectiveness of our method in proactively optimizing the performance and stability of in-home networks. The analysis of the results highlights the resounding success of our proactive clean-home algorithm, achieving a 70% improvement overall for the group of customers who received our targeted interventions. Notably, a significant majority of these customers experienced a remarkable 100% enhancement in WiFi performance after implementing the suggested actions.

The results of this work can be extended through the further optimization of our customer scoring method in future work. By delving deeper into the feedback from customers who encountered WiFi problems but chose not to participate in our program, we can establish meaningful correlations and craft a refined scoring method capable of accurately predicting a customer's inclination to engage in our proactive program. This provides a strategic advancement in our resource allocation and by directing our efforts towards customers who are most likely to benefit from, ensures a more targeted and efficient proactive approach.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AP	access point
GW	gateway
IoT	Internet of things
POD	point of delivery
PNM	proactive network maintenance
RSSI	received signal strength index
SINR	signal to interference plus noise ratio
STB	set-top box
UHD	ultra high definition
WiFi	wireless fidelity

Bibliography & References

[1] Volpe, B., Ottlik, B. (2021). Machine Learning and Proactive Network Maintenance: Transforming Today's Plant Operations. *SCTE 2021 Fall Technical Forum.*

[2] Cable Television Laboratories, Inc. (2016). *PNM Best Practices: HFC Networks (DOCSIS 3.0).*

[3] able Television Laboratories, Inc. (2022). *Primer for PNM Best Practices in HFC Networks (DOCSIS 3.1).*

[4] Zhu, J., Sundaresan, K., & Rupe, J. (2020). Proactive Network Maintenance using Fast, Accurate Anomaly Localization and Classification on 1-D Data Series. *2020 IEEE International Conference on Prognostics and Health Management (ICPHM).*

[5] Dugan, K., Evans, J. (2022). A Deep Learning Approach for Detecting RF Spectrum Impairments and Conducting Root Cause Analysis. *SCTE 2022 Fall Technical Forum.*

[6] Volpe, B., & Ottlik, B. (2021). Machine Learning and Proactive Network Maintenance: Transforming Today's Plant Operations. *SCTE 2021 Fall Technical Forum*.

[7] Virang, D., Chari, S., & Kraiman, S. (2021). Machine Learning for RF Impairment Classification. *SCTE 2021 Fall Technical Forum.*

[8] Ferreira, J., Harb, M., Subramanya, K., Santangelo, B., & Rice, D. (2020). Convolutional Neural Networks for Proactive Network Managment. *SCTE 2020 Fall Technical Forum.*

[9] Wolcott, L., Heslip, J., Thomas, B., & Gonsalves, R. (2016). A Comprehensive Case Study of Proactive Network Maintenance, *SCTE/ISBE Cable-Tec Expo 2016.*

[10] Rupe, J., Zhu, J. (2019). Kickstarting Proactive Network Maintenance with the Proactive Operations Platform and Example Application, *SCTE/ISBE 2019 Fall Technical Forum.*

[11] Shannon, C. E. (1949). The Mathematical Theory of Communication. *Urbana, IL: University of Illinois Press.*

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Artificial Intelligence (AI)-Based Cross-Platform Node Name Mapping (Cmap-NN)

A Technical Paper prepared for SCTE by

Jordan Kupersmith
Consultant
Cox Communications
6205-B Peachtree Dunwoody Road NE, Atlanta, GA 30328
404-915-6955
Jordan.kupersmith@cox.com

Wei Cai
Lead Data Scientist
Cox Communications
6205-B Peachtree Dunwoody Road NE, Atlanta, GA 30328
404-414-7876
Wei.cai@cox.com

Jeonpaolo Barvez, Cox Communications

Pavan Chandrashekar, Cox Communications

Richard Brown, Cox Communications

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Background and Literature Review 6
   - 2.1. Fuzzy String Matching and Related Work 6
   - 2.2. Haversine Distance 8
   - 2.3. Gaussian Mixture Model (GMM) Clustering 8
3. Methodology 9
   - 3.1. Levenshtein Distance Matching and ETL Scoring 10
   - 3.2. Haversine Distance 14
     - 3.2.1. Choice of the Cutoff Distance 14
     - 3.2.2. Calculation of the Node Similarity Score 15
   - 3.3. GMM Clustering 15
   - 3.4. Ensemble of the Methodologies 15
4. Results 16
   - 4.1. Levenshtein ETL Scoring Method 16
   - 4.2. Haversine and GMM Method 20
   - 4.3. Ensemble Method 21
5. Conclusion, Limitations and Further Research 23
   - 5.1. Conclusion 23
   - 5.2. Limitations 24
   - 5.3. Next Steps 25

Abbreviations 26

Bibliography & References 26

List of Figures

Title **Page Number**

Figure 1 - Levenshtein Distance Algorithm 7
Figure 2 - Haversine Distance Equation 8
Figure 3 - Gaussian Mixture Model Algorithm 8
Figure 4 - Diagram of the Matching System 10
Figure 5 - Weighted Mean Formula 13
Figure 6 – Node Similarity Score 15
Figure 7 - Node Geographic Distribution 16
Figure 8 - Confidence Distribution of Results for SmartPhy Nodes 17
Figure 9 - Confidence Distibution of Results for GNIS Nodes 18
Figure 10 - Confidence Distibution of Results for ICOMS Nodes 19
Figure 11 - Confidence Distibution Comparison Between Systems 19
Figure 12 - Mapped Node Name Counts by Levenshtein FuzzyString, Haversine Distance and GMM Clustering 21
Figure 13 - Mapped Node Name Counts by Levenshtein FuzzyString, Haversine Distance, GMM Clustering, Ensemble Method 22

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Tables

Title **Page Number**

Table 1 – Systems Overview 5
Table 2 – Smartphy by GNIS Count 13
Table 3 – Sample Nodes For Venn Diagram Clasification 20
Table 4 – Sample Node-Level Coordinates (eg Longitude and Latitude) 21
Table 5 – Ven Diagram Explanation 22

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The Strategic Network Planning and Automation team is the team at Cox Communications that is responsible for defining and applying strategic network growth policies, forecasting node level throughput, and creating a short to long-range plan for node actions From there, the plan is used by Cox's Outside Planning team, which plans anything downstream from the hub and Inside Planning team, which plans in the hub, to morph them into an operational plan. The plan drives millions of dollars in node actions annually for Cox. These node actions are vital for keeping the Cox customer connected and ensuring the quality of their connected experience. The node actions are also key for ensuring the continued growth and competitive viability for the company. Having the planning as accurate as possible so the right actions are being conducted at the right times and in the right places are necessary.

The team faces and has overcome many challenges and continues to work on the challenges they encounter daily. For example, the team has created highly predictive modeling using statistical and machine learning models to tune the model forecasts and provide greater accuracy over the years. Additionally, the team has created rules-based optimizations to create a more efficient planning process and limit the need for multiple touches on the node. This not only improves the customer experience, but it also saves money.

However, as the network grows more capable, it also grows more complex. With that, another set of challenges the team arises, which is the challenge of poor data quality. This is a growing problem as the complexity and size of the network continuously grows. With addresses mapped to the incorrect node, or node names not matching across databases the situation leads to errors in the field, and this can often lead to issues that impact customers.

As such, the team was challenged to couple its network expertise with their refined data science capabilities to define a machine learning algorithm that can not only identify mistakes in node attributes but predict the correct information. In this paper, we will dive into the technical aspects of the logic and the results of the proof of concept that one day will lead to greater accuracy in cross-system alignment, and practically eliminate errors, resulting in a more accurate planning process and avoid the costs associated with mistakes.

Cox relies on the outside plant (OSP) optical node name heavily as a major key in joining different systems together. In some instances, this can be the only key that may be used to join two or more systems together. As such, when there are discrepancies in one system for what the correct node name should be, this can cause a fallout in reporting or in some cases, customers receiving incorrect speed tiers or incorrect nodes being actioned for congestion relief and other upgrades. Three of the core systems Cox uses which has a significant influence on the node name are Geographic Name Information System. (GNIS), the collective Data Over Cable Service Interface Specification (DOCSIS) systems (CMTS and Cisco SmartPhy) and the Integrated Communications Operations Management System (ICOMS).

Cox uses a GNIS software system to maintain primarily outside plant network topology connectivity and assets. This system is critical to providing connectivity of customer to RF tap (and ultimately their optical node) as intended by the network design teams. The Hybrid Fiber-Coaxial (HFC) Design team is the originating stakeholders of which customers connect to which node. This name to topology association is ultimately passed to other systems such as ICOMS and inputted in the Cable Modem Termination System (CMTS). The topology information in GNIS can be associated with customers' billing information in ICOMS via their street address.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Commonly known as DOCSIS, this protocol is the international standard for IP traffic over coax. As mentioned in the context of this paper, it will collectively refer to the CMTS and SmartPhy data. The CMTS has a couple areas in the configuration where the node name may be written in the description field. Similarly, SmartPhy also allows for a "Remote Phy Device (RPD) Name" field which enables us to populate it with the Outside Plant (OSP) node name. The SmartPhy node name can be ensured it matches the node name in the service group description field by way of the chassis and Fiber Node ID field, which is a unique key for matching when combined. The customer identifiable information in the CMTS would be a list of customer premise equipment (CPE) MAC addresses reporting to each interface, which can be used to correlate to the MAC addresses in ICOMS and establish a physical location of each customer's device.

ICOMS is the third main source being leveraged. This is Cox's main billing system which contains certain customer identifiable data such as OSP node name, CPE, MAC address, and physical address. The MAC address can be mapped to a MAC address in the CMTS to match the node name per the CMTS via chassis and interface.

Unfortunately, data sources, especially large ones that are being constantly changed, are prone to discrepancies. While the interface to MAC address data in the CMTS is always correct, the device (node) name inputted in the CMTS/SmartPhy is made via manual entry. The node name in ICOMS is manual entry every time a new home becomes serviceable, or a node action causes a node name change. The GNIS OSP design data is all manual entry. Due to the ultimate human factor involved in node name entry for all core systems, this does pose some challenges to the machine learning factor (as it always has with the human factor).

Table 1 – Systems Overview

System	Node Name Entry Method	Level	Node originator
DOCSIS	Manual	Node	N
ICOMS	Manual	Customer	N
GNIS	Automatic	Node	Y

The current methodology for identifying discrepancies in the node name between systems is a dynamic reactive method. This is identified at two different granularities: one at the customer level and two, at the node level. The customer level tends to have the least impact traditionally of the two, with the primary impact of a single or staggered few customers within a node having the incorrect node name being subject to the possibility of services being sold that cannot be supported on their node. The node level granularity, which is when most of the customers within a node have the incorrect node name, can cause the incorrect CMTS interface to be associated to the OSP node name. This can result in node actions on a node that does not need one or higher/lower speed tiers being assigned to all customers in a node which cannot support those tiers of service. There are currently no teams nor anyone working or discovering these discrepancies as part of their official duties. Employees from various departments will discover these node name discrepancies accidently as they are impacted by them and submit them for a fix.

To address this problem, a Machine Learning based node name determination solution is proposed in this paper. If two systems have a street address, a fuzzy-string matching algorithm can establish linkage. For systems that can only be linked with location coordinates a, distance matrix and GMM clustering model are built.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The rest of this paper is organized as follows. In Section 2, we introduce the theoretical description of the chosen string similarity algorithm with the focus on the fuzzy string matching. The other algorithm – Haversine Distance used in this paper is also discussed. We further introduce theoretical principles behind the Gaussian Mixture Models (hereafter GMM) Clustering Algorithm, which is concluded with a summary of a related work. In Section 3 we describe the dataset used for our studies, together with the data sanitization used prior to the Extract Transform and Load (ETL) Scoring, Haversine Distance analysis and GMM clustering models, and present methodologies used in this paper along with our proposed hybrid scoring algorithm. In Section 4 we focus on presenting results and improvements accomplished by using the proposed model algorithm in mapping nodes across systems. Further research and next steps are summarized in Section 5.

2. Background and Literature Review

The way that street addresses are structured can vary, their components can be unstable, and they are prone to different human entry errors (Coetzee and Rademeyer 2009). When comparing them between two systems they are not always exact strings. While postal addresses are structured information since it contains easily identifiable and delimited fields such as city, postal code and street number, they can be written several ways through the use of abbreviations, synonymous and contractions, resulting in the possibility of non-exact strings for the same postal address between two systems (Perez et al 2008). Due to these constraints, postal addresses between two systems cannot be joined directly nor is it possible to generate a list from one to be used to filter the other. For this reason, Natural Language Processing (NLP) techniques must be used.

The fundamental aspect of the possibility of being able to match addresses is the algorithm used for measuring distance and dissimilarity between two postal addresses. There are several algorithms to choose from and they have a variety of techniques. There are algorithms that measure the changes required to convert an expression to the other that compare their textual structure, vocal structure, or a combination of both methods. Examples of textual methods are Guth, Hamming, Levenshtein, and Damerau–Levenstein, all of which measure the edit distance between two expressions counting the inserts, deletes, and substitutions required to transform them. Vocal methods such as Soundex, Metaphone and Phonex leverage the phonetics of words to analyze their similarity. Finally, there are methods like Jaccard and Cosine, which analyze a combination of string and vocal structures to provide probabilities to represent the similarity of the expressions (Rezayan et al 2018).

2.1. Fuzzy String Matching and Related Work

Fuzzy string matching is a technique used to find approximate matches between two strings. Algorithms may be divided into two categories due to the feature they measure:

- similarity algorithms: the match is found if S(X, Y) ≥ t_S,
- dissimilarity algorithms: the match is found if D(X, Y) ≥ t_D,

where $t_{S/D}$ is a string similarity/similarity threshold, S(X, Y) and D(X, Y) are the similarity and dissimilarity functions, X and Y are the two strings in question. When two datasets share only a single imperfect identifier, this is sometimes called the *fuzzy string matching problem* (Filipov and Varbanov 2019; Hall and Dowling 1980).

Previous research has shown that the best algorithms for detecting one road name written in two different ways are Mongo Elka soundex Jaro, Jaro Winkler, and Levenshtein in that the specified order when, for

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

example, comparing “Avenue J.F. KENNEDY” with “Avenue J-F KENNEDY”. However, when comparing two strings that are not the same such as “RUEDESARDENNES” and “RUEDESJARDINS”, all previously mentioned algorithms except for Levenshtein produce high similarity scores, meaning they have high false positive rates (Charif et al). In all classification problems, one must always choose whether it is more important to minimize false negatives or false positives, and the preference varies by the context. In the context of this paper, false positives are more detrimental than false negatives because failing to identify a match is not as determinantal as having false matches due to the quantity of matches in the system. If one match fails, there are others to stand in its place. For this reason, Levenshtein distance is the best suited algorithm for these purposes.

Furthermore, speed is just as important as accuracy. Fuzzy Matching on strings requires calculating the cartesian product of both datasets before calculating the similarity score between strings. The Cartesian Product is defined as the set consisting of all ordered pairs. For example, if $A = \{x, y\}$ and $B = \{3, 6, 9\}$, then $A \times B = \{(x, 3), (x, 6), (x, 9), (y, 3), (y, 6), (y, 9)\}$ (Encyclopedia Britannica). Given that there are millions of customers, it is important to be able to implement the algorithm quickly. Levenshtein Distance is not the fastest algorithm, but it is also not the slowest according to previous research (Christen 2006). However, Python, the programming language which is used for the majority of the study, has a library called FuzzyWuzzy that easily implements the Levenshtein distance in parallel, fully being able to harness the processing power of the 40 CPU system that it runs on. For the above-mentioned reasons, we selected Levenhstein Distance for string matching.

Levenshtein Distance is a mathematical measure used to calculate the difference or similarity between two strings of characters. It counts the minimum number of operations needed to transform one string into another, where each operation can be either an insertion (adding a character), a deletion (removing a character), or a substitution (changing a character). In short, the Levenshtein Distance tells you how many changes you would need to make to turn one word or phrase into another, giving you a measure of their overall similarity or dissimilarity. (Levenshtein 1966). For example, if the source string is “maple” and the target string is also “maple” then the distance is 0 because no transformations are needed, and the strings are identical to each other. On the other hand, if we compare Poplar ST and Popular AVE there is a Levenshtein Distance of 4 because to transform the source string into the target first we must add a “u”, then substitute an “s” with an “a”, then substitute a “t” with a “v” and finally add an “e”. Hence, the higher the Levenshtein Distance, the more different the strings are. The final distance is then displayed as 100-total distance count (Chen et al 2014).

$$\operatorname{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \operatorname{lev}_{a,b}(i-1,j) + 1 \\ \operatorname{lev}_{a,b}(i,j-1) + 1 \\ \operatorname{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

Figure 1 - Levenshtein Distance Algorithm

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.2. Haversine Distance

The Haversine Distance is a well-established method to compute the distance between two points represented as latitude-longitude coordinates. This formula was first presented by James Inman in 1835 (Inman, 1849), and serves as a foundational technique in numerous geographical information systems.

$$haversin(d/r) = haversin(\emptyset_{(}2-)\emptyset_1) + cos(\emptyset_1) + cos(\emptyset_{(}2))haversin(\lambda_{(}2-)\lambda_1)$$

Figure 2 - Haversine Distance Equation

Haversine Distance has been used to determine the shortest distance between two points on the surface of a sphere, given their longitude and latitude. This is especially useful in the context of finding location similarity, as it provides a method to calculate the distance between two geographic coordinates. Historically, the formula has been used in navigation and cartography, allowing sailors to determine the shortest path between two points on a globe before the advent of electronic navigational tools. There have been multiple research papers that discuss or implement the Haversine formula. For instance, research into efficient algorithms for geospatial data, location recommendation systems, and spatial network analyses often leverage the Haversine formula for distance computations.

2.3. Gaussian Mixture Model (GMM) Clustering

GMM is a mixed density model with a basic Gaussian distribution, which is equivalent to the weighted average of multiple Gaussian probability density functions. Each Gaussian density function is called a component, which represents a cluster. In the following description of GMM algorithm, α_ι denotes the weight of each Gaussian distribution in GMM, and μ_i, α_i^2 denote the mean and the covariance matrix respectively.

$$\mathrm{P(t)} = f(x) = \sum_{i=1}^{N} \alpha_i N(t; \mu_i, \alpha_i^2)(\alpha_i \geq 0, \Sigma_{i=1}^{N} \alpha_i = 1)$$

Figure 3 - Gaussian Mixture Model Algorithm

The robustness of GMM stands out when compared to traditional clustering techniques. While K-means clustering and Hierarchical clustering segment the data based on distance or hierarchical structures, and Density-based clustering (DBSCAN) operates on the density distribution of data points, the GMM lies in its ability to approximate continuous distributions with high accuracy, even when the component count escalates. This precision becomes particularly advantageous when addressing the spatial distribution of intricate datasets, like the ones examined in our study.

Our research sought to utilize the GMM's capabilities in generating cluster fingerprints that aptly represent the spatial distribution of our cross-platform nodes. Another intrinsic advantage of the GMM is its soft clustering feature. Unlike hard clustering methods that definitively assign data points to a specific cluster, GMM computes the probabilities of a data point belonging to different clusters. This probabilistic assignment becomes imperative when handling data points that are ambiguous and could potentially fit into multiple clusters. By leveraging these probabilities, our study derives the maximum likelihood for each data

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

point. This probabilistic output was subsequently transformed into a scoring component in our proposed algorithm, further aiding in the detailed interpretation and categorization of our data.

3. Methodology

The burgeoning field of computational techniques provides insights into intricate problems, one such being the identification of identical nodes and gauging the accuracy of node names. Guided by existing literature, particularly on fuzzy string matching, Haversine distance, and GMM clustering in spatial data, we create an innovative intersection. This nexus aims at harnessing computational prowess to mirror human cognitive thinking processes that intuitively recognize node similarities and discrepancies. There are numerous advantages, especially considering machine learning (ML) paradigms, renowned for swiftly discerning potential matches and mismatches-surpassing human speeds and possibly, accuracy.

As foundational steps, our methodology encapsulates both data sampling and subsequent data processing. Before diving deep into algorithms and ensemble learning, it becomes imperative to understand the data's structure, its origins, and the preliminary operations performed to make it amenable to advanced analytical processes.

Due to the volume of data and its processing time involved, only one geographic area was selected for the purpose of the study. Over 7,350 of nodes' data were selected as the dataset used in our analysis. Arizona was chosen because it has the greatest number of nodes that have gone digital, and thus resulted in a lot more node name changes.

In our methodology, we first procured low-level geographical data, specifically latitude and longitude coordinates, from each respective system under study. These granular data points were subsequently aggregated, culminating in a node-level dataset. This consolidated dataset then becomes pivotal as it serves as the input for our subsequent processes.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.1. Levenshtein Distance Matching and ETL Scoring

In this dataset, we follow the below data model to collect data from each system:

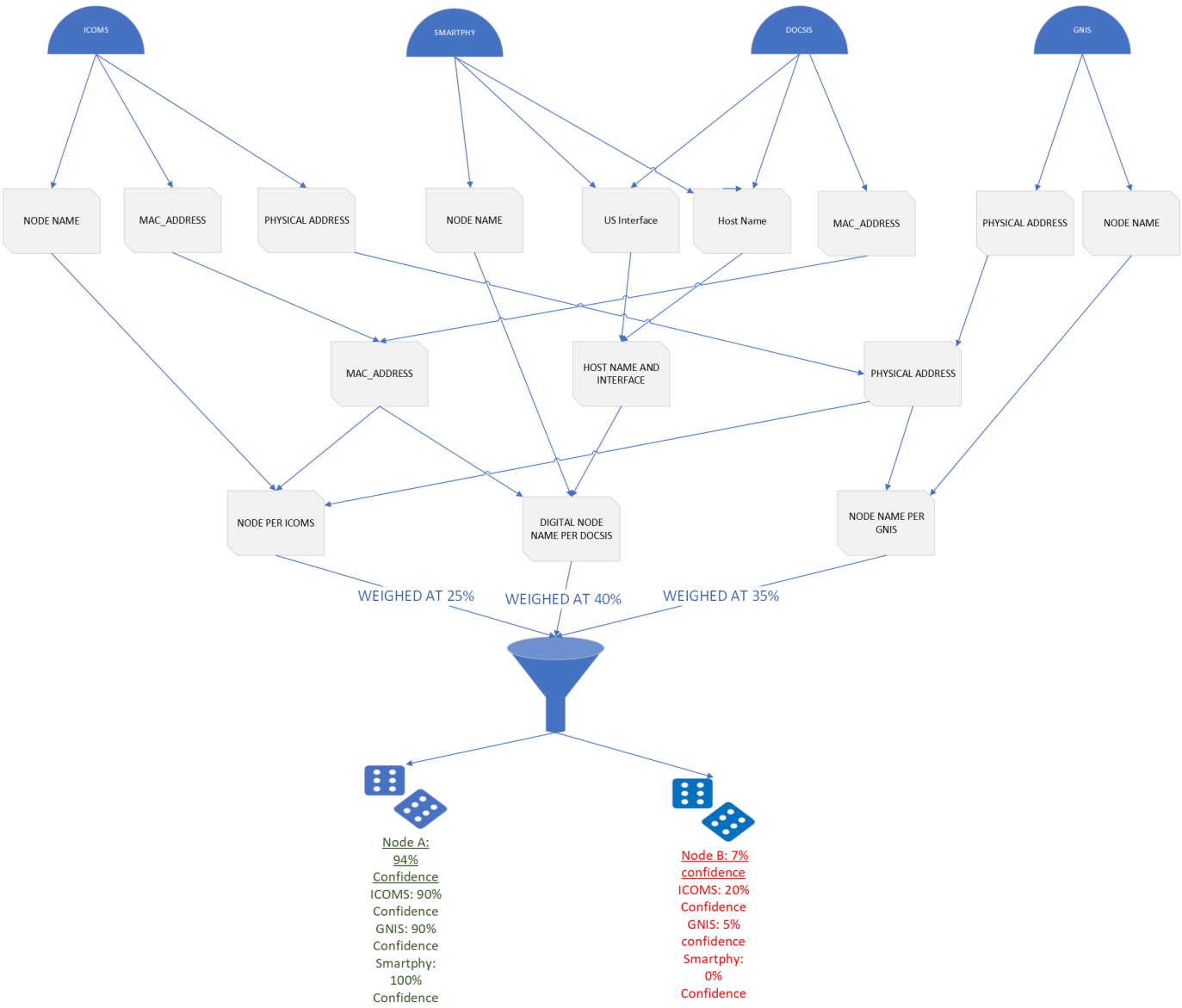


Figure 4 - Diagram of the Matching System

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

From the above data model diagram, we can see that while we cannot use node name to establishing links between the systems in all cases, there are ways to link them together. The main challenge between linking them together is that the links between each system are not always the same variable, and the format of these variables is not always the same due to different standards of writing street addresses between systems. In other words, an exact string join is not possible. Furthermore, these systems rely on many records of MAC addresses to be linked together, and while joining them would be straightforward if they were in the same database, these systems are not connected in any way, therefore it requires creative computational techniques to be able to link them.

The system begins by examining a node in SmartPhy. Each node has a host name and interface name associated with it, which is how this data source relates to DOCSIS. The tables are systems that are on the same database with slightly different text format, so after some string manipulation they can be easily joined together. Additionally, unlike subsequent steps, this join is done at the node level, meaning that it does not present any challenges of big data.

Nonetheless, all subsequent steps are done at the MAC address and street address level, meaning that millions of records need to be reconciled, and it is not possible to match these using simple database technology. Matching these records is only possible through programmatic techniques designed to meet the challenges posed both by big data and the fact that the data lies on disparate data systems.

Due to limits of RAM, this processing must be done by looping through nodes and writing output to the disk to minimize consumption of RAM at a given time. A list is generated of all of the nodes in the SmartPhy/DOCSIS dataset and that list is used to iterate through the dataset by node. In the loop, a temporary dataset is created that contains all MAC addresses per DOCSIS node, which is an average 1,500 in each node.

The MAC addresses from DOCSIS are then looked up in the data-frame for ICOMS data, resulting in the first cross-system node mapping of the program. This is because while we have a node name in DOCSIS and we have a nodename in ICOMS, we cannot trust that a given node name points to the correct combination of device and customer. In other words, let's say that Customer A has device 1. We have node XY1 associated with Customer A and we have node LM2 associated with device 1. While we can prove that device 1 belongs to customer A because they share a MAC address, we have no idea which node name is right. For this reason, we must line up customers and devices to know when node names coincide and when they do not. This process matches DOCSIS with ICOMS.

The next step takes our combined DOCSIS/ICOMS dataset and matches it to GNIS based on street address.

Given there are millions of customers, checking every single address in one system against every single address in another system is computationally too expensive. If Cox had only 1 million customers, this would be 1 trillion calculations. If each calculation took 1 second, this would take 31,709 years. While this could be parallelized, with 30 cores it only speeds it up to 1,000 years. If we were to quadruple the speed, then it would be reduced to 250 years. Even if it is done for just one node it would take 42 days per node at 1 second per address without multiprocessing. It is clear that NLP alone cannot solve this problem.

For this reason, it is necessary to pre filter the data before submitting it for NLP matching. For a given node in SmartPhy, all the zip codes in that node are extracted in the ICOMS dataset after passing it through the filter of MAC addresses in DOCSIS and these zip codes are then used as a filter to reduce the My World dataset to contain only entries that have the same zip codes which are present in ICOMS for this node. Subsequently, all the street numbers present in ICOMS are extracted and used to filter the My World dataset

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

to only contain records where the street number is present. By reducing the My World dataset to only a combination of zip codes and street numbers which are present in the ICOMS dataset, it can be filtered, which is possible because numeric data can easily be matched precisely, greatly reducing the number of records that need to be fed into NLP.

Given the large amount of data, this operation is run in parallel, using one less core than the maximum available number of cores and the amount of time this takes varies widely by node. There are nodes that can finish in 25 seconds and other nodes that take longer than 20 minutes. The result is a matrix of all address combinations from ICOMS, with the address candidates from My World and their score. This matrix is combined back onto the original ICOMS and My World datasets and they are merged based on the address string they contributed to the matrix which results in a combined dataset of all systems. In this dataset, there is a variable for node name for all 3 systems, the shared mac address between DOCSIS and ICOMS, the shared street address between ICOMS and My World as well as special attributes which are unique to only one system such as latitude and longitude which is only found in My World.

The matched data is then loaded into a loop where for each time a SmartPhy node appears per combination of MAC address and street address, each node name according to GNIS is counted. For example, if there are 100 combinations of MAC address and street address for node A and GNIS agrees that the node name is node A 70 times, but it says it is node B 40 times, the program returns a dictionary that attributes this count of 70/40 to the SmartPhy node in question. Nodes that make up less than 2% of the GNIS dictionary are removed to minimize noise. These counts are then translated into percentages and interpreted as the percentage that GNIS agrees with. In this example, the count of 70 and 40 can be translated to a percentage of 63.6% and 36.3%, respectively. These percentages can be interpreted as “GNIS agrees that node A is really called node A 63.6% of the time while GNIS believes 36.3% of the time that what SmartPhy calls node A should actually be called node B.” This same logic is repeated for each combination of MAC address and street address for all entries of SmartPhy’s node A for ICOMS to compare the ICOMS system to SmartPhy as well.

To examine a real world example, the third node in Table2 is 8ANPL as recorded by the SmartPhy system. For every MAC address that belongs to this node, we see that My World/GNIS calls this node by the same name only 32 times as opposed to calling it 8ANPJ 47 times. When also factoring in the lowest occurring node name, 8ANPK, this node is present in My World 106 times. The total percentage of confidence that My World gives to this node is 32/106=30.2%. This contrasts with the last row in the example that contains 8BSV2, where My World agrees 100% of the time resulting in a confidence score of 100% as well as the first row that contains 8EDG2 where My World does not cast even one vote of confidence in the node name, resulting in a confidence score of 40%, the minimum confidence for the SmartPhy system.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 2 – Smartphy by GNIS Count

SmartPhy	my_world	my_world_vote_for_SmartPhy	my_world_vote_pct_for_SmartPhy
8EDG2	{'01326': 18, '8EDG1': 328}	0	0.0%
8DWE1	{'00407': 252}	0	0.0%
8ANPL	{'8ANPJ': 47, '8ANPK': 27, '8ANPL': 32}	32	30.2%
8AEHB	{'8AEHB': 217}	217	100.0%
8BSV2	{'8BSV2': 298}	298	100.0%

All these percentages are recorded in a dictionary and stored in a dataframe to be associated with the SmartPhy nodes. For the example of 8ANPL, GNIS agrees with the node name in 30.2% of occurrences as shown in Table 2. Furthermore, ICOMS agrees with the node name in 29.5% of the occurrences. Since we are considering the information that SmartPhy presents, it automatically casts a vote for itself of 100%. These confidence votes are used in the weighted mean formula (Figure 5) with the weights of 40% for DOCSIS, 25% for ICOMS and 35% for GNIS as established in the introduction section of the paper. This results in an overall confidence of 57.9% that when SmartPhy claims it is the node name that it is indeed true.

$$\bar{X}_w = \frac{\Sigma w_i x_i}{\Sigma w_i}$$

Figure 5 - Weighted Mean Formula

While envisioning the framework of the machine learning model, a "weight" was decided to be assigned to each of the three main sources. A weight of 40% was assigned to DOCSIS, 35% to GNIS, and lastly 25% to ICOMS. A weight system was decided due to certain sources observed over the years that tend to have less discrepancies than others. The weights are also helpful being different between each source as this would especially help in situations where all three systems display a different node name. The node name in the DOCSIS-based systems (the CMTS and SmartPhy) tend to be the most accurate so we assigned it a weight of 40%. The GNIS system is meant to be the origination point of where all node names are created initially so one would assume that would have the greatest weight. However, the GNIS node name flowing through downstream databases first is not always the case as there can be a long delay before a design post a node action is pushed to a production state. This will result in all topology assets still indicating the old and now incorrect node name. The ICOMS system was given the least weight as it has been observed to have the most discrepancies over the years.

The same process of getting counts by systems is repeated two more times, counting the number of times a node name appears in GNIS and SmartPhy for each node name as recorded in the ICOMS system and counting the number of times a node name appears in ICOMS and SmartPhy for each node name according to GNIS. When examining the same node name from the perspective of the ICOMS and GNIS systems in our example of 8ANPL, it can result in difference confidence scores. For every instance where ICOMS claims a node to be named 8ANPL, GNIS agrees 100% of the time. However, SmartPhy only

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

agrees 33.3% of the time. Given that ICOMS votes for itself at 100% and its weight is 25%, the confidence score is 73.3% which can be interpreted as being a 73.3% confidence that when ICOMS claims a node is named 8ANPL, that it is true. When examining the same node from the perspective of GNIS/My World, ICOMS agrees with it 96.9% of the time and SmartPhy agrees with it 33.3% of the time. Given that GNIS votes for itself at 100% with a weight of 35%, it also results in a confidence score of 73.3%. This can be interpreted as being a 73.3% confidence that when GNIS claims a node is named 8ANPL, that it is true.

For each node in SmartPhy, the name of the node that occurs most frequently according to the systems is also extracted. For the example of 8ANPL, with 47 votes the most frequently occurring node according to My World is 8ANPJ. In this example, ICOMS provides the exact same confidence votes as My World, but this is not always the case. Since the most frequent node name is not the same across all systems, the confidence of this alternative node is name is also calculated, which is 44% for both My World and ICOMS and 0% for SmartPhy. The result is 26.7% confidence for this alternative node name. While these calculations are not conclusive for this node, this is not always the case. In the example of 8EDG2 from Table 2, this node name is not found in My World at all while My World has a 94.8% confidence that the actual node name is 8EDG1. ICOMS exhibits a confidence of 95.9% for this node name as well. The overall weighted mean for this alternative node name is 57.1%. The reason it is not higher is because SmartPhy, the most highly weighted system, attributes a vote of 0%. While not an extremely high confidence, it is still higher than the confidence of 40% for the default node name. 40% represents the minimum confidence for any node name according to SmartPhy and is assigned to node when the other systems have 0% agreement of the node name. If the weight of SmartPhy were to decrease, this minimum confidence would decrease, and the alternative node confidence rate would increase. The alternative node name calculations are repeated two more times as well, using ICOMS and GNIS as the reference point for nodes and using the other two remaining systems to cross check its validity.

The ETL scoring approach, while effective in many aspects, faces a notable limitation in its inability to capitalize on matching based on geographic latitude and longitude data. These coordinates hold significant predictive value in determining the accurate designation of node names, given the pivotal role that geography plays in distinguishing between nodes, irrespective of nomenclature. Neglecting these coordinates in the analysis could be seen as an omission, which calls for additional methodologies to be included in the algorithm.

3.2. Haversine Distance

In the distance analysis process, we need to consider two issues: how to choose the cutoff distance, and how to calculate the similarity score.

3.2.1. Choice of the Cutoff Distance

Selecting the cutoff distance to determine location similarity depends on the specific context and requirements of your application. In our schema, the cutoff distance defines the maximum distance between two locations for them to be considered similar. The choice of the cutoff distance will influence the granularity of the clustering or grouping of cross-system nodes. A cut off distance that is too high may lead to false positive cases where two different nodes are identified as the same node. Meanwhile, a cut off distance that is too low may lead to false negative cases where same nodes in two different systems mistakenly are classified as different nodes. In our scheme implementation, the network experts' insights were used to determine an appropriate cutoff distance range.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

However, selecting the cutoff distance is not always a straightforward process, and requires some trial and error. Hence, cross-validation was used to evaluate the performance of different cutoff distances on a hold-out dataset to achieve the best performance. We start with a small cutoff distance, and gradually increase it to observe how the similarity patterns change. This incremental approach allows us to explore different levels of granularity in the node mapping. The final cut-off value based on multiple trials was set to be 1.25 km.

3.2.2. Calculation of the Node Similarity Score

The node distance similarity score is based on the ratio of the Haversine Distance and cutoff distance. Namely, we express the percentage of similarity of two nodes as follows:

$$Similarity\ Score\ (\%) = abs\ (1 - \frac{\text{Haversine Distance}}{\text{Cutoff Distance}}) * 100\%$$

Figure 6 – Node Similarity Score

Here, the Haversine Distance, derived from the equation in Figure 6 specifies the distance in kilometers between nodes from distinct systems. Meanwhile, the Cutoff Distance represents the chosen threshold to adjudicate node similarity, as informed by our iterative assessment.

In the context of node mapping, a match pertains to nodes whose coordinates are either identical or proximal within the cutoff distance range. Our operational premise posits that nodes with minimal distances between them are more likely to be identical, and vice versa.

3.3. GMM Clustering

In the realm of data clustering, the Gaussian Mixture Model (GMM) has emerged as a pivotal tool, particularly when dealing with multifaceted distributions in datasets. The GMM method functions by approximating the underlying data distribution using a combination of multiple Gaussian distributions, typically identified via the Expectation-Maximization (EM) algorithms. Our research has incorporated this GMM clustering methodology, emphasizing its role in ascertaining node similarity scores, a key metric for analyzing and understanding node-related traffic pattern and strategic node action plan.

By harnessing the GMM's capabilities and incorporating a robust range of statistical measures, we believe our method offers a holistic view of node similarity, thereby enhancing the precision and relevance of our results in node mapping.

3.4. Ensemble of the Methodologies

As previously discussed, the ETL scoring method has limitations in that latitude and longitude are not considered by it. The Haversine and GMM methodologies also have limitations, particularly considering that the boundaries of nodes are often fluid, encompassing varying distances from a few yards to several miles. This complexity is compounded by the distribution of individual customers around node centers in diverse directions, leading to instances where a customer might be closer to a node center that is not their actual assigned node.

Such intricacies render geographic coordinates insufficient in unequivocally identifying a node; however, they serve as valuable tools in the elimination of potential node candidates. For instance, in Figure 7, a hypothetical example illustrates that geography alone might not definitively determine whether a depicted customer belongs to Node B or C, even though they are in closer proximity to the center of Node B, they

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

belong to Node C. Despite its inability to precisely pinpoint node designation, geographic data does aid in excluding certain nodes from consideration. For example, in Figure 7 it can be inferred that the customer does not belong to node A.

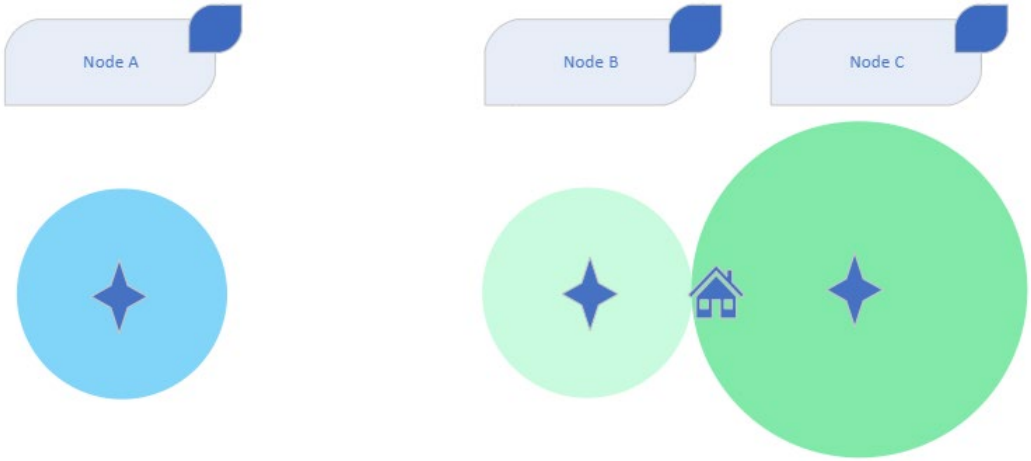


Figure 7 - Node Geographic Distribution

Due to the natural constraints of the cross geographic analysis, it is used as a supplemental analysis to either affirm or refute the conclusions from the ETL method, which was found by engineers to not always point to the correct node names. The resulting uncertainty is quantified through the confidence probabilities associated with these approaches, and when combined with the ETL method, the overall confidence score is modified. The final score is determined by taking a weighted average of all three scores, with the ETL method having a weight of 85%, the Haversine Method having a weight of 10% and the GMM method having a weight of 5%.

4. Results

In this section, we present a comprehensive demonstration of the enhanced efficacy achieved over existing best practices, substantiating our claims through discernible results and discoveries. Our objective here is to establish the superior capabilities of the ensemble learning algorithm in generating better-matched node names within the context of applied research.

4.1. Levenshtein ETL Scoring Method

Of the 5,582 Arizona SmartPhy nodes that were considered, 4,388 produced results. 3,190 nodes were found to be in complete agreement between all three systems. However, while highly correlated, agreement between systems is not synonymous with 100% confidence because agreement is defined as GNIS and ICOMS pointing to the same node name as the node name that has the most occurrences between all systems but there can still be other node names with nearly as high frequencies suggested by GNIS and ICOMS. 1,152 nodes of the SmartPhy nodes have 100% confidence and another 1,664 nodes have 90-99.9% confidence, all of which classify as agreeing. Of the nodes with under 90% confidence, only 373 of them classify as being in complete agreement.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Overall, regardless of whether a node is flagged as having complete agreement or not, there are 1,152 nodes with 100% confidence and 89 nodes with 40% confidence, which is the minimum confidence and is defined as not finding any support for the node name in the other two systems. The complete confidence distribution is displayed in Figure 8. Of these nodes, there are 54 nodes that have an alternative node suggested which have a higher confidence score than the SmartPhy node name and there are a total of 92 nodes that alternative nodes with higher confidence when combining more than one alternative node. The number of nodes that have a higher scoring alternative node name is limited given that they have a minimum confidence score of 40% due to the weights given to SmartPhy.

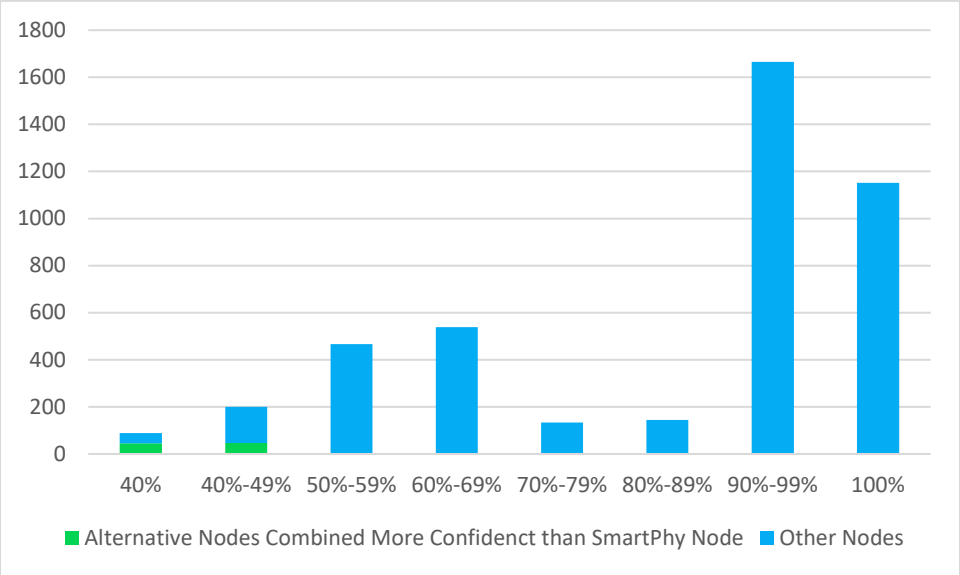


Figure 8 - Confidence Distribution of Results for SmartPhy Nodes

There are 6,483 nodes according to GNIS, which is slightly more than SmartPhy. With 377 nodes having the minimum confidence score of 35%, it is likely that there is a bit more noise in the node names in My World since there are only 89 nodes with the minimum score in DOCSIS. Of the nodes with a minimum score, 68% of them have alternative node names with a higher confidence score, suggesting the system is effective in suggesting what is likely the true node name when the default node name is in doubt. Looking towards the nodes with high confidence, there is a drop in the number of nodes that have 100% confidence from SmartPhy. In SmartPhy, these nodes are 26% of all nodes while in GNIS they are only 6% of all nodes. However, the 90-99% bin in GNIS represents 39% of all nodes, which is slightly more than the 37% of nodes in this bin in SmartPhy. Irrespective of confidence scores, 3,238 nodes, or 50% of nodes have agreement with the two other systems, meaning that the other systems cast a majority vote for the GNIS node name even if they have other alternative node names to suggest.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

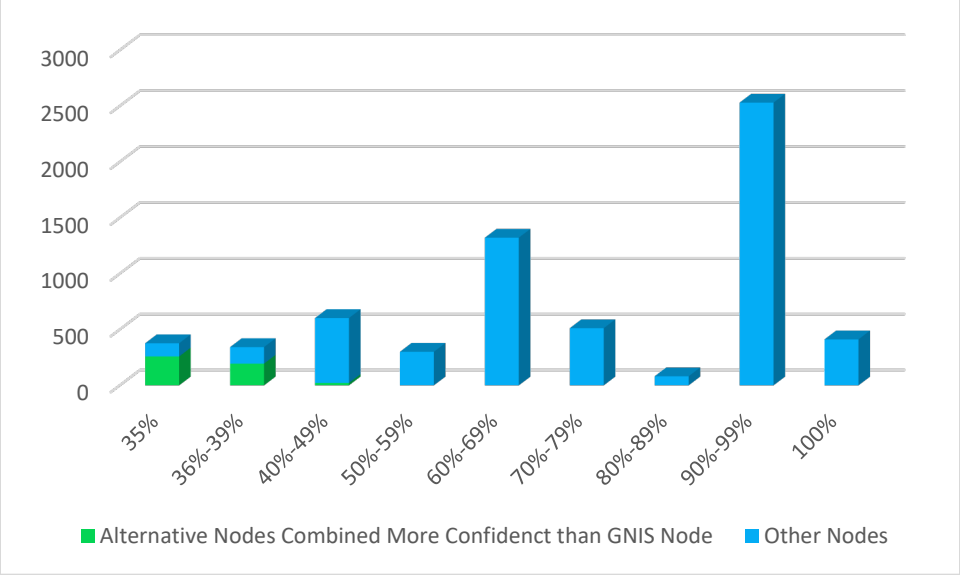


Figure 9 - Confidence Distibution of Results for GNIS Nodes

There are 10,292 nodes according to ICOMS, almost double the nodes of SmartPhy. Nonetheless, of these 10,292 nodes, 3,673 are likely noise as represented by having a percentage of 25%, which is the minimum for ICOMS. However, there are 1,037 nodes with a confidence of 25% or greater than the total amount of nodes in SmartPhy, suggesting that there are several nodes that the other systems are not capturing such as analog nodes which would not exist in SmartPhy or GNIS due to a node action. Furthermore, of 3,673 nodes with 25% confidence, the program suggests alternative node names with higher than 25% confidence in 3,091 of the nodes with 25% confidence. When examining all ICOMS nodes irrespective of their confidence score, there are 1,487 nodes whose alternative names have a confidence of 50% or greater, which is strong evidence that the system works. On the other side of the spectrum, there are 3,222 nodes where all node names are in complete agreement.

The complete confidence breakdown is displayed in Figure 10, which shows strong concentrations at the extremes of having complete confidence that the ICOMS node name is the correct node name as represented by 100% confidence as well as a confidence score of 25%, the minimum confidence, which represents 60% of all ICOMS nodes. There are not too many nodes at other confidence ranges except for a spike in the 60-69% confidence range. Of these 2,189 nodes with 60%-69% confidence, 1,874 of them have a very high confidence from GNIS and a very low confidence from SmartPhy. This may be indicative of a construction/splicing issue where the trunk and feeder cables were connected to the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

incorrect side of a 2x2 node.

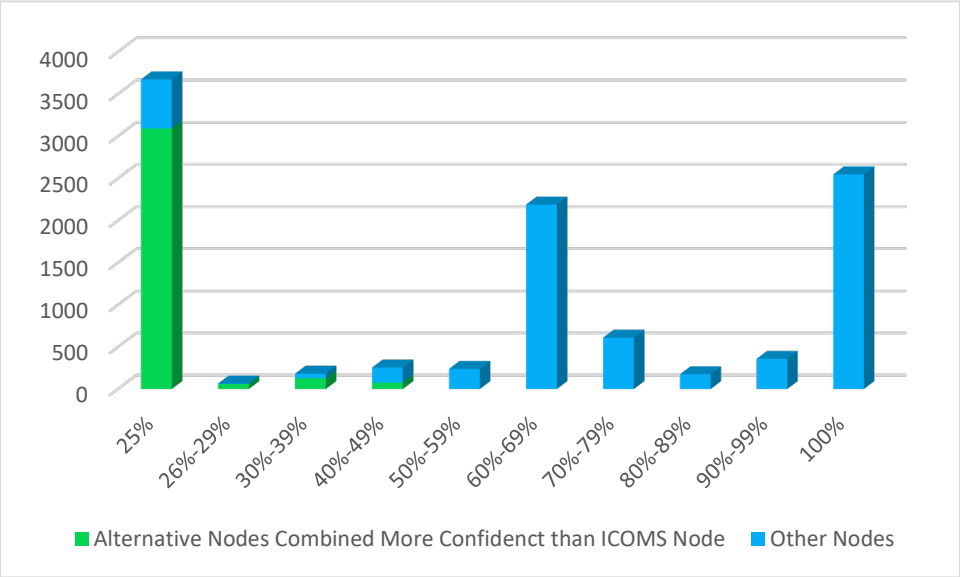


Figure 10 - Confidence Distibution of Results for ICOMS Nodes

Consistent with the weight selection, DOCSIS has the greatest node confidence followed by GNIS and finally ICOMS with 64.2%, 54.4% and 28.3% node confidence of 90% or greater in these systems respectively. DOCIS, GNIS and ICOMS have 6.6%, 20.4% and 40.4% node confidence of under 50%, respectively as well, showing consistency of the relative confidence of systems on both the high end and low end of confidence. While these confidence score differences are influenced by the weights, the gaps exceed the differences of weights between the systems which are 40% for DOCIS, 35% for GNIS and 25% for ICOMS, showing consistency with the weights and affirm the engineers' anecdotal observations about the difference of confidence between the systems.

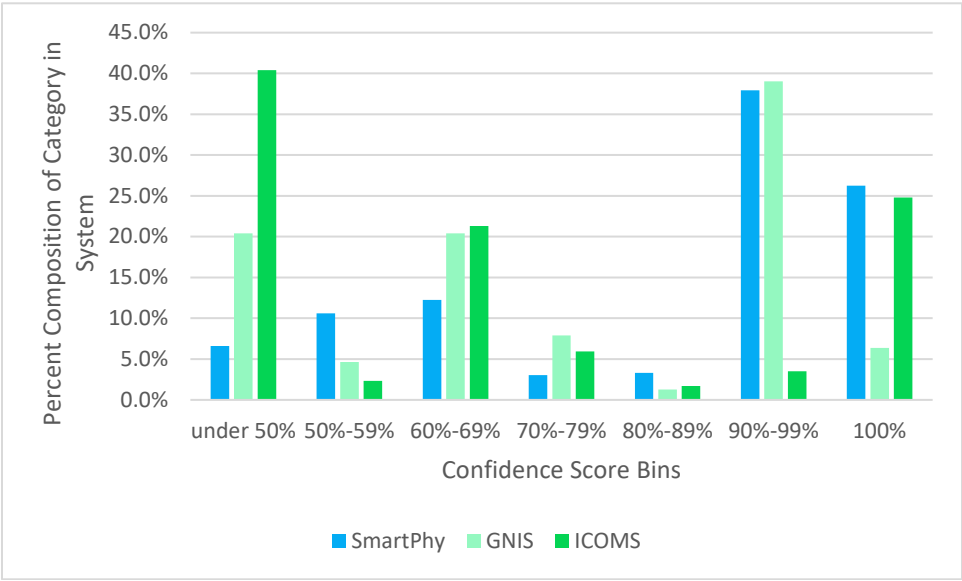


Figure 11 - Confidence Distibution Comparison Between Systems

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2. Haversine and GMM Method

As discussed in the methodology section 3.4, a scenario may emerge where each standalone method fails to provide engineers with sufficient information to definitively determine the valid node name across platforms, particularly in cases where a node bears multiple unmatched names across platforms. In response, we propose a hybrid machine learning approach to enhance confidence in the node name mapping process. To validate this proposed learning methodology, we incorporate distance matrix scores and GMM (Gaussian Mixture Model) probabilities. These scores are weighted using subject expert defined parameters across three systems, yielding distinct scoring sets for each node: the ETL/Levenshtein FuzzyString Score, Haversine Distance score and the GMM clustering score. With these scores and corresponding confidence datasets, we execute matching at the node level and tabulate the counts and percentages of node names exhibiting high similarity scores (≥ 70 out of 100) in each method. This alignment is illustrated through a Venn diagram in Figure 12, where individual methods are denoted by separate circles of distinct colors, facilitating clear differentiation. Sky-blue signifies Levenshtein FuzzyString, violet purple denotes Haversine Distance, and a slightly darker blue represents GMM Clustering.

In Table 3, we provide real life examples of nodes that are in Venn Diagrams for further clarification about what the Venn Diagrams represent.

Table 3 – Sample Nodes For Venn Diagram Clasification

Node	ETL/Fuzzy Score	Haversine Score	GMM Cluster Score
8ADBA	79.4%	88.9%	96.7%
8BXH1	100.0%	97.9%	66.7%
8AXR2	84.9%	66.5%	64.3%
8DTS1	99.4%	78.3%	91.8%
8VGR1	98.7%	90.3%	89.6%

Node 8ADBA is in the purple region of Diagram A, the dark blue region of Diagram B and the intersection region of Diagram C. Node 8BXH1 is in the intersection of A, the teal of B and the purple of C. 8AXR2 is in the teal of A, the teal of B and is not in C. 8DTS1 is in the purple of A, the blue (intersection) of B and the blue of C. Finally, 8VGR1 is in the intersection of all graphs.

In Diagram (A), out of 12,333 node names with high scores, the Levenshtein FuzzyString Method has 9,896 with high scores, the Haversine method has 6,863 with high scores and 4,426 (36%) nodes are in agreement by having high confidence scores in both methods. Similarly, in Diagram (B), out of 9,946 node names highly scored by Levenshtein FuzzyString and 5,565 with high GMM Clustering scores, both methods have high confidence in the node name in 3,672 (31%) of nodes. Diagram (C) shows that out of the 6,863 nodes with high confidence in Haversine Distance and 5,565 nodes with high scores in GMM Clustering, both methods give high confidence in the node name in 3,415 (38%) nodes.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

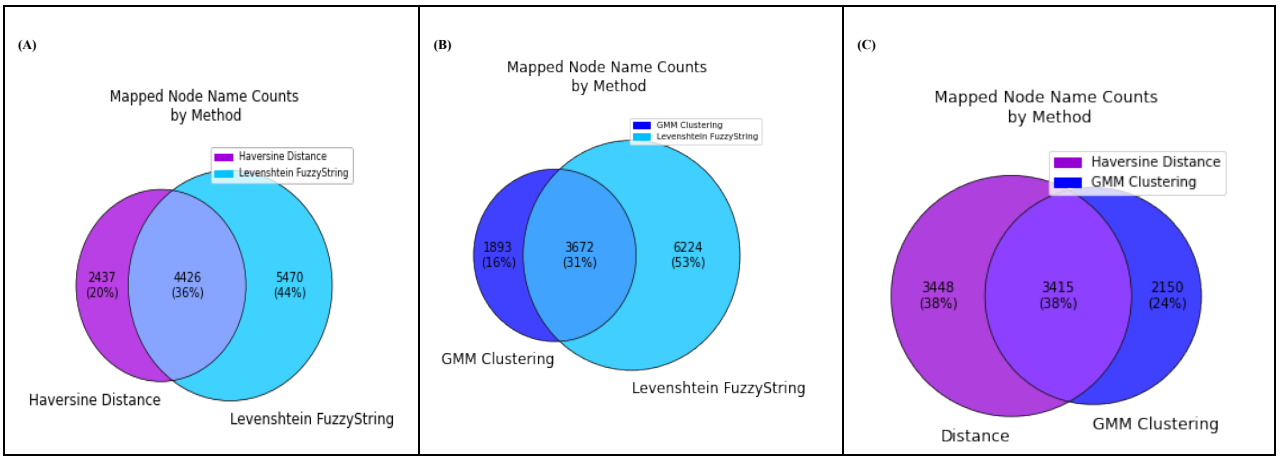


Figure 12 - Mapped Node Name Counts by Levenshtein FuzzyString, Haversine Distance and GMM Clustering

In our analysis, the overlapping region delineated by both GMM clustering and Haversine Distance is notably smaller than anticipated. We attribute this to two reasons:

(1) Amount of nodes with high scores: The count of nodes highly ranked by each method is relatively limited; the Haversine Distance method identified less than 7,000 nodes, while GMM clustering accounted for under 6,000. Nonetheless, having nearly 3,500 nodes recognized by both methods is still a significant overlap.

(2) Differences in Input Data: While both methods utilize the same raw data, there are certain disparities in the data inputs fed into each algorithm. For instance, consider the node labeled 8ARUC, which has 24 pairs of coordinates associated with it in the dataset. We engineered variables for the GMM model by calculating descriptive statistics including the mean, minimum, median, 95th percentile, and maximum of these 24 pairs, as detailed in Table 4.

Table 4 – Sample Node-Level Coordinates (eg Longitude and Latitude)

Node Name	Longitude Mean	Longitude Max	Longitude P50	Longitude P95	Longitude Min	Latitude Mean	Latitude Max	Latitude P50	Latitude P95
8ARUC	-111.927998	-111.926366	-111.927659	-111.926478	-111.932305	33.494212	33.494631	33.494228	33.494589

Only the median longitude and latitude (i.e., lon_p50 and lat_p50) were utilized to determine distance in the Haversine method. In contrast, GMM clustering employed the full array of these descriptive statistics. This discrepancy might illuminate why the overlapping region is smaller than initially projected.

4.3. Ensemble Method

While these individual methodologies demonstrate substantial alignment in node mapping, they also exhibit varying degrees of uncertainty when employed individually. To harness the strengths of each method, we propose an ensemble approach that amalgamates the three methods into a meta-learner. This ensemble

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

methodology seeks to yield a definitive consensus, and upon executing the same scoring and evaluation on equivalent datasets, we observe analogous performance results, as depicted in Figure 13.

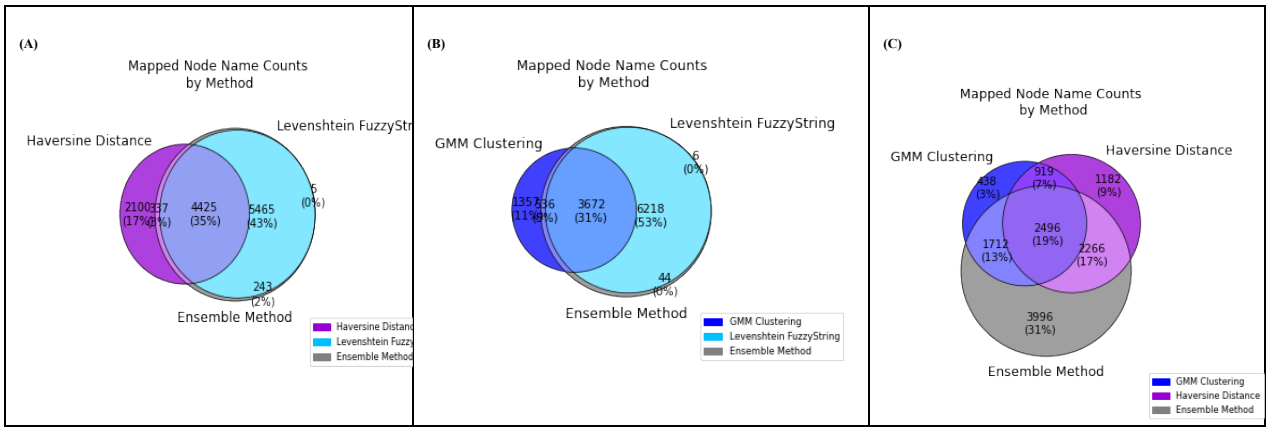


Figure 13 - Mapped Node Name Counts by Levenshtein FuzzyString, Haversine Distance, GMM Clustering, Ensemble Method

To aid in the understanding of the results, we have compiled a summary of the number of nodes captured by each method, accompanied by their respective percentages, in Table 5. It must be noted that in each Venn Diagram, it is comparing two methods with the combined method, meaning that one method is excluded in each Venn Diagram. For example, the gray area of Venn Diagram C shows that 31% of nodes belong to the ensemble method only category. However, these nodes are matches solely based on the Levenshtein Method which is not displayed in Venn Diagram C.

Table 5 – Ven Diagram Explanation

Method and Regions (Node Counts and Percentage)	(A)	Method and Regions (Node Counts and Percentage)	(B)	Method and Regions (Node Counts and Percentage)	(C)
Methods Used	(1) Haversine Distance; (2) Levenshtein FuzzyString, (3) Ensemble Method	Methods Used	(2) Levenshtein FuzzyString, (3) Ensemble Method; (4) GMM Clustering;	Methods Used	(1) Haversine Distance, (3) Ensemble Method; (4) GMM Clustering;
2,100 (17%)	Nodes with high similarity score unique to (1)	1,357 (11%)	Nodes with high similarity score unique to (4)	438 (3%)	Nodes with high similarity score unique to (4)
337 (3%)	Shared nodes with high similarity score in (1) and (3)	536 (5%)	Shared nodes with high similarity score in (3) and (4)	1,712 (13%)	Shared nodes with high similarity score in (3) and (4)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4,425 (35%)	Nodes with high similarity score that coexist in (1), (2) and (3)	3,672 (31%)	Nodes with high similarity score that coexist in (1), (2) and (3)	2,496 (19%)	Nodes with high similarity score that coexist in (1), (2) and (3)
				919 (7%)	Shared nodes with high similarity score in (1) and (4)
5,465 (43%)	Nodes with high similarity score in both (2) and (3)	6,218 (53%)	Nodes with high similarity score in both (2) and (4)	2,266 (17%)	Nodes with high similarity score in both (1) and (3)
				1,182 (9%)	Nodes with high similarity score unique to (1)
243 (2%)	Nodes with high similarity score unique to (3)	44 (0%)	Nodes with high similarity score unique to (3)	3,996 (31%)	Nodes with high similarity score unique to (3)

In Figure 13, a notable trend emerges: a reduction of 337 nodes that originally resided within the disagreement region between the Haversine Distance and FuzzString methods has been effectually shifted to the overlapping region between Haversine Distance method and our proposed Ensemble method. Although this yields a modest increase of just over 300 nodes in the shared overlapping region, the impact of this enhancement becomes more pronounced when considering the smaller population of nodes endowed with high-confidence scores in the Distance method.

Transitioning to Figure X1 diagram (B), a notable pattern emerges as 536 nodes now occupy the coveted category of targeted matched nodes within the ensemble framework. This outcome stands as a testament to the robustness and precision of the proposed approach in comparison with Fuzzy String Method.

Finally, the culmination of GMM clustering, Haversine Distance, and the Ensemble method offers intriguing insights. Upon closer scrutiny, a combined total of 5,457 nodes could be considered as nodes with high similarity score from three methods. Among these, 2,496 nodes are shared across three methods, 1,712 nodes are common between GMM and the Ensemble method, and 1,266 nodes coexist between Haversine Distance and Ensemble method. This shift exemplifies the capacity of the Ensemble method to bridge disparities and align node mappings in a more efficient manner.

5. Conclusion, Limitations and Further Research

5.1. Conclusion

With the objective of achieving automated node determination across disparate systems, we introduce a methodological framework underpinned by fuzzy postal address matching, location coordinate similarity, and unsupervised geographical clustering. This paradigm leverages an array of heterogeneous data sources,

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

encompassing mac addresses, home addresses, and longitudinal-latitude pairs, to meticulously delineate the identification of target matching nodes. Concomitantly, the methodology amalgamates ETL processes and an unsupervised machine learning model, thereby facilitating the cohesive evaluation of node similarity across heterogeneous platforms.

Through meticulously conducted experimental endeavors, the empirical findings notably underscore the appreciable amelioration in the mitigation of node misclassification discernible across distinct systems. It is very consequential in its implications and our proposed methodology demonstrates a performance trajectory that outpaces individual techniques when assessed within the context of cross-system node mapping.

The results of this paper serve as a proof of concept to demonstrate the potential of a Cross-Node Mapping System if resources are allocated to implementing it into production. By using Artificial Intelligence and modelling, we can continuously make improvements to the model and have the model self-learn to the point we can have node topology almost entirely AI-model driven and seamlessly incorporated with other tools and sources we have at our disposal once the system enters production. This will provide great benefits from a full automation and cost savings perspective.

5.2. Limitations

It is possible to improve the coverage of the system. While SmartPhy is regarded as the most reliable data source and is used as system that generates the list of nodes to consider for the system, it has a notable limitation in that only digital nodes are in it, meaning that for the purposes of this study analog nodes are excluded. It is estimated that there are 7,773 total nodes in Arizona but SmartPhy only has access to 5,872 nodes, meaning that there are approximately 1,901 analog nodes in Arizona that were excluded from this research. It is possible that these missing nodes could have associations with some of the 3,091 ICOMS nodes that appeared to be noise.

The lack of analog nodes is not the only cause of missing nodes in the study. Due to processing errors in the matching process loop, 4,388 out of 5,582 nodes were able to be considered. The code should be examined further to determine if it is capable of functioning on a portion of the 1,194 nodes that were not able to be processed.

One possible solution that can address missing nodes is to adjust the starting point of the intake process in the system. Currently, the proposed system takes the nodes that are in SmartPhy and loops through them to match to the other systems. As the ICOMS system is currently the system with the highest discrepancies, one possible adjustment is to loop through host name or interface name and feed those parameters into SmartPhy instead of simply the ICOMS or SmartPhy node name, which could be faster and capture more data. In Arizona, there are 142 CMTSs with each having 50-100 service grouping of line interfaces. Another potential approach is to start the system by looping through GNIS nodes generated directly by the design team when a node action occurs which could increase accuracy and trigger the system at the most critical times. Looping through the ICOMS nodes is another potential solution which could also provide greater coverage of nodes but could potentially be slower. Combined and shared-interface 2x2 nodes will pose a unique challenge at the moment when applying these nodes in the model as we currently cannot tell which node a modem is polling from. However, as modems on RPD nodes can be polled to know which US and DS port they are reporting back to even in the case of combined 2x2s, this will be something to refine the model logic and explore soon.

Finally, we found that latitude and longitude in the ICOMS database was not always accurate. To improve the Haversine Distance and GMM results it is necessary to improve the accuracy of these coordinates.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

This can be achieved through expanding the use of the google maps API. However, it costs $500 per each block of 100,000 requests.

5.3. Next Steps

To put the system into production, it is necessary to automate the data extraction process, which is currently done manually by querying the Oracle database where the data is housed, creating Comma-separated values files, reducing their size via bash and reading them into pandas. A possible solution is to write a series of more restrictive queries that pull data by market and create jobs to refresh the source files on a scheduled basis.

Another challenge that must be addressed is to speed up the ETL matching process, which currently takes 50 hours for Arizona alone. To scale up to the whole country, it could take one to two weeks of processing time under the current framework, which is not sustainable and means by the time the ETL process is done it could be outdated. Methods to improve the speed include a more efficient implementation of multiprocessing and further development of the fuzzy matching method, which is the subprocess with the longest compute time. Several possibilities for improvement include the exploration of more pre-filtering, a more efficient use of multiprocessing, the exploration of faster fuzzy matching algorithms and the exploration of different implementations within Python of Levenshtein distance. Studies have shown that the Jaro algorithm is 70% faster (Christen 2006) which could reduce the processing time from 50 to 15 hours. Furthermore, the program to create the confidence score takes 22 minutes in Arizona and should also take less time prior to scaling up.

Furthermore, it is possible to increase the efficacy of the Geographical Gaussian Mixture Model (GMM) spatial clustering. Particularly when applied in isolation, this mechanism evinces a diminished efficacy in scenarios involving nodes proximate to each other. A possible enhancement for future research is to construct a successive clustering scheme based on the GMM cluster results to further enhance node matching precision.

To be useful to engineers it is also necessary to establish a user interface that they can interact with to see the confidence of node names and alternative node name suggestions. A simple web application could be the first step, followed by integration into their existing systems and creating alerts to provide more easily available analysis and to allow for taking proactive measures.

Another enhancement is to introduce apartments into the string matching which is necessary for accuracy in urban areas as some apartment buildings can share the same address but have multiple nodes. Nonetheless, this has its challenges because My World has its apartments as ranges. For example, an address in My World could be 172 S Main ST 253-268 Phoenix, AZ. This single record represents all apartments in that building from 253 to 268. On the other hand, ICOM has a separate record for each of those unit numbers. The data must be adjusted, and additional preprocessing is required to be able to introduce apartment numbers into fuzzy matching.

After the successful implementation of the cross-node mapping system for informational purposes, there can be potential new opportunities to enhance customer service. The GNIS platform has a web interface called My World which has geocoding of data but is limited to data in the GNIS database. By matching to ICOMS, additional information can be made available such as if a customer is active or not, which can help improve customer service.

Another potential functionality of the system is identifying ICOMS nodes that have few customers associated with them for a given DOCSIS node which can be indicative of people who moved to another

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

jurisdiction and took their router with them. This system could aid in automatically updating their node information in ICOMS which would also improve customer service.

Abbreviations

CMTS	Cable Modem Termination System
CPE	Customer Premise Equipment
DBSCAN	Density-based clustering
DOCSIS	Data Over Cable Service Interface Specification
EM	Expectation-Maximization
ETL	Extract Transform and Load
GMM	Gaussian Mixture Model
GNIS	Geographic Name Information System.
HFC	Hybrid Fiber-Coaxial
ICOMS	Integrated Communications Operations Management System
ML	Machine Learning
NLP	Natural Language Processing
OSP	Outside Plant
RPD	Remote Phy Device
SmartPhy	Cisco SmartPhy platform for configuring RPD nodes

Bibliography & References

"Operations on Sets." *Encyclopædia Britannica*, www.britannica.com/science/set-theory/Operations-on-sets#ref387948. Accessed 2 Aug. 2023.

Charif O, Schneider M. "How to Geolocalize an Administrative File Basing on its Addresses Data? A Proposal to Achieve Better Geocoding" OMRANI H. CEPS/INSTEAD, DIFFERDANGE, LUXENBOURG

Chen, Yoke Yie, et al. "Email Hoax Detection System Using Levenshtein Distance Method." *Journal of Computers*, vol. 9, no. 2, 2014.

Christen, Peter. "A Comparison of Personal Name Matching: Techniques and Practical Issues." *Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06)*, 2006, https://doi.org/10.1109/icdmw.2006.2.

Coetzee, S., & Rademeyer, M. "Testing the spatial adjacency match of the Intiendo address matching tool for geocoding of addresses with misleading suburb or place name." Paper presented at the 24th International Cartographic Conference, Santiago, Chile, November 15–21 2009.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Filipov, L., and Z. Varbanov. 2019. “On Fuzzy Matching of Strings.” Serdica Journal of Computing 13(1–2):71–80.

Hall, P. A. V., and G. R. Dowling. 1980. “Approximate String Matching.” ACM Computing Surveys (CSUR) 12(4):381–402.

Levenshtein V. “Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. Soviet Physics Doklady10 (8); Feb 1966.p. 707–710.

Rezayan, Hani, et al. “Quality Evaluation of Postal Address Datasets Measuring Their Autocorrelation.” *GeoJournal*, vol. 84, no. 6, 2018, pp. 1617–1625, https://doi.org/10.1007/s10708-018-9940-x.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23 CELEBRATING 40 YEARS

2023 Fall Technical Forum SCTE • CABLELABS • NCTA

Best Practices for A/B Testing Machine Learning Models

An Operational Practice prepared for SCTE by

Piper Williams
Lead Data Scientist
Charter Communications
6380 S Fiddlers Green Circle, Greenwood Village, CO 80111
Piper.Williams@charter.com

Ryan Lewis
Manager, Data Science
Charter Communications
6380 S Fiddlers Green Circle, Greenwood Village, CO 80111
Ryan.M.Lewis@charter.com

Miranda Kroehl
Senior Director, Data Science
Charter Communications
6380 S Fiddlers Green Circle, Greenwood Village, CO 80111
Miranda.Kroehl@charter.com

Veronica Bloom
Senior Director, Data Science
Charter Communications
6380 S Fiddlers Green Circle, Greenwood Village, CO 80111
Veronica.Bloom@charter.com

Brock Bose
Vice President, Data Science
Charter Communications
6380 S Fiddlers Green Circle, Greenwood Village, CO 80111
brock.bose@charter.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. The Experimentation-Machine Learning Lifecycle Process ... 3
2.1. Model Discovery A/B Phase ... 3
2.1.1. Use Case Development ... 3
2.1.2. Experiment Objectives and Hypotheses ... 4
2.1.3. Experiment Design and Statistical Analyses ... 5
2.2. Productionalizing the Segment Phase ... 8
2.2.1. Validate Model Against Real-Time Data ... 8
2.2.2. Monitor Model Performance ... 8
2.2.3. Move Segments to Production ... 9
2.3. Model Implementation A/B Phase ... 10
2.3.1. Model Discovery vs. Implementation A/B Phases: Key Distinctions ... 10
2.3.2. Experiment Design and Statistical Analyses ... 10
2.4. Operationalization of Machine Learning Application Phase ... 11
2.4.1. Machine Learning Application Management ... 11
2.5. Continuous Improvement Phase ... 12
2.5.1. “Segment A” vs. “Segment B” Comparison ... 12
2.5.2. “Attribute A” vs. “Attribute B” Comparison ... 13
3. Conclusion ... 13

Abbreviations ... 14

Bibliography & References ... 14

List of Figures

Title **Page Number**

Figure 1 - Model Discovery Subgroups ... 6
Figure 2 - Regression Model for Model Discovery Analyses ... 7
Figure 3 - Results-Based Decision Framework for Model Discovery A/B Experiment ... 7
Figure 4 - 30-Day Model Performance Metric Trends Visualized in Tableau ... 9
Figure 5 - Model Implementation Subgroups ... 11

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Customization of support sites across self-service platforms can significantly enhance overall customer experience and satisfaction. At Charter Communications, one of our primary objectives in this area is to drive improvements to digital troubleshoot and support pages by optimizing digital support content, content "findability" and content rules. By customizing support content, we aim to reduce the time between our customers' questions and their resolutions, ease the navigation process for them by surfacing relevant content and eliminate the need for a customer service phone call. Machine learning offers powerful capabilities to customize content by leveraging vast amounts of data across several areas to predict support-seeking behaviors. These predictions enable us to proactively surface the right support articles pertinent to users at the right time.

It's not enough to build these machine learning models, which is a large feat in and of itself. Once a model is built and validated, monitoring model performance over time is necessary to make sure the models are performing as expected. Traditional model performance metrics, such as precision, recall and area under the receiver operating characteristic curve (AUC-ROC) allow data science teams to assess model accuracy and prevent model drift. In addition, a successful machine learning application also has a positive and significant impact on key performance indicators (KPIs), which requires experimentation (i.e., A/B testing) and robust statistical analyses. To assess the full impact of machine learning-driven experiences, it's critical to implement a systematic process that rigorously and iteratively tests, validates and optimizes these applications.

At Charter Communications, our Data Science teams have developed the Experimentation-Machine Learning Lifecycle Process, a standardized set of best practices for operationalizing machine-learning driven content rules. This process consists of five phases, which include 1) Model Discovery A/B, 2) Productionalization of the Model, 3) Model Implementation A/B, 4) Operationalization of the Machine Learning Application and 5) Continuous Improvement. In this paper, we walk through the Experimentation-Machine Learning Lifecycle Process, describe how to carry out each of the five phases for any machine learning application and provide examples for how these phases apply to our own teams' work. In doing so, we provide a set of guidelines that can be implemented by others to evaluate and optimize their machine learning applications.

2. The Experimentation-Machine Learning Lifecycle Process

2.1. Model Discovery A/B Phase

The first phase in the Experimentation-Machine Learning Lifecycle Process is the Model Discovery A/B Phase, where the primary goal is to assess the added value of customization for the target segment compared to other customers in the population. The first step is to develop a use case, which includes defining a relevant customer target segment and constructing a corresponding machine learning model that accurately predicts the customer target segment of interest. Our next step is to lay out the experiment objectives and statistical hypotheses needed to evaluate the impact of the variant on the full population and corresponding subgroups. Finally, we implement the A/B experiment to execute these analyses.

2.1.1. Use Case Development

A successful machine learning application requires purposeful development, and a strong use case helps to provide a clear purpose and direction for these applications. It also aligns the machine learning applications with specific business goals and customer needs. We recommend teams gather input from subject-matter experts that are a part of the stakeholder teams they are collaborating with to begin

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

building their use case. Additionally, exploratory data analyses that provide insight into customers' actions and their engagement patterns with the product/platform would supplement this by identifying areas that could be targeted by the machine learning application. This should then be weighed with what data is available and what machine learning models are feasible to develop that support the use case. Ultimately, defining the use case up front ensures teams are starting off on the right foot by answering preliminary questions such as:

- What is the specific problem that needs to be addressed?
- Who are the target users/beneficiaries?
- What data and resources are needed for development and implementation of the machine learning application?

At Charter Communications, our teams' use cases are centered around the support-seeking behaviors of our customers, or in other words, the information we think the customer needs that they might be having difficulties finding on their own. We collaborate with the Digital Service and Customer Experience (DSCX) team to explore potential segments of customers with shared support needs. Once those segments are defined, we determine what support content is relevant for those customer target segments that we'd like to test surfacing in the experiment. Together, the customer target segment and corresponding support content make up the machine learning-driven content rule.

2.1.2. Experiment Objectives and Hypotheses

The next step is to outline the objective of the Model Discovery A/B experiment, which subsequently informs the hypotheses. The objective provides context behind the overall goals of the experiment and should explain how the variant can solve the business problem at hand or address a specific need. A well-formulated objective statement may be structured like the following:

"[Variant] will accomplish [outcome] because [rationale]."

The variant is the new "feature" we are testing against the control (i.e., current "feature"). The outcome is the anticipated impact the variant will have on our KPIs. The rationale explains how the variant will impact our KPIs and, consequently, address the business outcome. For the Model Discovery A/B Phase, there are three objectives that need to be assessed:

1) [Variant] will accomplish [outcome] for the ***entire population*** because [rationale].
2) [Variant] will accomplish [outcome] for the ***customer target segment*** because [rationale].
3) [Variant] will have a more pronounced effect for those within the customer target segment compared to those not in the segment because [rationale].

In this phase of the lifecycle process, the goal is to understand the impact of the variant on the full population, which includes both the customer target segment (i.e., those that are predicted by the machine learning model) and those not in the segment (i.e., those that are not predicted by the machine learning model). In addition, it's critical to understand the impact of the variant on the customer target segment alone and compare the differential impact of the variant between the customer target segment and those not in the segment. The final comparison allows us to assess the incremental "lift" of the variant on the customer target segment.

Once the experiment objective is known, this can be translated into hypotheses for statistical testing. A good hypothesis states a clear, falsifiable relationship or outcome and should directly relate to the overall objective of the experiment. Statistical hypothesis testing is used to determine whether differences seen

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

between two or more groups are likely to be real differences, rather than a difference due to random chance. A hypothesis is often stated in the form:

"If [variant], then [results]."

The variant is the new "feature" we are testing against the control, and the results are the predicted change in our KPIs. This can then frame the null and alternative hypothesis. The null hypothesis typically reflects no impact or difference between control and variant, whereas the alternative reflects the impact we are trying to achieve. When translating the three primary objectives of this phase into hypotheses, we get the following:

1) If [variant], then [results] for the ***entire population***.
2) If [variant], then [results] for the ***customer target segment***.
3) If [variant], then the customer target segment will have a ***significantly larger difference*** in [results] compared to those not in the segment.

In the context of our teams' work, our three primary objectives are:

1) Surfacing [support content] to the [support page] will reduce calls for the ***entire population*** because surfacing support content that helps our customers to self-service efficiently provides them relevant information and resolutions for their inquiries.
2) Surfacing [support content] to the [support page] will reduce calls for the ***customer target segment*** because surfacing support content that helps our customers to self-service efficiently provides them relevant information and resolutions for their inquiries.
3) Surfacing [support content] to the [support page] will have a more pronounced effect for the customer target segment compared to those not in the segment because the surfaced support content is pertinent to the customer target segment's reasons for seeking support.

The corresponding hypotheses are:

1. If we surface [support content] to the [support page], then there will be a reduction in customer calls for the ***entire population***.
2. If we surface [support content] to the [support page], then there will be a reduction in customer calls for the ***customer target segment***.
3. If we surface [support content] to the [support page], then there will be significant difference in customer calls between control and variant for the customer target segment compared to those not in the segment.

2.1.3. Experiment Design and Statistical Analyses

The final steps in the Model Discovery A/B Phase are to design the experiment and execute the statistical analyses. Finalizing the experimental design includes implementing a randomization algorithm to randomly assign customers to the control or variant, determining the sample size (and ultimately, experiment duration) required to be adequately powered, mapping out the statistical analyses and creating a decision framework that allows teams to take action after drawing conclusions. At Charter Communications, our company has an in-house experimentation platform that enables us to run A/B tests. Within our experimentation platform, experimenters can define elements such as control and variant payloads, allocation units to determine the level of randomization and inputs required to determine the needed sample size. From there, we can readily launch an A/B test and pull this experiment data for our ad hoc analyses.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

As discussed previously, the three hypotheses to test in the Model Discovery A/B Phase measure 1) the impact of the variant on the full population, 2) the impact of the variant on the customer target segment and 3) the differential impact of the variant on the customer target segment compared to those not in the segment. Figure 1 below visually summarizes the different subgroups needed to answer these hypotheses:

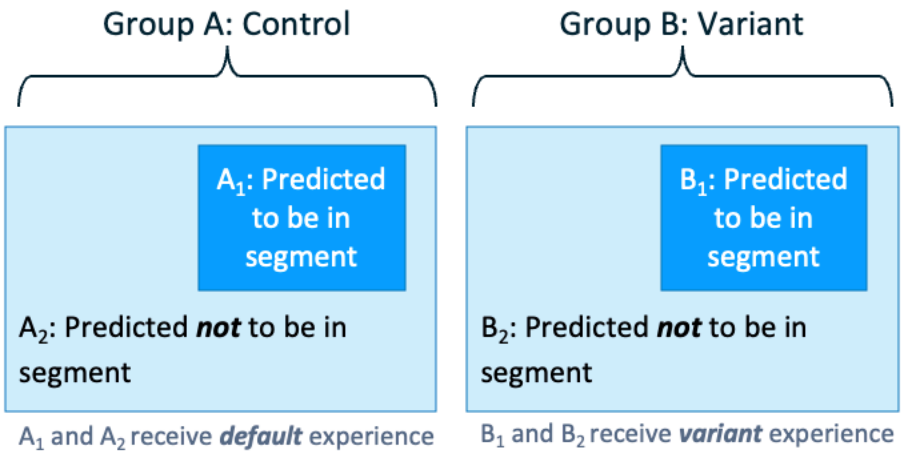


Figure 1 - Model Discovery Subgroups

In the case of two-variant experiments, groups A and B make up the entire population. Groups A and B can be further broken out into A_1/B_1 and A_2/B_2, each of which correspond to the individuals a part of the customer target segment and those not in the customer target segment, respectively. As a result, the three statistical comparisons to analyze and the methods to test these comparisons are:

Population-Level Impact of the Variant:

- H_0: $A - B = 0$
- H_A: $A - B \neq 0$
- *Statistical Test*: Two-sample t-test

Customer Target Segment-Level Impact of the Variant:

- H_0: $A_1 - B_1 = 0$
- H_A: $A_1 - B_1 \neq 0$
- *Statistical Test*: Regression model with an interaction term and contrast statements to conduct pairwise comparisons

Differential Impact of the Variant for the Customer Target Segment vs. Non-Target Segment:

- H_0: $(A_1 - B_1) - (A_2 - B_2) \leq 0$
- H_A: $(A_1 - B_1) - (A_2 - B_2) > 0$
- *Statistical Test*: Regression model with an interaction term and contrast statements to conduct pairwise comparisons

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In the case of a two-variant experiment, the regression model outlined in Figure 2 below can be used to evaluate the customer target segment-level impact and differential impact for the customer target segment vs. the non-target segment. We recommend teams consult with a Data Scientist to help conduct these analyses.

$$Y_i = \beta_0 + \beta_1 X_{Variant_i} + \beta_2 X_{Predicted_i} + \beta_3 X_{Variant_i} * X_{Predicted_i} + \varepsilon_i$$

Y_i: Outcome (i.e., KPI / Primary Success Metric) for the ith observation

β_0: Intercept (i.e., expected value of outcome for customers not in the target segment that were allocated to the control)

β_1: Slope for customers allocated to the variant

β_2: Slope for customers predicted to be in the target segment via the machine learning model

β_3: Interaction term

$X_{Variant_i}$: Indicator if i^{th} observation was allocated to the variant

$X_{Predicted_i}$: Indicator if i^{th} observation was predicted to be in the target segment via the machine learning model

ε_i: Residual error for the ith observation

Figure 2 - Regression Model for Model Discovery Analyses

The final step of this phase is to develop a results-based decision framework. The decision framework lays out all potential scenarios in which teams would declare success or failure of the variant(s). These scenarios should correspond with the selected experiment type and the burden-of-proof needed to justify releasing the change. For example, if the experiment is a move-the-needle A/B, the decision framework should reflect the requirements to demonstrate a statistically significant improvement in KPIs to justify the release. For the Model Discovery A/B Phase, a decision framework would look similar to the following:

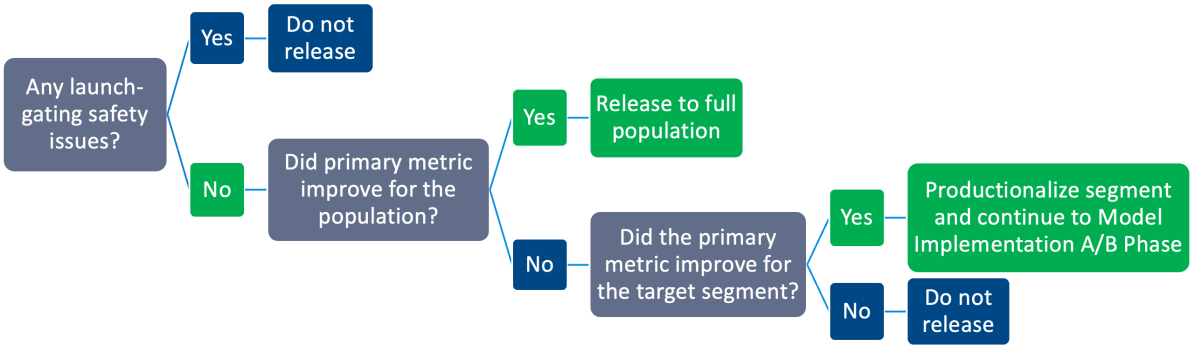


Figure 3 - Results-Based Decision Framework for Model Discovery A/B Experiment

This decision framework may be adjusted to include more steps as needed, although we recommend keeping the number of potential scenarios in a decision framework as minimal as possible to provide the most straightforward set of rules that is feasible.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.2. Productionalizing the Segment Phase

Following the Model Discovery A/B Phase, the next step is to validate the model and put the model predictions/segment into production. In this phase, it's important to test the model's accuracy on real-time data, monitor several model performance measures and define acceptance criteria for the corresponding model performance indicators. Once it's confirmed that the model is performing as expected, the model is productionalized, meaning a machine learning pipeline is constructed and model predictions that define the customer target segment are stored.

2.2.1. Validate Model Against Real-Time Data

If a certain level of model performance from the training and validation sets is expected, it's important to test whether the model, applied to real-time customer data, maintains that performance or not. To evaluate the model's predictions against customer data, we develop a pipeline that aggregates features in real time and then applies the machine learning model to those features. This outputs prediction data, which are then captured, stored and joined to the customer data post-mortem. By joining the prediction data onto customer data, teams can validate the accuracy of the model in predicting the desired customer target segment/customer behavior.

2.2.2. Monitor Model Performance

Regular monitoring to assess the model performance over time is critical for several reasons such as tracking seasonal trends in the model's predictive power, informing teams when to retrain the model and estimating the overall reach of the model. As part of productionalizing the model, it's important to consider which model performance metrics to examine. A few examples of key performance metrics include:

- ***Precision***: Proportion of correctly predicted positive instances
- ***Recall***: Proportion of actual positive instances correctly identified by the model
- ***Accuracy***: Overall correctness of the predictions
- ***F1-Score***: Harmonic mean of precision and recall
- ***Baseline Target Rate***: The baseline KPI for the full population (in the case of our teams' work, this would include baseline customer call rate for all customers visiting the support sections on Charter's self-service platforms)
- ***Lift Over Baseline*****:** Precision/Baseline Target Rate (i.e., an estimate of the incremental predictive power of the model over baseline)
- ***Percentage of Accounts in Prediction***: Percentage of customers predicted to be in the customer target segment against the entire experiment population (i.e., the overall reach of the model)
- ***AUC-ROC***: Probability that the model ranks a random positive example higher than a random negative example

Oftentimes, it's necessary to monitor multiple metrics to capture the comprehensive performance of the model. Selecting the appropriate model performance metrics depends on the specific problem being solved, the nature of the data and the overall objectives of the application. Following metrics selection, teams should determine how to effectively summarize their model performance measures. Our teams leverage reporting tools such as Tableau to visualize the daily and weekly trends of the model performance metrics. Ultimately, all summaries and visuals should be clear, informative and help stakeholders to understand the model's performance and possible areas for improvement.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

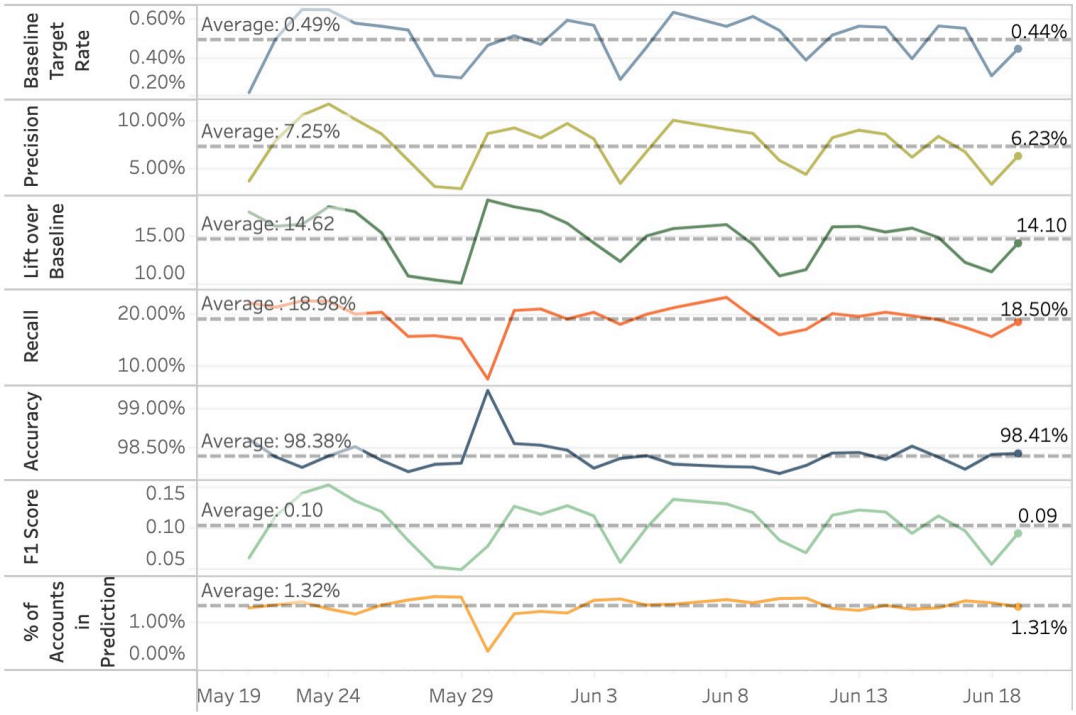


Figure 4 - 30-Day Model Performance Metric Trends Visualized in Tableau

Teams must consider which metrics the model should maximize and the "right" performance thresholds. Depending on the specific objectives of the machine learning application, the context of the problem, consequences of different outcomes and outcome prevalence, these decisions will likely differ across companies, as well as across products within each company.

2.2.3. Move Segments to Production

Once the model performance is fully validated, the segment is released into production. Through a series of well-defined steps, the model is transitioned from development to production. The core steps of a successful machine learning pipeline include data preparation, feature engineering, model training, prediction generation and storage. Data preparation consists of data preprocessing and data cleaning. Feature engineering aims to represent the data in a suitable format for the model, increasing the model's ability to make accurate predictions. Once the infrastructure is in place, the machine learning model is deployed to a production server or cloud platform. From there, the model receives real-time data and generates predictions at scale. The prediction data is typically saved in a database, flat file or cloud storage format. Storing the prediction data from the model allows for easy access and retrieval of the output.

The MLOps team within our organization orchestrates this entire pipeline and utilizes our company's in-house experimentation platform and a corresponding domain service to transfer stored predictions to our front-end experimentation tool. The model pipeline runs and generates daily predictions for the customer target segment. This prediction data is then stored and made available within our experimentation platform.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.3. Model Implementation A/B Phase

Once the machine learning models are in production, the next phase of the Experimentation-Machine Learning Lifecycle Process is the Model Implementation A/B Phase. In this phase, our primary goal is to evaluate the performance of the segment by implementing the customized experience as it would behave in real time. Additionally, this phase ensures the experience is optimized before updating the content rules to a new business-as-usual (BAU). This phase also utilizes A/B testing to assess the impact of the change for the customer target segment within the population. In this section, we will discuss the key differences between the Model Discovery and Model Implementation A/B Phases. We then review the main elements of this phase's experiment design and statistical analyses.

2.3.1. Model Discovery vs. Implementation A/B Phases: Key Distinctions

While an A/B experiment is performed for both the Model Discovery and Model Implementation A/B Phases of this lifecycle process, the experiments themselves are different from one another. The distinctions between these two phases are:

- ***Static vs. Dynamic Variant Allocation***: In the Model Discovery Phase, once a customer is allocated into the control or variant, their allocation remains constant throughout the duration of the experiment. However, this is not representative of the true experience. Depending on their prediction for a given day, a customer is either provided the default experience (i.e., control) or the customized experience (i.e., variant). In the Model Implementation Phase, we are mimicking the experience as if it were operationalized for our customers, meaning on any given day, a customer could be predicted by the machine learning model to be in the target segment or not and, thus, would be allocated to the control or variant accordingly.

- ***Variant Allocation Across Subgroups***: In the Model Discovery Phase, there are four subgroups that are analyzed in the experiment: 1) those in the customer target segment who are randomly allocated to the control (i.e., group A_1), 2) those in the customer target segment who are randomly allocated to the variant (i.e., group B_1), 3) those ***not*** in the customer target segment who are randomly allocated to the control (i.e., group A_2) and 4) those ***not*** in the customer target segment who are randomly allocated to the variant (i.e., group B_2). However, in the Model Implementation Phase, the B_2 subgroup receives the default experience that the control (i.e., Group A) receives.

- ***Questions Answered***: The questions each phase answers are different from one another. While the Model Discovery Phase assesses the impact of the variant for the customer target segment vs. the full population, the Model Implementation Phase assesses the impact of the variant for the customer target segment *within* the population. In other words, the Model Discovery Phase tests if the variant presents an added benefit for those a part of the customer target segment. During the Model Implementation Phase, we are evaluating the impact of the segment implemented dynamically as it will be in production, and asking whether implementation of this set of rules is optimized for the customer target segment or not.

2.3.2. Experiment Design and Statistical Analyses

In the Model Implementation A/B Phase, the same hypotheses that were tested in the Model Discovery A/B Phase outlined on Page 6 will be tested and analyzed. However, the questions answered from each set of hypotheses are different compared to the questions answered in the Model Discovery Phase due to

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the difference in the variant allocation process. The subgroups and their variant allocations are shown in Figure 5 below:

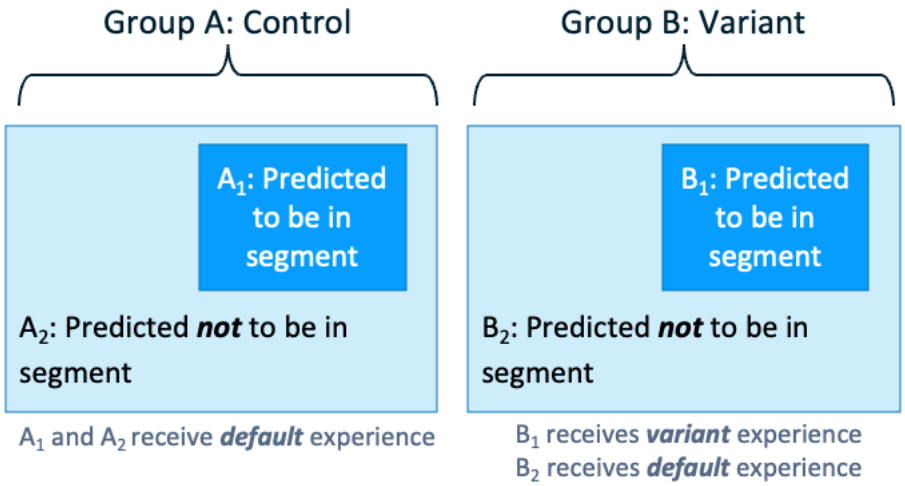


Figure 5 - Model Implementation Subgroups

As discussed previously, the main objective of the Model Implementation experiment is to test if the variant is optimized for the customer target segment within the population. It's also worth mentioning that the default experience can include multiple variants and multiple machine learning applications.

2.4. Operationalization of Machine Learning Application Phase

Following confirmation that the machine learning application (or in the context of our teams' work, the machine learning-driven content rule) is optimized and has a significant and positive impact on the customer target segment, the next step is to operationalize the machine learning application. In this phase, we update the BAU experience for the customer target segment. At Charter, updating the BAU experience includes implementing the machine learning-driven content rule that surfaces relevant support content to customer target segments predicted via the model. After operationalization, the models, their application and their corresponding business performance is monitored over time. There are a variety of methods teams can implement to monitor the business performance of the model and, ultimately, the application itself. In this next section, we outline the importance of managing the models and applications and propose two strategies teams can implement to holistically evaluate the application.

2.4.1. Machine Learning Application Management

Monitoring the impact of machine learning applications on KPIs is crucial for several reasons. First, managing machine learning applications through monitoring of KPIs allows teams to evaluate the overall business performance and effectiveness of the application. Machine learning applications are deployed to improve specific business outcomes and should continue to align with primary business objectives.

Second, monitoring KPIs can help identify potential issues such as data drift or model degradation. If there are significant shifts in KPIs over time, it may be a signal that the input data has changed, or the model needs to be retrained. By tracking the business performance of the application over time, teams can better detect early signs of model inaccuracies and take corrective actions as needed. In addition, monitoring KPIs provides valuable insights for making data-driven business decisions. This enables

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

stakeholders to make informed choices regarding resource allocation, project prioritization and future investment.

Proper management of the machine learning application requires consistent performance monitoring for the machine learning model and the business impact of the model application. The model monitoring processes outlined in the productionalization phase continues in this phase as well. It's also critical to keep track of the business performance of the machine learning application (i.e., the application's impact on KPIs). This requires real-time monitoring of KPIs associated with the machine learning application and both short and long-term evaluation of the impact of the application on business outcomes.

- ***Real-Time Monitoring of KPIs***: Creating dashboards that efficiently summarize and visualize KPIs over time, defining threshold(s) of acceptance and developing a notification system that alerts teams when the corresponding KPIs fall outside of those thresholds enables teams to proactively manage their operationalized machine learning applications.

- ***Short and Long-Term Evaluation of the Application***: By comparing business performance with an established baseline over short and long-term timespans, teams can determine if the performance of the model is consistent, improving or deteriorating. If teams find their KPIs are not performing as expected, teams can investigate the model, identify potential improvements to the model or attributes of the application and retest those changes to improve upon the current state.

2.5. Continuous Improvement Phase

The ultimate goal of Experimentation-Lifecycle Process is to make certain the deployed machine learning models and their applications are truly optimized for customers. In the final stage, the Continuous Improvement Phase, the primary objective is to iterate to continue improving the customized application/experience for customers. Continuous improvement is essential for machine learning applications for several reasons such as optimization of application effectiveness, drift detection and mitigation, maximization of business impact and identification of new application opportunities. Once a need for improvement is identified, updates to the model or application of the model are reevaluated via A/B testing.

2.5.1. "Segment A" vs. "Segment B" Comparison

The "Segment A" vs. "Segment B" comparison tests updates to the machine learning model itself. Some examples of model updates include engineering new input features, ensembling models or refining the definition of the target segment altogether and retraining the model. If the current model shows signs of drifting, the updated model can be tested against the current model to make sure the application and its impact on metrics are back on track. For example, let's say our teams previously operationalized a machine learning-driven content rule that surfaces a support article that helps users troubleshoot their internet connection for a customer target segment that is predicted to seek support for internet-specific issues (more specifically, call for internet-specific issues). In doing so, calls for the customer target segment are significantly reduced. However, our teams would like to assess the impact of adding new features to the model that are related to a customer's internet service. The addition of these features increases the F1 score. When we test the model with the added features compared to the old model, we find that the new model significantly reduces calls for the new customer target segment compared to the current customer target segment. As a result, the new model and corresponding customer target segment is operationalized and becomes the new BAU experience. For this comparison, we test the two segments via a Model Implementation experiment to understand the impact of the segments within the population.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.5.2. "Attribute A" vs. "Attribute B" Comparison

The "Attribute A" vs. "Attribute B" comparison tests updates to the application of the models themselves. In this case, there are no changes to the underlying machine learning model, but instead changes made to content, experience, etc. that are customized for the customer target segment predicted from the model. To use the previous example explained above, let's say we identify an opportunity to drive down customer calls even further by rewriting the support article surfaced to the customer target segment. An "Attribute A" vs. "Attribute B" comparison evaluates the impact of the rewritten support article surfaced to the customer target segment vs. the current support article. If results show a significant reduction in customer calls, the new content would be surfaced to the same customer target segment in place of the current content. In this case, because the machine learning model and associated segment were previously validated, a comparison of the two attributes would entail a simpler experiment where we'd assess the impact of "Attribute A" vs. "Attribute B" on the customer target segment only (i.e., A_1 vs. B_1 comparison).

3. Conclusion

In this paper, we outline best practices for A/B testing machine learning models and their applications. To ensure the applications developed are truly optimized for users of a company's products and platforms, we recommend implementing a systematic process that rigorously tests machine learning models and the business impact of their applications. The Experimentation-Machine Learning Lifecycle Process provides a general framework that companies can utilize to validate their models, evaluate whether the models' applications have a significant and positive impact for the customer target segments and business as a whole and identify areas for improvement.

While our teams apply this process to customize support content for Charter's customers, this framework is readily extensible to other areas. In the telecommunications industry, areas of potential application range from machine learning-driven segments and rules for proactive customer communications to television content recommendations. This lifecycle process creates a standardized workflow that guarantees our teams adhere to key best practices and, as a result, are truly optimizing the customization of Charter's products and platform experiences. By implementing a similar lifecycle process to test and validate machine learning applications, organizations can improve their product offerings and improve the experience for every customer.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AUC-ROC	area under the receiver operating characteristic curve
KPI	key performance indicator
DSCX	Digital Support and Customer Experience
BAU	business as usual

Bibliography & References

(2022, June 16). *Machine Learning Metrics: How to Measure the Performance of a Machine Learning Model*. Altexsoft: Software R & D Engineering. https://www.altexsoft.com/blog/machine-learning-metrics/

Ameisen, E. (2020). *Building Machine Learning Powered Applications*. O'Reilly Media, Inc.

Bright, J. (2021, July 09). *Dynamic A/B testing for machine learning models with Amazon SageMaker MLOps projects*. AWS. https://aws.amazon.com/blogs/machine-learning/dynamic-a-b-testing-for-machine-learning-models-with-amazon-sagemaker-mlops-projects/

Covalucci, V. (2020, May 14). The value of A/B testing in the era of machine learning. Capital One. https://www.capitalone.com/tech/machine-learning/a-b-testing-era-of-machine-learning/

Einblick Content Team. (2022, December 21). *MLOps: a guide to machine learning model management*. Einblick. https://www.einblick.ai/blog/machine-learning-operations-guide/

Kolltveit, A.B. and Li, J. (2022) *Operationalizing Machine Learning Models - A Systematic Literature Review*. IEEE/ACM 1st International Workshop on Software Engineering for Responsible Artificial Intelligence (SE4RAI). Pittsburgh, PA. doi: 10.1145/3526073.3527584.

Kuhn, M. and Johnson, K. (2013). *Applied Predictive Modeling*. Springer New York.

Steyerberg, E.W. (2009). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer New York.

Syed, W. (2022, February 14). *The Most Important Metric for Machine Learning Models*. Medium. https://towardsdatascience.com/the-most-important-metric-for-machine-learning-models-2c6a4c4b18ad

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

BGP Info Over DNS

A Technical Paper prepared for SCTE by

Tony Tauber
Engineering Fellow
Comcast

Charlie Helfinstine
Senior Principal Engineer
Comcast

Mark Feldman
Senior Principal Engineer
Comcast

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. The Problem – Sharing BGP (Border Gateway Protocol) information ... 3
   - 2.1. Internet routing basics ... 3
   - 2.2. Network Operations basics ... 4
3. Current Solutions for sharing BGP routing information ... 5
4. DNS (Domain Name System) is standard and universally available ... 7
5. Solution – BGP info served via DNS ... 9
   - 5.1. Delegation ... 10
   - 5.2. Governance ... 11
   - 5.3. Structure of information ... 13
   - 5.4. Caching ... 13
   - 5.5. Performance ... 13
6. More Details ... 13
7. Conclusion ... 14

Abbreviations ... 15

Bibliography & References ... 15

List of Figures

Title **Page Number**

Figure 1 – Router network core with users and servers at the edges ... 4
Figure 2 – Example BGP information retrieved via CLI ... 5
Figure 3 – Example BGP information details from Route Servers ... 6
Figure 4 – Example Looking Glass Web Interface ... 7
Figure 5 – DNS resolver iteratively queries more specific authorities to find an answer ... 8
Figure 6 – Using “dig” utility to look up a DNS record ... 8
Figure 7 – Using “dig” utility to look up a DNS TXT record ... 9
Figure 8 – The delegated structure of the DNS name space ... 11
Figure 9 – The delegated structure of the DNS “reverse” name space for IPv4 addresses ... 12

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Quickly knowing what is happening across one's network is a crucial, if challenging, task for the scale of any given modern network, particularly large service provider networks. Getting the BGP (Border Gateway Protocol) routing information about how a given node would reach a particular destination and comparing it against the same query gathered from various other vantage points can be transformational.

Logging into routers one at a time to execute the specific command for the given vendor and operating system is tedious, error-prone, and needs to be automated to reduce friction. Also, some groups can benefit from this information without needing general access to the devices. Generally, such access is, understandably, under tight control.

Our novel scheme exposes BGP routing information via DNS (Domain Name System) queries, which can be structured to query the information for a given router node or location. Network Operations staff can thus quickly gather details from around the network and synthesize it for the matter at hand. If a specific route or set of routes is of interest, one could retrieve that information regularly as a form of monitoring. The data sources inside a given operators network (BGP-speaking routers, Geolocation data stores, etc.) can push data to the DNS server(s) unidirectionally, meaning they are not subject to risk of compromise by outside users.

Furthermore, the delegated structure of the DNS namespace allows each network operator to tune the visibility of information according to their needs. Existing techniques for optimizing performance of DNS or other similar services (e.g., load-sharing, horizontal scaling) can be brought to bear.

Our paper and presentation will elaborate on this approach and explain the extensible nature of the technique. This technique of information sharing within and among Internet operators will make troubleshooting and problem resolution quicker and more accessible while maintaining infrastructure security and reliability.

2. The Problem – Sharing BGP (Border Gateway Protocol) information

To understand the need for gathering BGP information, we will give a brief overview of Internet routing.

2.1. Internet routing basics

The basic function of the Internet is to allow for the transfer of data. Consider a user retrieving information from a website. The user sends a request to the web server indicating what data they would like, and the server sends the data back in return. Each piece of this bi-directional flow of traffic between two endpoints has a "source" and a "destination". In the initial request from the user to the server, the user is the "source", and the server is the "destination". When the reply is returned, the server is the "source", and the user is the "destination".

The data for each the request and reply is encoded in one or more IP (Internet Protocol) packets (sometimes called "datagrams"). Between the user and the web server will be some number of intermediate "router" nodes which have some information about how to get to the destination.

Figure 1 shows an example national Internet backbone topology for purposes of illustration.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

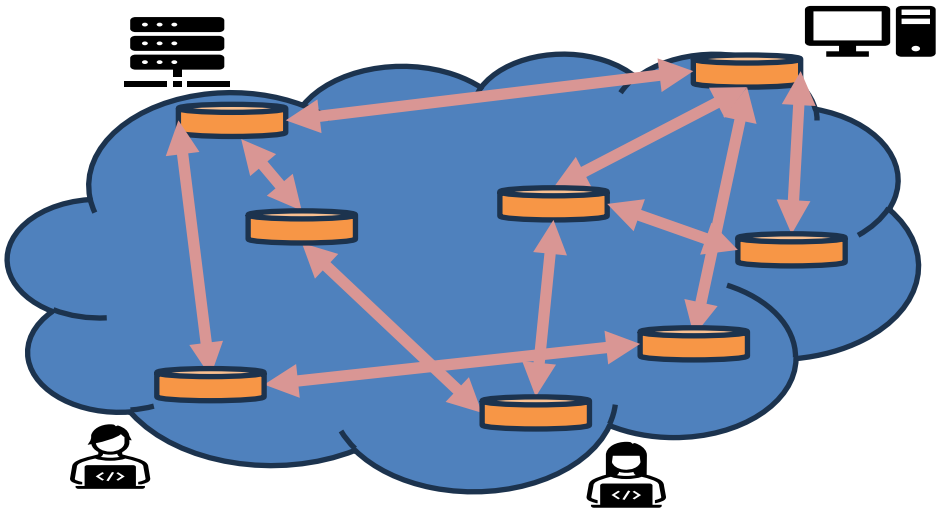


Figure 1 – Router network core with users and servers at the edges

When an IP packet arrives at a router for processing, the router examines the IP address in the "destination" field of the packet and searches its routing table for the "prefix" which is the best match to the destination address. This "answer" is known as the "Longest Prefix Match," sometimes abbreviated as "LPM" and the answer will be the "next-hop" to the destination. That "next-hop" is another router and this process continues until the last router before the destination (e.g., web server) is reached and the packet is finally delivered.

This routing and delivery process can be thought of much like a letter or package is delivered between cities or even countries using the postal network of postal stations and roads. Each "hop" along the way may only have general knowledge of how to reach a bit closer to the destination, getting more specific as things progress, e.g., Country to Country, State to State, City to City, City to Street, Street to Building.

2.2. Network Operations basics

In the course of network operations, it is often necessary to gather certain pieces of information about or from many different points and parts of a network.

The context could be within the infrastructure of a given operator, or could be from some other network(s) comprising the broader Internet. The reason often involves interrogating how a given router (and the user data traversing that router) would reach a given IP destination. Most often, this investigation happens in the case of troubleshooting a reported problem.

The traditional way of retrieving BGP routing information is via command-line interface to one or several network devices. Often it is necessary to gather the information quickly, and possibly submit it to further processing and cross-referencing.

Several problems arise from the need to use the command-line interface.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- The command structure and output format of these interfaces is non-standard and can vary by vendor, model, and software version. Figure 2 depicts an example of such CLI (Command-Line Interface) command output:

```
RP/0/RP0/CPU0:cs01.doraville.ga.ibone#show bgp 13.0.40.0/23
Mon Jul  3 03:27:04.080 UTC
BGP routing table entry for 13.0.40.0/23
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker          529060364   529060364
Last Modified: Jun 30 05:24:54.926 for 2d22h
Paths: (6 available, best #2)
  Advertised IPv4 Unicast paths to peers (in unique update groups):
    96.110.43.82    96.110.42.210   68.86.94.162
  Path #1: Received by speaker 0
  Not advertised to any peer
  15133 1321 33631, (received & used)
    96.109.22.227 (metric 4296) from 68.86.1.0 (96.109.22.227)
      Origin IGP, metric 0, localpref 275, valid, internal
      Received Path ID 1, Local Path ID 0, version 0
      Community: 7922:403 7922:3020
      Originator: 96.109.22.227, Cluster list: 68.86.1.0, 96.109.22.50
```

Figure 2 – Example BGP information retrieved via CLI

- Typically, such interfaces are considered privileged access to the routers in question.
  Such access is protected by authentication and authorization schemes as well as other controls.
  The reasons for these measures are legitimate:
  - Non-public data may be available by such means.
  - The invocation of the command-line interface itself may have cause some extra processing load to the devices which runs a risk, however minor, of compromising the critical network functions of the network devices.
- The overhead of the authentication and authorization can inject considerable lag into the information gathering process.
- For situations inside a given organization where different levels of access exist, but many people may have use for network information, the ability to query and retrieve such information in a dynamic manner that does not compromise the integrity of the devices for information that is deemed non-sensitive.
- Furthermore, where information is acceptable to be publicly available, no standard scheme for structuring and providing such information exists.

3. Current Solutions for sharing BGP routing information

For retrieval of BGP data, two main solutions exist:

- Route Servers - These are typically router CLI (Command-Line Interface) means.
  As described earlier, they have no standard query language nor result format.
  The need to handle authentication and authorization by third parties without explicit permission makes for a sub-optimal experience. Example outputs from two separate route servers are illustrated in Figure 3.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

```
RP/0/RSP0/CPU0:route-server.newyork.ny.ibone#show bgp 13.0.40.0/23
Sat Jul  1 04:55:11.784 utc
BGP routing table entry for 13.0.40.0/23
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker          633072720   633072720
Last Modified: May 14 12:03:50.037 for 2y06w
Paths: (1 available, best #1)
  Not advertised to any peer
  Path #1: Received by speaker 0
  Not advertised to any peer
  15133 1321 33631, (received & used)
    66.208.229.9 from 66.208.229.9 (96.109.22.227)
      Origin IGP, metric 0, localpref 275, valid, internal, best, group-best
      Received Path ID 0, Local Path ID 0, version 633072720
      Community: 7922:403 7922:3020
      Originator: 96.109.22.227, Cluster list: 96.109.22.250, 96.109.22.50
RP/0/RSP0/CPU0:route-server.newyork.ny.ibone#
```

```
rviews@route-server.ip.att.net> show route 13.0.40.0/23

inet.0: 911801 destinations, 14587004 routes (911801 active, 0 holddown, 0 hidden)
+ = Active Route, - = Last Active, * = Both

13.0.40.0/23       *[BGP/170] 3d 18:45:14, localpref 100, from 12.122.83.238
                      AS path: 7018 701 1321 33631 I, validation-state: unknown
                    > to 12.0.1.1 via em0.0
```

Figure 3 – Example BGP information details from Route Servers

- Looking Glasses - These are web servers which have some backend capability to query one or more network devices for the underlying information. (See Figure 4) Not only is the exact query language and output format not standard, but some back-end rate limits must be imposed to prevent adverse effects on the network devices themselves.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

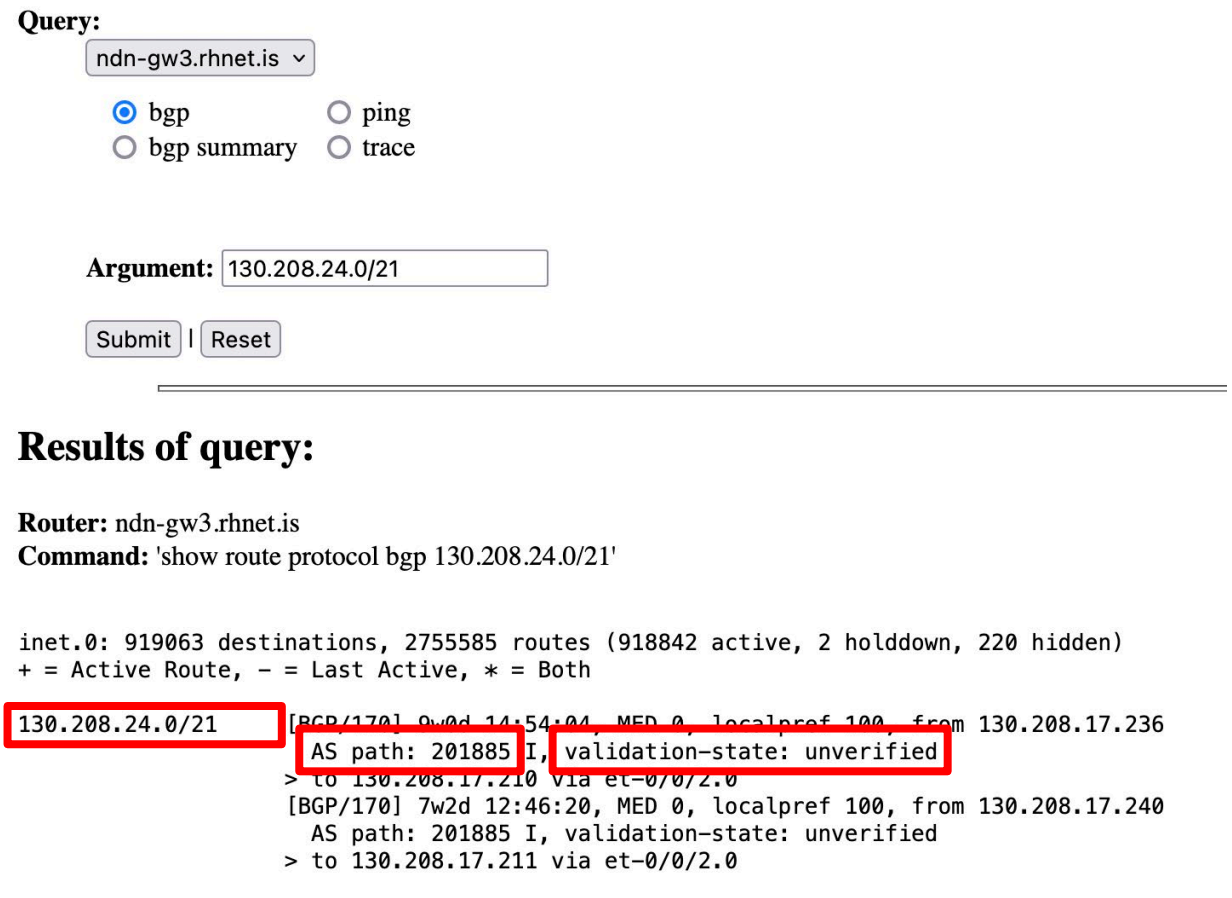


Figure 4 – Example Looking Glass Web Interface

4. DNS (Domain Name System) is standard and universally available

The DNS (Domain Name System) primarily provides a mapping between human-recognizable and "significant" names to IP addresses. The fundamental operation of a DNS "query" involves a resolver which recursively asks name servers, with presumed authority over the namespace, for an answer. These name servers respond to the best of their ability, which usually means undertaking the iterative function of asking another name server that has been delegated authority over a more specific part of the name space until one with a matching record is found. Figure 5 illustrates such a search.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

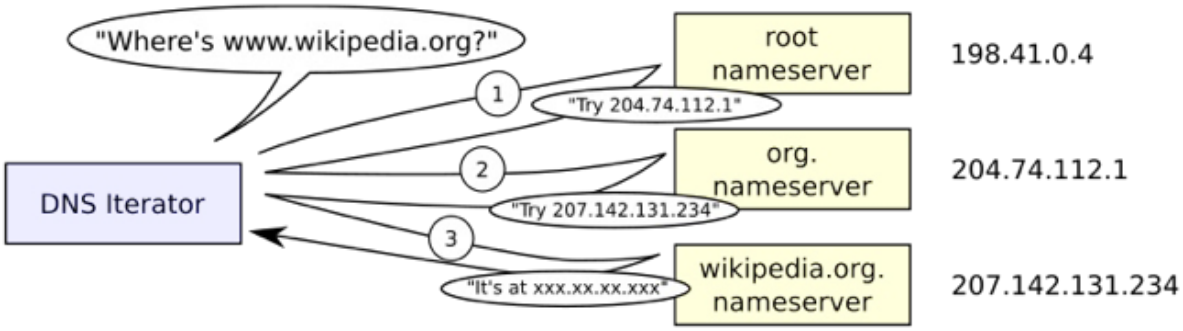


Figure 5 – DNS resolver iteratively queries more specific authorities to find an answer

These operations are done behind the scenes when using a web browser, for instance.

Using the "dig" command-line DNS utility allows for examining DNS details and can produce results such as those shown in Figure 6 below.

```
bash-3.2$ dig example.com

; <<>> DiG 9.10.6 <<>> example.com
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 5374
;; flags: qr rd ra ad; QUERY: 1, ANSWER: 1, AUTHORITY: 0, ADDITIONAL: 1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 4096
;; QUESTION SECTION:
;example.com.                   IN      A

;; ANSWER SECTION:
example.com.            74249   IN      A       93.184.216.34

;; Query time: 84 msec
;; SERVER: 69.252.80.80#53(69.252.80.80)
;; WHEN: Thu Aug 17 10:03:08 EDT 2023
;; MSG SIZE  rcvd: 56
```

Figure 6 – Using "dig" utility to look up a DNS record

Note the command and arguments we supplied at the top.
In the middle is the answer returned by the DNS.

The DNS has many defined Resource Record (RR) types.

We will focus on one in particular.

RFC1035 section 3.3.14 defines the TXT (pronounced "text") records as:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

TXT-DATA One or more <character-string>s.

TXT RRs are used to hold descriptive text.

The semantics of the text depends on the domain where it is found.

A very common current use relates to SPF (Sender Policy Framework) which is part of the global email system, but not germane to this discussion. We use it here to show an example of retrieving such a TXT record using the "dig" utility. Figure 7 shows the query at the top, the answer in the middle, and at the bottom we re-run the command adding the "+short" directive to give more terse output (just the "answer" without the verbose diagnostic information.)

```
bash-3.2$ dig example.com txt

; <<>> DiG 9.10.6 <<>> example.com txt
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 3299
;; flags: qr rd ra ad; QUERY: 1, ANSWER: 2, AUTHORITY: 0, ADDITIONAL: 1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 4096
;; QUESTION SECTION:
;example.com.                   IN      TXT

;; ANSWER SECTION:
example.com.            86400   IN      TXT     "wgyf8z8cgvm2qmxpnbnldrcltvk4xqfn
"
example.com.            86400   IN      TXT     "v=spf1 -all"

;; Query time: 65 msec
;; SERVER: 69.252.80.80#53(69.252.80.80)
;; WHEN: Wed Aug 23 16:46:56 EDT 2023
;; MSG SIZE  rcvd: 109

bash-3.2$ dig +short example.com txt
"v=spf1 -all"
"wgyf8z8cgvm2qmxpnbnldrcltvk4xqfn"
bash-3.2$
```

Figure 7 – Using "dig" utility to look up a DNS TXT record

5. Solution – BGP info served via DNS

Our proposed solution would be to use the DNS TXT ("text") record type to return BGP routing information when queried for a given IP address or range. Earlier we described the operational process of querying a router via CLI to retrieve information from or about the BGP table on that router. That process uses LPM to retrieve the relevant information. Our solution involves issuing a DNS query for an IP address or prefix to a server which performs LPM against a data store, which is similar to the BGP routing table and contains similar data which is then returned to the user.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The BGP table information can be exported from the router(s) to a server on a periodic or dynamic basis. An example of a periodic basis would be one where a daily process would:

1. Log into a router.
2. Use CLI commands to dump the routing table.
3. Parse the routing table contents into a format and data structure usable by the name server program.

An example of dynamic update might be:

1. A server runs a BGP software package.
    a. Examples: BIRD, Zebra, Quagga, openBGPd, FRR (Free Range Routing)
2. That server's BGP process is set to receive a BGP feed from router(s) in the network.
3. A process on the server converts the BGP information into a format and data structure usable by the name server program.

That server would then respond to queries which specify a given IP prefix (sometimes called a "route") and would then respond using the LPM mechanism to find the "best matching" prefix and then return the table information for that prefix in the form of one or more DNS TXT (text) record(s).

An example DNS query using the 'dig' DNS utility might be: ---
dig +short 1.1.4.1/24.get-ip-info.comcast.net
---

returns

'prefix: 1.0.4.0/22'
'as-path: 174 7545 2764 38803'
---
or alternately in JSON (JavaScript Object Notation) which is more suited to machine readability:
'{"prefix":"1.1.1.0/24", "as-path":"174 7545 2764 38803"}'

The solution benefits from the properties of the DNS:

5.1. Delegation

A key scaling feature of the DNS is achieved through delegation. No one set of servers could contain all the information about the Internet, so different pieces are delegated for administration at various levels. Returning to the analogy of postal delivery on the road system, a given country typically delegates some authority to states which, in turn, delegate more local authority to counties and/or cities and towns within their borders. Figure 8 illustrates part of the structure behind the delegated makeup of the DNS.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

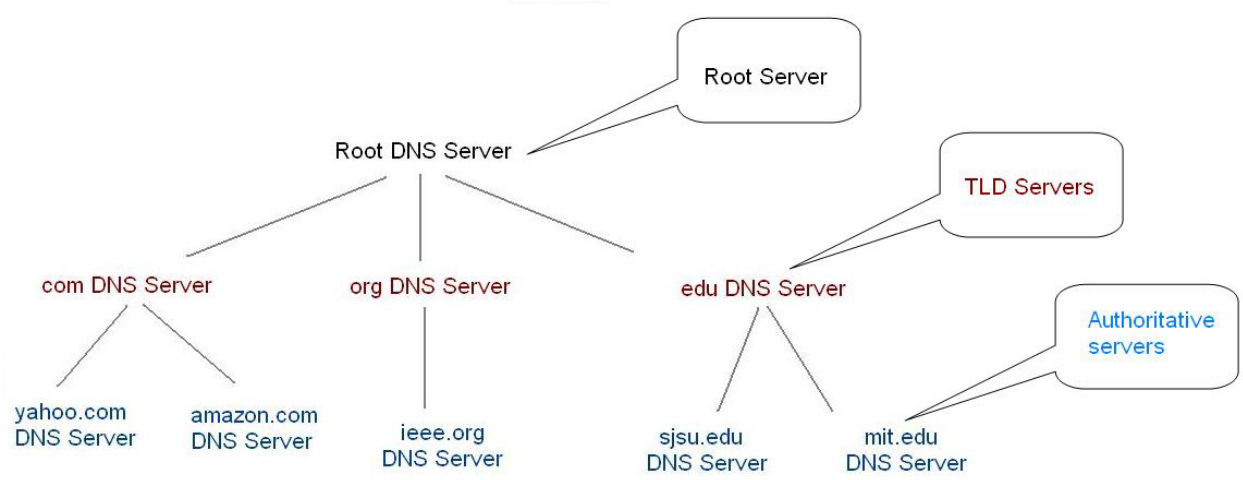


Figure 8 – The delegated structure of the DNS name space

5.2. Governance

As we introduced earlier, the most typical way the DNS is used is to provide mapping between a "name" and an IP address where one may retrieve information about that "name". In a related way, a "reverse" mapping between IP addresses and names is also possible within the DNS. As IP addresses are delegated from a "root" with increasing levels of specificity, those levels further delegate authority for their details.

RFC1035 section 3.5 describes the IN-ADDR.ARPA domain:

> The domain begins at IN-ADDR.ARPA and has a substructure which follows the Internet addressing structure.

and provides this example:

> A resolver which wanted to find the host name corresponding to Internet host address 10.0.0.6 would pursue a query of the form:
>
> QTYPE=PTR, QCLASS=IN, QNAME=6.0.0.10.IN-ADDR.ARPA, and would receive:
>
> 6.0.0.10.IN-ADDR.ARPA. PTR MULTICS.MIT.EDU.

RFC3596 describes the reverse mapping for IPv6 addresses under the IP6.arpa domain.

Figure 9 depicts the delegation structure for the in-addr.arpa domain which can be used to find "reverse" mapping of IP addresses to names.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

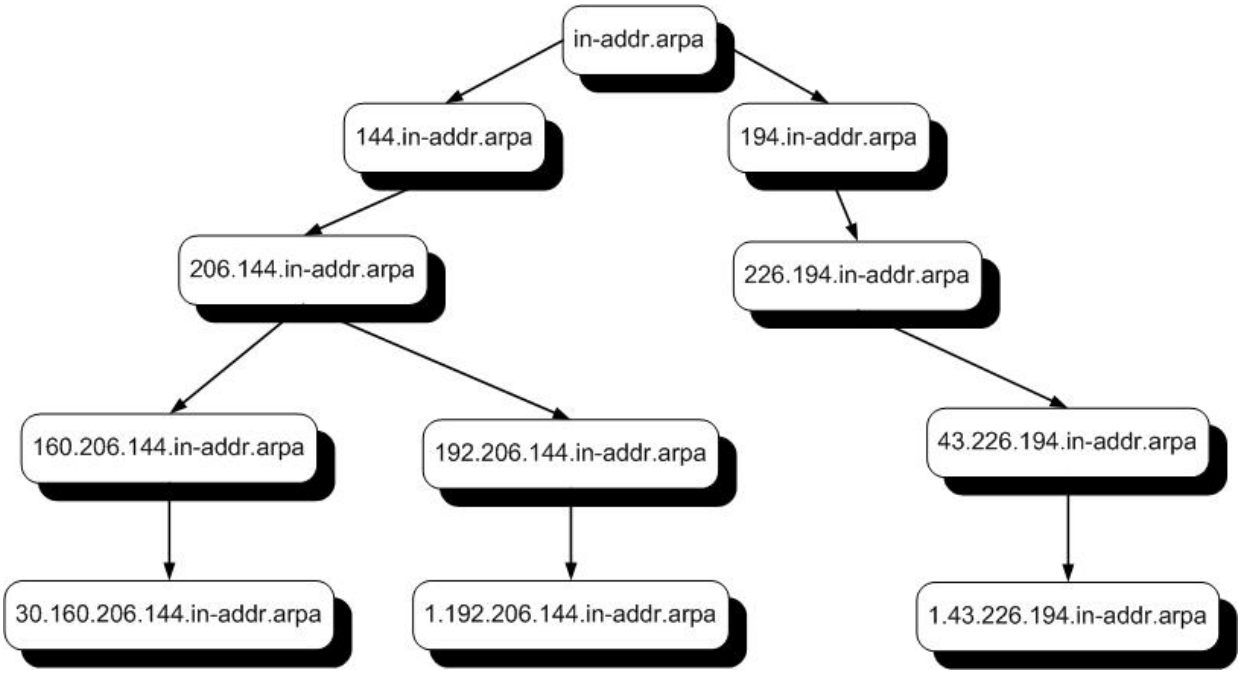


Figure 9 – The delegated structure of the DNS "reverse" name space for IPv4 addresses

The world of the Internet network operators is structured and reflected in BGP with each "network" as an AS (Autonomous System) having an ASN (Autonomous System Number) to uniquely identify it. Each AS, by definition:

> "…is a set of routers under a single technical administration ... [which] appears to other ASes to have a single coherent interior routing plan and presents a consistent picture of the destinations that are reachable through it."

[Source: RFC4271 (BGP - Border Gateway Protocol)]

In this sense, information for each ASN could be "located" in the DNS hierarchical namespace by reference to their ASN.

For example, the ICANN (Internet Corporation for Assigning Names and Numbers) and IANA (Internet Assigned Number Authority) could define a place in the DNS structure such as "network-info.arpa" where each delegation below (to the "left of") that domain could be delegated to the operator who has been assigned that ASN. Hence, Comcast may have a delegation for "7922.network-info.arpa" as well as each other ASNs which Comcast has been assigned. (The specific name "network-info.arpa" is used only for illustration.)

There is already an established delegation structure for AS numbers as well as DNS delegation so the "business rules" of such an invention would be straightforward to apply. Each AS operator would have a place to publish that could be found without any different "directory" than already exists today with ASNs being handled by the RIRs (Regional Internet Registries: ARIN, LACNIC, Afrinic, RIPE, APnic - and their delegates).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.3. Structure of information

With this name space delegated among the various "AS (Autonomous System)" operators, the operators could further break it up as they pleased by location or network device. They could "publish" the structure of their part of the namespace using a different structured TXT record which would enumerate the delegation below the root of their portion of the namespace.

5.4. Caching

DNS includes the concept and ability of caching by intermediate nodes. The length of time such records are cached (including "zero") can be controlled by the authority for the namespace.

5.5. Performance

Many performance enhancements have been undertaken by the industry in the form of software and hardware optimizations.

6. More Details

BGP data is just one example of network information that could benefit from such treatment. There are other parts of network inventory, network configuration, or network state which could be published and retrieved by these means. The concept and approach are flexible and extensible.

One example might be a DNS responder that replies with the client's source address in TXT records:

```
dig +short TXT what-is-my-ip.comcast.net @labweek.dns.server.comcast.net
        3600 IN TXT "ip6:2001:558:1438:21:8c3:3f13:8cfb:9ef3"
```

Another example might reply with administrative information about the address:

- Security Zone
- Location
- Edge Router Name
- Owner

```
dig +short TXT 10.0.0.1/32.get-ip-info.comcast.net @labweek.dns.server.comcast.net
        3600 IN TXT "containerPrefix:10.0.0.1/16"
        3600 IN TXT "securityGroup:PURPLEZONE"
        3600 IN TXT "site:OmahaDC"
        3600 IN TXT "lat:41.2438° N"
        3600 IN TXT "long:99.4874° W"
        3600 IN TXT "owner:lls_ops@comcast.com"
```

Or embed the prefix *and* the router name to get the view from that specific router:

- AS-PATH
- BGP Community
- Date/Time stamp

```
dig +short TXT 10.0.0.1/32.cs01.pittsburgh.pa.ibone.get-asn-info.comcast.net
@labweek.dns.server.comcast.net
        3600 IN TXT "AS-path: 22909 64922"
        3600 IN TXT "community: 7922:416 7922:999 7922:2800 7922:2900 22909:7000"
```

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

```
3600 IN TXT "date:Thu Mar 16 13:23:59.985 UTC"
```

7. Conclusion

Network operations often require synthesizing many sources of data from many directions.

In the realm of BGP routing, interrogating the routing information held within a given router and across several vantage points is sometimes needed to perform diagnosis and analysis.

The traditional method for getting such information from routers is using an authenticated and authorized CLI . Sharing or retrieval of such information among and between different network operators is done using the CLI of "route servers" or a GUI (Graphical User Interface) provided by a "Looking Glass" web server which issues commands on the users' behalf to routers to retrieve such information.

Significant drawbacks of these approaches are they lack standardization and are cumbersome to use. They are not easily amenable to automation and pose some risks to the routing infrastructure where they get their data.

Our solution is to obtain BGP information proactively from one or more routers within an operator's network and make it available via DNS queries which perform Longest Prefix Matching to return the relevant information via TXT ("text") records.

This scheme benefits from several features of the DNS including hierarchy, delegation, caching, and performance enhancements. Moreover, it can be used to encode other types of network metadata as well such as geolocation, etc.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AS	autonomous system
ASN	autonomous system number
BGP	border gateway protocol
CLI	command line interface
DNS	domain name system
GUI	graphical user interface
IANA	Internet Assigned Number Authority
ICANN	International Corporation for Assigned Names and Numbers
IP	internet protocol
LPM	longest prefix match

Bibliography & References

[RFC1034] Mockapetris, P., "Domain names - concepts and facilities", STD 13, RFC 1034, DOI 10.17487/RFC1034, November 1987, <https://www.rfc-editor.org/info/rfc1034>.

[RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, DOI 10.17487/RFC1035, November 1987, <https://www.rfc-editor.org/info/rfc1035>.

[RFC1101] Mockapetris, P., "DNS encoding of network names and other types", RFC 1101, DOI 10.17487/RFC1101, April 1989, <https://www.rfc-editor.org/info/rfc1101>.

[RFC2317] Eidnes, H., de Groot, G., and P. Vixie, "Classless IN-ADDR.ARPA delegation", BCP 20, RFC 2317, DOI 10.17487/RFC2317, March 1998, <https://www.rfc-editor.org/info/rfc2317>.

[RFC3596] Thomson, S., Huitema, C., Ksinant, V., and M. Souissi, "DNS Extensions to Support IP Version 6", STD 88, RFC 3596, DOI 10.17487/RFC3596, October 2003, <https://www.rfc-editor.org/info/rfc3596>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

Route Views Project <https://www.routeviews.org/routeviews/>

BGP Looking Glass Database <https://www.bgplookingglass.com/>

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23 CELEBRATING 40 YEARS

2023 Fall Technical Forum SCTE • CABLELABS • NCTA

Breaking the Barriers: Abstracting Traffic Management for Superior Quality of Experience in Multi-Technology Networks

A Technical Paper prepared for SCTE by

Sebnem Ozer, Ph.D.
Principal Architect III
Charter Communications
sebnem.ozer@charter.com

Ramneek Bali
Principal Wireless Engineer III
Charter Communications
ramneek.bali@charter.com

Kamran Yousuf, Principal Engineer P.ENG | PMP, Charter Communications

Moutaz Elkaissi, Director, Product Management, Charter Communications

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. The Diverse Network and Service Technologies Within ISP Architecture ... 3
3. Everchanging Landscape of Quality of Experience Framework ... 6
4. Case Study: Supporting IETF L4S and 5G Midhaul Over DOCSIS Networks ... 11
5. Future-Proofing the Network and Service Platforms ... 18
6. Conclusion ... 21

Abbreviations ... 21

Bibliography & References ... 22

Acknowledgments ... 23

List of Figures

Title **Page Number**

Figure 1- MSO Architecture With Diverse Networking Technologies ... 4
Figure 2- DOCSIS DAA Example ... 5
Figure 3- PON DAA Example ... 5
Figure 4- vRAN Architecture Example ... 6
Figure 5- IoT Usage ... 10
Figure 6- DOCSIS Serving Group Utilization Characteristics per Observation Time Windows (source: NetScout Packet Inspection Tool) ... 11
Figure 7- Downstream IETF L4S application throughput and RTT when home network is saturated. Top: Without LLD; Bottom: With LLD ... 12
Figure 8- IETF PHB-NQB Traffic One Way Latency Top: 300/10Mbps ; Bottom: 1G/200Mbps ... 13
Figure 9- RAN Architecture - Backhaul ... 14
Figure 10- vRAN Architecture - Midhaul ... 14
Figure 11- Unloaded DOCSIS Throughput in the Lab Setup ... 15
Figure 12- DOCSIS Latency Benchmark in the Lab Setup ... 15
Figure 13- “Without LLD” Use Case ... 16
Figure 14- “With LLD” Use Case ... 17
Figure 15- Multiples of the Number of UEs Supported for Different DOCSIS Load Percentages with Additional LLD Firmware in a Midhaul Use Case ... 17
Figure 16- Traffic Abstraction for Automated and Programmable Systems ... 19
Figure 17- End-to-End Traffic Management ... 20
Figure 18- End-to-End and per hop QoS Monitoring: KPI examples ... 20

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Today, cable service providers deploy a variety of network technologies to offer a wider range of services to their customers and to tailor their services to specific geographic areas or customer needs. These technologies include Hybrid Fiber-Coaxial (HFC), Passive Optical Networks (PON) and Mobile Networks, with WiFi as the final leg of the connection to end devices or as a hotspot network. Deploying different network technologies can be challenging due to the increased complexity and cost of managing multiple networks. Operators may also face interoperability issues, which can result in poor service quality or higher maintenance costs. Distinct standards govern each of these technologies, which are further characterized by disparate network and service components. Consequently, the development of Quality of Experience (QoE) platforms becomes inherently tailored to specific technologies, creating barriers to achieving a cohesive and unified service framework.

Within this paper, we present a compelling use case for Internet Engineering Task Force - Low Latency, Low Loss, and Scalable Throughput (IETF L4S) and 5G midhaul support [1] over DOCSIS networks in order to optimize the customer's QoE for emerging and future use cases. Our focus delves into the exploration and evaluation of the mapping of 5G Stream Control Transmission Protocol (SCTP) signaling traffic onto the low latency Data Over Cable Service Interface Specification (LL DOCSIS/LLD Service Flow, substantiating the capacity of MSOs to accommodate this emerging service with the LLD support. Furthermore, we address the support for the IETF L4S traffic over multiple network segments. The study encompasses crucial facets like traffic classification, resource mapping and monitoring metrics, each pertaining to the corresponding network segments.

The LLD test system described in this paper uses manual processes for traffic mapping and key performance indicator (KPI) monitoring in a controlled lab environment. Building upon this foundation, we propose an automated solution to support end-to-end QoE by abstracting traffic management with a high-level orchestration platform and a digitized and hierarchical component structure. Ultimately, our paper provides guidelines to ensure optimal QoE delivery while simultaneously curbing the complexities and costs associated with traffic management.

2. The Diverse Network and Service Technologies Within ISP Architecture

Multiple System Operators (MSOs) employ an array of network technologies to broaden the spectrum of services accessible to subscribers, adapting offerings to specific geographical locations or consumer needs. The main technologies [15] are wireline networks such as Hybrid Fiber-Coaxial (HFC), Passive Optical Networks (PON) and Mobile Networks, with WiFi constituting the final link between the connectivity and end-user devices (depicted in Figure 1). Historical architectural design is predominantly centered on purpose-built infrastructure featuring centralized and integrated data, control and management systems. However, propelled by advancements in Software Defined Networking (SDN), Network Function Virtualization (NFV), cloud integration, computational capabilities, switching and data analysis technologies, MSOs have transitioned towards deploying distributed and virtualized configurations. Notable among these are Remote PHY/Virtual Cable Modem Termination System (vCMTS) for HFC (Figure 2), Remote Optical Line Terminal (rOLT)/Virtual OLT for Passive Optical Networking (PON) (Figure 3), Remote Radio Unit (RRU) in tandem with Virtual Radio Access Network (vRAN) (Figure 4) and Open RAN (ORAN) for Mobile Networks. These innovations are also being harnessed within the core and WiFi network segments, where NFV and SDN are being strategically applied.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

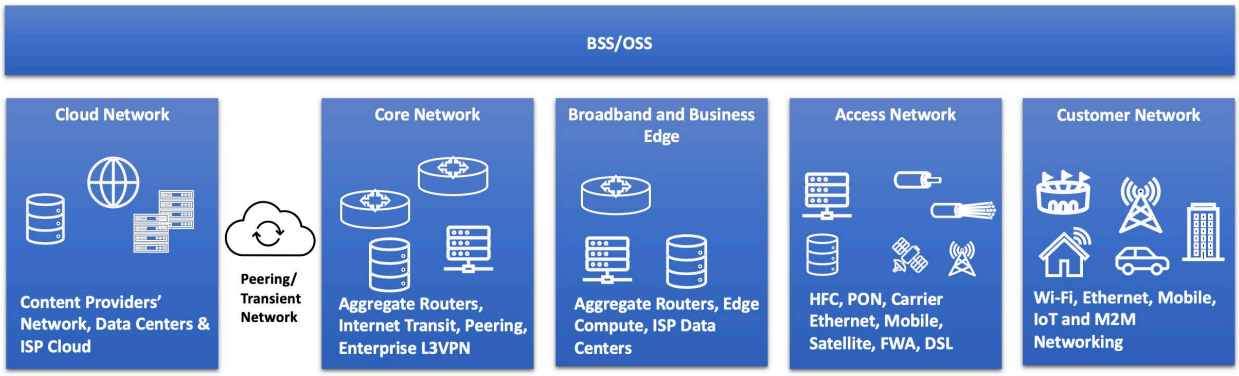


Figure 1- MSO Architecture With Diverse Networking Technologies

In these network paradigms, a prevalent strategy involves leveraging commercial-off-the-shelf commercial (COTS) hardware to deploy containerized applications and microservices, all seamlessly orchestrated by a management platform. Figure 2 displays an example of DOCSIS DAA where Remote PHY (RPD) is connected to vCMTS POD over fiber via a switch.

Here, containerized software materializes vCMTS and other access VNFs. While initial implementations encompassed a monolithic vCMTS data and control paradigm, recent design strategies emphasize microservices. Data forwarding functionality includes classifiers, traffic shaping, scheduling, queueing, filtering, relaying and routing, while control features include subscriber management, service provisioning, device configuration and activation.

Virtual routing interconnects numerous vCMTS PODs with the core network. Similarly, Figure 2 illustrates the PON DAA scenario, where Remote OLT (ROLT) interfaces with vOLT PODs. ROLT integrates PHY and lower MAC functionalities, while PON network provisioning and management functionality is virtualized. The modular separation of monolithic functions offers MSOs the avenue to disentangle data and control VNFs, harmonizing common attributes across DOCSIS, PON and ethernet technologies, while remote PHY/MAC components nestle within a remote switch node [3]. The orchestration of services and resources encompasses lifecycle management, automation and data analysis, all empowered by cloud-based telemetry.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

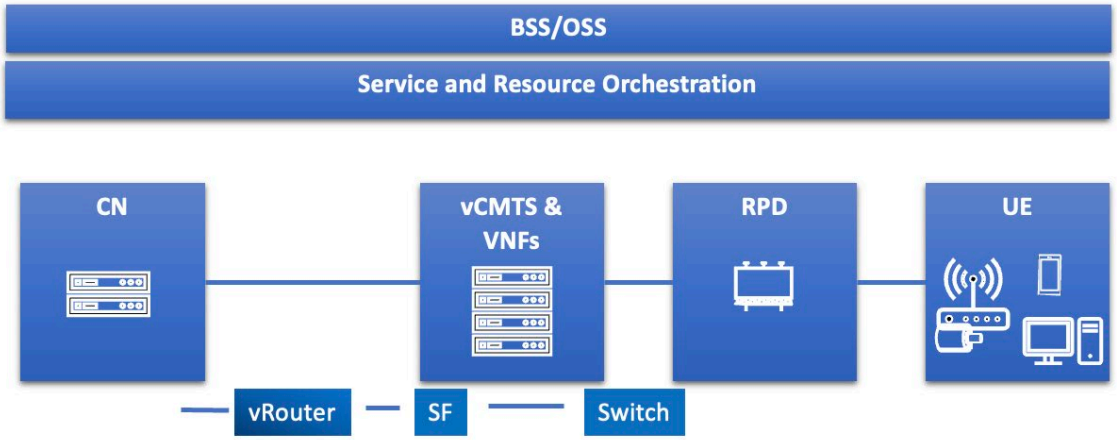


Figure 2- DOCSIS DAA Example

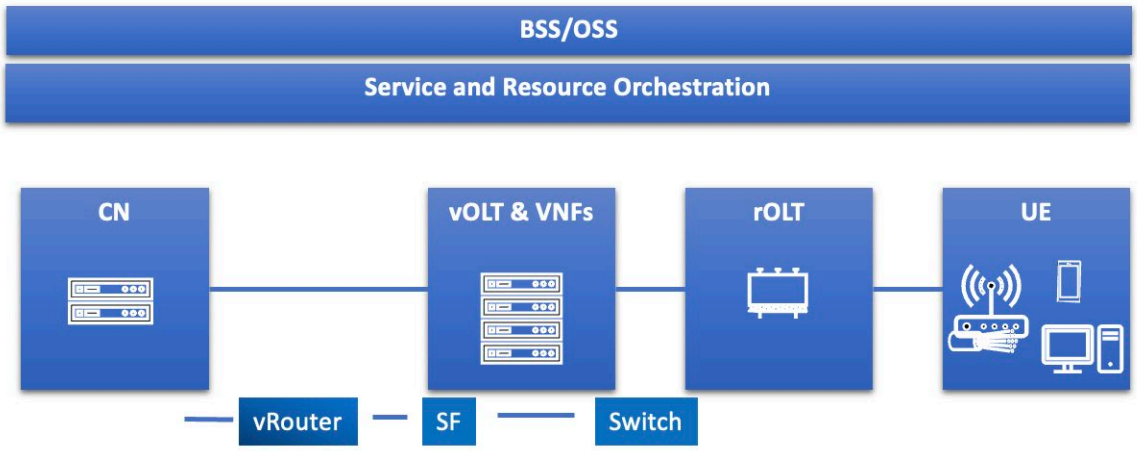


Figure 3- PON DAA Example

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

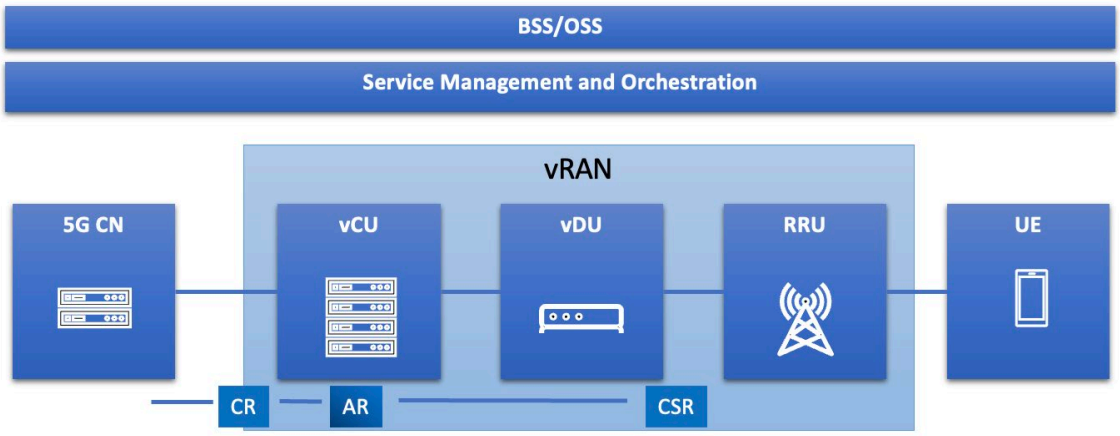


Figure 4- vRAN Architecture Example

Figure 4 shows a VRAN architecture. Virtual Centralized Unit (vCU) provides non-real-time processing and access control. It manages higher layer protocols including Radio Resource Control (RRC) from the control plane, and Service Data Adaptation Protocol (SDAP) and Packed Data Convergence Protocol (PDCP) from the user plane. The vCU is connected between the 5G core network and the Virtual Distributed Unit (vDU). One vCU can be connected to multiple vDUs. The vDU provides real-time processing and coordinates lower layer protocols including Physical Layer, Radio Link Control (RLC) and Media Access Control (MAC). Virtualization shifts the vCU and vDU from dedicated hardware to software components, allowing for flexible scaling, as well as rapid and continuous evolution. With vDU, all the baseband functions of real-time RLC/MAC/PHY layers are executed over the commercial-off-the-shelf (COTS) server Remote Radio Unit (RRU), which does the physical layer transmission and reception, supporting technologies such as Multiple Input Multiple Output (MIMO). Session management, usage metering, common policy, authentication and network slicing are implemented in the virtualized core network. Service management and orchestration (SMO) includes life cycle management, automation, analytics and non-real time intelligent controller. ORAN architecture defines open interfaces between RRU and DU and between DU and CU.

Amidst these breakthroughs in ISP network and service architecture innovation, concurrent efforts are still underway to enhance subscribers' quality of experience regardless of location (e.g. in the home/office or on the go), time (e.g. peak and off-peak utilization times) or device (e.g. wired and wireless devices), while simultaneously streamlining expenditures by enhancing operational and business support systems. These endeavors collectively harmonize diverse access and service technologies through unification and convergence. In this paper, we address this framework, with a concentrated focus on the facilitation of low-latency services. Notably, while our discourse employs the terminology of "low latency services," our underlying perspective interprets this as addressing the Quality of Service (QoS) requirements across distinct services via a comprehensive approach aimed at bolstering Quality of Experience (QoE). Our investigation thus addresses how MSOs can channel investments into future-proof network and service architectures.

3. Everchanging Landscape of Quality of Experience Framework

In the ever-evolving landscape of the Internet, extensive research revolves around the critical factors of Quality of Service (QoS) that affect subscribers' Quality of Experience (QoE), which is a

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

multidimensional and multi-disciplinary phenomenon. One QoS metric that has been monitored and managed extensively is bandwidth, although it's widely used interchangeably with speed, capacity, throughput or bit rate [2]. The QoS criteria, including parameters such as throughput, latency, jitter and packet loss, are contingent upon the specific characteristics of the service, the application in question, and the unique requirements of the customer use cases.

Embedded within this narrative is a fundamental premise: the intricate interplay of bandwidth, latency and reliability necessitates a holistic approach. This complexity arises from the cyclic interdependency encompassing network and service technologies, application evolution and evolving customer expectations, all propelled by induced demand. Technological enhancements usher in higher speeds, reduced latency and augmented reliability. This, in turn, spurs novel applications that mold consumer behavior and aspirations, thereby fostering increased demand for advanced services and network technologies, forming a self-perpetuating cycle. Reflecting upon the last decade's Internet development and its assimilation into diverse demographics, we glean insights into the path that the ensuing 10 years may traverse.

While projections regarding bandwidth requirements may differ, consensus emerges around the ascendant significance of attributes beyond mere speed. Reliability, consistent low latency and responsiveness occupy the forefront. To complement advancements in self-healing and automated solutions, customer service that adeptly resolves issues will wield equal significance.

Within the scope of Quality of Experience (QoE), the foundational elements remain steadfast, yet the dynamic consumer behavior and expectations continue to shift in tandem with the rapid evolution of technology. A compelling illustration of this phenomenon emerges when examining the statistics related to the abandonment rate of subpar streaming content on connected TVs and virtual Multichannel Video Programming Distributors (vMVPDs) in the United States during the years 2017 and 2018 [5]. Evidently, there was a significant surge in this rate over one year, underscoring a discernible decline in consumer tolerance for videos that exhibit glitches or compromised quality.

Extending our focus to a distinct study conducted in early 2021, a revelation unfolds – half of U.S. consumers opt out of online purchases when confronted with a slow and complex checkout process [5]. It is noteworthy that research has brought to light the impact of network delays on the loading time of web pages and associated activities. A revealing example is a study [6] which indicates that a mere 10ms increase in network latency engenders an additional 0.8 seconds for a page to load. A 60ms delay prolongs loading time to 1.7 seconds, while a 510ms lag culminates in a 17-second delay.

As expectations grow, so do the number of users. Anticipating a burgeoning future, it's anticipated that approximately 187.5 million U.S. users will have conducted at least one online purchase via web browsers or mobile apps on their mobile devices by 2024 [5]. This projection, a notable climb from the 167.8 million mobile U.S. buyers recorded in 2020, augments the prospect of a digitally immersive future.

A glimpse into the rapidly expanding realm of Internet of Things (IoT) devices, as depicted in [5], emphasizes the transformative role of digital innovation in every facet of modern existence. Notably, consumer internet and media devices, followed closely by smart grids and connected vehicles, are at the forefront of IoT applications, as portrayed in Figure 5. This extensive survey epitomizes the broad spectrum of services that necessitate dynamic quality of experience due to the ever-shifting landscapes of residential, enterprise and industrial services.

In the wake of cutting-edge developments, novel services like Augmented Reality/Virtual Reality (AR/VR) usher in fresh paradigms for sensory evaluation, while the realm of Machine-to-Machine

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

(M2M) services introduces new dimensions of criteria. Another thought-provoking survey [5] encapsulates consumers' unchanging expectation for consistent service quality, regardless of the underlying technology. Thirty-nine percent of surveyed gamers in 2022 anticipated that mobile adaptations of games would rival their PC or console counterparts in terms of storyline and narrative quality.

The widespread demand for seamless experiences across devices is palpable, as evidenced by a survey from Q3 2022, revealing that a staggering 81.9 percent of global internet users actively engage in video gaming across diverse platforms. A staggering 66.2 percent of respondents endorsed smartphone gaming, crowning it as the most popular form of gaming worldwide. In a close second, PC gaming via laptops or desktops claimed a substantial 37.9 percent share.

The far-reaching ramifications of latency, jitter and packet loss on various applications, including traditional and cloud-based gaming, as well as videoconferencing, is explained in [2]. The impending advent of immersive applications like AR/VR across domains such as entertainment, education, healthcare and more underscores their pronounced sensitivity to latency, jitter, packet loss and throughput performance. In tandem with these Quality of Service (QoS) metrics, reliability, security, privacy and accessibility collectively shape subscribers' satisfaction with services. A pressing imperative for cable operators lies in adapting to the flexible QoE frameworks as discerning subscribers increasingly demand services that transcend temporal, spatial and device constraints.

As technology advances and network and service technologies elevate user expectations, the tolerance for disruptions diminishes. This intricate interplay underscores the central importance of staying attuned to evolving performance standards. The subjective nature of QoE requires personalized solutions, but MSOs need to support it without requiring custom solutions. This is why network and service programmability with automated platforms that enable self-optimization continues to drive design improvements.

Pioneering strategies in network monitoring and operations come to the forefront, aimed at precisely measuring QoS metrics encompassing latency and jitter behaviors for a diverse range of traffic characteristics. The crucial significance of the IETF L4S and Non-Queue-Building Per-Hop Behavior (NQB-PHB) specifications and drafts lies in their mandate for measuring scalable TCP and NQB-PHB traffic, both at idle and working latency levels, in conjunction with classic traffic measurements [7]. Hop-by-hop measurements are used to find bottlenecks and other impairments at each network segment, while end-to-end measurements map all the findings to QoE.

Strategically opting for both active and passive latency measurement platforms, catering to both hop-by-hop and end-to-end analyses, emerges as a pivotal pursuit. It's important to capture network and service states during specific timeframes that precipitate latency fluctuations, as this enables the identification of key correlations and causations. For instance, latency sources in DOCSIS networks primarily stem from queuing and media access, while factors such as propagation, serialization, encoding and switching also contribute to latency. PON networks exhibit notable distinctions, especially from a media access and queuing standpoint, but the major latency and jitter factors are similar. Although propagation delay differs between coaxial and fiber mediums, the introduction of DAA has brought fiber closer to the subscriber site in HFC networks, resulting in an increased portion of the distance being covered by fiber.

While 5G stands apart due to its distinct medium and protocols, the same underlying factors impact both latency and jitter. In the case of WiFi, inter-AP and external interference, as well as downstream and upstream contention, the dynamics of latency and jitter factors are magnified. An important consideration is that detecting latency variations and their origins is challenging until in-packet measurements become feasible, given the potentially sub-millisecond or lower durations of these events.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Embracing automation and data analytics emerges as a promising avenue for effectively fine-tuning measurement and analysis protocols, ushering in a more efficient era of QoS evaluation. Delving into the heart of network segments, and keenly understanding their inherent characteristics, proves indispensable in tailoring QoS measurement tools. A notable instance is the nuanced role of core networks, which wield a more pronounced influence on latency than jitter, while WiFi, on the other hand, significantly impacts jitter. This comprehension paves the way for precision-tailored tuning strategies, enhanced by judicious application of data analytics and automation.

As depicted in Figure 6 where an aggregate view of internet usage is collected for network analysis, a DOCSIS service group's utilization at various resolutions showcases microbursts occurring at different aggregate levels. However, most traditional and modern distributed MSO architectures typically gather data in 15- to 300-second measurement windows, potentially missing microbursts and their impact on jitter. Similarly, unless Received Signal Strength Indicator (RSSI) or other link issues, along with microbursts arising from home network applications, are measured at smaller resolutions, their influence on WiFi latency and jitter might remain unnoticed. MSOs can leverage data analytics to detect such events.

The dawn of new Simple Network Management Protocol Management Information Bases (SNMP MIBs), encompassing latency histograms and congestion metrics (i.e. docsQosSfLatencyStats and DocsQosSfCongestionStats) within CableLabs DOCSIS specifications, enhance the QoS monitoring for MSOs. Employing similar measurements in other network segments can also improve hop-by-hop and end-to-end analyses if executed in a coordinated manner. This coordination entails synchronized measurements, standardized metrics and integrated dashboards alongside comprehensive data analysis.

In conclusion, the unceasing evolution of technology beckons us to explore the ever-shifting terrain of Quality of Experience (QoE). As technology reshapes expectations, it's imperative for stakeholders to navigate this landscape with agility, embracing the tools of analytics, automation and adaptability to ensure that the march towards optimal quality remains uninterrupted.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

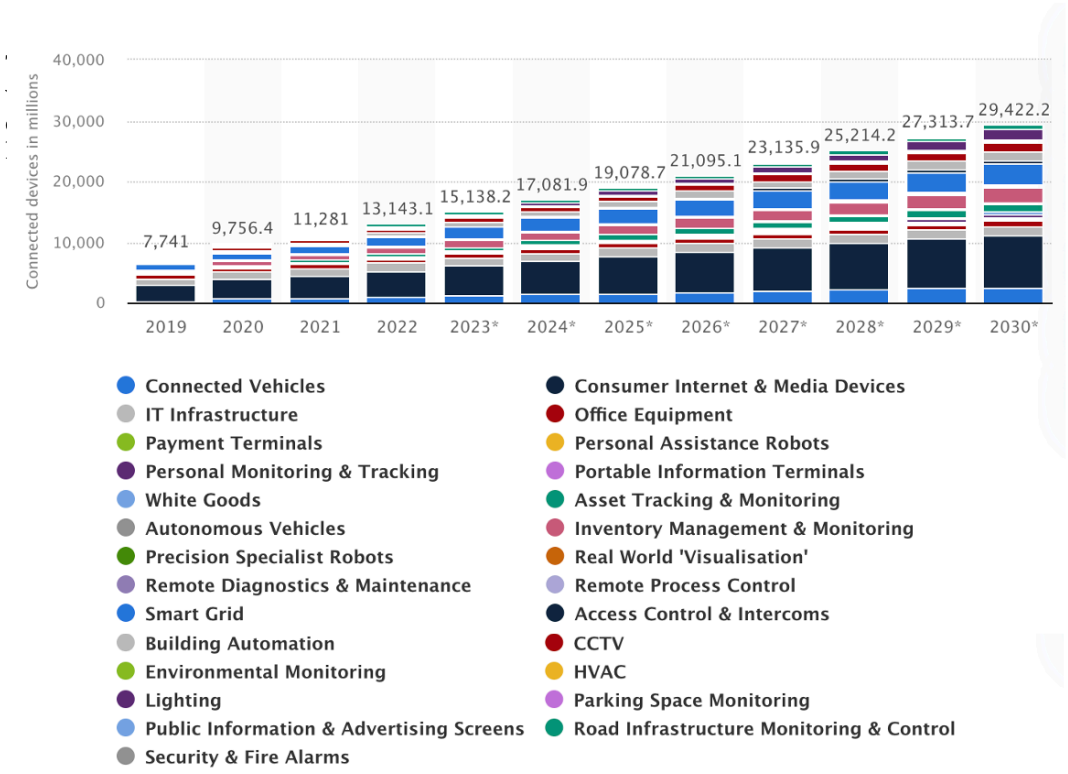


Figure 5- IoT Usage

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 6- DOCSIS Serving Group Utilization Characteristics per Observation Time Windows (source: NetScout Packet Inspection Tool)

4. Case Study: Supporting IETF L4S and 5G Midhaul Over DOCSIS Networks

The LLD improvements are first tested for Low Latency traffic (i.e. IETF L4S and PHB-NQB) [10,11] to assess the wireline quality improvements over DOCSIS. Figure 7 shows a lab result where LLD ensures downstream L4S traffic to keep consistent low latency and stable throughput changes during bursts created by other household utilization that saturates the home network capacity. Even when full home bandwidth is used, IETF L4S traffic RTT stays low while total capacity is efficiently used. Figure 8 illustrates the one-way latency percentiles of upstream and downstream low rate UDP traffic when home network is saturated for different bandwidth rates. The latency and jitter improvements for LL marked traffic can be observed when compared to unmarked traffic results. The findings indicate that increased bandwidth does not necessarily ensure consistent reductions in latency and jitter, particularly when a home network's bandwidth becomes saturated. The phenomenon of induced demand introduces scenarios in which microbursts may occur, even within non-latency-sensitive applications that can derive advantages from transmitting substantial traffic volumes. One important aspect is that this may happen in any network segment candidate to be the bottleneck over the service path.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

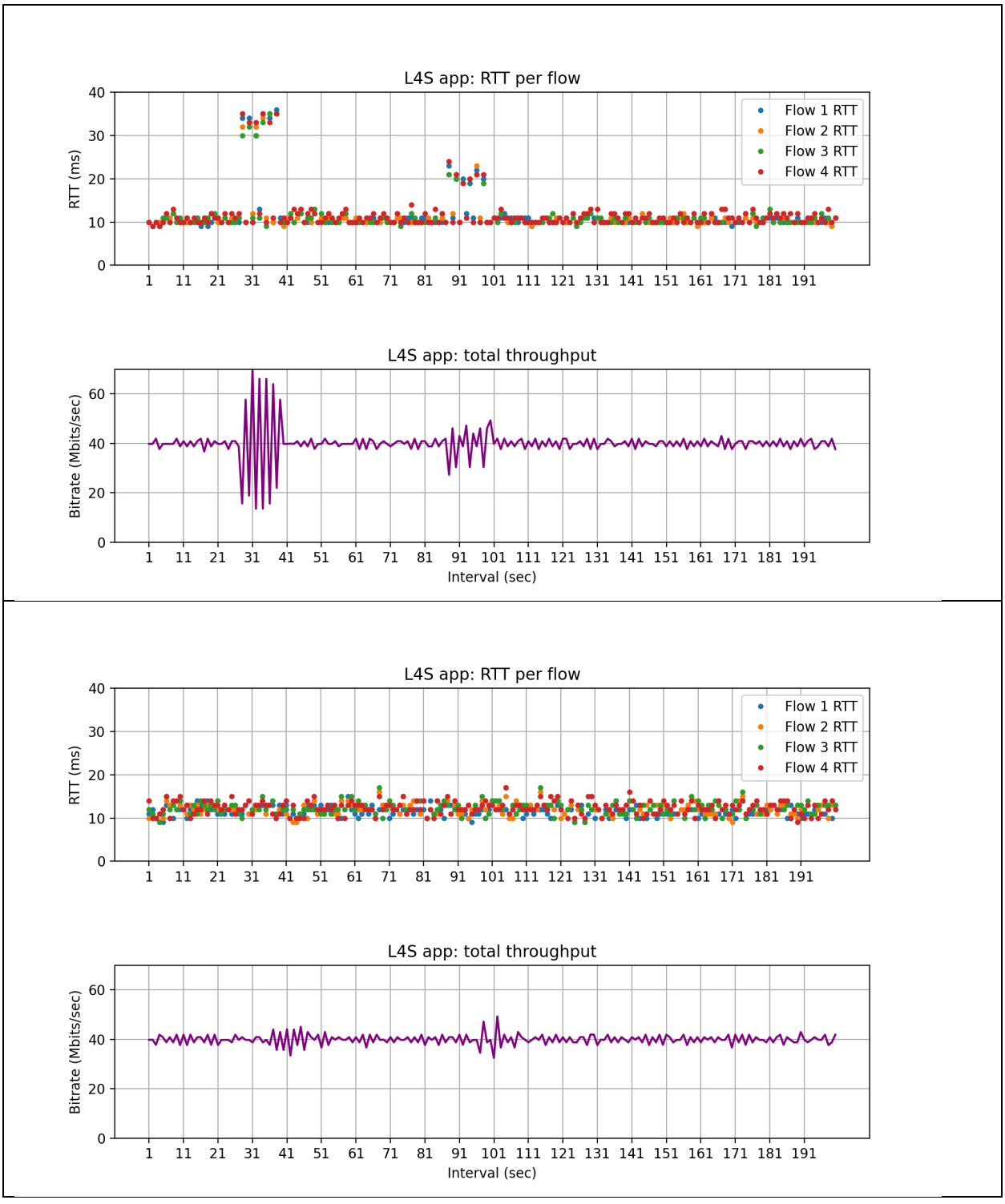


Figure 7- Downstream IETF L4S application throughput and RTT when home network is saturated. Top: Without LLD; Bottom: With LLD

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

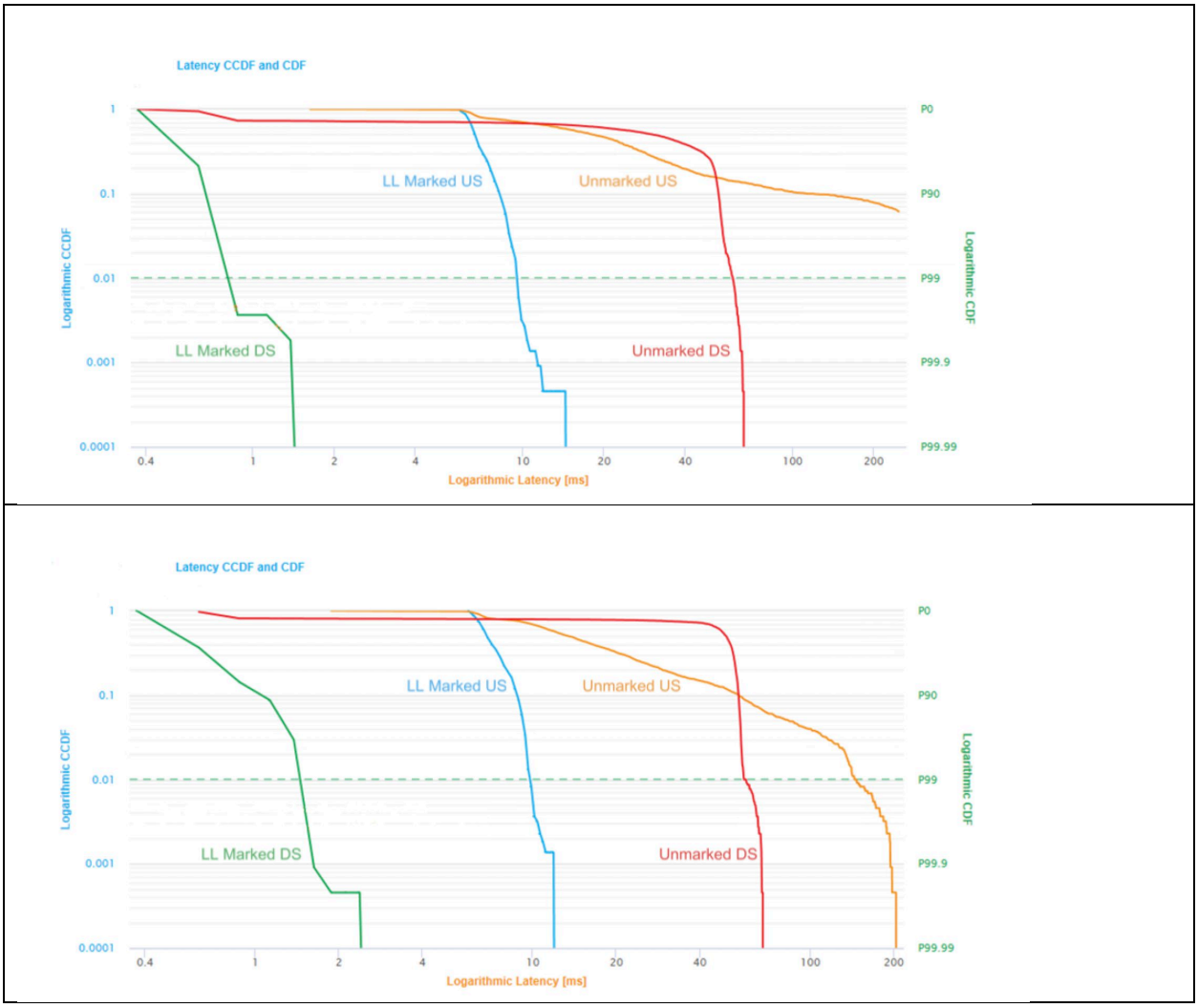


Figure 8- IETF PHB-NQB Traffic One Way Latency Top: 300/10Mbps ; Bottom: 1G/200Mbps

Virtualization of RAN offers the flexibility of varied deployment options based on the latency requirements. The deployment possibilities for MSOs are discussed in [1]. MSOs can leverage their HFC network to build a network of 5G small cells to offload MVNO traffic, thereby reducing service cost. Today most of the HFC networks can support backhaul use case (Figure 9) without additional upgrades or architecture changes. With the LLD support, the feasibility of midhaul (Figure 10) use cases becomes also

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

achievable by only firmware changes [4], which is studied in this section. Fronthaul options would require specific architectural design and protocols, like LLX.

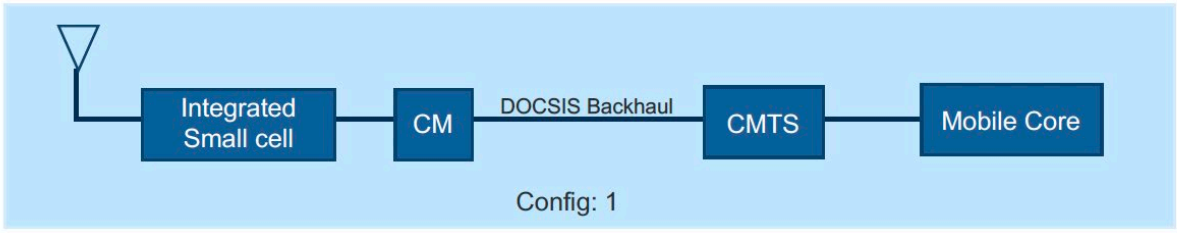


Figure 9- RAN Architecture - Backhaul

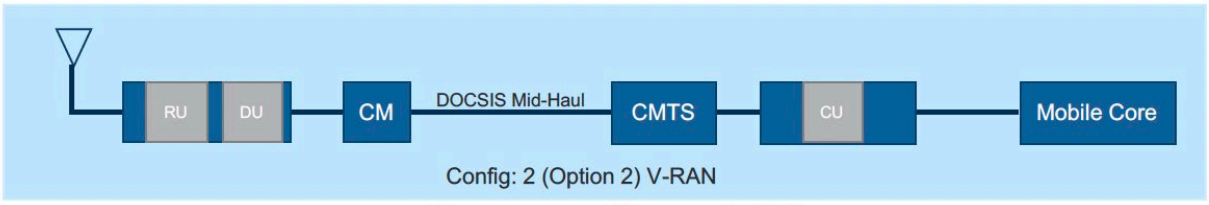


Figure 10- vRAN Architecture - Midhaul

The F1 interface (midhaul) connects a gNB vCU to a gNB vDU. This interface is applicable to the vCU-vDU Split gNB architecture. The control plane of the F1 (F1-C) allows signaling between the vCU and vDU, while the user plane of the F1 (F1-U) allows the transfer of application data. The latency performance of DOCSIS with LLD is best suited for this type of deployment with vDU at the edge and vCU centralized or at the edge.

A lab setup with 5G virtual core, V-RAN on open platforms and HFC network was done to evaluate the performance of the midhaul use case. The main challenges of the lab setup were interoperability issues with the integration of OS, vCU and vDU SW, along with routing and messaging on standard interfaces, while underlaying DOCSIS network setup and configuration was relatively easy. Unloaded DOCSIS throughput performance is shown in Figure 11 for benchmarking purposes of achievable peak VRAN (100Mhz) rates with selected configuration parameters. Total channel capacity is ~180Mbps in upstream and ~2.25Gbps in downstream. Figure 12 shows latency values when utilization is close to channel capacity.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

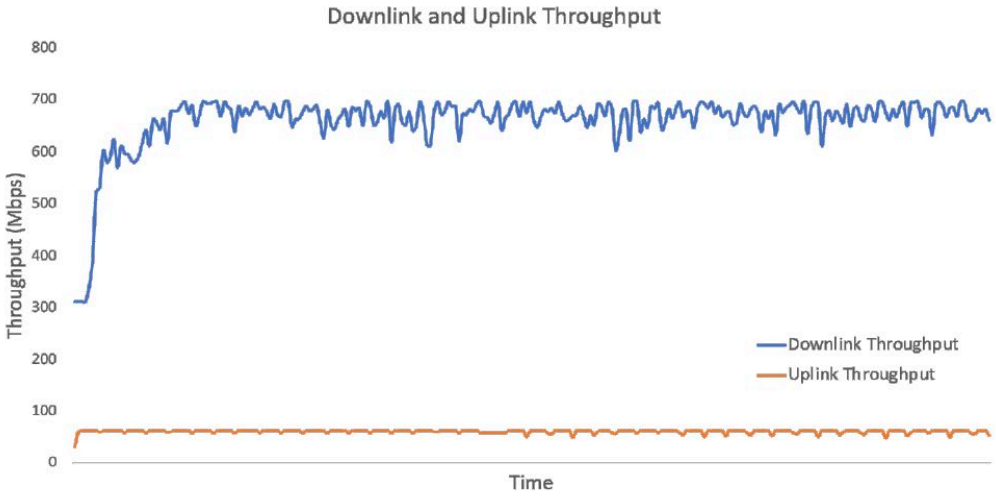


Figure 11- Unloaded DOCSIS Throughput in the Lab Setup

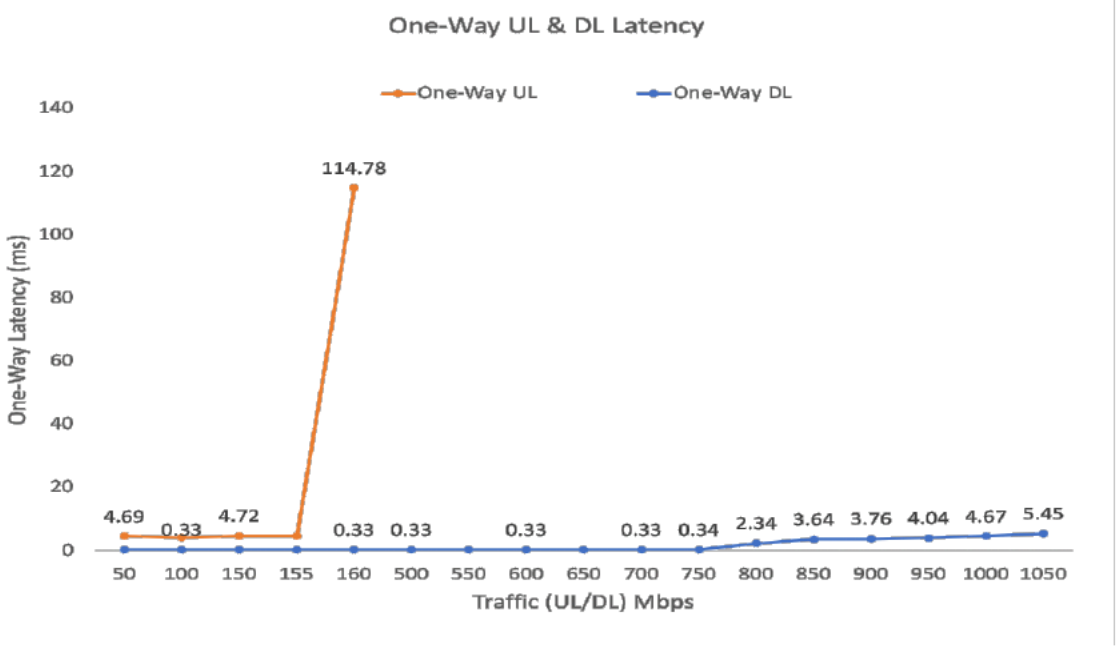


Figure 12- DOCSIS Latency Benchmark in the Lab Setup

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Two use cases are used for each test suite. “Without LLD” refers to the case where signaling traffic is not classified to DOCSIS LL SF. In this case, both signaling and user data are sent through the same DOCSIS service flow as illustrated in Figure 13 [10, 11]. Both flows of traffic use the same queue with classic Active Queue Management (AQM) and they are subject to the same scheduling decisions. “With LLD” refers to the case where signaling is classified as LL SF and user data as classic SF in the same Aggregate Service Flow (Figure 14). In this case, user data is sent through Classic AQM while signaling traffic is forwarded through IAQM defined in [10, 11]. A weighted intra-SF scheduling is applied in the CMTS for upstream and downstream traffic. Queue protection detects and sanctions misbehaving traffic while coupling aims throughput fairness between classic and LL traffic.

The emphasis was given on the upstream performance due to the limiting factor in terms of latency and jitter, which varies with the utilization and other factors. Figure 15 displays the results for the multiples of the number of simultaneously active UEs per small cell, which increases with the LLD deployment. When SCTP is classified as LL traffic, more users can be supported in the small cell in a reliable way even when DOCSIS load percentage increases without requiring architecture or hardware changes as long as DOCSIS 3.1 is deployed with LLD.

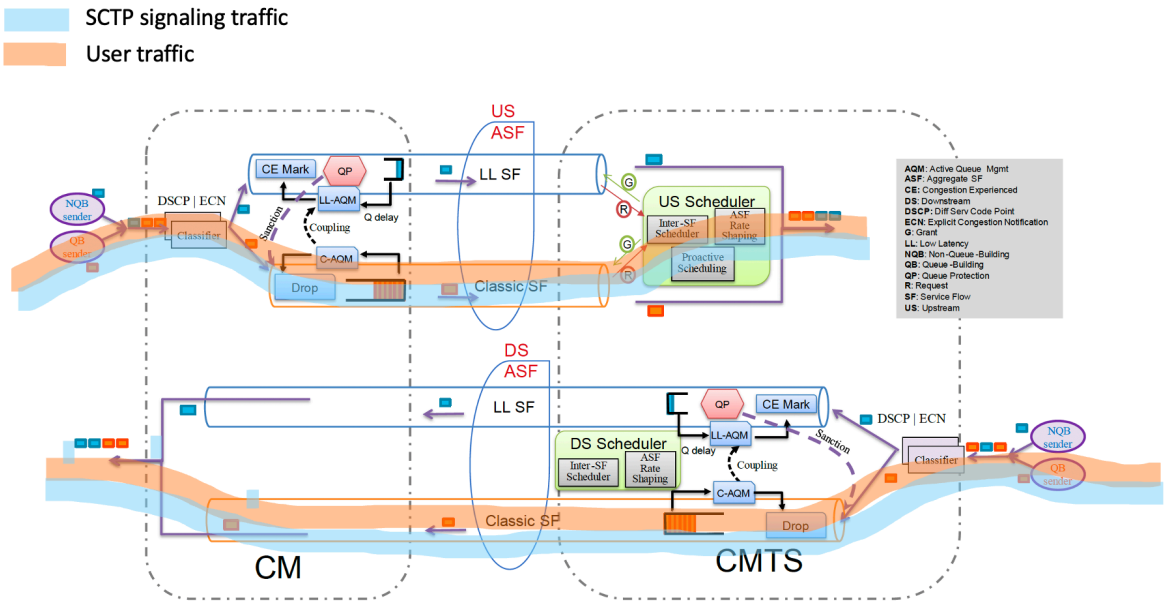


Figure 13- “Without LLD” Use Case

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

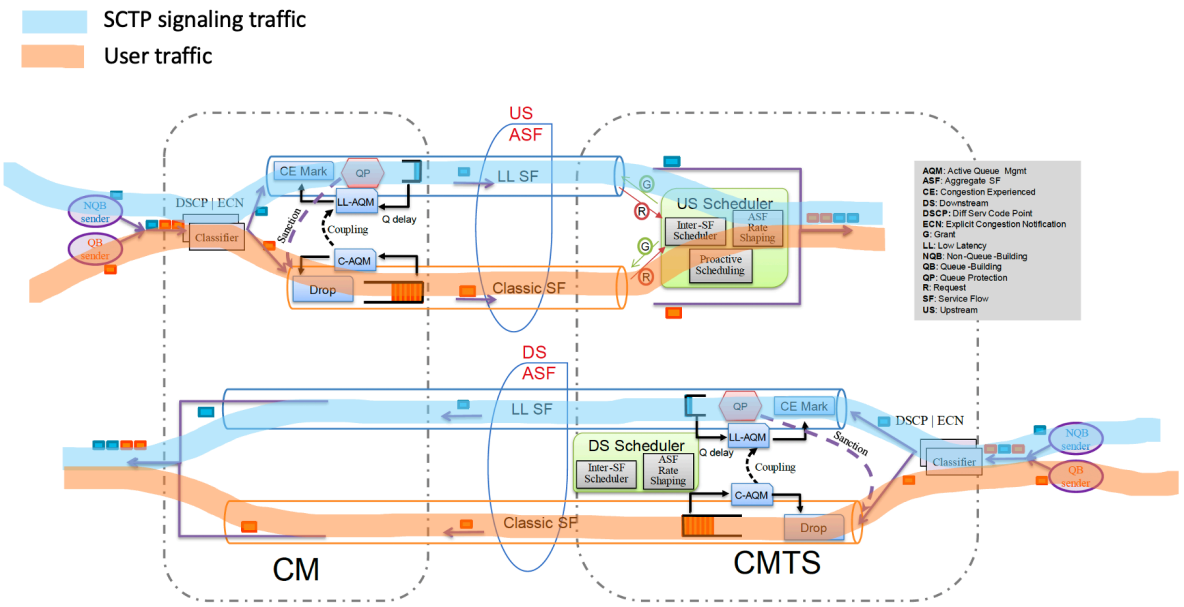


Figure 14- "With LLD" Use Case

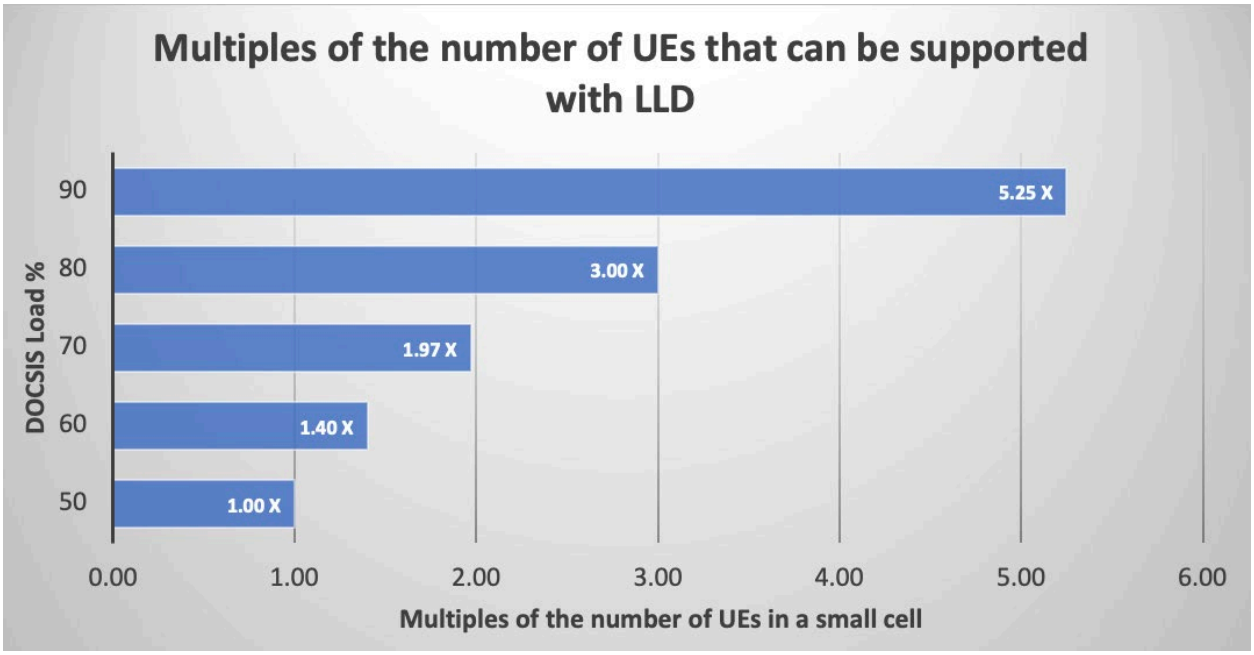


Figure 15- Multiples of the Number of UEs Supported for Different DOCSIS Load Percentages with Additional LLD Firmware in a Midhaul Use Case

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Creating this proof of concept required several critical considerations:

- Mapping SCTP signaling to DOCSIS LL Service Flow
  - SCTP signaling is differentiated in the 5G domain, logically from user data, but normally it is aggregated in the same DOCSIS SF that carries all xhaul traffic
- Measurement of QoS metrics and mapping them to subscriber experience
  - Employing throughput, latency, jitter and other QoS measurement methodologies capable of distinguishing between various traffic categories
  - Analyzing end-to-end KPIs (e.g., service availability/connectivity for UEs)
- Understanding QoS functionalities at different network segments
  - Forwarding of signaling traffic in 5G versus forwarding of DOCSIS Service Flows (e.g., queueing and scheduling)
- Analysis of factors contributing to QoS impairments
  - Identifying impairments specific to 5G versus DOCSIS networks, including considerations like utilization ratios

Our laboratory analysis holds the potential for broad applicability across other traffic categories and their corresponding demands for an enhanced quality of experience. Although our lab procedures included manual steps for testing purposes (e.g., classifying SCTP signaling traffic as LL SF and using different monitoring metrics from vRAN and DOCSIS), MSOs can extend support to generalized use cases by integrating a traffic abstraction layer within the framework of service and network orchestration, a topic covered in a subsequent section.

5. Future-Proofing the Network and Service Platforms

In this section, we discuss several design items that MSOs should consider for future-proof QoE support in heterogeneous networks and service architectures. Although convergence of multiple access technologies and BSS/OSS systems are being investigated [8], functionalities like traffic abstraction can be supported before full convergence is feasible and incorporated to the convergence in the future phases.

One critical design choice is the digitization of network and service components defined with a unified set of metrics so that they can be defined through models that are agnostic of access technology and abstracted from low layer attributes (Figure 16). Network components can be various Virtualized Network Functions (VNF) and Physical Network Functions (PNF) [9]. The service request is then conveyed by using an inventory system of mapped digitized components. Resource attributes are maintained in a common knowledge center for each digitized network resource component. Based on service and resource data, an orchestration layer creates configuration models that can be adaptive to network and resource states. Such a system can provide automated rules to support 5G backhaul service by configuring CM and vCMTS components with corresponding classifiers that are mapped to 5G traffic classification by using programmable and automated systems. Data analytics can be part of the orchestration system that evaluates the service and network states, e.g., by observing UEs that have reliable service with required metrics versus UEs that are denied service or lost connectivity.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

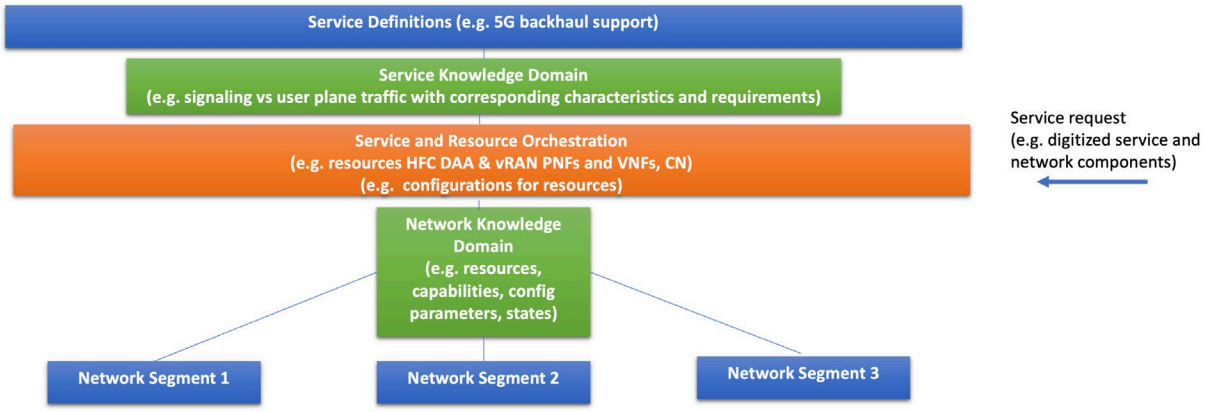


Figure 16- Traffic Abstraction for Automated and Programmable Systems

An essential element in supporting a system, as depicted in Figure 15, is the alignment of standards to ensure convergence. Our focus delves into three key facets of standardization endeavors that guide the mapping from QoS to QoE:

- End-to-end QoS path (Figure 17): Automating traffic classification mapping and translation at network boundaries
    - This translation encompasses various standards such as DOCSIS and PON Service Flows (SFs) and Aggregate SFs, 5G QoS identifiers (5QI), IEEE 802.11's user priority (UP) and similar traffic classifications. Multiple standardization efforts have tackled the concept of end-to-end traffic classification. IETF L4S and NQB-PHB specifications and drafts propose the use of ECN and DSCP bits for low latency service flows, endorsing end-to-end marking utilization. For instance, Low Latency (LL) traffic can be a subset within a DOCSIS ASF or WiFi MultiMedia (WMM) Access Category, correlated to a UP while mapped to a single 5QI. Such, mapping at network segment boundaries, while preserving LL markings, streamlines traffic abstraction. The WiFi Alliance [12] suggests QoS translation/mapping to bolster low latency traffic across wired, WiFi and 5G networks. DSCP mapping is proposed to convert WiFi QoS (UP and WMM ACs) to 5QI. However, the proposal missed an important feature for LL services, which is effective AQM with granular congestion notification.
- End-to-end QoS layering (Figure 17): Employing application marking and MAC-to-transport layer congestion notification
    - Each layer, spanning from PHY and MAC to routing, transport and application, integrates a suite of features aimed at accommodating QoS metrics tailored to diverse traffic characteristics and requirements. The absence of interlayer visibility within this vertical dimension constitutes a barrier to automated traffic management and robust QoE support. The deployment of Low Latency Active Queue Management (LL AQM), paired with precise congestion notification and scalable transport protocols as outlined in IETF L4S specifications, seeks to remove this obstacle. Cablelabs DOCSIS 3.1 specifications encompass L4S support, while both Cable Modem (CM) and Cable Modem Termination System (CMTS) vendors have realized effective implementations. 3GPP Release 18 specifications include L4S ECN marking and API-driven exposure to radio congestion information, with ongoing initiatives to extend this support. Emerging proposals within IEEE 802.11 [13, 14] include L4S integration within the WiFi domain. Achieving end-to-

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

end L4S support holds the potential to exert tangible influence on QoE, given that bottlenecks may manifest across distinct network segments and temporal intervals. To illustrate, if WiFi acts as the bottleneck and only WMM is applied for L4S traffic, accommodating congestion dynamics proves challenging, potentially amplifying fluctuations and diminishing QoE for latency-sensitive, high-rate NQB applications.

- End-to-end QoS monitoring (Figure 18): Harmonizing Key Performance Indicators (KPIs) and telemetry platforms across the entire path and layering
    - Orchestrating traffic abstraction hinges on the efficient and timely collection and analysis of hop-by-hop and end-to-end metrics such as throughput, latency, jitter, packet loss and reliability metrics. Coordinated monitoring of latency and jitter at each network segment serves to discern and forecast the end-to-end state, complemented by scrutinizing additional network and service KPIs for detecting impairments and optimizing performance. This necessitates harmonized telemetry systems that span both dimensions of the network architecture.

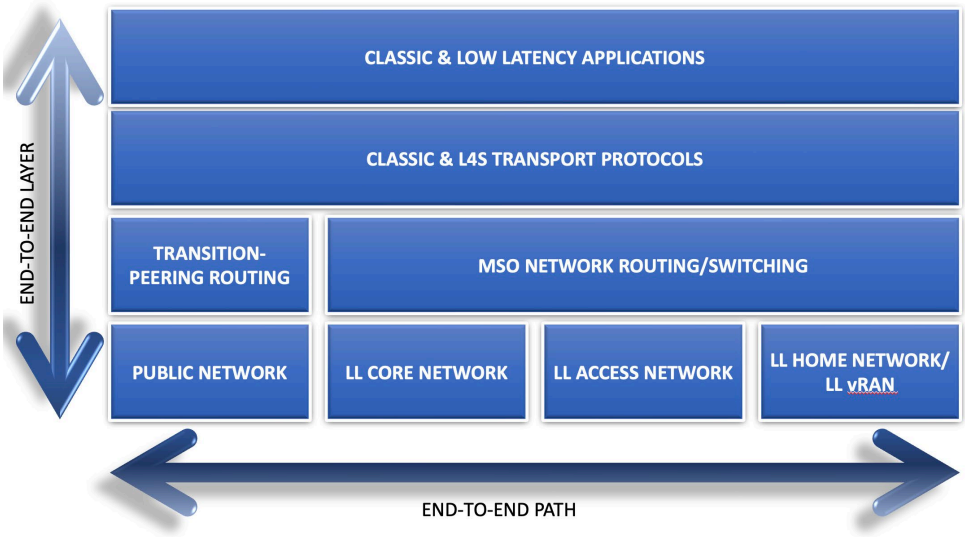


Figure 17- End-to-End Traffic Management

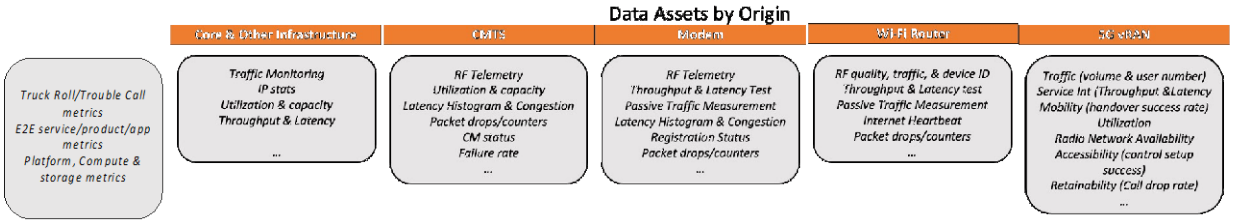


Figure 18- End-to-End and per hop QoS Monitoring: KPI examples

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Conclusion

Distributed architectures incorporating SDN, NFV and cloudification empower MSOs to implement programmable and automated network and service platforms. However, while MSOs utilize a range of network technologies to expand the array of services available to subscribers, these platforms are currently not access technology agnostic (i.e. tailored for specific network such as HFC, PON or Mobile). Conversely, the diverse service landscape mandates a dynamic quality of experience, owing to the ever-evolving nature of residential, enterprise and industrial services.

Consequently, MSOs must embrace access-agnostic, converged platforms that can ensure optimal QoE for a diverse array of services. To address this concern, we delved into the issue by examining low-latency support within MSO networks. We specifically analyzed the viability of IETF L4S and 5G midhaul services over DOCSIS networks, showcasing that MSOs can effectively deploy such services using LLD specifications. This deployment is facilitated by the design of a traffic abstraction architecture, enhancing automation and programmability for efficient operations and a superior QoE.

Abbreviations

AQM	Active Queue Management
ASF	Aggregate Service Flow
CMTS	Cable Modem Termination System
CU	Centralized Unit
DOCSIS	Data Over Cable Service Interface Specification
DU	Distributed Unit
gNB	Next-Generation Node B (5G Base Station)
HFC	Hybrid Fiber-Coaxial
IAQM	Integrated Active Queue Management
IETF	Internet Engineering Task Force
ISP	Internet Service Provider
L4S	Low Latency, Low Loss, and Scalable Throughput
LL	Low Latency
LLD	Low Latency DOCSIS
MIB	Management Information Base
MSO	Multiple System Operators
NFV	Network Function Virtualization
NQB	Non-Queue-Building
ORAN	Open Radio Access Network
OS	Operating System
PHB	Per-Hop Behavior
PNF	Physical Network Function
PON	Passive Optical Networks
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RSSI	Received Signal Strength Indicator
rOLT	Remote Optical Line Terminal

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

RRU	Remote Radio Unit
SCTP	Stream Control Transmission Protocol
SDN	Software Defined Networking
SF	Service Flow
UP	User Priority
vCMTS	Virtual Cable Modem Termination System
vCU	Virtual Centralized Unit
vDU	Virtual Distributed Unit
VNF	Virtualized Network Function
vOLT	Virtual Optical Line Terminal
vRAN	Virtual Radio Access Network

Bibliography & References

[1] *Virtual RAN (V-RAN) Over DOCSIS: DOCSIS As A Midhaul*, Ramneek Bali, SCTE 2023

[2] https://www.bitag.org/documents/BITAG_latency_explained.pdf

[3] *A Virtual Broadband Network Gateway (vBNG) Approach for Cable Operators in a Distributed Access Environment*, Jason Combs, SCTE 2020

[4] *Configuring and Deploying Low Latency DOCSIS Networks*, Greg White and Karthik Sundaresan, SCTE 2021

[5] https://www.statista.com

[6] https://www.excentis.com/blog/latency-in-a-nutshell-and-what-makes-it-so-important/

[7] *Fastest Path to Low Latency Services How Can Cable Operators Deliver Consistent Latency by Following an Efficient and Future-Proof Design Path?*, Sebnem Ozer et. al., SCTE 2021

[8] *How VCMTS Paves The Way For 5G Over DOCSIS Exploring Software-centric Solutions for 5G Xhaul and FMC,* by Brendan Ryan et. al. , SCTE 2021

[9] The Future of Operations: Building a Data-Driven Strategy by Sebnem Ozer, Sinan Onder and Nagesh Nandiraju, SCTE Journal of Network Operations, 2019

[10] IETF RFC 9330 Low Latency, Low Loss, and Scalable Throughput (L4S) Internet Service: Architecture

[11] CableLabs DOCSIS Specification CM-SP-MULPIv3.1-I25-230419

[12] Analyzing operator QoS deployment scenarios between WiFi and 5G, WiFi Alliance, 2022

[13] https://mentor.ieee.org/802.11/dcn/23/11-23-0650-01-0uhr-qos-re-visited.pptx

[14] https://mentor.ieee.org/802.11/dcn/23/11-23-0679-00-0uhr-low-latency-qos-based-on-l4s.pptx

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

[15] https://www.fcc.gov/consumers/guides/getting-broadband-qa

Acknowledgments

The authors would like to thank and acknowledge all those who helped to make this paper possible. This paper includes several tests done by the lab supported by Philip Anderson, Lei Zhou, Jiaxin Zhao, Timothy Welch, Deependra Rawat, and Daniel Lynch.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bringing Us Together: Impacts of Learning Cohorts in Technician Onboarding

An Operational Practice prepared for SCTE by

Abbie E. O'Dell
Senior Director, Learning Services: Field Operations
Charter Communications
6399 S Fiddler's Green Cir
Greenwood Village, CO 80111
720-482-4205
Abbie.odell@charter.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Abstract ... 3
2. Introduction ... 3
3. Literature Review ... 4
   - 3.1. Adult Learning Theories and Practices ... 4
   - 3.2. Learning Communities ... 5
   - 3.3. Learning Transfer ... 5
   - 3.4. Conclusions ... 5
4. Research Methods ... 6
   - 4.1. Participants ... 6
   - 4.2. Qualitative Data Collection ... 8
   - 4.3. Quantitative Data Collection ... 8
   - 4.4. Qualitative Analysis ... 8
   - 4.5. Quantitative Analysis ... 9
5. Results ... 9
   - 5.1. Qualitative Survey ... 9
   - 5.2. Quantitative Attrition and Performance ... 11
6. Discussion ... 12
   - 6.1. Study Limitations ... 12
   - 6.2. Future Research ... 13
7. Conclusion ... 13

Abbreviations ... 13

Bibliography & References ... 13

8. Appendices ... 16
   - 8.1. Appendix A: Cover Letter with Informed Consent ... 16
   - 8.2. Appendix B: Survey Questions ... 16

List of Figures

Title **Page Number**

Figure 1 - Characteristics of Adult Learners ... 4
Figure 2 – Response Summary – Recommendations for Future Use ... 11

List of Tables

Title **Page Number**

Table 1 - Population Data – Total Headcount by Job Title ... 7
Table 2 - Sample Data – Total Headcount for Selected Regions ... 7
Table 3 – Summary of Responses – Question 1 ... 9
Table 4 – Summary of Responses – Question 4 ... 10
Table 5 – Summary of Attrition Rates ... 11
Table 6 – Summary of Scorecard Averages ... 12

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Abstract

The COVID-19 pandemic drastically shifted the landscape of work, transforming how and where employees performed their daily activities. Remote work caused many employees to report feeling a sense of isolation, and organizations have sought to proactively create a sense of community by using technologies such as group chats and video meetings to recreate traditional workplace social interactions. The adult learning experience also changed, with some organizations embracing online learning activities to replace the traditional classroom. Post-pandemic, traditional classroom-based learning has re-emerged, yet training organizations may realize benefits of continuing to use techniques and technologies employed to connect learners during remote work. This mixed method study investigated the experience of learners who were provided a structured cohort chat group (CCG) during their onboarding program and evaluated learners' self-reported satisfaction using this learning method. The study also explored these learners' overall performance and attrition rate compared against a control group of learners who were not provided access to a CCG. The findings indicate a positive relationship between the presence of a CCG and learner performance, attrition rate and overall perception of their onboarding program, suggesting potential for further inquiry into use of this method.

2. Introduction

The impact of the COVID-19 pandemic on learning and education was unprecedented. According to a policy brief from the United Nations (2020), nearly 1.6 billion learners across 190 countries were impacted by closures of schools and learning spaces, impacting 94 percent of the world's student population. Workplace learning was similarly affected, with an early estimate indicating that over one-half of all in-person learning was cancelled or postponed in the early months of the pandemic (Kshirsagar et al, 2020). For learning organizations to remain vital and continue to help learners develop critical skills, most shifted to digital learning methodologies, including virtual delivery, e-learning, videos and online discussion groups. On-the-job training (OJT) increased in 2020, with 62 percent of organizations indicating that they were emphasizing this method; and online delivery methods increased from 26 percent in 2019 to 32 percent in 2021 (ATD Research, 2021).

The learning organizations of major telecommunications operators were similarly impacted by the transition to digital delivery methods because of the pandemic. To create a collaborative and conversational learning environment while still adhering to social distancing protocols, many trainers began to employ the use of instant-messaging groups to enable team conversation amongst class cohorts. These CCGs were formed to support the formal learning experience, and many learners continued to utilize the groups post-training as a social learning space.

Social learning and learning communities have been widely researched and are considered a fundamental aspect of adult learning (Chao, 2009; Knowles, 1980, Wlodkowski & Ginsberg, 2017; Merriam et al, 2007; Kasworm et al, 2010). Connectedness and a sense of belonging is necessary to create an atmosphere for learning, and a strong sense of community within a group of learners creates an inclusive environment where learners can co-construct knowledge (Wlodkowski & Ginsberg, 2017; Lawrence, 2002; Khoo & Cowie, 2020; Abawajy, 2012; Zajac et al, 2022).

Most telecommunications operators have structured classroom training programs for onboarding frontline field technician roles that cover the core skills and knowledge needed to be successful on the job, yet some learners still struggle with learning transfer after completion of the program. When learners do not receive support after formal training completion, the effectiveness of the program may diminish. Organizations often rely upon OJT provided by more tenured or senior employees after completion of the core program, but these types of experiences can have negative consequences if the experienced

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

employee performs contrary to what was presented in training (Kasworm et al, 2010). Enabling a post-training support mechanism for students in the same learning cohort can help overcome this issue by empowering classmates to support each other.

Exploring the relationship between the presence of a structured online communication group used during and after the onboarding training can help workplace learning professionals identify the value of this practice. Instructional design teams can further customize the learning activity design for onboarding programs to include activities and guidance for trainers on the use of these methods. Operational leaders may not recognize the value these methods bring to the learning space, so an examination of the relationship between structured learning cohorts and post-training performance may help justify their use.

3. Literature Review

3.1. Adult Learning Theories and Practices

When asked, most employees in field technical roles would state their learning preference is hands-on, real-world, practical training; and most operational leaders would agree. How then can a chat group in an online format bring value to a learner? Why would such a tool be expected to improve on the job performance, or influence these individuals to stay with a job? An answer may be found in early concepts of andragogy and adult learning theories first explored by Knowles, who presented a series of five foundational characteristics of the adult learner, which have since been incorporated into the collective understanding of adult learning theory (see Figure 1).

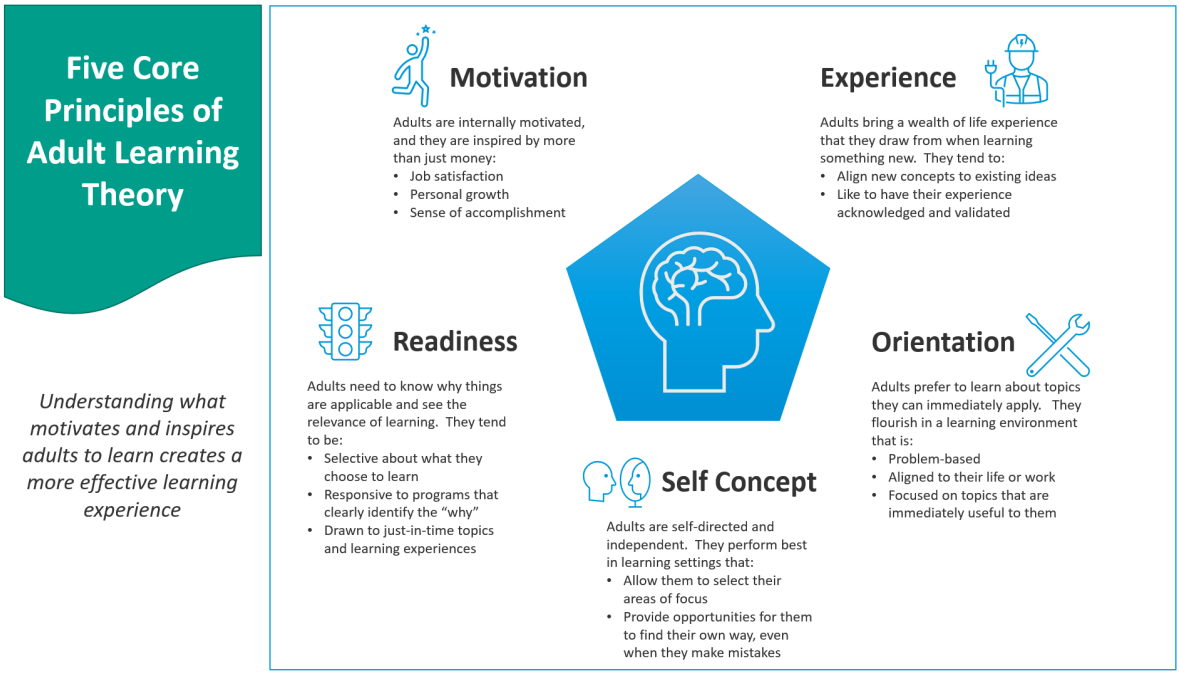


Figure 1 - Characteristics of Adult Learners

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Among these principles are that adults are "problem-centered and interested in an immediate application of knowledge" (Merriam, 2001, p. 5). This concept, known as orientation, provides clear justification for why a new employee would find value in a CCG, where they can quickly connect with classmates to find answers and solve problems they encounter on the job. Learners are motivated to seek out information and utilize learning tools when they believe it will help solve a real-world problem (Wlodkowski & Ginsberg, 2017).

3.2. Learning Communities

A second theory of adult learning that aligns with this research is the concept of learning communities, which has its origins in 1927 with an experimental program at the University of Wisconsin, where students attended a series of courses together over a period of two years and functioned as an intact cohort throughout that time. Since these early forays into this learning method, the concept of learning communities has continued to develop in academic settings and is currently understood to mean any of a variety of different course structures that give learners an opportunity to deepen their understanding of the material through interacting and co-constructing meaning (Gabelnick et al, 1990). Learning communities have been found to benefit learners in a myriad of ways, including improving academic achievement, program retention and positive perception of the learning institution (Cross, 1998; Khoo & Cowie, 2020; Zajac et al, 2022). Despite the benefits of these types of programs, institutions still may struggle to effectively implement such programs, encountering challenges with low adoption rate of tools or resistance from faculty or instructors (Cross, 1998; McPherson & Nunes, 2004).

3.3. Learning Transfer

The fundamental purpose of an onboarding training program is to provide the necessary learning experience that translates into on-the-job performance. In adult learning theory, this concept is called learning transfer: the ability to put the new skills learned in the classroom environment into practice (Roumell, 2019). Foley and Kaiser (2013) define the different levels of learning transfer, including near and far, positive and negative, and high- and low-road transfer, and explore the concept of scaffolding. Scaffolding is a learning concept that provides learners structured tools to enable them to construct their learning. This aligns with the CCG, in that the learner is empowered to work with classmates to construct their own learning experience while still receiving support and guidance (scaffolding) throughout the learning process.

One of the challenges of workplace learning is ensuring that what was learned in the formal classroom environment can be transferred back to the job, and often the design of the program does little to ensure that far transfer occurs after the structured learning. Facilitators often do not have an opportunity to create structure beyond the classroom, and the skills fail to transfer to the real-world environment (Roumell, 2019). Roumell also suggests a key technique to overcoming the challenges of far transfer is to create an environment where learners can engage in "solution-seeking discourse" with classmates in order to deepen their understanding of concepts (p. 20).

3.4. Conclusions

Existing research provides insights into methods that can be used to help ensure that skills and knowledge transfer back to the job. Core adult learning theories on problem-orientation and learning communities have a long history, and more recent works support the value of social learning and highlight some of the challenges with its implementation. Beyond the theoretical considerations, current research describing

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

training methods or practices provides additional perspective, but there is still a gap in the collective body of work. Specifically, most research has been done in the academic setting, where learning communities are used in different learner contexts. Limited research has been done relative to learning communities in the workplace. In addition, most of the exploration of learning communities does not address the virtual environment or, more specifically, the use of cohort-based chat groups.

The purpose of this study is to examine the relationship between the presence of a structured cohort-based learning technology during the onboarding program, and learners' self-reported satisfaction, overall performance and attrition rate. To do so, we pose the following questions:

- Does the presence or absence of a cohort-based learning technology solution impact learners attending field technician onboarding after completion of the course?
    - What are the learner impressions of the use of this technology?
    - What is the relationship to attrition after course completion?
    - What is the relationship to overall job performance (i.e., scorecard)?

4. Research Methods

As detailed in the literature review, there is a lack of current research that directly addresses the topic of CCGs in workplace learning. To investigate the impact of this type of technology on the learner experience, an open-ended qualitative inquiry was needed. This provides an opportunity to better understand the attitudes of learners who experienced the use of a CCG during training. Qualitative study enables insight into the human experience and perspective, which is a foundational element of what this research sought to identify (Creswell, 2013). To provide additional insights into the potential effectiveness of CCGs, quantitative data related to employee retention and job performance was also investigated. Quantitative data provides a more concrete picture of impacts to job performance, helping to validate whether organizations should seek to invest resources in CCGs.

4.1. Participants

The participants in the qualitative portion of the study were frontline field technicians in a large telecommunications organization who completed their onboarding program in the first quarter of 2023. These employees attended a multi-week course that blended online asynchronous coursework with hands-on instructor-led activities. The course provided training on technical topics (e.g., installation and troubleshooting practices), safety topics (e.g., electrical safety and ladder handling) and customer service and support skills.

These individuals were a diverse group of males and females in various locations across the 41 states in which the company operates, with a wide range of ages. Since this study is being performed in a workplace setting where age, gender and race are considered sensitive data, participants were not asked to provide this specific demographic information as part of the study.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 - Population Data – Total Headcount by Job Title

Month Hired	Total Count
January 2023	571
February 2023	478
March 2023	542
	1591

The total population of employees in the field technician role who were hired and attended their onboarding program in the first quarter of 2023 is 1591 (see Table 1). Because this research seeks to explore perceptions of learners who experienced an online cohort group during their onboarding period, the concept sampling method was used. This method allows for intentional selection of study participants to gain a deeper understanding of the specific topic being researched (Creswell, 2003; Creswell & Guetterman, 2019).

Two geographic regions were selected to participate in this study. One was selected because of its status as the only area in which a CCG was used throughout the first quarter of 2023. The second region was selected as the control since they had not implemented any structured or unstructured learning cohorts as part of the field technician onboarding program. For the purposes of this study, the region using the CCG will be labeled as Region A and the region without will be labeled as Region B. It is also important to note that the two regions selected for the study are located in the same geographic area of the United States, where participants can be assumed to have similar social demographics and attitudes.

The training managers for the two selected regions were notified via email of the study opportunity. For Region A, the training manager was given the criteria related to dates of onboarding program attendance and requirement for the existence of a CCG. Using the criteria, the manager provided data for eligible participants to be surveyed on the topic. Region B was not surveyed, since the learners in that area did not experience cohorts; instead, solely the attrition and performance data for learners attending onboarding during the same time period was used. This represents 22 percent of the total staff population that met the study criteria: hired in first quarter of 2023 in the field technician job family. Of that 22%, 12% are from Region A and 9% are from Region B (see Table 2).

Table 2 - Sample Data – Total Headcount for Selected Regions

Region	Headcount	Percent of Total
Region A	197	12.4%
Region B	150	9.4%
All Regions	1591	

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

.

4.2. Qualitative Data Collection

Study respondents from Region A were sent a cover letter which provided background on the importance and purpose of the study, as well as language that fulfilled the need for informed consent and confidentiality considerations based on examples from Creswell & Guetterman (2019). Respondents were given 10 days to complete an online survey using SurveyMonkey® after which time the survey was closed and no new responses were accepted. The survey included closed-ended Likert scale questions and one open-ended question (see Appendix B). Online surveys are an effective method for gathering impression data, since responses need not be captured via a live in-person interview requiring transcription of notes. Secondly, the group being surveyed work differing shifts and an asynchronous survey method avoids scheduling issues and delays in data gathering.

The survey began with an overall query about the participants' relative level of satisfaction with their onboarding experience, using a Likert scale to gather impressions of training material/content, trainer and training facility. This initial question was designed to help the respondents recall their impressions of the overall onboarding experience to prepare them to answer questions more specific to their experience using cohorts.

Participants were then asked whether they were provided a structured chat group via WebEx®. While it had already been confirmed that all surveyed learners were in fact provided a CCG, inclusion of this question allowed for the possibility of a learner who did not recognize or understand that such tool was being provided, which could uncover communication issues within the structured program. Subsequent questions explored learners' perceptions of the value of the chat group to their learning experience, whether they experienced technical issues and if they felt that future learners would benefit from such a tool. The survey concluded with an open-ended question seeking any other feedback on the course overall, giving an opportunity for deeper understanding of the trends from earlier questions (Creswell, 2003; Creswell & Guetterman, 2019).

4.3. Quantitative Data Collection

Quantitative data on attrition and performance for both Region A and Region B were collected from internal systems used to track employee data and performance data. To preserve anonymity of employees, all sensitive or personally identifiable information was removed from the data set prior to analysis. For employment and attrition, data were limited to employees hired during the first quarter of 2023 in the specific job title family investigated. The performance data were filtered using the same criteria. Aggregate data for all regions were prepared to enable comparison of Regions A and B, not only against one another, but also against the company-wide performance in these areas.

4.4. Qualitative Analysis

After the survey concluded, the data was exported into a spreadsheet from the survey system. Each respondent was assigned a number (e.g., Respondent 1), which was used throughout the rest of the study to notate and track that individual's response. The closed-ended questions were coded using Likert scale responses (e.g., 5 – Strongly agree, 4 – Agree, 3 – Neither agree nor disagree, 2 – Disagree, 1 – Strongly disagree) (see Appendix B for full question and answer details). Each closed-ended question also included an optional open-ended question (e.g., "Please explain"), and those responses were reviewed and coded for themes.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.5. Quantitative Analysis

To investigate trends related to attrition (i.e., job departure), employee data was exported into a spreadsheet and all personally identifiable information was removed, leaving only the data fields for job title, hire date, departure date (if applicable) and region. This dataset was then organized using pivot tables to calculate the numbers of employees hired during the specific period and those who were still currently employed at the time of the research. The number of employees from the sample who departed the company was converted into a percentage, indicating an attrition rate.

For performance data, individual employee scores relative to the standard company scorecard were used. The standard scorecard for this job family uses multiple discrete performance measures (e.g., quality, productivity) that are combined into a tiered 1-5 overall score. Employee performance scores were calculated for each of the two regions considered in this study, and an aggregate company-wide comparison score was also prepared.

5. Results

5.1. Qualitative Survey

Average time to complete the survey was three minutes and nineteen seconds; the longest time was five minutes and thirty-nine seconds for Respondent 3; and the shortest time was one minute and thirty seconds for Respondent 2. Of the 40 participants invited to complete the survey, five responded, for an overall response rate of 12.5%. All respondents indicated that they were provided a structured chat group via WebEx®.

Participants were asked to provide overall feedback on their experience in their onboarding course, using a Likert scale (5 – Very Satisfied, 4 – Satisfied, 3 – Neutral, 2 – Somewhat Satisfied, 1 – Not Satisfied) in the categories of training material and content, trainer, training facility and classmates. The mean score across all categories was 4.55, with the highest categories tied at 4.6 between trainer, training facility and classmates (see Table 4).

Table 3 – Summary of Responses – Question 1

Question Text: Consider the experience of your Field Technician Onboarding program as a whole. Please provide your feedback on the aspects of the program listed here.				
Respondent ID	**Training Material and Content**	**Trainer**	**Training Facility**	**Classmates**
5	5	5	5	5
4	5	5	5	5
3	4	4	4	5
2	4	5	5	4
1	4	4	4	4
	4.4	4.6	4.6	4.6
				4.55

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Question 3 asked the participants to indicate their perception of the value of the structured group chat (CCG). The mean response was 4.6, with three participants responding "Extremely valuable" and the remaining two responding "Somewhat valuable." Respondent 4 provided a comment in the open-ended field for question 3, stating, "Everyone has a different experience in the ride along portion of training and sharing those individual experiences with class mates (sic) has taught me about things I didn't get that chance to see in person."

In question 4, participants were asked to consider their experience using the CCG and indicated their perception of the importance of factors including staying connected to classmates during and after class, asking questions of classmates or trainer and feeling part of a learning community. The mean score for all categories was 4.32, with "asking questions of the trainer" the lowest at 3.8 and "feeling part of a learning community" the highest at 4.6 (see Table 4). Respondents 4 and 2 both indicated a greater importance on staying connected to classmates after class, while Respondent 1 rated usage after class significantly lower than during class. The lowest score in any category (2 – Somewhat important) was from Respondent 3, in relation to the value of being able to ask questions of their trainer.

Table 4 – Summary of Responses – Question 4

Question Text: Consider your experience using the chat group with your classmates. Please review the list below and indicate the level of importance to you.					
Respondent ID	**Staying connected to my classmates during class**	**Staying connected to my classmates after class**	**Being able to ask questions and get answers from my classmates**	**Being able to ask questions and get answers from my trainer**	**Feeling part of a learning community**
5	5	5	5	5	5
4	4	5	5	4	5
3	4	4	4	2	4
2	4	5	4	4	5
1	5	3	4	4	4
	4.4	**4.4**	**4.4**	**3.8**	**4.6**
					4.32

Question 5 sought to determine whether participants experienced any issues or problems with the chat group, including whether the group was distracting, if usage amongst classmates was limited or if there were problems with the software. All participants indicated they did not experience problems in these categories, with the exception of Respondent 3 who noted that they did experience software or connection problems but did not elaborate in the comments.

Participants were then asked whether they would recommend the group chat feature be added for future classes, which is a modified version of a promoter question. Eighty percent (or four out of five) participants indicated yes, while Respondent 3 indicated they were not sure (See Figure 2). Respondent 5 added a comment "Very helpful vreates (sic) a good work environment and needing the trainers as well for structure and conversation control." Respondent 4 commented "Sups can get busy and it's always good to have other people to reach out to for help."

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

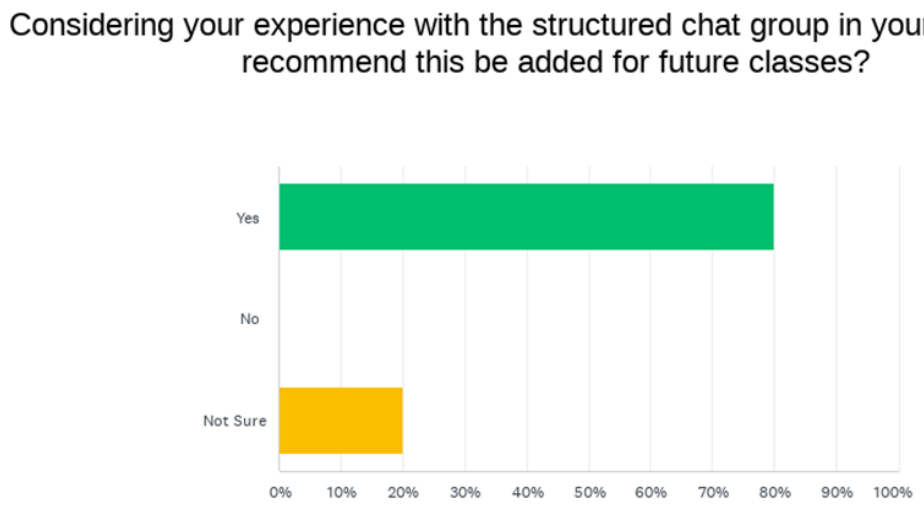


Figure 2 – Response Summary – Recommendations for Future Use

5.2. Quantitative Attrition and Performance

An attrition rate was calculated by dividing the number of former employees (those having departed the company at the time of the study) into the total number of employees hired during the first quarter of 2023, expressed as a percentage. This was separated into an aggregate score for all regions, and two individual scores for the two regions considered in this study. Region A, the region that included CCGs as part of the onboarding experience, had a 23% attrition rate for the sample group. Region B, without such groups, had a 33% attrition rate, or 10 percentage points higher loss of employees for this same time period and group. Comparatively, the overall aggregate rate for the full sample of all regions was 25%, meaning Region A outperformed the overall company score by 2% (see Table 5).

Table 5 – Summary of Attrition Rates

	Current	Former	Grand Total	Attrition Rate
Region A	151	46	197	23%
Region B	100	50	150	33%
All Regions (Aggregate)	1188	403	1591	25%

Performance data were compiled using standard scorecard measures, which uses a tiered 1-5 scoring system for employee performance. This was separated into an aggregate score that included performance data for individuals in all regions, and two individual scores for the two regions considered in this study. Region A, the region using CCGs during the onboarding program, had an average of 3.0 for employees in the sample group, while Region B had an average of 2.9 (Table 6). Similar to the findings in the attrition data, Region A not only outperformed the control (Region B), but also outperformed the aggregate score for all regions.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 6 – Summary of Scorecard Averages

Region	Scorecard Average
Region A	3.0
Region B	2.9
All Regions	2.9

6. Discussion

A clear pattern of favorability emerged from the qualitative survey responses, with the overall mean of all questions scoring greater than 4. Cross (1998) indicates the greatest value in the learning community is that students have the ability to connect with their classmates and foster a strong sense of connection. Respondents to the study indicated the greatest value in their ability to connect with classmates, both during and after the formal classroom period. Respondents also indicated that feeling part of a learning community was even more valuable than being able to get answers from their trainer. These responses align to the concepts and ideas found in the literature, and specifically with findings of the efficacy of learning communities in academia, where learners indicate a more positive opinion of their learning institution and program overall when provided with this type of support (Khoo & Cowie, 2020; Zajac et al, 2022). The quantitative data also indicates that the presence of a CCG could have a positive relationship to attrition rates, with Region A significantly outperforming not only the control region considered in the research, but also the aggregate score for the company-wide sample. The literature indicates when learning communities are used in academic settings, learners are more likely to stay in their program of study and less likely to drop out (Price, 2005).

Roumell (2019) suggests learning professionals must create structure in the post-training environment in order to facilitate effective learning transfer, and the positive impact to the scorecard of learners in Region A may be a result of their structured CCG. While the difference between Region A and Region B is comparatively much smaller than the attrition rate, it still does indicate a possible positive relationship. This performance improvement is also supported by research, where learners in academic settings who have this type of community with their classmates consistently demonstrate higher grades and overall academic performance (Zajac et al, 2022).

6.1. Study Limitations

There are at least two potential limitations related to the scope of this research. A first limitation relates to the number of respondents to the survey, which represents only a small sample of the overall population of employees in the job role. Intentional selection of a smaller sample of learners who were known to have experienced a structured chat group cohort provided greater understanding of the central phenomenon, but surveying a larger group, including both those who have and have not experienced this type of tool, could uncover additional perspectives. A second potential limitation relates to the respondents of the survey, and the selection of only learners. By using the concept of triangulation and gaining perspectives from different sources (e.g., supervisors or trainers), the validity of the data analysis could be improved.

Despite these limitations, the results suggest both a learner preference for the inclusion of this type of structured learner cohort using chat groups, and a potential positive impact on both performance and attrition rate. The implication to organizations who provide onboarding training is that the presence of this type of tool can help create durable learning experiences that improve trainees' confidence post-training and increase their desire to stay with the organization.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6.2. Future Research

Although the findings of this particular study support the value of structured learning cohorts in the field technician onboarding space, the most important contribution may be that it raises awareness of this instructional method and creates an opportunity for further inquiry. If, as this study suggests, the use of a CCG improves learner performance, attrition and overall experience of onboarding, organizations would be wise to consider incorporating such methods. A recommendation for future studies would be to broaden the inquiry to a larger group of participants and consider investigating the experience of employees beyond the field technician role. Learning organizations seeking to implement such measures will need to prepare a clear and compelling business justification in order to obtain support from business leaders; and further research by learning professionals will be critical to help guide future use of these methods.

7. Conclusion

The findings of this research indicate that learner experience during onboarding is positively impacted by the presence of a structured learning cohort using chat groups, and that performance and employee retention may be positively influenced as well. Concepts from the literature, including orientation, learning transfer and learning communities, were identified in the participant responses, and quantitative performance and attrition data align with existing research in academic settings. Although future study will be needed to further explore this topic, the present study has enhanced the understanding of the relationship between structured learning cohorts in onboarding and provided clear support for the value of this instructional method.

Abbreviations

CCG	Cohort chat group
OJT	On-the-job training
ATD	Association for Talent Development

Bibliography & References

ATD RESEARCH. (2021). State of the industry 2021.

Abawajy, J. (2012). Analysis of asynchronous online discussion forums for collaborative learning (Version 1). Deakin University. https://hdl.handle.net/10536/DRO/DU:30049653

Bandura, A. (1977). Social learning theory. Prentice-Hall.

Creswell, J. W. (2003). Research Design: Qualitative, quantitative, and mixed methods approaches (2nd ed.). Sage Publications, Inc.

Creswell, J. W. (2013, February 19). What is mixed methods research? [Video]. YouTube. https://https://www.youtube.com/watch?v=1OaNiTlpyX8

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Creswell, J. W. & Guetterman, T. C. (2019). Educational research: Planning, conducting, and evaluating quantitative and qualitative research. Pearson Education, Inc.

Cross, K. P. (1998). Why learning communities? Why now?. About campus, 3(3), 4-11.

Gabelnick, F., MacGregor, J., Matthews, R. S. & Smith, B. L. (1990). Learning communities: Creating connections among students, faculty, and disciplines. New Directions for Teaching and Learning, 41. San Francisco, Jossey-Bass, Inc.

Kasworm, C.E., Rose, A.D., & Ross-Gordon, J.M. (Eds.). (2010). Handbook of adult and continuing education. Thousand Oaks, CA: Sage Publishing.

Khoo, E. & Cowie, B. (2020). A framework for developing and implementing an online learning community. Journal of Open, Flexible, and Distance Learning, 24(1), 17–29.

Knowles, M. S. (1980). The modern practice of adult education: From pedagogy to andragogy (2nd Edition) . New York: Cambridge Books.

Knowles, M. S., Holton III, E. F., & Swanson, R. A. (1998). The adult learner: The definitive classic in adult education and human resource development. Gulf Publishing.

Kshirsagar, A., Mansour, T., McNally, L., & Metakis, M. (2020, March 17). Adapting workplace learning in the time of coronavirus. McKinsey & Company. https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/adapting-workplace-learning-in-the-time-of-coronavirus#/

Lawrence, R. L. (2002). A small circle of friends: Cohort groups as learning communities. New Directions for Adult and Continuing Education, 2002(95), 83–92. https://doi.org/10.1002/ace.71

Maslow, A. H. (1943). A Theory of Human Motivation. Psychological Review , 50, 370-396.

Merriam, S. B. Caffarella, R. S. & Baumgartner, L. M. (2007). Learning in Adulthood: A Comprehensive Guide (3rd Edition). San Francisco: Jossey-Bass.

Merriam, S., & Bierema, L. (2014). Adult learning: Linking theory and practice. John Wiley & Sons, Incorporated.

McPherson, M., & Nunes, M. B. (2004). The failure of a virtual social space (VSS) designed to create a learning community: Lessons learned. British Journal of Educational Technology, 35(3), 305-321.

Mitsakis, F., Karageorgakis, T. (2020). E-learning: A Temporary 'By-Product' of Covid-19 Pandemic or a Contemporary Solution to Workplace Training and Learning?. In: Loon, M., Stewart, J., Nachmias, S. (eds) The Future of HRD, Volume I. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-52410-4_7

Pokhrel, S., & Chhetri, R. (2021). A Literature Review on Impact of COVID-19 Pandemic on Teaching and Learning. Higher Education for the Future, 8(1), 133–141. https://doi.org/10.1177/2347631120983481

Price, D. V. (2005). Learning communities and student success in postsecondary education. A background paper. MDRC.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Roumell, E. A. (2019). Priming adult learners for learning transfer: Beyond content and delivery. Adult Learning, 30(1), 15-22.

Wlodkowski, R. J. & Ginsberg, M.B. (2017). Enhancing adult motivation to learn: A comprehensive guide for teaching all adults (4th ed.). San Francisco, CA: Jossey Bass.

Yarberry, S., & Sims, C. (2021). The Impact of COVID-19-Prompted Virtual/Remote Work Environments on Employees' Career Development: Social Learning Theory, Belongingness, and Self-Empowerment. Advances in Developing Human Resources, 23(3), 237–252. https://doi.org/10.1177/15234223211017850

Zajac, S., Randall, J., & Holladay, C. (2022). Promoting virtual, informal learning now to thrive in a post-pandemic world. Business and Society Review, 127(S1), 283– 298. https://doi.org/10.1111/basr.12260

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

8. Appendices

8.1. Appendix A: Cover Letter with Informed Consent

As a Field Technician, you have valuable insights on the methods used during training to connect classmates with real-time convenience.

We are conducting a research study to investigate whether having a group WebEx® chat with classmates has a positive or negative impact on the learning experience.

To gather data on this topic, I am using an online survey in SurveyMonkey®, and your leadership recommended you as a great candidate to provide your perspective on this! Your participation in this research is, of course, completely voluntary. Responses to the survey are totally anonymous and will not be linked back to you directly.

If you choose to participate in the survey, simply use the link below to access and provide your responses. The survey should take about 5-10 minutes to complete, and you can candidly provide as much detail as you feel appropriate. The survey will be available for completion through Monday, July 24th, after which it will be closed.

Click here to access survey.

Should you have any questions about the study, the survey, or the research in general - please feel free to contact me and I would be happy to provide additional information!

Thank you for your time and consideration!

8.2. Appendix B: Survey Questions

Question #	Question Text	Answer Options
1	Consider the experience of your Field Technician Onboarding program as a whole. Please provide your feedback on the aspects of the program listed here. • Training Material and Content • Trainer • Training Facility • Classmates	5 - Very Satisfied 4 – Satisfied 3- Neutral 2 - Somewhat Satisfied 1 - Not Satisfied

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2	As part of the onboarding experience, were you provided a structured new hire chat group with your classmates via WebEx®?	2 – Yes 1 - No
3	You indicated that you WERE provided a structured new hire chat group with your classmates. Please rate your opinion on how helpful/useful you found this.	5 - Extremely valuable 4 - Very valuable 3 - Somewhat valuable 2 - Not so valuable 1 - Not valuable at all
4	Consider your experience using the chat group with your classmates. Please review the list below and indicate the level of importance to you. • Staying connected to my classmates during class • Staying connected to my classmates after class • Being able to ask questions and get answers from my classmates • Being able to ask questions and get answers from my trainer • Feeling part of a learning community	5 - Extremely important 4 – Important 3 – Neutral 2 - Somewhat important 1 - Not important at all
5	Did you experience any of the following problems or issues with the chat group? Please provide feedback on each item. • Software or connection problems • Chat messages were distracting to me • Could not get the answers I needed in a timely fashion • Little or no use of the chat group by classmates	2- Yes 1 - No
6	Considering your experience with the structured chat group in your class, would you recommend this be added for future classes?	2 – Yes 1 – No 0 - Not sure
7	Considering your training experience as a whole, are there any suggestions/recommendations you have to improve the experience for others? Please provide as much detail as you like.	Open-ended

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Cable Modem Transmit Headroom Resiliency Management

A Technical Paper prepared for SCTE by

Ron Hranac
Chair, SCTE Network Operations Subcommittee
rhranac@aol.com

Roger Fish, Master Software Engineer, Broadcom

Tom Kolze, R & D Advanced Technology Development Engineer, Broadcom

Satish Mudugere, Senior Principal Engineer, MaxLinear

Alexander Podarevsky, CTO, Promptlink Communications

Jason Rupe, Distinguished Technologist, CableLabs

Foad Towfiq, CEO, Promptlink Communications

Sheldon Webster, Lead Architect – DOCSIS, CableLabs

Larry Wolcott, Fellow Engineer, Comcast

Lei Zhou, Principal Engineer, Charter

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 5
2. Cable Modem Upstream Transmit Power 6
2.1. Principal DOCSIS 3.1 Cable Modem Transmit Power Requirements 6
2.2. Dynamic Range Window 7
3. What is Cable Modem Transmit Headroom? 7
3.1. Cable Modem Total Composite Power Headroom 7
3.2. Cable Modem Minimum Channel Headroom 8
3.3. DOCSIS 3.0 Example 9
3.4. DOCSIS 3.1 Example 11
4. Factors That Can Affect Transmit Headroom 13
4.1. Transmitted RF power 13
4.1.1. Total Composite Power 13
4.1.2. Cable Modem Termination System Set Point 13
4.1.3. Transmit Channel Set Occupied Bandwidth 13
4.2. Dynamic Range Window 14
4.3. Long Loop Automatic Level Control 14
4.3.1. Additional Loss in Return Path 15
4.3.2. Coaxial Cable Attenuation Versus Temperature 16
4.4. Upstream Pre-Equalization 17
4.5. Transmit Channel Set Occupied Bandwidth (Power Spectral Density Versus RF Bandwidth) 21
4.5.1. Example: Adding Channels to the Transmit Channel Set 22
5. Management of Modem Transmit Power in Cable Network Operation 24
5.1. Ways to Manage Cable Modem Transmit Headroom 24
5.2. Headroom Management Knobs 26
6. Create Your Own Spreadsheet 26
6.1. Simple Headroom Calculation for DOCSIS 3.0 26
6.2. Mixed Channels Headroom Calculation for DOCSIS 3.1 27
7. Additional Work Underway 29
8. Conclusion 29

Appendix A 31
9. What is Power Spectral Density? 31
9.1. Power and Power Spectral Density 31

Appendix B 33
10. Dynamic Range Window 33

Appendix C 35
11. Why Pre-Equalization Does Not Impact Transmit Power, When All is Working as Intended 35
11.1. Cable Modem Transmit Power Adjustment in DOCSIS PHYv3.1 35
11.2. The Pre-Equalization Vector is Normalized to Unity Power and Application of Pre-Equalization Does Not Alter the Cable Modem Transmit Power 36

Abbreviations 38

Bibliography & References 39

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Figures

Title **Page Number**

Figure 1 - TCP distribution curve for non-OFDMA enabled modems. 5
Figure 2 - TCP distribution curve for SC-QAM plus OFDMA-enable modems. 5
Figure 3 - Cable modem Min Ch headroom is 7 dB in this example. 9
Figure 4 – Here, the modem's transmitted per-channel signal level is equal to the maximum per-channel signal level P_{hi_n}, resulting in 0 dB Min Ch transmit headroom. 10
Figure 5 - In this example the Min Ch transmit headroom is 7 dB and the TCP headroom is 10 dB (see text). 10
Figure 6 - The TCP headroom and Min Ch headroom in this example are 6.8 dB and 5.2 dB respectively; see text. 11
Figure 7 - The TCP headroom and Min Ch headroom in this example are 11.6 dB and 3.0 dB respectively; see text. 12
Figure 8 - Baseline upstream signal level scenario for Internet-only subscriber. Here the net upstream attenuation between the modem and CMTS is 43 dB. 15
Figure 9 - The newly-installed four-way splitter adds 7 dB of attenuation to the drop, which reduces downstream and upstream levels. Note that the input to the CMTS is now –7 dBmV instead of the desired 0 dBmV. The net upstream attenuation between the modem and CMTS has increased to 50 dB. 15
Figure 10 - The station maintenance process commands the modem to increase its upstream transmit level by 7 dB to overcome the 7 dB attenuation of the newly-installed splitter. The net upstream attenuation between the modem and CMTS remains at 50 dB. 16
Figure 11 - Channel frequency response before and after onset of an impairment. 18
Figure 12 - Burst receiver input power per subcarrier, a) before onset of an impairment; b) after the impairment impacts the channel but prior to a ranging, where about 5 dB of loss occurs across most of the channel and up to 25 dB loss in the suckout; c) the average power per subcarrier after the impairment occurs; and d) after the first ranging response, which increases the transmit power by 4.6 dB and equalizes the channel with pre-equalization. 19
Figure 13 - The cable modem transmit power per subcarrier a) prior to the onset of the impairment; and b) after the first ranging response following the onset of the impairment, with the average power per subcarrier indicated by the dotted line. Also shown is the normalized pre-equalization of the first ranging response, seen to vary from about –12 dB to about 12.5 dB, with most of the channel pre-equalization hovering around –10 dB. 20
Figure 14 - The total power of four 6.4 MHz-wide SC-QAM channels in this figure, each with a per-channel signal level of 46.02 dBmV, is 52.04 dBmV. 22
Figure 15 - In this example a 25.6 MHz-wide OFDMA channel has been added to the upstream spectrum. 23
Figure 16 - Example mid-split distribution design using conditioned taps 25
Figure 17 - Simple spreadsheet to calculate DOCSIS 3.0 cable modem TCP headroom and Min Ch headroom. 27
Figure 18 - Another example spreadsheet for calculating modem transmit headroom. 28
Figure 19 - Example showing four SC-QAM channels with different bandwidths but identical PSD 32
Figure 20 - Example showing four SC-QAM channels with different bandwidths but identical per-channel power. 32

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Tables

Title **Page Number**

Table 1 - Before and after headroom performance. 14
Table 2 - Attenuation versus temperature for 1500 feet of 0.500 inch hardline cable. 17
Table 3 - Attenuation versus temperature for 100 feet of Series 6 drop cable. 17
Table 4 - Before and after headroom performance. 23
Table 5 - Formulas for spreadsheet in Figure 17. 27
Table 6 - Formulas for spreadsheet in Figure 18. 29

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Cable operators are extending the operating frequency range of their cable networks, including the upstream. Expanding the upstream operating bandwidth and adding more channels brings with it a variety of challenges, including an impact on cable modem (CM) upstream transmit power capability and headroom. In particular, the modem's available transmit power spectral density (PSD) is reduced because it must be spread over a wider radio frequency (RF) bandwidth. Cable operators face new challenges for managing upstream power in the cable network, because of a complicated system of transmitted RF power, dynamic range window (DRW), long loop automatic level control (ALC), pre-equalization settings, and channel bandwidth. If a CM's transmitted power is insufficient, then forward error correction (FEC) errors can result, and bit loading may need to be reduced. As well, modems can go into partial service, resulting in those modems being unable to achieve advertised upstream speeds.

Figure 1 illustrates the total composite power (TCP) distribution for a population of CMs transmitting only single carrier quadrature amplitude modulation (SC-QAM) channels in the upstream. The horizontal axis in this and the next figure is TCP in units of decibel millivolt (dBmV), increasing right-to-left.

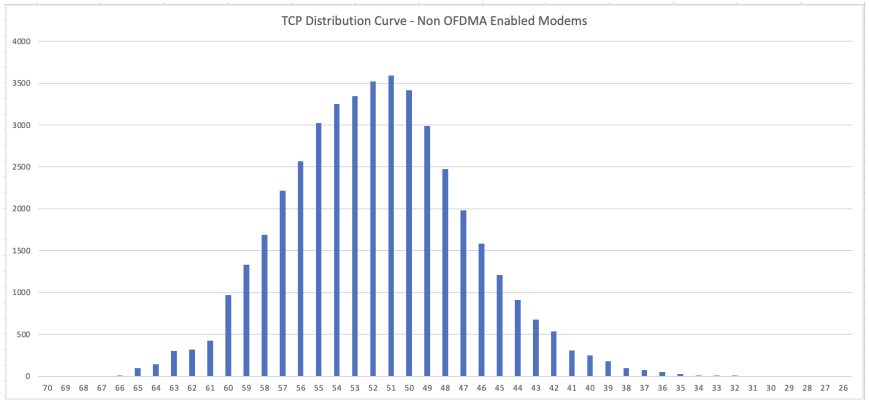


Figure 1 - TCP distribution curve for non-OFDMA enabled modems.

Figure 2 shows the TCP distribution for a population of CMs transmitting SC-QAM *and* orthogonal frequency division multiple access (OFDMA) signals. Note the shift of the curve to the left; why is this important? As the TCP increases, the transmit headroom can decrease because more modems are transmitting closer to their maximum capability.

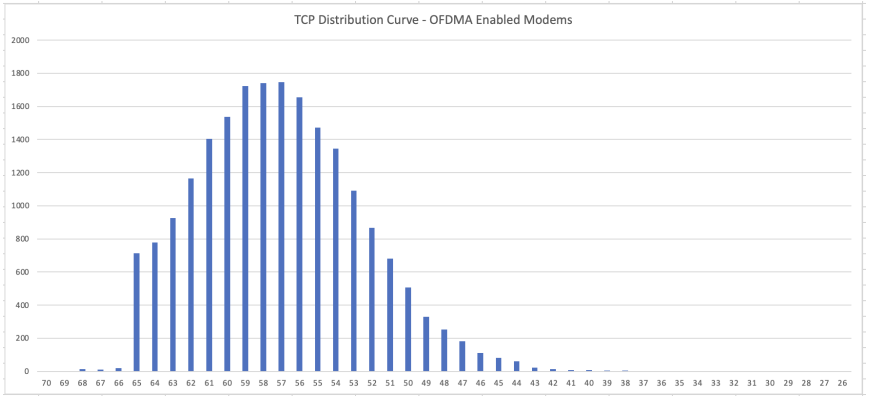


Figure 2 - TCP distribution curve for SC-QAM plus OFDMA-enable modems.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Therefore, it is incumbent upon cable operators to manage CM transmit power so that service quality is maintained. Understanding the impact of impairments in the spectrum and how they relate to the amount of transmit power "reserve" available to the CM, operators are able to not only grasp the urgency and prioritization of proactive repair in these cases, but also the impact on service, so they can manage this important resiliency mechanism.

This paper and its companion presentation discuss the latest understanding of the dynamic system that is the CM's upstream RF transmission operation and how it can impact service quality. Also included is guidance to create a spreadsheet for modeling available cable modem transmit headroom.

2. Cable Modem Upstream Transmit Power

Data-Over-Cable Service Interface Specifications (DOCSIS®) CM upstream transmitters are designed to accommodate a wide range of net attenuation[1] between each modem and the upstream burst receiver circuitry in a cable modem termination system (CMTS), converged cable access platform (CCAP), or remote PHY device (RPD). Net attenuation in this context includes the combination of all upstream active device gains and coaxial cable and passive device losses in the RF signal path. Further complicating net attenuation is the impact of ambient temperature changes on coaxial cable attenuation from day to night, season to season, and so forth.

2.1. Principal DOCSIS 3.1 Cable Modem Transmit Power Requirements

The minimum highest value of the total power output of the cable modem P_{max} is 65 dBmV, although higher values are allowed.

A CM is typically configured to transmit on a number of channels, but the number of channels and amount of spectrum in those channels varies from node to node and possibly even modem to modem. Upstream occupied bandwidth is defined in [D3.1PHY] as spectrum the modem may be required to transmit within its current upstream channel configurations. A number of "equivalent DOCSIS channels" is calculated from the occupied bandwidth (BW_{legacy} and BW_{OFDMA}), which is, informally, the number of 1.6 MHz "chunks" of spectrum that is configured for possible transmission by the modem. The total number of DOCSIS equivalent channels configured for a CM is denoted as N_{eq}.

The maximum PSD for the CM derives from when the maximum modem transmit power is used, and is a function of how much spectrum the modem may be required to transmit within its current configuration (that is, N_{eq}). In DOCSIS 3.1 the cable modem PSD is dealt with on a "per 1.6 MHz" basis, and is referred to as the equivalent channel power.

The maximum PSD for the CM, depending on its configured upstream occupied bandwidth, is its maximum equivalent channel power, and is denoted as $P_{1.6hi}$ in DOCSIS 3.1. $P_{1.6hi}$ is calculated as follows:

$$P_{1.6hi} = P_{max}[dBmV] - 10\log_{10}(N_{eq})$$

Each channel in the transmit channel set (TCS) is described by its reported power $P_{1.6r_n}$, which is the channel power when it is fully granted, divided by the occupied bandwidth of the channel (in units of MHz), multiplied by 1.6. The result is called "the power per 1.6 MHz" of the channel.

[1] Net upstream attenuation is the combination of all gains and losses between a modem and the input to the CMTS. For instance, if a cable modem's upstream transmit level is +45 dBmV and the input to the CMTS is 0 dBmV, the net upstream attenuation is 45 dBmV – 0 dBmV = 45 dB.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.2. Dynamic Range Window

The concept of DRW was introduced in [D3.0PHY] and carried over in later versions of DOCSIS (refer to Appendix B for more information about DRW). It could be stated that the primary purpose of DRW is to prevent the channel in the TCS with the lowest PSD from being "too far" below the channel with the highest PSD. This is important because the modem has a limited dynamic range capability.

The CMTS is expected to manage the modem's commanded transmit power levels and DRW. Regardless of a DRW violation, the CM has to attempt to comply with what the CMTS commands. Typically, a CMTS may drop from a CM's TCS a channel with a PSD that is too low. It is always possible that a CMTS might operate the high PSD channel at a lower transmit power or the low PSD channel at a higher transmit power to prevent DRW violation. This will result in those channels reaching the burst receiver at a lower or higher PSD than the set point (aka the burst receiver target PSD).

Much more detail is given about the DRW in Appendix B.

3. What is Cable Modem Transmit Headroom?

Generally speaking, a CM's upstream transmit headroom is the difference, in decibels, between that modem's maximum transmit power capability and its actual transmit power. But in DOCSIS the CM transmit headroom really has two different interpretations or definitions: 1) CM TCP headroom, and 2) what we call CM minimum channel (Min Ch) headroom. These two headroom values comprise two different figures-of-merit to characterize a CM's commanded channel power compared to its transmit power capability. Note: In a plant with tilt and two or more upstream channels, the Min Ch headroom will be the "stop" before the TCP headroom. Additionally, each channel of a CM has its own figure-of-merit in this regard: CM channel headroom. These figures-of-merit are explained in the following sections.[2]

3.1. Cable Modem Total Composite Power Headroom

The CM TCP headroom is the difference between the maximum total composite power *capability* of the CM, P_{max} in [D3.1PHY], and the total composite power currently commanded (called "reported power" in the DOCSIS specs) to the CM's TCS by the CMTS. In [D3.1PHY] the transmit power commanded to the nth channel is the product of the PSD commanded to the nth channel, $P_{1.6r_n}$, and the occupied bandwidth of the nth channel divided by 1.6 MHz:[3]

$$Power\ in\ n^{th} channel\ (dBmV) = PWR_n$$

$$= P_{1.6r_n}(dBmV) + 10log_{10}\left(BW_{legacy\ or\ OFDMA}(MHz)/1.6\ MHz\right).\ \{D3.1\}$$

$$= P_{r_n}\ \ \{D3.0\}$$

The commanded TCP of the CM in both [D3.1PHY] and [D3.0PHY] is the sum of the commanded power of all of the channels:

[2] Several of the equations in this section include a bracketed term such as "{D3.1}" at the end of the equation. That bracketed term is not a mathematical operation; rather, it indicates the version of DOCSIS to which the equation applies.

[3] P_{r_n} is a well-defined term in the D3.0 spec, referring to the *reported* power for the nth channel. It is not a term explicitly used in the D3.1 spec. We use PWR_n in this paper as a generic term to denote the power in the channel for D3.0 or D3.1.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

$$Commanded\ TCP\ (dBmV) = 10log_{10}\left(sum\ over\ all\ n\ of\ 10^{(PWR_n/10)}\right).\ \{D3.0\ and\ D3.1\}$$

For the DOCSIS 3.1 CM, the CM TCP headroom is given by:

$$CM\ TCP\ headroom\ (dB) = P_{max}(dBmV) - Commanded\ TCP\ (dBmV).\ \{D3.1\}$$

For DOCSIS 3.0 modems, the calculation of the Commanded TCP is straightforward when the constellations are 32-QAM or larger, so we will consider this case for an introductory explanation. The legacy TDMA channels are all commanded in terms of the channel power, where PWR_n is the commanded ("reported") power of the nth channel, in units of dBmV.

The determination of the CM's TCP headroom capability in [D3.0PHY] can be complicated, unless we invoke the assumption of 32-QAM TDMA or larger constellation in all channels, which we will retain for now. The DOCSIS 3.0 CM TCP capability is 57 dBmV with these conditions, for any number of channels (one to four).[4] Thus, for DOCSIS 3.0, with 32-QAM or higher constellations with TDMA channels:

$$CM\ TCP\ headroom\ (dB) = 57\ dBmV - Commanded\ TCP(dBmV).\ \{D3.0, 32\text{-}QAM\ or\ larger\}$$

3.2. Cable Modem Minimum Channel Headroom

In addition to the CM TCP headroom, there is also the CM Min Ch headroom. The CM Min Ch headroom is important because it quantifies how much increase can occur in the channel with the largest PSD [D3.1PHY] or largest power, PWR_n [D3.0PHY] (with the assumption of 32-QAM or larger constellation using DOCSIS 3.0 technology).

In DOCSIS 3.1 the CM Min Ch headroom is the difference between the maximum PSD for the CM, which is $P_{1.6hi}$, and the largest PSD commanded in the TCS (largest value of $P_{1.6r_n}$, over all N channels):

$$CM\ Min\ Ch\ headroom\ (dB) = P_{1.6hi}(dBmV) - max\left(P_{1.6r_n}, over\ all\ N\ channels\right).\ \{D3.1\}$$

In DOCSIS 3.1 the CM Min Ch headroom is described by the term $P_{1.6load_1}$, the loading of the highest loaded channel (see [D3.1PHY]):

$$CM\ Min\ Ch\ headroom\ (dB) = P_{1.6load_1}.\ \{D3.1\}$$

In DOCSIS 3.0 the CM Min Ch headroom is described by the term P_{load_1}, the loading of the highest loaded channel (see [D3.0PHY]):

$$CM\ Min\ Ch\ headroom\ (dB) = P_{load_1}.\ \{D3.0\}$$

In DOCSIS 3.0, since P_{load_1} equals P_{hi_1} minus P_{r_1}, and since for our simplifying assumption of TDMA with at least 32-QAM constellations, P_{hi_1} is 57 dBmV with one channel in the TCS, 54 dBmV with two channels in the TCS, and 51 dBmV with three (or four) channels in the TCS, we can also write:

$$CM\ Min\ Ch\ headroom\ (dB) = P_{hi_1} - P_{r_1}.\ \{D3.0, 32\text{-}QAM\ or\ larger\}$$

[4] P_{max} represents the CM's TCP capability in [D3.1PHY] but represents an individual channel power limit in [D3.0PHY]. There is not a term relating to TCP capability in [D3.0PHY], which is why we show how to calculate it in this paper.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Note that in both DOCSIS 3.0 and DOCSIS 3.1 each upstream channel of each CM can be described with its own figure-of-merit: CM Channel headroom. For DOCSIS 3.0 modems, the nth channel's CM channel headroom is P_{load_n}, which also equals P_{hi_n} minus PWR_n. For DOCSIS 3.1 modems, the nth channel's CM channel headroom is $P_{1.6hi}$ minus $P_{1.6r_n}$.

3.3. DOCSIS 3.0 Example

According to [D3.0PHY], the maximum total transmit power that the CM can support, P_{max}, for a DOCSIS 3.0 modem transmitting four 64-QAM channels is 51 dBmV per channel (57 dBmV total power).[5] If the modem is transmitting 44 dBmV for each channel (50 dBmV total power), and P_{hi_n} for each channel is 51 dBmV, then the CM Min Ch headroom is 51 dBmV – 44 dBmV = 7 dB. In this example, the CM TCP headroom is the same as the CM Min Ch headroom, since from a total power perspective, 57 dBmV – 50 dBmV = 7 dB. See Figure 3.

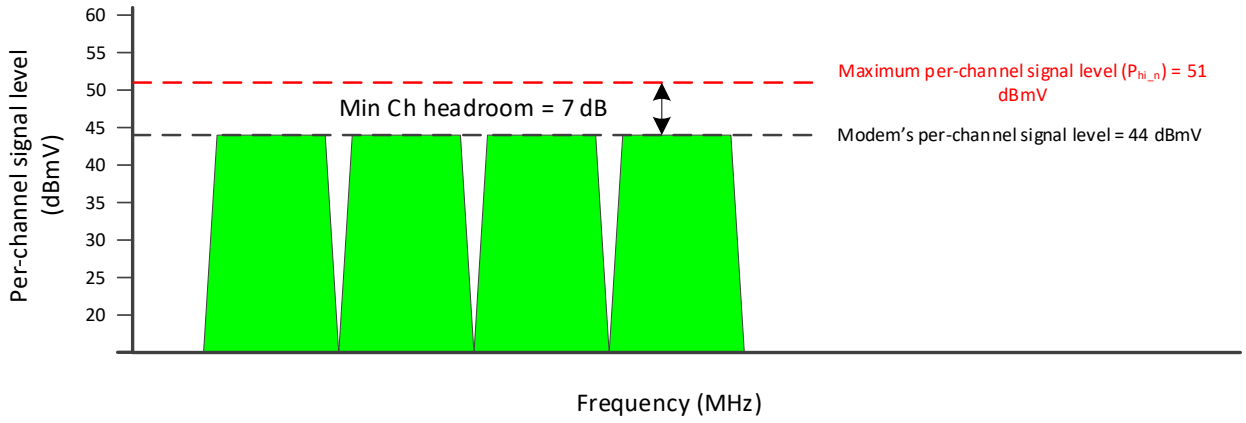


Figure 3 - Cable modem Min Ch headroom is 7 dB in this example.

If, however, that same modem is transmitting 51 dBmV per channel, so that PWR_n equals P_{hi_n} for each channel, the CM Min Ch headroom is 51 dBmV – 51 dBmV = 0 dB, and the CM TCP headroom is the same, 57 dBmV – 57 dBmV = 0 dB, as shown in Figure 4.

[5] Note that [D3.0PHY] supports higher per-channel transmit power for lower modulation orders such as QPSK, than what is shown in this example. For more information the reader is encouraged to review Section 6.2.19 of [D3.0PHY].

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

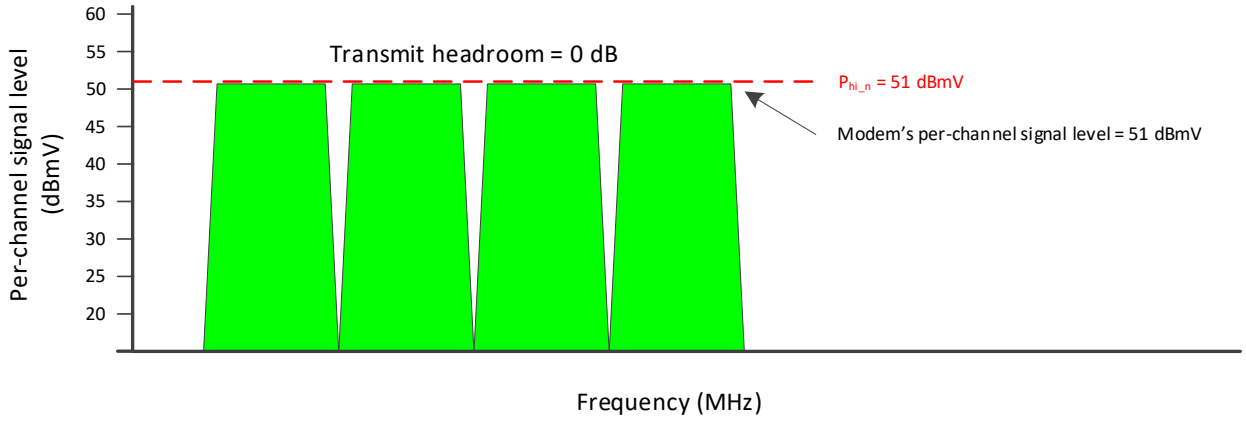


Figure 4 – Here, the modem's transmitted per-channel signal level is equal to the maximum per-channel signal level P_{hi_n}, resulting in 0 dB Min Ch transmit headroom.

For another example (refer to Figure 5), consider that the same modem is transmitting 44 dBmV on one channel, 41 dBmV on another channel, and 38 dBmV each on the third and fourth channels. In this case P_{r_1} equals 44 dBmV, and P_{hi_n} for each channel is 51 dBmV, so the CM Min Ch headroom is 51 dBmV – 44 dBmV = 7 dB.

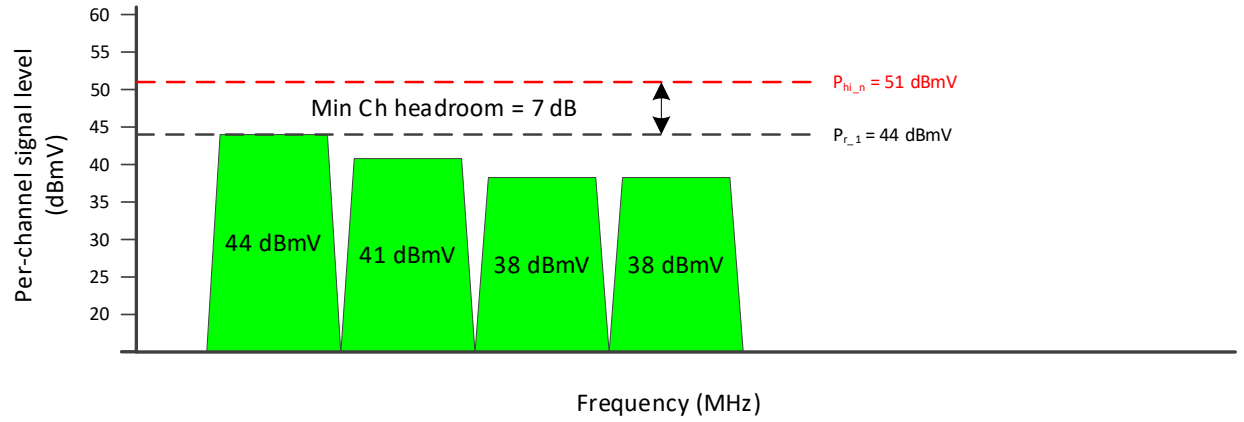


Figure 5 - In this example the Min Ch transmit headroom is 7 dB and the TCP headroom is 10 dB (see text).

The CM TCP for this uneven per-channel transmit power example is 47 dBmV. The CM TCP capability is still 57 dBmV, so the CM TCP headroom is 57 dBmV – 47 dBmV = 10 dB.

In this example the CM TCP headroom is larger than the CM Min Ch headroom, 10 dB compared to 7 dB. While the CM total transmit power can be increased by 10 dB according to the CM TCP headroom, there is the practical constraint to consider, especially if the channel already transmitting at 44 dBmV is a channel where additional power is desired. This channel can only increase its transmit power by 7 dB, the amount of the CM Min Ch headroom; however, if this channel is commanded to increase by 7 dB, the two lower powered channels will be 13 dB lower than this channel, and this will violate the conditions of the 12 dB range of the DRW for [D3.0PHY]. The second channel is now much more than 4 dB lower than the first channel, so if the channel power changes are implemented, the spurious emissions requirements will be significantly relaxed (see [D3.0PHY]). If the lower power channels are increased in power and

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

arrive at the burst receiver above the target power for the channels, then the spurious emissions contributions budgeted are larger due to the increased channel power; and also, because they arrive higher than the target power, the spurious emissions from this CM for these channels are larger than intended in making the requirements. In a practical sense, if the channel with the largest power is a channel needing even more power, then the CM Min Ch headroom may be a more important consideration than the CM TCP headroom. But both must be managed.

3.4. DOCSIS 3.1 Example

As with DOCSIS 3.0 modems, an ideal case for DOCSIS 3.1 CMs is with all channels having the same PSD, $P_{1.6r_n}$ the same for all n. In these cases, the CM TCP headroom and CM Min Ch headroom are the same. Whenever any $P_{1.6r_n}$ is equal to $P_{1.6hi}$, then the CM Min Ch headroom will be 0 dB.

Now consider some further examples of CM transmit headroom for DOCSIS 3.1 modems.

In the first example, assume one SC-QAM channel of 1.6 MHz bandwidth and three 6.4 MHz-wide SC-QAM channels, and two OFDMA channels, with 16 MHz modulated spectrum (occupied bandwidth) and 40 MHz modulated spectrum; see Figure 6. The number of equivalent DOCSIS channels, N_{eq}, is 48. $P_{1.6hi}$ equals 48.2 dBmV. Let the PSD for the six channels be, respectively: 37 dBmV, 37 dBmV, 37 dBmV, 37 dBmV, 40 dBmV, and 43 dBmV.

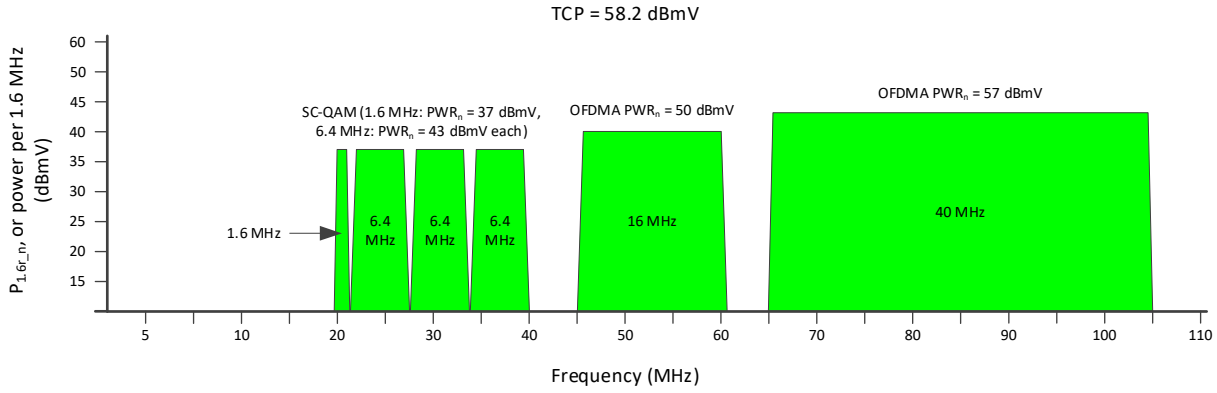


Figure 6 - The TCP headroom and Min Ch headroom in this example are 6.8 dB and 5.2 dB respectively; see text.

The powers for the six channels, PWR_n, are: 37.0 dBmV, 43.0 dBmV, 43.0 dBmV, 43.0 dBmV, 50.0 dBmV, and 57.0 dBmV. The TCP of the TCS sums to 58.2 dBmV.

The CM TCP headroom is P_{max} – 58.2 dBmV = 65 dBmV – 58.2 dBmV = 6.8 dB.

The CM Min Ch headroom is $P_{1.6hi}$, 48.2 dBmV, minus the largest $P_{1.6r_n}$, 43.0 dBmV, which yields 5.2 dB.

The CM's TCP headroom and Min Ch headroom are fairly close, 6.8 dB compared to 5.2 dB.

For another more extreme example, illustrated in Figure 7, consider one SC-QAM channel of 1.6 MHz bandwidth and one SC-QAM channel of 6.4 MHz bandwidth, and two OFDMA channels, with 16 MHz modulated spectrum (occupied bandwidth) and 40 MHz modulated spectrum. The number of equivalent

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

DOCSIS channels, N_{eq}, is 40. $P_{1.6hi}$ equals 49.0 dBmV. Let the PSD for the six channels be, respectively: 46 dBmV, 40 dBmV, 36 dBmV, and 36 dBmV.

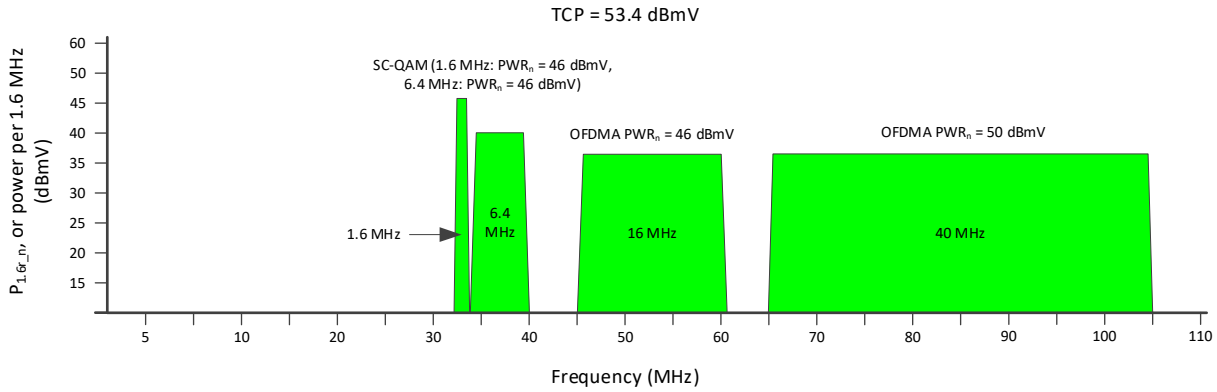


Figure 7 - The TCP headroom and Min Ch headroom in this example are 11.6 dB and 3.0 dB respectively; see text.

The powers for the six channels, PWR_n, are: 46.0 dBmV, 46.0 dBmV, 46.0 dBmV, and 50.0 dBmV. The TCP of the TCS sums to 53.4 dBmV.

The CM TCP headroom is P_{max} – 53.4 dBmV = 11.6 dB.

The CM Min Ch headroom is $P_{1.6hi}$, 49.0 dBmV, minus the largest $P_{1.6r_n}$, 46.0 dBmV, which yields 3.0 dB.

The CM's TCP headroom and Min Ch headroom are very different in this example, 11.6 dB compared to 3.0 dB. This is attributable to the 1.6 MHz-wide SC-QAM channel having a relatively high PSD, much higher than the other three channels, and very close (within 3 dB) to the max PSD allowed with the given TCS occupied bandwidth, but also, the channel is only a small portion of the TCS occupied bandwidth. This leaves little room to increase this channel PSD. However, since the other channels have much larger bandwidth, their total channel power is significantly larger than with the 1.6 MHz SC-QAM channel.

Note that the fidelity requirements (both the spurious emissions and the TxMER, in [D3.1PHY]) are significantly relaxed compared to when all channels have the same $P_{1.6r_n}$. The spurious emissions and TxMER requirements are relaxed by almost 10 dB in this example, due to the channels with the majority of the occupied bandwidth (the OFDMA channels) having their $P_{1.6r_n}$ at 10 dB lower than the channel with the highest PSD.

If there is a need to raise the PSD of the 1.6 MHz-wide SC-QAM channel further, more than 3 dB, there is not enough CM Min Ch headroom to provide that increase. However, there is ample CM TCP headroom available. And already the fidelity requirements are significantly relaxed with the relatively large-bandwidth OFDMA channels operating at much lower PSD than the narrow-bandwidth SC-QAM channel. Also, raising the PSD of the 1.6 MHz-wide SC-QAM channel by 3 dB will violate the 12 dB DRW limit, if the OFDMA channels' PSD is not increased. Due to the large amount of CM TCP headroom available, and the large relaxation in fidelity requirements already occurring in the example, it may be preferred to not increase the 1.6 MHz SC-QAM channel power, and let it arrive below the receiver target PSD, or perhaps even drop the channel entirely from the TCS. The latter will result in tightened fidelity requirements, and minimal reduction in occupied bandwidth in the TCS.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. Factors That Can Affect Transmit Headroom

At first glance it may seem that understanding and managing CM transmit headroom is fairly straightforward. If the net attenuation between a modem and CMTS changes, the CMTS will command the modem to adjust its transmit power as needed. If only it were that easy! Cable modem transmit headroom is affected by a complicated combination of factors including transmitted RF power, DRW, long loop automatic level control (ALC), pre-equalization settings, and channel bandwidth.

4.1. Transmitted RF power

4.1.1. Total Composite Power

As stated in Sections 2.1 and 3.1, [D3.1PHY] specifies P_{max} for the TCS. The TCP headroom is the difference between the TCP and P_{max}. If TCP in the TCS exceeds P_{max} supported by the CM, the modem does not have enough power to maintain such transmission. In this case the CM may be forced to drop one or more of the channels and fall back to partial service, so TCP is within the supported range. The headroom indicates how much reserve is available in the current TCP.

4.1.2. Cable Modem Termination System Set Point

The configured CMTS target set point[6] has a direct impact on the TCP of the CM's TCS. Configuring the CMTS target set point to a lower or higher level effectively decreases or increases the TCP of the CM's TCS, which in turn increases or decreases the effective CM headroom.

4.1.3. Transmit Channel Set Occupied Bandwidth

Assuming a CM is not already transmitting at its maximum capability, increasing the TCS occupied bandwidth also increases the TCS's TCP (see Figure 1 and Figure 2). Consider the following example. The TCS initially comprises four 6.4 MHz-wide SC-QAM channels (64-QAM) transmitting with $P_{1.6}$= 32.75 dBmV, 32.75 dBmV, 34.75 dBmV, and 35.75 dBmV respectively. The power per 6.4 MHz-wide channel (PWR_n) = 38.77 dBmV, 38.77 dBmV, 40.77 dBmV, and 41.77 dBmV. As part of plant expansion, a new 52 MHz-wide OFDMA channel is added to the TCS. This channel transmits with $P_{1.6}$= 33 dBmV. Table 1 shows the before and after headroom performance:

[6] The target set point is the configured upstream receive signal level at the CMTS or CCAP. This parameter was called commanded nominal receive power in [D3.0PHY] and earlier.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 - Before and after headroom performance.

Channel	Channel width	Initial TCS Occupied Bandwidth: 25.6 MHz		New TCS Occupied Bandwidth: 77.6 MHz	
		$P_{1.6}$, dBmV	Headroom, dB	$P_{1.6}$, dBmV	Headroom, dB
1	6.4 MHz	32.75	20.21	32.75	15.35
2	6.4 MHz	32.75	20.21	32.75	15.35
3	6.4 MHz	34.75	18.21	34.75	13.35
4	6.4 MHz	35.75	17.21	35.75	12.35
5	52 MHz			33.00	15.10
Summary					
N_{eq}		16		49	
TCP, dBmV		46.24		50.29	
$P_{1.6hi}$, **dBmV**		52.96		48.10	
TCP Headroom, dB		18.76		14.71	
Min Ch Headroom, dB		17.21		12.35	

Table Summary:

- Bandwidth increased by 3.03 times, or logarithmically, $10\log_{10}(77.6\text{ MHz}/25.6\text{ MHz}) = 4.82$ dB
- TCP headroom degradation: 18.76 dB – 14.71 dB = 4.05 dB
- Min Ch headroom degradation: 17.21 dB – 12.35 dB = 4.86 dB

The bandwidth expansion (logarithmic) is 4.82 dB, and the degradation in Min Ch headroom is 4.86 dB. The two values are very close, and would be the same except for the quantization involved in using the ceiling function when computing N_{eq}.

See Section 4.5.1 for an additional illustrated example of occupied bandwidth expansion.

4.2. Dynamic Range Window

As mentioned in Section 2.2 and Appendix B, one of the factors that limits CM transmission power is DRW. DRW does not affect headroom directly as it limits the difference between the least and most loaded channels, and the headroom for the most loaded channel stays the same. If the DRW is violated, the fidelity requirements for the TCS may be compromised and eventually most loaded channel(s) will be dropped from the TCS, forcing the CM to partial service. The resulting partial TCS may have different (better) headroom for the remaining channels. But because the resulting TCS is impacted, directly comparing their headroom is incorrect.

4.3. Long Loop Automatic Level Control

The content in this section is excerpted and adapted from [LLALC], and is used with permission of the author.

A surprising amount of behind-the-scenes interaction occurs between DOCSIS cable modems and the CMTS (or CCAP or RPD). One important process is called station maintenance, during which the received upstream channels are evaluated by the CMTS's burst receiver. The burst receiver looks at the received signal level, the channel's center frequency, its timing, and other parameters. If any of those are out of tolerance, the CMTS will command the affected modem to adjust its transmitted signal as needed.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In particular, the received upstream signal level is supposed to match the CMTS's target set point, which has a default value of 0 dBmV in most CMTSs (that value is user configurable). If the received upstream signal level doesn't match the target set point, the CMTS instructs the modem via a ranging response (RNG-RSP) message to adjust its upstream transmit power as needed so the received signal level at the CMTS is correct. This upstream signal level management is called long loop ALC.

Anything that causes the net upstream attenuation between modems and the CMTS to change will result in the CMTS telling the affected modems to change their upstream transmit power accordingly. Examples include coaxial cable attenuation changes over temperature, installation of components in the drop that add loss (e.g., replacing a two-way splitter with a four-way splitter), changing padding in the headend or hub site upstream combiner, and so on.

4.3.1. Additional Loss in Return Path

Consider a scenario in which a subscriber has only high-speed Internet service. Figure 8 shows what the typical upstream signal levels might be. The downstream RF input to the modem is +3 dBmV, the modem's upstream transmit power is +43 dBmV, and that upstream channel's RF level at the input to the tap (after drop cable loss) is +41 dBmV. By the time the channel reaches the node, it's +15 dBmV. Assume that this configuration gives the desired 0 dBmV at the CMTS upstream port.

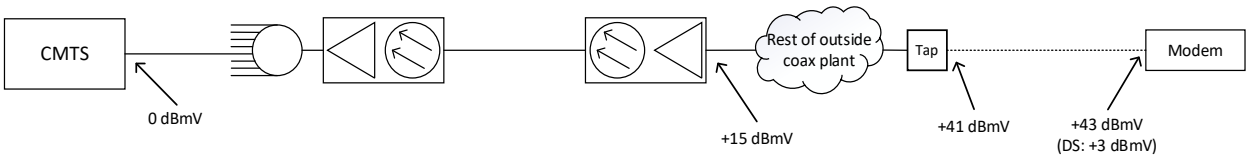


Figure 8 - Baseline upstream signal level scenario for Internet-only subscriber. Here the net upstream attenuation between the modem and CMTS is 43 dB.

Now assume the same subscriber decides to have three set-top boxes connected along with the existing CM. An installer is dispatched to the subscriber premises to install the additional outlets, cabling, and four-way splitter. What happens to the upstream channels with the additional 7 dB of loss from the splitter? As illustrated in Figure 9, for a brief amount of time the modem's upstream signal level is reduced by 7 dB, giving +34 dBmV at the tap, +8 dBmV at the node, and –7 dBmV at the CMTS upstream port. The modem's transmit level is still at its original +45 dBmV, but the downstream input to that modem has dropped by 7 dB from +3 dBmV to –4 dBmV.

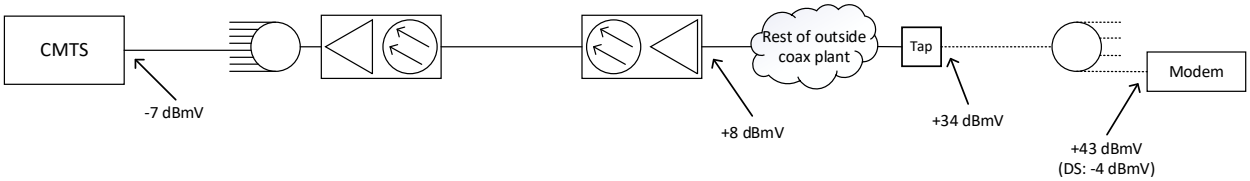


Figure 9 - The newly-installed four-way splitter adds 7 dB of attenuation to the drop, which reduces downstream and upstream levels. Note that the input to the CMTS is now –7 dBmV instead of the desired 0 dBmV. The net upstream attenuation between the modem and CMTS has increased to 50 dB.

The station maintenance process measures the lower RF input to the CMTS, and commands the modem via a RNG-RSP message to increase its transmit power by 7 dB (see Figure 10). Now the modem's upstream transmit level is +50 dBmV. Note that the input to the tap is back to the original +41 dBmV, the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

node input is +15 dBmV, and the CMTS input is 0 dBmV. The downstream level at the modem input is still 7 dB lower than its original value, though.

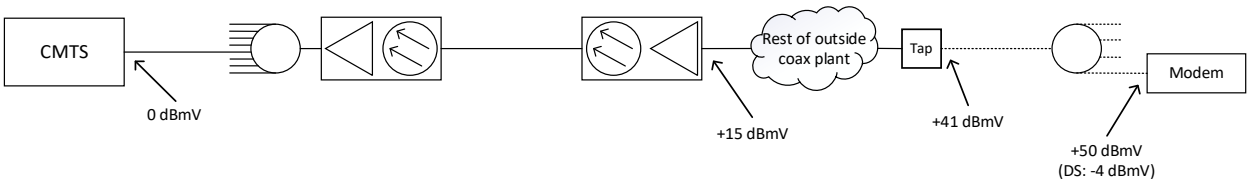


Figure 10 - The station maintenance process commands the modem to increase its upstream transmit level by 7 dB to overcome the 7 dB attenuation of the newly-installed splitter. The net upstream attenuation between the modem and CMTS remains at 50 dB.

4.3.2. Coaxial Cable Attenuation Versus Temperature

Coaxial cable attenuation in decibels changes about 1% for every 10 °F ambient temperature change. As the temperature increases, cable attenuation increases; as the temperature decreases, cable attenuation decreases. Both downstream and upstream RF signal levels are affected by temperature-related cable attenuation variations.

Long loop ALC can help to compensate for temperature-related coax attenuation changes in the upstream. For instance, assume 100 feet of overhead Series 6 drop cable at the subscriber premises, 1500 feet of overhead 0.500 inch diameter hardline feeder cable between the tap and the node, and a worst-case seasonal temperature variation from –10 °F to +105 °F. Table 2 and Table 3 summarize the upstream attenuation versus temperature at 5 MHz, 55 MHz, and 85 MHz.[7]

[7] The values in the tables were calculated using published 0.500 coax attenuation specifications of 0.16 dB/100 feet at 5 MHz, 0.55 dB/100 feet at 55 MHz, and 0.69 dB/100 feet at 85 MHz; and published Series 6 coax attenuation specifications of 0.58 dB/100 feet at 5 MHz, 1.60 dB/100 feet at 55 MHz, and 1.97 dB/100 feet at 85 MHz. All published attenuation values are at a reference temperature of 68 °F. See section 20.3.4. in [SCTEMath] for the equation used to calculate attenuation at other temperatures.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 2 - Attenuation versus temperature for 1500 feet of 0.500 inch hardline cable.

	–10 °F	68 °F	+105 °F	Maximum variation over temperature
5 MHz	2.19 dB	2.4 dB	2.5 dB	0.31 dB
55 MHz	7.54 dB	8.25 dB	8.59 dB	1.05 dB
85 MHz	9.46 dB	10.35 dB	10.77 dB	1.31 dB

Table 3 - Attenuation versus temperature for 100 feet of Series 6 drop cable.

	–10 °F	68 °F	+105 °F	Maximum variation over temperature
5 MHz	0.53 dB	0.58 dB	0.60 dB	0.07 dB
55 MHz	1.46 dB	1.60 dB	1.67 dB	0.21 dB
85 MHz	1.80 dB	1.97 dB	2.05 dB	0.25 dB

To account for the combined temperature related attenuation changes in the 100 feet of Series 6 drop cable and 1500 feet of 0.500 inch diameter feeder cable, the modem's transmit power would vary up to 1.56 dB (at 85 MHz) from –10 °F to +105 °F.

4.4. Upstream Pre-Equalization

Upstream pre-equalization is used to compensate for non-flat frequency response in the channel caused by an impairment between the modem and the burst receiver. For instance, if a frequency response impairment in the network causes a suckout in the upstream channel, the pre-equalization process "pre-distorts" the modem's transmitted signal to have essentially the opposite frequency response of the channel impairment. Ideally, the pre-equalized channel will be received with a flat frequency response at the CMTS's burst receiver.

It is important to note that pre-equalization is normalized so that it will not change the CM transmit power. The CM transmit power remains as commanded. After application of a non-flat pre-equalization, generally the receive power at the burst receiver is reduced because more power was shifted into the parts of the channel with the greatest insertion loss. The overall effect is that the CM needs more transmit power to satisfy the target PSD at the receiver. The target PSD for the receiver is termed the "set point" in [D3.1PHY].

Figure 11 provides an example of an OFDMA channel frequency response which is flat and provides 45 dB of insertion loss between the CM transmitter and the upstream burst receiver. The example OFDMA channel has 240 subcarriers covering 12 MHz, and in the healthy channel each subcarrier is transmitted with approximately 30 dBmV and received at approximately –15 dBmV. These values correspond to 45 dBmV per 1.6 MHz and 0 dBmV per 1.6 MHz respectively. Figure 11 then also shows an example impairment which begins impacting the channel, with 6 dB peak-to-peak ripple, 4.45 dB additional insertion loss, and a 20 dB suckout covering 2 MHz of the channel.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

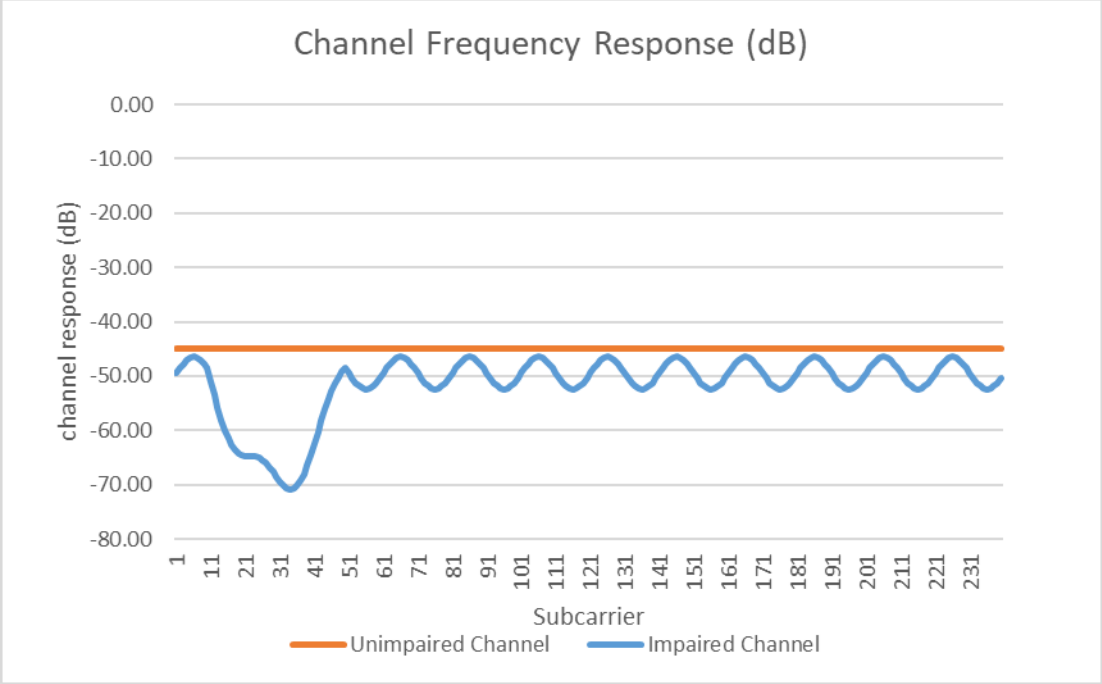


Figure 11 - Channel frequency response before and after onset of an impairment.

Figure 12 shows the power per subcarrier received at the burst receiver before the impairment occurs, which is about –15 dBmV per subcarrier. Figure 12 also shows the received power at the burst receiver after the suckout has occurred, which varies from about –16 dBmV per subcarrier at a maximum to below –40 dBmV per subcarrier in the deepest part of the suckout. Figure 12 also shows the average of the received power per subcarrier, in the dotted line; this average power per subcarrier after the impairment is 4.6 dB lower than the received power per subcarrier before the impairment. Figure 12 also shows the resulting received power per subcarrier after the first ranging which occurred after the impairment impacted the channel. After the first ranging the received subcarriers are now the same power (flat) across the channel at the receiver again, by virtue of the pre-equalization. The first ranging also called for an increase of the CM transmit power by 4.6 dB to compensate for the insertion loss, based on the received average power per subcarrier being 4.6 dB low after the impairment. The average of the power per subcarrier after the impairment onset includes the effect of the power diminished in the relatively small percentage of the channel impacted by the large suckout. While the suckout is large, it covers less than 20% of the channel. It can also be seen in Figure 12 that after the first ranging following the impairment onset the subcarriers are all received at the same power as a result of the pre-equalization; however, it can also be seen that the "per subcarrier" received power is now reduced from –15 dBmV per subcarrier with the healthy channel, to about –24 dBmV per subcarrier. More precisely, the subcarriers are 8.8 dB lower after the first ranging following the impairment onset.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

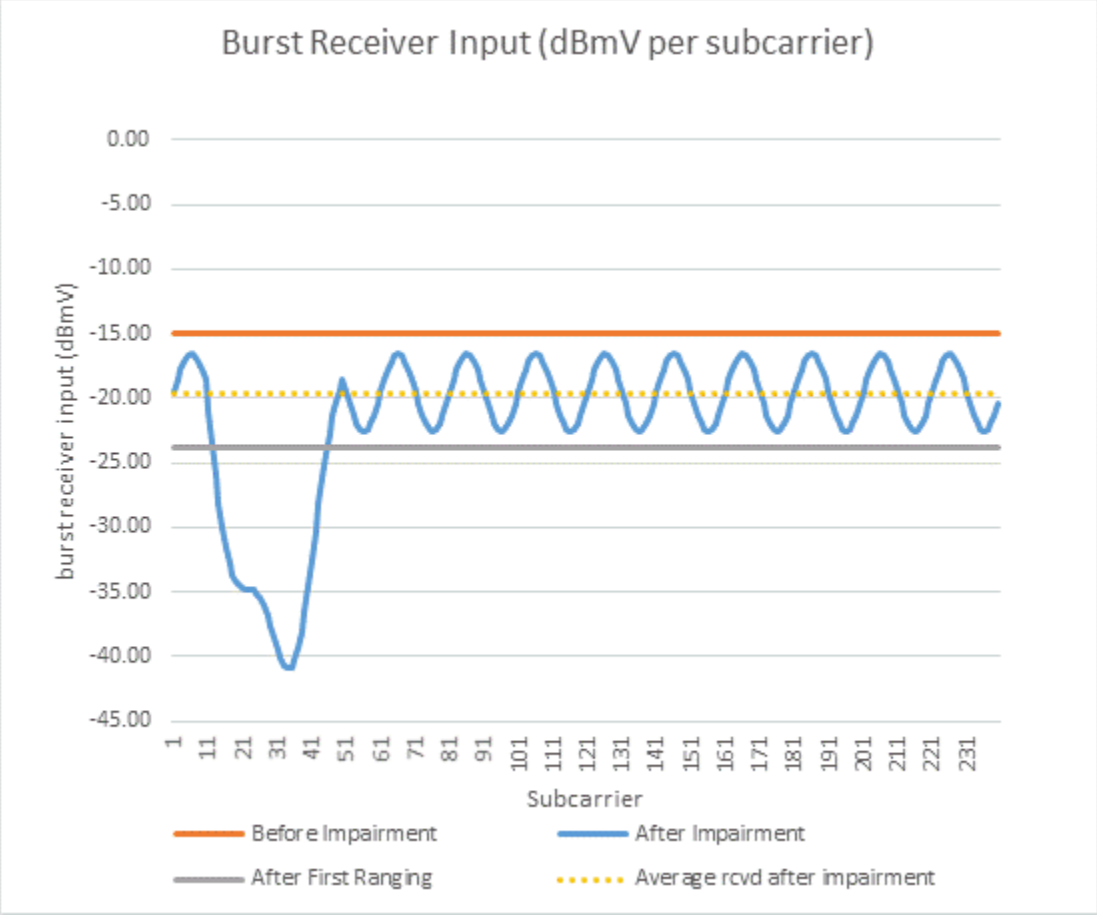


Figure 12 - Burst receiver input power per subcarrier, a) before onset of an impairment; b) after the impairment impacts the channel but prior to a ranging, where about 5 dB of loss occurs across most of the channel and up to 25 dB loss in the suckout; c) the average power per subcarrier after the impairment occurs; and d) after the first ranging response, which increases the transmit power by 4.6 dB and equalizes the channel with pre-equalization.

Figure 13 illustrates why the burst receiver input is reduced by almost 9 dB after the first ranging following the onset of the impairment. Figure 13 shows the normalized pre-equalization vector (normalized to unity power). It can be seen that the normalized pre-equalization vector varies from about –12 dB to +12.5 dB, with most of the channel having the value hovering around –10 dB and only a few MHz in the suckout-impacted region hovering around 10 dB. Since the averaging is performed with linear power, and not the decibel values, the small amount of spectrum with 10 dB pre-equalization is drawing so much power that the transmit power outside the suckout is suppressed by almost 10 dB. Figure 13 shows the power transmitted per subcarrier prior to the onset of the impairment, and also the power per subcarrier transmitted after the first ranging, which included the 4.6 dB commanded increase in transmit power (mentioned in discussion of Figure 12) and the pre-equalization shaping. The average power per subcarrier after the first ranging is indicated in Figure 13 with the dotted line, showing that it is 4.6 dB higher than the (flat) power per subcarrier transmitted prior to the impairment. The pre-equalization at the CM is pushing down the transmit power per subcarrier in most of the channel, causing the power at the burst receiver to be reduced from prior to the impairment onset, even with the commanded power increase to the CM of 4.6 dB. The average PSD of the CM transmission after the first ranging is increased from

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the initial (flat) transmitted PSD solely because of the 4.6 dB command to increase the CM transmit power, where the pre-equalization is not impacting the CM transmit power. Another 8.8 dB of increased transmit power is needed in a second ranging response, as illustrated in Figure 12, to bring the power per subcarrier back up to the desired amount, –15 dBmV, which is 0 dBmV per 1.6 MHz.

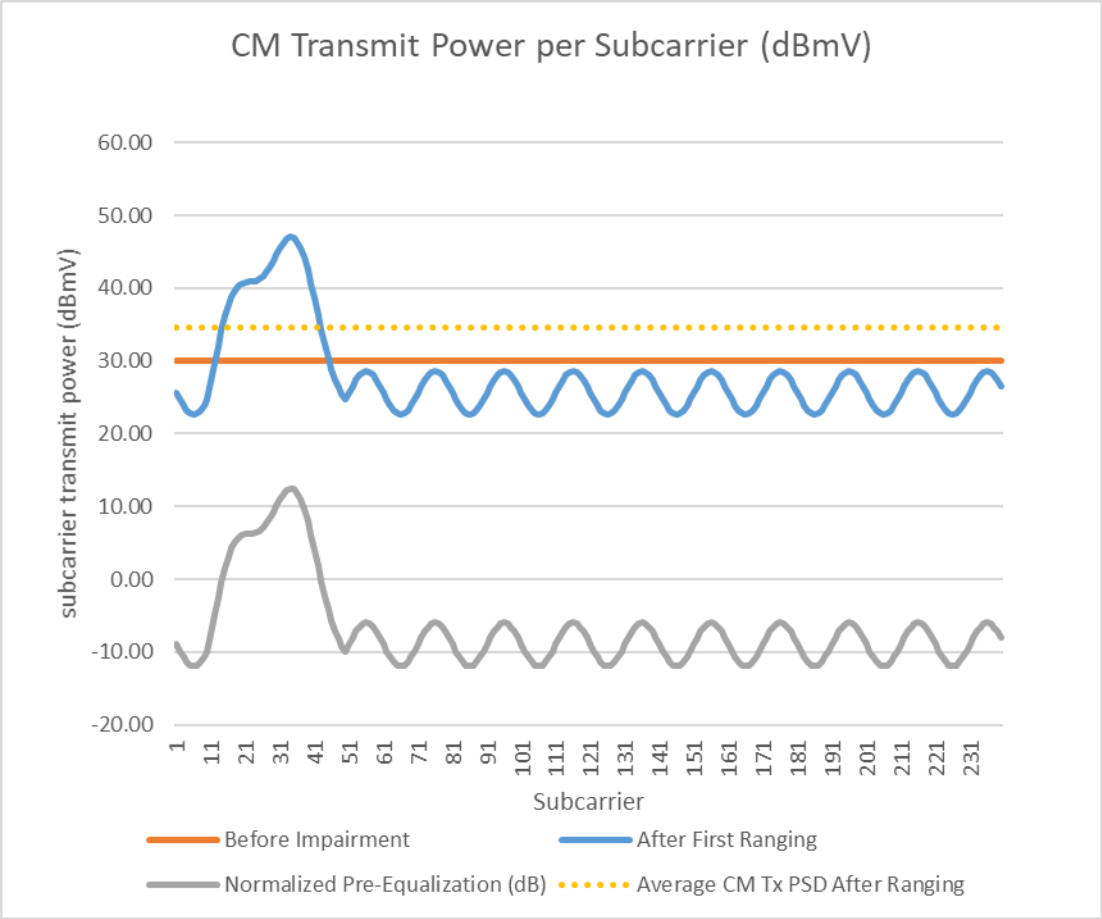


Figure 13 - The cable modem transmit power per subcarrier a) prior to the onset of the impairment; and b) after the first ranging response following the onset of the impairment, with the average power per subcarrier indicated by the dotted line. Also shown is the normalized pre-equalization of the first ranging response, seen to vary from about –12 dB to about 12.5 dB, with most of the channel pre-equalization hovering around –10 dB.

If the CM is already transmitting near or at its maximum transmit PSD prior to when the impairment occurs, the desired increase of 8.8 dB for the CM transmissions noted in the example above (in the second ranging response), may not be possible for the CM. The initial commanded increase of 4.6 dB may not even be achievable, partly or completely. The DOCSIS 3.0 and 3.1 specifications anticipate such occurrences and provide a) requirements[8] for the burst receiver to operate with a minimum dynamic range about its target PSD, thus providing operation when channels reach the burst receiver at a lower PSD than

[8] From [D3.1PHY], Section 7.4.14.1, "REQ25088 The CMTS Upstream demodulator MUST operate within its defined performance specifications with received bursts within the ranges defined in Table 17 - Upstream Channel Demodulator Input Power Characteristics of the set power." Table 17 shows a range about the Set Point of –9 dB up to +3 dB for constellations 256-QAM and lower. For the higher constellation densities of 512-QAM to 4096-QAM the required supported dynamic range is –3 dB to +3 dB.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the target; and b) requirements allowing different upstream bit loading profiles, allowing the CMTS to make grants to the CM with lower bit loading to accommodate a lower SNR. (With lower input channel PSD to the burst receiver, the SNR at the input of the burst receiver due to the upstream noise present at the receiver input will be lower, too, which might require a lower bit loading in data grants than if the transmissions were able to reach the receiver at the target PSD.)

A mathematical treatment of how the normalization process of the pre-equalization maintains the CM's transmit power for the channel is included in the appendix.

4.5. Transmit Channel Set Occupied Bandwidth (Power Spectral Density Versus RF Bandwidth)

Cable modem transmit power is a limited resource in the connection between each modem and the CMTS. A CM is limited in the amount of power it can transmit across all channels, and the power is fairly allocated across the channels; essentially, the maximum power that a modem can transmit is divided among the various channels, which includes SC-QAM (both TDMA and S-CDMA) and OFDMA channels. This allocation of transmit power determines the requirements for the dynamic range window.

As cable operators migrate from sub-split band plans to mid-split or high-split band plans, additional spectrum becomes available to carry more upstream channels (see [SPLIT] for more information about cable network band plans). Those additional upstream channels support greater data capacity. One potential constraint to manage is the impact of the newly added channels to a modem's transmitted total power.

Recall from Section 2.1, in DOCSIS 3.1 the maximum PSD for the CM depends on its configured upstream occupied bandwidth, and is denoted as $P_{1.6hi}$ in DOCSIS 3.1. $P_{1.6hi}$ is calculated as:

$$P_{1.6hi} = P_{max}[dBmV] - 10\log_{10}(N_{eq})$$

In DOCSIS 3.1 the CM Min Ch headroom is the difference between the maximum PSD for the CM, which is $P_{1.6hi}$, and the largest PSD commanded in the TCS (largest value of $P_{1.6r_n}$, over all N channels):

$$CM\ Min\ Ch\ headroom\ (dB) = P_{1.6hi}(dBmV) - max(P_{1.6r_n}, over\ all\ N\ channels).\ \{D3.1\}$$

If the TCS is commanded an increase in its occupied bandwidth by an amount corresponding to N_{eq_added} (additional "chunks" of 1.6 MHz bandwidth), this will reduce $P_{1.6hi}$ by

$$P_{1.6hi}\ decrease(dB) = 10log_{10}[(N_{eq} + N_{eq_added})/N_{eq}]$$

$$= 10log_{10}[1 + (N_{eq_added}/N_{eq})]$$

If the Min Ch headroom is smaller than $P_{1.6hi}\ decrease\ (dB)$, that is,

$$CM\ Min\ Ch\ headroom\ (dB) < 10log_{10}[1 + (N_{eq_added}/N_{eq})]$$

then either a) you cannot add all of the desired new occupied bandwidth, or b) the highest PSD channel(s) will have to have its (or their) transmit power(s) per 1.6 MHz decreased.

Thus, the reader may be interested in knowing how much occupied bandwidth can be added maintaining the existing channel powers; solving some algebra shows that

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

$$N_{eq_additional_possible} = \left(\left[10^{(CM\ Min\ Ch\ headroom\ dB/10)}\right] - 1\right) * N_{eq}$$

Of course, alternatively, the PSD of the highest loaded channel could be lowered instead, to allow the addition of more new occupied bandwidth. The same principles apply with DOCSIS 3.0.

4.5.1. Example: Adding Channels to the Transmit Channel Set

Consider the example shown in Figure 14, which illustrates four upstream 6.4 MHz-wide SC-QAM channels. Assume the power per 6.4 MHz channel is 46.02 dBmV (or 40 dBmV per 1.6 MHz). The total power of the four SC-QAM channels is $P_{total(SC\text{-}QAM)} = 46.02\ dBmV + 10log_{10}(4) = 52.04\ dBmV$. In this example N_{eq} = 16 and $P_{1.6hi}$ = 52.96 dBmV, and CM Min Ch headroom is 12.96 dB. The additional bandwidth that can be added to the TCS based on the current CM Min Ch headroom is

$$N_{eq_additinal_possible} = \left(\left[10^{(CM\ Min\ Ch\ headroom\ dB/10)}\right] - 1\right) * N_{eq} = 64$$

That is, 64 * 1.6 MHz = 102.4 MHz.

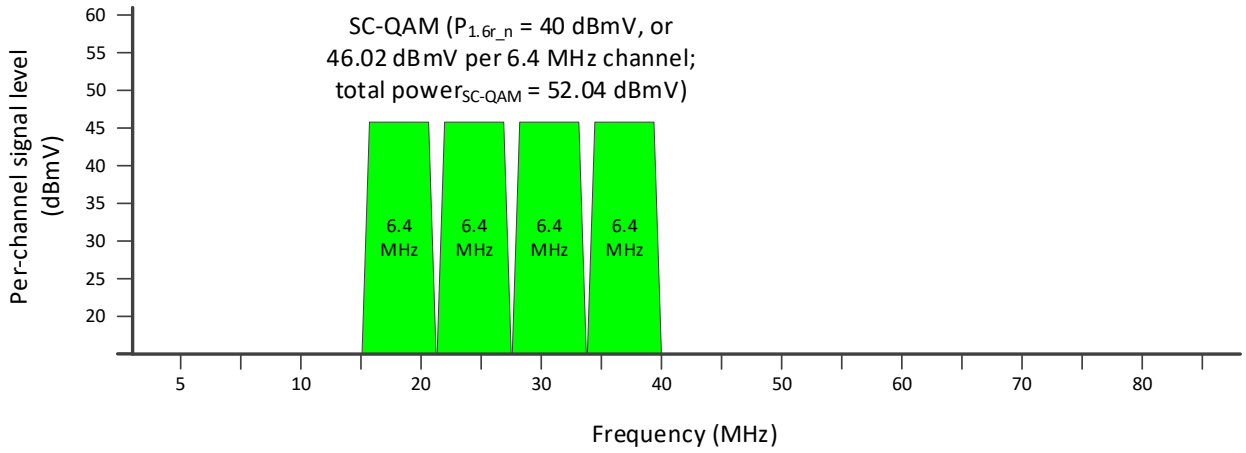


Figure 14 - The total power of four 6.4 MHz-wide SC-QAM channels in this figure, each with a per-channel signal level of 46.02 dBmV, is 52.04 dBmV.

Next, consider a scenario in which a 25.6 MHz-wide OFDMA channel (N_{eq_added} = 16) is added to the upstream spectrum along with the original four 6.4 MHz-wide SC-QAM channels, as shown in Figure 15. Assume for this example the OFDMA's power is 40 dBmV per 1.6 MHz. The total power of the OFDMA channel is $P_{total(OFDMA)} = 40\ dBmV + 10log_{10}(16) = 52.04\ dBmV$. When the total power of the four SC-QAM channels and the total power of the OFDMA channel are added together (logarithmically), the combined total power has doubled and now is

$$P_{total(SC\text{-}QAM+OFDMA)} = 10log_{10}\left[10^{(52.04/10)} + 10^{(52.04/10)}\right] = 55.05\ dBmV$$

The new $P_{1.6hi}$ is 49.95 dBmV, and the new CM Min Ch headroom is 9.95 dB.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

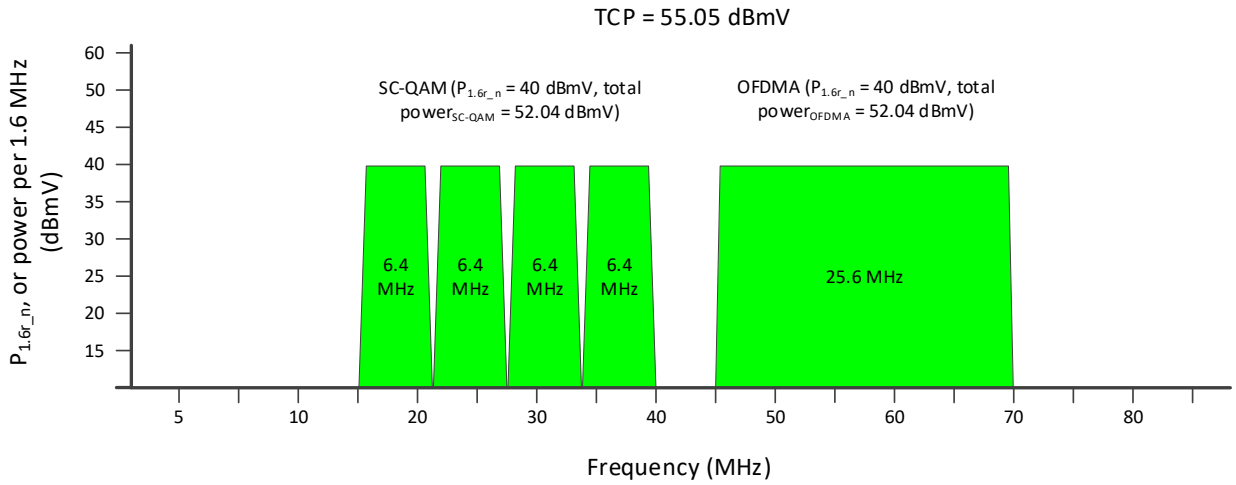


Figure 15 - In this example a 25.6 MHz-wide OFDMA channel has been added to the upstream spectrum.

Table 4 summarizes the before and after headroom performance for the examples in this section.

Table 4 - Before and after headroom performance.

Channel	Channel width	Initial TCS Occupied Bandwidth: 25.6 MHz		New TCS Occupied Bandwidth: 51.2 MHz	
		$P_{1.6}$, dBmV	Headroom, dB	$P_{1.6}$, dBmV	Headroom, dB
1	6.4 MHz	40.00	12.96	40	9.95
2	6.4 MHz	40.00	12.96	40	9.95
3	6.4 MHz	40.00	12.96	40	9.95
4	6.4 MHz	40.00	12.96	40	9.95
5	25.6 MHz			40	9.95
Summary					
N_{eq}		16		32	
TCP, dBmV		52.04		55.05	
$P_{1.6hi}$, dBmV		52.96		49.95	
TCP Headroom, dB		12.96		9.95	
Min Ch Headroom, dB		12.96		9.95	

Table Summary:

- Bandwidth increased by two times, or logarithmically, $10\log_{10}(51.2\text{ MHz}/25.6\text{ MHz}) = 3.01$ dB
- TCP headroom degradation: 12.96 dB – 9.95 dB = 3.01 dB
- Min Ch headroom degradation: 12.96 dB – 9.95 dB = 3.01 dB

The bandwidth expansion (logarithmic) is 3.01 dB, and the degradation in Min Ch headroom is 3.01 dB. The two values are the same.

As the size of a CM's TCS increases to take advantage of the expanded upstream spectrum, the total combined power of those transmitted channels increases. If the modem was already transmitting at or near its maximum power with a smaller channel load, it may not be able to support the transmit power requirements of added channels.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Management of Modem Transmit Power in Cable Network Operation

The previous sections discuss the interaction of spectrum bandwidth and transmit power headroom, and the impact and conditions associated with adding more upstream occupied bandwidth to a CM. The impact of adding a channel could be much greater in the presence of impairments. Knowing the actual transmit power that a modem is using before adding a new upstream channel (or channels) is therefore important to measure, and is useful when estimating the impact to the transmit power of adding a channel to the system. All of this emphasizes the importance of properly managing CM transmit headroom.

As explained in Section 4.5, as operators add channels to address increasing demand, added upstream occupied bandwidth will result in a reduction of $P_{1.6hi}$ (with DOCSIS 3.1 as the example), which may mean some modems cannot participate in the expansion if they do not have enough available transmit power to use the additional channels. These CMs will not be able to provide higher data rates reliably as a result. When this occurs, proactive network maintenance (PNM) may be required to allow these modems to use the new channels.

When transitioning a network to a higher band split, legacy sub-split components – especially those with diplex filters, equalizers, etc. – may affect transmission in mid-split or high-split frequencies. More power may be required from impacted modems as a result, which further limits the power available to add occupied bandwidth to the TCS.

Adding channels is not the only reason to manage the CM transmit power limits. A modem will be commanded to adjust its transmit power levels periodically to address changes in the network's transmission capabilities due to temperature changes (see Section 4.3.2), and the appearance of impairments which often change over time as well. Plant or drop problems will affect signal transmission, and require more power from the modem to try to compensate. A CM that is already close to its transmission limits is likely to exhaust its transmit power budget on occasion due to natural changes in the network. When the CM has insufficient power headroom to transmit at the necessary signal level (to reach the burst receiver at the target PSD), it can potentially benefit from a lower modulation order for some or all channels, otherwise service may be impacted. Calculating CM transmit power compared to the upper limit of its capability (the CM transmit power headroom, P_{load_1} and $P_{1.6load_1}$, in DOCSIS 3.0 and 3.1 respectively), as shown in Section 3.2, is therefore a useful tool for PNM.

The CM's transmit power headroom is one of the figures-of-merit describing a CM's operating margin.

5.1. Ways to Manage Cable Modem Transmit Headroom

When managing CM transmit power headroom, the cause of any issues impacting that headroom must first be considered. If the plant was properly designed and built, degradation may be the main factor to address. But in addition, changes to the network to accommodate evolution, such as converting to a gateway architecture at the subscriber premises, or migrating to higher band splits, can lead to a mix of plant conditions that can affect modem transmit headroom.

If impairments are present and they are causing excessive net attenuation between the modem and CMTS, then long loop ALC (see Section 4.3) will compensate with higher modem transmit power and thus may exhaust the transmit headroom for some CMs in the system. PNM should be conducted first in this case. Poor drop conditions, including water in the cable, or other damage and degradation, can lead to higher CM transmit levels to compensate. Likewise, unused splitters and other passives at the subscriber

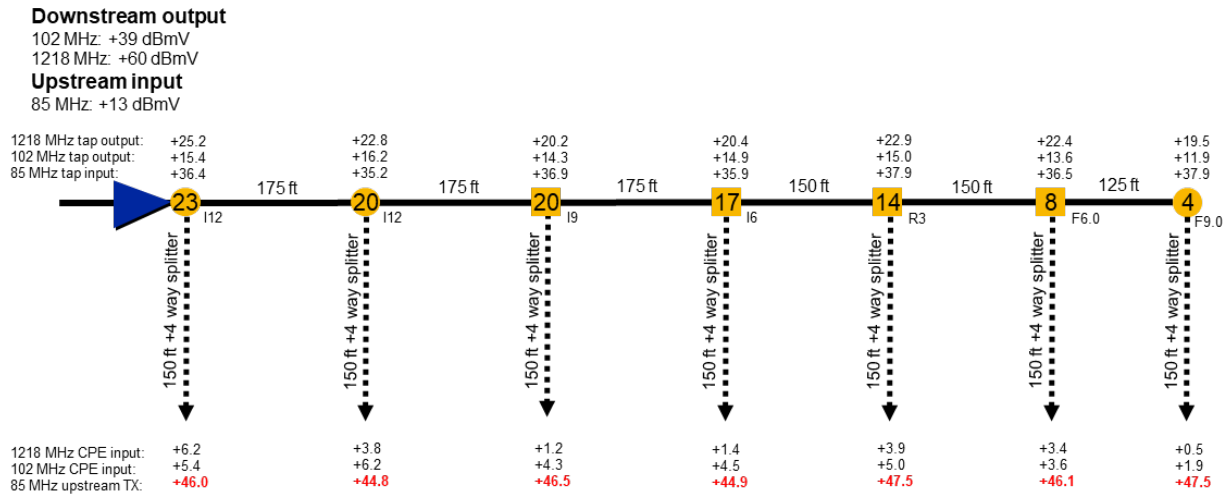
© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

premises can contribute to the problem, and should be removed if not needed. Improperly installed drop components such as backwards splitters should be fixed. In the distribution network, correct alignment of amplifiers and nodes is critical to ensuring proper modem transmit headroom.

In an otherwise unimpaired plant, adding capacity or changing band splits can lead to the need to address CM transmit headroom. Legacy sub-split components in a mid- or high-split network should be removed or replaced. In older network designs it was common to use high-value taps (29 dB and even 32 dB!) at the outputs of actives, which cause modems connected to those taps to transmit at higher levels. Modern HFC networks often use 26 dB or lower value taps at active device outputs, but some of the older higher value taps may have been missed during upgrades and rebuilds. Those high value taps should be replaced with correct values.

Any distribution components that have integrated sub-split filtering need to be replaced or upgraded with the correct filtering when moving to a mid- or high-split architecture. This includes in-line and drop equalizers (and so-called step or reverse attenuators), and drop amplifiers. By removing or replacing equipment intended for a previous architecture and ensuring correct tap values, the pressure on CM transmission levels can be eased, and thus managed to relieve headroom.

Some newer network architectures take advantage of conditioned taps,[9] which can help to narrow the window of modem transmit levels, further helping with modem transmit headroom. See Figure 16 for an example of a mid-split feeder design using conditioned taps.



Note: Design uses taps with plug-in reverse attenuator (e.g., "R3"), forward equalizer (e.g.," F6.0"), or forward inverse equalizer (e.g., "I12") as needed.

Figure 16 - Example mid-split distribution design using conditioned taps

At the headend/hub CMTS or R-PHY node, operators can optimize the configured target set point. As well, headend or hub upstream combining and splitting (and use of external padding at the CMTS upstream ports) should be evaluated and modified if necessary to ensure the overall net attenuation between the modem and CMTS isn't excessive.

[9] Conditioned taps use plug-in pads or equalizers to fine tune downstream and upstream losses in the tap, allowing the modem transmit window to be narrowed and upstream levels optimized.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

A balanced system provides the most dynamic range for the CMs to transmit at lower levels overall so that all CMs have some transmit power headroom for resiliency. If a system is designed so that one or a few CMs are already transmitting at high levels without impairments, then there is less headroom for these CMs to use for resiliency.

5.2. Headroom Management Knobs

There are three "knobs" to manage transmit headroom:

- Reduce CMTS upstream receive set point
- Reduce the TCS occupied bandwidth
- Fix net attenuation problems (see Section 5.1)

The next section includes guidance to create a spreadsheet that can be used as a tool for better understanding the impact of the three transmit headroom management knobs.

6. Create Your Own Spreadsheet

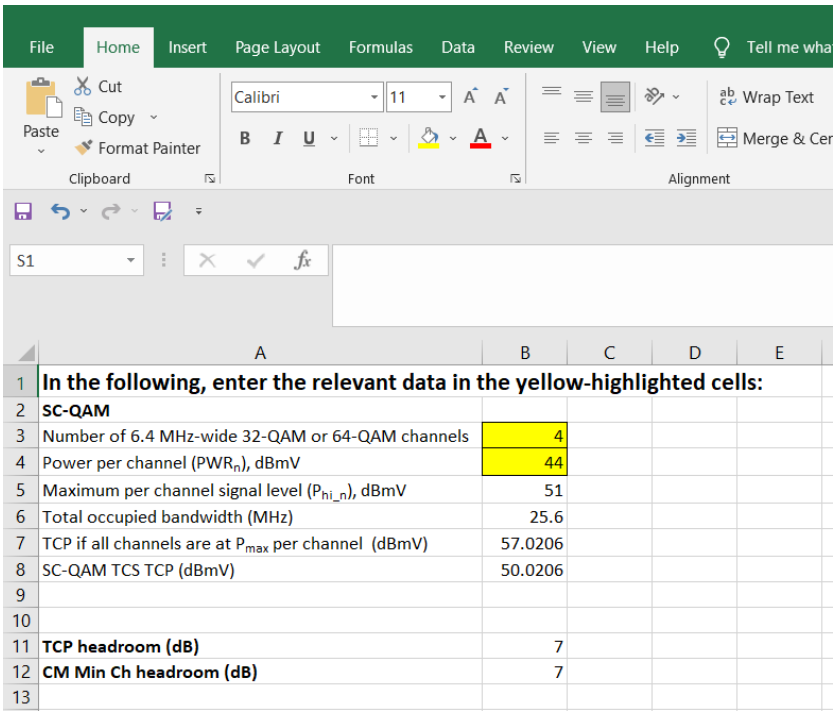
This section includes guidance to create a spreadsheet for calculating and modeling available CM transmit headroom.

6.1. Simple Headroom Calculation for DOCSIS 3.0

Figure 17 shows an example of a simple spreadsheet to calculate a DOCSIS 3.0 CM's TCP headroom and Min Ch headroom. The user enters data in the yellow-highlighted cells. That data includes the number of SC-QAM channels (four shown), and power per channel in dBmV (44 dBmV shown). The assumptions in this simple example are that the SC-QAM channels are either 32-QAM or 64-QAM, are all 6.4 MHz wide, and have the same per-channel transmit power.

The spreadsheet calculates total occupied bandwidth (25.6 MHz shown); the maximum per channel signal level P_{max} per the DOCSIS 3.0 specification (51 dBmV shown); the TCP if all channels are at P_{max} per channel (57.02 dBmV shown); and the TCS's TCP (50.02 dBmV shown). The example spreadsheet's final output in cell B11 is the CM's TCP headroom (7 dB shown), and in cell B12 the Min Ch headroom (7 dB shown). Note that TCP headroom and Min Ch headroom are equal because all four channels have the same PWR_n.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



	A	B	C	D	E
1	**In the following, enter the relevant data in the yellow-highlighted cells:**				
2	**SC-QAM**				
3	Number of 6.4 MHz-wide 32-QAM or 64-QAM channels	4			
4	Power per channel (PWR_n), dBmV	44			
5	Maximum per channel signal level (P_{hi_n}), dBmV	51			
6	Total occupied bandwidth (MHz)	25.6			
7	TCP if all channels are at P_{max} per channel (dBmV)	57.0206			
8	SC-QAM TCS TCP (dBmV)	50.0206			
9					
10					
11	**TCP headroom (dB)**	7			
12	**CM Min Ch headroom (dB)**	7			
13					

Figure 17 - Simple spreadsheet to calculate DOCSIS 3.0 cable modem TCP headroom and Min Ch headroom.

Table 5 summarizes the formulas for various cells in the spreadsheet in Figure 17.

Table 5 - Formulas for spreadsheet in Figure 17.

Cell	Formula
B5	=57-3*CEILING(LOG(B3,2),1)
B6	=B3*6.4
B7	=B5+(10*(LOG(B3)))
B8	=B4+(10*(LOG(B3)))
B11	=B7-B8
B12	=B5-B4

6.2. Mixed Channels Headroom Calculation for DOCSIS 3.1

The next example, shown in Figure 18, is a spreadsheet that can handle up to six SC-QAM channels and two OFDMA channels. As before the user enters data in the yellow-highlighted cells.

The spreadsheet calculates a variety of parameters. Examples include channel TX headroom (row 14); channel RX difference from the CMTS set point (row 15); adjusted channel TX headroom (row 16); TCS TCP headroom (cell B24); combined TCS RX difference from the CMTS set point (cell B25); adjusted TCS TCP headroom (cell B26); $P_{1.6hi}$ (cell B30); $P_{1.6r_1}$ (cell B31); Min Ch headroom (cell B32); worst RX channel difference from the CMTS set point (cell B33); worst adjusted Min Ch headroom (cell B34); and effective DRW (cell B36). This particular spreadsheet was customized with conditional formatting to color certain cells with respect to pre-defined thresholds (beyond the scope of this paper).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

	A	B	C	D	E	F	G	H	I
1	**Transmit Channel Set (TCS)**	**Legacy Channel 1**	**Legacy Channel 2**	**Legacy Channel 3**	**Legacy Channel 4**	**Legacy Channel 5**	**Legacy Channel 6**	**OFDMA Channel 1**	**OFDMA Channel 2**
2	Channel Center Frequency (MHz)	16.4	22.8	29.2	35.6			64	
3	Channel Width (MHz)	6.4	6.4	6.4	6.4	-	-	42.000	-
4	Bandwidth in which power is measured (MHz)	6.4	6.4	6.4	6.4	-	-	1.6	-
5	Reported Modem Transmit Power (dBmV)	44.0	47.0	53.0	53.0	-	-	47.0	-
6	Actual Receive Power at CMTS (dBmV)	0.0	0.0	-1.0	0.0	0.0	0.0	0.0	0.0
7	CMTS's Set Point Receive Power (dBmV)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8									
9	**Per channel report**	**Legacy Channel 1**	**Legacy Channel 2**	**Legacy Channel 3**	**Legacy Channel 4**	**Legacy Channel 5**	**Legacy Channel 6**	**OFDMA Channel 1**	**OFDMA Channel 2**
10	Linear Transmit Power	25118.86432	50118.72336	199526.2315	199526.2315	0	0	1315616.488	0
11	Channel N_{eq_n}	4	4	4	4	-	-	27	-
12	Channel TX Power, (dBmV)	44.00	47.00	53.00	53.00	-	-	61.19	-
13	Channel Power Spectral Density $P_{1.6r_n}$ (dBmV)	37.98	40.98	46.98	46.98	-	-	47.00	-
14	Channel TX Headroom (dB)	**10.69**	**7.69**	**1.69**	**1.69**	**0.00**	**0.00**	**1.67**	**0.00**
15	Channel RX Difference from CMTS Set Point (dB)	0.0	0	-1	0	0	0	0	0
16	Adjusted Channel TX Headroom (dB)	10.69	7.69	0.69	1.69	0.00	0.00	1.67	0.00
17									
18									
19	**TCS report**	**Whole TCS**	**Legacy (SC-QAM)**	**OFDMA**					
20	N_{eq} (number of 1.6 MHz equiv)	43	16	27					
21	Occupied US Spectrum (MHz)	67.600	25.6	42					
22	TCP (dBmV)	62.53	56.76	61.19					
23	P_{max} (dBmV)	65.00							
24	TCS TCP Headroom (dB)	2.47							
25	Combined TCS RX Difference from CMTS Set Point (dB)	-0.09							
26	Adjusted TCS TCP Headroom (dB)	2.38							
27									
28	**Worst channel report**		**Identified channel**						
29			**Description**	**Center Frequency**					
30	$P_{1.6hi}$ (dBmV)	48.67							
31	$P_{1.6r_1}$ (dBmV)	47.00	OFDMA Channel 1	64					
32	Min Ch Headroom (dB)	1.67	OFDMA Channel 1	64					
33	Worst RX Channel Difference from CMTS Set Point (dB)	-1.00	Legacy Channel 3	29.2					
34	Worst Adjusted Min Ch Headroom (dB)	0.69	Legacy Channel 3	29.2					
35									
36	**Effective DRW**	9.02							
37									

Figure 18 - Another example spreadsheet for calculating modem transmit headroom.

Table 6 summarizes some of the formulas for various cells in the spreadsheet in Figure 18. (The formulas in rows 10-16 of columns C-I are similar to those in cells 10-16 of column B. For example, the formula in cell C10 is =10^(C13/10)*C3/1.6 and so on.

An online tool for calculating CM transmit headroom is available at https://promptlink.com/tools/headroomcalc3.1

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 6 - Formulas for spreadsheet in Figure 18.

Cell	Formula
B10	=10^(B13/10)*B3/1.6
B11	=CEILING(B3/1.6,1)
B12	=IF(B3>0,B5-10*LOG(B4/B3),"-")
B13	=IF(B3>0, B5-(10*LOG(CEILING(B4/1.6,1))),0)
B14	=IF(B3>0,B30-B13,0)
B15	=B6-B7
B16	=B14+B15
B20	=CEILING(B21/1.6,1)
B21	=SUM(B3:I3)
B22	=10*LOG(SUM(B10:I10))
B24	=B23-B22
B25	=SUMPRODUCT(B15:I15,B3:I3)/B21
B26	=B24+B25
B30	=B23-10*LOG(B20)
B31	=MAX(B13:I13)
B32	=B30-B31
B33	=MINIFS(B15:I15,B3:I3,">0")
B34	=MINIFS(B16:I16,B3:I3,">0")
B36	=MAXIFS(B13:I13,B3:I3,">0")-MINIFS(B13:I13,B3:I3,">0")

7. Additional Work Underway

As cable operators deploy DOCSIS 3.1 OFDMA in the upstream RF spectrum, better understanding of the CM's TCS performance is needed to ensure reliable service. While this paper's in-depth analysis of CM transmit headroom is useful to provide additional insight for deploying, activating, and maintaining additional occupied bandwidth, there is more work to be done. For instance, the CableLabs PNM Working Group is looking at additional improvements to the existing PNM measurements:

- **OFDMA in-channel frequency response** – Consistent with the definition of ICFR carried over from DOCSIS 3.1 PNM but adding additional context to how the information is used, having significantly higher frequency resolution.
- **OFDMA channel tilt** – Used to quantify important plant setup, filter performance and response impairments.
- **OFDMA channel tilt residual** – Important to distinguish impairment magnitude when separating the effects of tilt, which are often unrelated.
- **Frequency amplitude ripple periodicity** – Provides improved distance accuracy when measuring the length of echo cavities caused by impedance mismatches.
- **Frequency response signature matching** – Determining frequency response commonalities to provide localization, which is useful in the troubleshooting process.

8. Conclusion

At first glance CM transmit headroom seems relatively simple: generally speaking, it is the difference, in dB, between a modem's maximum transmit power capability and its actual transmit power. But as discussed in this paper, CM transmit headroom is more complicated and really has two different

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

interpretations or definitions: 1) CM TCP headroom, and 2) what we call CM Min Ch headroom. These two headroom values comprise two different figures-of-merit to characterize a CM's commanded channel power compared to its transmit power capability. Additionally, each channel of a CM has its own figure-of-merit in this regard: CM channel headroom.

CM transmit headroom is related to a combination of transmitted RF power, DRW, long loop ALC, pre-equalization settings, CMTS receive power set point, and channel bandwidth. All of these factors are even more important as operators migrate to an expanded upstream spectrum, and increase the occupied bandwidth of transmitted signals to support greater data rates. Insufficient CM transmit headroom is a good indication that service quality could be impacted, making it all the more important that operators properly understand and manage CM transmit power.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Appendix A

9. What is Power Spectral Density?

9.1. Power and Power Spectral Density

The terms "power" and "power spectral density" are used throughout this paper, but what do those terms mean? First, we need to understand what the power of a channel is. RF power refers to the amount of energy carried by the channel (its signal level) – more specifically, power is the rate at which work is done, or energy per unit of time. 1 watt of power is equal to 1 volt causing a current of 1 ampere. Note: While the watt is the International System of Units (SI) unit of power, the cable industry uses dBmV (or decibel microvolt, dBµV, in some parts of the world), which is an expression of power in terms of voltage. If we plot the spectrum of the channel (i.e., plot the PSD), power can be represented as the complete area under the PSD covered by the channel over its occupied bandwidth.

From a high-level perspective, power spectral density "describes how power of a signal … is distributed over frequency." PSD is commonly expressed in units of power over bandwidth, typically (but not always) as power per hertz (Hz).[10] For more information see [TPPSD]

Assuming the power of a channel is distributed evenly across its occupied bandwidth (i.e., its relative amplitude is uniform across the spectrum), then PSD can be calculated as:

$$PSD_{[dBmV/Hz]} = Power_{[dBmV]} - 10log_{10}\left(Bandwidth_{[Hz]}\right)$$

Note: The examples in this appendix focus on SC-QAM channels, but the concepts are the same for OFDM, OFDMA, and other RF signal types. What matters is channel power, occupied bandwidth and PSD, and the relationship between them.

Consider an example in which a single 1.6 MHz-wide SQ-QAM channel is being transmitted with a digital channel power of 45 dBmV. The PSD of that channel can be calculated as:

$$PSD_{[dBmV/Hz]} = 45\ dBmV - 10log_{10}(1.6 * 10^6[Hz]) = -17.04\ dBmV/Hz$$

Another way to look at PSD in this example is that each hertz of the 1.6 MHz-wide SC-QAM channel's occupied bandwidth carries –17.04 dBmV of power.

For the next two examples, assume that we have four SC-QAM channels of different bandwidths: two at 1.6 MHz, and one each at 3.2 MHz and 6.4 MHz.

For the first example (see Figure 19) the four channels are transmitted with the same PSD = –17.04 dBmV/Hz. All the channels will have the same relative PSD, as illustrated in the figure. But the power of each channel varies depending on its occupied bandwidth. The power of each 1.6 MHz-wide channel is 45 dBmV. The 3.2 MHz-wide channel has the same amount of power as *two* 1.6 MHz-wide channels together ($3.2\ MHz = 2 * 1.6\ MHz$), so its power is twice as much ($10 * log_{10}(2) =$ 3.01 dB more) as a single 1.6 MHz-wide channel: $45\ dBmV + 10log_{10}(2) = 48.01\ dBmV$. The 6.4 MHz-wide SC-QAM channel has four times the power (or 6.02 dB more) of one 1.6 MHz-wide channel ($6.4\ MHz = 4 * 1.6\ MHz$), so its power is $45\ dBmV + 10log_{10}(4) = 51.02\ dBmV$.

[10] $P_{1.6hi}$ and similar spec terms in [D3.1PHY] have units of power (dBmV) and are not literally a PSD, but represent power measured in a 1.6 MHz bandwidth and function in the specification much like a PSD "per 1.6 MHz."

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

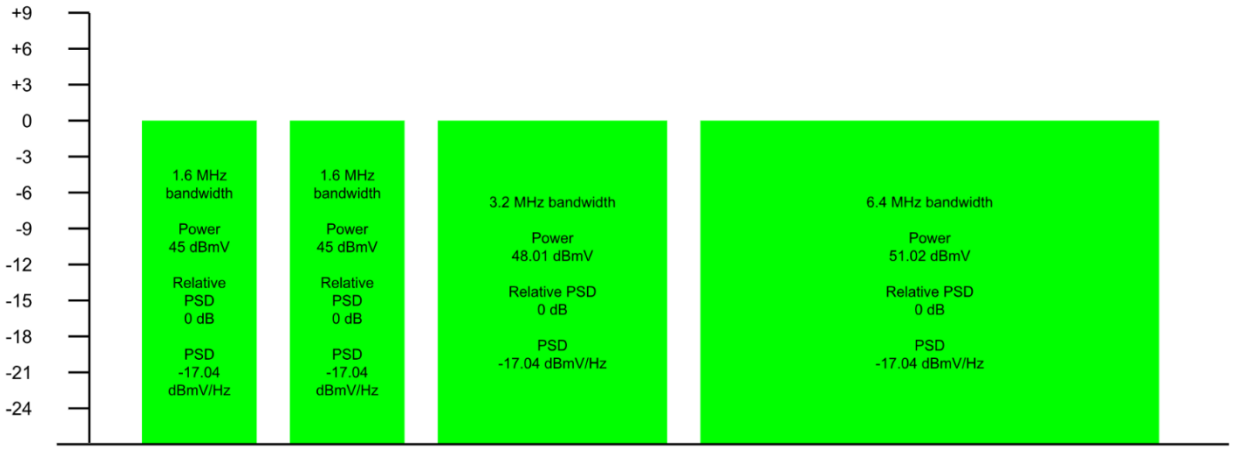


Figure 19 - Example showing four SC-QAM channels with different bandwidths but identical PSD.

For the next example (see Figure 19) the four channels are transmitted with identical power, each at 45 dBmV. There are no changes compared to the previous examples for the 1.6 MHz-wide channels, so they maintain the same power and PSD as before. However, for the twice-as-wide 3.2 MHz-bandwidth channel, in order to maintain the same power the PSD needs to be *decreased* by a factor of two (by $10log_{10}(2) = 3.01$ dB) and will be

$$PSD = -17.04\ dBmV/Hz - 10log_{10}(2) = -20.05\ dBmV/Hz$$

For the four-times-as-wide 6.4 MHz bandwidth channel, the PSD needs to be reduced even further, by $10log_{10}(4) = 6.02$ dB:

$$PSD = -17.04\ dBmV/Hz - 10log_{10}(4) = -23.06\ dBmV/Hz$$

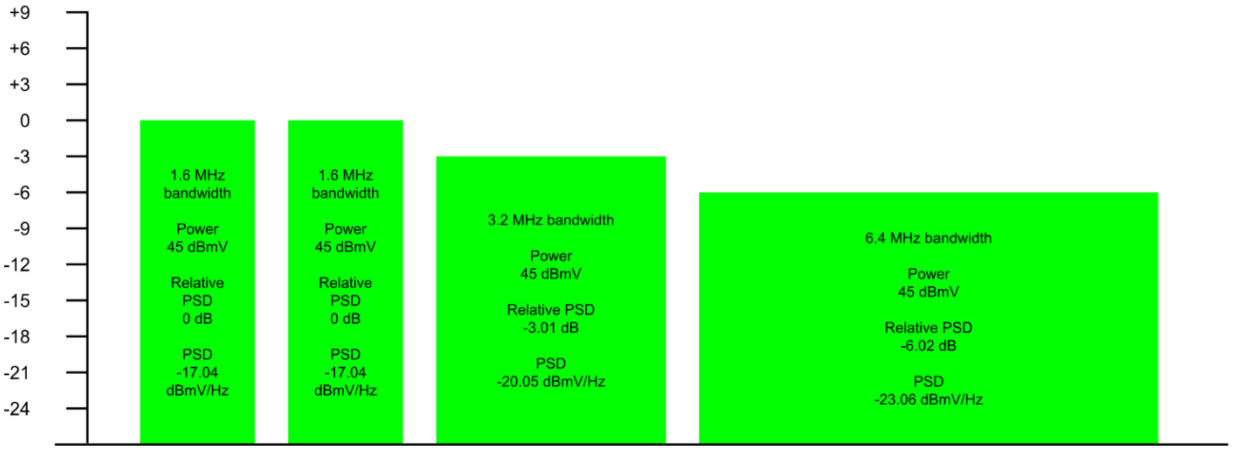


Figure 20 - Example showing four SC-QAM channels with different bandwidths but identical per-channel power.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Appendix B

10. Dynamic Range Window

From DOCSIS 3.0 onwards, cable modems started supporting multiple upstream and downstream channels, which were logically bonded to provide higher data rates than were available with the single-channel operation in DOCSIS 1.0, 1.1, and 2.0 technology. To accommodate multiple upstream channels, new upstream power control parameters were introduced. When a CM is using multiple upstream channels, typically the power spectral density varies from channel-to-channel. This is because generally the upstream receiver operates to control the PSD to be flat at its input, and the frequency response of the plant from most CMs back to the receiver is typically not perfectly flat. Some channels generally transmit with a larger average PSD than others, even from the same modem.

When a cable modem is using multiple upstream channels, typically the PSD varies from channel-to-channel. This is because generally the upstream receiver operates to control the PSD to be flat at its input, and the frequency response of the plant from most CMs back to the receiver is typically not perfectly flat. Some channels generally transmit with a larger average PSD than others, even from the same CM. To accommodate multiple upstream channels a DRW is defined. The DRW in DOCSIS 3.0 defines a 12 dB range of transmit power levels for the cable modem. It is expected that the CMTS will maintain each of the upstream channels in the CM's TCS within the CM's DRW. The cable modem learns the TCS from the CMTS during registration. The CMTS manages the DRW for the modem, ensuring that the cable modem is not ranged at a value that would result in a violation of the DRW. If the CMTS commands the modem to use a transmit power level, that would result in a violation of the DRW; the modem performs the commanded adjustment and indicates an error to the CMTS (the modem must obey the CMTS all the time). The DRW is controlled by the CMTS and communicated to the CM in the ranging or in the registration (TCS) or the dynamic bonding change (DBC) message. Occasionally the plant conditions cause the modem to go out of the DRW, which eventually result in the CM dropping the channel and indicating partial service to the CMTS. Partial service impacts subscribers' upstream speed.

The new parameters introduced from DOCSIS 3.0 are $P_{load_min_set}$, and P_{load_n} (e.g., P_{load_1}, P_{load_2}, etc.). The top of the DRW is defined as $P_{load_min_set}$ and is expressed as some number of dB below P_{hi} for each channel. The P_{hi} for each channel depends on the multiplexing technology – for instance, time division multiple access (TDMA), commonly known as A-TDMA in DOCSIS 2.0 and later – and the modulation order (e.g., 16-QAM, 64-QAM, etc.) for the channel as well as the total number of active channels. The P_{hi} is defined in the PHY specification.

The parameter $P_{load_min_set}$ is commanded to the modem by the CMTS and sets in place the restriction on P_{load_n} that $P_{load_min_set} \leq P_{load_n} \leq P_{load_min_set} + 12$ dB, for all channels in the TCS.

Since $P_{load_n} = P_{hi_n} - P_{r_n}$, the value of $P_{load_min_set}$ places the limit on P_{r_n} for each channel. The P_{r_n} provides the transmission power level, as an offset from the reference channel that the modem is to use on the new channel such that transmissions arrive at the CMTS at the desired power.[11]

$$\left(P_{hi_n} - P_{load_min_set}\right) \geq P_{r_n} \geq \left(P_{hi_n} - P_{load_\min_set} - 12\ dB\right)$$

For each channel in the TCS, the range of P_{r_n} for that channel is restricted to operating over a window of 12 dB by the value commanded by the CMTS. This is the DRW for the modem for the nth channel.

[11] PWR_n is used in other parts of this paper to denote P_{r_n} in [PHY3.0] and $P_{1.6r_n}$ in [PHY3.1].

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

For example, with 64-QAM TDMA the maximum power level allowed is 57 dBmV, 54 dBmV, and 51 dBmV for one, two, and three (or four) channels, respectively, in the TCS. A "fully loaded" 64-QAM TDMA SC-QAM channel with one channel in the TCS would be transmitted at 57 dBmV while a "fully loaded" 64-QAM SC-QAM channel with three channels in the TCS would be transmitted at 51 dBmV. A 64-QAM TDMA SC-QAM which is "underloaded" by 15 dB would transmit at 57 dBmV – 15 dB = 42 dBmV with one channel in the TCS, and a 64-QAM TDMA SC-QAM channel which is "underloaded" by 15 dB would transmit at 51 dBmV – 15 dB = 36 dBmV with three channels in the TCS. The parameter P_{load_n} is used to convey the amount by which the nth channel is underloaded. Thus, in the case of the single channel in the TCS, with 42 dBmV for the 64-QAM TDMA SC-QAM channel's transmit level, the channel is underloaded by 15 dB, therefore, P_{load_1} = 15 dB. Similarly, in the case of three channels in the TCS, with three 64-QAM TDMA SC-QAM channels transmitted at 41 dBmV, 38 dBmV, and 36 dBmV, we have P_{load_1} = 10 dB, P_{load_2} = 13 dB, and P_{load_3} = 15 dB, where P_{load_1} corresponds to the channel with the lowest value of P_{load}, or equivalently, the highest or more full loading. The concept of "loading," and the values of P_{load_n}, are only indirectly tied to the absolute transmit power, with the absolute transmit power equaling the maximum transmit power (for the modulation and number of channels in the TCS) minus P_{load_n} for the nth channel.

Consider a modem that has a TCS with three 6.4 MHz wide 64-QAM TDMA upstream SC-QAM channels. The theoretical capacity is 3 * 5.12 megasymbols per second * 6 bits per symbol = 92.16 Mbps.

If that modem lost one of its SC-QAM channels due to a DRW violation, then it would operate using two TDMA SC-QAM channels, result in lower upstream capacity: 2 * 5.12 megasymbols per second * 6 bits per symbol = 61.44 Mbps.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Appendix C

11. Why Pre-Equalization Does Not Impact Transmit Power, When All is Working as Intended

This appendix describes the "basics" of the operation of the CM transmit power in the DOCSIS long loop automatic level control of a DOCSIS HFC network (see [LLALC] for more information on long loop ALC). The specified details of updating the CM transmit power are presented in the first section (the [D3.1PHY] details for OFDMA are shown for illustration). In the second section it is shown that the normalization of the pre-equalization (Pre-Eq) coefficients provides that the CM transmit power is not altered by the pre-equalization. The third section of this appendix is a reproduction of the definitions of terms for upstream transmit power control in [D3.1PHY].

In an OFDMA channel in [D3.1PHY], some upstream subcarriers may be excluded, but all other subcarriers in the channel are active subcarriers, and are used in probes from a CM in that channel. This appendix defines a set of pre-equalization coefficients *for all active subcarriers* for a given CM as "the pre-equalization vector" for the CM. Similarly, when writing "the Pre-Eq coefficients," it is intended to mean the set of coefficients for the CM, for *all* the active subcarriers of the channel.

11.1. Cable Modem Transmit Power Adjustment in DOCSIS PHYv3.1

The CM determines its target transmit power per channel $P_{1.6t_n}$, as follows, for each channel which is active. Define for each active channel, for example, upstream channel n:

$P_{1.6c_n}$ = commanded power for channel n. (TLV-17 in RNG-RSP)

$P_{1.6r_n}$ = reported power level (dBmV) of the CM for channel n.

$P_{1.6hi} = P_{max}$ dBmV $- 10\log_{10}(N_{eq})$

The CM updates its reported power per channel in each channel by the following steps:

$$\Delta P = P_{1.6c_n} - P_{1.6r_n}$$

$$P_{1.6r_n,new} = P_{1.6r_n,previous} + \Delta P$$

Add power level adjustment (for each channel) to reported power level for each channel. Written another way, to avoid recursion but to clarify the assignment of value (as in programming),

$$P_{1.6r_n} := P_{1.6r_n} + \Delta P$$

Now, we introduce the pre-equalization portion. For OFDMA, the CM then transmits each data subcarrier with the target power:

$$P_{t_sc_i} = P_{1.6r_n} - P_{1.6delta_n} + Pre\text{-}Eq_i - 10 * \log_{10}(number\ of\ subcarriers\ in\ 1.6\ MHz\{32\ or\ 64\})$$

where Pre-Eq$_i$ is the magnitude of the ith subcarrier pre-equalizer coefficient (dB), and $P_{1.6delta_n}$ equals 0 dB for non-boosted channels, 0.5 dB for boosted channels with 25 kHz subcarrier spacing, and 1 dB for boosted channels with 50 kHz subcarrier spacing.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

11.2. The Pre-Equalization Vector is Normalized to Unity Power and Application of Pre-Equalization Does Not Alter the Cable Modem Transmit Power

In this section we illustrate the normalization of the Pre-Eq vector which is required to be performed in the CM, and we illustrate how as a result of the normalization the CM transmit power is not impacted by the application of the Pre-Eq vector.

Let's (for here) call the ith coefficient of the normalized Pre-Eq vector as:

$$pe_i = I_i + jQ_i$$

We can then let:

$P_{pre\text{-}eq_i}$ = power of ith Pre-Eq coefficient, but in linear power rather than dB,

$$P_{\text{pre-eq}_i} = \text{power of the } i^{th} \text{ Pre-Eq coefficient, but in linear power rather than dB}$$

so:

$$P_{\text{pre-eq}_i} = |pe_i|^2 = I_i * I_i + Q_i * Q_i$$

Recall we had *Pre-Eq_i* as the power of the ith pre-eq vector expressed in units of dB:

$$\begin{aligned}\text{Pre-Eq}_i[dB] &= 10 * \log_{10}(|pe_i|^2) \\ &= 10 * \log_{10}(I_i * I_i + Q_i * Q_i)\end{aligned}$$

And thus,

$$\text{Pre-Eq}_i[\text{dB}] = 10 * \log_{10} P_{\text{Pre-Eq}_i}$$

And also note, due to the normalization of the Pre-Eq coefficients, the following condition is ALWAYS satisfied (presuming spec-compliant operation of the CM):

-With the number of active (that is, non-excluded) subcarriers in the OFDMA channel being N,

-Sum of $\{P_{pre\text{-}eq_i} / N\}$ over all non-excluded subcarriers is 1.

Consider a pre-equalization vector which is not yet normalized, $P_{pre\text{-}eq_unnormalized_i}$:

Total_Pre-Eq_Power_unnormalized = Sum of $\{P_{pre\text{-}eq_unnormalized_i}\}$ over all active (non-excluded) subcarriers

Then to NORMALIZE the Pre-Eq coefficients to unity power, they should be divided by a scaling factor:

$$\text{Scaling_Factor} = \sqrt{\text{Total_Pre-Eq_Power_unnormalized}}$$

So that the scaled pre-equalization normalized values are computed as:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

$$pe_{normalized_i} = pe_i = \frac{pe_{unnormalized_i}}{\text{Scaling_Factor}}$$

$$= \frac{(I_{unnormalized_i} + jQ_{unnormalized_i})}{\text{Scaling_Factor}}$$

$$= (I_i + jQ_i)$$

With the above notation, we can formally walk through the calculations applying the pre-equalization coefficient values to each subcarrier, collect terms, and mathematically show that the CM transmit power after application of normalized pre-equalization values does not alter the CM transmit power.

We desire to add up the impact of *Pre-Eq*$_i$ (dB) over all non-excluded subcarriers (N), which we do in linear power.

$$10^{P_{t_sc_i}/10} = \frac{10^{P_{1.6r_n}/10} * P_{\text{Pre-Eq}_i}}{\left(N * 10^{P_{1.6delta_n}/10}\right)}$$

And so, to sum over the non-excluded subcarriers, we have

$$\sum_{i=1}^{N} 10^{P_{t_sc_i}/10} = \sum_{i=1}^{N} \frac{10^{P_{1.6r_n}/10} * P_{\text{pre-eq}_i}}{(N * 10^{P_{1.6delta_n}/10})}$$

$$\sum_{i=1}^{N} 10^{P_{t_sc_i}/10} = \frac{10^{P_{1.6r_n}/10}}{10^{P_{1.6delta_n}/10}} \sum_{i=1}^{N} \frac{P_{\text{pre-eq}_i}}{N}$$

$$\sum_{i=1}^{N} 10^{P_{t_sc_i}/10} = \frac{10^{P_{1.6r_n}/10}}{10^{P_{1.6delta_n}/10}}$$

Where the last step is because the sum of $\{P_{pre\text{-}eqi} / \text{N}\}$ over all non-excluded subcarriers is 1. So,

$$P_{t_sc} = P_{1.6r_n} - P_{1.6delta_n}$$

Thus, the pre-equalization coefficients have no impact on the channel power!

Note: Many of the definitions of the terms used in this analysis are discussed in this paper, and are also included in Section 7.4.12.3.1 of [D3.1PHY].

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

ALC	automatic level control
CCAP	converged cable access platform
CM	cable modem
CMTS	cable modem termination system
dB	decibel
DBC	dynamic bonding change
dBmV	decibel millivolt
DOCSIS	Data-Over-Cable Service Interface Specifications
DRW	dynamic range window
e.g.	exempli gratia (for example)
°F	degree Fahrenheit
FEC	forward error correction
HFC	hybrid fiber/coax
Hz	hertz
i.e.	id est (that is)
kHz	kilohertz
MHz	megahertz
Min Ch	minimum channel [headroom]
OFDMA	orthogonal frequency division multiple access
PHY	physical layer
PMA	profile management application
PNM	proactive network maintenance
PSD	power spectral density
QAM	quadrature amplitude modulation
RF	radio frequency
RNG-REQ	ranging request
RNG-RSP	ranging response
RPD	remote PHY device
S-CDMA	synchronous code division multiple access
SC-QAM	single carrier quadrature amplitude modulation
SCTE	Society of Cable Telecommunications Engineers
TCP	total composite power
TCS	transmit channel set
TDMA	time division multiple access

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

[D1.0RFI] DOCSIS 1.0 Radio Frequency Interface Specification SP-RFI-C01-011119

[D3.0PHY] DOCSIS® 3.0 Physical Layer Specification CM-SP-PHYv3.0-C01-171207

[D3.1PHY] DOCSIS® 3.1 Physical Layer Specification CM-SP-PHYv3.1-I20-230419

[D4.0PHY] DOCSIS® 4.0 Physical Layer Specification CM-SP-PHYv4.0-I06-221019

[LLALC] Hranac, R., "Long Loop ALC," Fall 2019 issue of *Broadband Library*, https://broadbandlibrary.com/long-loop-alc/

[SPLIT] Hranac, R., "Understanding Band Splits In Two-Way Networks," Spring 2020 issue of *Broadband Library*, https://broadbandlibrary.com/understanding-band-splits-in-two-way-networks/

[TPPSD] Hranac, R., "Total Power and Power Spectral Density," Summer 2019 issue of *Broadband Library*, https://broadbandlibrary.com/total-power-and-power-spectral-density/

[SCTEMath] SCTE OPERATIONAL PRACTICE, SCTE 270 2021r1 Mathematics of Cable

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Case Studies From 5G Wireless Private Networks in the Mining Sector

A Technical Paper prepared for SCTE by

Andy Striegler
Sr. Director, Enterprise Customer Solutions
Rogers Communications Inc.
8200 Dixie Road, Brampton, Ontario, Canada
+1 647-747-0644
andy.striegler@rci.rogers.com

Kamaljit Bal
WPN Architect, Enterprise Customer Solutions
Rogers Communications Inc.
8200 Dixie Road, Brampton, Ontario, Canada
+1 416-825-7084
Kamaljit.bal@rci.rogers.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Private Network Definition ... 3
3. Why Private Networks ... 4
4. Private Networks Architectures ... 5
   - 4.1. Standalone Private Network Deployment ... 5
   - 4.2. Hybrid cloud private deployment ... 6
   - 4.3. Network Slice based private network deployment ... 7
   - 4.4. Spectrum related requirements ... 8
5. Wireless Private Network: Coal Mine in British Columbia ... 8
   - 5.1. Introduction ... 8
   - 5.2. Solution Architecture ... 8
   - 5.3. Outcome ... 10
6. Wireless Private Network: Open Pit Gold Mine in Northern Ontario ... 11
   - 6.1. Introduction ... 11
   - 6.2. Solution Architecture ... 11
   - 6.3. Outcome ... 12
7. Conclusion ... 13

Abbreviations ... 14

Bibliography & References ... 14

List of Figures

Title **Page Number**

Figure 1- On-premises full deployment Model for WPN ... 6
Figure 2- Hybrid cloud Model for WPN ... 6
Figure 3- Network Slice Model for WPN ... 7
Figure 4- Customer Mine Site ... 9
Figure 5- high Level Architecture ... 10
Figure 6- Customer Mine Site ... 11
Figure 7- high Level Architecture ... 12

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

A new generation of private 5G networks is emerging to address critical wireless communication requirements in public safety, industrial operations, and critical infrastructure. Advanced network features and services are needed to serve mission-critical and business-critical use cases.

Within the industrial market segment, the mining sector has been an early adopter of wireless private technology and has been investing heavily in wireless private networks (WPNs) around the globe to enhance overall productivity, operational efficiency, and safety.

In recent years, Rogers has deployed multiple WPNs in the mining sector in Canada. In 2020, Rogers launched Western Canada's first private LTE network with one of the latest coal mine in British Columbia and later deployed 5G-Ready private network solutions at an Open Pit Gold Mine in Northern Ontario. The WPNs were designed to provide reliable, low-latency wireless communications primarily for autonomous haul truck operations and other critical in-pit applications.

On-premises deployment architecture model was chosen as a standalone private network. To enable a fully operational 5G Wireless Private Network at the mine, Rogers has deployed multiple cellular network towers with its full range of spectrum frequency bands to support a diverse set of use cases and applications throughout the mine site. Radio Access Network (RAN) Sites were configured to enhance efficiency, increase Bandwidth, reduce latency to less than 20ms. Wireless Private Network sites are connected to fully redundant High Availability (HA) Mobile Cores deployed locally on mines Data Centers. Each site has fiber transport backed up by Microwave link. The network has been built with a full failover backup system across the site. The system is designed to guarantee automated deployment and configuration, and efficient operation and maintenance throughout its life.
The authors of this paper and presentation will discuss the lesson learned, key system requirements, use cases, design considerations, operations, and network performance.

2. Private Network Definition

The definition of a private mobile network used in this report is a 3GPP-based 4G LTE or 5G network intended for the sole use of private entities, such as enterprises, industries, and governments. The definition includes MulteFire or Future Railway Mobile Communication System. The network must use spectrum defined in 3GPP, be intended for business-critical or mission-critical operational needs. Non-3GPP networks such as those using Wi-Fi, TETRA, P25, WiMAX, Sigfox, LoRa and proprietary technologies are excluded from the Private network list. Furthermore, network implementations using solely network slices from public networks or placement of virtual networking functions on a router are also excluded. Where identifiable, extensions of the public network (such as additional sites deployed at a location, as opposed to dedicated private networks), are excluded. These items may be described in the media as a type of private network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Why Private Networks

The demand for private mobile networks based on 4G LTE (and increasingly 5G) technologies is being driven by the spiraling data, security, digitization and enterprise mobility requirements of modern business and government entities. Organizations of all types are combining connected systems with big data and analytics to transform operations, increase automation and efficiency or deliver new services to their users. Wireless networking with LTE or 5G enables these transformations to take place even in the most dynamic, remote, or highly secure environments, while offering the scale benefits of a technology that has already been deployed worldwide.

Private mobile networks are also often part of a broader digital transformation program in an organization. This could include the introduction or development of cloud networking and other digital technologies such as artificial intelligence and machine learning, and data analytics. More and more applications of the private mobile network will use these capabilities combined with mobile connectivity.

There are several reasons for the growing demand for 4G and 5G private networks. Here are some key factors driving this trend:

Enhanced Security: Private networks provide an extra layer of security compared to public networks. Industries that deal with sensitive data, such as government agencies, financial institutions, healthcare organizations, and critical infrastructure sectors, require robust security measures to protect their data and operations. Private networks offer increased control over network access and data management, reducing the risk of unauthorized access and data breaches.

Customized Solutions: Private networks allow organizations to tailor their network infrastructure according to their specific requirements. They can optimize network performance, allocate resources efficiently, and prioritize critical applications. This customization ensures reliable and predictable network connectivity, reducing latency and improving overall efficiency.

Industrial Internet of Things (IIoT) Applications: The rise of the Industrial Internet of Things has led to an increased demand for private networks. IIoT applications, such as smart manufacturing, autonomous vehicles, remote asset monitoring, mining, and precision agriculture, often require low latency and high bandwidth connections. Private networks can offer dedicated and reliable connectivity to support these demanding applications.

Mission-Critical Communications: Certain industries, such as public safety, defense, and emergency services, require uninterrupted and reliable communication for their critical operations. Private networks can ensure dedicated and prioritized connectivity, enabling real-time communication and coordination in emergencies or mission-critical situations.

Improved Performance and Capacity: 4G and 5G private networks offer superior performance and increased capacity compared to public networks. These networks can handle a larger number of connected devices simultaneously and provide faster data transfer rates, enabling bandwidth-intensive applications and services.

MEC: 5G Multi-Access Edge Computing (MEC) brings compute power closer to the network edge, eliminating the time it takes to run data to a centralized data center. This reduces communication latency, increases system performance, and leads to new business opportunities in the private network space.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Overall, the private network market will experience substantial growth over the next several years. These networks offer organizations greater control, reliability, and tailored services to meet their specific requirements. As a Canadian leader in 5G innovation, Rogers has been developing 5G use cases in several sectors in partnerships with the Canadian universities and industry consortiums. In these following sections, we will present case studies from two different private network deployments done by Rogers.

4. Private Networks Architectures

5G technology holds immense importance for industrial use cases due to its transformative capabilities that revolutionize the way industries operate. With its ultra-low latency, high data rates, and massive device connectivity, 5G enables a range of applications that significantly enhance productivity, efficiency, and safety in industrial settings. Multiple architecture and partnership options can be considered when implementing private wireless solutions. Choosing the best approach depends on multiple factors such as the character of various sites, the use cases, and the Private enterprise capabilities. In certain use cases, it is necessary for the data to remain within the confines of the organization to maintain privacy and allay security worries. In such cases, some (if not all) of the network entities need to reside locally in the organization's premise and/or be owned by the organization.

4.1. Standalone Private Network Deployment

In this model, a small standalone cellular network with associated management and operations tools and applications, user end points, and radios are integrated into the enterprise LAN and managed by the enterprise IT department or a service provider. The security and administration of this private cellular network are often handled separately, and there is typically little to no integration of this network with other enterprise network components. Figure 1 shows the high-level architecture of On-premises private network deployment. This concept has been used to create private 4G LTE networks in locations that require cellular access but do not have coverage from major providers, mining is a typical example.

In many new private 5G deployment scenarios, this model is being looked at as the first alternative. This model is typically used in early proof-of-concept and trial deployments to assess the maturity of use cases and deployment options. Larger more complex deployments, for example multi-site enterprises with multiple use cases for private cellular, may find this model limiting and too complex.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

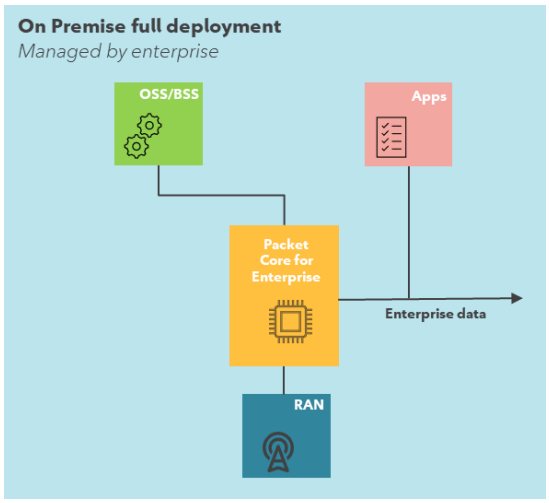


Figure 1- On-premises full deployment Model for WPN

4.2. Hybrid cloud private deployment

As a result of business demands for less complicated management of a stand-alone private cellular network, cloud hosted private deployment options have been emerging in recent years.

These models involve placing a portion of the private cellular network on the premises of the business, on an edge platform, and managing it remotely via the cloud. Depending on the architecture, the precise mix of on-premises to cloud infrastructure can change. In the simplest scenario, the radio and packet core are on-site at the enterprise, while the management applications are hosted in the cloud. Another option is to place the packet core, or key components of a virtualized packet core such as the User Plane Function (UPF), is on premise. The ultimate deployment architecture is determined by the demand for throughput and latency as data moves across the cloud. Figure 2 shows the high-level architecture of hybrid cloud private network deployment

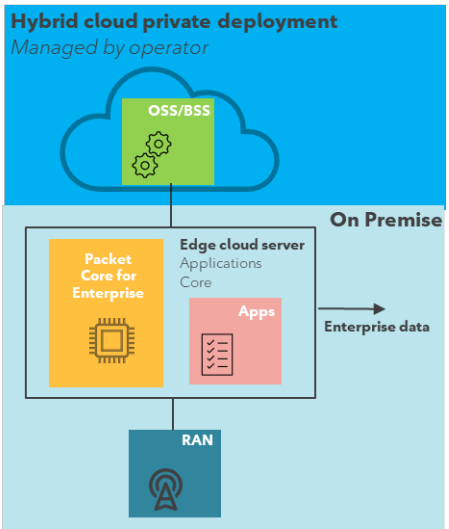


Figure 2- Hybrid cloud Model for WPN

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

These models can be created in novel ways thanks to the flexibility brought about by the 5G virtualized packet core and edge platform advancements. The cloud hosted options are attractive for enterprises as they can offload complexity of operations of a standalone cellular network to the cloud provider operator. However, they suffer from the same drawback as the standalone models, as they depend on leased or shared/local spectrum availability, unless enterprises or system integration partners own spectrum that can be used in these models.

4.3. Network Slice based private network deployment

The network slice deployment model is where a Mobile Network Operator (MNO) dedicates a "slice" of their existing commercial cellular network to an enterprise. A slice can include a set of radios, spectrum bands, fiber network capacity, 5G Core, and other collaterals as defined by the MNO. In this model, the operator will continue to operate the slice of the network that is dedicated to the enterprise for a cost and will integrate the slice into the enterprise network enabling the enterprise to co-manage the slice through MNO operations portal that will be made available. The minimum core elements and applications are deployed on the premises to ensure reliability, low latency, and preserve the confidentiality of the data. The rest of the core elements and applications run in the service provider cloud and rely on existing cloud core management.

Figure 3 shows the high-level architecture of slice based private network deployment

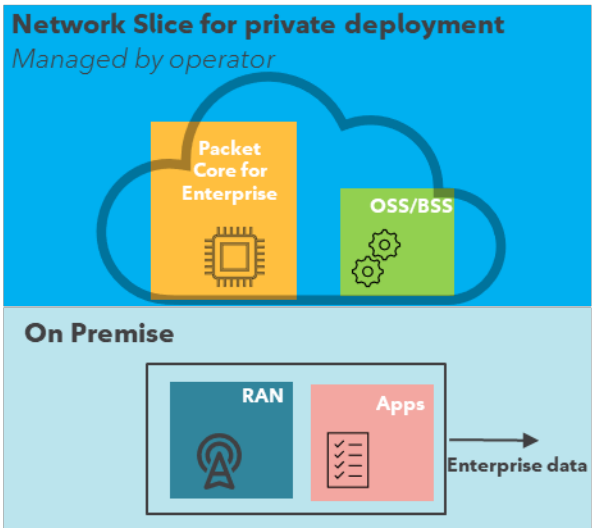


Figure 3- Network Slice Model for WPN

The level of control, security, and management of the slice is specified and agreed to through Service Level Agreements (SLA) that will be made between the provider and the enterprise. This network slice model is extremely appealing as it can open the vast resources of cellular providers: spectrum, footprint, coverage, & expertise, to enterprises. However, it remains to be seen how providers can successfully monetize this model as it might directly compete with the profitable consumer cellular use cases that providers have hitherto prioritized.

In conclusion, the complexity of implementation and operational costs are highest for autonomous private architecture and lowest for core slicing. Data confidentiality, security and reliability typically decreases

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

with as-a-service and core slicing solutions. A hybrid private–public architecture could be a more appropriate choice if the operations are spread over a wider area, need connectivity with off-site and mobile assets, or prefer to outsource this part of the enterprise IT operation. Dedicated private wireless solution could be more suitable if enterprise need to have full control of their network and ensure the reliability and availability of the local operations when failures occur in the outside world.

4.4. Spectrum related requirements

Spectrum is crucial to any wireless network deployment. A key factor influencing the uptake of wireless solutions is the question of how to handle spectrum for industrial purposes, since reliable connectivity demands licensed spectrum. Some countries provide spectrum dedicated for industrial use, whereas others do not. Communication Service Providers (CSP) are in the prime position to optimally address these industrial connectivity needs with powerful 5G networks and business models focused on industries. There is a huge opportunity for CSPs to address industrial connectivity needs with 3GPP based cellular technologies. The opportunity encompasses a range of industries, including diverse segments with diverse needs, such as those in the manufacturing, mining, port, energy, and utilities, automotive and transport, public safety, media and entertainment, healthcare, and education industries, amongst others.

The more available spectrum, the more enriched use cases can be enabled. In this context, private mobility is an accelerator, as it allows service provisioning even before availability of the public network. Private mobility also enables tailoring an optimized wireless solution to match the target application requirements. Government officials and bodies recognize the benefits widespread 5G deployment can bring to various industries and public services. There has been a notable shift in focus towards dedicating spectrum for the deployment of private cellular networks.

5. Wireless Private Network: Coal Mine in British Columbia

5.1. Introduction

Customer has recognized that mining companies are at a tipping point. As they strive to boost productivity and efficiency, attain safety and eco-sustainability, and deliver higher shareholder value, customer reimagines their operations paradigms and embraces new digital innovations and technologies.

Customer has taken a giant step to modernize its company with a critical infrastructure upgrade built for demanding applications, including Autonomous Haulage Systems (AHS), Internet of Things (IoT), Push-to-Talk (PTT), advanced analytics and many more.

Customer’s vision to create a genuinely Private LTE (pLTE) infrastructure to enable their digital transformation strategy called for an innovative partnership with Rogers. Since then, pLTE deployment has become a foundational component for customers digital transformation. By design, it is robust, scalable and application ready. pLTE is an enabler for various end devices, user equipment (UE) and sensors that will drive rich data analytics and decision support tools.

5.2. Solution Architecture

Rogers launched Western Canada’s first private LTE network at one of the Coal Mine in British Columbia. This pLTE were designed to provide reliable, low-latency wireless communications primarily for autonomous haulage truck operations and other critical in-pit applications.

The Network design involves geo-redundancy to design highly available applications. Customer chose geo-redundant architectures to avoid downtime by distributing applications and infrastructure across significant

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

distances. For improved reliability and capacity, the two identical systems are placed in a group, and their configurations are synchronized to prevent a single point of failure. A typical application of geo-redundant architecture is dealing with natural disasters, which may cripple certain areas and ensure 0% migration downtime.

The continuous terrain changes in mining industry makes design phase challenging. The design should consider Mine Evolution "terrain changes". Reliable and real-time propagation loss modeling play a significant role in the efficient planning, development, and optimization of macro cellular communication networks in each terrain. Thus, the need to create/adapt or regularly tune an existing design to enhance its signal propagation accuracy in a specified terrain becomes imperative.

Below are pictures from the customer mine location showcasing the radio install and coverage plan using fixed communication towers and CoWs due to changing terrain.



Figure 4- Customer Mine Site

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- On-premises deployment architecture model was chosen as a standalone private network to enable a fully operational pLTE Network at the mine.
- Rogers has engineered the network with dedicated multiple spectrums to address peak usage and future requirements.
- RAN Sites are designed to consider terrain changes, and configured to support enhance efficiency, increase Bandwidth, reduce latency.
- The Core sites are designed with fully geo-redundant HA architecture.
- The network has been built with a full failover backup system with the site.
- Considering the entire network is deployed behind the Customers LAN network which means, it is fully secured from application/IT/Management Perspective.
- Network support a diverse set of use cases and applications such as AHS, PTT
- Enable Customers digitalization innovation strategy which includes Autonomous Haulage System, PTT, manned mining apps and UE manufacturers (AVI).

Figure 5 shows the high-level architecture of Customer Ltd pLTE network deployment

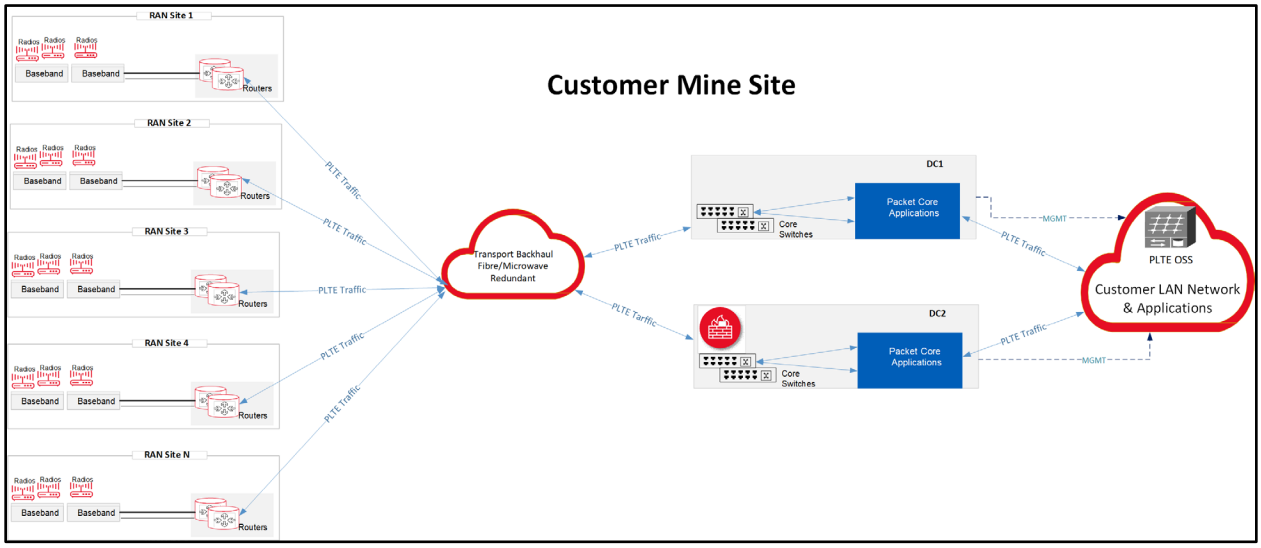


Figure 5- high Level Architecture

5.3. Outcome

- Considering the entire network is deployed behind the Customers LAN network which means, it is fully secured from application/IT/Management Perspective.
- Reliable and real-time propagation loss modeling play a significant role in the efficient planning, development, and optimization of Radio networks in different terrains.
- Enable customer's digitalization innovation strategy which includes Autonomous Haulage System, PPT, manned mining apps and UE manufacturers (AVI).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Reliable coverage and backhaul connectivity increased the operational accuracy of the field assets and devices
- Production Process efficiency has improved by broader network coverage for assets and devices using fleet management application.
- Allow workers/miners to spend more time in the field and access/process applications data remotely hosted in the central facility.

6. Wireless Private Network: Open Pit Gold Mine in Northern Ontario

6.1. Introduction

The Rogers Wireless Private Network is deployed in Open bit Gold Mine in Northern Ontario, it is to provide the private LTE/5G coverage at the mine with high capacity, low-latency solution to improve Mine Safety & increase Mining Operations Efficiency.

6.2. Solution Architecture

Gold Mine becomes the first open pit mining operation in Canada to be fully connected over LTE/5G Wireless Private Network. Dedicated LTE/5G network key to transform into Smart Digital Mine of the future.

Rogers has deployed multiple cellular network towers with its full range of spectrum frequency bands to support a diverse set of use cases and applications throughout the mine site. RAN Sites were configured to enhance efficiency, increase Bandwidth, reduce latency to less than 20ms.

Wireless Private Network sites are connected to fully redundant HA Mobile Cores deployed locally on mines Data Centers. Each site has fiber transport backed up by Microwave link. The network has been built with a full failover backup system across the site. The system is designed to guarantee automated deployment and configuration, and efficient operation and maintenance throughout its life.

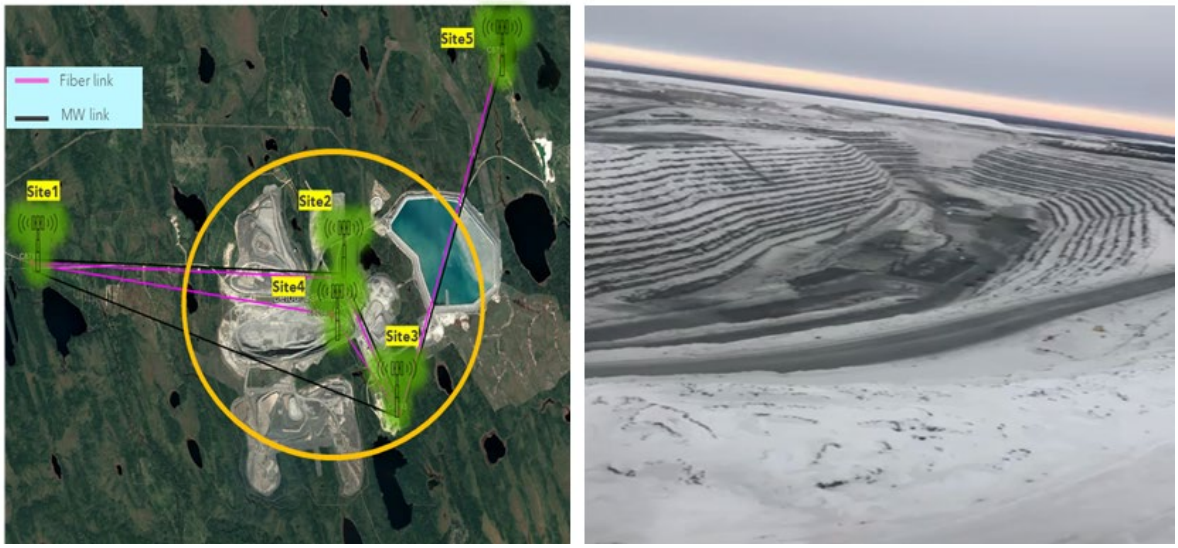


Figure 6- Customer Mine Site

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- On-premises deployment architecture was chosen with Cloud based management network to enable a fully operational 5G Wireless Private Network at the mine:
- Rogers has deployed multiple cellular network towers with full range of spectrum frequency bands.
- RAN Sites were configured to enhance efficiency, increase Bandwidth, reduce latency to less than 20ms.
- Wireless Private Network sites are connected to fully redundant HA Mobile Cores.
- The network has been built with a full failover backup system across the site.
- Network support a diverse set of use cases and applications throughout the mine site.
- Enable digitalization of mining use cases such as Autonomous Haulage and Drilling System, teleremote operations which require low latency less than 20ms.

Figure 7 shows the high-level architecture of Open Pit Gold network deployment

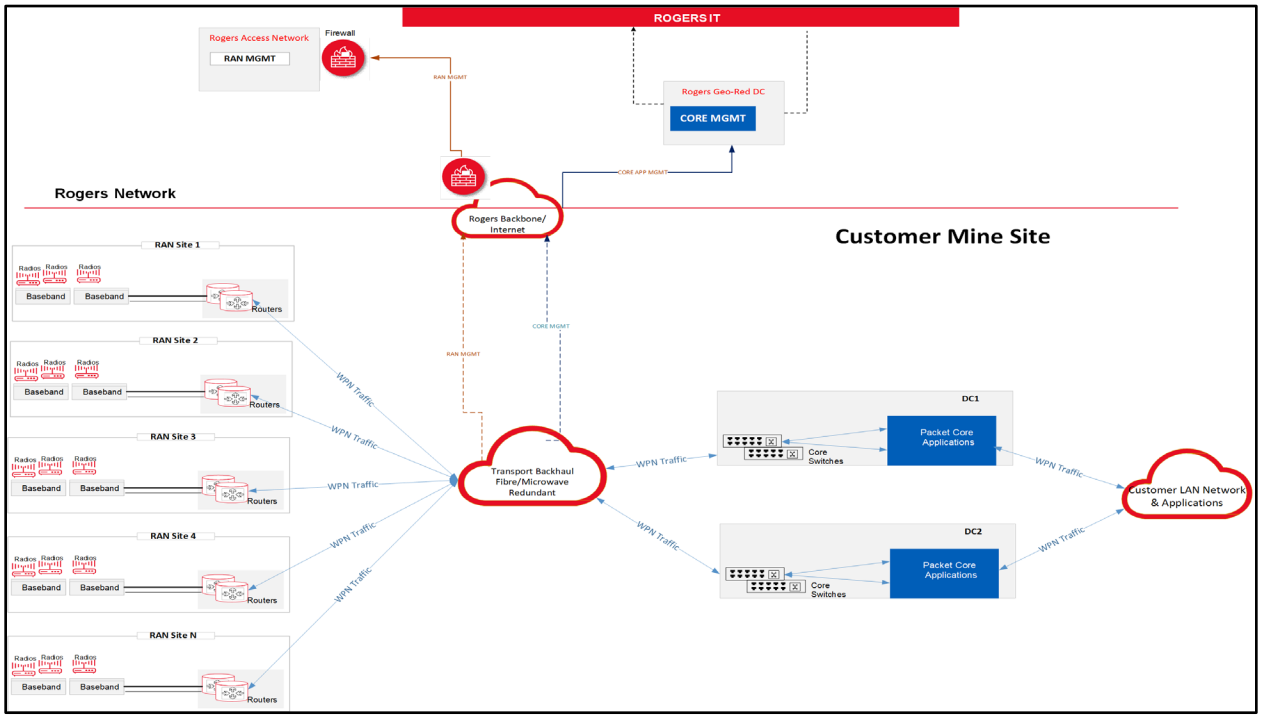


Figure 7- high Level Architecture

6.3. Outcome

- Production Process efficiency has improved by broader network coverage for assets and devices using fleet management application.
- Enable digitalization of mining use cases such as Autonomous Haulage and Drilling System, Blast loading SW which require low latency less then 20ms and some other high latency tire pressure sensor application.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- The round-trip latency between the devices and applications are measured less than 20ms.
- Wireless Private Networks enables high-capacity real time video surveillance system attached to Autonomous Haulage vehicles to upload around 60 Mbps data.
- Reliable coverage and backhaul connectivity increased the operational accuracy of the field assets and devices
- Allow workers/miners to spend more time in the field and access/process applications data remotely hosted in the central facility.

7. Conclusion

Wireless Private Networks resolve strategic challenges in the mining industry such as mine safety, operational efficiency, Mission Critical use cases, Data accuracy, Privacy, and capacity issue (Bandwidth/latency).

Canadian mining industry is gradually adopting the latest technology to deploying, integrating a managed, modular, and customizable Wireless Private Network to ensure that mining sector benefits from next-generation innovation

- Wireless Private Networks are key to the success of Mining Industrial evolution worldwide.
- Enable digitalization of mining use cases such as Autonomous Haulage System, and Drilling System, Blast loading SW and teleremote operations which require low latency less than 20ms and some other high latency tire pressure sensor application.
- The round-trip latency between the devices and applications are measured less than 20ms.
- Wireless Private Networks enables high-capacity real time video surveillance system attached to Autonomous Haulage vehicles to upload around 60 Mbps data.
- Considering Wireless Private Networks are deployed dedicated for the customer with industry grade HW, the packet drop within the network is less then 0.25%.
- Reliable and real-time propagation loss modeling play a significant role in the efficient planning, development, and optimization of Radio networks in different terrains.
- Production Process efficiency has improved by broader network coverage for assets and devices using fleet management application.
- Allow workers/miners to spend more time in the field and access/process applications data remotely hosted in the central facility.
- Wireless Private Network meet all the requirements laid by the mining industry such as operational efficiency/Mine Safety/coverage/Services Availability.
- CSP’s along with their vendor partners are contributing to the Wireless Private Network Market in Canada with innovation and research to help drive revenue

As mining industry continues to face multiple complex challenges, from uncertainty across geopolitical landscape to the disruption of digital economy, resulting in increased pressure as productivity in operation

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

needs to improve. To capitalize on digital revolution, mining companies would benefit from 5G and be the catalyst for this transformation:

- Up to 25% increase production
- Up to 40% drill operations
- Up to 20% energy savings

5G networks have the potential to provide opportunities to the mining industry such as automation and remote operations due to improved coverage, lower latency, and higher reliability at every stage of its business operations.

Abbreviations

AP	Access Point
LTE	Long Term Evolution
IIOT	Industrial Internet of Things
MNO	Mobile Network Operator
CSP	Communications Service Provider
AHS	Autonomous Haulage System
IOT	Internet of Things
PTT	Push to talk
UE	User Equipment
COW	Cell on Wheel
RAN	Radio Access Network
HA	High availability
LAN	Local Area Network
Mbps	Megabits per second
HD	high definition
SCTE	Society of Cable Telecommunications Engineers

Bibliography & References

The ARRL Antenna Book, 20th Ed.; American Radio Relay League

Code of Federal Regulations, Title 47, Part 76

Reflections: Transmission Lines and Antennas, M. Walter Maxwell; American Radio Relay League

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Cloud Versus MEC for Telcos and MSOs

MEC Use Case Study

A Technical Paper prepared for SCTE by

Yassar Abbas
5G Innovation Lead Architect
Rogers Communications Inc.
8200 Dixie Road, Brampton ON
+14166252051
Yassar.abbas@rci.rogers.com

Brad Tizard, Rogers Communications Inc.

Probir Khaskel, Rogers Communications Inc.

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
1.1. Edge Compute Benefits 5
1.1.1. Latency 5
1.1.2. Bandwidth 5
1.1.3. Security 5
2. Edge Compute Definitions 5
2.1. Private MEC 5
2.2. Public MEC 6
2.3. Public Cloud 6
2.4. Cloud Edge 6
2.5. Hybrid Cloud 6
2.6. Near-Edge and Far-Edge 6
2.7. Content Distribution Network 6
3. Tesbed Architecture 7
3.1. Performance Results 9
3.1.1. Latency 9
3.1.2. Jitter 10
4. AR based Sports Fan Experience 11
4.1. Overview 11
4.2. Architecture 11
4.3. Conclusion 13
4.3.1. Latency 13
5. MEC based AR/VR Video Rendering 14
5.1. Introduction 14
5.2. Architecture 14
5.3. Conclusion 15
5.3.1. Battery 15
5.3.2. Bandwidth 15
5.3.3. Remarks 16
6. Cross Operator Music Jam 16
6.1. Overview 16
6.2. Architecture 16
6.3. Conclusion 18
7. Last Mile Food Delivery using Autonomous Robot 19
7.1. Overview 19
7.2. Architecture 19
7.3. Conclusion 20
8. Overall Conclusion 21

Abbreviations 22

Bibliography & References 22

List of Figures

Title **Page Number**

Figure 1 – Edge Compute Locations 7
Figure 2 – Rogers MEC Infra Deployment for POCs 8
Figure 3 – MEC and Public Cloud Locations 9

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 4 – Latency Improvements Public Cloud versus Rogers MEC 10
Figure 5 – Jitter Improvements Public Cloud versus Rogers MEC 10
Figure 6 – Immersiv.io Application Architecture 11
Figure 7 – AR Based Sports Fan-Experience POC Architeture 13
Figure 8 – Impact of Latency on User Experience 13
Figure 9 – MEC based AR/VR Video Rendering Architecture 15
Figure 10 – Technical Architecture for 5GFF Music Festival 17
Figure 11 – Edge Discovery Service API deployment 18
Figure 12 – Last Mile Food Delivery Robot POC Architecture 20

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The need for lower latency applications, decentralized processing power and increased adoption of IoT and connected devices are giving rise to the shift of data processing closer to the source commonly known as edge computing or MEC (Multi-access Edge Computing).

[1]Multi-access Edge Computing (MEC) offers application developers and content providers cloud-computing capabilities and an IT service environment at the edge of the network. This environment is characterized by ultra-low latency and high bandwidth as well as real-time access to network information that can be leveraged by applications.

This brings a unique opportunity for Telcos and cable operators to deploy compute resources leveraging their distributed facilities across their geographic footprint. Edge computing is particularly interesting in this era for MSO and Telcos with virtualization and decentralization of network functions moving user plane workloads closer to edge co-located with third-party applications to reduce latency.

The opportunities that edge computing presents are attractive; however, network operators need to identify the right use cases and applications that are candidates for edge computing.

In this paper, authors will share lessons learned from some of the MEC use cases like Cross-operator music jam session from different parts of the world, AR (Augmented Reality) based in-stadium fan experience during a live NHL game, last mile food delivery using autonomous robots and VR (Virtual Reality) banking experience.

These use cases have been tested and demonstrated in Roger's network leveraging MEC as part of the 5G Innovation and Partnership program.

This paper also provides a brief description of edge computing terminology and how network edge or MEC is related to a hyper-scaler public cloud. Please note that there is no standard terminology, and you may see other terms used to refer to the same thing in other material. However, the focus is really to highlight the difference between network edge, public cloud, and cloud edge.

[1]MEC initiative is an industry-standard specification group within ETSI to define standard and open environment which can allow interoperability across mobile operators, application developer, Independent Software Vendors (ISV), telecom equipment vendors and technology providers. However, the use of MEC terminology has been applied much more broadly than solely to ETSI standards and interoperability.

Although the focus of MEC use cases has mainly been on wireless access-based services but as the name implies, it is inclusive of all access network technologies including fixed and WLAN.

In some literature, MEC has also been called Mobile Edge computing due to its early focus on 5G and wireless use cases. The use of Edge computing terminology is very generic across the industry and subject to interpretation. For example, for user device manufacturers edge is on the device while for car OEMs edge is in the car. Our focus in this writing is the network edge use cases and not on standards and interoperability.

MEC infrastructure can be used to deploy operator's network functions (VNF) workloads or to host third-party applications. Operator own network workloads or VNFs are typically deployed on private MEC or private cloud however the use of public cloud has also been seen in the case of some greenfield deployments.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1.1. Edge Compute Benefits

There are three main benefits of edge computing when it comes to differentiating from the cloud.

1.1.1. Latency

Latency improvements are probably the most talked about attribute of MEC. Cloud computing needs processing of data in centralized data centres sent over the internet resulting in higher latency not suitable for applications like such as industrial applications or cloud gaming. MEC or edge computing in general is ideal for applications requiring low latency since data is processed closer to the source and edge of the network.

1.1.2. Bandwidth

Cloud computing requires transporting data over the internet far from data sources. This requires backhauling of the data traffic and increased bandwidth demands. In comparison to cloud computing, MEC can save bandwidth by computing at the edge and not requiring data to be transported to the cloud.

1.1.3. Security

With cloud computing data is processed and stored in remote data centres owned by cloud providers. MEC can alleviate those concerns by processing data locally and not transmitting it over the internet. With cloud computing data in transit can be protected using encryption and tunneling techniques at additional cost. Also, in some countries, there are data residency requirements that mandate local data processing and storage requirements.

2. Edge Compute Definitions

There are many different forums defining the specifications and standards for IT platforms enabling the use of MEC to locate the closest MEC and instantiate compute, storage, and other resources to host the application. It is not the purpose of this paper to delve into standards and MEC platform details but rather to describe MEC as a general concept. The use of MEC and edge computing has been used interchangeably in this document. However, for the sake of clarity to readers, we would define some key terms and nomenclature in the context of this paper.

2.1. Private MEC

Any edge compute infrastructure deployed dedicated to a single Enterprise is called private MEC or cloud. It is typically a single tenant serving only one organization or Enterprise. It can be deployed on customer premises aka on-premises (factory floor, warehouse, etc.) or on operator facilities very close to the network edge (e.g., cell tower location) to serve ultra-low latency use cases. Enterprises can combine the benefits of on-premises wireless private networks and private MEC to optimize latency and throughput while leveraging technologies like computer vision, AR/VR/XR, and machine learning.

For Telco and MSOs, private MEC or cloud is typically hosting the VNF (virtual network functions). With CUPS (control and user plane separation), user plane functions can be deployed closer to the edge resulting in latency improvement and bandwidth saving by not backhauling traffic to the central cloud data centre.

Azure Stack Edge and AWS outpost family are examples of Hyperscaler private MEC. Operators can also choose to deploy their own private MEC/cloud infrastructure and services.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.2. Public MEC

Public MEC is an extension of the Public Cloud and supports multi-tenancy. AWS Wavelength zones are an example of Public MEC deployed at network operator facilities in core data centres. Public MEC is used to host third-party application providers or ISV (Independent Software Vendor) workloads closer to the network edge. Anyone can subscribe to Public MEC just like anyone can subscribe to Public Cloud.

AWS wavelength zones are an example of Public MEC offered in partnership with HCP and network operators.

2.3. Public Cloud

Public cloud is a cloud offering by hyper-scale cloud providers such as AWS, Azure, and GCP (Google Cloud Platform). Public cloud providers' data centres are centralized but geographically spread across regions over a wide area providing global coverage. When it comes to latency and bandwidth optimization, the physical placement of public cloud providers data centres plays an important role. Public cloud is ideal for applications needing large-scale compute and latency-tolerant applications.

2.4. Cloud Edge

Cloud edge is edge locations or POPs owned by HCP which are closer to users in comparison to public cloud providers' regions and availability zones. AWS local zones or edge locations POP is an example of cloud edge. A distinction should be made between HCP edge services versus cloud edge. HCP can provide edge services and infrastructure anywhere at the network edge or on-premises in partnership with network operators at POPs owned by network operators at the far-edge or near-edge.

2.5. Hybrid Cloud

A hybrid cloud is a mix of on-prem compute, private or public MEC, and public cloud environments. Workloads are deployed based on application latency, security, and data residency requirement.

2.6. Near-Edge and Far-Edge

Near-Edge and Far-Edge are the terms describing how far Edge is from the Cloud service provider. Network operators' central office or Cable Hub site can be classified under near-edge. An example of Public MEC like AWS wavelength zones is deployed at near edge sites like Telco central office or core data centres and cable Hub site.

Far-Edge is furthest away from the cloud and closer to the data source. Far-edge would include Network operator's sites deeper into the network, for example cell tower sites would fall into the far-edge category. Private MEC can be deployed at near-edge or far-edge.

2.7. Content Distribution Network

CDNs or content distribution networks are caches that are deployed closer to a user being served. CDNs are deployed at edge sites owned by caching providers like Akamai, HCP, or network operators.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

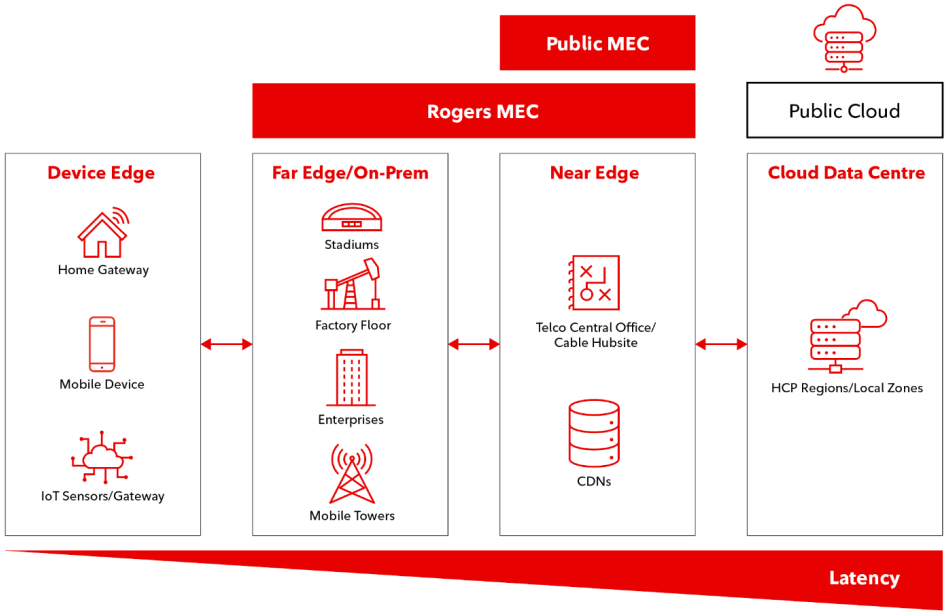


Figure 1 – Edge Compute Locations

The Figure 1 above shows possible edge compute locations and how they are categorized. MNO or Telco facilities fall under near-edge or far-edge categories and are good candidates for private or Public MEC. As applications are moved from the public cloud to the edge, latency improves. The actual edge cloud services or infrastructure can be built by the Network operator using a private cloud or in partnership with a Hyperscale cloud provider.
Roger's implementation of MEC for the tested use cases described is near-edge private MEC, and for the purpose of this paper, it will be referred to as Rogers MEC.

3. Tesbed Architecture

Figure 2 describes the generic architecture for Rogers MEC deployment common for all the use cases discussed in following sections.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

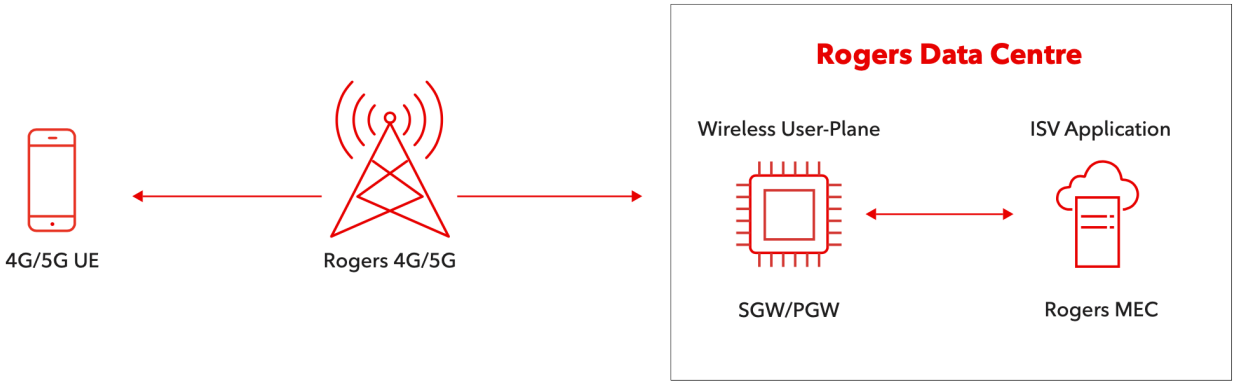


Figure 2 – Rogers MEC Infra Deployment for POCs

Rogers partnered with AWS to deploy Private MEC infrastructure collocated with wireless user plane functions (SGW/PGW) in one of Rogers's core data centres. AWS Private MEC called Rogers MEC hereafter was only hosting third-party applications and not wireless user plane network functions or VNFs. This resulted in optimizing the latency between 4G/5G UE (user equipment) and 3rd party applications hosted on the MEC server. For some use cases, a hybrid cloud model was used to deploy most latency-critical applications on MEC and more latency-tolerant workloads in the public cloud (AWS Canada central region).

5G Non-Stand Alone (NSA) mode was used in all Proof of Concepts (POCs).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 3 – MEC and Public Cloud Locations

Figure 3 shows the location of Rogers data centre and public cloud for the purpose of use cases and benchmarking tests discussed in this whitepaper.

3.1. Performance Results

3.1.1. Latency

Figure 3 shows the location of Rogers's data centre and public cloud for the purpose of use cases and benchmarking tests discussed in this whitepaper.

When comparing latency between the public cloud (AWS Canada central region) and Rogers MEC, there is a reduction in network latency by one-fourth. The impact of latency have for particular use cases has been discussed in individual sections relating to the use case

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

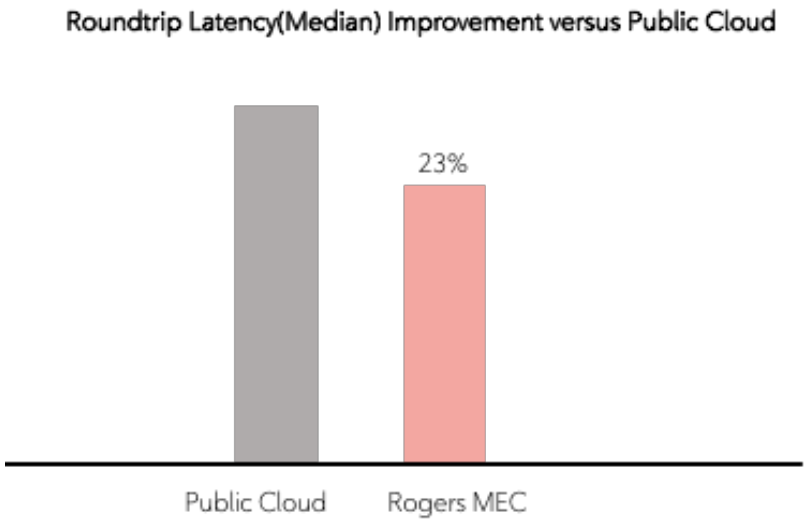


Figure 4 – Latency Improvements Public Cloud versus Rogers MEC

Note: Roundtrip latency measurements were performed by sending continuous ICMP test packets from wireless UE to server deployed in Rogers MEC and Cloud. Results may vary depending on circumstances.

3.1.2. Jitter

We saw around one tenth improvement in jitter while comparing Rogers MEC with public cloud.

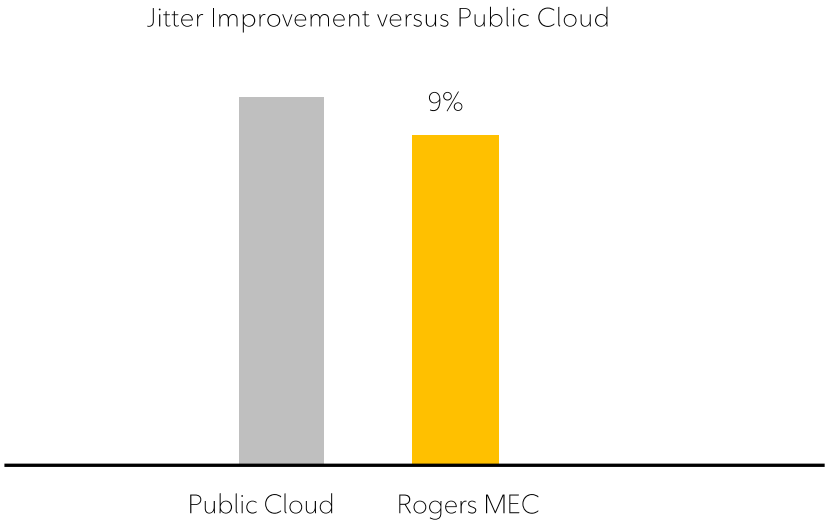


Figure 5 – Jitter Improvements Public Cloud versus Rogers MEC

Note: Jitter measurements were performed by data collected using application logs

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. AR based Sports Fan Experience

4.1. Overview

AR and VR technologies have been around for a while; however, their mass adoption has been somewhat limited due to various reasons including cost, customer experience, and battery life. Advancement of compute and cloud technologies along with more responsive transmission networks like 5G has brought a renewed interest in these technologies and made their adoption more realistic.

AR enhances a user's experience by superimposing digital assets onto what they are viewing in the physical world. With AR glasses or a smartphone, users can play games like Pokémon Go, envision how outfits looks like, or enhance fans sports viewing experience in a stadium or at home.

Rogers partnered with AWS, AR-based sports fan experience company called Immeriv.io, NHL, and MLSE to showcase the value of 5G and MEC technologies during a live NHL game at Scotiabank Arena in Toronto.

Users were able to watch live game in the arena on their smartphone phones or by wearing AR glasses with augmented features like player tracking, puck speed, and viewing player stats overlaid on the live game view. Puck and player tracking features have previously been demonstrated for in-home game-watching experience on TV but in arena fan experience during a live game is a more challenging environment.

This is because the user experience is negatively impacted if the AR view is not in sync with a real-life view of the stadium. This kind of application makes it a perfect candidate for AR content to be fetched over low latency network like 5G and hosted on MEC closer to the user being served.

4.2. Architecture

This use case was enabled by NHL puck and player tracking data collected through sensors deployed in the stadium, inside the puck, and in player jerseys.

Puck and player tracking data is generated, ingested, and stored on the premises server in the arena and also sent to cloud servers for analytics and other application consumption. To display game analytics in near real-time on the user's mobile phone or AR glass, puck and player tracking data needs to be sent from the in-stadium tracking application server to Immersiv.io tracking server on Rogers MEC and then to the AR application on the user device in low latency over Rogers 5G network.

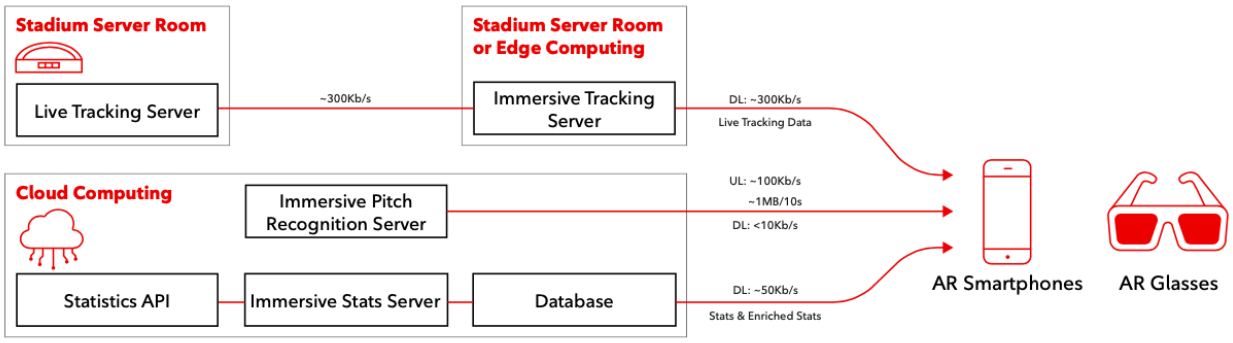


Figure 6 – Immersiv.io Application Architecture

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Immersiv.io solution has the following main software components:

Pitch Recognition Server: Pictures of the ice rink or pitch in the stadium are uploaded periodically to the pitch recognition server by smartphone in a stadium.

Stats Server: Statistics server stores the league stats in the database by querying 3rd party providers and sends it to an in-stadium user.

Tracking Server: This server ingests puck and player tracking data from the in-stadium tracking server and sends live tracking data to smartphones in real-time. This data is most latency-sensitive and so the tracking server was the only candidate to be deployed on MEC infra.

To reduce the latency of Immersiv.io live tracking server data sent to UE, we considered three different options described below.

1. Host tracking application server on MEC co-located with wireless user plane functions (SGW&PGW) in the stadium.
2. Host tracking application server on MEC deployed in Rogers wireless data centre collocated with wireless user plane functions (SGW&PGW)
3. Host tracking Application server in Hyper-scaler closest regional cloud data centre with wireless user plane function (SGW&PGW) still deployed in Roger's data centre.

Option 1 was not chosen due to the cost and effort of deploying wireless user plane functions (SGW&PGW) in the stadium outweighing the benefits of latency improvement for this particular use case.

So, we deployed a tracking server application on MEC infra using Option 2. A tracking server was also deployed (Option 3) in AWS's geographically closest regional data centre (AWS Canada central) only to benchmark MEC performance versus the cloud.

In both cases, there were application workloads (like database, stats server etc.) which did not demand strict latency and were deployed in the public cloud.

The following diagram shows the high-level architecture of the demo with wireless packet core user plane functions co-located with MEC hosting immersiv.io tracking server while other application functions are deployed in the public cloud.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

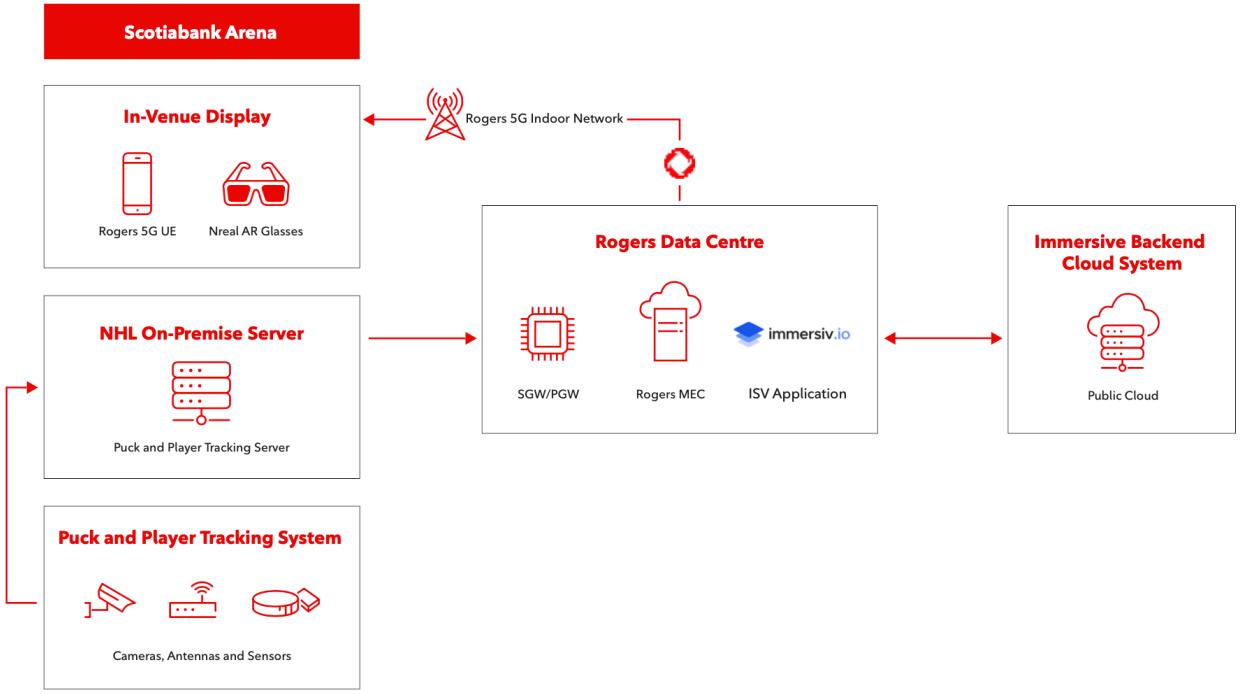


Figure 7 – AR Based Sports Fan-Experience POC Architeture

4.3. Conclusion

4.3.1. Latency

When comparing latency between the tracking server and smartphone UE in the stadium, there was a reduction in network latency by one fourth as mentioned in section3.1.1

If the latency is too high for this use case, then AR overlays, the circles around the players, would be lagging the real positions of the players. The following Figure 8 shows an example of latency impact for AR-based application if latency is too high. The left-hand side of Figure 8 shows when the puck is tracked accurately while the right-hand side shows when the puck is inside the ice rink, but AR tracking shows it outside.

This kind of impact can be experienced if the public cloud region is geographically (or fibre distance is too long) too far from the network edge or network operator data centre.

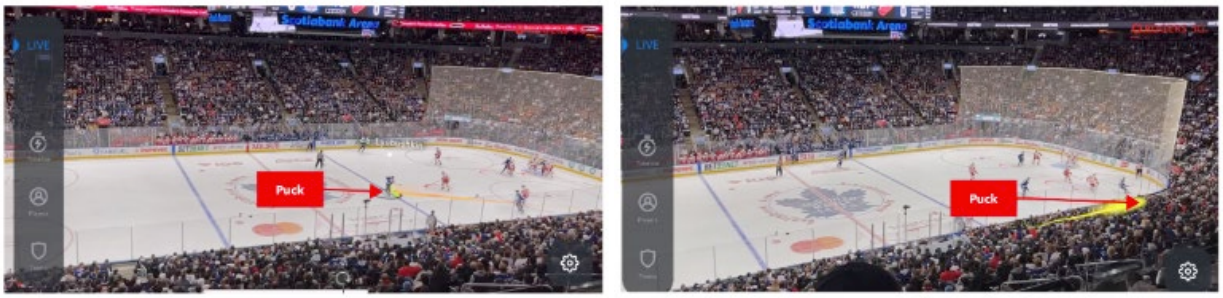


Figure 8 – Impact of Latency on User Experience

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. MEC based AR/VR Video Rendering

5.1. Introduction

Imagine being free to explore a virtual world, play games, attend events, or buy a new home as if you were actually there, without leaving the comfort of your home or office using a headset that covers your eyes. There has been a resurgence in the popularity of Augmented Reality (AR) and Virtual Reality (VR) in recent years. Thanks to improvements in processing power and cloud computing, these technologies have the potential to enable new, unique experiences for both consumers and businesses.

To begin bringing this imagination into reality, we searched for a best-in-class real-time 3D platform and collaborated with Unity to help us build a tangible VR prototype. Our next hurdle was meeting the demands of heavy Unity 3D graphics. As 3D visuals increase in quality to enhance AR and VR experiences, first, they require additional graphic processing power which, secondly, uses more battery power, and neither of which is abundantly available on headsets.

The solution was to completely move all major computational processing to the cloud. Cloud-hosted VR applications make use of the head-mounted displays in goggles for streaming ultra-realistic, high-fidelity, immersive virtual experiences instead of relying on ever-smaller headsets for rendering and processing. By doing this, only the pixels stream back to the headset, similar to a TV in traditional broadcasting. This will allow users to collaborate with others in much more immersive and extensive virtual workspaces than VR traditionally promises.

Streaming from the cloud, however, can create one hurdle: there is usually some physical and logical distance between the cloud and the headset which produces latency, which can stifle the experience. Latency is a key performance indicator in any modern viewing experience: from film actors being out of sync with their voices, news broadcasts waiting extensively before the interviewee responds, or Netflix reducing quality or buffering extensively. Everyone has knowingly or not experienced high latency, but in VR, high latency not only results in poor performance, low-fidelity graphics, and significant buffering like we traditionally expect – here, it is a recipe for nausea.

5.2. Architecture

In situations where Public Cloud is geographically very distant with high latency, a Private MEC can meet the low latency requirements of VR streaming. Rogers tested this use case, by offloading the graphic processing power from the VR headset onto a private telco MEC, using the architecture originally show in Figure 2. The entire solution added the components seen in Figure 9, and we were able to reduce the overall network roundtrip time and Jitter, as seen in Figure 4 and Figure 5 respectively, and in VR this improvement can be visually perceived by the average consumer.

3D Unity content for use inside the headset was developed separately offline and sideloaded onto the Rogers MEC server using out-of-band channels. Later, on demand, real-time rendering on the Rogers MEC streamed high-fidelity visuals over 5G back to the VR headset. Streaming with 5G minimizes RF latency while increasing bandwidth, and hosting MEC servers logically close to the client optimized the network latency. This allowed for more pixels and higher-definition graphics at a more stable rate, and thus a better VR experience overall.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

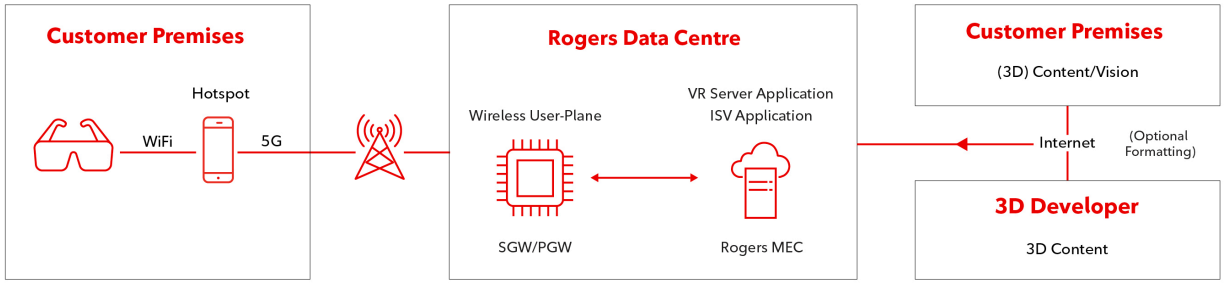


Figure 9 – MEC based AR/VR Video Rendering Architecture

5.3. Conclusion

Our architecture leveraged a VR headset connected to our Rogers MEC server via the internet, using Wi-Fi from a 5G cellular hotspot as last mile access technology. This end-to-end solution enabled us to leap over hardware limitations imposed by the headset and showcase a vivid and wireless VR streaming experience thanks to exceptionally high bandwidth, and low latency. It is reasonable conclude that the results, and VR experience overall, could be further improved by leveraging either 1) a headset with embedded SIM (eliminating the Wi-Fi hop), 2) a far-edge MEC deployment or 3) both of the aforementioned.

5.3.1. Battery

As developers realize higher fidelity graphics from on-board processors there is an intrinsic trade-off: power – including heat, weight, and capacity. Manufacturers constantly consider ergonomics, and with a VR headset balancing battery capacity with the physical weight and strain on a person's head is critical. Ironically, larger batteries produce more heat by-product and as processors compute more complex calculations they too heat up. Uncomfortably warm components inches from your face encourage users to remove their VR headset. This problem quickly grows as the battery becomes more and more consumed.

These small factors altogether increase the chances you will take the headset off sooner, but more important decrease the immersion. As an application developer that is the opposite of what you want –you want to encourage retention by having your audience use the product as much as possible and 'get lost in the experience'. VR streaming significantly reduces the processing loads on the headset allowing processors to stay cool which reduces battery drain– rivalling battery life as if the device was just simply watching movies and encouraging users to finish enjoying themselves instead abruptly stopping to charge.

5.3.2. Bandwidth

An on-premises solution and –in the right situation– Public Cloud can stream an identical VR solution over Wi-Fi, but each VR streams consumes a large bandwidth with preferred graphics. This requires a properly provisioned connection. Even still, any single access point on an average/properly provisioned Wi-Fi deployment is likely to be overloaded by multiple VR streams, and with real potential to overload the backbone IP network - especially in combination with everyday usage.

Here, 5G presents a unique solution with the opportunity to have dedicated high-capacity channels for each headset by pairing each headset with its own SIM card. For businesses with average capacity LAN networks even the slightest impact on the rest of the business can be a critical factor for adopting VR workloads.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.3.3. Remarks

It is much more enjoyable to work and collaborate remotely, play games, and explore environments with the incredible detail, lighting, scale, and lifelike physics that unrestricted GPU power provides. VR headsets have already evolved very quickly over the past decade, and as the industry matures and expands with new technology, we can expect a wireless and lifelike experience– unlocking and making practical use cases from: business to home entertainment; from indoors to outdoors; from single player to massively-multiplayer, emergencies, navigation, and more that we have yet to imagine.

6. Cross Operator Music Jam

6.1. Overview

Minimizing audio latency (or audio delay) between a group of musicians is critical for musicians to hear, play, and react to each other's performances. Musicians who play together in the same physical location, experience little to no audio latency between each other.

Musicians who have attempted to remotely perform online with widely used digital collaboration solutions such as video & audio meeting applications and audio-only communications applications have experienced challenges with synchronously performing together due to the significant audio latency within these solutions (+250 ms).

In this demonstration, three world-renowned guitarists performed a medley of rock ‘n roll selections, each sitting in separate countries through the low-latency compute environment using 5GFF’s Edge Discovery Service (EDS) API, with Open Sesame Inc. SyncStage ultra-low latency audio pipeline, with the 5G connectivity of Rogers, Verizon, and Vodafone.

SyncStage is an audio pipeline that is optimized for 5G and made available to application developers via Software Development Kit (SDK) to integrate with their application. Whenever a synchronized audio session is requested, OpenSesame’s platform infrastructure powers the audio connections and communications between a group of synchronized audio users.

Edge discovery API lets the application select the closest MEC server in a country thus making sure there is a reduction in latency and jitter between UE and application. This helps the traffic path optimizations in case there are multiple MEC locations.

This demo not only achieved multi-country, glass-to-glass latency that enables the guitarists to jam together online, but also unlocked a new reference pattern for multi-edge, multi-region applications using an interoperable EDS API.

6.2. Architecture

The OpenSesame SyncStage platform consists of three major components:

SyncStage Backend: Built on Amazon API Gateway, the Backend enables session control across all the participating devices during a session. At the MWC demo, all the devices connected to the Backend were deployed in AWS us-east-1 region. The Backend checks with the 5GFF EDS APIs and provides optimal MEC service endpoint address and connection information to the devices in real-time.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Studio Server: Deployed on MEC infrastructure, the Studio Server streams content to and from the connected devices. An instance of Studio Server is deployed at each MEC location, and the devices connect to their nearest instance via 5G network.

SyncStage SDK: A software-only SDK for application developers to integrate into their application that provides ultra-low audio latency to a group of users to enable a vast range of synchronized audio experiences.

Mobile Application using SyncStage SDK: During the demo, the musicians plugged in their instruments and monitor to a 5G connected smartphone, running the application.

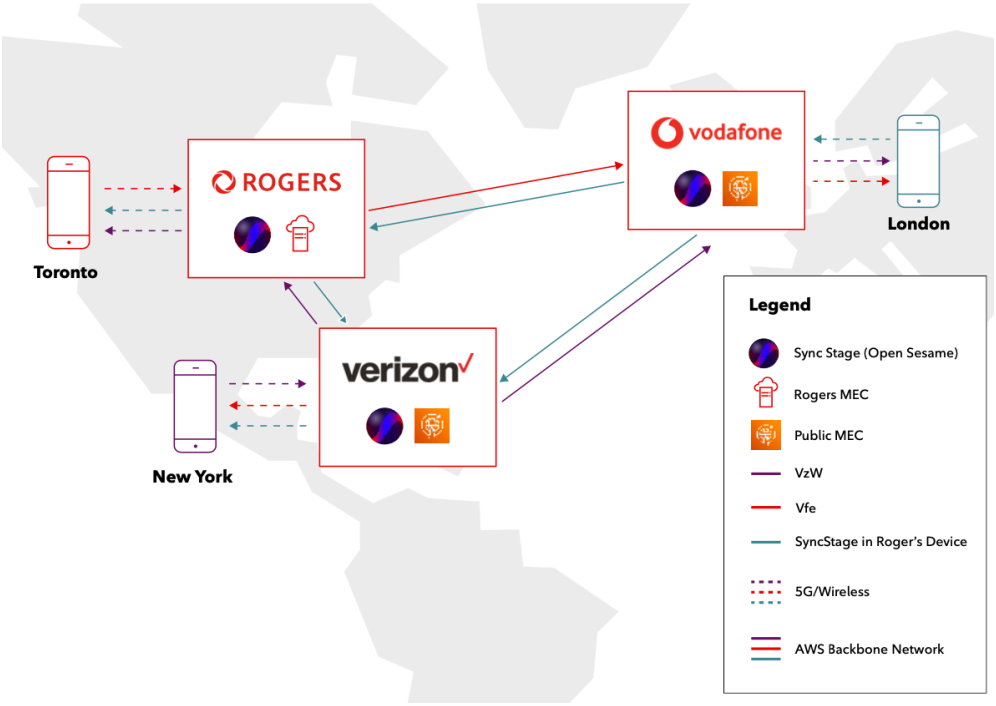


Figure 10 – Technical Architecture for 5GFF Music Festival

In New York and London, OpenSesame deployed the Studio Server instances at the Public MEC (local Wavelength Zones). In Toronto, the Studio Server was deployed on Rogers MEC, connected to the Rogers 5G network to resemble the setup in the other locations.

OpenSesame application running in 5G devices of the participating partner operators, got connected by IPv4 private network established among the three geographically distant MEC nodes through the AWS backbone network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

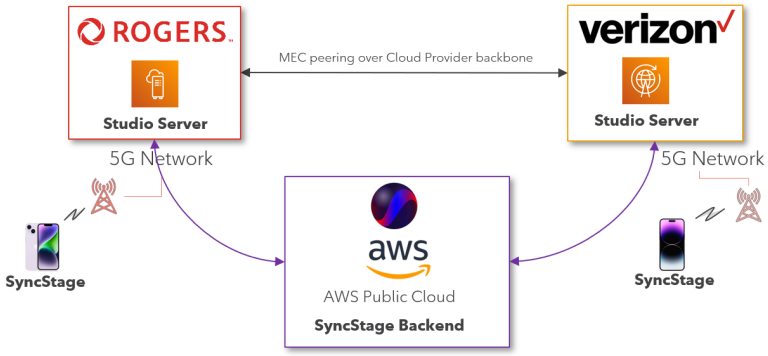


Figure 11 – Edge Discovery Service API deployment

When SyncStage mobile application initiates a call, on the control plane side, it queries to SyncStage Backend for optimal MEC service endpoint in a multi-MEC deployed network. The Backend checks with 5GFF Edge Discover Service (EDS) APIs and provides optimal MEC service endpoint address and connection information to the devices in real-time.

Once the optimal MEC endpoint is known, SyncStage application in the originating device requests a user-plane connection to the selected MEC endpoint in the home network that in turn establishes VPN tunnel with the partner MEC endpoint by using a private IP address and connects to SyncStage application in the target device.

While testing the demo setup, we observed musician-to-musician network latency. Some of this latency is attributed to the physical distance that the audio needs to travel to reach each performer. We experienced lower latency values between Toronto and New York, attributed to the shorter physical distance between these two cities. The longer distances between North America and the United Kingdom incurred greater latency. However, these latencies were neither noticeable to the live performers, nor was it evident in the session recording.

6.3. Conclusion

- MEC makes a good sense for the latency, jitter, and packet loss sensitive audio applications while keeping the traffic local in the MEC Zone where Local Zone and/or Region is geographically away. It measured close to one-fourth roundtrip latency reduction in Rogers' test environment setup by using Rogers MEC vs. public cloud for musician in Toronto. This test result is very much aligned with the findings presented in sub-section 3.1 Performance Results.

- Point-to-point physical direct connection between multi-operators MEC nodes would be beneficial to support latency, jitter, and packet loss sensitive use-cases, given there are use-cases to compensate for such cost. The estimated latency reduction in this specific application and use-case is 11-21% over the current MEC to MEC peering connection. This could be beneficial especially for intra-continental MEC deployment, favoured by distance and the law of physics.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

7. Last Mile Food Delivery using Autonomous Robot

7.1. Overview

Rogers' objective has been to showcase the benefits of 5G network along with the capabilities of MEC. Rogers and OVIN (Ontario Vehicle Innovation Network) launched a challenge for SMEs (small to medium Enterprises) to come up with innovative 5G use cases in smart transportation and CAV (connected and autonomous) vehicle sector leveraging technologies like MEC.

LoopX, a start-up specializing in vehicle autonomy was selected to demonstrate the idea of last-mile food delivery using autonomous robot in Rogers' 5G powered smart campus testbed.

LoopX's last-mile food delivery solution uses electric-powered robots providing efficient delivery of food in an autonomous, contactless, and low carbon footprint manner. These delivery robots are designed to operate on sidewalks or bike lanes.

Delivery robots are equipped with LIDAR sensors, ultrasonic sensors, cameras for perception, GNSS receivers and IMU sensors for localization. Each robot is equipped with two onboard NVIDIA GPUs computers to host LoopX robot controller autonomy software.

Local sensors generate perception data on the robot to feed into its autonomy control software hosted on-board computer and GPU.

While there is a lot of buzz around autonomous or self-driving vehicles but, fully autonomous solutions would take time to mature given the regulatory and ethical aspects associated. In today's world, hybrid driving techniques involving a mix of autonomy and human driving (teleoperated and autonomous) are more practical and will continue to play an important role in the near future. we discuss MEC-based cooperative operation of multiple robots in this paper using Hybrid operation (autonomous and teleoperation with teleoperator driver intervention when required).

7.2. Architecture

For this use case, the MEC server in Rogers' data centre was hosting LoopX video session manager software and Mapping engine.

The mapping engine hosted on MEC plays a critical role in the collaborative operation of delivery robots sharing the same route. In case of an obstruction en route, robots can dynamically update the map on MEC enabling other robots to download an updated map and avoid routes blocked by an obstruction. Without a centralized mapping engine, Teleoperator intervention is needed to reroute the robot after detecting obstruction on the sidewalk or road. It is important to note that maps used for the autonomous operation of vehicles are not the same maps used for turn-by-turn navigation in use by human drivers. Autonomous vehicle operation needs HD maps which are more precise and specifically built for the safe operation of autonomous vehicles. Those maps need to be constantly updated.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

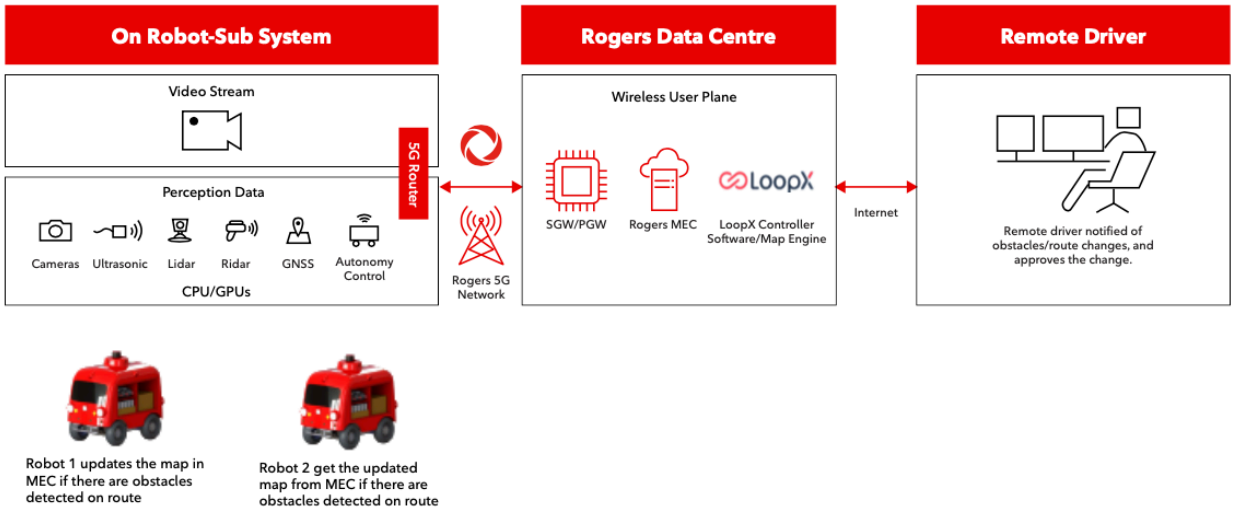


Figure 12 – Last Mile Food Delivery Robot POC Architecture

The other function which MEC software performed in this case was brokering the video session between Teleoperator and robots equipped with HD cameras before video streaming began. This improved the response time to start video streaming in comparison to controlling traffic using the public internet. Also, by hosting the video control on MEC, the overall security of the solution is improved.

Robots were communicating over Rogers's 5G network to application functions hosted on the MEC server in Roger's data centre. In this particular POC, the teleoperator driver was using wireline access over the internet to connect to MEC. The teleoperator driver ideally would connect to the MEC server over a dedicated wireline connection so that end-to-end control traffic does not cross the internet and the connection is private. This would enhance the overall security of the solution and provide more predictable service quality.

7.3. Conclusion

By hosting the controller software and mapping engine on the MEC server, the overall responsiveness and security of the solution was improved.

Due to project time constraints, the team was only able to offload only certain functions from onboard computers to MEC.

This particular use case can be further expanded to offload compute-intensive and more latency tolerant (comparing latency with on-prem versus MEC) workloads from Robots to MEC potentially extending battery life on robots and reducing cost for the compute and GPU resources.

It is worth noting that 5GAA has published various use cases related to autonomous vehicles area with corresponding SLR (service level requirements) which can be candidates for MEC. However, most of those use cases are uplink heavy and at scale, deployment of those use cases is a bit challenging at this point.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

8. Overall Conclusion

As transport latency is a function of geographic and fibre distance between source and destination endpoints, latency benefits of operator MEC are largely defined by how close or far are public cloud provider regions and data centres are located to the user endpoint. However, traffic traversing over best-effort internet from Network operators to cloud provider data centre hosting applications results in higher jitter values not suitable for some applications although geographic distances are not that far.

Routing optimization and direct peering between Network operators and hyper-scale cloud providers also improve the overall latency for workload hosted in cloud data centres minimizing the benefits of MEC infra hosted relatively closer to public cloud provider data centres.

Regardless of the above factors, today's applications are built based on microservices architectural principles and can be deployed in a highly distributed way. Applications workloads which are latency or jitter-sensitive and bandwidth-heavy can be deployed at the network edge while large-scale workloads not so sensitive to latency can be deployed at hyper-scale cloud data centres.

The adoption of operator MEC would also largely depend upon the economic feasibility of moving latency-sensitive device edge workloads to the network edge.

There are a lot of opportunities to offload compute-intensive workloads like video rendering, video analytics and computer vision from the device edge or on-prem edge to the operator MEC or network edge. As most of these use cases are upstream-heavy, so the savings in compute cost should offset the network costs to build capacity and transport data to the network edge at increased volume and higher peak rates.

Uplink transport costs and peak capacity will become lesser of an issue with increased 5G spectrum availability and costs of data transmission going down.

It would also be important to note that low latency mission-critical use cases like remote surgeries and C-V2X need ultra-reliable network and latency reduction itself is not enough to host applications at the network edge.

Another area which can help with the wider adoption of MEC and make it commercially viable for Network operators is video caching at the edge using Network operator MEC infrastructure. Network operators are already hosting caching solutions provided by Akamai and other caching providers at their facilities.

Adoption of MEC is at a slow pace due to various reasons like the development of low latency use cases, commercial viability, lack of wider upstream spectrum, reliability of network (for mission-critical use case) and lack of 5G device ecosystem.

However, as some of those challenges ease with time. MEC has the potential to grow as a vertical for Network operators to be a source of additional revenue.

There are areas which need to be developed and need to be looked into further for MEC commercialization. For example, just-in-time workload orchestration, optimization of compute and GPU resources, scaling and supporting MEC for ISVs and facilities upgrades. Some of those are being addressed through ETSI, 5GFF and other forums and topics of further studies.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AWS	Amazon Web Services
AR	Augmented Reality
C-V2X	Cellular vehicle to anything
CAV	Connected and Autonomous Vehicles
DOCSIS	Data over cable system Interface specification
EDS	Edge Discovery Service
ETSI	European Telecommunications Standard Institute
GNSS	Global Navigational Support System
GPU	Graphical processing unit
HCP	Hyper-scale cloud provider
ISV	Independent software vendor
MEC	Multi-access Edge computing
MWC	Mobile World Congress
OVIN	Ontario Vehicle Innovation Network
PGW	Packet Gateway
POC	Proof of concept
SME	Small to medium Enterprises
SGW	Serving Gateway
VNF	Virtual Network Function
VR	Virtual Reality
5G	5th Generation
5GAA	5G Automotive Association
5GFF	5G Future Forum
SCTE	Society of Cable Telecommunications Engineers

Bibliography & References

1. https://www.etsi.org/technologies/multi-access-edge-computing

2. https://www.smt.com/articles2019/puck-and-player-tracking-now-part-of-nhl-data-and-technology-offering

3. https://5gff.org/work-weve-done/?filter=whitepaper

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Coherence After Turbulence:

Coherent Links for Hurricane Relief in the Sunshine State

A Technical Paper prepared for SCTE by

Venk Mutalik
Fellow
Comcast
1800 Arch Street, Philadelphia, PA 19103
+1 (860) 262-4479
Venk_Mutalik@Comcast.com

Amarildo Vieira, Principal Optical Engineer

Dan Rice, Vice President

Eric Moore, Engineer

John Chrostowski, Executive Director

Scott Wood, Engineer

Elad Nafshi, Chief Network Officer

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 3
2. Background: Comcast Architectures 3
   2.1. Distributed Access Architecture (DAA): 3
   2.2. Continuous and Pervasive Monitoring (CPM) of Optical Assets 4
3. Coherent Optical Links in the Access Domain 6
   3.1. Dual Laser Single Fiber Bi-Directional (BiDi) Coherent Links in Access Networks 6
   3.1. Converging the Coherence 9
4. Ground Zero for Hurricane Ian 10
   4.1. After Hurricane Ian 10
5. Solving the Supply Chain: Equipment for Florida 13
   5.1. Redesigning Analog Links to DAA 13
   5.2. Generators to the Rescue 14
   5.3. A Word on Optical Connectivity and Monitoring 15
6. Conclusions 16
7. Acknowledgements 16

Abbreviations 17

Bibliography & References 17

List of Figures

Title **Page Number**

Figure 1 – Illustrating the DAA Network 4
Figure 2 – Illustrating Continuous Pervasive Monitoring 5
Figure 3 – Illustrating Continuous Pervasive Monitoring Coverage 5
Figure 4 – Illustrating Bi-Directional Coherent Systems 7
Figure 5 – Illustrating Coherent Muxponder solution for DAA 8
Figure 6 – Illustrating the Converged Access Network 9
Figure 7 – Convergence of Services, Optics and Tooling 10
Figure 8 – Before and After Pictures of the Sanibel Causeway [6] 11
Figure 9 – Fiber Availability on the Causeway and from Pine Island 12
Figure 10 – Extension Site on Sanibel Island 12
Figure 11 – Pedestal Node Location – Post-Hurricane 14
Figure 12 – Illustrating Monitoring of the Coherent Mux Ponder Solution 15

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Travel sites are filled with glowing reviews of Sanibel and Captiva Islands on the West Coast of Florida, just south of Fort Myers. Home to the famed J.N 'Ding' Darling National Wildlife Refuge, Sanibel Island is connected to the mainland via the Sanibel Causeway, its lifeline to the mainland [1].

So powerful was Hurricane Ian that when it made landfall in late September of 2022 the winds and storm surge that followed collapsed the Causeway. With it, communication lines between Sanibel Island and the mainland were also cut. Of course, widespread destruction on the Island also destroyed much of the infrastructure. Most heartbreaking of all, the human toll approached 150 lives, with nearly half of those in Lee County alone [2].

With sustained winds approaching 155 mph, the winds and storm surge caused by Hurricane Ian caused parts of Florida to be without power for weeks. Cleanup took even longer. During this time, it was all hands-on deck at Comcast, as Comcast focused on restoring connectivity to locations impacted by the storm. This paper is a case study of our efforts to provide connectivity to Sanibel Island in the aftermath of Hurricane Ian.

After initial rescue efforts, power restoration, and cleanup were in full swing, it became clear that connectivity to the region was just as important as well. With respect to Sanibel Island, an initial survey of fiber infrastructure showed limited accessibility from the mainland to the island, but that the secondary site on the island could be brought up with backup power. To conserve limited fiber and available critical infrastructure, Coherent optics were used to light up the trunk link and bring up the nodes one by one with backup power, thus, restoring connectivity to affected customers as quickly as possible. This paper will focus on innovative technologies deployed and the great efforts undertaken by Comcast teams to provide connectivity to this part of Florida.

2. Background: Comcast Architectures

Providing capacity while also enhancing reliability is at the heart of innovation at Comcast. With two years of Covid behind us at the time, Comcast had already begun implementation of the Distributed Access Architecture (DAA) and fine-tuned many aspects of deployment that aided in providing unprecedented insight into our network health in real time.

2.1. Distributed Access Architecture (DAA):

In previous years [3], we have reported on how Comcast has pioneered the DAA that has virtualized CMTS functionality and separated the PHY layer from the CMTS and distributed it out into nodes in the field. These are now called Remote PHY Device (RPD) nodes.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

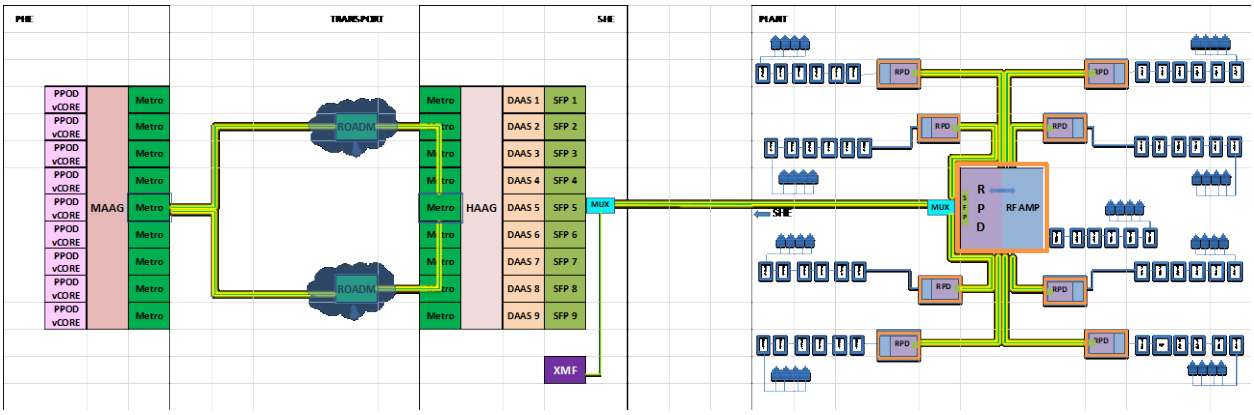


Figure 1 – Illustrating the DAA Network

Shown above is a simplified version of the DAA system. While the Primary to Secondary connections are based on 100G or higher capacity Coherent optics, the connections from the DAAS to the DAA nodes (RPD nodes) in the access domain, are always 10Gbps DWDM-based signals. vCMTS (virtual Cable Modem Termination System) cores from the Primary headend are connected to DAA switches (DAAS) via Highly Aggregated Switches (HAAGs) in the Secondaries, typically via Comcast's own Metro networks. Once out of the DAAS, both the US and DS 10G wavelengths are multiplexed on a single strand of fiber using a novel passive with consolidated test ports and sent over into the field, where they are demultiplexed at an Outside Plant (OSP) enclosure location. From there, dual strands of fiber carry the 10Gbps signals over to the various RPDs in the immediate location. The distance from the Secondary to the field OSP mux is called the trunk fiber, while the fiber from the OSP Mux location to the various RPDs is called the distribution fiber. On the trunk fiber, Comcast has a 100GHz ITU WL (wavelength) Plan standard that has adjacent odd and even pairs of ITU channels serving each node beginning at ITU 61/60 pair to ITU 15/14 pair. Thus, all in all, 24 pairs of wavelengths are available on each trunk fiber capable of serving 24 different remote devices. While the DAA program is on pace to upgrade all traditional nodes and CMTSs to the vCMTS based digital nodes in the next few years. In Sanibel Island, when the hurricane hit, however, all the nodes were still analog nodes served via a Hub in South Fort Myers hub.

2.2. Continuous and Pervasive Monitoring (CPM) of Optical Assets

Our current network base is a mix of single and dual fiber CWDM (Coarse Wavelength Division Multiplexing) and DWDM (Dense Wavelength Division Multiplexing) connections. Many of our current business customers that require fiber to their premises require 1G to 10G services and could be in the vicinity of existing OSP Mux enclosures. For a go-forward plan, since these are digital signals, these customers can be served by the same strand of trunk fiber that would connect analog and digital nodes. This is an important way to relieve capacity on our fiber assets by enabling more efficient use. In addition, it enables a common way to monitor optical assets continuously and pervasively, which is described in a subsequent paragraph.

Figure 2 illustrates a converged system that is a combination of residential and commercial services, incorporating Coherent and direct detect systems, all capable of running on a single fiber. Since all services can be multiplexed on a single fiber, the current series of optical passives in Comcast has consolidated test ports that provide access to view forward and return wavelengths of live links. Furthermore, the test ports also allow unrestricted 1611 nm (nano meters) access for the use of Optical Time Domain Reflectometers (OTDRs) to scan live links and provide fiber impairment and cut information.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

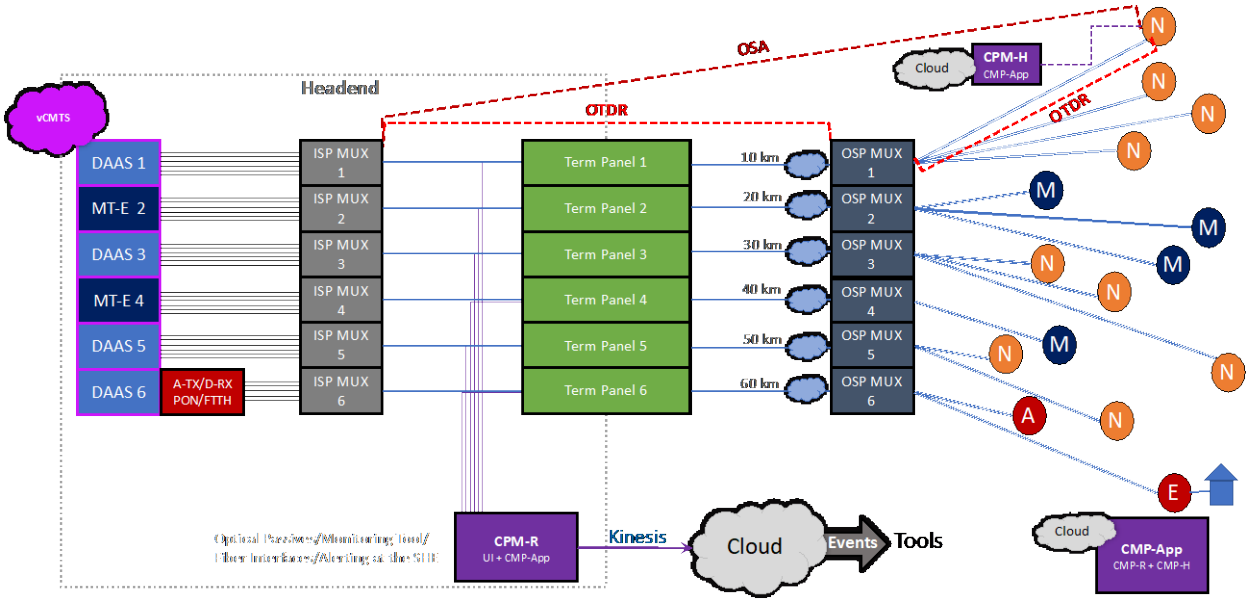


Figure 2 – Illustrating Continuous Pervasive Monitoring

The continuous and pervasive monitoring paradigm in Comcast was described in earlier SCTE papers [4]. By way of summary, all optical passives are connected to a continuous monitor comprising an OSA and OTDR as well as an optical switch. This arrangement continuously monitors connected fibers in a round-robin fashion and generates fault information comprising either fiber impairments or individual wavelength impairments on average within 90 seconds. The entire information set is sent into the cloud for storage to aid in longer term analysis, but alarms and events observed are sent to fix agents in real time.

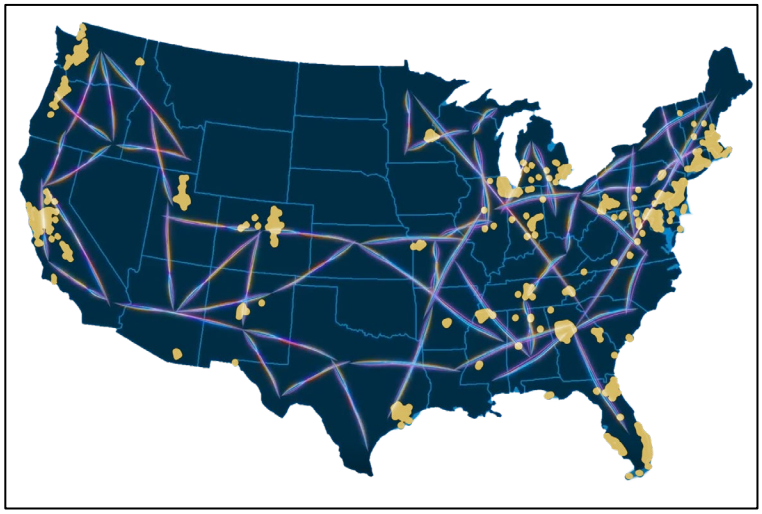


Figure 3 – Illustrating Continuous Pervasive Monitoring Coverage

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

To promote a common language between headend and field technicians across the Comcast footprint, a version of the monitor is available as a handheld device that has the same cloud connectivity as the headend version. Connecting the headend unit, the hand-held unit, or both to the cloud allows the entire region, division, and company to view all events in real time and resolve important issues impacting customers.

3. Coherent Optical Links in the Access Domain

Not only is fiber expensive to deploy, but it also takes a long time to do so. One way to be efficient about fiber usage is to be able to use fibers bi-directionally. Not only is this efficient, but also it promotes resiliency as a fiber cut can be repaired by just one splice repair. This is easier to do with 10G optics and indeed the entire DAA footprint is designed with bi-directional links. But in cases where fiber is limited and higher capacity is needed, commonly used Coherent Optics, which can provide 100Gbps or more of capacity, usually require a dual fiber option. We describe in this section, a single fiber Coherent Optical link option that enables the use of high capacity (100Gpbs or more) and works well with an active 10Gbps fiber plant. Such an option needs a new set of dual laser bi-directional Coherent pluggable transceivers and a Coherent Mux-ponder (CMP) terminating device [5]. We describe these in more detail in this section.

3.1. Dual Laser Single Fiber Bi-Directional (BiDi) Coherent Links in Access Networks

Coherent optics rely on having a laser on the transmitting end to send out phase and amplitude information in the form of QAM (Quadrature Amplitude Modulation) constellations to a receiver. At the receiver, the incoming signal is ‘beat’ up against a laser that has, in effect, the exact wavelength of the transmitting laser and it teases out the phase and amplitude information. Since information is coded in phase and amplitude as opposed to only amplitude in direct detect systems, the information carrying capacity of the Coherent system is vastly superior to as that of direct detect systems for the same amount of bandwidth available.

To reduce cost and to ensure that the wavelengths are identical, most of the pluggable transceivers available in the market use just one laser in each plug and split its light to do double duty, in effect to act as the encoding transmitters and the decoding receiver. This is illustrated in the top half of Figure 4 below.

The industry has gotten quite good at this type of design and has successfully used it to provide 100s of Tbps (Tera bits per second) over 1000s of km in core and metro networks. Unfortunately, this approach always requires two separate fibers, one for downstream and another for upstream information. When a fiber can be fully loaded with data, say for 100s of Tbps, this approach seems reasonable. But when the fibers are unable to be fully loaded, like what happens in access networks, an approach of this kind is very inefficient and can result in severe fiber exhaustion.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

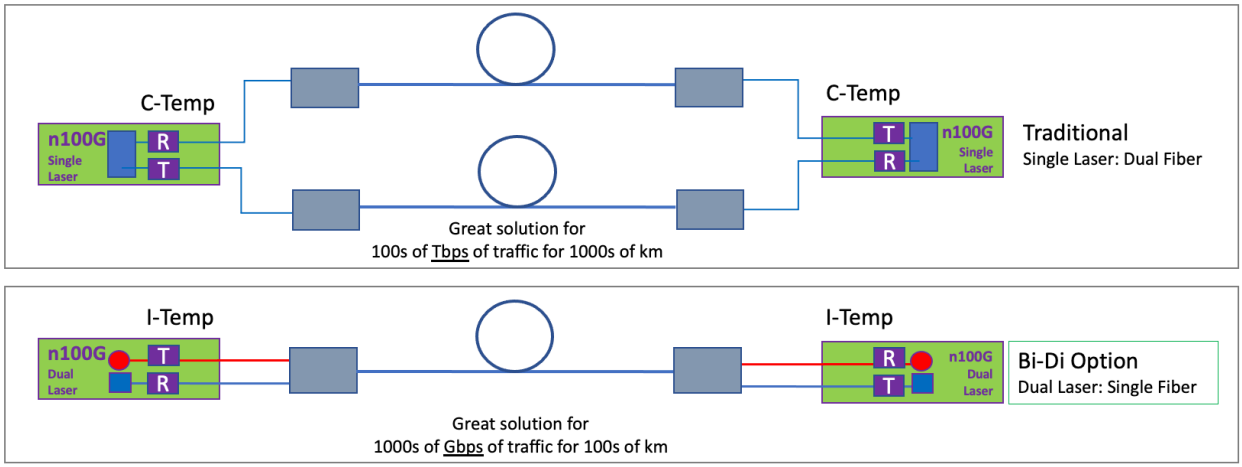


Figure 4 – Illustrating Bi-Directional Coherent Systems

Therefore, a novel approach to accommodate Coherent optics in the access domain needed to be developed. This approach enabled both high-capacity wavelengths that feed our residential DAA systems and support our business customers as standalone high-capacity wavelengths as well.

While the use of optical circulators is a valuable option and could potentially allow the same wavelength to traverse in opposite directions, that approach is prone to fiber plant reflections and Rayleigh backscattering which could limit the link budget. A more robust solution is needed for wide bi-directional deployments that envision the use of existing fibers with other bi-directional signals already running on them.

A robust solution to this problem could be to have a separate laser for the transmitter and the receiver at each end. The dual laser approach enables the two wavelengths to be multiplexed (or combined) together and transmit bidirectionally (Bi-Di) on the same single strand of fiber. In fact, many Coherent optical wavelengths could be multiplexed on the same fiber along with direct detect wavelengths, leading to a converged system described in the previous section.

Interestingly, low earth orbit (LEO) satellite communications systems have also come up with the same requirement of dual laser bi-directional transmission. In free space, Coherent optical modules communicate with adjacent satellites and hop signals from satellite to satellite to minimize the need for ground stations. But the vast speeds of the LEO satellites and their relative speeds to other satellites create the familiar doppler shift in frequency sufficient to thwart the use of a single laser design. So, in this effort of defining a dual laser plug, we have had the unusual conjoining of space requirements as well.

With this in mind, Comcast has specified dual laser, bidirectional, fully C-Band tunable, industrial temperature C Form-factor Pluggable 2 (100Gbps) (CFP2) plugs for use in access networks, much aligned with the CableLabs point-to-point Coherent specification [4]. Current systems support 100Gbps rates with Quadrature Phase Shift Keying (QPSK) modulation at 32GBaud with a 25dB link budget, enough to support point-to-point links up to 100kms. Due to their limited reach requirements, there are no ROADMs (Reconfigurable Optical Add Drop Multiplexers) in the access plant and there are no non-reciprocal elements in the system, this allows for the relaxation of certain specifications and needs for chromatic dispersion compensation that enable these new innovative devices to be cost-effective as well.

While 100G Coherent systems can exist on their own and support individual wavelength services for our commercial customers in the converged access architecture described in the previous section, a

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

particularly important application of the dual laser BiDi system is to support DAA systems. Often these DAA systems have to operate out of smaller extension sites or secondary cabinets that make direct metro transport system connections difficult due to limited critical infrastructure. Adding additional fiber to the trunk line is an expensive and time-consuming effort. In such cases, we have developed a system that takes in 10 of the 10Gbps streams and electronically converts it to a 100Gbps stream that is modulated in the CFP2 which is then multiplexed to a single fiber and sent over to the secondary. Such a device is called the Coherent Mux Ponder (CMP) and two such devices are bookended one each at the Primary - after the DAAS for aggregating the 10Gbps streams to 100Gbps - and the Secondary extension site – for disaggregating the 100Gbps streams into its constituent 10Gbps streams. Once the 10Gbps streams are recovered, the CMP uses standard full band tunable 10Gbps SFPs and optical passives to serve DAA nodes out in the field. It is as if the DAAS has now been extended in a compact form factor into a much smaller extension side enclosure.

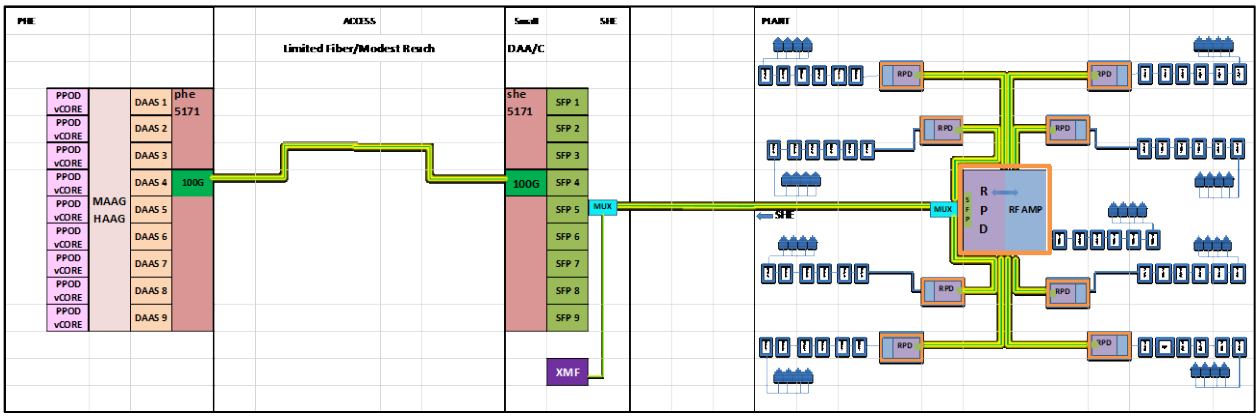


Figure 5 – Illustrating Coherent Muxponder solution for DAA

The CMP system designed can support up to 40 of the 10Gbps streams in 2RU of space and is shallow enough to be placed in tight extension sites. The entire system can be scaled up to serve up to 240 of the 10Gbps services all on a single strand of fiber. Converged Optical Architecture in the Access Domain

Section 2 provided a simple description of Comcast's network. Although the path towards digitization and distributed access is evident, we still need to address multiple legacy architectures, including analog DS wavelengths and potentially analog or digital US wavelengths, during the DAA transition. In addition, one interesting aspect of the access plant is that most of the businesses that require separate optical wavelengths are also interspersed among residential customers. Techniques to intersperse analog and digital WLs on the same fiber has been described in earlier papers and not repeated here.

Converged architectures are an important way to relieve capacity on our fiber assets by enabling their more effective use. In addition, it enables a common way to continuously and pervasively monitor optical assets which is described in a subsequent section.

But what is interesting in the Converged Access Network in Figure 6 is that we have now been able to incorporate Coherent optics in the access plant in much the same way as direct detect has been deployed. This type of convergence requires the innovative technology of bidirectional dual laser pluggable optics used at Comcast in recent years and is described in this section.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

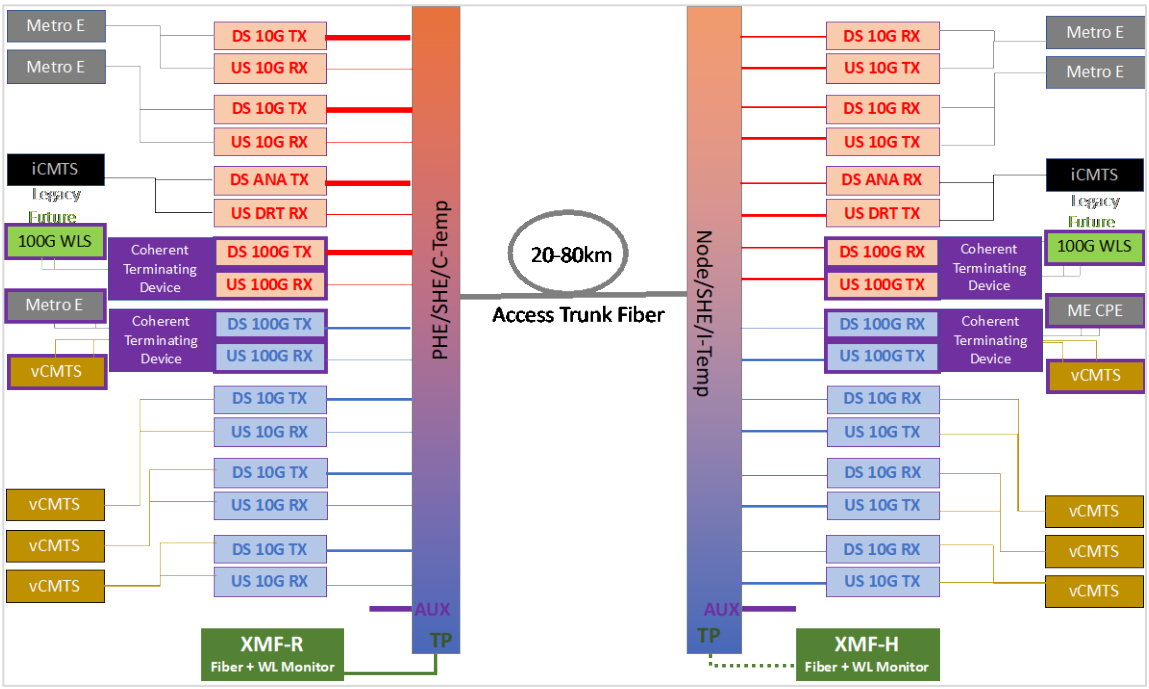


Figure 6 – Illustrating the Converged Access Network

Since ROADM cascades are not used in our DAA access networks due to modest fiber lengths, the design of high-capacity (100G – 800G) solutions could be elegant as there is no additional degradation due to these devices. Furthermore, these links can support bidirectional transmission quite easily since there are no non-reciprocal elements in the fiber path. Indeed, the relaxed relatively modest fiber reach required in DAA enables us to innovate on optical devices as well as reduce cost, oftentimes by innovating on the otherwise saturated edge of technology.

3.1. Converging the Coherence

While the CMP is a muxponder and intentionally does not participate in any switching or stat muxing regimes, it is an additional active element in the system. As such, it is a perfect vehicle to converge multiple services such as 10Gbps metro-e services, 5G cell back hauls and the DAA services which also use 10Gbps line rates not only on one fiber, but also on the same wavelength. Additionally, even 100Gbps services could be offered to those that needed it. Such convergence enables commonality of infrastructure and makes the network strong and elastic.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

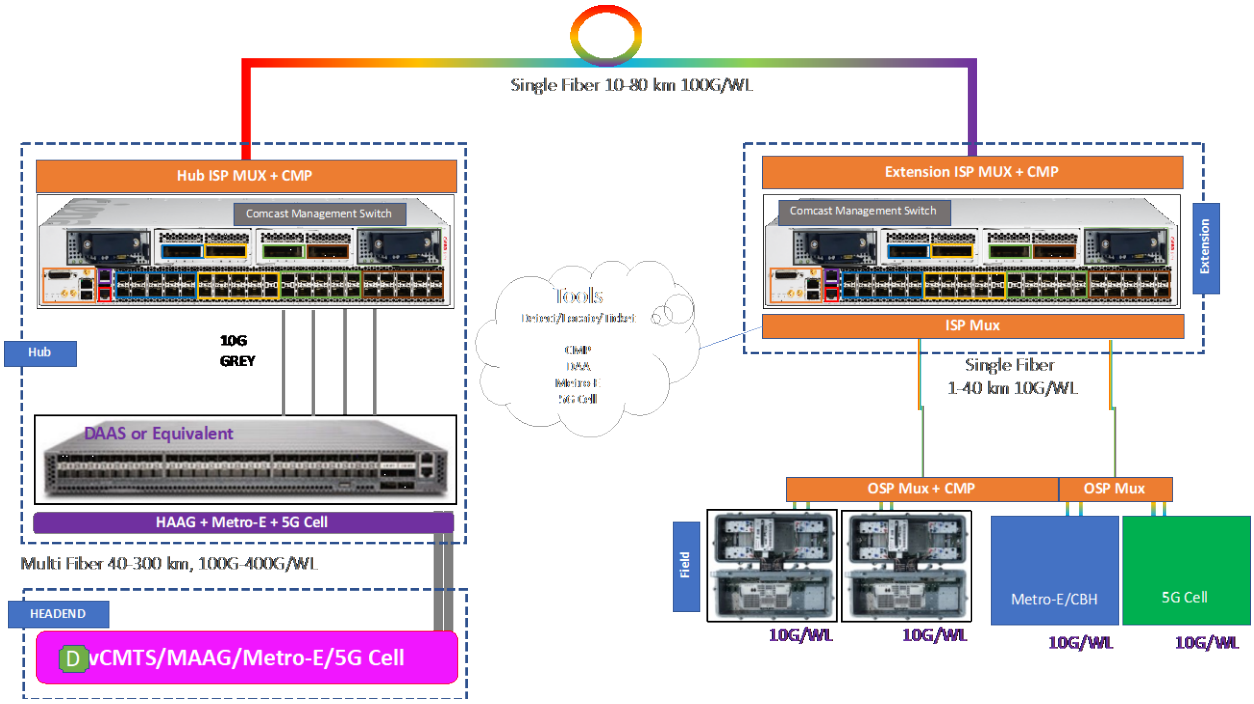


Figure 7 – Convergence of Services, Optics and Tooling

4. Ground Zero for Hurricane Ian

As Hurricane Ian barrelled towards the country, Comcast implemented its storm-mode protocol. Under this protocol, Comcast's Network Operations Center (NOC) postpones non-critical scheduled work on applicable equipment infrastructure. Postponing non-critical scheduled maintenance activities decreases other noise that may appear on the NOC's (Network Operating Center) monitoring dashboards. The Storm Mode keeps critical components available to help support the unknown stress the storm will unleash. Accordingly, key support teams can focus on the remediation process necessary to restore services to our customers once the impacted area has been deemed safe to work in. Storm Preparation mode concludes by identifying the areas affected by the storm, assessing the safest areas to start restoring services to and prioritizing the equipment restoration to allow for the best impact for our customers.

4.1. After Hurricane Ian

In the aftermath of Hurricane Ian, restoring power and communications was a top priority of Sanibel Island. At Comcast, this was an 'all hands-on deck' moment. Comcast personnel had to be ferried to and from the island every day to perform any work on the island. Comcast responded by placing Wi-Fi hotspots on trucks in major landmarks so folks could come in get connections and possibly charge their phones. Comcast is the only non-wireless internet service provider for Sanibel Island. Getting services restored to municipal facilities was critical for search and rescue services as well as providing additional resources for those stranded on the island.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 8 – Before and After Pictures of the Sanibel Causeway [6]

The existing Sanibel Island facility was a cabinet that housed a DC power plant, some optoelectronics and fiber patch panels. The 6-foot storm surge completely flooded the cabinet and ruined everything inside. Before the cabinet could be re-used, it would have to undergo a thorough cleaning and have all electrical components replaced.

In the meantime, a check on fiber connections showed that the main route of fiber from South Ft. Myers along the Causeway was cut (see above figure) and there was no secondary route available. Through local partnerships, Comcast secured limited fiber between the mainland and a location near the Pine Island hub. Comcast was then able to restore a damaged fiber path that could be used to restore services to Pine Island and eventually to Sanibel Island.

The issue now was to supply signals to the Sanibel Extension from a new site and make all the nodes operational. At this time, all of Sanibel Island was supplied by conventional analog optics, which had limited reach and was based in the South Ft. Myers hub. Accordingly, fiber availability from Pine Island did nothing to improve connectivity due to the long distances from Pine Island to Sanibel Island and the limited reach of analog optics. It was therefore decided to move the entire Sanibel Island footprint to the DAA to supply connectivity. A more detailed description of DAA has already been given and it is good to recall that DAA pre-supposes 2 10G wavelengths of DWDM light for each node lit up.

A detailed survey of Sanibel from design maps showed us that there would be up to 57 fiber nodes in Sanibel Island serving over 11,000 passings. We also found that critical sections of the fiber path had single strands of fiber and also that the extension site in Sanibel -would not support the form factors of DAA leaf and spine access switches as well as routers and the transport gear needed (see Figure 10).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

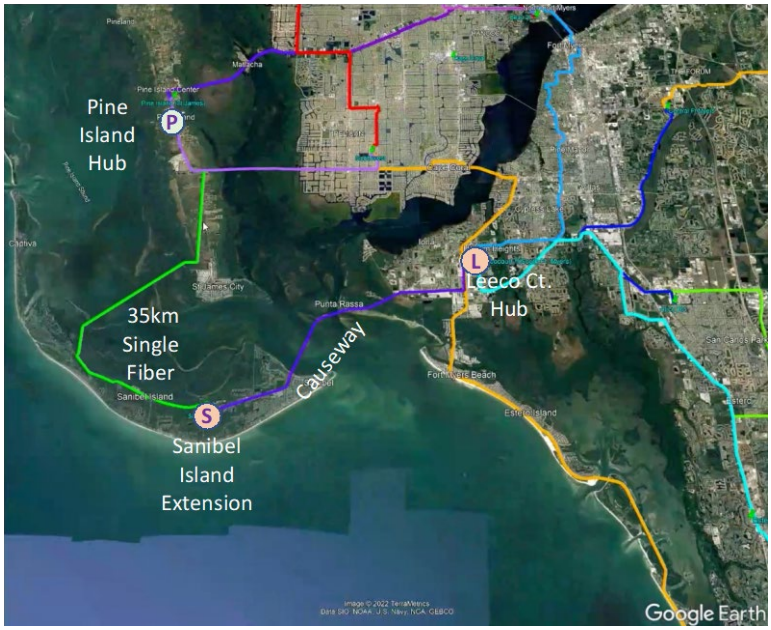


Figure 9 – Fiber Availability on the Causeway and from Pine Island



Figure 10 – Extension Site on Sanibel Island

It was at this point that division leadership reached out to the National organization with a request for technology that would enable lighting up to 120 digital nodes as a practical maximum on one single strand of fiber, and also ideally having a low profile so as to fit in extension cabinets with minimal critical infrastructure requirements. This extremely compact and critical-infrastructure-conserving solution would also ideally serve 10G Metro E customers and 5G cell towers on the same strand.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Coherent mux Ponder (CMP solution fitted this bill and furthermore, Central and West Division Leadership had seen this setup in action during the SCTE Show at the Comcast Technology Center a month before.

At this point, our Leadership decided to support this request and participate in Hurricane Relief as a priority and shipped out the needed CMPs, the 100Gbps BiDi transmitters, the tunable and grey SFPs and optical passives necessary along with management switches and pervasive optical monitoring equipment and shipped these out as well to Central. With the necessary preparations complete, it was time to proceed with the task of organizing and installing the equipment, as well as undertaking the meticulous but crucial process of activating digital nodes throughout the entire island.

5. Solving the Supply Chain: Equipment for Florida

While Covid-19 lockdowns had been lifted in many parts of the world by 2022, we continued to face its impact for a significant portion of the year. Additionally, global supply chain disruptions persisted, leading to ongoing delays in procuring materials and equipment across various industries, even at present. Despite these challenges, we managed to secure an adequate amount of equipment, including spare CMPs and BiDi CFP2 modules for deployment in West and Central divisions.

Our first step was to stage the solution to make sure everything was properly configured and set before the equipment was sent out to the deployment areas. The staging consisted of uploading the latest approved software, the required switch licenses, the approved switch configurations for the local and remote sides and setting up IP addresses (these could be changed by the local teams). Once this was completed, the available CFP2 modules were inserted in the switches and the 100G link availability was verified. After that, the approved gray optics SFP+s (small formfactor pluggables) were inserted in the local and remote CMPs. Using a 10Gbps Viavi T-BERD connected to these SFP+s, we were able to pass error-free traffic for a short period of time (30 min) to validate the systems. The next step was to pack and ship the systems into the region and start the deployment phase.

That was the time when the Storm Mode lifted, and we moved to recovery mode. Since the equipment was already staged, configured, and tested, the only local changes to be made were providing console port and management port connectivity. Such configuration could serve up to 40 optical nodes bundles very quickly. As the RPDs in the optical nodes began to successfully light up and were provisioned, it was the beginning steps towards bringing back connectivity to the Island.

5.1. Redesigning Analog Links to DAA

To provide immediate connectivity, Comcast opened Wi-Fi access points for the public use for free, across the island thus helping people find critical resources like food, water, and social services.

Our first step was to light up the Sanibel extension site with the CMP, and have it connected to the Pine Island hub (also called the St. James Hub). The St. James Hub CMPs were all connected to DAAS at that location, while the Sanibel Extension site CMP was made ready with all the optics to connect out to individual DAA nodes. But before the DAA nodes started lighting up, as mentioned once before, since the existing nodes were all analog nodes, Comcast's Planning and Design department had to re-design all the Sanibel nodes as DAA nodes.

The Sanibel extension site was powered up and made functional even while the redesign process of designing DAA nodes in place of the existing analog nodes was being carried out. However, connecting said nodes was an intricate process. The following picture shows what the remediation team saw when they arrived at where a node was supposed to be on a pedestal.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 11 – Pedestal Node Location – Post-Hurricane

First the links from the individual node locations to the Sanibel extension site had to be restored. Once that was done, the node was then to be set up and activated and could be put into a full production state with only the addition of a fiber jumper into the Coherent Muxponder Chassis. The Coherent Muxponder allowed for everything at the source site up to the extension site Coherent Muxponder Chassis to be connected in advance of the node turnup. Due to the elegance and simplicity of this platform, each node was ready to be configured and set up extremely quickly. Each pair of CFP2 optics enabled the team to activate 10 additional nodes. Eventually, all nodes were restored much sooner than if a legacy technology had been used.

5.2. Generators to the Rescue

A major issue however was that a majority of the island was without power, and so, activating nodes was not really possible even if connectivity with the Sanibel extension site was established. But Comcast had deployed many portable generators that were used to power Wi-Fi hotspots.

By enabling Wi-Fi in the affected areas, emergency restoration crews were able to establish a communication network. This network played a crucial role in aiding and assisting in a more strategic remediation plan, as well as providing progress reporting to facilitate effective response and recovery efforts. For the island's residents, it provided a comforting sense of connection to friends, family, and the outside world for status and news updates [7].

Comcast team followed that playbook and deployed many portable generators to power connected nodes and went about ensuring that they were up and running even before the power was restored so that when

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

power did come back, the nodes were ready for operation. Each DAA node was brought up, activated, verified and then brought down and connected to line power when power began to get restored.

5.3. A Word on Optical Connectivity and Monitoring

When Comcast initially verified fiber connectivity between Sanibel and Pine Islands, a 1x48 DWDM mux was installed on a single fiber on each end of the link. These muxes were used for the Coherent optics connecting the CMP chassis. Once the link connectivity and back-office connectivity were established node activations described above could begin [8]. At the Sanibel side, multiple Muxes were used to connect to clusters of DAA nodes as determined by the optical design team. All of these Muxes were monitored by pervasive monitoring technology within Comcast and real time data of the fiber network was available to technicians in the field. The importance of real time fiber monitoring especially in times of disaster recovery cannot be overstated. If fiber connectivity is well established and known in real time, other barriers to connectivity - such as power outages or equipment connections - can be more easily isolated thus ensuring a more stable restoration environment.

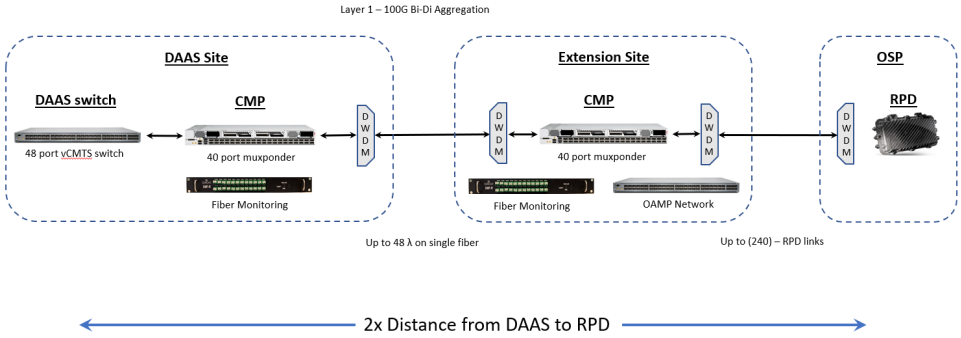


Figure 12 – Illustrating Monitoring of the Coherent Mux Ponder Solution

Thus, fiber assets across Pine Island to Sanibel Island were located, Coherent Optics then spanned the across these assets and lit up Sanibel Island providing great capacity and reach. Services were then converted from analog to DAA nodes, and the nodes lit up using portable generators while awaiting power restoration. Once power was restored, the nodes were converted to line power and connectivity was restored all in record time.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Conclusions

In this paper, we have provided a glimpse into the extraordinary efforts of Comcast personnel, from line technicians to senior leadership, to implement recovery efforts on Sanibel Island after Hurricane Ian. While this exceptional effort highlights disaster recovery, such efforts display the norm for technology advancements with a tight loop of learning and feedback that helps pass information from the field to divisions to leadership.

Coherent optics are an essential part of access technology, especially in fiber-starved or extension sites where space is limited. At the heart of the innovation is a dual laser single fiber pluggable transceiver design that transforms every access link into a high-capacity and highly reliable converged link Moreover, this development opens up possibilities for our converged access architecture, allowing for the integration of services like residential, commercial, and 5G onto a single unified platform.

7. Acknowledgements

It is not often that work done on a technical level find use so quickly in the field and at such a critical juncture so as to support our customers in their time of need.

Our thanks go to the division leadership for their innovativeness, hard work, and timely authorization of Coherent link deployment. We acknowledge all our teammates and vendors that worked with us on this project.

We end this paper again with our thoughts on the Hurricane Ian victims and our thanks to the first responders. Our thanks are also due to our own Comcast personnel that enabled connectivity with grit and grace as soon as they did in the affected areas.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

Gbps	GigaBits per second
BERT	Bit Error Rate Test
BiDi	Bi-Directional
C Band	Conventional Band (1530 to 1565 nm)
CMP	Coherent Mux Ponder
CFP2	C Form-factor Pluggable 2 (100Gbps)
CWDM	Coarse Wavelength Division Multiplexing
DAAS	Distributed Access Architecture Switch
DAA	Distributed Access Architecture
DS	Down Stream
DWDM	Dense Wavelength Division Multiplexing
Hz	Hertz
ISP	Inside Plant
LEO	Low Earth Orbit
nm	Nano meter
NOC	Network operating Center
OSP	Outside Plant
OTDR	Optical Time Domain Reflectometry
QAM	Quadrature Amplitude Modulation
QSFP	Quad Small Form-factor Pluggable transceiver (40, 100Gbps)
QPSK	Quadrature Phase Shift Keying
RFC 2544	Benchmarking Methodology for Network Interconnect Devices
ROADMs	Reconfigurable optical Add Drop Multiplexing
SCTE	Society of Cable Telecommunications Engineers
SFP+	Small Form Factor Pluggable +
SCTE	Society of Cable Telecommunications Engineers
SMF	Single Mode Fiber
Tbps	Terabits per second
US	Up Stream
WL	Wavelength
vCMTS	Virtual Cable Modem Termination System

Bibliography & References

[1] https://www.nationalgeographic.com/travel/article/sanibel-island-florida-hurricane-ian#:~:text=Everything%20changed%20when%20the%20hurricane,car%20traffic%20to%20the%20island.

[2] https://www.tampabay.com/news/environment/2022/12/22/hurricane-ian-florida-landfall-sanibel-island-causeway-lee-county/

[3] Accelerating the Virtualization: Introducing Hybrid Fiber Shelf into the Mix, Venk Mutalik et. al., SCTE Expo 2020

[4] It's 10PM, do you know where your wavelengths are? Venk Mutalik et. al., SCTE Expo 2020

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

[5] Photon Avatars in the Comcast Cosmos, Venk Mutalik et., al., SCTE Expo 2022
[6] https://abcnews.go.com/US/hurricane-ian-images-show-destruction-sanibel-island-fort/story?id=90919034
[7] https://corporate.comcast.com/press/releases/comcast-free-xfinity-wifi-hotspot-network-hurricane-ian
[8] https://santivachronicle.com/news/comcast-service-restored-well-ahead-of-schedule-on-sanibel/
[9] Fiber QAM vs. RF QAM, Amarildo Viera et. al., SCTE Expo 2023

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Connection Manager for Automotive- Enabling a Shift in Connectivity to Support the Evolution of the Car

A Technical Paper prepared for SCTE by

Ken Mathias
VP, Tech Strategy and Planning
Charter Communications
Ken.mathias@charter.com

Will Logan
VP, Wireless Device Engineering
Charter Communications
will.logan@charter.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Digitization of the Automotive Industry: An Imperative for OEMs ... 3
3. Software Revolution of the Vehicle ... 3
4. Automotive Connectivity of the Future ... 5
5. Rethinking Automotive Connectivity – A Complimentary MSO Broadband Solution ... 7
6. Technical Overview – Connection Manager Switching Between Cellular, WiFi and Beyond (Satellite) ... 8
7. Integration Points ... 10
8. Conclusion ... 12

Bibliography & References ... 12

List of Figures

Title **Page Number**

Figure 1: SAE Levels (Source: SAE, 2021, May 3) ... 3
Figure 2: Hardware vs Software Defined Architecture[6] ... 4
Figure 3: Comparing ICE and EV Average Component Price[5, 9] ... 5
Figure 4: Data Transport Needs[6] ... 6
Figure 5: WiFi vs Cellular Experience Comparison[16, 17] ... 8
Figure 6: Connection Manager High Level Design ... 9
Figure 7: Connection Manager Components ... 10
Figure 8: Connection Manager Integration Points ... 11

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The automotive industry has been on a transformational journey of digitization since the early 2000s. With the recent introduction of Advanced Driver Assistance Systems (ADAS) and the growth of electric vehicles (EVs), software defined architecture for vehicles is demonstrating expanding connectivity needs. This whitepaper explores the evolution of the automotive architecture towards a software-based model driven by vehicle electrification and the need for a digital chassis. It also discusses how an application like Connection Manager can solve the future automotive connectivity needs by enabling on-device-based switching, connecting to MSO networks to help accelerate software-defined vehicle architecture.

2. Digitization of the Automotive Industry: An Imperative for OEMs

The automotive industry has undergone a remarkable transformation since the early 2000s with the introduction of ADAS. "Connected Vehicles" are now ubiquitous among new cars. Historically, connected vehicles, although existing since the mid-1990s with providers like OnStar, were primarily focused on basic low data and voice connections for emergency and concierge services. However, the rapid growth of electric vehicles (EVs), along with the increasing prevalence of ADAS features requiring sophisticated software, has driven the industry towards digitization.

According to Juniper Research, the number of connected vehicles in service will reach 367 million globally in 2027, up from 192 million in 2023[1]. The rapid growth of almost 91% between 2023 and 2027 will be driven primarily by ADAS and infotainment.[2] By 2025 in the US, over 50% of cars sold are projected to have L1 driving automation, 35% will have L2 technology, 9% with L3 features and just 1% with L4/L5 technology[3]. Figure 1 below goes into further detail of ADAS and levels of driving automation.

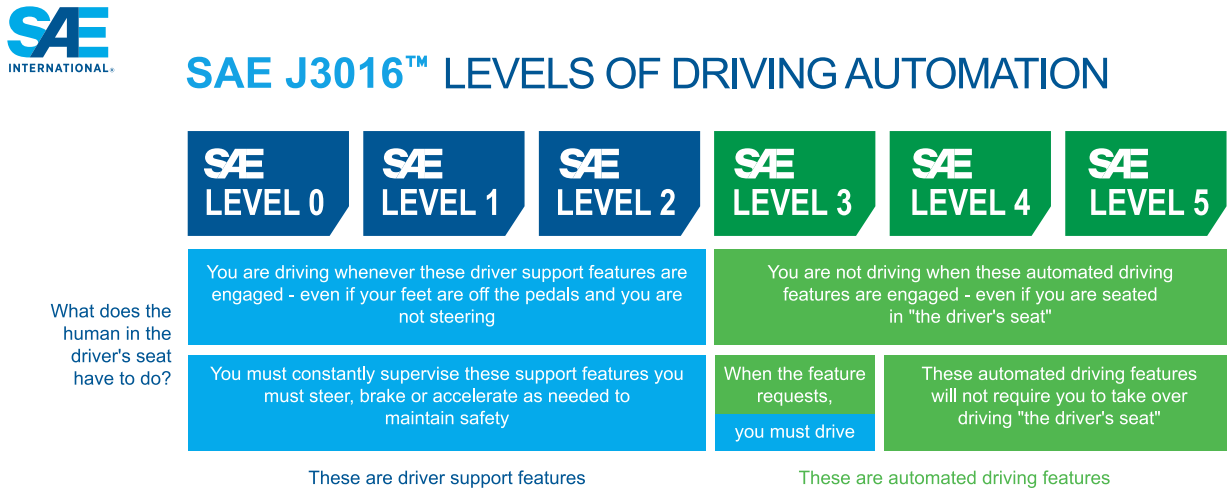


Figure 1: SAE Levels (Source: SAE, 2021, May 3)

3. Software Revolution of the Vehicle

So why is there an acceleration in digitization and driving automation now? For one, the electrification of the vehicle and its sudden and rapid rise in popularity.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The transportation industry is seeing several factors driving this electrification spike, but chief among them is the push from various regulatory jurisdictions to meet new ESG requirements. Most large regulatory bodies within the US and EU have put into place emissions restrictions on new vehicle sales. For reference, the US has a stated goal that half of all new passenger cars and light trucks sold will be zero-emission vehicles by 2030 while the EU will require all new cars sold in the EU to be zero-emission vehicles starting in 2035[4]. The industry is also seeing a major historical hurdle to EV adoption – cost – begin to erode. In 2023, the average internal combustion engine (ICE) vehicle in the US was $48k (up from $38k in 2020), allowing the cost gap to close compared to $53k for an EV[5] and resulting in EVs being price accessible. In conjunction with the price gap closing, regulatory bodies are continuing to subsidize EVs (Federal $7,500, e.g., Colorado $5,000), sometimes making the effective EV price lower than ICE vehicles, further accelerating demand of consumers' vehicle preferences towards EVs and accelerating Original Equipment Manufacturers (auto 'OEMs') shift toward EVs.

This electrification of automobiles and shift to autonomy threatens to halve the demand for vehicles worldwide over the next 20 years, due to increasing lifespan of EVs (more discussion follows on lifespan). This means OEMs must earn (higher margin) digital revenues if they are to maintain profitability. If they fail to drive usage of the infotainment unit, which assumes the ability to unlock digital revenue through application-based content and consumption, they will be unable to capitalize on the opportunity and it is anticipated that additional participants will come in over the top.[6]

To capture infotainment usage, OEMs need to abandon their legacy platforms and start again, providing an opportunity to redesign the relationship between hardware and software in the vehicle. The functionality of most vehicles today is defined by the hardware that is installed at the point of manufacturing. This means the specification of the vehicle is crystalized at the time of manufacture and there is limited opportunity for the OEM to make upgrades or changes once the vehicle has been manufactured. This new software-defined architecture will enable software to be updated remotely enabling new features and functionality as well as a full penetration of digital services into a model that generates digital revenue which OEMs need to survive.[6]

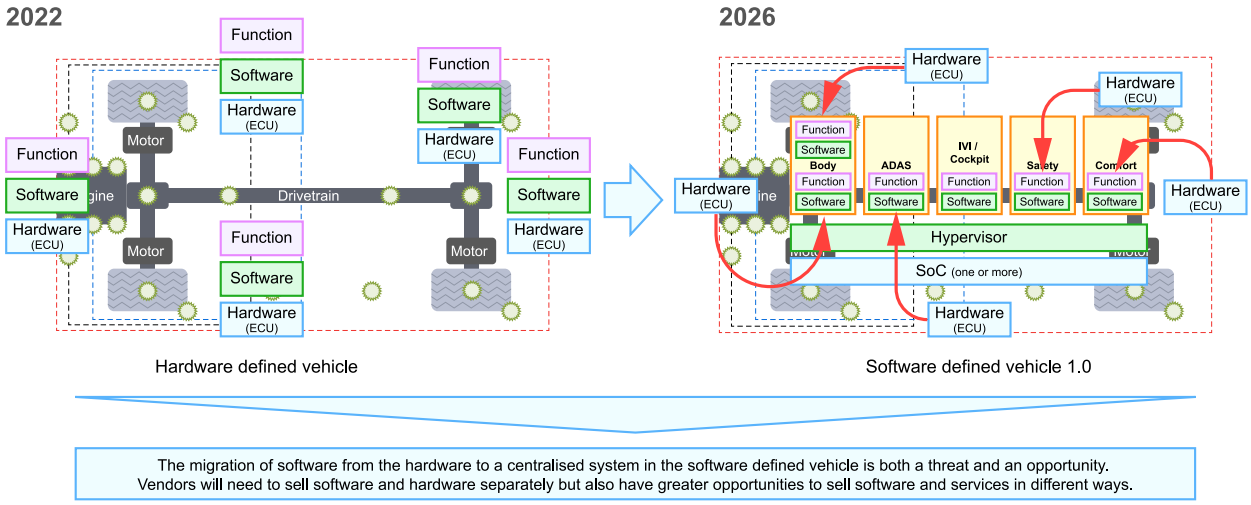


Figure 2: Hardware vs Software Defined Architecture[6]

With EVs, the R&D and component investment paradigm has flipped towards software vs. physical components.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

With the increased investment in software integrated components, the average cost of vehicle-based technology for EVs has increased compared to ICE vehicles. Figure 3 shows a high-level breakdown of this component level cost change where investment capacity allows a shift to technology-based components, including software. OEMs are also shifting their approach to lifecycle management as this cost paradigm shifts. An industry that was originally based on planned obsolescence is transforming to one that places increased importance on the maintenance of software and with less components, will release resource capacity to reinvest in software[7]. Furthermore, the average lifespan of an ICE vehicle is ~14.8 years compared to the EV, which could be ~22.2 years[8]. Since there are no mass market EVs that are 22+ years old, the implied life of an EV is likely much longer. With this increase in the life of an EV as compared to that of an ICE vehicle, OEMs revenue model must change to capture service-based revenue through software services, further necessitating a shift towards software models.

Average Price / Component - 2023			
	ICE Price		**EV Price**
Avg Vehicle Components	$48,000.00		$53,438.00
30,000 (ICE)	$1.60		(For comparison) $1.78
20,000 (EV)	(For comparison) $2.53		$3.56
Applying ICE Component Cost to EV			
Vehicle Components	Cost per Component	Price	SW Investment Capacity
20,000	$1.60	53,438.00	$10,562
Software Investment (& Technology) Capacity represents current margin headroom to invest in software development, notwithstanding EV component cost increases			

Figure 3: Comparing ICE and EV Average Component Price[5, 9]

As vehicles' lifespan increases and the focus shifts from hardware to software, automotive manufacturers are redefining their software strategies. Vehicle functionality is being digitized and, where possible, cost is being reduced. The maintenance of connected digital vehicles now requires continuous updates, similar to OS & application software patches and improvements on smartphones and PCs.

The advent of software-defined architecture has become crucial to the automotive industry and vehicles using this architecture will use large amounts of data, requiring significant connectivity capacity to be effective.[6]

4. Automotive Connectivity of the Future

The software revolution of the vehicle is something that OEMs must quickly master, but that is not the only race for competitive parity. The competitive advantage of the future will be moving up the curve on autonomous driving as the winners of the multi-hundred-billion-dollar ($700B+[10]) US automotive market will be the OEMs that demonstrate real autonomous L4/L5 driving. The pre-seeding of sensors with software upgradable vehicles is happening currently in Tesla and other auto manufacturers, and it is anticipated that capturing this data is needed to evolve autonomous driving algorithms and will result in exponentially growing data needs.

There will be two main pools of data that software-defined vehicles will generate. The first is digital data from digital experiences in the vehicle (digital cockpit and infotainment domain) and the other is the data

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

generated by the driver assist and autonomous systems of the vehicle (ADAS domain). They are different in both size as well as time dependency and should therefore be assessed separately.[6]

The digital cockpit domain is where infotainment and the instrument cluster reside and is where all the data driving the digital experience in the vehicle will be generated. This could be smaller in total than generated by the ADAS domain (see below), but it will be time sensitive meaning that a good proportion of it will need to be delivered real-time via a cellular connection. Other aspects such as software updates (e.g., OS updates or app updates like Zoom in select 2024 Mercedes models) will be less time sensitive and could be delivered via WiFi when parked or charging without degrading the user experience or increasing reliance on cellular networks.[6]

The ADAS domain is where all driver assistance and autonomous features reside and where data from cameras, lidars and radars will be collected and processed. As L1 through L5 driving automation is developed and deployed in vehicles, the need for continuous and reliable processing, storage, and transport of data will become paramount. This will lead to very large volumes of data in its raw form, however the bulk will undergo processing in the vehicle in order to enable the ADAS functions to work. For example, raw video feeds will be processed into semantic scene descriptions (what the camera sees and where the objects are), which will be far smaller in size than the raw video feeds from which they were derived but still represent substantial amounts of data. How much of this data is uploaded will depend on how the machine vision algorithms in the cloud are trained, regardless, it is anticipated that every training scenario will improve the autonomous driving system and will result in significant amounts of data consumption if uploaded to the cloud. This training data will not be time dependent which, combined with its size and volume, makes it very likely that it will need to be uploaded over WiFi when the vehicle is parked and/or charging.[6]

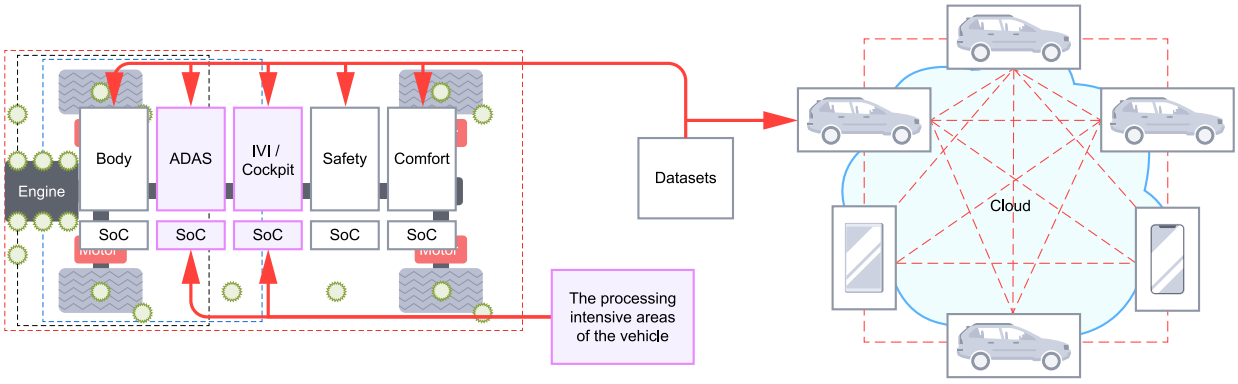


Figure 4: Data Transport Needs[6]

For reference on potential data payload sizes, Level 1 vehicles today can process up to three gigabytes per hour of data. As sensors and functionality increase in vehicles, they could process up to fifty gigabytes per hour[12]. Taking autonomous driving pioneer Waymo as a proxy for the L5 vehicle of tomorrow, a car will use up to eight cameras, three lidars and five radars. It is not unrealistic to think a fully autonomous vehicle could generate a continuous stream of data of up to 18TB per hour in complex driving conditions (e.g., snowy conditions)[12].

At some point after 2026, autonomous vehicles are likely to reach commercial grade and be deployed into the marketplace. It is at this point that data captured by vehicle sensors is likely to be uploaded far more

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

frequently and in much greater volumes. Up until this point, ADAS system will be uploading and downloading data, but these will be mostly HD map and software system updates[6].

While it is unlikely that these payloads will be transferred back to the OEM or other cloud provider for processing due to cellular capacity and wireless wholesale costs, it is anticipated that corresponding data logs for this information will grow, and possibly be stranded on the vehicle[11]. 'Stranded data' translates to lost value from the digitization of the vehicle, not just for evolving active ADAS functionality, but for the future continuous improvement cycle inherent in good software development and next generation safety solutions.

Cellular connectivity alone has supported the connected vehicle to date; however, to support the connectivity needs of the future, such as software updates and ADAS training logs, OEMs will need to leverage high throughput, low-cost solutions. MSO broadband networks, including WiFi, are deployed across the majority of US households and can easily support automobiles of the future. And with wireline capacity being upgraded across all MSOs, connectivity can be offered at a price and throughput now and into the future that still allows OEMs to efficiently manage the ongoing costs with sufficient throughput for the needed software maintenance of the vehicle.

5. Rethinking Automotive Connectivity – A Complimentary MSO Broadband Solution

The increasing data needs of vehicles will require complimentary connectivity beyond just existing cellular networks. As ADAS features become more prevalent and vehicle autonomy increases, the data consumption will surge. With the high data rates and large volume of vehicle (250M+ vehicles in the US, similar in scale to smartphones of ~300M+), cellular networks require complimentary connectivity, like WiFi, which has been utilized by smartphones for years, and also CBRS. This use case will demonstrate the need for MSO solutions utilizing unlicensed spectrum such as WiFi and sharing scenarios like CBRS to offer a high throughput and cost-effective vehicle connectivity.

Looking to proven high throughput connectivity solutions in conjunction with mobile data might unlock the data growth needed to move the automotive industry forward. Broadband internet, WiFi and planned CBRS MSO networks are the missing element that automotive manufacturers should consider. In 2022 alone, the US and Canada had a total internet bandwidth of ~200 Tbps[13] with the average subscriber using ~587 GB per month[14] in their home. And WiFi has already demonstrated itself as a complimentary use case for smartphones as we have seen that MSO supported WiFi makes up a high majority of data traffic on smartphones today. Total data traffic on smartphones in 2022 was ~136 GB per month, of which, ~116 GB per month, or 85%, was over WiFi[15]. The smartphone of today and the automobile of tomorrow could end up looking very similar, and MSO Broadband supported through WiFi are poised to support the future growth of data consumption for cars as they do today for smartphones.

As autonomous vehicles are expected to generate a massive amount of data, car manufacturers will need to continuously increase bandwidths for both point-to-point data pipes and distributed network structures to meet these new data demands. Given the challenges in achieving Level 4 and 5 autonomy, complementary high-capacity networks such as broadband, WiFi and CBRS networks will play a pivotal role in enabling seamless data transfer of ADAS training logs, improving autonomous driving software models.

To meet customer expectations for a future software maintained vehicle full of applications, and digital experiences, automotive connectivity needs to be rethought. The seamless and error-proof maintenance of vehicles and continuously enhanced applications, including ADAS and L4/L5 autonomy, require

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

additional capacity beyond current "mobile" connectivity solutions. Broadband and WiFi networks will play a complementary role, supporting high-throughput scenarios and satisfying data-intensive applications, while delivering the right customer experience.





Connectivity Solution	WiFi	Cellular
Avg. Speed	1 Gbps	40 Mbps
Transfer Time – 10GB Payload	1 min, 20 sec	33 min, 20 sec

Figure 5: WiFi vs Cellular Experience Comparison[16, 17]

A comprehensive approach, leveraging a combination of solutions, is necessary to address the evolving data needs of the automotive industry. However, auto OEMs will need to contend with a disparate set of mobile and broadband/WiFi networks, and Connection Manager emerges as a crucial bridge between these connectivity solutions, enabling smart on-device switching to optimize traffic flow. By managing multiple connectivity solutions with Connection Manager, automakers can drive a substantial portion of traffic to WiFi, ensuring an improved customer experience over a cellular connection.

6. Technical Overview – Connection Manager Switching Between Cellular, WiFi and Beyond (Satellite)

Connection Manager is a device-based connection application that can automatically and intelligently, authenticate broadband/WiFi and CBRS networks based on pre-configured thresholds and scenarios. Its primary function is to use logic to connect to the right connectivity solution based on the needs. Utilizing a Connection Manager on-device client and cloud-based solution will enable intelligent data offload with improved quality of service (QoS), ensuring least cost routing for high throughput, delay tolerant workloads (e.g., OTA updates, ADAS training log transfers).

Connection manager has been utilized successfully by MSOs in the US as an Android mobile software client that drives network connectivity decisions for WiFi and other cellular network (primarily MSO owned CBRS spectrum) connectivity. It helps to enable auto-authentication to configurable purpose-built WiFi and CBRS networks, supported by MSO broadband networks, while also capturing wireless connection usage data to improve QoS on a continuous basis.

Using MSO mobile offload as a use case, one can start to project how Connection Manager could drive an impact for the automotive industry, accelerating L4/L5 ADAS development by offloading stranded data logs through lower cost connections and higher throughput of broadband networks. Utilizing this mobile model, OEMs should consider on-device software connectivity solutions like Connection Manager and support it with hardware architecture to utilize complimentary MSOs networks.

The industry is already seeing the hardware architecture for the automotive industry evolve towards this multi-connectivity solution reality. Qualcomm's Snapdragon Automotive Platform now supports a dual SIM dual active use case. This means two cellular (or other) connections can be utilized simultaneously and movement across subscriptions can happen without interruption. This platform also has WiFi 6 dual band support, which helps to demonstrate how chipset manufactures are viewing the future connectivity and bandwidth needs of the vehicle[18]. In this new hardware reality, one could envision the vehicle of the future connecting to a 5G MNO cellular network for emergency or mapping services, a MSO CBRS

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

network for standard data streaming and offload while in motion, and, when stopped, a nearby WiFi access point for a low latency gaming or entertainment session. And as the vehicle begins to move again, these connections can be evaluated in real time by Connection Manager to connect to the most appropriate network option available for the use case and scenario, all while being seamless to the customer.

Connection Manager Server: Hosted (e.g., AWS) server backend that stores and communicates connectivity policies and credentials to clients. Integrates to a separate mobile billing/activation (BSS) platform and a separate public WiFi authentication infrastructure (AAA).
Hosted PostgreSQL: Hosted database server that stores all client device information, policies and manages CM application binaries for self-update process.
Connection Manager Clients: Register with Connection Manager Server and receive connectivity policy information, network information, and credentials from server. Responsible for implementing policy and network auto-connections with host OS as well as collecting analytics and reporting information for product intelligence.

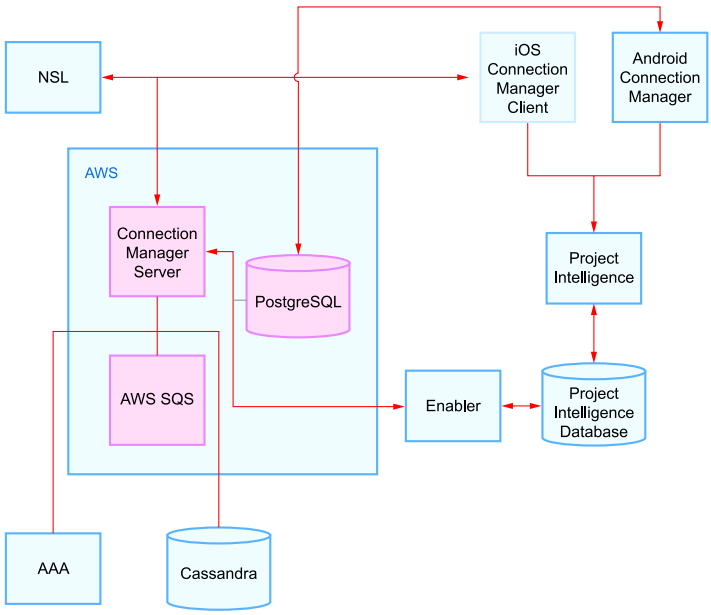


Figure 6: Connection Manager High Level Design

The Android mobile software client is currently embedded in Samsung, LG, and Motorola device firmware. It contains pre-granted and whitelisted Android permissions. Connection Manager also features an over-the-top (OTT) client Android SDK that is embedded within host applications, subject to permission limitations by user selection.

A connection manager application currently utilized by MSOs includes the following components:

- Built on open-source stack
- Elastic bean stalk instance with Auto-Scaling group to support on demand scaling
- Defined public private subnet, public private route table, NAT gateway, etc.
- Configured security group for Elastic load balancer, CM server and database, etc.
- Instance spread across multiple availability zones
- End-to-end TLS encryption (certificates managed by AWS certificate manager)
- Route 53 managed DNS to instance mapping
- VPS with AWS best practices
- Defined AWS WAF rules and Cloud Alerts
- Cloud formation script-based infrastructure creation and destruction

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

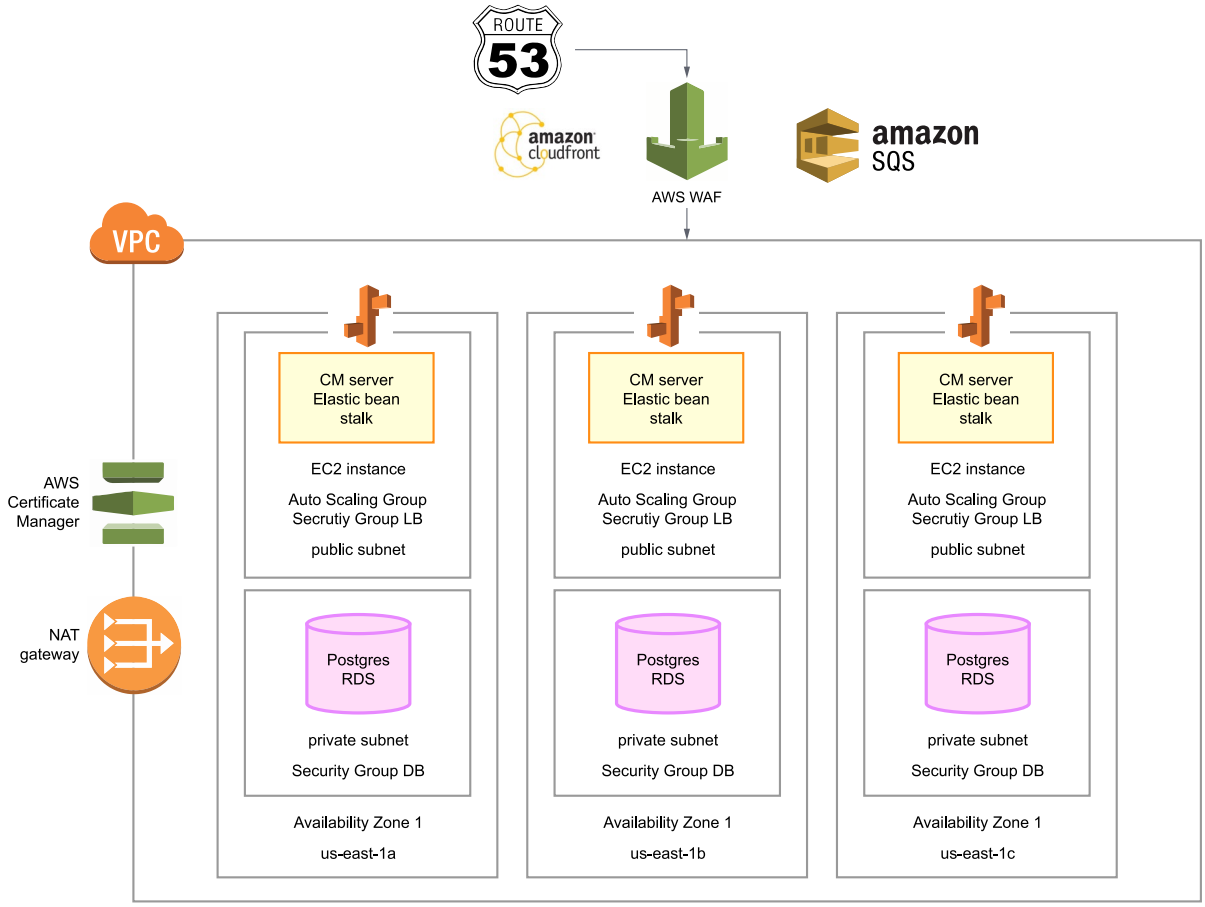


Figure 7: Connection Manager Components

7. Integration Points

Provisioning Platform (BSS) and CM Server: Account management operations (activate, deactivate, suspend) are sent in real-time to the CM server to update entitlements and policy per business rules. Upon next sync of policy to the device, the business rules are applied.

CM Server & Client: Clients register with CMS and receive all policy information assigned to the device per business rules.

Connection Manager Server & IDP Systems: CM Server must provision/deprovision unique per-device credentials into the IDP/AAA platforms to ensure the network parameters sent to the device are setup properly for the user.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

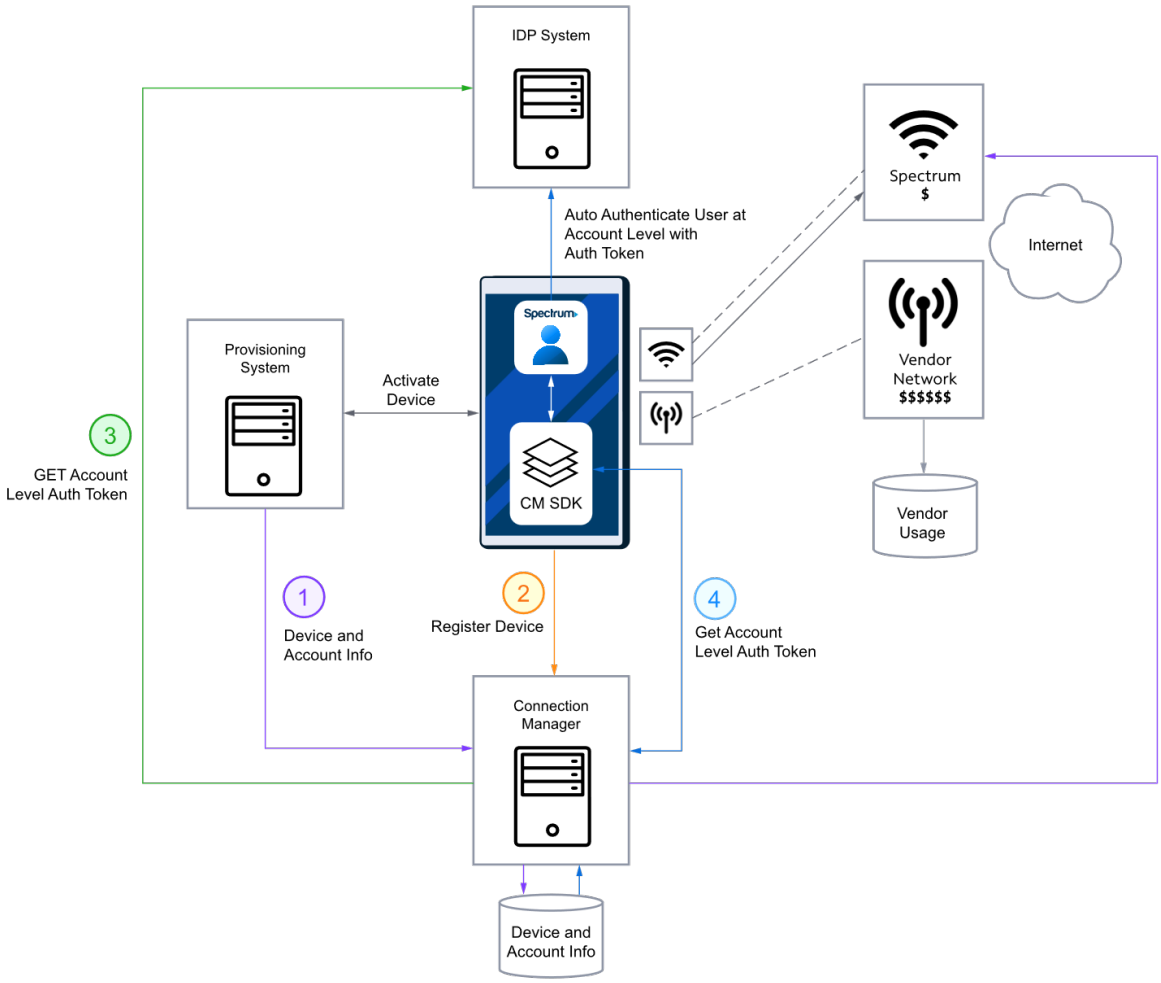


Figure 8: Connection Manager Integration Points

While Connection Manager was purpose built for a mobile offload use case, the current and future features can support the automotive use cases of the future. Current features include: network quality check, auto-connect, network management (including banned list), WiFi on/off control, user opt-out, network planning, and competitive analytics. As the application of Connection Manager broadens, the roadmap for logic will continue to evolve. Two key features on the roadmap are 1) support for satellite connectivity in conjunction with mobile and WiFi and 2) least cost routing intelligence to manage data traffic timing based on use case and cost considerations.

As OEMs contend with various connectivity options, the Connection Manager roadmap will support a real time, seamless management of all these connections and, based on traffic markings, enable more efficient cost management and higher throughput, resulting in improved outcomes (customer – faster updates, and OEM – vehicle maintenance and ADAS training log uploads). Utilizing a technical solution such as Differentiated Service Code Point (DSCP) markings, utilized by the MSO industry for years, combined with a QoS Competitive Index (QCI), will allow Connection Manager to ensure quality of service for connections and steer data as needed based on the speed and cost of the network, whether cellular (MNO or MSO), WiFi, or even satellite. Real-time payloads like telematics and entertainment can still be serviced in a meaningful way via cellular (MNO or MSO, based on availability), mobile connectivity. Emergency updates such as urgent vehicle maintenance updates (e.g., malware or known

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

defect) can be transferred via the most appropriate network, such as higher cost, lower throughput cellular connection, if needed. Delay tolerant payloads such as training logs from ADAS systems can be queued at the right time and right priority on WiFi or MSO CBRS, based on speeds, and to ensure maximum compliance. The connectivity solution will be tuned to address the needs, always balancing cost and throughput.

8. Conclusion

The evolution of the automotive industry towards digitization, electrification, and software-defined architecture has highlighted the need for a transformative shift in automotive connectivity. As the reliance on high-bandwidth connectivity grows with the proliferation of software-defined vehicles, ADAS and the development of true autonomous driving, automotive manufacturers must rethink their approach to connectivity in order to stay competitive. The adoption of a Connection Manager, complementing existing cellular networks and new MSO CBRS networks with broadband and WiFi solutions, will be critical in addressing the increasing data needs, ensuring seamless maintenance and updates, and providing enhanced experiences for customers in the connected and autonomous vehicles of the future.

Bibliography & References

1. Juniper Research - Global Connected Vehicles
2. Cubicle Telecom - Connected Car Growth
3. Roland Berger - Driving Automation Penetration by Level
4. EU Zero Emission Plan; US Zero Emission Target
5. ICE Vehicle Pricing - Source #1; Source #2, EV Pricing - KBB
6. Dr. Richard Windsor, Radio Free Mobile
7. ICE Planned Obsolescence
8. Average Lifespan
9. Motor and Wheels - # of parts
10. Total automotive market calculated assuming ~15M cars sold per year at a price of $48k
11. Dgtl Infra - Data Consumption
12. Euro Tech - Data Consumption
13. Total Internet Bandwidth
14. Open Vault Broadband Insights
15. Circana Data Consumption Reports; Ericsson Mobile Data Traffic Outlook (2021-22E)
16. WiFi speed offered by leading MSO in 99% of footprint
17. Simple Avg of Big 3 MNO download speed
18. Qualcomm DSDA

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Converged Service Management: Wi-Fi Speed Boost

A Technical Paper prepared for SCTE by

Rahil Gandotra
Ph.D., Lead Architect
CableLabs
858, Coal Creek Circle, Louisville, CO - 80027
(303) 661-3439
r.gandotra@cablelabs.com

Yunjung Yi
Ph.D., Principal Architect & Director of Wireless Standardization
CableLabs
858 Coal Creek Circle, Louisville, CO - 80027
(303) 661-3849
y.yi@cablelabs.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Converged Service Management Layer ... 4
3. Wi-Fi speed boost ... 8
4. Conclusion ... 11

Abbreviations ... 13

Bibliography & References ... 13

List of Figures

Title **Page Number**

Figure 1 - High-level illustration of CSML ... 4
Figure 2 - ETSI NFV MANO architecture ... 5
Figure 3 - 3GPP-defined network slice management with interface to NFV-MANO ... 6
Figure 4 - O-RAN SMO framework ... 6
Figure 5 - High-level CSML architecture ... 7
Figure 6 - Network topology of Wi-Fi speed boost ... 9
Figure 7 - End-to-end interactions of CSML with different xNFs ... 10

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Two network paradigms are transforming service provider networks. First, is the transition of network functionalities being implemented as legacy physical appliances to virtualized functions running on commercial-off-the-shelf (COTS) hardware. Deploying or updating physical appliances forces operators to substantially increase both capital and operating expenditures due to the need of specialized hardware that are expensive, have high energy costs, and have limited scope for adding new functionalities [1]. Virtualizing network functions as software-based applications allows modularity and isolation of each function enabling enhanced management, network optimization, and cost reduction. Second, is the decoupling of control plane from the user plane, and the shift of closed inter-function interfaces to open standards-based interfaces. This enables the network control to be directly programmable, agile, and centrally managed, and a unified interconnection standard for white-box hardware and open-source software elements from different vendors [2].

This transformation benefits the operators by allowing for faster time to deployment, enabling enhanced flexibility, and giving them the ability to offer novel differentiated services to their users. An important requirement to realize this new paradigm is the capability to manage different access domains, such as cable and mobile, from a central management platform. While traditionally different network domains have operated in their own siloes, enabling operations convergence involves developing a common framework for deploying, configuring, and managing network functions constituting a service. Employing network domain-specific solutions with dedicated teams to provision and manage each different access service leads to a disjointed model of network management which is challenged by operational economics.

Typically, domain-specific management systems, such as for cable or mobile networks, do not have the visibility or control over other domains for fault, configuration, accounting, performance and security (FCAPS) tasks, and a central management entity can enable that inter-domain communication to achieve multi-access convergence. The Converged Service Management Layer (CSML) project aims to harmonize the management of multi-domain services by providing a single comprehensive framework to model end-to-end services and abstracting and automating the control and management of physical and virtual resources [3]. The primary goal of the project is to demonstrate the importance of converged service operations by developing novel use cases on differentiated service offerings enabled through enhanced operational agility.

CSML is envisioned to act as a master Element Management System (EMS) communicating with various domain-specific infrastructure layers, EMS, and network functions in order to develop converged services. CSML consists of a service orchestrator which leverages different workload and network orchestrators to deploy and manage services comprising of multiple functions operating in different network domains. Figure 1 illustrates the high-level concept of CSML. Different types of network functions – Physical Network Function (PNF) or Cloud-native Network Functions (CNF) – can be deployed in different zones – centralized cloud or headend or customer premises – in different network domains – cable or mobile. CSML enables the deployment of new services that converge different accesses allowing for improved quality of service for operators and seamless user experiences for subscribers.

As part of the CSML project, multiple proof-of-concepts (PoC) were developed, described in section 2, to demonstrate the converged management capability in novel use cases. This paper presents details on one use case on Wi-Fi speed boost. The motivation behind this use case was to provide the operators with the ability to incentivize their subscribers who have purchased both of their mobile and home Internet services. We extend the last PoC developed on dynamic cable speed boost by incorporating 5G core (5GC) and Wi-Fi access point (AP) into the framework to make the triggering of the speed boost more dynamic. The paper

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

examines the mechanisms developed for identifying mobile devices connecting to an operator's Wi-Fi network and activating cable speed boost for that device in an automated way.

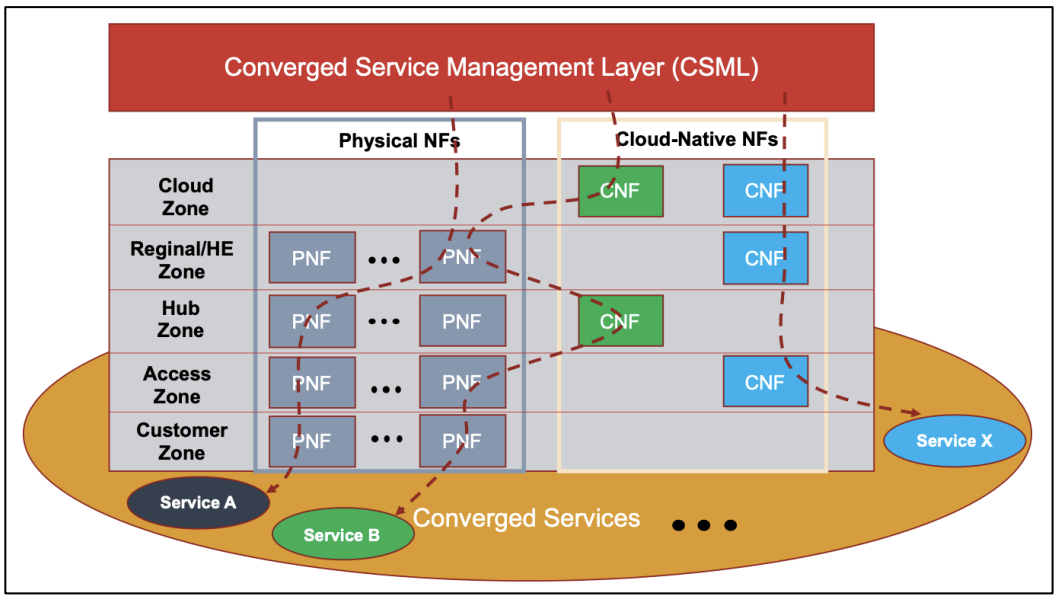


Figure 1 - High-level illustration of CSML

2. Converged Service Management Layer

Management and orchestration functionalities have been defined by multiple different standards development organizations (SDO) such as ETSI, 3GPP, and O-RAN, applicable to their domains. Certain features defined by them are independent while some are overlapping. An overarching framework that abstracts domain-specific management notions will enable a simplified approach with enhanced capabilities of cross-domain interactions without which piecemeal management solutions would need to be integrated leading to high operational complexities.

The Network Functions Virtualization Management and Orchestration (NFV-MANO) architecture specified by ETSI is shown in Fig. 2 [4]. It consists of a virtualized infrastructure manager (VIM) that is responsible for the control and management of the network functions virtualization infrastructure (NFVI) compute, storage, and network resources, usually within one operator's infrastructure domain. The Virtualized Network Function Manager (VNFM) is responsible for the lifecycle management of individual virtualized network functions (VNF). The NFV Orchestrator (NFVO) acts as the global resource manager responsible for managing services composed of multiple VNFs. Reference points with different functionalities have been specified for the interfaces between two entities. ETSI has also leveraged the NFV MANO architecture for specifying the multi-access edge (MEC) architecture in an NFV environment [4]. Interfaces between the MEC entities and MANO entities have been specified as reference points. ETSI has also developed an open-source MANO (OSM) software stack as reference implementation aligned with the NFV MANO architecture. CSML enables enhanced MANO acting as an NFVO and VNFM. It includes the capability to manage VNFs using either its built-in generic-VNFM (gVNFM) module or by interacting with external specific-VNFM (sVNFM), including an interface to the VIM layer for virtual resource management over the NFVI. Additionally, CSML can manage PNFs either via external EMS or by directly interacting with PNFs using its PNF Manager (PNFM) module over well-known management protocols.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

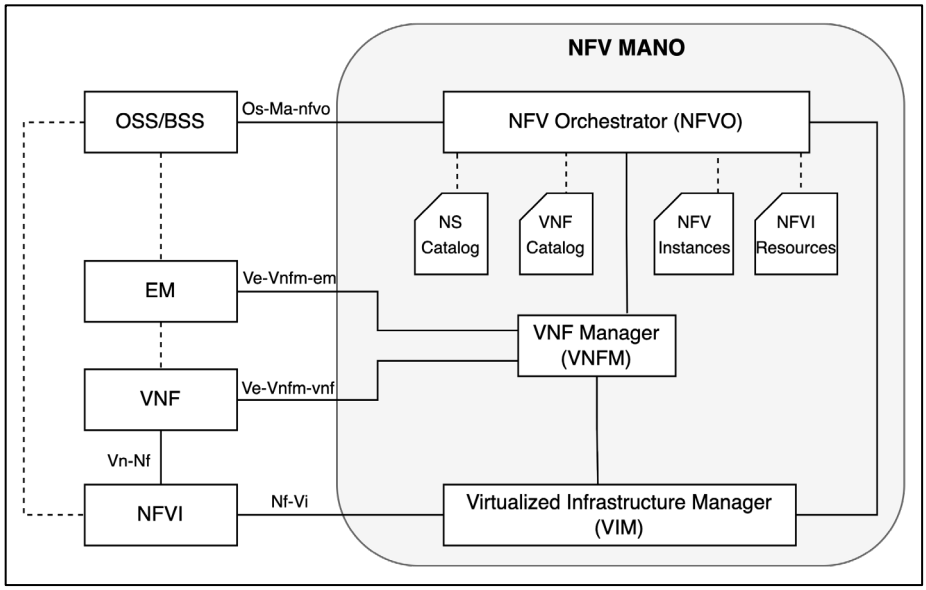


Figure 2 - ETSI NFV MANO architecture

Within the 3GPP's Technical Specification Group for Service and System Aspects (TSG SA), the working group 5 (SA5) is responsible for the management and orchestration capabilities in 3GPP systems. SA5 covers aspects such as operation, fulfillment, assurance, and automation, which includes management interactions with entities external to the network operator (e.g., service providers and verticals). The 5G management system includes features such as self-organizing networks, management data analytics, a service-based management architecture, intent-based management. Additionally, management of new 5G-enabled capabilities such as AI/ML, network slicing, URLLC, edge computing, and energy efficiency are specified. SA5 also coordinates with other SDOs such as ETSI and GSMA in the specification work pertinent to management and orchestration. For instance, Fig. 3 illustrates how the network slicing management functions defined in 3GPP utilize the ETSI NFV-MANO interfaces for configuration, fault, performance, and life-cycle management of network slices [5]. CSML architecture has the capability to incorporate such network domain-specific management functions, like the 3GPP-defined Communication Service Management Function (CSMF) and Network Slice Management Function (NSMF), to realize new use cases which help operators play a key role in catering to the needs of industry verticals apart from the traditional broadband service needs in an optimal manner.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

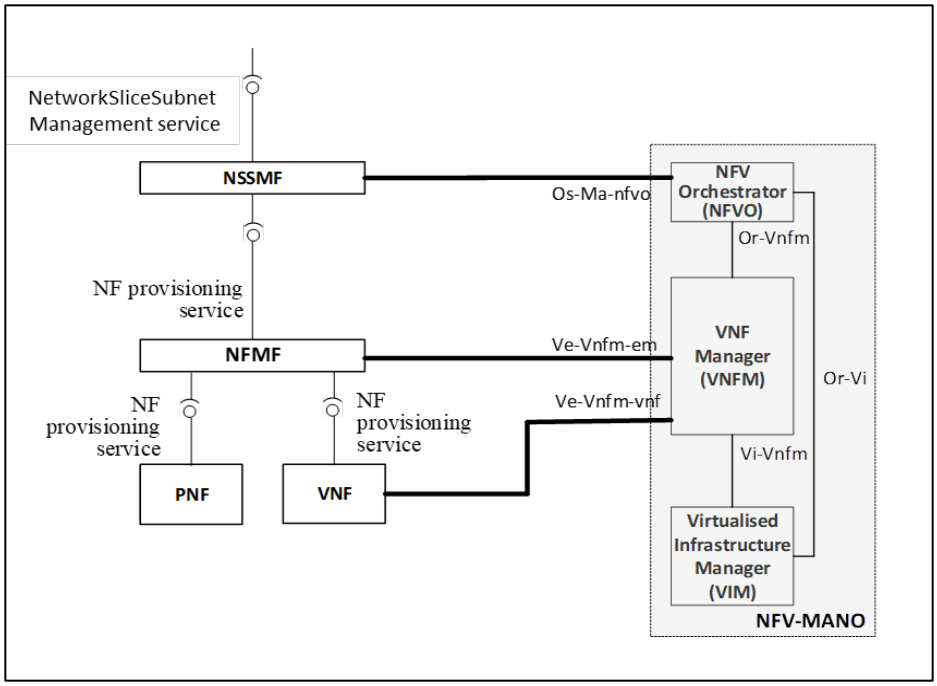


Figure 3 - 3GPP-defined network slice management with interface to NFV-MANO

The Open RAN Alliance (O-RAN) defines technical specifications and interfaces related to RAN's service management and orchestration (SMO) [6]. The SMO framework can be understood as a RAN domain controller enabling an automation platform for O-RAN radio resources. Fig. 4 depicts the SMO components and interfaces. The capabilities enabled by SMO include infrastructure management of O-Cloud, lifecycle management of the O-RAN network functions, and intelligent RAN optimization in non-real-time by providing policy-based guidance using data analytics and AI/ML models. The O-RAN O2 interface is functionally similar to the NFV-MANO Nf-Vi interface for a VIM to manage a NFVI and the O1 interface is similar to the NFV-MANO Ve-Vnfm-vnf interface for a VNMF to manage VNFs. CSML enables leveraging existing capabilities as well as incorporating new features in order to include RAN management within its purview.

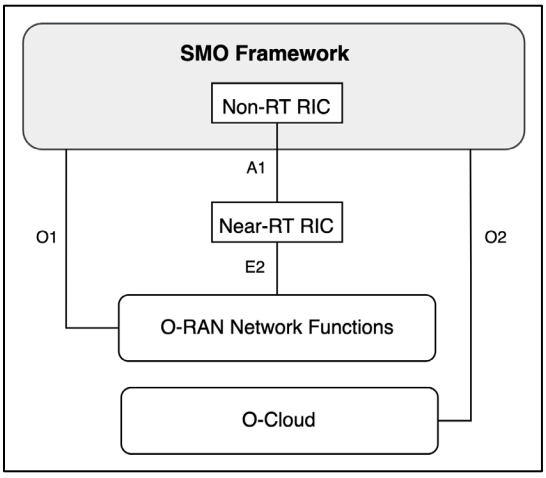


Figure 4 - O-RAN SMO framework

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The primary motivation of the CSML project is to create a coherent strategy for operators to integrate these different domain-specific management functions as these technologies are introduced into their networks by enabling harmonious inter-operation of these distinct orchestration functions.

Fig. 5 illustrates the high-level architecture of CSML. The two primary frameworks, namely the design-time framework and the run-time framework, enable three types of lifecycle management (LCM) activities:

1. Service design: Allows for composing end-to-end services consisting of multiple xNFs, along with specifying their Day-0 and Day-N configuration parameters and policy rules, enabling elastic management of the service.
2. Service deployment: Involves instantiating the modeled services on to the target infrastructures (both physical and virtual) and supervising scale-in/out as needed. For deployment of VNFs and CNFs onto NFVIs, CSML interacts with the VIM APIs such as OpenStack or Kubernetes, and for provisioning of PNFs, CSML enables their plug and play by discovering the deployed PNFs and managing them using standard management protocols such as NETCONF or Ansible.
3. Service assurance: Involves monitoring the deployed services and taking closed-loop actions using analytic tools to make the framework self-healing and self-optimizing. CSML enables telemetry data collection both directly from the xNFs or from external collectors such as Prometheus or NetFlow.

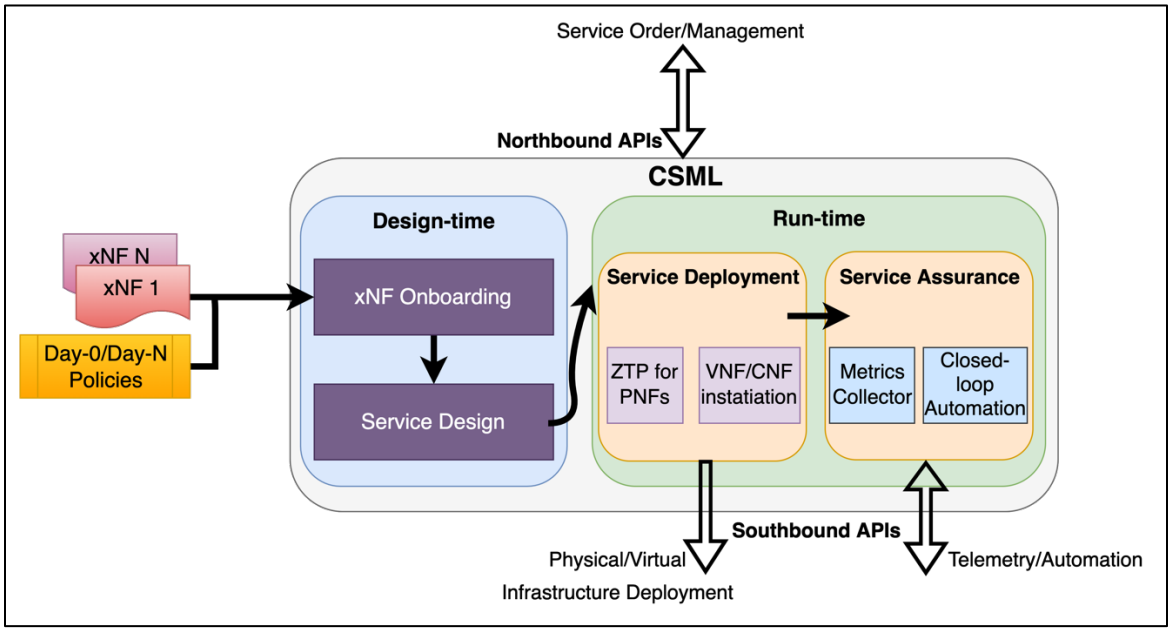


Figure 5 - High-level CSML architecture

The first use case developed during the runtime of the project was on orchestration and assurance of a wireless steering application [7]. The application consisted of eight containerized server-side functions to be deployed on three different cloud locations. CSML communicates with both public cloud and private cloud VIMs to instantiate the CNFs appropriately based on the service requirements - user-plane functions needed to be geographically closer to the user, while the control-plane functions needed to be centralized in the public cloud. Additionally, CSML enabled closed-loop automated assurance of the service by collecting near-real-time operational data, checking with pre-loaded management policies, and taking the corresponding action enabling self-healing of the service. This use case demonstrated the capability of

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

CSML to orchestrate CNFs-based network services on different virtualized environments by providing a common abstracted framework for intelligent deployment and management.

The second use case developed using CSML was on dynamic cable speed boost. Typically, in cable networks, the process of requesting for and implementing a change in subscriber bandwidth is manual and requires a reboot of the cable modem (CM) for the change to take effect. This process was automated by leveraging the PacketCable Multimedia (PCMM) technology that provides a mechanism to create dynamic DOCSIS service flows [8]. A virtual PCMM domain controller was onboarded into CSML to enable dynamic, closed-loop, application-specific QoS control over the legacy CMTS and CMs to support existing and future QoS-enhanced services. With this use case, CSML demonstrated how the management of physical network elements can be harmonized with virtual elements to preserve existing network investments. CSML can also act as the Telco Service Exposure Platform as part of the Linux Foundation CAMARA project enabling harmonizing of technology-specific APIs facilitating the application-to-network integration [9].

3. Wi-Fi speed boost

While the prior use cases developed using CSML focused on physical and virtual network functions orchestration, they were limited to domain-specific domains. The first use case targeted virtualized functions in the mobile and Wi-Fi domains, while the second use case focused on incorporating HFC-domain virtual and physical functions into the framework to demonstrate hybrid orchestration. The third use case, described in this section, was developed to facilitate operations convergence for operators that provide both cable and mobile services.

In order for an operator to incentivize users who have subscribed to both of their mobile and home internet services, a speed boost could be enabled for those mobile devices while connected via Wi-Fi to their cable network. This requires automated –

(i) Identification of mobile devices connecting to the Wi-Fi access points,
(ii) Verification that those mobile devices are registered with their mobile network, and
(iii) Initiating a speed boost for that device over the cable network.

To enable this, a management entity is needed to interact with multiple core and access network functions in an automated manner. CSML can act as the central orchestrator in this scenario to realize this use case by leveraging its capability to control both virtual and physical functions in mobile and cable networks.

Fig. 6 illustrates the high-level network topology of the use case implemented. CSML is based on a microservices-based architecture deployed in the public cloud in Azure Kubernetes Service (AKS). It employs Representational state transfer (REST) APIs in the southbound direction to communicate with different network functions and controllers. Interfaces were developed for CSML to interact with a 5G core, a PCMM controller, a CMTS, and a Wi-Fi AP. The 5G core was based on the open-source project free5GC [10]. The core is based on 3GPP Rel-15 standalone specifications and supports service-based interfaces. UE and RAN were simulated using open-source UERANSIM [11]. It consists of an 5G-SA UE and gNodeB implementation. The open-source OpenDaylight (ODL), a Linux Foundation Networking (LFN) software-defined controller project, was used as the PCMM controller in this use case which includes an implementation of the PCMM service as an additional module [12]. ODL was running as a virtual machine (VM) in the public cloud. ODL exposes northbound REST APIs to provision a CMTS and service flows and uses the Common Open Policy Service (COPS) protocol as transport to communicate with the CMTS. CMTS uses MAC-layer DOCSIS Dynamic Service Add/Change/Delete (DSA/DSC/DSD) messaging for service flow management for a specific CM. An Arris E6000 was used as the CMTS, a Netgear CM1000

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

was used as the CM, and a Flint GL-AX1800 running the OpenWrt operating system was used as the Wi-Fi AP. Another laptop was connected to the AP for throughput-testing purposes.

The design-time work involved onboarding the 5GC, RAN, and UE CNFs within CSML and composing a service. The 5GC consisted of Access and Mobility Management Function (AMF), Session Management Function (SMF), User Plane Function (UPF), Authentication Server Function (AUSF), NF Repository function (NRF), Unified Data Management (UDM), Unified Data Repository (UDR), Policy Control Function (PCF), Network Slice Selection Function (NSSF), Non-3GPP Inter-Working Function (N3IWF), and a web-UI function. CSML onboarded the 5GC, RAN, and UE CNFs packaged using Helm charts. Helm charts allow for managing complex Kubernetes applications through a collection of files that describe a related set of resources such as a virtual function, virtual network, configuration, and storage-related information. Additionally, blueprint models for ODL and CMTS were reused from the previous PoC on dynamic cable speed boost in order for their interaction with CSML. The southbound interface used to configure ODL is a REST-based environment that enables pipelining multiple API request-responses defined in a workflow, while the southbound interface used for the CMTS was a Python-based environment that allowed reusing existing scripts for managing the CMTS over SSH.

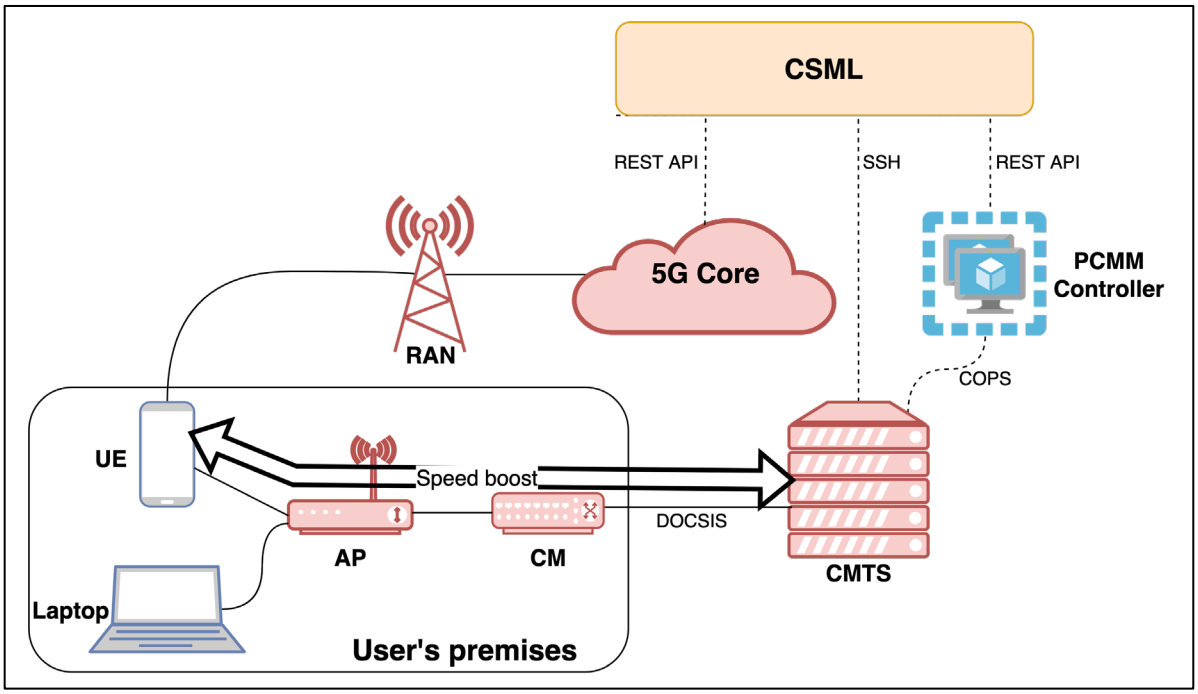


Figure 6 - Network topology of Wi-Fi speed boost

The first step in the run-time was to orchestrate the 5GC, RAN, and UE CNFs onto different Kubernetes clusters. 5GC CNFs were instantiated onto an AKS cluster, while the UE and RAN CNFs were instantiated onto a Kubernetes cluster running on a laptop acting as a mobile phone and the RAN node. This involved registering the Kubernetes clusters with CSML using cluster configuration files, employing the Kubernetes plugin within CSML to invoke Kubernetes APIs to deploy the CNFs, and configuring their interconnections [12].

To detect and identify a device connecting to the Wi-Fi AP, a mechanism was needed on the AP to interact with the device to identify if it is a mobile device, fetch its 5G identifier, and send this information to CSML. Custom scripts on the AP were employed for this purpose. A bash script was developed to identify

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

any new device connecting to the AP via DHCP. The script continually monitored the wireless interface on the AP for any new device connecting over Wi-Fi by tracking the MAC addresses and then obtained the IP address of the connecting device via the DHCP lease information. In case of MAC randomization, since the DHCP leased IP address changes if the MAC address changes, the mechanism on the AP to detect and obtain the current IP address for the device would suffice. Next, the AP used REST to invoke an API on the new device to fetch its International Mobile Subscriber Identity (IMSI). An API was developed to run on the UE CNF which provided the IMSI as a response to the API request from the AP. While this API on the UE was a simulated API, an operator application running on the UE could expose this information for use by the AP. For example, Android SDK has a package called TelephonyManager that provides access to information about the telephony services on the device. The getSubscriberId method within this package returns the unique subscriber ID, for example the IMSI, to a calling app that has carrier privileges.

Since the UE connects to the CM using the Wi-Fi AP, a challenge while implementing this PoC was how would the CM identify traffic from the UE, since the Wi-AP NATs all LAN IP addresses to a public IP address. With respect to the CM, the identity of the UE is hidden behind the AP, while the packet filtering and service flow enforcement for that UE's traffic will occur at the CM. The solution employed for this PoC is the AP marking all traffic matching the UE LAN IP address using a DSCP value before forwarding it to the CM, and a classifier on the CM matching that DSCP value can be used to redirect traffic from the UE to a higher-bandwidth service flow. On the AP, the mangle table of the iptables rules was used to alter the IP headers by marking all traffic matching the LAN IP address of the UE with a DSCP value, such as 56, to be used locally by the CM.

Fig. 7 depicts the multiple interactions between CSML and different xNFs. The end-to-end flow works as:

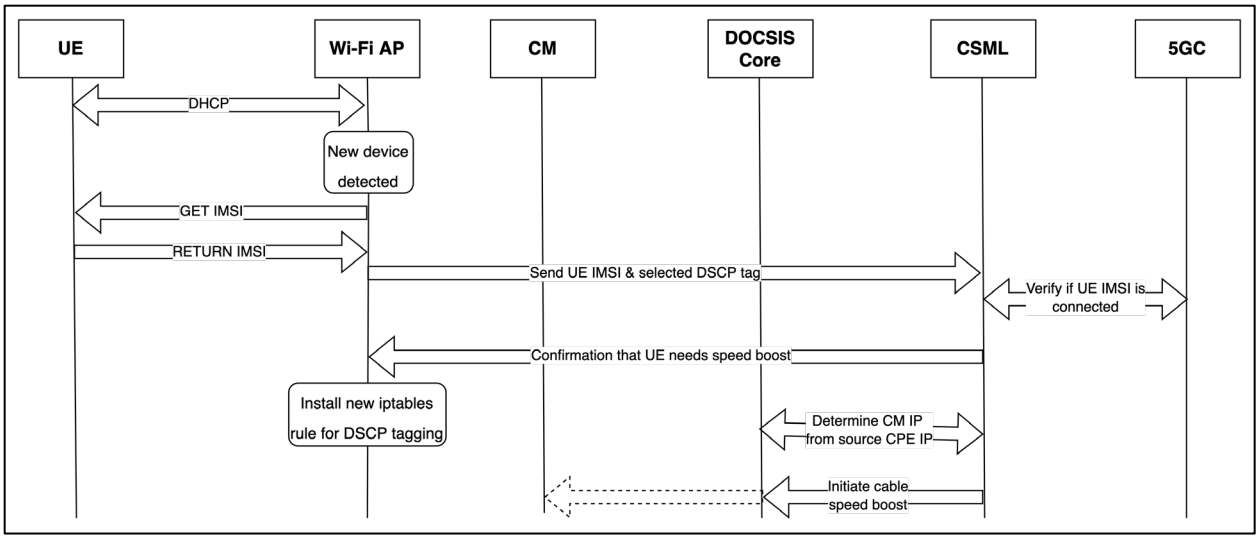


Figure 7 - End-to-end interactions of CSML with different xNFs

1. UE connects to the Wi-Fi AP using DHCP.
2. A bash scripting running on the AP detects a new device connecting over its wireless interface and acquires the LAN IP address of the UE.
3. A Python script running on the AP attempts to fetch the 5G identification, IMSI, from the UE using REST API.
4. An API running on the UE returns the IMSI to the AP. In case the device connecting to the AP is a non-mobile device, such as a laptop, the IMSI collection attempt by the AP will fail since no IMSI-exposure API will be running on the laptop.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Next, the AP selects a local DSCP value to be assigned to all the traffic matching the UE's LAN IP address, and sends the UE IMSI and the selected DSCP value to CSML using a REST API.
6. Once the API server running within CSML receives a request from the AP, it needs to check whether the IMSI of the UE is registered with the 5GC. It does this by invoking a REST API on the 5GC with the IMSI being sent as a payload in the HTTP request.
7. If the IMSI is registered with the 5GC using the 3GPP access, CSML then proceeds with the next steps to initiate a cable speed boost for that UE.
8. CSML first returns a confirmation to the AP that the UE requires a speed boost.
9. The AP then installs iptables rules to mark all traffic matching the UE's LAN IP address with the selected DSCP value by adding the corresponding line to the mangle table.
10. To push a PCMM policy to the correct CM, CSML needs to identify the CM IP. To do this, it uses the source IP address of the API request from the AP, connects to the CMTS over SSH, and checks which CM does that CPE IP belong to.
11. After obtaining the CM IP address, CSML sends a REST API request to ODL with the new service flow information. This request consists of the CM IP, the classifier matching all traffic from the UE's LAN IP address, and the name of the pre-configured higher-bandwidth DOCSIS service class.
12. ODL uses the COPS protocol to communicate with the CMTS about the new service flow requested, and the CMTS uses DOCSIS to push the service flow to the specific CM.
13. Once ODL returns the confirmation to CSML that the flow has been pushed, as part of verification, CSML initiates a SSH connection to the CMTS requesting the list of all service flows active for that CM IP.

The end-to-end flow described above demonstrates the level of abstraction offered by using a converged service orchestrator. Once the UE connects to the cable network through the Wi-Fi AP, the various steps needed to detect, identify, and initiate a speed boost are automated by using CSML. This enables eliminating any human intervention needed, either by the subscriber or by the operator, to achieve this. CSML enables integrations with multiple different network functions and controllers, thereby allowing for the implementation of novel use cases with the orchestrator handling much of the operational complexities associated with them.

To test the overall functioning of this use case, once the UE connects to the AP and CSML installs the higher-bandwidth service flow, throughput testing was performed on the UE and the laptop connected to the AP. While the tests performed for the UE indicated throughput up to ~450 Mbps, throughput on the laptop was capped at the subscribed bandwidth – 100 Mbps. This highlights the capability enabled by CSML for the operators to provide differentiated services to their bundled subscribers.

While the same use case could be achieved by developing various scripts to achieve individual xNF-to-xNF communication, using a centralized orchestrator provides additional benefits. Firstly, developing scripts for interaction with each vendor-specific xNF is not trivial, and the form of automation needed to achieve this without any human intervention is complex. CSML abstracts and automates various workflows to help composing an end-to-end service comprising of multiple functions and exchanges. Secondly, once a function and its associated interaction has been onboarded within CSML, the same could be used in other service designs, thereby eliminating the need to start from scratch. For example, in this PoC, the workflows defined in the previous use case on dynamic cable speed boost with ODL and CMTS were reused and incorporated with the 5GC and Wi-Fi AP to enable multi-domain management.

4. Conclusion

Network convergence allows for unifying disparate access technologies to deliver seamless connectivity enabling operators to intelligently use both mobile and fixed networks to deliver new services. To benefit

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

from this new paradigm, a converged service operator needs to have the ability to model end-to-end multi-domain services and to abstract and automate the control of physical and virtual network functions. This paper presented the CSML framework, highlighted the different use cases developed using it on multi-cloud orchestration, closed-loop assurance, and dynamic cable speed boost. Details were provided on the Wi-Fi speed boost PoC wherein a mobile device belonging to a subscriber of both broadband and mobile services from an operator is provided a higher broadband bandwidth. This involved detecting a mobile device connecting to the operator's AP, checking if this mobile device is registered with the operator's mobile core, and initiating a speed boost for that device over the broadband network. The paper demonstrates the value of operations convergence by demonstrating the novel capability enabled by using a centralized orchestrator for multi-domain network service comprising of different types of network functions and controllers.

In addition to the use cases described in this paper, O-RAN SMO capability was also incorporated into CSML. The O-CU, O-DU, and O-RU software was based on simulators developed by the O-RAN software community. The CNFs Helm charts were onboarded within CSML and instantiated onto a Kubernetes cluster. Configuration, fault, and performance management was implemented over the O1 interface using NETCONF and YANG models. This further augmented CSML's capability to manage a wide variety of network functions belonging to mobile and cable cores and access networks. Employing CSML in other novel use cases working with different stakeholders would help drive the adoption of the concept of converged orchestration.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

5GC	5th Generation mobile network core
AKS	Azure Kubernetes Service
AP	Access point
CM	cable modem
CMTS	Cable Modem Termination System
CNF	cloud-native network function
COPS	Common Open Policy Service
COTS	commercial off-the-shelf
CSML	Converged Service Management Layer
DOCSIS	Data-Over-Cable Service Interface Specification
EMS	element management system
ETSI	European Telecommunications Standards Institute
FCAPS	fault, configuration, accounting, performance, security
HFC	hybrid fiber-coaxial
IMSI	international mobile subscriber identity
MANO	management and orchestration
NF	network function
NFV	network functions virtualization
NFVI	network functions virtualization infrastructure
NFVO	network functions virtualization orchestrator
ODL	OpenDaylight
PCMM	PacketCable Multimedia
PNF	physical network function
QoS	quality of service
SDN	software-defined networking
SDO	standards development organization
VIM	virtualized infrastructure manager
VM	virtual machine
VNF	virtual network function
VNFM	virtual network function manager

Bibliography & References

[1] J.G. Herrera and J.F. Botero, "Resource allocation in NFV: A comprehensive survey," in *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, Sep. 2016, pp. 518-532.
[2] N. McKeown et al., "OpenFlow: Enabling network innovation in campus networks," in ACM SIGCOMM Computer Communications Review, vol. 38, no. 2, pp. 69-74, Mar. 2008.
[3] R. Gandotra, S. Khan, and Y. Yi, "Converged service orchestration: Dynamic cable speed boost," in *SCTE Cable-Tec Expo Fall Technical Forum*, Philadelphia, Sep. 2022.
[4] ETSI, "NFV Release 4; Management and Orchestration; Architectural Framework Specification," *ETSI GS NFV 006*, v4.4.1, Dec. 2022. [link]

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

[5] 3GPP, "Management and orchestration; Architecture framework," *TS 28.533*, v17.3.0, Mar. 2023. [link]
[6] O-RAN, "Architecture description," *WG1.OAD-R003*, v09.00, 2023.
[7] Cable Television Laboratories, Inc., "IWiNS—An Informed Approach to Mobile Traffic Steering", January 2021. [link]
[8] Cable Television Laboratories, Inc., "PacketCable™ Multimedia Specification", PKT-SP-MM, November 2019. [link]
[9] CAMARA: The Telco Global API Alliance, The Linux Foundation Projects, https://camaraproject.org/.

[10] free5GC, https://github.com/free5gc/free5gc.

[11] UERANSIM, https://github.com/aligungr/UERANSIM.

[12] OpenDaylight, The Linux Foundation Projects, https://www.opendaylight.org/.

[13] Edge Multi-Cluster Orchestrator (EMCO), The Linux Foundation Projects, https://project-emco.io/.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Convergence With 5G Applications and Use Case Monetization

The Path to 5G Monetization

A Technical Paper prepared for SCTE by

Amitabh Chhabra
Architect
Rogers Communications Inc.
8200 Dixie Road, Brampton, Ontario, Canada
+14163055943
Amitabh.chhabra@rci.rogers.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Why 5G ... 3
3. Intelligent Transportation POC at Aurora testbed ... 5
   3.1. Introduction ... 5
   3.2. NoTraffic System Overview ... 5
   3.3. NoTraffic System Deployment ... 6
   3.4. AI based detection ... 7
   3.5. Results and data collection ... 8
   3.6. Summary from NoTraffic POC ... 9
4. Mixed Reality Human Teleoperation ... 11
   4.1. What is a Teleoperation ... 11
   4.2. Current Challenges ... 11
   4.3. Human Teleoperation System Architecture ... 12
   4.4. Communication network requirements and performance results ... 13
   4.5. Summary from Human teleoperation ... 15
5. Conclusion ... 15

Abbreviations ... 17

Bibliography & References ... 18

List of Figures

Title **Page Number**

Figure 1- Non-Roaming 5G System Architecture (Source 3GPP : 23501.i10) ... 4
Figure 2- NoTraffic system Overview ... 6
Figure 3- NoTraffic deployment at the 9 locations at UBC ... 7
Figure 4- NoTraffic typical installation at Wesbrook-Thunderbird at UBC ... 7
Figure 5- NoTraffic Pedestrian and Vehicle detection ... 8
Figure 6- Pedestrian and Vehicle data Figure 7- Vehicle vol bef/after optimization ... 9
Figure 8- Overview of Human Teleoperation system ... 12
Figure 9- Overview of communication system ... 14
Figure 10- Aspects of expert-follower spatial registration ... 15

List of Tables

Title **Page Number**

Table 1- Pedestrian delay improvements at 4 Intersections ... 9
Table 2- Vehicle delay improvements at 4 Intersections ... 9
Table 3- City-Wide estimates based upon intersection results ... 10
Table 4. Worst-case uplink and downlink data throughput requirements on the follower side ... 13
Table 5. Latency(round trip) measurements with different network technologies ... 14

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

4G was a great leap forward, allowing people to stream music and video on the go. 5G takes it to a different level and is designed to connect many more types of devices than smartphones – anything really. This opens a plethora of 5G applications and use cases ranging from intelligent transportation, smart cities, automated factories, telemedicine and a new wave of enterprise communications.

As a Canadian leader in 5G innovation, Rogers has been developing 5G use cases in several sectors in partnership with Canadian universities and industry consortiums. The objective is to pursue research discoveries and develop new applications of 5G to address societal and economic challenges.

In this paper, we will present case studies from two different sectors to demonstrate how 5G networks have enabled Rogers to accomplish its strategy and objectives. These use cases focused in the sectors of smart transportation and health care.

Taking signalized intersections technology to a new level, Rogers deployed NoTraffic autonomous traffic management system at seven live signalized intersections and two roundabouts at the University of British Columbia (UBC) campus in Vancouver. Reducing pedestrians' delay in crossing intersections, was one of the priorities of this project. The deployment optimized the flow of both vehicles and pedestrians. After two weeks, the project's KPIs indicated it had delivered CO2 reduction of 2.8 tons, with a pedestrian delay improvement of 92 hours and vehicle delay reduction of 182 hours. Total economic value over two weeks was about $6,500.

In the second application, the research team developed an ultrasound system which can perform a remote ultrasound operation under the guidance of an expert sonographer. The sonographer remotely "tele-operates" an untrained person (the follower) wearing a mixed reality headset by controlling a virtual ultrasound probe projected into the person's scene. The pose, force, video, ultrasound images, and 3-dimensional mesh of the scene are fed back to the expert in real time. 5G teleguidance is more precise and fast than verbal/video guidance, yet more flexible and inexpensive than robotic teleoperation. Remote ultrasound teleoperation has the potential to revolutionize healthcare access for first nations and remote communities, where receiving medical services can be a significant challenge. In these areas, where it can take months to schedule and conduct an ultrasound examination, this innovative technology can bridge the gap by enabling real-time, long-distance tele-ultrasound imaging.

Together, these two use case scenarios show how several technologies, such as IOT, 5G, cloud, and sensing, are converging and capable of significantly enhancing both public safety and quality of life.

2. Why 5G

5G technology holds immense importance for industrial use cases due to its transformative capabilities that revolutionize the way industries operate. With its ultra-low latency, high data rates, and massive device connectivity, 5G enables a range of applications that significantly enhance productivity, efficiency, and safety in industrial settings. One key advantage of 5G for industrial use cases is its ability to support real-time communication and control. This enables industries to implement remote monitoring and control systems, facilitating operations in hazardous or hard-to-reach areas without putting human workers at risk. Additionally, the low latency of 5G ensures minimal delays in transmitting critical data, allowing for quick decision-making and response times.

The 3GPP (3rd Generation Partnership Project) service-based architecture (SBA) is a key architectural framework which differentiates 5G from previous generations of technologies. Figure 1 below shows the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

reference 5G system architecture. It introduces a modular and flexible approach, where network functions are decoupled and communicate with each other using standardized interfaces. This SBA brings several benefits to the 5G ecosystem, including scalability, interoperability, and efficient service delivery. 5G improves on the 4G services over several dimensions:

- Enhanced Mobile Broadband (eMBB): According to the ITU's requirements, 5G should be able to provide peak data rates of up to 20 gigabits per second (Gbps) for downlink and up to 10 Gbps for uplink.
- Critical Communications (CC) and Ultra Reliable and Low Latency Communications (URLLC): In some contexts, extremely high reliability is required. For instance, for remote control of process automation, a reliability of 99.9999% is expected, with an end-to-end latency of <10 ms.
- Massive Internet of Things (mIoT). Several scenarios require the 5G system to support very high traffic densities of devices. The Massive Internet of Things requirements include the operational aspects that apply to the wide range of IoT devices and services anticipated in the 5G timeframe.
- Flexible network operations. These are a set of specificities offered by the 5G system. It covers aspects such as network slicing, network capability exposure, scalability, and diverse mobility, security, efficient content delivery, and migration and interworking.

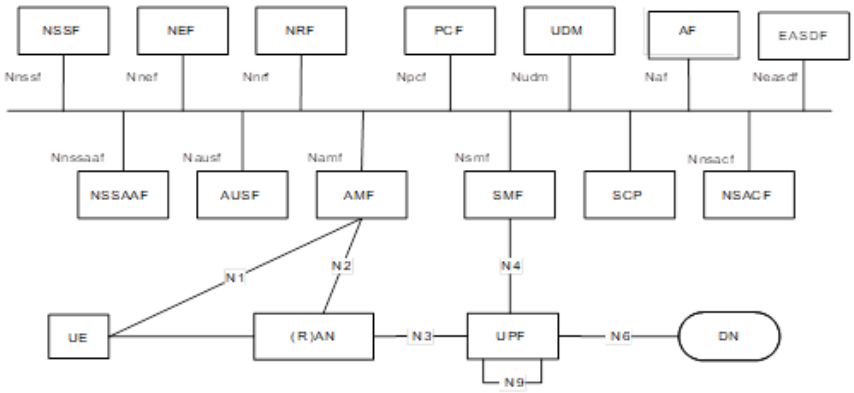


Figure 1- Non-Roaming 5G System Architecture (Source 3GPP : 23501.i10)

One of the important functions in the 5G SBA is the User Plane Function (UPF). The UPF is responsible for handling and processing user data traffic, acting as a data plane anchor in the 5G network. It plays a crucial role in enabling low latency use cases such as gaming, video streaming, smart factory and autonomous driving. UPF is also useful for quality of Service (QoS) Enforcement , Network Slicing, Efficient Data Offloading etc.

With the service based architecture, NFs capabilities are exposed via Representational State Transfer (REST) APIs which are based out of HTTP2.0 protocol. This helps the Application-driven Digital Service Providers which require agility and quicker time-to-market for the products/services rollouts such as advanced automation and robotics in industrial processes.

In conclusion, 5G technology is indispensable for industrial use cases as it enables real-time communication, advanced automation, and massive device connectivity. By leveraging the power of 5G, industries can unlock unprecedented levels of efficiency, safety, and productivity, paving the way for a new era of smart and connected industries.

Rogers plays a significant role in driving and fostering the culture of innovation in the country. Rogers actively collaborates with universities, startups, entrepreneurs, and incubators to support the growth of

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

innovative ideas and businesses. With our partnership with UBC, we are advancing research on several 5G use cases. In the next sections, I am going to present case studies from three different sectors to demonstrate how 5G networks have enabled Rogers to accomplish its strategy and objectives. These use cases are focused in the sectors of smart transportation, public safety and healthcare.

3. Intelligent Transportation POC at Aurora testbed

3.1. Introduction

Rogers and UBC have been working alongside public, private and academic partners with the goal of leveraging 5G to solve problems ranging from traffic congestion and climate change to wildfires and earthquakes. One of the projects is related to intelligent transportation, leveraging UBC's AURORA connected vehicle testbed. The researchers are using the testbed, powered by Rogers 5G, to explore ways to reduce traffic congestion and improve road safety, while also enhancing commercial vehicle operations and efficiencies. As traffic congestion is improved, it also reduces harmful emissions, improving air quality and helping in the fight against climate change.

In a spring 2019 report, Guidehouse Insights forecast that just 7% of North America's signalized intersections would use so-called adaptive traffic-control technology in 2020. On a global level, that number was 2.3%. This means 97% of traffic in the U.S. is not reacting to what's happening on the road.

NoTraffic solved this problem and is enabling the shift from analog traffic-light controls to digital, adaptive systems using artificial intelligence (AI). NoTraffic has developed the first AI-powered traffic signal platform that connects road users to the city grid, solving today's traffic challenges while unlocking smart mobility benefits for the cities, and creating an entirely new way of life.

3.2. NoTraffic System Overview

Figure 2 describes the system overview of the NoTraffic solution. NoTraffic has developed a hardware and software solution to manage traffic in real-time using a network of cloud-linked sensors deployed at intersections, corridors, or grid networks. Intersections may run in a passive data collection mode, detection mode, or optimization mode. The sensors are vendor agnostic and can work with any existing infrastructure. They are compatible with most traffic cabinets, traffic controllers, and power sources. The sensors fuse video and radar for object detection and classification and have a built-in Road-Side Unit (RSU) for Connected Vehicle applications. The sensors communicate to the cabinet equipment wirelessly, so only power is required. The Control Unit is installed in the cabinet near the intersection and interfaces with the traffic signal controller.

All sensors are connected to the cloud using wireless communications and accessed anywhere using the Virtual Management Center (VMC) dashboard. The VMC monitors the proper functioning of the traffic controller and provides real-time alerts if it detects any problems with the traffic controller (E.g., signal in flash, stuck pedestrian calls, communications failure). It also provides alerts from events picked up by the sensors, such as accidents or stuck vehicles.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

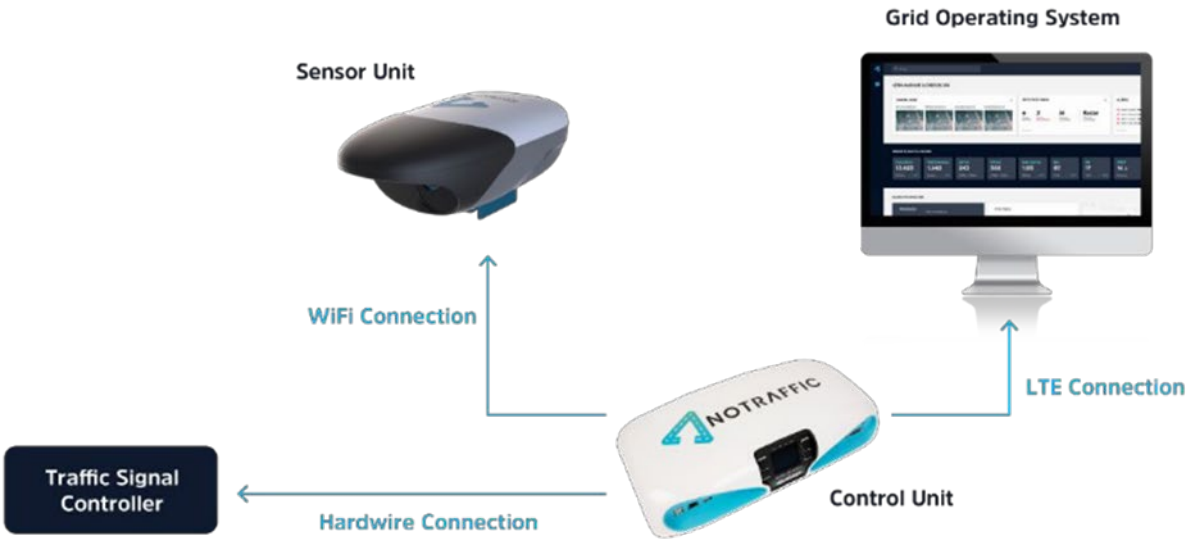


Figure 2- NoTraffic system Overview

3.3. NoTraffic System Deployment

The AURORA testbed is home to the Detection & Traffic Management pilot from NoTraffic. The sensors were installed at seven signalized intersections and two roundabouts across UBC's Vancouver campus. The goal of this project was to digitize the intersections using AI driven detection technology and reduce the traffic congestion and pedestrian delay along the Wesbrook Corridor at UBC campus in Vancouver by using advanced optimization features in the NoTraffic platform. Figure 3 shows the nine locations at UBC Aurora testbed where NoTraffic system was deployed.

NoTraffic system can be run in three different modes of operations:

Passive Mode: In Passive mode, the sensor collects data, but does not actuate the traffic signal. This is a typical application for a roundabout or pedestrian signal. The sensors do still pass traffic demand information to downstream intersections if the corridor/grid is in Optimization mode.

Detection Mode: In Detection mode, the NoTraffic sensors detect vehicles, and provide inputs to the existing traffic signal controller to operate the intersection. The key benefit is nearly 100% accuracy, reduced maintenance costs, and increased reliability of vehicle detections. The intersection still operates using the timing and detector plans programmed in the traffic controller.

Optimization Mode: In Optimization Mode, NoTraffic uses AI to autonomously optimizes traffic signal operations in real-time based on actual demand, and by predicting two minutes into the future. Rather than adhering to a fixed cycle or historical estimates, a predictive system changes, or adapts, based on actual traffic demand in real time. The software uses advanced AI algorithms to track and count vehicles, analyze incoming data, and respond appropriately regardless of intersection geometries or traffic demand changes.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

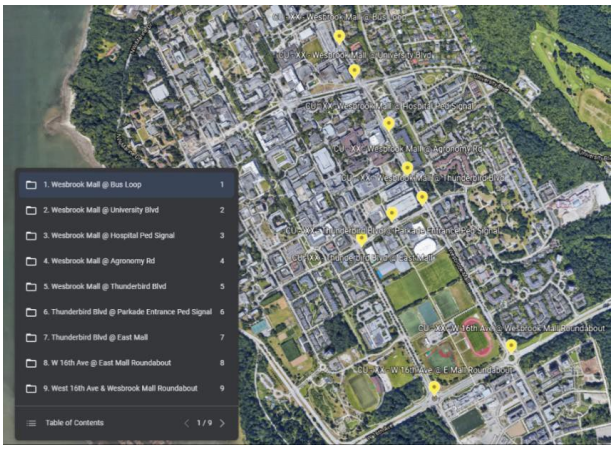


Figure 3- NoTraffic deployment at the 9 locations at UBC

Figure 4 shows a typical installation at one of the locations at Wesbrook and Thunderbird Blvd.

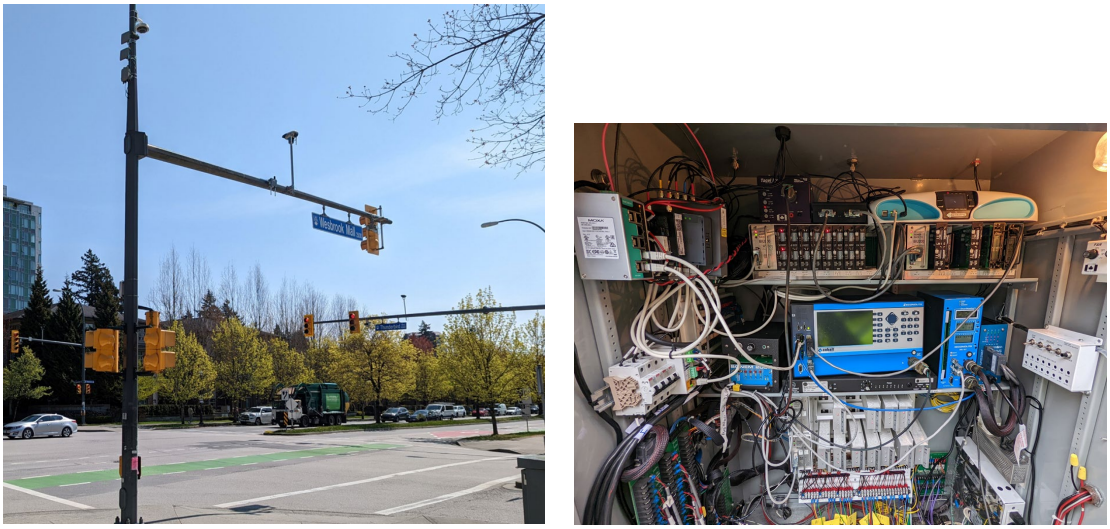


Figure 4- NoTraffic typical installation at Wesbrook-Thunderbird at UBC

3.4. AI based detection

The AI-Based Sensor collects object-based data. These sensors are mounted on each approach to provide a full view of the intersection. This data is collected and sent several times per second, leading to a rich and extensive data set used for optimization and analysis. The data includes:

- Object classification (light vehicle, heavy vehicle, pedestrian, bicycle, and truck)
- Object position (lane, phase, and distance to stop-bar)
- Object trajectory (direction and speed)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The platform also has the capability to place a call into the controller based on object classification. For example, if an agency only wants bicycles detected in a specific detection zone, the object classification of a vehicle will not allow the system to place an output in that detection zone to the traffic signal controller. Figure 5 shows the Pedestrian and Vehicle detection as seen from NoTraffic dashboard.

The AI-powered platform automatically controls traffic light grids, helping to reduce road congestion, coordinate transit and emergency services, and improve safety for drivers, cyclists and pedestrians alike.

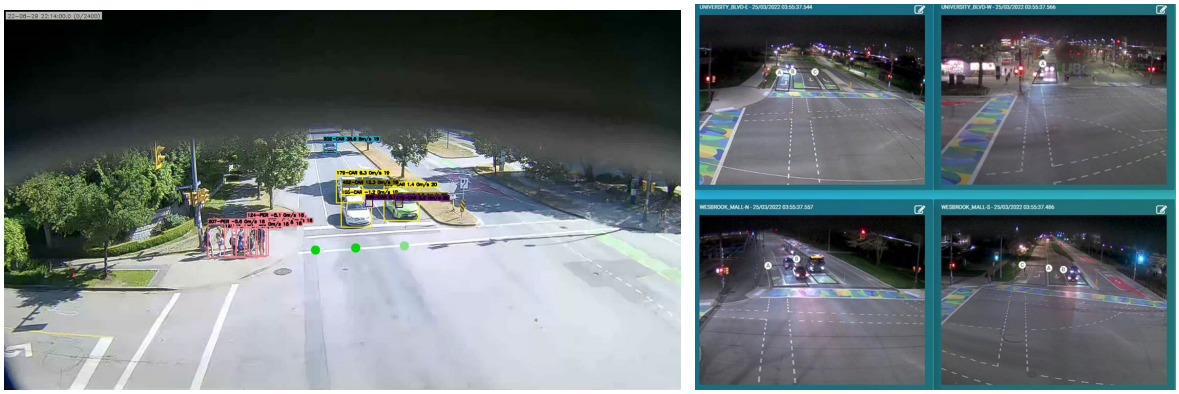


Figure 5- NoTraffic Pedestrian and Vehicle detection

3.5. Results and data collection

Figure 6 shows the counts of different traffic types going through the intersection. The AI-Based Sensor collects object-based data. These sensors are mounted on each approach to provide a full view of the intersection. This type of data supplements connected vehicle technology to improve on the current low penetration of in-vehicle communications devices (DSRC, C-V2X). Each object within the sensor field-of-view is assigned a unique identifier, a position, and a delay. Performance measures can be derived from this rich data set directly from the NoTraffic system, which can reduce costs and the number of different systems to effectively manage traffic for an agency. Figure 7 shows the Vehicle volume data before /after turning of optimization mode. The after scenario has slightly higher volume, which can have an impact on the final results.

Pedestrian delay was the major measure of effectiveness for this project. The research is still new in the industry, but the latest information shows that pedestrians will reduce compliance with walk signals at around 25 to30 seconds. Non-compliance is a surrogate measure for safety. For example, if a pedestrian crosses when they are not supposed to, it may surprise a vehicle and cause a conflict. This reduces the walkability of an area and induces more vehicular demand as pedestrian will not feel safe or adequately served. Furthermore, the higher the pedestrian delay, the higher the difference between “actual” and “perceived” delay, which happens around 25to 30 seconds. The point of this is that each second counts for pedestrians, and their delay is sometimes more important than a vehicle delay.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

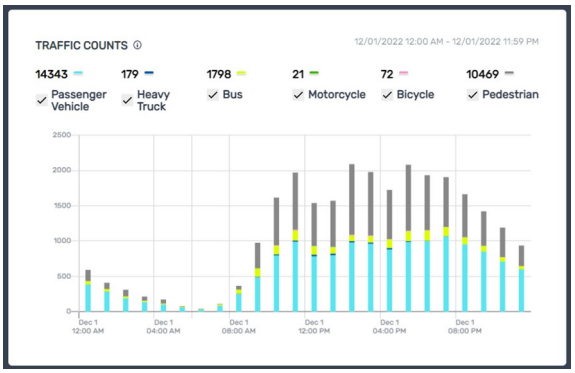


Figure 6- Pedestrian and Vehicle data

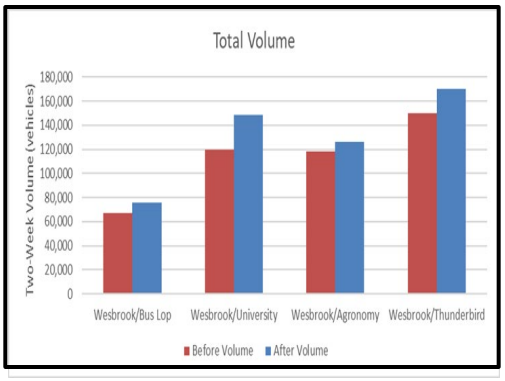


Figure 7- Vehicle vol bef/after optimization

For this project, pedestrian delay was reduced by 16% systemwide, or about six seconds per pedestrian. The major mode of travel on campus is via bike or walking. This reduction encourages more trips by foot or non-motorized mode and reducing vehicle-miles-travelled, thereby continuously improving the area and reducing greenhouse gas emissions. Table 1 shows the average improvements for each intersection.

Table 1- Pedestrian delay improvements at 4 Intersections

Intersection	Before, Average Delay (sec/ped)	After, Average Delay (sec/ped)	Improvement (sec/ped)	Improvement Change
Bus Loop	10.56	8.65	1.92	18%
University	20.85	18.96	1.89	9%
Agronomy	18.34	10.43	7.91	43%
Thunderbird	16.64	12.31	4.34	26%

On an annual basis, pedestrians are saved 2,429 hours. In the most critical areas, pedestrian delay was decreased by over 40%.

3.6. Summary from NoTraffic POC

Overall delay was improved across the study area by 7%, with both the non-coordinated and coordinated movements improving. Table 2 shows the vehicle delay improvements across the corridor.

Table 2- Vehicle delay improvements at 4 Intersections

Segmentation	Before, Average Delay (sec/veh)	After, Average Delay (sec/veh)	Improvement	Improvement Change
Main-street (Wesbrook)	16.74	15.49	1.25	7%
Side street	12.42	11.07	1.18	10%
Overall	**28.06**	**26.30**	**1.76**	**6%**

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Overall, the NoTraffic system works well in a small context but can have scalable and major improvements on a city-wide basis. Since the system removes legacy requirements, it can have outstanding impacts when deployed across a City, Province, or Region.

The benefits of the system for the scenarios scaled from five intersections over the course of a year are shown in table 3 below along with the benefits if no traffic deployment is scaled and deployed across the full city of Vancouver. The results clearly demonstrate that the deployment of AI-based systems at intersections brings significant economic value through the savings generated from reduced emissions and decreased vehicle delays. By leveraging advanced algorithms and real-time data analysis, these intelligent systems optimize traffic flow, resulting in smoother and more efficient movement of vehicles. As a result, there is a substantial reduction in the time spent idling at intersections, which translates into lower fuel consumption and emissions. The monetization potential arises from collaborations with municipalities and traffic management agencies by digitizing the intersections and generating useful data. The advanced data analysis from the intersections can provide real time traffic and accident insights which can potentially help municipalities to monetize with the insurance companies.5G technology effectively addresses the challenge of dense sensor deployments by offering increased capacity, higher bandwidth, and lower latency. To increase spectral efficiency, 5G networks employ cutting-edge techniques like massive MIMO (Multiple Input Multiple Output) and beamforming. This means that more devices can be connected simultaneously within a given area without experiencing significant degradation in performance. In addition functionalities like network slicing can be customized to meet specific QoS requirements of the ITS industry.

Table 3- City-Wide estimates based upon intersection results

Metric	2-week results on UBC Campus	…Scaled Annualized UBC Campus	…if scaled across Vancouver (city)
Number of Intersections	5	5	727
[1]Total Economic Value ($CAD)	$6,371	$165,654	$84,108,701
CO2 Reduced (Tons)	2.8	74	52,726
Pedestrian Delay Reduction	93.42 (hours)	2,429 (hours)	55,164 (days)
Vehicle Delay Reduction	181.67 (hours)	4,723 (hours)	139,921 (days)

[1] *Benefit numbers are calculated based upon UBC data with inputs from several NoTraffic deployments across North America.*

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. Mixed Reality Human Teleoperation

4.1. What is a Teleoperation

In today's world, mobile connectivity is a lifeline for many communities. It helps some of the most vulnerable people in places where timely medical assistance cannot be obtained. 5G promises to bring high speeds with low latencies which could advance telemedicine applications to a new level.

Telemedicine allows healthcare professionals to assess, diagnose, and treat patients from a distance using technology and other wireless communication devices. With telemedicine, patients don't have to go to their health care professional as frequently because health care specialists can make diagnoses and other specific treatments remotely. One popular way of providing telemedicine is through online video chat.

Current teleguidance methods include verbal guidance and robotic teleoperation, which present tradeoffs between precision and latency versus flexibility and cost. We present a novel concept of human teleoperation which bridges the gap between these two methods. We developed an ultrasound system which can perform a remote ultrasound operation under the guidance of an expert. The Expert (sonographer) could be connected to a wired network (in clinic) or a wireless network (in ambulances) to guide an untrained person (follower) at a remote community connected to a wired or wireless network.

An expert (sonographer) remotely "tele-operates" an untrained person (the follower) wearing a mixed reality headset by controlling a virtual ultrasound probe projected into the person's scene. The follower matches the pose and force of the virtual device with a real probe. The pose, force, video, ultrasound images, and 3-dimensional mesh of the scene are fed back to the expert in real time. In this control framework, the input and the actuation are carried out by people with near robot-like latency and precision. This allows teleguidance that is more precise and fast than verbal/video guidance, yet more flexible and inexpensive than robotic teleoperation.

4.2. Current Challenges

Though some robotic tele-ultrasound (US) system has been used in clinical trials, commercial success has been limited despite the robots' ability to provide precision, low latency, and haptic feedback. This is likely due to time consuming set-up, restricted workspaces, cost and complex maintenance and operation. The cost is notably significant when contrasted to relatively affordable US devices making such systems difficult to deploy in small communities. Despite this, a large body of literature has studied autonomous robotic US-using force-based positioning, depth camera-based planning , and reinforcement learning. However, guaranteed robustly safe human-robot interaction is an issue, particularly for regulatory bodies, and robotic tele-US remains relatively impractical.

On the other hand, there are several commercially available video conferencing-based mobile systems. Butterfly Network, Clarius Mobile Health Corp., and Philips use a point of care US (POCUS) device with live imaging and video conferencing available via a cloud interface on a mobile phone or tablet. Some visual guidance can be given by overlaying arrows or pointers on the US image. Though accessible and inexpensive, these systems are designed rather for quick expert review of a capable sonographer's captured images. The resulting interaction is thus very inefficient leading to low precision and high latency.

The existing options are neither flexible and accessible nor accurate and efficient. However, recent advances in extended reality (XR) research may solve this issue. There are many definitions and classifications pertaining to the approach of augmenting the user's view, including Milgram and

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Kishino's "reality-virtuality continuum" , and more recent additions. We refer to our system as mixed reality (MR) in comparison to AR where overlays are applied to videos. In MR visual guides can be located within the real environment itself through the use of optically transparent headsets and waveguides. The ability of MR to project 3D visual information seamlessly into the real world is the key enabling technology in a new concept we call "human teleoperation", introduced in , which leverages MR, haptics, and high-speed communication to bridge the gap in remote guidance techniques.

In this system, a human follower is controlled as if they were a flexible, cognitive robot through an MR interface. In this way, both the input and the actuation are carried out by people, but with tight coupling, leading to latency and precision more similar to a tele-robotic system. This enables remote guidance that is more intuitive, accurate, and efficient than existing audiovisual systems, yet less expensive, more accessible, and more flexible than robotic teleoperation. To accomplish this, the communication system must have a high throughput, low latency, and be adaptable to various networks and signal conditions. The visual control system should be user friendly for the follower and lead to good accuracy and little lag. Similarly, the expert should have a sensation as close as possible to carrying out the procedure in person, called teleoperation transparency, which involves visual, positional, and force feedback.

4.3. Human Teleoperation System Architecture

Figure 8 shows the overview of human teleoperation system. Traditionally in teleoperation systems, there is a local agent, often a robot, which interacts with its environment, and a remote operator who receives feedback from the local agent and provides instructions on what actions to carry out.

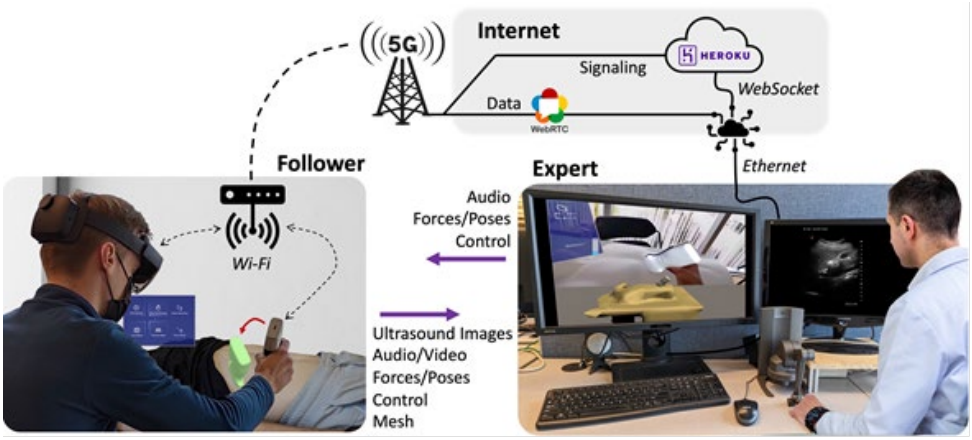


Figure 8- Overview of Human Teleoperation system

In human teleoperation, the local agent is a human, the "follower", while the remote operator is experienced in the task being performed, and is referred to as the "expert". For teleultrasound, the expert is a sonographer or radiologist with ultrasound experience, and the follower is an inexperienced novice, whose interactions with his/her environment constitute moving an US probe on a patient, as instructed by the expert.

The follower wears an MR headset, the Microsoft HoloLens 2, which projects a virtual US probe into their field of view. To perform the desired procedure, they align their real US probe with the virtual one

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

and track it as it moves around on the patient. In this way, the desired position and orientation (pose) of the probe are achieved. The desired force is similarly reached through a visual control system.

The desired pose and force of the probe are set in real time by the expert, who manipulates a haptic device (Touch X, 3D Systems, Rock Hill, SC). This is a small, 6-degree-of-freedom (DOF) serial manipulator that measures the 6-DOF pose of its pen-like handle and applies 3-DOF forces to the handle's tip for haptic feedback. As the expert grasps the end-effector and carries out his/her desired motion, the probe pose is transmitted to the virtual handle guiding the follower. The haptic device applies forces back to the expert. By pushing against them, the expert also inputs his/her desired force.

One of the primary objectives of the system is teleoperation transparency, or making the expert feel as if they are performing the procedure in person by matching the expert and follower positions and forces. To this end, a three or four-channel architecture is required, in which force and velocity are sent from expert to follower and vice versa at a high rate. Hence, the desired forces and poses are sent from expert to follower, along with an audio stream and some control packets. Conversely, the measured force and pose of the follower are returned to the expert, along with an audio stream, a video stream captured by the HoloLens which includes the virtual objects in place (called an MR capture), the live ultrasound images, control packets, and occasional mesh data.

In this way, the expert sees the patient, ultrasound probe, and ultrasound images live, and feels the applied forces. She/he consequently decides how to move and performs the motion on the haptic device, which updates the input signal to the follower through the visual control system, which the follower tracks. At the same time, the expert and follower are in verbal communication. Preliminary experiments with carrying out a remote ultrasound examination by a novice follower when teleoperated by an expert have shown promising results.

4.4. Communication network requirements and performance results

The throughput requirements on the follower side are outlined in Table 4. As the expert is stationary and attached via a wired connection, their throughput is of less concern.

Table 4. Worst-case uplink and downlink data throughput requirements on the follower side

Type	Uplink	Contents	Downlink	Contents
Timing	1.28 Kbps	64 bit long int × 20 Hz	3.84 Kbps	3 64 bit long ints × 20
Force	16 Kbps	3 32 bit floats + 1 64bit timestamp × 100 Hz	16 Kbps	Same as uplink
Pose	28.8 Kbps	7 32 bit floats + 1 64 bit timestamp × 100 Hz	28.8 Kbps	Same as uplink
Video	≈1-2 Mbps	960 × 540 px H.264 encoding, 25 Hz, Variable quality	0	No
Audio	128 Kbps	Typical MP3 bitrate (Part of MPEG-4 stream)	128 Kbps	Same as uplink
US	4.64 Mbps	58 KB JPEG image (worst-case) × 10 Hz	0	No US
Mesh	2.3 Mbps	≈12k mesh triangles × 3 points and 3 indices × 32bit floats	0	No downlink mesh
Total	6.81 Mbps	Mesh sent rarely on demand. Peak throughput 9.11 Mbps	180 Kbps	Sum

The follower side is mobile and could be installed anywhere, including a moving ambulance or a remote location, it is designed to work over mobile networks. A tight coupling between expert and follower is required. The communication system was designed to minimize latency using the WebRTC (Web Real Time Communication) standard. As shown in Figure 9, WebRTC is a peer- to-peer architecture which eliminates server-related delays. It is built on stream control transmission protocol and real time transport protocol, both of which are related to the user datagram protocol (UDP) which prioritizes speed over reliability.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

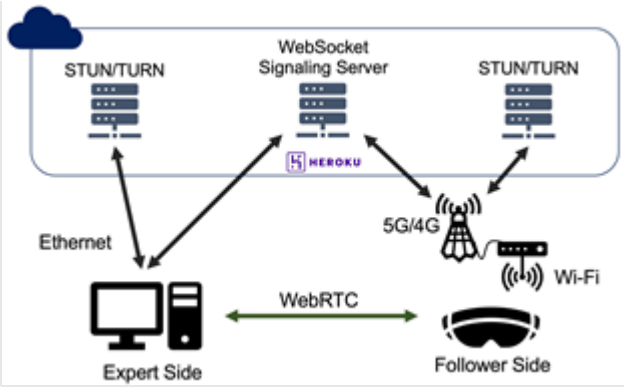


Figure 9- Overview of communication system

WebRTC is built upon Datagram Transport Layer Security (DTLS), so all communication is encrypted and secure, which is important for medical applications.

We carried out tests of the communication performance over WiFi, LTE and 5G networks, in different signal conditions. LTE and 5G tests were carried out from our public network with best effort QoS. Some of the key results are shown in Table 5. We found that the communication performed sufficiently well over 4G or 5G with good signal conditions up to about 5 Mbps continuous throughput, and well beyond this for WiFi or Ethernet. Typical latencies for 5G were 25 to 50 milliseconds which is appropriate for teleoperation with direct force feedback in the case of a relatively soft contact environment like a patient. In future with 5G standalone (SA) network deployments, Ultra-Reliable Low Latency Communication (URLLC) would further enable reduced latencies and a more reliable communication link which is necessary for real-time teleoperation scenarios.

Table 5. Latency(round trip) measurements with different network technologies

Throughput	Ethernet(ms)	WiFi(ms)	LTE(ms)	5G NR(ms)
1.28 Kbps	1.07±0.57	5.80±3.30	38.41±6.63	26.95±7.72
46.08 Kbps	0.94±0.61	5.90±2.77	38.77±7.92	27.67±6.21
2.17 Mbps	0.93±0.59	5.87±1.75	43.49±30.18	39.61±6.14
4.97 Mbps	-	-	52.30±59.30	47.64±22.81

The expert is presented with a live video stream from the follower's perspective, and the ultrasound images of the patient. Based on these, he/she decides how to move. However, the follower coordinate frame is positioned by the HoloLens and varies every time the application is started, even in the same location. Therefore, a registration between the expert and follower spaces is required to ensure that directions in both frames correspond intuitively.

A small sensing element embedded on the probe measures the field from an electromagnetic transmitter to determine the 6-DOF pose. To compare this to the expert's desired pose, the reading must be transformed from the transmitter frame to the HoloLens frame.

Furthermore, the HoloLens 2 continuously captures a 3D spatial mesh of its environment for SLAM purposes. This mesh can be accessed and sent to the expert side to provide 3D visual feedback as well as a haptic interaction.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

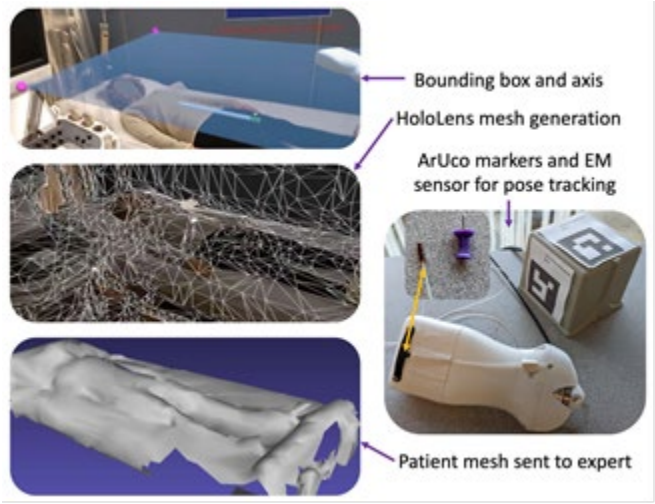


Figure 10- Aspects of expert-follower spatial registration

To extract the mesh of only the patient, not the rest of the room, and to enable the spatial registration, the position and orientation of the patient and electromagnetic transmitter must be determined. The ArUco markers are placed in four corners of the test surface, as shown in Figure 10. The HoloLens automatically determines the poses of the markers and thus finds the patient bounding box, orientation, and the transmitter transform.

4.5. Summary from Human teleoperation

The convergence of augmented reality/virtual reality (AR/VR) and 5G technologies presents a promising possibility for performing ultrasound teleoperation, revolutionizing medical procedures and remote healthcare. This paper has described a novel system for teleoperating a human with precision and latency similar to that of a robot through a mixed reality interface. With this technology, sonographers can perform ultrasound examinations remotely and can guide a novice in real time. This is a crucial service for remote areas where the absence of qualified sonographers could result in simple ultrasound procedures taking several weeks to complete. Monetization avenues arise through offering specialized telehealth packages for clinics and patients, who can benefit from expert consultation without geographical constraints. Additionally, AR/VR and 5G can enhance medical training by enabling trainees to practice ultrasound procedures in a virtual environment, fostering skill development and reducing the need for physical equipment. Ultimately, the integration of AR/VR and 5G technologies in ultrasound teleoperation holds immense potential to improve healthcare accessibility, enhance diagnostic capabilities, and advance medical education.

5. Conclusion

Advancements in 5G, sensor technologies, and augmented reality/virtual reality (AR/VR) are poised to enable transformative use cases across various domains. AI-powered systems can leverage sensor data from cameras, radar, lidar, and other sources to analyze real-time traffic patterns, predict potential collisions, and optimize traffic flow. Traffic management had not changed much for over 50 years. Usually, Traffic lights are timed on a fixed grid, disconnected, not dynamic and cannot react in real-time. With our proof of concept implementation at UBC, we have demonstrated that this can not only safe lives but also brings economic benefits in the savings of fuel costs, emissions and improved productivity and

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

cost savings for businesses and individuals. The monetization potential arises from collaborations with municipalities and traffic management agencies by digitizing the intersections and generating useful data. The advanced data analysis from the intersections can provide real time traffic and accident insights which can potentially help municipalities to monetize with the insurance companies.

5G and AR/VR technologies has the potential to revolutionize multiple industries, offering immersive experiences, real-time connectivity, and enhanced productivity. In healthcare, 5G and AR/VR facilitate remote consultations, surgical assistance, and medical training. Surgeons can use AR/VR to visualize patient data and perform complex procedures with enhanced precision. In this project, we have demonstrated a successful teleoperation which could transform the lives in rural areas. Monetization potential arise through offering specialized telehealth packages for clinics and patients, who can benefit from expert consultation without geographical constraints.

Overall, the high-speed, low-latency and massive device connectivity offered by 5G will unlock new possibilities, drive innovation, and improve efficiency across various sectors, paving the way for a more connected and technologically advanced future.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

ADC	analog-to-digital converter
AI	artificial intelligence
AR	augmented reality
C-V2X	cellular vehicle-to-everything
DOF	degree-of-freedom
DSRC	dedicated short-range communication
DTLS	datagram transport layer security
eMBB	enhanced mobile broadband
FR1	frequency range 1
FR2	frequency range 2
GPIO	general-purpose input output
IOT	internet of things
M2M	machine-to-machine
MEC	multiaccess edge computing
MR	mixed reality
NF	network functions
QoS	quality of service
RAN	radio access network
RSU	road-side unit
SBA	service-based architecture
SLAM	simultaneous localization and mapping
SPI	serial peripheral interface
UDP	user datagram protocol
UPF	user plane function
URLLC	ultra reliable and low latency communications
US	ultrasound
VMC	virtual management center
VR	virtual reality
XR	extended reality

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

*System architecture for the 5G System (5GS),*3GPP : 23501.i10

Rogers NoTraffic Implementation report at Aurora testbed at UBC , Feb 2023 Justin Healey

Human teleoperation - a haptically enabled mixed reality system for teleultrasound. Human Computer Interaction, 2022 D. Black, Y. O. Yazdi, A. H. H. Hosseinabadi, and S. Salcudean.

Human-as-a-robot performance in augmented reality teleultrasound. TechRxiv, Oct 2022. doi: 10. 36227/techrxiv.21432012.v1 D. Black, H. Moradi, and S. Salcudean

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Converging User Planes in Modern Networks: A Story of GUPpies

A Technical Paper prepared for SCTE by

Jennifer Andreoli-Fang, PhD
Senior Manager, Cloud Architecture
AWS
nextjen@amazon.com

John T. Chapman
CTO Broadband Technologies, Cisco Fellow
Cisco
jchapman@cisco.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Separating the Control and User Planes ... 3
   2.1. 5G Network Architecture ... 3
   2.2. 5G CUPS ... 4
   2.3. CUPS in Other Access Networks ... 5
   2.4. Motivation for CUPS ... 6
3. The Protocols Behind CUPS ... 7
   3.1. PFCP ... 7
   3.2. Remote PHY vs. Mobile RAN ... 7
   3.3. Converging the User Planes ... 8
   3.4. GUP – Generic User Plane ... 9
   3.5. P4 ... 10
4. Practical Use Cases ... 12
   4.1. CMTS Disaggregation ... 12
   4.2. Cloud Primer ... 12
   4.3. CUPS and the Cloud ... 13
5. Business Considerations ... 15
6. Conclusion ... 16

Abbreviations ... 16

References ... 17

List of Figures

Title **Page Number**

Figure 1 – Key 5G Core Network Functions ... 4
Figure 2 – 5G CUPS Simplified ... 5
Figure 3 – Disaggregated BNG ... 6
Figure 4 – Packet Processing Workflow in UPF ... 7
Figure 5 – Remote PHY vs. Mobile RAN ... 8
Figure 6 – Convergence via W-AGF ... 9
Figure 7 – Truly Converged User Planes ... 10
Figure 8 – NPU Programming Model ... 11
Figure 9 – vCMTS Functional Blocks ... 12
Figure 10 – Cloud Infrastructure ... 13
Figure 11 – Partial Offload – CP in the Cloud ... 14
Figure 12 – Full Offload – Disaster Recovery in the Cloud ... 15

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Many modern broadband networks today are built with three areas of enhancement:

1. containerized implementation,
2. serviced-based architecture (SBA), and
3. control and user plane separation (CUPS).

5G networks are an example network that have all three traits. DOCSIS networks, on the other hand, have yet to define and benefit from SBA and CUPS. This paper focuses on CUPS. We discuss why it is useful and how to implement it. We also cover practical use cases that CUPS can enable.

While we focus on DOCSIS, PON, and 5G networks, the ideas presented in this paper are extendable to Wi-Fi and fixed wireless access (FWA). These ideas build upon the authors' previous white paper on cable and mobile convergence [1].

2. Separating the Control and User Planes

Communication service providers in our industry operate many access networks:

- DOCSIS networks, where the cable modem termination system (CMTS) is the subscriber termination system
- 5G networks, where the 5G core is the subscriber termination system
- Fiber-based passive optical networks (PON), where the BNG (broadband network gateway) is the subscriber termination system
- Wi-Fi and fixed wireless access (FWA), where the BNG is also the subscriber termination system

DOCSIS and 5G were designed by different standards groups. Some of the key differences are: 5G needs to provide mobility management, an unnecessary functionality in DOCSIS. 5G establishes and tears down user sessions, whereas DOCSIS maintains an always-active primary service flow. Most DOCSIS secondary service flows are also active even though the specification allows the ability to dynamically establish, change, and delete them.

The user authentication processes differ significantly. Different terminologies are also adopted. For example, the term "user plane" in 5G is equivalent to "data plane" in DOCSIS. While most of the 5G implementations are containerized, DOCSIS implementations are on the path to containerization.

DOCSIS and 5G are both access technologies, one over the wire, and one over air. While different, there are also many similarities. In the following, we will start by providing a short overview of the 5G network architecture while pointing out similarities to DOCSIS to help the reader orientate.

2.1. 5G Network Architecture

5G has a flat and highly disaggregated architecture. The core network consists of a set of building blocks, or network functions (NFs), offering services needed to execute the functionality of a mobile network. These NFs are interconnected via the "service-based architecture" (SBA) [1].

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Prior to 5G, network functions talked to each other via point-to-point interfaces standardized by the 3GPP. The idea behind SBA is to migrate from telecom-style protocol-leaden interfaces to standard web-based application programming interface (API). With SBA, network functions can register themselves and subscribe to each other's services. Doing so allows operators to deploy new services without waiting for standards bodies to define and vendors to develop new protocols. 5G control plane functions are connected via HTTP/2 from the IETF, with JSON as data format, while user plane functions use the classic point-to-point protocols.

Key functional blocks for 5G, as shown in Figure 1, are:

- Access and mobility management function (AMF) – controls the radio access network (RAN) and the connections to the user equipment (UE), as well as UE mobility management
- Session management function (SMF) – manages user sessions and user plane functions
- User plane function (UPF) – handles packet processing, traffic aggregation, and tunneling
- Unified data management (UDM) – manages and provides network user data to the services that request it
- Authentication server function (AUSF) – handles authentication process for a subscriber
- Policy control function (PCF) – works with the SMF and supports policy and charging
- Radio access network (RAN) – last mile wireless network to the user

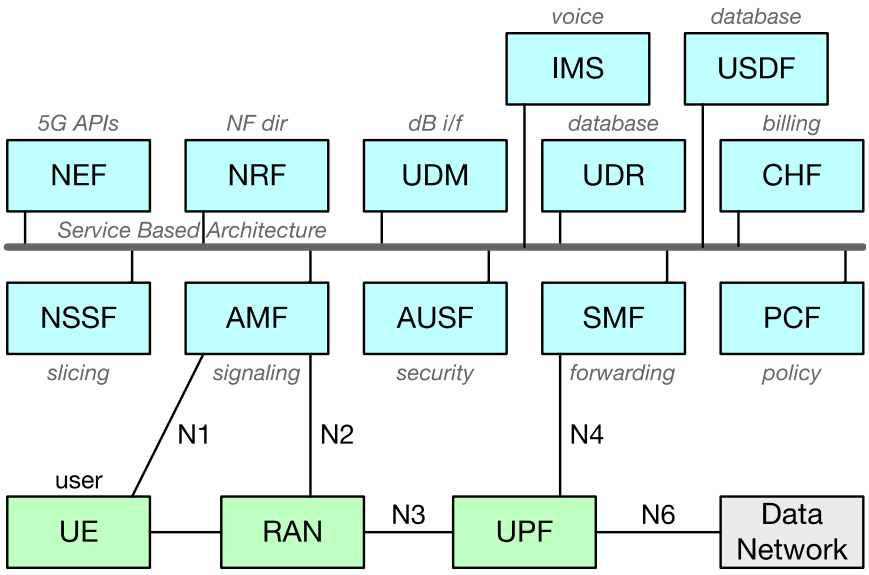


Figure 1 – Key 5G Core Network Functions

2.2. 5G CUPS

Control and user plane separation, or CUPS, was first introduced in 4G / LTE and defined in 3GPP Rel-14. The idea was to separate the packet gateway into control and user planes to allow for resource sharing, flexible deployment, and independent scaling to achieve benefits in CapEx and OpEx. CUPS is now required for 5G standalone architecture.

Figure 2 shows the 5G network architecture simplified with CUPS. The control point to access the 5G user plane across N4, the UPF, is the SMF. Multiple SMFs can share the same UPF and a SMF can interface

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

with multiple UPFs. More importantly, the UPF is mostly protocol agnostic – it just needs to support a 3GPP packet tunneling protocol (N3) and a control protocol (N4) that it uses to talk to the SMF. The SMF is the network function that interfaces with other 5G control plane functions.

The 5G control plane functions including the SMF support diverse and complex set of functions all related to the essential functions of the mobile network and subscribers, while leaving the user plane to handle tasks solely related to packet handling. The UPF specializes in packet processing that is not unique to 5G, but also to other communication networks. This opens the possibility for the UPF to be reused in a multi-access network scenario. We will expand on this point later in the paper.

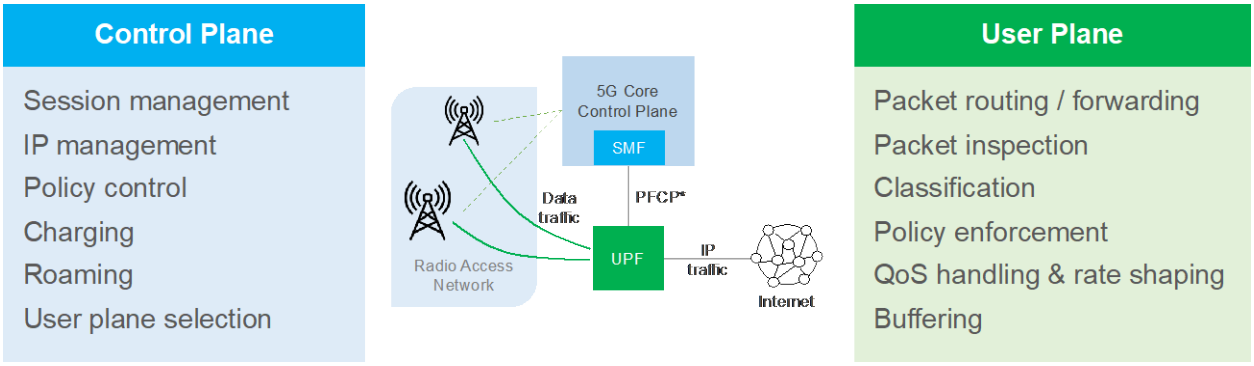


Figure 2 – 5G CUPS Simplified

2.3. CUPS in Other Access Networks

Many cable operators also operate passive optical network (PON) for their fiber network. The broadband network gateway (BNG) is the subscriber termination system in PON. BNG is specified by the Broadband Forum (BBF), and has gone through several iterations, from monolithic architecture, to virtualized BNG that leverages commodity hardware in lieu of proprietary hardware, to disaggregated BNG (DBNG) that is based on the CUPS principle [3].

Figure 3 shows DBNG architecture. The DBNG control plane (CP) provides subscriber management functions including authentication, authorization, and accounting (AAA), IP address management, and policy. The DBNG CP interfaces with external management services with the northbound control interfaces. The CP is also responsible for programming data traffic forwarding rules for the DBNG UP. CUPS enables the option of having a CP function to control multiple user plane (UP) functions. This allows the CP can have the ability to steer a subscriber session to a particular UP. The steering function enables this functionality and is currently being defined by the BBF.

There are strong similarities between the DBNG and the 5G core architectures. The steering function is a combination of the 5G's AMF and SMF. The information flow between the DBNG CP and UP adopted and extended two protocols standardized by the 3GPP: packet forwarding control protocol (PFCP) and general packet radio services (GPRS) tunneling protocol user plane (GTP-U)[1]. The BNG architecture is

[1] Note that GTP-U in DBNG architecture is used to tunnel the control packets between the CP and CPE with the UP as the intermediary. In 5G, the control packets between the CP (AMF) and CPE (UE) are carried on the N1 interface.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

also set to evolve into a new network function, the aggregation gateway function, or AGF (to be covered later) to support wireless-wireline convergence.

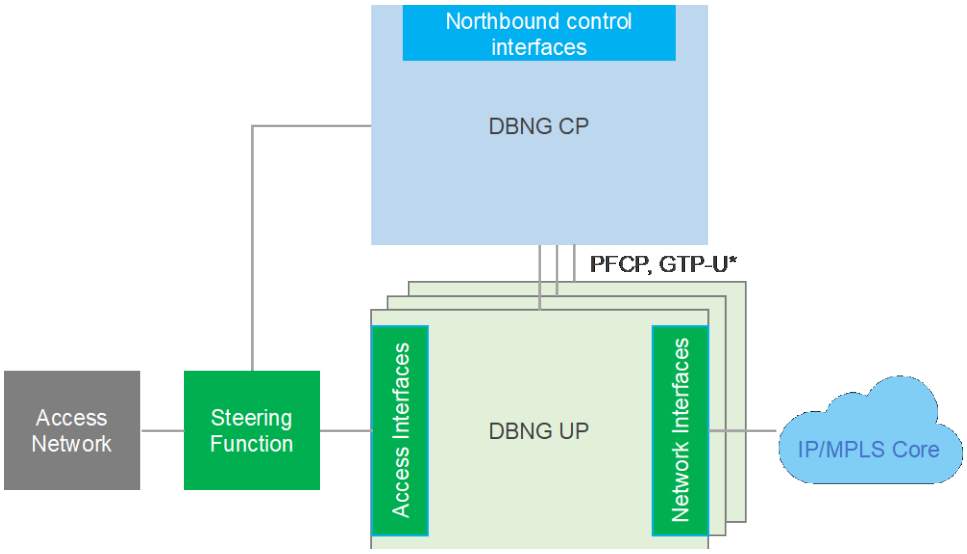


Figure 3 – Disaggregated BNG

2.4. Motivation for CUPS

Many argues that a key benefit of 5G CUPS is that it allows the network operator to distribute resources throughout the network, with the control plane (CP) sitting in a centralized location and the user plane (UP) closer to the application.

More importantly, CUPS architecture provides several benefits in operating a modern network.

Operational considerations: CUPS allows control and user plane software to be updated independently. Considering the complexity of the control plane and the multitude of functionalities it supports, the CP should be updated frequently. Operators can leverage DevOps techniques to update the CP. On the other hand, user plane software updates potentially create network downtime and outages, and ideally should be updated infrequently. By decoupling the control and user planes, operators will have more system uptime and results in better high availability.

Implementation and deployment optimization: The Remote PHY architecture is essentially a DOCSIS radio in the fiber node with the CMTS core software in the hub. With CUPS, the DOCSIS control plane can be deployed in the hub or in the cloud, while the DOCSIS data plane can remain in the hub. The DOCSIS data plane can further be implemented as dedicated central processor unit (CPU) or network processor unit (NPU) cores.

Enabling rapid adjustment of compute and network resources: Unplanned outages or expected capacity bursting require operators to react quickly to deal with massive surge of control and/or user plane traffic. An area outage can cause a large number of cable modems (CM) to reboot and re-register at the same time. In the case of a registration storm, only the control plane is impacted. CUPS enables the independently

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

scaling of the CP – in this case, the registration function only, without the need to scale the UP. Another example is busy hour or holiday-related uptick in capacity demand that requires the increase in UP capacity without necessarily increasing the capacity of the CP.

3. The Protocols Behind CUPS

3.1. PFCP

The packet forwarding control protocol (PFCP) is the session layer protocol that runs on the 5G's N4 interface between the SMF and UPF, and is used by SMF to associate with and configure the UPF. An SMF in 5G can control multiple user plane functions. The SMF needs to select and associate with a UPF. The SMF configures a user session by sending packet detection rules (PDRs) for each session. Figure 4 illustrates what happens when a GTP-U encapsulated packet is received by the UPF and eventually egresses as an IP packet to the data network.

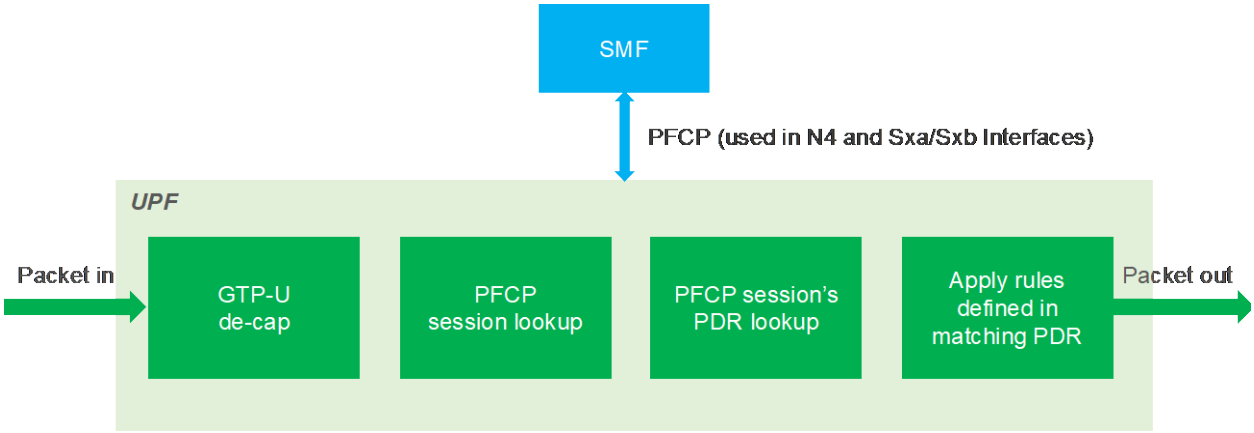


Figure 4 – Packet Processing Workflow in UPF

3.2. Remote PHY vs. Mobile RAN

The Remote PHY architecture was invented by John Chapman of Cisco in 2001 [4] [10] [11] [12] and was originally part of the Modular Headend Architecture (MHA). The main protocols of Remote PHY are:

- **R-DEPI**, the remote downstream external PHY interface, a pseudowire that connects the CMTS Core to the RPD [5],
- **R-UEPI**, the remote upstream external PHY interface, a pseudowire that connects the RPD upstream to the CMTS Core [6],
- **GCP**, the generic control plane, that is the management protocol for the RPD [7], and
- **R-DTI**, the remote DOCSIS Timing Interface, that uses IEEE 1588 to manage the DOCSIS timestamp between the CMTS Core and the RPD [8].

DEPI, UEPI, and DTI were written in 2004. GCP was written in 2012 at a Pete's Coffee shop. DEPI and DTI were standardized as part of MHAv1 at CableLabs in 2005 [13]. UEPI and GCP, along with a "Remote" prefix, were standardized at CableLabs as part of CDOCSIS in 2014 [14] and MHAv2 in 2015 [15].

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Remote PHY is a fundamental enabler of the DOCSIS CUPS architecture. Remote PHY disaggregated the integrated CMTS into MAC and PHY components. The MAC is placed into a CMTS Core and can be virtualized [9]. The PHY is placed into a remote PHY device (RPD) that is located in the HFC fiber node.

The RPD and node together act as a "DOCSIS radio" and are analogous to the mobile radio access network (RAN). In fact, the Remote PHY protocol was the influence for the development of the Small Cell Forum (SCF) network functional application programming interface (nFAPI) split 6 protocol [16].

In 5G, the RAN is managed by the AMF. In Remote PHY, the RPD is managed by the principal core. There is much more to both stories, but the analogy holds true. The vCMTS contains the Remote PHY tunnel generation and termination as well as the DOCSIS MAC. As a containerized virtual implementation, the DP and CP should be separate containers and could be located on different servers, or even as diverse as moving the CP to the regional or far edge cloud.

Today, the protocols between the DOCSIS DP and CP are proprietary and internal to each CMTS manufacturer. PFCP should make an excellent baseline as a protocol, just as it has for mobile (4G and 5G) and PON (BNG). There will likely be extensions required for DEPI and UEPI, and for any DOCSIS functionality that does not map directly into existing PFCP.

Note that the DOCSIS upstream scheduler is a real-time function that would likely stay with DOCSIS DP or stay close by. This is also true in mobile and PON where the upstream scheduler stays close to the RAN.

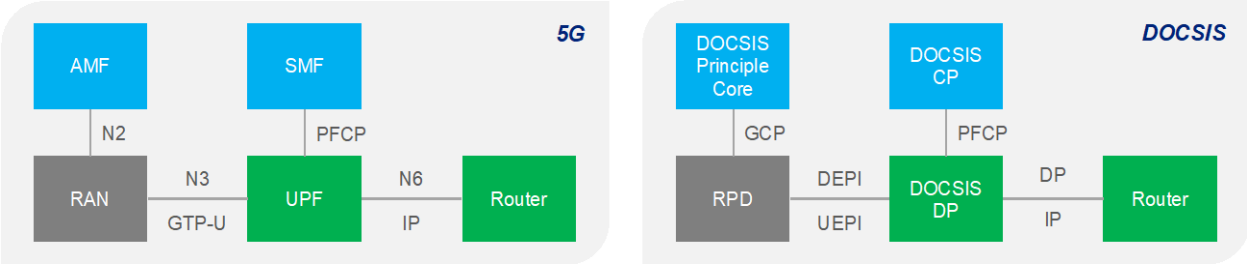


Figure 5 – Remote PHY vs. Mobile RAN

3.3. Converging the User Planes

If DOCSIS can be implemented with a CUPS architecture that is very similar to 5G and PON/BNG, then could the networks be converged? Convergence does not mean that all the protocols must be changed, but it does imply reuse of major blocks, with the intent of achieving a common management and provisioning system.

One early example of convergence was the use of DOCSIS Provisioning over EPON (DPoE) [17] that reused a DOCSIS provision system to provision a passive optical network (PON) system. DPoE specified a signaling gateway that converted between the two worlds.

The 5G architecture defines the access gateway function (AGF) that connects a wireline termination system to a 5G core. Jennifer Andreoli-Fang, while at CableLabs, defined a cable version of this called the cable fixed access gateway function (CFAGF) in the 3GPP technical report on wireless-wireline convergence [20]. CFAGF was later renamed as wireline AGF (W-AGF) in the specification stage [21].

CableLabs adopted the W-AGF terminology in the converged core technical report (TR) [22], where the W-AGF was shown to connect an integrated CMTS to a 5G core. The role of W-AGF in a Remote PHY system was not discussed in the CableLabs TR. However, just as in the I-CMTS case, W-AGF can connect

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

R-PHY to a 5G system with an additional functionality of converting between DEPI/UEPI and GTP-U. This is all shown in Figure 6.

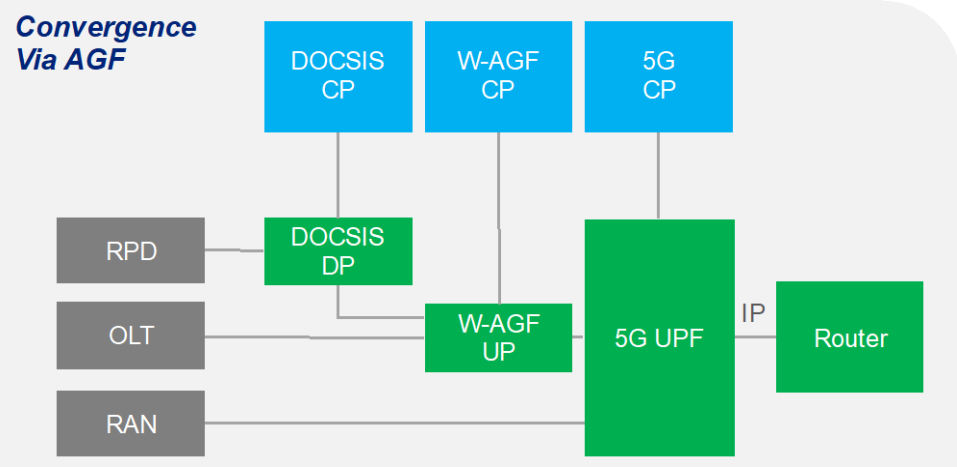


Figure 6 – Convergence via W-AGF

Care must be taken in the interpretation and implementation of this functionality. The W-AGF as defined in 5G connects the DOCSIS and 5G worlds together though a gateway. Without any optimization, the result may end up with two completely separate data planes that are now interconnected. Such an interconnection couples the two data planes and can make the combination redundant and more complicated that two separate systems, without achieving cost reduction and simplicity goals. For example, there are now two quality of Service (QoS) and service level agreement SLA control points and data planes.

3.4. GUP – Generic User Plane

A more pragmatic implementation would be something described in Figure 7. The DOCSIS RPD, the PON OLT, Federated Wi-Fi, and the 5G RAN would all send their traffic to a generic user plane (GUP). The GUP would have a pseudowire function (PWF) that terminated and generated pseudowires of interest, specifically:

- GTP-U for mobile
- VLAN over L2TP/PPPoE for PON
- DEPI/UEPI (L2TPv3) and DOCSIS Framer for DOCSIS
- L2TP (or equivalent) for Federated Wi-Fi

It is worth noting that Remote PHY uses a pseudowire called L2TPv3 that is derived from the BNG pseudowire L2TP. L2TPv3 is the base protocol for DEPI and UEPI.

The interface between the pseudowire function (PWF) and the packet services function (PSF) would be an IP over Ethernet packet coupled with context information containing metadata like a generic service identifier (GSID) and port interface IDs.

It would then apply a common set of subscriber management processes and QoS tools. In DOCSIS, this would be packet classification into service flows and the associated rate shaping. There are equivalently defined operations in mobile and PON. The key is to have this all done once, on the same hardware and software user plane, not a separate operation per service domain.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

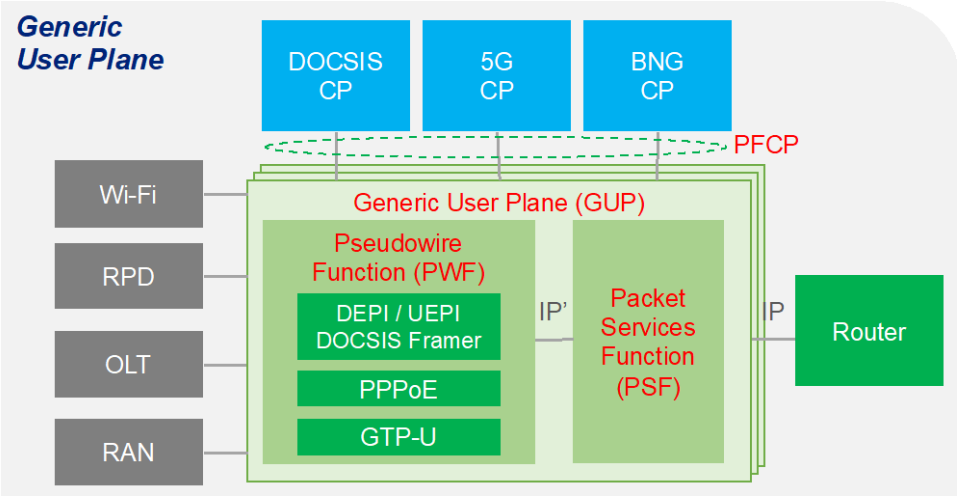


Figure 7 – Truly Converged User Planes

The GUP function could be a larger centralized model or a distributed model. And yes, a distributed GUP system would be a bunch of GUPpies.

We have spoken strictly about the user plane in this study. There are also incredible opportunities in the control plane and management planes to converge workloads even though the access protocols may differ. But this is for another whitepaper. Meanwhile, let's go down one more level on the user plane implementation.

3.5. P4

P4 stands for "Programming Protocol-Independent Packet Processors". P4 is a programming API for programming a network processor unit (NPU).

What is an NPU? Let's compare to what we know. Central Processing Units (CPU) are used to run generic software and are programmed with a machine specific assembly language. Graphical Processing Units (GPU) are massively parallel processors that specialize in vector operations and can be used for graphical rendering, and more recently bitcoin mining and artificial intelligence. They also have a machine specific programming interface. An NPU is used for switching packets. Routers have NPUs. An NPU can be implemented in a hardware application-specific integrated circuit (ASIC) or as a software core on an CPU. Hardware NPUs typically had a proprietary machine specific programming language. P4 changes that by standardizing the interface to an NPU.

PFCP is a layer 3 control plane protocol that is intended to be run between two software stacks on two different network elements. The GUP software stack would convert between PFCP and P4. This is shown in Figure 8.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

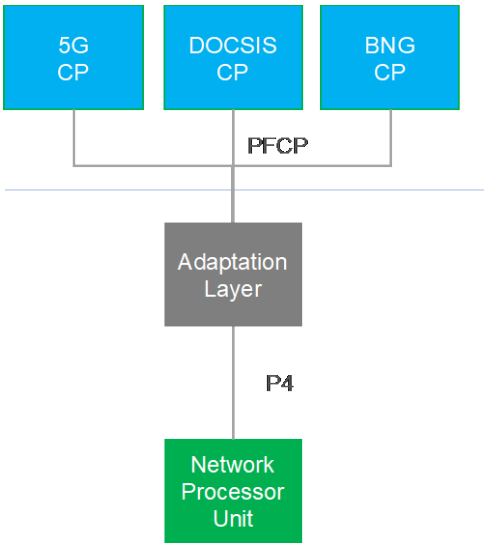


Figure 8 – NPU Programming Model

P4 is agnostic to all protocols. P4 works on bytes and tables. P4 allows fields to be defined within a packet. Fields can be extracted and inserted, and those fields can used as indexes into tables and be retrieved from tables.

For example, in P4, you could define a packet type called “Ethernet”. The first six bytes would be the “Destination Address”, the next six bytes would be the “Source Address”, and the next two bytes would be the “Type”. For layer 2 forwarding:

- a forwarding operation would be performed by telling the hardware to use the Destination Address as a lookup into the forwarding table.
- a learning operation would be performed by storing the “Source Address” as a new forwarding table entry along with the port number the packet came in on.

A router would use the Type field to understand what the next protocol is, such as an Internet Protocol (IP) packet, and P4 would have a description of what to do. When a layer 3 packet is forwarded, P4 would have a task to rewrite the Ethernet Destination Address.

Why is all this important?

It is important that all the functionality that is defined by PFCP must be implementable by P4.

And why is that also important?

Because once P4 can explain everything to an NPU in agnostic terms, then a common user plane becomes realistic. And once that user plane is common, then it could be converted into an ASIC to drastically increase performance, decrease costs, and distribute user planes across the IP network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. Practical Use Cases

As a bonus, the CUPS architecture enables the partial or full offload of the CMTS core to the cloud. We will start with a primer on CMTS disaggregation and the cloud, before describing the use cases.

4.1. CMTS Disaggregation

Virtual CMTS functions can be split into service domains based on timing requirements. Figure 9 is adopted from [19] and illustrates the functional blocks in the DOCSIS management plane, control plane, and data plane.

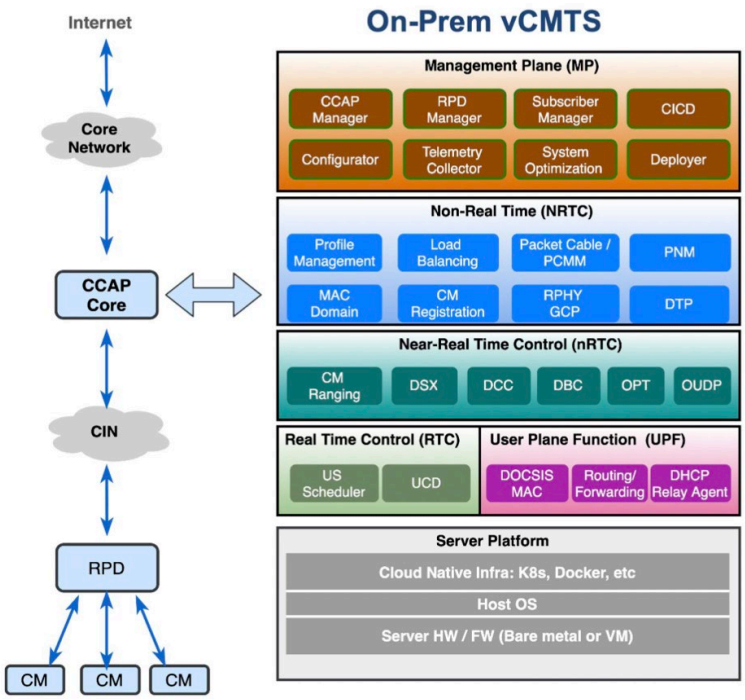


Figure 9 – vCMTS Functional Blocks

4.2. Cloud Primer

This section provides a primer on general cloud terms and concepts. To describe the use cases with sufficient detail, we will use some Amazon Web Services (AWS) specific terms as examples, although we will keep the terminologies agnostic whenever possible. Other cloud providers may have similar concepts and naming conventions.

Every communication service provider (CSP) is likely familiar with what cloud providers call the "region." It's best to relate a cloud region with a CSP's regional data center. Each region is typically consisted of multiple, isolated locations where data centers sit to provide a resilient infrastructure. But the cloud infrastructure is much more.

As shown in Figure 10, the cloud infrastructure extends from cloud region to cloud edge and to the cloud far edge to support applications with different timing constraints or local residency requirements. Cloud edge is typically located in major metro centers and the cloud far edge can sit in the communication service providers (CSP) facilities. In AWS, region, edge, and far edge are inter-connected and have the same "look

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

and feel" for cloud infrastructure services such as networking, compute, storage, and deployment automation.

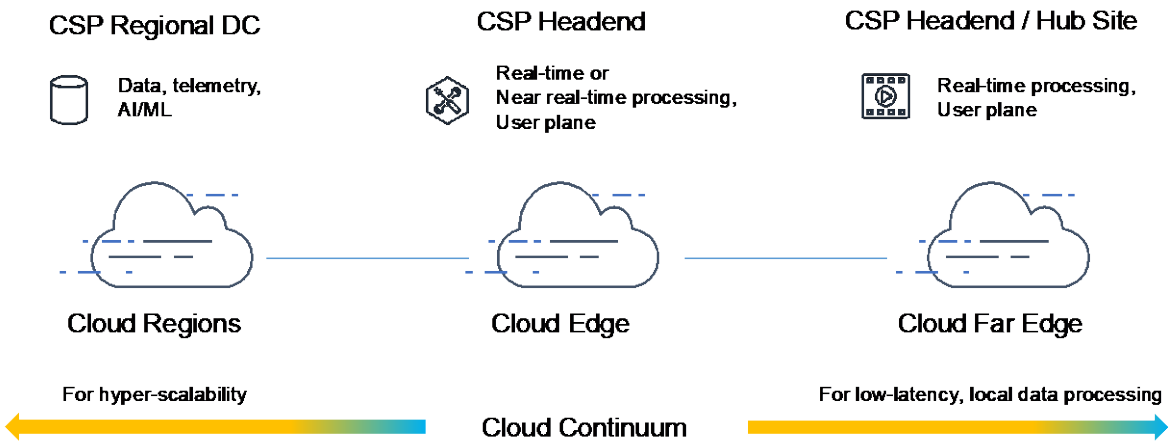


Figure 10 – Cloud Infrastructure

Putting containerized applications in the cloud allows an operator to leverage some of the fundamental benefits that come with the cloud.

- **Resiliency**: Due to its massive scale and design, the cloud infrastructure is inherently more resilient compared to a CSP's on-prem data centers.
- **Pay-what-you-use model**: With on-prem data centers, CSPs need time, CapEx, and OpEx to plan, build, and maintain data centers for peak capacity. Large amounts of resources planned for peak usage sit idle during normal operating conditions. With cloud, CSPs only pay for normal network capacity and pay for peak capacity only on the occasions when capacity demand surges.
- **Horizontal scaling**: Modern software applications are architected to enable horizontal scaling, or scaling out, which involves spinning up additional containers used for a functionality when demand arises. (In contrast, vertical scaling, or scaling up, refers to scaling by adding more power like CPU or memory.) Horizontal scaling and cloud go hand-in-hand. With cloud's near-infinite compute resources, applications can automatically scale out to deal with a surge in demand.

4.3. CUPS and the Cloud

This section focuses on DOCSIS CUPS use cases, although the same concept is extendable to BNG and mobile. DOCSIS management plane applications such as subscriber management or remote PHY orchestrator are well-suited to be deployed in the cloud due to their traffic and latency characteristics. CUPS enables a service provider to partially offload the vCMTS application to the cloud. As shown in Figure 11, the vCMTS control plane functions can be deployed in the region. When containerized, the CP can leverage the cloud's native ability for horizontal scaling to scale out when required in cases such as registration storm. CUPS allows the DOCSIS data plane to be deployed on-prem. A dedicated connection between CSP on-premise infrastructure and cloud region (shown in Figure 11 as AWS Direct Connect) carries control plane traffic, which can be PFCP.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

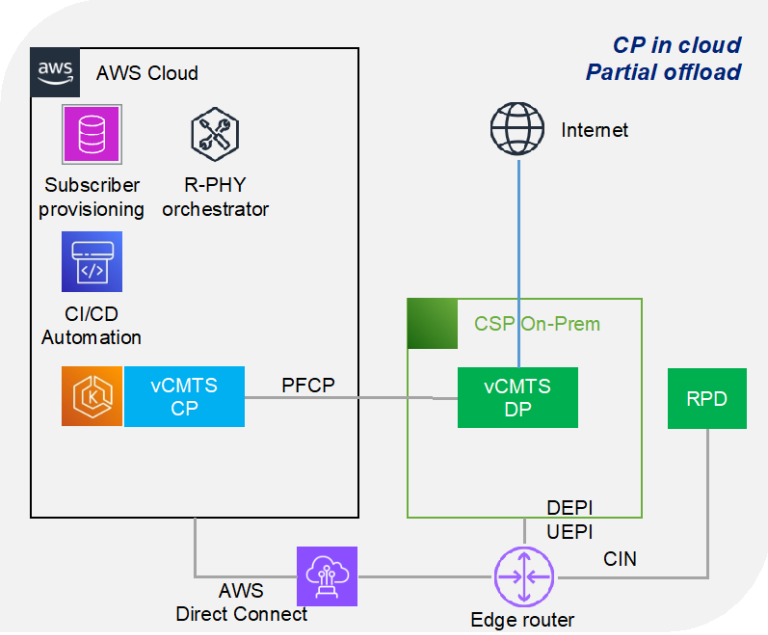


Figure 11 – Partial Offload – CP in the Cloud

In a temporary network outage, deploying the DOCSIS DP in the cloud can accelerate disaster recovery. Figure 12 shows the basic architecture where both the DOCSIS CP and DP run in the cloud. The DP may be run in the cloud edge such as a metro center to minimize latency from the outside plant. The dedicated connection between cloud region and on-prem would now carry DEPI and UEPI traffic. The ingress and egress point of DOCSIS data plane traffic to the internet is directly through cloud.

Another use case not shown is cloud-based capacity bursting. During normal operations, DOCSIS DP runs on CSP premise. When capacity demand increases, additional DP can spin up in the cloud. This way, CSPs don't need to build and operate their data center for peak capacity, where a large number of resources sit idle for most of the time.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

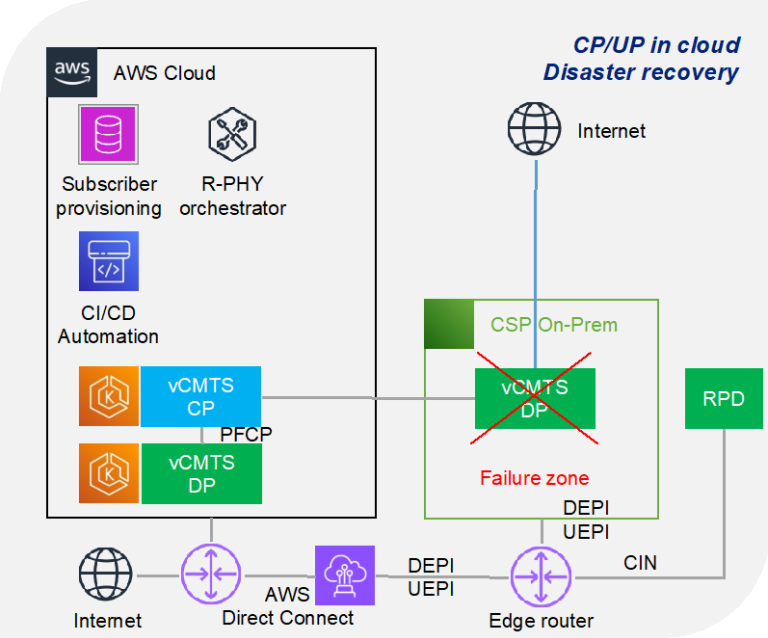


Figure 12 – Full Offload – Disaster Recovery in the Cloud

5. Business Considerations

As with all business decisions for adopting of new technologies, there needs to be a strong business case or the product may be technically brilliant but fail in the marketplace. There must be a return on investment (ROI). This means either an increase in revenue or a decrease of capital and/or operational costs for the operator.

The pro-side of convergence is a single system that does everything. There could be less redundant spend across multiple systems. There could be one team to manage that one system. The customer can seamlessly move between the different access systems.

The con-side of convergence is that by combining the systems together, the entire system becomes more complex. It is more complex in implementation, more complex to test, and more complex to update software. It requires cross domain knowledge just to operate in one of the access domains which make it more complex to staff.

There is a further deployment risk. The 5G architecture seems to be the most attractive to converge to, but it is also the least deployed architecture compared to PON and DOCSIS. 5G systems also have the least bandwidth requirements when compared to 10G/25G/100G links for PON and 10G links for DOCSIS. 5G also has the most complex feature set when it comes to roaming, hand-offs, deep packet inspection for flow discovery, and more.

A practical move to convergence should prioritize a move to simplicity, more common features, and less legacy access-specific features.

In this white paper, we took a simple approach of just converging the user plane. All user planes classify and forward IP over Ethernet packets that are contained in an access-specific encapsulation. This is a contained problem. An optimized user plane that could process packets at higher speeds and be usable across all access market segments should also be lower cost per bit.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In addition, a common user plane technology could be removed from the centralized access systems and moved into the network to allow for distributed user plane termination. This would then allow distribution of edge processing functions like media and application caches.

6. Conclusion

A generic user plane, GUP, enables scale. Scale of the business case. Scale of the actual NPU. A larger market for a common device. An option of a software-based or hardware-based user plane. The opportunity to move the user plane from gigabits to terabits and petabits of data throughput.

Convergence may not be an all-or-none proposal. It may be just one step at a time, and the data/user plane looks to be one of the first steps.

Abbreviations

AAA	authentication, authorization, and accounting
AGF	access gateway function
AMF	Access and mobility management function
API	Application programming interface
ASIC	Application-specific integrated circuit
AUSF	Authentication server function
BBF	Broadband network Forum
BNG	Broadband network gateway
CFAGF	cable fixed access gateway function
CM	Cable modems
CMTS	Cable modem termination system
CP	Control plane
CPU	Central processing unit
CSP	Communication service provider
CUPS	control and user plane separation
DevOps	Development and operations
DEPI	Downstream external PHY interface
DOCSIS	Data over cable system interface specification
DBNG	Disaggregated BNG
DPoE	DOCSIS provisioning of EPON
EVPN	Ethernet Virtual Private Network
FWA	Fixed wireless access
GCP	Generic control plane protocol
GPRS	General packet radio services
GTP	GPRS tunneling protocol
GTP-U	GTP user plane
GUP	Generic user plane (first defined in this paper)
IP	Internet protocol
LAN	Local area network
NF	Network function
NPU	Network processing unit

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

PCF	Policy control function
PDF	Packet detection rules
PON	Passive optical network
PSF	Packet services function (first defined in this paper)
PWF	Pseudowire function (first defined in this paper)
QoS	Quality of service
RAN	Radio access network
RPD	Remote PHY Device
SBA	service-based architecture
SMF	Session management function
TR	Technical report
UDM	Unified data management
UEPI	Upstream external PHY interface
UP	User plane
UPF	User plane function
VLAN	Virtual LAN
W-AGF	Wireline AGF

References

[1] Jennifer Andreoli-Fang, John Chapman, et. al, "Cable and Mobile Convergence: A Vision from the Cable Communities Around the World", *SCTE Cable-Tec Expo Fall Technical Forum*, Denver, Oct, 2020. [link]

[2] "System architecture for the 5G System (5GS)", 3GPP TS 23.501.

[3] "Multi-Service Disaggregated BNG with CUPS", BBF TR-459, 2020.

[4] John T Chapman, "Packet Fiber Node", US Patent 7,209,442, filed Jun 27, 2001, issued April 24, 2007. [link]

[5] John T Chapman, "Downstream Remote Physical Interface for Modular CMTS (DEPI)", US Patent 7,688,828, filed May 17, 2005, issued March 30, 2010. [link]

[6] John T Chapman, "Upstream physical interface for modular cable modem termination system (UEPI), US Patent 7,639,617, filed May 17, 2005, issued Dec 29, 2009. [link]

[7] John T. Chapman, "Generic Control Protocol (GCP)", US Patent 9246700, filed Sept 12, 2012, issued Jan 26, 2016. [link]

[8] John T. Chapman et. al, "Managing time offset and frequency drift in asynchronous docsis remote phy network environments (DTI)", US Patent 9,722,739, filed April 15, 2015, issued Aug 1, 2017. [link]

[9] John T Chapman, Alon Berstein, et. al., "Virtual Cable Modem Termination System", US Patent 9,369,339, filed Sept 30, 2013, issued Jun 14, 2016. [link]

[10] John T Chapman, "Modular CMTS Architecture", NCTA Spring Technical Forum, 2005. [link]

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

[11] John T. Chapman, "DOCSIS Remote PHY", *SCTE Cable-Tec Expo Fall Technical Forum*, Atlanta, GA, Oct 2013.

[12] John T. Chapman, "Remote PHY for Converged DOCSIS, Video and OOB", *NCTA/SCTE Technical Forum*, Los Angeles, Jun, 2014. [link]

[13] CableLabs, Modular Headend Architecture Specifications, 2005. [link]

[14] CableLabs, CDOCSIS Specifications. [link]

[15] CableLabs, Remote PHY Specifications (MHAv2). [link]

[16] "nFAPI Specifications", Small Cell Forum. [link]

[17] CableLabs, DPoE Specifications, 2012. [link]

[18] R. MacDavid et al., "A P4-based 5G User Plane Function", Proceedings of the ACM SIGCOMM Symposium on SDN Research (SOSR), Oct 2021. [link]

[19] John T. Chapman, Tong Liu, "Unleash the Power of Cloud Computing for CMTS", SCTE Expo 2021. [link]

[20] "Study on the Wireless and Wireline Convergence for the 5G system architecture", 3GPP TR 23.716.

[21] "Wireless and Wireline Convergence Access Support for the 5G System," 3GPP TS 23.316.

[22] "5G Wireless Wireline Converged Core Architecture Technical Report", CableLabs, June 2020.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Cost Measurement and Optimization of Cloud Computing While Enabling Virtualization of IP Network Functions

A Technical Paper prepared for SCTE by

Ajay Gavagal
Data Solutions Architect
Comcast
1800 Arch Street, Philadelphia, PA 19103
303.658.7826
ajay_gavagal@cable.comcast.com

Mehul Patel
Distinguished Architect
Comcast
183 Inverness Dr. W, Englewood, CO 80112
303-658-7826
mehul_patel@cable.comcast.com

Sinan Onder
Vice President, Software Architecture
Comcast
1800 Arch Street, Philadelphia, PA 19103
267-260-0964
sinan_onder@comcast.com

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Death by a thousand cuts ... 5
3. Strategy Outline ... 7
4. Cloud cost measurement ... 7
   - 4.1. Measurement & Observability End State ... 8
   - 4.2. Sources of cloud cost data & its related elements ... 9
   - 4.3. Guidelines for establishing foundational datasets ... 10
   - 4.4. Business Capability Dataset ... 10
   - 4.5. Software Service Dataset ... 11
   - 4.6. Business Capability to Software Service Mapping Dataset ... 11
   - 4.7. Service Component dataset ... 11
   - 4.8. Cloud service dataset ... 12
   - 4.9. Cloud service component cost dataset ... 12
   - 4.10. Financial budget ... 12
   - 4.11. Cloud Usage Dataset ... 13
5. Data transformation, storage & accessibility ... 13
6. Issues arising while consolidating cloud cost data ... 14
7. Optimizing Cloud Costs ... 15
   - 7.1. Scenario - 1 ... 15
   - 7.2. Scenario - 2 ... 16
8. Conclusion ... 17

Abbreviations ... 18

Bibliography & References ... 18

List of Figures

Title **Page Number**

Figure 1 - Overview of Virtualization of Network Functions ... 3
Figure 2 - Small Charges Leading to a Large Bill ... 6
Figure 3 - Vision for Cloud Cost Measurement and Observability End State ... 8
Figure 4 - A Look at End-to-End Datapipeline for Cloud Costs ... 14
Figure 5 - End State Application View ... 17

List of Tables

Title **Page Number**

Table 1 – Sources of Cloud Cost Data & Its Related Elements ... 9
Table 2 – Business Capability Dataset ... 10
Table 3 – Software Services Dataset ... 11
Table 4 – Business Capability to Software Service Mapping Dataset ... 11
Table 5 – Service Component Dataset ... 11
Table 6 – Cloud Service Dataset ... 12
Table 7 – Cloud Service Component Cost Dataset ... 12
Table 8 – Financial Budget ... 12
Table 9 – Cloud Usage Dataset ... 13

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

IP networks are designed to support a wide range of services, including video streaming, high-speed data transmission, voice communication, and more. However, any advancements or updates in either software or hardware necessitate the costly and time-consuming process of upgrading vendor-specific appliances deployed in these networks, posing significant challenges for network operators.

In recent years, network operators have critically reassessed the conventional approach of relying on vendor-specific appliances that tightly integrate hardware and software to deploy and operate their networks. As a result, the concept of virtualizing network functions has gained widespread acceptance among network operators. Virtualization allows specific functions of the network to be decoupled from proprietary hardware and instead implemented as software-based on virtual instances. These virtual instances utilize commercial off-the-shelf (COTS) hardware and open new possibilities for network operators. The virtual instances could either be in a private or a public cloud (consider on-demand scaling, disaster recovery scenarios). Management and operation of networks requires a whole host of software applications that can be run in a private or a public cloud as well.

Network operators have gained several advantages while virtualizing network functions. The advantages of this approach encompass both hardware and software deployments. However, in this paper, our focus will solely be on the software component, specifically on its cloud cost measurement and optimization. The authors have decided to reserve an in-depth examination of the hardware component for a separate study.

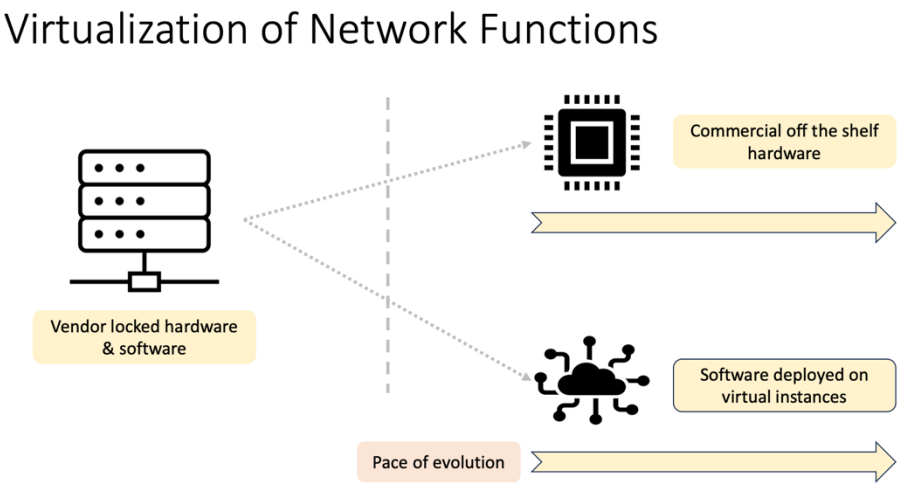


Figure 1 - Overview of Virtualization of Network Functions

Virtualizing network functions provides network operators with enhanced agility and flexibility. Operators can now rapidly deploy and scale software by creating or removing virtual instances as per demand. On the other hand, hardware deployment is also faster due to the use of the COTS model. Network operators use simple and multipurpose hardware. This approach of scaling software and hardware individually significantly improves the responsiveness of network operators to quickly adapt to changing customer needs and market demands.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

With the virtualization of network functions, network operators can fine-tune their networks to specific business demands and capabilities allowing for diversification among operators. Examples of diversification include scalability and elasticity, vendor choices for hardware, and flexibility to manage software release rate without the rigidity of hardware refresh cycles. This diversification also enables the pace of innovation and empowers operators to respond to market trends and customer preferences in a more nuanced approach, ensuring they can stay competitive in the dynamic telecommunications landscape. Imagine if all network operators were stuck with only vendor-dictated options for both hardware and software, this would impact their level of fine-tuning and the level of diversification in the market they can achieve.

Furthermore, the virtualization of network functions offers substantial cost efficiencies since bulk cashflow outlays can be avoided. By decoupling the network functions from dedicated hardware, operators can speed up or slow down software and hardware upgrades separately and at their own comfortable pace. Also, avoiding vendor lock-in, giving network operators a broader range of selection for both software and hardware vendors while avoiding costly upgrade cycles.

Network operators also gain a lot of speed and agility specifically in software development and deployment. Virtual instances can be provisioned dynamically, which automatically reduces the overall time for software releases introducing new features or fixing bugs. This does not mean we are excluding other contributing factors to the increase in speeds of software deployments such as the adoption of agile methodologies, the introduction of various new technologies, and hyperscaler's introducing higher order services.

Now that we have seen the benefits of virtualizing network functions, let's focus on some of the challenges that have been brought in through this model for network operators, specifically in the software component when using virtual instances in either a private or public cloud.

The introduction of virtualization brought about a shift in the pace at which software could be deployed when using cloud. However, this advancement also resulted in the removal of certain advantageous constraints that were previously imposed on software deployment due to financial considerations. To gain a more comprehensive understanding of this matter, let us delve further into the traditional approach that was prevalent before the use of the cloud.

In the past, software development teams, typically centralized within an IT department, engaged in extensive negotiations with the finance department to obtain the necessary funding for developing new software or upgrading existing releases. These resources, once approved, went through a typically long cycle of identification, purchase, and deployment prior to any software development activities. Software development teams were required to present a compelling case to the finance department, justifying the introduction or modification of a business capability and the requirement of resources for the same. This intricate process, although time-consuming, ensured that the deployment of resources was subject to meticulous evaluation, considering questions like "how does this addition or change in business capability impact the overall strategic value?"

However, with the advent of provisioning virtual instances in a cloud for development and deployment purposes, software development teams, and more specifically engineers, gained a considerable degree of autonomy away from finance and justification. They were now empowered to freely provision instances as needed to meet their software requirements, without giving the same degree of consideration to the value associated with each deployment with respect to the business capability or eventually the business activity it is supporting.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

This predicament was further exacerbated for finance. Suddenly, the traditional capital expenditure associated with resource provisioning began to transform into operational expenditure, requiring the finance department to adapt and prepare for this notable shift. Additionally, the visibility and control that finance teams once had, enabling them to identify the purpose of a provisioned instance for a specific project, became exceedingly challenging to attain from private or public cloud environments. Consequently, when the strategic viability of a project diminishes, finance teams encounter difficulties in enforcing rigorous clean-up or repurposing measures for the associated instances. This lack of visibility and control is prompting finance teams to seek answers regarding where to initiate their efforts, sparking the rise of a new discipline known as FinOps, or financial operations, which has gained substantial traction within the industry. Aptly so, FinOps Foundation executive director J. R. Storment says, noting that the "cloud removes finance from the buying process and hands the credit card to cloud engineers."

This paper partly concentrates on the impact of transitioning from a capital expenditure model to operational expenditure model on finance during software operations, neglecting any discussion regarding the impact on finance during software development. However, it is worth noting that the problem of transitioning from capital expenditures (CapEx) to operational expenditures (OpEx) due to the adoption of cloud also persists during software development, necessitating a comprehensive study on its own.

When considering the convergence of two significant trends, the challenges faced by network operators become evident. In a landscape where virtualizing network functions and embracing cloud deployments for enhanced speed, adaptability, and innovation are standard practices, the questions arise:

1. How can a network operator effectively manage and control the software operational expenses that accumulate?
2. What strategic value can network operators add to their portfolio by being on top of such operational expenses?
3. Is all this effort to go after these expenses even worth the time & effort?

In this paper, the authors dedicate some amount of time to discuss all the issues there are in cloud cost measurement, interpretation, and optimization in a cloud environment and suggest some strategies on how best to get a hold on the situation to avoid overrun of operational expense.

2. Death by a thousand cuts

Before delving into the crux of this white paper on cloud cost measurement and optimization, it is imperative that we draw a compelling parallel between the financial dynamics of cloud expenditure and the purchasing tendencies exhibited by an average consumer when acquiring software applications through subscription models, often resulting in underutilization or complete non-usage of resources.

The new landscape of consumer behavior has witnessed a significant shift, with consumer product companies opting to adopt the subscription-based paradigm over the conventional approach of selling software products at a one-time price which is not affordable to a sect of consumers. This change can be observed across a spectrum of software apps, spanning from operating systems tailored for smartphones, laptops, and tablets to a vast array of applications encompassing video streaming platforms, instant messaging services, video conferencing tools, gaming applications, and an extensive repository across other categories.

Under a software subscription-based model, consumers are frequently offered an enticing assortment of software applications to cater to their needs. In this environment, the cost of individual applications seems inconsequential when contrasted with the recurring stream of modest subscription fees. However, this cost-

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

effective shopping spree can quickly lead to an accumulation of unwanted, duplicate applications that are either utilized scarcely or, in certain instances, not used at all.

Drawing a direct parallel to the world of cloud computing and its costs, organizations often find themselves with similar purchasing habits, except on a large scale. The benefit of the cloud, and its promise of scalability, flexibility, and pay-as-you-go pricing, make organizations provision a multitude of resources, services, and infrastructure without extensive thought into their strategic value, actual utilization, or the full cycle from provisioning to decommission. Just as an average consumer may accumulate a whole host of unused or underused software apps, organizations can find themselves with an unwieldy cloud bill with no insight into redundancy, inefficiency, and financial wastage.

The procurement process for cloud services is equally intricate compared to the traditional software and hardware purchase model. While the pricing structure may appear transparent and based solely on operational expenditures (OpEx), cloud providers offer diverse options to assist consumers, which can nonetheless result in cost opacity. For instance, consumers may be able to opt to make an upfront payment for a subscription fee, enabling them to secure a significant discount on the service. Or bulk purchases may be available, and which may result in substantial discounts. Furthermore, consumers frequently have a wide range of choices when it comes to purchasing licenses on a subscription model or accessing higher-order services for a higher price to meet their specific requirements. All these factors significantly complicate the decision-making process of consumers during both software development and operations. The decision-making process of choosing the right resources falls entirely on the software engineering team due to the autonomy of cloud provisioning.

The authors call the problem of having to deal with a thousand details of choosing and using the right resources as “death by a thousand cuts” to an average consumer or enterprise.

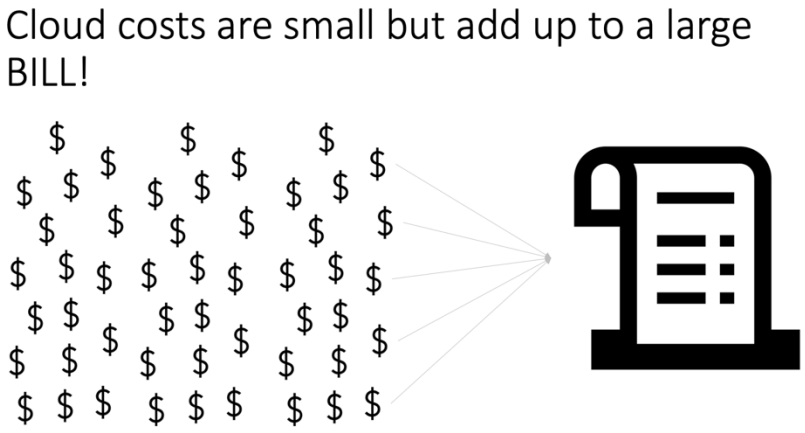


Figure 2 - Small Charges Leading to a Large Bill

Looking at this striking similarity between consumer app subscriptions and cloud costs, it becomes evident that organizations must exercise extreme caution and adopt a proactive approach toward cloud cost measurement and optimization. Failure to do so can result in substantial financial resources that are squandered on underutilized or unnecessary cloud resources, negating the benefits and advantages offered by the cloud.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Strategy Outline

Addressing cloud expenditure shares similarities with other strategies for cost optimization. Organizations can pursue several key initiatives to regain control over their cloud costs:

1. Measure cloud costs per business capability and eventually per activity
2. Minimize wastage, including the mitigation of underutilization.
3. Engage in comprehensive evaluations with cloud providers, which involve:
    a. Negotiating pricing for services and infrastructure.
        i. Capitalizing on bulk purchase discounts.
    b. Engaging in best practice discussions.
    c. Tracking the roadmap to prevent redundant efforts.
    d. Submitting feature requests that facilitate cost reduction.
4. Re-architect applications and their components to diminish redundant computing and storage.
5. Perform accurate forecasting to anticipate future requirements.
6. Embrace a proactive optimization approach for future needs.

When examining the above overarching strategies, it becomes evident that **"Measurement and Observability"** of cloud costs are fundamental prerequisites for successfully implementing any of these initiatives. Another crucial factor for achieving success in a substantial undertaking like this is to ensure unwavering commitment from every individual involved throughout the organization. Furthermore, it is important to emphasize that this commitment should not be treated as a one-time event, but rather as a continuous and ongoing effort.

4. Cloud cost measurement

Before delving into the measurement of cloud costs, let us first take a step back and examine the types of observabilities we can establish during software design, development, and deployment. The authors propose that there are primarily two types of observabilities that can be established:

1. Observability during the design phase: This approach takes a proactive stance, treating cloud cost observability as a design principle and a fundamental pillar of sound software design. By incorporating observability into the design stage, it becomes possible to comprehend software behavior, identify areas for optimization and debugging, determine where and how to scale, and more. Employing approaches such as enhancing telemetry to help finance measure activities during development vs. operations and their respective cloud costs is one way to achieve this. Cost tagging is another approach. However, achieving this proactive observability approach is challenging in reality due to factors such as priorities, escalating requirements, tight deadlines, and limited budgets. Consequently, the authors defer a comprehensive examination of implementing proactive observability for cloud cost measurement to a future study.

2. Observability after development and deployment: This approach is more reactive in nature, as it involves responding to cloud costs after they have already been incurred rather than anticipating and addressing them proactively. Unfortunately, many enterprises still lack a solid foundational approach to measuring and observing cloud costs even in a reactive approach, resulting in uncoordinated responses. Nevertheless, there is good news as establishing such an observability platform is relatively straightforward to establish with existing tools and processes. Therefore, the authors choose to focus on establishing a robust measurement and observability platform—in a

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

reactive approach focusing only on software operations within this paper. Additionally, the authors plan to conduct a more detailed study on how to react to observed cloud costs in the future.

Now let's explore the intricacies of cloud cost measurement. Through the upcoming sections, we will examine:

1. The different sources of information available for cloud cost measurement.
2. Challenges that arise during the data collection process.
3. Analyzing the extent to which this information is digitized.
4. How this information is interpreted within the organization.
5. Finally, we will delve into the incentives that drive participation in cost measurement and optimization efforts, as well as the importance of effective communication channels and the frequency of communication.

4.1. Measurement & Observability End State

The authors have a vision for the end state of the measurement and observability platform, which is depicted below. However, before we present the desired end state, it is crucial to establish a clear understanding of all the terminologies employed in the accompanying diagram. We recommend linking these terminologies with the ones familiar to your enterprise, as the terminology might differ in wording but ultimately converge on the same meaning.

Terminologies & their meaning:

a. Business Activity: Any activity that helps make money for the business.
b. Financial Budget: Spending plan established based on income and expenses of the business.
c. Business Capability: Capabilities or functions required by a business to achieve a business activity.
d. Service: One or more software services established to achieve a business capability.

The remaining terminologies are straightforward and will not be elaborated upon. Now that we have introduced the terminologies, we dive into what the measurement and observability platform end state might look like:

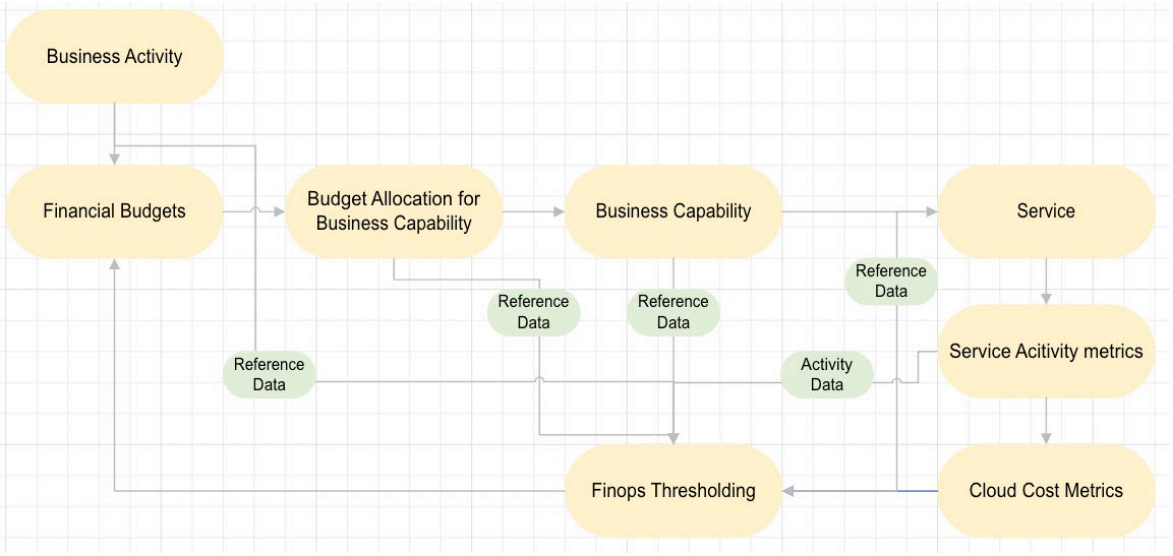


Figure 3 - Vision for Cloud Cost Measurement and Observability End State

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2. Sources of cloud cost data & its related elements

Looking at different sources of information available to enable a successful cloud cost measurement.

Table 1 – Sources of Cloud Cost Data & Its Related Elements

Data Name	Source of Data	Usage of data in cloud cost measurement	Size of Data	Data available for consumption in digitized form?
Software Service Name, Owner, Contact, Purpose, etc.,	Service portfolio management systems	Categorization of cloud costs per service within an organization	Small	Only for software services that are mature
Business Capability Name, Owner, Contact, Purpose, Software application supporting the business capability, etc.,	Business Capability Maps	Categorization of cloud cost data per business capability	Medium	Only for mature business capabilities
Business Activity Name, Owner, Business Capabilities supporting the specific business activities, etc.,	Enterprise Resource Planning systems (ERP)	Categorization of cloud costs per business activity	small	Yes
Cloud costs in detail per service used	Cloud provider usually provides a tool that tracks costs per service, tag, etc.,	Total cloud cost reporting by service, tags, etc.,	Large	Yes
Financial budgets allocated for cloud spend per account and per organization	Some sort of ERP systems	Tracking actual spend vs. budget to influence cash flow requirements	Small	Yes
Thresholds on cloud spend per service, service	In-house built systems to store such information	To provide notifications in a reactive	Large	Maybe

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

limits, contractual agreements etc.,	or distributed across individuals systems	environment on cloud costs to finance, engineers etc.,		
Contact information in case of exceeding thresholds, escalation paths etc.,	Workflow management software	Notify the right contacts to resolve breaches in cloud service limits, spend etc.,	Medium	Yes
Cloud resource inventory	Systems that maintain all the cloud resources deployed within an organization	Cloud centers of excellence to understand cloud usage patterns across an organization	Large	Yes

By referring to the preceding table, it becomes evident that a wide range of services and information systems are essential for facilitating the collection, ingestion, and persistence of data pertaining to cloud costs.

4.3. Guidelines for establishing foundational datasets

If you are embarking on the journey of cloud cost measurement for the first time, laying the groundwork with foundational datasets can significantly impact your ability to analyze data and identify opportunities for optimization. The authors suggest a selection of foundational datasets presented in tabular form, which serves as a starting point rather than an exhaustive list.

4.4. Business Capability Dataset

Table 2 – Business Capability Dataset

Column Name	Description
Business Capability Identifier	Identifier for a Business Capability
Business Capability Name	Name of the Business Capability powering a business activity
Business Capability Description	A short description of the Business Capability
Business Capability Owner	Owner of the Business Capability
Business Capability Owner Contact Information	Contact information is meant to identify how a Business Capability owner should be notified

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.5. Software Service Dataset

Table 3 – Software Services Dataset

Column Name	Description
Service Identifier	Identifier for a software Service
Service Name	Name of the Service as identified and recognized by consumers
Service Description	A short description of the Service, its intent
Service Owner	Owner of the Service
Service Owner Contact Information	Contact information is meant to identify how a Service owner/s should be notified

4.6. Business Capability to Software Service Mapping Dataset

Table 4 – Business Capability to Software Service Mapping Dataset

Column Name	Description
Service Identifier	Identifier for a software Service
Business Capability Identifier	Identifier for a Business Capability

4.7. Service Component dataset

Table 5 – Service Component Dataset

Column Name	Description
Service Identifier	Identifier for the Service that component belongs to
Service Component Identifier	Identifier for the Service within the application
Service Component Name	Name of the Service component as identified and recognized by developers
Service Component Description	A short description of the Service Component, its intent

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.8. Cloud service dataset

Table 6 – Cloud Service Dataset

Column Name	Description
Cloud service Identifier	Identifier for the cloud service
Cloud service name	Name of the cloud service used
Cloud service description	A short description of the cloud service

4.9. Cloud service component cost dataset

Table 7 – Cloud Service Component Cost Dataset

Column Name	Description
Cloud service identifier	Identifier for the cloud service
Cloud service component Identifier	Components of the cloud service where cost is assigned
Unit cost for the component	Unit cost here refers to the kind of unit the cloud provider has assigned to measure the usage of a service component

4.10. Financial budget

Table 8 – Financial Budget

Column Name	Description
Financial budget identifier	Identifier for the set budget
Business Capability Name	Name of the Business Capability
Business Capability financial budget	Set budget for the Business Capability
Maximum allowed deviation	Deviation allowed from budget beyond which triggers need to be kicked off
Finance owner	Owner from finance who is in charge of budget allocation
Finance owner contact information	Contact information for finance owner

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.11. Cloud Usage Dataset

Table 9 – Cloud Usage Dataset

Column Name	Description
Cloud Service Usage identifier	Self-explanatory
Cloud Service Usage Metric Name	Self-explanatory
Cloud Service Usage Metric	Brief definition of the cloud service usage metric
Free tier	Cloud service providers might offer free tier usage to promote their services
Minimum usage required	Minimum usage required by the cloud service provider to ensure the rates provided justifies their cost

5. Data transformation, storage & accessibility

In order to establish the foundational datasets, as mentioned previously, a critical step involves identifying the appropriate source systems within the enterprise. Once these systems are identified, the next requirement is to implement a robust big data processing and storage platform. It is important to note that when dealing with metrics related to cloud service usage and costs, the volume of data can reach staggering proportions, easily amassing billions of records. Given the enormity of these datasets, processing them effectively necessitates a substantial number of clusters and considerable processing power to handle the computational demands.

Below is an application-level view of the entire pipeline when set up:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

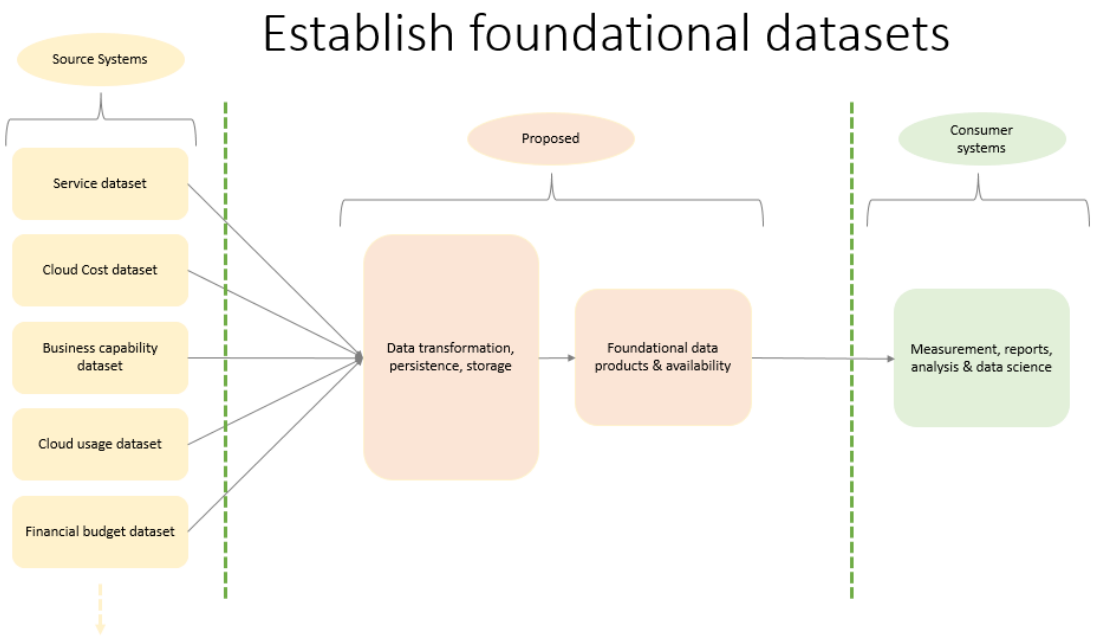


Figure 4 - A Look at End-to-End Datapipeline for Cloud Costs

Once such an environment is established, network operators gain the capability to measure cloud costs from multiple perspectives.

Let's explore some of the questions that can be addressed through this setup:

1. For Leaders/Finance:
    - What does the overall cloud cost look like across the organization?
        1. Breakdown of costs by business capability and business activity.
    - Which business capability requires focused attention due to escalating cloud costs?
        1. Identification of business capability owners
    - Determination of services associated with this business capability.
2. For Architects/Engineers:
    - Which components of applications are contributing to exceeding of cost thresholds?
    - Can we consolidate components from different applications to create higher-order services? (Cost serving as a justifiable factor for strategic discussions).
    - What are the most significant opportunities for optimization?
3. For Product Owners/Program Managers:
    - Which services within my product/portfolio account for the highest cost expenditure?
    - Can the cost breakdown be specified by individual application components, aiding in the formulation of user stories for further investigation?

These questions not only pave the way for individual optimization efforts but also open avenues for optimization from various angles.

6. Issues arising while consolidating cloud cost data

Having familiarized ourselves with the wide array of diverse information accessible for establishing a robust framework for measuring cloud costs, it is now imperative to shift our focus towards the potential

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

challenges that may arise when consolidating such data to derive meaningful insights, which can aid in optimizing efforts and conducting benefit analysis.

1. Determining the appropriate level of resolution in data is consistently a challenging problem to solve. This challenge becomes particularly evident when comparing the resolutions between business capabilities and the underlying software that enables them, as opposed to the measurable cloud cost per software.
2. Digitized and readily accessible mappings between software services and the products they support are rarely available.
3. Data sources may be dispersed across different organizations, teams, silos, and geographical locations. This inherent distribution complicates the process of establishing a robust measurement framework.
4. Data sources can exist in diverse systems with varying storage patterns, including legacy systems with limited data accessibility and raw data formats.
5. The availability of contextual information alongside the data may be limited or entirely absent. For example, when a cloud cost for utilizing a serverless function is reported in "X" units, the interpretation and meaning of those "X" units might be unclear or undisclosed.
6. Cloud providers frequently alter their pricing models and the metrics they employ for charging. Failure to account for these modifications in the measurement process, especially when using non-digitized input methods, can lead to a significant increase in the effort required for cloud cost measurement.

7. Optimizing Cloud Costs

Cloud cost optimization is a vast and intricate topic that necessitates thorough examination. Although this paper only scratches the surface, the authors have tried to explore key concepts related to the cloud cost optimization process. They discuss the challenges that may be encountered during such efforts and offer a high-level insight into automating the entire process—from source to measurement to optimization. Without further ado, let us delve into the topics.

The optimal approach for conducting an optimization exercise involves leveraging measured data and the insights it offers. But after leveraging the measured data, there is a whole hosting of convincing the engineers to act. To illustrate this point further, the authors present two examples:

7.1. Scenario - 1

In one typical scenario, software engineers are assigned user stories to develop features for Project X. These engineers promptly set up the necessary infrastructure to deploy their workloads. As they progress through the development process and conduct load testing, they realize that using larger compute resources with increased storage capacity enhances workload performance. The cost difference between the two environments, in terms of compute and storage, is merely $0.02 per day. Considering the marginal cost increase against the significant boost in workload processing speed, the engineer decides to opt for the higher compute and storage options.

Let's assume there are one hundred engineers in this organization, each developing and deploying one workload. Furthermore, let's assume that 20% of the engineers do not perceive this slight cost increase as problematic. By performing some calculations:

Total amounts increase per year = (% of engineers with larger compute/storage) * (number of engineers) * (cost increase per workload per day) * 365

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

This results in a grand total of $146, which is approximately 7,300 times $0.02. At first glance, this may not appear significant. However, several factors have not been considered to make a full assessment:

a. Does the increase in compute/storage truly justify the enhanced output speed?
b. Are the consumers of these workloads satisfied with lower frequency outputs?
c. Who has the authority to make decisions regarding increased compute/storage requirements?
d. Has any cost modeling been conducted beforehand to assist engineers in making such decisions?

When engineers solely focus on individual workloads and their cost wastage, the problem seems small enough to even address.

On the other hand, when leadership or finance teams solely focus on the aggregate numbers, they may fail to recognize a crucial aspect: the diminishing value of the product relative to the amount of money being spent. In other words, as time progresses, the organization is receiving a progressively decreasing level of output in relation to the financial resources invested.

The above scenario is one of the many situations where wastage becomes a crucial issue to tackle for optimization efforts. Let's review another scenario.

7.2. Scenario - 2

In this scenario, software engineering teams have made the decision to deploy an application that requires a certain amount of compute resources (X) and storage (Y) to generate a specific output. As multiple teams within the organization develop similar applications to perform comparable functions (a common situation in large enterprises with isolated structures), the total compute and storage needed becomes dependent on the level of duplication.

The total expenditure for these applications can be calculated as: Total $ spent = N (X Compute $'s) + N (Y Storage $'s), where N represents the number of applications with similar functionality.

At this point, organizational leadership steps in to address the issue of silos and initiates a consolidation effort. However, during such consolidation processes, it is often inevitable that some loss in workforce is expected. The newly onboarded engineers are now faced with the challenge of dealing with numerous compute instances and storage setups as they undertake the consolidation effort.

In this context, instances and storage layers that remain unused during the consolidation process and incur relatively lower expenses, such as a few hundred dollars, often go unnoticed.

Now, let's expand the scope of this scenario to encompass the entire enterprise and consider the cumulative impact of the ongoing transformations, as well as the accumulation of unused cloud infrastructure capacity over time. While the overall cost is substantial when viewed as a whole, individual engineers may perceive the individualized resource cost as relatively insignificant or fragmented for the problem of underutilization.

The authors have emphasized only two scenarios, yet the issue of optimization becomes unmistakably apparent. Specifically, it raises the question of how to persuade engineering teams that seemingly insignificant expenses resulting from the unused or underutilized cloud infrastructure can accumulate into a significant sum if left unchecked.

The authors propose a suggestion to tackle this issue by implementing a set of measures. Please note that the following topics are mentioned briefly, but they require a comprehensive study on their own. These include establishing robust measurement mechanisms, defining thresholds using micro budgets, implementing alarm systems, and creating a feedback loop for engineers. Additionally, the authors

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

recommend that engineers who take the initiative to optimize and address the problem should be duly rewarded for their efforts.

The proposed end-to-end flow is outlined below:

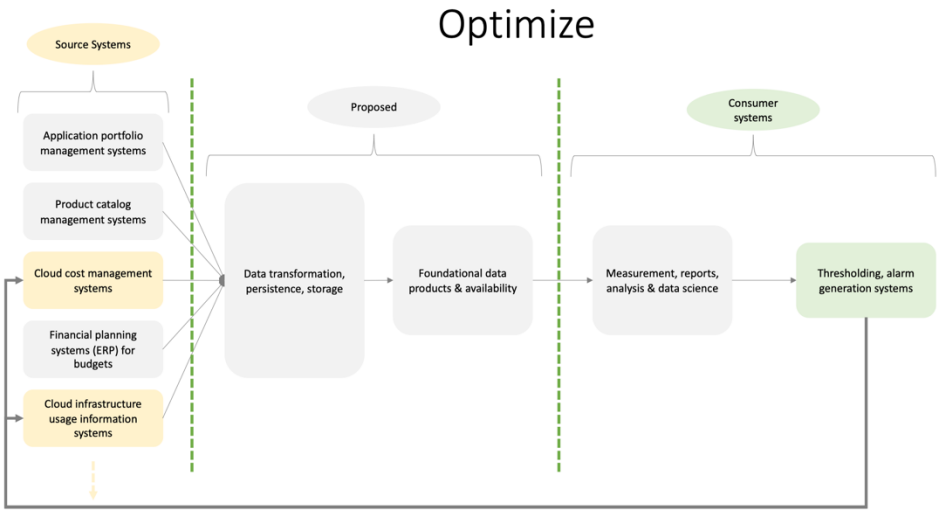


Figure 5 - End State Application View

In a planned subsequent paper, the authors deliberately defer the exploration of several crucial aspects related to establishing an automated feedback loop, implementing gamification techniques involving rewards for engineers, setting precise thresholds and alarms, as well as employing proactive approaches to cloud cost management. However, it is anticipated that the concepts presented herein serve as a catalyst for the generation of innovative ideas and inspire further investigations into the diverse range of possibilities in constructing a comprehensive and resilient framework encompassing the entirety of cloud cost measurement, optimization, and actionable strategies, all while leveraging the potential benefits derived from gamification principles.

8. Conclusion

Network operators have fully embraced the virtualization of network functions, recognizing its benefits in terms of speed, flexibility, and innovation. Simultaneously, the adoption of cloud technologies for deploying software through virtual instances is gaining momentum. These two trends are interconnected but also pose new challenges for network operators. One significant challenge is effectively managing cloud costs during software development and deployment. Organizations now confront complex cloud bills comprising billions of individual line items with small amounts. Finance and leadership teams face the daunting task of comprehending how these various monetary amounts contribute to business capabilities while minimizing waste and redundancy.

This paper addresses measures required to tackle such a situation, emphasizing the importance of a robust measurement and observability platform as a foundation. Additionally, the authors offer insights into potential issues that may arise in such a setup. Finally, the paper concludes by outlining the future of cloud cost management, advocating a proactive approach over a reactive one.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

COTS	Commercial off the shelf
VNF	Virtualization of network functions

Bibliography & References

J. R. Storment, executive director of the FinOps Foundation says, "Cloud removes finance from the buying process and hands the credit card to cloud engineers."

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Cox Next Generation 400G IP+OLS Architecture for Maximum Network Optimization and Cost Benefits

A Technical Paper prepared for SCTE by

Saurabh Patil
Principal Architect
Cox Communications
6305 Peachtree Dunwoody Rd, Atlanta, GA, 30328
404-269-8364
Saurabh.patil@cox.com

Jason Bishop
Senior Manager Design Engineering
Cox Communications
6305 Peachtree Dunwoody Rd, Atlanta, GA, 30328
404-269-3994
Jason.Bishop@cox.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Cox Evolution of Optical Infrastructure ... 3
   - 2.1. Cox Optical Footprint ... 4
   - 2.2. Problem Statement ... 4
3. Open Line System (OLS) ... 6
   - 3.1. OLS Requirements ... 7
   - 3.2. OLS Benefits ... 9
4. 400G ZR++ Pluggables ... 9
5. Live Production testing and Details ... 10
   - 5.1. Results of the test ... 11
6. Architecture Changes ... 13
   - 6.1. IP Layer Architecture Change ... 14
7. Operational Challenges ... 14
8. Conclusion ... 15

Abbreviations ... 16

List of Figures

Title **Page Number**

Figure 1 – Current Transport Network ... 5
Figure 2 – OLS Transport Network ... 6
Figure 3 – Flex Grid or Gridless Example ... 7
Figure 4 – Open Line System with Disaggregated Transponders (OLSDT) ... 8
Figure 5 – 400G ZR ++ plug QSFP-DD form factor ... 9
Figure 6 – 400G ZR ++ in Prod ... 10
Figure 7 – Route LA-PHX Production test ... 11
Figure 8 – Results LA-LV ... 11
Figure 9 – Results LA-PHX ... 12
Figure 10 – Target implementation of 400G ZR ++ in Production ... 14

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Cox Communications has extensive optical fiber infrastructure comprising over 25,632 route miles of National Optical backbone and in Metro Markets. Cox Communications relies on Backbone to connect all their Metro markets, urban centers, and peering centers to keep everything running smoothly and efficiently. It's an essential tool that ensures seamless connectivity and enables them to serve their customers better. Fiber is an integral part of Cox Communication's assets, and we in Engineering work constantly to optimize and strategically utilize our fibers for the lowest Total Cost of Ownership (TCO) and better customer experience to transmit Terabits of data.

As we approach the Shannon limit to exhaust fiber capacity, what other options do we have as a service provider to keep up with ever-increasing bandwidth demands and capacity pressure from innovation and customer consumption? There is another 15-20% scope to improve spectral efficiency before fiber exhaustion, but that is just a stop-gap for 1-2 years for capacity growth.

This problem needs not only engineering innovation from optical vendors and Internet Protocol (IP) routers but also from service providers to rethink how we architect our network and use surgical techniques to use appropriate technology in certain parts of the networks where it's needed.

In Cox, we use embedded highly expensive coherent transponders extensively in Backbone, metro, and some cases in access networks as a solution for adding capacity, even if it's short distances. The saying "When you only have a hammer, every problem can start to look like a nail" is undeniably true. Having limited tools or resources can lead to a narrow-minded way of thinking. It's crucial to acknowledge that there are multiple approaches to a problem, and sometimes, it's necessary to step out of our comfort zone and explore unfamiliar tools and routines to find the optimal solution. Here, we use high-power transponder solutions for all our problems with heavy reliance on over-engineering. The consequence is paying a very high significant cost and increasing power consumption. With the introduction and maturity of new coherent pluggable optics, we at Cox Communications are learning to use pluggable optics in Backbone, Metro, and Access to leverage lower pricing and new capabilities. Utilizing these new coherent pluggable optics on high-capacity short spans can offer exceptional spectral efficiency at a reasonable cost. This approach is a strategic method to enhance performance while staying within budget.

In this paper, we will discuss how we are re-architecting Cox's Backbone and metros to overcome Shannon limits and use coherent pluggable optics in Cox's Next generation Open Line System (OLS) that supports both pluggable (400G ZR, ZR+, ZR++) and traditional transponder architectures for maximum network optimization and cost benefits.

Also, we share our experience of doing live production field testing of 400G ZR++ coherent pluggable optics on very challenging spans and its operational impacts and results.

2. Cox Evolution of Optical Infrastructure

The Cox Communications network backbone was established circa 2001, consolidating leased OC-48, OC-12, and OC-3 circuits. Previously, each Cox market operated independently with its transit peering connections. Over the years, Cox experienced tremendous growth in markets and data, which made financial sense to build our own national Backbone to control our capacity planning without using leased circuits for better customer experience. The goal was to connect all metro markets to peering transit locations, reduce leased circuit costs, and provide resiliency.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.1. Cox Optical Footprint

In 2007, a fiber optic network was established using Dense Wavelength Division Multiplexing (DWDM) technology, gradually expanding to cover all metropolitan markets and peering locations. Initially, the DWDM network utilized 80 channels with a fixed spacing of 50Ghz in the C Band, with each channel capable of carrying 10G/OC-192/1G. To carry non-return-to-zero (NRZ) 10G signals, dispersion compensation modules (DCM) were employed on the line systems. It wasn't until 2012 that Cox made the move to 100G Ethernet technology on the same fixed 80-channel DWDM system.

The national optical backbone links all national peering centers (NPC), national data centers (NDC), regional caching centers (RCC), and Cox regional data centers (RDC) using an optical mesh topology. Meanwhile, a metro core optical network connects all Cox hub locations as a hub and spoke topology, which aggregates at the RDC.

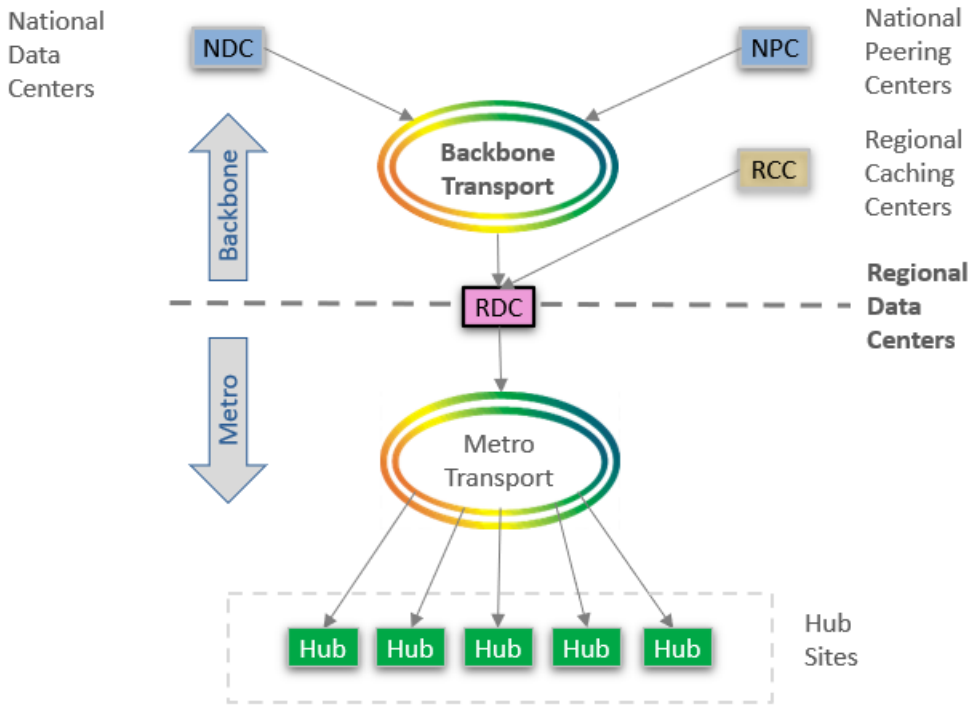


Figure 2 – Cox Backbone/Metro as of 2023

2.2. Problem Statement

The current Backbone and Metro core network architecture will not support future capacity requirements and service demands. To simplify the deployment and its ongoing support, re-architecting the entire optical transport system is needed, especially the line system, to accommodate future technological evolution, e.g., 400G/600G/1Tbps/1.2Tbps/1.6Tbps single wavelengths and pluggable optics such as 400GZR/ZR+/ZR++ for internet protocol over dense wavelength division multiplexing (IPoDWDM) architecture.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In summary, the current architecture in Cox's metro cores and national optical Backbone cannot handle advanced data rates. Many current line systems have limited technologies, such as fixed channel Reconfigurable Optical Add Drop Multiplex (ROADM) filters with a capacity of 100Ghz/50Ghz and 40, 80, or 96 channels. Some systems also use Fixed Optical Add Drop multiplex (FOADM) for optical transport, while others still require Dispersion Compensation Modules (DCM) for 10G wavelengths.

Approximately 70% of the optical Backbone and metro infrastructure require replacement to support 400G to 1.2Tbps waves and 400G ZR+ pluggable optics, while the remaining 30% needs updates in software and hardware to facilitate pluggables.

Let's dive into what the current transport network looks like.

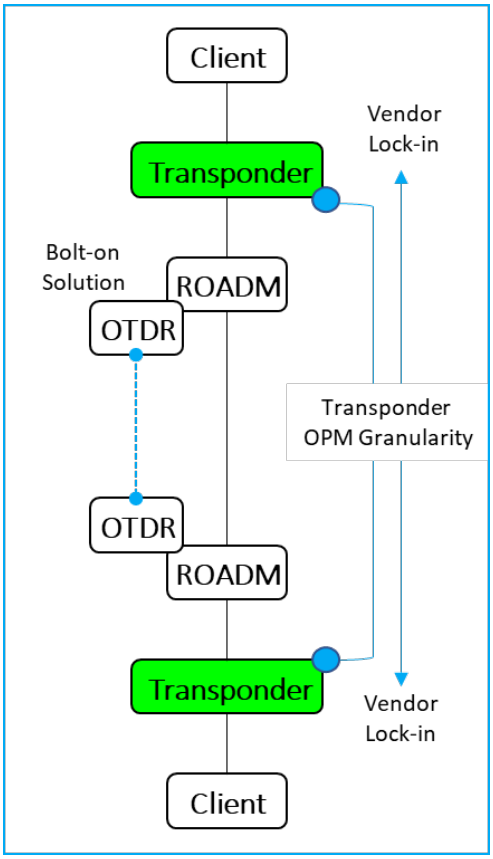


Figure 1 – Current Transport Network

Above is a typical example of a transportation network wherein a single vendor offers a comprehensive system encompassing the optical line system, with in-line amplifiers (ILA), ROADM, and transponders that convert client traffic to DWDM wavelengths on the line system. The advantage of a single vendor is that it has tighter power management between the line system and transponders, along with an element management system (EMS) for end-to-end visibility. This architecture allows a single vendor complete control over technology and innovation in service provider networks, leading to slow functional improvement and limited cost compression opportunities.

Introducing new technology on the transponder where most of the innovation is happening to maximize the spectral efficiency fully will require a whole line system upgrade or change to an entirely new vendor, which is cost-prohibitive and time-consuming.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Open Line System (OLS)

As the name suggests, an Open Line System (OLS) is an optical transport Dense Wavelength Division Multiplexing (DWDM) system without technology and vendor lock-in. The line system is meticulously designed to provide reliable and efficient support for high-capacity signal transponders. It also has pluggable optics that enable seamless access and facilitate the smooth operation of metro and long-haul backbones.

There are multiple interpretations of OLS, but at Cox, we use the OLS Architecture which means that a single vendor provides the optical and physical line system with open APIs, while different vendors supply the transponders and coherent pluggable optics with their open APIs. This type of architecture is sometimes referred to as partial disaggregation.

Without the technology lock-in, our line system can quickly adapt to support a wide range of technologies. It can scale up to accommodate future speeds of 400/800/1.6T coherent signals.

Furthermore, by avoiding vendor lock-in, the line system can support various transponder technologies from different vendors, even those that may be more advanced than the current vendor's technology. This flexibility allows service providers to choose the best vendors for their needs and mix and match them to take advantage of each vendor's strengths.

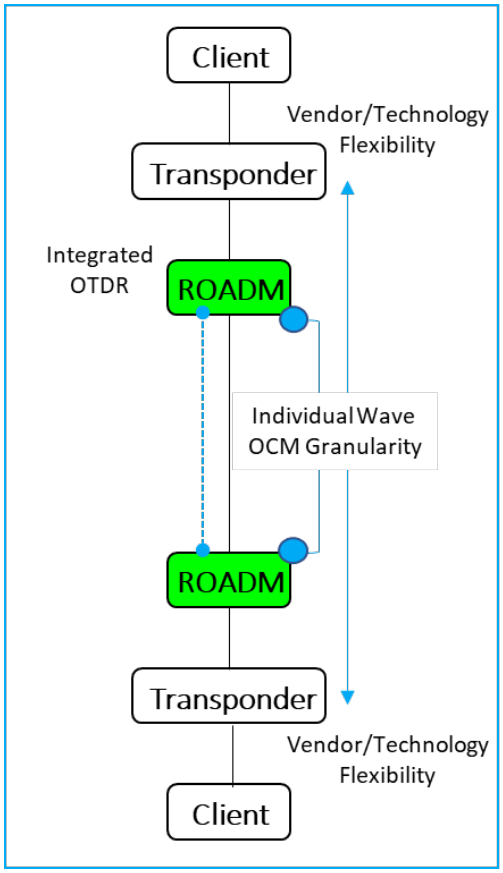


Figure 2 – OLS Transport Network

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.1. OLS Requirements

The requirements for implementing OLS in various networks may differ, but at Cox, we ensure the inclusion of necessary components in our network. These components are mandatory to ensure the smooth and efficient functioning of the OLS system.

Reconfigurable Optical Add Drop Multiplexer ROADM-based architecture must have flex grid capability. ROADMs are capable of wavelength routing, power balancing between channels, and spectral bandwidth allocation. To achieve this, they require the ability to measure fine granular optical channels using precise Optical Channel Monitors (OCMs). These OCMs can measure external alien wavelengths and power, mainly when using pluggable optics in OLS.

Flex Grid or Gridless is central technology that allows complete flexibility in packing any signal width (i.e., 50 GHz, 75 GHz, 112.5 GHz, or 150 GHz) as closely as possible. The current grid structure has been modified to allow for greater flexibility in frequency allocation. The grid can be adjusted in increments of either 6.25 GHz or 12.5 GHz, providing more excellent customization options for users.

Spectral efficiency is a critical factor in determining the extent to which available spectrum is utilized. This measure is indicative of the effectiveness with which fiber is being utilized, which is of utmost importance. Higher spectral efficiency allows service providers to assess how well transponders can be used to cram more bandwidth into a minimal fiber spectrum. This, in turn, expands the use of single-line systems, reducing overall network costs while simultaneously increasing capacity. Therefore, it is crucial to maintain and improve spectral efficiency as it contributes significantly to the network's overall performance.

$$\text{Spectral Efficiency (SE)} = \frac{\text{Information rate (bit/sec)}}{\text{Bandwidth (Hz)}}$$

A 10G signal sent over 50 GHz fixed spectrum has a SE of 0.2 b/s/Hz.

A 100G signal sent over 50 GHz has a much better SE of 2 b/s/Hz.

A 400G signal sent over 75 GHz has a better SE of 5.33 b/s/Hz.

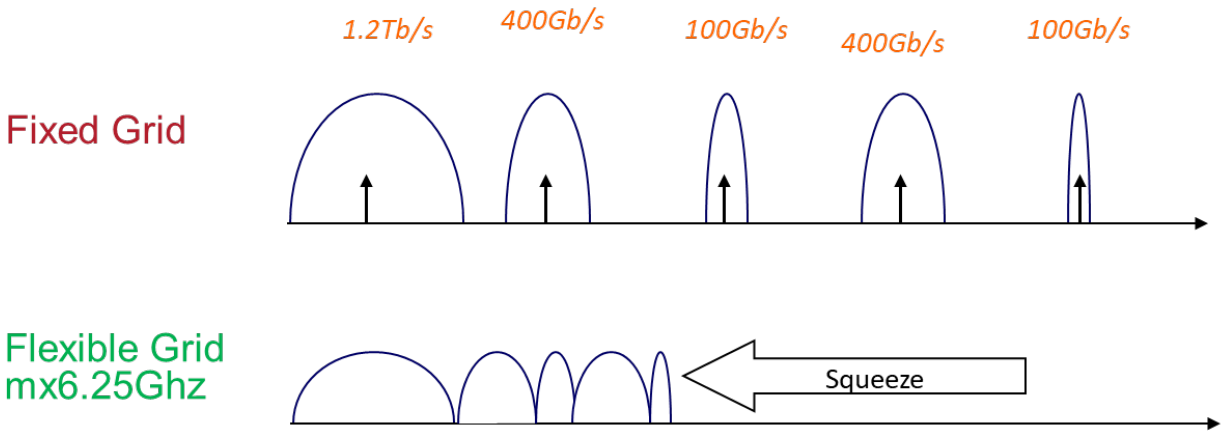


Figure 3 – Flex Grid or Gridless Example

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Our simulations show a 30% improvement in capacity as channels only use as much bandwidth as required and nothing additional, as seen in the figure above. The Flex Grid feature is needed in the OLS architecture, which makes the optical transport modulation signal agnostic and a secure investment in the ROADM-based architecture.

An Optical Time Domain Reflectometer (OTDR) is a highly efficient and advanced device that is utilized for the purpose of detecting any fiber measurements and cuts in real-time. It provides the operator with real-time loss measurements and any possible deviations, which are essential for ensuring accurate design parameters. The OTDR is an indispensable tool for professionals working in fiber optic communication, as it enables them to quickly and easily identify any issues that may arise and take necessary corrective actions expeditiously. This device has proven to be highly reliable and effective, making it a popular choice among professionals in the industry. In the new optical line systems, OTDRs are integrated into the ROADMs, which give higher integration.

C+L band support starts with C-Band and then expands to L-Band to double the fiber capacity for future-proofing networks and being scalable.

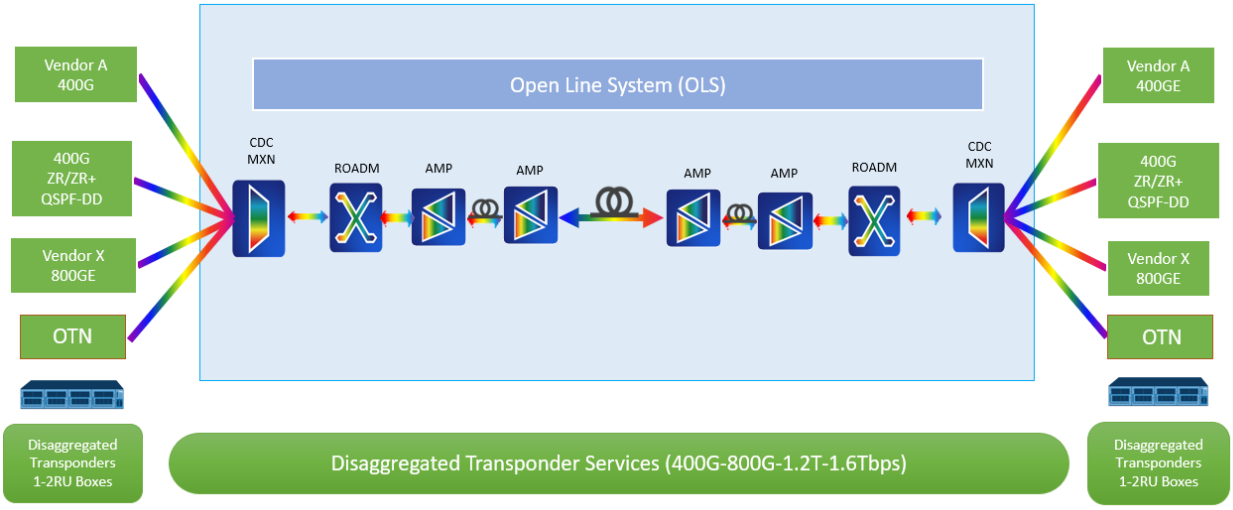


Figure 4 – Open Line System with Disaggregated Transponders (OLSDT)

The above figure shows the OLS with disaggregated transponders from different vendors, including OTN and pluggable optics. The line system is disaggregated from transponders and technology, giving Cox flexibility to scale and use each vendor's technology strengths.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.2. OLS Benefits

As seen in the above section, a summary of the benefits are listed below:

- Maximize metro and core fiber pair bandwidth with C+L.
- Client and line disaggregated (different vendors)
- Separate transponder and ROADM development
- Highest opportunity for cost compression.
- Remove traditional client and line integration.
- Remove traditional IP and Optical boundaries (400G ZR/ZR+)
- Create maximum flexibility with client/service aggregation.
- Reduce client/service aggregation costs.

4. 400G ZR++ Pluggables

The history of optical transmission has seen many technical milestones, such as 40G and 100G coherent optical transmission. However, the latest innovation of the 400G ZR coherent optical plug is significant and is expected to revolutionize the optical industry. Service providers must adjust their architecture to take advantage of the plug's small form factor and reduced cost-per-bit savings. The 400G ZR is a highly advanced and reliable coherent optical plug that has been ingeniously modified to fit the compact QSFP-DD form factor. The innovative Optical Interoperability Forum (OIF) has designed this cutting-edge standard for data transmission as one of the first standards to establish an interoperable 400G interface. This cost-effective and practical solution is rapidly gaining popularity in the industry due to its unparalleled performance and efficiency in transmitting data.

OIF 400G ZR is limited to 80 to 120 km pt-pt links and best suits the Data Center Interconnect DCI environment. The typical launch is -10dbm, and this is an issue launching into ROADM-based architecture without external amplification.

To take a step further, 400G OpenZR+ MSA expanded 400ZR reaches far beyond 120km by utilizing higher-gain forward error-correction (FEC) coding and increased compensation for chromatic and polarization mode dispersion. This enabled the use of 400G pluggables not only in hyperscaler environments but also in metro/regional service provider network environments. Service providers have thus focused on these "plus" versions of 400G modules for their network needs.

The 400ZR++ is a pluggable solution that makes use of coherent optics. It is capable of extending beyond 120km, with the only difference being a higher Tx power range of -1 dBm to +5 dBm. Cox plans to utilize this solution as it is perfect for both Brownfield and Greenfield ROADM architecture. The power output is sufficient to counteract high insertion loss of Add/Drop Colorless -Directionless-Contentionless (CDC) architecture and transmit almost 1000km on the same OLS line system as vendor-specific embedded transponders.



Figure 5 – 400G ZR ++ plug QSFP-DD form factor.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

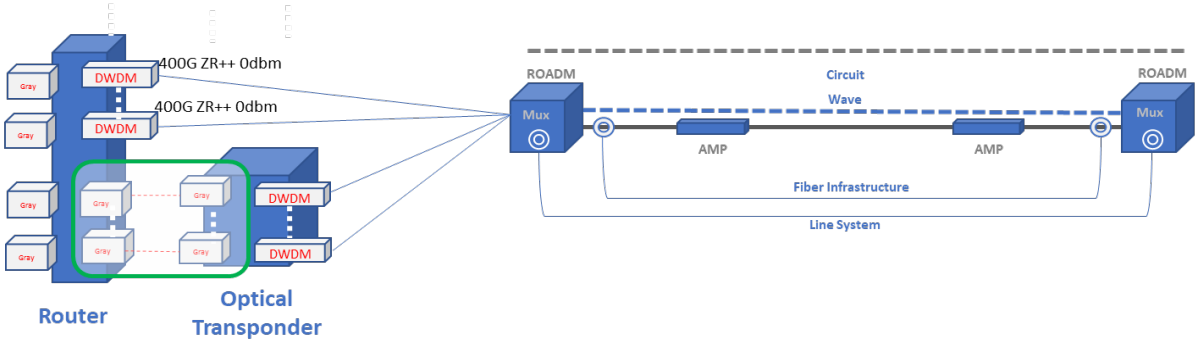


Figure 6 – 400G ZR ++ in Prod

In the Cox production, the implementation displayed in the figure above demonstrates where the 400G ZR++ pluggables will be utilized along with the Open Line System (OLS). The 400G ZR++ will be directly plugged into the line card of the routers, eliminating the need for an optical transponder and grey 400GE Ethernet client back-to-back plugs.

5. Live Production testing and Details.

In Cox, we put the 400G ZR ++ high power optics with -1dbm to +5dbm to test in a live Backbone production environment on very challenging routes over two different vendor line systems: one is OLS with Cox standard, and the other is a fixed older system with a Multicast Switch (MCS) add-drop architecture.

1. LA-LV route 522 km

On the Los Angeles to Las Vegas route, we have a combination of Large Effective Area Fiber(LEAF) fiber type and Single Mode Fiber (SMF) fiber type with distances between amplifiers of 110km each with the combination of Erbium Doped Fiber Amplifier (EDFA) and RAMAN amplifier chains which use stimulated Raman scattering.

The line system used was vendor A with OLS line system and twin Wavelength Selective Switches (WSS) which is commonly named as MXN add drop architecture.

Line System vendor B with MCS add/drop architecture with fixed spectrum.

This route was over 4 ROADM hop architecture.

2. LA-PHX route 1130 km

On the Los Angeles to Phoenix route, a Combination of TWC, LEAF, and SMF-28 fiber with Optical Ground Wire (OPGW) aerial fibers with 8 ROADM hops.

The same line system characteristics are utilized as in the previous LA-LV link.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

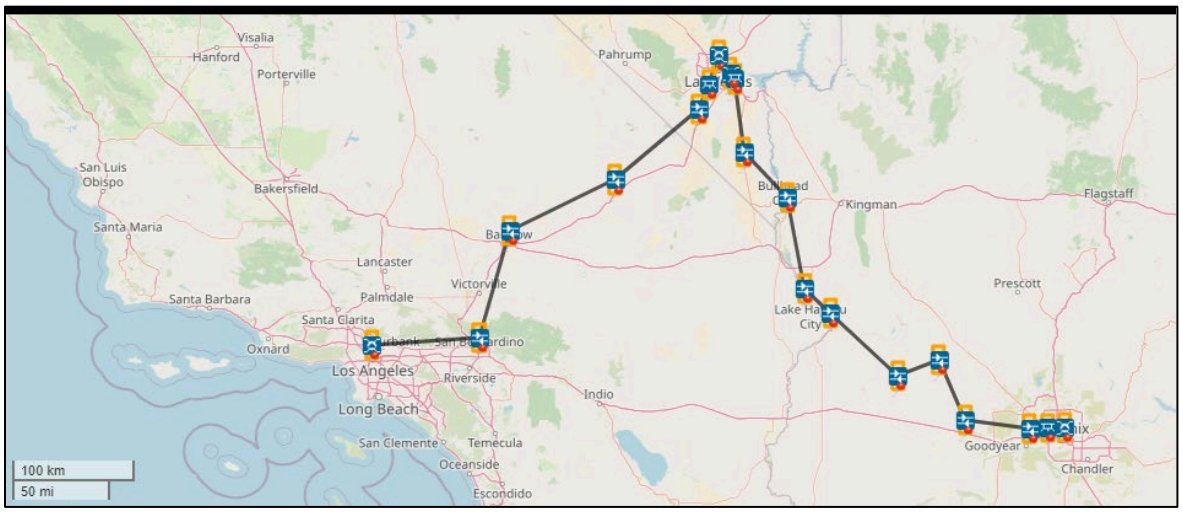


Figure 7 – Route LA-PHX Production test

Each pluggable vendor was allowed to bring their own host (i.e., the router-like devices where the 400G ZR++ are inserted), which they know will operate without issues and can be configured.

We used a 100G test set in Las Vegas and Phoenix to measure the ethernet frames for data integrity with a loopback at the LA side.

The testing methodology used a designated port on the Add/drop structure on each line system and tested frequencies in the different ranges – starting in the lower frequencies around 193Thz, gradually moving to the center C band frequencies, and then to higher frequencies around 196Thz – while keeping power constant at around -1 dBm to 0 dBm for optimal performance.

Also, we kept the spectral width – including guard bands – constant at 75Ghz, although some vendors needed an extra width of 87.5Ghz with their proprietary FEC and higher baud rate to get the highest performance from the plug.

5.1. Results of the test

Vendor 400G ZR++ 0dbm	Freq Low 193-194Thz	Freq Center 194-195Thz	Freq High 195-196Thz	Ckt Width (GHz)	Notes
Q	Pass	Pass	Fail	75	300G Test pass on higher freq
Y	Pass	Pass	Fail	87.5	300G Test pass on higher freq
G	Pass	Pass	Fail	87.5	300G Test pass on higher freq

Figure 8 – Results LA-LV

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Vendor 400G ZR++ 0dbm	Freq Low 193-194Thz	Freq Center 194-195Thz	Frequency High 195-196Thz	Ckt Width (GHz)	Notes
Q	Fail	Fail	Fail	75	300G Test pass
Y	Fail	Fail	Fail	87.5	300G Test pass
G	Pass	Pass	Fail	87.5	300G pass on failed High freq

Figure 9 – Results LA-PHX

In the test over a live production environment, we observed that the pluggable vendors could close the span consistently. We were very impressed with the performance of every vendor we tested.

Before we dive into the results, let's review a couple of important metrics that Cox uses to determine its field deployable standard. The Pre-FEC BER threshold is the highest error rate at which the FEC algorithm can still provide nearly error-free communication after decoding. A Pre-FEC BER of 1.8 E-3 or higher is considered good and approved for field deployment. The Optical Signal to Noise Ratio (OSNR) measures the ratio of signal power to noise power in an optical channel. A higher OSNR is better, and generally, OSNR values above 24 dB are acceptable and approved for field deployment.

The Los Angeles to Las Vegas span was short in distance, and all vendors passed the lower and center frequencies with good Pre-Fec BER and OSNR margins. The results could have been more consistent in the higher frequencies, with some vendors passing the test error-free, but the OSNR values were not up to Cox standard to be field deployable.

The challenging span between Los Angeles to Phoenix is 1130km, where only one vendor could close the link with acceptable Q value but failed at higher frequencies, i.e., 195Thz and above. Then, the data rate was shifted to 300G, and all plugs could close the link with acceptable Pre-fec BER.

The biggest issue we discovered over our both-line system is that the 4000G ZR++ failed over higher frequencies, i.e., 195Thz range and above. The Pre FEC and Q values were not production-ready to be deployed and were very unstable, with inconsistent data integrity.

This issue can occur due to greater nonlinear penalties and ripple effects at higher frequencies, which causes Optical Signal to Noise Ratio (OSNR) to drop.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Architecture Changes

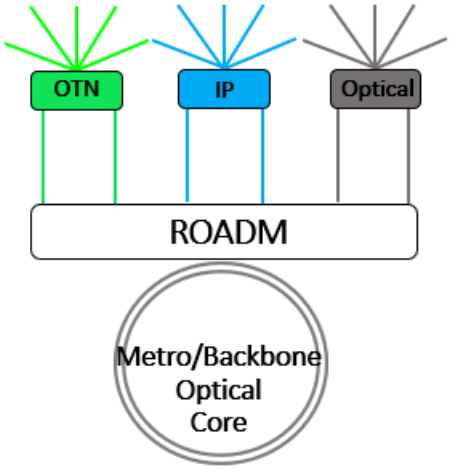


Figure 8 – Cox OLS Network Architecture

Cox's architecture has undergone a major change with the introduction of the Open Line System (OLS) at the foundational optical transport layer Zero (L0). The OLS must be scalable and future-proof to support the next generation of high-capacity, higher baud rate embedded transponders and pluggable optics. OLS includes key architecture components, such as ROADMs with Flex grid capabilities that can accommodate any data rates and modulation format. Additionally, C+L technology is used to future-proof the network from the ever-increasing bandwidth demands, providing 2x the bandwidth in the same fiber strands.

We will keep the OTN layer, IP layer with routers, and Optical transponders layer separate to eliminate any single point of failure while maintaining service termination and delivery separate.

We can optimize each layer separately and scale it according to its lifecycle and capacity demands. This gives Cox the advantage of being flexible and highly competitive if the demands or scope changes.

A ROADM optimized core optical layer is the best architecture for Cox multi-layer and multi-service network, which warrants flexibility and not vendor lock-in or a particular architecture locked in compared to Router HOP-HOP architecture.

400G ZR is a technology option and not an architecture change, and it's another transponder technology that will be used wherever the distances are less than 1000km in Backbone depending on the ROADM hops. Embedded transponders, which are highly designed with long haul links for distances above 1000 km and can overcome multiple ROADM hops filtering penalties, will be used for distances above 1000km and challenging multiple ROADM links.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

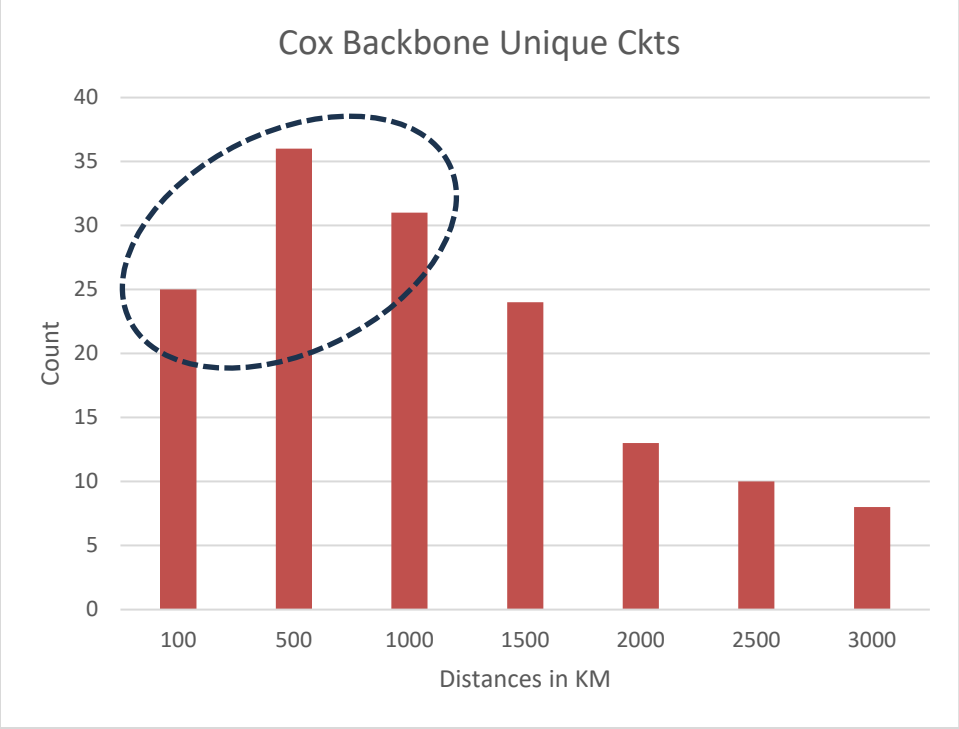


Figure 10 – Target implementation of 400G ZR ++ in Production

As we have tested, we can close the links within close to 1000km, but it all depends on the filtering penalties and the number of ROADM hops it takes.

The above Graph shows that close to 55% of the Cox Backbone circuits are candidates to close with 400G ZR++. It can significantly change how we design our circuits going forward because of the price compression, space, and power reduction we get using pluggable optics.

6.1. IP Layer Architecture Change

With the move to 400G optics, Cox is reevaluating its backbone router network topology. Today, the routed Backbone mixes point-to-point links and express circuits. All the links comprise various quantities of bundled 100G interfaces, with the highest around 30x100G (3Tbps). Moving to 400G interfaces will reduce the number of physical links by 75%. Some of the express circuits may be removed to take advantage of the space, power, and cost savings of using 400G ZR+. Another benefit of eliminating express circuits will be reducing the complexity of Shared Risk Link Group (SRLG) configurations on the routers. When you have multiple A-Z router bundles traversing the same physical span, your SRLG configurations need to be 100% accurate so you can avoid issues during failure events.

7. Operational Challenges

We have observed that in live production testing on different frequency spectrums, the performance is different because of nonlinear penalties at higher frequencies and ripple effects at different frequencies. This makes engineering alien 400G ZR++ waves through the OLS system challenging and will limit the distances it operates.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Line system vendors will need to develop robust engineering planning tools to provide engineering feasibility of these optical trunks with provided PSD (power spectral density).

End-to-end connectivity visibility is lost between the transponder, 400G ZR++ pluggable optics, and ADD/Drop card, making it harder for the operational team to troubleshoot. Also, there is no LLDP support since the grey ethernet optics with back-to-back are eliminated.

Currently, there exists no algorithm to manage power distribution between the line system and pluggable optics. This poses a challenge in adjusting power levels automatically in the event of a loss due to fiber issues and patch panel challenges. To minimize any fiber impairments, it is advisable to position the pluggable optics in the router as close as possible to the OLS transport.

Standards need to be accepted and implemented by both Optical and IP router vendors and work together to expose Open Config Common Management Interface Specification (CMIS) standards consistently.

Service provisioning in OLS and IP config is a challenge where the end-to-end to point and click experience through EMS is lost.

Service and Fault management is broken in optical vendor EMS, and it needs to be stitched together through an orchestrator/controller or home-grown tool to make it operational and workable.

Host router application selection (appsel) data for 400G ZR++, i.e., vendor-specific configuration for maximum performance with proprietary FEC and baud rate, must be supported and verified by the router vendor. Hence, it is part of the software configuration.

8. Conclusion

The optimal strategy for Cox is a ROADM architecture with a Flex grid and C+L system with optimized ROADM bypass. Using 400G ZR++ in the router directly will enable IPoDWDM, saving space and power and reducing cost per bit.

Cox will continue to deploy OLS ROADM infrastructure and use 400G ZR++ where applicable, i.e., for distances below 1000 km.

400G ZR QSFP-DD form factor pluggables are mature and ready for production deployment. Cox is confident that this will be deployed in large numbers in the coming years and integrated into the Cox Backbone and Metro networks.

This architecture with OLS gives Cox communications maximum flexibility by using embedded optics transponders for long-haul challenging links over 1000 km with multiple ROADMs and 400GZR ++ for short-haul links, which make the bulk of Cox Backbone and Metro links for maximum optimized gains and lowest Total Cost of Ownership.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

appsel	application selection
bps	bits per second
CMIS	common management interface specification
DCM	dispersion compensation module
DWDM	dense wavelength division multiplexing
DCI	data center interconnect
EMS	element management system
EDFA	erbium doped fiber amplifier
FEC	forward error correction
Gbps	gigabits per second
Ghz	gigahertz
HD	high definition
IP	internet protocol
IPoDWDM	internet protocol over dense wavelength division multiplexing
ILA	inline amplifier
L0	layer 0
LEAF	large effective area fiber
MCS	multicast switch
NRZ	non return to zero
NPC	national peering center
NDC	national data center
OLS	open line system
OSNR	optical signal to noise ratio
OTN	optical transport network
OPGW	optical ground wire
RDC	regional data center
ROADM	reconfigurable optical add-drop module
RCC	regional caching center
SCTE	society of cable telecommunications engineers
SMF	single mode fiber
THz	terahertz
Tbps	terabit per second
TCO	total cost of ownership

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Data Collection, Interpretation Methodologies, and Challenges for Proactive Network Maintenance

A Technical Paper prepared for SCTE by

James Medlock
Founder & CEO
Akleza, Inc.
18695 Pony Express Dr 2350, Parker, CO 80134
+1-303-670-7951
jmedlock@akleza.com

Ron Hranac, SCTE

Allen Maharaj, Rogers

Alexander Podarevsky, Promptlink Communications

Jason Rupe, CableLabs

Foad Towfiq, Promptlink Communications

Brady Volpe, Volpe Firm

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 4
2. Data Sources and Protocols ... 4
3. Data Collection ... 5
   - 3.1. Overview ... 5
   - 3.2. Cable Modem Data Collection ... 6
   - 3.3. CMTS Data Collection ... 7
4. Downstream Data Use Cases ... 8
   - 4.1. Cable Modem Spectrum Capture ... 8
   - 4.2. OFDM RxMER per Subcarrier ... 9
     - 4.2.1. LTE/5G Ingress Detection ... 10
     - 4.2.2. Profile Management Application ... 12
5. Upstream Data Use Cases ... 14
   - 5.1. Pre-equalization ... 14
   - 5.2. UTSC ... 15
     - 5.2.1. Overview ... 15
     - 5.2.2. freeRunning Mode of Operation ... 16
     - 5.2.3. idleSID Mode of Operation ... 16
     - 5.2.4. Dynamic Triggers ... 17
     - 5.2.5. Smart Amplifiers ... 17
   - 5.3. OFDMA RxMER per Subcarrier ... 18
     - 5.3.1. PMA ... 18
6. Bulk Data Collection Challenges and Recommendations ... 19
   - 6.1. Operational Considerations ... 20
   - 6.2. Security and Reliability Considerations ... 20
   - 6.3. Deployment Considerations ... 21
   - 6.4. New Data Transfer Methods ... 21
7. Conclusion ... 22

Abbreviations ... 23

Bibliography & References ... 24

List of Figures

Title **Page Number**

Figure 1: DOCSIS 3.1 PNM data collection sequence diagram ... 6
Figure 2: Example of spectrum capture display and common elements (courtesy Akleza) ... 8
Figure 3: Examples of RF spectral impairments ... 9
Figure 4: OFDM RxMER per subcarrier channel from a live plant (courtesy Nimble This) ... 9
Figure 5: LTE/5G interference impacting OFDM RxMER per subcarrier (courtesy Akleza) ... 11
Figure 6: Localizing ingress faults using RxMER per subcarrier data (courtesy Akleza) ... 12
Figure 7: Downstream RxMER per subcarrier of OFDM channel with suggested PMA profile (purple line) (courtesy: OpenVault) ... 13
Figure 8: Example UTSC spectrum trace (courtesy Akleza) ... 16
Figure 9: UTSC during channel quiet time (courtesy Comcast) ... 17
Figure 10: Upstream RxMER per subcarrier of OFDM channel with suggested PMA profile (purple line) (courtesy: OpenVault) ... 19

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 11: TFTP upload process 20

List of Tables

Title **Page Number**

Table 1 – Data, Sources and Protocols 5
Table 2 – Mapping downstream RxMER to supported QAM modulation order 10

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Proactive network maintenance (PNM) measurements have been a part of DOCSIS® specifications for over a decade now, but data collection and interpretation are still a challenge for many, both vendors and operators. The DOCSIS specifications define how to request measurement data and the format within which it is to be returned but stop short of properly outlining their use. Furthermore, vendor implementations vary, and newer DOCSIS architectures have added their own challenges to this problem. As PNM tools evolve, we have learnt that our ability to gather the data can be impacted by the collection mechanism in ways we as an industry are unable to qualify in documentation without defining use cases.

The PNM Working Group (PNM-WG) at CableLabs, in conjunction with SCTE Network Operations Subcommittee Working Group 7 (NOS WG7) on PNM, continues to evolve the recommended best practices for this data to assist with troubleshooting network issues. As the network ecosystems are further investigated, gaps are identified that create additional concerns to be addressed. One such gap is that all the information available does not guarantee network equipment behaves as designed when faults occur. Some uses for this data require more than has been specified or described.

To help the industry take full advantage of PNM data, the PNM-WG at CableLabs continuously develops use cases, guiding interested parties in the capture and utilization of PNM data to the benefit of both vendors and operators alike.

In this paper we review the available data collection methods and discuss several PNM use cases to help the reader understand how PNM data can be used in the real world, identify challenges based on current implementations, and propose improvements to the collection methods and PNM data where needed.

2. Data Sources and Protocols

PNM was originally developed to derive in-channel frequency response (ICFR) using upstream pre-equalization coefficient data to identify response impairments. The pre-equalization coefficient data is collected for upstream channels from the cable modems and CMTS using SNMP and then processed by a PNM server. For more information on PNM's use of pre-equalization data see the PNM Best Practices document [4].

A few years after the launch of DOCSIS 3.0 technology, support for cable modem spectrum capture was added to the PNM data available from cable modems, providing a downstream spectrum capture capability.[1] For more information in the use of cable modem spectrum capture see SCTE 280 2022 Understanding and Troubleshooting RF Spectrum [4].

The development of DOCSIS 3.1 added more PNM data, based on tests that can be executed on the cable modem and CMTS. The primary PNM tests supported and in use today provide RxMER per subcarrier data for OFDM and OFDMA channels, and upstream triggered spectrum capture data providing return path spectrum data. Additional tests and data are defined, and more details can be found in the PNM sections of CM-OSSIv31 [1], CCAP-OSSIv31 [2], and CM-SP-PHY3.1 [9]. Unfortunately, not all the defined tests and data are available yet. As the data becomes generally available, additional uses are expected to be developed but are not discussed in this paper which is focused on data readily available in today's network.

[1] Cable modem spectrum analysis functionality was first defined in version I20 of the DOCSIS 3.0 Operations Support System Interface Specification (CM-SP-OSSIv3.0-I20-121113) in November 2012

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 details the primary PNM data, its source, and collection protocol for commonly collected PNM data types.

Table 1 – Data, Sources and Protocols

Data Type	Source	Protocol
Pre-Eq	CM & CMTS	SNMP
Spectrum Capture (D3.0)	CM	SNMP
Spectrum Capture (D3.1)	CM	TFTP
OFDM RxMER	CM	TFTP
OFDMA RxMER	CMTS	TFTP[2]
OFDMA RxPower	CMTS	TFTP
UTSC	CMTS	TFTP or L2TP pseudowire from RPD

3. Data Collection

3.1. Overview

One of the first challenges in developing PNM applications is the collection of data from a potentially large number of sources in the network. CableLabs developed a DOCSIS Common Collection Framework (DCCF), part of a common collection framework [7], that can be used as a starting point for operators to build their own collection application. Several vendors have also developed and provide their own collection frameworks as part of commercially available PNM applications.

Developing scalable SNMP collection applications has been fundamental to network management platforms for decades. DOCSIS 3.1 introduced a new bulk-data transfer mechanism that uses TFTP to upload PNM data to a destination TFTP server. A PNM application first uses SNMP to configure the data being requested and configures the destination TFTP server where the data is sent. This configuration interface is defined in the DOCSIS OSSIv3.1 specifications [1], [2].

Figure 1 shows an SNMP configuration and TFTP transfer sequence diagram for the retrieval of OFDM RxMER data from a cable modem. A similar sequence would be used to request other types of PNM data.

[2] At least one CMTS implementation implements a Kafka bus streaming telemetry mechanism for bulk upload of OFDMA RxMER data.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

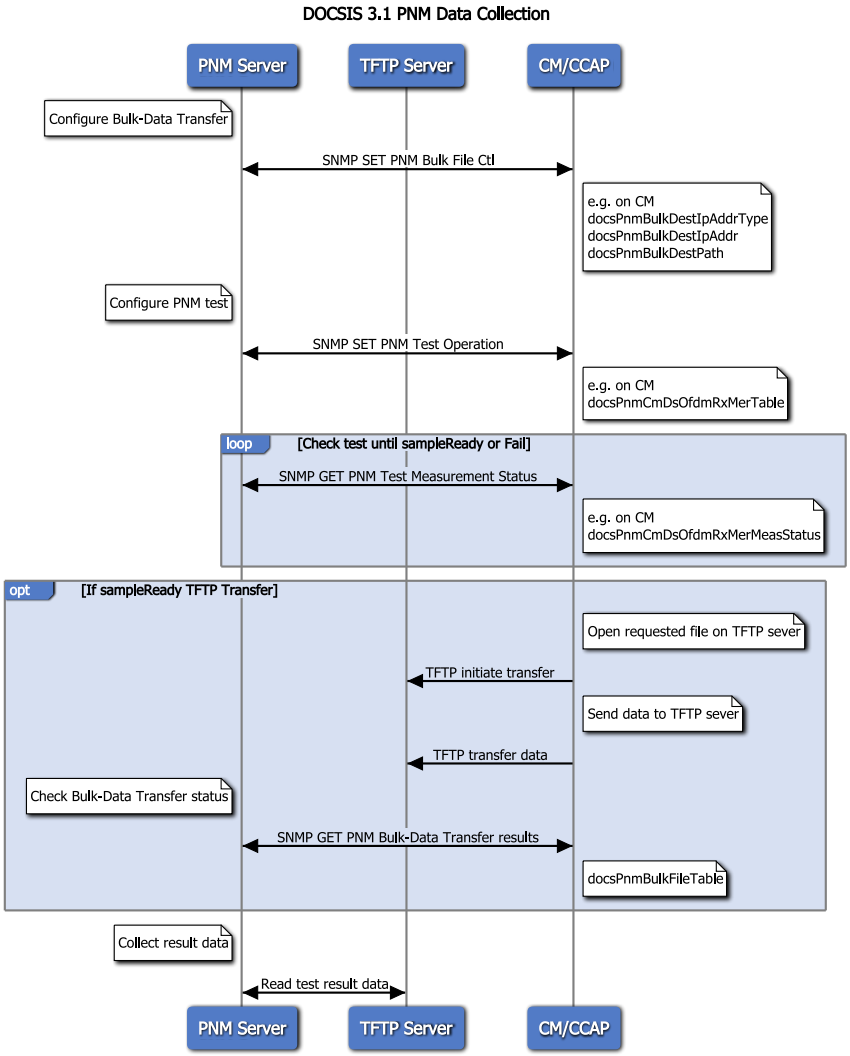


Figure 1: DOCSIS 3.1 PNM data collection sequence diagram (from [3])

3.2. Cable Modem Data Collection

When scaling the collection of PNM data from cable modems for pre-equalization, cable modem spectrum capture, and OFDM RxMER data, we are collecting small amounts of data from many individual endpoints. This allows the creation of a collection framework that scales horizontally, by adding compute resources that handle multiple groups of devices simultaneously. This scaling provides multiple parallel threads within a single compute resource and allows resources to be added as needed. In today's virtualized compute environment, it is also possible to dynamically add, modify, and delete compute resources as required to handle the workload. By implementing a multi-threaded, and multi-compute engine approach, it is possible to scale the data collection of cable modem PNM data from thousands to millions of devices, multiple times a day without creating undue risks or impacting the network negatively.

For DOCSIS 3.1 data such as OFDM RxMER per subcarrier, the cable modem pushes data via TFTP. The same multi-threaded, multi-compute engine environment will also work. However, consideration of

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

system level resources to handle the incoming TFTP requests is required. Because each compute engine has a limited ability to handle multiple incoming TFTP requests, a larger number may be required to support this functionality. However, an individual CM cannot run the test for RxMER per subcarrier and spectrum capture at the same time.

3.3. CMTS Data Collection

Scaling PNM data collection from CMTSs presents a different challenge as we are now collecting a large amount of data from a small number of endpoints. The collection framework must be cognizant of CMTS resources, and CMTS implementations put limits on the number of requests and/or data types that can be retrieved simultaneously.

For OFDMA RxMER per minislot or subcarrier data, the DOCSIS OSSI configuration interface is defined such that data is requested for a single cable modem on an OFDMA interface at a time, by specifying the MAC address of the modem. This means that, for the OFDMA channel, a PNM application must serially request and wait for the data for each cable modem. Depending on the number of cable modems on the channel, this may take a significant amount of time. It may be possible to request the data for cable modems on different channels in parallel, but different CMTS implementations may limit the ability to do this, or require requests be from channels on different line cards.

As described later in this paper, RxMER data used for performance analysis and PMA applications requires the collection of data on common channels to be collected as closely as possible chronologically, so that the results can be compared at nearly the same time, under the same conditions. In this case, the CMTS becomes a bottleneck which a PNM application cannot directly mitigate to scale its data collection. To alleviate this limitation, alternative streaming telemetry mechanisms are being investigated by some CMTS vendors to provide substantially more data at an increased frequency.

For upstream triggered spectrum capture (UTSC) data, the PNM section of the CCAP-OSSIv3.1 specifications defines a configuration interface that allows a PNM application to request upstream spectrum data from an upstream port which is then pushed via TFTP to a PNM server. Based on available resources, the CMTS restricts how many simultaneous captures can run on both a line card and the overall chassis. As with OFDMA RxMER per subcarrier data, and based on constraints around the number of simultaneous requests active on the CMTS, the load and requirements on the PNM application infrastructure are easily managed. If required, additional compute engines can be deployed to handle groups of CMTSs, receiving the UTSC capture files for requested data.

While the CCAP-OSSIv3.1 specification does not distinguish between ports on integrated line cards and RPDs from a configuration perspective, the R-PHY specification does describe an alternate delivery mechanism for RPD upstream ports. For RPDs, the specification defines use of a dynamic pseudowire directly to transfer data from the RPD to the PNM application using an L2TP protocol. The L2TP data stream can result in a data transfer rate of 100 Mbps to 120 Mbps per actively monitored port. As a result, the PNM application, and the data links between the application and the RPDs, must be engineered to handle the aggregate rate from multiple L2TP data streams. For this specific use case, multiple compute engines across a DAA network may be required to effectively receive and process the data.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. Downstream Data Use Cases

4.1. Cable Modem Spectrum Capture

Cable modem spectrum capture allows operators to capture and analyze the downstream RF spectrum at the cable modem. A spectrum capture capable modem provides the equivalent of a spectrum analyzer in every home, a powerful tool which allows remote RF power measurement without deploying a technician or requiring access to the cable modem's physical location.

Figure 2 shows how spectrum capture data may be displayed, and common elements typically seen in a downstream spectrum.

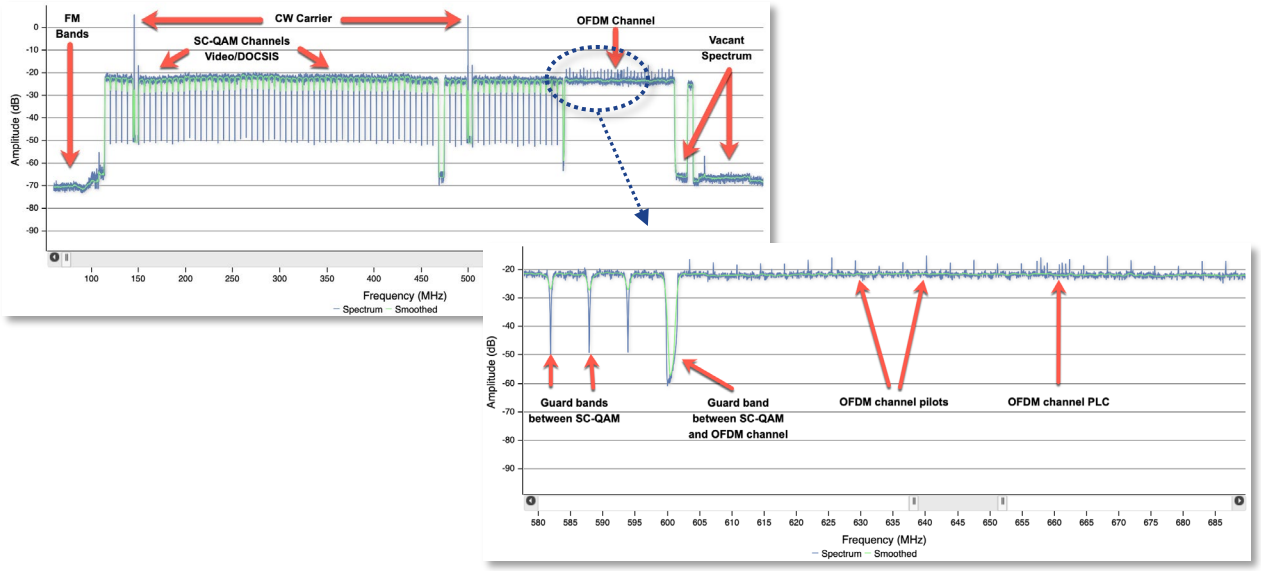


Figure 2: Example of spectrum capture display and common elements (courtesy Akleza)

Within the CableLabs PNM-WG, some members have collaborated on a software library that automatically detects common RF impairments. Additionally, some PNM application vendors have developed their own libraries and techniques offered as part of their commercial PNM applications. Commonly, this software is known as spectral impairment detection (SID) and can assist field technicians in identifying the type and potential cause of impairments impacting the downstream spectrum.

Figure 3 shows several common impairments and how they appear in a spectrum capture trace.

SCTE's NOS WG7 has developed an industry reference, SCTE 280 2022, Understanding and Troubleshooting RF Spectrum [8], that discusses the fundamentals of cable RF spectrum and provides an in-depth analysis of these common cable impairment types and their typical causes and resolution.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

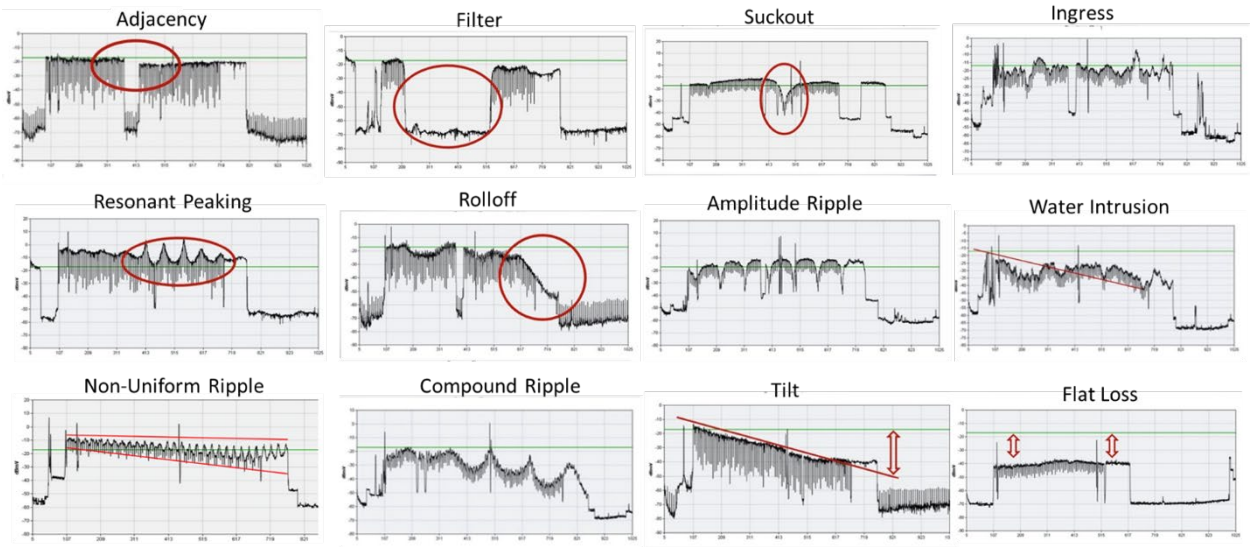


Figure 3: Examples of RF spectral impairments

4.2. OFDM RxMER per Subcarrier

With the release of the DOCSIS 3.1 specifications, there were many useful PNM tests added. One very powerful test added for OFDM (and OFDMA) is called RxMER per subcarrier. RxMER per subcarrier, provides a modulation error ratio measurement for each subcarrier in the OFDM or OFDMA channel. This means if an OFDM channel has 7600 subcarriers, data collectors will receive 7600 RxMER results estimating the signal quality of each subcarrier. When plotted, it creates a visualization like that of a spectrum analyzer but has a different meaning. See Figure 4 for an example graph of OFDM RxMER per subcarrier.

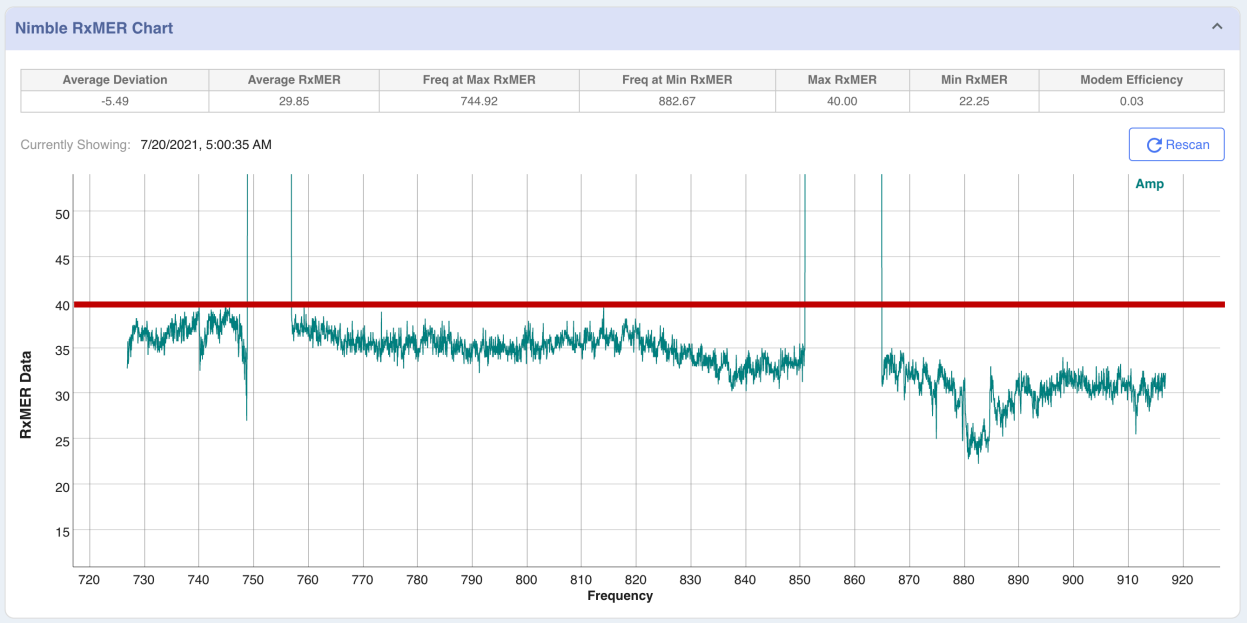


Figure 4: OFDM RxMER per subcarrier channel from a live plant (courtesy Nimble This)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In Figure 4, there are a couple of unique observations one can make.

First, the average RxMER is 29.85 dB, as shown in the top of the chart. A red line is added at 39 dB RxMER, which is the minimum threshold for 4096-QAM. Ideally, all RxMER per subcarrier datapoints should be above the red line to provide the highest level of data speed to the subscriber. If the CMTS was configured to provide 4096-QAM, this modem would not be able to receive any data. See Table 2 for the mapping of RxMER per subcarrier to OFDM modulation order.

Secondly, there are two sections of the RxMER per subcarrier plot where the chart goes above 50 dB. One is from 749 MHz to 757 MHz and the other between 851 MHz and 865 MHz. These are called exclusion bands which have been configured by the operator to exclude use of areas of the spectrum that may have interference or poor performance.

RxMER per subcarrier can be directly mapped to the modulation order which can be supported by a receiving cable modem. These mappings are defined in [PHYv3.1] (*see Table 46 - CM Minimum CNR Performance in AWGN Channel*) and [1] (*see Table 72 - CmDsOfdmRequiredQamMer Object*).

Table 2 shows these mappings up to 16384-QAM, but field testing has indicated that these values are conservative. As indicated in Figure 4, 39 dB is often used as the lower RxMER limit for 4096-QAM rather than the CableLabs recommendation of 41 dB for 4096-QAM.

Table 2 – Mapping downstream RxMER to supported QAM modulation order

Constellation/Bit Loading	CNR/MER (dB)
16-QAM	15.0
64-QAM	21.0
128-QAM	24.0
256-QAM	27.0
512-QAM	30.5
1024-QAM	34.0
2048-QAM	37.0
4096-QAM	41.0
8192-QAM	46.0
16384-QAM	52.0

4.2.1. LTE/5G Ingress Detection

While it is possible to deploy one or more OFDM channels anywhere within the downstream frequency spectrum, it is common for at least one OFDM channel to be located at the higher end of the spectrum, typically above 600 MHz. This frequency range is also where cellular LTE and 5G signals are transmitted over-the-air, so any cable shielding integrity faults may result in interference on the OFDM channel from the cellular signals.

Like cable modem spectrum capture impairment detection, it is possible to identify impairments in RxMER per subcarrier data, including the presence of interference caused by LTE or 5G ingress.

Figure 5 shows an example of a cellular 5G signal that is leaking into the cable network and reducing the RxMER for an approximately 20 MHz span of the OFDM channel, reducing these subcarriers' RxMER and their ability to utilize higher modulation orders.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

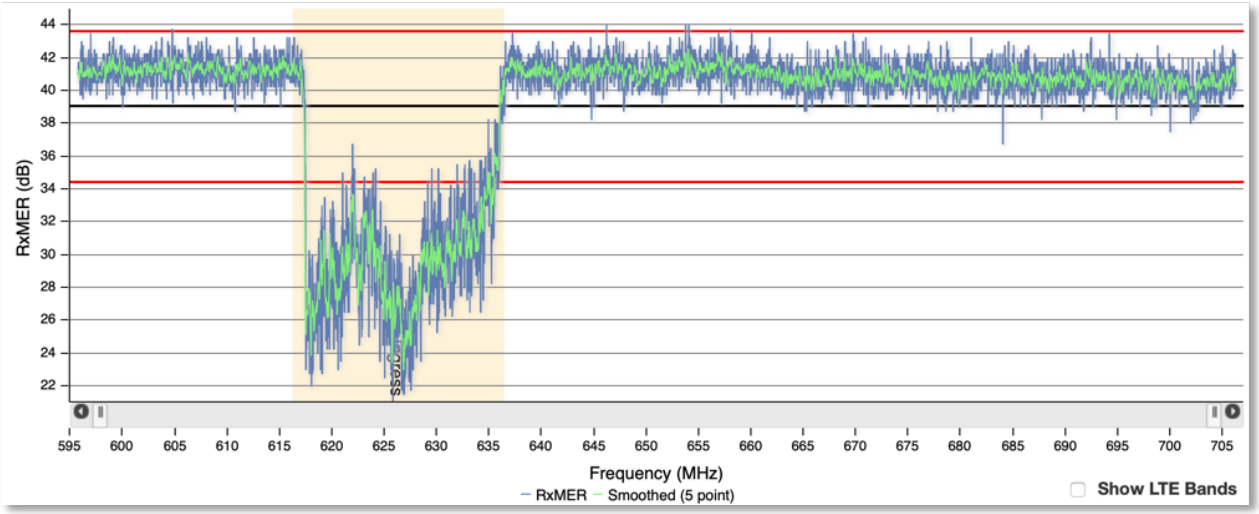


Figure 5: LTE/5G interference impacting OFDM RxMER per subcarrier (courtesy Akleza)

To isolate the location of the ingress further, RxMER per subcarrier data can be examined from several cable modems.

Figure 6 shows an example where RxMER per subcarrier data was analyzed from several cable modems, working upstream from where the ingress was detected.

When an upstream cable modem unaffected by the ingress is identified, the fault can be isolated to the segment of plant between this cable modem and the first cable modem downstream that is impacted. In this example, the ingress was caused by an amplifier housing issue that allowed LTE to leak into the downstream signal path.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

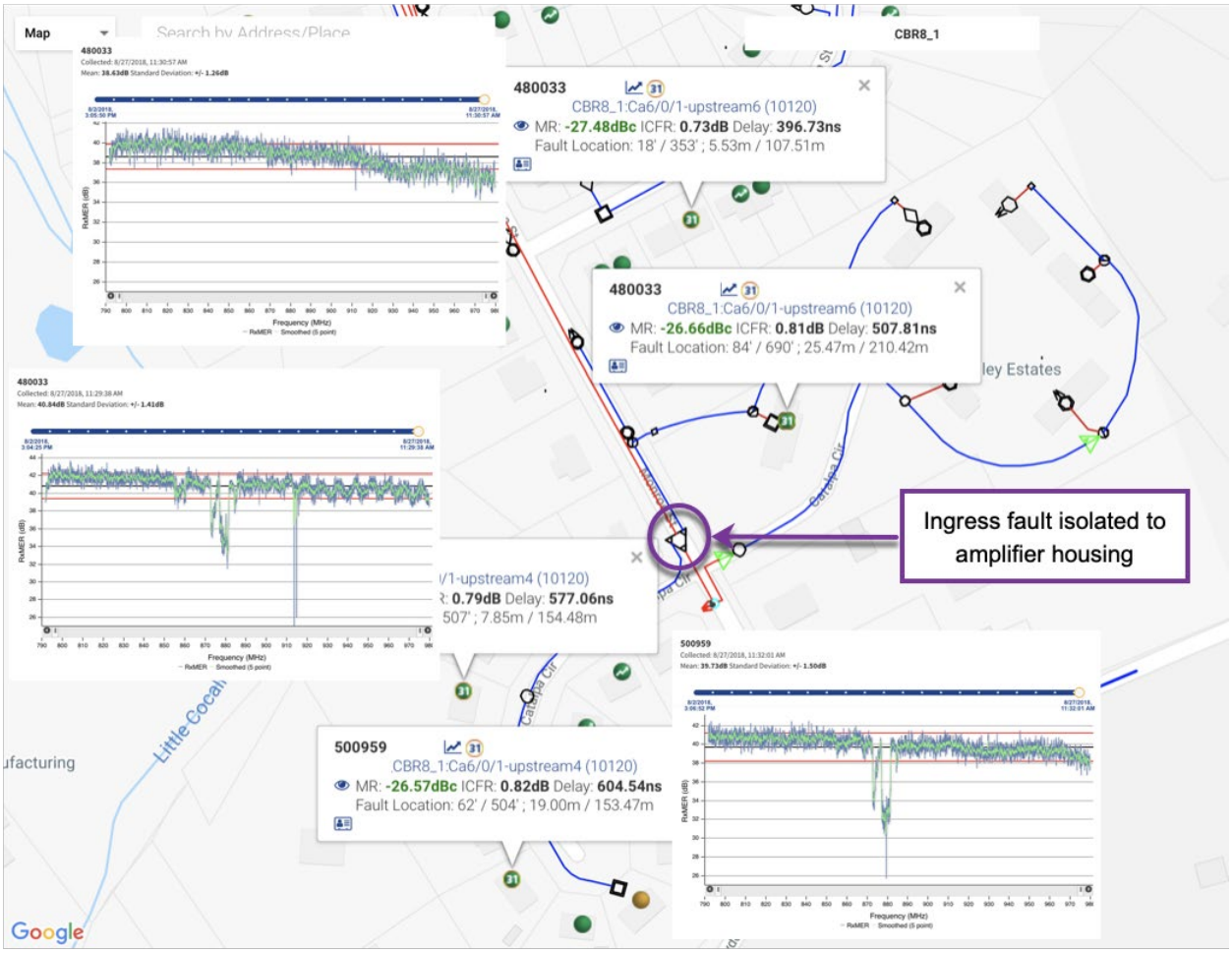


Figure 6: Localizing ingress faults using RxMER per subcarrier data (courtesy Akleza)

4.2.2. Profile Management Application

A profile management application (PMA) typically functions by ingesting RxMER per subcarrier data from a designated node's service area, conducts a detailed analysis of the data, and generates an optimal modulation profile for each cable modem. This procedure aims to augment both the capacity and the robustness of the network.

The improvement in network capacity and robustness is achieved by selectively reducing the modulation order in impaired regions of the OFDM channel, while maintaining higher modulation order(s) within the rest of the channel. This adjustment facilitates cable modems to consistently synchronize with the OFDM channel (contributing to robustness) and to utilize the highest order modulation profiles whenever feasible (boosting capacity).

With this approach, cable operators employing PMA discover a significant increase in the usage of the OFDM channel by DOCSIS 3.1 modems compared to systems not utilizing PMA. This results in the liberation of capacity on the SC-QAM channels for non-DOCSIS 3.1 modems. Subscribers equipped with DOCSIS 3.1 modems often report significantly enhanced download speeds, typically in concordance with their provided speed tiers.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

PMA is a processing engine which analyzes the RxMER data from every cable modem in each node. After analyzing the RxMER data, the PMA engine will provide the optimal modulation profile for each subcarrier and the optimal frequency location for the PLC. This takes the guess work out of determining which profiles and PLC location each node's service area requires. Further, PMA can be fully automated, so that every node's service area can be optimized multiple times per day to compensate for the ever-changing nature of plant conditions and RF impairments.

Figure 7 shows a representation of an RxMER per subcarrier of a downstream OFDM channel starting at 853 MHz and ending at 994.495 MHz. One can observe there are lower frequency RF impairments and a high frequency rolloff as indicated by the light yellow and reddish lines. The more frequent the subcarrier dB reading, the darker the color. These lines represent nine cable modems which have been clustered together, having similar downstream RxMER per subcarrier values.

The PMA engine analyzes this data and creates a near-ideal profile given the impairments observed. This profile is indicated by the purple line, which has a stepped profile. Each step in the profile can be used to generate a command which will be run on the CMTS either manually or automatically to compensate for the impairments.

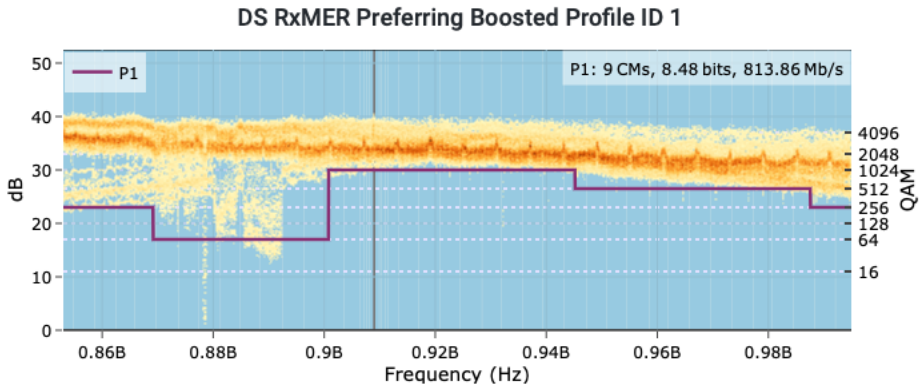


Figure 7: Downstream RxMER per subcarrier of OFDM channel with suggested PMA profile (purple line) (courtesy: OpenVault)[3]

The challenge of modifying OFDM and OFDMA profiles on a CMTS during the PMA process is three-fold, however.

There's a lack of a standardized protocol for carrying out these modifications using SNMP commands, meaning we must use Secure Shell (SSH) to access the CMTS command line interface (CLI). While SSH usage brings certain security considerations, it isn't the core issue.

Each make and model of CMTS often has vendor specific CLI commands for updating the OFDM and OFDMA profiles. This lack of uniformity demands a detailed understanding of the individual commands and procedures for each system. With the necessary SSH credentials in hand, vendors and cable operators face the daunting task of learning the unique syntax and command structure of each respective CMTS.

Once these hurdles have been cleared, an additional challenge emerges. The entire system must be automated to create a closed loop method of operation. In this approach, the PMA engine generates optimal profile recommendations, and the CMTS is then automatically updated with these profiles via the

[3] These charts are heat maps of the RxMER subcarriers for a 24-hour period with multiple poll cycles. The darker the color (red) the more frequent the reading. If there were no variations over time, this would result a straight dark red line.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

CLI. This involves careful scripting, thorough testing, and continuous monitoring to ensure the loop operates flawlessly, further enhancing the complexity of the entire process.

As an initial step and until CMTS technicians are comfortable with the changes that will be made, a PMA can make recommendations which would then be manually applied by the CMTS technician.

PMA has several benefits, which are immediately realizable as follows:

- PMA will provide the optimal profile for a node:
  - PMA even provides modulations on a subcarrier level, though this is not yet supported by CMTS vendors.
- Through profile optimization per node, each node's service area will always see the maximum data throughput:
  - Many cable operators run OFDM conservatively, meaning they may not use 4096-QAM or even 2048-QAM and opt for lower order modulation orders to ensure data throughput to their subscribers. PMA will allow operators to run at the highest modulation possible with confidence.
- PMA can provide PLC placement recommendations to ensure the PLC is not operating in an area with impairments:
  - This is often overlooked as a critical feature of PMA. If the PLC is placed in a region where impairments exist (think about a suckout or LTE ingress), this will result in an outage. PMA will aid in preventing such instances from occurring.

These immediate benefits may seem obvious and exciting, but they have a bottom-line impact to total system capacity, which results in OPEX savings and preventive CAPEX investment.

In 2019, Comcast developed a PMA system for generating and transacting DOCSIS 3.1 downstream profiles tailored to the conditions of each OFDM channel in its network. Some point-in-time metrics from Comcast's deployment of PMA which indicates its realized value [5]:

- 34.3% capacity gain in OFDM profiles (Division A)
- Raw gain of 6020 Gbps for Division A
- 91.0% CM success rate (percent of CMTSs that were successfully configured with updated profiles)

Comcast considered PMA a huge success and subsequently issued a public press release [5].

5. Upstream Data Use Cases

5.1. Pre-equalization

PNM was originally developed based on the availability of pre-equalization coefficient data and the ability to detect and localize frequency response impairments in the return path signal by analyzing this data. The PNM Best Practices document [4] covers the use of pre-equalization in detail including data collection, analysis, and fault localization.

In contrast to SC-QAM channels, an OFDMA channel utilizes a separate pre-equalization coefficient for each subcarrier. Using OFDMA pre-equalization coefficients in conjunction with PNM active/quite probe PNM tests, allows applications to better estimate transmission line characteristics.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The pre-equalization coefficient data available consists of two sets: actual pre-equalization coefficients, and the latest updates as received by the cable modem. Actual pre-equalization coefficients are used by the cable modems adaptive pre-equalizer to adjust the transmitted signal, so it will be received by the CMTS in nearly optimal conditions. For more details see section 6.2 of [3].

Significant changes in the latest update data indicate adjustments to the pre-equalization coefficients instructed by the CMTS. This may be an indication of rapid changes in the transmission line characteristics and may indicate significant noise, ingress interference or cable defects caused by environmental factors (wind, precipitation, vibrations, etc.).

Support for the collection and analysis of OFDMA pre-equalization coefficient data still has some challenges.

First, not all vendors implement this in all firmware versions so coverage can be inconsistent.

Second, requests for the data may be blocked due to other outstanding PNM data requests, such as RxMER per subcarrier and spectrum capture data. It is therefore important for the PNM application to manage these data requests and schedule appropriately to minimize any conflict.

5.2. UTSC

5.2.1. Overview

Return path spectrum monitoring and analysis are important tools that cable operators rely on to troubleshoot RF interference in the upstream spectrum. However, the return funnel[4] effect in cable networks means that locating ingress sources continues to be a challenge, and is largely inference-based, which is manual and tedious. Locating the source of interference is still typically achieved through a “divide and conquer” technique, where return paths are isolated while validating if the interference is still visible.

Traditional return path monitoring platforms have also been hardware based, located in headend facilities, and measuring the RF signal split out using an RF switch or similar. As distributed access architectures (DAAs) are deployed, the RF connection back to the headend has been replaced by a digital fiber connection and therefore no RF signal is available to be measured.

UTSC provides a software solution for monitoring the return path and operates across both traditional RF and DAA environments.

A UTSC application provides a wideband spectrum analyzer function using the burst receiver within the CMTS or RPD to capture the RF signal. Since the receiver has knowledge of the scheduler it can be triggered to examine desired upstream transmissions as well as underlying noise and interference during a quiet period when no devices are scheduled to transmit. It can also be triggered to capture on specific timestamps or time slots granted to a specific MAC address. To date, availability of these trigger capture modes is not consistently implemented across all CMTS vendors for UTSC, but support is being rapidly developed and deployed based on operator and PNM application vendor priorities and requests.

[4] Also referred to as noise funneling. Downstream signal transmission in a cable network is point to multipoint (headend or hub site to multiple subscribers); upstream transmission is the opposite, multipoint to a single point. Upstream interference such as noise or ingress can enter the network at one or more locations and "funnel" back to the headend or hub, hence the reverse funnel effect.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.2.2. freeRunning Mode of Operation

Currently all CMTS vendors implementing UTSC support the "freeRunning" trigger mode. This mode provides a continuous spectrum capture across a specified frequency range and resolution like that of existing visualizations provided by hardware-based platforms and field meters. Using this trigger mode, a UTSC application can support a real-time, live, spectrum analyzer function updating several times per second.

In addition to having scheduler information available, the CMTS also has demodulators that run continuously on active upstream channels. These demodulators can provide additional information for each of the channels to the UTSC application, such as RxMER and FEC statistics. This is useful for determining the impact of impairments and for troubleshooting.

Figure 8 shows an example trace from a UTSC application in a sub-split node. In this example you can clearly see the presence of common path distortion (CPD) with a raised noise floor across the entire spectrum. By integrating the channel metrics from the burst receiver, RxMER, total codewords, correctable codewords, and uncorrectable codewords, we can see the impact the impairment is having on the upstream transmissions.

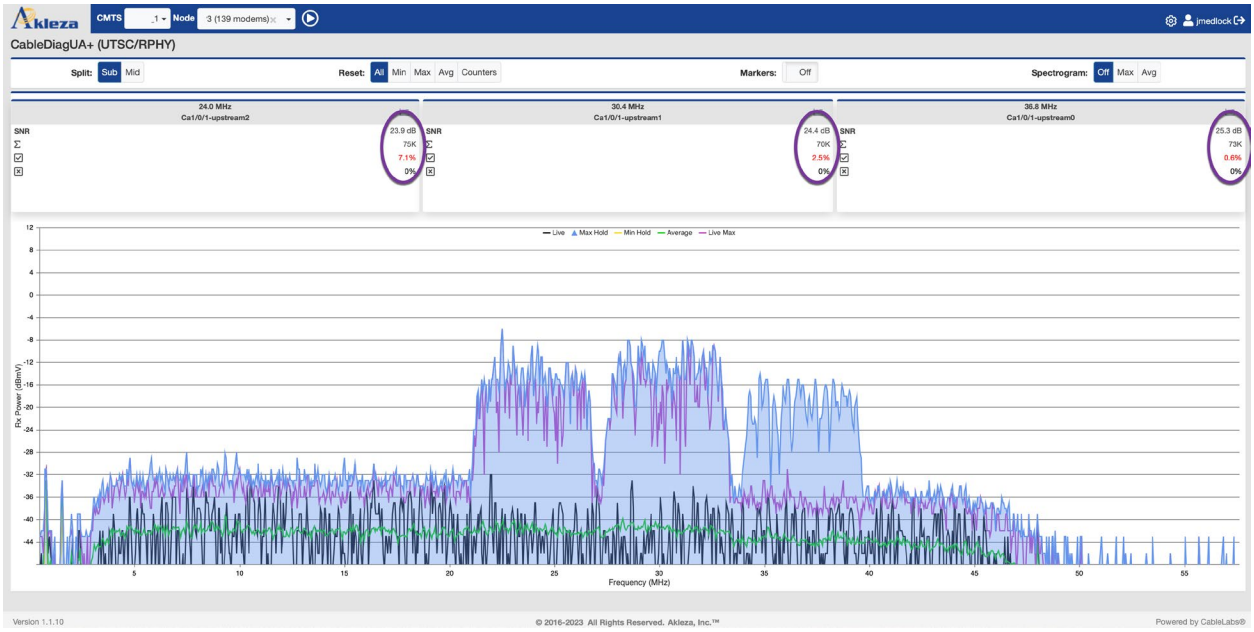


Figure 8: Example UTSC spectrum trace (courtesy Akleza)

5.2.3. idleSID Mode of Operation

While not available from all CMTS vendors, some implementations do support the "idleSID" or quiet time trigger mode. This captures the configured frequency range during a period when a channel is not transmitting which allows you to see any potential noise or interference under a carrier.

Figure 9 shows a UTSC example during a quiet time for the first channel. As is highlighted, this ability allows us to see the ingress under the carrier and again the burst receiver metrics highlight the impact the ingress is having on the channel transmission.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 9: UTSC during channel quiet time (courtesy Comcast)

As a companion to the SCTE 280 2022, Understanding and Troubleshooting RF Spectrum informational report, SCTE NOS WG7 is developing another informational report focused on understanding and troubleshooting upstream impairments. This is expected to be released later this year.

5.2.4. Dynamic Triggers

UTSC can be triggered based on several different factors. Automation and orchestration are key to triggering, capturing, and analyzing upstream spectrum data and making intelligent decisions about the impairments and deployment of resources to repair them. Some options for triggering UTSC captures include:

- Periodic time-based captures as a proactive measurement tool during peak usage times to validate that service level agreements are being met.
- Quality of Service (QoS) metrics-based triggers including but not limited to latency spikes, SNR/MER degradation, cumulative packet loss increases, jitter variations, bandwidth utilization and congestion.

Implementation of these trigger mechanisms requires the necessary compute and data storage infrastructure along with the testing and optimization to make them truly effective. Other considerations include communication protocols available, hardware and software compatibility, security and privacy, error handling, and user feedback, all to consistently improve the algorithms being used to identify and investigate upstream impairments.

5.2.5. Smart Amplifiers

Smart amplifiers, also known as adaptive or intelligent amplifiers, are relatively new to the industry. They utilize advanced technologies and techniques to optimize performance in ever changing environments. They are designed to provide efficient and high-quality amplification while adapting to changing RF conditions, interference, and other factors that can affect signal quality.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Smart amplifiers include features like interference management, detecting the presence of other RF signals, and enhancing the signal quality by optimizing amplification parameters and managing that interference, maintaining higher signal quality, minimizing distortion, and contributing to better overall system performance.

Utilizing UTSC with smart amplifiers involves using event-based triggers that monitor incoming RF signal characteristics and triggering the capture of spectrum data when specific conditions are met. Some of these conditions include signal strength, signal-to-noise ratio, interference levels, or any other relevant QoS metric.

Using the spectrum capture data, adaptive adjustments can be made to the amplification parameters in real time to optimize performance. The smart amplifier can modify gain, frequency selection, and other parameters available to ensure the best signal quality possible.

By allowing the amplifier to make informed decisions using spectrum analysis, it can mitigate potential customer impacts, either negating the need to send a technician, or allowing them the time to arrive onsite and investigate further without additional load on call centers or technical support representatives.

The UTSC technique is particularly useful in environments with rapidly changing RF conditions or where interference sources can impact signal quality. By capturing and analyzing the RF spectrum on demand, the smart amplifier can respond in real-time, ensuring consistent performance, and enhancing the overall quality of communication or signal processing.

This approach may require integration of specialized hardware such as spectrum analyzers or software defined radios (SDRs), and software within the smart amplifier system to operate effectively.

5.3. OFDMA RxMER per Subcarrier

RxMER per subcarrier data for OFDMA channels is important for fault identification, localization, troubleshooting, node health scoring, KPI reporting, and a host of other metrics critical to plant operations. Use of this data is consistent with the use of RxMER per subcarrier data for OFDM channels.

5.3.1. PMA

OFDMA RxMER per subcarrier data can be used by a PMA to optimize modulation profiles for upstream OFDMA channels. Upstream impairments vary frequently in many aspects including intensity and impact, and cable operators may need to reconfigure their OFDMA profiles as much as once per hour which requires obtaining that data at the same frequency. Running RxMER data collection and other PNM tests and queries concurrently is required to meet this business objective; but as discussed previously, this data collection frequency has some challenges.

Figure 10 shows a representation of RxMER per subcarrier data for an upstream OFDMA channel starting at 10 MHz and ending at 22 MHz. One can observe there is a significant amount of RF noise across the spectrum, preventing the modems from operating at 1024-QAM as observed by the light yellow and reddish lines. The more frequent the subcarrier dB reading, the darker the color. These lines represent 123 cable modems which have been clustered together, having a similar reported upstream RxMER per subcarrier.

The PMA engine analyzes this data and creates a near-ideal profile given the impairments observed. This profile is indicated by a purple line, which has stepped profiles. Each step in the profile can be used to generate a command which will be run on the CMTS either manually or automatically to compensate for

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the impairments. In Figure 10, one can observe that the optimal profile modulations are between QPSK and 32-QAM[5] given the level of severe upstream noise. While this is not the ideal 1024-QAM, it will allow cable modems to remain locked to the OFDMA channel and enable those modems to take advantage of the capacity of the OFDMA channel, thus alleviating load from the SC-QAM channels.

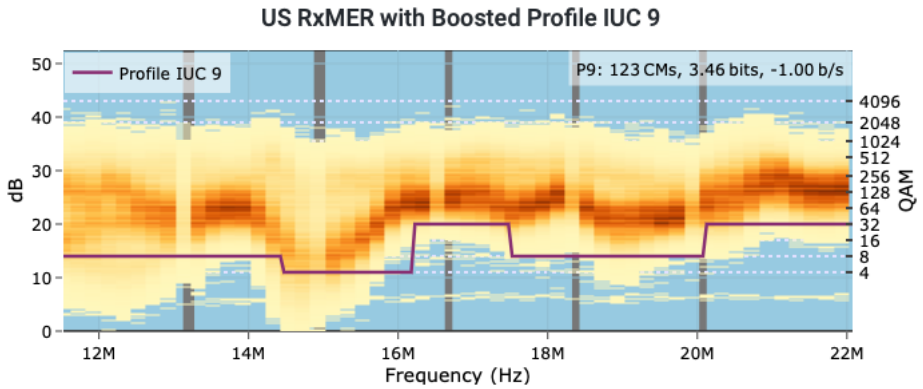


Figure 10: Upstream RxMER per subcarrier of OFDM channel with suggested PMA profile (purple line) (courtesy: OpenVault)[6]

CMTS vendors are currently exploring alternative mechanisms to deliver RxMER per subcarrier data in bulk, including streaming telemetry mechanisms such as a Kafka bus to periodically publish data, making it available to a subscribed PNM collection application. Such streaming mechanisms would benefit applications like PMA and help to meet the requirements of the cable operator.

6. Bulk Data Collection Challenges and Recommendations

There are several challenges with obtaining large amounts of rapidly sampled upstream data. Some examples include:

- Obtaining RxMER per subcarrier data while running any other PNM tests or queries, such as UTSC, or while another PNM application is running, can prevent data delivery. For example, on some platforms one can obtain RxMER but not UTSC data at the same time, and vice versa. This can limit troubleshooting in that more than one problem cannot be analyzed simultaneously, and only a single technician can work on the platform at a given time. Fault management can be urgent, and technician time is valuable.
- Further, on some vendor platforms, it is only possible to obtain data for one modem at a time (single threaded or per modem basis). This results in a significant amount of time necessary to query all needed data on a single CMTS.
- On some CMTSs, obtaining some forms of data can introduce an out-of-service event to end users, which is clearly unacceptable.

[5] This customer was looking for robustness/resiliency/stability on the DS and capacity on the US, resulting in very robust (e.g., low order) modulation profiles

[6] These charts are heat maps of the RxMER subcarriers for a 24-hour period with multiple poll cycles. The darker the color (red) the more frequent the reading. If there were no variations over time, this would result a straight dark red line.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6.1. Operational Considerations

The use of TFTP by the DOCSIS 3.1 PNM bulk-file transfer mechanism may require some network configuration changes if there are any firewalls between the CM or CMTS and the destination TFTP server. TFTP is a UDP-based protocol that uses ephemeral ports to transfer data.

Figure 11 shows the upload sequence in which Ports X and Y are ephemeral port numbers.

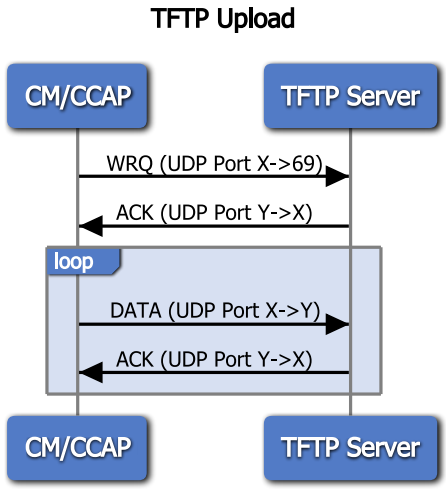


Figure 11: TFTP upload process (from [3])

For the TFTP upload process to complete, any network firewalls between the CM or CMTS and the TFTP server must be configured to allow UDP traffic from the ephemeral port range between the CM or CMTS and TFTP server. Though it may be possible to use source and destination addresses to configure an explicit rule for CMTS devices, it is not practical for CM subnets. As such, rules allowing UDP Port 69 and all UDP traffic in the ephemeral port range destined for the TFTP server are required to support the PNM bulk-file transfer mechanism.

6.2. Security and Reliability Considerations

TFTP and SNMP are currently not required to be encrypted; however, one can add application layer security if desired. For example, an HTTPS proxy can be added in front of data collectors like CCF, which allows encrypted posts and gets.

UDP, which is low priority, is the assumed transport mechanism. As such, a capacity limit is reached if there is too much traffic, and the UDP packets can be dropped, causing a loss of PNM information. Further, there is a general assumption that if a test fails or if information is corrupted in the transfer process, then a retry of the test is warranted. The application layer (in this case the PNM application) is assumed to manage the information reliability.

In addition, CMs can be limited in their processing of PNM information. If a PNM request takes five minutes to complete and is only initiated when requested, then the CM will not likely respond to other requests until the first is completed. Consider this limitation when requesting data from CMs.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6.3. Deployment Considerations

The network needs to have a TFTP server to receive and hold the TFTP files that are responses to the data requests.

Also, firewall rules on the network must allow the TFTP packets to travel from the CM to the server and to any applications that need the data.

Similarly, SNMP requires firewall rules to be configured to allow the traffic in both directions.

Virtual machines (VMs) are commonly used to deploy PNM data collectors such as CCF. The VM should have the compute power and configuration required to adequately support its deployment. It is common to have one PNM data collector instance for each CMTS, but different models are possible and should be considered based on expected use.

In at least one implementation, data and files generated by CMTS can remain in the local memory and would need to be removed via script or command on a regular basis to prevent it from becoming overutilized and unavailable.

6.4. New Data Transfer Methods

Defined in 2021, there is a new bulk data transfer mechanism described in section 7.5 of CCAP-OSSIv31 [2] This mechanism is defined to better facilitate obtaining the increased amounts of disparate data from the CMTS resulting from PNM tests; it is also useful for other large data acquisition requirements.

Based on the experience of previous methods, a small group formed from the CableLabs PNM-WG to define improved methods that would allow more effective data transfer and thus more useful PNM testing involving the CMTS.

In the configuration of a bulk data transfer, the destination for the resulting file is set, along with the transfer protocol of choice: TFTP, HTTP, or HTTPS. When using TFTP, the CMTS must have a TFTP client to complete the file transfer, initiated at the request of the PNM server. gNMI is also an option specified,[7] but it is not supported by DOCSIS 3.1 devices for PNM. However, there is nothing preventing a CMTS vendor from offering it.

---

[7] gNMI (gRPC Network Management Interface) is a protocol used for network management and configuration. It's a protocol developed by Google and is part of the larger gRPC framework, which is a high-performance, open-source remote procedure call (RPC) framework that uses HTTP/2 for transport and Protocol Buffers for serialization. It is specifically designed for network devices and systems to provide a standardized way of managing and configuring them. This method is often used in conjunction with other technologies like YANG models (Yet Another Next Generation) which define the structure and semantics of network data. YANG models provide a common language to describe configuration and operational data for network elements.

gNMI provides many improvements over TFTP and L2TP, including higher efficiency and scalability, secure communication, and asynchronous notifications, allowing devices to send updates to the management system in real-time when changes occur.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Though the working group is also moving toward streaming telemetry using YANG, which significantly improves methods for data acquisition, the updated method discussed here can be implemented in DOCSIS 3.1 through a software update.

For R-PHY networks, UTSC data is delivered to a PNM server over an L2TP pseudowire. While efficient, and a necessary difference due to the nature of R-PHY, vendors have not yet aligned on how to configure data collection on RPDs.

7. Conclusion

As defined throughout this paper, there are many options for collecting PNM related data. While they all have their own requirements and present their own challenges, none are out of reach, and only need the necessary constructs to be deployed.

Downstream data collection provides valuable insight into what is going on and which customers are impacted, especially when utilizing the spectrum capture functionality of D3.0 and above modems and applying OFDM RxMER calculations within a PMA application to create appropriate profiles for each customer to get the most of their service tier.

Upstream data collection challenges seem daunting, being limited by CMTS implementation and a singular device supporting thousands of customers, but new data transfer mechanisms are being developed to allow the required data to be transferred are rates required for real time measurement and network adaptation.

UTSC, a novel feature that allows operators and vendors alike to capture and analyze upstream data and use it to their advantage without having eyes on glass during impairments, changes the landscape significantly. It allows for the possibility of remote configuration changes within node housings that support digital RF boards and smart amplifiers, enabling frequency optimization and adaption to impairments, potentially mitigating customer impacts and allows technicians the time to troubleshoot.

While collection frameworks and data transfer mechanisms have their unique considerations, as we evolve so will they. Also, as new use cases are defined, and new best practices are published, our collective efforts will benefit both ourselves and our customers experience.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

4G	fourth generation [mobile telecommunications technology]
5G	fifth generation [mobile telecommunications technology]
ADC	analog-to-digital converter
AGC	automatic gain control
CCAP	converged cable access platform
CMTS	cable modem termination system
CPE	customer premises equipment
CSV	comma separated values
CW	continuous wave
DAA	distributed access architecture
dB	decibel
dBmV	decibel millivolt
DC	direct current
DFT	discrete Fourier transform
DOCSIS	Data-Over-Cable Service Interface Specifications
DSP	digital signal processing
FFT	fast Fourier transform
FM	frequency modulation
HFC	hybrid fiber/coax
I	in-phase (real)
Hz	hertz
kB	kilobyte
kHz	kilohertz
log	logarithm
LTE	long term evolution
MHz	megahertz
MIB	management information base
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiplexing access
OSSI	operation(s) support system interface
PLC	1) physical layer link channel; 2) PHY link channel
PNM	proactive network maintenance
Q	quadrature (imaginary)
QAM	quadrature amplitude modulation
RBW	resolution bandwidth
RF	radio frequency
RPD	remote PHY device
SC-QAM	single carrier quadrature amplitude modulation
SCTE	Society of Cable Telecommunications Engineers
SNMP	Simple Network Management Protocol
TDR	time domain reflectometer
TFTP	Trivial File Transfer Protocol
TV	television
UHF	ultra-high frequency
UTSC	upstream triggered spectrum capture

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

[1] Data-Over-Cable Service Interface Specifications DOCSIS® 3.1 Cable Modem Operations Support System Interface Specification CM-SP-CM-OSSIv3.1 (Cable Television Laboratories)
[2] Data-Over-Cable Service Interface Specifications DOCSIS® 3.1 CCAP™ Operations Support System Interface Specification CM-SP-CCAP-OSSIv3.1 (Cable Television Laboratories)
[3] PNM Current Methods and Practices in HFC Networks (DOCSIS® 3.1) CM-GL-PNM-3.1 (Cable Television Laboratories)
[4] DOCSIS® Best Practices and Guidelines PNM Best Practices: HFC Networks (DOCSIS® 3.0) CM-GL-PNMP (Cable Television Laboratories)
[5] Full Scale Deployment of PMA, Lessons Learned from Deploying the Profile Management Application System at Scale and Considerations for Expanding the System Beyond OFDM, A Technical Paper prepared for SCTE•ISBE by Maher Harb, Bryan Santangelo, Dan Rice, Jude Ferreira, Comcast
[6] Comcast Tech Chief Werner: Peak Traffic Up 60% in Some Cities, But Network Is Handling It (*Multichannel News*)
[7] Combined Common Collection Framework Architecture Technical Report (Cable Television Laboratories)
[8] SCTE 280 2022, Understanding and Troubleshooting RF Spectrum
[9] DOCSIS 3.1 Physical Layer Specification CM-SP-PHYv3.1

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Deploying a Low Latency Service on the Cable Edge

Virtual Reality Streaming

A Technical Paper prepared for SCTE by

Omar Alami
Principal Engineer I
Charter Communications
6399 S Fiddlers Green Cir #100, Greenwood Village, CO 80111
720-699-3433
Omar.Alami@Charter.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Emerging Network Powered Use Cases ... 4
3. Network Demands of the Future ... 5
4. Cloud Virtual Reality ... 5
5. Cloud Virtual Reality Field Trial ... 8
   - 5.1. Field Trial Application Technology ... 8
   - 5.2. Field Trial Infrastructure ... 10
   - 5.3. Field Trial Results ... 10
     - 5.3.1. Network Metrics ... 11
     - 5.3.2. Client and Server Performance ... 12
     - 5.3.3. Survey Results ... 12
6. Conclusion ... 12

Abbreviations ... 13

Bibliography & References ... 13

List of Figures

Title **Page Number**

Figure 1 - Cloud Virtual Reality ... 6
Figure 2 - Cloud Virtual Reality Latencies ... 7
Figure 3 - Over-rendering and Head Motion Prediction ... 9
Figure 4 - Cloud VR Video Quality Measurement ... 9
Figure 5 – Cloud Virtual Reality Field Trial ... 10

List of Tables

Title **Page Number**

Table 1 – Experience Rating vs Network Latency ... 11

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

In this paper, Charter's Emerging Technology group will share its findings from the field trial of a low-latency virtual reality (VR) gaming service built on its edge computing premises. High-powered virtualized servers were configured for streaming interactive VR content to Spectrum Internet customers. Key metrics (latency, jitter, packet loss) were used to evaluate and benchmark network performance for high bandwidth, low latency services deployed at the Charter edge.

Recent trends in connectivity have accelerated use cases in which customers utilize high bandwidth but also require low latency. Emerging immersive use cases, such as virtual reality and augmented reality (AR), are inherently latency-sensitive due to the nature of being tethered to a user's head. This demand for low latency presents a significant challenge for hardware, especially in form factors that are consumer friendly and affordable to a mass market. Today, devices priced in line with gaming consoles suffer from low fidelity and devices with high fidelity require tethering to a gaming computer that usually costs thousands of dollars. The cable industry is poised to accelerate these technologies by offloading compute to high-powered servers on the network edge. With the industry's commitment to 10G service, higher levels of bandwidth can be utilized to enable experiences through the network that traditionally require high local compute.

To better understand the latency demands of these emerging use cases, Charter aimed to launch a field trial that tested our network more than typical existing use cases. Charter's Cloud Virtual Reality field trial is a system in which high-powered virtualized servers stream high fidelity interactive VR content from a regional data center (RDC) to customers on the Charter network. The aim of this field trial was to understand how an immersive use case that demands high bandwidth and low latency would perform in our existing network. Network and performance metrics were collected to draw conclusions on what performance our edge delivers in terms of latency, packet loss and jitter. Additional metrics of the client's connection were collected such as WiFi signal (RSSI), WiFi frequency and WiFi band. Metrics were collected on the server to understand how the infrastructure required to serve this experience performed, including render latency, encode latency, GPU utilization and CPU utilization. All together, these metrics establish key parameters and expectations on enabling low latency services through the network.

Any game streaming experience must consider the network latency as a key factor of the overall latency. Total input latency (i.e., the latency between a user action and the action being reflected in the display) is made up of the client-side latencies, the network latency and server-side latencies. When compared to a traditional gaming system, streaming a game also adds encode and decode latencies. A major challenge service providers face is how to utilize high powered compute to help make up for the additional latencies introduced by streaming. Techniques that utilize additional bandwidth were explored to help hide and minimize the impact of latency. These tradeoffs can be adjusted to allow for different levels of latency and fidelity depending on the use case and distance to users.

Learnings from this field trial help Charter understand how low latency services can perform on the network today and how this can change with future network enhancements such as Low Latency DOCSIS (LLD), Active Queue Management (AQM) and Low Latency, Low Loss and Scalable Throughput (L4S).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2. Emerging Network Powered Use Cases

The increasing availability of high-speed internet, including gigabit speeds, has enabled high bandwidth use cases such as 4K video streaming. Many households now have the capacity to handle multiple concurrent 4K streams without negatively impacting other internet applications within the household. Beyond 4K video streaming, it is unclear if 8K streaming will become a popular use case in the near term.

One domain poised to increase the demand for 4K and beyond streaming is VR and AR. Modern VR displays are 4K, or near 4K, and are an obvious use case for streaming 4K video. Head-mounted displays also have the ability to view 360-degree videos, in which 8K streaming is needed. This emerging use case is poised to continue growing with the increase of VR device sales led by the Meta Quest (formerly known as Oculus Quest) headsets and the announcement of the first headset from Apple. Use cases showcased by Apple include streaming high-definition content from streamers like Disney+, including immersive assets rendering within a user's environment. The sheer size of immersive video and content creates an obvious need for high bandwidth streaming.

It is possible that consumers determine 4K is indistinguishable enough from 8K on 2D TV panels so that 8K remains a niche use case. More immersive media consumption would be reserved for head-mounted displays. What would be next for the household TVs? One use case that is looking to challenge traditional TV displays are light field displays. Light field displays aim to create a field of light that shows distinct colors depending on the direction it is being viewed from. This is similar to how we take in light in the real world, in which light rays reflect off of objects and the angle of the rays determines the color and depth we perceive. A traditional TV display emits a color for each pixel regardless of the angle of the viewer. A light field display enables a sense of 3D and depth since the image is dynamic based on view angle. Modern light field displays are a major upgrade from traditional 3D TVs that failed to see adoption in peoples' homes. Some forms of light field displays require eye tracking, a similar problem to requiring 3D glasses. Mass adoption would require the same viewing convenience as traditional TVs. Displays need to not require special glasses or eye-tracking from cameras, while also covering a large field of view (FOV). Once this becomes the case, the potential for replacing household TVs will be very real. Supporting streaming to light field displays requires significantly more bandwidth than traditional TVs. Light field displays require multiple camera angles to create its image, so instead of streaming a single video stream, it will be equivalent to streaming up to 100 video streams, depending on the display and format. This could easily require hundreds of megabits per second to stream to a single display.

The need for more bandwidth will continue to grow, however providing high bandwidth alone to customers is not enough for many of these use cases, such as VR and AR. The network must be able to provide services with low latency at the same time. Low latency has always been desired for traditional use cases like live broadcasting. Viewers grow frustrated if they hear their neighbor cheer for a score in a sports game before their stream showed it to them. In the context of immersive use cases like VR and AR, low latency moves beyond just convenience and becomes a strict requirement. Streaming content to head-mounted displays requires low latency to prevent VR sickness, which can be caused by a delay in displaying updated images after head motion. Given these emerging use cases, the network will need to not only provide high bandwidth, but ensure it can be delivered with low latency.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Network Demands of the Future

Given these emerging use cases, Charter looks to find ways the network can be leveraged to enable better services. If consumer-grade light field displays were available today, could our existing networks support streaming to them? If the cost of hosting compute in the cable edge suddenly dropped, what kind of network-powered services could be supported over existing networks? Challenging the Charter network requires looking at emerging use cases and understanding the network demands of the future.

The recent increase in the adoption of VR headsets can be attributed to the affordability of standalone headsets. Traditional VR headsets are hundreds of dollars alone and require being tethered to a computer. The computer runs all of the content and requires a high-end GPU and CPU, often costing thousands of dollars. This has reserved VR setups for a niche market willing to spend the high upfront cost. Meta Quest headsets removed the tether and are available for only a few hundred dollars. While this significantly increased adoption, it came at the cost of visual fidelity. Standalone headsets suffer from tradeoffs between size, power, thermal dissipation and battery life. Keeping a headset compact and minimizing heat to the user's face means the GPU power that can be used in a standalone headset is limited. Traditional VR content can run at 4K+ with very high visual fidelity meeting the current standards of AAA video game graphics. Standalone VR content was forced to adopt lower fidelity graphics, with a lot of developers opting for content with simpler cartoon-like graphics rather than photorealistic content.

These tradeoffs offer a unique opportunity for service providers to leverage their networks. One tradeoff that standalone headsets haven't made compared to tethered headsets is screen quality and resolution. Despite having less compute power, the displays of standalone headsets can be just as good as tethered headsets. This means that the issue can be summarized as a compute location problem. Having compute outside of the headset could enable standalone headsets to display the same high-fidelity graphics of a tethered headset.

This opportunity is where Charter looked to understand the compute required to provide a service like this, as well as the network requirements. Could the existing Charter network support streaming video beyond 4K and with the low latency required for VR?

4. Cloud Virtual Reality

Cloud virtual reality is a service in which virtual reality content is rendered on a server and streamed to a head-mounted device. The most applicable example of a similar service is NVIDIA GeForce Now, which is a cloud gaming service that offers the ability to run high-end PC (personal computer) games on a cloud computer and have it streamed to a device that couldn't normally run those games. As shown in Figure 1 - Cloud Virtual Reality, Cloud VR works by having a standalone VR headset act as a thin client that sends out its inputs, including headset position/orientation and controller position/orientation with button states, to a cloud or edge server. The server, which consists of hardware capable of rendering high fidelity VR graphics, receives the input data, executes the content, renders and encodes frames sent back to the client. The client then decodes the video frames and displays it to the user. This concept can be achieved within a user's home if they have a computer that is capable of rendering PC VR and performs very well due to the low latency of a local network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

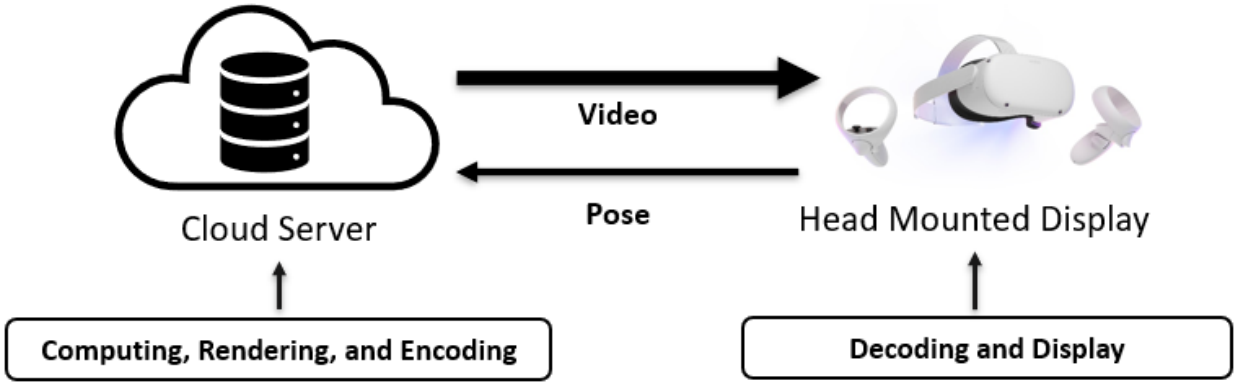


Figure 1 - Cloud Virtual Reality

Running a service like this from a traditional cloud is limited by latency. Cloud VR is a use case for utilizing an edge server, in which case the server is physically located closer to users than a traditional cloud server would be. The three main components to a service like this are the application, the infrastructure and the content. The application includes a client application and server application that communicate with each other. The infrastructure includes an edge server, consisting of servers with high-power GPUs and CPUs capable of running high fidelity VR games, and an edge network with efficient routing to users. The content includes the VR games and/or experiences that would run on the servers.

At first glance, introducing additional latency via the network to a system that is highly dependent on latency as a key performance indicator seems unwise. However, Charter developed a system in which additional bandwidth can be used in conjunction with VR software techniques to help mask a lot of the additional latency. Motion-to-photon (MTP) latency is the delay between a motion and that motion being reflected in a frame shown on a display. MTP latency is one of the most important performance indicators of a VR system. Traditionally, a MTP latency of 20 ms or less is considered good. This means that when a user moves their head, having an updated frame within 20 ms is critical to ensure a smooth experience and avoid VR sickness. Ideally, a VR system is able to maintain a MTP latency that matches the desired framerate. In VR systems with 72 frames per second (FPS), the latency between each frame would be about 13.9 ms and about 11.1 ms for 90 FPS.

VR providers have developed software techniques such as reprojection to ensure the MTP latency is low even when a system cannot render at the target frame rate. Meta's (Oculus) implementation of this feature is called Asynchronous TimeWarp and Valve's (Steam) implementation is called Asynchronous Reprojection. The feature works by taking the latest motion data available and applying it to a rendered frame in order to reproject or warp the frame to a view that is closer to the user's current head location. The technique works because motion data from the headset's sensors are always available to the headsets with low latency (1-2ms). This latency is usually much lower than the render time, and in cases where the renderer is unable to hit the target framerate, reprojection allows a system to appear more responsive. Without this technique, users would always have to wait the entire render loop latency to see an updated view, but with it, the view can be updated more consistently. This technique has limitations including introducing artifacts in a frame, especially in cases where motion is high. Additionally, this technique only modifies the view; the actions of the game will still require the full render loop latency to display to a user. However, this limitation is well worth it since the main cause of VR sickness is not delays in viewing game actions (such as firing a gun), but instead delays in seeing an updated view after moving your head.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Providing a cloud VR service requires understanding the different latencies in the system and balancing tradeoffs between bandwidth, compute and latency. Figure 2 - Cloud Virtual Reality Latenciesshows the different latencies in a cloud VR system.

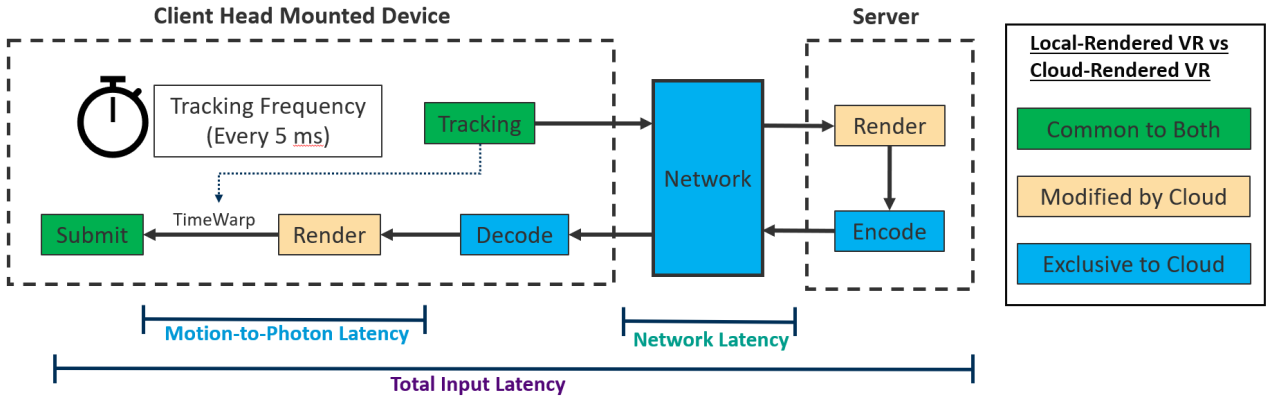


Figure 2 - Cloud Virtual Reality Latencies

The network latency is simply the round-trip time (RTT) between the client and server. The total input latency is the delay between any action (such as a button press) and when that action's result is reflected in the display. This is the main latency that suffers from cloud VR, however it is critical to separate this latency from the MTP latency. Traditional multiplayer gaming (PC or console) latencies range from 30-100 ms. Many gamers can tolerate latencies up to 100 ms depending on the type of game. The total input latency is most relatable to traditional multiplayer gaming latencies since those also include the system latency (PC or console) as well as the network latency. As previously mentioned, VR cannot tolerate latencies that high due to the nature of being head-mounted. This is why the distinction between total input latency and MTP latency is so important. MTP latency takes advantage of reprojection on the client to maintain a low perceived latency from head motion, regardless of total input latency.

Understanding these latencies allowed Charter to move forward with exploring how a cloud VR system could be delivered on our network. Beyond latency, delivering a 4K stream from a real-time rendered system requires a robust system and network capable of reliable high bandwidth. Not only is the resolution high, but the framerates used in VR (starting at 72 FPS) are higher than traditional video streaming. Netflix recommends at least 15 Mbps for 4K video streaming which is usually 24 FPS. This means a 4K VR stream could utilize approximately 50 Mbps. Additionally, real-time interactive content means the use of video techniques such as buffering is limited.

The cloud VR use case is a clear challenge for any network, requiring high bandwidth and low latency. This challenge is what motivated Charter to understand if our network is capable of handling this emerging use case or other use cases with similar requirements.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Cloud Virtual Reality Field Trial

Charter aimed to launch an employee field trial in 2023 in which select employees would be given a Meta Quest 2 headset that would stream VR games from a Charter RDC. The first step to achieving this was to develop the application and test it within a lab.

5.1. Field Trial Application Technology

The core technology of the application is relatively simple. The client application takes the position, orientation and button states of the controllers and headset and sends them over the network. The server runs VR games on a Windows operating system (OS) and encodes the rendered frames for transport back to the client. The client then decodes the frame for display.

Testing in the lab revealed the limitations and challenges of a system like this. Although reprojection enables a low MTP latency, it only works to a certain point. The more latency in the system, including network latency, the older the frame is that the client is reprojecting. This is dependent on the total latency and speed of the head motion. A frame that moved very little compared to the previous frame can tolerate more latency. If the difference between frames is too high, reprojection creates artifacts. This is because the reprojection only changes the view and cannot create new pixels. One example of an artifact is black bars on the side of the frame. For example, if a user is looking forward and suddenly rotates their head to the right, reprojection will move the view to the right and black pixels will be shown on the right edge of the frame. These black bars can become visible to the user and if bad enough, cause VR sickness due to lack of smooth motion.

Given this issue, Charter looked into how bandwidth could be used in conjunction with reprojection to minimize the latency of a cloud VR system. One technique used to help mitigate the artifact issue is called over-rendering. This is the concept of rendering a frame larger than what can actually be displayed. Combining a larger frame with reprojection allows the reprojection to have additional pixels that can be displayed to the user in the case of higher latency or higher motion. This was a key feature that enabled Charter to utilize more bandwidth to help reduce the perceived latency. While the Charter network was able to handle additional bandwidth, this bandwidth comes at the cost of higher compute. By rendering at a higher resolution, the server GPU requirements are increased.

Over-rendering was initially configured to render 20% more pixels, divided equally among all four sides of the frame. However, most of the time, additional pixels are only needed on one or two sides of the frame, depending on the direction of the head motion. To help alleviate the additional compute demands, another technique called head motion prediction (HMP) was developed. This technique utilizes a Kalman filtering algorithm to predict the direction of a user's head motion. In general, when a user moves their head to look in a direction, the movement will be in a single direction between most frames until they change directions. This provides a good opportunity for prediction in which errors are only expected in frames where the direction is changed. Combining prediction with over-rendering allows the server to over-render from a predicted head position. This enables a better use of rendered pixels, minimizing the percentage of over-rendering needed and lessening the compute requirements. This behavior is shown in Figure 3 - Over-rendering and Head Motion . The artifacts present are a function of head motion speed and network latency. The scenarios in the figure show the same scene with the same network latency and head motion speed and demonstrates how artifacts are reduced with over-rendering and prediction.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 3 - Over-rendering and Head Motion Prediction

Utilizing these techniques in a field trial required a quantifiable way to measure artifacts and understand what quality of service can be delivered from the Charter network at different points. Charter created a turntable capable of rotating a headset at precise speeds in a lab environment where latency could be injected. This setup is shown in Figure 4 - Cloud VR Video Quality Measurement, below.

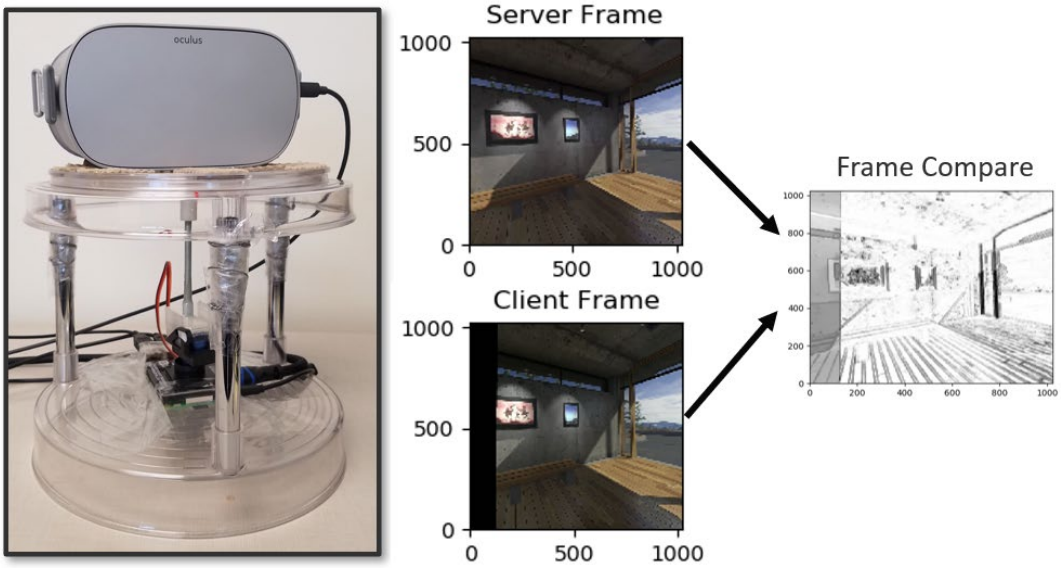


Figure 4 - Cloud VR Video Quality Measurement

This system included frame captures on both the client and server applications. A video quality metrics analysis was then done to automatically detect the artifacts introduced between the client and server. Aside from unavoidable frame differences caused by the transcoding, black pixels introduced from head motion would cause dramatic changes in quality scores. The lab testing provided key guidelines on the tolerable network latencies for different over-rendered resolutions at different head motion speeds. A game was considered high motion if it required head rotation of about 400 degrees per second. We found that at our target resolution, a network latency as high as 35 ms would still enable a good experience for most users.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.2. Field Trial Infrastructure

A 35 ms target network latency enabled Charter to deploy servers in an RDC in which employees within a few hundred miles would achieve latencies well below the target. The field trial infrastructure included an Infrastructure as a Service (IaaS) made up of several server racks with data center-grade GPUs. The IaaS ran Windows 10 virtual machines (VM) which the content ran on. Connecting the client application to the server application utilized an orchestration layer that was developed for the field trial. Since these microservices were not latency sensitive, they were hosted in a Charter national data center (NDC) as part of Charter's Container as a Service (CaaS). The field trial architecture is shown in Figure 5, below.

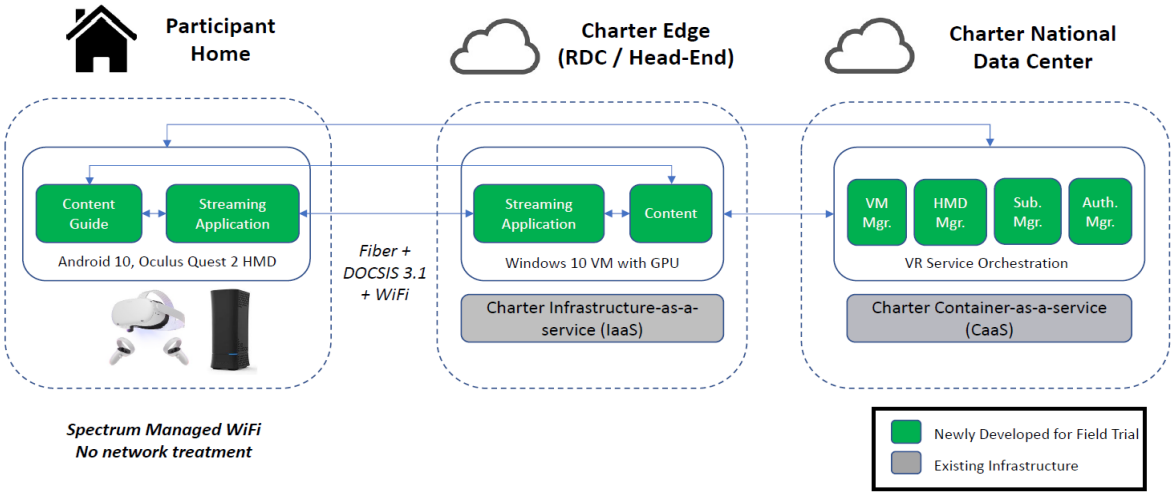


Figure 5 – Cloud Virtual Reality Field Trial

The client application utilized the orchestration layer to authenticate employees via the CaaS microservices and connect them to a VM hosted in the RDC. Once connected, the client is able to start streaming from the server to their device on the home WiFi.

5.3. Field Trial Results

The field trial included 35 employees over a 90-day period. The first phase of the field trial had no network treatment and the second phase had active queue management (AQM) activated for the users. The users combined for over 180 hours of data. Built into the application was a metrics gathering system that collected network and performance data of the application, home network and servers. After each session, the application prompted users to give a 1-5 star rating of the experience and content. The experience rating was meant to rate the quality of the stream while the content rating was meant to separate the user's opinion of the content. Users reported an average of 4.3 for the experience rating and 4.5 for the content rating.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.3.1. Network Metrics

Sessions with lower network latency correlated to a higher experience rating. Low latency (<27 ms) sessions had 88% good experience ratings (4+) while high latency sessions (30+ ms) had 61% good experience ratings. Table 1 – Experience Rating vs Network Latency below shows the median network latency for each experience rating. Overall, the field trial averaged a network latency of 29 ms.

Table 1 – Experience Rating vs Network Latency

Rating	5	4	3	2	1
Median Network Latency (ms)	27.6	28.3	29.9	28.9	30

The average and median latency were lower in the second phase with AQM, but given a different set of users in each phase, it can't be determined if this was because of AQM conclusively. Additionally, network latency increased during peak times (7PM – 11PM) without AQM, but actually decreased during peak time when AQM was on.

Jitter was measured based on the client utilizing the method described in internet standard RFC 1889/3550. The average and median jitter was about 3.5 ms. Lab testing determined that jitter of 5 ms or less provided a good experience for most users. We found a correlation between jitter and network latency, in which higher network latency usually led to higher jitter.

We found that packet loss was arguably more important than network latency in terms of being a key performance indicator. Ideally, packet loss would be 0% at all times. However, given this field trial relied on home WiFi connections, some levels of packet loss were expected. Lab testing indicated an average of less than .1% of packet loss was ideal. The application did utilize forward error correction (FEC) to help mitigate higher levels (acceptable up to .5%). The field trial demonstrated a median of .03% packet loss, however we found that packet loss would happen in bursts. These bursts included periods of high packet loss (1%+) within the home which could cause visual impairments. These bursts were rare and, in most cases, happened in isolation within the user's home WiFi network. Users that experienced sustained levels of high packet loss were instructed to upgrade and/or replace their modem/router.

WiFi metrics were also gathered by the client since most internet issues in a home occur on the WiFi network. All users were instructed to utilize 5GHz WiFi, and the application would warn users that were connecting with 2.4GHz. However, since this was only a warning, some users proceeded on 2.4GHz regardless and demonstrated higher network latency, jitter and packet loss. The WiFi signal strength was measured with the network latency, which were found to be correlated. Sessions with higher signal strength (closer to the router) were found to have lower network latencies.

The target streaming bitrate was 50 Mbps and averaged about 45 Mbps during the field trial. The actual bandwidth used is determined by the framerate that could be achieved by the server and was dependent on the rendering performance of each game. The upstream bandwidth used on average was 600 Kbps, higher than required for a system like this due to the metrics being uploaded from the client.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.3.2. Client and Server Performance

The average total input latency was 74 ms and 88% of sessions performed under 80 ms. Other than network latency, the largest latencies were the render and encode latencies on the server. Optimizations and improvements to the application can reduce the total input latency further, which would allow for a larger tolerance for higher network latency.

GPU utilization averaged less than 50% however the game with the highest fidelity graphics averaged 65%. Encode performance was arguably the bottleneck of server performance in the field trial, with encode utilization averaging 53%. Encode performance was also dependent on content, where games with higher motion had longer encode times.

Overall, head motion prediction performed well. Only 7% of frames reported a prediction error, in which the predicted render origin was not close enough to the actual origin to prevent any black pixel artifacts. Prediction in the Y direction (up and down) was more accurate than in the X direction (left and right). This is likely because left to right head movement is more common than up and down head movement in VR. Prediction performance was less accurate for content with higher motion.

5.3.3. Survey Results

Employees were surveyed after the field trial to gather feedback on the service. Twenty-five out of 27 respondents said they were satisfied or more than satisfied with the performance; 22 out of 27 users said they would likely subscribe to a service like this; 17 out of 22 users said the service was consistent or very consistent; and 13 out of 18 users (with previous VR experience) said this was equal to or better than previous VR experiences.

6. Conclusion

Charter aimed to test the network against a demanding use case that utilized very high bandwidth and required consistent performance with low latency. The results showed that Charter's network is already capable of providing quality Cloud Virtual Reality service from a Regional Data Center to users nearby and software improvements and reductions to access network latency will expand the service to more users.

The physical range between the server and users remains a constraint that can be further reduced via software improvements and optimizations, more efficient future hardware and reduction of network latencies. More efficient hardware is critical for making services like this more economical. Currently, the upfront cost of the hardware (especially GPUs) makes a service like this a challenge to scale.

Network key takeaways included the importance of the in-home WiFi network, ensuring users are near their router and that packet loss in the home, caused by congestion, is minimized as much as possible. Sessions with lower network latency and jitter resulted in higher ratings. More so than bandwidth, this field trial demonstrated the importance of understanding network latencies and reducing them as much as possible.

Services like cloud VR that demand low latency stand to benefit from new network technology, such as LLD and L4S. Cloud VR technology will continue to be improved and be tested against the latest network technologies.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

VR	virtual reality
AR	augmented reality
RDC	regional data center
NDC	national data center
RSSI	received signal strength indicator
GPU	graphics processing unit
CPU	central processing unit
LLD	low latency DOCSIS
DOCSIS	data over cable service interface specification
AQM	active queue management
L4S	low latency, low loss, and scalable throughput
FOV	field of view
PC	personal computer
MTP	motion-to-photon
FPS	frames per second
RTT	round-trip time
OS	operating system
HMP	head motion prediction
IaaS	infrastructure as a service
CaaS	container as a service
VM	virtual machine
AQM	active queue management
FEC	forward error correction

Bibliography & References

"Internet connection speed recommendations," Netflix, https://help.netflix.com/en/node/306

"Under the hood: Building 360 video," Engineering at Meta, Oct. 15, 2015. https://engineering.fb.com/2015/10/15/video-engineering/under-the-hood-building-360-video/

"How 10G Networks Will Power Holographic Technologies | NCTA — The Internet & Television Association," www.ncta.com, Dec. 08, 2020. https://www.ncta.com/whats-new/how-10g-networks-will-power-holographic-technologies

"Keeping the virtual world stable in VR | Qualcomm," www.qualcomm.com, Jun. 28, 2016. https://www.qualcomm.com/news/onq/2016/06/keeping-virtual-world-stable-vr

"Asynchronous TimeWarp (ATW) | Oculus Developers," developer.oculus.com. https://developer.oculus.com/documentation/native/android/mobile-timewarp-overview/

"Introducing SteamVR Motion Smoothing," Valve Corporation, Nov. 27, 2018. https://steamcommunity.com/games/250820/announcements/detail/1705071932992003492

H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RFC 3550 - RTP: A Transport Protocol for Real-Time Applications," Internet Engineering Task Force, Jul. 2023. https://www.ietf.org/rfc/rfc3550.txt

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Deploying Segment Routing for PON Aggregation in Cox's Metro Network

A Technical Paper prepared for SCTE by

Deependra Malla
Lead Network Design Engineer
Cox Communication Inc.
6305 Peachtree Dunwoody Road, Atlanta, GA 30328
602-694-4429
deependra.malla@cox.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Cox Metro Network Overview ... 3
3. Label Distribution Protocol (LDP) Limitations ... 4
4. Segment Routing ... 5
   - 4.1. SR-MPLS Operation ... 5
   - 4.2. Segement Routing Global Block (SRGB) ... 6
   - 4.3. SRGB Design and Management ... 6
   - 4.4. SR-MPLS Data Plane Operation ... 8
   - 4.5. Segment Routing IS-IS Control Plane ... 8
5. Topology Independent Loop Free Alternate (TI-LFA) ... 9
   - 5.1. TI – LFA Protection Options ... 9
6. LDP to SR Migration ... 10
7. Conclusion ... 11

Abbreviations ... 12

Bibliography & References ... 12

List of Figures

Title **Page Number**

Figure 1 – Cox Communication High Level Metro Network Architecture ... 4
Figure 2 – Cox metro SRGB design and index assignment example ... 7

List of Tables

Title **Page Number**

Table 1 – Segment list mapping to MPLS label stack operation ... 6
Table 2 – Protocol preference value for various vendors ... 11

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The continuous advancement of IP networks marked by increased capacity requirements, increased network assets, and elevated customer expectations on service continuity and quality of service, has placed substantial demands on the network infrastructure of Multi-Service Operators (MSOs), Internet Service Providers (ISPs), and large enterprises. To help meet these demands and improve network performance, network operators have started adopting Segment Routing (SR) in the various parts of their networks - access, metro, and backbone.

The evolution of access technology from cable Internet to Passive Optical Network (PON) involves a transition in the technology and architecture used to deliver high-speed broadband services to end-users. Both cable Internet and PON are broadband access technologies, but they differ significantly in their underlying infrastructure and methods of data delivery. The primary concept behind PON is to bring fiber optic cables directly to the end-users, eliminating the need for intermediate coaxial cables or copper lines. This is often referred to as Fiber-to-the-Home (FTTH) or Fiber-to-the-Premises (FTTP). Although cable access is Cox's legacy access technology, it introduced FTTH to its customers in 2014 under the brand name "Gigablast". Since then, Cox has continuously invested in network infrastructure and improving the quality of Gigablast services offered to its customers. Adoption of Segment Routing in the Cox's PON aggregation and metro network is a crucial step towards making Cox's access network more optimized, scalable, and resilient.

As the networking landscape evolves, the industry is leaning towards more flexible and software-driven networking paradigms. So, transitioning to Segment Routing positions Cox metro network to embrace these future changes and innovations. Adopting the SR also helps Cox to provide improved network performance and service continuity to its customers.

The Cox metro networks is an MPLS enabled network and runs LDP as the label switching protocol. LDP is used for following purposes in Cox metro networks:

- Label switch BGP traffic,
- Transport label for L2VPN, and
- Transport label for L3VPN

In the Cox metro network, Segment Routing is used to distribute MPLS labels and accomplish all the above services. Following are the motivations for enabling SR in Cox metro networks:

- Protocol simplification and unified control plane
- Achieving <50ms failover during link failure
- Tactical traffic engineering capability
- New protocol knowledge acquisition

Since Segment Routing is a more recent approach to label switch traffic, it is not available in legacy hardware. Hence, during the initial phase of SR deployment in Cox, LDP and SR will co-exist. Wherever the SR is enabled along with the LDP, the SR is made a preferred label switching protocol.

2. Cox Metro Network Overview

PON network in Cox metro is a multi-vendor hub-and-spoke topology with hub routers aggregating multiple pairs of PON aggregation routers. These PON aggregation routers aggregate multiple Optical Line Terminals (OLTs). The Cox metro network runs IS-IS as an IGP protocol, LDP as a label distribution protocol and hierarchical BGP. The high-level topology of Cox metro network is shown in figure 1.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

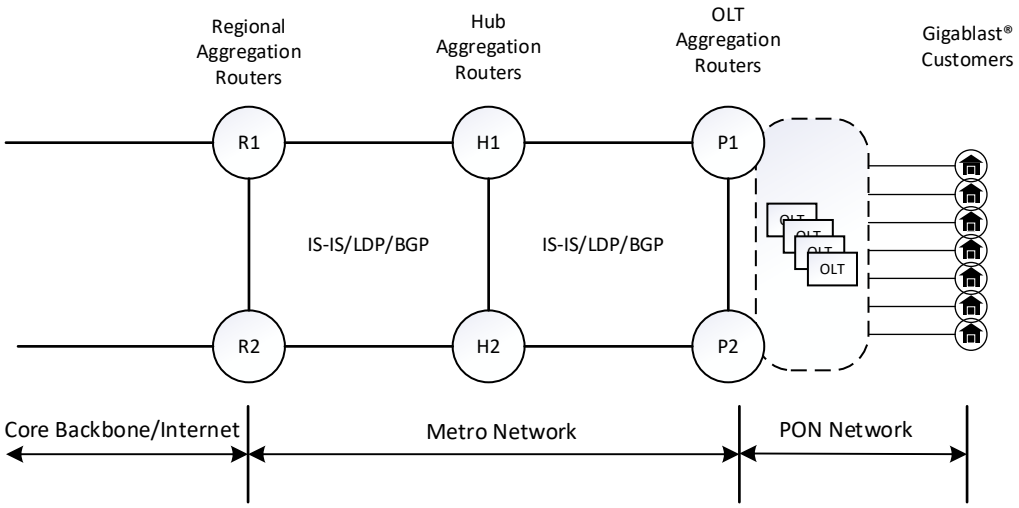


Figure 1 – Cox Communication High Level Metro Network Architecture

Regional aggregation routers are the gateway to the Internet and other Cox markets, and they aggregate several hub site's hub aggregation routers. Hub aggregation routers further aggregates several edge routers such as PON routers, CIN routers, local service/management routers, CMTS/CCAP etc. These hub aggregation routers function as Provider (P) routers in the Cox metro MPLS domain with no services provisioned on them. Regional aggregation routers and edge routers function as Provider Edge (PE) routers. The Internet traffic to and from OLT aggregation routers to regional aggregation routers are LDP label switched.

3. Label Distribution Protocol (LDP) Limitations

The Cox metro network is an MPLS enabled network with LDP used to distribute labels for loopback interfaces of metro routers. LDP is a traditional method for assigning and distributing labels in an MPLS network. It relies on label distribution from the headend router to downstream routers, establishing label-switched paths (LSPs) along which packets are forwarded based on labels. Today, LDP generated labels are used to label switch customer Internet traffic in Cox metro network. It is also used to provide a transport label for services like L2VPN and L3VPN. Although, LDP has its own advantages such as its widespread adoption, simplicity of basic MPLS and predictable behavior, there are certain limitations too.

Disadvantages of LDP:

Label Distribution Protocol (LDP) is a widely used technology for label assignment and distribution in Multiprotocol Label Switching (MPLS) networks. However, like any technology, LDP has its disadvantages and limitations. At Cox, we have noticed following limitations of LDP:

- **LDP-IGP Synchronization Requirement**: LDP must be fully established and synchronized with the underlying interior gateway protocol (IGP) before an IGP path is used for forwarding traffic. If the LDP and IGP are not synchronized, packet loss can occur. During link or node failures, we must wait until LDP-IGP synchronizes before putting traffic into LDP LSP.
- **Limited Traffic Engineering:** LDP provides limited capabilities for fine-grained traffic engineering and path optimization. It lacks the flexibility to specify explicit paths and perform dynamic traffic steering based on real-time network conditions. This can be a limitation in networks that require efficient resource utilization and optimized routing.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- **Scalability Concerns:** As networks grow larger and more complex, LDP's signaling overhead can become a scalability concern. The protocol generates a significant amount of control plane traffic to establish label-switched paths (LSPs), which can lead to increased processing and memory requirements on routers.
- **Limited Support for Programmability**: LDP's lack of programmability and flexibility can hinder its compatibility with modern network automation and software-defined networking (SDN) principles.
- **Incompatibility with IPv6 Transition**: LDP does not inherently support IPv6 traffic over an IPv4 infrastructure, making the transition to IPv6 more challenging.

Segment Routing (SR) and Label Distribution Protocol (LDP) are both networking technologies used in the context of MPLS (Multiprotocol Label Switching) networks. Each technology has its own advantages and use cases. However, transitioning from Label Distribution Protocol (LDP) to Segment Routing (SR) in Cox network can have following added benefits:

- Ability to introduce traffic engineering and path optimization,
- Simplicity and reduced network states,
- Improve scalability,
- Flexibility and Programmability,
- Unified Control Plane
- Faster traffic re-routes during link and node failure,
- Integration with SDN and automation

4. Segment Routing

Segment Routing (SR), as defined in RFC 8402,(Clarence Filsfils et al., 2018) is a modern network architecture and forwarding paradigm that simplifies the way packets are routed through networks. It leverages source-routing principles and enables efficient traffic engineering, optimal path selection, and seamless integration with software-defined networking (SDN) principles. SR is particularly suited for Multiprotocol Label Switching (MPLS) and IPv6 networks. There are two types of Segment Routing technologies - SR with MPLS data plane (SR-MPLS) and SR with IPv6 data plane (SRv6). The current deployment of Segment Routing in Cox metro network uses SR-MPLS and is primarily targeted for label switching IPv4 traffic. This document will focus on SR-MPLS deployment in Cox metro network.

In SR, the concept of a "segment" is central. A segment can be thought of as an ordered list of instructions that guides the path a packet should take through the network. These segments are represented as a stack of instructions embedded in the packet header. The segments define the specific nodes and links that the packet should traverse, allowing for explicit path control and optimization.

4.1. SR-MPLS Operation

SR-MPLS, also known as Segment Routing with MPLS data plane, is an implementation of Segment Routing (SR) within Multiprotocol Label Switching (MPLS) networks. It combines the benefits of SR's explicit path control and MPLS's label-switching capabilities. SR-MPLS offers efficient traffic engineering, simplified network management, and seamless integration with existing MPLS infrastructure.

In SR-MPLS, segments are represented as labels in the MPLS label stack. Each label corresponds to a specific segment or node in the path. The stack of labels in the packet header is known as the "segment list." The top label indicates the next hop, and subsequent labels indicate the subsequent nodes or segments in the path. At the source node, the path is encoded into the packet header as a stack of labels (the segment

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

list). The labels in the segment list represent the desired path that the packet should follow through the network. As the packet traverses the network, routers perform MPLS label swapping according to the label stack in the packet header. Each router examines the top label, determines the next hop or segment, and forwards the packet accordingly.

The SR MPLS data plane operations use the existing MPLS forwarding operation. The segment list operations (PUSH, CONTINUE, POP) are mapped to the MPLS data plane operations in Table 1.

Table 1 – Segment list mapping to MPLS label stack operation

SR Segment List Operation	MPLS Label Stack Operation
PUSH	PUSH
CONTINUE	SWAP
POP	POP

As described earlier, the concept of segment in central to SR architecture. A segment is an instruction that an ingress node inserts on the packet header and is identified by a Segment Identifier (SID). The segment can be a global segment known as Segment Routing Global Block (SRGB) or a local segment known as Segment Routing Local Block (SRLB). The implementation of SID in Cox network for IPv4 FEC is an MPLS label (32-bits). Although there are many types of SIDs, for Cox's SR deployment, IGP Node Segment (Node-SID) and IGP Adjacency Segment (Adj-SID) are discussed in this paper. Node-SID is globally significant and should be unique in a SR-MPLS domain whereas Adj-SID is locally significant to the router.

Although, SR architecture does not assume a specific data plane implementation, for SR deployment in Cox metro network we use existing MPLS data plane to forward SR label switched traffic.

4.2. Segement Routing Global Block (SRGB)

The SRGB is a range of labels reserved for Segment Routing global segments when using MPLS data plane for segment routing (Clarence Filsfils et al., 2017). SRGB needs to be configured on every router that participates in the segment routing domain. Varying size of label spaces on various vendors poses a significant challenge on designing of SRGB. If the SRGB space is allocated from the already used label space in a router, then the router needs hard reboot for that label to take effect. To minimize the number of router reboots in the Cox network, Cox has allocated 300,000 through 559,999 label space for SRGB. This range of SRGB will be used homogenously and contiguously across the entire Cox network.

4.3. SRGB Design and Management

For Cox deployment, SRGB is used for global labels like Node-SIDs. Node-SIDs are globally unique identifiers that require static assignments. To make the allocation of Node-SID unique and easy, each market in Cox network is allocated a block of SID indices from the global range. This range is used for Node-SID index for IPv4, IPv6 address families and any other SR algorithm Node-SID requirements. Cox is not using "absolute" SID assignment for Node-SID.

For Cox's deployment, SRGB parameters are defined as follows:

- SRGB Base value = 300,000
- SRGB Block Size = 260,000
- SRBG Global Range = 300,000 – 559,999 (Across all Cox domains)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Each router in SR domain requires a unique Node SID which is generated by using unique node SID index assigned to each router in each market. Formally, a Node SID is a unique index in the SRGB. Any node in the SR domain derives the local label associated to the Node-SID as SRGB + index.

Different vendors implement SRGB in different ways – global allocation and/or per protocol allocation. The SRGB design and implementation in a network depends on network operators design requirement, topology, and size of the network.

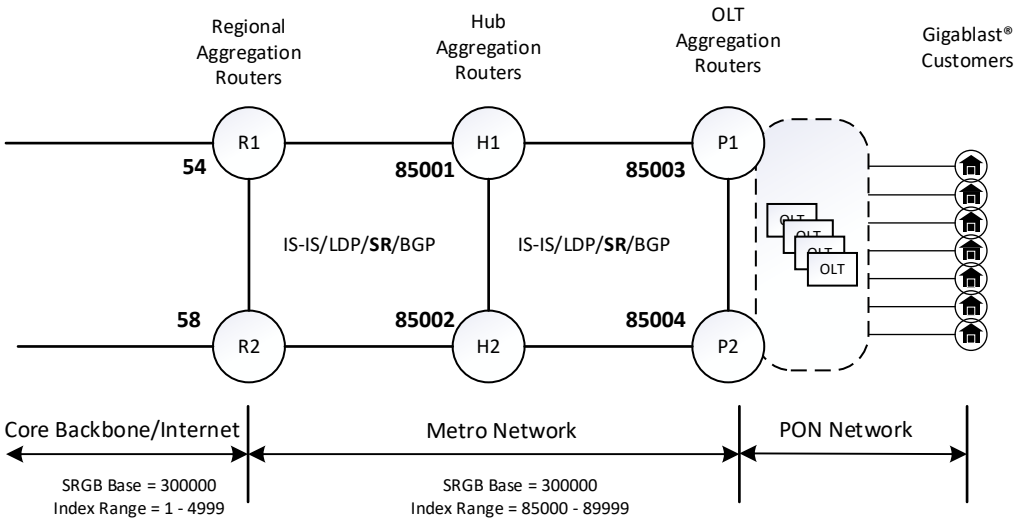


Figure 2 – Cox metro SRGB design and index assignment example

Example:

If the node-sid index for the regional aggregation router R1 in figure 1 is 54, then hub aggregation (H1, H2) and OLT aggregation routers (P1, P2) will calculate local label for R1 as follows:

Local label on P1 router for R1 = SRGB base on P1 + R1's node-sid index

= 300000 + 54

= 300054

The outgoing label is the local label that the downstream neighbor (the next-hop router on the shortest path to destination router) allocated for the Node-SID.

Outgoing label on P1 router for R1 = SRGB base advertised by H1 + R1's node-sid index

= 300000 + 54

= 300054

If other routers in the network are configured for same SRBG base value, all remaining routers in figure 1 topology will have label 300054 programmed in their LSD (Label Switching Database) for R1.

To avoid duplication of SID index, Cox will either create user defined field (UDF) in IPAM or create a database from where SID index can be assigned and stored for each router in the Cox network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.4. SR-MPLS Data Plane Operation

The traditional MPLS data plane is reused without any modification for Segment Routing in Cox metro network. Label operations (Push, Swap/Continue, Pop), PHP behavior, MPLS TTL processing, EXP bits processing, ECMP load-balancing techniques and MTU (Maximum Transmit Unit) handling are used similarly for Segment Routing like its being used for LDP today.

4.5. Segment Routing IS-IS Control Plane

Link state routing protocols like IS-IS and OSPF (Open Shortest Path First) have been extended to support distribution of SR information in IGP domain. RFC 8667 added several new sub-TLVs to IS-IS to signal various SR capabilities (Clarence Filsfils et al., 2019). The SR information includes Prefix-SID index, Adjacency-SID, SR data plane capabilities and label range used by SR. These new sub-TLVs are used to attach Prefix-SIDs and Adj-SIDs to various prefix and adjacency advertisement TLVs. Following are some useful TLV/sub-TLVs used by IS-IS:

- TLV = 22, Extended IS reachability
  - Sub-TLV = 31, Adjacency Segment Identifier
- TLV = 135, Extended IP reachability
  - Sub-TLV = 3, Prefix Segment Identifier
- TLV = 242, Router capability
  - Sub-TLV = 2, SR capability

The following example (output from Cisco IOS-XR) shows the IS-IS LSP flooded by R1 in figure 2 into the network. The example shows that both IPv4 and IPv6 address families are enabled in multi-topology mode. We can see several valuable information regarding segment routing information being advertised into the LSP such as:

- Router capability TLV (`Router Cap:  10.0.0.1, D:0, S:0`)
- IPv4 Node-SID index, default algo, and flags (`Prefix-SID Index: 54, Algorithm:0, R:0 N:1 P:0 E:0 V:0 L:0`)
- IPv6 Node-SID index, default algo, and flags (`Prefix-SID Index: 654, Algorithm:0, R:0 N:1 P:0 E:0 V:0 L:0`)
- Various Adj-SID and flags (`ADJ-SID: F:0 B:0 V:1 L:1 S:0 weight:0 Adjacency-sid:16`)
- SRGB Range (`SRGB Base: 300000 Range: 260000`)

This information is useful in validating the design and troubleshooting of SR if there are any issues.

```
RP/0/0/CPU0:H1#sh isis database R1.00-00 verbose
Wed Aug 16 19:18:43.915 UTC

IS-IS ISIS (Level-1) Link State Database
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime  ATT/P/OL
R1.00-00              0x00000e37   0x1f53        753             0/0/0
  Area Address:   49.0090
  TLV 14:         Length: 2
  NLPID:          0xcc
  NLPID:          0x8e
  MT:             Standard (IPv4 Unicast)
  MT:             IPv6 Unicast                                  0/0/0
  Hostname:       R1
  Metric: 40         IS-Extended R2.00
    Interface IP Address: 192.168.10.1
    Neighbor IP Address: 192.168.10.2
    Local Interface ID: 330, Remote Interface ID: 330
    ADJ-SID: F:0 B:0 V:1 L:1 S:0 weight:0 Adjacency-sid:16
```

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

```
:
Metric: 40          MT (IPv6 Unicast) IS-Extended R2.00
  ADJ-SID: F:1 B:0 V:1 L:1 S:0 weight:0 Adjacency-sid:17
:
Metric: 0           IP-Extended 10.0.0.1/32
  Prefix-SID Index: 54, Algorithm:0, R:0 N:1 P:0 E:0 V:0 L:0
:
Metric: 0           MT (IPv6 Unicast) IPv6 2001:BEEF:1:0:10::1/128
  Prefix-SID Index: 654, Algorithm:0, R:0 N:1 P:0 E:0 V:0 L:0
Router Cap:     10.0.0.1, D:0, S:0
  Segment Routing: I:1 V:1, SRGB Base: 300000 Range: 260000
  SubTLV 19 Length: 1
```

Maximum Segment Depth (MSD)

In SR, maximum segment depth refers to the number of segments (labels in SR-MPLS) a router is capable of imposing on a packet and varies across the vendors and the platforms (software & hardware). SR will increase the depth of label stacks if SR features such as TI-LFA, SR-TE, and traffic accounting are enabled on the routers. The topology and metrices of a network determine the segment depth of the labeled packets. Since Cox metro topology is a hub-and-spoke topology with static metric, the number of SIDs required for TI-LFA link-protection is 2.

5. Topology Independent Loop Free Alternate (TI-LFA)

Cox metro network provides critical services to customers such as Metro Ethernet over HFC (Hybrid Fiber Coaxial), voice and streaming services apart from Internet traffic. These critical services are the driving factors for enablement of fast reroute in metro network. In SR enabled network, TI-LFA provides the desired protection mechanism to such critical traffic. Topology Independent Loop-Free Alternate (TI-LFA) is a fast reroute (FRR) mechanism used in networks to provide protection against link and node failures. It is specifically designed to work in IP networks, including those running Segment Routing (SR), and aims to quickly restore connectivity in case of failures while avoiding the creation of forwarding loops. TI-LFA enhances network resilience and ensures that traffic continues to flow smoothly even during network disruptions. There are three types of FRR – Classic Loop Free Alternate (cLFA), Remote Loop Free Alternate (rLFA) and Topology Independent Loop Free Alternate (TI-LFA). cLFA and rLFA do not provide 100% coverage and TI-LFA does. So, at Cox we have deployed TI-LFA with link protection option. TI-LFA provides 100% link and node protection and micro loop avoidance. It always routes protected traffic on the post convergence path. Since Cox metro topology is dual egress hub-and-spoke topology, the primary benefit of TI-LFA implementation is micro-loop avoidance rather than post-convergence optimization.

Since TI-LFA uses Segment Routing for the repair path, SR must be deployed in network for TI-LFA to work. Following are the benefits of TI-LFA:

- TI-LFA provides less than 50ms link, node and SRLG (Shared Risk Link Group) protection with 100% coverage.
- The repair path is automatically computed by IGP.
- TI-LFA uses a post-convergence path as backup path.
- TI-LFA can be incrementally deployed; it is locally significant.
- TI-LFA also protects LDP and IP traffic in addition to SR traffic.

5.1. TI – LFA Protection Options

TI-LFA protection options ensure a repair path along the post convergence path in any topology. TI-LFA offers following protection options:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- **Link Protection**: Link protection is the most basic form of the options that is available for TI-LFA and all it needs is to be just enabled for all the links that need protection. Link protection is the only protection option that is enabled in Cox metro network. All interfaces that are running MPLS/LDP are enabled for link protection.
- **Node Protection**: In the node protection, the neighbor node is excluded during the post convergence backup path calculation.
- **SRLG protection**: SRLG refers to the scenario where network links share the same fiber or a common physical attribute. Such links are susceptible to collective risk. TI-LFA SRLG protection attempts to find the post-convergence backup path that excludes the SRLG of the protected link. All local links that share any SRLG with the protecting link are excluded.

The following output on H1 router shows the post convergence backup path calculated by H1 to reach R1 in figure 2 using TI-LFA.

```
RP/0/0/CPU0:H1#show isis fast-reroute 10.0.0.1/32
Thu Aug 17 06:30:28.144 UTC

L1 10.0.0.1/32 [16/115]
     via 100.120.100.36, GigabitEthernet0/0/0/1, R1, SRGB Base: 300000, Weight: 0
         Backup path: TI-LFA (link), via 100.120.100.39, GigabitEthernet0/0/0/0 H2,
SRGB Base: 300000, Weight: 0
           P node: R2.00 [10.0.0.2], Label: 300058
           Prefix label: 300054
```

6. LDP to SR Migration

Migrating label switching mechanism in a network from Label Distribution Protocol (LDP) to Segment Routing (SR) involves a well-planned process to transition the way traffic is routed and forwarded. This migration aims to leverage the benefits of SR, such as fast reroute, enhanced traffic engineering and path control, while ensuring minimal disruption to ongoing network operations. Before migrating LDP to SR, it is important to assess the network's current architecture, topology, and traffic patterns. Identify the specific areas or services that would benefit from SR's capabilities and then plan the migration strategy, considering factors such as network complexity, critical services, and potential impact on end-users.

Cox chose to deploy Segment Routing along with the existing LDP in its metro network. Adopting a parallel and incremental deployment approach allows for a gradual transition without affecting ongoing operations. This approach is also called “ships-in-the-night” and avoids the use of Segment Routing Mapping Server (SRMS) as Cox will not be enabling “interworking” between LDP and SR. The "ships-in-the-night" method is a gradual and non-disruptive approach for migrating from Label Distribution Protocol (LDP) to Segment Routing (SR) within a network. It is designed to minimize service disruption and ensure a smooth transition from one technology to another. The concept behind the ships-in-the-night migration is to allow both LDP and SR to coexist temporarily, with traffic being migrated gradually over time without affecting ongoing network operations.

Below is the high-level migration plan to enable Segment routing and removal of LDP protocol. The implementation will be split into three phases.

Phase 1: SR and LDP co-existence by configuring SR and letting LDP be the preferred label imposition method.

Phase 2: Preferring SR over LDP as a label imposition method.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Phase 3: Remove LDP.

Protocol Preferences

Like different routing protocols have different preferences, label switching protocols (LDP, SR, RSVP-TE etc.) also have different protocol preferences. When an operator changes the label preference on the node, the LSR will reprogram its FIB to "push" labels based on the preferred preference value (usually lower value is considered preferable). Thus, LSPs for which the LSR is the headend are impacted by this change. When a LSR terminates a service and has the preference changed, a service interruption can occur. This scenario is applicable to the PE routers in the network. Since LSPs are unidirectional, the remote LSR's LSP is not impacted by the change. The following table shows how vendors can use protocol preference values for LDP and SR to influence which label switching protocol to use to switch labeled traffic in the network.

Table 2 – Example protocol preference values

Vendor	Protocols and default preference values
Vendor A	LDP = 9 SR (L-ISIS) = 14
Vendor B	LDP = 55 IS-IS SR = 65
Vendor C	LDP = - (default preferred label switching protocol) SR = sr-prefer (makes SR preferable over LDP)
Vendor D	LDP = 9 SR-ISIS = 11

Network operators can lower the preference value of SR than that of LDP or raise the preference value of LDP to be greater than SR to make SR a preferred label switching protocol in the network.

7. Conclusion

As the networking landscape continues to evolve, transitioning from LDP to SR can offer significant advantages to network operators. Segment Routing is a flexible and efficient source routing mechanism that would simplify the control plane of a network. It can provide several benefits such as less than 50ms fast reroute for traffic in link and node failure scenario, deterministic and efficient traffic engineering capabilities, and network programmability that are otherwise not available using LDP. As Cox introduces SR in Cox metro network, the experience and knowledge acquired can be easily transformed into other part of the network such as core backbone, data center network, and business network. Cox considers the introduction of SR in Cox metro network as a solid foundation for the evolution of programmable network, and it will have a tremendous impact on Cox's next generation network evolution and modernization.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

SR	Segment Routing
LDP	Label Distribution Protocol
LSD	Label Switching Database
MPLS	Multi-Protocol Label Switching
IS-IS	Intermediate System-Intermediate System
TI-LFA	Topology Independent - Loop Free Alternate
SRLG	Shared Risk Link Group
EXP	MPLS Experimental Bits
ECMP	Equal Cost Multi Paths
MTU	Maximum Transmission Unit
PON	Passive Optical Network
BGP	Broader Gateway Protocol
LSR	Label Switch Router
SRMS	Segment Routing Mapping Server
MSD	Maximum Segment Depth

Bibliography & References

Clarence Filsfils, Ahmed Bashandy, Hannes Gredler, & Bruno Decraene. (2019). IS-IS Extensions for Segment Routing—RFC 8667. *Internet Engineering Task Force (IETF)*. https://doi.org/10.17487/RFC8667

Clarence Filsfils, Kris Michielsen, & Ketan Talaulikar. (2017). *Segment Routing Part I*. Cisco Systems.

Clarence Filsfils, Stefano Previdi, Les Ginsberg, Bruno Decraene, Stephane Litkowski, & Rob Shakir. (2018). Segment Routing Architecture—RFC 8402. *Internet Engineering Task Force (IETF)*. https://doi.org/10.17487/RFC8402

Ventre, P. L., Salsano, S., Polverini, M., Cianfrani, A., Abdelsalam, A., Filsfils, C., Camarillo, P., & Clad, F. (2021). Segment Routing: A Comprehensive Survey of Research Activities, Standardization Efforts, and Implementation Results. *IEEE Communications Surveys & Tutorials*, *23*(1), 182–221. https://doi.org/10.1109/COMST.2020.3036826

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Detection and Classification of OFDMA Spectrum Impairments by Machine Learning

Developing Machine Learning Models to Detect and Classify Impairments in OFDMA Channels

A Technical Paper prepared for SCTE by

Jude Ferreira
Principal Data Scientist
Comcast
215.286.4070
jude_ferreira@cable.comcast.com

Kevin Dugan
Senior Machine Learning Engineer
Comcast
719.493.2600
kevin_dugan2@comcast.com

Maher Harb
Distinguished Engineer
Comcast
267.260.1846
maher_harb@comcast.com

Mike O'Dell
Distinguished Engineer
Comcast
412.417.0481
michael_odell@cable.comcast.com

Larry Wolcott
Fellow
Comcast
267.260.1846
larry_wolcott@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. OFDMA Impairments ... 3
   - 2.1. Known Impairments ... 4
     - 2.1.1. VHF Ingress ... 4
     - 2.1.2. Analog Modulator ... 4
     - 2.1.3. RFoG Ingress ... 5
   - 2.2. Unknown Impairments ... 5
3. Training Data/Labeling UI ... 8
4. Model Architecture ... 10
5. Model Training/Performance ... 11
6. Machine Learning Operations ... 14
7. Conclusion ... 15

Abbreviations ... 16

Bibliography & References ... 17

List of Figures

Title **Page Number**

Figure 1 - Normal OFDMA RxMER Examples ... 4
Figure 2 - VHF Ingress ... 4
Figure 3 - Analog Modulator Ingress ... 5
Figure 4 - RFoG Ingress ... 5
Figure 5 - Unclassified OFDMA Impairments. Each row represents an unclassified pattern example, and columns represent RxMER samples from different devices on the same node segment. ... 6
Figure 6 - RxMER Heatmap for a device with an unknown impairment - 1 ... 7
Figure 7 - RxMER Heatmap for a device with an unknown impairment - 2 ... 8
Figure 8 - Example of an MER plot in the labeling tool. The tool is used to both assign initial labels and conduct periodic reviews of the model's predictions. ... 9
Figure 9 - Convolutional Neural Network (CNN) Components ... 10
Figure 10 - Training/Validation ROC ... 12
Figure 11 - Holdout ROC ... 12
Figure 12 - Training/Validation Confusion Matrix ... 13
Figure 13 - Holdout Confusion Matrix ... 13
Figure 14 - False negative (FN) examples from the Holdout dataset ... 14
Figure 15 - False positive (FP) examples from the Holdout dataset ... 14
Figure 16 - ML Ops workflows for scaling the model and monitoring performance ... 15

List of Tables

Title **Page Number**

Table 1 - Number of samples by impairment classification. Sample counts below 50 are not shown. ... 9
Table 2 – Hyperparameters and Ranges evaluated during training. ... 11

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Automated Proactive Network Management (PNM) is no longer an afterthought or a luxury but is considered table stakes when it comes to maintaining Comcast's vast HFC (hybrid fiber-coaxial) network. Our networks experience a wide range of conditions that can degrade their performance over time. These conditions include issues such as loose connections between components, cracks and breakages in lines, disruptive energy, and signal impediments, all of which are inherent challenges in maintaining our constantly evolving network. The process of early detection and efficient mitigation minimizes service disruptions, reduces downtime, and leads to a better customer experience. Identifying the specific nature and root cause of network impairments also enables us to route repair technicians to the appropriate location, and reduces Mean Time to Repair (MTTR), thereby driving operational efficiency.

Deploying OFDMA (Orthogonal Frequency Division Multiple Access) in the mid-split region of the spectrum has allowed us to offer ~3-10x higher upstream speeds to customers and is also an important steppingstone toward offering multi-gig symmetrical services using FDX (Full Duplex) under our 10G roadmap. D3.1 OFDMA allows the use of higher modulation levels up to 4096-Quadrature Amplitude Modulation (QAM) and provides up to 2x efficiency increases when compared to Single Carrier QAM levels of 64-QAM. Profile Management Application (PMA) systems are being used to manage OFDMA Profiles as described in our previous SCTE contribution [1]. However, PMA can also mask network impairments at a cost to capacity as an inherent component of its functionality. Therefore, it is critical to develop systems to perform PNM to reduce the impact of network impairments and enable the highest possible capacities and speeds.

In this paper, we describe the use of Convolutional Neural Networks (CNNs) to identify network impairments within the mid-split region and share the current performance of our machine learning (ML) models. This effort is similar to our previous efforts to identify network impairments in OFDM and Downstream Single Carrier QAM (DS-SC-QAM) sections of the spectrum as described in our previous SCTE contributions [2,3].

2. OFDMA Impairments

Receive Modulation Error Rate (RxMER) is an extremely effective metric for understanding network impairments as it picks up both core signal-to-noise (SNR) characteristics and signal imperfections. For OFDMA channels, RxMER at a granular mini-slot resolution is available. RxMER for D3.1 devices using OFDMA channels is polled at frequent intervals and stored in our data lake. The methods described to identify various impairments for OFDMA Channels in this paper are based on this high-resolution RxMER.

Figure 1 represents RxMER samples where no impairments exist. These are characterized by the RxMER curves that are essentially flat over the entire width of the OFDMA channel with values over 40 dB.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

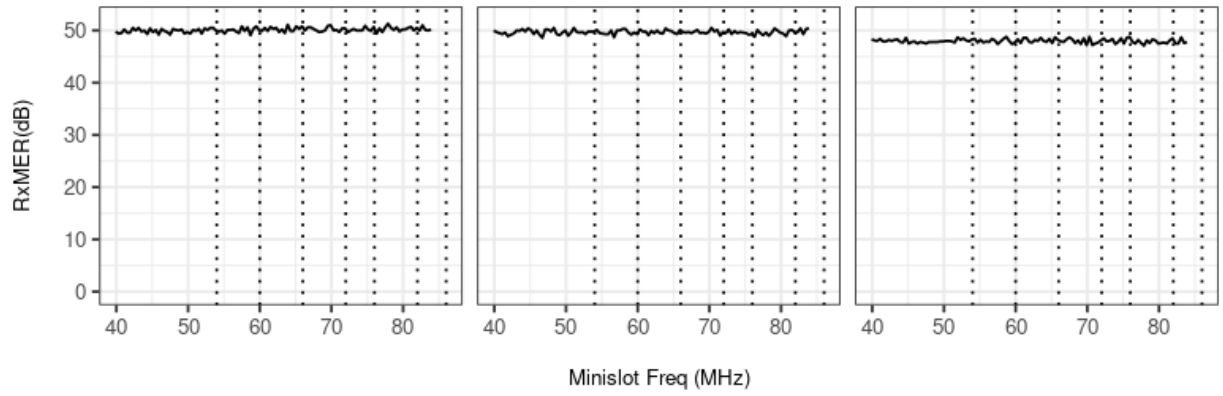


Figure 1 - Normal OFDMA RxMER Examples

2.1. Known Impairments

While some sources of interference in the mid-split region such as VHF (Very High Frequency) were known beforehand, we have also encountered and identified additional interference types over the course of our mid-split deployments.

2.1.1. VHF Ingress

VHF over-the-air (OTA) ingress is one of the more common ingress sources in our OFDMA deployments. Depending on the location of TV Transmitters within a geographical area, one or more channels may be impacted as seen in Figure 2.

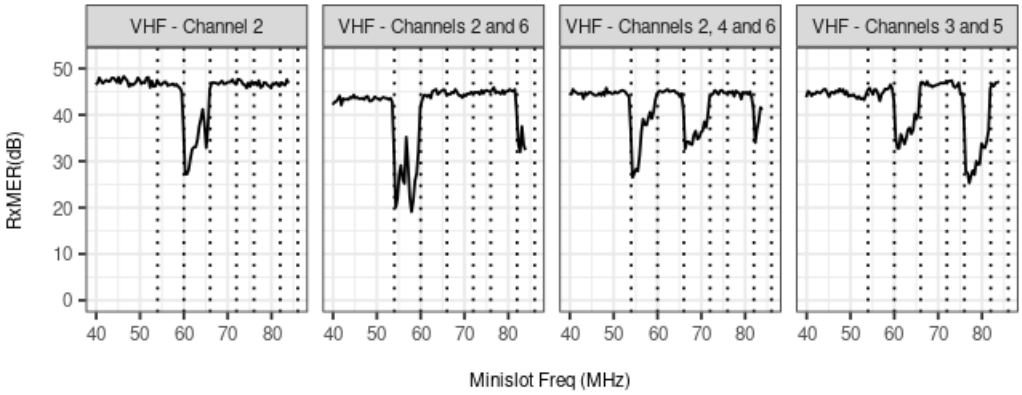


Figure 2 - VHF Ingress

2.1.2. Analog Modulator

The Analog Modulator impairments are narrow band ingressors caused by an old VCR, a gaming console, or the wrong connector on an older set-top box connected to an outlet in the home. As seen in the image below, this impairment impacts the mini-slots at either 61.1 MHz or 66.1 MHz and is classified as Analog Modulator Channel 3 and Analog Modulator Channel 4 respectively.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

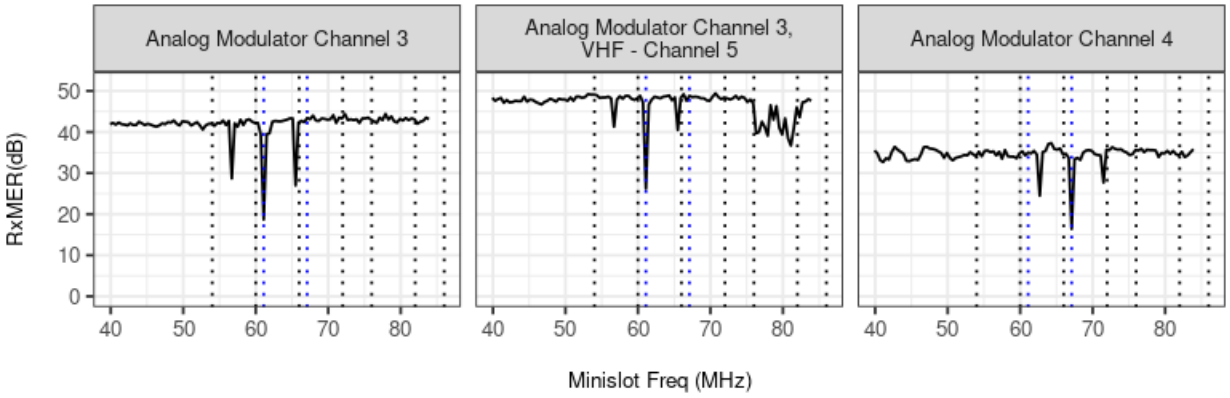


Figure 3 - Analog Modulator Ingress

2.1.3. RFoG Ingress

RFoG (Radio Frequency over Glass) impairments are caused by customers who have disconnected their service and moved to another Fiber broadband provider but are still connected to our network. As compared to VHF Ingress and Analog Modulator Ingress, RFoG ingress impacts a much wider portion of the OFDMA spectrum because a set of downstream channels from the Fiber broadband provider interferes with our OFDMA spectrum.

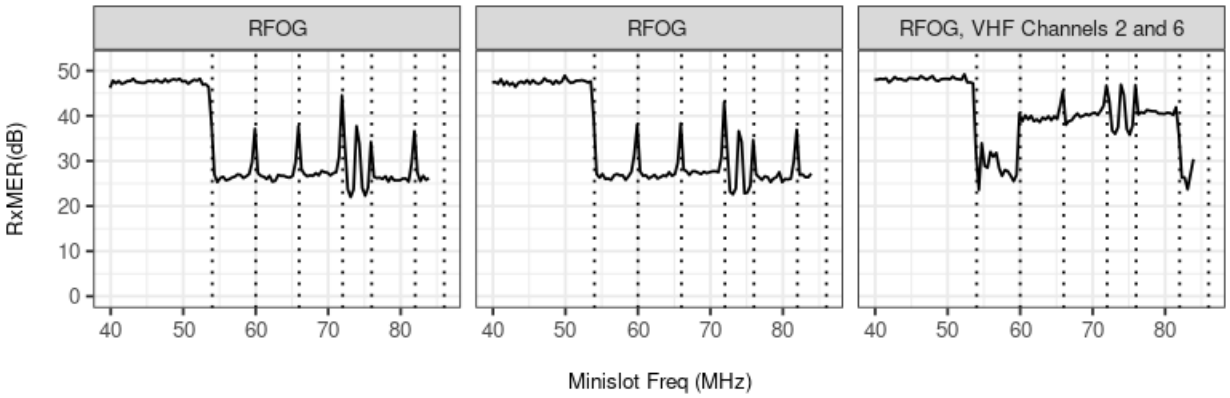


Figure 4 - RFoG Ingress

2.2. Unknown Impairments

In addition to the above impairments, we have also encountered a multitude of diverse impairment signatures that are currently under investigation for identification. This task presents significant challenges due to the intermittent nature of many of these impairments, posing obstacles in troubleshooting and accurately classifying them. We look forward to engaging with the broader PNM community in a collaborative approach to expand upon these efforts, benefiting the industry as a whole and advancing our collective understanding of these impairments.

We have grouped the unclassified impairments based on their RxMER signatures and a selection of these unclassified impairments are displayed in Figure 5 below. The individual examples listed in each row

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

below come from the same node segment and thus increase confidence that they are caused due to the same underlying issue.

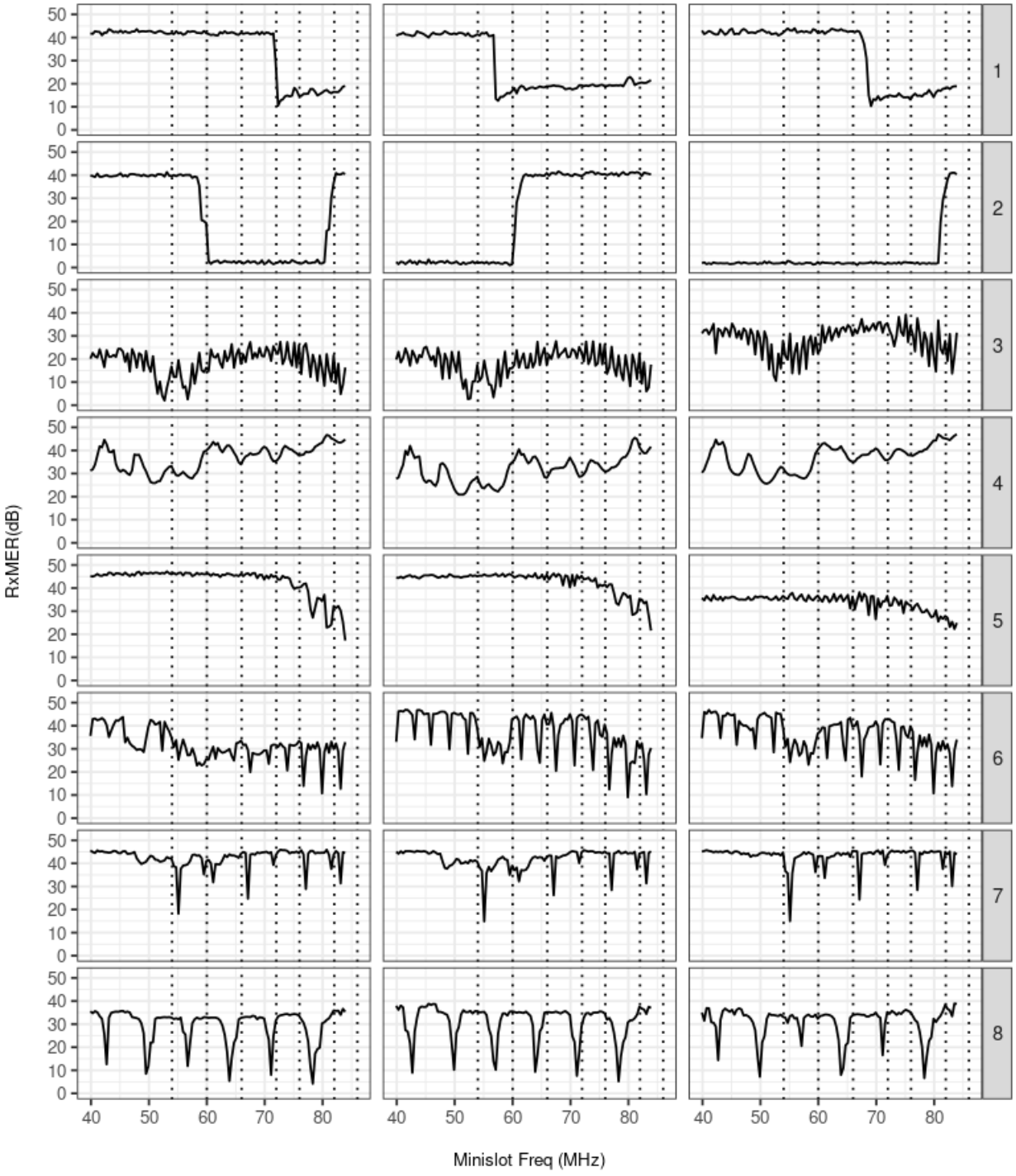


Figure 5 - Unclassified OFDMA Impairments. Each row represents an unclassified pattern example, and columns represent RxMER samples from different devices on the same node segment.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

There is some inherent noise in the RxMER measurements at an individual device level and even if the inherent quality of the spectrum doesn't change, repeated measurements will show some variation. Outside of this expected variation, an analysis of RxMER over multiple time samples illustrates the transient nature of some of the above impairments. While many devices that exhibit the unclassified patterns were examined, we consider two devices, one of which exhibited Pattern 3 from Figure 5, and the other which exhibited Pattern 8 from Figure 5 to demonstrate the transient nature of some of the unclassified impairments. The individual RxMER measurements every ~5 minutes from these two devices over the course of 6 days are plotted as heat maps in Figure 6 and Figure 7. In the images, the colors scale from lighter blue, which represents RxMER values that happen infrequently, to darker orange, which represents values that occur repetitively.

In Figure 6, the device exhibits the impairment pattern only a few times each day over the six days of observation.

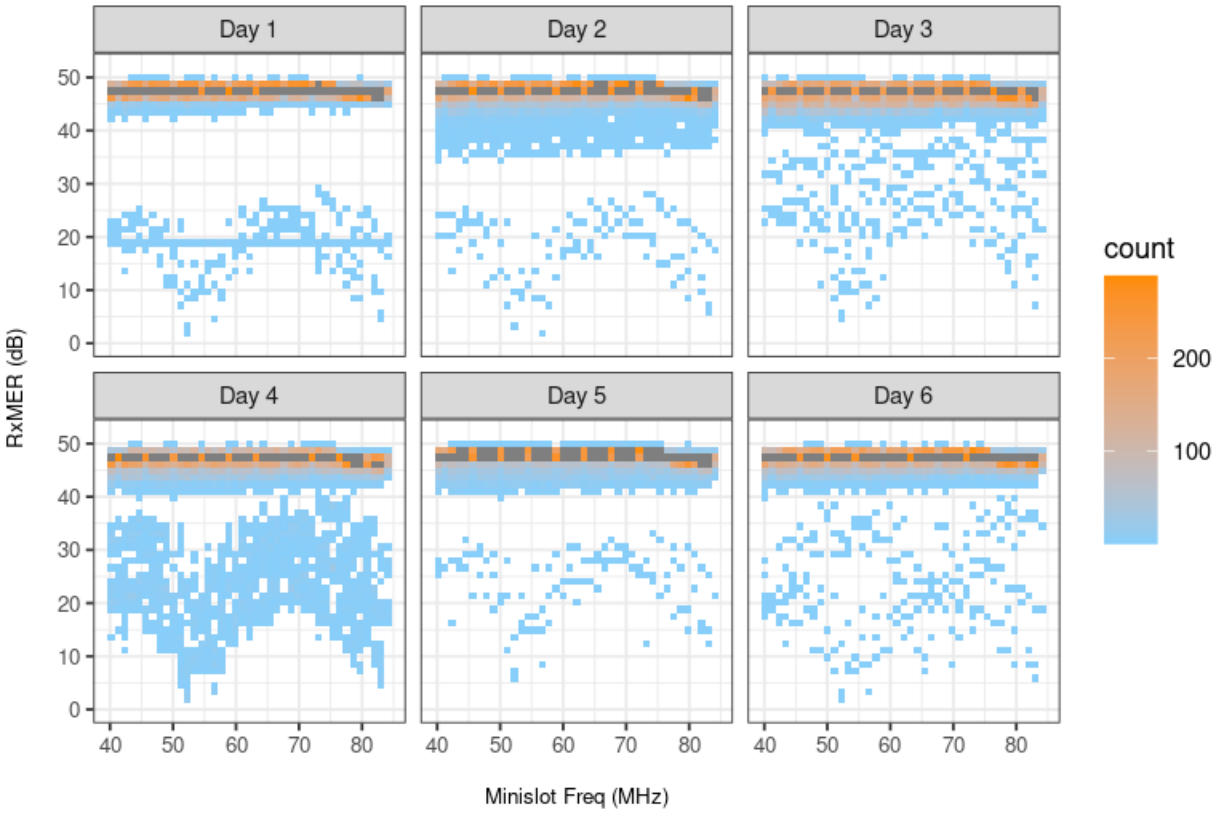


Figure 6 - RxMER Heatmap for a device with an unknown impairment - 1

In Figure 7, we see that the impairment pattern only shows up a few times on two out of six days of observation. This device also appears to have VHF Ingress on channel 5, which is more prevalent.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

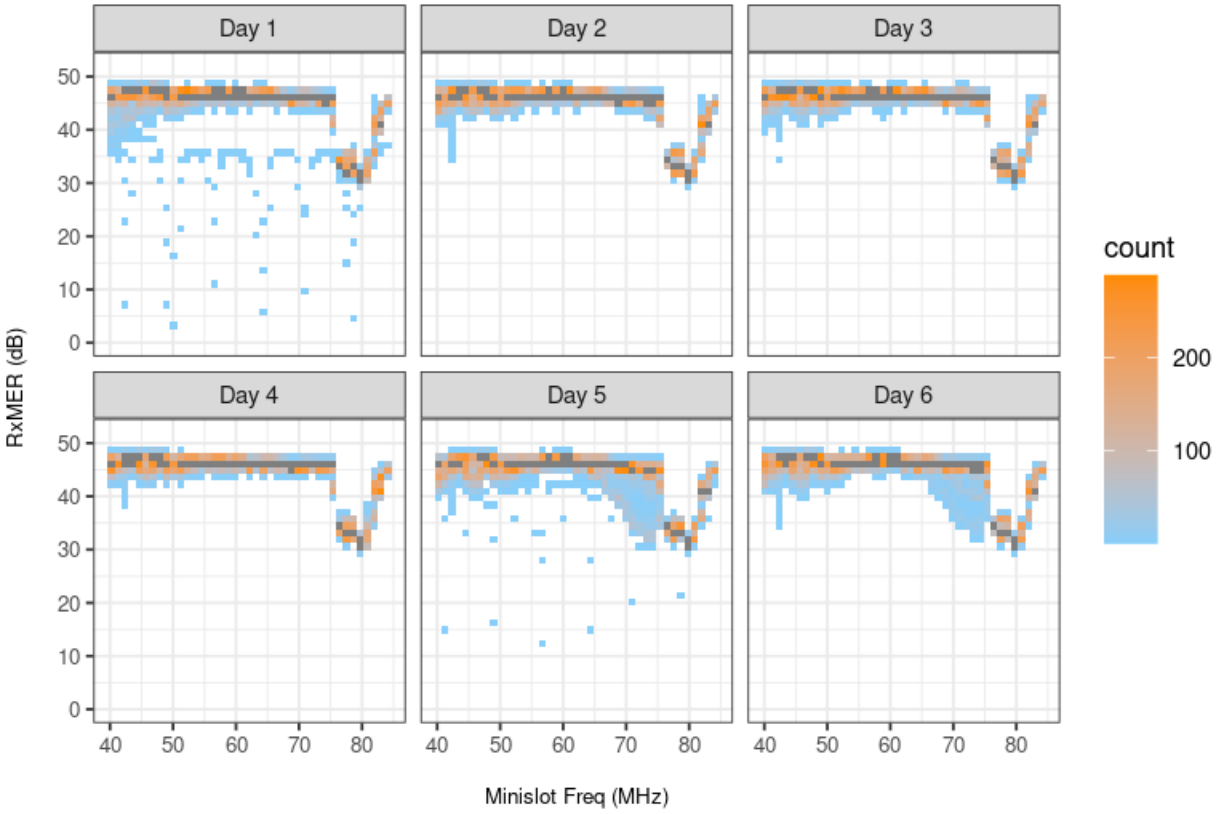


Figure 7 - RxMER Heatmap for a device with an unknown impairment - 2

3. Training Data/Labeling UI

Generating labeled data for supervised machine learning is a labor-intensive activity that requires subject matter experts (SMEs) to carefully examine and classify impairments. To generate a larger population of labeled data, we employed a hybrid approach. We, as data scientists, labeled the easier-to-classify samples, while the more challenging samples were assigned to field technicians and other SMEs for classification.

To help capture impairments, and given that a majority of RxMER samples do not have impairments, the sampling strategy focused on capturing samples with high variance over OFDM subcarriers. In addition, rule-based methods were utilized to identify potential VHF, Analog Modulation, and RFoG impairments, which underwent further validation before inclusion in the labeled dataset.

We developed a custom UI tool for gathering labeled data from SMEs. In addition to the impairments, the pattern locations in the spectrum were also captured to support the model's requirements. The locations refer primarily to the standard SC-QAM channels between 54MHz and 88MHz (channels 2 – 6).

Each plot in the UI represents a single modulation error rate (MER) capture from a cable modem. The MER values are inverted over the y-axis to align the visual representation with internal noise monitoring tools that our experts are accustomed to reading. The example in Figure 8 illustrates the user experience in the labeling tool for an impaired device with VHF ingress located on channels 2, 4, and 6. Once a collection of samples has been labeled by experts, the collection can be added to the population of available training and validation data.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 8 - Example of an MER plot in the labeling tool. The tool is used to both assign initial labels and conduct periodic reviews of the model's predictions.

We generated around 19.5k samples using the hybrid approach described above. Table 1 below shows the number of samples by impairment classification.

Table 1 - Number of samples by impairment classification. Sample counts below 50 are not shown.

Labels	Sample Count
Normal	7,733
VHF - Channel2	1,446
VHF - Channel2, VHF - Channel6	1,298
Other	1,264
VHF - Channel3	1,209
VHF - Channel4	726
VHF - Channel5	714
VHF - Channel3, VHF - Channel5	489
RFoG, VHF - Channel2, VHF - Channel6	461
Analog Modulator - Channel 3	421
VHF - Channel2, VHF - Channel6, Other	388
VHF - Channel2, Other	341
VHF - Channel3, VHF - Channel4, VHF - Channel6	333
VHF - Channel2, VHF - Channel4, VHF - Channel6	326
VHF - Channel4, VHF - Channel5, VHF - Channel6	288
VHF - Channel6	218
Analog Modulator - Channel 3, VHF - Channel2, VHF - Channel6	209

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Analog Modulator - Channel 3, Other	171
VHF - Channel2, VHF - Channel4	170
RFoG	158
Analog Modulator - Channel 4	147
Analog Modulator - Channel 4, VHF - Channel2, VHF - Channel6	103
VHF - Channel3, VHF - Channel4, VHF - Channel5, VHF - Channel6, Other	73
VHF - Channel5, Analog Modulator - Channel 3	72
VHF - Channel2, VHF - Channel4, VHF - Channel6, Other	70
VHF - Channel4, VHF - Channel5	65
Analog Modulator - Channel 4, VHF - Channel2, VHF - Channel4, VHF - Channel6	62

4. Model Architecture

Similar to our previous efforts to classify impairments in OFDM Channels [2] and SC-QAM Channels[3], the model architecture used to classify OFDMA impairments is a 1-D Convolutional Neural Network (CNN) which has the general architecture shown in Figure 9.

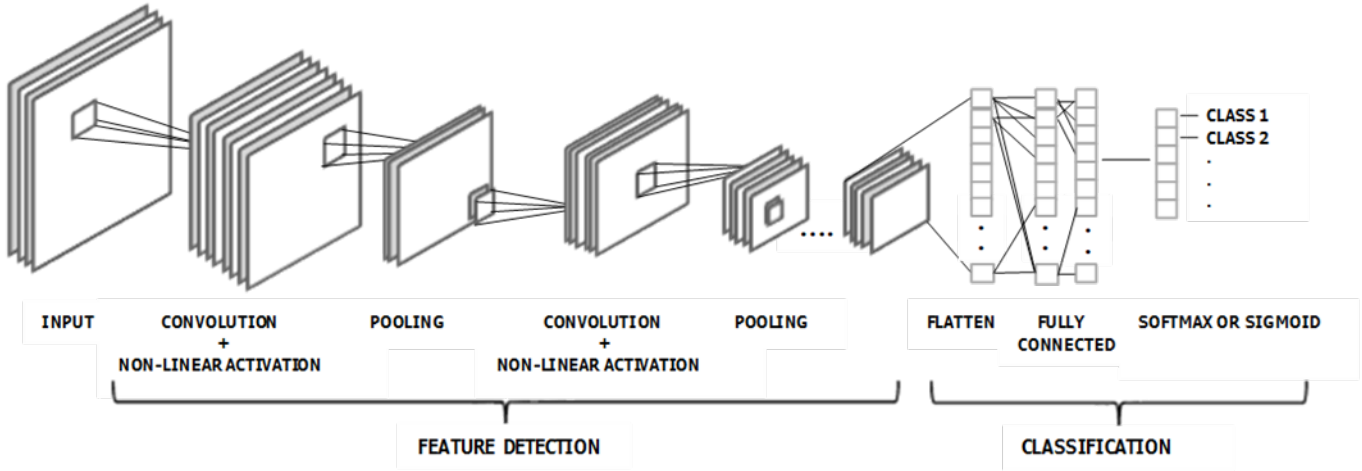


Figure 9 - Convolutional Neural Network (CNN) Components

CNNs typically have a series of convolutional and pooling layers that are stacked together. Each convolutional layer consists of several filters. As the data flows through the network, the learned features increase in complexity [4]. The convolutional layers are often followed by pooling layers which reduce the spatial dimensions of the data while preserving the most important features. Pooling helps to make the model more robust to variations in the input, and it also helps reduce the computational requirements.

After the convolutional and pooling layers, CNNs usually include one or more fully connected layers. These take the learned features from the earlier layers to perform tasks such as classifications or predictions.

In our specific use case, the input to the CNN consists of one-dimensional arrays composed of RxMER samples. Each sample represents a RxMER per mini-slot capture for a single device. The output from the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

model for each input sample is an array containing the probabilistic predictions for each impairment category.

We experimented with 1-D CNN architectures that had between 1-3 convolution blocks and 1-3 fully connected layers. A grid search was performed on the following hyper-parameters before selecting the final model. The results displayed only slight variations across numerous hyperparameter combinations, indicating that the size of the training data may have a more significant impact on performance than the specific parameters employed.

Table 2 – Hyperparameters and Ranges evaluated during training.

Hyperparameter	Range	Hyper-parameter Type
Number of filters in convolutional layers	[32, 64, 96, 128]	Network Structure
Kernel size in convolutional layers	[3,5,7,9]	Network Structure
Pooling Size	[2,3,4,5]	Network Structure
Fully connected hidden layer size	[32, 64, 96, 128, 160]	Network Structure
Dropout	[0.2, 0.25, 0.3, 0.4, 0.5]	Network Structure
L2 Regularization	[0, 0.0001, 0.0005, 0.001, 0.005, 0.01]	Network Structure
Learning Rate	[0.00005, 0.0001, 0.0003, 0.0005, 0.001, 0.003, 0.005, 0.01, 0.03]	Network Training
Batch Size	[16, 32, 64, 128, 256]	Network Training

5. Model Training/Performance

Out of the ~19.5k labeled samples, ~90% were used for training and validation with the remaining ~10% used as a holdout dataset to estimate performance of the model in production. Initially, 5-fold cross-validation was used, and validation/training loss was used to determine the hyperparameters that had the best performance. The top-performing model was then trained on the complete set of training and validation samples to yield the final model.

The models were evaluated on receiver operating characteristic (ROC), precision, and recall during training as well as on the holdout dataset for the individual classes.

ROC for all classes is very close to 1 during training as well as on the holdout dataset as seen in Figure 10 and Figure 11. This indicates that the model can almost exactly distinguish between the positive and negative samples for each class. This is also reflected in the very high precision and recall values seen in **Error! Reference source not found.** Figure 12 and Figure 13.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

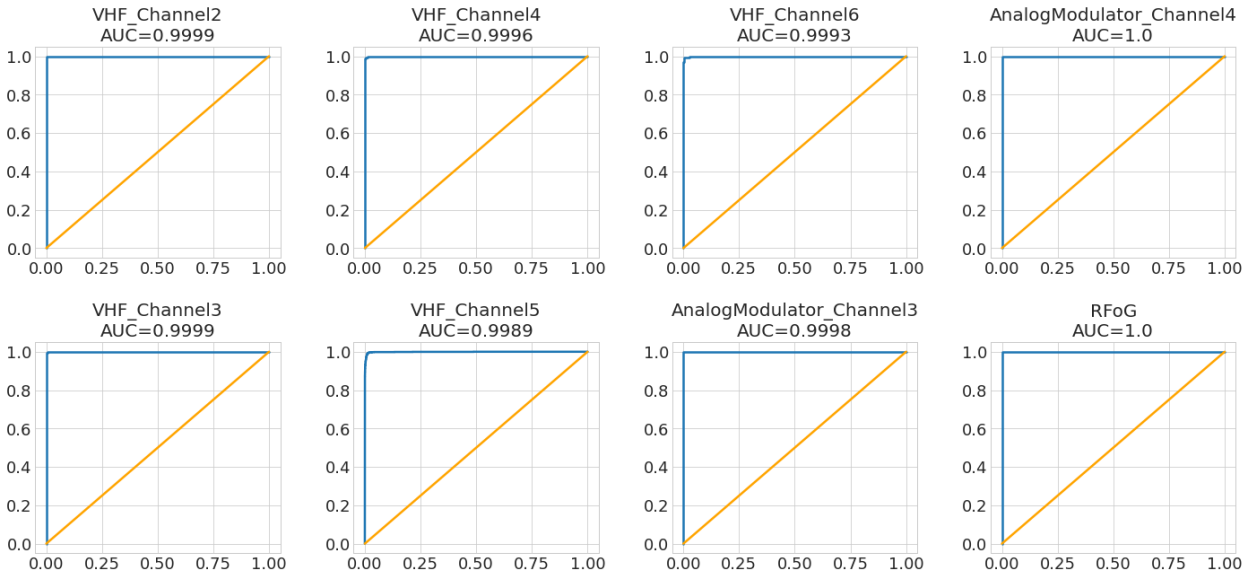


Figure 10 - Training/Validation ROC

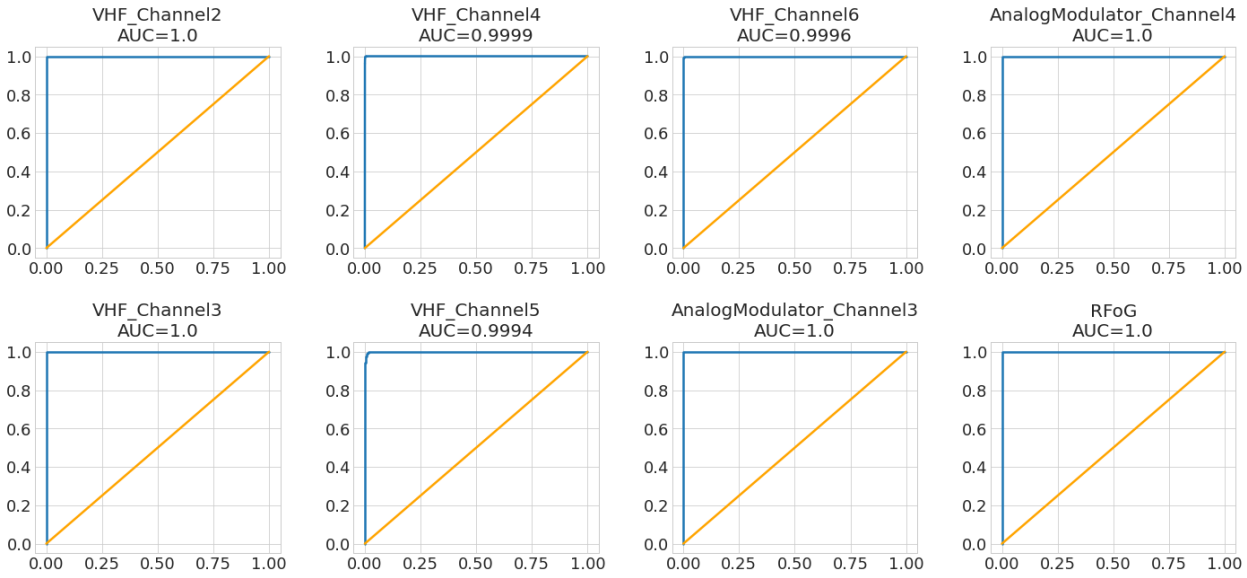


Figure 11 - Holdout ROC

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

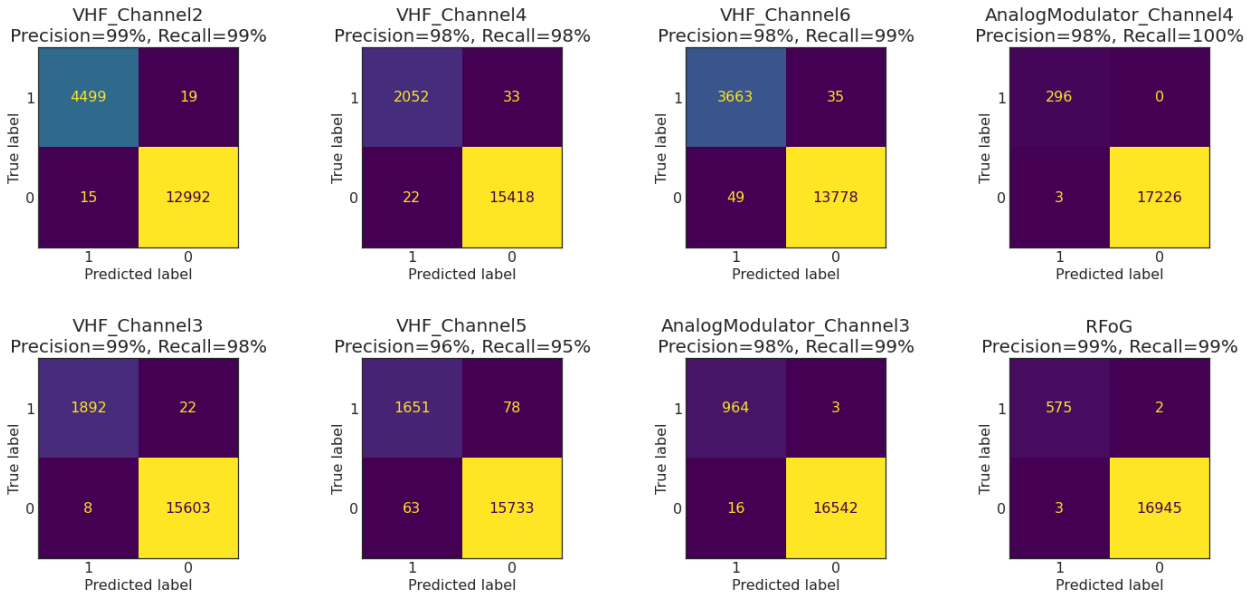


Figure 12 - Training/Validation Confusion Matrix

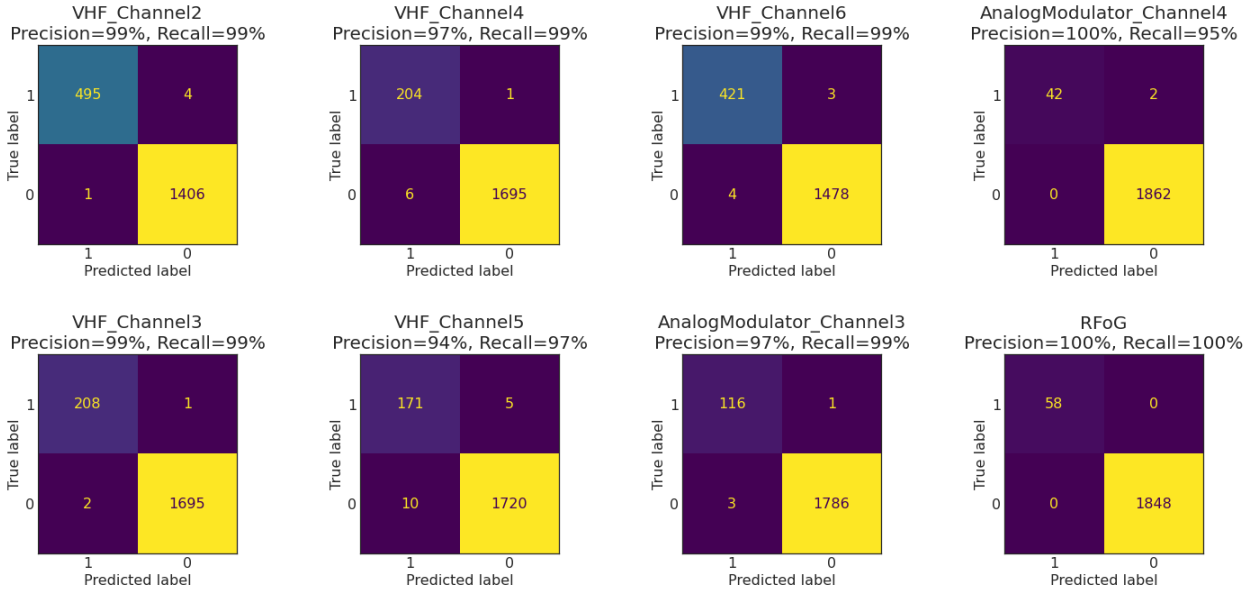


Figure 13 - Holdout Confusion Matrix

These initial results are extremely promising. We believe that the effectiveness of the model stems from both the CNN-based architecture and the substantial number of training samples.

We investigated the samples whose labels were predicted incorrectly by the model. A couple of examples of false negatives are shown in Figure 14 below. In both examples, multiple impairments exist, and the model did not predict the impairment highlighted in red. Note that 'Other' indicates the presence of an unknown impairment that the current model is not attempting the classify. In the first example, 'VHF – Channel 4' is not predicted by the model and it may be due to its severity being low. In the second

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

example, 'VHF – Channel 2' is not predicted by the model. We see that unclassified impairment(s) also exist in this example that interferes with the typical signature of a 'VHF – Channel 2' impairment. These missed classifications can usually benefit from additional training samples of the same genre.

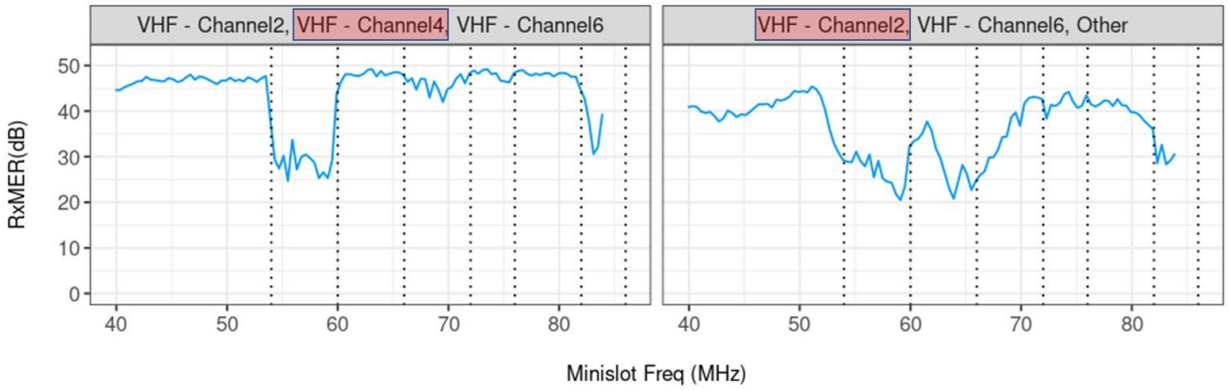


Figure 14 - False negative (FN) examples from the Holdout dataset

A couple of examples of false positives are shown in Figure 15 below along with their original and predicted classifications. In the first example, the model predicts the existence of 'VHF – Channel 2' in addition to 'Analog Modulator – Channel 3'. On closer inspection, some elements of a typical VHF signature are discernable in this example. In the second example, the sample was labeled with 'Other' due to the presence of unknown Suckouts that appear to occur periodically. One of the Suckouts is predominantly contained within Channel 4 that may have caused the model to predict it as 'VHF – Channel 4'.

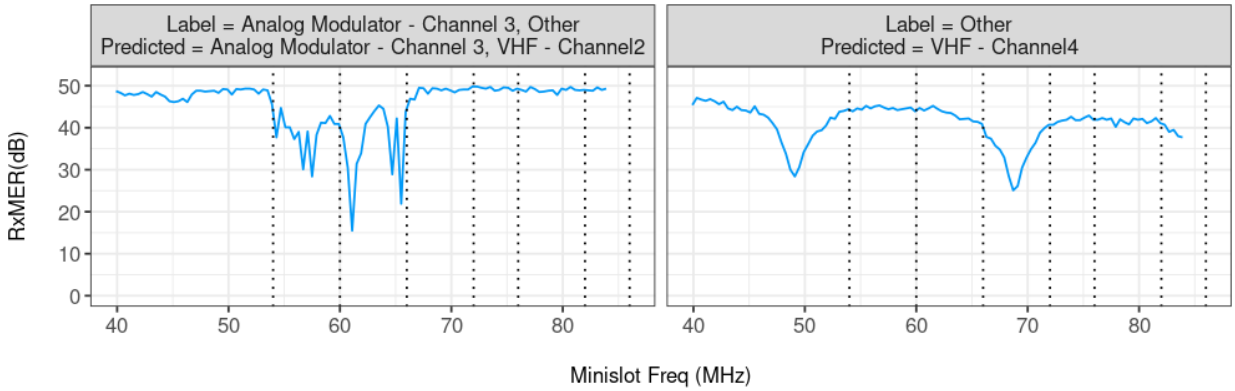


Figure 15 - False positive (FP) examples from the Holdout dataset

6. Machine Learning Operations

A machine learning pipeline was developed using Apache Spark to scale the model for the entire footprint of OFDMA-enabled devices on our network. Each record of the source data is comprised of a single MER capture for a single cable modem, with each capture being comprised of 111 data points. Over the course of 24 hours, each modem's MER is captured every 5 minutes for a total of 288 records per modem, per day. Processing each MER sample allows us to determine, with high confidence, how persistent or transient a pattern is manifested throughout the day.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

As of this writing, the pipeline needs to be capable of processing over 800 million records daily to achieve the most comprehensive coverage. As such, the pipeline's architecture was designed with horizontal-scaling capabilities in mind. Even at this volume, CPU-based compute machines are sufficient to process the data in a timely, cost-effective manner through a distributed compute cluster.

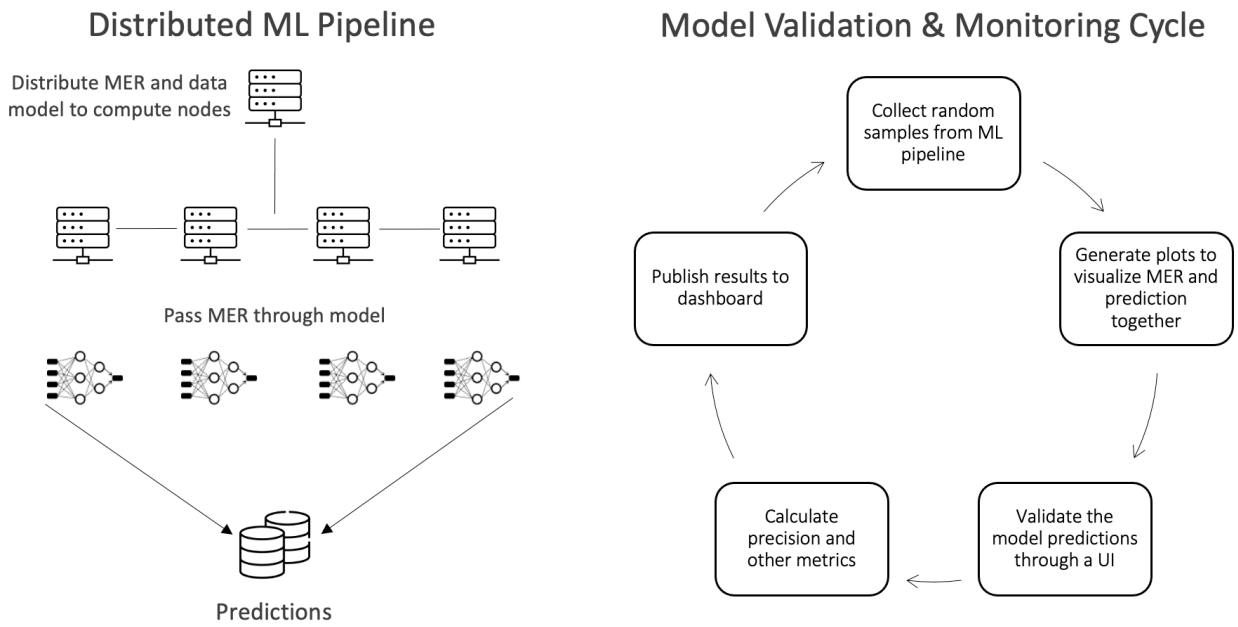


Figure 16 - ML Ops workflows for scaling the model and monitoring performance.

The other important workflow in MLOps is tracking model performance over time. We can understand performance by calculating the model's precision for each of the pattern types. The standard precision calculation is expressed as:

$$\frac{Number\ of\ True\ Positives}{Number\ of\ True\ Positives + False\ Positives}$$

To accomplish this, periodic reviews are completed on random samples of the model's predictions on real-time data. The samples are loaded into the same labeling tool used to build the training data,and an expert manually reviews the model's prediction alongside the plot of the MER. Samples with an incorrect prediction are flagged accordingly, and the results of the review are used to calculate the precision metric, along with other standard model metrics.

From an operations perspective in this situation, any false positive detection could contribute to a technician being mistakenly dispatched to solve a non-existent problem. For this purpose, the precision metric is closely monitored over time and evaluated against a threshold.

7. Conclusion

We have so far seen excellent results in being able to identify impairments such as VHF, Analog Modulator and RFoG Ingress. As part of ML Ops, we'll continue to monitor and validate the performance of the current model over time and take the necessary steps to ensure model performance remains optimal.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

We believe that the model performance seen for known patterns will also translate to some of the unknown impairments seen in OFDMA once we are able to classify them and generate a sufficiently large, labeled dataset.

We have made significant improvements to our impairment detection in the Downstream Single Carrier QAM (DS-SC-QAM) sections of the spectrum by incorporating Root Cause Analysis (RCA) algorithms. These algorithms utilize a graph representation of the network topology, allowing us to pinpoint the source of the impairment more accurately. Despite the phenomenon of Upstream (US) noise funneling, where impairments impact multiple devices on a node segment, the actual root cause might be attributed to a single subscriber or a specific network element. Identifying methods to determine the root cause of US impairments would greatly reduce Mean Time to Repair (MTTR). We are actively exploring various approaches to isolate upstream impairments and are open to insights and expertise from industry experts in this area.

Abbreviations

1-D	one dimension
CM	cable modem
CMTS	cable modem termination system
CNN	convolutional neural network
CPU	central processing unit
dB	decibels
D3.0	data over cable service interface specification 3.0
D3.1	data over cable service interface specification 3.1
DS	downstream
DS-SC-QAM	downstream single carrier quadrature amplitude modulation
FDX	full duplex
FN	false negative
FP	false positive
HFC	hybrid fiber-coaxial
MER	modulation error rate
ML	machine learning
MHz	Megahertz
ML Ops	machine learning operations
MTTR	mean time to repair
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
OSS	operational support systems
OTA	over the air
PMA	profile management application
PNM	proactive network maintenance
RCA	root cause analysis
RFoG	radio frequency over glass
RxMER	receive modulation error ratio
ROC	receiver operating characteristic
QAM	quadrature amplitude modulation

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE	Society of Cable Telecommunications Engineers
SME	subject matter expert
SNR	signal-to-noise ratio
TN	True Negative
TV	Television
TP	True Positive
UI	User Interface
US	Upstream
VCR	Video Cassette Recorder
VHF	Very High Frequency

Bibliography & References

1. "Deploying PMA-Enabled OFDMA in Mid-Split and High-Split", Maher Harb, Dan Rice, Kevin Dugan, Jude Ferreira, Robert Lund, and Ramya Narayanaswamy, NCTA Technical Paper, 2022.
2. "Convolutional Neural Networks for Proactive Network Management", Jude Ferreira, Maher Harb, Karthik Subramanya, Bryan Santangelo, and Dan Rice, NCTA Technical Paper, 2020.
3. "A Deep Learning Approach for Detecting RF Spectrum Impairments and Conducting Root Cause Analysis", Kevin Dugan, Justin Evans, and Maher Harb, NCTA Technical Paper, 2022.
4. "Visualizing and Understanding Convolutional Networks", Mathew Zeiler, and Rob Fergus, https://arxiv.org/pdf/1311.2901.pdf

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

DNS Cowboys, On the Edge of a New Frontier

Lowering Customer DNS Latency with vCMTS Edge Compute

A Technical Paper prepared for SCTE by

Charlie Helfinstine
Sr. Principal Engineer
Comcast
1800 Arch St. Philadelphia, PA
267-858-8002
charles_helfinstine@comcast.com

De Fu Li, Distinguished Engineer, Comcast

Eric Stonfer, Sr. Principal Engineer, Comcast

Joe Crowe, Principal Engineer, Comcast

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. The Drive for Lower Application Latency ... 3
   - 2.1. Emerging Markets for Lower Latency ... 3
   - 2.2. Emerging Technologies for Lower Latency ... 4
   - 2.3. Lower Latency Network Services ... 5
3. Shrinking DNS Latency in the Network ... 5
   - 3.1. DNS Customer Caching Locations in Provider Networks ... 5
   - 3.2. Current DNS Infra and the Latency Provided to Customers ... 6
4. Leveraging vCMTS Edge Compute for Lower DNS Latency ... 7
   - 4.1. vCMTS Edge Cloud Background Information ... 8
   - 4.2. Deploying DNS on vCMTS Edge Cloud Platform (High Level Design) ... 8
5. Conclusion ... 10

Abbreviations ... 11

Bibliography & References ... 11

List of Figures

Title **Page Number**

Figure 1 – Low Latency Technology Areas ... 4
Figure 2 – DNS Caching Locations ... 5
Figure 3 – DNS Service Locations ... 6
Figure 4 – DNS Service Locations with vCMTS ... 7
Figure 5 – vCMTS Edge Coud (Simplified View) ... 8
Figure 6 – vCMTS Namespace Partitions ... 9
Figure 7 – DNS Utilization per CM ... 10

List of Tables

Title **Page Number**

Table 1 – Comcast DNS Latency ... 7

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

With cable wireline users demanding low-latency services for new applications like hosted gaming and virtual reality (VR), Multiple System Operators (MSO) are lowering the latency of critical network services like Domain Name System (DNS) by hosting them on edge compute infrastructure, such as Virtual Cable Modem Termination Systems (vCMTS).

vCMTS platform is a multi-site, Kubernetes edge compute platform deployed at scale to deliver the Xfinity 10G Network experience. It has been focused on delivering best-in-class network access services to residential and business customers but is being expanded to support new network service capabilities such as DNS. No other computing platform has been deployed this close to customers inside the Comcast Network. If edge computing is the new frontier, then vCMTS platforms are literally at the edge.

Our paper and presentation will outline work underway to host DNS residential caching resolvers on the vCMTS and enhance the multi-tenant capabilities of the platform to host third-party applications. In addition, the integration of vCMTS DNS into the residential DNS cache operations, engineering, and future scaling will be described. Moving network functions like DNS caching closer to the network edge will lower overall DNS latency and enhance the DOCSIS 10G Network experience.

2. The Drive for Lower Application Latency

Since the advent of Transmission Control Protocol / Internet Protocol (TCP/IP), latency has been critical for determining user experience with network applications. From the responsiveness of online games and interactive media to the needs of video conferencing, limiting the network delay is a key component for a modern network. The overall latency is an aggregation of component delays and propagation delays between network devices spanning from end-to-end on every packet transit. Limiting the number of components, the delay through network devices, and the distance between them all play a part in reducing the latency experienced by the end user.

Applications attempt to limit the number of round-trips traffic must take across the network; if there are ways to shorten the path, the effect can be multiplied across the frequent traffic needed for continuous interactive applications. Ongoing research in low latency technologies is opening the opportunity for new applications such as edge-cloud hosted gaming.

Customer applications that strive for lower latency may be optimized to have quick responses, but they are reliant on common network services like DNS to not act as a latency bottleneck and a source of unexpected experience. The effects of DNS latency can impact any application, and care must be taken to enhance the network latency of DNS to keep pace with the push for lower latency.

2.1. Emerging Markets for Lower Latency

Basic responsiveness is important for all applications, but there are widespread and growing new needs for low network latency from online gaming and teleconference meeting systems. Recent advances in network latency enable hosted gaming environments that compete with the console and mobile gaming markets. The time between a user's controller input and the response of the game, alongside interactions with other players, are critical for the entertainment value of a game. In online meetings, the added delay between participants in a conversation contributes to communication difficulties and loss of the natural rhythm of speech.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Longstanding low-latency applications like content delivery networks and video-on-demand systems are also driving lower latencies due to the competitive market for the services. Any improvement in network latency for a provider's services can be a differentiator in a crowded space.

2.2. Emerging Technologies for Lower Latency

A number of recent network initiatives are driving lower latency at almost every point in the network. Whether the work is in response to the growing needs of applications or has enabled the viability of new uses, there is a focus on optimizing queuing and congestion management to push the most throughput and lowest latency out of TCP and QUIC transport protocols.

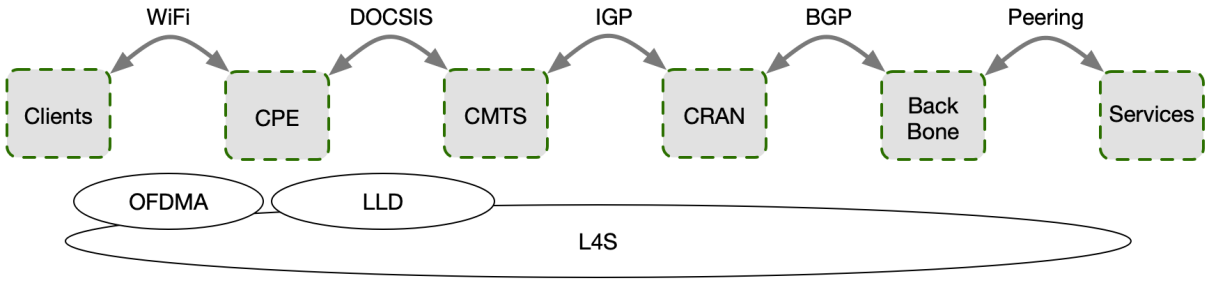


Figure 1 – Low Latency Technology Areas

The Low Latency, Low Loss, Scalable Throughput Internet Service (L4S) architecture described in RFC9330 is one such project that is developing a systemic approach for lowering end-to-end latency in congested networks. It describes the use of Explicit Congestion Notification (ECN) congestion flow-control signaling as a way to hold off packet drops and limit the retransmit and TCP window size scaling effects on overall throughput. However, to feel the benefits of ECN signaling, care must be taken to ensure that appropriate queue management is enforced on the ECN flows end-to-end to ensure optimal delivery of latency-sensitive traffic mixed among all traffic. There is significant engineering work needed to implement L4S services across large networks with standard queuing paths, with the goal of lower latency achieved when complete end-to-end paths are upgraded to support the queue management features required.

Another project driving down latency closer to the network edge is Low Latency Data Over Cable Service Interface Specification (LLD). As part of DOCSIS 3.1, LLD lowers the queueing delay and media acquisition delay for application traffic from sources that are termed Non-Queue-Building applications. Such applications include interactive gaming systems and online meeting tools that attempt to send lower capacity time-sensitive flows across the network and are less likely to overflow queues in the path. LLD will support these apps with a low-latency queue. Measurements show that low-latency queue performance can lower latency from the standard 10ms down to 1ms.

Even the last link to client devices, Wi-Fi, is lowering latency with Orthogonal Frequency-Division Multiple Access (OFDMA) modulation included in Wi-Fi 6 and 6E. OFDMA provides subchannel access between wireless endpoints that can allow better utilization of available spectrum and queuing of traffic. It's possible to lower application traffic latency over OFDMA for lightweight flows that can fit into the subchannel multiple access. The same apps that may be considered Non-Queue-Building applications in LLD are the types of applications that may best utilize the OFDMA subchannels.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Across L4S, LLD, and Wi-Fi 6, there is a shared strategy of low-latency queueing. Each technology focuses on a different segment of the network, but together they can enable better end-to-end latency for applications like hosted gaming or online meetings.

2.3. Lower Latency Network Services

Alongside application data, network services like DNS benefit from lower latency. Most network interactions start with DNS resolution, which needs to complete before further application flows may start, making the DNS resolution time additive to the overall application performance. DNS latency can be improved with caching located close to customer devices and by leveraging the lower latency queueing capabilities of the network in the future.

3. Shrinking DNS Latency in the Network

Providing DNS customer caching services at MSO service provider scale with low latency is a challenge and must be pursued with a goal of zero downtime. The needs of low-latency applications are pushing service providers to embrace new approaches for DNS that can lower latency at the network edge while integrating into existing DNS architectures and maintaining Service Level Agreement (SLA) expectations.

With the emerging LLD and L4S protocols providing a path forward for low-latency applications to traverse the network, DNS service appears to be a good candidate to use the low-latency queues when available. Standard DNS services are low-bandwidth flows using Non-Queue-Building protocols with short packet sizes that will lend well for minimizing queue depth in low-latency queues. As LLD and L4S adoption increases, DNS's use of the low-latency systems should increase in tandem.

In the interim, DNS latency may be improved by managing the location of caching services provided in the network. Moving the customer caching services closer to the end users provides a clear path to lowering the latency observed for DNS services.

3.1. DNS Customer Caching Locations in Provider Networks

DNS caching services for residential and business customers can be hosted at a number of different locations with differing effects on latency. Service providers who operate a network edge can deploy DNS services centrally off their backbone, in main data centers, in regional network sites or datacenters closer to the network edge, or in Customer Premises Equipment (CPE) on site with the end user.

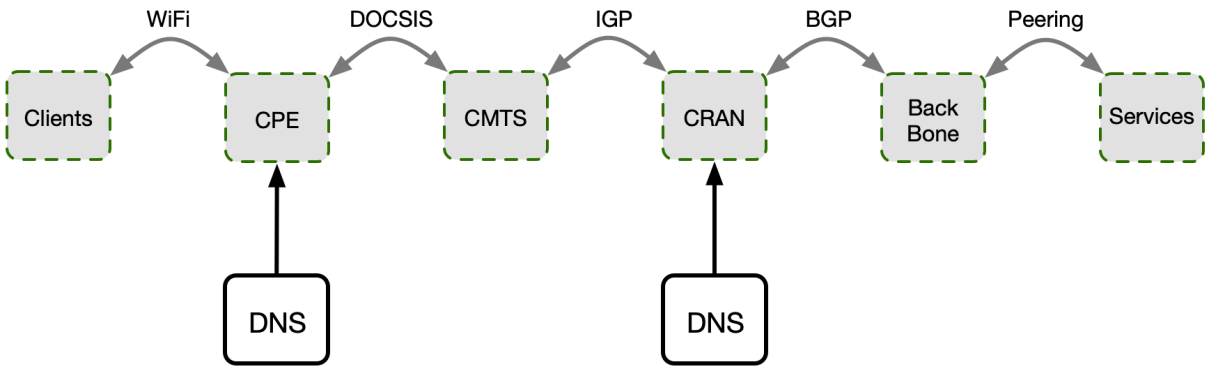


Figure 2 – DNS Caching Locations

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

By counting hops and considering the speed of light through fiber, the natural assumption is that the closer to the end user, the lower the network latency possible for the DNS service. However, there is a complexity to different deployment strategies. DNS cache hit rates are affected by the number of subscribers sharing a given cache view, and placing a cache too close to a few subscribers will benefit the resolution of domains that are widely used by almost all clients, but there is a long tail of DNS entries supporting important applications that are less frequently queried yet need fast resolution.

Placing the physical infrastructure for DNS services also increases in cost the closer you get to the edge of the network. When a service can be deployed in a limited number of sites, the cost per unit of DNS service is lower than a corresponding deployment of the same capacity that is located in exponentially more sites that are geographically closer to customers - due to the expense of installation and deployment being roughly similar per site independent of capacity.

Due to the automation of changes and upgrades, the number of sites supported operationally is less important than in the past, but the development effort and the room for error still increase with the number of sites. In an environment where little or no operational intervention is needed for DNS services, the number of DNS sites also has less relevance. For instance, hosting DNS caching services on customer premises equipment is a common approach that is very close to customer devices, however, conducting operational DNS functions such as cache flushes or deploying config changes becomes very difficult.

3.2. Current DNS Infra and the Latency Provided to Customers

MSOs commonly provide residential and commercial DNS resolution services across their footprints with a combination of regional sites and centralized sites. As an example, the Comcast network is subdivided into Converged Regional Access Networks (CRAN) that serve differing numbers of subscribers ranging from hundreds of thousands to millions based on geographical location, and each CRAN hosts a corresponding DNS site. The CRAN resolving service is advertised with anycast routing on the well-known addresses of 75.75.75.75 and 2001:558:FEED::1 as the primary DNS address given by DHCP, and a centralized copy is advertised on 75.75.76.76 and 2001:558:FEED::2 as the secondary DNS address.

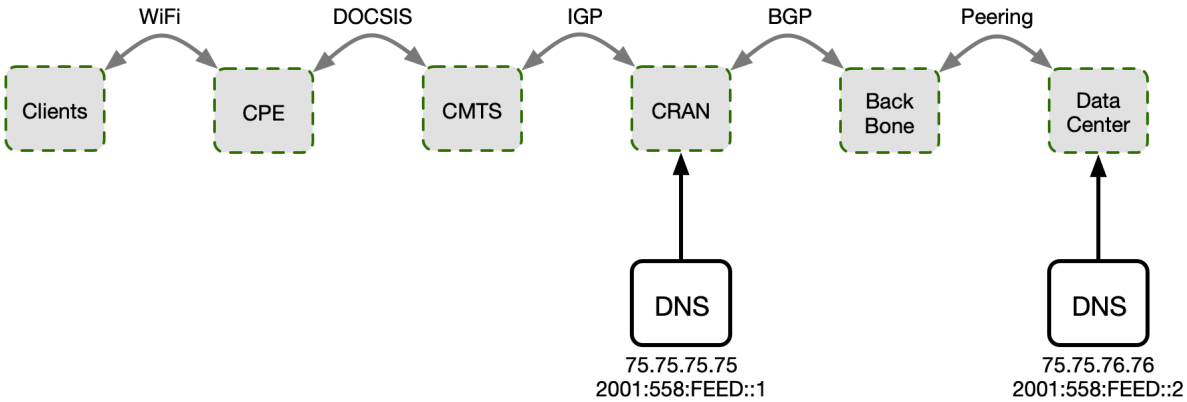


Figure 3 – DNS Service Locations

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The regional site deployment model provides close, low-latency DNS service to customers and high cache hit rates from the client scale that maps to a given site. DNS latency data has been shared in more technical detail before, but for context, Comcast customers commonly see DNS latencies ranging as shown.

Table 1 – Comcast DNS Latency

DNS Anycast Address	Typical Latency (ms)
75.75.75.75	18
75.75.76.76	25

Source: RIPE Atlas

It should be noted that Comcast does not cache DNS records on CPE devices due to scaling concerns around flushing stale cache entries. Additionally, Comcast resolvers provide Domain Name System Security Extensions (DNSSEC) validation of signed domains and encrypted DNS services that do not scale well at the CPE level.

4. Leveraging vCMTS Edge Compute for Lower DNS Latency

With the drive to low-latency applications, MSOs are working to lower the DNS latency seen by customers by placing new DNS service components closer to the network edge. The virtual CMTS (vCMTS) platform, deployed as a key component of the Xfinity 10G network, offers a computing environment located as close as possible to the customer prem without being on-site and can host DNS services with significantly lower latency.

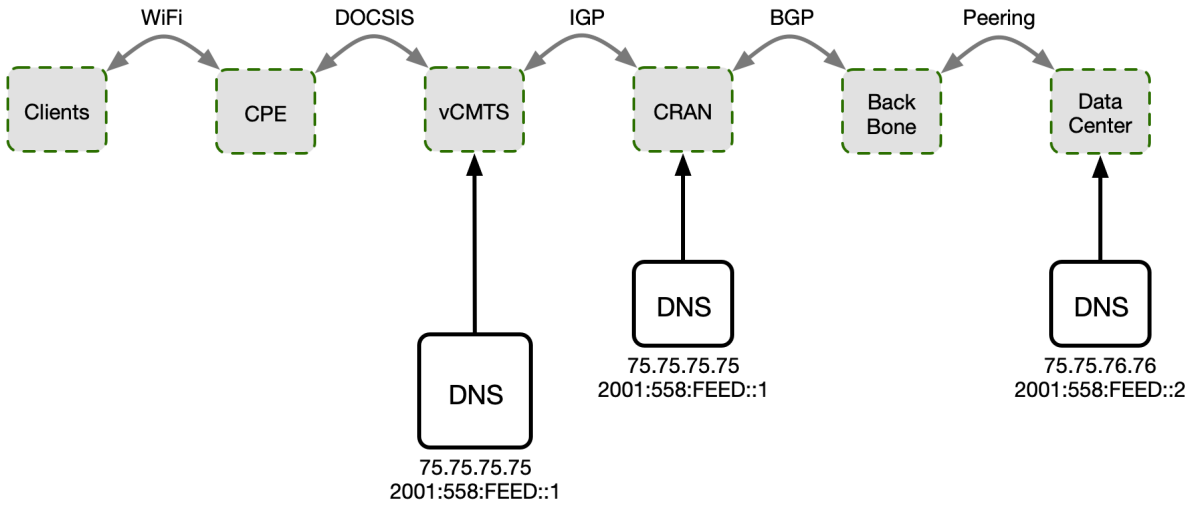


Figure 4 – DNS Service Locations with vCMTS

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.1. vCMTS Edge Cloud Background Information

The broadband access over the cable access medium is defined by the Data Over Cable Service Interface (DOCSIS) specifications. The cable modem is the customer-premises equipment (CPE). The cable modem termination system (CMTS) is the headend equipment. Earlier generations of the CMTS were built on custom hardware platforms. As part of the DOCSIS specification umbrella, the distributed access architecture (DAA) specification splits the CMTS into two distinct components: the physical function and the core function. The Remote Physical Layer (PHY) Device (RPD) provides the PHY function, while the core functions consist of CMTS operating on the MAC and upper layers. With the advance of data center technologies, the CMTS core function is realized in software running on Commercial Off-The-Shelf (COTS) servers and virtualized. Hence, the virtual CMTS (vCMTS) is referred to as a software implementation of the DAA CMTS core function.

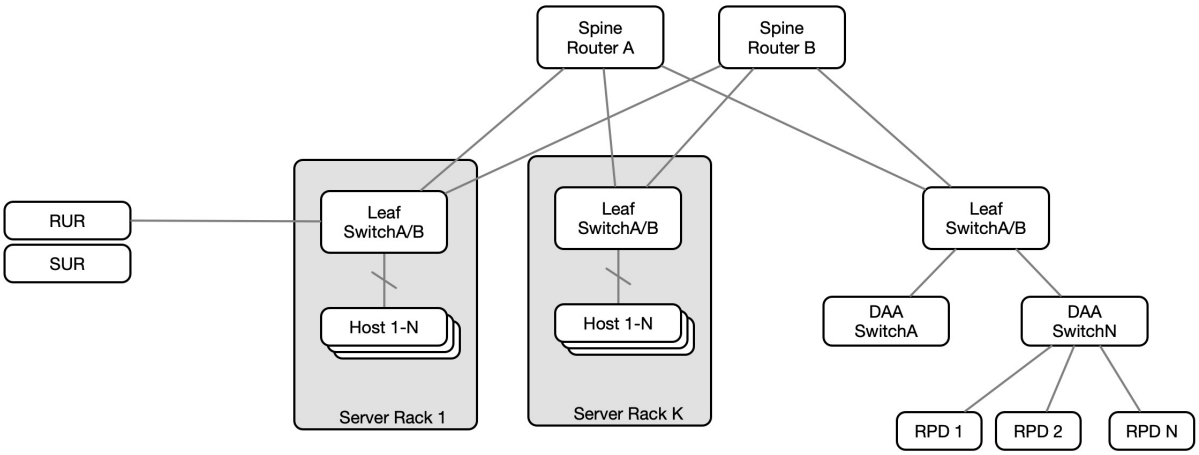


Figure 5 – vCMTS Edge Coud (Simplified View)

To host the vCMTS software applications, private edge clouds physically located in selected headends are required. Figure 5 shows a simplified diagram of the vCMTS edge cloud platform. The RPDs are connected to the DAA aggregation switch. There are multiple racks of servers that host the vCMTS and supporting applications. The vCMTS software is instantiated or orchestrated to run on the host; there are hundreds and thousands of these vCMTS workloads (instances). The network traffic flows between the workloads on the host and RPDs interconnected via the datacenter leaf-spine switch fabric.

4.2. Deploying DNS on vCMTS Edge Cloud Platform (High Level Design)

As mentioned earlier, the virtual CMTS (vCMTS) platform is a multi-site, Kubernetes edge compute platform deployed at scale to deliver the DOCSIS 10G Network experience. It has been focused on delivering best-in-class network access services to residential and business customers but is being expanded to support new network service capabilities such as DNS.

The vCMTS applications are deployed and orchestrated using Kubernetes. The microservice workloads are logically partitioned into Kubernetes namespaces. This is illustrated in Figure 6. The workloads for the vCMTS core function are orchestrated to its own namespace. Similarly, there are many software infrastructure services, such as a REDIS cluster for in-memory databases.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

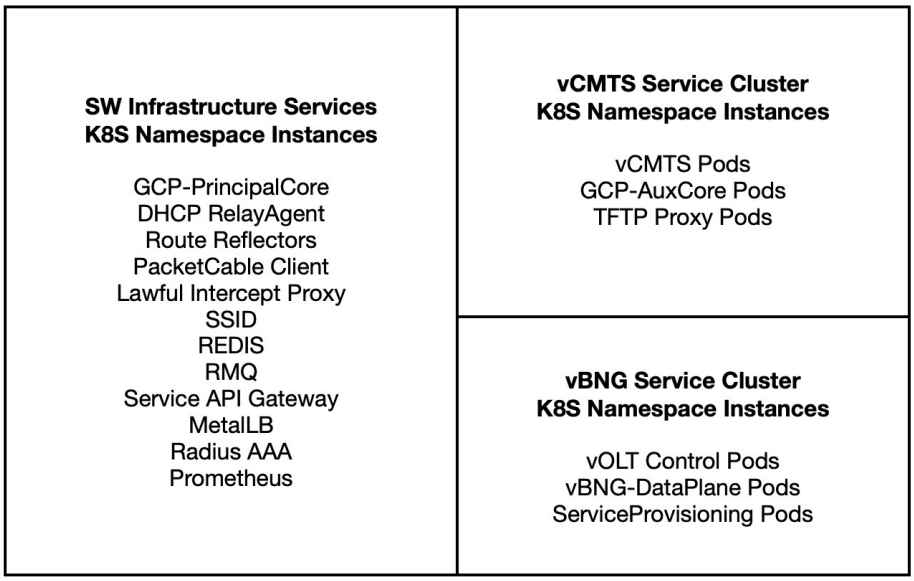


Figure 6 – vCMTS Namespace Partitions

The bulk of vCMTS workloads perform network function virtualization (NFV) tasks for DOCSIS core functions which are typically compute and network bandwidth-intensive Data Plane Development Kit (DPDK) applications. In contrast, the infrastructure microservices like REDIS have very different CPU and network bandwidth requirements, more in line with traditional application services.

Despite being a network service, the DNS cache workload functions more like a traditional application than an NFV component that is better suited for the infrastructure microservice category. Hence, one consideration is deploying it in the infrastructure namespace, where it can be managed as other shared components.

With the vCMTS disaggregated forwarding plane and Kubernetes container network interface, careful integration of the DNS service is required. Existing anycast DNS addresses will be advertised by vCMTS route reflectors locally in the routing table of the vCMTS to forward client DNS traffic to the ingress service for the local DNS caching pods. Customer devices will use the same DNS service addresses of 75.75.75.75 and 2001:558:FEED::1, however, the vCMTS DNS pods will perform the DNS caching and recursive resolution instead of the regional network DNS anycast services.

The local anycast route advertisement is subject to status monitoring performed on the DNS pod, and any interruption in DNS service will cause the local anycast route to be withdrawn. Route filtering in the vCMTS table ensures that the best route to the anycast addresses after a local anycast route is withdrawn will be the regional CRAN resolvers that provide DNS services at scale today. vCMTS instances in the same site will not be used as fail-over resources for DNS services inside the site. Typically, capacity is reserved for failover scenarios and growing demand in anycast systems, but by planning the fail-over paths, the capacity reservations can be absorbed by the scaled systems in the CRAN.

When creating load models for the vCMTS DNS service, an analysis was performed to predict the DNS traffic volume expected from a set number of subscribers. This volume is expected to grow with the numbers of devices in households and use characteristics of popular applications, however, as a baseline calculation, measurements were taken through 2022 and 2023 on observed queries per second counts over Comcast's DNS customer caching service per region and indexed to the subscriber counts per region. The

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

results were remarkably consistent across the footprint and yielded a capacity model that is scaled for implementing DNS services on the vCMTS.

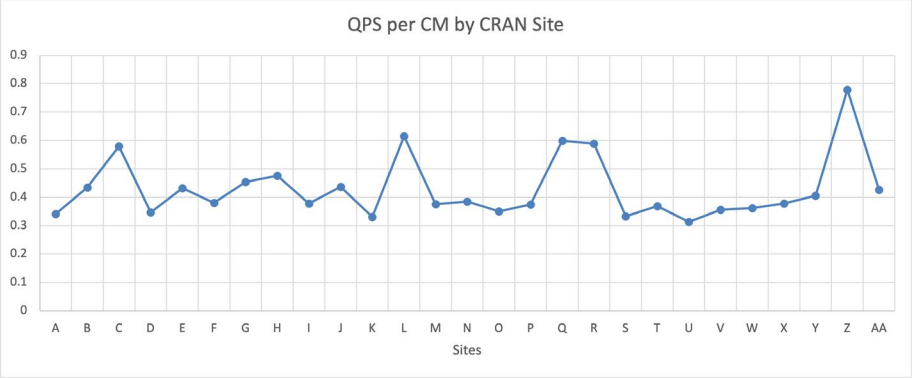


Figure 7 – DNS Utilization per CM

Functionally, the vCMTS DNS service supports the same feature-set as the CRAN customer caching resolvers with DNSSEC validation for signed domains and dual-stacked services among many others. The uniformity of the service requires that operational commands such as DNSSEC negative trust anchors are applied in both the CRAN locations and the vCMTS locations when needed. Operational reliability expectations are the same for DNS services hosted on vCMTS as the services hosted in CRAN locations. DNS system performance, status monitoring, and alerting, capabilities need support and equal vigilance.

5. Conclusion

To meet the push for low-latency services, MSOs can expand DNS services to the vCMTS edge compute platform. Development work is ongoing to expand the DNS service closer to end users and support rising next generation low-latency applications.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

CM	Cable Modem
CMTS	Cable Modem Termination System
COTS	Commercial Off-The-Shelf
CPE	Customer Premises Equipment
CRAN	Converged Regional Access Network
DPDK	Data Plane Development Kit
DOCSIS	Data Over Cable Service Interface Specification
DNS	Domain Name System
DNSSEC	Domain Name System Security Extensions
ECN	Explicit Congestion Notification
IP	Internet Protocol
L4S	Low Latency, Low Loss, Scalable Throughput Internet Service
OFDMA	Orthogonal Frequency-Division Multiple Access
RPD	Remote PHY Device
SLA	Service Level Agreement
TCP	Transmission Control Protocol
vCMTS	Virtualized Cable Modem Termination System

Bibliography & References

[RFC9330] Briscoe, B., Ed., De Schepper, K., Bagnulo, M., and G. White, "Low Latency, Low Loss, and Scalable Throughput (L4S) Internet Service: Architecture", RFC 9330, DOI 10.17487/RFC9330, January 2023, https://www.rfc-editor.org/info/rfc9330.

CableLabs: Low Latency DOCSIS: Technology Overview 2019

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

DOCSIS 4.0 and Wi-Fi 7

Perfect Together for Multiple Gigabits per Second Speed

A Technical Paper prepared for SCTE by

David John Urban
Fellow
Comcast
One Comcast Center Philadelphia
610-476-2596
david_urban@cable.comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. DOCSIS 4.0 4
3. Wi-Fi 7 6
4. Channel Width: EHT ups the ante to 320 MHz 9
5. EHT Preamble 12
   5.1. L-STF to detect the signal 13
   5.2. L-LTF 13
   5.3. L-SIG 13
   5.4. RL-SIG 13
   5.5. U-SIG Introducing version independent fields 13
   5.6. EHT-SIG 13
   5.7. EHT-STF 13
   5.8. EHT-LTF for receiver MIMO channel estimation 13
   5.9. PE packet extension field 14
6. Let's talk about spectrum. 14
7. Modulation and Coding; Introducing 4k-QAM to Wi-Fi 19
8. The hidden node problem 25
9. A quick word on beacons and probes 26
10. DCM+DUP: How to use 320 MHz channel width to improve range 28
    10.1. EHT DUP transmission 29
11. Dual Lane PCIe3: The interface between DOCSIS 4.0 and Wi-Fi 7 30
12. Conclusion 30

Abbreviations 31

Bibliography & References 32

List of Figures

Title **Page Number**

Figure 1 – EHT MU PPDU fields and typical durations 8
Figure 2 – Range is improved with increased channel width for a required throughput 11
Figure 3 – Improvement in range for 320 MHz with constant power spectral density 12
Figure 4 – U-NII band 15
Figure 5 – 6 GHz band preferred scanning channels and center frequencies in MHz 15
Figure 6 – Spectrum analyzer plot of 6g133/320-2 over 6 GHz band 16
Figure 7 – Spectrum analyzer plot of 6g37/320-1, 6g101/320-1, 6g165/320-1 18
Figure 8 – Spectrum analyzer plot of 6g69/320-2, 6g133/320-2, 6g197/320-2 18
Figure 9 – Spectrum analyzer plot to 8 GHz with EHT 6g37/320-1 max download 19
Figure 10 – MCS13 QAM-4096 LDPC 5/6 320 MHz 802.11be Wi-Fi 7 signal 20
Figure 11 – BPSK signal showing spectrum, constellation, time domain signal, EVM. 21
Figure 12 – QPSK modulation maps 2 bits to each subcarrier. 22
Figure 13 – 16-QAM modulation maps 4 bits to each subcarrier. 22
Figure 14 – 2x2 320 MHz EHT device data rate including impact of preamble 25
Figure 15 – 2x2 320 MHz EHT device data rate including impact of preamble and RTS CTS Block Ack 26

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 16 – The measured beacon duration of a 320 MHz 6GHz band EHT signal. 27
Figure 17 – The beacon is repeated over the full 320 MHz channel width 28

List of Tables

Title **Page Number**

Table 1 – EHT PPDU fields with typical durations 8
Table 2 – Subcarrier allocation of 320 MHz EHT modulated fields 10
Table 3 – EHT-LTF symbol duration options 14
Table 4 – USA FCC Unlicensed National Information Infrastructure Frequency in MHz 14
Table 5 – 6 GHz band 320 MHz channels 16
Table 6 – Modulation 2x2 320 MHz Wi-Fi 7 station 23
Table 7 – PHY rate in Mbps and SNR for 20, 40, 80, 160, 320 MHz channel width 24

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

It is a thing of beauty when two interdependent technologies arrive at the same time wherein each needs the other to deliver service to customers. Such is the case with the newest cable modem standard for Data-over-Cable Service Interface Specification DOCSIS 4.0 and Institute for Electrical and Electronic Engineers (IEEE) 802.11be standard Wi-Fi 7. Two technologies under development for many years coming out of the lab and into customer homes. Wi-Fi 6 and 6E wireless adapters in many phones, computers and tablets have a maximum physical symbol, PHY rate of 2400 Mbps, (million bits per second). With a 2400 Mbps PHY rate throughput of file downloads, speed tests can reach 1.2 gigabits per second (Gbps). Wi-Fi 7 increases the modulation to 4096-QAM, quadrature amplitude modulation, and the channel width to 320 MHz, million cycles per second, resulting in a 2x2 client wireless adapter with a maximum PHY rate of 5764 Mbps, enough PHY rate for a throughput of 4.6 Gbps. About one and a half years ago, the first laboratory trials of DOCSIS 4.0 demonstrated speeds faster than 4 Gbps upload and download. This paper will describe the underlining technology propelling Wi-Fi 7 based on the IEEE 802.11be standard that will allow the speeds delivered by the DOCSIS 4.0 cable modem and the DOCSIS 4.0 infrastructure sending the signal to the cable modem to be delivered all the way to customer phones, tablets, and computers. In addition to higher order modulation and wider channel width, Wi-Fi 7 adds key technology to improve the range in the 6 GHz, billion cycles per second, band. Improving the range in the 6 GHz band is just as important as improving the speed. After all, high speed in the 6 GHz band is not much good to customers if they need to use another band to maintain coverage. Dual carrier modulation, duplication of dual carrier modulation, increased transmit power at 320 MHz channel width, and improved receiver sensitivity are key aspects of Wi-Fi 7 to improve the 6 GHz band range. This paper explains how modulating two carriers within a 20 MHz block of spectrum and duplicating these 20 MHz blocks results in improved dynamic range and better coverage. The paper provides measured results in both the lab and customer homes showing DOCSIS 4.0 and Wi-Fi 7 are perfect together.

2. DOCSIS 4.0

Cable operators are building next generation networks to provide multigigabit upload and download speeds at scale. Increased reliability, performance, and lower latency are benefits to customers provided by the next generation 10G network. The 10G network is a natural evolution of the original hybrid fiber coaxial cable (HFC) concept.

In an HFC architecture radio frequency (RF) signals in the 5-42 MHz upstream band and the 54-750/860/1002 MHz downstream band are created at the head end and carried over long distance with a low loss fiber optic cable using modulated transmitter and receiver laser to a fiber node. The extent of the downstream frequency varies. The fiber node converts the RF signal from optical to electrical, to distribute to cable modems and set top boxes in customer homes over coaxial cable, amplifiers, directional couplers, taps, and cable drops to homes. The signals to the cable modem (CM) are created at the head end by the cable modem termination system (CMTS). The CMTS converts 10 Gbps Ethernet signals for high-speed data to RF signals in the 5-42 MHz upstream and 108-750/860/1002 MHz downstream spectrum to feed optical receivers and transmitters that feed the HFC network. Over the years the downstream spectrum has been steadily expanded to 750, 860, and 1002 MHz as needed on a per fiber node basis and even the upstream in some cases has been expanded to 40 to 42 to 85 MHz upper end upstream spectrum. The ability of HFC to incrementally add capacity as needed with advanced technology with small changes to filters, amplifiers, and spacing has been key to the success of HFC to deliver voice, video, and phone to customer homes, continually meeting demand in a cost-effective way.

The 10G network replaces the CMTS with a virtual cable modem termination system (vCMTS). The vCMTS is located in the head end or even deeper in the operator's core network of routers linked with

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

100 Gbps connections distributed nationwide. The vCMTS provides the functionality of the CMTS without the RF converters that create the upstream signals in the 5-85 MHz band and downstream signals in the 108-1200 MHz band. The RF converters are moved to a remote physical interface node (RPHY) node. The RPHY node converts a 10 Gbps Ethernet interface carried over optical fiber to upstream RF signals in the 5-85 MHz band and downstream signals in the 108-1200 MHz band. The RF signals from the RPHY node are delivered to cable modems and set top boxes in customer homes over coaxial cable with trunk feeds, directional couplers, splitter taps, and drop coaxial cables. For a frequency division duplex system (FDD), the boundaries of the upstream and downstream allocated spectrum can be adjusted as upload and download demand from customers changes over time. For a full duplex DOCSIS system (FDX) the spectrum from 108 to 684 MHz within the downstream band of 108-1200 MHz can be used for upstream as well as downstream signals. This network architecture is still a hybrid of fiber optical cable and coaxial cable, thus an HFC network. To distinguish the new HFC from the old, the new 10G network is called a distributed access architecture (DAA) in order to highlight the virtualization of the CMTS.

10G and DAA deliver greater reliability by providing visibility into Internet performance from the core to the gateway and every element in between. Download speeds of 8 Gbps and upload speeds of 5 Gbps have been demonstrated over an HFC cable network to a DOCSIS 4.0 cable modem. (Comcast Demonstrates Fastest-Yet Speeds Over a Complete 10G Connection on a Live Network, 2022) Lab tests of FDX CM have delivered upload and download speeds faster than 4 Gbps using 10G network technology. (Comcast Announces World-First Test of 10G Modem Technology Capable of Delivering Multigigabit Speeds to Homes, 2022) A key component of the 10G network is low latency DOCSIS (LLD) based upon Internet Engineering Task Force's (IETF) open standards to mark latency sensitive traffic for interactive applications (Comcast Kicks Off Industry's First Low Latency DOCSIS Field Trials, 2023). All the advances in the 10G wide area network (WAN) require a next generation wireless local area network (WLAN) to match. Wi-Fi 6 will not suffice; Wi-Fi 7 comes to the rescue.

The focus of this paper is how Wi-Fi 7 can deliver the speeds of DOCSIS 4.0 to customer phones, tablets, computers, watches, and televisions. This section will provide a summary of the speeds of DOCSIS 4.0 as a WAN connection to a home router with WLAN (wireless local area network) radio to customer devices.

To get a rough idea of the speed capability of DOCSIS 4.0, over 1,000 MHz of spectrum is utilized from 5 MHz up to 1,200 MHz and perhaps beyond. 1,000 MHz of bandwidth with 10 bits/sec/Hz spectral efficiency yields a data rate of 10 Gbps. 4096-QAM modulates each OFDM (orthogonal frequency division multiple access) subcarrier with 12 bits, allowing for error correction and overhead to still get to a spectral efficiency of around 10 bits/s/Hz. Coaxial cable is bi-directional, meaning signals carried by the coaxial cable in both forward and reverse directions at the same time using the same frequency spectrum can be separated for demodulation in the receivers without filtering using a directional coupler. Thus, in theory a 1,000 MHz coaxial distribution system can support 10 Gbps in the upstream and downstream direction at the same time.

Directional couplers effectively separate the upstream and the downstream of a coaxial cable without the need for filtering of distinct upstream and downstream spectrum. The transmitted signal reflects back into the receiver due to impedance mismatches in the HFC plant. The reflected forward signal propagates in the coaxial cable in the same direction as the reverse signal. The directional coupler cannot separate the reverse signal from an echo of the forward signal since these two signals travel inside the coaxial cable in the same direction. An echo canceller is needed to feed samples of the transmitted signal back into the receiver with just the right time delays, amplitude, and phase to cancel the unwanted echoes of the transmitted signal inside the receiver.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The details of DOCSIS 4.0 are described in reference (Robert Howald, 2022) as well as many other excellent references. The key takeaway from reference (Robert Howald, 2022), for the purposes of this paper, is that 4 Gbps upload and download speeds become practical with DOCSIS 4.0 cable modems. 4 Gbps is not a speed that a Wi-Fi 6 or Wi-Fi 6e phone, tablet, or computer today can deliver. Wi-Fi 7 is needed to match the speed of DOCSIS 4.0 with customer devices.

3. Wi-Fi 7

Wi-Fi 7 is based on the IEEE 802.11be standard (IEEE 802.11be D3.0, 2023) for extremely high throughput. The increase in data rate is primarily accomplished with a doubling of the channel width in the 6 GHz band to 320 MHz and a quadrupling of the number of points in the constellation of the highest modulation from 1024 to 4096.

The IEEE 802.11be standard introduces a new physical layer interface (PHY) called extremely high throughput (EHT). EHT defines many new parameters, some borrowed from older standards and some novel. A definition of the timing-related parameters new to EHT PHY is needed to calculate the speeds of Wi-Fi 7 and determine if Wi-Fi 7 can carry the speeds of DOCSIS 4.0 to customer devices.

Subcarrier frequency spacing for the pre-EHT modulated fields is defined in equation (1).

$$\Delta_{F,Pre-EHT} = 312.5kHz \tag{1}$$

All transmissions are not EHT, even for EHT access points (AP) and stations (STA). Management and control signals as well as preambles of EHT signals mostly use 802.11a signals with data rates of 6, 12, and 24 Mbps. For EHT signals these are called pre-EHT. When OFDM was first introduced to WLAN with IEEE 802.11a, the subcarrier spacing was called, well, the subcarrier spacing, since there was only one. With the introduction of 802.11n, the subcarrier spacing did not change but the number of subcarriers increased for better spectral efficiency and a 40 MHz channel width option and 256-QAM modulation was added. The coexistence of the two PHY types necessitated the need to designate 802.11a as non-HT fields. And that technique and terminology is still with us today.

Subcarrier frequency spacing for the EHT modulated fields is defined in equation (2). This is the subcarrier spacing for EHT data symbols.

$$\Delta_{F,EHT} = \frac{\Delta_{F,Pre-EHT}}{4} = 78.125kHz \tag{2}$$

An OFDM transmitter applies modulation to frequency domain subcarriers and then performs an inverse discrete Fourier transform (IDFT) to create a time domain signal that provides input to an analog-to-digital converter (ADC). IDFT/DFT (discrete Fourier transform) period for the pre-EHT modulated field is defined in equation (3). It takes 3.2 μs time duration to create an 802.11a symbol DFT. An OFDM receiver accepts the time domain samples created in the transmitter and performs a DFT that converts the time domain samples into the frequency domain modulated symbols.

$$T_{DFT,Pre-EHT} = \frac{1}{\Delta_{F,Pre-EHT}} = 3.2\mu s \tag{3}$$

IDFT/DFT period for the EHT Data field is defined in equation (4). In calculating the PHY data rate in Mbps of EHT symbols, the denominator will be 12.8 μs plus the guard interval (GI). Using μs units for the symbol time conveniently results in data rate units of Mbps.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

$$T_{DFT,EHT} = \frac{1}{\Delta_{F,EHT}} = 4 \cdot T_{DFT,Pre-EHT} = 12.8\mu s \quad (4)$$

$$T_{GI,Pre-EHT} = 0.8\mu s \quad (5)$$

Guard interval duration for the pre-EHT modulated fields excluding the legacy long training field (L-LTF) is defined in equation (5).

$$T_{GI,L-LTF} = 1.6\mu s \quad (6)$$

The guard interval of the L-LTF field is defined in equation (6).

The base guard interval duration for the Data field is defined in equation (7). For indoor residential environments, the base guard interval is plenty long enough to protect against inter-symbol interference due to multipath reflections.

$$T_{GI1,Data} = 0.8\mu s \quad (7)$$

The double guard interval duration for the Data field is defined in equation (8).

$$T_{GI2,Data} = 1.6\mu s \quad (8)$$

The double guard interval duration for the Data field used to protect against reflections from objects 800 feet from the AP is defined in equation (8). The double guard interval duration is useful for the long echoes of outdoor applications with an elevated AP.

The quadruple guard interval duration for the Data field is defined in equation (9). The quadruple guard interval is also useful to protect against long echoes for outdoor access points mounted high above the ground to increase coverage.

$$T_{GI4,Data} = 3.2\mu s \quad (9)$$

The quadruple guard interval duration for the Data field used to protect against reflection from objects 1600 feet away from the AP is defined in equation (9).

$$T_{GI,Data} = T_{GI1,Data} \ or \ T_{GI2,Data} \ or \ T_{GI4,Data} \quad (10)$$

The guard interval of the Data field defined in equation (10) is either the base, double, or quadruple guard interval duration. For indoor applications in residential homes and apartments, the base guard interval provides sufficient protection from multipath reflections. The longer guard intervals provide additional redundant information so that rate adaptation algorithms will sometimes resort to the double and quadruple guard intervals in special cases that need help with demodulation.

$$T_{GI,EHT-LTF} = T_{GI,Data} \quad (11)$$

The guard interval duration for the EHT-LTF is the same as the guard interval duration of the Data field as shown in equation (11).

The OFDM symbol duration with base guard interval is defined in equation (12).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

$$T_{SYM1} = T_{DFT,EHT} + T_{GI1,Data} = 1.0625 \cdot T_{DFT,EHT} = 13.6\mu s \quad (12)$$

The OFDM symbol duration for the base interval is the sum of the EHT DFT duration defined in equation (4) and the base guard interval defined in equation (7).

The subcarrier frequency spacing for pre-EHT modulated fields is 312.5 kHz, making the DFT period 3.2 µs as shown in equation (13). The subcarrier frequency spacing for EHT modulated fields is 78.125 kHz, one fourth that of pre-EHT signals, making the DFT period 12.8 µs, four times longer than the pre-EHT period. EHT subcarrier spacing and DFT period is the same as high efficiency mode (HE) introduced in IEEE 802.11ax. The reasons for the longer DFT period and the closer subcarrier spacing used for EHT and HE are explained in reference (Urban, The Importance Of Wi-Fi 6 Technology For Delivery Of GBPS Internet Service, 2019).

$$\Delta_F = \frac{1}{T_{DFT}} \quad (13)$$

The signal that a Wi-Fi 7 receiver demodulates is called a physical layer protocol data unit (PPDU). The physical layer (PHY) preamble of the PPDU is processed by the receiver to aid in detection and demodulation of the PPDU.

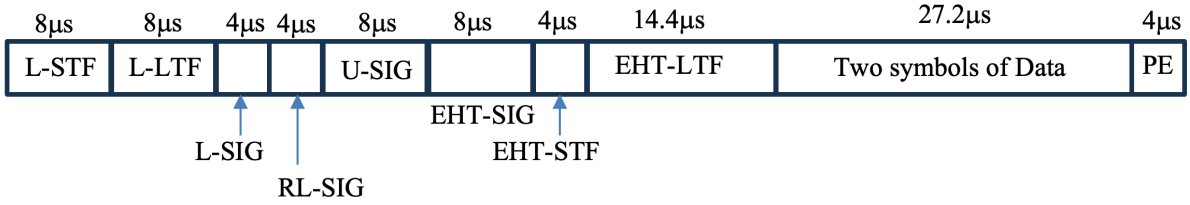


Figure 1 – EHT MU PPDU fields and typical durations

The maximum PHY rate of a 2x2 320 MHz Wi-Fi 7 wireless adapter is 5764 Mbps. The PHY rate is the number of data bits carried in one OFDM data symbol divided by the symbol period. Data symbols are only part of a PPDU field. A preamble is needed in order to detect the signal, lock onto the frequency, estimate the channel, determine the signal parameters, and adjust for amplitude and phase in order to demodulate the data symbol.

The EHT PPDU fields are listed in Table 1 along with a description and duration in μs.

Table 1 – EHT PPDU fields with typical durations

Field	Description	Duration μs
L-STF	Non-HT short training field	8
L-LTF	Non-HT long training field	8
L-SIG	Non-HT SIGNAL field	4
RL-SIG	Repeated Non-HT SIGNAL field	4
U-SIG	Universal SIGNAL field	8
EHT-SIG	EHT SIGNAL field	4
EHT-STF	EHT short training field	8

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Field	Description	Duration μs
EHT-LTF	EHT long training field	14.4
Data	The data carrying the PSDU(s)	13.6 per symbol
PE	Packet Extension field	0

An EHT transmission begins with a legacy short training field (L-STF). A legacy symbol is also referred to as a non-HT symbol. HT mode stands for high throughput mode, which was introduced with 802.11n, thus a non-HT symbol uses the PHY mode of 802.11a. 802.11a was the first Wi-Fi standard to use OFDM with a 312.5 kHz subcarrier spacing and a 6 Mbps data rate symbol using binary phase shift keying (BPSK) and ½ binary convolutional coding (BCC) code rate in a 20 MHz channel width. The L-STF is followed by a legacy long training field (L-LTF) the legacy signal field (L-SIG). After the 20 μs legacy part of the preamble, a repeated L-SIG (RL-SIG) and universal SIG (U-SIG), precedes the EHT preamble fields. EHT-SIG, EHT-STF, EHT-LTF precede the Data symbols followed by a packet extension (PE) field.

4. Channel Width: EHT ups the ante to 320 MHz

Wi-Fi 7, based upon the IEEE 802.11be standard, increases the maximum channel width from 160 MHz to 320 MHz of spectrum. A fast Fourier transform (FFT) is a mathematically efficient algorithm to calculate a discrete Fourier transform (DFT). The DFT of a time domain signal sampled every 3.125 ns will demodulate tones covering a 320 MHz channel width as shown in equation (14).

$$B = \frac{1}{T_{sampling}} = \frac{1.0}{3.125E-9} = 320E6 \tag{14}$$

The guard interval protecting two symbols from interfering with each other due to multipath reflections for EHT symbols can be one of three selections 0.8, 1.6, or 3.2 μs. The 800 ns guard interval protects against reflections that are 400 feet from the access point, AP, since light travels one foot in one ns. The 800 ns guard interval is sufficient for indoor applications while the 1.6 and 3.2 μs guard intervals are mostly applicable for outdoor applications with a km or greater coverage range.

The DFT period and the guard interval form a symbol period. The DFT period is selected to be 16 times the guard interval of 800 ns to increase efficiency, putting the DFT period at 12.8 μs as shown in equation (15). The 12.8 μs DFT period was introduced in 802.11ax and continued in 802.11be, this is what puts the efficiency in high efficiency (HE) mode.

$$T_g = 800 \cdot 10^{-9} \qquad T_{DFT} = 16 \cdot Tg = 12.8 \cdot 10^{-6} \tag{15}$$

The DFT size is the number of time domain samples and the number of frequency domain tones that make up the useful part of an OFDM symbol. The DFT size can be determined by dividing the DFT period by the time domain sampling rate or equivalently dividing the channel width by the subcarrier spacing as shown in equation (16).

$$N_{DFT} = \frac{T_{DFT}}{T_{sampling}} = \frac{12.8 \cdot 10^{-6}}{3.125 \cdot 10^{-9}} = 4096 = \frac{\mathrm{B}}{\Delta_{\mathrm{F}}} = \frac{320\mathrm{MHz}}{78.125\mathrm{kHz}} \tag{16}$$

The DFT size of an EHT 320 MHz channel width OFDM symbol is 4096. A transmitter creating a 320 MHz OFDM EHT symbol will perform an IDFT on 4096 frequency domain tones to produce the 4096

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

time domain samples of the DFT period portion of an OFDM symbol. An OFDM symbol is created by appending a cyclic prefix to the DFT period to provide a guard time between OFDM symbols to prevent inter-symbol interference due to multipath reflections. The OFDM EHT symbol period is 13.6 µs as calculated in equation (17).

$$T_{SYM} = T_{GI} + T_{DFT} = 0.8 + 12.8 = 13.6\mu s \quad (17)$$

For the purposes of understanding data rates and throughput of Wi-Fi 7 signals as it relates to customer experience, coverage, range, device, and application requirements, the key takeaway of an understanding of the DFT used in forming an OFDM symbol is the number of data subcarriers in a symbol and the symbol duration. These are used in the calculation of PHY rate, which is one of the important parameters impacting throughput.

Table 2 – Subcarrier allocation of 320 MHz EHT modulated fields

parameter	CBW320 tones	Description
N_{DFT}	4096	DFT subcarriers or tones
$N_{SD,total}$	3920	Total number of data subcarriers
N_{SP}	64	Number of pilot subcarriers
N_{SR}	2036	Highest data subcarrier index
N_{DC}	23	Direct current null subcarriers (DC)
$N_{Guard,Left}$	12	Number of low frequency guard subcarriers
$N_{Guard,Right}$	11	Number of high frequency guard subcarriers

The 4096 subcarriers that make up an EHT 320 MHz OFDM symbol consist of three types: null, pilot, and data subcarriers. An EHT OFDM 320 MHz symbol has 3,920 data subcarriers as shown in Table 2.

An EHT signal with 320 MHz channel bandwidth is composed of four identical blocks of 80 MHz of spectrum. An 80 MHz block of spectrum contains 16 pilot subcarriers. The receiver needs pilot subcarriers to demodulate the signal. The frequency response of the received signal can be estimated and corrected by comparing the pilot subcarriers at different frequencies within a symbol time. The center frequency of the received signal can be locked by determining the phase change between two pilot subcarriers at different symbol times. The symbol time is known so the frequency offset can be calculated as the change in phase divided by the change in time.

Null subcarriers have zero energy and are used to protect the receiver from transmit center frequency leakage, receiver DC offset, and interference from neighboring adjacent signals. Null subcarriers are added to the left and right guard band and baseband DC. Baseband DC corresponds to the center frequency of a radiated signal on a particular 6 GHz band channel. The left guard band of an 80 MHz channel is protected by 12 null subcarriers while the right guard band is protected by 11 null subcarriers, and the baseband DC offset is protected by 5 null subcarriers.

Range is improved with increased channel width. A common misperception is that narrower channel width provides better range for WLAN systems. As with many myths, a truth taken out of context is the root of the misunderstanding. It is certainly true that in the 5 GHz band WLAN, as one walks further and further from the access point, the rate adaptation algorithm will adjust the channel width from 160 to 80 to 40 and finally to 20 MHz as the receive level decreases. However, the coverage range is only improved at reduced channel width when the only factor considered is connectivity without considering throughput. Wi-Fi is a high-speed data transfer system, so high speed must be an important requirement. When one looks at the range of coverage for a given required amount of throughput, the picture looks quite different.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The plot in Figure 2 illustrates the range advantage of increased channel width. The total conducted transmit level is the same at each channel width of 20, 40, 80, 160, and 320 MHz of spectrum. This is often the case in the 2.4 and 5 GHz Wi-Fi bands due to convention and regulatory limits. The Federal Communications Commission (FCC) allows 1 watt of conducted power into the antenna in the 2.4 GHz industrial, scientific, medical (ISM)band, and unlicensed national information infrastructure (U-NII) 1 and 3 bands with some restrictions in power spectral density, bandwidth, and out-of-band emissions.

Since the transmit level is the same for a 20, 40, 80, 160, and 320 MHz channel width, the signal to noise ratio improves by 3 dB when the channel width is cut in half since the receive level remains the same, while the noise level decreases by 3 dB with the halving of the equivalent noise bandwidth. This improves coverage and is reflected in the plot at the very low PHY rates.

However, as the PHY rate is increased, the range advantage of wider channel widths becomes more apparent. The range advantage of wider channel width increases as the PHY rate increases, seen as the widening gap for different channel widths moving to the right in the plot as PHY rate increases. The 3 dB advantage in the lower noise floor due to halving the channel width becomes small compared to the stronger signal level needed for higher modulation level to double the spectral efficiency. Increased modulation rate is needed for small channel width to make up for the greater number of data carrying subcarriers of wider channel width signals.

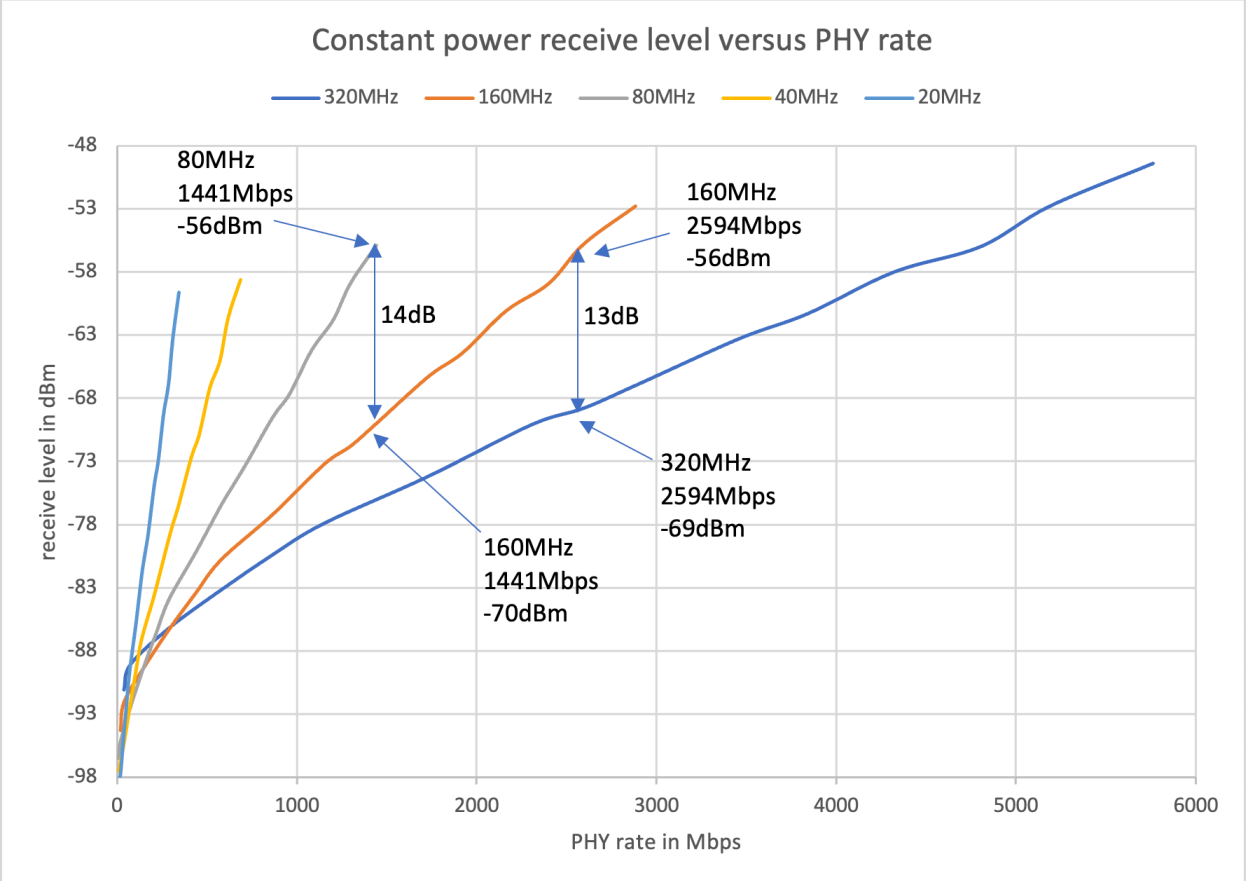


Figure 2 – Range is improved with increased channel width for a required throughput

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

When power spectral density is restricted, as is the case with the 6 GHz band for indoor operation, the advantage of wider channel width for range improvement becomes more acute as shown in Figure 3. A plot of PHY rate versus distance, comparing 320 MHz and 160 MHz using the 802.11 path loss model for indoor residential, is shown. The cell edge is neither improved nor degraded by 320 MHz channel width compared to 160 MHz as seen in the converged lines beyond 35 meters. At any given PHY rate above about 200 Mbps, the distance is greater for the 320 MHz channel width than the 160 MHz channel width. This is the range advantage of 320 MHz versus 160 MHz or lower channel width. As a result, the rate adaptation algorithm will always use the highest available channel width when range is the limiting factor. There can still be reasons for using narrower resource units, such as channel availability and OFDMA, but the link budget will never be a reason to reduce the channel width. The higher the required PHY rate, the greater the distance advantage of 320 MHz channel width over 160 MHz channel width.

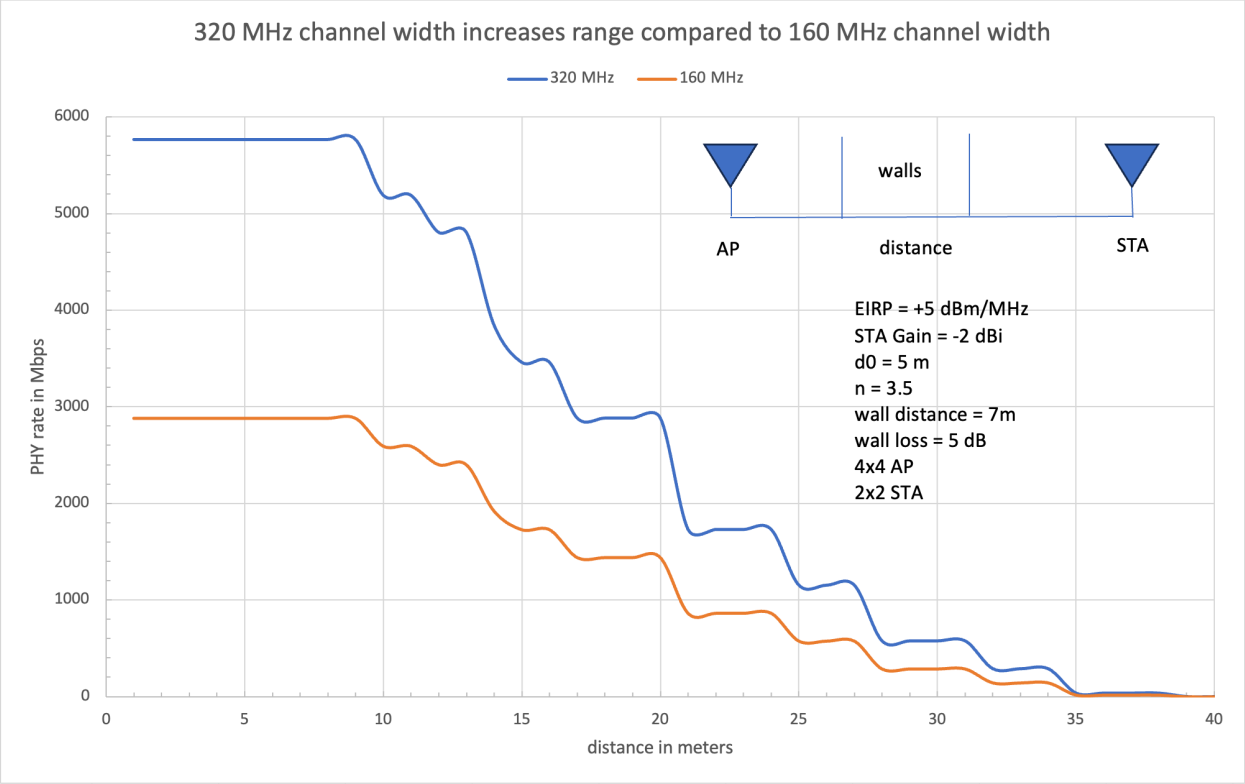


Figure 3 – Improvement in range for 320 MHz with constant power spectral density

5. EHT Preamble

Before an OFDM data symbol can be transmitted or received, a preamble must precede the OFDM data symbol for the receiver to detect the presence of the signal, lock into the frequency of the signal, estimate the channel, determine the characteristics of the signal, and use automatic gain control to adjust for signal level.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.1. L-STF to detect the signal

An EHT transmission begins with a legacy short training field, L-STF. A legacy symbol is also referred to as a non-HT symbol. HT mode stands for high throughput mode, which was introduced with 802.11n, thus a non-HT symbol uses the PHY mode of 802.11a. 802.11a was the first Wi-Fi standard to use OFDM with a 312.5 kHz subcarrier spacing and a 6 Mbps data rate symbol using BPSK and ½ BCC code rate in a 20 MHz channel width.

A DFT with a time domain input sampling rate of 3.125 ns and 4096-time samples will output 4096 frequency domain tones, having a channel width of 320 MHz. The DFT period is 12.8 μs. The L-STF consists of 172 tones over 320 MHz channel width. The tones can be created with a 256-point IDFT using a 3.125 ns sampling interval since 172 subcarriers are greater than 128 but less than 256. The L-STF symbol time is 256 times 3.125 ns, which equals 800 ns or 0.8 μs. The 0.8 μs duration of the L-STF field consists of 10 repeated symbols without any guard interval. The repetition of ten symbols in the L-STF is important since it allows a self-correlation function to detect the reception of an 802.11 signal. Wi-Fi is an asynchronous time division duplex (TDD) system so a receiver never knows a priori when a signal will arrive that needs to be demodulated. A self-correlation function that detects the repeated 10 symbols of the L-STF is critical in detecting the presence of an incoming signal.

5.2. L-LTF

Following the L-STF in the EHT preamble is the L-LTF.

5.3. L-SIG

Following the L-LTF in the EHT preamble is the L-SIG.

5.4. RL-SIG

Following the L-SIG in the EHT preamble is the RL-SIG.

5.5. U-SIG Introducing version independent fields

Following the RL-SIG in the EHT preamble is the U-SIG. The U-SIG is composed of two 4 µs symbols. The duration of the U-SIG is 8 µs. U-SIG introduces version independent fields designed to improve compatibility with future versions of Wi-Fi beyond Wi-Fi 7. Wi-Fi 8 is already in the works.

5.6. EHT-SIG

Following the U-SIG in the EHT preamble is the EHT-SIG.

5.7. EHT-STF

Following the EHT-SIG in the EHT preamble is the EHT-STF.

5.8. EHT-LTF for receiver MIMO channel estimation

The EHT-LTF helps the receiver with multiple input multiple output (MIMO) antennas channel estimation. The duration of the EHT-LTF is not always the same, $N_{EHT-LTF}$ is the number of EHT-LTF symbols in the EHT preamble. Most client wireless adapters are 2x2 so that most often two spatial

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

streams are used, $N_{SS} = 2$. When $N_{SS} = 2$ the initial number of EHT-LTF symbols is two, $N_{EHT-LTF} = 2$.

The EHT-LTF duration of each OFDM symbol without a guard interval can be set to 3.2, 6.4, or 12.8 μs.

Table 3 – EHT-LTF symbol duration options

Parameter	Value	Description
$T_{EHT-LTF-1X}$	3.2 μs	Duration of each 1xEHT-LTF OFDM symbol without GI
$T_{EHT-LTF-2X}$	6.4 μs	Duration of each 2xEHT-LTF OFDM symbol without GI
$T_{EHT-LTF-4X}$	12.8 μs	Duration of each 4xEHT-LTF OFDM symbol without GI

2xEHT-LTF is the most common setting, so this will be used in calculations for this paper. Using a T_{GI} of 0.8 μs and two EHT-LTF symbols of 6.4 μs each yields an EHT-LTF symbol duration of 7.2 μs as shown in equation (18) .

$$T_{EHT-LTF-SYM} = T_{EHT-LTF} + T_{GI,EHT-LTF} = 0.8 + 6.4 = 7.2\mu s \quad (18)$$

For two spatial streams, a most common situation, the duration of the EHT-LTF field will be 14.4 μs for a setting of EHT-LTFx2.

5.9. PE packet extension field

Following the EHT preamble and the Data field in the EHT PPDU is the packet extension field (PE). The duration of the packet extension field denoted T_{PE} in the IEEE 802.11be standard (IEEE 802.11be D3.0, 2023) can be 0, 4, 8, 12, 16, or 20 μs depending on the actual packet extension duration used.

6. Let's talk about spectrum.

The rules and regulations for Unlicensed National Information Infrastructure (U-NII) in the United States are defined in part 15.407 (FCC). Table 4 – USA FCC Unlicensed National Information Infrastructure Frequency in MHz lists the number of the band, the start and stop frequency in MHz and the bandwidth in MHz of each band. U-NII 2B is listed in the table but usage is not allowed for WLAN application.

Table 4 – USA FCC Unlicensed National Information Infrastructure Frequency in MHz

Band	Start	Stop	Bandwidth
U-NII 1	5150	5250	100
U-NII 2A	5250	5350	100
U-NII 2B	5350	5470	120
U-NII 2C	5470	5725	255
U-NII 3	5725	5850	125
U-NII 4	5850	5925	75
U-NII 5	5925	6425	500
U-NII 6	6425	6525	100
U-NII 7	6525	6875	350
U-NII 8	6875	7125	250

In the United States the FCC has allocated 1,200 MHz of spectrum for unlicensed use in U-NII bands 5,6,7,8. While one can always look up the channels and frequencies, it is sometimes handy to have a mnemonic trick that can aid in developing a better appreciation of the capacity of the allocation. The

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

spectrum from 6 GHz to 7 GHz spans 1,000 MHz so that 200 MHz of addition spectrum is required outside of the 6 GHz band. 75 MHz is allocated below 6 GHz and 125 MHz is allocated above 7 GHz. That puts the range of spectrum from 5925-7125 MHz.

Thus, 6 GHz band spectrum in the United States for WLAN begins at 5925 MHz and ends at 7125 MHz, covering 1200 MHz total bandwidth. U-NII 5 spans 500 MHz of spectrum beginning 75 MHz below 6 GHz. The lowest frequency is 5925 MHz, and the highest frequency is 6425 MHz. The first 20 MHz is not used, being too close to U-NII 4. The first 20 MHz channel thus begins at 5945 MHz and ends at 5965 MHz with a center frequency of 5955 MHz. This 20 MHz channel is designated as channel 1, leading to the formula to calculate the center frequency given the channel number as shown in equation (19).

$$f_c = 5950 + 5 \cdot n_{ch} \tag{19}$$

The lowest frequency 20 MHz block of spectrum is centered at 5955 MHz, designated as channel 1. The highest frequency 20 MHz block of spectrum is centered at 7115 MHz, designated as channel 233. With 233 possible primary channels of 20 MHz channel width to choose from, preferred scanning channels are defined to reduce the time and resources it takes to find a signal to connect. If the process of sending a probe request leading to the connection on a primary channel takes 20 ms and there are 233 primary channels to scan, then it will take 4.66 seconds to scan all the channels. 15 primary channels are defined as preferred scanning channels (PSC) spaced by 16 channels. The list of PSC in the 6 GHz band is 5,21,37,53,69,85,101,117,133,149,165,181,197,213,229.

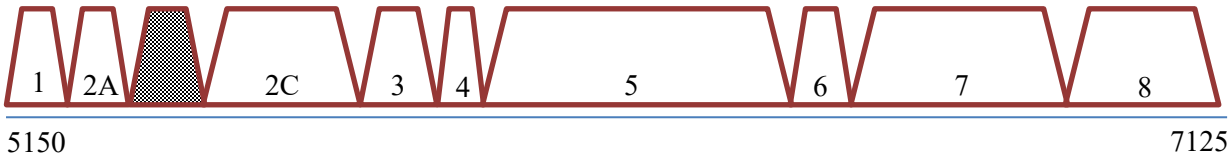


Figure 4 – U-NII band

The U-NII band begins, and 5150 MHz ends at 7125 MHz. The 2.4 GHz band has three 20 MHz channels, the 5 GHz band has three 160 MHz channels, and the 6 GHz band has three 320 MHz channels.

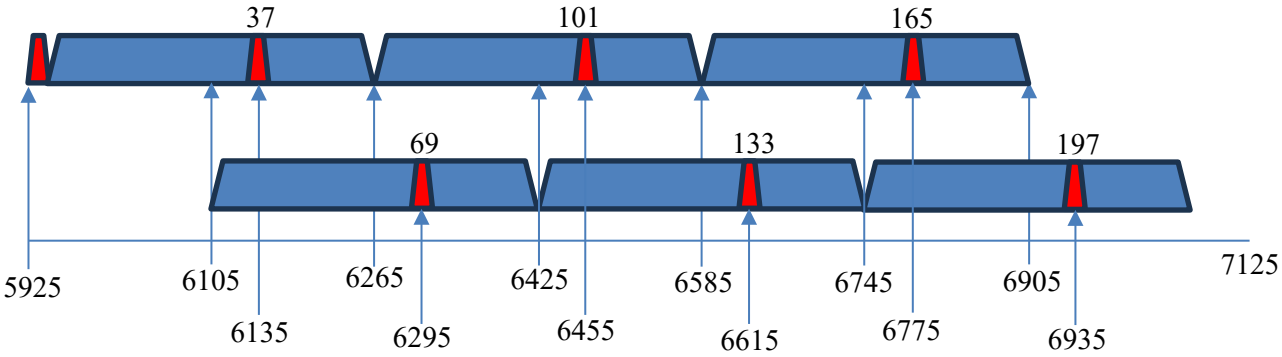


Figure 5 – 6 GHz band preferred scanning channels and center frequencies in MHz

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The 6 GHz band has three 320 MHz channel width signals that can be placed beginning at the lowest frequency or ending at the highest frequency as shown in Figure 5 and listed in Table 5 – 6 GHz band 320 MHz channels.

Table 5 – 6 GHz band 320 MHz channels

Channel	f_c MHz	Center channel	PSC Channel	PSC f_c MHz
6g37/320-1	6105	31	37	6135
6g101/320-1	6425	95	101	6455
6g165/320-1	6745	159	165	6775
6g69/320-2	6265	63	69	6295
6g133/320-2	6585	127	133	6615
6g197/320-2	6905	191	197	6935

The list of 320 MHz channels shown in Table 5 – 6 GHz band 320 MHz channels details the channel naming scheme, the center frequency in MHz of the 320 MHz wide channel width, the number of the primary scanning channel, and the center frequency in MHz of the 20 MHz wide primary scanning channel. Since channel 37 is also a 5 GHz band channel number, 6g37 is used to indicate that this refers to channel 37 in the 6 GHz band and not the 5 GHz band. The 6g37/320 notation indicates the primary scanning channel is 37 in the 6 GHz band with a channel width of 320 MHz where the three 320 MHz channels begin at the lower end of the 6 GHz band.

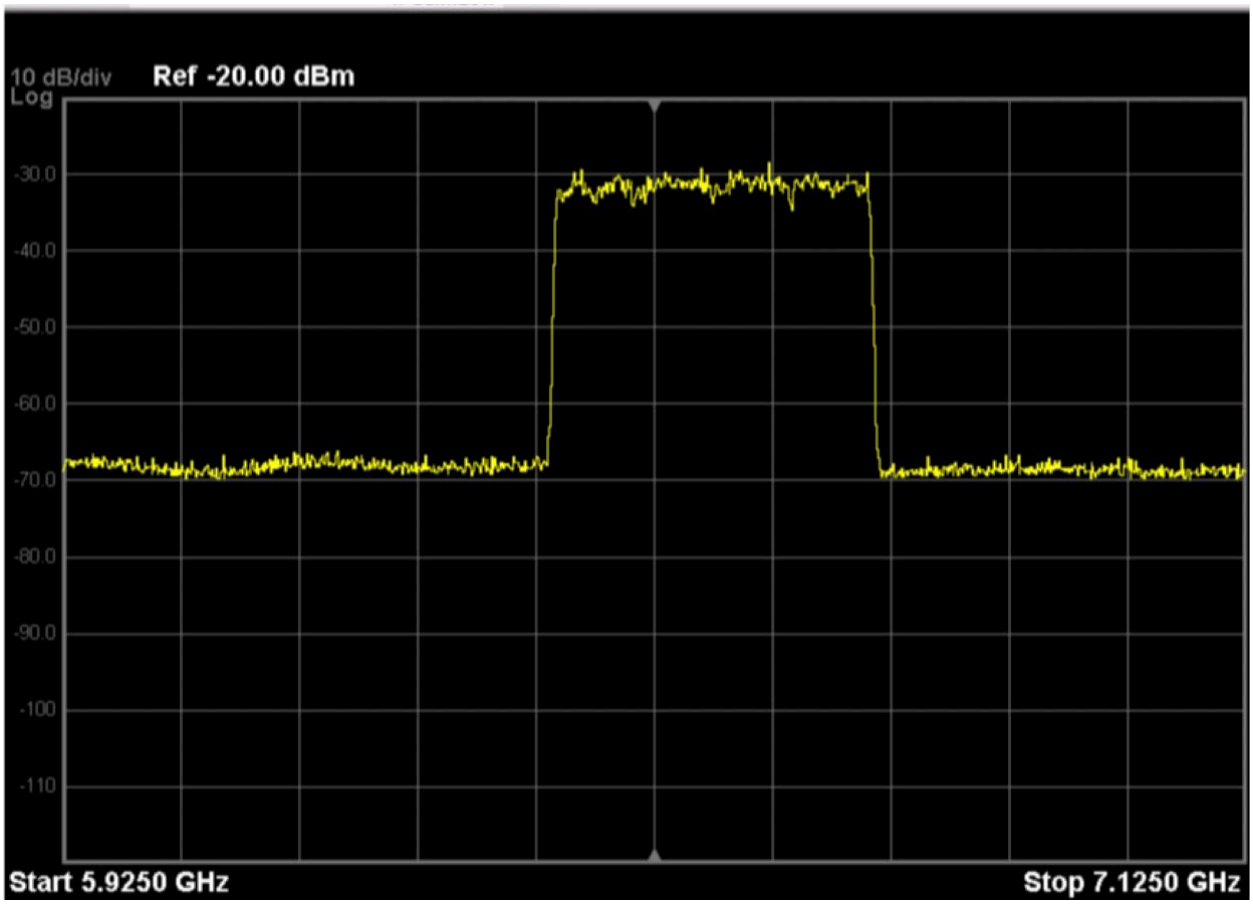


Figure 6 – Spectrum analyzer plot of 6g133/320-2 over 6 GHz band

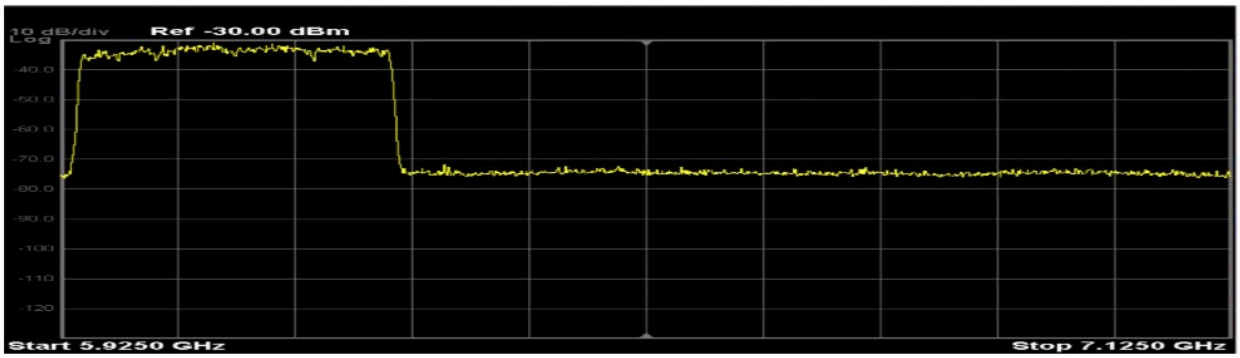
© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

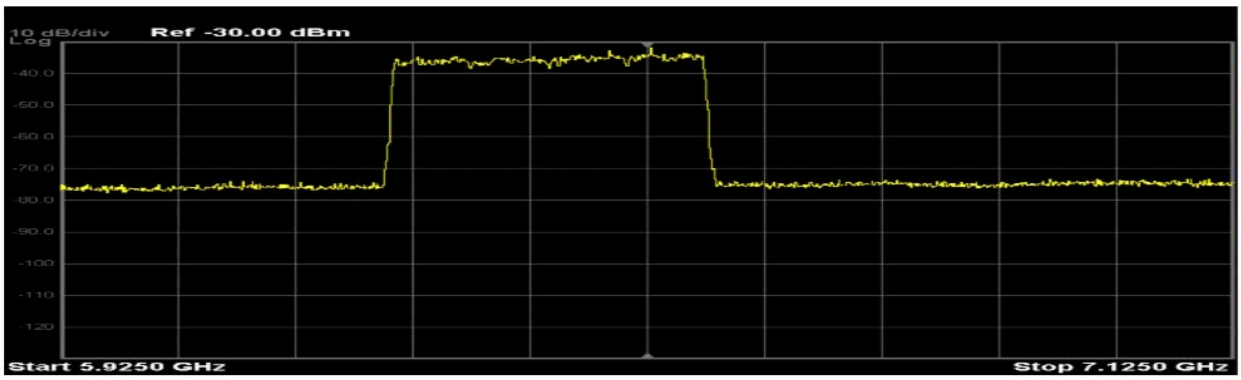
A spectrum analyzer sweep is shown in Figure 6 with peak hold to reveal the full 320 MHz channel width of 6g133/320-2 swept from the 5925 MHz lowest part of the 6 GHz band to the 7125 MHz highest part of the 6 GHz band. The frequency sweep over the whole 6 GHz band spectrum shows the relative position of 6g133/320-2 within the full 6 GHz band.

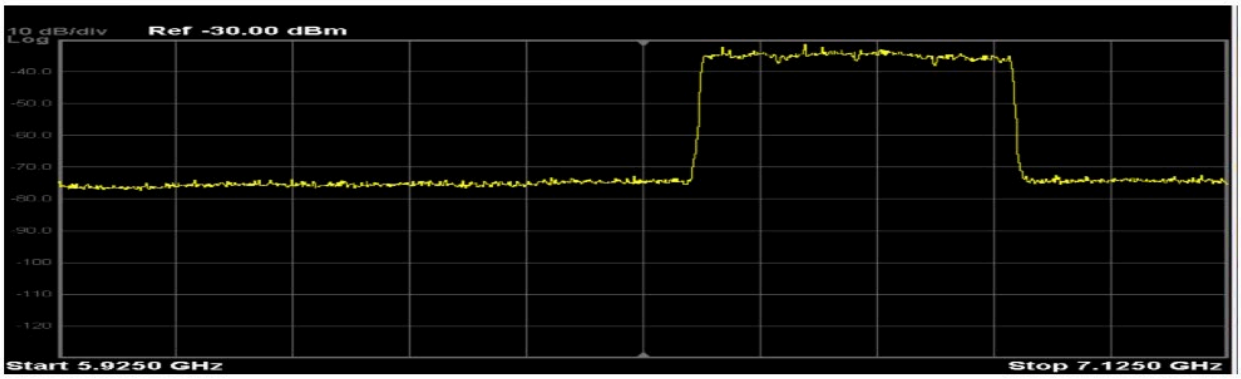
Figure 7 – Spectrum analyzer plot of 6g37/320-1, 6g101/320-1, 6g165/320-1 shows a sweep of all three of the 320 MHz channel width signals set towards the lower end of the 6 GHz band.

Figure 8 – Spectrum analyzer plot of 6g69/320-2, 6g133/320-2, 6g197/320-2 shows a sweep of all three of the 320 MHz channel width signals set at the upper end of the 6 GHz band.

The peak hold spectrum analyzer measurements of all six 320 MHz channel width signals in the 6 GHz band helps illustrate the channel selections that overlap in spectrum and those that do not. When trying to keep channels from overlapping within a cluster of access points stick with either the "-1" scheme or the "-2" scheme. Don't mix the dashes.

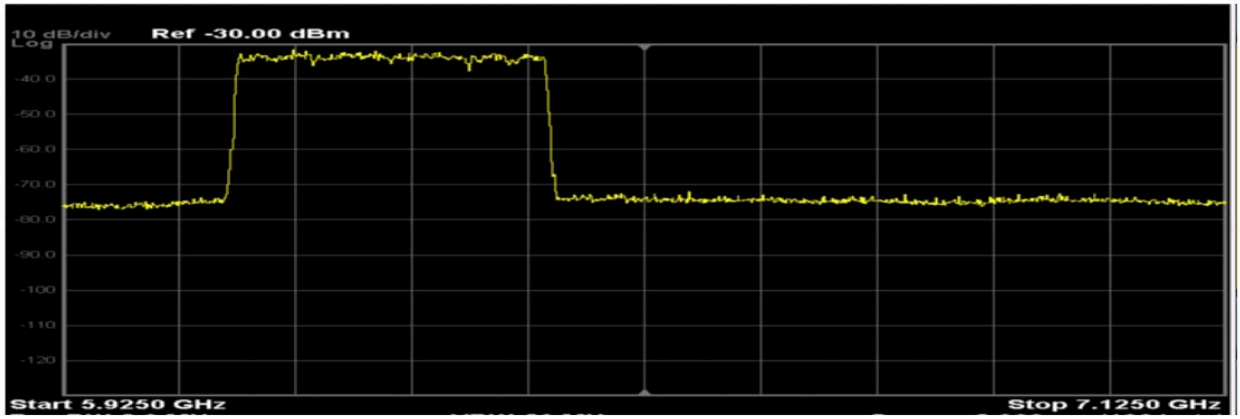


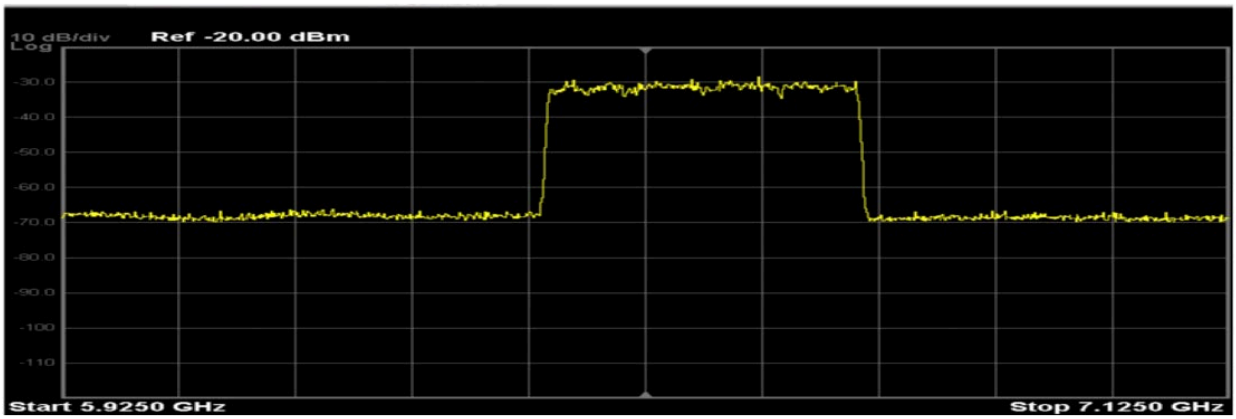




© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 7 – Spectrum analyzer plot of 6g37/320-1, 6g101/320-1, 6g165/320-1





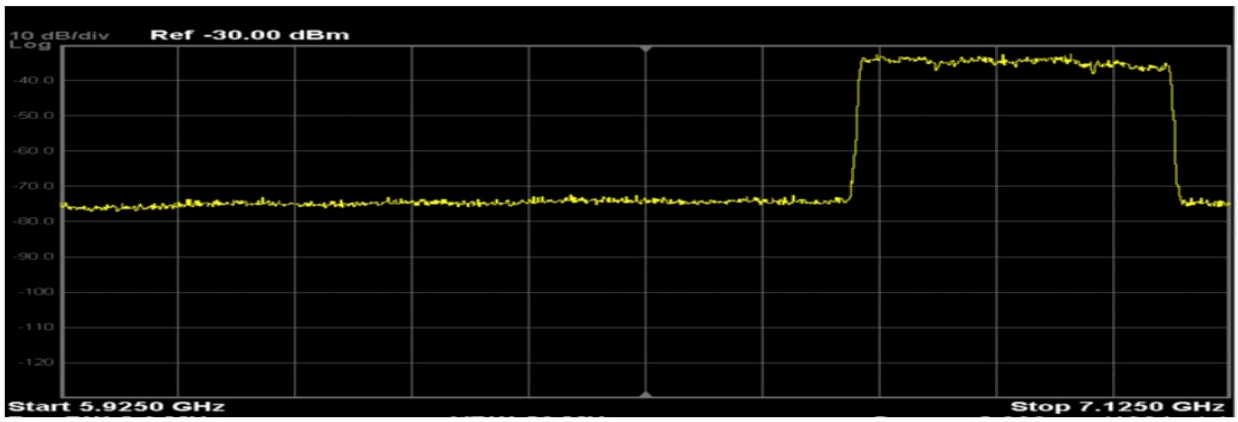


Figure 8 – Spectrum analyzer plot of 6g69/320-2, 6g133/320-2, 6g197/320-2

Figure 9 – Spectrum analyzer plot to 8 GHz with EHT 6g37/320-1 max download shows a spectrum analyzer sweep with a span of 8.39 GHz and a center frequency of 4.205 GHz while a TCP download is run on channel 6g37/320-1. It is always important to widen the sweep outside of the channel width and even the U-NII band to make sure that no unwanted emission falls outside of the desired frequency spectrum.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

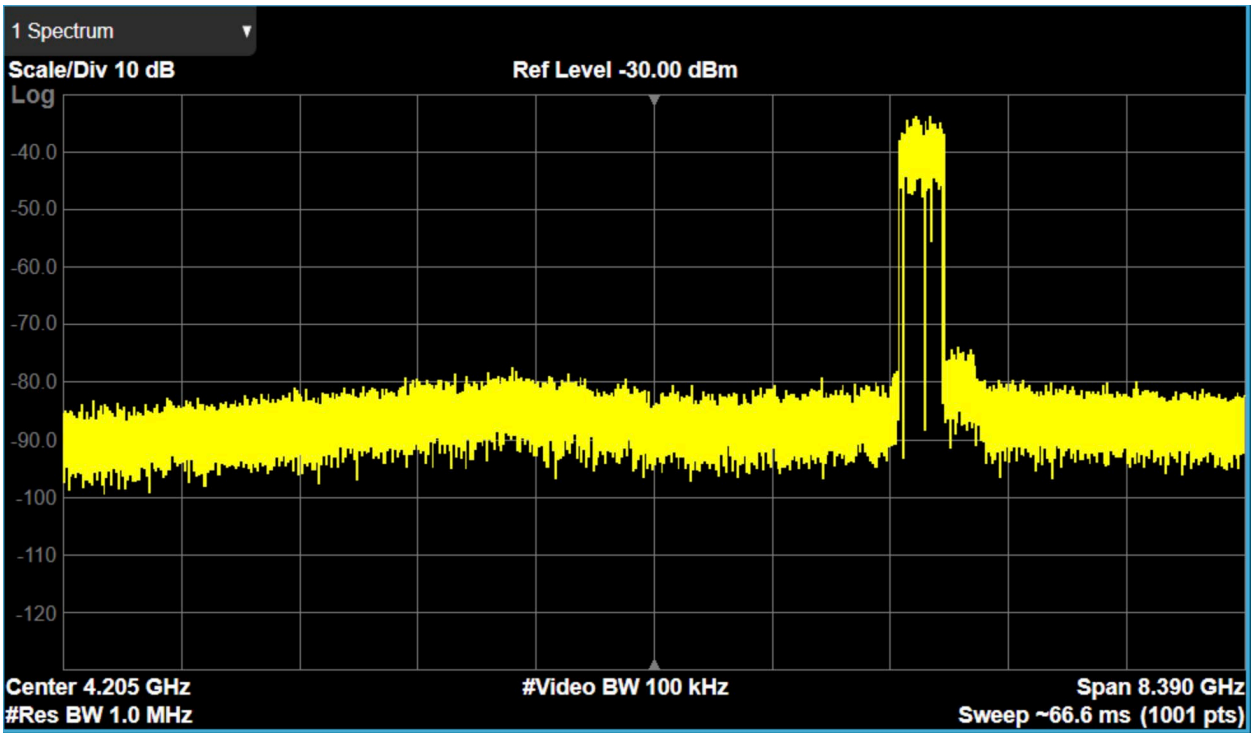


Figure 9 – Spectrum analyzer plot to 8 GHz with EHT 6g37/320-1 max download

7. Modulation and Coding; Introducing 4k-QAM to Wi-Fi

Wi-Fi 7, based upon the IEEE 802.11be standard, adds two modulation and coding (MCS) levels with higher spectral efficiency. MCS12 employs 4096-QAM with ¾ low density parity check coding (LDPC) coding. MCS13 employs 4096-QAM with 5/6 LDPC coding. 4096-QAM has a constellation size of 4096 points, where each point represents 12 bits as calculated in equation (20).

$$2^{12} = 4096 \tag{20}$$

Claude Shannon published a mathematical theory of communications in the Bell System Technical Journal in 1948. A key insight in this seminal paper was that methods of modulation exchange bandwidth for signal to noise ratio (Shannon, 1948). The capacity of a channel of band W perturbed by white thermal noise power N when the average power is limited to P is given by equation (21).

$$C = B \cdot log_2\left(1 + \frac{P+N}{N}\right) \tag{21}$$

Solving for signal to noise ratio (SNR) in dB as a function of bits is shown in equation (22).

$$SNR = 10 \cdot log_{10}\left(\frac{P+N}{N}\right) = 10 \cdot log_{10}\left(2^{\left(\frac{C}{B}\right)} - 1\right) \tag{22}$$

Applying a C/B of 10 bits per second per Hz for 1024-QAM modulation in (22) calculates 30 dB SNR. Applying a C/B of 12 bits per second per Hz for 4096-QAM modulation in (22) calculates 36 dB SNR. These are reasonably close to the measured SNR threshold for MCS11 and MCS13, and the difference of 6 dB higher SNR requirement between 1024-QAM and 4096-QAM is quite close to the IEEE 802.11be

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

receiver sensitivity requirement differences between 1024-QAM and 4096-QAM for the same LDPC code rate and the measured sensitivity level differences. Applying the actual data rate and channel width to Shannon's capacity formula seems to indicate that there is still some room for improvement. Shannon's capacity formula is good to keep in mind in order to appreciate the relationship between adding capacity to channel width compared to adding capacity with modulation in relation to the impact of SNR requirement and receive level requirement.

The PHY data rate, R_d can be calculated using the number of spatial streams, N_{SS}, the number of data subcarriers, N_{SD}, the number of coded bits per OFDM symbol per subcarrier, N_{CBPSS} , the code rate, R, and the symbol period, $T_{SYM} = T_{DFT} + T_{GI}$. The formula is shown in equation (23) and calculated for the maximum PHY rate of a 4x4 320 MHz EHT AP.

$$R_d = \frac{N_{SS} \cdot N_{SD} \cdot N_{CBPSS} \cdot R}{T_{DFT} + T_{GI}} = \frac{4 \cdot 3920 \cdot 12 \cdot 5/6}{12.8 + 0.8} = 11{,}529\ Mbps \qquad (23)$$

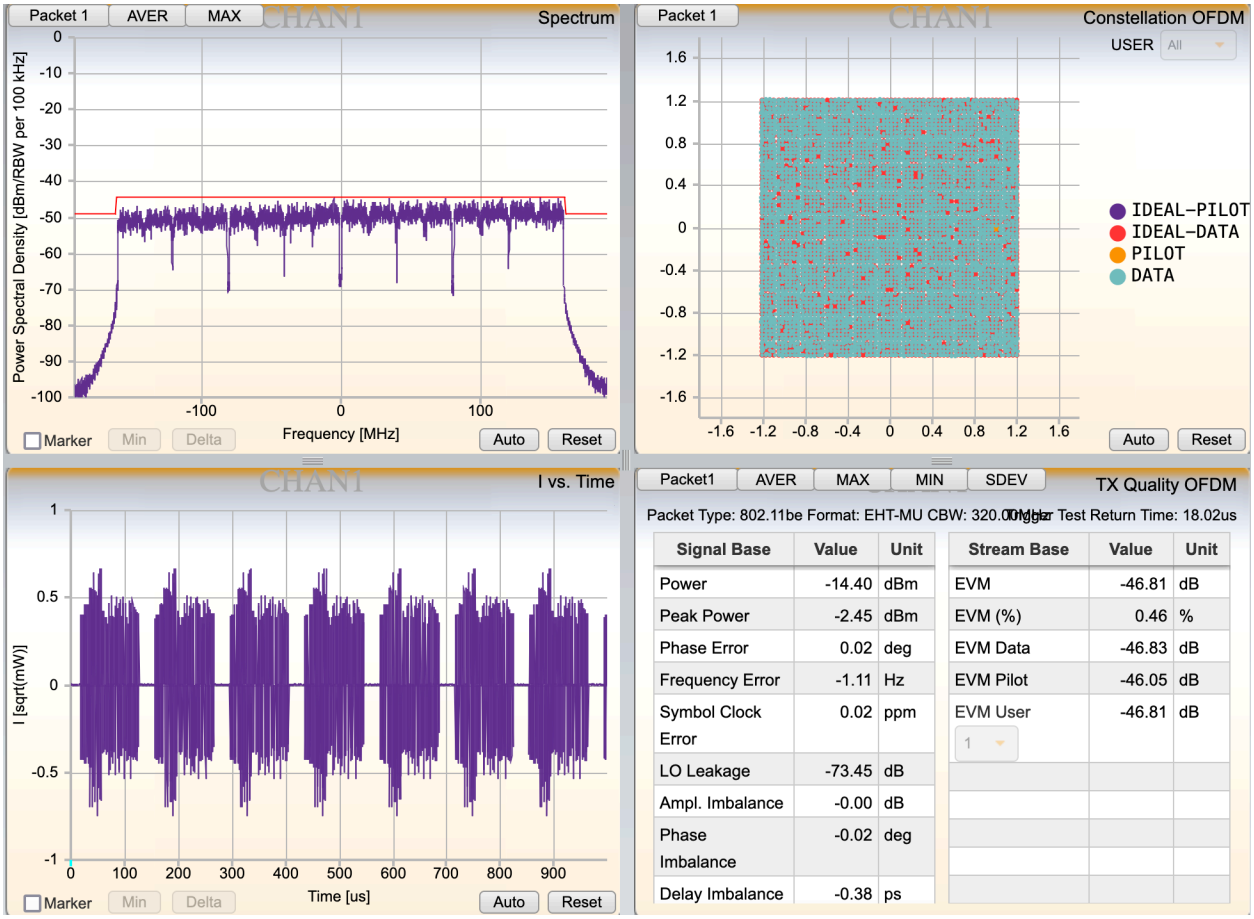


Figure 10 – MCS13 QAM-4096 LDPC 5/6 320 MHz 802.11be Wi-Fi 7 signal

Figure 10 – MCS13 QAM-4096 LDPC 5/6 320 MHz 802.11be Wi-Fi 7 signal show the 320 MHz spectrum of the signal, the 4096 point constellation diagram of MCS16, the power versus time waveform of the signal, and the error vector magnitude (EVM) of -46 dB.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

BPSK modulation assigns 1 bit to each subcarrier with two points in the constellation diagram. The measured BPSK constellation is shown in Figure 11 along with the time waveform power, spectrum, and EVM per subcarrier.

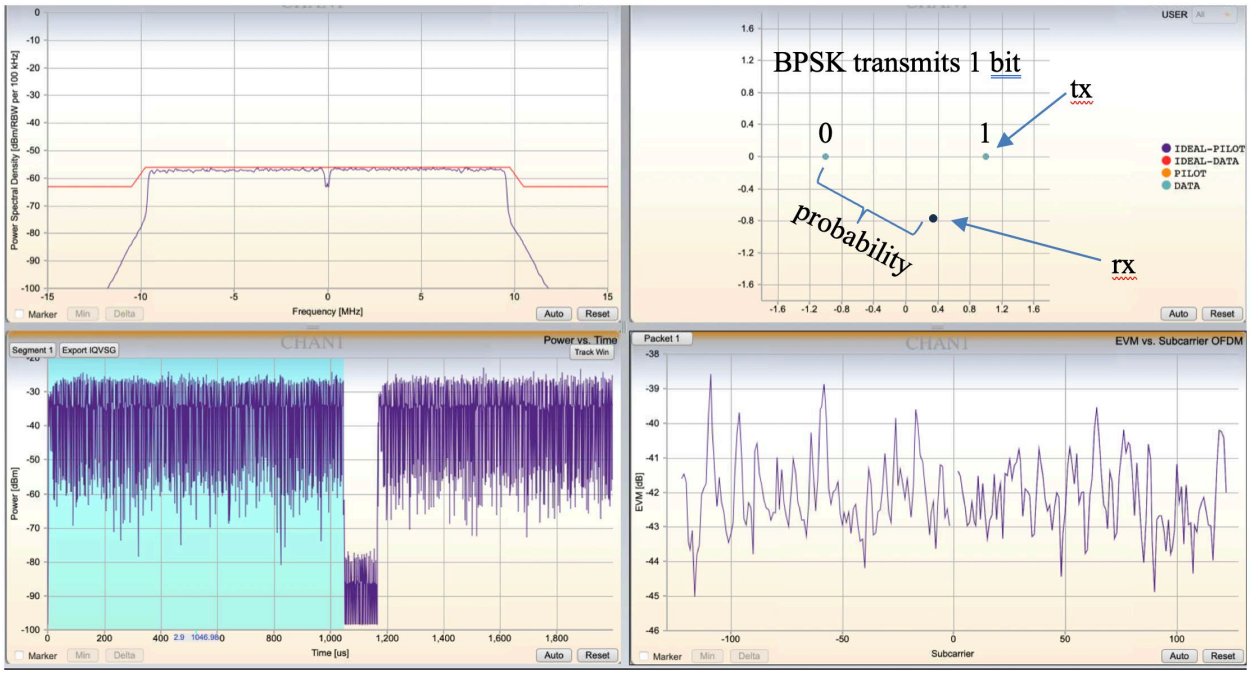


Figure 11 – BPSK signal showing spectrum, constellation, time domain signal, EVM.

The markup in the BPSK constellation diagram illustrates that BPSK carries 1 bit of information, one dot represents a 0 and the other a 1. The point labeled “tx” is the transmitted signal and the added point labelled “rx” represents an example receive point that deviates from the transmitted signal due to noise, interference, and distortion. The BPSK demodulator will send two probabilities to the LDPC decoder for the message passing algorithm. A probability that the transmitted bit was a 1 given the received vector. and a probability that the transmitted bit was a 0 given the received vector is sent to the message passing algorithm of the LDPC decoder based upon the distance from the two possible transmitter vectors and the received vector using a probability distribution model.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

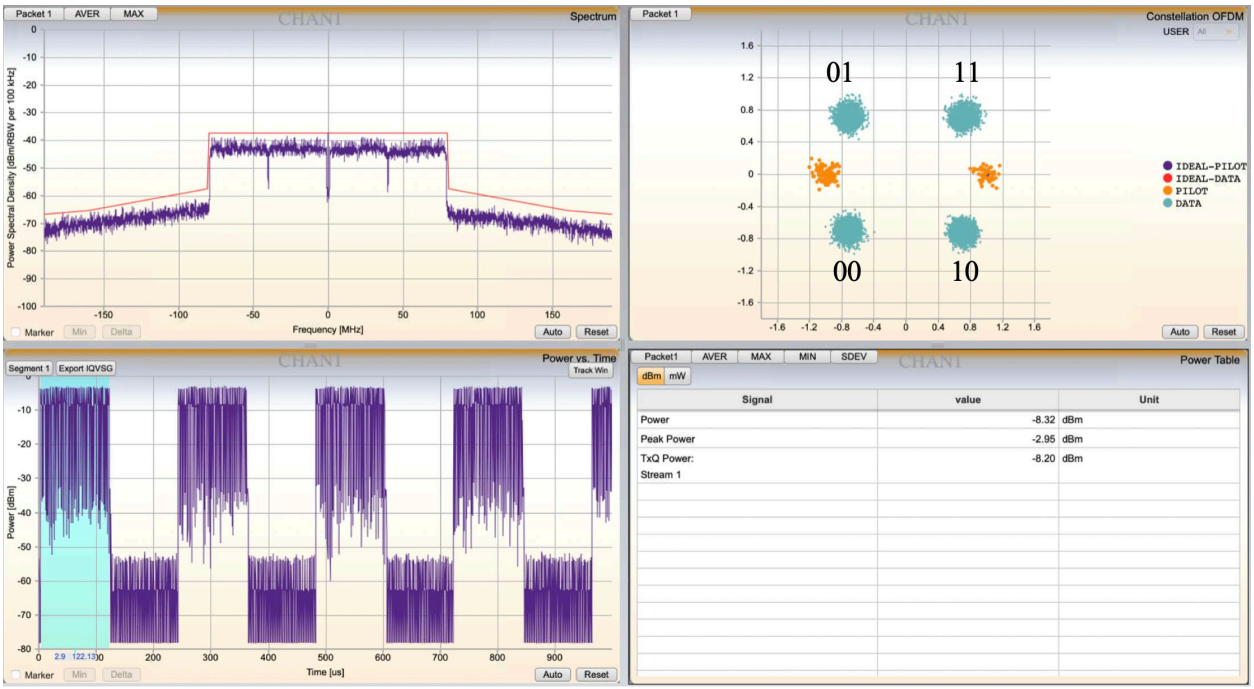


Figure 12 – QPSK modulation maps 2 bits to each subcarrier.

QPSK modulation maps two bits to each subcarrier. QPSK has four constellation points. This is illustrated in Figure 12 – QPSK modulation maps 2 bits to each subcarrier.

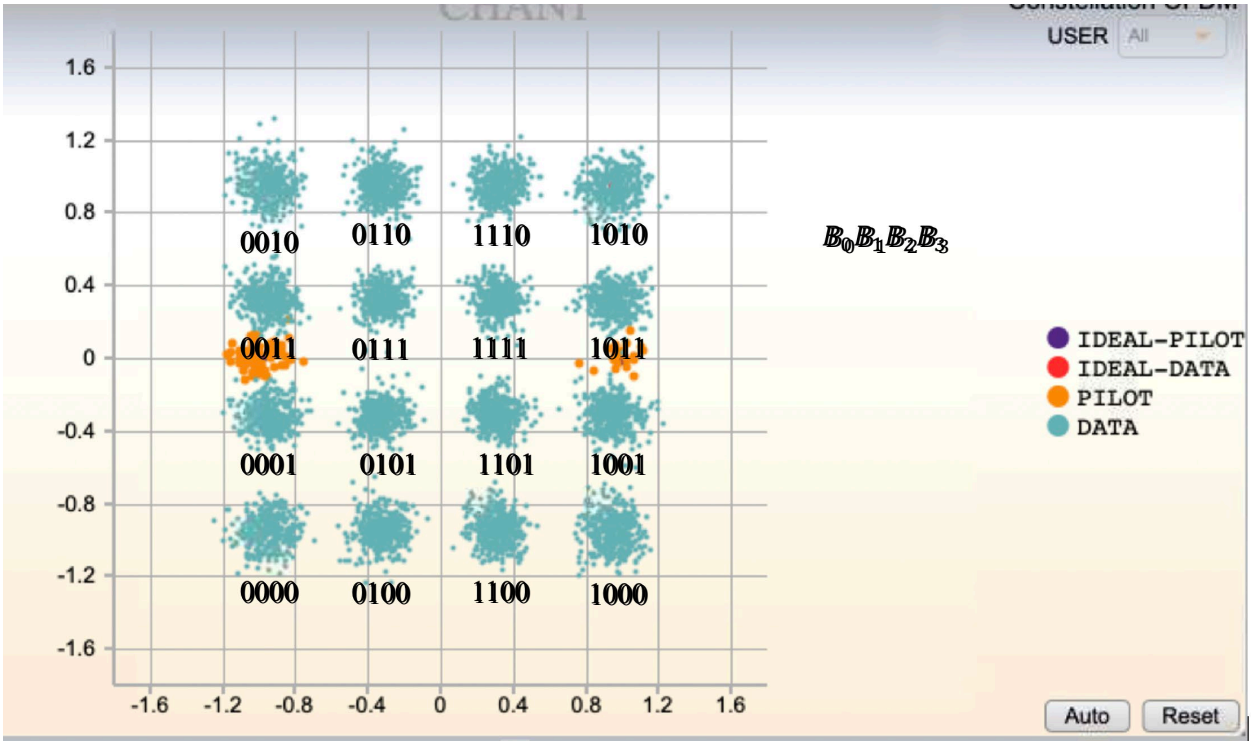


Figure 13 – 16-QAM modulation maps 4 bits to each subcarrier.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 13 – 16-QAM modulation maps 4 bits to each subcarrier. The markup shows how 4 bits are mapped to each constellation according to the IEEE 802.11 standard.

Table 6 – Modulation 2x2 320 MHz Wi-Fi 7 station

MCS Index	Modulation	Code rate	Subcarriers	PHY Mbps
14	BPSK	1/2	980	36
15	BPSK	1/2	1960	72
0	BPSK	1/2	3920	288
1	QPSK	1/2	3920	576
2	QPSK	3/4	3920	864
3	16-QAM	1/2	3920	1152
4	16-QAM	3/4	3920	1729
5	64-QAM	2/3	3920	2305
6	64-QAM	3/4	3920	2594
7	64-QAM	5/6	3920	2882
8	256-QAM	3/4	3920	3458
9	256-QAM	5/6	3920	3843
10	1024-QAM	3/4	3920	4323
11	1024-QAM	5/6	3920	4803
12	4096-QAM	3/4	3920	5188
13	4096-QAM	5/6	3920	5764

The PHY rate in Mbps is shown in Table 6 for all MCS indices from 0 to 15 for a 2x2 320 MHz EHT STA. The modulation and code rate for each MCS index is also shown. MCS 14 and MCS 15 only use one spatial stream since these rates are intended to trade off throughput to get the best range at cell edge. As a result, the PHY rate for MCS 15 is 72 Mbps while MCS 14 is 36 Mbps. All the other MCS indices calculate the PHY rate for two spatial streams since the EHT STA is 2x2 and the EHT AP is 4x4. Two spatial streams require sufficient multipath reflections for the required MIMO channel capacity. Indoor Wi-Fi channels have proven to easily support two spatial streams with 4x4 AP and 2x2 STA.

Table 7 – PHY rate in Mbps and SNR for 20, 40, 80, 160, 320 MHz channel width shows the PHY rate of each MCS from lowest PHY to highest PHY for all channel widths. The SNR requirement is the same for all the channel widths. The noise floor will change for each channel width but not the required SNR. Since the noise floor changes with channel width, the input level required for each MCS rate increases along with the channel width. For a noise figure of 3 dB, the noise floor in a 320 MHz channel width will be -86 decibel above a milliwatt (dBm). A 39 dB SNR will require an input level of -47 dBm. MCS13 for a 320 MHz channel width signal has a PHY rate of 5764 Mbps and needs an input level of -47 dBm. The SNR requirement, the number of spatial streams, the receiver noise figure, the maximum ratio combining effectiveness, and the quality of both the AP and STA radios will vary. Still these estimates are helpful in determining requirements and understanding how Wi-Fi 7 will work in customer homes. An SNR of about 1 dB is required for MCS14, indicating that an input level of -85 dBm will be required for a PHY rate of 36 dBm. Measurements of MCS14 show a throughput of about 30 Mbps with input level of about -85 dBm.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 7 – PHY rate in Mbps and SNR for 20, 40, 80, 160, 320 MHz channel width

STA	2x2		AP	4x4		
Channel Width in MHz		20	40	80	160	320
MCS	SNR in dB	PHY Mbps	PHY Mbps	PHY Mbps	PHY Mbps	PHY Mbps
14	1	N/A	N/A	9.0	18.0	36.0
15	3	4.3	8.6	18.0	36.0	72.1
0	3	17.2	34.4	72.1	144.1	288.2
1	6	34.4	68.8	144.1	288.2	576.5
2	8	51.6	103.2	216.2	432.4	864.7
3	11	68.8	137.6	288.2	576.5	1152.9
4	15	103.2	206.5	432.4	864.7	1729.4
5	19	137.6	275.3	576.5	1152.9	2305.9
6	20	154.9	309.7	648.5	1297.1	2594.1
7	21	172.1	344.1	720.6	1441.2	2882.4
8	26	206.5	412.9	864.7	1729.4	3458.8
9	28	229.4	458.8	960.8	1921.6	3843.1
10	31	258.1	516.2	1080.9	2161.8	4323.5
11	33	286.8	573.5	1201.0	2402.0	4803.9
12	36	309.7	619.4	1297.1	2594.1	5188.2
13	39	344.1	688.2	1441.2	2882.4	5764.7

Figure 14 – 2x2 320 MHz EHT device data rate including impact of preamble shows that the duration of the PPDU needs to be about 1 ms in duration to get close to the symbol data rate of 5.7 Gbps when factoring in the preamble. This is why the preamble field was detailed with particular attention to the duration of each field. The PHY rate only translates into high throughput when many data symbols are packed into a single PPDU. This tells us two things. First, aggregation will be just as important of a factor when measuring throughput as PHY rate. Second, the traffic demand must be very high to feed the high PHY rates so that PHY rate results in high throughput.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

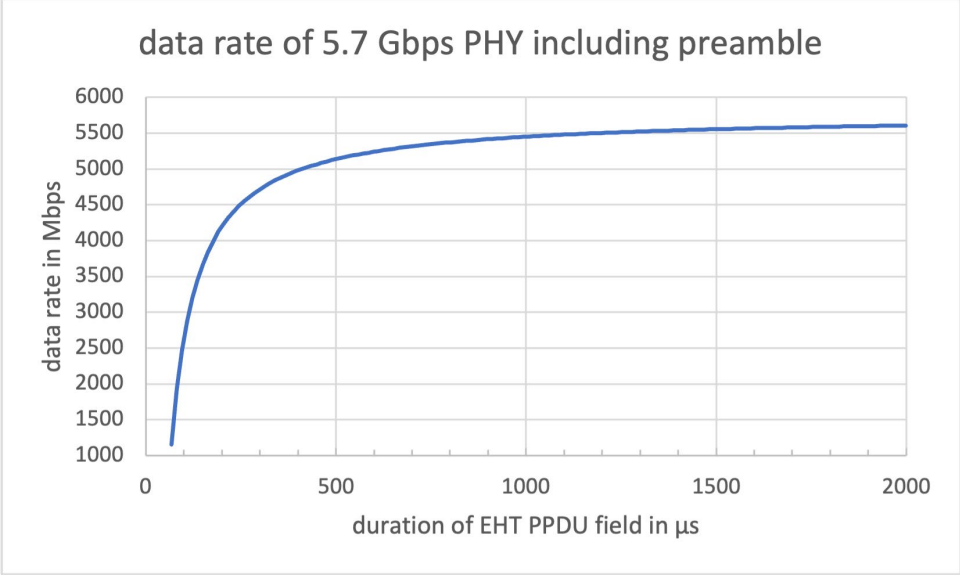


Figure 14 – 2x2 320 MHz EHT device data rate including impact of preamble

8. The hidden node problem

AP and STA check for channel busy before transmitting. This works as long as all radios here all signals. The hidden node problem is common and occurs when a radio checks for channel busy and finds all clear only because the interference at the receiver is blocked from the transmitter. Request to send (RTS), clear to send (CTS), and block acknowledgment (block Ack), prevent the hidden node problem by first asking the receiver to check for channel busy rather than just relying on the transmitters assessment.

The RTS from an AP to STA has a length of 76 bytes. The RTS data rate is 24 Mbps. The RTS duration is 28 μs with preamble of 20 μs and interframe space (IFS) of 83 μs. The CTS response from the STA to the AP has a length of 70 bytes. The CTS data rate is 24 Mbps. The CTS duration is 28 μs with 20 μs preamble and 14 μs IFS. The RTS and CTS packets are followed by a series of quality of service (QoS) Data packets and then a Block Ack packet. The Block Ack from the STA to the AP after a download of packets has a length of 88 bytes. The Block Ack data rate is 24 Mbps. The Block Ack duration is 32 μs with a preamble of 20 μs.

Figure 15 – 2x2 320 MHz EHT device data rate including impact of preamble and RTS CTS Block Ack plots the data rate as a function of duration from RTS to block Ack. At least two ms is required to get the effective data rate to 5 Gbps.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

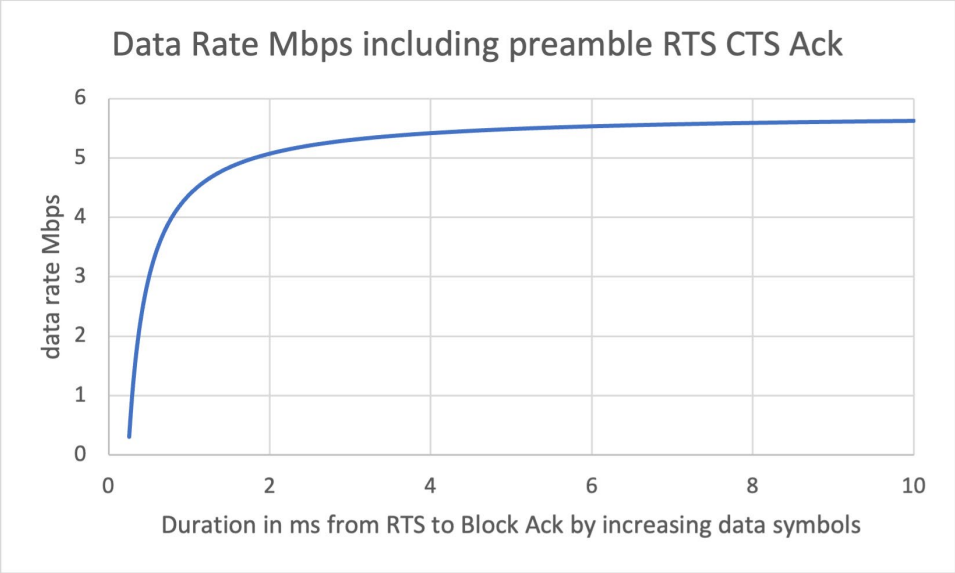


Figure 15 – 2x2 320 MHz EHT device data rate including impact of preamble and RTS CTS Block Ack

9. A quick word on beacons and probes

Often the most impactful variable in Wi-Fi performance in customer homes is beacon pollution, that is, too many beacons making it hard for clients to find the service set identifier (SSID) they are searching for and competing for airtime when trying to transmit and receive data packets.

The measured beacon duration of 0.6 ms is shown in Figure 16. Beacons are repeated every 102.5 μs so that a single beacon occupies only 0.59% of the airtime. Inverting the beacon duty cycle reveals that 171 beacons within earshot would try to fill all of the airtime. Access points often transmit multiple beacons for different services provided. If each AP transmits three beacons, then only 57 access points with overlapping coverage beacons attempt to transmit 100% of the time. Beacons must check for channel busy like any other transmission before transmitting so that beacons will not actually fill all the airtime. However, beacon pollution will still impact the data capacity of the spectrum and impact clients trying to connect to a desired SSID. Excessive beacons and probes, particularly with overlapping yet offset primary channel, trick the self-correlator into detecting signals that cannot be demodulated. This is called false detection and will result in automatic gain control to respond by increasing the effective receiver noise floor.

The 20 MHz 802.11a beacon signal is repeated 16 times to cover the full 320 MHz channel width as shown in Figure 16 – The measured beacon duration of a 320 MHz 6GHz band EHT signal.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

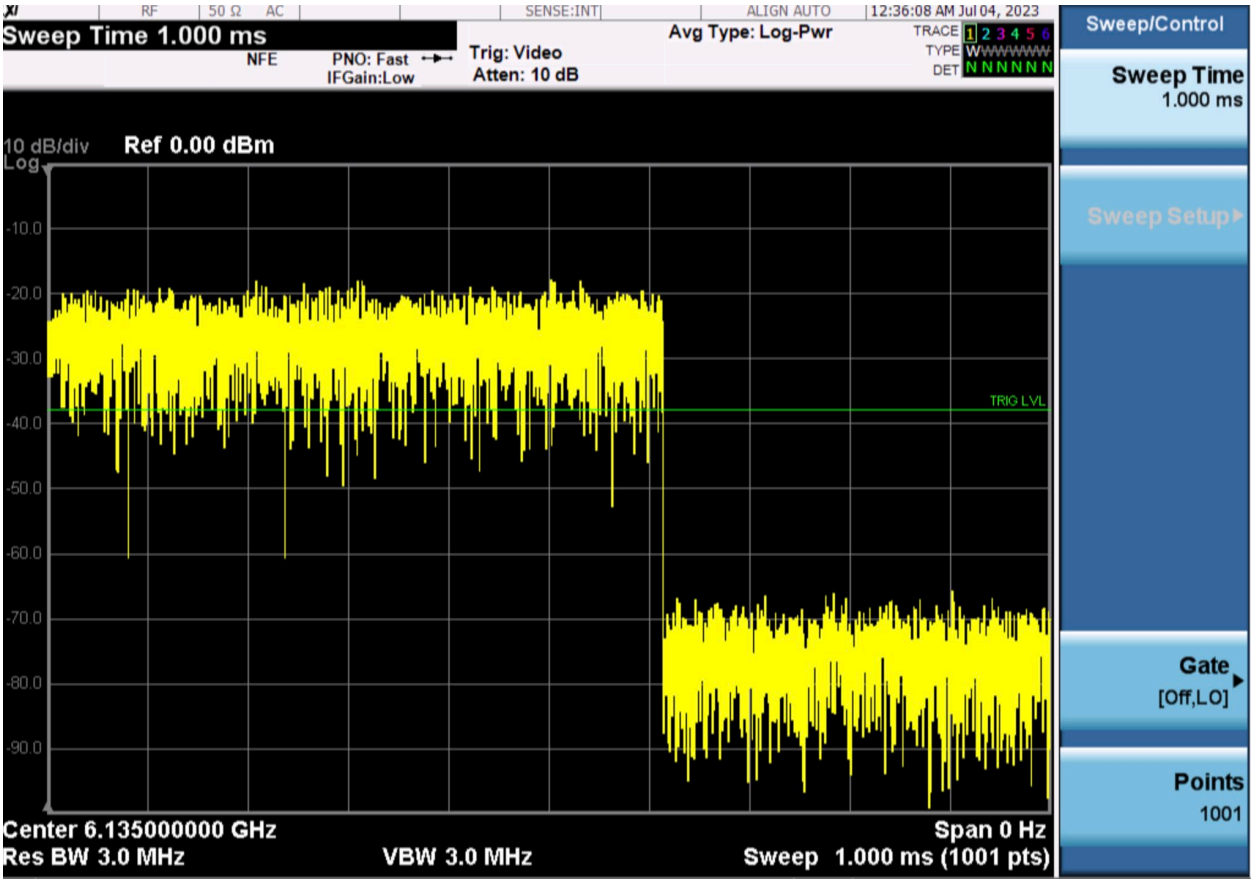


Figure 16 – The measured beacon duration of a 320 MHz 6GHz band EHT signal.

The peak power of the signal is only about 6 dB higher than the average power in this example, peak to average power ratio (PAPR) reduction is working. The modulation of the signal is BPSK as indicated by the constellation of two points separated by 180 degrees of phase shift.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

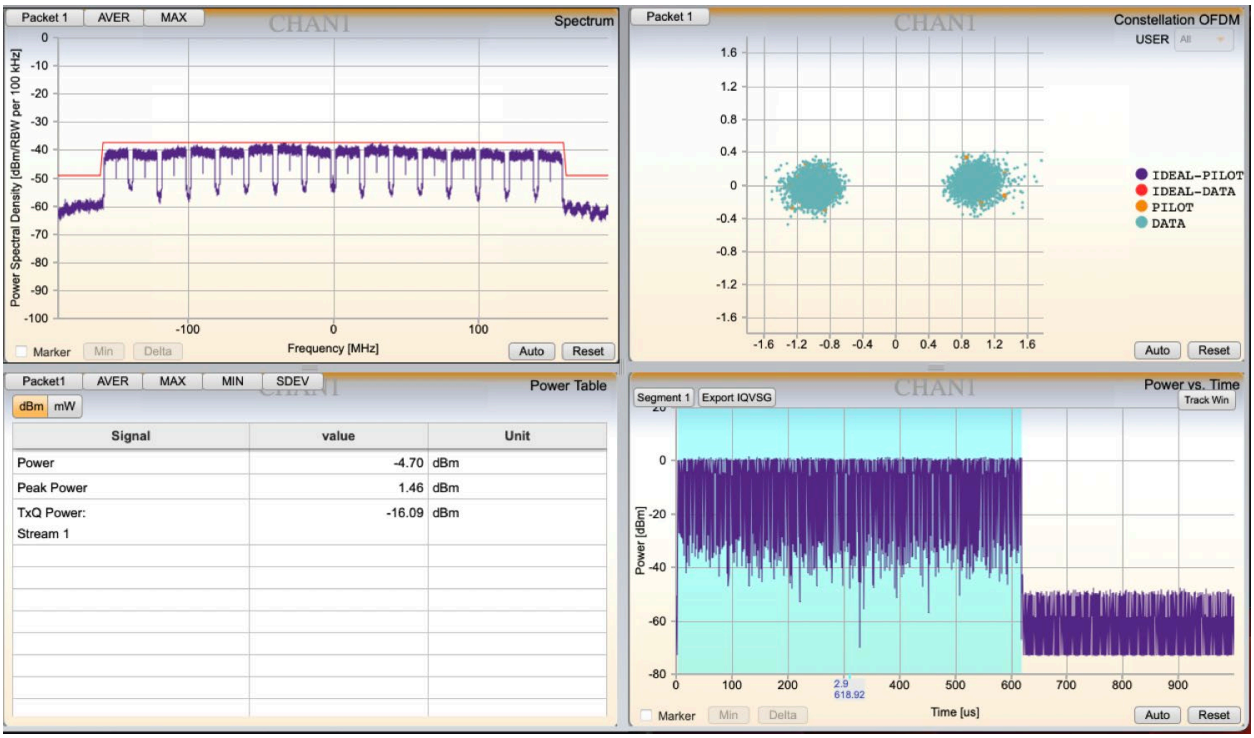


Figure 17 – The beacon is repeated over the full 320 MHz channel width

10. DCM+DUP: How to use 320 MHz channel width to improve range

The increase to 320 MHz channel width can be used to double the throughput with only a 3 dB penalty in input signal level. Since the 6 GHz band for indoor operation rules limit effective isotropic radiated power (EIRP) per MHz, doubling the channel width allows 3 dB increase in transmit power level which counteracts the 3 dB increase in noise with a doubling of the equivalent noise bandwidth. In general, wider channel width is the closest thing to a free lunch in wireless communications with a doubling of the channel width doubling the throughput with a small price of 3 dB increase in receive level. With EIRP per MHz as the regulatory limit, increasing the channel width is truly a free lunch, doubling the channel width doubles the throughput without reducing the link budget. The receive level needs to be 3 dB higher but the allowed transmit level is 3 dB greater.

Doubling the throughput by increasing the channel width from 160 MHz to 320 MHz is great and all for extremely high throughput, but how does this help with coverage range? Spread spectrum techniques have been utilized often to make a tradeoff between spectrum use, data rate, and range. The first cellular systems used wideband frequency modulation (FM) to improve the link budget by 10 dB by FM modulating a 3 kHz analog voice signal with 30 kHz FM.

Dual carrier modulation (DCM) and duplicate (DUP) combine to improve range for MCS14 in IEEE 802.11be. How does repeating modulation on two subcarriers within a 20 MHz channel width and then repeating the 20 MHz DCM subcarriers twice within the 320 MHz channel width improve range? To help us understand, let's begin with the old approach to enhancing range with narrow channel widths at higher power spectral density.

In the 2.4 and 5 GHz WLAN bands, the main regulatory restriction on transmit level is the total conducted power. There are also restrictions on out-of-band emissions, antenna gain, and power spectral

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

density as well as practical limits to DC power consumption and thermal constraints. However, in many cases, the conducted transmit level can get very close to the regulatory limit, without being impacted by the channel width.

Consider a 160 MHz channel width in the 5 GHz band transmitting at +24 dBm total conducted power. The noise level with a 3 dB noise figure receiver is -89 dBm for a 160 MHz channel width signal as calculated in Table 7 in reference (Urban, How Broadband Customers Can Benefit from Newfangled Wi-Fi Multiple User Features, 2022). As the receiver level gets close to -89 dBm the MCS rate will go to zero and spatial streams will go to one. As the level approaches -89 dBm at 160 MHz channel width, the rate adaptation algorithm can reduce the channel width from 160 MHz to 80 MHz to 40 MHz to 20 MHz. The conducted power into the antenna can remain at close to +24 dBm as the noise level goes to -92 dBm for 80 MHz channel width, and -95 dBm for 40 MHz channel width, and -98 dBm for 20 MHz channel width. At the very cell edge the MCS level will be zero with one spatial stream and the channel width will be 20 MHz.

In the 6 GHz band for indoor operation with a restriction of peak EIRP per MHz when the channel width is halved, the conducted transmit level must be reduced by 3 dB. Reducing the channel width from 320 MHz to 160 MHz will reduce the noise level from -86 dBm to -89 dBm. However, the transmit level must be reduced by 3 dB so that the signal to noise ratio of the receiver will not change. This is, in fact, one of the advantages of indoor limits on peak EIRP per MHz, as wider channel width is encouraged resulting in less concentrated interference levels to other users of the spectrum.

So, if going from 320 MHz channel width to 160 MHz channel width provides no advantage in SNR then what can be done when the receive levels approach the noise level of -86 dBm?

Consider packet detection, using self-correlation to detect the repeated symbols in the L-STF field of the preamble. A single L-STF has 16 samples to perform a cross correlation while the cross correlation of 256 samples can be used for signal detection when the L-STF is repeated 16 times over the full 320 MHz channel width. Detecting a signal with 256 samples is much easier than trying to detect a signal with only 16 samples.

Maximum ratio combining (MRC) can be employed on the L-SIG, RL-SIG, and U-SIG and EHT-SIG to provide a 12 dB range improvement compared to a 20 MHz PPDU since these signals are repeated 16 times over the 320 MHz channel width.

10.1. EHT DUP transmission

EHT duplicate (EHT-DUP) transmission repeats the payload portion of a PPDU in frequency. EHT-DUP is only for the 6 GHz band with 80, 160, and 320 MHz channel width and $N_{SS} = 1$ spatial streams.

MCS14 uses both DCM+DUP to get a 36 Mbps PHY rate with a 320 MHz channel width. The measured throughput using MCS14 and 320 MHz channel width is around 30 Mbps. The improved sensitivity between MCS0 and MCS14 for a 320 MHz channel width EHT signal measured between 2 and 6 dB.

While it proved hard to pinpoint the sensitivity difference between MCS0 and MCS14 due to the hysteresis and variability in establishing and maintaining connectivity at cell edge, it was certainly the case that when connectivity was lost at MCS0, connectivity and 30 Mbps of steady throughput could be restored by enabling MCS14.

30 Mbps throughput is not much in a multiple Gbps world. Yet it is important to understand the variability and fading that results in connectivity loss or a move to a band or extender not able to deliver

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

many Gbps speed. The advantage of the MCS14 PHY rate is to maintain connectivity in the 6 GHz band during temporary and transient fades.

11. Dual Lane PCIe3: The interface between DOCSIS 4.0 and Wi-Fi 7

Both DOCSIS 4.0 and Wi-Fi 7 are capable of throughput greater than 10 Gbps. Importantly, the maximum PHY rate of a 4x4 AP using 320 MHz channel width is over 11 Gbps. Experience has shown that bottlenecks feeding the Wi-Fi radio less than the PHY rate wreak havoc on Wi-Fi rate adaptation resulting in erratic behavior and excessive re-transmission.

The transactions per second of peripheral component interconnect express (PCIe) 3 have been increased to 8 Gbps compared to PCIe 2 transactions per second of 5 Gbps. Both standards use two bits of parity for error correction. However, the older standard uses eight information bits while the newer standard uses 128 information bits. The older PCIe 2 standard has a raw information rate of 4.0 Gbps, while PCIe 3 has a raw information rate of 7.8 Gbps. The raw information rate of dual lane PCIe 3 is 15.7 Gbps.

Feeding a Wi-Fi 6 radio with a maximum PHY rate of 4800 Mbps using a single lane PCIe 2 measured TCP throughput of close to 2.5 Gbps with a window size of 4 Mbytes and 8 TCP streams. Assuming equal efficiency between PHY rate and throughput indicates that the throughput will be about 9.8 Gbps with a 15.7 Gbps dual lane PCIe 3 raw information rate and an 11.4 Gbps PHY rate. Indeed, this matches measured results on early Wi-Fi 7 radios.

12. Conclusion

DOCSIS 4.0 and Wi-Fi 7 are indeed perfect together.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

Ack	acknowledgment
ADC	Analog-to-digital converter
AP	access point
bps	bits per second
BPSK	binary phase shift keying
BCC	binary convolutional coding
CM	cable modem
CMTS	cable modem termination system
CTS	clear to send
DAA	distributed access architecture
dBm	decibels above a milliwatt
DC	direct current, baseband representation of center frequency
DCM	dual carrier modulation
DFT	discrete Fourier transform
DOCSIS	data over cable service interface specification
DUP	duplicate DCM
EHT	extremely high throughput
EHT-DUP	extremely high throughput duplicate transmission
EHT-LTF	EHT Long Training field
EHT MU PPDU	EHT multi-user PPDU
EHT-SIG	EHT SIGNAL field
EHT-STF	EHT Short Training field
EHT TB PPDU	EHT trigger based PPDU
EIRP	effective isotropic radiated power
EVM	error vector magnitude
FCC	Federal Communications Commission (USA)
FDD	frequency division duplex
FDX	full duplex DOCSIS
FEC	forward error correction
FFT	fast Fourier transform
FM	frequency modulation
Gbps	gigabits per second
GI	guard interval
HE	high efficiency
HFC	hybrid fiber coaxial cable
HT	high throughput
Hz	hertz
IDFT	inverse discrete Fourier transform
IEEE	Institute of Electrical and Electronic Engineers
IETF	Internet Engineering Task Force
IFS	interframe space
ISM	Industrial scientific medical band
K	kelvin
LDPC	low density parity check coding
LLD	low latency DOCSIS

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

L-LTF	Non-HT Long Training field
L-SIG	Non-HT SIGNAL field
L-STF	Non-HT Short Training field
LTF	long training field
Mbps	Megabits per second
MIMO	multiple input multiple output
MRC	maximum ratio combining
OFDMA	orthogonal frequency division multiple access
PAPR	peak to average power reduction
PCIe	peripheral component interconnect express
PSC	preferred scanning channel
PE	Packet Extension field
PHY	physical layer interface
PPDU	physical layer protocol data unit
QAM	quadrature amplitude modulation
QoS	quality of service
QPSK	quadrature phase shift keying
RF	radio frequency
RPHY	remote physical interface
RL-SIG	Repeated Non-HT SIGNAL field
RTS	request to send
SCTE	Society of Cable Telecommunications Engineers
SNR	signal to noise ratio
SSID	service set identifier
STA	station
STF	short training field
TDD	time division duplex
U-NII	Unlicensed National Information Infrastructure
U-SIG	Universal SIGNAL field
vCMTS	virtual cable modem termination system
WAN	wide area network
WLAN	wireless local area network

Bibliography & References

Comcast Announces World-First Test of 10G Modem Technology Capable of Delivering Multigigabit Speeds to Homes. (2022, January 13). Retrieved from https://corporate.comcast.com/press/releases/world-first-test-10g-modem-technology-multigigabit-speeds-to-homes

Comcast Demonstrates Fastest-Yet Speeds Over a Complete 10G Connection on a Live Network. (2022, April 28). Retrieved from https://corporate.comcast.com/press/releases/comcast-fastest-speeds-complete-10g-connection-live-network

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Comcast Kicks Off Industry's First Low Latency DOCSIS Field Trials. (2023, June 16). Retrieved from https://corporate.comcast.com/stories/comcast-kicks-off-industrys-first-low-latency-docsis-field-trials

FCC. (n.d.). Part 15.407.

IEEE 802.11be D3.0. (2023, January). *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,*.

Robert Howald, J. U. (2022). Collision-Free Hyper-Speeds on the Bi-Directional FDX Highway . *SCTE CABLE-TEC EXPO 22 2022 Fall Technical Forum.* Philadelphia.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *27*.

Urban, D. J. (2019). The Importance Of Wi-Fi 6 Technology For Delivery Of GBPS Internet Service. *Cable-Tec Expo Fall Technical Forum.* New Orleans.

Urban, D. J. (2022). How Broadband Customers Can Benefit from Newfangled Wi-Fi Multiple User Features. Philadelphia: SCTE Cable-Tec Expo 22 Fall Technical Forum.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

DOCSIS® Waveforms Designed for Improved Upstream RF Leak Detection and Localization

A Technical Paper prepared for SCTE by

Richard A Primerano
Principal Engineer
Comcast
Richard_Primerano@comcast.com

Benny Lewandowski
Engineering Architect IV
Comcast
Benny_Lewandowski@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Background ... 3
   - 2.1. Using the OUDP Signal for Upstream Leak Detection ... 3
     - 2.1.1. The Anatomy of the OUDP Pulse ... 3
   - 2.2. Detection via Matched Filtering ... 5
3. The OUDP Pulse Train ... 5
   - 3.1. Detection using Autocorrelation ... 5
   - 3.2. Demonstrating the Detector in a Cabled Test ... 6
   - 3.3. Over-the-Air Testing ... 7
   - 3.4. Incorporating Doppler Processing for Leak Localization ... 8
4. Choosing OUDP Pulse Train Parameters ... 9
   - 4.1. Choosing Parameters based on Environment ... 10
5. Conclusion ... 10

Abbreviations ... 11

Bibliography & References ... 11

List of Figures

Title **Page Number**

Figure 1 - DOCSIS 3.1 Cable modem upstream transmitter block diagram [2] ... 4
Figure 2 - Pilot pattern 11 has the highest density of pilot symbols [2] ... 4
Figure 3 – A multi-pulse OUDP burst ... 5
Figure 4 - An autocorrelation-based detector for detecting multi-pulse bursts ... 6
Figure 5 - A synthesized 5-pulse OUDP burst ... 6
Figure 6 - Output from correlation-based detector (magnitude only) ... 6
Figure 7 - Baseline capture showing strong interference ... 7
Figure 8 - Correlator output showing OUDP peaks and strong interference ... 8
Figure 9 - A particle filter converging on the true location of a leak over time ... 9

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

RF signal leakage occurs in the coaxial distribution system due to cable damage, loose connections, and other issues. FCC regulations limit allowable leakage field strengths and require cable providers to routinely survey the plant for RF signal leakage. In standard-split plants, the primary spectrum of interest falls in the 108 MHz to 137 MHz VHF aeronautical band, and leaks are detected by having headend equipment generate a test signal that is easily detected by specially designed meters. In high-split and full duplex (FDX) plants, critical bands now fall in the upstream, and the cable modems must generate the probe signal. The upstream OFDMA signal is noise-like, making it difficult to detect directly. The OFDMA frame contains a known preamble pilot pattern that can be detected with a traditional matched filter approach, but since the pilot is a small portion of the overall frame, an equally small portion of the leak energy can be captured for detection.

In this paper, we present a new upstream leak detection technique built around a specially designed DOCSIS ODFMA upstream data profile (OUDP) testing burst. By sending two identical pulses in a row, with known spacing between them, we can build a *correlation-based detector* that captures the energy of the full leak pulse. As a result, we have demonstrated a sensitivity increase of greater than 20 dB over the preamble detection approach.

The output of the correlation-based detector has both magnitude (indicating the strength of the leak) and a phase (indicating the relative velocity of the detector with respect to the leak) components. We demonstrate in the paper how this Doppler velocity measurement can be used to localize leaks in space using particle filtering techniques borrowed from the radar and image processing fields.

2. Background

2.1. Using the OUDP Signal for Upstream Leak Detection

With the introduction of high-split and FDX plants, critical FCC-regulated bands that were previously in the downstream now exist in the upstream. This raises new leakage detection challenges because of the busty nature of upstream signals and the noise-like nature of OFDMA signals. The DOCSIS specification requires both DOCSIS 3.1 and DOCSIS 4.0 cable modems to support generation of the OUDP signal. The signal is used for measuring channel performance and performing sounding. In addition, the industry is settling on using this pulse to support upstream leak detection. The current technique relies on building a matched filter on a portion of the OUDP pulse that is known [3,4]. Next, we look more closely at the structure of the OUDP pulse to understand the level of achievable processing gain using this matched filter approach.

2.1.1. The Anatomy of the OUDP Pulse

The OUDP pulse passes through the same upstream transmission block as regular data transmissions. This block, shown in Figure 1, contains error control coding (LDPC encoder), scrambling, and data interleaving. This cascade of operations converts the incoming data to a noise-like signal. Unless the incoming data, scrambler state, and randomizer state are all known, we will not know what signal is being transmitted over the wire, preventing us from building a matched filter to detect it.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

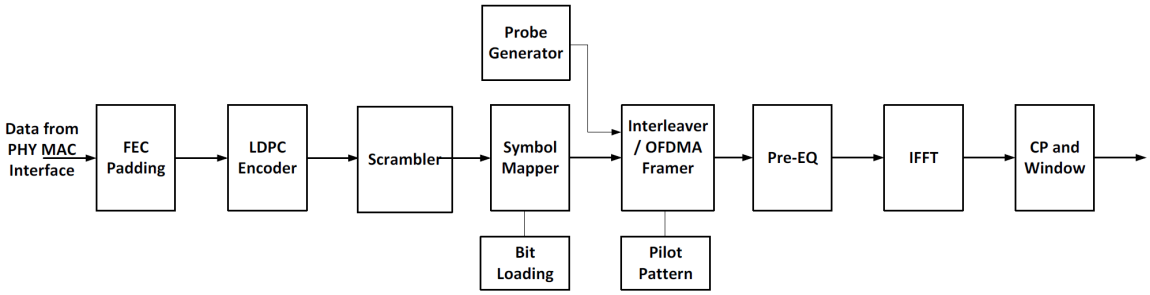


Figure 1 - DOCSIS 3.1 Cable modem upstream transmitter block diagram [2]

Note that the upstream signal path also contains a *Pilot Pattern* block. This block inserts a known data pattern into the minislot just before transmission. In normal operation, this pattern is used for upstream channel equalization. Some pilot patterns defined by the DOCSIS PHY specification [2] are shown in Figure 2. Those shown in the figure are for the 4k subcarrier OFDMA configuration. While we can't build a matched filter on the payload data subcarriers (white squares in the figure), we can build one around the pilots (red/green squares). Pilot pattern 11 has the highest density of pilot subcarriers compared to other patterns.

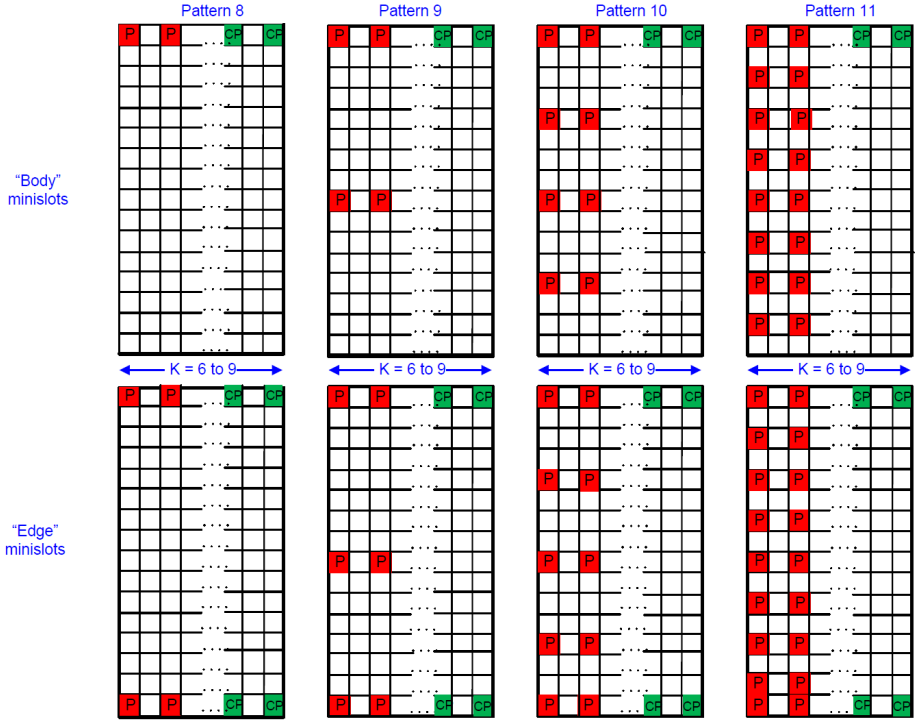


Figure 2 - Pilot pattern 11 has the highest density of pilot symbols [2]

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.2. Detection via Matched Filtering

With a deterministic portion of the OUDP signal identified, a matched filter can be built to detect it. The matched filter has an impulse response that is a conjugated version of the signal to be detected. Mathematically, the (discrete time) matched filter is given by

$$y[n] = \sum_{k=0}^{T} h[n-k]x[k]$$

where $x[k]$ is the leak signal, $h[k]$ is the matched filter impulse response, and T is the support of the matched filter. The filter output $y[n]$ is then sampled and if the signal exceeds some threshold, a leak detection is declared. The matched filter is optimal in that it maximizes the signal-to-noise ratio (SNR) of the output signal in the presence of a leak. The SNR of the matched filter can be expressed as

$$SNR = \frac{\sum_{k=0}^{T} x[k]}{kBT}$$

which is the ratio of the energy in the signal to be detected to thermal noise energy. In the denominator, k is the Boltzmann constant, B is the detector bandwidth, and T is the temperature. Note that the SNR of this filter can be increased either by increasing signal energy or by reducing detector bandwidth. In the case of DOCSIS leak detection, bandwidth reduction will not help because any reduction in bandwidth gives an equal reduction in signal energy. Furthermore, the fixed pilot pattern in the OUDP pulse places a fixed value on the numerator of the above equation.

3. The OUDP Pulse Train

The matched filter detector relies on knowing some portion of the transmitted signal. Other detectors can be built to leverage known statistical properties of the signal. Consider a signal like the one in Figure 3, representing a burst of multiple OUDP pulses, evenly spaced. The waveform is defined by three parameters: OUDP pulse width (T_p), burst repetition period (T_r), and number of transmitted bursts (n).

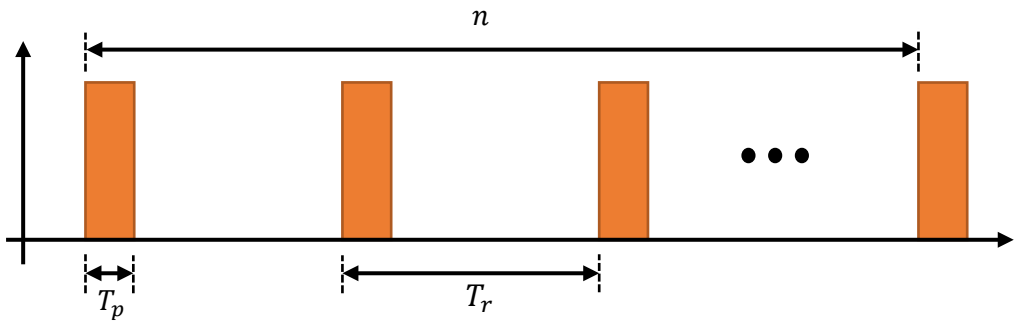


Figure 3 – A multi-pulse OUDP burst

3.1. Detection using Autocorrelation

A repeated waveform like the one in Figure 3 can be detected through *autocorrelation*. In this process, the signal is compared to a delayed copy of itself. If the spacing between pulses is known, the autocorrelation can be computed at exactly that lag, resulting in a low-complexity detector. The block diagram in Figure

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4 shows the implementation of the detector. The incoming signal $x[n]$ is multiplied by a delayed and conjugated version of itself. The delay is equal to the pulse repetition interval T_r in Figure 3. The resulting signal is averaged over a period of T_p, the OUDP pulse width in Figure 3.

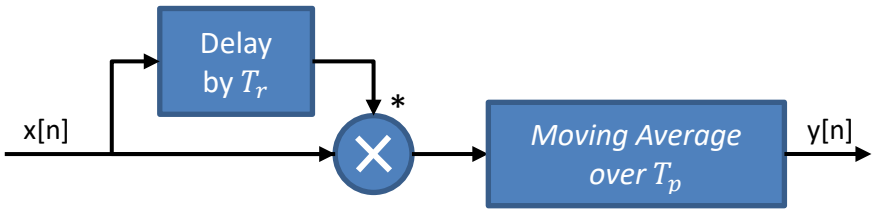


Figure 4 - An autocorrelation-based detector for detecting multi-pulse bursts

3.2. Demonstrating the Detector in a Cabled Test

To demonstrate the approach, we generated an OUDP multi-pulse burst by capturing a single OUDP burst from a cable modem and repeating it five times with known spacing. The pulse, along with its parameters, is shown in Figure 5.

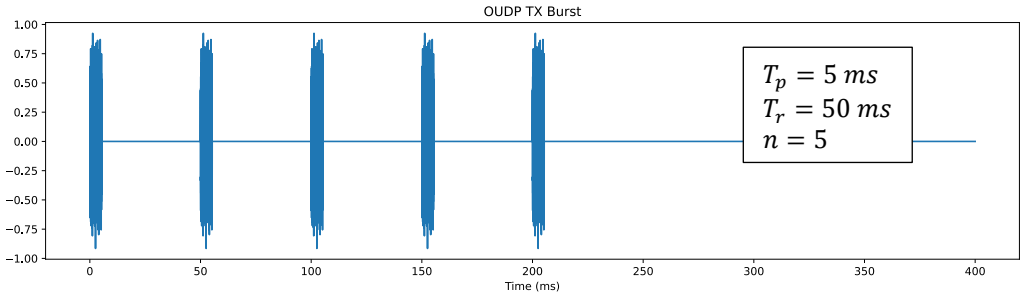


Figure 5 - A synthesized 5-pulse OUDP burst

The burst was transmitted from one software-defined radio (SDR), over cable to another SDR and recorded. Finally, the detector in Figure 4 was implemented in Python and used to process the recorded signal. The output of the detector is shown in Figure 6. Note that the received signal is not time-aligned to the transmitted one, resulting in a time shift. As expected, the detector output contains four triangular-shaped pulses for the five received pulses.

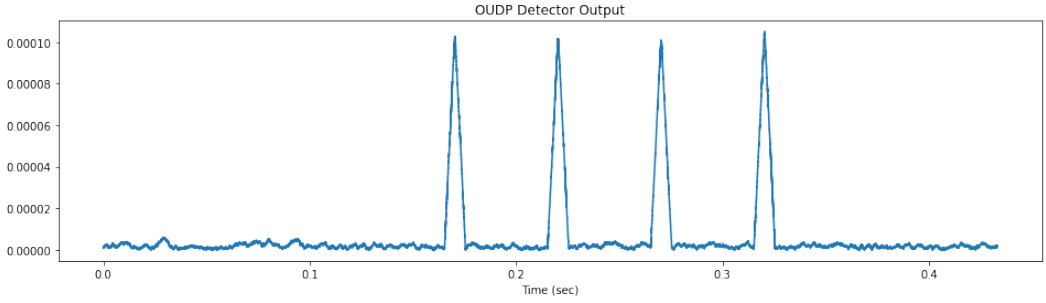


Figure 6 - Output from correlation-based detector (magnitude only)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.3. Over-the-Air Testing

Using the same five-pulse burst shown above, we conducted indoor over-the-air (OTA) testing. The test was conducted in the 433 MHz portion of the 70-centimeter amateur radio band, but the same principles apply for VHF aeronautical (108 MHz to 137 MHz) and lower LTE (600 MHz) bands. Transmitter and receiver were 50 meters apart and the transmit power was set to a representative DOCSIS OUDP pulse.

Because the test was done in an amateur radio band, we expect that at times there will be significant interference. One weakness of the correlation-based detector is its susceptibility to correlated interference signals. Any interference that has the same correlation properties as our OUDP burst will pass through the detector. Figure 7 shows the output of the detector when to leak signal was being transmitted. This baseline capture shows the effect of correlated interference on the detector. This could have been from licensed amateur radio communications or from non-licensed Part 15 devices[1] operating in the vicinity of 433 MHz.

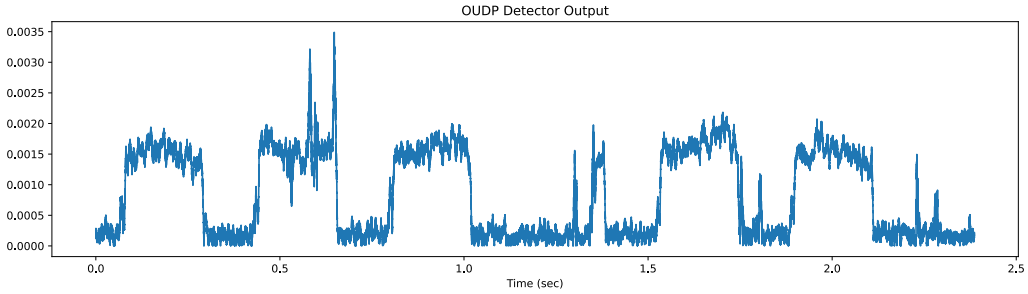


Figure 7 - Baseline capture showing strong interference

Next, we performed the same capture while the OUDP leak signal was being transmitted continuously on a loop. The capture in Figure 8 shows the detector output in this case. Four pulse groups are clearly visible among the interference energy. During quiet periods, we observe post-detector signals with an SNR in excess of 20 dB. Later in the paper we discuss techniques for rejecting interference.

[1] A "Part 15 device" is a low-power, non-licensed transmitter that complies with the regulations in Part 15 of the FCC rules. Part 15 transmitters typically use low power, often less than a milliwatt. Part 15 transmitters are non-licensed because their operators are not required to obtain a license from the FCC to use them. Some examples of Part 15 devices include automotive, garage, and consumer remote controls. Per the FCC Rules, Part 15 devices must not cause harmful interference, and must accept any interference caused by licensed OTA services. Some Part 15 devices operate in the vicinity of 433 MHz, which is in the 70-centimeter amateur radio band, a licensed over-the-air radiocommunication service.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

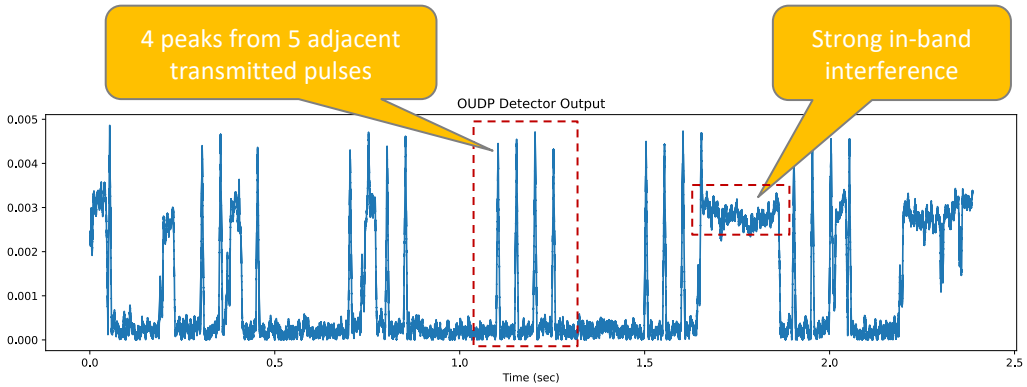


Figure 8 - Correlator output showing OUDP peaks and strong interference

3.4. Incorporating Doppler Processing for Leak Localization

The detector output in Figure 8 shows magnitude only. If the receiver is mounted in a moving vehicle, we can use the phase of the detector output to measure the relative velocity of the leak with respect to the vehicle. The relationship between velocity and phase is given by:

$$v = \frac{\theta c}{T_r f_c}$$

where θ is the phase of the detected peak (in radians), c is the speed of light, T_r is the pulse repetition interval, and f_c is the center frequency of the OUDP pulse.

If a leak is observed over a span of several minutes as a vehicle moves through an area, we will build a time-history of Doppler velocity estimates between the vehicle and leak. Furthermore, if we know the position and velocity of the vehicle for each measurement, we can apply tracking techniques to help localize the leak. This is important because detected leaks must generally be repaired. Figure 9 shows a simulation of a particle filter being used to localize a leak (gold star at x=700 m, y=250 m) using velocity measurements captured from a vehicle-mounted leak sensor (blue star starting at x=400 m, y=50 m). The arena is 1 km wide by 500 meters high and the four frames shown below would represent a collection spanning several minutes.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

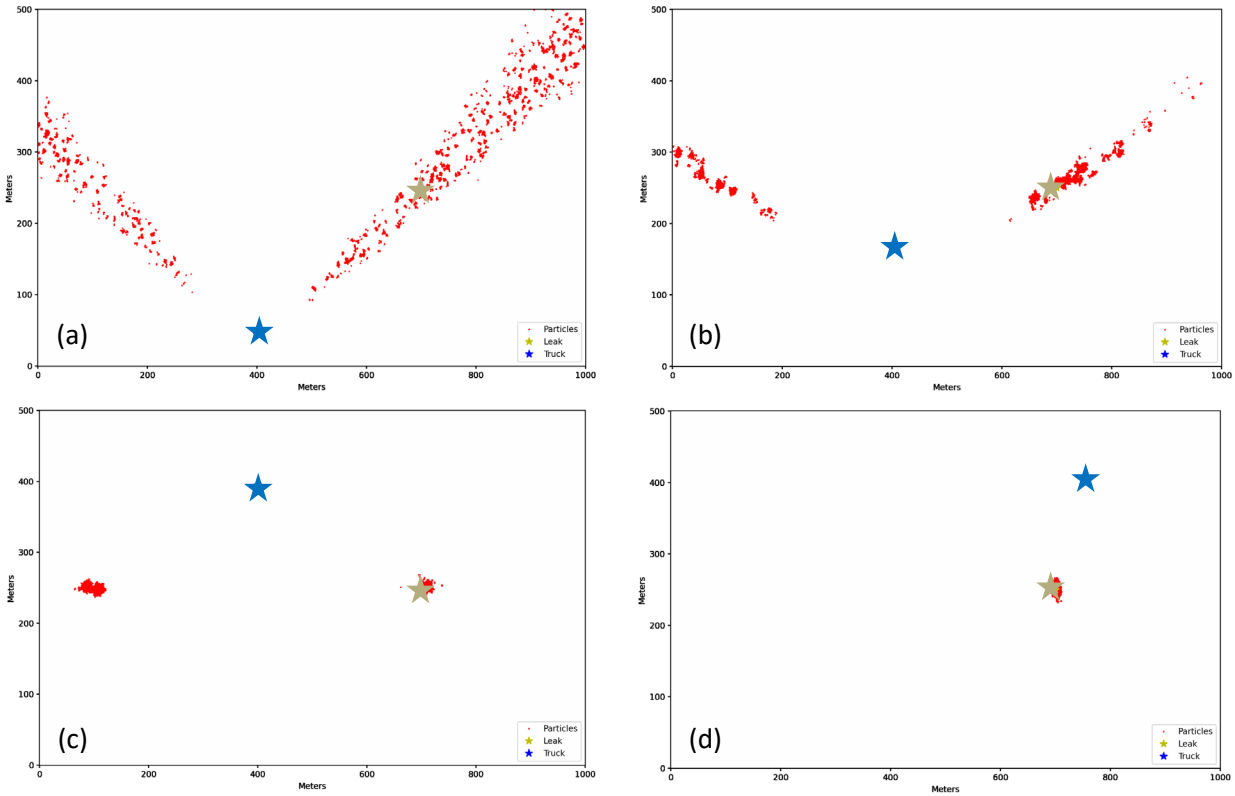


Figure 9 - A particle filter converging on the true location of a leak over time

The red particles represent assumed locations for the leak. The particle filter begins by randomly distributing them in the arena. As the vehicle moves forward (a), it measures a velocity relative to the leak indicating the leak is in front of the vehicle at a bearing of approximately 45 degrees. As the vehicle continues to move (b), it collects additional measurements indicating that the leak is almost stationary relative to the vehicle, i.e., the leak is almost at a right angle to the vehicle's motion. Further in time (c), the vehicle observes a change in sign of the velocity, indicating the leak is moving away from the vehicle. The set of measurements taken between (a) and (c) are all in a single direction. This is not enough data for the particle filter to determine if the leak is to the left or right. As the vehicle changes bearing, the velocity relative to the leak shows the leak is moving toward the vehicle. This immediately allows the particle filter to determine which of the candidate locations from (c) is the true leak location. Although we have described the process above as the tracker making decisions about leak location at discrete points in time, the actual operation happens continually. The particle filter is solving a maximum-likelihood problem iteratively such that the particles coalesce around the point(s) that best agree with the velocity measurements the filter has seen over time.

4. Choosing OUDP Pulse Train Parameters

The three pulse-burst parameters in Figure 3 can be freely adjusted by the designer to optimize system performance. Some considerations when selecting pulse parameters are discussed next.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Pulse Width (T_p)

Pulse width is the main parameter that sets detection sensitivity. The relationship between pulse width (T_p), signal bandwidth (BW), and processing gain (PG) is approximately $PG = 10 \log (T_p * BW)$. Using typical values of $T_p = 5\ ms$ and $BW = 1.6\ MHz$, we get $PG = 10 \log(8000) = 39\ dB$.

Pulse Repetition Interval (T_r)

When both the transmitter and receiver are stationary, the phase of the autocorrelator's output peak is zero. If there is relative motion between the two, the phase will be non-zero. The phase of the peak encodes the distance (in wavelengths) the receiver has traveled between pulse. The value of T_r sets the maximum unambiguous velocity (v_{max}) that can be measured, $v_{max} = \lambda/2T_r$. The receiver can travel no more than one half of a wavelength between pulses for unambiguous measurement. Using typical values of $T_r = 50\ ms$ and $\lambda = 2.3\ meters$ ($f = 130\ MHz$), we get $v_{max} = 23\ m/s = 51\ mph$.

Number of Pulses (n)

The minimum number of pulses that can be transmitted is two, resulting in a single peak at the detector output. Transmitting *n* pulses results in $n - 1$ detection events. If the relative motion between transmitter and receiver is constant over the duration of these pulses, the detection events can be coherently averaged to form a single effective detection. In this case, the overall processing gain becomes $PG = 10 \log ((n - 1) * T_p * BW)$.

4.1. Choosing Parameters based on Environment

The autocorrelation-based detection is *blind*. It requires no knowledge of the transmitted signal and is sensitive only to the repetition within the signal. One drawback of this approach is that any signal exhibiting repetition of length T_r will pass through the detection and appear as interference. We may tune the three OUDP leakage burst parameters based on background interference characteristics to minimize their impact on detector performance.

Waveform parameter selection can also be used to tune system performance based on the environment. Examples include:

- **A dense urban environment:** *many cable modems in RF range, close to the receiver's path of travel*
  - A narrow pulse width is used. This results in low sensitivity, but a larger number of devices can be scheduled in a fixed time interval.
  - Since homes are near the road, RF path loss is low and the sensitivity loss is acceptable.
- **A sparse rural environment:** *few cable modems in RF range, far to the receiver's path of travel*
  - A wide pulse width is used. This results in high sensitivity, but fewer devices can be scheduled in a fixed time interval.
  - Since homes are far from the road, the high sensitivity pulse overcomes RF path loss.

5. Conclusion

In this paper, we have presented a new leak detection approach built on the same OUDP waveforms currently defined in the DOCSIS specification. By sending multiple pulses at known spacing, we can employ an autocorrelation-based detector that has high processing gain to detect weak signals. This approach has a drawback: specific forms of interference have the potential to adversely affect

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

performance. By integrating Doppler processing and adjusting pulse parameters based on environment, we expect some of these issues can be overcome.

Abbreviations

BW	Signal Bandwidth
dB	decibel
DOCSIS	Data-Over-Cable Service Interface Specifications
FCC	Federal Communications Commission
km	kilometer
m	meter
v_{max}	maximum unambiguous velocity
MHz	megahertz
Ms	millisecond
N	number of pulses
OFDMA	orthogonal frequency-division multiple access
OTA	over-the-air
OUDP	OFDMA upstream data profile
PG	processing gain
RF	radio frequency
SDR	software-defined radio
SNR	signal-to-noise ratio
T_p	pulse width
T_r	pulse repetition interval

Bibliography & References

[1] *Data-Over-Cable Service Interface Specifications DOCSIS® 3.1, MAC and Upper Layer Protocols Interface Specification, CM-SP-MULPIv3.1-I25-230419*; Cable Television Laboratories, Inc.
[2] *Data-Over-Cable Service Interface Specifications DOCSIS® 3.1, Physical Layer Specification, CM-SP-PHYv3.1-I20-230419*; Cable Television Laboratories, Inc.
[3] *Leakage Detection in a High-Split World, Industry Progress Toward a Viable Solution*; R. Coldren et al., SCTE Fall Technical Forum, October 2021
[4] *Leakage Detection in Full Duplex DOCSIS*; B. Lewandowski et al., SCTE Fall Technical Forum, September 2022

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Dynamic Deep Cycling Testing

The Use of Dynamic Deep Cycling Testing to Predict Battery State-of-Health in Outside Plant Environments

A Technical Paper prepared for SCTE by

Michael Nispel
Sr. Principal Engineer
Comcast Cable
Downingtown, PA
+1 (610) 952-3783
Michael_Nispel@comcast.com

Alexander Falcon
Product Manager
Comcast Cable
Downingtown, PA
+1 (484) 999-7762
alexander_falcon@comcast.com

Kang Lin
Manager Quality Assurance
Comcast Cable
Downingtown, PA
kang_lin@comcast.com

Paul Schauer
Distinguished Engineer
Comcast Cable
Denver, CO
paul_schauer@cable.comcast.com

Cory Thompson
Technical Project Manager
Comcast Cable
Denver, CO
cory_thompson@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Background ... 3
   - 2.1. Identification of the Proposed Solution ... 3
   - 2.2. State-of-Health Algorithm ... 4
3. Data Collection ... 6
4. Results ... 7
   - 4.1. Training ... 7
   - 4.2. Verification Runs ... 8
5. Implementation ... 10
6. Conclusion ... 10

Abbreviations ... 11

Bibliography & References ... 11

List of Figures

Title **Page Number**

Figure 1 - Points of Interest in Voltage Discharge Curve ... 5
Figure 2 – Coup-de-fouet of a VRLA Battery ... 6
Figure 3 – Collection of Data by Time ... 7
Figure 4 – Battery Voltages During Discharge Testing ... 8
Figure 5 – Dataset from Verification Testing ... 9
Figure 6 – Calculated Metric Values for Batteries Tested SOH ... 9

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Battery health is critical to the reliability of the outside plant network for cable internet providers. Unfortunately, the health of a battery is continuously being degraded due to use, environment, abuse, and many other known and unknown factors. This makes the change in battery performance variable and difficult to predict, with a battery's calendar age an ineffective method to determine performance. Although discharge events can be tracked readily, the challenge is that the vast majority of outages result in partial battery discharges, with power restored at inconsistent and variable times. To address this challenge, Comcast set out to use a partial discharge analysis to determine the state-of-health (SoH) for each battery powering the over 250 thousand outside plant power supplies in its network. SoH is defined here as the percentage capacity (or runtime) a battery can deliver as compared to its new or original rating. This was done without removing the battery from the site, without the need for any external equipment, without the need to visit the site, and without ever having the downstream load unprotected.

Previously, the replacement approach for power supply batteries was based on the calendar age of the battery. This did not account for the many factors that may have degraded the battery prematurely. As such, it may have resulted in the early replacement of batteries in good health or the delay in removing batteries in poor health, which translates directly to a false sense of reliability. The absence of a view into battery health made capital planning more difficult, as battery attributes (model, count, age, etc.) were the only information available to justify replacement. By providing a health score for batteries, long-range capital planning can focus on the true condition of batteries rather than their calendar age. This view into asset condition allows for the most effective use of resources by targeting the locations of greatest need. Views into future years' battery replacement quantities are much more predictable using current battery health metrics and their degradation rate.

The replacement of the batteries of greatest need also significantly improves the reliability of the outside plant powering network. Power supplies can withstand commercial power interruptions more effectively when poor performing batteries are identified and replaced. Improved power supply reliability directly impacts customers who may have a backup generator available or may continue to have commercial power during isolated power outages. Providing front line maintenance technicians with a view into each battery's SoH promotes an effective, efficient, and proactive maintenance strategy for power supply battery replacements.

A SoH assessment was accomplished by developing a unique algorithm used to predict the performance of outside plant batteries when subjected to a controlled, partial discharge event, specific to the power supply's unique load. In this way, Comcast will be improving the reliability of the outside plant network by advancing beyond a calendar age replacement cycle. Access to this information aids in the continued effort to improve infrastructure reliability. The use of a dynamic deep cycle battery discharge test with a prediction of the battery state-of-health will continue to improve Comcast's best-in-class powering network.

2. Background

2.1. Identification of the Proposed Solution

In traditional power backup applications to the utility grid, batteries are continuously on standby, ready to be discharged when utility power fails. In today's Hybrid-Fiber Coax network, these batteries are primarily valve-regulated lead-acid batteries, which have an underlying chemistry whereby the SoH is impacted over time by various influences including charge/discharge cycles, temperature, time and other

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

factors. Partial discharge events are common, where power is restored at completely inconsistent and variable times before the battery has been completely depleted. In contrast, full discharge events (where a battery is fully depleted) are undesirable. If a full outage continues to the point of depleting a battery, the site then fails and customers experience a service interruption as the outside plant equipment necessary for delivering services is no longer powered. In summary, backup batteries are designed to accommodate numerous partial discharges but are oversized and configured such that full discharges are uncommon by design.

The goal of the proposed algorithm is to determine the SoH of a battery which can be used for battery replacement planning to ensure optimal uptime during the discharge. The existing methods for the SoH determination in literature can be divided into two categories [1]:

1. Determining the SoH in a laboratory by changing effective parameters, such as temperature, in a wide range.
2. Determining the SoH using the AC impedances and conductance measurements.

Pascoe and Anbuky have proposed a model for Valve Regulated Lead Acid (VRLA) batteries based on the discharge rate, ambient temperature, charge rate, initial state-of-charge (SoC), and SoH degradation [2]. This model was further developed in different operating conditions by Jossen [3]. These methods, and others similar to them are admittedly effective, but require external equipment, as well as the removal of batteries for testing, leaving the load unprotected. As an indirect process, two methods using the Alternating Current (AC) conductance and impedance have been proposed for assessing the SoH of VRLA batteries. These methods are less accurate and are most effective in identifying outliers, i.e., the failed battery in a battery bank [4, 5]. In short, there was no accurate method uncovered that could assess a battery's SoH without performing offline checks or using external equipment.

2.2. State-of-Health Algorithm

This paper describes a method to use partial discharge events to predict the SoH of a battery while the battery is in operation. By using a unique combination of measured and derived metrics that are collected only during the initial portion of the battery's discharge, the battery health can be predicted without relying on a total discharge of the battery. Using a thorough knowledge of chemistry and the electrochemistry occurring within the lead-acid battery, critical metrics were selected at very specific timestamps. The timestamps were selected to capture specific modes of known failure and degradation. The metrics of interest are listed below. It is important to note that the same metric taken at different timestamps can reflect different internal mechanisms within the lead-acid chemistry.

- Voltage
- First derivative of voltage
- Second derivative of voltage
- Change in voltage from optimal
- Change in voltage from charge (prior to discharge event)
- Recovery of voltage from discharge to charge (after discharge event)
- First and second derivatives of the recovery voltage

The figure below shows the voltage discharge curves of a set of VRLA batteries as captured by an actual power supply that has switched to battery back-up mode and is using these batteries for backup power. This set of voltage curves indicates a range of battery SoH, with one battery in particular much worse

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

than the others. This is not an unusual situation found in the field. An additional complication is that this site has two sets of batteries connected in parallel to increase the required runtime for that site. This figure shows the primary items of interest in relation to the voltage discharge curve that were considered in the analysis.

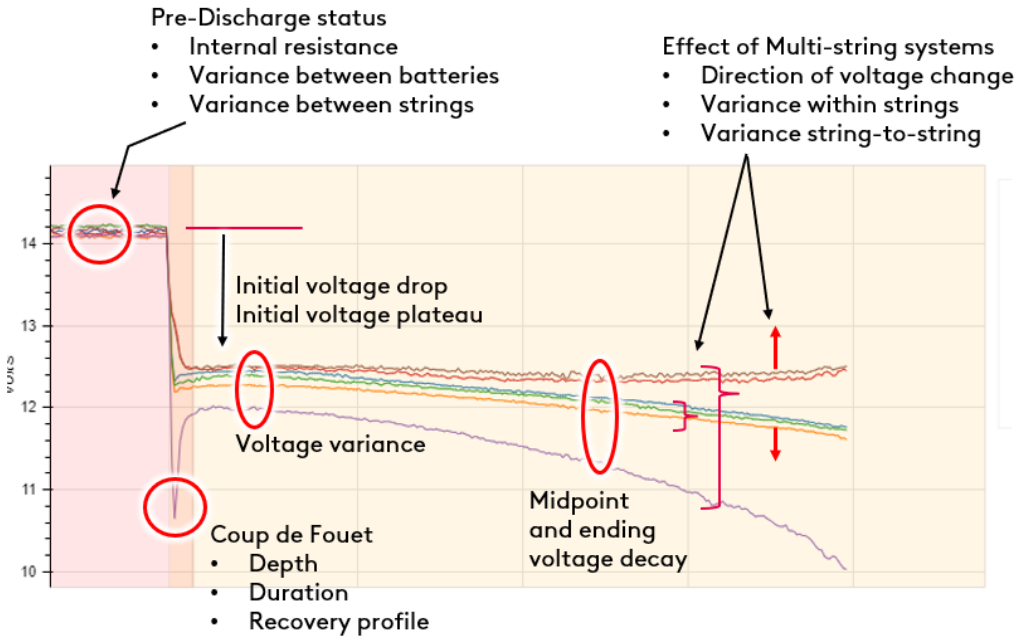


Figure 1 - Points of Interest in Voltage Discharge Curve

In addition to these metrics determined from the voltage curve, there are additional items useful to the predictive algorithm. This includes the following items.

- Temperature
- Number of batteries
- Output load and/or current
- Immediate vs. distant history
- Cycle history – depth, number, accumulated energy
- Charge status – time on recharge, time from last discharge, stability of battery voltage
- Battery information – manufacturer, model, age/install date

An important part of algorithm creation is the correlation of the metric to the electrochemistry of the battery. For example, the battery voltage drop due to the coup-de-fouet (the initial voltage drop seen in lead-acid batteries) is used to indicate the state-of-charge, electrolyte strength, and plate health, primarily the negative plate [6, 7]. Another example is the first derivative of the battery voltage due to the coup-de-fouet. This value is used to indicate the degradation of plate core health, primarily the negative plate and primarily due to cycling. Other metrics and their timestamps were selected because of their correlation to

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

other internal battery mechanisms, such as active material crystal size, positive grid corrosion, negative plate sulfation, and electrolyte purity [8, 9, 10].

A further consideration for this algorithm was the time scale. Discharge events in the field have traditionally been measured in time (hours and minutes) with a minimum runtime depending on the criticality of the downstream equipment. A conventional discharge curve is shown in Figure 2 below.

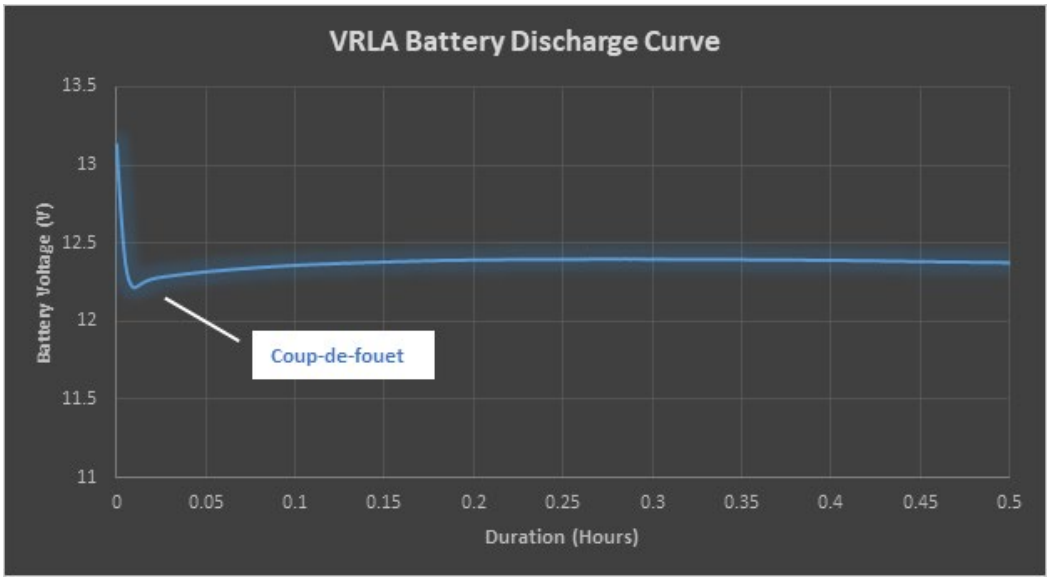


Figure 2 – Coup-de-fouet of a VRLA Battery

However, because the downstream load varies with every power supply, the time value will vary. That is, a five-hour discharge on a *lightly* loaded battery is not the same as a five-hour discharge on a *highly* loaded battery and the voltages are not comparable. To normalize the runtimes due to varying downstream loads, the battery discharge curves were plotted against 'Percent of Original Capacity Removed.' In this way, discharge curves and voltages were made to be directly comparable. An important implication of this is that when a discharge was made to a predetermined percent capacity, the discharge time would vary based on the downstream load on that power supply. In this way, the discharge times for each power supply are dynamic and fully time varying.

3. Data Collection

The dynamic deep cycle testing leverages the ANSI/SCTE 38-4 outside plant power supply management information base (MIB) for measurements. This long-established standard contains all the data points necessary to measure a deep cycle battery test.

The MIB exposes key power supply voltage and configuration information in the psDeviceTable. The MIB reports each battery's voltage in each battery string in the psBatteryTable. Each individual output current is reported in the psOutputTable. The power supply itself must instrument and report all these values.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The data collection interval is key for deep cycle testing. The practical limit for Simple Network Management Protocol (SNMP) data collection in an operational outside plant network is one sample per second. Polling the power supply data at slower than five minutes for these tests reduces the fidelity of the curve fitting and requires longer test periods which then become detrimental to battery health. Collecting the data at intervals between one second and two minutes per sample has been shown to provide sufficient accuracy for the curve fitting.

Another benefit of this dynamic testing is that high-rate data polling does not need to be sustained. Contemporary collection systems may already be capable of ingesting these data points at rates that can be used directly in the curve fits. If not, running higher speed collection for a brief period before the start and after the end of the deep cycle test is sufficient to collect the data necessary to analyze battery health. This allows operators to manage the overall amount of data the polling system retains.

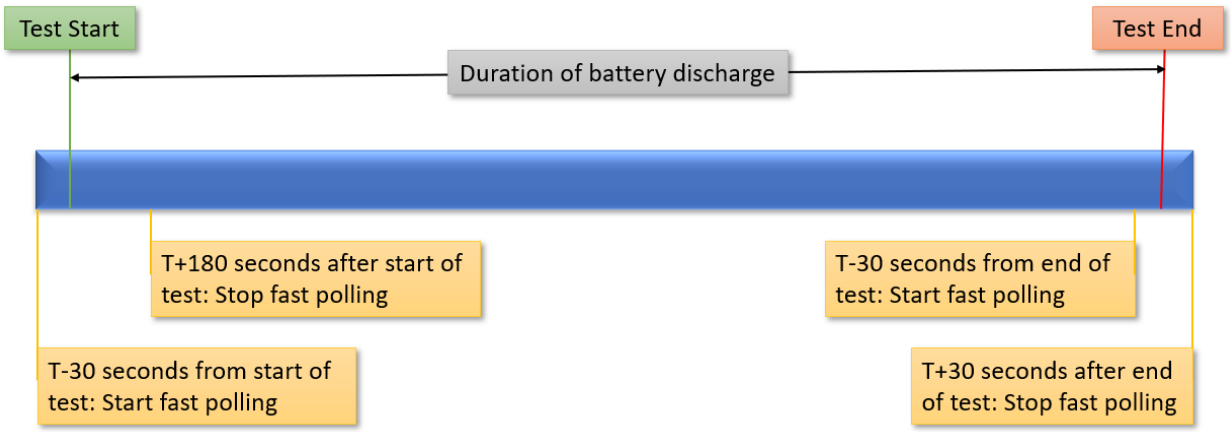


Figure 3 – Collection of Data by Time

By utilizing this well-known SCTE standard, operators can reuse much, if not all of their existing power supply data collection infrastructure to support the deep cycle battery testing analysis.

4. Results

As described in the earlier sections, there are a large combination of metrics that can be selected for a predictive algorithm. For example, the battery's voltage during the discharge is a single metric but can be measured in many ways. It can be recorded at numerous different times, which will represent different chemical actions within the battery. How it deviates from the norm, the magnitude of the change over time and the rate of that change over time can also be measured, as each of these can indicate a completely different mechanism within the battery. High-resolution discharge events were carried out in the lab to determine which metrics were most critical to a SoH prediction.

4.1. Training

A representative set of batteries was recorded under a controlled discharge in the lab for initial training development. The batteries had a SoH range of less than 25% to over 100%. As described previously, the discharges were normalized to a 'Percent of Original Capacity Removed' scale. The voltages were analyzed for the full discharge, down to 100% capacity removed. Additionally, a very high-resolution

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

discharge was run to highlight the voltage curves during the initial 5% of the discharge. Representative curves are shown in the figures below.

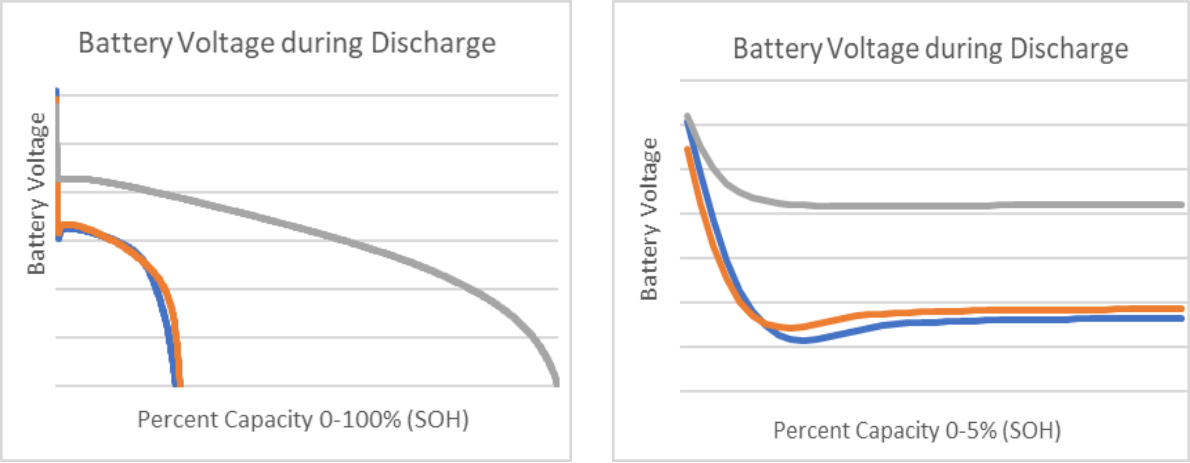


Figure 4 – Battery Voltages During Discharge Testing

Metrics were selected to differentiate the high performing batteries from the low performing batteries using these curves as a basis for 'good' vs. 'bad.' In particular, the voltage, the deviation of the voltage from an optimal value, its first derivative, and its second derivative, each at selected points throughout these discharge curves were selected. All derivatives were time based. After some review, ten total metrics were selected, and each were normalized to range from 0 to 1, so they would be mathematically comparable. They were then combined using a weighting factor based on experiential knowledge of the lead-acid battery chemistry. The final output of this algorithm provides a predicted state-of-health between 0% and 100% based on four major metric classifications: voltage, voltage deviation from optimal, first derivative of voltage, and second derivative of voltage.

[Normalized **voltage** at selected intervals between 0 and 25%] x [weighting factor(s)] +

[Normalized **voltage deviation from optimal** at selected intervals] x [weighting factor(s)] +

[Normalized **first derivative of voltage** at selected intervals] x [weighting factor(s)] +

[Normalized **second derivative of voltage** at selected interval(s)] x [weighting factor(s)] =

Percentage State-of-Health

4.2. Verification Runs

As a preliminary test of the algorithm, batteries from ten separate field locations were removed from their installations and brought to the lab. They were capacity discharged to determine their actual states-of-health (plotted as 'Actual' in the figure below). The algorithm was then run on the initial 25% of each discharge curve of these batteries to determine the prediction vs. the actual SoH (shown as 'Prediction' in the figure below). This comparison is shown in the table below and in the scatter plot as shown with the average difference between the predicted and actual less than 13%. Admittedly very small, these preliminary, non-optimized results are extremely promising. It is premature to apply an Analysis of

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Variance (ANOVA) analysis to these comparisons because of the sample size, but the f-statistic value and the P-value both strongly indicate that the two sets represent the same group.

Average % difference: 13%

Prediction	Actual
18%	26%
27%	20%
31%	19%
34%	35%
38%	44%
65%	31%
79%	100%
82%	78%
86%	100%
91%	91%

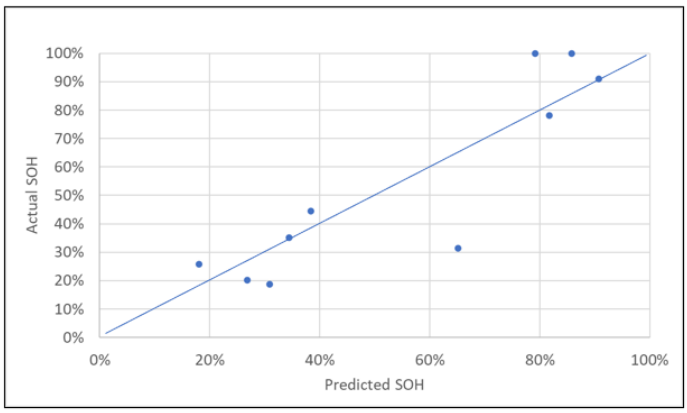


Figure 5 – Dataset from Verification Testing

Several of the battery predictions were well correlated to their actual battery health. However, others displayed prediction variations larger than desired. To better understand this, each separate metric was looked at individually. The ten metrics used are shown as calculated and normalized in the algorithm. When each metric was looked at individually, it is evident that some are better at differentiating between the battery's health condition than others. The differentiation can be affected somewhat by the constants originally selected for the calculation of that metric. A nominal effort was begun to optimize the algorithm by manually adjusting these constants. Also, the same effort was expanded to manually adjust the weighting factor of each metric and the normalizing factor of the compilation of all metrics. The optimization of this algorithm has progressed but is far from complete. This effort appears very promising and will continue in earnest in the next phase of this project when a much larger number of sample sets will be examined.

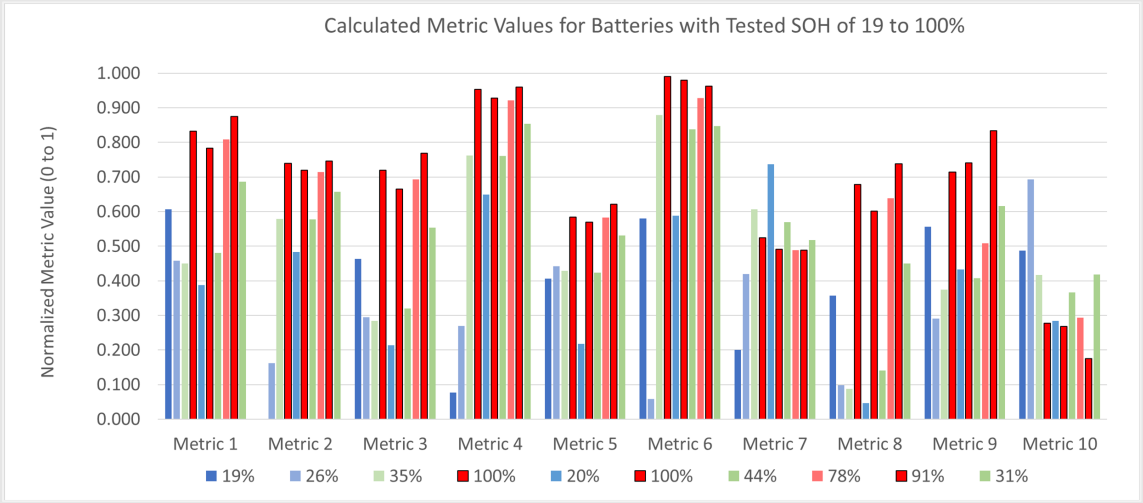


Figure 6 – Calculated Metric Values for Batteries Tested SOH

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Implementation

This algorithm can be implemented within the Power Supply Notebook (PSNB) so that each battery will generate a health score. The PSNB is an application that manages the inventory and tracks the status of Comcast's outside plant power supplies. There are over 250 thousand power supplies that are managed using the software. Power supplies provide telemetry each minute, reporting of the device status, and select parameters, all of which are displayed within the PSNB. Maintenance tickets can be generated from within the application based on the information available. Incorporating the results from the dynamic deep cycle test of the battery health score into the PSNB application ensures that battery replacements are managed effectively, and the correct power supply maintenance is performed.

Maintenance technicians utilize the PSNB to review power supplies' telemetry and to document maintenance work completed. Importantly to this project, it also allows remote testing of power supplies to ensure their effective operation. This feature will be used to implement a dynamic testing profile based on the algorithm of this work. Within this tool, battery health scores will be initiated, tracked, and recorded, all through the PSNB application. A detailed history of power supply discharge events is also available as a reference and verification method for the resulting battery health score. The real-world power supply performance history and battery health score are imperative in determining maintenance plans for battery replacement.

The ability to view current power supply status, historical events, and the results of the battery testing in one location is critical to the planning process. A poor battery health score will prioritize the power supply in the maintenance plan. The ease of creating maintenance plans is greatly increased by utilizing the battery health score as a key driver, and the ability to access the information for all power supplies in a single location is a key new feature. Additionally, this will enable the tracking of proactive maintenance progress and the monitoring of battery replacements and logged work within the PSNB.

6. Conclusion

The preliminary results presented of the dynamic duration deep cycle testing and algorithm to determine battery health have shown to be extremely promising. By normalizing the discharge data, the downstream load can be utilized to generate battery health values that can be compared directly, despite variations in the downstream loads. The current algorithm developed is applied to only the initial portion of the discharge and a battery health prediction is made. A key characteristic of this method is that only a partial discharge is necessary, and as such, the power supplies are never left unprotected due to a fully depleted battery. Using the Power Supply Notebook to implement the testing capabilities allows for complete integration of the power supply monitoring, maintenance tracking, and battery health ratings in one unified application. The PSNB also allows remote testing of any site with full data collection and access. The use of this functionality is expected to improve the reliability of the outside plant network and the effectiveness of the capital replacement program for the worst performing batteries. It is believed that a dedicated machine learning effort would significantly improve the accuracy of this model and is already planned as a future improvement.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AC	Alternating current
ANSI	American National Standards Institute
SCTE	Society of Cable Telecommunications Engineers
MIB	Management information base
ANOVA	Analysis of Variance
PSNB	Power Supply Notebook
SoC	State-of-charge
SNMP	Simple Network Management Protocol
SoH	State-of-health
VRLA	Valve Regulated Lead Acid

Bibliography & References

[1] M. Shahriari and M. Farrokhi, "State of health estimation of VRLA batteries using fuzzy logic," 2010 18th Iranian Conference on Electrical Engineering, Isfahan, Iran, 2010, pp. 629-634, doi: 10.1109/IRANIANCEE.2010.5506994.

[2] P. E. Pascoe and A. H. Anbuky, "A VRLA battery simulation model," Energy Conversion and Management vol. 45, pp. 1015-1041 2004.

[3] A. Jossen, "Fundamentals of battery dynamics," Journal of Power Sources, vol. 154, issue 2, pp. 530-538, 2006.

[4] I. Damlund, "Analysis and interpretation of AC-measurements on batteries used to assess state-of-health and capacity-condition," in Proc. 1995 IEEE Telecommunications and Energy Conference, The Hague Netherlands, vol. 17, pp. 828-833.

[5] D. C. Cox and R. Perez-Kite "Battery state of health monitoring combining conductance technology with other measurement parameters for real-time battery performance analysis," in Proc. 2000 IEEE Telecommunications and Energy Conference, Phoenix Arizona, USA, vol. 22, pp. 342-347.

[6] Pascoe, P.E.; Anbuky, A.H. The behavior of the coup de fouet of valve-regulated lead–acid batteries. J. Power Sources 2002, 111, 304–319.

[7] Yuan, S.-K.; Cheng, L. Estimation of SOH of battery based on Coup de fouet. Chin. LABAT Man 2018, 5, 65–68.

[8] Bode, H., (1977), *Lead-Acid Batteries*, Wiley-Interscience Publication.

[9] Linden, D., Handbook of Batteries 2nd Edition, McGraw-Hill, Inc.

[10] Berndt, D., (2003), *Maintenance-Free Batteries 3rd Edition*, Research Studies Press, Ltd.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Eating the Energy Elephant One Byte at A Time

Managing Energy Consumption Across the Cable Network

A Technical Paper prepared for SCTE by

Katie Flynn
Technical Project Manager, Energy Planning & Implementation
Comcast Cable
Philadelphia, PA
267.593.7385
Katie_Flynn2@Comcast.com

Budd Batchelder
Senior Director, Energy
Comcast Cable
Foxboro, MA
617.659.3686
Budd_Batchelder@Comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Starting an Energy Management System ... 3
   - 2.1. Defining an Energy Management System ... 3
   - 2.2. Establishing Need for an Energy Management System ... 4
   - 2.3. Top-down Commitment ... 4
3. Understanding the Elephant ... 5
   - 3.1. Determining How to Measure ... 5
   - 3.2. Assessing Performance ... 7
     - 3.2.1. Baselines & Benchmarking ... 7
     - 3.2.2. Energy Efficiency ... 7
     - 3.2.3. Energy Consumption and Expense ... 8
4. Tackling the Elephant ... 8
   - 4.1. Setting Goals ... 8
   - 4.2. Identifying Opportunities and Prioritizing Energy Conservation Measures ... 8
   - 4.3. Executing Efficiency through Empowerment ... 10
   - 4.4. Setting Energy Targets and Action Plans ... 11
5. It Never Ends: Continuous Improvement ... 11

Abbreviations ... 12

Bibliography & References ... 12

List of Figures

Title **Page Number**

Figure 1 - Visualization of EMS Elements Driving One Another ... 4
Figure 2 - SCTE Power Pyramid (Source: SCTE) ... 6
Figure 3 - Energy Conservation Measure Tracking ... 9
Figure 4 - Energy Accountability through Collaboration ... 10
Figure 5 - Evolution of Energy Management Program ... 12

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Cable Operators consume energy in unique ways relative to other industries, making a structured, data-driven approach to energy management a requirement to meet Carbon Neutrality and Network Energy Efficiency goals in an optimized manner. Managing the unique energy footprint of a network can often feel like a massive challenge (like the Desmond Tutu metaphor of eating an elephant) but implementing an energy management system (EMS) that can evolve over time makes controlling network energy possible. The foundation for an energy management program is data. Energy consumption must be measured at actionable levels so that performance can be quantified.

Comcast is implementing a data-driven EMS like never before. In 2022, Comcast set a goal to double network energy efficiency (as measured by electricity per consumed byte) by 2030. To advance on this goal, we have developed a holistic data collection system that quantifies energy consumption at the asset level (i.e., facilities and the outside plant, measured at power supplies). We are empowering our stakeholders to control their energy consumption by providing self-service dashboards to visualize progress and performance.

With the right tools in place to measure and evaluate performance, we now tackle the challenge of eating the energy elephant, one byte at a time. Key influencers of energy consumption are identified and engaged to drive opportunities to execution. Teams can validate and prioritize energy opportunities based on potential impact and available resources. Execution of energy conservation measures is verified. Collectively, these activities are a systematic implementation of the traditional plan, do, check, act continuous improvement cycle. This paper and presentation discuss the evolution of Comcast's energy management program from grassroots, pilot-style ideation to network-wide embedded effort where our team of experts, field personnel, and management collectively make tackling the overall, elephant-sized challenge manageable.

2. Starting an Energy Management System

Starting to develop an EMS can feel overwhelming and there are many aspects to consider. The following steps help to start to define and justify creating an EMS.

2.1. Defining an Energy Management System

When people hear the terms "energy management system" or "energy performance program," images of an online dashboard of internal energy statistics and reports often come to mind. However, a formal EMS is much more than a single analysis or project. The International Organization for Standardization (ISO) 50001 Energy Management Standard states:

> *Development and implementation of an energy management system includes an energy policy, objectives, energy targets, and action plans related to its energy efficiency, energy use, and energy consumption while meeting applicable legal requirements and other requirements. An energy management system enables an organization to set and achieve objectives and energy targets, to take actions as needed to improve its energy performance, and to demonstrate the conformity of its system to the requirements of this document.*

An EMS should include all the above elements and should take a holistic, data-driven, and systematic approach that is aligned with an operator's overall strategy. Every company's EMS will be unique to its own goals, opportunities, and operations that continue to evolve.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.2. Establishing Need for an Energy Management System

Many companies implement energy efficiency projects without establishing a formal EMS. The need for one should be articulated in the context of efforts to date to make the difference between efficiency projects and an EMS clear. Taking a project-based approach to energy conservation and efficiency may result in energy reductions but will not ensure that investments in efficiency are strategically executed to maximize impacts to reduce energy consumption, expense, and emissions.

An EMS is required to drive a culture of accountability around energy performance and continuous improvement. A data-driven, systematic approach must be taken to focus efforts where the greatest impacts can be made and ensure that results are achieved.

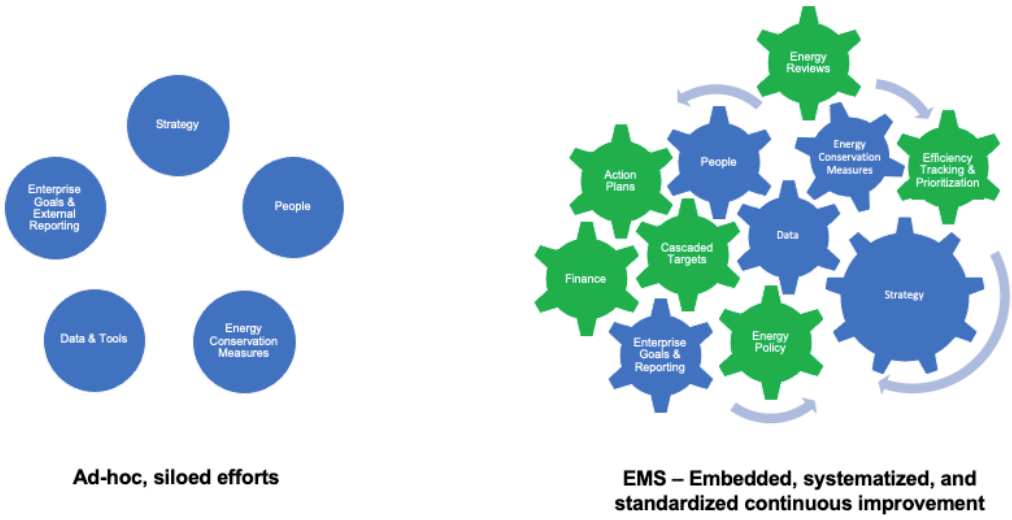


Figure 1 - Visualization of EMS Elements Driving One Another

Comcast set a goal to be carbon neutral in our scope 1 and 2 emissions by 2035. There are many paths to becoming carbon neutral, but those that do not include energy efficiency and performance as part of the strategy will be more expensive and miss opportunities to unlock additional internal value, such as improved reliability.

Communicating the business value of having a formal EMS can be difficult as a program drives, optimizes, and assures return for investments in energy performance and is dependent on the available opportunity and support. To develop a business case for a program, review efforts to date, adoption rates of best operational practices, and potential opportunities. A sense of the financial value of a program can be derived by looking at the opportunity of driving widespread adoption of energy conservation measures (ECMs) and energy efficiency measures (EEMs). Begin by mapping out deliverables for the first few years of a program to get a sense of the need of the program and show what the program will look like.

2.3. Top-down Commitment

Communicating a top-down commitment to treat energy as a controllable consumable by measuring and improving energy performance is an important starting point for an EMS. Energy consumption and expense have been traditionally thought of as overhead. Eliminating this preconceived notion is critical to the success of an EMS.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

An internal energy policy documents the top-down commitment to improving energy performance. Many companies have concise policies to commit to continually improving energy performance and strategy. ENERGY STAR has found that successful energy policies:

- Set an objective
- Establish accountability
- Ensure continuous improvement
- Promote goals

A policy must clearly state the objectives of energy management and how it is aligned with the overall organization. Examples of energy management objectives include improving efficiency, embedding consideration for energy impacts in decision-making, and reducing carbon emissions associated with energy consumption. Establishing accountability around energy management is critical as many areas have an impact on overall performance. Clear roles and responsibilities empower teams to drive an energy management program. Committing to continuous improvement is necessary to ensure that energy performance is optimized in operations and considered in long-term strategy. The policy should connect to specific energy, environmental, or overall financial goals of the company to demonstrate how energy management supports other organizational objectives. The policy should be shared with all key stakeholders and made readily available within an organization.

3. Understanding the Elephant

Once the commitment and the preliminary justification of creating an EMS are received, an energy review should be conducted to start to understand the energy footprint.

3.1. Determining How to Measure

Determining the current state of tools and data is a key first step in quantifying energy consumption, expense, and performance. Where data is and is not available should be documented with consideration for the following:

- Types of energy consumed: electricity, natural gas, gasoline, diesel, etc.
- Where energy is used: outside plant, critical facilities, data centers, non-technical facilities, and fleet
- Volume of energy consumed, associated expense, and other variables such as greenhouse gas emissions
- History of available data

A plan should be made to address gaps and estimates starting with the most impactful areas. Ideally, an organization should have a central, single source of truth for tracking energy consumption and expense and should be made readily available to all stakeholders.

Unlike other industries, where energy is primarily consumed in metered buildings, an outside plant is the largest end use of energy for cable operators. Many cable operators have unmetered power supplies in the outside plant which can challenge quantifying and tracking energy consumption. Power supply telemetry can be leveraged where available and is becoming more prevalent and sophisticated. The number of electric accounts associated with the outside plant can also make tracking difficult.

After mapping out the availability of energy data, energy consumption and expense should be calculated for the operator and broken out into significant energy users (SEUs). SEUs can be unique for each cable operator and describe facilities by type, systems, processes, or equipment. An operator's energy footprint is complex and spread out over retail sites, data centers, inside plants, and outside plant. Company

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

vehicles can also be a significant consumer of energy for an operator. Creating a program that encompasses the entire network may seem daunting. Not every end use of energy needs to be in the scope of an EMS at the start. Focus on the largest energy consumers and areas of opportunity first.

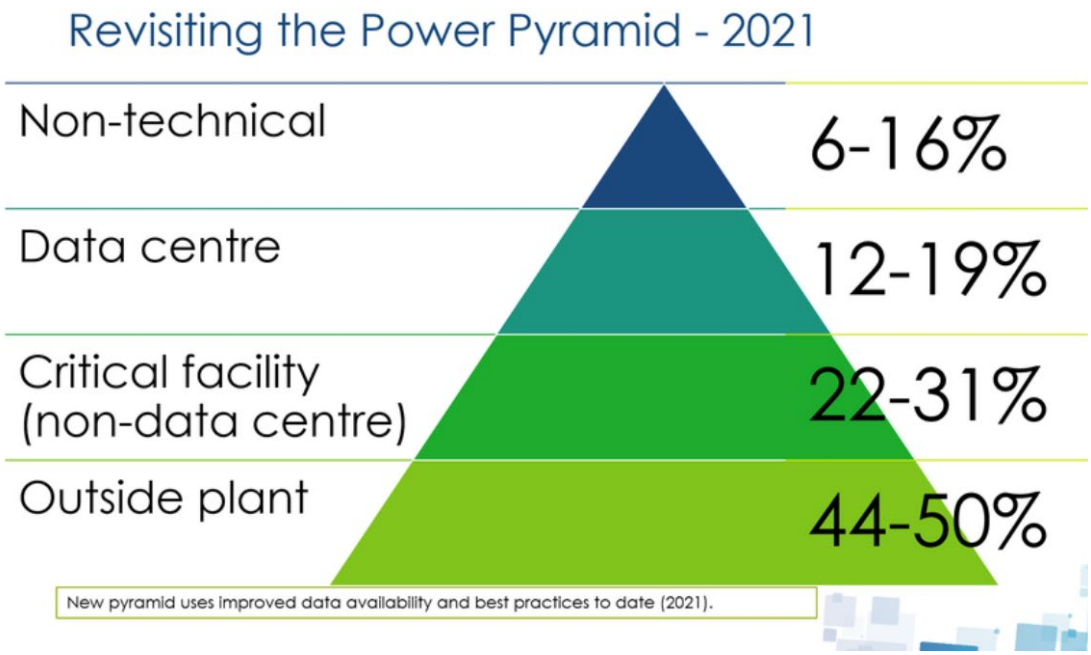


Figure 2 - SCTE Power Pyramid (Source: SCTE)

An operator could choose to review energy consumption by device types, such as cooling systems, power supplies, or IT equipment, or by type of site, such as data centers, headends and hubs, and outside plant. SEUs should align with company structure and how different SEU efforts will be focused. Data should be estimated using documented, repeatable, and logical methodology where data is not available. Expense and consumption data should both be analyzed. Electricity prices can vary by SEU with outside plant electricity rates often higher than critical infrastructure facility rates. Energy rates, especially electricity, can often vary dramatically by geographic location and over time.

Determining where energy is used across the network and how expense varies is key in informing where to prioritize efforts. An informed energy management program focuses efforts on the most impactful areas and can be leveraged to evaluate opportunities more quickly. For example, knowing the range in outside plant electricity rates and how much energy is consumed by power supplies can enable an energy manager to determine the impact of improving the efficiency of power supplies and develop a business case.

Historical data can be leveraged to further inform the energy understanding. Measuring how consumption has increased, decreased, or remained flat over time in different SEUs can provide feedback on the energy impact of different major initiatives.

A model of future network energy consumption should be developed to understand where energy use is expected to grow and shrink. Potential drivers of future energy consumption should be flagged to

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

prioritize energy opportunities that can reduce future energy consumption and for planning purposes, such as energy procurement.

3.2. Assessing Performance

Once the overall energy landscape is quantified, the next step is understanding energy performance and how well consumption, efficiency, and expense are being managed. Creating a baseline for measuring performance in these key areas enables the tracking of key metrics over time. Assessing performance is not a one-time activity and should be done on an ongoing basis with results being communicated to key stakeholders to validate efforts and reinforce that energy is a controllable consumable.

3.2.1. Baselines & Benchmarking

A baseline year should be created to measure energy performance over time. Reviewing performance over time will show trends and provide high-level feedback on the impact of current strategies on energy consumption.

Energy performance should be benchmarked internally, externally, and over time to inform the energy assessment. Where possible sites and networks should be compared to others internally to understand the range of energy efficiency and opportunities across the network.

Publicly available peer energy and sustainability report data can provide an external comparison and a sense of potential to improve performance.

3.2.2. Energy Efficiency

Measuring energy efficiency allows an operator to measure how effectively the company consumes energy over time in relation to the output of customer services while normalizing for growth. Energy efficiency metrics drive innovative thinking about how a network delivers customer bytes with fewer kilowatt-hours of electricity, which provides better visibility into performance than measuring strictly absolute consumption and expense metrics.

Efficiency can be measured at the site level in addition to the overall network level. ANSI/SCTE 213 2020 *Edge and Core Facilities Energy Metrics* defines Electricity per Consumed Byte (EPCB) for a facility. Additionally, SCTE is currently developing additional enterprise energy-related metrics. Leveraging industry-standard efficiency metrics enables organizations to benchmark their efficiency to peers knowing that metrics are consistently calculated.

By measuring efficiency across all sites with consistent metrics, direct comparisons can be made. Sites with high efficiency could be indicative of following best practices in terms of operations and equipment in place. Sites with high intensity or low efficiency could be indicative of candidates for deploying energy efficiency opportunities.

Relevant efficiency metrics for other areas of consumption, such as retail or office spaces, should also be developed. Consideration must be given to the behaviors that will be driven as a result of the leading metrics chosen. For example, Power Usage Effectiveness (PUE) is the ratio of total facility power to IT power and is often used as a measure of efficiency in data centers. Leveraging PUE may drive focus on reducing energy consumed for cooling rather than an overall energy conservation plan and may drive focus toward cooling system optimization rather than IT equipment decommissioning.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.2.3. Energy Consumption and Expense

Tracking both consumption and expense separately is important as the price of energy can vary over time and solely tracking consumption or expense may tell an incomplete story. Often energy expense is the main data point reviewed and if fluctuations in price are not taken into account, results of energy management efforts cannot be verified. For example, if an operator upgrades lighting at office buildings to LEDs over the same period that electricity rates increase, electricity expense stays flat post-upgrade and the project may appear to be unsuccessful. However, if the project had not been completed, electricity expense would have increased. This is why both energy variables need to be tracked and progress in each area quantified.

Additional data may be needed to contextualize energy performance over time. Homes passed, number of sites, and ambient temperature are all additional data points that can help to assess energy performance over time. Complimentary metrics and data can be included in energy performance reports, but leading metrics around efficiency, consumption, and expense data should drive overall performance assessments.

4. Tackling the Elephant

Once the size and current state of the elephant are known, plans need to be made for how to start to tackle and manage the energy elephant.

4.1. Setting Goals

Overarching goals for energy management should be created and explicitly shared. Energy goals should be aligned with sustainability or other related goals. Goals should be SMART- (specific, measurable, attainable, realistic, and timebound) and drive the focus of a program. For example, goals related to reducing costs will drive different results than goals focused on reducing absolute consumption.

Comcast set the goal to double our network efficiency by 2030 to ensure that efficiency is a key part of the strategy to progress on its carbon neutral goal and demonstrate that energy efficiency is a priority internally and externally. The mission of the Comcast energy management program is to optimize our processes, systems, and resources to maximize energy efficiency and support our goals to be Carbon Neutral by 2035 and Double Network Efficiency by 2030 in a fiscally responsible manner.

4.2. Identifying Opportunities and Prioritizing Energy Conservation Measures

Opportunities for reducing energy consumption, expense, and associated emissions should be considered on an ongoing basis. The rapid rate of equipment refreshes and innovation in the cable operating space means that new opportunities are always developing. A framework for capturing, comparing, and tracking potential ECMs and EEMs is necessary for prioritizing efforts based on their impact and alignment with other initiatives. ECMs should be tracked throughout their lifecycles. Tracking prioritized measures from ideation through adoption ensures continued progress and may highlight opportunities for automation.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 3 - Energy Conservation Measure Tracking

The first step in evaluating ECMs and EEMs is to document the business cases and measured savings for what has already been implemented. Efficiency projects and studies may have been done prior to establishing an energy program and should not be overlooked. Document the costs, time, and benefits of each project. Identify what worked well, areas for improvement, and lessons learned. Understanding what has worked in the past will help inform future efficiency project deployments and increase the likelihood of success.

Low-cost and low-effort ECMs and EEMs should not be overlooked. Implementing projects that target "low-hanging fruit" first has been the guidance for decades when starting on an energy management journey. All too often these opportunities are taken for granted as having already been implemented. Simple, best practices, such as standard temperature setpoints, timely decommissioning, and airflow management have proven returns and are included in internal standards. However, adherence to these best practices can be overestimated. When evaluating ECMs and EEMs, the adoption of best practices should be quantified where possible, so that "low-hanging fruit" is not overlooked. Additionally, low-cost and low-effort opportunities are a way to embed energy efficiency as a part of the culture around energy management.

Capital opportunities should be reviewed, and the business cases developed to complete the list of opportunities and begin to prioritize and plan. The business cases should be holistic and account for the overall impact on energy consumption, expense, and carbon emissions in addition to the capital and non-energy operation expense impact. Opportunities will have varying impacts across these business case variables which is why holistic business cases must be developed and variables prioritized based on a company's specific energy management objectives.

The ability to implement ECMs and EEMs will also be dependent on other factors including:

- Financial, people, tool, and other resource requirements
- Alignment with other priorities such as reliability, tech-refresh, network expansion, and branding
- Timeline and/or feasibility to deploy

The outcome of this prioritization should be a preliminary roadmap for how energy performance will be improved over time. Opportunities that are not ready for implementation should remain on the list and periodically reevaluated to assess feasibility.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.3. Executing Efficiency through Empowerment

The number of employees who are dedicated to energy management is small relative to the size of the organization. An energy management program needs to engage with key leaders, managers, site operators, and network technicians to gain traction to implement solutions and improve performance. An elephant cannot be eaten alone and a team approach must be formed.

An energy management program should map out the key influencers across the network to understand who makes decisions that impact the energy efficiency of existing and future networks. Identifying all energy influencers can feel overwhelming, but focusing on SEUs and who has the most influence can help. While key influencer business units will vary and be dependent on an individual company's structure, typical areas that are key influencers of energy performance include:

- Finance
- Engineering
- Operations
- Sustainability
- Design & Construction
- Procurement

When energy influencers have been identified, a framework for engagement needs to be created to ensure that opportunities are evaluated, action plans are created, and performance is reviewed compared to targets. The value of managing energy proactively needs to be communicated. Public commitments around energy and carbon can be leveraged as a proof point of a company's charge to improve energy performance. Internal education and engagement programs are needed to show employees how their work connects and impacts goals.



Figure 4 - Energy Accountability through Collaboration

No one person is entirely responsible for the energy consumption and efficiency of the overall network or an individual site. Many stakeholders make decisions that impact energy performance, from what type of equipment to install, to the setpoint for an air conditioning system. Accountability around energy performance must be shared through educating influencers on how their role impacts energy performance, empowering them to act, and providing feedback on energy performance. Energy consumption forecast

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

models should be developed and shared with influencers who impact future energy consumption and highlight the energy implications of decisions.

4.4. Setting Energy Targets and Action Plans

Energy action plans must be created to ensure funding, secure required resources, and develop a timeline for implementing prioritized energy opportunities. The plans must be approved by the appropriate leaders and communicated.

Energy targets must be developed to hold people accountable for energy performance. Target development should be informed by the energy consumption forecast model and energy action plans. Annual targets should align with overall energy goals. At Comcast, we are creating site targets that are aligned with our overarching goal to double network efficiency by 2030.

Creating targets at an actionable level is important so that people can be held accountable for performance within their control. Energy targets reinforce the priority of implementing the energy action plan and highlight where energy performance is not as expected. If energy targets are not achieved, the root cause of the energy performance variation should be identified and corrected if possible. Variables such as weather can impact actual performance and some allowances should be made where reasonable and out of the accountable influencer's control.

Energy performance should be measured against targets and regularly communicated to appropriate energy influencers. A monthly review provides an appropriate cadence to evaluate performance and take corrective action if necessary. Depending on the scale and timing around the implementation of ECMs or EEMs, a separate measurement and verification plan may be needed in addition to measuring performance against targets.

5. It Never Ends: Continuous Improvement

Energy management is not a program that is completed in a year. Once the initial elements are stood up, the energy management program follows a plan, do, check, act continuous improvement cycle. Each element of the program is critical to systematize and move the program forward with the result being improved energy performance.

Over time, efficiency should be embedded into the organization to create a culture of responsible energy use. Tools should be leveraged to reduce manual work and intervene as much as possible. The more energy management becomes embedded within an operator's ways of working, the more optimized the network will become over time. The energy management program will not feel like an additional effort, rather it will become the business as usual.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 5 - Evolution of Energy Management Program

As progress is made, new opportunities should be considered to continue to optimize energy performance and challenge the organization to improve. If the network becomes more innovative there will be new opportunities to improve performance.

Abbreviations

ECM	energy conservation measure
EEM	energy efficiency measure
EMS	energy management system
PUE	Power usage effectiveness

Bibliography & References

ISO 50001: *Energy management systems – Requirements with guidance for use 2nd Edition 2018-08*

SCTE ENERGY 20/20 *Revisiting the Power Pyramid - 2021*

ENERGY STAR *Guidelines for Energy Management*

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Enabling Transport Network Resiliency with Unified Access Transport

A Technical Paper prepared for SCTE by

Ivan Bogdanovic
Manager, Transport and IP Access Network Engineering
Rogers Communications Inc.
8200 Dixie Road, Brampton ON L6T 0C1
416-508-9109
ivan.bogdanovic@rci.rogers.com

Giancarlo Urbani
Director, Cable Access and Transport Network Engineering
Rogers Communications Inc.
8200 Dixie Road, Brampton ON L6T 0C1
416-457-1099
giancarlo.urbani@rci.rogers.com

Jordan Tontini, Principal Engineer, Rogers Communications Inc.

Alex Talinga, Principal Optical Transport Engineer, Rogers Communications Inc.

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. What is Unified Access Transport? 4
3. UAT Drivers 5
   - 3.1. Network Simplification and Resiliency 5
   - 3.2. Wireline Network Modernization and Uplift 6
   - 3.3. Wireless Network Modernization and Uplift 7
   - 3.4. Service Expansion 7
4. Engineering Design Considerations for UAT 8
   - 4.1. Topology Options 8
   - 4.2. Active Optical Transport 10
   - 4.3. Passive Optical Transport 12
     - 4.3.1. Passive DWDM Multiplexers in Secondary UAT Rings and Optical Spurs 12
     - 4.3.2. Passive DWDM Multiplexers in Primary UAT Collector Rings 14
   - 4.4. Optical Transceivers in UAT 16
     - 4.4.1. Optical Transceivers in Secondary UAT Rings and Optical Spurs 16
     - 4.4.2. Optical Transceivers in Primary UAT Collector Rings 17
5. What's Next for UAT? 18
6. Conclusion 23

Abbreviations 23

Bibliography & References 24

List of Figures

Title **Page Number**

Figure 1 – Unified Access Transport Network 4
Figure 2 – Network Topology Before and After UAT Transformation 5
Figure 3 – Network Resiliency Before and After UAT Transformation 6
Figure 4 – Mobile Transport Capacity Uplift and Latency Improvement 7
Figure 5 – Legacy FOADM to ROADM UAT Uplift in Service Expansion Areas 8
Figure 6 – UAT Ring Physical Topology 9
Figure 7 – Hub-and-Spoke Topology 9
Figure 8 – Hop-by-Hop Topology 10
Figure 9 – Protected Bi-Directional Single-Fiber System 12
Figure 10 – Unprotected Bi-Directional Single-Fiber System 12
Figure 11 – Cascaded Bi-Directional Single-Fiber Configuration 13
Figure 12 – Hub-and-Spoke Bi-Directional Single-Fiber Configuration 13
Figure 13 – Channel Allocation for Passive Multiplexers Used in the UAT Collector Rings 14
Figure 14 – Greenfield Deployment with 1310+1550 nm Passive Coupler 15
Figure 15 – Brownfield DWDM-over-CWDM Upgrade 15
Figure 16 – Brownfield ROADM-over-cWDM Upgrade 16
Figure 17 – Relative Cost per Port Comparison for the Primary UAT Rings 18
Figure 18 – UAT Plan of Record 19
Figure 19 – OTDR on a Plug [1] 20

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 20 – Coherent Transceivers in UAT [2] 21
Figure 21 – EDFA on a Plug 22
Figure 22 – Coherent Termination Device 22

List of Tables

Title **Page Number**

Table 1 – Metro versus Long-Haul ROADM Line System Comparison 11
Table 2 – Optical Transceivers in Secondary UAT Rings and Optical Spurs 17

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

This paper introduces the concept of Unified Access Transport (UAT), a common access optical transport network for Residential, Business and Wireless services at Rogers Communications. The paper not only presents UAT but also puts forth recommendations for its effective deployment.

Unified Access Transport is a key enabler to Rogers Communications' network modernization and expansion programs. UAT enables core consolidation, Hybrid Fiber-coaxial (HFC) and 5G modernization, as well as Wireless and Wireline service expansions into underserved areas.

Within the subsequent sections of this paper, design considerations for urban and rural deployments are explored with a primary emphasis on bolstering network resiliency and scalability. Network enhancements such as automated optical protection switching to minimize customer impact, and automated Optical Time Domain Reflectometer (OTDR) to reduce Mean Time-to-Repair (MTTR) are examined. Network deployment accelerators such as Reconfigurable Optical Add-Drop Multiplexer (ROADM) over existing passive multiplexers and optical disaggregation are also discussed.

This paper provides insight into lessons learned from deployments to date and previews what's next for UAT.

2. What is Unified Access Transport?

UAT is a common access optical transport network that transports aggregated traffic from homes, businesses and wireless cell sites to the core network. It connects network elements in the field – Remote PHY Device (RPD), Passive Optical Network (PON) Optical Line Terminals (OLT), Enterprise Network Interface Devices (NID) and Radio Base Stations (RBS) – to core locations via common optical multiplexers and Points of Presence (POPs) as outlined in Figure 1.

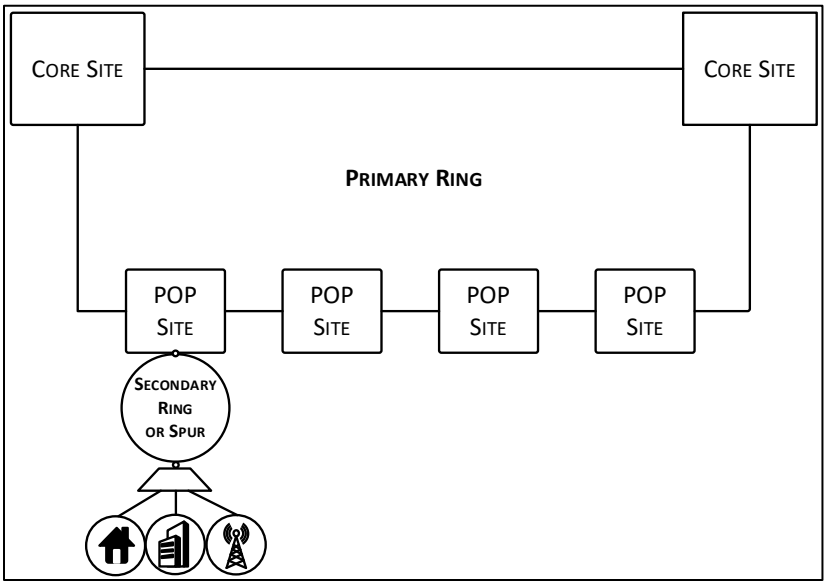


Figure 1 – Unified Access Transport Network

UAT is composed of primary collector rings connecting POP sites to Core sites and secondary rings or spurs connecting devices in the field to the POP. Optical transport for the primary collector rings uses

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

ROADM line systems or dual-fiber Dense Wavelength Division Multiplexing (DWDM) couplers. Secondary access rings of UAT are comprised of single-fiber DWDM multiplexers in the field and POPs. POP site optical switches and field splitter/couplers are used to protect the DWDM multiplexer trunks.

At Rogers Communications, UAT complements Unified Fiber Plan (UFP), a holistic methodology for planning, designing and building fiber links that optimizes fiber routing and dimensioning to connect all sites of interest. It optimizes the fiber use in the network. UFP is outside of the scope of this paper.

3. UAT Drivers

The following sub-sections discuss four key drivers for the UAT rollout.

3.1. Network Simplification and Resiliency

The first driver is network simplification and resiliency. Figure 2 illustrates the network before and after the UAT transformation.

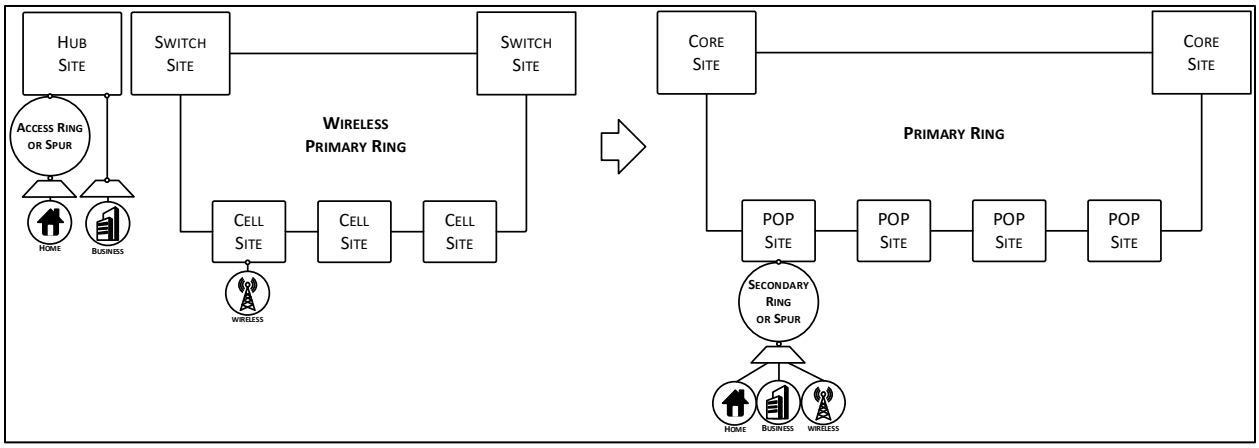


Figure 2 – Network Topology Before and After UAT Transformation

Prior to UAT introduction, Rogers Wireline, Enterprise and Wireless access networks used different topologies, did not share fiber and often did not share the core facilities. Although both Wireline and Enterprise access networks were based on a hub-and-spoke topology where a device in the field has a dedicated optical link to the core node located at the hub site, services did not share fiber. Wireline RPD and 10 Gigabit Symmetrical PON (XGS-PON) traffic traversed single-fiber DWDM multiplexers while Enterprise IP network and Layer 1 wave services traversed dual-fiber Coarse Wavelength Division Multiplexing (CWDM) multiplexers. Note that all wireline and enterprise services from the large geographical area terminated at the same centralized core locations and as such were exposed to catastrophic facility failure.

Rogers Wireless IP Radio Access Network (IPRAN) mobile transport network topology is much more distributed. It is based on a set of collector aggregation rings connecting IPRAN cell sites deeper in the access plant to two geographically diverse wireless switch sites. As such, wireless service, connected via compliant topology, is not exposed to catastrophic failure at the switch site facility. Wireless aggregation routers at IPRAN cell sites are connected via dedicated dual-fiber CWDM multiplexers in urban areas or legacy Fixed Optical Add-Drop Multiplexer (FOADM) based networks in more rural areas. As facilities housing IPRAN core routers are often not collocated with wireline primary hubs, wireless CWDM multiplexers and ROADM optical networks were rarely shared with other services. Furthermore, traffic

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

from most of the wireless sites connects to wireless aggregation routers at the IPRAN cell sites via pair of dedicated linear fiber strands making for the inefficient use of fiber resources.

To effectively utilize fiber resources, UAT first had to align different network topologies so that all services shared common add and drop locations. Aligning on a common set of UAT POPs and core sites, facilitates the use of common passive and active optical systems which improves fiber utilization. Adopting a distributed topology where traffic is aggregated closer to the edge of the network further improves fiber utilization and enhances network resiliency against the fiber cuts and core facility outages. The concept is illustrated in Figure 3 where a recent headend relocation and HFC modernization project uplifted approximately 100,000 homes passed (HP) area to the UAT architecture. The project deployed a UAT primary ring with four UAT POPs that aggregate traffic deep in the access plant, eliminating fiber segments greater than 2,000 linearly fed customers. Wireless locations were leveraged for UAT POPs drastically reducing cost and implementation time. Rather than establishing a new hub site location, Converged Cable Access Platform (CCAP) and Broadband Network Gateway (BNG) devices were split between the existing two hub site locations connected to the UAT ring allowing for significant Capex and Opex savings. As such, UAT provides the opportunity to reduce the Rogers fiber diversity threshold to less than 2,000 customers, instead of fiber architecture augmentation alone.

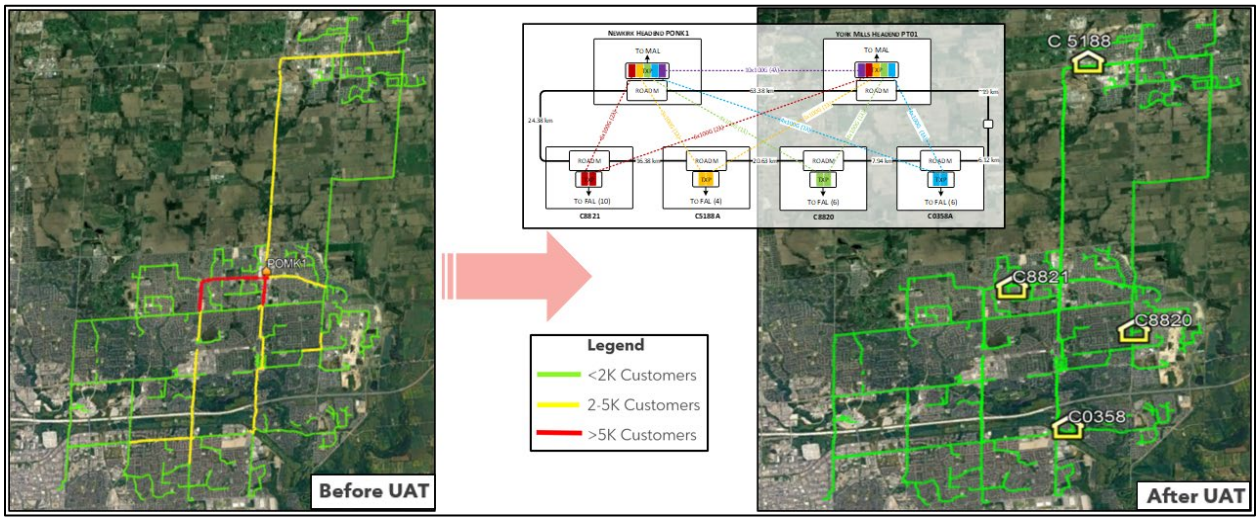


Figure 3 – Network Resiliency Before and After UAT Transformation

3.2. Wireline Network Modernization and Uplift

The second driver is wireline network modernization and uplift of HFC networks to support Data Over Cable Service Interface Specification (DOCSIS) 4.0 and Fiber to the Home (FTTH) uplift in select parts of the wireline network. An essential part of the HFC modernization is migration to Distributed Access Architecture (DAA) where analog optical nodes are replaced with digital optical RPD nodes. Rogers Communication IP Wireline Access Network (IPWAN) enables wireline service aggregation at the UAT POPs further improving fiber utilization and enhances network resiliency by enabling RPD and OLT connectivity to core devices at geographically diverse locations. Longer distances associated with digital optics on RPDs are further extended with active optical transport at UAT POPs. Together with virtual CMTS and DAA, UAT enables core node consolidation by up to 50%, that results in a favorable estimated 10-year total cost of ownership.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.3. Wireless Network Modernization and Uplift

The third driver is modernization of the wireless access network, which introduces new Base Band Units (BBU), Remote Radio Units (RRU) and antennas that support multiple frequency bands, higher order carrier aggregation combinations and Multiple Input Multiple Output (MIMO) technologies. This along with the rollout of newer wireless terminal devices significantly increases demand per customer. Higher traffic demand is further amplified by the introduction of the new 5G frequency bands under the wireless uplift program. Wider channels and even higher order MIMO antennas in mid and millimeter frequency bands further increase traffic demand and put a strain on mobile transport network.

To meet the increased demand, UAT enables metro and regional collector ring upgrades from 10 Gigabit per Second (Gbps) to 100 Gbps. Capacity per site and latency improvements are particularly noticeable in urban areas where wireless traffic terminates at the UAT POPs with ROADM based optical transport.

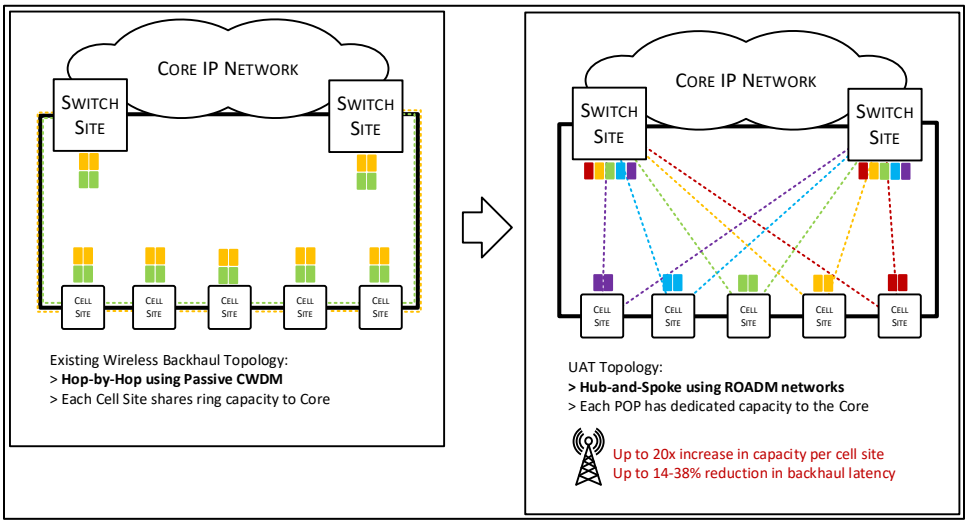


Figure 4 – Mobile Transport Capacity Uplift and Latency Improvement

To meet the increased customer demand in urban areas and improve coverage in the areas where macro towers are not permitted, Rogers Communications is densifying networks by deploying small cells where RRUs mounted on street furniture are connected to BBU hotels at macro sites via passive DWDM multiplexers.

3.4. Service Expansion

The fourth driver is fiber network expansion to out of footprint greenfield and brownfield underserved communities. Expanding wireline XGS-PON and Fixed Wireless Access (FWA) services to rural underserved areas outside of traditional cable footprint requires active optical transport to backhaul services to the existing core locations. Meeting aggressive occupancy timelines from government subsidy programs is critical. Uplifting legacy 10 Gbps FOADM-based optical transport networks used to backhaul mobile traffic in rural areas to high-capacity ROADM based coherent systems enables faster service turnup across wide areas of opportunity. The concept is illustrated in Figure 3 where uplifting existing optical systems used to backhaul IPRAN reduced new fiber construction to less than 10% of the total fiber distance used for the primary rings.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

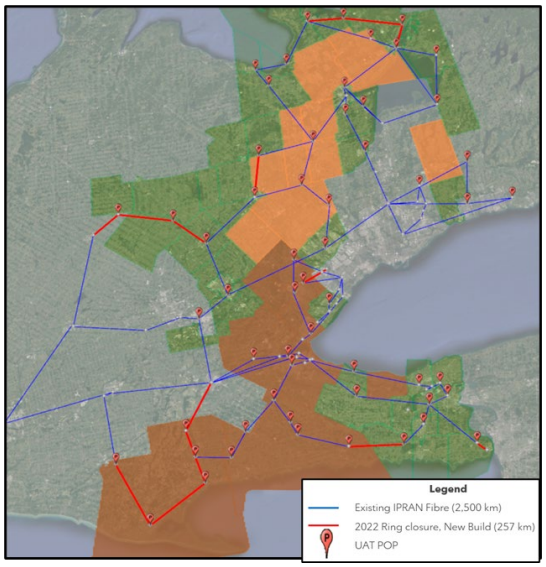


Figure 5 – Legacy FOADM to ROADM UAT Uplift in Service Expansion Areas

Uplifting legacy optical systems to UAT, provides backhaul capacity not only for the newly connected XGS-PON and FWA customers but also uplifts the transport to meet the customer demand for newly added 5G bands in rural areas. Furthermore, in the absence of cable headend locations in rural areas, uplifting existing wireless shelters to UAT POPs that house XGS-PON OLTs and IPWAN aggregation routers, proved to be more cost effective and less time-consuming than setting up standalone wireline facilities and connecting them to optical rings.

4. Engineering Design Considerations for UAT

The following sections review engineering design standards taken into consideration when building the UAT network.

4.1. Topology Options

The physical topology of a UAT ring is characterized as a classical ring. POP sites are daisy-chained together by fiber plant to form a primary ring. The fiber plant must be diverse, so a single fault cannot interrupt multiple spans and isolate POP sites. The primary ring terminates at two geographically diverse core sites. To close the ring, the core sites are connected via a backbone fiber system as shown in Figure 6. The backbone and primary ring must also be fiber diverse. A single backbone can support multiple primary rings so long as the fiber diversity criteria is met.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

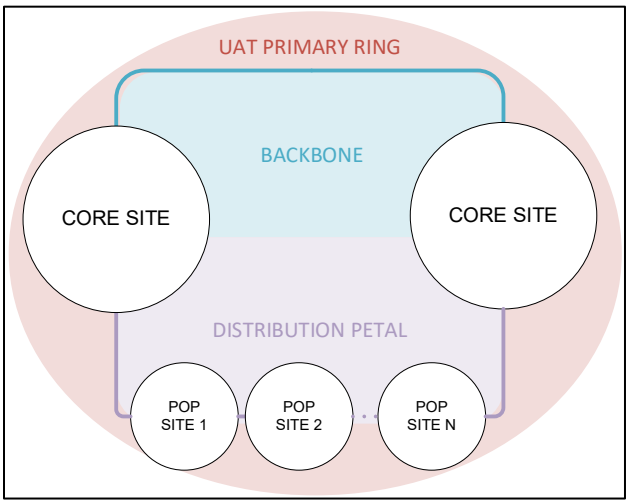


Figure 6 – UAT Ring Physical Topology

After a new primary ring is physically established, an IP system planner has two options to connect each POP site to the two core sites:

1) Hub-and-Spoke Topology
2) Hop-by-Hop Topology

Hub-and-Spoke topology in Figure 7 connects each POP site with dedicated capacity to the core sites. This is commonly achieved with dedicated wavelength circuits over a ROADM line system. Hub-and-spoke topology provides scalable high capacity to each POP site. It is best suited for high to medium density metro deployments and/or high growth deployments.

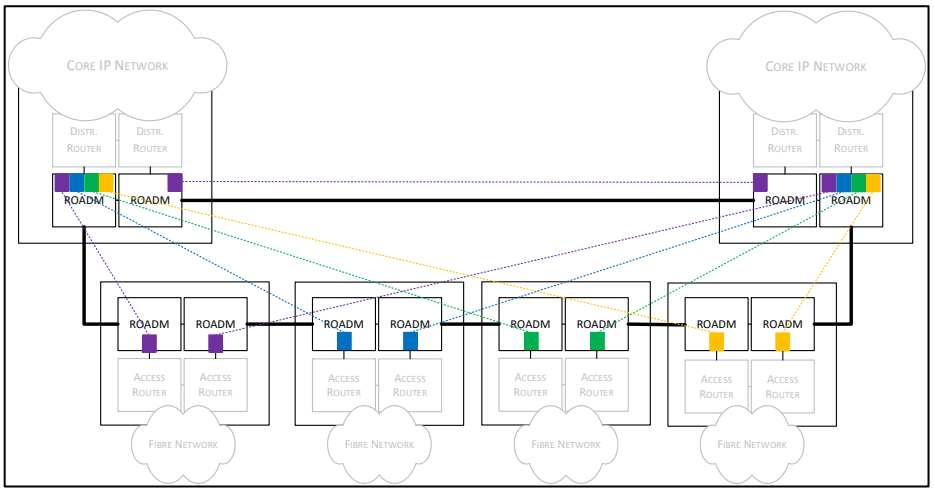


Figure 7 – Hub-and-Spoke Topology

Hop-by-Hop topology in Figure 8 connects each POP site with shared capacity to the core sites. The capacity is shared amongst all POP sites in the primary ring. Fiber span distance is the primary determining factor on whether to deploy Hop-by-Hop using active or passive transport; more on this in the following section. Hop-by-Hop topology is best suited as a cost-effective alternative for low density, low growth deployments.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

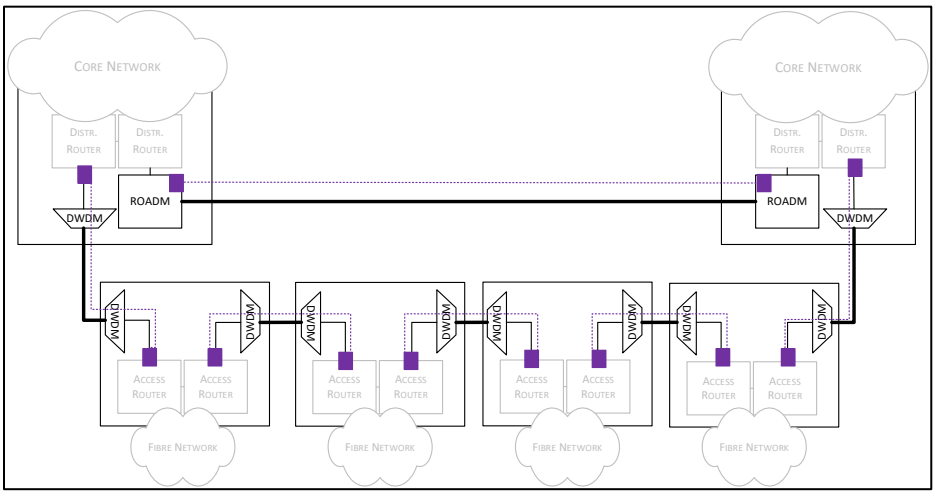


Figure 8 – Hop-by-Hop Topology

4.2. Active Optical Transport

For high-capacity metro UAT applications or distances that exceed the optical budget of a passive design, an active DWDM system is required to provide transport support between UAT POPs and core locations. The active system should be optimized, in features and in cost, for access applications.

While long-haul DWDM systems are designed to maximize utilization of scarce inter-city fiber resources, metro DWDM systems should be designed to minimize equipment cost as fiber is usually more readily available. Such cost optimization is achieved by using hardware specialized for the metro-regional applications, while maintaining the resilience and operational features deployed in long haul systems.

Transponders used for the long-haul applications are usually based on proprietary fixed lasers and optoelectronics, tailored for maximum capacity and reach performance. Conversely, metro-regional transponders typically use more cost-effective pluggable line side DWDM optics that offer adequate performance for the application while allowing operators to benefit from continuous cost and performance improvements in the pluggable coherent transceivers.

Flexible grid support in long-haul DWDM systems is required to ensure compatibility of the DWDM photonic layer with the future generation of transponders, that will use higher baud rate optics allowing capacity per wavelength exceeding 1.2 Terabits per second (Tbps). The ROADMs are usually equipped with expensive optoelectronics including Route and Select (R&S) twin Wavelength Selective Switch (WSS) cards with larger number of ports (20-32), Erbium-Doped Fiber Amplifier (EDFA) arrays, and MxN multicast switches/WSS.

ROADMs for the access use flexible grid Broadcast and Select (B&S) WSS cards, with a lower number of ports (4-9) and passive 100 Gigahertz (GHz) flat top filters that support DWDM wavelengths up to 400 Gbps, at a much lower price point. This architecture allows the deployment of a mix of 10 Gbps non-coherent DWDM wavelengths for low-capacity application, with coherent 100, 200 and 400 Gbps wavelengths for higher capacity applications. Long-haul DWDM systems typically only support coherent wavelengths.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Due to their lower complexity, the photonic and transponder cards for access applications consume less power and are fitted in a more compact form factor. This is an important consideration for UAT POPs where space and power are sometimes restricted.

The ROADMs should support full hardware redundancy at shelf and card level to maximize traffic resiliency in case of hardware failure or during equipment software upgrades.

OTDR support, implemented on long-haul corridors to quickly detect and locate the fiber breaks, is based on dedicated hardware per span. Long-haul tailored hardware has a large dynamic range to accommodate fiber spans in excess of 100 km. In metro networks, where the spans are usually much shorter (15-20 km), the same hardware can be leveraged across multiple spans by bypassing the wavelength used by the OTDR between neighbor spans as illustrated in Figure 9. This implementation preserves the OTDR functionality at a lower cost than long-haul OTDR implementations.

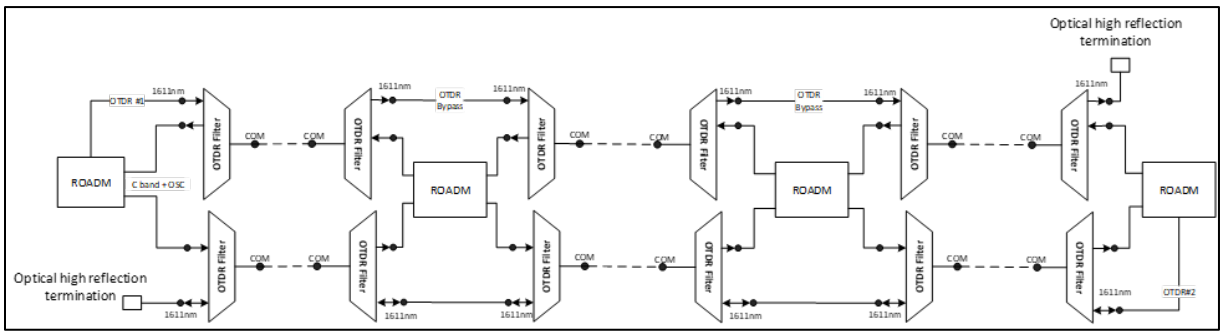


Figure 9 – OTDR Implementation on a Metro-Regional Optical Systems

Metro-regional DWDM systems are also more suitable for the foreign wavelength applications. The large OSNR margin of the coherent wavelengths can accommodate the penalty introduced by the third-party open line systems. Foreign wavelengths enable lower cost IP-over-ROADM solutions where transponders are eliminated and pluggable DWDM coherent optics are connected directly in the routers. More on this in Section 4.4.2.

Table 1 highlights the main differences between the metro-regional and long-haul DWDM active optical systems.

Table 1 – Metro versus Long-Haul ROADM Line System Comparison

Feature	Metro DWDM	Long-Haul DWDM
Transponders	Pluggable Optics $$	Proprietary Optics $$$
Flexible Grid Support	Yes	Yes
ROADM WSS Type	1x4 B&S/1x9 R&S $	1x20/1x32 R&S $$$
Low Space and Power Consumption	Yes	No
Hardware Diversity (Cards and Shelf)	Yes	Yes
OTDR Support	Multiple Spans $	Single Span $$$
CDC ROADM	No	Yes
Layer 0 Control Plane	No	Yes
800 Gbps+ Wavelength Support	No	Yes
Native 10 Gbps Wavelength Support	Yes	No

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Feature	Metro DWDM	Long-Haul DWDM
Mix of 10 Gbps and 100/200/400 Gbps DWDM on Same Photonic Layer	Yes	No
Mix of Different ROADM Types on Same Photonic Layer	Yes	No
Transponder Direct Access to WSS for Low Cost/Low-Capacity drop	Yes	No
Foreign Wavelength Support	Yes	Optional

4.3. Passive Optical Transport

Passive optical transport networks play a prominent role in the UAT and are used in both primary and secondary UAT rings and optical spurs – see Figure 1.

4.3.1. Passive DWDM Multiplexers in Secondary UAT Rings and Optical Spurs

Single-fiber bi-directional 40 channel passive DWDM multiplexers are used in UAT secondary rings and spurs to connect up to 20 devices in the field – RPD, XGS-PON OLT, Cell Site Routers (CSR), Enterprise NIDs – to aggregation routers at UAT POPs or hub sites. Figure 9 shows a protected configuration where an optical switch at the UAT POP is used to protect against fiber cuts that would impact more than 5,000 wireline or 10,000 wireless subscribers.

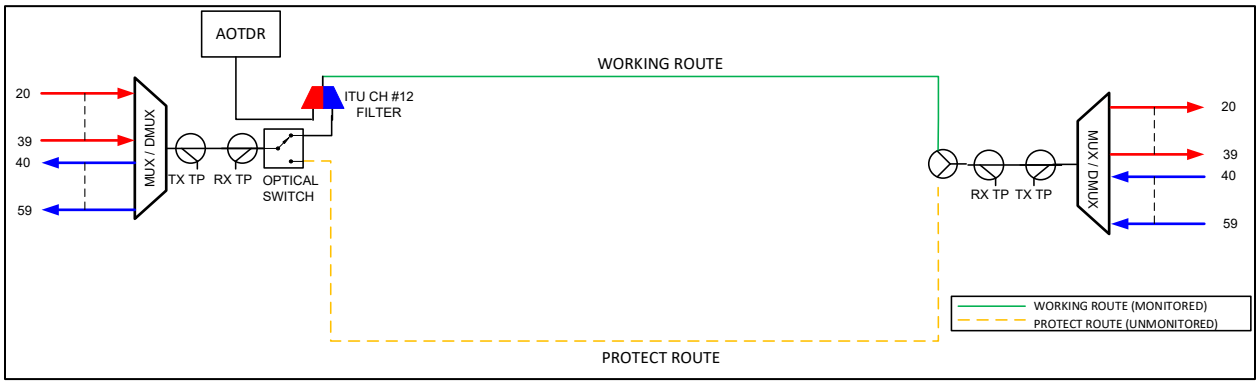


Figure 9 – Protected Bi-Directional Single-Fiber System

If the fiber cut impact is less than 5,000 wireline and 10,000 wireless subscribers, an unprotected configuration – as shown in Figure 10 – can be used to connect devices in the field to an appropriate aggregation router in the UAT POP.



Figure 10 – Unprotected Bi-Directional Single-Fiber System

The section between the DWDM multiplexer in the field and the UAT POP, which on average accounts for approximately 80% of the access fiber distance to the field device, is monitored with automated OTDR

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

instrument. Automated OTDRs reduce MTTR of the fiber cut by approximately 2 hours when compared to unmonitored fiber. Shortening MTTR is the result of improved fault detection, notification, and localization.

Single-fiber bi-directional DWDM multiplexers are also used to connect wireless small cell RRUs mounted on the street furniture and BBUs at wireless macro sites. Two configurations illustrated in Figures 11 and 12 are used to connect up to six small cell locations to a BBU hotel at the macro site. Each small cell location has an 8-channel DWDM multiplexer that connects to a 48-channel DWDM multiplexer at the macro site allowing for up to 6 small cell locations per single fiber strand. Configuration selection is based on the fiber availability and layout of the fiber cables in the specific deployment area.

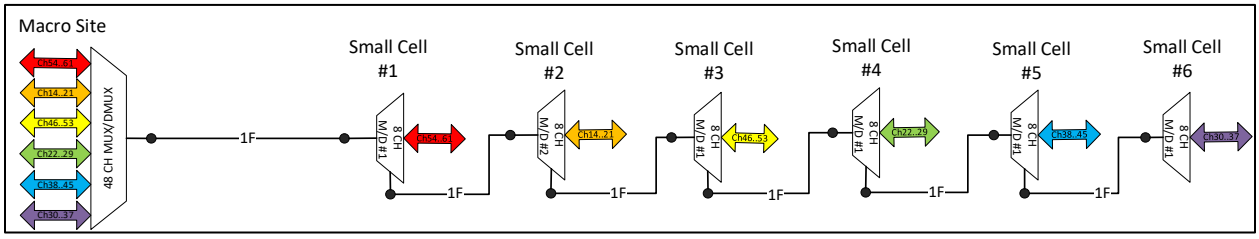


Figure 11 – Cascaded Bi-Directional Single-Fiber Configuration

The cascaded topology in Figure 11 is recommended for areas with scarce fiber and small cells installed along the architectural fiber cable. Fronthaul for small cells deployed along the major boulevards usually leverages cascaded topology.

Hub-and-spoke configuration in Figure 12 connects up to 6 small cell locations to a BBU hotel at the macro site via intermediate 6-band single-fiber DWDM multiplexer. This configuration is often used to connect small cells deployed in residential neighborhoods to architectural fiber where the band coupler is installed at the neighborhood entrance.

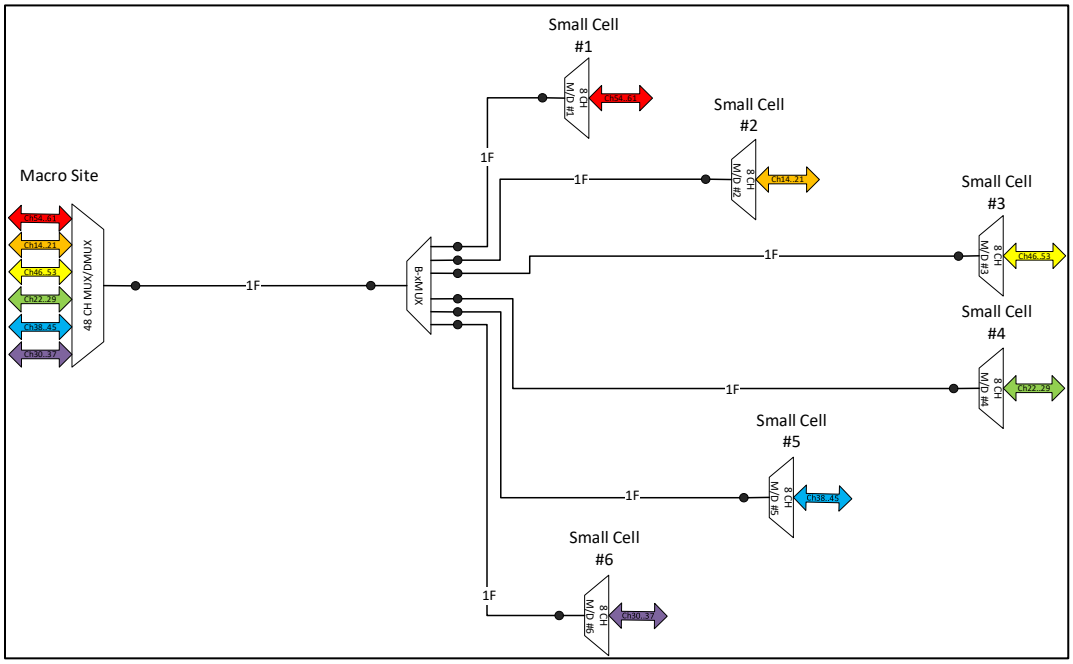


Figure 12 – Hub-and-Spoke Bi-Directional Single-Fiber Configuration

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Maximum fiber distance between the BBUs at the macro site and RRUs at the small cell locations is dependent on the optical specifications of the end transceivers, optical impairments of the transmission medium (optical loss, chromatic dispersion, etc.) and latency limitation imposed by the Common Public Radio Interface (CPRI) protocol. Although detailed fiber distance calculations are outside of the scope of this paper, practical deployments have shown that 10 Gbps fronthaul link distance is limited to 15 kilometers (km) by the (e)CPRI protocol while 25 Gbps fronthaul link distance is limited by chromatic dispersion (CD) to 13-15 km. Interestingly, fronthaul configuration and number of passive DWDM multiplexers between the BBU and RRU are often not the limiting factor.

4.3.2. Passive DWDM Multiplexers in Primary UAT Collector Rings

Passive DWDM multiplexers are not only used in the UAT secondary rings and spurs to connect devices in the field to the UAT POPs. Many of the UAT collector rings in urban areas, where distances between the POPs are shorter, are also built using passive DWDM and CWDM multiplexers. Channel allocation plan for the three passive multiplexers commonly used for the UAT primary collector rings is illustrated in Figure 13.

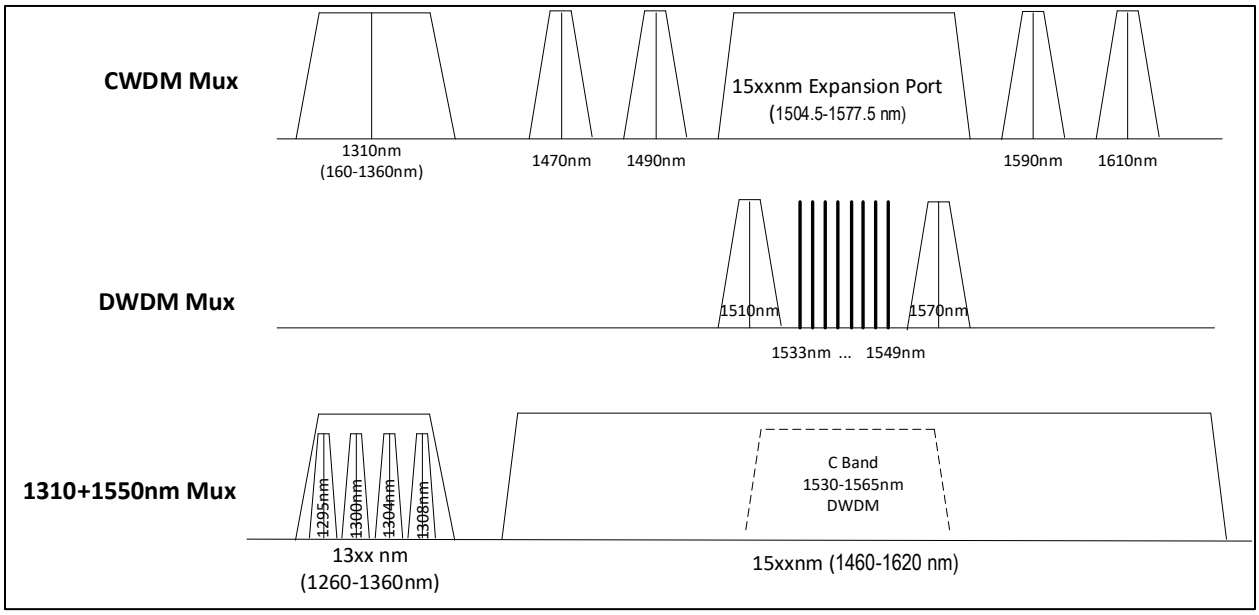


Figure 13 – Channel Allocation for Passive Multiplexers Used in the UAT Collector Rings

Figure 14 illustrates a greenfield deployment scenario where aggregation routers at the UAT POPs are connected to the 1310 nm port of the 1310+1550 passive coupler. The 1310 nm port fits four channels used by the 100 Gbps ER4 and ZR4 optics thus enabling connectivity between the routers with QSFP28 interfaces often deployed in the service provider access aggregation networks. The 1550 nm port of the coupler spans the full Conventional (C)-band which allows future system scalability. Cascading the DWDM multiplexer to the 1550 nm expansion port facilitates access to eight DWDM channels with 112.5 GHz passband that support line rates up to 800 Gbps. As per Figure 13, the 1550 nm expansion port also includes the 1510 nm channel used for the Optical Supervisory Channel (OSC) or OTDR monitoring of transport network elements, and the 1570 nm channel used for Optical Transfer Channel (OTC), required to build a high-precision timing network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

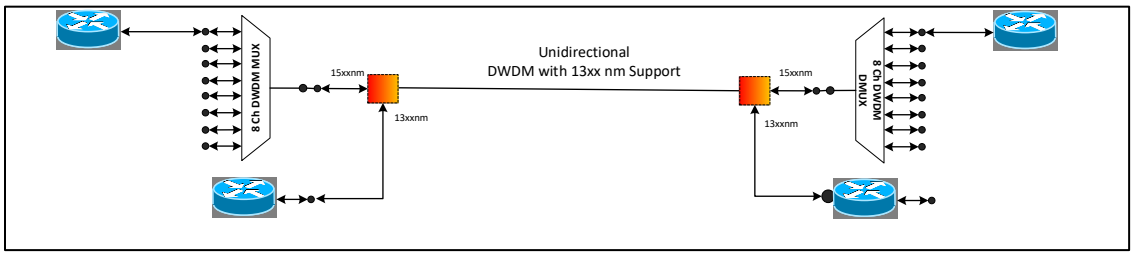


Figure 14 – Greenfield Deployment with 1310+1550 nm Passive Coupler

Dual-fiber unidirectional CWDM multiplexers are used extensively to interconnect cell sites in urban areas. The procedure to insert new high capacity DWDM systems into the 1550 nm expansion port of existing passive CWDM systems enables easier and more cost-effective rollout of UAT networks with all the benefits outlined in the Section 3.1.

Figure 15 shows a typical brownfield span connecting two IPRAN cell sites on an urban passive CWDM system. The CWDM multiplexer deployed in IPRAN has 1310 nm, 1550 nm and four CWDM ports - 1470, 1490, 1590 and 1610 nm. 1310 nm port is used for legacy Synchronous Optical Networking (SONET) traffic while one or more of the CWDM ports are used for 10 Gbps IPRAN links. Similar to 1310+1550 nm coupler, the 1550 nm expansion port of the CWDM multiplexer covers the full C-band. The addition of the DWDM multiplexer, as shown in Figure 15, allows for an easy addition of high capacity 100-400 Gbps DWDM channels used by newer QSFP-DD and CFP2 based aggregation routers and UAT active optical transport equipment without disruption to legacy SONET and 10 Gbps IPRAN channels. DWDM multiplexer also allows for access to 1510 nm OSC/OTDR channel and 1570 nm OTC channel.

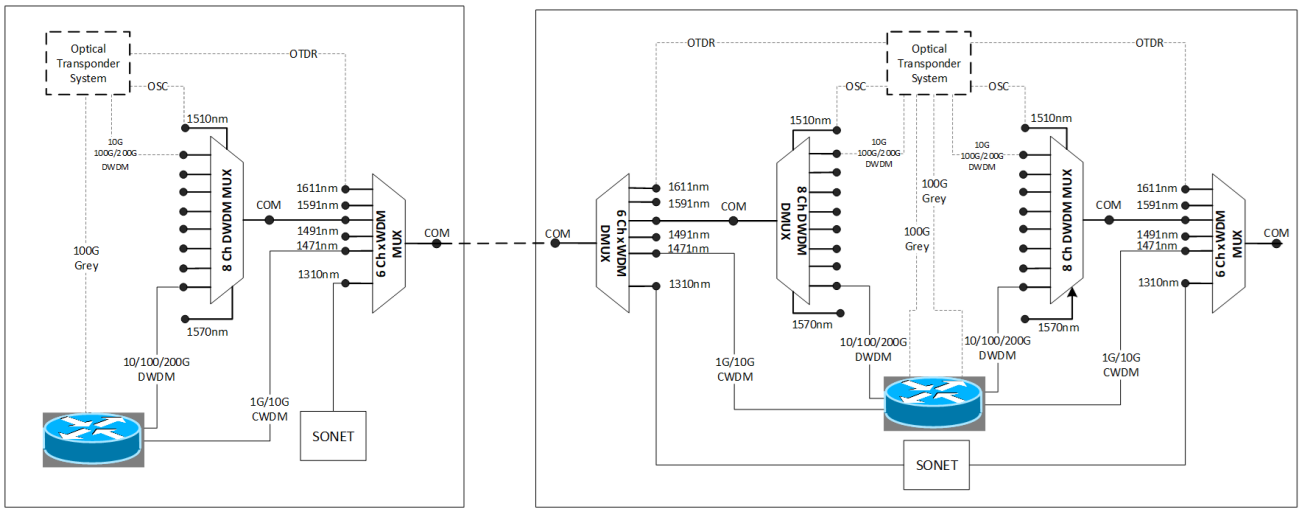


Figure 15 – Brownfield DWDM-over-CWDM Upgrade

Solutions outlined in Figures 14 and 15 are intended to be used for the hop-by-hop systems described in Section 4.1 where topology is static and optical passthrough of intermediate sites is not required. If optical passthrough capability through a single POP is required, the system designer would first consider using the higher power transceivers to overcome the cumulative optical loss from cascaded passive multiplexers. The designer can also reduce optical loss by eliminating standalone DWDM multiplexers and connecting the DWDM transceivers directly to the 1550 nm express port of the 1310+1550 nm coupler or CWDM multiplexer. The downside with this approach is a lack of future system scalability. If expressing wavelengths through multiple intermediate sites is required, coupling amplified ROADM system to the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1550 nm express port of the CWDM multiplexer or 1310+1550 nm coupler shown in Figure 16 is required. With this approach, the designer can build hub-and-spoke topology described in Section 4.1 over brownfield CWDM passive optical system.

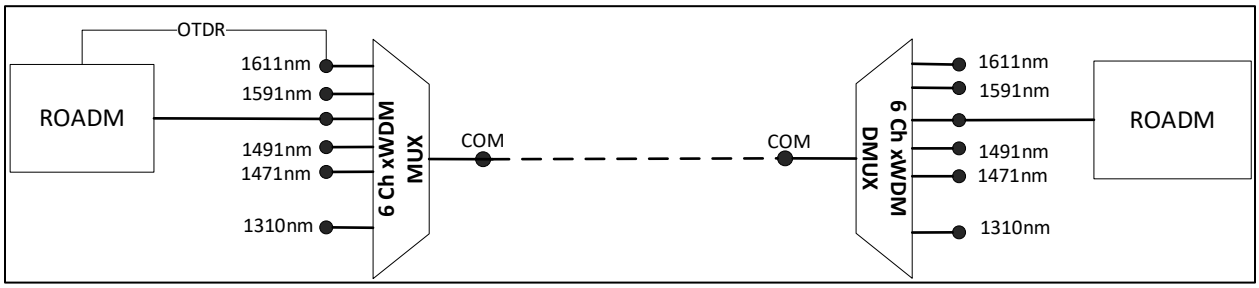


Figure 16 – Brownfield ROADM-over-cWDM Upgrade

4.4. Optical Transceivers in UAT

Optical transceivers play a prominent role in the UAT network. The following sections discuss optical transceiver requirements for the primary and secondary UAT rings.

4.4.1. Optical Transceivers in Secondary UAT Rings and Optical Spurs

Optical transceivers in the UAT secondary rings and spurs should be optimized for low cost, high volumes and simple turnup and commissioning. Pluggable optics with simple direct detect receivers that rely on a non-return-to-zero (NRZ) modulation to provide 1-25 Gbps line rates over non-amplified DWDM passive systems are tailored for such an environment. To simplify the commissioning process and minimize sparing by having a single module type that spans the full C-band, tunable DWDM transceivers have become the norm. As outlined in Table 2, some of the client devices in the access plant are not able to tune the transceiver wavelength. For such devices, external tuning devices are used to configure the channel for the tunable optics. Alternatively, a new class of tunable transceivers that automatically tunes to the channel determined by the DWDM filter, without the host equipment or external devices involvement, has recently been introduced. Although auto-tunable optics hold significant promise, current implementations of the tuning mechanism complicate optical power measurements along the optical path using conventional power meters making them impractical for the large-scale deployment. Potential solutions that address some of the limitations of existing auto-tunable modules are discussed in Section 5.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 2 – Optical Transceivers in Secondary UAT Rings and Optical Spurs

	Client Devices					
	RPD Node	**OLT Node**	**Enterprise NID**	**IPRAN Router**	**Baseband Unit**	**Remote Radio Unit**
Line Rate (Gbps)	10	10	1 and 10	10	10 and 25	10 and 25
Reach (km)	80	80	80	80	15	15
Host Device Tuning	No	Yes	Yes	Yes	No	No
Frequency Grid (GHz)	100	100	100	100	100	100
I-Temp Support	Yes	Yes	No	No	No	Yes
Form Factor	SFP+	SFP+	SFP+	SFP+	SFP+; SFP28	SFP+; SFP28
Power Consumption (W)	<2	<2	<2	<2	<2.5	<2.5

For 10 Gbps line rates typically encountered in the secondary UAT rings or optical spurs, optical power loss induced by the fiber transmission path – fiber, passive multiplexers and couplers, connectors, splices – is the dominant limiting factor. As the line rates in the access plant increase to 25 Gbps or above, optical reach of the NRZ direct detect optics is chromatic dispersion (CD) limited to 15 km. To leverage the existing passive single-fiber DWDM network and overcome the CD distance limitation, a new type of direct detect transceiver based on Pulse Amplitude Modulation 4-level (PAM4) modulation is required. To support even higher 100 Gbps line rates, more complex and costly coherent transceivers are required. Coherent module types considered for the access part of the UAT and their impact on the currently deployed single-fiber DWDM passive multiplexers are briefly discussed in Section 5.

4.4.2. Optical Transceivers in Primary UAT Collector Rings

Standard line rates for UAT primary rings range between 100 Gbps for the aggregation routers directly connected to the passive multiplexers and 200-400 Gbps for the transponder-based ROADM systems.

For the passive hop-by-hop primary rings, aggregation routers are connected directly to the optical multiplexer. To leverage existing QSFP28-based routers in the brownfield networks, 100 Gbps QSFP28 ER4 or ZR4 optics are directly connected to the 1310 nm port of the WDM couplers described in Section 4.3.2. 100 Gbps ER4 and ZR4 transceivers multiplex signals from four 25 Gbps NRZ direct detection optics operating in the 1310 nm range. As CD impact in 1310 nm band is minimal, 100 Gbps ER4 and ZR4 modules provide optical reach between 40 and 80 km not including loss from the WDM coupler which is sufficient for most of the primary collector rings.

If the 1310 nm port of the passive multiplexer is unavailable in the brownfield network applications or additional channels are required for scalability, operators have two options: (1) uplift the aggregation router to support CFP2 or QSFP-DD transceivers that can operate in the C-band, or (2) install active optical transport system to “convert” the grey 1310 nm QSFP28 interface to DWDM. As both options are costly and intrusive, the QSFP28-based coherent transceiver examined in Section 5 is urgently needed.

For collector rings with longer distances between sites or high-capacity metro applications where hub-and-spoke topology is required, active optical ROADM-based systems are recommended. As ROADMs typically require high input power of at least 0 decibel milliwatts (dBm), existing aggregation routers are connected to the optical line systems through a transponder card with grey 100 Gbps FR QSFP28 client ports and high power 200-400 Gbps CFP2 lines. Until recently, 0 dBm transmit power was achievable only with larger CFP2 modules that have enough space to house micro EDFAs and dissipate heat produced by the transceiver. Larger form factor and higher power consumption of CFP2 necessitates bigger and more expensive routers, making the IP-over-ROADM model with coherent CFP2 transceiver cost unattractive.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

With the recent introduction of smaller form factor 0 dBm QSFP-DD ZR+ transceivers that can be hosted in smaller and less costly aggregation routers than CFP2, the IP-over-ROADM model has finally become more cost attractive than the traditional transponder-based model. In addition to reducing capital expenditure, eliminating the transponder in the IP-over-ROADM model reduces space and power requirements leading to operational savings. Eliminating the transponder also means fewer elements to plan, install and maintain, leading to quicker deployment times and more resilient overall transport system. As high power QSFP-DD ZR+ transceivers become more common and router vendors endorse pay-as-you-grow model for the 400 Gbps ZR+ transceivers, IP-over-ROADM costs are expected to become even more attractive.

Introduction of the 0 dBm 100G QSFP28 ZR transceivers mentioned in Section 5 is expected to have similar benefits as the high power QSFP-DD ZR+ modules. Figure 17 compares the IP-over-ROADM and IP-over-passive DWDM cost per port with the traditional transponder-based model.

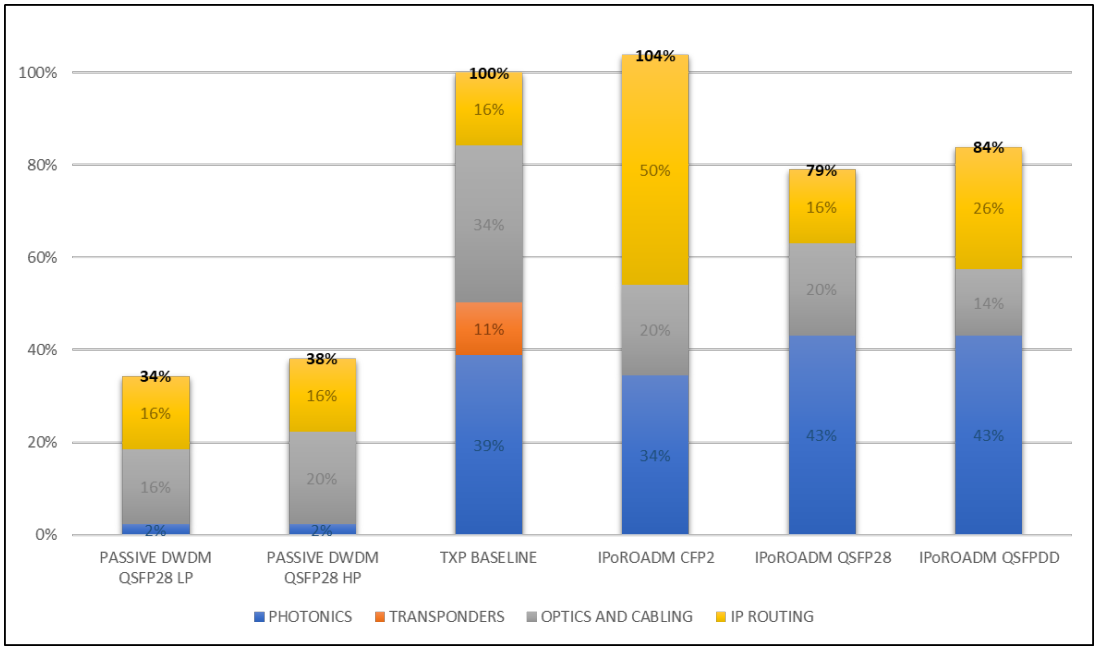


Figure 17 – Relative Cost per Port Comparison for the Primary UAT Rings

5. What's Next for UAT?

The following section previews the Plan of Record (POR) activities for the UAT. The focus of the activities is continued evolution of UAT to meet growing traffic demand while reducing costs and time-to-market without sacrificing network reliability.

Figure 18 shows directional POR for the UAT activities.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

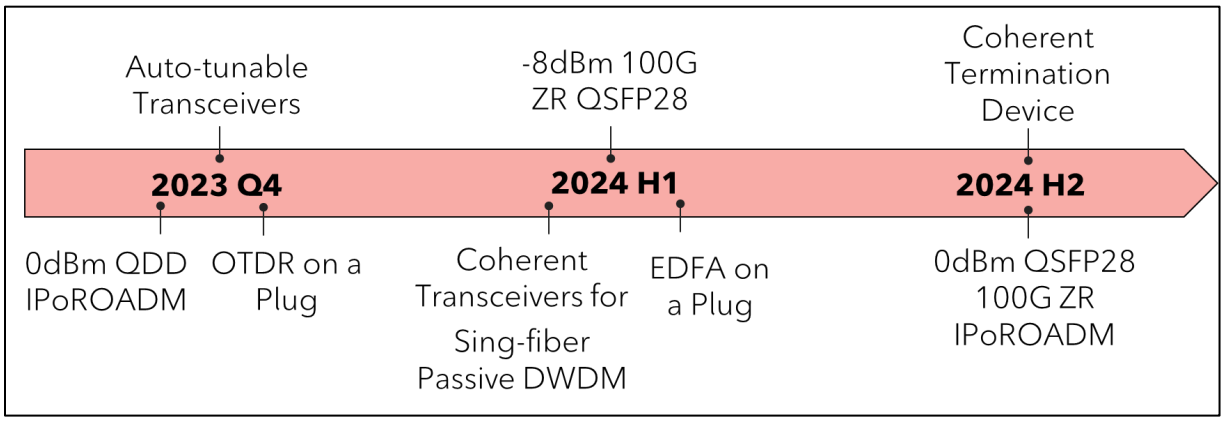


Figure 18 – UAT Plan of Record

Auto-tunable Transceivers. As discussed in Section 4.4.1, auto-tunable transceivers eliminate the need for the host equipment or an external tuning device to configure the transceiver's operating wavelength. This is particularly important for the access plant where host equipment is simpler and usually not able to change the transceiver's wavelength. Two main challenges exist with the current generation of auto-tunable transceivers: (1) the tunning mechanism engages shortly after receiver detects loss of signal which complicates troubleshooting procedures with conventional power meters that operate at a fixed wavelength, and (2) the tuning-mechanism lasts between 3.5 and 7 minutes prolonging the commissioning time. The first challenge can be resolved by having a troubleshooting mode that disables the scanning mechanism by fixing the wavelength with the external tuning box. Rogers Communications is working with equipment and transceiver vendors to certify this solution. The second challenge is addressed with the recent introduction of narrow-band auto-tunable transceivers that reduce the tuning mechanism duration to 1-1.5 minutes by covering a smaller set of adjacent channels. In addition to faster tuning, narrow band tunable transceivers are simpler and lower cost than full-band auto-tunable transceivers. The downside of the narrow-band tunable modules is that multiple part numbers are required to cover the full C-band, thus complicating sparing.

High Power QSFP-DD and Optical Disaggregation. Although introduction of 0 dBm QSFP-DD ZR+ transceivers and Open Line Systems enable optical disaggregation, improved automation and management tools are required before the IPoROADM model is rolled out in production networks. Rogers Communications engineering and Operations and Support Systems (OSS) teams are working collaboratively to introduce a vendor agnostic hierarchical multidomain controller that communicates with individual IP and Optical domain controllers to provide real time topology and service visibility, network analytics and service assurance through a common user interface. In addition to being able to provision and manage services end-to-end, the hierarchical controller provides a multilayer view of the network enabling easier correlation between optical and IP layer events.

OTDR on a Plug. Sections 4.2 and 4.3.1 outline the benefits of OTDR monitoring for the secondary rings and active optical systems used in UAT primary rings. To provide low cost OTDR monitoring for optical links on the passive collector rings, Rogers is working with routing and optics vendors to integrate a micro-OTDR transceiver directly into the aggregation router. Upon the loss of signal detected by the receiver, the transmitter switches into the OTDR mode and sends optical pulses down the impacted fiber link. The router then extracts the distance to the fiber fault from the OTDR transceiver and sends an SNMP message to the element management system as shown in Figure 19. The Micro-OTDR SFP transceiver is a CWDM transceiver that can operate in 1470, 1490, 1510, 1570, 1590 and 1610 nm channels and is thus perfectly suited for the passive CWDM and DWDM multiplexers on the UAT primary rings.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

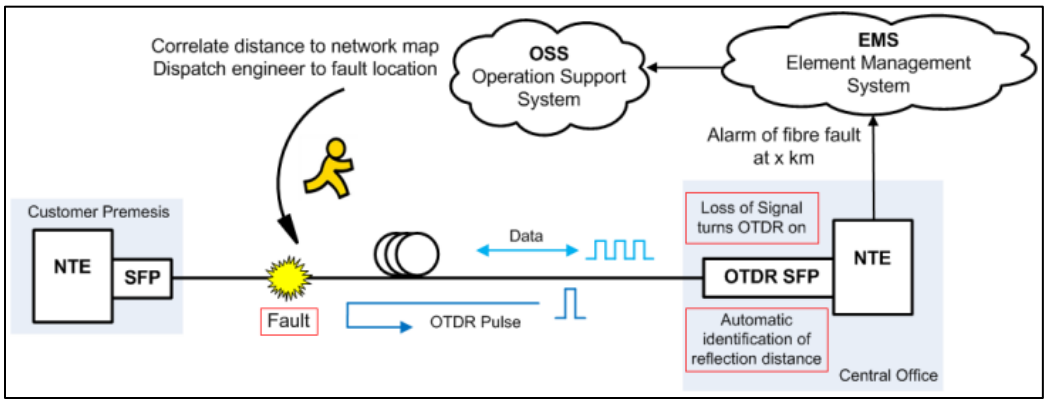


Figure 19 – OTDR on a Plug [1]

100 Gbps QSFP28 ZR. Figure 17 in Section 4.4.2 illustrates the benefit of using 100 Gbps ZR QSFP28 optics for the 100 Gbps DWDM primary rings. In addition to offering lowest cost per port for the greenfield deployment, 100 Gbps ZR transceivers allow for easy and cost-effective upgrade of the brownfield primary rings with widely deployed aggregation routers. While -8 dBm 100 Gbps QSFP28 ZR will be used for the passive primary rings, a high power 0 dBm version of the transceiver is predominantly expected to enable IPoROADM deployments. A critical challenge for transceiver vendors is to design 0 dBm optics that fit into the QSFP28 power budget supported by older routers. As primary rings evolve to higher speed line rates with the mass use of high power 400 Gbps QSFP-DD optics, 100 Gbps QSFP28 ZR will proliferate in the access part of the UAT and is expected to be one of the key enablers for the wider adoption of the Coherent Termination Device (CTD) discussed later in this section.

Coherent Transceivers for Single-fiber Passive DWDM. Conventional coherent transceivers use the same wavelength in both transmit and receive directions. Therefore, standard coherent modules require dual-fiber unidirectional passive multiplexers shown in Figure 20a. As discussed in Section 4.3.2, these multiplexers are deployed in the UAT primary rings. To enable the use of CTD over single-fiber bi-directional DWDM multiplexers used in the UAT secondary rings and spurs, a new approach is required. Figures 20b and 20c show two potential solutions for single-fiber unidirectional DWDM multiplexers: (1) coherent transceiver with two separate lasers operating at different wavelengths and (2) conventional single-laser coherent transceiver connected to a single fiber common line via optical circulators connected at both ends of the link. Table 4 shows high level pros vs cons of options illustrated in Figure 20. A more in-depth study is planned for the first half of 2024.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

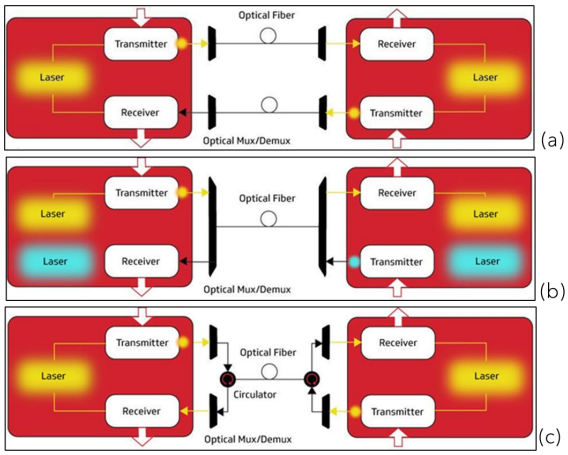


Figure 20 – Coherent Transceivers in UAT [2]

Table 4 – Coherent Transcievers over Passive DWDM Multiplexers

	Advantages	Disadvantages
Dual-Fiber Unidirectional DWDM Multiplexer	1. Works with conventional single laser transceiver that fits in a cost effective QSFP28 form factor. 2. Small form factor allows for lower cost host device.	1. Fiber inefficient.
Single-Fiber Bidirectional DWDM Multiplexer	1. Fiber efficient.	1. Requires large form factor costly CFP2 module to fit two lasers for the Bi-Di transceiver. 2. CFP2 form factor drives up the host device cost.
Single-Fiber Bidirectional DWDM Multiplexer with Circulator	1. Fiber efficient. 2. Works with conventional single laser transceiver that fits in a cost effective QSFP28 form factor. 3. Small form factor allows for lower cost host device.	1. Transmission performance is highly dependent on the quality of the fiber. Low reflection splices and connectors required.

EDFA on a Plug. Optical amplification is usually required for the longer spans connecting UAT POPs in rural primary rings. In this scenario active ROADM-based optical systems with WSS for expressing wavelengths and amplifier module are present. Amplification might also be required for urban primary rings to overcome losses from multiple cascaded passive multiplexers. This is applicable to brownfield DWDM-over-CWDM scenarios where optical signal from aggregation routers is patched through several POPs as shown in Figure 21. If space and power at the UAT POPs are constrained for standalone amplifier transport shelf, one option being considered is to connect a pluggable EDFA amplifier module directly to the aggregation router. While several routing vendors offer native EDFA-on-a-plug in the QSFP-DD form factor, integration of pluggable amplifiers in the SFP form factor is still very uncommon. An EDFA amplifier can be used as a booster, pre-amplifier or in-line-amplifier. A more in-depth study of pros and cons of integrating SFP EDFAs with existing brownfield routers is planned for the first half of

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2024.

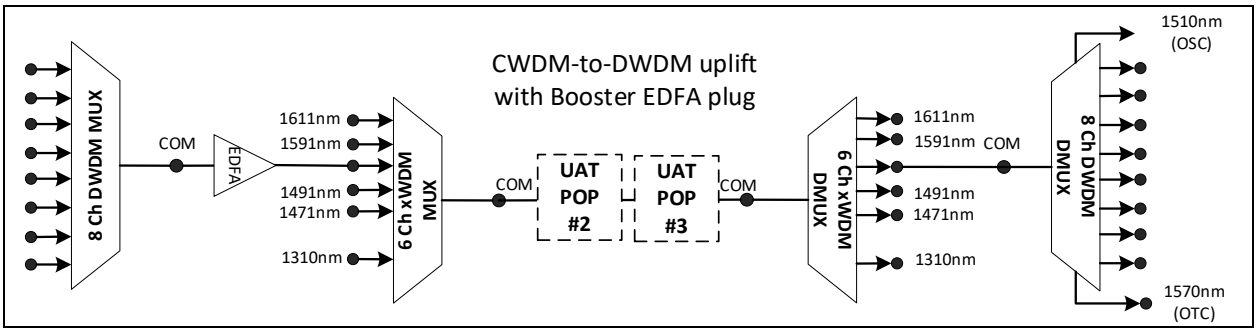


Figure 21 – EDFA on a Plug

Coherent Termination Device. The goal of the Coherent Termination Device (CTD) is to aggregate client devices on an outside plant router installed deeper into the access plant. As illustrated in Figure 22, CTD uses coherent transceivers to connect to aggregation routers located at the UAT POPs and direct-detect 10 or 25 Gbps transceivers to feed client end devices – RPD, OLT, Enterprise NIDs or CSRs. To keep the CTD costs and power consumption down, trunk coherent links should use the QSFP28 transceivers; this would also minimize upgrades of existing QSFP28-based aggregation routers at the UAT POPs. In fiber rich areas, CTD QSFP28 trunk ports would connect to aggregation router over dual-fiber unidirectional multiplexers. To avoid upgrading currently used single fiber bidirectional DWDM multiplexers, CTD QSFP28 trunk would connect through the optical circulator.

One of the advantages of aggregating traffic from client devices deeper in the plant via a CTD device instead of a passive DWDM multiplexer is that costly tunable 10 and 25 Gbps DWDM transceivers can be replaced with significantly lower cost grey optics. Additionally, unlike the topology in Figure 9 where client devices are connected to a single POP location, the CTD enables diverse coherent feeds to two geographically diverse UAT POPs. The topology shown in Figure 22 reduces the failure domain by eliminating a UAT POP site failure as a service affecting incident.

Despite the previously outlined advantages, the CTD adds another active point of failure in the plant that was not present with the passive multiplexer solution in Figure 2. A more in-depth study of the CTD's role in the access network is planned for the second half of 2024.

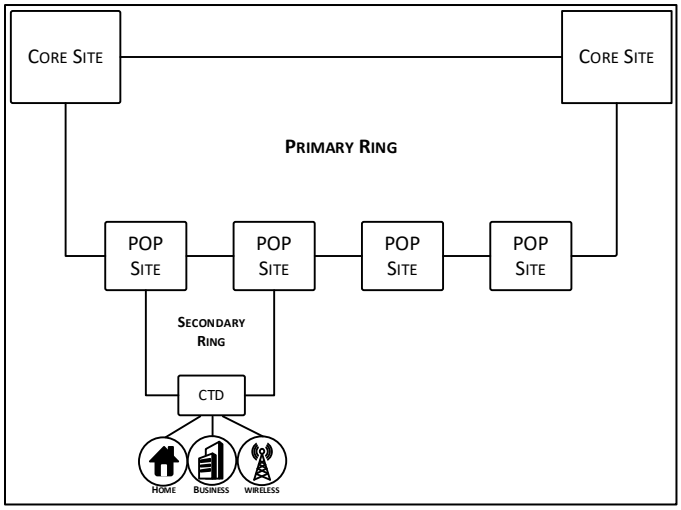


Figure 22 – Coherent Termination Device

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Conclusion

This paper takes a closer look at Unified Access Transport, a common access optical transport network for all services at Rogers Communications. It explains the major drivers for the UAT – Network Simplification and Resiliency Improvements, Wireline and Wireless Modernization and Uplift, and Service Expansion – and presents how UAT enables them. After examining UAT topology options used in the urban and rural deployments, this paper explores specific requirements for the active optical solutions suitable for the metro-regional systems before investigating passive DWDM solutions for primary and secondary rings. Innovative network designs that minimize customer impact, such as optical switching between DWDM multiplexers in the plant and at the UAT POP, and monitoring solutions that reduce MTTR, such as OTDR, are also examined. In addition, transceiver options used in the UAT, both direct detect transceivers typically used in the secondary rings and coherent transceivers predominantly used in the primary collector rings are investigated. The analysis concludes that recent introduction of the 0 dBm 400 Gbps QSFP-DD ZR+ is a key enabler for the wider adoption of IPoROADM solutions in the primary collector rings. The paper concludes by reviewing near to medium term POR which indicates that transceiver advancements play a prominent role in the UAT evolution.

Abbreviations

BBU	baseband unit
BNG	broadband network gateway
B&S	broadcast and select
CCAP	converged cable access platform
CPRI	common public radio interface
CSR	cell site router
CTD	coherent termination device
CWDM	coarse wavelength division multiplexing
DOCSIS	data over cable service interface specification
DWDM	dense wavelength division multiplexing
DAA	distributed access architecture
dBm	decibel milliwatts
EDFA	erbium-doped fiber amplifier
FOADM	fixed optical add-drop multiplexer
FTTH	fiber to the home
FWA	fixed wireless access
Gbps	gigabit per second
GHz	gigahertz
HP	homes passed
HFC	hybrid fiber coaxial
IP	internet protocol
IPRAN	ip radio access network
IPWAN	IP Wireline Access Network
km	kilometer
MTTR	mean time-to-repair
MIMO	multiple input multiple output
NID	network interface device
OLT	optical line terminal
OSNR	optical signal to noise ratio

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

OTDR	optical time domain reflectometer
PON	passive optical network
POP	point of presence
RBS	radio base station
ROADM	reconfigurable optical add-drop multiplexer
R&S	route and select
RPD	remote phy device
RRU	remote radio unit
SONET	synchronous optical networking
Tbps	terabit per second
UAT	Unified Access Transport
UFP	Unified Fiber Plan
VCCAP	virtual converged cable access platform
WSS	wavelength selective switch
XGS-PON	10 gigabit per second symmetrical passive optical network

Bibliography & References

[1] *Gigabit SFP Transceiver with Integrated Optical Time Domain Reflectometer for Ethernet Access Services*; Neil Parkin, Meir Bartur, Derek Nesset, David Jenkins; ECOC 2013

[2] *Doubling up on Fiber Capacity: A Winning Strategy for Full Duplex Coherent Optics*; Steve Jia; https://www.cablelabs.com/blog/doubling-fiber-capacity-winning-strategy-full-duplex-coherent-optics

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Exploring Programmatically Generated HFC Plant Topology

Using Data and AI/ML to Automate Your Network Documentation Process

A Technical Paper prepared for SCTE by

Lin Cheng
Principal Architect
CableLabs
858 Coal Creek Cir, Louisville, CO 80027
l.cheng@cablelabs.com

L. Alberto Campos
Fellow
CableLabs
858 Coal Creek Cir, Louisville, CO 80027
a.campos@cablelabs.com

Justin Riggert
Engineer 5, Product Development Engineer
Comcast Corporation
183 inverness Dr W, Englewood CO 80112
justin_riggert@cable.comcast.com

Larry Wolcott
Comcast Fellow Engineer
Comcast Corporation
183 inverness Dr W, Englewood CO 80112
larry_colcott@cable.comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 4
2. Background and Scope ... 4
3. Use Cases, Benefits, and Examples ... 5
   - 3.1. Generating a Topology With No As-Built Maps ... 5
   - 3.2. Audit and Update Existing As-Built Maps ... 5
   - 3.3. Tap-to-Home Association ... 6
   - 3.4. Drop Length ... 6
   - 3.5. Verification of Link Budget, Loss, and Performance Targets ... 7
   - 3.6. Real-Time Repair and Construction Verification ... 8
4. MIND™ Overview ... 8
   - 4.1. Multiple Sources of Information ... 9
   - 4.2. Analysis Mechanisms ... 10
   - 4.3. Normalization, Calibration, and Correlation ... 10
     - 4.3.1. Timing Offset ... 10
     - 4.3.2. Group Delay ... 11
     - 4.3.3. Power Level ... 11
   - 4.4. Relative versus Absolute Metrics ... 12
   - 4.5. Sequential and Iterative Discovery Processes ... 13
   - 4.6. Network Modelling ... 13
5. Analysis ... 17
   - 5.1. Rule-based Analytics ... 19
   - 5.2. ML-Based Analytics ... 21
   - 5.3. Data Processing ... 22
6. Implementation ... 22
7. Future Work ... 24
8. Conclusion ... 24

Abbreviations ... 25

Bibliography & References ... 27

Appendix A ... 28

Appendix B ... 30

Appendix C ... 33

List of Figures

Title **Page Number**

Figure 1 – Analyzed coaxial segment with hardline cable (blue) and drop cable (black). ... 11

Figure 2 – Relative CM US power versus frequency for 80’ drop (dashed), 100’ drop (solid), 120’ drop (dashed-dot) a) at 23 dB tap drop b) at 14 dB tap drop c) at 8 dB tap drop all with no upstream tilt. ... 12

Figure 3 – CM Rx power vs. frequency for 80’ drop (dashed), 100’ drop (solid), 120’ drop (dashed-dot) a) at 23 dB tap drop b) at 14 dB tap drop c) at 8 dB tap drop in a 1dB/100 MHz uptilt. ... 12

Figure 4 – Amplifier S-parameters a) reverse transmission and reflection b) forward transmission and reflection. ... 14

Figure 5 – Amplifier reverse transmission S-parameters a) magnitude and phase b) magnitude and group delay. ... 14

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 6 – Amplifier forward transmission S-parameters a) magnitude and phase b) magnitude and group delay. 15
Figure 7 – 17 dB Tap S-parameters a) main path b) coupled port path. 15
Figure 8 – 17 dB tap main path S-parameters a) magnitude and phase b) magnitude and group delay. 15
Figure 9 – Coaxial path within a fiber node serving area 44 to CM 99. 16
Figure 10 – Analysis flow of MIND. 17
Figure 11 – Example of loss prediction of 100 meter coaxial cables. 17
Figure 12 – Example flow of rule-based analytics in MIND. 19
Figure 13 – Laboratory example of HFC network topology. 23

List of Tables

Title **Page Number**

Table 1 – Sources describing distance/length. 9

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Hybrid fiber/coax (HFC) network element and connectivity information in operator databases is sometimes inaccurate, out-of-date, or even missing. Many operational, administrative, and business functions rely on accurate plant data. A solution for programmatically generating coaxial plant topology using Data-Over-Cable Service Interface Specifications (DOCSIS®) network telemetry, spatial and address information, and deployment practices is presented. The approach proposed enables cable operators to keep their system designs up-to-date in real time without relying on manual processes, thereby reducing delays and associated manual burden. By automating the process of map recording, the proposed solution offers significant time savings while ensuring that accurate plant design information is available for efficient field operations, tool development, network planning, and effective service activation and delivery. This approach leverages machine learning (ML) and spatial analysis techniques to extract network topology from DOCSIS network telemetry, deployment practices, and geodata. The effectiveness of the approach is demonstrated through simulations and experiments on real-world data. This solution has the potential to revolutionize how cable operators manage their coaxial plant infrastructure, evolve their networks, and improve overall network efficiency.

2. Background and Scope

In HFC network designs, technological advancements have been met with numerous challenges, resulting in a constant evolution of system designs. Traditionally, these designs are prepared by trained engineers using computer aided design (CAD) systems and specialized tools. This paper proposes a revolutionary automated system for generating HFC network topology and geographic information system (GIS) data, focused on enhancing the efficiency and accuracy of these designs while grappling with their inherently dynamic nature.

The engineering and design of HFC networks can be a complex process, with inherent difficulties caused by several issues. During the initial network design, various factors can lead to the network designs not being implemented as intended. For example, construction teams may lack specific equipment, necessitating substitutions not accounted for in the initial design. There are also environmental factors and unforeseen infrastructure elements, such as pre-existing conduits, that may also compel modifications to the initial design. There are also legal and regulatory elements such as easement access, local permit processes, and homeowner association (HOA) regulations that could potentially force alterations to the design.

Data integrity is a crucial aspect of the design process. Inaccuracies can result in flawed network parameters, jeopardizing the construction process and compromising the effectiveness of the network. Post-activation, the network continues to evolve, with countless factors contributing to divergence from the original design. Field conditions, such as unanticipated drop cable lengths and the addition of home splitters, can cause technicians to alter tap values and conditioning. Technicians may modify designs by adding taps where needed, a necessity not foreseen during the initial design. Amplifier setup, a frequent point of deviation from the initial design, can be altered due to a multitude of reasons, including natural temperature changes or cable impairments.

Major changes to the system to accommodate damaged cables, additional capacity, or improved redundancy can further exacerbate the disparity between design and implementation. This divergence necessitates constant updates to the design, known as the red line process, which can be a laborious and time-consuming task. System upgrades, such as node splitting, often come with their own set of challenges. Additional nodes can intensify existing issues and introduce new complexities, leading to

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

delays in construction and activation of new services, thus impacting the efficiency of ongoing network maintenance.

Given this extensive array of factors that contribute to the dynamic nature of HFC network designs, the traditional methods of design and maintenance can be fraught with inefficiencies. This paper proposes an automated approach, addressing these challenges and paving the way for a more streamlined and effective way of maintaining HFC network designs.

3. Use Cases, Benefits, and Examples

3.1. Generating a Topology With No As-Built Maps

In older cable networks, particularly those built before the advent of modern CAD software and GIS, there can be a significant lack of documentation about the network's topology. This can cause a wide range of issues, from difficulty identifying and fixing network issues to problems planning upgrades or expansions.

The problem compounds when we consider that these networks may have been operating for several decades. Over time, modifications, repairs, and upgrades are likely to have been made to these networks that deviate from any original designs. Furthermore, the physical geography in which these networks are deployed may have also changed over time due to new construction or other factors.

This presents a challenge: How can we obtain accurate, up-to-date network topology information from these older systems that may not have any digital documentation? The answer lies in a process often referred to as reverse-engineering the network topology. However, without any data to start with, this can prove to be a challenging task.

With traditional methods, creating an accurate topology from these systems would require a combination of physical inspection (which could be both time-consuming and costly, especially for larger networks), interviews with long-term staff (who may or may not remember the specifics), and potentially sorting through old, possibly outdated paper documentation.

Therefore, the challenge is to generate a current as-built topology for older, undocumented systems with the potential for significant deviations from any existing documentation. This involves the identification of all nodes and connections within the network, their characteristics, and their geographical locations, which is a substantial task without initial data or documentation.

3.2. Audit and Update Existing As-Built Maps

In this scenario, a system design is already in place, but due to a multitude of reasons previously outlined, the actual network might have deviated from the original design. These reasons can include but are not limited to:

- Infrastructure modifications and repairs over time that weren't correctly documented.
- Construction that differed from the original design due to inaccuracies or misinterpretation of design parameters.
- Physical and geographical changes in the environment around the infrastructure.
- Technological upgrades and network expansions that may have been inadequately recorded.

Despite having an initial system design, these discrepancies can result in the as-built network topology diverging significantly from the as-designed or "as is" map. The challenge here is to validate the existing

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

system design and generate an accurate, up to date as-built topology that reflects the network's current state. One option would be for the operator to conduct a full physical audit of the network performed either by in-house staff or contracted field engineers to generate up to date as-built maps. This is a labor-intensive approach that would have to be conducted continuously to maintain validity.

The solution to this problem could be an automated auditing tool. The auditing tool would cross-verify the actual network performance and topology against the existing design, effectively producing an audit score that quantifies the level of accuracy or confidence in the original designs compared to the current network status.

By implementing such an audit tool, operators would gain a more reliable and accurate understanding of their network topology and be better equipped to make informed decisions on future work, prioritization of physical audits, and improvements in documentation accuracy.

3.3. Tap-to-Home Association

A typical issue in cable networks is the uncertainty in associating customer locations to specific network taps. A tap is a connection point in a network that provides a junction between the primary network cable and individual drop connections to households or businesses. They are often organized in a sequential manner along a feeder line. However, taps can often be fed in ways that are counter-intuitive to a casual observer.

Given their configuration, it is not uncommon for multiple taps on a feeder to be approximately equidistant to a service location. This presents a challenge in accurately associating a given service location with its respective tap based on distance and sequence.

Furthermore, the choice of a tap for a specific location isn't always driven by proximity. There can be various practical or environmental factors that influence the selection. For instance:

- Physical barriers: Trees, buildings, and other structures can impact the choice of tap. For example, it might be more feasible to connect a household to a tap that is farther away but has a clear line-of-sight, rather than a closer tap that requires navigating around a building.
- Access issues: Fences, easements, and other access restrictions can also influence the choice of tap. Taps that are easier to access, even if they're slightly farther away, might be preferred over closer but less accessible taps.

Given these considerations, the choice of tap for a service location may not be immediately evident from the network design or based solely on geographical proximity. This can result in a level of uncertainty in the network topology, making network management and troubleshooting more challenging.

Therefore, there's a need to accurately determine the correct association between service locations and physical taps based on actual network configuration and installation considerations, rather than just relying on geographical proximity or sequence. This would lead to a more accurate network topology, which in turn would improve network management, performance, and planning for future upgrades.

3.4. Drop Length

Cable network designs typically include the main network infrastructure, outlining the main lines of communication and significant hardware components such as amplifiers and nodes. However, one area often not included or lacking detail in these designs is the drop connection – the final link that connects the network to individual users or households.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

This lack of drop connection information in network designs primarily arises because these connections are typically the responsibility of other teams or are installed and activated well after the initial construction of the plant. This practice often results in a network design that doesn't include the accurate drop lengths and types. This poses a significant problem as drop connections are a crucial part of the network. Their length, type, and quality can significantly impact network performance parameters, such as signal strength, error rates, and overall link quality.

Moreover, standard design parameters and design guidance that inform the selection of tap parameters may not always align with the realities of drop installations. There may be deviations in drop lengths or types due to on-site realities, end-user requirements, or other factors not considered in the initial design stages.

This presents an incomplete and possibly inaccurate picture of the actual network. Without an accurate understanding of drop connections, the effectiveness of network management, troubleshooting, optimization, and planning for future upgrades can be impacted. Therefore, there's a need to find ways to infer these missing details and complete the network design based on the actual parameters of the network.

3.5. Verification of Link Budget, Loss, and Performance Targets

In cable network operations, a key aspect is being able to assess whether the network is performing as it was designed. This involves validating various network parameters, such as frequency response, signal strength, and error rates, against the theoretical values outlined in the system design.

One specific area of interest is the radio frequency (RF) link budget, a calculation that considers all the gains and losses from the transmitter, through the coaxial and passive network elements, to the receiver in a system. This includes factors like transmission power, cable losses, tap losses, connector and splitter losses, and noise figures.

The RF link budget forms an integral part of the network design, outlining the expected performance of each link in the network. However, due to a multitude of factors, the actual network performance can deviate from these theoretical expectations.

This could be due to changes in network components over time, including impairments, connector degradation, and even environmental factors like water, temperature, and humidity fluctuations affecting signal propagation. Alternatively, the actual drop lengths and types might deviate from the standard design parameters, impacting the RF link budget.

Being able to validate the actual network performance against the designed RF link budget is important for cable operators. This is because it enables them to:

- Confirm Service Delivery: By verifying that the network's actual performance aligns with the design, operators can confidently confirm the level of service they can deliver to their customers.
- Identify Network Issues: Deviations between actual performance and the design can highlight potential network issues, like faulty equipment or connections, which can be proactively addressed.
- Plan for Network Upgrades: Understanding the real-world performance of the network can guide future planning for network upgrades and improvements, ensuring that the network continues to meet customer service requirements.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Therefore, an important use case for DOCSIS network telemetry is to verify network performance against the system design, including validating the actual RF link budgets against their designed values. This allows for effective network management and optimizes service delivery based on the network's actual, rather than theoretical, performance.

3.6. Real-Time Repair and Construction Verification

Cable operations involve regular repair and construction activities, ranging from routine maintenance to infrastructure upgrades. However, validating the completion and efficacy of these operations often requires manual inspection or relying on field reports, which can be time-consuming, labor-intensive, and possibly prone to error.

This use case involves leveraging DOCSIS network telemetry in real-time to automate the validation of repair and construction activities. The process would be akin to a digital "red line" procedure that provides instant feedback and validation of completed tasks.

For instance, if a work ticket outlines a construction job that requires adding a specific length and type of cable to a segment, network telemetry can be used to validate the job completion. After the cable installation, the network's RF levels and tilt can be checked through telemetry data. If these values align with the expected results for the added cable's type and length, it validates that the job has been completed as intended.

This real-time verification not only helps ensure accuracy in construction and repair tasks but also may reduce the need for subsequent manual checks.

Additionally, the same real-time telemetry data can be used to close the loop on repair recommendations made by other artificial intelligence (AI)/ML systems used by cable operators. For example, if an AI system suggests a specific repair to address a detected network issue, the telemetry data can be used post-repair to verify if the recommended action has effectively resolved the problem.

Therefore, another potential use case of network telemetry is real-time repair and construction verification. This application can greatly enhance the accuracy, efficiency, and effectiveness of network maintenance, construction, and repair activities, contributing to optimal network performance and reliability.

4. MIND™ Overview

The origin of the MIND or Methodology for Intelligent Network Discovery concept was driven to answer the previously discussed challenges existing from the partial or total unavailability of HFC network data. From the realization that just a single source of data generation may not be sufficient to discover HFC network elements, MIND leverages and integrates multiple data sources and a diverse set of tools in a coordinated fashion to enhance:

- capabilities in type and number of HFC elements discovered,
- granularity in determining HFC element values or characteristics,
- sensitivity in detecting common traits for grouping or clustering.

MIND targets to discover all possible HFC network elements, from their types, values, and connectivity relations to other elements as well as their location in a programmatic fashion. (A list of network elements targeted for discovery is provided in Appendix A.) MIND assumes that no prior knowledge of HFC network topology and elements is necessary as these will be determined through the MIND process. If

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

they are available, MIND can be used to verify the characteristics and location within the fiber node topology of the HFC network elements. While the discovery in MIND is processed one RF domain at a time, which is typically a fiber node serving area at a time using DOCSIS service groups that link cable modems (CMs) to fiber nodes. As machine learning is leveraged, the learnings and tools derived from the larger population of RF service groups are applied to the fiber node serving area being processed.

4.1. Multiple Sources of Information

A single source or single tool may not uncover with certainty topology or HFC network element values. But when multiple sources of information are intelligently combined, the accuracy and confidence level in defining the HFC network topology and its network elements are increased.

The sources of information considered in MIND include:

- **DOCSIS and proactive network maintenance (PNM) metrics** – Available through DOCSIS management information base (MIB) and/or command line interface, PNM collection tools, etc. [1-5] Initial DOCSIS metrics considered in MIND are included in Appendix B.
- **Geodata** – Streets, lot boundary, home construction perimeter, pedestals, attachment poles, aerial, or underground network identification, etc.
    - **CM association to street address** - CM latitude and longitude estimate
- **HFC network deployment rules and guidelines** – Includes guidelines on the placement of pedestals, aerial versus underground practices, practices when distribution coaxial cable is deployed in front of homes or when it runs in rear easements, multiple dwelling unit (MDU)/building deployment practices, taps deployed in decreasing value and operator specific practices such as range of tap values, etc.
- **Non-DOCSIS instrumentation results** – Includes those obtained through RF tools, optical time domain reflectometer (OTDR)/ metallic time domain reflectometer (TDR), alternating current (AC) voltage readings at different actives, etc.
    - **Network element type characteristics** – Network element models or specification datasheets.
    - **Existing as-built plant data**- If available for verification of accuracy.

An example of how multiple sources of information can be used to assess a specific parameter is given when we measure distance or length. We can represent distance measurements from fiber node to CM, by comparing metrics at these two network devices or by adding the individual segment lengths of the cascaded elements between the fiber node and CM. These metrics can be based on DOCSIS, PNM, geodata or others (Table 1)

Table 1 – Sources describing distance/length.

Metrics	Coverage	Source
Timing Offset	End-to-end	DOCSIS
Group Delay	End-to-end	PNM - Ch. Estimate / S_{21}
Reflection Cavity Ripple	Segment	PNM
Dist. between Pedestals/Poles	Segment	Geodata / Deployment practices
Power Level Difference	Segment	DOCSIS Rx Power/Attenuation
AC Voltage Drop	Segment	Non-DOCSIS - Attenuation / Pwr. Consumption

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2. Analysis Mechanisms

The mechanisms for uncovering HFC network data fall into two main categories:

The first category is rules-based mechanisms such as direct analysis, which can be used to estimate the values that characterize the network. For example, a micro-reflection ripple is indicative of a reflection or echo cavity between two interfaces, likely between two HFC network elements. From the ripple and coaxial cable characteristics, one estimates the length between these two interfaces.

The second category is machine learning-based mechanisms where, through training, CMs with common characteristics can be clustered together to uncover bifurcations, topology, or CM-tap association.

Both rules-based and machine learning-based mechanisms can be used to discover the same network feature or characteristic, combining the different analysis mechanisms increases the confidence level in approaching "truth" (approaching "as is"). Section 6 discusses in detail the types of analysis.

4.3. Normalization, Calibration, and Correlation

MIND extensively leverages correlation tools. In order to compare and differentiate component types, component values, component connectivity, and location, these comparisons and/or differentiation exercises must be done using components described following the same rules, formats, and definitions, to compare apples with apples. Once the data is deemed to be consistently defined then discrimination and correlation processes can take place. Curating the data can range from a simple formatting process to a comprehensive and elaborate calibration process. Parameters such as timing offset, group delay and power level that could benefit from calibration and normalization are discussed next.

4.3.1. Timing Offset

In one case in particular, curation of timing offset data takes the form of a calibration process. In DOCSIS, timing offset is defined as the compensation delay the CM must apply so that from a timing perspective it appears to be located right next to the cable modem termination system (CMTS). Successful timing offset compensation results in CM transmissions to be time aligned as they are received by the CMTS. The original purpose for such an alignment was to minimize guard time between transmissions. In DOCSIS 3.1, the granularity of timing offset was refined to meet the tighter orthogonal frequency division multiplexing (OFDM)/ orthogonal frequency division multiple access (OFDMA) synchronization requirements. The mandatory timing offset resolution in time division multiple access (TDMA) has a granularity of 6.25 µs/64 or 97.65625 ns. A TDMA optional/ synchronous code division multiple access (SCDMA) mandatory higher resolution of 1/(256*10.24 MHz) or 381 ps is available, while an optional granularity of high-resolution timing offset (OFDMA) of 1/(256*204 MHz) or about 19 ps is also available. This high-resolution granularity is defined by the timing adjust fractional part in Section 6.4.6 of DOCSIS MULPI specification [6]. Assuming coaxial transmission with a velocity of propagation of 87% of the speed of light, regular TDMA resolution timing offset provides a distance granularity of 25.47 meters, SCDMA results in a distance granularity of about 10 cm, while OFDMA high resolution timing offset provides a distance granularity of 4.974 mm or about half a cm. This higher resolution option in principle makes timing offset a powerful tool for clustering and discrimination of HFC network components.

The challenges are that timing offset was designed to align transmissions of a specific CM and the CMTS. Internal delays at the CM such as processing delays that might be included in the timing offset need to be calibrated out so that the timing offset metric can be used for comparison across different CMs.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

MIND conducts a timing offset calibration exercise to remove this CM model and chip dependent variability to leverage this powerful tool.

4.3.2. Group Delay

Another parameter that can benefit from calibration is group delay. In single carrier quadrature amplitude modulation (SC-QAM) signals, equalization parameters are provided in the time domain and main tap information is typically fixed at tap number 8, providing a good common time reference to calculate and compare equalization information as well as parameters that are derived from equalization information such as group delay. In DOCSIS 3.1 equalization coefficients are in the frequency domain. Tom Williams described a calibration process to compare OFDMA equalization across CMs [7]. This mechanism can also be used to compare group delay across CMs which helps in discrimination based on distance in addition to distortion. Metrics that at first glance may not appear useful become powerful tools with the proper normalization and calibration.

4.3.3. Power Level

Power level has always been an important metric in the verification of the proper operation of the CM. Nevertheless, it has not yet been used in the discovery of HFC plant characteristics and its elements. One important change with the introduction of DOCSIS 3.1 is that both the downstream as well as the upstream spectrum have been increased. In the upstream DOCSIS is covering up to 204 MHz while in the downstream 1.2 GHz and 1.8 GHz upper frequency limit options are feasible. This means that with much wider bandwidth, power variation effects are easier to detect, characterize, and compare with other CMs. PNM is also playing a role since full band capture (FBC) is becoming more widely used and operators are recording spectrum readings at amplifiers and taps, either from embedded CMs or test ports at amplifiers or at tap drop ports. This provides useful reference information that can be compared with the CM FBC and to extract the delta performance to gain detailed information on the drop-home portion of the network. Along with FBC channel estimate or equalization information at the tap are powerful network discovery metrics.

The coaxial segment in Figure 1 has been analyzed to highlight the use of power level signatures from CMs.

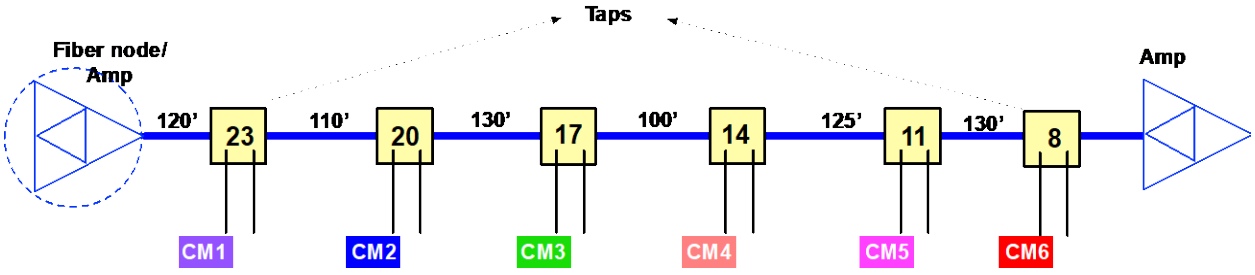


Figure 1 – Analyzed coaxial segment with hardline cable (blue) and drop cable (black).

In this coaxial segment scenario, CMs connected at different taps with different drop lengths are evaluated. Figure 2 shows CMs connected at the first tap (a), fourth tap (b) and sixth tap (c).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

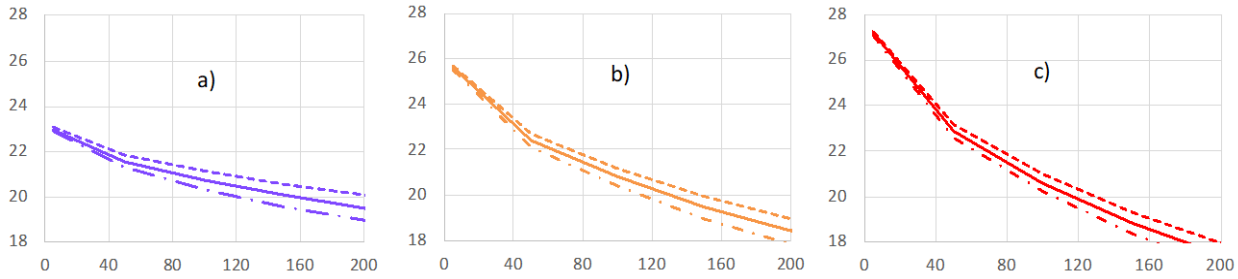


Figure 2 – Relative CM US power versus frequency for 80' drop (dashed), 100' drop (solid), 120' drop (dashed-dot) a) at 23 dB tap drop b) at 14 dB tap drop c) at 8 dB tap drop all with no upstream tilt.

Even though CMs have some variability in reporting absolute power for the behavior versus frequency, the delta power level between highest and lowest frequency and the shape of the curve is very telling of the tap it is connected to and of the tap-to-CM cable length. If tap measurements are available, then greater insight of the drop-home becomes available. A calibration process would have to take place as you would need to subtract the RF contributions up to the tap. Similar analysis can take place in the downstream (Figure 3).

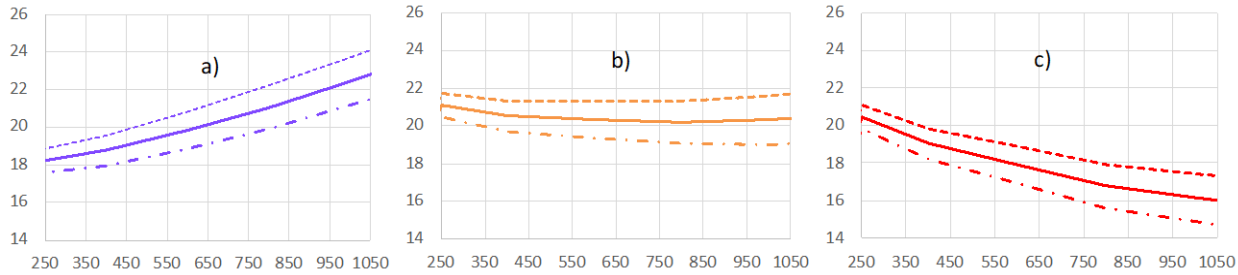


Figure 3 – CM Rx power vs. frequency for 80' drop (dashed), 100' drop (solid), 120' drop (dashed-dot) a) at 23 dB tap drop b) at 14 dB tap drop c) at 8 dB tap drop in a 1dB/100 MHz uptilt.

Comparing the responses of CMs at the different taps, one can easily observe how the tilt changes as you move along the coaxial segment. One implementation complexity is that to maximize the US and DS spectrum coverage, CMs may have to move to MAC domains that include the edges of the US and DS bands.

4.4. Relative versus Absolute Metrics

In some cases, absolute metrics may be available, while in most other cases metrics relative to a reference point may be everything that is needed to perform valuable analysis. For example, if a fiber node is used as a known reference point, finding relative distances from each of the components to that reference point/fiber node is all that is needed for an accurate representation of the coaxial network covered by that fiber node. MIND uses the fiber node as a reference point or anchor point to provide latitude-longitude reference to the rest of the network elements.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.5. Sequential and Iterative Discovery Processes

In MIND, different analytical and ML-based processes take place in a sequential manner. This means that after each process new network elements and their values are discovered or updated improving our level of confidence in the network. This improved knowledge is refined after running the algorithms that follow. As the network is not static and new information becomes available or as one accumulates greater knowledge of the network, the overall process is run again and in a continuous fashion to attain further refinement of network knowledge and to detect changes that may have happened in the network so that knowledge of the network is always current, thereby moving from an as-built to an "as is" knowledge paradigm of the network.

4.6. Network Modelling

Network models are useful tools in estimating the performance of a network. MIND uncovers the network components within a fiber node serving area to a large degree by understanding the performance and behavior measured at the CMTS and CM located at the edges of that HFC network portion or serving area. Knowledge of the behavior of HFC network elements within that fiber node serving area, even in a partial or approximate fashion, is useful in determining what components are there, how they are connected, and their characteristics. Network models can help us make sense of the readings measured at the edge by the CMs and CMTS.

Since the HFC network targeted for discovery operates in a stable RF environment, a useful linear RF frequency characterization mechanism is provided by S-parameters or scattering parameters [8, Appendix C], which are designed to capture the transmission and reflection characteristics of RF devices.

Even if actual S-parameter measurements of specific devices are not available, specification datasheets with return loss and transmission loss versus frequency datapoints or formulas describing behavior versus frequency can be converted into S-parameter matrices needed for modelling. S-parameters describe system reflection and transmission characteristics using complex numbers (using amplitude and phase information). Phase information versus frequency allows network operators to derive group delay versus frequency, enabling delay versus frequency estimation across elements and end-to-end system.

Historically, the cable industry has described and specified network components using magnitude information. Components used at higher frequencies such as microwave frequencies are specified using S-parameters with phase information. As our industry operates at higher frequencies, it is only natural to evolve into using phase. The modelling of cable networks proposed by Narayanaswamy, Prodan and team [9] will achieve accurate results if we characterize and specify our components in magnitude and phase. In fact, our industry, through the CMTS and CM in-phase (I) and quadrature (Q) information already gathered, leverages amplitude and phase. This additional information of network elements will not just enable accurate modelling and simulation but will also improve detection, localization, and resolution of plant problems.

The two most prevalent element types in HFC networks are the amplifier and the tap. S-parameter characterization measured on a vector network analyzer (VNA) of a sample amplifier is shown in Figure 4, Figure 5, and Figure 6, and S-parameter characterization of a sample tap is shown in Figure 7 and Figure 8.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

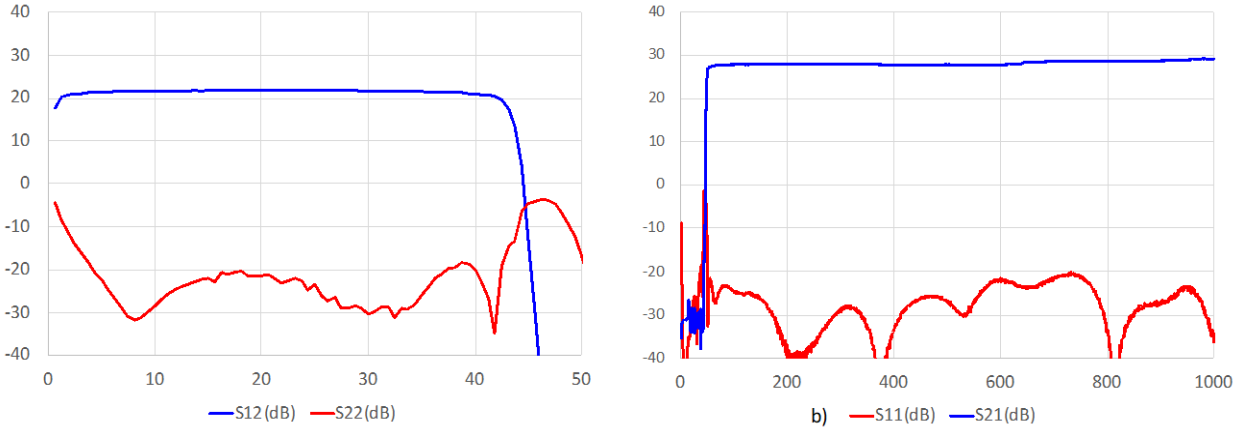


Figure 4 – Amplifier S-parameters a) reverse transmission and reflection b) forward transmission and reflection.

Figure 5 shows S-parameter characteristics of an amplifier upstream band. From the phase information in Figure 5a we can derive group delay shown in Figure 5b. This group delay is the delay in traversing the amplifier on the reverse path. While instrumentation will provide a good assessment of group delay, DOCSIS devices provides relative group delay information that may include additional elements such as the front end of a CM or CMTS.

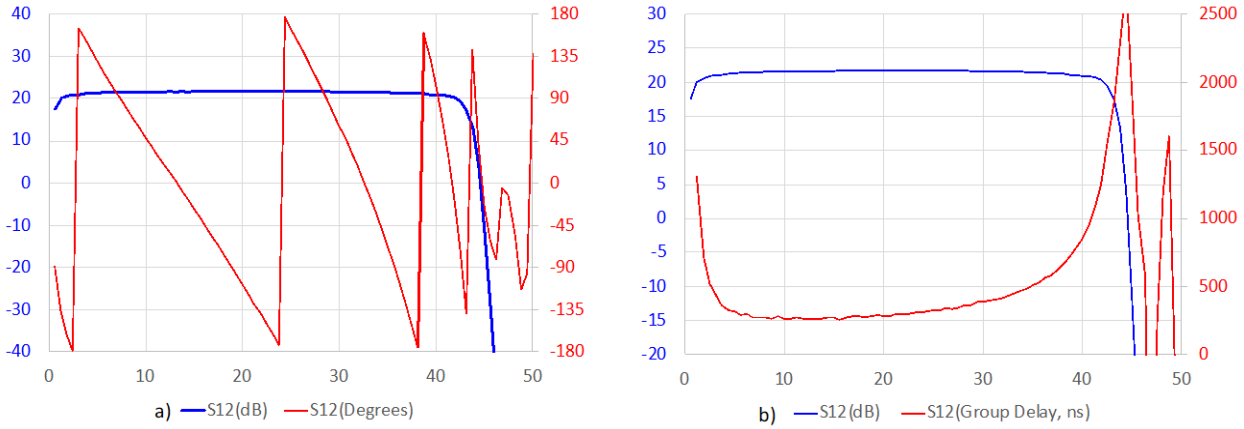


Figure 5 – Amplifier reverse transmission S-parameters a) magnitude and phase b) magnitude and group delay.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

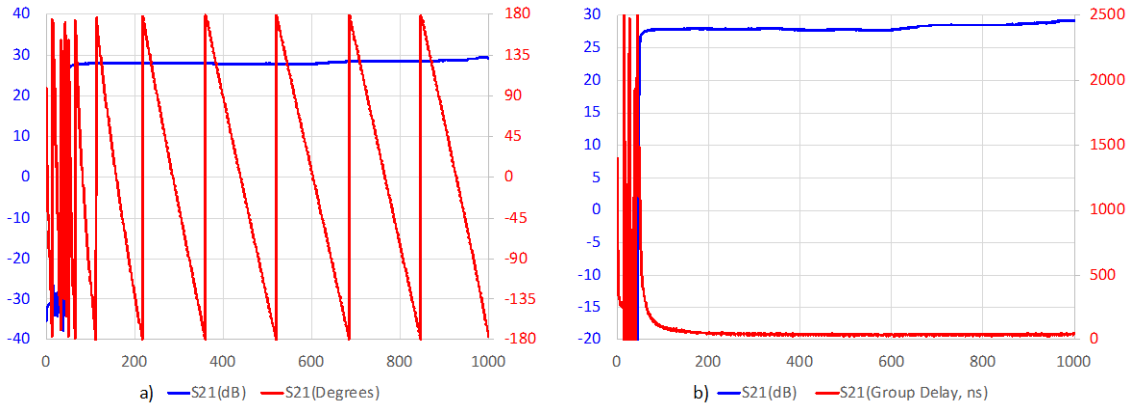


Figure 6 – Amplifier forward transmission S-parameters a) magnitude and phase b) magnitude and group delay.

A 17 dB four port tap main path insertion loss can be compared with coupled port loss (Figure 7).

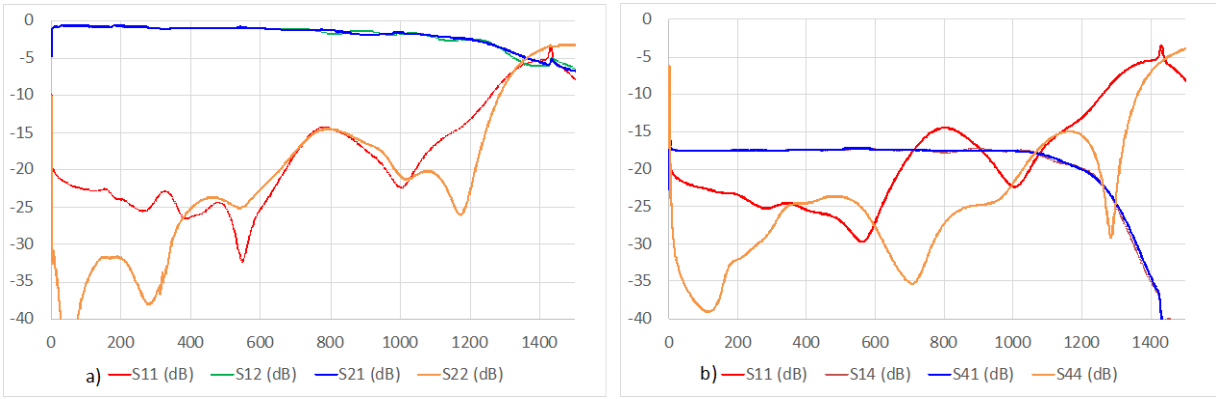


Figure 7 – 17 dB Tap S-parameters a) main path b) coupled port path.

In this example, the passive tap rated for 1 GHz can operate way beyond 1 GHz (Figure 8), introducing a transmission delay of about 9 ns.

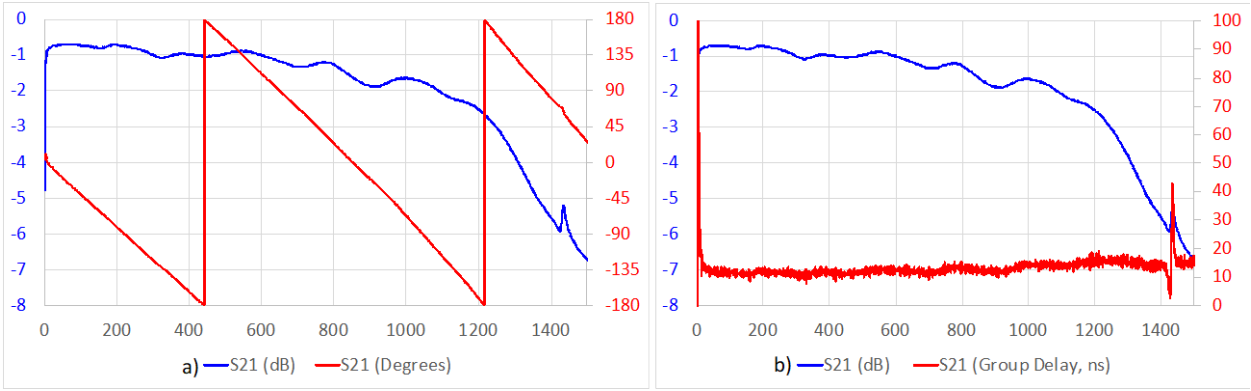


Figure 8 – 17 dB tap main path S-parameters a) magnitude and phase b) magnitude and group delay.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

DOCSIS channel estimation by CM or by CMTS through complex equalization coefficients provide transmission S-parameters assessment for the DOCSIS channel frequency range. Tom Williams [7] also shows a way to also obtain reflection S-parameter by inserting a CM probe into the network. One must keep in mind that CM and CMTS-based estimates will include internal distortions such as front-end CM/CMTS receiver distortion which would have to be calibrated out for an accurate delay assessment. Nevertheless, even without calibration, comparison of signal distortion from the same CM models can provide good insights of network characteristics.

While an S-parameter model is good for assessing an individual component, when assessing components that are connected in cascade, related and more practical parameters are defined, they are called the T-parameters. The relationship between S-parameters and T-parameters is shown in Appendix C.

If you have HFC elements A and B in cascade, each expressed in their T-parameter representation, the T-parameter equivalent of the cascaded system is given by the matrix product of the individual T-parameters matrices A and B.

$$[T_{Equiv}] = [T_A][T_B]$$

So, if you have a coaxial segment represented by cascading HFC network elements, they can be modelled using T-parameters.

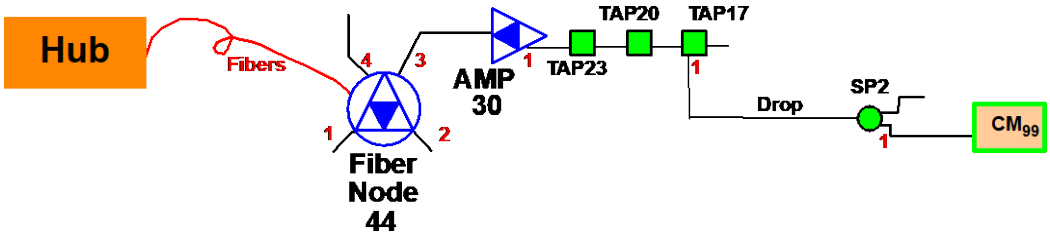


Figure 9 – Coaxial path within a fiber node serving area 44 to CM 99.

The coaxial segment using the long name convention described in [10] is represented as:

FN44$_3$-C.5(200)-AMP30 -TAP23-C.5(125)-TAP20 C.5(102)-TAP17$_1$-RG6(90)-SP2$_1$-RG6(30)-CM99

The port numbers in red in Figure 9 are used as subscripts in the long name representation, where the half inch hardline cable is depicted as C.5() with the number in parenthesis being the cable length in feet. The coaxial segment in Figure 9 is modeled using T parameters by cascading its individual T-parameters resulting from the following matrix product.

$$[T_{Equivalent}]=[T_{C.5(200)}][T_{AMP30}][T_{TAP23}][T_{C.5(125)}][T_{TAP20}][T_{C.5(102)}][T_{TAP171}][T_{RG6(90)}][T_{SP21}][T_{RG6(30)}]$$

So given a known input/output of FN44, the input/output at CM99 can be derived using the $T_{Equivalent}$ matrix representing the HFC elements in cascade. This analysis can be used when comparing measurement at the CM to measurements at the tap for estimating drop/home cable distance, or when comparing elements that share a common portion of a coaxial segment or to estimate the number of actives in cascade, etc.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Analysis

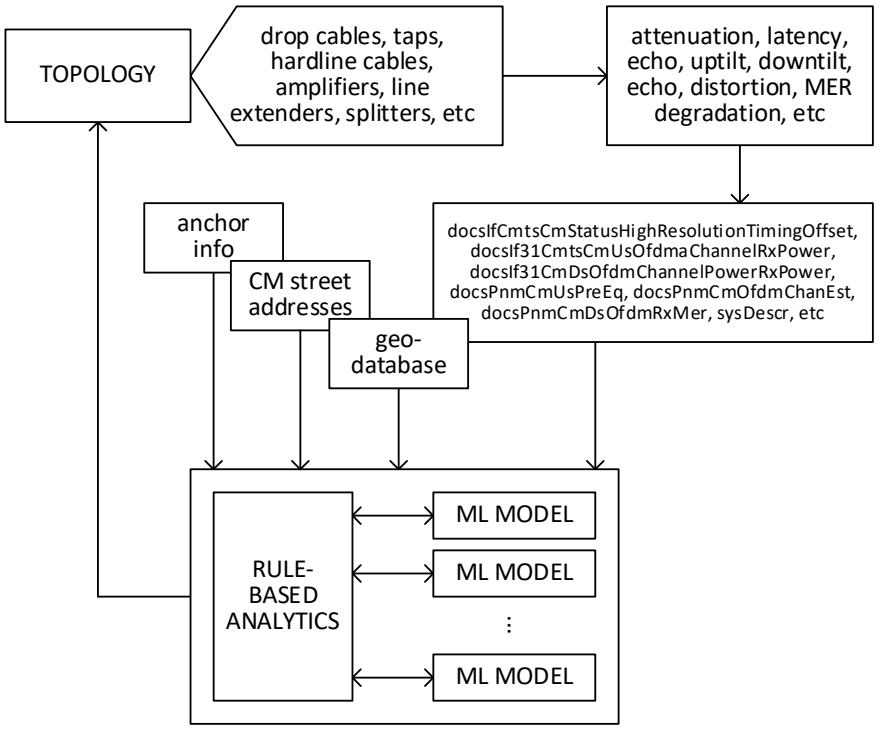


Figure 10 – Analysis flow of MIND.

Figure 10 presents the analysis flow of MIND in our preliminary study. The process begins by utilizing latitude-longitude information from an HFC network element, typically a fiber node, as an anchor point. Additionally, the analysis relies on street addresses of CMs and leverages lot parcel data and street layout information. To further enhance the understanding of the network and differentiate between different topologies, geodata information and common network deployment practices are incorporated. This includes factors like the typical locations of pedestals, the choice between underground and aerial deployment, and other relevant details. This information is represented in a logical format that can be embedded in the analytics.

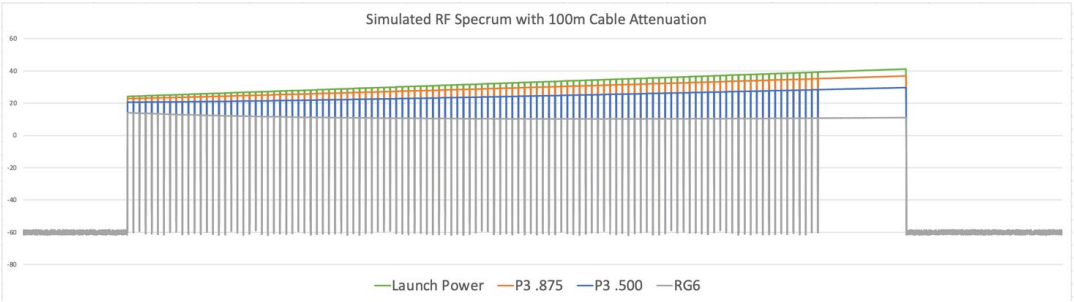


Figure 11 – Example of loss prediction of 100 meter coaxial cables.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

As DOCSIS signals traverse the network, they are influenced by various plant characteristics, leaving distinctive signatures on the signals (Figure 11). By analyzing and correlating these signal signatures, we can discover network elements, their connectivity, and characteristics. In the HFC network, every cable modem communicates with the CMTS through multiple network components, forming a unique topology. Each component contributes to the signal's channel response. For example, a drop cable causes greater power loss compared to a same-length hardline cable, and a tap introduces loss and potential spectrum uptilt in scenarios of taps with conditioning plug-in modules. Hardline cables introduce signal latency, where, for instance, a 1,000-ft cable with 87% velocity of propagation results in a 1170 ns latency. Active components introduce dispersion and roll-off, distorting the signal's magnitude and phase. Different CMs accumulate distinct channel responses induced by the network components and the topology.

DOCSIS facilitates the recording of this information. MIND leverages the granular monitoring tools available in DOCSIS, such as the DOCSIS MIB information from the CMTS and CMs within the analyzed fiber node serving area. DOCSIS 3.1 has introduced new tools like PNM with improved resolution, offering additional information that can expedite convergence to a decision and increase confidence in the results.

Using these sources of input information, the core analysis consists of two approaches: rule-based analytics and ML models. Rule-based analytics draw upon engineering knowledge to make decisive judgments rather than soft decisions. ML models enhance MIND's performance by detecting features that may not be immediately apparent through direct analysis or observation of management metrics. Supervised ML requires extensive training data for a complex HFC network. Therefore, comprehensive datasets encompassing the aforementioned information, as well as a logical representation of the topology, are necessary to support the ML training process.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.1. Rule-based Analytics

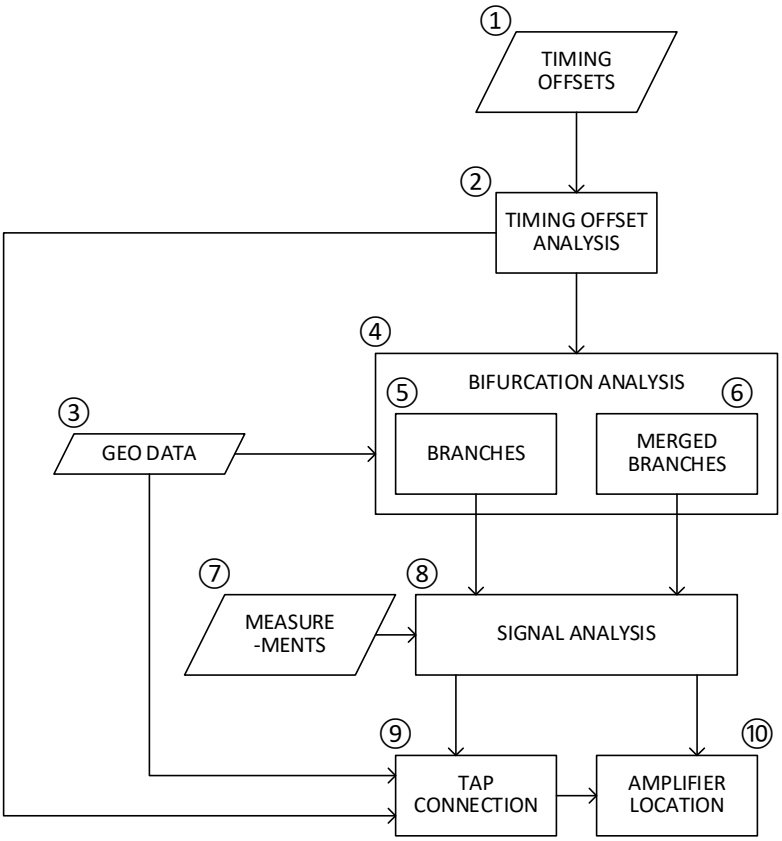


Figure 12 – Example flow of rule-based analytics in MIND.

As previously mentioned, the objective of the algorithm is to provide a comprehensive visualization of the cable network's layout on a map, including the precise locations of all network components. Straightforwardly, the algorithm follows rule-based analytics that align with the layout of the network. For instance, in a tree-like network, branches can be segmented by amplifiers; if two CMs have similar signatures except for a higher latency in one, the algorithm can infer that they likely belong to the same cable segment, with one being farther downstream than the other; additionally, the presence of different distortions suggests the presence of an amplifier between the two CMs.

In rule-based analytics, measurements that explicitly reflect topology relations are utilized. Timing offsets reflect relative CMTS-CM distances, street addresses reflect both location and candidate anchoring point, and signal-related measurements and their variations reflect channel conditions, among others. By adhering to these rules, the algorithm can hierarchically depict the network's structure and identify the relative positions of components. Figure 12 illustrates the following steps, which demonstrate an example of rule-based analytics.

Step 1: Timing Offset Information Collection

This initial step gathers the timing offset, which serves as latency information, from each CM. Tools like MIB, network management interface (NMI), and Simple Network Management Protocol (SNMP) are employed for this purpose.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Step 2: Timing Offset Analysis

Building upon the information obtained in Step 1, this step analyzes the relative latency between each cable modem and the CMTS by comparing and calibrating a large data set of CMs with redundant attribute combinations. By interpreting the latency as a measure of distance, the algorithm derives an estimate of the spatial arrangement.

Step 3: Geodata Collection

In this step, the algorithm collects geographical data for each cable modem, including details such as street address, latitude/longitude, and building parcels.

Step 4: Bifurcation Analysis

While all CMs analyzed are associated with the same fiber node, they may belong to various branches of coaxial cables connected to the node. Using the distance offset information obtained in Step 2 and the geodata from Step 3, this step determines which cable modems belong to the same cable branch.

Step 5: Branch Grouping

The outcome of Step 4 is a set of cable modems grouped based on their respective branches.

Step 6: Merged Branches

Considering the nature of cable connections, certain branches identified in Step 5, despite being separate, might belong to the same branch. This step identifies pairs of branches that are potentially cascaded head-to-tail and merges them into a single branch.

Step 7: Measurement Collection

Utilizing tools like MIB, NMI, and SNMP, this step gathers various measurements from each cable modem, including pre-equalization data, channel estimates, transmitted and received power, modulation error ratio (MER), spectrum density, and non-linearity measurements.

Step 8: Signal Analysis

This step examines the branches one-by-one. For each branch, it clusters the CMs belonging to that branch into small groups based on their measurements collected from Step 7. After adjusting the clustering conditions, these small groups are then clustered into larger groups.

Step 9: Tap Connection

Using the group clustering results from Step 8, this step determines which cable modems are connected to the same tap. The distance offset results from Step 2 guide the placement of these taps, which are anchored to pedestal or utility pole candidates identified in Step 3.

Step 10: Amplifier Location

Building on the large-group clustering results from Step 8 and the tap identification from Step 9, this final step estimates the existence and locations of amplifiers. Each amplifier's location is potentially associated with multiple adjacent pedestal or pole candidates that share a similar likelihood.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.2. ML-Based Analytics

In contrast to the conventional analysis flow of rule-based analytics, ML methods are employed to leverage the implicit information hidden within the measurements, rather than relying on explicit utilization. By applying ML techniques, these methods can extract valuable insights and patterns from the data that may not be easily discernible through explicit rule-based approaches. This allows for a more comprehensive and nuanced utilization of the underlying information contained within the measurements.

ML methods can primarily be categorized based on their degree of supervision or reinforcement. In our preliminary study, considering the limited availability of training data, we have investigated unsupervised clustering methods, including K-means, density-based spatial clustering of applications with noise (DBSCAN), and even as simple as manually drawing a threshold line in between data points. Among these methods, we have found that agglomerative clustering proves to be the most effective. It is used in Step 8 in Section 5.1 to cluster US power values (~10-dimension space), DS power values (~10-dimension space), and spectrum measurements (~1000-dimension space).

Agglomerative clustering initially treats each cable modem as an individual cluster and then progressively merges the closest pairs of clusters based on proximity. Various metrics, such as signal strength, latency, or spatial distance (or to be precise, the difference on a GIS grid), can be employed to determine the proximity of clusters. In our agglomerative clustering implementation, we adopted Ward linkage defined as

$$\frac{|A| \cdot |B|}{|A \cup B|}\|\mu_A - \mu_B\|^2 = \sum_{x \in A \cup B}\|x - \mu_{A\cup B}\|^2 - \sum_{x \in A}\|x - \mu_A\|^2 - \sum_{x \in B}\|x - \mu_B\|^2$$

where A and B are the two clusters to be merged, μ is the centroid of a cluster, and $\|\cdot\|$ expresses the difference between two values if scalar measurements or the 2-norm distance if vector measurements. Utilizing Ward linkage for cluster merging offers the advantage of minimizing variance growth. However, this approach has a time complexity of $\mathcal{O}(n^3)$, which is relatively slow when compared to alternative methods. Nonetheless, this should not pose significant issues since the algorithm is designed to run infrequently and with a limited number of CMs in each clustering operation.

When performing clustering, we have the option to utilize either multiple types of measurements or a single type as input. Using multiple types of measurements is often ideal in ML, as it allows us to consider potential dependencies or correlations among observations. However, encoding multi-type input data can be challenging when we have limited data available. In our preliminary work, we opted for using a single type of measurement, which resulted in multiple clustering results. To make joint decisions in this scenario, we employed Silhouette scores, which are calculated based on the clustering results. Specifically, for each value x within the set C of a measurement, assuming x is clustered into C_x by the clustering of $\{C_i\}$, we define the Silhouette score at x as

$$s_x = \frac{\min_{k \neq x} \frac{1}{|C_k|}\sum_{y \in C_k}\|x - y\| - \frac{1}{|C_x| - 1}\sum_{y \in C_x, y \neq x}\|x - y\|}{\max\left(\min_{k \neq x} \frac{1}{|C_k|}\sum_{y \in C_k}\|x - y\|, \frac{1}{|C_x| - 1}\sum_{y \in C_x, y \neq x}\|x - y\|\right)}.$$

The average score of the clustering for that measurement is determined by calculating the mean of all such scores, $\sum_{x \in C} s_x / |C|$. Finally, topology-related decisions are made based on the scores obtained from multiple measurements.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

On top of clustering models, it becomes possible to uncover even more complex and implicit patterns and relationships between cable modems and their associated network components by applying supervised learning techniques, such as classification or regression algorithms, to the available data. For instance, a decision tree can be employed to determine if a CM is adjacent to an active component; long short-term memory (LSTM) or a transformer can be used to ascertain the length of the next cable segment; support vector machine (SVM), or a kernel method can segmentate cable branches; and multilayer perceptron (MLP) can classify CMs with specific types of impairments and determine their proximity. These algorithms necessitate labeled datasets where cable modem measurements and locations are paired with known information about the network topology and component locations, or other suitable representations. The ML problem posed by a cable network is highly complex due to the vast number of potential latent spaces, correlations, and intricate features. Therefore, it is essential to explore a wide range of methods, models, and transformations. Equally important is the acquisition of a substantial amount of labeled data for training.

5.3. Data Processing

Both rule- and ML-based analytics require the calibration and normalization of input data before utilization. However, this calibration process can sometimes be tedious and ambiguous, primarily because some of the measurements collected from the plant were not originally designed for network discovery. For instance, the timing offset measurement, which serves as an intermediate parameter for ranging, is highly influenced by multiple factors. To illustrate this, in our tests, we observed that when two CMs of different models are in media access control (MAC) domain 1, CM A reports a timing offset 2.1531 microseconds (549 meters) greater than CM B for the same plant location. Similarly, when both CMs are in MAC domain 2, CM A reports a timing offset 2.1299 microseconds (543 meters) greater than CM B at the same plant location. Furthermore, MAC domain 2 exhibits a timing offset 0.0068 microsecond (1.73 meters) higher than MAC domain 1 for the same CM B and plant location. Conversely, MAC domain 2 reports a timing offset 0.0164 microsecond (4.18 meters) lower than MAC domain 1 for the same CM A and plant location. To make this data usable, calibration of these biases is necessary, requiring the collection of an extensive amount of data that covers all possibilities and is labeled with the ground truth values. This calibration process constitutes a major task within our project.

Normalization is also an essential step, particularly for MLP models in ML. As an example, we utilize the group delay of a signal to estimate various topological information, such as the presence of an amplifier. Group delay is derived from the DS channel estimate and US pre-equalization data through unwrapping and removing the linear phase shift of the complex data. This process necessitates a proper normalization practice to strike a balance between revealing and erasing the data's features while maximizing the accuracy of the ML model.

6. Implementation

The MIND prototype was implemented over an HFC network in the laboratory that mimics a network in the field (Figure 13). The network is designed to have the dimensions of a portion of a network located close to CableLabs' facilities in Louisville, Colorado. The advantage of building the network is that, for verification purposes, we know the exact dimensions and characteristics of each component and we have control of the CMTS and CMs used for this proof of concept. So even though we start with very basic information, we have all the DOCSIS CMTS/CM and PNM data as well as geodata from the replicated coaxial network. This re-created network has approximately 5,000 feet of coaxial cable and 47 DOCSIS 3.1 CMs of different models.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

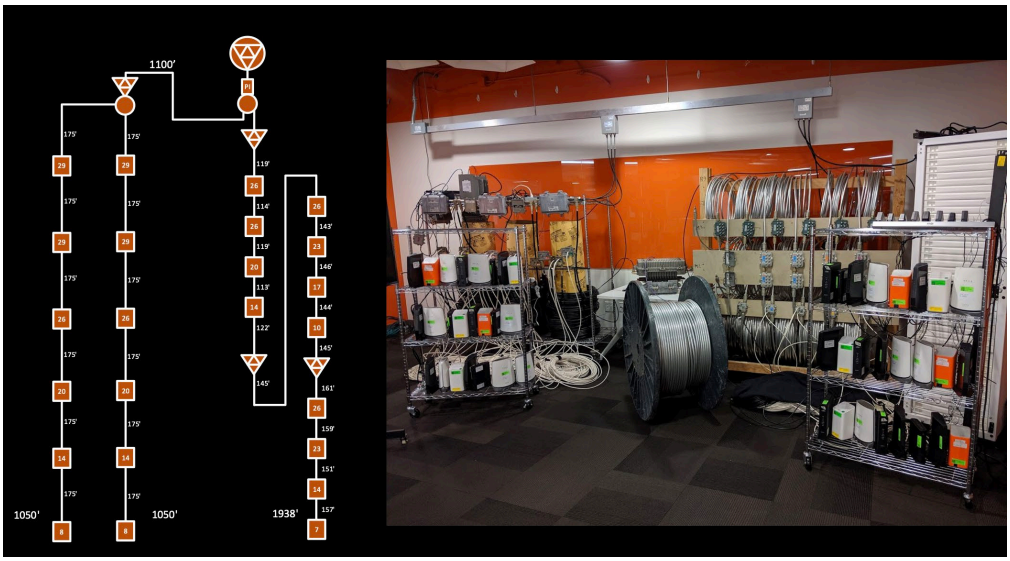


Figure 13 – Laboratory example of HFC network topology.

As the DOCSIS signals traverse the different HFC network components, they are modified or marked according to the characteristics of each traversed component, which by the time the accumulated signature is collected at the edge, it has the history of components traversed.

The diversity of metrics MIND collects provides a rich set of composite signatures that are analyzed to deduce network characteristics. As indicated at the beginning of the process, we assume very limited plant map knowledge. We use an anchor point with latitude and longitude information. A convenient anchor point is the fiber node. Using DOCSIS service group information we determine which CMs are associated with the fiber node. We also assume that we have access to street addresses that correspond to the subscriber's CMs. MIND translates street addresses to latitude and longitude. Even though this lat-lon information is approximate, it allows discovery of topology features. In the next step MIND gathers geodata information corresponding to parcel boundaries, construction footprint and street data. Even though there are exceptions, pedestal locations or pole locations are typically at the corner of the property lot to have easier access to homes served. This is one of the deployment rules that we leverage and provides a clue where pedestals or poles might be located. So, at this point we just have candidate pedestal or pole locations. Later processes will associate which amplifiers and taps are housed in which pedestal or mounted on which pole. These processes will also allow us to discard certain candidate pedestal/pole locations until we are left just with the actual locations. The deployment rules that we follow will vary if the plant is aerial or if the plant is underground. At this point we assume that we have this information, as we refine MIND algorithms this is something that could also be discovered.

One parameter that we take advantage of is timing offset. In DOCSIS 3.0, a higher resolution option was made available for TDMA and in DOCSIS 3.1 the timing offset granularity was increased even further to support subcarrier orthogonality. One challenge with timing offset is that it has been designed to adjust ranging between CMTS and a specific CM. There is an internal processing delay at the CM that changes depending on model and chipset. We need a process that normalizes timing offset across all CMs in a fiber node serving area so that delay comparison and therefore cable lengths can be estimated. Since we built this network, we know our coaxial cable lengths and we use this information to build a calibration table that removes this model/chipset variability. This way the enhanced resolution that can be obtained through timing offset can be realized.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The next process is to figure out where the bifurcations or branches are in the coaxial network. Different coaxial branches have different signatures. Analysis of these signatures allows MIND to cluster or group CMs that are connected through one branch or another. Distortion metrics such as equalization information are key in this clustering exercise.

The next task is to determine pedestal/pole locations. Remember that, at this point, we only have candidate pedestal/pole locations. Through analysis, MIND discards certain candidate pedestals/poles because if they would be in a specific candidate location, the timing offset would not be consistent with that pedestal/pole location. Clustering along with street address information can help determine if the coaxial distribution cable runs behind properties or in front of the homes.

The next process allows you to identify amplifier location relative to CMs. In the upstream, group delay at the band edges is very indicative if signals have traversed one, two, or more amplifiers. Downstream tilt characteristics, transmit and received power-level analysis, and other metrics allows you to determine how close you are to the amplifier, which can help to figure out if amplifier and tap share the same physical location (e.g., pedestal/cabinet or pole). After amplifier-to-pedestal/pole association, a similar process takes place to associate splitters and couplers to tap.

The final network topology discovered that we observed to be a good match was between the programmatically derived and the built network.

In this prototype implementation, not all possible tools have been leveraged. In an actual field scenario, it is expected the need for additional discovery algorithms that are only possible after training with a much larger data set. A limited set of different CM models allows us to generate a timing-offset calibration database. In the field you have a combination of DOCSIS 3.0 and DOCSIS 3.1 CMs. In the lab we had the luxury of using the higher capability DOCSIS 3.1 CMs. We must weigh in the pros and cons of whether to develop tools for DOCSIS 3.0 or wait for DOCSIS 3.1 to reach a higher percentage of deployment. Limiting MIND to DOCSIS 3.1 devices makes timing offset calibration easier as the population of different CM models and chipsets is more manageable. As we investigate field deployment, we also must verify that different CMTSs and CMs provide all the MIBs populated and formatted according to the DOCSIS specifications. There are still a good number of remaining tasks to migrate MIND into the field and as our networks cover higher frequencies the benefits are greater.

7. Future Work

This paper describes a complex process which involves significant data collection, processing, and analysis. While a significant amount of work has been done to explore the efficacy of the approaches described, there is much more work that needs to be done to deliver solutions to the use cases that were discussed in Section 3.

8. Conclusion

Programmatic discovery of HFC network topology and elements is becoming feasible as we have more and more accurate tools to deduce the HFC network characteristics. We have described an approach and have demonstrated it through the MIND prototype tool. MIND programmatically discovers the HFC network leveraging DOCSIS, PNM, geodata and deployment rules, without a pre-existing as-built map. MIND relies on a very comprehensive set of data and the calibration of the DOCSIS timing offset among other metrics. This paper discusses the challenges in using these techniques in the field and the benefits of the different use cases it enables. Prevalence of DOCSIS 3.1 technology in the network along with its full set of MIBs and capabilities provides better insight and higher resolution in the discovery process. As we

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

evolve to network discovery in the field, machine learning will play a greater role both in the generation of new discovery algorithms as well as in the increase in sensitivity to distinguish the more subtle features of the HFC network.

Abbreviations

AI	artificial intelligence
AC	alternating current
CAD	computer aided design
cm	centimeter
CM	cable modem
CMTS	cable modem termination system
dB	decibel
DBSCAN	density-based spatial clustering of applications with noise
DOCSIS	Data-Over-Cable Service Interface Specifications
DS	downstream
FBC	full band capture
GHz	gigahertz
GIS	geographic information system
HFC	hybrid fiber/coax
HOA	homeowner association
I	in-phase
LSTM	long short-term memory
MAC	media access control
MDU	multiple dwelling unit
MER	modulation error ratio
MHz	megahertz
MIB	management information base
MIND	Methodology for Intelligent Network Discovery
ML	machine learning
MLP	multilayer perceptron
mm	millimeter
NMI	normalized mutual information
ns	nanosecond
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
OTDR	optical time domain reflectometer
PNM	proactive network maintenance
pre-EQ	pre-equalization
ps	picosecond
Q	quadrature
RF	radio frequency
Rx	receive
SCDMA	synchronous code division multiple access
SC-QAM	single carrier quadrature amplitude modulation
SCTE	Society of Cable Telecommunications Engineers
SNMP	Simple Network Management Protocol

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SVM	support vector machine
TDMA	time division multiple access
TDR	time domain reflectometer
US	upstream
VNA	vector network analyzer
µs	microsecond

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

[1] Data-Over-Cable Service Interface Specifications DOCSIS® 3.1 Physical Layer Specification (CM-SP-PHYv3.1-I20-230419); Cable Television Laboratories

[2] Data-Over-Cable Service Interface Specifications DOCSIS® 3.1 Cable Modem Operations Support System Interface Specification (CM-SP-OSSIv3.1-I24-230613); Cable Television Laboratories

[3] Data-Over-Cable Service Interface Specifications DOCSIS® 3.1 Converged Cable Access Platform Operations Support System Interface Specification (CM-SP-CCAP-OSSIv3.1-I26-230613); Cable Television Laboratories

[4] "Pre-equalization based Pro-active Network Maintenance Methodology", SCTE Cable Tec Expo 2008 by Alberto Campos, Eduardo Cardona, and Lakshmi Raman (June 2008, Philadelphia, Pennsylvania)

[5] PNM Current methods and Practices in HFC Networks (DOCSIS® 3.1) CM-GL-PNMP-V04-230213, February 13, 2023, Cable Television Laboratories, Inc.

[6] Data-Over-Cable Service Interface Specifications DOCSIS® 3.1 MAC and Upper Layer Protocols Interface Specification (CM-SP-MULPIv3.1-I25-230419); Cable Television Laboratories

[7] "OFDMA Predistortion Coefficient and OFDM Channel Estimation Decoding and Analysis" SCTE Cable Tec Expo 2021 by Tom Williams, Alberto Campos, Lin Cheng, James Lin, Jason Rupe, and Jay Zhu (October 2021, Atlanta, Georgia)

[8] "Microwave Engineering" by David Pozar, (2011 Edition Chapter 4), Wiley and Sons, ISBN-13 978-0470631553

[9] "When Physical Layer Simulation Gets Real" SCTE Cable Tec Expo 2022 by Ramya Narayanaswamy, Karthik Subramanya, Dr. Richard Prodan & Larry Wolcott (October 2022, Philadelphia Pennsylvania)

[10] "Cable Network Management Infrastructure Evolution", L. Alberto Campos, Jennifer Andreoli-Fang, Vivek Ganti, NCTA 2014 (May 1 2014, Los Angeles California)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Appendix A

RF Network Elements

Abbreviated – full name (description/comments)

Attributes

SPLC – RF Distribution Splice (Connects coaxial segments together)

Diameter

Insertion Loss (dB) / Insertion Loss vs. Frequency

SP – RF Distribution Splitter (Splits/Combines RF energy from one port to two ports, sometimes 3 ports)

Number of Ports

Nominal Loss Value (dB)

Loss vs. Frequency

DC – RF Directional Coupler (Single port RF Coupler, Taps RF energy from main path to secondary path)

Value (dB)

Type

Coupling Loss vs. Frequency

TAP – RF Distribution Tap (Multiport RF Coupler, Taps RF energy from main path to drop ports)

Nominal Value (dB)

Number of Ports

Type ()

Insertion Loss vs. Frequency

Coupling Loss vs. Frequency

Term – Termination (End of line device to avoid reflections, terminates coaxial path)

Coax Cable (a.k.a Hardline) (Rigid coaxial transmission line carrying RF energy and AC power)

Length

Type – (e.g., 715QR, 540QR, 500P3)

Diameter

Velocity of Propagation

Atten vs. Frequency

AMP – Amplifier (Amplifies signal to compensate attenuation in cable and devices)

Housing

Number of Ports

Port(i)

Lo/Md/Hi Power

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Gain vs. Freq

FN – Fiber Node (Converts optical signal to RF signal, typical multiple RF ports)

Number of RF Ports

Port(i)

Lo/Md/Hi Power

Power vs. Freq

EQ – External Equalizer (compensates for cable's higher losses at higher frequencies, typically after traversing some coaxial cable distance)

Value (dB)

Max Freq

Drop – Coaxial Drop Cable (flexible coaxial cable that extends from the tap drop port to the premise/home point of entry)

Length

Type –Series 6, Series 11

Diameter

Velocity of Propagation

Atten vs. Frequency

PI – Power Inserter (Inserts AC power into coaxial hardline cable to power fiber node amplifiers and other devices)

Home AMP – Drop Amplifier

Home Splitter –

DOCSIS Network Elements

CMTS

Type (Integrated, RMD, RPD)

Configuration

CM

Model, Silicon, Version

Facility Elements

Pedestal

Pole

Cabinet

Splice box

Power Supply

Hub

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Appendix B

MIND PNM & DOCSIS MIBs

PNM

docsPnmCmUsPreEq
docsPnmCmUsPreEqPreEqCoAdjStatus
docsPnmCmOfdmChanEst
docsPnmCmDsOfdmRxMer`
docsPnmCmDsOfdmSym
docsPnmCmDsOfdmSymTrigGroupId
docsPnmCmDsOfdmSymPlcExtTimestamp
docsPnmCmDsOfdmSymFftLength
docsPnmCmDsOfdmFec
docsPnmCmDsOfdmFecSumType
docsPnmCmDsHist
docsPnmCmDsHistSymmetry
docsPnmCmDsHistDwellCnts
docsPnmCmDsHistHitCnts
docsPnmCmDsHistCntStartTime
docsPnmCmDsHistCntEndTime
docsPnmCmtsDsOfdmSym
docsPnmCmtsDsOfdmSymTrigGroupId
docsPnmCmtsDsOfdmSymFirstActSubCarIdx
docsPnmCmtsDsOfdmSymLastActSubCarIdx
docsPnmCmtsDsOfdmSymPlcExtTimestamp
docsPnmCmtsDsOfdmSymFftLength
docsPnmCmtsUsOfdmaAQProbe
docsPnmCmtsUsOfdmaAQProbeCmMacAddr
docsPnmCmtsUsOfdmaAQProbeUseIdleSid
docsPnmCmtsUsOfdmaAQProbePreEqOn
docsPnmCmtsUsOfdmaAQProbeTimeout
docsPnmCmtsUsOfdmaAQProbeNumSymToCapt
docsPnmCmtsUsOfdmaAQProbeMaxCaptSymbols
docsPnmCmtsUsOfdmaAQProbeNumSamples
docsPnmCmtsUsOfdmaAQProbeTimeStamp
docsPnmCmtsUsHist
docsPnmCmtsUsHistSymmetry
docsPnmCmtsUsHistDwellCnts
docsPnmCmtsUsHistHitCnts
docsPnmCmtsUsHistCntStartTime
docsPnmCmtsUsHistCntEndTime
docsPnmCmtsUsOfdmaRxMer
docsPnmCmtsUsOfdmaRxMerCmMac
docsPnmCmtsUsOfdmaRxMerPreEq
docsPnmCmtsUsOfdmaRxMerNumAvgs

CM

docsIf3CmSpectrumAnalysisMeasFrequency
docsIf3CmSpectrumAnalysisMeasAmplitudeData
docsIf31CmDsOfdmProfileStatsProfileId
docsIf31CmDsOfdmProfileStatsTotalCodewords
docsIf31CmDsOfdmProfileStatsCorrectedCodewords
docsIf31CmDsOfdmProfileStatsUncorrectableCodewords
docsIf31CmDsOfdmChannelBandIndex
docsIf31CmDsOfdmChannelPowerCenterFrequency
docsIf31CmDsOfdmChannelPowerRxPower
docsIf31CmUsOfdmaChanTxPower
docsIf31CmUsOfdmaChanSubcarrierZeroFreq
docsIf31CmUsOfdmaChanFirstActiveSubcarrierNum
docsIf31CmUsOfdmaChanLastActiveSubcarrierNum
docsIf31CmUsOfdmaChanNumActiveSubcarriers
docsIf31CmUsOfdmaChanSubcarrierSpacing
docsIf31CmUsOfdmaChanCyclicPrefix
docsIf31CmUsOfdmaChanRollOffPeriod
docsIf31CmUsOfdmaChanNumSymbolsPerFrame
docsIf31CmUsOfdmaMinislotCfgStateStartMinislotNum
docsIf31CmUsOfdmaMinislotCfgStateFirstSubcarrierId
docsIf31CmUsOfdmaMinislotCfgStateNumConsecutiveMinislots
docsIf31CmUsOfdmaMinislotCfgStateMinislotPilotPattern
docsIf31CmUsOfdmaMinislotCfgStateDataSymbolModulation
docsIf31RxChStatusPrimaryDsIndicator
docsIf31CmDsOfdmChanChannelId
docsIf31CmDsOfdmChanPlcFreq
docsIf31CmDsOfdmChanNumPilots
docsIf31CmDsOfdmChanTimeInterleaverDepth
docsIf31CmDsOfdmChanPlcTotalCodewords
docsIf31CmDsOfdmChanPlcUnreliableCodewords
docsIf31CmDsOfdmChanNcpTotalFields
docsIf31CmDsOfdmChanNcpFieldCrcFailures
docsIf31CmDsOfdmChanChanIndicator
docsIf31CmDsOfdmChanSubcarrierZeroFreq
docsIf31CmDsOfdmChanFirstActiveSubcarrierNum
docsIf31CmDsOfdmChanLastActiveSubcarrierNum
docsIf31CmDsOfdmChanNumActiveSubcarriers
docsIf31CmDsOfdmChanSubcarrierSpacing
docsIf31CmDsOfdmChanCyclicPrefix
docsIf31CmDsOfdmChanRollOffPeriod

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

CMTS

docsIfCmtsCmStatusMacAddress
docsIfCmtsCmStatusIpAddress
docsIfCmtsCmStatusInetAddress
docsIfCmtsCmStatusDownChannelIfIndex
docsIfCmtsCmStatusHighResolutionTimingOffset
docsIfCmtsCmStatusTimingOffset
docsIf3CmtsCmUsStatusChIfIndex
docsIf3CmtsCmUsStatusHighResolutionTimingOffset
docsIf3CmtsCmUsStatusIsMuted
docsIf3CmtsCmUsStatusRangingStatus
docsIf3CmtsCmUsStatusModulationType
docsIf3CmtsCmUsStatusRxPower
docsIf3CmtsCmUsStatusSignalNoise
docsIf3CmtsCmUsStatusMicroreflections
docsIf3CmtsCmUsStatusEqData
docsIf3CmtsCmUsStatusUnerroreds
docsIf3CmtsCmUsStatusCorrecteds
docsIf3CmtsCmUsStatusUncorrectables
docsIf31CmtsDsOfdmChanChannelId
docsIf31CmtsDsOfdmChanNumActiveSubcarriers
docsIf31CmtsDsOfdmChanSubcarrierSpacing
docsIf31CmtsDsOfdmChanLowerGuardbandWidth
docsIf31CmtsDsOfdmChanUpperGuardbandWidth
docsIf31CmtsDsOfdmChanCyclicPrefix
docsIf31CmtsDsOfdmChanRollOffPeriod
docsIf31CmtsDsOfdmChanTimeInterleaverDepth
docsIf31CmtsDsOfdmChanNumPilots
docsIf31CmtsDsOfdmChanPilotScaleFactor
docsIf31CmtsDsOfdmChanNcpModulation
docsIf31CmtsDsOfdmChanLowerBdryFreq
docsIf31CmtsDsOfdmChanUtilization
docsIf31CmtsDsOfdmChanUpperBdryFreq
docsIf31CmtsDsOfdmChanLowerBdryEncompSpectrum
docsIf31CmtsDsOfdmChanUpperBdryEncompSpectrum
docsIf31CmtsDsOfdmChanPlcFreq
docsIf31CmtsDsOfdmChanSubcarrierZeroFreq
docsIf31CmtsDsOfdmChanFirstActiveSubcarrierNum
docsIf31CmtsDsOfdmChanLastActiveSubcarrierNum
docsIf31CmtsDsOfdmSubcarrierStatusStartId
docsIf31CmtsDsOfdmSubcarrierStatusEndId
docsIf31CmtsDsOfdmSubcarrierStatusModulation
docsIf31CmtsDsOfdmChanPowerBandIndex
docsIf31CmtsDsOfdmChanPowerCenterFrequency
docsIf31CmtsDsOfdmChanPowerTxPower
docsIf31CmtsUsOfdmaChanTemplateIndex
docsIf31CmtsUsOfdmaChanPreEqEnable
docsIf31CmtsUsOfdmaChanFineRngGuardband
docsIf31CmtsUsOfdmaChanFineRngNumSubcarriers
docsIf31CmtsUsOfdmaChanFineRngPreambleLen
docsIf31CmtsUsOfdmaChanInitRngGuardband
docsIf31CmtsUsOfdmaChanInitRngNumSubcarriers
docsIf31CmtsUsOfdmaChanInitRngPreambleLen
docsIf31CmtsUsOfdmaChanProvAttribMask
docsIf31CmtsUsOfdmaChanTxBackoffStart
docsIf31CmtsUsOfdmaChanTxBackoffEnd
docsIf31CmtsUsOfdmaChanConfigChangeCt
docsIf31CmtsUsOfdmaChanRangingBackoffStart
docsIf31CmtsUsOfdmaChanRangingBackoffEnd
docsIf31CmtsUsOfdmaChanUtilization
docsIf31CmtsUsOfdmaChanId
docsIf31CmtsUsOfdmaChanTargetRxPower
docsIf31CmtsUsOfdmaChanLowerBdryFreq
docsIf31CmtsUsOfdmaChanUpperBdryFreq
docsIf31CmtsUsOfdmaChanSubcarrierSpacing
docsIf31CmtsUsOfdmaChanCyclicPrefix
docsIf31CmtsUsOfdmaChanNumSymbolsPerFrame
docsIf31CmtsUsOfdmaChanRollOffPeriod
docsIf31CmtsUsOfdmaDataIucStatsDataIuc
docsIf31CmtsUsOfdmaDataIucStatsMinislotPilotPattern
docsIf31CmtsUsOfdmaDataIucStatsMinislotModulation
docsIf31CmtsUsOfdmaDataIucStatsTotalCodewords
docsIf31CmtsUsOfdmaDataIucStatsCorrectedCodewords
docsIf31CmtsUsOfdmaDataIucStatsUnreliableCodewords
docsIf31CmtsUsOfdmaDataIucStatsInOctets
docsIf31CmtsUsOfdmaDataIucStatsCtrDiscontinuityTime
docsIf31CmtsUsOfdmaDataIucDetailStatusLowerFreq
docsIf31CmtsUsOfdmaDataIucDetailStatusUpperFreq
docsIf31CmtsUsOfdmaDataIucDetailStatusMinislotPilotPattern
docsIf31CmtsUsOfdmaDataIucDetailStatusMinislotModulation
docsIf31CmtsUsOfdmaRangingIucStatusIuc
docsIf31CmtsUsOfdmaRangingIucStatusGuardband
docsIf31CmtsUsOfdmaRangingIucStatusNumSubcarriers
docsIf31CmtsDsOfdmSubcarrierTypeStartSubcarrierId
docsIf31CmtsDsOfdmSubcarrierTypeEndSubcarrierId
docsIf31CmtsDsOfdmSubcarrierTypeSubcarrierType

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

CMTS cont'd

docsIf31CmtsUsOfdmaSubcarrierTypeStartSubcarrierId
docsIf31CmtsUsOfdmaSubcarrierTypeEndSubcarrierId
docsIf31CmtsUsOfdmaSubcarrierTypeSubcarrierType
docsIf31CmtsCmRegStatusAssignedEmIds
docsIf31CmtsCmRegStatusPartialChanState
docsIf31CmtsCmRegStatusDsProfileIdList
docsIf31CmtsCmRegStatusUsProfileIucList
docsIf31CmtsCmRegStatusTcsPhigh
docsIf31CmtsCmRegStatusTcsDrwTop
docsIf31CmtsCmRegStatusMinUsableDsFreq
docsIf31CmtsCmRegStatusMaxUsableDsFreq
docsIf31CmtsCmRegStatusMaxUsableUsFreq
docsIf31CmtsCmRegStatusPartialSvcState
docsIf31CmtsCmUsOfdmaChannelRxPower
docsIf31CmtsCmUsOfdmaChannelRangingStatus
docsIf31CmtsCmUsOfdmaChannelCurPartialSvcReasonCode
docsIf31CmtsCmUsOfdmaChannelLastPartialSvcTime
docsIf31CmtsCmUsOfdmaChannelLastPartialSvcReasonCode
docsIf31CmtsCmUsOfdmaChannelNumPartialSvcIncidents
docsIf31CmtsCmUsOfdmaChannelNumPartialChanIncidents
docsIf31CmtsCmUsOfdmaChannelMeanRxMer
docsIf31CmtsCmUsOfdmaChannelStdDevRxMer
docsIf31CmtsCmUsOfdmaChannelRxMerThreshold
docsIf31CmtsCmUsOfdmaChannelThresholdRxMerValue
docsIf31CmtsCmUsOfdmaChannelThresholdRxMerHighestFreq
docsIf31CmtsCmUsOfdmaChannelMicroreflections
docsIf31CmtsCmUsOfdmaChannelHighResolutionTimingOffset
docsIf31CmtsCmUsOfdmaChannelIsMuted
docsIf31CmtsCmUsOfdmaProfileTotalCodewords
docsIf31CmtsCmUsOfdmaProfileCorrectedCodewords
docsIf31CmtsCmUsOfdmaProfileUnreliableCodewords
docsIf31CmtsCmUsOfdmaProfileLastPartialChanTime
docsIf31CmtsCmUsOfdmaProfileLastPartialChanReasonCode
docsIf31CmtsCmUsOfdmaProfilePartialChanReasonCode
docsIf31CmtsCmDsOfdmChannelCurPartialSvcReasonCode
docsIf31CmtsCmDsOfdmChannelLastPartialSvcTime
docsIf31CmtsCmDsOfdmChannelLastPartialSvcReasonCode
docsIf31CmtsCmDsOfdmChannelNumPartialSvcIncidents
docsIf31CmtsCmDsOfdmChannelNumPartialChanIncidents
docsIf31CmtsCmDsOfdmProfilePartialChanReasonCode
docsIf31CmtsCmDsOfdmProfileLastPartialChanTime
docsIf31CmtsCmDsOfdmProfileLastPartialChanReasonCode

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Appendix C

S-parameters and T-parameters

Using S-parameter definitions, a multiport network can be expressed in terms of the incident voltage waves and the outgoing or reflected voltage waves. A two port S-parameter network has an incident wave a_1 and an outgoing or reflected wave b_1 on port 1 and an incident wave a_2 and an outgoing or reflected wave b_2 on port 2 (Figure Y)

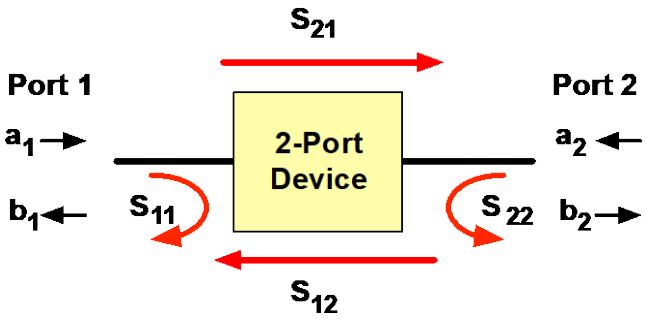


Figure Y - Two-port scattering or S-parameters defined by incident and reflected or outgoing waves.

The ratio of the reflected wave on port 1 and the incident wave on port 2 when there is no incident wave on port 2 is S_{11} parameter

$$S_{11} = \left.\frac{b_1}{a_1}\right|_{a_2=0}$$

The ratio of the outgoing wave on port 2 and the incident wave on port 1 when there is no incident wave on port 2 is S_{21} parameter

$$S_{21} = \left.\frac{b_2}{a_1}\right|_{a_2=0}$$

The ratio of the outgoing wave on port 1 and the incident wave on port 2 when there is no incident wave on port 1 is S_{12} parameter

$$S_{12} = \left.\frac{b_1}{a_2}\right|_{a_1=0}$$

The ratio of the reflected wave on port 2 and the incident wave on port 2 when there is no incident wave on port 1 is S_{22} parameter

$$S_{22} = \left.\frac{b_2}{a_2}\right|_{a_1=0}$$

These S-parameters in matrix representation are described as:

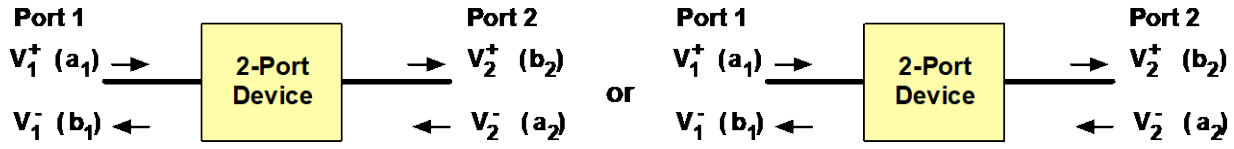
© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

$$\begin{vmatrix} b_1 \\ b_2 \end{vmatrix} = \begin{vmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{vmatrix} \cdot \begin{vmatrix} a_1 \\ a_2 \end{vmatrix}$$

$$S_{11} = \frac{b_1}{a_1}\Big|_{a_2=0} \quad S_{21} = \frac{b_2}{a_1}\Big|_{a_2=0} \quad S_{12} = \frac{b_1}{a_2}\Big|_{a_1=0} \quad S_{22} = \frac{b_2}{a_2}\Big|_{a_1=0}$$

$$\Rightarrow \quad \text{2-Port S-Parameter Matrix} = \begin{vmatrix} S_{11} & S_{21} \\ S_{12} & S_{22} \end{vmatrix}$$

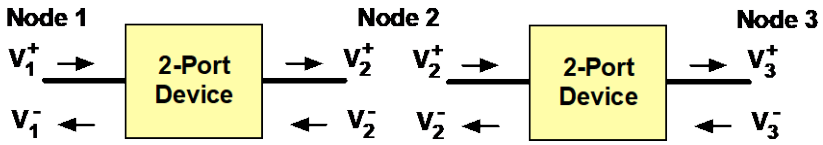
Also using incident and reflected or outgoing voltage waves, T-parameters are defined but is a slightly different fashion.



Where the + and – notation indicates the direction of the wave and leads to the following matrix representation

$$\begin{vmatrix} a_1 \\ b_1 \end{vmatrix} = \begin{vmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{vmatrix} \cdot \begin{vmatrix} b_2 \\ a_2 \end{vmatrix} \qquad \begin{vmatrix} V_1^+ \\ V_1^- \end{vmatrix} = \begin{vmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{vmatrix} \cdot \begin{vmatrix} V_2^+ \\ V_2^- \end{vmatrix}$$

The second matrix shows the usefulness in cascading elements to obtain the equivalent system representation



This way of defining T-parameters allows for representing systems comprised of cascading elements.

$$\begin{vmatrix} V_1^+ \\ V_1^- \end{vmatrix} = \begin{vmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{vmatrix} \cdot \begin{vmatrix} T_{22} & T_{23} \\ T_{32} & T_{33} \end{vmatrix} \cdot \begin{vmatrix} V_3^+ \\ V_3^- \end{vmatrix}$$

From the T-parameter definition we can obtain its relationship with S-parameters that are typically obtained from the individual characterization of network element.

$$T_{11} = \frac{-\det(S)}{S_{21}} \qquad T_{12} = \frac{S_{11}}{S_{21}}$$

$$T_{21} = \frac{-S_{22}}{S_{21}} \qquad T_{22} = \frac{1}{S_{21}}$$

$$\det(S) = S_{11}\,S_{22} - S_{12}\,S_{12}$$

Providing us the flexibility to go back and forth between T- and S-parameters.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Exploring the Benefits of Network Intelligence Applications to Optimize HFC Networks Using a Data-Driven Design Approach

A Technical Paper prepared for SCTE by

Diana Linton
Principal Engineer I
Charter Communications
14810 Grasslands Drive Englewood, CO 80112
720-536-1027
Diana.linton@charter.com

Esteban Sandino
Distinguished Engineer
Charter Communications
14810 Grasslands Drive Englewood, CO 80112
720-518-2318
Esteban.Sandino@charter.com

Nader Foroughi
CTO Americas
Technetix Inc.
8490 Upland Drive, Suite 200, Englewood, CO 80112
+1 403 606 7427
nader.foroughi@technetix.com

Keith Auzenne, Principal Engineer III, Charter Communications

Premton Bogaj, VP Product Management, Technetix Inc.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 3
2. Factors Influencing RF Level Optimization in Hybrid Fiber Coax (HFC) Networks 3
   2.1. RF Levels 4
   2.2. Signal Quality 4
   2.3. Noise and Nonlinear Distortions 4
   2.4. Modulation Technologies 5
3. HFC Network Optimization Opportunities 5
   3.1. Coexistence with Legacy Systems 6
   3.2. Interference and Noise Mitigation 6
   3.3. HFC Network Reliability and Redundancy 6
   3.4. Capacity Expansion and Future-Proofing 6
   3.5. Network Monitoring 6
   3.6. Upstream and Dowstream Power Balance 7
4. An Updated HFC Network Upgrade Design Approach 7
   4.1. Statistical Analysis Tools to Enhance HFC Network Design 9
5. Incremental Benefits of Optimizing HFC Networks with AI and ML 11
   5.1. Improve Signal Quality and Performance 11
   5.2. HFC Predictive Maintenance 11
   5.3. Enhance Network Capacity and Efficiency 12
   5.4. Minimization of Interference and Noise 12
   5.5. Extending the Reach of HFC Networks 12
   5.6. Power Efficiency 13
6. Future Directions and Emerging Technologies 13
7. Case Study 15
8. Conclusion 18

Abbreviations 19

Bibliography & References 20

List of Figures

Title **Page Number**

Figure 1 – Gaussian distribution [7] 8
Figure 2 – CCN (≈MER) with a 20log addition factor for CIN vs Cascade Depth 10
Figure 3. Traditional HFC segment 14
Figure 4. HFC segment with SDN 15
Figure 5. CM Rx level histogram and cumulative curve 16
Figure 6. CM Tx level histogram and cumulative curve 16
Figure 7. CM Rx MER Histogram and cumulative curve 17
Figure 8. CM US SNR histogram and cumulative curve 18

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The ever-increasing demand for high-speed internet connectivity has led to the continuous evolution of Hybrid Fiber-Coaxial (HFC) access networks. To meet these demands efficiently, the adoption of data-driven design and operational approaches has become crucial. This paper delves into the benefits of data-driven design in HFC access networks, outlining how it can lead to improved efficiencies. Through the integration of advanced analytics, predictive modeling and optimization techniques, data-driven design empowers network planners and operators to make informed decisions, allocate resources more efficiently, and enhance network performance, which can lead to additional benefits including: reducing HFC networks' operational costs and their environmental impact through improved energy efficiency; extending the lifespan of network components, enhancing their reliability, and reducing the need for maintenance; improved signal quality and performance for end-users; and creating a more harmonious coexistence with adjacent frequency bands and other networks and services, enhancing overall network stability and customer satisfaction.

In the era of digital transformation, access networks play a pivotal role in delivering high-speed internet services to end-users. HFC networks, a combination of fiber and coaxial cable technologies, have long been a cornerstone of broadband infrastructure. However, the growing demand for bandwidth-intensive applications and the advent of new technologies necessitate innovative approaches particularly for network design. Data-driven design, which involves utilizing data analytics and modeling techniques to guide decision-making, emerges as a powerful tool to optimize HFC access networks.

The paper begins with an overview of traditional HFC architecture. Then we discuss key factors that influence Radio Frequency (RF) level optimization, including signal quality, noise, power levels and modulation schemes. Next, we explore challenges related to RF level optimization. Following that, we present the benefits of implementing Artificial Intelligence and/or Machine Learning (AI/ML) tools. Here we examine ways to leverage Cable Modem (CM) data to drive AI/ML tools to optimize HFC network design, performance, and efficiency. We will explore using CM data to dynamically optimize amplifier gain, output level, and performance while staying well above amplifier thermal noise floors and avoiding operation within the amplifier non-linear distortion region.

Then we explore the use of a data-driven approach using machine learning techniques for identifying RF signal issues and developing automated solutions for the detection and resolution of signal-impacting events.

We conclude with an analysis of signal data collected from approximately 14,000 CMs. The analysis examines the range of RF transmit and receive levels across all CMs to determine available operating margins for a given desired performance. We will also explore how CM RF operating margins can be used to help set the appropriate RF gain for outside plant amplifiers. This approach will seek to take advantage of the smart amplifier capabilities expected to be part of the next generation of 1.8 GHz amplifiers. Optimizing amplifier gain to better reflect actual network conditions may reduce the maximum gain requirements for individual amplifiers and lead to more optimal RF designs and deployments.

2. Factors Influencing RF Level Optimization in Hybrid Fiber Coax (HFC) Networks

The HFC network has been a foundation for the cable industry for decades. It combines optical fiber and coaxial cable technologies to deliver broadband services. Fiber optic and coaxial technologies enable operators to deliver multiple services to numerous users simultaneously.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

As demand for high data rates and reliable communication increases, operators assess factors that affect connectivity. Thus, factors such as RF levels and signal quality are critical because they directly impact the internet connection's performance and stability. These factors also influence range and coverage for active devices. For instance, a strong and stable RF signal allows active equipment in the HFC network to communicate over longer distances without significant signal degradation. Conversely, weak or low-quality signals can limit the effective transmission range and may result in areas with poor connectivity.

2.1. RF Levels

In an HFC network, each active component receives the proper RF signal level from the previous active device and passes the correct signal level on to the succeeding active device. If the RF signal level is too low, the signal can be obscured by noise or interference. On the other hand, if the RF signal level is too high, the active device performance is nonlinear, which may cause distortions or, eventually, damage. Therefore, RF levels and signal quality are crucial for CMs' optimal operation. However, the strength of the RF signal alone does not necessarily determine how well the CM will perform in an HFC network. Thus, signal level into the CM should be carefully managed to prevent RF energy from being wasted.

Later in the paper we will explore the benefits of network intelligence applications that implement machine learning techniques to dynamically adjust RF parameters in active devices to optimize and better control RF input levels.

2.2. Signal Quality

RF signal quality is one of the most influential factors affecting the overall performance and reliability of HFC networks. It will dictate the highest levels of modulation orders the HFC network will support. Therefore, maintaining high-quality RF signals ensures downstream data remains accurate and undistorted when reaching the CM. Similarly, in the upstream direction, data integrity is crucial to optimizing performance at the input of the CMTS. Poor signal quality can lead to data errors, corruption, and even complete information loss, which may result in degraded performance or even system failure. How well the quality of the signal is managed will determine the service tiers and data rates that the network will support.

Later we will examine how network operators can determine correlations using a data-driven approach. For instance, how signal quality, CM received power levels, and performance (Modulation Error Rate (MER) in the downstream) combine to affect overall downstream and upstream throughput. We will then discuss how to leverage these correlations for the optimal configuration and set up of CM devices.

2.3. Noise and Nonlinear Distortions

RF signals are susceptible to various types of noise, which may significantly degrade signal quality. Noise and impairments that can degrade RF signals include: thermal noise, impulse noise, flicker noise, phase noise and intrinsic noise. These noise types may lead to poor signal reception, data corruption and communication errors. Thus, to provide high speed internet services that are reliable and efficient, network operators must understand the different types of noise, their sources, and how to minimize their impact. For example, optimal performance can be achieved by managing the input level into the amplifiers and staying as far away as possible from the thermal noise floor of all active devices. However, to avoid overdriving the power amplifier and CM devices, the optimal input level must be balanced against amplifier gain and output levels so as not to create unacceptable distortion levels.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In HFC networks, nonlinear distortion limits RF output levels and determines RF signals' maximum reach. It also determines the maximum number of RF carriers supported by active devices. Now that HFC networks use only digital carriers, the resulting nonlinear distortions are noise-like signals, described as Composite Intermodulation Plus Noise (CIN), which are generally indistinguishable from random noise. CIN products are signals generated by nonlinear interactions in RF amplifier components and changes in the input signals that may affect the frequency and amplitude of these intermodulation products [1]. This is why it is critical to carefully manage Total Composite Power (TCP) in active devices.

This kind of optimization requires continuous data analysis which is why network intelligence tools are extremely beneficial. Intelligent networking tools can help operators identify noise in HFC networks and take appropriate corrective action.

2.4. Modulation Technologies

Network operators have implemented advanced Orthogonal Frequency Division Multiplexing (OFDM) modulation technologies to increase transmission rates and overall throughputs. However, as the number of symbols and QAM modulation orders increases, sensitivity to noise and distortions also increases. Consequently, symbols are more difficult to isolate at the CM receiver [2]. OFDM uses low-rate subcarriers (long symbol times) orthogonal to each other and closely spaced with a cyclic prefix to ensure robust and efficient transmission. Moreover, OFDM subcarriers are modulated and spread over a wide frequency range. This allows a large amount of data to be transmitted in parallel over a single channel. The cyclic prefix acts as a guard interval that ensures that the transmitted symbols are orthogonal and the signal is faded [3]. OFDM technology allows operators to provide a variety of services to a wide range of users and applications, and allows operators to prevent narrowband interference by nulling sub-carriers.

However, OFDM can have a high crest factor, which is defined as the ratio of the peak to the average value of a signal. At the amplifier output stage, it contributes to determining the power margin to be reserved. Because of this, OFDM systems require efficient power amplification techniques to mitigate the impact of the crest factor.

Statistical functions, such as the complementary cumulative distribution function (CCDF), will be critical to specifying the power characteristics of the signals that will be amplified, mixed, and decoded in communication systems. For instance, CCDF can help reveal how an RF power amplifier may distort a complex modulated signal such as an OFDM channel. By comparing CCDF plots with different amplifiers at different average power levels, it could validate the linearity of the device. It can also reveal the introduction of data errors that can be caused by amplifier compression. The power of machine learning will help us leverage the use of CCDF plots as another tool for network operators to characterize and optimize device performance [4].

3. HFC Network Optimization Opportunities

As network operators begin 10G upgrades, complexities related to the expansion of the spectrum need to be addressed for HFC networks to perform optimally.

The complexities presented in this section have traditionally been addressed by network operators through a combination of technical expertise, manual processes, and collaboration between network engineers, operation teams, and vendors. We will present later how network operators can tackle complexities more efficiently by implementing network intelligence tools.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.1. Coexistence with Legacy Systems

Over time, upgrading an HFC network to 10G will require operators to replace Customer Premises Equipment (CPE), which cannot be completed for all subscribers instantaneously. For a period of time, operators will need to support both legacy CPE and DOCSIS 4.0 CMs. To support legacy CPE while upgrading the HFC network, tap output power levels in the legacy spectrum must remain close to their pre-upgrade values. During the RF spectrum expansion period, optimal performance of both legacy CPE and CM could be at risk. RF spectrum expansions will also drive tap replacements and updated faceplate values to ensure proper RF input to CMs [5]. CM performance is also affected by the drop. It is noteworthy, however, that cable length and type will vary across installations.

3.2. Interference and Noise Mitigation

Operators focus on minimizing distortion and noise to improve HFC network alignment, allowing CMs to use higher-order modulation schemes. In an ideal HFC network, there would only be downstream and upstream modulated RF signals. There are unwanted signals, however, that are a consequence of real-world applications, such as HFC networks that transmit RF signals using coaxial cables that are subject to noise and distortion. Challenges in RF level optimization include identifying the sources of interference and noise, designing a system that is resilient to interference and noise, and maintaining a constant level of signal power across the length of the coaxial cable. Thus, when network operators implement network intelligence tools, interference and noise mitigation should be considered as part of the optimization process. Some AI tools are already being deployed to predict and adapt to changes in the environment.

3.3. HFC Network Reliability and Redundancy

Identifying and resolving potential network issues fast and accurately is imperative for network operators. However, doing so in remote areas may be particularly difficult as gaining access to the physical network can be time-consuming and costly. This concern is why network operators implement redundant systems such as power backup to ensure uninterrupted operations in case of failure. In a later section, we will discuss the benefits of deploying AI-based network analytics and monitoring tools to proactively identify and address network issues. Such tools can reduce network maintenance and operations costs, particularly in remote areas.

3.4. Capacity Expansion and Future-Proofing

As demand for bandwidth grows, network operators may implement several strategies to expand capacity and enhance their coverage. These include adding additional spectrum, upgrading HFC infrastructure and raising modulation orders. Modulation orders in particular can be optimized to fit those portions of the network that can reliably support higher transmission rates by utilizing AI/ML tools.

3.5. Network Monitoring

Due to a large number of components in the HFC network, implementing monitoring tools has sometimes been a challenge for network operators in that identifying issues maybe complex and time-consuming and network operators have often relied on manual debugging and troubleshooting. This method may delay the identification and resolution of network issues. We will present some of the benefits of implementing the right AI tools to help network operators automate the detection, diagnosis, and resolution of network issues to reduce operational time and improve the customer experience.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.6. Upstream and Dowstream Power Balance

Because residential consumers are typically consumers versus producers of information, HFC networks operate with asymmetric data traffic, mimicking consumer behavior with downstream traffic being much higher than upstream. Though outside the scope of this paper, this asymmetry may be challenging when higher upstream capacity is needed since it traditionally requires a physical change of diplex filters in active devices. However, emerging technologies such as full duplex (FDX) seek to implement dynamic allocation of upstream and downstream bandwidth without the use of diplex filters and leveraging AI/ML tools. There are several interesting papers and articles to explore this topic further [11], [12].

4. An Updated HFC Network Upgrade Design Approach

Generally, operators' primary objective when planning HFC capacity upgrades will be to expand usable RF spectrum while maintaining as many existing network components as possible. HFC network upgrades will typically focus on maintaining amplifier spacing and legacy levels to minimize customer impact. Thus, HFC RF spectrum expansions require updating RF parameters such as gain, tilt, input, and output RF levels for all actives, as well as updating tap output level specifications to overcome higher drop losses at higher operational frequencies.

Traditional network operators implement a drop-in approach to maintain amplifier legacy locations by installing a new amplifier module with higher downstream and/or upstream bandwidth, helping minimize upgrade downtime and cost. As higher attenuation values are introduced at 1,794 MHz, amplifier output levels must be raised. This requires low noise, high gain amplifiers with a high degree of linearity that can handle large signal swings without distorting the signal, while maintaining efficiency so the amplifier does not exceed existing power consumption. Operators have mainly relied on semiconductor technology to maximize amplifier TCP and performance over existing coaxial cable spans while maintaining a quality output. This also ensures consistent amplifier RF gain and output levels to maintain unity gain and facilitate amplifier alignment. Advances in power amplifier (PA) technology are allowing operators to expand their bandwidth up to 1.8 GHz while leveraging their existing infrastructure. Despite the semiconductor industry's continuous efforts to increase TCP, it might not be enough to sustain a linear tilt due to power amplifier limitations. Thus, TCP has become a limiting factor, which may become increasingly important as operators implement higher frequencies. To ensure TCP is managed properly, operators must control amplifier parameters tightly and optimize output levels and gain levels.

Another challenge in traditional HFC designs is that RF parameters are conservatively set to accommodate a wide variety of deployment scenarios. RF amplifier gain is maximized to ensure enough RF budget to overcome the longest coaxial spans in the HFC network and maintain a unity-gain based design. However, this approach could lead to excess RF levels into amplifiers deployed over shorter spans. The excess RF would then need to be padded down at the input of the RF amplifier to target prescribed operational input levels and to avoid excessive distortions after RF amplification. However, this classic approach to HFC design fails to take advantage of input and output operating margins that may be available in RF amplifiers. Depending on desired minimum performance, RF input levels to an amplifier have the ability to vary over a range of values and still maintain performance, so lower-gain amplifiers may be just as effective. Tap levels that are universally designed to accommodate a "standard length" RF drop may similarly lead to unnecessarily high RF levels into home CMs connected to shorter drops.

Operators have traditionally relied on static conditions to determine the best design for an area. As an example, the process for downstream design can generally be summarized as the following:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Determine a worst case or reasonable worst-case scenario for span losses.
- Determine a worst case or reasonable worst-case scenario for drop losses.
- Determine the optimal receive level of the end devices (modems and set top boxes).
- Determine the maximum output power and gain of amplifiers and nodes to meet the criteria mentioned above.

Although this has served operators well, it could lead to inefficiencies and higher construction costs if we were to analyze RF statistics typically available from subscriber terminal devices today. It should be noted that historically operators have not had full access to these data points which now include not only terminal devices but newer smart amplifiers as well. This is the primary reason why selecting a worst-case scenario has worked well.

Statistically the reason a worst-case design scenario works can be explained using the histogram below. Such histograms can be used to represent the range of CM input levels in a typical HFC network. We know almost every similar distribution can be represented as a normal or Gaussian type of distribution with a bell curve shape. Although the shape and skewness of the distribution can vary from operator to operator, for this example let us assume an ideal normal distribution.

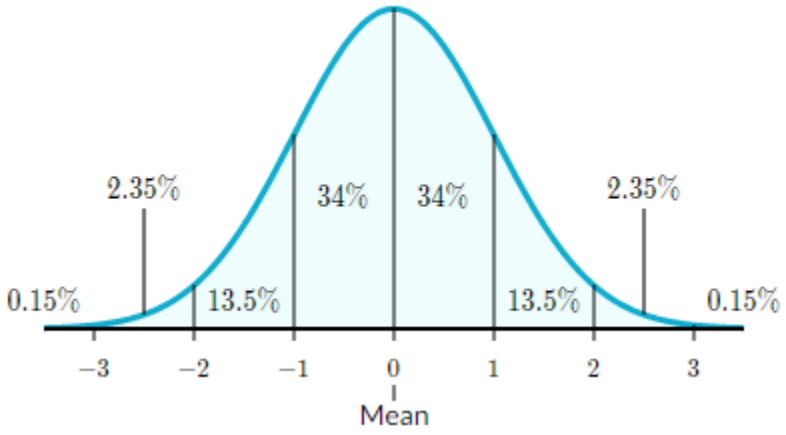


Figure 1 – Gaussian distribution [7]

To select a reasonable worst-case scenario in this case, we must choose an area that statistically is somewhat of an outlier, but not as far as three standard deviations from the mean. The reason for this approach is to cover as much of the farther ends of the network during the design phase and avoid making outliers of those areas, which would result in too much operational burden. While it might seem intuitive to maximize the output power of each amplifier and node to maximize reach, this can lead to over-engineered plants, a reduction in the useful life of amplifiers, and an overall reduction of the dynamic range of the system.

HFCs are always designed with an end device in mind. It is inherent that any design should be oriented toward an end device; in this case, a CM. This is true even if the outside plant designer might not design specifically for it. The complexity lies in the number of parameters to be accounted for, including (but not limited to): cable length, cable type, passive type, number of passives, and amplifier type. The HFC network design can be optimized with these factors in mind.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Because each segment of the network can be unique, it is not possible to account for every potential variation in the outside plant. A network is comprised of a series of insertion losses throughout, from the port of each active device to the port of each receiving device. This high variability makes outside plant design very challenging and can result in over-engineered HFC networks.

An outside plant amplifier's function is to provide adequate RF to the next amplifier in line to ensure enough level and to allow the signal to be repeated with the highest Signal-to-Noise Ratio (SNR). However, if an HFC network was designed to account for a certain span loss, but the actual loss is lower, extra actives are unnecessarily added to the network, which may result in amplifiers with high-value pads in their input ports and which can indicate excessive RF signal levels and, potentially, an overdesigned HFC network. Therefore, a different – and arguably more efficient – way of designing would be to allow each amplifier to use enough output power to reach the next amplifier with enough level. A detailed discussion of these benefits will be presented later, both from the perspective of signal quality and plant stability.

In addition, intelligent tools can be used to establish and leverage optimal RF operating margins for amplifiers and CPE to increase design effectiveness. CM data, including downstream and upstream signal strength and MER performance and errors, can also be analyzed to glean information about RF signals' health and performance through the HFC network. Similarly, telemetry from smart amplifiers and nodes can be collected and combined with CM data to select optimal amplifier gain profiles. Such optimized profiles will reach all CMs within a given geography while minimizing excessive RF levels into the home.

4.1. Statistical Analysis Tools to Enhance HFC Network Design

As mentioned above, TCP is a limiting factor when expanding downstream spectrum to 1.8 GHz. Expanding the downstream spectrum to 1.8 GHz requires the introduction of a step down in the output power spectral density (PSD) in the Outside Plant (OSP). As amplifiers cannot introduce this step down, the step down must be introduced in the node. With every dB of step down in the node output PSD, a one-to-one reduction in SNR can be expected in the node.

This approach results in a significant amount of capacity wasted. Figure 2 illustrates the difference between a high starting MER versus a low one, representing a 3-dB and 6-dB step down scenario. Note that carrier to composite noise ratio has been used instead of MER. The reason behind this is the fact that CCN is the true measure of plant performance, as it is not tied to any measuring device. While MER is directly tied to the measuring device's performance, making it a limited measure of performance.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

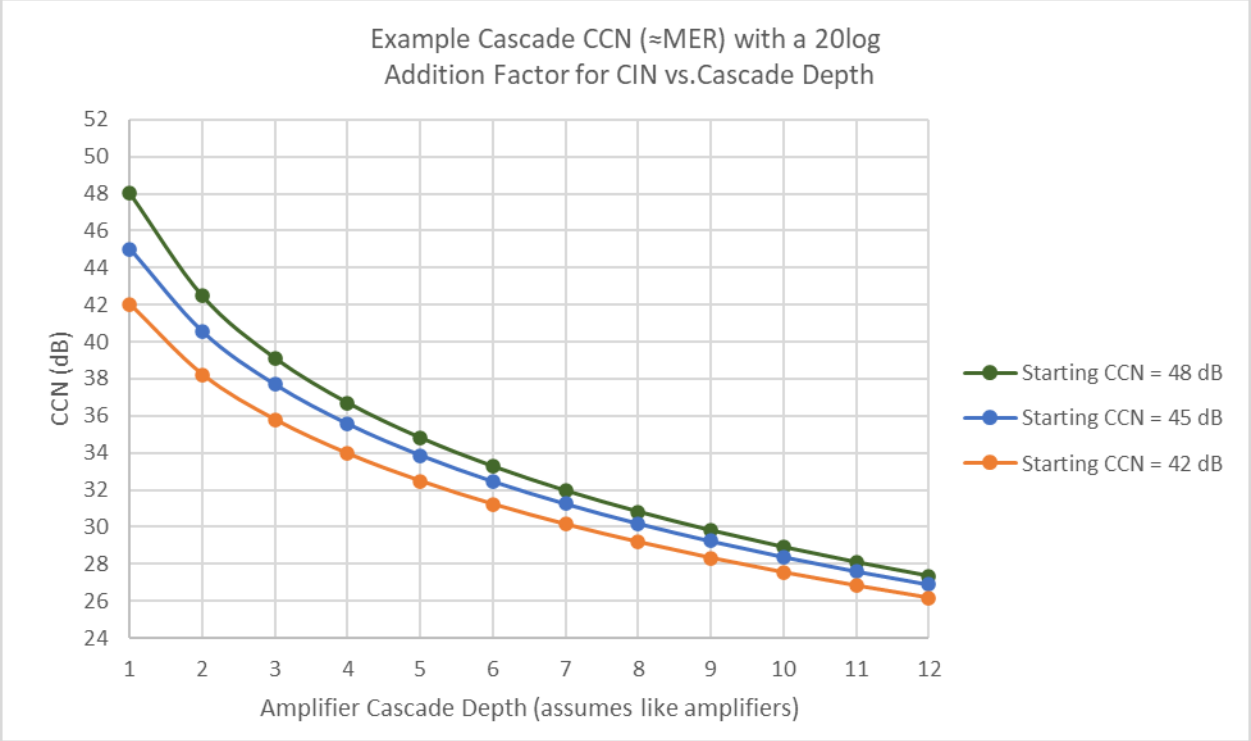


Figure 2 – CCN (≈MER) with a 20log addition factor for CIN vs Cascade Depth

Figure 2 is incredibly informative when it comes to design decisions and optimizations. In the bottom right hand of the graph, it can be seen that, as amplifier cascade length increases, a slow convergence occurs around a common value for CCN performance, regardless of the starting CCN or MER. This result is primarily due to the fact that noise and distortion products in the cascade dominate the signal quality aspect. By contrast, focusing on the left side of the graph, see that there is a wider gap in signal quality up to N+3. This result is also further emphasized knowing that the performance of each span is dictated by the active device, being the node or the amplifier.

As a result of implementing step downs, and the corresponding one-to-one SNR reduction for those frequencies impacted by the step down, quite a bit of aggregate capacity is sacrificed in the node as modulation orders for the higher frequency carriers are reduced in the stepped down areas. It can be argued that the use of step-downs to better manage TCP inadvertently sacrifices performance over shorter cascades for the sake of reaching the most distant ends of a network. Although historically designers may have focused on the last tap of the last amp, the results shown here further support the thesis that worst-case design methodologies can still be improved upon for even more optimal RF designs.

HFC networks are complex systems that often require time-consuming and costly simulations to optimize design and performance. AI-powered simulations can significantly speed up this process. Based on predefined performance metrics, automation can identify the most efficient setups by iterating through different scenarios. HFC systems can be fine-tuned for optimal performance using this accelerated design process not only by saving resources, but also by considering all factors that affect performance.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Incremental Benefits of Optimizing HFC Networks with AI and ML

A key component of operators' efforts to improve performance, reliability, and efficiency is optimizing their HFC networks. Monitoring system parameters, such as RF signal quality and CM received and transmitted levels, provides operators with valuable information for enhancing service reliability and user experience. Operators can benefit from network intelligence applications to analyze and optimize network performance based on CM data. Network intelligence applications that leverage data and automation can enhance service reliability and the user experience. Automation can help identify and resolve issues faster, while network intelligence can provide insight into HFC plant health by monitoring CM and amplifier behavior. This can allow for proactive optimization and quicker troubleshooting. Automation and network intelligence can also help reduce operational time and cost, while providing faster response times and improved reliability. Finally, integrating advanced analytics and automation tools into the network design process can help to ensure the best performance and reliability of the network.

HFC systems' inherent inter-complexities make manual optimization and troubleshooting daunting tasks. With AI, analysis of vast and intricate datasets becomes manageable. ML models can identify patterns, anomalies, and correlations that human operators might overlook, allowing for proactive identification of potential issues. In addition to improving overall efficiency and reducing downtime, ML-driven systems can help identify noise and interference, enhance HFC network capacity and potentially extend HFC network reach.

5.1. Improve Signal Quality and Performance

Robust RF signal quality ensures consistent and dependable communication while minimizing connectivity failures. The RF signal quality has an impact on the modulation schemes supported by CM, as mentioned previously. High-quality signals result in faster data transfers. RF signal quality cannot be overemphasized. It directly influences HFC networks' performance, reliability, and efficiency, making it a fundamental consideration in design and operation. With the use of network intelligence applications in HFC networks, it is possible for network operators to regularly monitor and diagnose signal quality with real-time data from CMs and active devices. Using network intelligence applications, operators can continuously monitor and analyze signal quality metrics such as SNR, MER, and bit error rate (BER) to diagnose signal quality problems. Using network intelligence, for instance, can help operators detect a range of impairments, from simple channel interference to more complex ones such as micro reflections. By continuously learning and adapting AI algorithms based on network data, operators can pinpoint the approximate location of the impairment and recommend appropriate solutions.

HFC systems that are AI-driven can dynamically adapt to changing network conditions, which is one of their major advantages. HFC network performance is influenced by factors such as user demand, signal noise, and signal degradation over amplifier cascades. Implementing AI algorithms that can continuously analyze these variables and adjust signal amplification, modulation orders, and profiles ensures optimal signal quality throughout the network. Adaptability results in consistent and better end-to-end performance, minimizing disruptions and improving the user experience.

5.2. HFC Predictive Maintenance

AI-driven automated management and monitoring systems can help operators quickly identify and address issues without having to physically access the network. Furthermore, AI-driven design introduces a level of proactive maintenance that is invaluable in HFC systems. This foresight enables operators to replace or repair components before they lead to network disruptions. As a result, system downtime is

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

minimized and maintenance becomes a planned, rather than reactive, process. Predictive maintenance allows operators to use data analytics to anticipate potential problems. AI-driven analysis enables network operators to detect subtle changes in the system that would otherwise remain unnoticed. This proactive approach can help operators identify and address potential issues before they occur. This will reduce system downtime and encourage them to plan for maintenance and repairs in advance.

5.3. Enhance Network Capacity and Efficiency

Advances in network optimization and virtualization also help operators manage their networks more effectively. Virtualization is being promoted deeper in the access network to bring services, resources, and intelligence closer to subscribers in the HFC plant. A virtualization platform that applies numerous applications and software-driven technologies powered by AI enables operators to expand revenue streams and optimize operations across their networks. Operators can also gain insights from multiple sources and make informed decisions efficiently by using data from multiple sources. Furthermore, it helps minimize risks associated with deploying new applications by scaling operations with increased demand [6].

For instance, virtualized cable access networks can dynamically manage network traffic more efficiently, resulting in increased throughput. With ML, operators will be able to identify and analyze more efficient network paths, thereby optimizing their network resources and maximizing bandwidth usage. With ML, operators can identify and predict traffic patterns to ensure that network demands are met during peak periods, for example.

Another benefit is that network intelligence can also be used to forecast demand and capacity needs, helping operators plan for the future.

5.4. Minimization of Interference and Noise

Intermodulation products and noise can be predicted through careful analysis, so their impact can be minimized. This reduces signal noise and increases network efficiency. Based on historical data and usage patterns, AI models can predict potential noise and interference-related problems in active RF devices. In addition, network intelligence applications implement AI algorithms to extract relevant features from the data. These features may include frequency, amplitude, phase, and time-domain characteristics. By analyzing these features, AI algorithms can identify abnormal patterns indicative of noise, interference, or distortions. Machine learning algorithms can also be used to learn from datasets containing both clean and noisy examples of RF signals. The classification of signals as noise-free or noisy can be done using supervised learning techniques including support vector machines (SVMs) [8], decision trees, random forests, and convolutional neural networks (CNNs) [9], as well as unsupervised learning to identify anomalous patterns or deviations from expected behavior that may indicate noise or distortion [13,14].

5.5. Extending the Reach of HFC Networks

By optimizing the amplifiers, HFC networks can be improved to provide better coverage and higher data speeds, in addition to ensuring the signal is strong enough to reach the user without being excessively strong to avoid distortions that may compromise the signal quality. Improved distortion and noise analysis are imperative to ensuring CMs can operate with higher modulation orders.

Operators can take advantage of CM received level margins to optimize amplifier performance by dynamically adjusting signal strength across the network, mitigating both over-amplification and the risk of creating distortions.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.6. Power Efficiency

The HFC outside plant can benefit significantly by operating at peak efficiency, meaning lowering amplifier output power and reducing power draw. The first benefit of this strategy is that it can **reduce HFC networks' operational costs and their environmental impact through improved energy efficiency**. Amplifiers may consume a considerable amount of power in HFC systems, which can be decreased by reducing their output power and power draw. This may result not only in cost savings, but also in contributing to a greener and more sustainable infrastructure, which aligns with the growing emphasis on energy conservation in the telecommunications industry [10].

The second benefit of reducing amplifier output power is that it can **extend the lifespan of network components, enhancing their reliability and reducing the need for maintenance**. Increasing output power may result in overheating components, signal distortions, noise or other impairments, which may degrade the quality of transmitted signals and increase error rates. Operating amplifiers at lower power levels allows for signal integrity to be maintained, leading to more robust network performance which may result in fewer service interruptions, less downtime and, ultimately, improved customer satisfaction.

Third, reduced amplifier output power can lead to **improved signal quality and performance for end-users**. Lower power levels reduce signal distortion and noise, leading to cleaner and more reliable signals reaching subscribers' homes, meaning cable TV and internet service providers can deliver higher-quality video and audio.

Lastly, maintaining tighter control over signal power through lower output power levels can mitigate issues related to signal leakage and ingress, giving HFC operators **a more harmonious coexistence with adjacent frequency bands and other networks and services, enhancing overall network stability and customer satisfaction**.

Using AI algorithms, amplifier bias can be adjusted to optimize amplifier power efficiency in real time, allowing for dynamic changes to the amplifier and enabling it to operate at peak efficiency, saving energy and reducing operating costs.

6. Future Directions and Emerging Technologies

AI-driven design in HFC systems holds immense promise for enhancing efficiencies and end-to-end performance due to the intricate inter-complexities of the variables inherent in such systems. HFC networks, which combine optical fiber and coaxial cable, are the backbone of modern broadband services. The complexity of a network arises from the need to balance factors like signal quality, bandwidth allocation and network maintenance. With AI's ability to analyze vast amounts of data and optimize these variables in real-time, HFC systems can operate with greater precision.

The integration of AI-driven design into HFC systems marks a transformative leap in efficiency and end-to-end performance. The complex interplay of variables inherent in these networks is effectively managed through real-time adaptation, proactive issue identification, proactive maintenance, and accelerated design optimization. AI empowers HFC systems to operate at their peak potential, delivering reliable and high-quality services to users while minimizing downtime and operational costs. As technology continues to evolve, AI's role in enhancing HFC systems' efficiency and performance is bound to become increasingly indispensable.

AI has also emerged as a transformative force in the realm of technology, and its application within Software-Defined Networks (SDNs) in HFC systems holds immense promise. SDNs have revolutionized

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

network management by decoupling the control plane from the data plane, allowing for centralized control and dynamic configuration. When integrated with AI capabilities, SDNs in HFC systems can greatly enhance network performance and efficiency. AI algorithms can analyze massive datasets generated by HFC networks, identifying patterns and anomalies that might be missed by traditional network management systems. This data-driven insight enables predictive maintenance, where potential issues are detected in advance and network resources are optimally allocated, leading to minimized downtime and improved user experience.

Moreover, AI can play a pivotal role in optimizing bandwidth utilization within HFC systems. By leveraging machine learning algorithms, SDNs can intelligently allocate bandwidth based on real-time demand, ensuring that network resources are dynamically allocated where they are needed the most. This capability is particularly beneficial in HFC systems, which serve as a critical infrastructure for broadband internet and cable television services. With AI-powered SDNs, operators can adapt to varying traffic patterns, allocate bandwidth for streaming services during peak hours and prioritize critical applications for seamless performance. As a result, end-users enjoy smoother streaming experiences, while service providers can make more informed decisions on capacity planning and network upgrades.

Figure 3, below shows the traditional silos associated with each segment of the HFC:

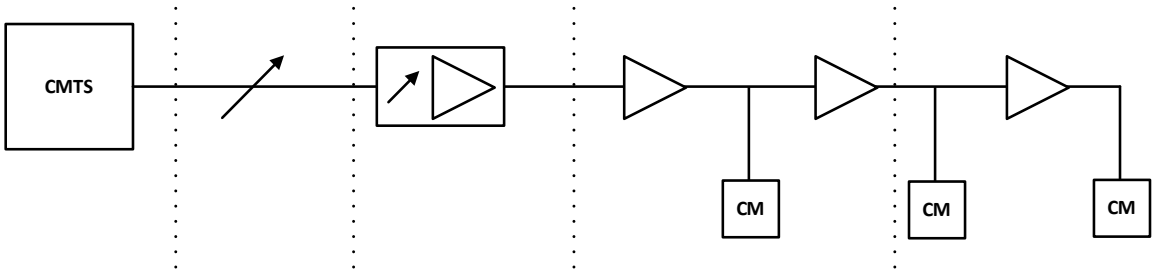


Figure 3. Traditional HFC segment

In an SDN environment, the HFC network is treated as an end-to-end system, which dramatically increases the efficiency and quality of the system as shown in Figure 4:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

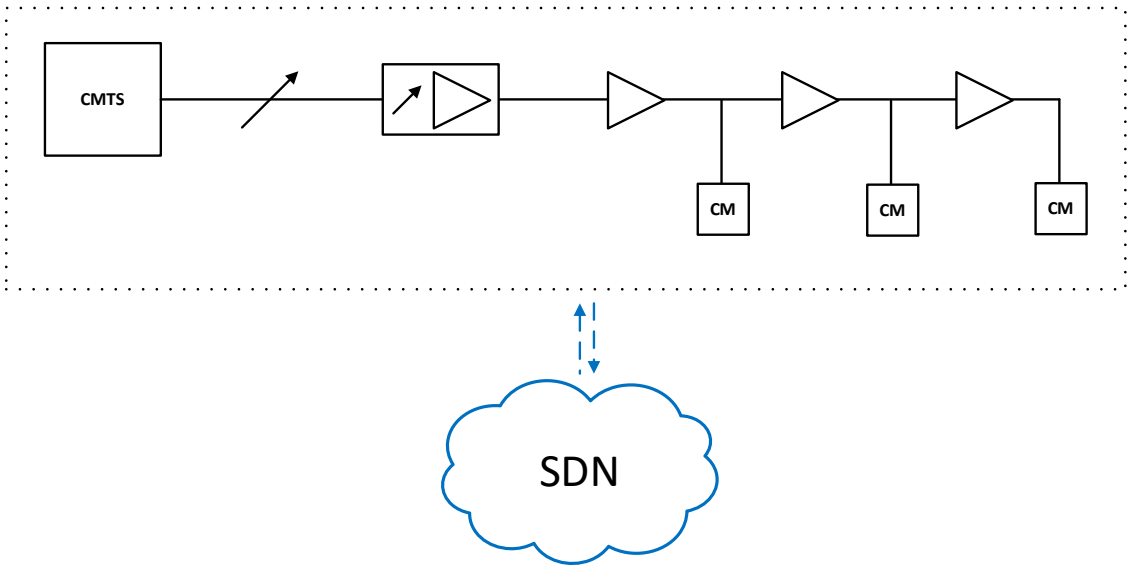


Figure 4. HFC segment with SDN

As AI algorithms become more sophisticated, they can also be used to predict future user demand and network requirements. This allows proactive optimization of HFC networks and can also automate complex manual tasks, freeing up personnel to work on innovating tasks that can drive the organization forward.

As AI algorithms evolve and network convergence is implemented, operators can enable self-healing capabilities in their networks. AI algorithms can, for instance, reroute traffic when noise or interference is detected. They can also dynamically adjust RF parameters or perform other corrective actions to maintain optimal performance.

7. Case Study

The following analysis is based on data collected from approximately 14,000 production CMs. From received (Rx) level histogram in Figure 5, it is observed that there is a wide range between the lowest and highest CM Rx levels; with few CMs as low as -17 dBmV/6.4 MHz, and CM Rx levels as high as +17 MHz/6.4 MHz. While CM received levels span a wide range, roughly 85% of them are above -5 dBmV/6.4 MHz. This is an indicator that traditional HFC design does not always result in a converged set of CM Rx levels. This confirms that HFC networks usually produce levels above the minimum specifications due to reasons previously mentioned.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

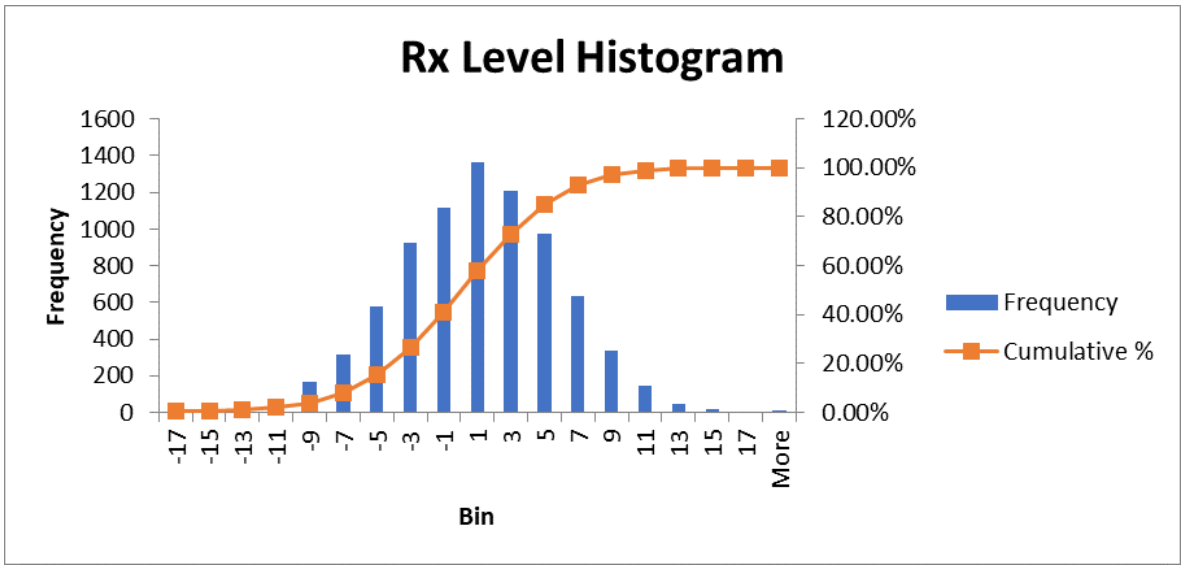


Figure 5. CM Rx level histogram and cumulative curve

The same holds true for CM transmit (Tx) levels. As shown in Figure 6, the Tx level histogram has a wide range. Despite the expectation that all CM Tx levels should converge based on target tap input if drop engineering assumptions are followed in the field, the conventional HFC design using static numbers to maintain unity gain results in a wide range of transmission levels due to the factors mentioned above.

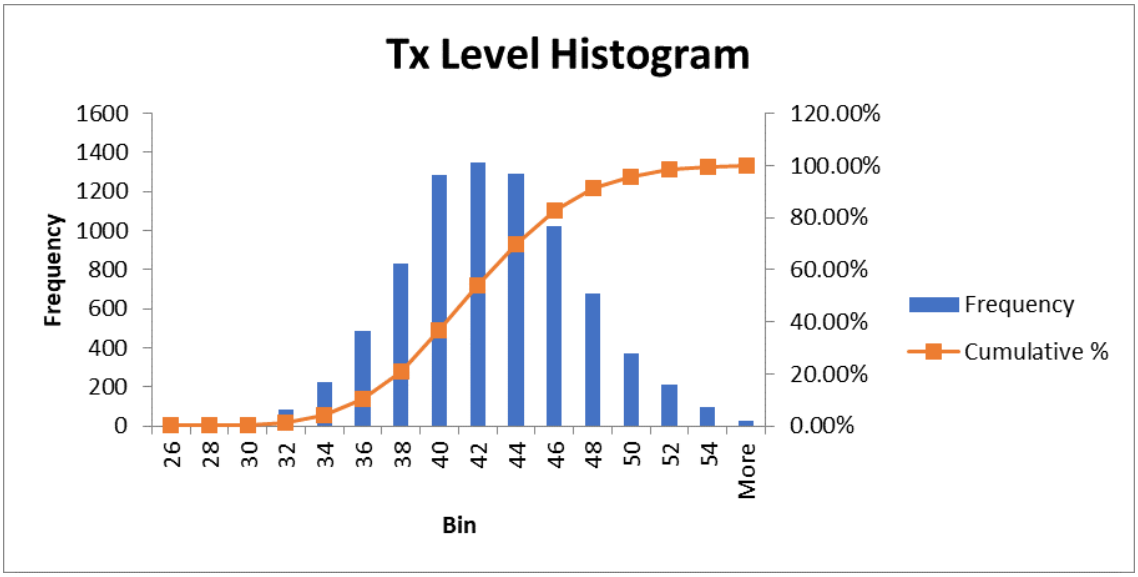


Figure 6. CM Tx level histogram and cumulative curve

It can be seen that the two distributions in Figures 5 and 6 are very close to perfect Gaussian distributions. Some interesting observations can be made. Let's begin by focusing on the CM receiving levels shown in Figure 5. There is a wide range between the lowest and highest RF receiving levels that CMs report. This is an indicator that the historical approach to design does not result in a converged set of receive RF

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

levels. Remember that on paper many designers and engineers would consider a modem receive level range of between -5 dBmV to +5 dBmV to be ideal. In this example, however, the range is much wider, indicating the need for further optimization of the distribution network than can be achieved through a traditional approach.

The same observation holds true for transmit levels. The wide range indicates that setting a static number at the unity gain point for the return path will result in a wide range of transmit levels. The traditional design for modems would consider a transmit level range of between +40 dBmV and +48 dBmV to be ideal. This also offers opportunities for further optimization.

To assess the receive signal quality for each modem, let's consider both the upstream and downstream levels. Figures 7 and 8 show the histograms for downstream and upstream signal quality, measured at the cable modem and the CMTS, respectively.

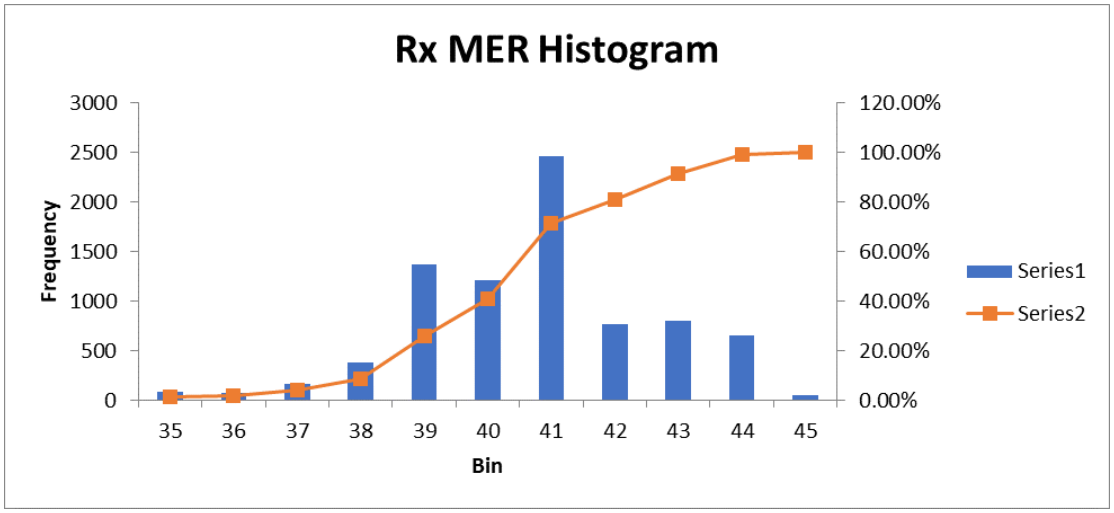


Figure 7. CM Rx MER Histogram and cumulative curve

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

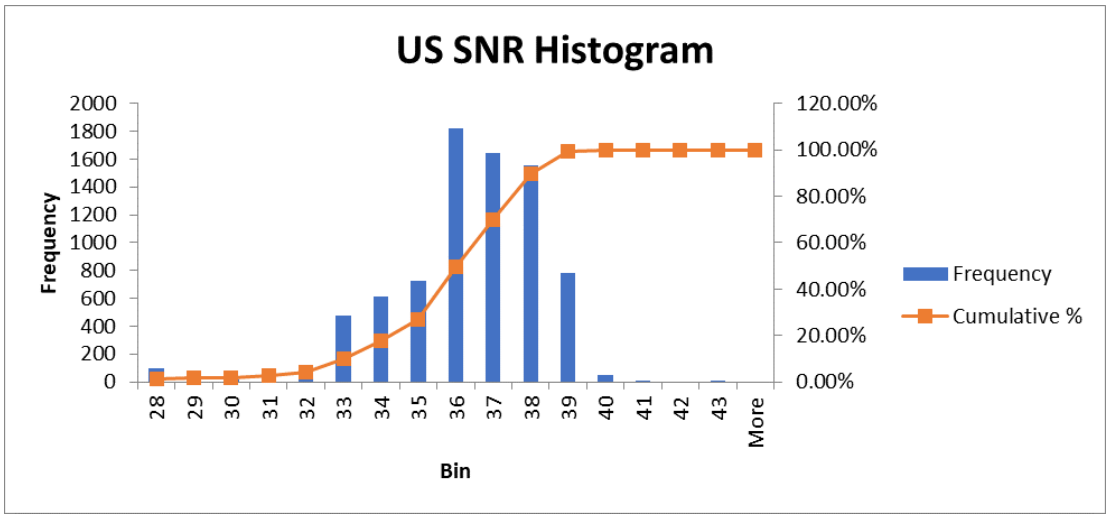


Figure 8. CM US SNR histogram and cumulative curve

It can be observed that both distributions are heavily skewed to the left with downstream MER skewness measured at -1.65 and upstream measured at -1.86. Although this may seem surprising, it is a perfect example of the design artifacts discussed in previous sections. In a cascade, particularly after the 2nd and 3rd amplifiers, downstream performance is dictated by cascade depth (noise and distortion) rather than signal level. This is further confirmed in the upstream by the fact that from the data set, there is a loose correlation between upstream transmit level and upstream MER (0.14).

Furthermore, the analysis of this study suggests that higher transmit levels from the cable modem in the upstream and higher receive levels in the downstream do not necessarily correlate with higher SNR/MER as predicted earlier in this paper. This indicates that the SNR/MER values are not necessarily dependent on the transmit/receive levels. Therefore, it is important to consider other factors when determining the optimal SNR/MER values for a modem.

8. Conclusion

AI and ML based tools have begun to play a key role in the management of HFC networks and should also be considered to help design and optimize HFC networks to improve efficiency. The power of AI and ML tools comes from the following.

- Collection of device telemetry in real time.
- Data analysis and identification of performance trends in real time.
- Recognition and diagnosis of potential network impairments in real time.
- Identification of optimal operating margins for amplifiers and CPE devices.
- Selection and application of optimal gain profiles for active devices.

Data-driven design revolves around the collection, analysis and utilization of network-related data to drive better decision-making processes. In the context of HFC networks, this encompasses data on subscriber usage patterns, network topology, equipment performance and environmental factors. By harnessing this wealth of information, network planners can gain valuable insights into network behavior, capacity requirements, and potential bottlenecks, which allow for operations at peak efficiency.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Operating at peak efficiency delivers a number of benefits, including: reducing HFC networks' operational costs and their environmental impact through improved energy efficiency; extending the lifespan of network components, enhancing their reliability and reducing the need for maintenance; leading to improved signal quality and performance for end-users; and creating a more harmonious coexistence with adjacent frequency bands and other networks and services, enhancing overall network stability and customer satisfaction.

Abbreviations

AI	Artificial intelligence
BER	Bit error rate
bps	Bits per second
CCDF	Complementary cumulative distribution function
CIN	Composite intermodulation plus noise
CNN	Carrier-to-composite noise ratio
CNNs	Convolutional neural network
CM	Cable modem
CPE	Customer premises equipment
FDX	Full duplex
FEC	Forward error correction
HD	High definition
HFC	Hybrid fiber coaxial
Hz	Hertz
MER	Modulation error rate
MHz	Mega hertz
ML	Machine learning
OSP	Outside plant
PA	Power Amplifier
PSD	Power spectral density
Rx	Received
TCP	Total Composite Power
Tx	Transmitted
SCTE	Society of Cable Telecommunications Engineers
SDN	Software defined networks
SNR	Signal-to-Noise Ratio
SVMs	Support vector machine

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

[1] *RF performance and plan design considerations for all-digital HFC networks,* D. Job; Cisco Systems, Inc.

[2] *RF Fundamentals Seminar Part 5: Digital Modulation and Vector Signal Analysis,* D. Rick. Rohde & Schwarz.

[3] *RF Fundamentals Seminar Part 6: Intro to OFDMA (Orthogonal Frequency Division Multiple Access)*, D. Rick. Rohde & Schwarz.

[4] *Complementary Cumulative Distribution Function for Performance Analysis of OFDM Signals And Implement PAPR Reduction Techniques In OFDM System using CCDF function on Matlab*, U Lekhraj & K Sudhir. International Journal of Electronics and Communication Engineering.

[5] *The road to 10G Migrating Today's HFC network to meet tomorrow's demand,* M Cooper, D Job & B Wall. SCTE Expo 2021.

[6] *A roadmap for virtualization in HFC networks,* A Bender. SCTE Expo 2019

[7] *What is a normal distribution?* Khan Academy: https://www.khanacademy.org/math/statistics-probability/modeling-distributions-of-data/normal-distributions-library/a/normal-distributions-review

[8] *Why you use SVMs*, NetApp: https://library.netapp.com/ecmdocs/ECMP1354558/html/GUID-ABC30BB1-222D-4663-B878-52D4C6BFC191.html.

[9] *A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way*, S Saha. Data Science & ML. December 15, 2018.: https://saturncloud.io/blog/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way/

[10] *World Energy Outlook 2020, Part of World Energy Outlook*. Iea: https://www.iea.org/reports/world-energy-outlook-2020

[11] *FDX Amplifier for Supporting N+M Network.* Hang, J & Chapman, J. SCTE EXPO 2019.

[12] *Operational Considerations & Configurations for FDX & Soft-FDX. A Network Migration Guide To Converge The Cable Industry.* AI-Banna, A; O'Keeffe, F & Cloonan, T. SCTE EXPO 2019.

[13] *Improving Operational Intelligence for Maintaining Cable Networks*. M. Spaulding, L. Wolcott, J. Rupe. Prodeedings of the Cable-Tec Expo., 2022.

[14] *Proactive Network Maintenance using Fast, Accurate Anomaly Localization and Classification on 1-D Data Series*, J. Zhu, K. Sundaresan, J. Rupe. 2020 IEEE International Conference on Prognostics and Health Management (ICPHM), 2020.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Fiber Network Planning Options for Rural Communities

Remote OLT (rOLT) Use Case Comparison and Analysis

A Technical Paper prepared for SCTE by

Adam Soccorsy
Director of Sales FTTX Systems
Technetix
Technetix Ltd.,
Innovation House
Technetix Business Park
Albourne, West Sussex
BN6 9EB
United Kingdom

Adam.Soccorsy@technetix.com

Luc Absillis Ph. D
Chief Operating Officer
First Principles Innovation
111 Fieldbrook Court
info@fpinno.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Section 1 - Reimagining the network with rOLT ... 4
3. Section 2 - Lamoille County, Vermont ... 6
4. Section 4 - Power consumption ... 10
5. Conclusion ... 10

Abbreviations ... 11

Bibliography & References ... 12

List of Figures

Title **Page Number**

Figure 1 - Rackmount OLT topology map of Washington Country, North Carolina ... 5
Figure 2 - rOLT topology map of Washington County ... 5
Figure 3 - FWA topology map of Washington County ... 5
Figure 4 - HLD of Lamoille County, Vermont ... 6
Figure 5 - LLD Eden Mills, Vermont (courtesy of Vetro) ... 7
Figure 6 - Estimated cost per HHP in Eden Mills, Vermont ... 8
Figure 7 - CAPEX comparison Lamoille County ... 9
Figure 8 - Mixed synthetic and real project costs per HHP with speeds ... 9

List of Tables

Title **Page Number**

Table 1 – Basic assumptions for GIS computer modelling ... 4
Table 2 - Initial comparison of 10-year total cost of ownership (TCO) and details ... 6
Table 3 - Construction cost comparison ... 8
Table 4 - Power required to serve 19M homes with symmetrical 10 Gbps for one year ... 10

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Advances in semiconductor technology and the relentless progression of Moore's Law have enabled ever more capable smart devices in every aspect of our lives. Chips enabling network functions to be miniaturized and compressed into pluggable devices are encouraging innovation in the access network space. Being able to move a single optical line terminal port (OLT) out of its normal position inside a rackmount chassis and place it in a small, hardened node enclosure on a telephone pole, or other remote exterior mounting, is a new concept that was not possible a few years ago. These same advances in semiconductor technology have also benefitted other network construction types such as Fixed Wireless Access (FWA).

Multiple vendors have launched remote OLT (rOLT) products with new dimensions of price and capability that take advantage of the new lower power, higher capability chips.

This paper uses the generic dimensions of a synthetic remote OLT and combines them with several computer models that utilize geographic information system (GIS) software to assess the overall impact of constructing greenfield fiber connectivity projects in the United States with remote OLT technology.

What is a Remote OLT (rOLT)?

There is more than one standard covering the technology. The International Telecommunication Union (ITU) is the body that covers the passive optical network (PON) standards under their remit. However, the standard for XGS-PON, G.987[1] does not cover the physical characteristics of an OLT device but rather describes the architecture and communication protocol. The American National Standards Institute (ANSI) also has a standard entitled the Generic Access Platform (GAP)[2] which partially covers rOLTs; but this is more concerned with mixing DOCSIS and PON in a modular format. It does not cover PON-only devices that were not technically feasible at the time of writing. Remote OLTs operate in a grey area between two standards; standards that are effectively defined by the technology vendors operating in this space. There are several vendors now selling or designing remote OLTs operating solely with xPON, but also some that are mixing DOCSIS and xPON.

The rest of this paper will focus exclusively on remote OLTs that only serve XGS-PON technology.

Taking a wider look at trends, we will assess the remote OLT across some broader goals broken down into three areas: lowering the cost of connecting our rural communities; lowering the complexity of the network; and ensuring fundamental digital equality across society. Specifically, we reference the Federal Communications Commission (FCC)[3] finding that there are more than 19 million American homes without access to broadband. This represents a basic number: it does not include the many millions more of underserved connections.

Looking internationally, we highlight a recent report by Gigaclear in the UK[4], which studied the technical and commercial aspects of using remote OLTs in their network. The report established that smaller, more integrated solutions are required for serving rural communities. This paper will go into more depth about these requirements.

---

[1] ITU G.987: 10-Gigabit-capable passive optical network (XG-PON) systems: Definitions, abbreviations, and acronyms

[2] ANSI/SCTE 273-2 2021: Generic Access Platform (GAP) Modules Specification

[3] FCC Website. www.fcc.gov

[4] Gigaclear, Small Coverage OLT Report (Internal), Tim Durkin B. Sc, M. Sc, MIET, 2023

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2. Section 1 - Reimagining the network with rOLT

A rack-mounted OLT is normally constructed to fit into a 19" rack, and usually contains several slots where cards can be installed. A typical minimum installation would contain one card, consisting of eight ports, with expansion slots capable of accommodating up to a total 32, 64, or 128 OLT ports. Although different network designs utilize different split ratios, a typical 64-way split architecture would imply a minimum capacity of 2,048 end user homes that could be connected within a 20-mile radius of the device. Using a 32-way split architecture would imply a capacity of 1,024 end user connections for a minimum capacity rackmount installation (eight ports of a 32-port chassis). At the time of writing, a consensus indicates that the approximate cost of installing a field installation OLT cabinet, with all the necessary accessory equipment, currently costs somewhere around $70,000 - $100,000 for the capability to connect between 256 and 2,000 homes. This includes installation cost.

A rOLT by contrast follows a different set of dimensions. It can have between one and four OLT ports but does not need to be installed in a cabinet. However, using the 32- and 64-way architectures, it is possible that the rOLT can serve between 32-256 (1x32- 4x64) homes. Since rOLTs require far fewer accessories and much less power, we assume a cost of $20,000 to acquire and install a rOLT that can serve between 32-256 homes within a 20km radius. Since this paper is addressing new construction areas, we would also like to suggest the following comprise part of the 'wish-list' for an ideal device: low power consumption, minimal installation complexity, robustness, and subsidy-awareness.

The compact size of the newer rOLTs enables many more mounting options for active equipment. In this computer model we assume equal parts aerial and underground construction. Roughly $15k of the total $20k deployment cost (see Table 1 below) is attributed to a remotely mountable device kit, with $5K for its installation. Cabinet mount options require more kit, planning and concrete costing - around $45K and roughly $25K to mount. These costs were benchmarked across several tools and market information sources.

Table 1 – Basic assumptions for GIS computer modelling

Technology Type	Homes	Power	OLT Ports	Cost to Deploy (est.)
Rackmount OLT	256-2048	1050 Watts	8-64	$70,000
rOLT	32-256	48 Watts	1-4	$20,000

Taking these parameters of the equipment into consideration, we can create a template profile in a GIS tool that is used to calculate the cost of constructing and operating different types of networks. Washington County, North Carolina was chosen as a rural county for some initial comparisons. The county contains around 11,000 households passed (HHP) in around 76 square miles. The graphics below show the estimated topology locations of the equipment for the network types: Rackmount OLT, rOLT, and FWA.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

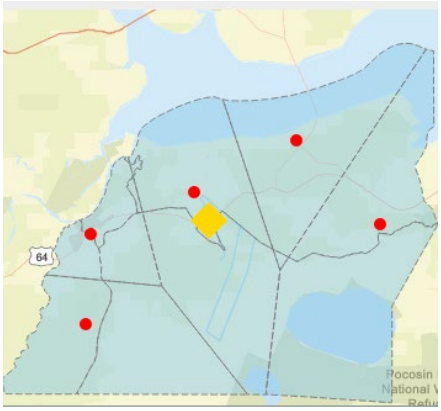


Figure 1 - Rackmount OLT topology map of Washington Country, North Carolina

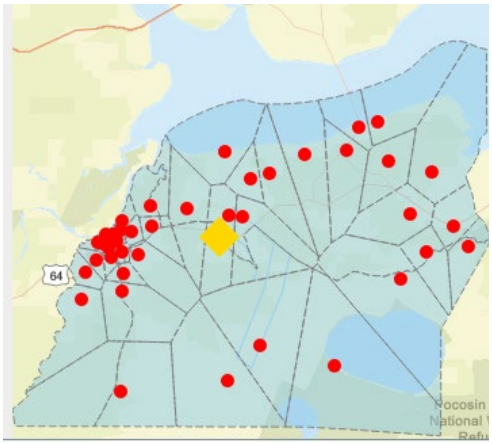


Figure 2 - rOLT topology map of Washington County

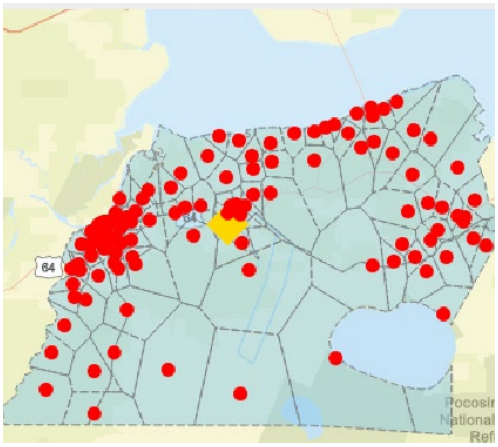


Figure 3 - FWA topology map of Washington County

Figures 1, 2 and 3 demonstrate the difference in topology types using different types of network construction. In this model, we also take future traffic growth into account when looking at the long-term cost. As Table 2 shows, FWA initially costs substantially less to construct ($28M). However, it ultimately

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

could cost substantially more ($65M) if a constant combined annual growth rate (CAGR) for bandwidth growth is assumed.

Our comparisons conducted in Washington County established two significant findings. Firstly, the behavior of the GIS systems in plotting the equipment against the census derived geodata works and appears to follow common sense assessment of location. Secondly, this first analysis changed the perception that rOLT construction would be more expensive in large HHP areas. Washington County showed that, when starting from scratch, the costs are on a par.

Table 2 - Initial comparison of 10-year total cost of ownership (TCO) and details[5]

Technology Type	Number of Units	TCO 10 Year
Rackmount OLT	5	$39M
rOLT	59	$40M
FWA	FWA + 59 rOLT	$65M ($28M)

3. Section 2 - Lamoille County, Vermont

Lamoille County, Vermont has been selected as the second area for comparison of network construction projections. Using the GIS software, we will select a smaller rural sample area to understand if there is economic threshold, and where it becomes cheaper to build a network using rOLT rather than with a rackmount OLT. Bringing in another GIS tool that can create low level designs (LLD), we will conduct both high level design (HLD) and LLD analysis of the same area to study the economics.

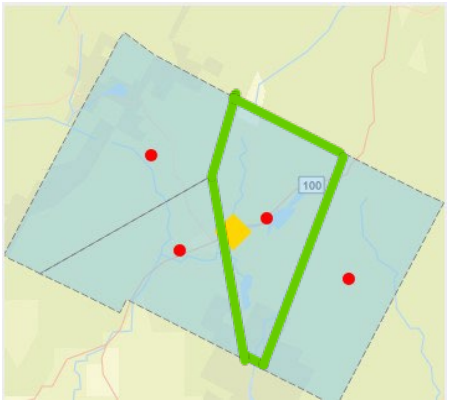


Figure 4 - HLD of Lamoille County, Vermont

[5] AP-Jibe, www.fpinno.com

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Lamoille County contains around 912 households passed (HHP) and is a predominantly rural county with some lakes and a nature reserve. Figure 5 shows a yellow square in the middle: this represents the location where a rackmount OLT could be installed to serve this area. The small red dots represent the locations of the rOLT devices that would require installation to serve this area. Each of these rOLT devices sits within an area that has been created to allow equal division into blocks of roughly 228 homes. Each one of these 228 home segments has its own rOLT serving symmetrical 10 Gbps.

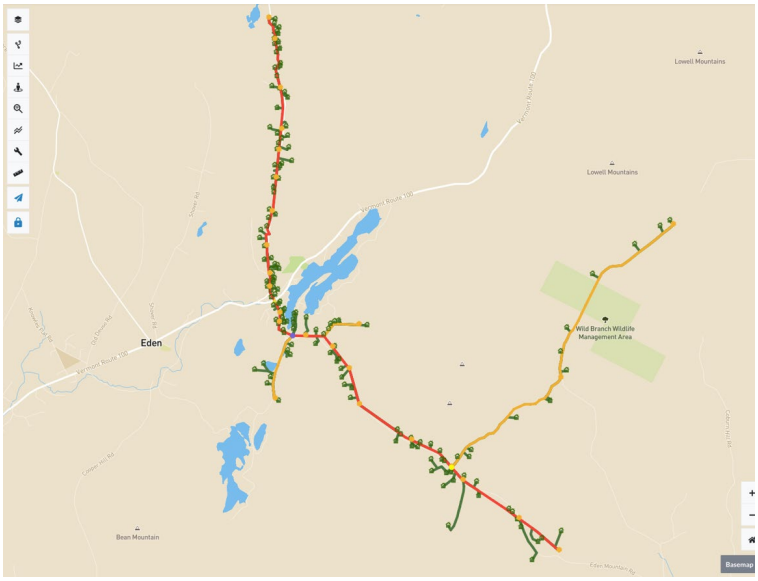


Figure 5 - LLD Eden Mills, Vermont (courtesy of Vetro[6])

Both the HLD and the LLD designs contain projections of the costs of constructing a greenfield FTTX network, including projections for cash flow and revenue. The LLD area is roughly equivalent to one section (the highlighted green section of Figure 4) of the HLD design; essentially, 224 homes and, which provides robust comparison. In addition, we can model the cost of constructing with a greenfield FWA, and a DOCSIS network.

Combining the HLD and LLD of fiber construction, along with FWA and DOCSIS construction, allows us to calculate the following results:

[6] Vetro fibermap, www.vetrofibermap.com

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

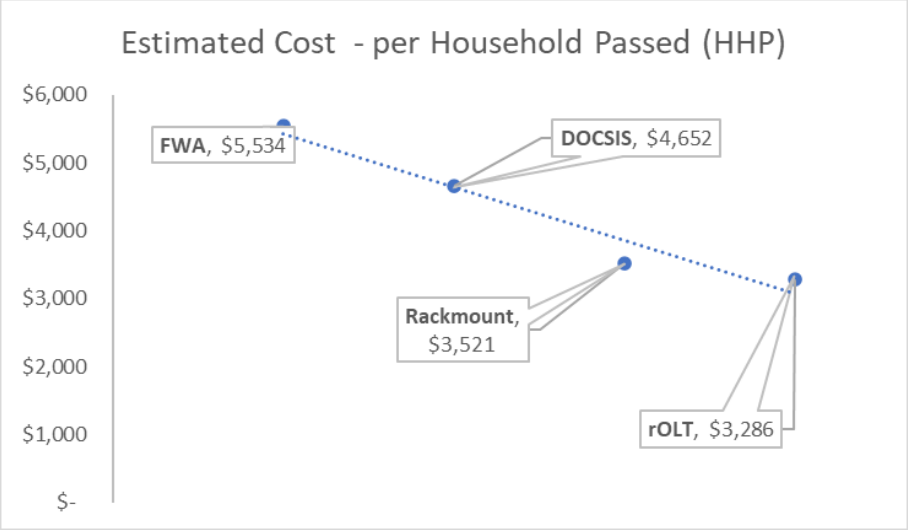


Figure 6 - Estimated cost per HHP in Eden Mills, Vermont

Figure 6 above shows the trend graph with the different build types listed in descending order showing a modest advantage of 6.6% for building with rOLT. In addition, we note that placing four rOLTs in different locations should also save miles of plant construction: we call this 'topology saving'. However, we do not endeavor to prove topology saving here.

Table 3 below provides a more detailed examination of total project cost for Eden Mills:

Table 3 - Construction cost comparison

Equipment Type	HHP	Cost to Construct
rOLT	228	**$806,070.90**
rOLT	912	$3,011,253.00
Rackmount OLT	912	$3,218,741.00

When looking deeper into the detailed bills of materials that sit behind the calculations, it is apparent that the cost of building rural network is largely driven by a few variables, including the number of feeder and trunk miles that need to be constructed. For clarity, our model assumes 50% aerial and 50% underground construction.

Looking at the first row of Table 3 based on our previous assumption that the equipment costs $20k, most of the remaining balance ($786,070) applies to the miles of construction required to reach end users.

This same financial data is reflected in the color bars on the charts (Fig. 7). The equipment requirements are represented by the pale orange segment at the top of each bar; the black section directly below represents the customer premises equipment (CPE); with the miles of construction below in green.

The left-hand chart (Fig 7) contains four roughly equal bars that are each valued at around $750K. The Eden Mills LLD project would represent one bar (1/4) of this potential four-part project.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

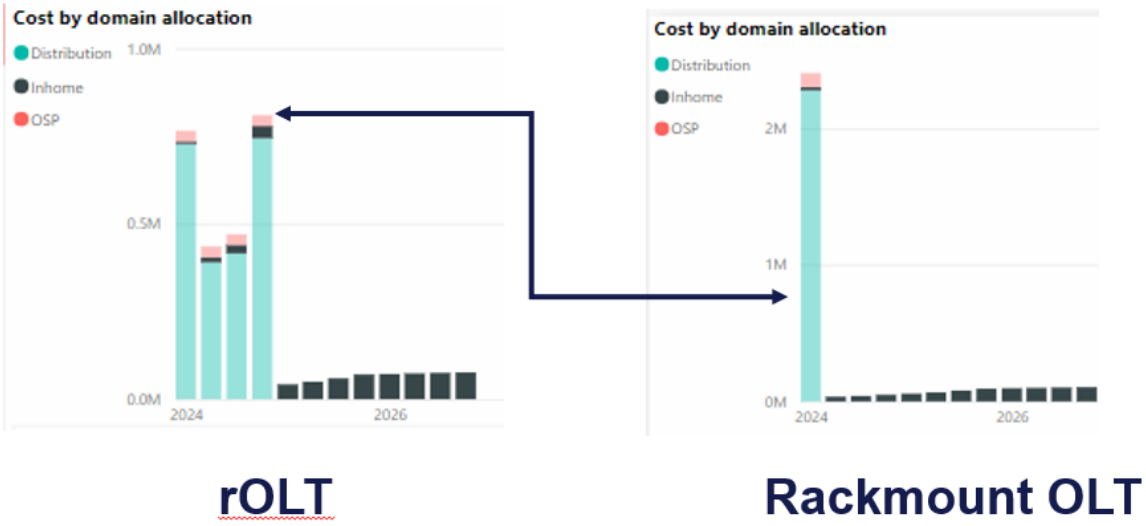


Figure 7 - CAPEX comparison Lamoille County

Imagine for a moment a community backed project raising funds for their own fiber network, they would now be able to raise a much smaller amount of money to begin their rOLT based project. If this same project succeeds, it can then continue expanding into the 3 other areas of their county. In contrast, a rackmount OLT project would need to account for the full construction cost at the beginning of the project. This cash flow improvement is potentially immensely valuable for projects that struggle with funding or require greater financial or technical flexibility for completion. Community projects looking to fund or construct their own network can gain a good advantage with a smaller capital expenditure (CAPEX) profile.

Looking more broadly within the United States then, how does our HLD and LLD analysis compare with real projects already in process?

Award	HHP	Cost HHP	DS	US
Navajo Tribal Utility	20287	$ 2,506	1000	1000
Synthetic Calculation VT	224	$ 3,216	10000	10000
Jicarilla	1192	$ 5,819	1000	1000
Dilkon Chapter	3643	$ 9,126	25	3
Ak-Chin	255	$ 12,081	100	100
Kewa	680	$ 18,788	25	3
San Ildelfonso	255	$ 19,316	1000	1000
Cocopah Indian Tribe	210	$ 24,832	25	3
Synthetic Calculation AZ	867	$ 27,664	10000	10000
Mescalero	1200	$ 36,619	1000	1000
Havasupai Tribe	135	$ 52,423	100	20

Figure 8 - Mixed synthetic and real project costs per HHP with speeds

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 8 is a sample of real projects costs in Arizona and New Mexico published on the Government's Internet For All website[7]. The yellow rows are synthetic calculations for both Vermont and Arizona using the same methodology used earlier in this paper. Looking closely at the detail of these projects, it is apparent that, in many cases, the connections created are between vital institutions and places of learning - not solely residential. Using a cost-per-HHP as the benchmark, there are compelling indicators that rOLT would be a highly viable option for projects like the above, depending on the topology.

4. Section 4 - Power consumption

Although some variance can be observed in the power consumption of rOLT products already on the market, it is worth noting that some of the most recent models have significantly lower power consumption than Rackmount OLT versions. This is because newer rOLTs are passively cooled. Indeed, the power-saving in a rOLT is because the device does not consume energy running fans, since moving parts have been eliminated from the OLT design.

If we connected the 19M unconnected homes in the USA now, what would the power consumption be?

Table 4 - Power required to serve 19M homes with symmetrical 10 Gbps for one year

	Watts	OLT Ports	Watt/Port	GB/Watt	kWh 1 year
Rackmount OLT	1050	64	16.4	0.61	42,666,503.91
rOLT	48	4	12.0	0.83	31,207,500.00

Another possible benefit may be since the rOLTs are spread out geographically. This could be ideal for solar panel power generation, producing an ecologically friendly method of connecting homes and reducing further the carbon footprint. Directly comparing the Gigabit (GB) per Watt column in Table 4, we understand that we are pushing over 30% more GB through the network for the same amount of power consumed.

5. Conclusion

Circling back on the wider trends in networking mentioned at the beginning of the paper, let us review.

In terms of lowering the cost of connecting rural communities, there is a strong case that rOLT could be the cheapest method of connecting many for these rural communities with the lowest cost per HHP on some topology types, particularly low density. Looking at the question of CAPEX, having a device that is aptly sized for serving 256 homes, lowers the financial bar for many projects.

Lowering the complexity of the network is difficult to prove. However, we can demonstrate that a large-scale deployment of rOLTs would create a measurable and desirable improvement in the power consumption needed to close the digital divide without compromising on network speed. Lowering the year-on-year power consumption lowers the future complexity of owning and operating any network, in addition to lowering the environmental cost.

Digital equality is a significant goal of the government broadband stimulus. The role of industry is to support the initiative with customer-centric innovation that lower the cost per HHP, and enables the spending to reach more places, faster, and at the best cost.

[7] Internet for All – U.S. Government Website www.internet4all.gov

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

rOLTs present a strong solution for network planners and operators looking to construct areas of network of 1000 HHP or less. For a small community project, rOLT provides an affordable starting point to bring fiber to small numbers of isolated or difficult-to-reach homes. For the network planners of the future, the ability to connect communities at a lower dollar and environmental cost are both tangible benefits. rOLTs bring with them the potential to reimagine a lower power, lower impact network helping us build a better future for all.

Abbreviations

ANSI	American National Standards Institute
AZ	Arizona
CAGR	Combined Annual Growth Rate
CAPEX	Capital Expenditure
CPE	Customer Premises Equipment
DOCSIS	Data Over Cable Service Interface Specification
FCC	Federal Communications Commission
FWA	Fixed Wireless Access
FTTX	Fiber to the x
GAGR	Compound Annual Growth Rate
GAP	Generic Access Platform
GB	Gigabits per second
Gbps	Gigabits per second
GIS	Geographic Information System
HHP	Households Passed
HLD	High Level Design
ITU	International Telecommunication Union
gWh	Gigawatt Hours
kWh	Kilowatt Hours
mWh	Megawatt Hours
LLD	Low Level Design
OLT	Optical Line Terminal
OPEX	Operating Expenditure
PON	Passive Optical Network
rOLT	Remote Optical Line Terminal
SCTE	Society of Cable Television Engineers
TCO	Total Cost of Ownership
VT	Vermont
xPON	Any version of PON technology

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

ITU G.987: 10-Gigabit-capable passive optical network (XG-PON) systems: Definitions, abbreviations, and acronyms

ANSI/SCTE 273-2 2021: Generic Access Platform (GAP) Modules Specification

FCC Website. www.fcc.gov

Gigaclear, Small Coverage OLT Report (Internal), Tim Durkin B. Sc, M. Sc, MIET, 2023

First Principles Innovations, AP-Jibe®, www.fpinno.com

Internal for All – U.S. Government Website www.internet4all.gov

Vetro fibermap, www.vetrofibermap.com

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Fiber QAM vs RF QAM

Demystifying Coherent Optics

A Technical Paper prepared for SCTE by

Amarildo Vieira, Ph.D.
Principal Optical Engineer
Comcast Connect
1800 Arch Street, Philadelphia PA
+1 (215) 301-7480
amarildo_vieira@comcast.com

Venk Mutalik, Fellow/Comcast

Dan Rice, Vice President/Comcast

John Chrostowski, Executive Director/Comcast

Saif Rahman, Distinguished Engineer/Comcast

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. QAM is QAM ... 3
   2.1. SNR vs OSNR ... 4
3. RF QAM Brief Review ... 5
4. Fiber QAM Review ... 6
   4.1. The Form Factor Dilemma ... 6
   4.2. Coherent Optics Implementation ... 7
   4.3. Forward Error Correction (FEC) ... 9
5. Plant & Operations Implications ... 10
   5.1. Plant Optical Return Loss ... 10
   5.2. Cascaded WDM Power Penalty ... 11
   5.3. Fiber Characterization ... 13
6. Fiber QAM Industry Initiatives ... 13
   6.1. Application Scenario ... 14
7. DOCSIS Implications ... 15
8. Capacity Implications ... 17
9. Conclusions ... 18

Abbreviations ... 19

Bibliography & References ... 20

List of Figures

Title **Page Number**

Figure 1 – Symbol constellation for different Modulation Orders ... 4
Figure 2 – Residual BER Prediction [2] ... 4
Figure 3 – BER vs SNR and OSNR correction factor plots for AWGN & no FEC ... 5
Figure 4 – Pictures of the CFP2-DCO (left) and QSFP-DD-DCO (right) ... 6
Figure 5 – Generic Coherent Optical Receiver Block Diagram (after [7]) ... 8
Figure 6 – Soft Decision concept (after [8]) ... 9
Figure 7 – Power penalty caused by the ORL in a 100Gbps QPSK BiDi system ... 11
Figure 8 – Modulated linewidth at different data rates and modulation orders ... 12
Figure 9 – Sensitivity Power Penalty caused by the WDM passband ... 12
Figure 10 – DAAS Extension CMP solution ... 14
Figure 11 – Bremerton Primary and Poulsbo Secondary in WA ... 15
Figure 12 – XR Transmitter Optical Spectrum for 4, 8 and 16 bidirectional subcarriers ... 16

List of Tables

Title **Page Number**

Table 1 – Coherent Optics Pluggable Comparison ... 7
Table 2 – Metrics for standardized FEC in optical fiber transmission systems ... 10
Table 3 – Current CableLabs Coherent Optics related Standards ... 13
Table 4 – Total Capacity Estimation per Fiber for Different Modulation Orders ... 17

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Quadrature Amplitude Modulation (QAM) transmission has been the mainstay of Radio Frequency (RF) networks for the better part of two decades. During that period, we have witnessed modulation complexity rise from 64-QAM to 4096-QAM. With each increase in complexity, we have adapted network configurations to higher SNR and spectrum to deliver higher and higher capacity. Similar is the case for optical transmission as operators move to higher modulation order Coherent Optics to keep up with the ever-increasing capacity needs in the transport, metro, and access networks. There is a need to get beyond contemplating the many similarities and challenges coherent optics are currently facing, however, and use the lessons learned in the RF domain for the journey with Fiber QAM.

100Gbps QPSK optics have been routinely used in different segments of Comcast's network and 800Gbps 32-QAM and beyond are not far on the horizon. In the access network, 100Gbps bi-directional optics have been instrumental in supporting different initiatives beyond capacity increases [1]. This includes fiber conservation, pushing Distributed Access Architecture (DAA) further down close to the node, and faster and lower-cost deployment and upgrades.

In this paper we first share our experience in preparing for higher order m-QAM modulation in terms of Coherent Optics terminology and tracing parallels with the steps the industry took during their m-QAM RF transition. Concerns about the fiber infrastructure, managing fiber non-linearity, and network architecture are then addressed with our recommendation to operators on how to get ready for Fiber QAM evolution. Lastly, as part of the envisioned evolution, the concept of Orthogonal Frequency-Division Multiplexing (OFDM) and Orthogonal Frequency-Division Multiplexing Access (OFDMA) in the optical domain leveraging the current cable operator experiences are explored. This includes a review of current industry initiatives and some predictions based on the uncovered parallels.

2. QAM is QAM

QAM is a technique that mixes both amplitude and phase variations in a carrier at the same time. This technique allows higher data rates within the same bandwidth. The higher the modulation order, the higher the spectral efficiency although higher signal fidelity is required to support such capacity gains. Figure 1 shows the constellation for QAM signals for different modulation orders.

The first observation about m-QAM signals and its detection process is that a constellation diagram is a geometrical problem that is valid for both RF and Fiber (Optical) QAM signals. Without getting into the mathematical details [2], the Bit Error Ratio (BER) can be estimated based on the symbol distance to the threshold levels defined by the boundary spacing. The received symbol being interrogated by the decision circuit will be a combination of the actual imperfect symbol and different noise sources, oscillator phase noise, system linearity, and other types of distortion. The probability of error for a symbol, assuming a Gaussian distribution, is estimated by computing the complementary error function of the effective (RMS) distance of the symbol to the boundary. Non-linear effects change the symbol cluster shape and the effective symbol distance to the boundary needs to be corrected for that. The energy (Power x Time) of the symbols in constellation affect the distance between the data points in the constellation. The higher the energy of the symbols the larger the distance between symbols.

The impact of such limitations and impairments are well known in the RF industry and can be identified in the constellation using the notorious HP/Agilent illustration [3] shown in Figure 2. This has been well documented [4,5] in the industry. The theoretical waterfall curves defining the minimum Signal to Noise Ratio (SNR) required for a maximum BER is well documented in the literature and in the industry.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

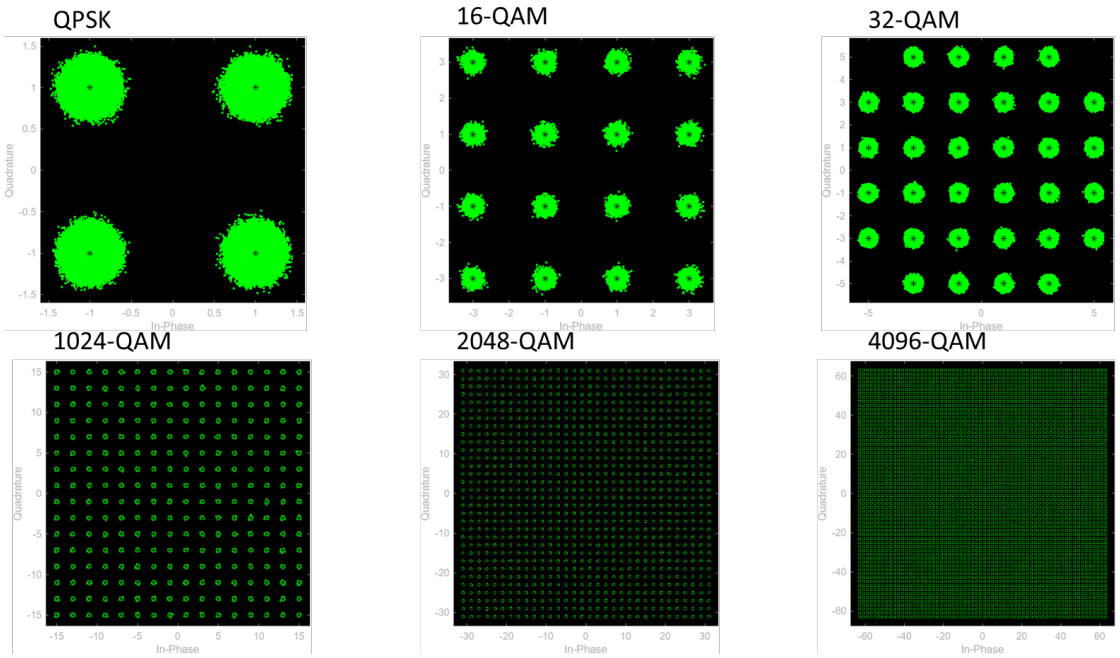


Figure 1 – Symbol constellation for different Modulation Orders

Also, for comparison purposes, Fiber-QAM state of the art is currently using 64-QAM, limited by the Baud rates the modules can handle. Meanwhile, RF-QAM is currently commonly using 4096-QAM on live networks, with the possibility of moving to 16k-QAM in the future.

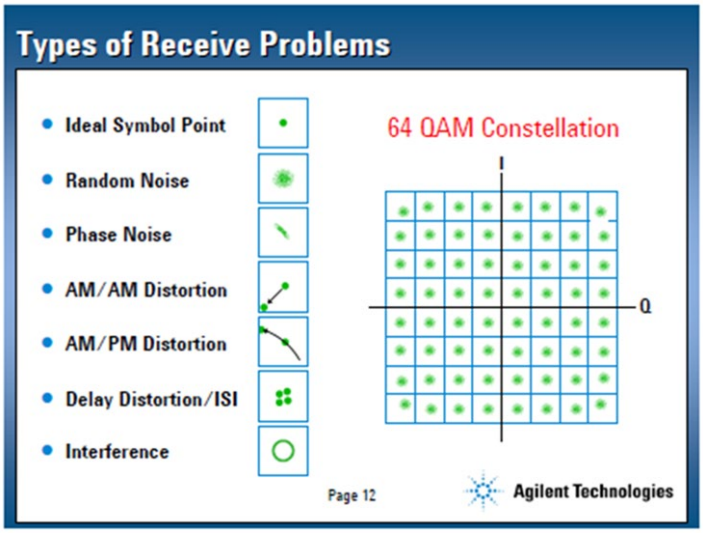


Figure 2 – Residual BER Prediction [2]

2.1. SNR vs OSNR

Figure 3 shows a few examples of BER vs. SNR curves for different modulation orders based on the geometric model [6]. The model applies to both RF and Fiber QAM, but as we will discuss next, there

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

are some terminology differences particularly when the industry refers to SNR and Optical SNR (OSNR). Many of us would think that OSNR is simply twice the SNR but that is not the case. The optics industry defines OSNR as the SNR referenced to 0.1nm (or 12.5GHz @ 1550nm) optical resolution bandwidth. Therefore, an OSNR, an optical signal with 30 GHz bandwidth, needs to be corrected by 3.8dB with respect to the SNR value. For example, a QPSK signal with 30GHz baud rate would require 8dB SNR to produce a BER around 2e-2 which, as we will see in the Forward Error Correction (FEC) section, would lead to error free operation after Soft-Decision Forward Error Correction (SD-FEC). The OSNR correction factor for this condition is 3.8dB as indicated in the Figure 3, and would lead to a minimum 11.8dB OSNR.

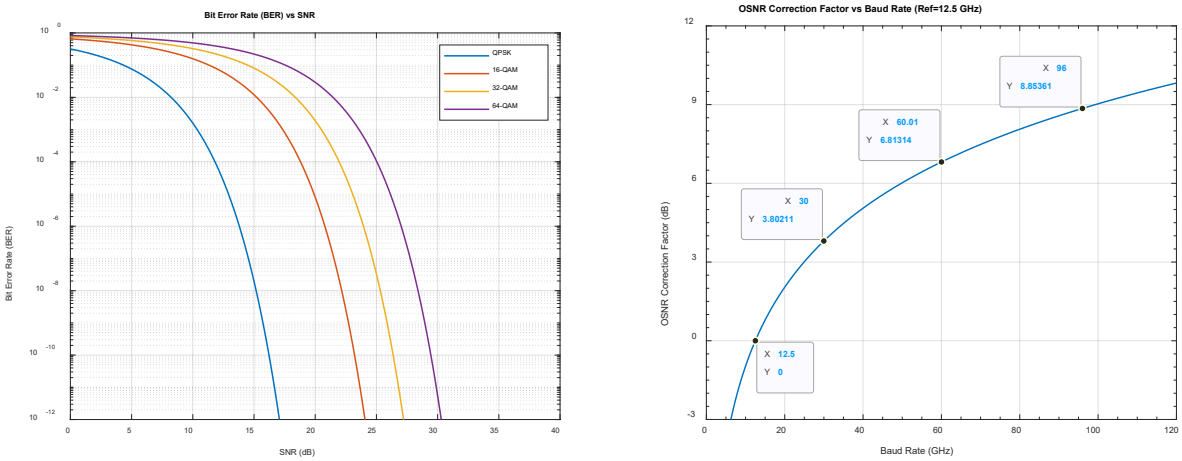


Figure 3 – BER vs SNR and OSNR correction factor plots for AWGN & no FEC

Further improvements in error rates are obtained in both RF and Optical QAM systems using different error correction techniques such as Low-Density Parity Check (LDPC) and other Shannon Capacity approaching codes and are discussed later in this paper.

Note: we will be loosely equaling baud rate or symbol rate with signal bandwidth across this paper but understand that spectral shaping and roll-off factor need to be considered in this relationship.

3. RF QAM Brief Review

As a quick review of the RF QAM history, reference [2] provides a great review of all the work that was done with digital communication systems that started in the late 1950s where the concept of coherent and non-coherent detection schemes were devised. By the 1990s, coherent communications took off thanks to the availability of components driven by satellite communications initially followed by wireless/mobile communications. Topics such as optimum constellation, coherent vs. non-coherent detection, clock recovery, filtering, channel model, equalization, besides many others, were explored and lead to a wide range of products, as well as standards, enabling different applications. Those are the very same topics in the forefront of the Fiber QAM developments.

Particularly with respect to the cable industry, the adoption of digital communication followed the communications industry, benefiting from the devices being developed for wired and mobile communications. In fact, the cable and broadband industry was first using 16, 64, and 256 QAM in 1997 when the wireless industry was still using QPSK and GMSK. This initial use of QAM modulation included a variety of symbol rates in the upstream for DOCSIS to increase speeds. Now cable broadband

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

has capabilities up to 16k QAM on OFDM, and 4k QAM on OFDMA when moving to DOCSIS 3.1 in 2013. These same capabilities exist today with spectrum used in both directions with Full-Duplex DOCSIS (FDX).

4. Fiber QAM Review

Fiber QAM has been following a similar path to the RF QAM development steps. The need for higher capacity has driven the industry to look for ways to increase spectral efficiency and symbol rate. Although there are many ways to increase capacity, one primary approach, particularly when fiber and wavelengths are scarce, is to use a higher order of modulation. That has driven the optical industry to develop a wide range of coherent optics modules in the last 10 years. For many years, coherent optics has been a topic with dedicated technical sections at the Optical Fiber Conference (OFC) since 2008 and at SCTE since 2017, and different standard driven efforts across the industry.

The main enabling technologies are the maturity of the optical manufacturing capabilities that are effectively moving optics from art to craft. The development of manufacturing capabilities is making it possible to, cost effectively, build more complex optical devices, particularly pluggable optics. A key technology particular to coherent optics is the development of powerful Digital Signal Processing (DSP) devices that makes these complex devices possible. 7nm, and now 5nm manufacturing resolution allows for higher density with more digital logic, higher speeds, and, potentially, lower power consumption devices.

4.1. The Form Factor Dilemma

Not long ago, coherent optics modules were only available in a line card form. Now, there are many form factors that have been proposed and standardized to a certain extent that allow for predictable integration efforts. Particularly in the coherent optics front, there are different form factors and applications currently available in the market. However, for the purpose of this paper, we will be focusing on two main contenders based on Comcast's current needs, the C Form-factor Pluggable type (CFP2) and Quad Small Form-factor Pluggable – Double Density (QSFP-DD), although the OSFP modules are also a potential option.

Figure 4 depicts the two modules while Table 1 provides a brief comparison of their main characteristics.



Figure 4 – Pictures of the CFP2-DCO (left) and QSFP-DD-DCO (right)

The CFP2 is the C form-pluggable and is the larger module with a larger area that can provide better thermal management for some of the Hybrid Fiber Coaxial (HFC) industry applications that Comcast started to investigate. These applications included outside plant (OSP) where the module is exposed to

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

wider temperature ranges. Initial industrial operating temperature range (-40 to +85^{0}C) requirements were reduced to take into consideration the temperature rise in certain applications. However, cold start under this requirement is still a must.

QSFP-DD is the double-density quad small form-factor pluggable. Its main advantages are density, due to its compact size, and lower power consumption. However, power consumption may be also reduced to other form factors by careful selection of features that the device supports.

Table 1 – Coherent Optics Pluggable Comparison

	CFP2	OSFP		QSFP-DD	
Form Factor	C Form-Pluggable	Octal Small Form-factor Pluggable		QSFP Double Density	
Specification	OIF-CFP2-DCO	OIF 400ZR	OpenZR+	OIF 400ZR	OpenZR+
Main Application	DCI, Metro, regional/long haul	Metro	Long Haul	Metro	Long Haul
Line Rate (Gbps)	100/400	400	100-400	400	100/400
Baud Rate (Gbaud)	28	32	60/63	60	30/60
Modulation Order	QPSK, 8QAM, 16QAM	16QAM	QPSK-16QAM PCS	16QAM	QPSK, 8QAM, 16QAM
FEC	SCFEC, SDFEC	SCFEC,GFEC	oFEC	cFEC	oFEC, CFEC
Target Reach (km)		<120	>120km	<120	450@400Gbps
Dimensions (mm)	41.5×12.4x106	22.6x13.5x108		18.4x8.5x89.4	
Power Consumption (W, max)	24	18	22	18	22

It would be ideal if interoperating different form factors were possible. In that way operators would be able to use high-density form factors in controlled environments, such as head-ends, while more thermally robust devices could be used in the OSP. Currently, that is far from reality due to the somewhat immature state of coherent optics.

4.2. Coherent Optics Implementation

DSP has been used in the CATV industry for a long time. Cable modem equalization, digital return, and edge QAM channel generation are all good examples of how the technology can help us to cost effectively deliver services to our customers. The semiconductor industry, driven by a wide range of technological innovations, has developed very high-density (5nm process), low-cost computing power, ADC performance, and higher electrical speeds. This 5 nm technology also results in a power reduction of 20% by reducing core voltages thus reducing leakage power.

Coherent optics is a complex communication system whose performance is degraded by a range of linear and non-linear distortions. These distortions include transceiver limitations such as frequency response, E/O and O/E non-linearity, and even manufacturing tolerances. There are also optical channel distortions such as Chromatic Dispersion (CD), Polarization Mode Dispersion (PMD), fiber non-linearity (XPM, SPM, FWM), optical filtering, and noise.

In general, most of the DSP to compensate for these challenges is done on the receiver side with the goal of improving system performance. The compensation is done in the form of equalization at different points of the coherent optical receiver for each of the optical polarization components. The length of the equalizer (number of taps) defines the compensation range. The wider the required compensation, the larger the device is with higher power consumption. Defining and limiting the requirements for the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

application is key. Figure 5 shows a generic block diagram for a coherent optical receiver indicating the DSP algorithms.

Digital equalization of chromatic dispersion is straightforward and understood. Applications such as Access and Metro optical links, with typically 80 and 120km, require relatively short tap lengths since the required compensation is under 1920 ps/nm, which can be accomplished with 256 taps up to 60 Gbaud rates [7].

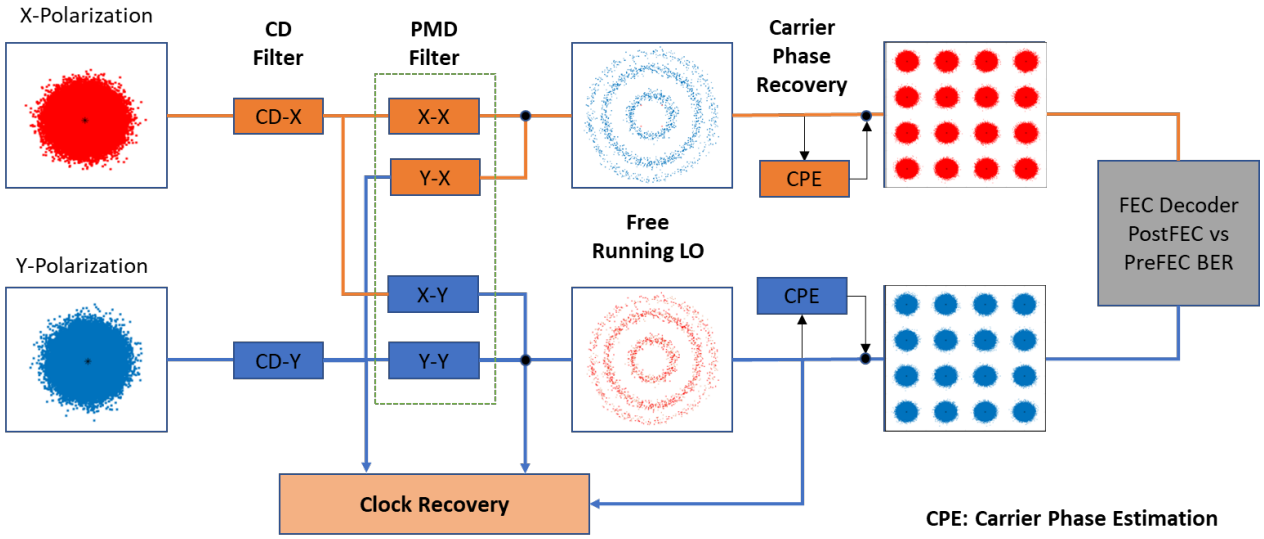


Figure 5 – Generic Coherent Optical Receiver Block Diagram (after [7])

It is important to note that Fiber QAM takes advantage of its powerful DSP capabilities to use the X and Y polarizations to double the channel capacity, something RF QAM should consider as well. Differential Group Delay (DGD) is produced by the optical fiber refractive index variation along their principal optical axis, typically denoted as X and Y. The refractive index variation, termed birefringence, is a random physical phenomenon that is temperature and wavelength dependent, and present in all fibers. As a note, most communication fibers are designed to introduce low birefringence, but some fibers are purposely made as highly birefringent, as in the case of the Polarization Maintaining (PM) fibers. The term DGD is sometimes exchanged with the related Polarization Mode Dispersion (PMD). However, DGD is a statistical representation of the random variations while PMD is its average value. The butterfly configuration of the PMD Filter shown in Figure 5 provides the means for tracking polarization variation. There are many DSP algorithms that were developed for such a purpose, using time or frequency domain implementations capable of compensating a wide range of dynamic environmental caused by temperature variation (rad/s) and vibrations (krad/s) and even lightning (Mrad/s) [7].

Carrier phase recovery algorithms in the optical domain are necessary to minimize the phase and frequency mismatch between Tx and Rx lasers in a coherent system. The process is the same as in the RF domain where laser-integrated phase noise needs to be minimized to minimally impact the system performance. Besides the Tx and Rx laser phase noise, channel impairments such as Self-Phase Modulation (SPM) and Cross-Phase Modulation (XPM) also need to be considered. There are many carrier phase recovery algorithms used for this purpose.

Although we focused here on these three main DSP components, we should also mention that non-linear compensation and timing recovery are also part of the channel compensation. The main trade-offs to keep

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

in mind are that the longer the filters used, the higher the latency and power consumption. Also, for the short-reach applications we are concerned with in the cable access network (<120km), we can optimize cost based on the shorter distances. Also, since the potential solutions exceed the required link budget for such application, designers could consider absorbing some of the power penalties imposed by such distortions. That is due in part to the powerful Forward Error Correction (FEC) currently used in the industry and that will be reviewed in the next section.

4.3. Forward Error Correction (FEC)

Forward Error Correction (FEC) is a method or algorithm whereby the original signal is encoded with error detection and correction redundant information. With this approach, RF or Optical receivers can detect and correct errors that occur in the transmission path and dramatically lower the BER and extend the distances that signals can be transmitted without regeneration. FEC is also commonly considered as an attractive, cost-effective tool to recover the lost sensitivity due to the transition to higher data rates.

There are several FEC algorithms to choose from that vary in complexity, strength, and performance across the industry. The most common and standardized first-generation FEC is Reed-Solomon (255, 239) that adds slightly less than 7% overhead and provides approximately 6 dB net coding gain.

Second and third-generation FECs became a reality as cost-effective computer power became more readily available. Second-generation FEC includes LDPC (Low Density Parity Check) used in DOCSIS 3.1 and Ultra FEC (UFEC) and Enhanced FEC (EFEC) in 10 and 40Gbps optics. These algorithms still use 7% overhead, but implement stronger, more complex encoding and decoding algorithms, which provide an additional 2–3 dB coding gain over Reed-Solomon.

Third generation FEC extends the performance and overall optical distances even further. Based on even more powerful encoding and decoding algorithms, iterative coding, and SD-FEC. The concept of SD-FEC is illustrated in Figure 6.

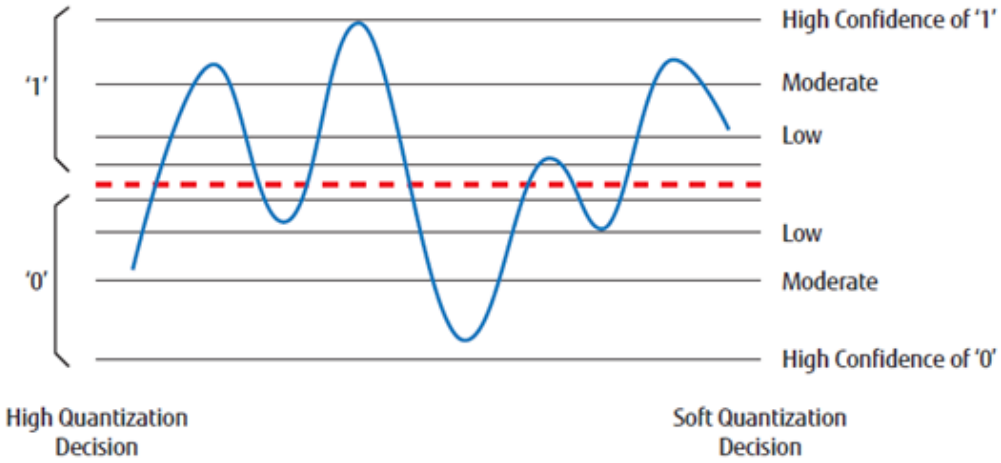


Figure 6 – Soft Decision concept (after [8])

Reference [9] provides a comprehensive review of FEC algorithms being currently used in coherent optics. It includes the definition of the metrics, Net Coding Gain (NCG), Coding Overhead, and the Pre-FEC BER threshold. Another metric that is not covered in this reference and that has become more

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

important nowadays is latency, which became a key design parameter for operators due to stringent end-to-end design guidelines.

To illustrate the trade-off between FEC gain, implementation latency, and power consumption, Table 3 shows the performance for a 100Gbps module operating with some different FEC modes [7, 10]. The table shows an increase of 19us in latency and ~2.5% in power consumption for a 2.1dB code gain, when moving from EFEC to oFEC. That corresponds to approximately 10km additional reach with a modest power consumption increase.

Table 2 – Metrics for standardized FEC in optical fiber transmission systems

Mode	Type	Coding Overhead (%)	Estimated NCG	Pre-FEC BER Limit	Latency (us)	Est. Power Consumption Variation
100G GFEC (Generic, RS)	HD	6.7	6.2	8.5E-05	11	0 (ref)
100G EFEC (Enhanced)	HD	12.3	9	4.5E-03	26	~+10%
100G oFEC (open)	SD	15.3	11.1	2.0E-02	45	~+12.5%
100G CFEC (Concatenated)	SD	14.8	11.2*	1.2E-01	45	~+15%

There is also proprietary FEC implementations in the industry that is part of the maturity concern of the coherent optics industry. As an example, XR optics, a sub-carrier based coherent optics technology for point-to-point and point-to-multipoint topologies, has data rates ranging from 25Gbps up to 400Gbps. Such technology has great potential in moving coherent optics to fulfill the potential of using coherent optics in the access networks by providing flexibility to operators to adjust capacity to their HFC, P2P Ethernet and PON customers, on demand. With the intent to address the proprietary concerns in the industry, the Open XR Forum [14] was formed in 2021 to promote open, standards-based adoption of XR optics.

5. Plant & Operations Implications

One of the most basic tasks of a cable operator's daily activities is the upkeep of their coaxial network. The reason for that is quite simple - the cleaner the HFC network is, the higher spectral efficiency is attained. That leads to a wide range of benefits including service experience and CAPEX deferral, among others. Addressing ingress, distortion, and micro-reflection issues is key to a well-run operation. However, RF-QAM is extremely adaptable and forgiving, with features such as DOCSIS Profile Management Application (PMA).

5.1. Plant Optical Return Loss

The very same concerns apply to the fiber network as well. Attention is needed to make sure the optical plant supports a higher order of modulation. The quality of the optical network needs to be monitored and managed as the coaxial plant. As operators start moving to a higher order of modulations, it is necessary to manage the optical reflections from the components used in the network, active and passive. When specifying the Optical Return Loss (ORL) of these devices, careful consideration should be taken, similar to what is done for the ORL of HFC analog optics in the not-so-distant past. Numbers in the 60 dB RL in the RF domain would translate to 30 dB ORL, which may be an issue if an Ultra Physical Contact (UPC)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

port is left unterminated and producing ~ 14 dB ORL. Keep in mind that losses before unterminated ports could improve ORL though that needs to be evaluated for each case.

Reflected light in an optical system, as we learned from our analog days, can create unwanted feedback to the lasers, adding non-linearities by changing frequency or modulation responses, degrading system performance in terms of Relative Intensity Noise (RIN) optical frequency variations, and even laser line-width variations. The system performance degradation can be expressed as power penalties, which can be as high as 2 dB depending on the number of reflection points. Figure 7 shows ORL power penalty for a 100Gbps Bi-Directional (BiDi) system. The expectation is that even higher power penalties could be produced at higher data rates and higher modulation orders.

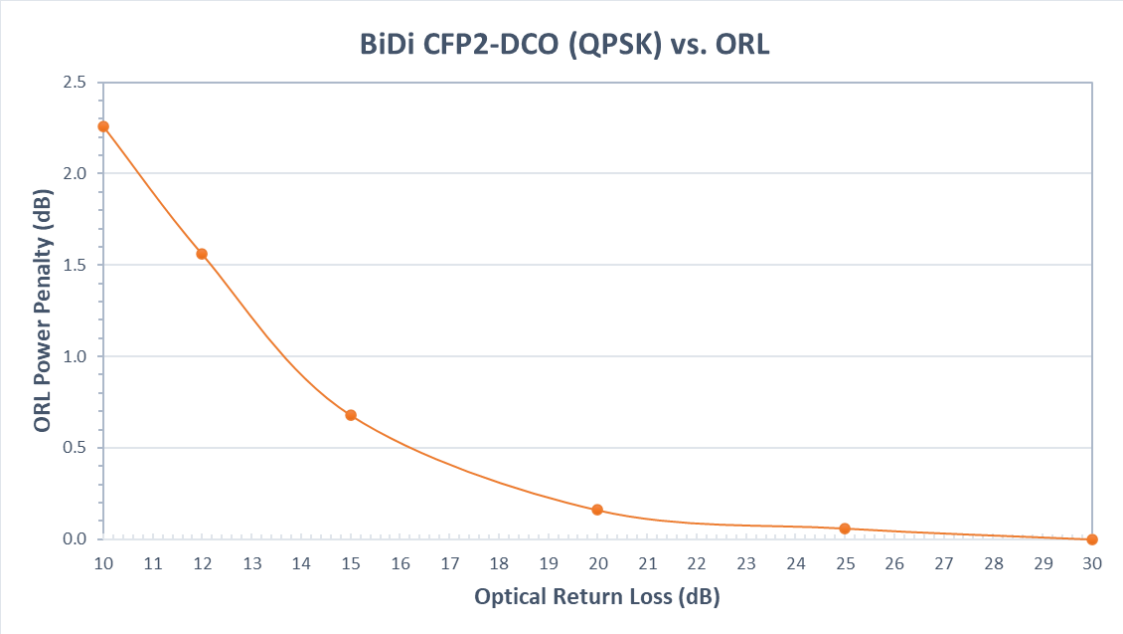


Figure 7 – Power penalty caused by the ORL in a 100Gbps QPSK BiDi system

It is strongly recommended to use Angled Physical Contact (APC) connectors instead of the UPC connectors commonly used in digital systems. That is again another parallel with the analog days of the industry when SC/APC connectors were a must. Lesson learned. It is important to note that current coherent optical modules do use a UPC connector and that is fine for the concerns above since they will always be terminated.

Another reflection phenomena that APC connectors will not address is Rayleigh Backscattering produced by the fiber. Light propagating in one direction is scattered by fiber imperfections (scatters) in the fiber core as one of the loss mechanisms in fiber. Part of the scattered light is coupled back in the fiber, but in the reverse direction. The level of Rayleigh Backscattering is very wavelength dependent (inverse fourth power), being more severe at shorter wavelengths.

5.2. Cascaded WDM Power Penalty

As we increase the data rates by increasing the baud rate, there is a concern that the interaction of wider bandwidth and the passband of passive optics like muxes and demuxes along the link may start to impact

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the performance. That impact could be quantified as a power penalty, which defines how much margin we need to add to a system to guarantee error free operation. Figure 8 illustrates the modulated linewidth of different types of pluggable devices at different data rates: Pink line for a10 Gbps SFP+, Yellow for 100 Gbps QPSK, Purple for 200 Gbps 8-QAM, and Green for 400 Gbps 16-QAM.

To verify the impact of the muxes as we increase the bandwidth and/or the modulation order of the transceivers, we measured a system-sensitivity change with no muxes, with a pair of muxes, and, finally, with 2 pairs of muxes. The measurements were done with 100Gbps/QPSK, 200Gbps/QPSK, 300Gbps/8QAM, and 400Gbps/16QAM signals and the results are summarized in Figure 9.

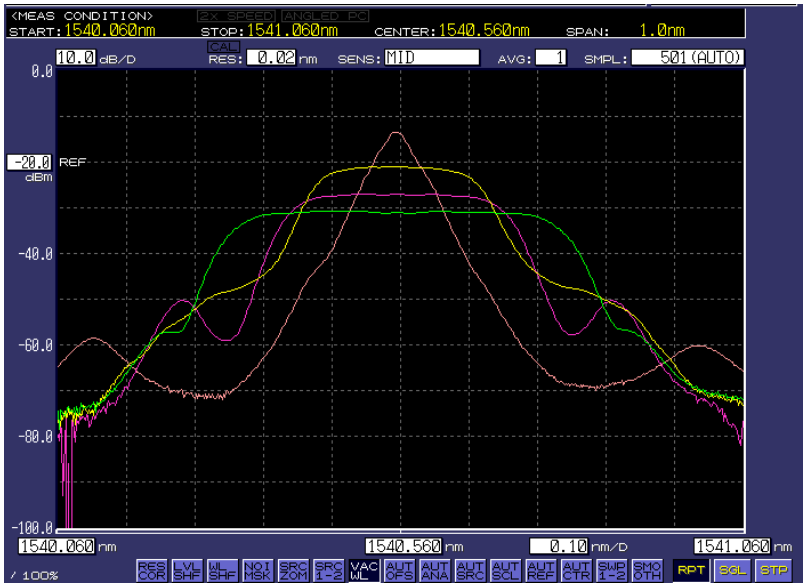


Figure 8 – Modulated linewidth at different data rates and modulation orders

Since the signal bandwidths, shown in Figure 8, are narrower than the WDM pass band, 100GHz, we don't see conclusive indication of WDM impact in the system performance, perhaps at higher data rate.

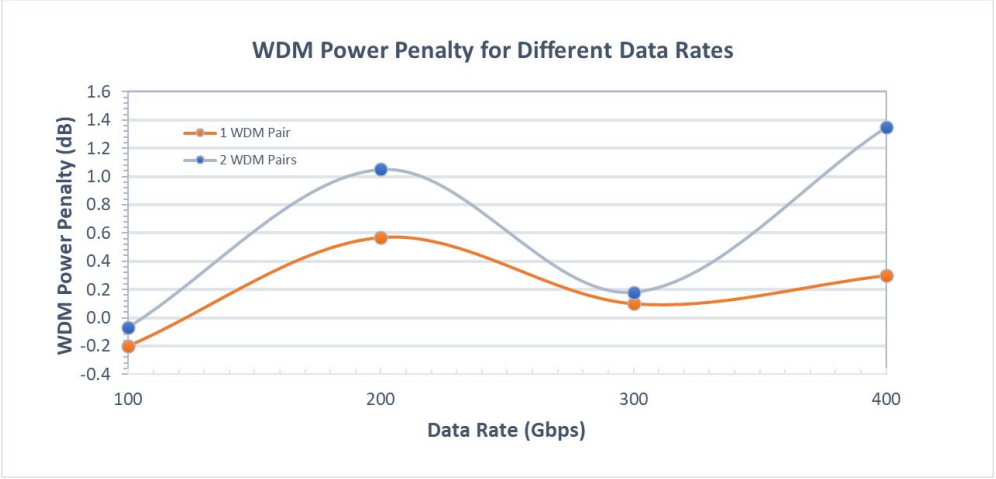


Figure 9 – Sensitivity Power Penalty caused by the WDM passband

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

We expect the impact would be more evident at higher data rates such as 800Gbps/64-QAM with expected 120GHz linewidths unless we keep increasing the modulation orders as it was done with RF-QAM. Another option is Probabilistic Constellation Shaping (PCS) where, by modifying the distribution of the data symbols to match the channel, the modulated linewidth takes a more Gaussian shape, which better endures the impact of channel bandwidth limitations. This technique also has the potential to conserve energy by using low-energy symbols more frequently, reducing signal distortion and allowing lower voltage operation. That may be an opportunity for RF-QAM as well.

5.3. Fiber Characterization

Comcast already has a wide range of tools to monitor their fiber infrastructure, both lighted-up (active) and dark (unused), using tools such as XMF-R and Link-It. However, as we start to use a higher data rate there is a higher modulation order.

We briefly discussed chromatic dispersion (CD), Differential Group Delay (DGD, and Polarization Dependent Loss (PD) in the DSP section. The impact of such parameters is due to the light-to-fiber interaction. The impact of such interaction, such as CD and DGD, are intrinsic to the fiber and can be managed by limiting the fiber length. Since Comcast has a wide range of legacy fiber in their networks, there is a need to keep track of such parameters. The same applies to PDL, but depending on how the fiber is installed, it may be necessary to make sure fiber birefringence produced by bending pressure points varying with temperature are not creating PDL issues.

6. Fiber QAM Industry Initiatives

There are currently many initiatives related to the adoption of coherent optics in the cable industry. That includes many CableLabs' specifications described in Table 3 below.

Table 3 – Current CableLabs Coherent Optics related Standards

Title	Published	Version	ID
CPON Architecture Specification	05/03/23	I01	CPON-SP-ARCH
Coherent Optics Termination Device OSSI Specification	01/27/23	I03	P2PCO-SP-CTD-OSSI
Coherent Termination Device Requirements Specification	06/09/21	I01	P2PCO-SP-CTD
P2P Coherent Optics Physical Layer 1.0 Specification	05/01/20	I03	P2PCO-SP-PHYv1.0
P2P Coherent Optics Physical Layer 2.0 Specification	05/01/20	I02	P2PCO-SP-PHYv2.0

Besides that, CableLabs has been sponsoring other activities such as the CableLabs Point-to-Point (P2P) Coherent Optics Interoperability Event that took place in December 2018 and July 2022, in addition to periodic optics-related calls (CableLabs Envision Vendor Forum: FTTP & Optical Day). It is important to note that interoperability efforts have been focused on the line side but the same should be done on the client side.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Meanwhile, Comcast started developing and deploying their own coherent optics-based solutions. The next section briefly describes those initiatives.

6.1. Application Scenario

One pressing issue that operators are facing is the lack of fiber available in the access networks for capacity expansion across its extensive access networks. The capacity expansion includes activities such as new installs, expansion, and node-splitting initiatives. That drove the Comcast Connect engineering team to consider new solutions. At the same time, there was a need to address some limitations of the Distributed Access Architecture (DAA) that Comcast has embraced, particularly in long-reach systems served by small secondary hubs.

Initial analysis comparing the deployment cost and deployment time for laying new fiber compared with the cost of adding a coherent optics solution was carried out. The comparison included the design, construction, licensing, and deploying of new fiber compared with the deployment of a 400Gbps system aggregating up to 40 x 10 Gbps links serving up to 40 optical nodes. After a few iterations and discussion with the deployment team and the optics industry, it was clear that the Coherent MuxPonder (CMP) solution was extremely competitive.

Figure 10 shows a block diagram of the current CMP solution currently being deployed by Comcast. It basically consists of a pair of Ciena 5171 switches configured as muxponder switches where each one of the 40 x 10Gbps ports on one side is mapped directly to the same port on the other side. Each group of 10 x 10 Gbps ports is aggregated into a dedicated 100Gbps BiDi CFP2 optics. This solution is also internally called DAAS Extension solution since it is equivalent to moving the DAAS switch ports further down in the OSP closer to the customers. Figure 11 exemplifies the typical secondary hub location that the solution is intended for to illustrate its physical and environmental challenges.

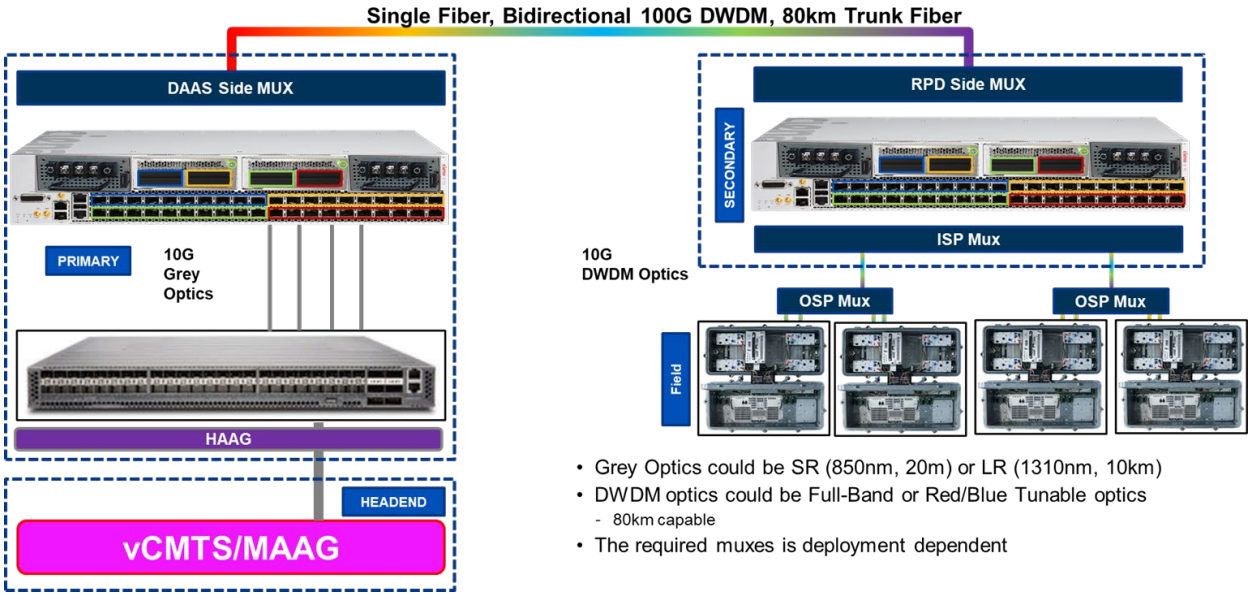


Figure 10 – DAAS Extension CMP solution

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.





Figure 11 – Bremerton Primary and Poulsbo Secondary in WA

This solution is currently deployed in two different regions that started in 2022. The first deployment took place in the West division. Its main objective was to provide an alternative solution for remote areas currently served with very long reach analog links with marginal performance. Since the CMP solution can extend the DAAS reach to more than 80km, not including the additional reach of the DWDM optics (additional 80km) at the secondary hub, making more than 160km links possible, depending on the optical loss of the fiber networks and passives. (It is important to note that the DOCSIS 150km limitation takes place from the demarcation point provided by the RPD.) The cutover results were very promising with average MER performance improvement around 6dB.

The second deployment in the Central division was also very successful since it was part of the Hurricane Ian recovery effort when Comcast was able to quickly reestablish service to the devastated Sanibel and Captiva islands area when only limited fiber was available. This recovery effort is better described in another paper being presented at this Technical Forum. Please check reference [13] for more details of this CMP success story.

Another effort using coherent optics is in the core network transport, where there are many ways of increasing the total fiber capacity. Comcast is already using 100 Gbps optics as their backbone and it is currently evaluating the progression to 400 and even 800Gbps.

From the longer-term point of view, Comcast also started looking at ways of potentially using coherent optics to match our DOCSIS capacity needs. This is explored in the next section while Section 8 discusses the nuances of capacity planning when related to coherent optics.

7. DOCSIS Implications

As stated before, capacity is the driving force behind the constant upgrades that operators need to provide to keep up. With Compound Annual Growth Rate (CAGR) between 20 and 25%, it is critical that operators work ahead to provide cost-effective solutions. That is true not only to the access network but, consequently, also to the metro (C-RAN) and backbone networks feeding the access networks. The goal is to avoid bottlenecks that limit the current demands and quality of service.

DOCSIS success and longevity is rooted in delivering the right capacity in a flexible, gradual, and cost-effective way to their customers. There are different ways to add capacity in the access, including increasing the modulation order, number (or width) of channels, or reducing the size of the service group

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

(number of subscribers served by a group of channels). Besides that, DOCSIS is very adaptable, allowing the modulator order to be adjusted depending on the channel signal quality. Features such as PMA provide such flexibility and allow the system to optimize spectral efficiency to the conditions. That is something Fiber-QAM does not support right now.

Capacity planning is an end-to-end exercise. As higher speed new service tiers start to be offered as FDX DOCSIS becomes part of our access options, proper capacity allocation needs to planned not only for the back-haul side, but also on the access network. Since FDX provides dynamic bandwidth allocation for both downstream and upstream, flexible capacity allocation serving FDX nodes/RPDs would be a great synergy. Optical solutions such as XR subcarrier technology [12,13] could provide some flexibility in the optical domain as well. The XR technology, intended for point-to-point and point-to-multipoint application, has been pioneered by Infinera and is now part of the Open XR Forum [14], which is one of the ways to address concerns of a proprietary technology.

Figure 11 shows the XR transmitter output spectrum for 3 conditions, 100Gbps (4 x 25 Gbps subcarriers), 200Gbps (8 x 25 Gbps subcarriers) and 400Gbps (16 x 25 Gbps subcarriers) using 16-QAM subcarriers. In this scenario, all subcarriers are being used to provide the selected bandwidth with each sub-carrier configured as an independent data pipe. This feature could be very useful for point-to-point and point-to-multipoint single-fiber applications. Note that the total power is the same for all the conditions.

The subcarrier utilization could also be set individually as downstream (downlink) and upstream (uplink) for all subcarriers in the same optical ITU channel. That enables different types of applications such as P2P, P2MP, and PON equivalent applications.

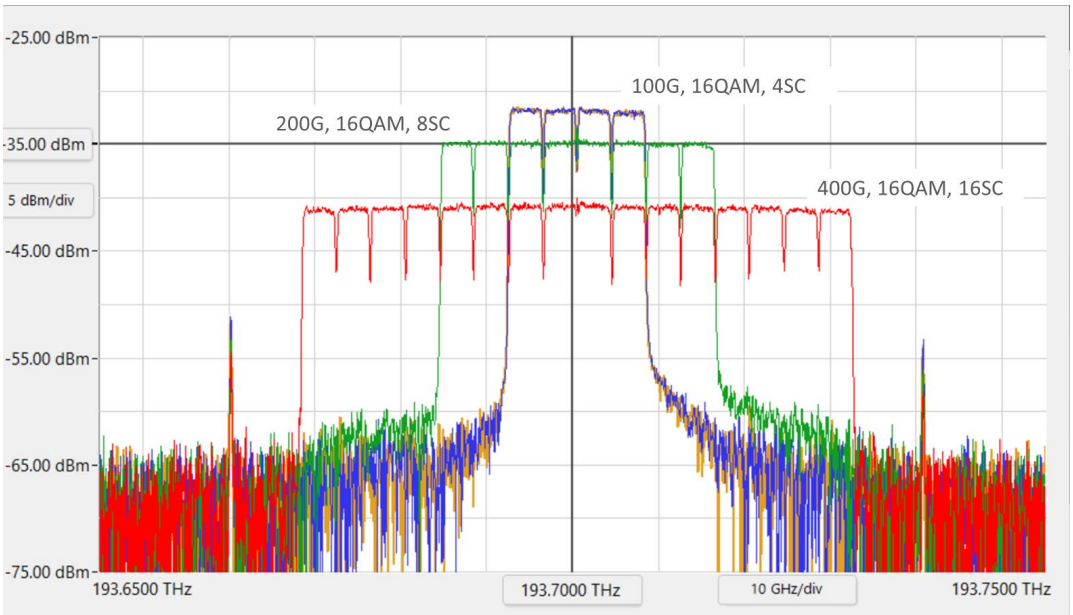


Figure 12 – XR Transmitter Optical Spectrum for 4, 8 and 16 bidirectional subcarriers

It is important to note that the functionality and flexibility the subcarrier technology offers could be replaced with less granular bidirectional channel multiplexing instead, since coherent optics devices are tunable. That is similar to the RF-QAM progression, which moved from granular SC-QAM to much larger OFDM blocks. The OFDM blocks are, however, composed of many subcarriers.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Such flexibility could be used to better manage the available bandwidth allocation across different end users, including fiber nodes, in a more advanced distributed access architecture (DAA). That may be possible if we can map the control signals currently used in a FDX DOCSIS. That is a departure from the current static capacity management to a more dynamic approach.

The main concern, from the operator's point of view, is that you may be paying extra for something that you may not use. However, the expectation is that using such a device over a wide range of applications may bring economy of scale which eliminates that expectation.

8. Capacity Implications

All this work has a simple objective: to make sure we can deliver the necessary bandwidths to our customers. The goal is to have a path to continuously increase the capacity in our core, metro, and access networks as demands require. The transition path must be smooth and cost effective.

Table 4, below, shows the estimated sensitivity and total capacity delivered by fiber for different data rates and modulation orders. The estimated sensitivity assumes a thermal noise limited system that is valid only when the optical power levels are low but provides an indication of the link budget requirements as we increase modulation order and bandwidth. Frequency roll-off and overhead due to FEC is assumed to be 20%. The total capacity assumes an Extended C-band system with 48 x 100GHz spaced DWDM system. Moving to a Super C + L band extension with 60 channels would add 25% capacity to the estimated numbers.

Table 4 – Total Capacity Estimation per Fiber for Different Modulation Orders

Data Rate	Modulation Order	Modulation Rate	# of Pol.	Symbol Rate	SNR @ 1e-3 PreFEC BER	OSNR @ 1e-3 PreFEC BER	Sensitivity (Th. Noise)	Total Capacity
Gbps/Ch	QAM	bits/S		Gbd or GHz	dB	dB	dBm	Tbps
100	QPSK	2	2	30	10.3	14.1	-31.7	4.8
200	QPSK	2	2	60	10.3	17.1	-28.7	9.6
200	8-QAM	3	2	40	14.1	19.2	-26.7	9.6
300	8-QAM	3	2	60	14.1	20.9	-24.9	14.4
400	8-QAM	3	2	80	14.1	22.2	-23.7	19.2
400	16-QAM	4	2	60	17.3	24.1	-21.7	19.2
400	32-QAM	5	2	48	20.3	26.1	-19.7	19.2
400	64-QAM	6	2	40	23.5	28.6	-17.3	19.2
800	32-QAM	5	2	96	20.5	29.4	-16.5	38.4
800	64-QAM	6	2	80	23.5	31.6	-14.3	38.4

Looking at the sensitivity requirements listed in the table, it is easy to notice the expected linear relationship with the bandwidth. Doubling the bandwidth would increase the sensitivity requirement by 3dB for the same modulation order. The same is not true about increasing the modulation order.

In terms of migration path, this table offers a few options. The first one is the default approach of adding more channels as necessary. Once all the wavelengths are taken, we can start the process of upgrading the modulation order if the switch and the pluggable modules support that. Upgrades up to 400Gbps should be easily adjusted by properly setting margins during the design or by adding amplification. Pluggable optical amplifiers are readily available and could be easily operationalized as other devices. That would support a smooth 800Gbps upgrade as well.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

9. Conclusions

As we show in this paper, the QAM challenges we are facing right now with coherent optics are similar to what RF-QAM followed over the many years it has been around and we believe we can learn a lot from that experience.

We have covered a few aspects that remind us of such parallels. Minimizing plant reflections, offering bandwidth granularity with optical subcarriers like SC-QAM, and concern with the pass-band of optical passives are a few of the many technical aspects that we need to further define and standardize across the industry.

However, there are many other technical aspects that we did not cover here that need more attention. The first is optical non-linearities and its impact on the system performance as we move to higher modulation orders. That is connected to other Comcast efforts related to Hollow Core Fiber (HFC) that we reported at last year's SCTE technical forum [15]. The hollow core approach reduces not only latency due to lack of glass and light interaction, but also has many linearity advantages that makes this type of fiber very attractive for many applications, including access networks.

A second aspect is making sure we don't degrade performance as we start to aggregate a large number of channels and subchannels to provide the capacities we need. That may lead to some opportunities to be better defined and a possible cost reduction in the optics that we are currently using. We believe that will be something similar to the Downstream RF Interface Specification (DRFI) [16] exercise that enabled the development of very dense Edge-QAM devices used in the industry.

The third aspect is to consider proactive network maintenance aspects of the coherent optical network by defining the key performance indicators (KPIs) to proactively address issues impacting service experience. We expect that many of the RF QAM KPIs would be used for that purpose.

In conclusion, coherent optics is maturing and ready to be used in operator links across the world. Let's use the lessons learned from RF-QAM to make Fiber-QAM another successful optics technology.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

ADC	Analog to Digital Converter
BER	Bit Error Ratio
BiDi	Bi-Directional
CAGR	Compound Annual Growth Rate
CD	Chromatic Dispersion
CFP2	C Form-factor Pluggable type 2
ComCMP	Coherent MuxPonder
DAA	Distributed Access Architecture
DAAS	DAA Switch
DCO	Digital Coherent Optics
DGD	Differential Group Delay
DOCSIS	Data Over Cable Service Interface Specification
DRFI	Downstream RF Interface
DSP	Digital Signal Processing
DWDM	Dense Wavelength Division Multiplex
Gbps	Giga bits per second
FDX	Full Duplex (DOCSIS)
FEC	Forward Error Correction
FWM	Four Wave Mixing
GMSK	Gaussian Minimum-Shift Keying
Hz	Hertz
LDPC	Low Density Parity Check
MER	Modulation Error Ratio
NCG	Net Coding Gain
OFDM	Orthogonal Frequency-Division Multiplexing
OFDMA	OFDM Access
OFC	Optical Fiber Conference
OSFP	Octal SFP
OSP	Outside Plant
ORL	Optical Return Loss
OSNR	Optical SNR
P2P	Point-to-Point
P2MP	Point-to-MultiPoint
PM	Polarization Maintaining
PMA	Profile Management Application
PMD	Polarization Mode Dispersion
PON	Passive Optical Network
QAM	Quadrature Amplitude Modulation
QSFP	Quad Small Form-factor Pluggable
QPSK	Quadrature Phase Shift Keying
QSFP-DD	QSFP – Double Density
RF	Radio Frequency
SC	Subcarrier
SCTE	Society of Cable Telecommunications Engineers
SFP	Small Form-factor Pluggable
SNR	Signal to Noise Ratio

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SPM	Self-Phase Modulation
WDM	Wavelength Division Multiplex
XPM	Cross Phase Modulation

Bibliography & References

[1] https://www.nctatechnicalpapers.com/Paper/2022/FTF22_CONV03_Mutalik_3796

[2] Modern Quadrature Amplitude Modulation – Principles and Applications for Fixed and Wireless Channels, W. T. Webb and L. Hanzo, IEEE Press & J. Wiley, 1994

[3] Residual BER Prediction – June 27, 2001, presented by Kent Johnson, Agilent – Also consider the HP AN355, Digital Radio Theory and Measurements, 1988

[4] The Grown-Up Potential of a Teenage PHY, R Howald, A Vieira, NCTA Convention and Exposition, 2012

[5] New Megabits, Same Megahertz: Plant Evolution Dividends, R Howald, A Vieira and M Aviles, Spring Technical Forum Proceedings, 2009

[6] Broadband Hybrid Fiber/Coax Access System Technologies, Winston I. Way, Academic Press, 1999

[7] SC393 – Digital Signal Processing for Coherent Optics Optical Transceivers, Chris Fludger, OFC Short Course Notes, 2023

[8] Soft-Decision FEC Benefits for 100G, White Paper, Fujtsu, 2012

[9] Steve Jia, “Forward Error Correction (FEC): A Primer on the Essential Element for Optical Transmission Interoperability”, CableLabs Wired, April 2019

[10] R. Nagarajan et al, “Low Power DSP-Based Transceivers for Data Center Optical Fiber Communications (Invited Tutorial)”, IEEE JLT, VOL. 39, No. 16, August 15, 2021

[11] Coherence After Turbulence: Coherent Links for Hurricane Relief in the Sunshine State, Venk Mutalik et al., SCTE CableTec Expo 23

[12] Coherent Optics at 400G, 800G, and Beyond, Heavy Reading White Paper, November 2021

[13] ICE XR Demonstration: Introduction to XR Optics, SCTE CableTec Expo 2022

[14] https://www.openxrforum.org/

[15] https://www.nctatechnicalpapers.com/Paper/2022/FTF22_WLINE09_Mutalik_3920

[16] https://www.cablelabs.com/specifications/CM-SP-DRFI

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Fighting FTTH with DOCSIS

A Technical Paper prepared for SCTE by

Mike Darling
Principal Engineer
Rogers Communications
mike.darling@sjrb.ca

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. The DOCSIS Ecosystem 4
   2.1. Comparison to DSL 5
3. Traffic Growth 8
   3.1. IPTV transition 9
   3.2. Higher Resolutions 9
   3.3. Increased Coding Efficiency 10
4. DOCSIS Capacity Calculations 11
5. Traffic Patterns 12
6. FTTH Competition 13
7. HFC Upgrade Options 14
   7.1. Sub-split 14
   7.2. Mid-split 15
   7.3. High-split 16
   7.4. Ultra-high-split 17
8. Strategies 18
   8.1. Maintaining a Sub-split Plant 18
      8.1.1. Capacity Perspective 18
      8.1.2. Node Splits 18
      8.1.3. IPTV Transition 20
      8.1.4. DOCSIS CPE Upgrades 20
   8.2. Mid-split Upgrade 20
   8.3. High-split Upgrade 21
   8.4. Ultra-high-split Upgrade 22
   8.5. FTTH Upgrade 23
9. Conclusion 24

Abbreviations 26

Bibliography & References 27

List of Figures

Title **Page Number**

Figure 1 – Broadband Subscriptions by Technology - OECD [1] 4
Figure 2 – DSL Speed Evolution [2] 5
Figure 3 – Crosstalk in DSL Networks 5
Figure 4 – Speed vs Distance in DSL Networks [3] 6
Figure 5 – Subscriber Speeds Over Time 7
Figure 6 – Downstream Peak Network Traffic and YoY Growth in 2010s 8
Figure 7 – Downstream Peak Network Traffic and YoY Growth 8
Figure 8 – Upstream Peak Network Traffic and YoY Growth 9
Figure 9 – Optimal Viewing Distance [4] 10
Figure 10 – Coding Efficiency of Video Standards [5] 11
Figure 11 – Downstream SNR 12
Figure 12 – DS:US Traffic and Provisioned Ratios 13

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 13 – Passive Optical Network Coexistence 14
Figure 14 – Sub-split HFC Network 14
Figure 15 – Mid-split HFC Network 15
Figure 16 – High-split HFC Network 16
Figure 17 – Ultra-high-split HFC Network 17
Figure 18 – Node Split Example 19
Figure 19 – Peak Hour Traffic Distribution 19
Figure 20 – Tavg vs Nsub 20
Figure 21 – High-split and UHS Capacities 22
Figure 22 – TCO Comparison Between FTTH and HFC Upgrades 24

List of Tables

Title **Page Number**

Table 1 – Available Spectrum in Sub-split Plant 15
Table 2 – Capacity in Sub-split Plant 15
Table 3 – Available Spectrum in Mid-split Plant 15
Table 4 – Capacity in Mid-split Plant 16
Table 5 – Available Spectrum in High-split Plant 16
Table 6 – Capacity in High-split Plant 16
Table 7 – Available Spectrum in Ultra-high-split Plant 17
Table 8 – Capacity in Ultra-high-split Plant 18

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Cable operators are preparing to, or have already begun, upgrading their hybrid fibre-coax (HFC) networks in order to offer higher-speed service tiers. There are many upgrade paths and corresponding strategies that cable operators could take, each with different considerations. In this paper each upgrade path will be analyzed from a strategic point of view contemplating topics such as capacity, tier capability, and traffic growth. These circumstances could be compared to those faced by telcos a decade ago when they had to decide whether to upgrade their digital subscriber line (DSL) networks to maintain competitiveness with HFC, although unlike DSL, HFC networks have a long life ahead of them. This paper will consider if lessons can be learned by cable operators from those circumstances. This round of network upgrades represents a significant investment, underlying the importance of a well-thought-out strategy to optimize return on investment.

2. The DOCSIS Ecosystem

The share of global broadband subscribers served with HFC technology has plateaued in recent years. This is primarily due to three reasons:

1. Most cable operators have moved to fibre-to-the-home (FTTH) in their greenfield builds, limiting the number of new HFC homes passed built each year.
2. Government subsidy programs are geared toward FTTH in underserved areas, increasing the overall share of FTTH homes passed.
3. A subset of HFC operators are upgrading their HFC plant to FTTH.

This trend can be seen in the OECD broadband data [1] in Figure 1. While both FTTH and HFC had been increasing market share until 2020 at the expense of DSL, in the last two years HFC's share has been reduced as well.

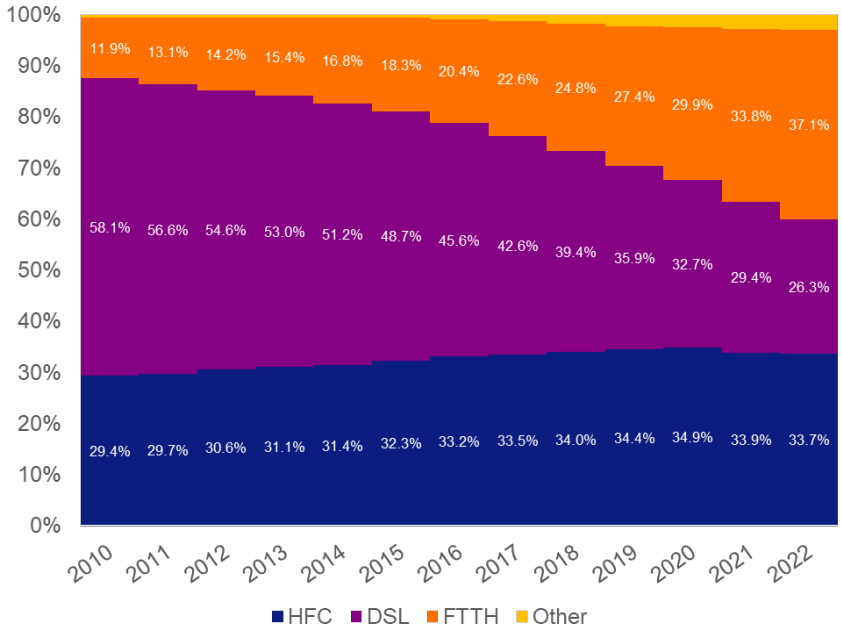


Figure 1 – Broadband Subscriptions by Technology - OECD [1]

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Looking at this trend, some are questioning the future of HFC technology and whether there will be continued innovation in the ecosystem. The myriad potential upgrade options and publicly announced plans by cable operators have added to the complexity faced by cable operators considering upgrade paths. This is similar to the situation telcos found themselves in years ago as the DSL market began to contract, with a few very important differences. An analysis of this situation has the potential to inform the current discussion.

2.1. Comparison to DSL

Telcos in Canada began to transition away from their legacy DSL networks approximately 10 years ago. It is worthwhile examining the circumstances that led them to that decision. Just as data over cable service interface specifications (DOCSIS) technology improved through DOCSIS 1.0, 1.1, 2.0, 3.0, 3.1 and most recently DOCSIS 4.0, broadband DSL had its own path. Starting with Asymmetric DSL (ADSL), the technology moved to very-high-speed DSL (VDSL), and finally G.fast [2], adding variants that improved speeds along the way, as shown in Figure 2.

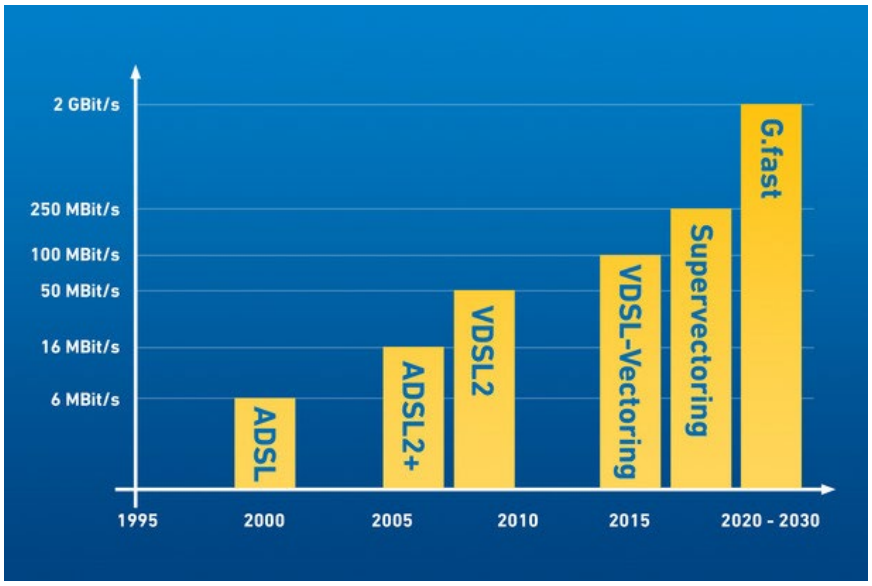


Figure 2 – DSL Speed Evolution [2]

As with HFC technology, DSL standards incorporated additional spectrum over time. Unlike HFC technology, one of the main challenges for DSL is crosstalk, or interference between subscriber lines, which are carried in bundles as shown in Figure 3.

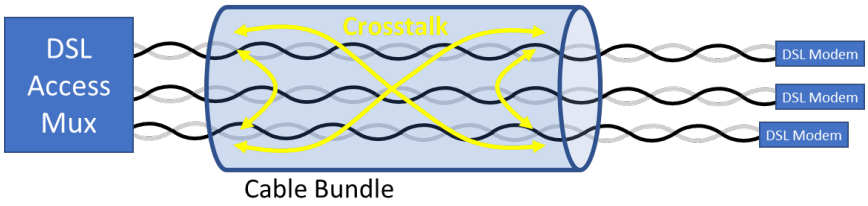


Figure 3 – Crosstalk in DSL Networks

To increase speeds over the DSL network, crosstalk had to be minimized. To this end, vectoring, which pre-distorts the signal before being sent through the carrier bundle, began being used to balance out the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

impact of crosstalk. There is only so much crosstalk reduction to be accomplished through vectoring however, and to support higher speeds the distance between the access node and CPE had to be minimized, as shown in Figure 4 – Fibre-to-the-curb (FTTC) Speed Graph.

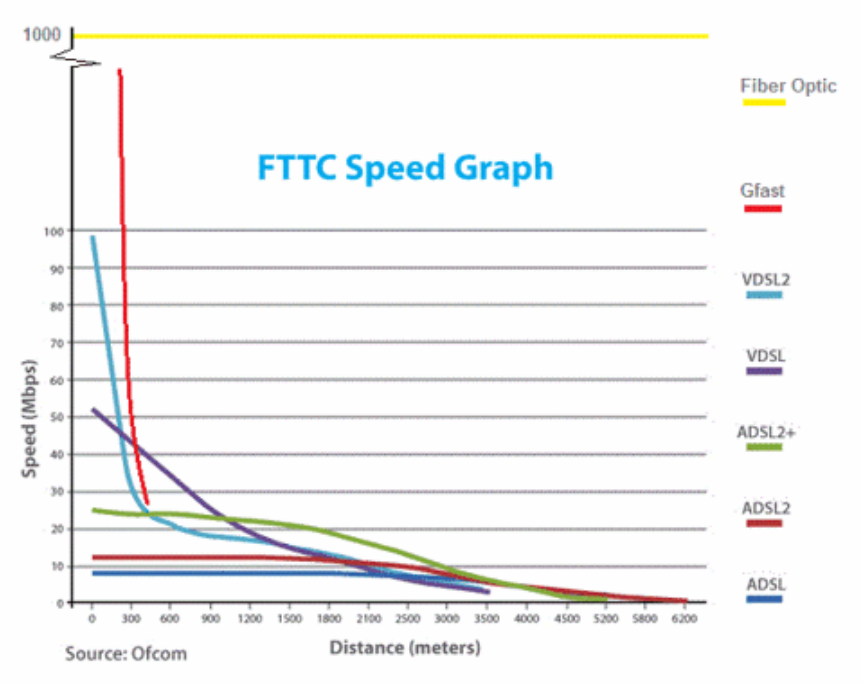


Figure 4 – Speed vs Distance in DSL Networks [3]

To reduce distances, telco operators were required to change their outside plant architectures, which began with loop bundles originating from central offices (COs) and travelling several kilometers to subscriber premises. The new architecture made use of DSL access multiplexers (DSLAMs) installed in the outside plant, which brought fibre connectivity closer to the subscriber, limiting the distance of the bundle and thus reducing crosstalk.

Over time, subscriber speed requirements increased due to internet protocol (IP) video adoption and competition from HFC operators who could offer speeds higher than VDSL2 technology could support. This forced the telcos to make a decision whether to upgrade their DSL networks to G.fast, which could increase speeds offered, but required DSLAMs to be moved even closer to subscribers; or to move directly to FTTH.

In Canada the telcos decided to forego a large-scale upgrade to G.fast, deciding instead to begin upgrading their networks directly to FTTH. The dynamic for both cable and telco operators is shown in Figure 5.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

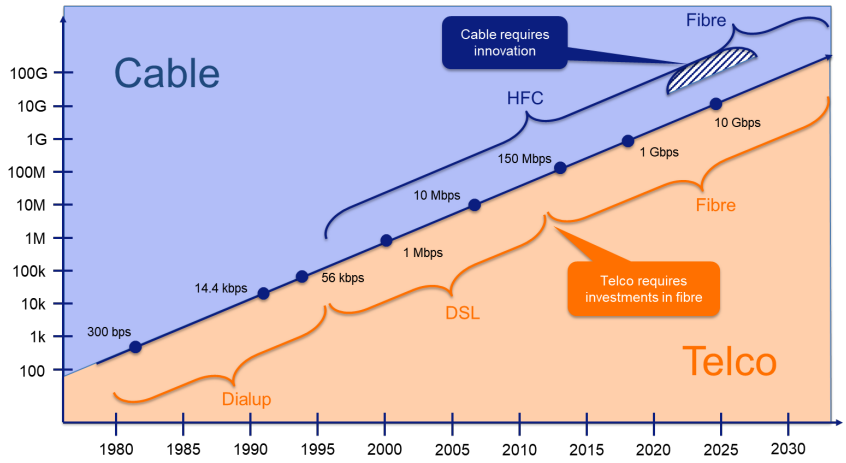


Figure 5 – Subscriber Speeds Over Time

There are some who would argue that cable operators are in the same place now as the telcos were approximately 10 years ago, having to decide whether to perform another upgrade to their current technology, or make the move to an all-fibre network. There are important differences in this situation, however.

A significant difference is that a G.fast upgrade to telco DSL networks would require a large number of net-new DSLAMs to be installed in the network, along with the temperature controlled cabinets that house them in order to reduce the loop lengths between DSLAMs and subscribers. On the HFC side, a DOCSIS 4.0 upgrade only requires the exchange of equipment to newer versions.

Another difference is the additional speed achieved by the upgrade. The speed increase realized by G.fast would strongly depend on the loop distances, but it is likely to have topped out at several hundred Mbps. This would not have been sufficient to match DOCSIS 3.0 speeds beginning to be available at that time, which had a downstream capacity of greater than 1 Gbps. DOCSIS 4.0 has the potential to upgrade downstream speeds to over 10 Gbps which is on par, or slightly better than, 10 Gbps Passive Optical Network (XGS-PON), the current top of the line FTTH technology being deployed.

The last difference is the traffic growth environment. In the mid-2010s year-over-year (YoY) traffic growth was routinely over 50%, as shown for the Western Canadian cable footprint in Figure 6, causing network demand to double in less than two years. The solid line is peak traffic, shown for reference but without an axis, while the dashed line is YoY growth, measured monthly. This is drastically different from today, where growth has slowed to below 20% and is forecasted to drop even further.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

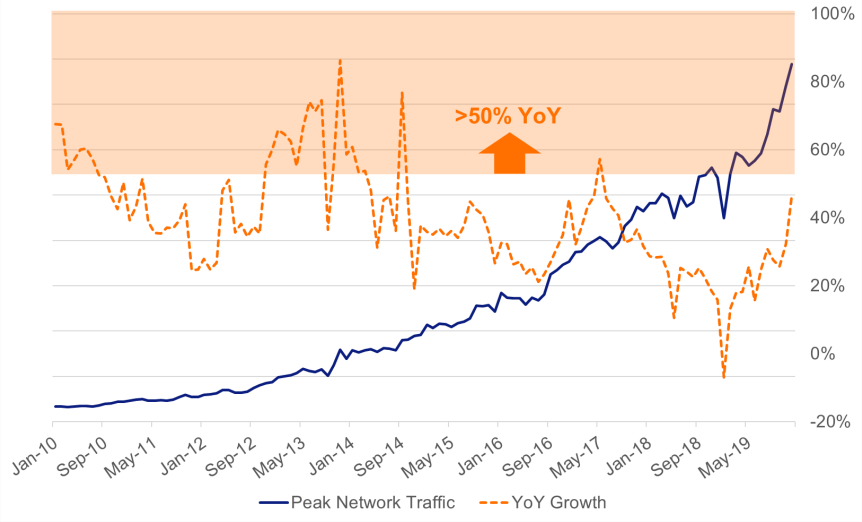


Figure 6 – Downstream Peak Network Traffic and YoY Growth in 2010s

The important question that cable operators must answer is which upgrade path is optimal. The answer will depend on what cost per home passed the operator forecasts for each type of upgrade, and what kind of traffic growth each operator has forecasted for the coming years. The former topic has been the subject of much debate, but will be specific to each operator's context, while the latter topic will be explored in the next section.

3. Traffic Growth

The COVID-19 pandemic generated a large amount of network traffic in a short time due to remote-learning, work from home, and people spending more time in their residences. As the pandemic started to wane, traffic growth slowed, resuming a trend that had started in the years before COVID-19. Figure 7 shows the total network traffic trend in the Western Canadian cable footprint in the last six years. The large growth due to COVID-19 can clearly be seen in the higher YoY growth in the months from roughly March 2020 to March 2021. In the two years since that time, the YoY growth rate has hovered around 10%.

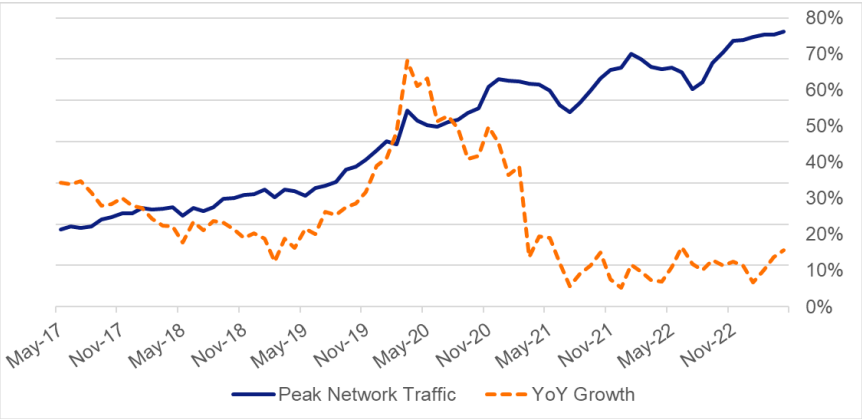


Figure 7 – Downstream Peak Network Traffic and YoY Growth

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In the upstream, the story is much the same. Pre-pandemic YoY growth was slowing, a trend that returned at the end of the pandemic, with YoY growth below zero for some specific months as shown in Figure 8.

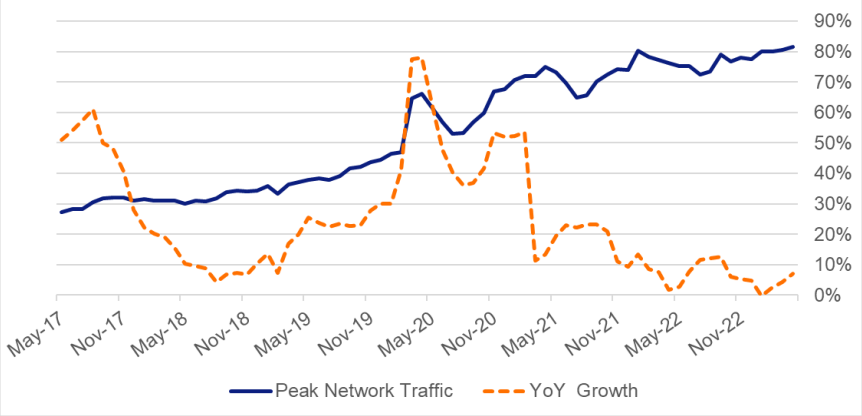


Figure 8 – Upstream Peak Network Traffic and YoY Growth

Forecasting network traffic growth is difficult, as it tends to come from new and unpredicted network use. However, most network growth has been triggered by some variety of IP video. In the mid-2010s this was Netflix, YouTube, and the greater adoption of IP Television (IPTV), while during the COVID-19 pandemic it was video conferencing that especially increased upstream growth rates. Given this we can analyze the potential impacts of video in the coming years.

3.1. IPTV transition

While IPTV adoption is high, cable operators continue to offer broadcast video based on Quadrature Amplitude Modulated (QAM) signals. Most operators have chosen to move to a unicast or multicast IPTV platform as there is no limitation on the number of channels carried. As subscribers are transitioned from the QAM infrastructure to IPTV, additional traffic is added to the DOCSIS network. The impact of the transition depends to what extent the subscribers were watching broadcast video and to what extent they were watching over-the-top (OTT) IPTV services that were already being carried over the DOCSIS network. In the worst case, subscribers can be considered to have increased their network use from a non-IPTV user to match the average IPTV subscriber. Most broadband video subscribers watch OTT IPTV services prior to switching to linear IPTV, thus reducing the impact of the transition. In addition, operators can influence the speed at which the IPTV transition happens through marketing strategies. Lastly, the upside of the IPTV transition is when all subscribers are transitioned, the spectrum currently used to carry broadcast video can be freed up for DOCSIS signals, increasing the capacity available. At current DOCSIS serving group sizes, this transition ends up using spectrum more efficiently, although creating an accordion effect where both IPTV and broadcast video must be supported until the latter is no longer required.

3.2. Higher Resolutions

In addition to increased IPTV subscriber counts, the resolution of IPTV signals, and thus the quantity of traffic, is increasing over time. 4K video signals create roughly four times the traffic as high-definition (HD) signals. Although 4K television adoption is high and increasing, content continues to be limited. This is partially due to 4K having limited perceived benefit over HD in most viewing environments. As shown in Figure 9, 4K or Ultra HD only has a perceived benefit either for large televisions or for viewers sitting close to their TV.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

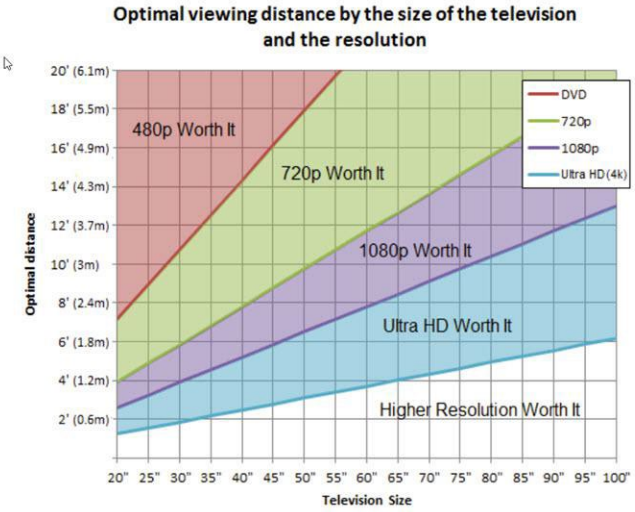


Figure 9 – Optimal Viewing Distance [4]

The first content category that will drive more 4K adoption is sports, which benefits from higher resolutions and higher bitrates in general due to quick motion. This could be observed in IPTV network traffic statistics during the National Hockey League (NHL) playoffs this year, where games were carried in 4K. Especially when local teams were involved, 4K viewership drove increased peak network traffic.

3.3. Increased Coding Efficiency

Video bitrates for the same resolution have decreased over time, owing to the increased coding efficiency of new compression standards. Many operators are still carrying Motion Picture Experts Group Version 2 (MPEG-2) signals in their broadcast video platforms but have moved to MPEG-4 for IPTV. As shown in Figure 10, the high efficiency video coding (HEVC) standard will allow for a further approximately 50% reduction in bitrates when compared to MPEG-4.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

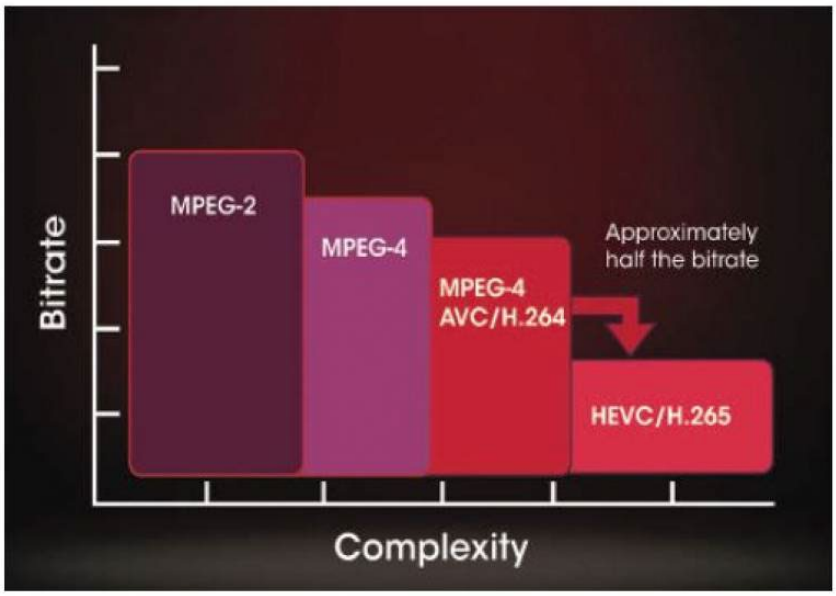


Figure 10 – Coding Efficiency of Video Standards [5]

The two trends of increased video resolution and increased video coding efficiency are likely to mostly cancel each other out over time. Future network growth is likely to be driven by some form of video. Luckily for network operators, they have influence over video bitrates and speed of IPTV transition and thus some measure of control over network traffic growth.

4. DOCSIS Capacity Calculations

Network capacity calculations can be done using several different methods. There is the physical data rate (PHY rate), which includes all overhead, there is also the theoretical maximum which takes the PHY rate and subtracts out the theoretical overhead, and finally there is the practical data rate observed in real networks. While marketing material may make use of PHY rates or theoretical maximums, network planners need to use real-world capacity calculations.

In calculating DOCSIS capacity, achievable modulation rates must be considered. While in most circumstances a modem may receive a signal with high enough signal-to-noise ratio (SNR) to be able to use 4096-QAM signals, there may be a material number which cannot. In Figure 11, the SNR distribution of several hundred thousand modems in Western Canada is displayed.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

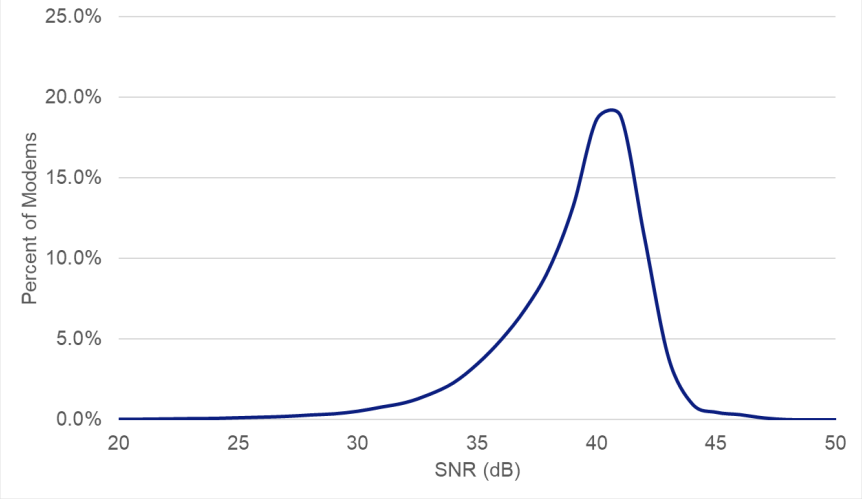


Figure 11 – Downstream SNR

Assuming 4096-QAM is achieved when SNR is greater than 38 dB, the majority of the network would support that modulation rate. However, a significant portion of the network has an SNR below 38 dB. Care must be taken when declaring the capacity of the network, as using 4096-QAM will overcount the capacity to some of the network and using 1024-QAM would undercount the majority. Overcounting may lead to circumstances where operational teams are pushed to increase SNR to achieve 4096-QAM in areas of the network where it is difficult and expensive. One method of estimating capacity is to decide what percentage of the network is less than 4096-QAM acceptable. In this case capacity calculations can be done using the 90th percentile, or what performance 90 percent of the network is able to achieve. In this paper 9 bps per Hertz (bps/Hz) is used in the downstream and 7 bps/Hz in the upstream.

5. Traffic Patterns

ADSL was the first version of DSL technology aimed at multimedia access, including video. Early standards had downstream to upstream (DS:US) capacity ratios of 10:1 or greater. It was only with the introduction of VDSL2 that high-speed symmetric services were offered.

FTTH technologies have symmetric and asymmetric variants, with the most common in use in Canada currently being Gigabit Passive Optical Network (GPON) with 2.5 Gbps downstream and 1.25 Gbps upstream PHY rates. While this is asymmetric, the telcos offer mostly symmetric services, only offering asymmetric tiers when the upstream cannot match the downstream, such as 1.5 Gbps downstream by 940 Mbps upstream. The next generation of FTTH technology is XGS-PON, with a PHY rate of 10 Gbps symmetric, which will allow operators to offer symmetric tiers greater than 1 Gbps. Offers up to 8 Gbps symmetric over XGS-PON are in market today.

Starting with VDSL2, telcos have tried to highlight their symmetric services as an advantage over HFC technology, which has always been asymmetric. The success of this strategy has been limited as network traffic is inherently asymmetric.

In the Western Canadian cable footprint the network has exhibited a DS:US traffic ratio of approximately 15:1 since the mid-2010s. This has remained stable despite some major changes in the DS:US ratio of offered tiers, which aggregated for all subscribers makes up the network provisioned ratio. Figure 12 shows both the DS:US traffic ratio and the DS:US provisioned ratio. There were two major changes in DS:US provisioned ratio, first when downstream provisioned rates for several popular tiers were doubled

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

overnight, and next when upstream provisioned rates were increased as a mid-split upgrade was close to completion. As can be seen, the DS:US traffic ratio stayed the same, despite the provisioned ratio changing quite drastically. This points to the traffic ratio being an inherent product of subscriber behaviour rather than a product of network characteristics.

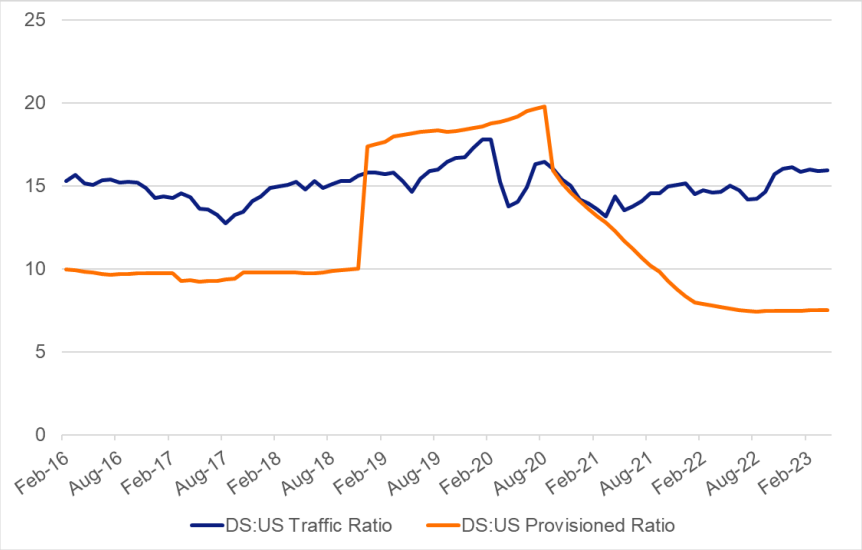


Figure 12 – DS:US Traffic and Provisioned Ratios

This leads to the question as to whether symmetric tiers are necessary, and if not, whether they will become so in the future. As with traffic growth it is not easy to predict what applications might cause a change in DS:US traffic ratios, but it is difficult to envision a wide-spread scenario where subscribers share more information with the world than they consume.

6. FTTH Competition

Canadian telcos continue to aggressively upgrade their DSL networks to FTTH. As of the end of 2022, Bell's network was 70% FTTH [7], and Telus' network was 91% FTTH [8]. They currently offer speeds of 1.5 Gbps downstream and 940 Mbps upstream over GPON, while quickly upgrading their platforms to XGS-PON. They are able to upgrade their networks to XGS-PON by adding new network equipment that can coexist with installed GPON equipment as shown in Figure 13 [6].

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

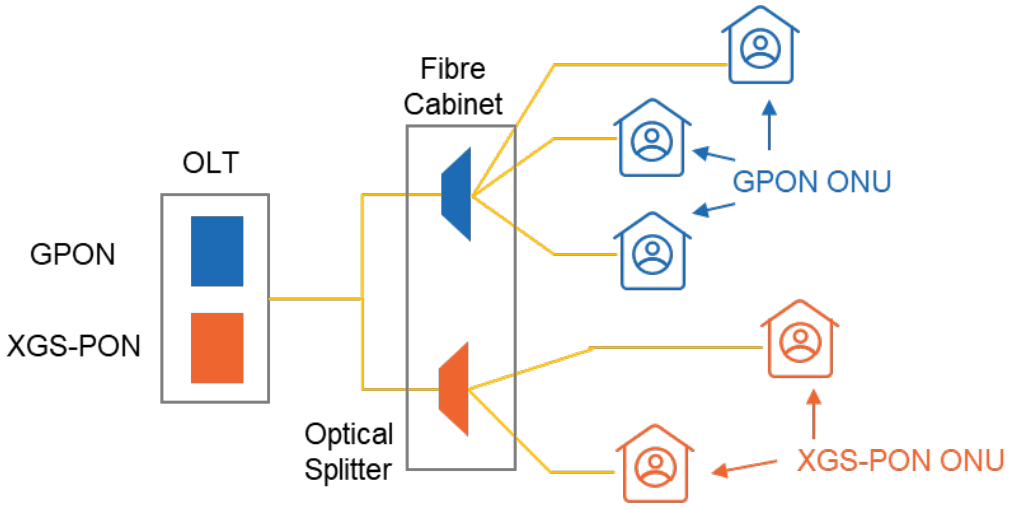


Figure 13 – Passive Optical Network Coexistence

In the future, 50 Gbps PON and subsequent versions will coexist with GPON and XGS-PON, allowing for higher speed offerings.

7. HFC Upgrade Options

HFC networks have long operated in a frequency division duplex (FDD) mode using a sub-split, where the upstream is carried from 5 to 42 MHz with the downstream starting at 54 MHz. Previous HFC network upgrades were focused on adding downstream spectrum, 550 MHz then 750 MHz, 860 MHz and 1 GHz. Using DOCSIS 3.1 cable modems (CMs), gigabit speeds are possible with this network setup, allowing cable operators to offer competitive speed tiers. From here operators have several options to upgrade the HFC network, including remaining with a sub-split plant.

7.1. Sub-split

As mentioned, a sub-split plant uses upstream spectrum to 42 MHz and downstream spectrum from 54 MHz to either 750 MHz, 860 MHz or 1 GHz. Figure 14 shows an example with a top frequency of 750 MHz along with a possible breakdown in spectrum use.

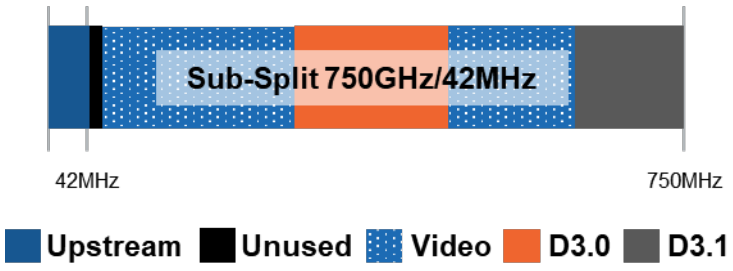


Figure 14 – Sub-split HFC Network

The capacity of a sub-split network depends on how much of the spectrum is used for DOCSIS vs broadcast video, and whether and how much spectrum is being used for DOCSIS 3.1 signals, which use spectrum more efficiently than DOCSIS 3.0 signals. Assuming all spectrum is allocated to DOCSIS, the spectrum available is shown in Table 1.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 – Available Spectrum in Sub-split Plant

Upper Frequency (MHz)	Upstream Spectrum (MHz)	Downstream Spectrum (MHz)
750	37	696
860	37	806
1,002	37	948

Using a capacity estimate of 9 bps/Hz in the downstream and 7 bps/Hz in the upstream as discussed in section 4, we arrive at the capacities shown in Table 2, with related DS:US ratios.

Table 2 – Capacity in Sub-split Plant

Upper Frequency (MHz)	Upstream Capacity (Mbps)	Downstream Spectrum (Mbps)	DS:US Ratio
750	259	6,264	24:1
860	259	7,254	28:1
1,002	259	8,532	33:1

A sub-split plant in combination with DOCSIS 3.1 allows operators to offer services fulfilling the requirements of the vast majority of subscribers. The DS:US capacity ratios in a fully utilized sub-split plant, however, are higher than the peak traffic ratios which, as discussed earlier, have stabilized at approximately 15:1.

7.2. Mid-split

A mid-split plant uses upstream spectrum to 85 MHz and downstream spectrum from 108 MHz to either 1 GHz or 1.2 GHz. Figure 15 shows an example with a top frequency of 1 GHz.

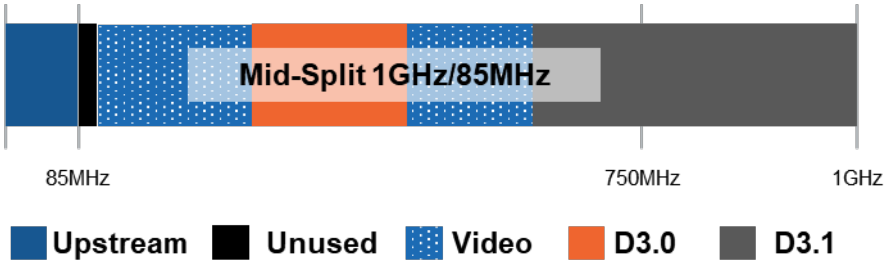


Figure 15 – Mid-split HFC Network

Assuming all spectrum is allocated to DOCSIS, the spectrum available is shown in Table 3.

Table 3 – Available Spectrum in Mid-split Plant

Upper Frequency (MHz)	Upstream Spectrum (MHz)	Downstream Spectrum (MHz)
1,002	80	894
1,218	80	1,110

Again, using a capacity estimate of 9 bps/Hz in the downstream and 7 bps/Hz in the upstream, we arrive at the capacities shown in Table 4, with related DS:US ratios.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 4 – Capacity in Mid-split Plant

Upper Frequency (MHz)	Upstream Capacity (Mbps)	Downstream Capacity (Mbps)	DS:US Ratio
1,002	560	8,046	14:1
1,218	560	9,990	18:1

A mid-split upgrade to 85 MHz in the upstream and to 1 GHz or 1.2 GHz in the downstream allows operators to offer services that meet the requirements of virtually all subscribers and has a DS:US capacity ratio in the same range as DS:US traffic ratios.

7.3. High-split

A high-split plant uses upstream spectrum to 204 MHz and downstream spectrum from 258 MHz to either 1 GHz, 1.2 GHz, or 1.8 GHz. Figure 16 shows an example with a top frequency of 1.2 GHz.

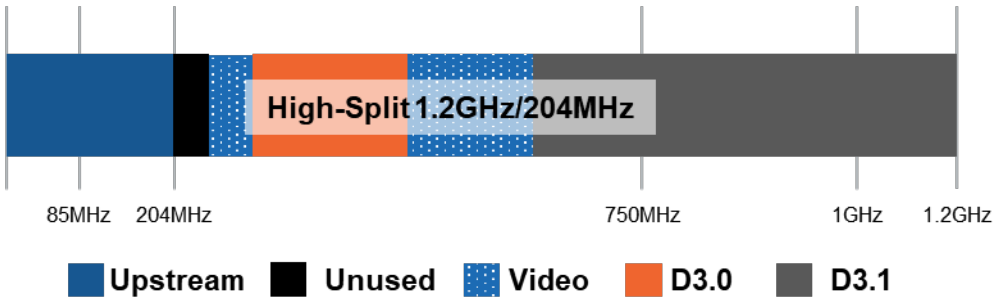


Figure 16 – High-split HFC Network

Assuming all spectrum is allocated to DOCSIS, the spectrum available is shown in Table 5.

Table 5 – Available Spectrum in High-split Plant

Upper Frequency (MHz)	Upstream Spectrum (MHz)	Downstream Spectrum (MHz)
1,002	199	744
1,218	199	960
1,794	199	1536

Again, using a capacity estimate of 9 bps/Hz in the downstream and 7 bps/Hz in the upstream, we arrive at the capacities shown in Table 6, with related DS:US ratios.

Table 6 – Capacity in High-split Plant

Upper Frequency (MHz)	Upstream Capacity (Mbps)	Downstream Capacity (Mbps)	DS:US Ratio
1,002	1,393	6,696	5:1
1,218	1,393	8,640	6:1
1,794	1,393	13,824	10:1

A high-split upgrade allows operators to offer services that meet the requirements of most subscribers and has a lower DS:US ratio than is seen in DS:US traffic ratios. A milestone that a high-split upgrade potentially enables is a 1 Gbps upstream service. This represents greater than 70% of the capacity of the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

upstream which is higher than operators are generally comfortable with, and any interference from over the air (OTA) radio or TV signals may reduce this capacity. This is similar to FTTH operators offering a 1 Gbps upstream service on asymmetric GPON, which has a PHY rate of 1.25 Gbps.

An additional challenge with high-split upgrades is regarding broadcast video platforms using SCTE 55-1 or 55-2 standards, where the downstream telemetry does not operate in the frequencies required with a high-split plant. This requires that all SCTE 55-1 and 55-2 broadcast video platforms be swapped out for DOCSIS set-top gateway (DSG) platforms which use the DOCSIS network for command and control, or IPTV platforms which use the DOCSIS network for command and control as well as to transfer video content.

7.4. Ultra-high-split

An ultra-high-split (UHS) plant uses upstream spectrum to 300 MHz, 396 MHz, 492 MHz, or 684 MHz and downstream spectrum from 372 MHz, 492 MHz, 606 MHz, or 834 MHz to 1.8 GHz. Figure 16 shows an example with upstream to 396 MHz and a top frequency of 1.8 GHz.

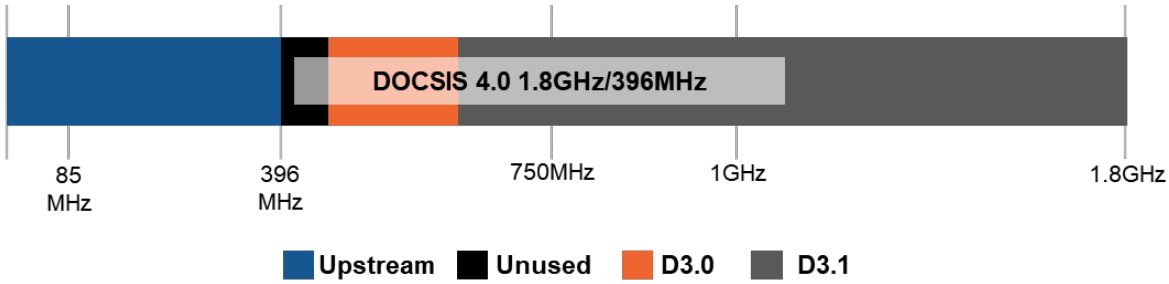


Figure 17 – Ultra-high-split HFC Network

Assuming all spectrum is allocated to DOCSIS, the spectrum available is shown in Table 7.

Table 7 – Available Spectrum in Ultra-high-split Plant

UHS Variant	Upper Frequency (MHz)	Upstream Spectrum (MHz)	Downstream Spectrum (MHz)
UHS-300	1,794	295	1,422
UHS-396	1,794	391	1,302
UHS-492	1,794	487	1,188
UHS-684	1,794	679	960

Again, using a capacity estimate of 9 bps/Hz in the downstream and 7 bps/Hz in the upstream, we arrive at the capacities shown in Table 8, with related DS:US ratios.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 8 – Capacity in Ultra-high-split Plant

UHS Variant	Upstream Capacity (Mbps)	Downstream Capacity (Mbps)	DS:US Ratio
UHS-300	2,058	12,798	6:1
UHS-396	2,737	11,718	4:1
UHS-492	3,409	10,692	3:1
UHS-684	4,753	8,640	2:1

As can be observed, the services as well as the DS:US ratio vary significantly depending on the choice of diplex frequency. Crucially, all UHS variants have a downstream capacity in excess of the estimated 8,600 Mbps that XGS-PON is practically capable of [10].

8. Strategies

8.1. Maintaining a Sub-split Plant

Operators may choose to maintain their sub-split network and perform node splits to control congestion. This strategy may be used as an end-goal if the operator believes that network traffic has saturated, or will soon, and that upstream tiers supported over sub-split are sufficient. It also makes sense if investment needs to be minimized for various reasons, or to buy time to allow the network to be upgraded to FTTH, a capital intensive and slow process.

8.1.1. Capacity Perspective

There are two major considerations when analyzing capacity - the service tiers that can be offered and the ability to manage congestion. If an operator desires to offer symmetric tiers to compete with FTTH it is better to have symmetric capacity, while if an operator desires to manage network congestion most efficiently it is better to have asymmetric capacity, in-line with observed DS:US ratios, discussed earlier. The capacity calculations performed in section 7.1 assume that all spectrum is allocated to DOCSIS and reveal a DS:US ratio from 24 to 33:1 depending on the high frequency. This DS:US ratio is not optimal for either offering symmetric services or managing congestion. In practice, however, all downstream spectrum is not allocated to DOCSIS, with a portion continuing to support broadcast video, meaning that the actual DS:US ratio is lower. The number varies from operator to operator depending on how much spectrum is dedicated to video. In addition, the upstream and downstream capacity will depend on whether DOCSIS signals are DOCSIS 3.0 or DOCSIS 3.1, as the latter are capable of higher orders of modulation and thus are more spectrally efficient. This is in turn dictated by the mix of DOCSIS customer premises equipment (CPE) in the network.

8.1.2. Node Splits

A major benefit of node splits is that resources can be targeted where they are needed. A node split is when one or more logical serving groups (SGs) are added to a pocket of homes connected to an existing serving group, decreasing the number of homes sharing DOCSIS capacity. This can involve adding DOCSIS ports in a hub site, adding additional optics feeding the optical node, or building fibre to a subset of homes, usually down a trunk run, and installing a new optical node. The latter scenario is shown in Figure 18.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

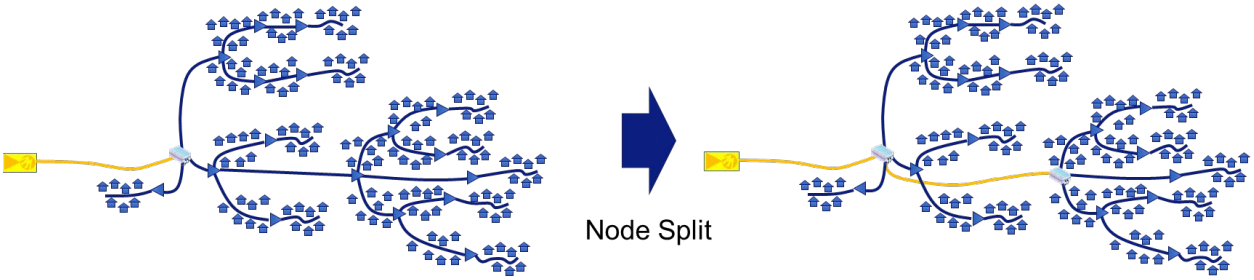


Figure 18 – Node Split Example

The aim of a node split is to precisely divide up the homes from a single serving group into two or more serving groups, theoretically equally dividing up the traffic. In practice, node splits are never perfectly balanced due to physical characteristics of the outside plant, and to the extent that they are balanced in terms of numbers of homes this does not necessarily equate to traffic balance. Figure 19 shows traffic distribution by modem at the network daily peak, averaged over a month, which illustrates that most users' traffic demands are small, even at network peak. The larger the serving group, the more likely its distribution looks like the network aggregate distribution and the more likely a node split will result in two balanced nodes.

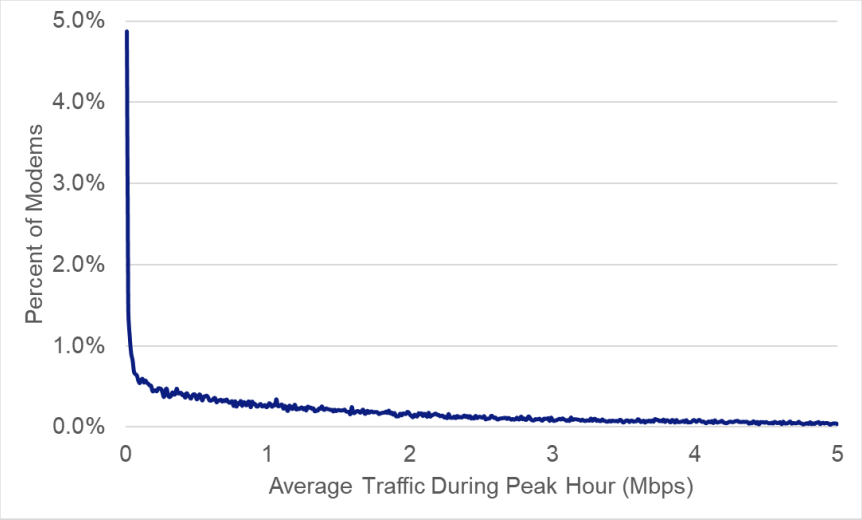


Figure 19 – Peak Hour Traffic Distribution

Due to the statistical nature of traffic use, it is not uncommon to split a congested serving group only to find that one of the new serving groups remains congested. This is more likely to occur as serving group sizes shrink, as a small serving group that is congested is more likely to be dominated by a small number of heavy users. From a capital perspective, node splits also increase in terms of cost per home passed as serving groups shrink, as the costs are primarily labor costs incurred during fibre builds. The result is that node splits have increasing costs and diminishing returns over time as serving groups shrink. At this point, the strategy is no longer cost efficient, and another approach should be explored.

Especially with larger serving group sizes, node splitting is a simple but effective strategy which allows for capacity to be added where it is needed, reducing wasted effort. The eventual outcome of node-splitting is small node sizes in areas where demand is great, and large node sizes where demand is small. In recent years the Cloonan Formula [9] has been used to estimate the effort required to upgrade the network over time. The simple version is shown below:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Serving Group Capacity >= Nsub*Tavg + K*Tmax_max

Where Tavg is the average traffic per subscriber in bits per second during the network busy hour, Nsub is the number of subscribers in the serving group, Tmax_max is the maximum provisioned traffic rate in bps and K is a quality of experience (QoE) factor.

Using the Cloonan Formula with a known serving group capacity, Tavg and Tmax_max, a target number of subscribers per node can be calculated. This however is only useful to the extent that Tavg and Nsub are independent variables, which after years of adding capacity where it is required, has largely ceased to be the case. This may lead operators to the conclusion that they need to split larger nodes to hit a target, even in the absence of congestion, reducing the strategic benefit of targeting capacity where it is needed. Figure 20, based on data from Western Canada, shows the relationship between Tavg and Nsub, which shows that Tavg is smaller for larger Nsub.

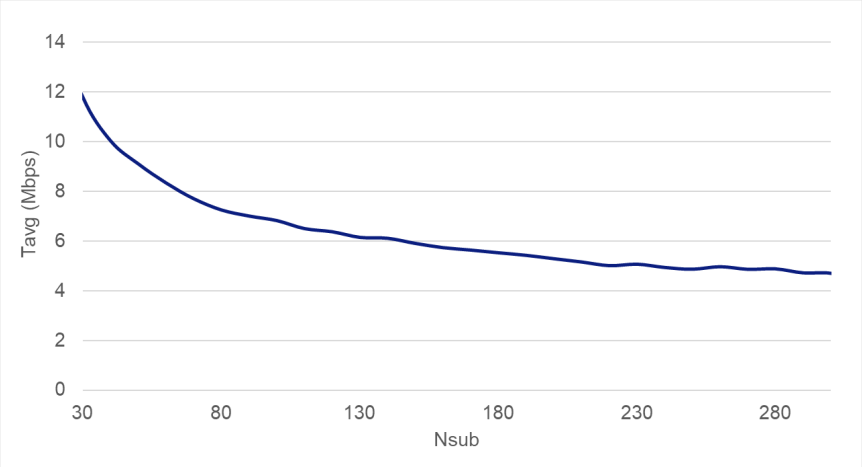


Figure 20 – Tavg vs Nsub

8.1.3. IPTV Transition

Transitioning broadcast video subscribers over to an IPTV platform can free up spectrum that can be allocated to DOCSIS. If the subscriber count is not too high, there will be a net benefit in efficiency by making the transition. The transition requires swapping out subscriber CPE and can be done on a system-by-system or node-by-node basis. By transitioning on a node-by-node basis capacity upgrades can be targeted where they are needed, but the complexity of the upgrade is increased.

8.1.4. DOCSIS CPE Upgrades

Transitioning DOCSIS 3.0 signals to DOCSIS 3.1 will increase capacity and can be done on a node-by-node basis, as with the IPTV transition. Depending on service tiers supported over DOCSIS 3.0 CPE, a small number of DOCSIS 3.0 signals may be left to support low tier services, as well as low-bandwidth services such as phone and broadcast video, allowing legacy phone and video CPE to be kept in the network for longer.

8.2. Mid-split Upgrade

Operators may choose a mid-split upgrade if node-splits cease being an efficient means to control congestion, or if there is desire to increase upstream services. Assuming all spectrum is allocated to

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

DOCSIS, the DS:US ratio of a mid-split network is either 14:1 or 18:1, which matches the observed DS:US traffic ratio well. Although Gbps symmetric services are not supported with a mid-split network, upstream services of 200 Mbps are in market currently, leveraging mid-split networks.

A mid-split upgrade was completed in the Western Canadian cable footprint, allowing for new higher tiers, increased speeds in current tiers, and congestion to be brought to near-zero. Mid-split upgrades have the advantage that the majority of broadcast video hardware continues to function, limiting upgrade costs to the video platform. The outside plant upgrade can be performed as a drop-in upgrade, meaning that re-spacing amplifiers, a costly endeavor, is avoided. Depending on the status of the network, much of the sub-split network hardware can be retained. Costs can be limited if amplifier and node housings can be reused, and analog node optics maintained. Similarly, tap and passive housings can be reused, and only the faceplates changed out.

If an operator has a large installed base of broadcast video CPE that makes use of SCTE 55-1 or SCTE 55-2 signaling, then a mid-split upgrade may be an optimal strategy that can eliminate congestion and allow for increased tiers without a large increase in capital intensity and in a relatively short time frame.

8.3. High-split Upgrade

If operators do not have a large installed base of SCTE 55-1 or SCTE 55-2 based broadcast video, or want to offer Gbps upstream services, then a high-split upgrade makes sense.

There are a few decisions to be made when upgrading to high-split. The first is whether to move to a DAA platform at the same time. While it is possible to continue using analog optics in a high-split network up to 1.2 GHz, the upstream transmitter in the optical node may be a performance bottleneck. Upstream HFC systems are designed to maximize SNR without saturating the upstream transmitter. If signal levels into the upstream transmitter are too high, the laser will saturate or clip, reducing the SNR or causing bit errors. If the signal levels into the transmitter are too low, the SNR will not support higher order modulations. For this reason, it may make sense to move to a DAA architecture which replaces the analog optical link with a digital link. DOCSIS 4.0, and by extension 1.8 GHz plant, assumes the use of DAA optics. Transitioning to DAA can cause challenges, as current analog nodes may be using a CWDM or DWDM system which does not interoperate with digital optics. In this case all nodes using the same multiplexing system might have to be upgraded in a similar timeframe.

The next decision is whether to upgrade the taps and passives. This will depend on whether the current devices are 1 GHz or lower, and whether spectrum over 1 GHz is free to be utilized. If the devices are lower than 1 GHz, then they must be replaced, either to 1 GHz, 1.2 GHz, 1.8 GHz or higher if available and cost effective. If the current devices are already 1 GHz, and spectrum above 1 GHz is used by Multimedia over Coax Alliance (MoCA) in-home video signals, then it is likely not worth upgrading the passives. If current devices are 1 GHz and the spectrum above 1 GHz is free, then it may be worth an upgrade to unlock the additional spectrum.

Another decision is whether to upgrade to 1.8 GHz in the downstream, which is likely to be dictated more by equipment availability than by strategy. Assuming the equipment cost is only incrementally more than 1 or 1.2 GHz, it makes sense to upgrade to the highest bandwidth amplifiers, taps and passives, even if spectrum above 1 GHz will not be used until a later date.

A high-split 1.8 GHz network has the advantage of being able to offer downstream service tiers competitive with, or better than, XGS-PON and a 1 Gbps upstream. As mentioned earlier, however, upgrading to an ultra-high-split variant provides more upstream headroom and is likely to be a better strategy in this case.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

8.4. Ultra-high-split Upgrade

A high-split 1.8 GHz plant offers downstream capacity greater than 10 Gbps and upstream capacity in excess of 1 Gbps. An upgrade to UHS from high-split 1.8 GHz trades off downstream capacity for upstream capacity, and so the decision depends upon what maximum tier an operator wishes to offer and what percent of the total capacity the operator is willing to offer as a service. The capacity tradeoff is shown in Figure 21, using 9 bps/Hz in the downstream and 7 bps/Hz in the upstream.

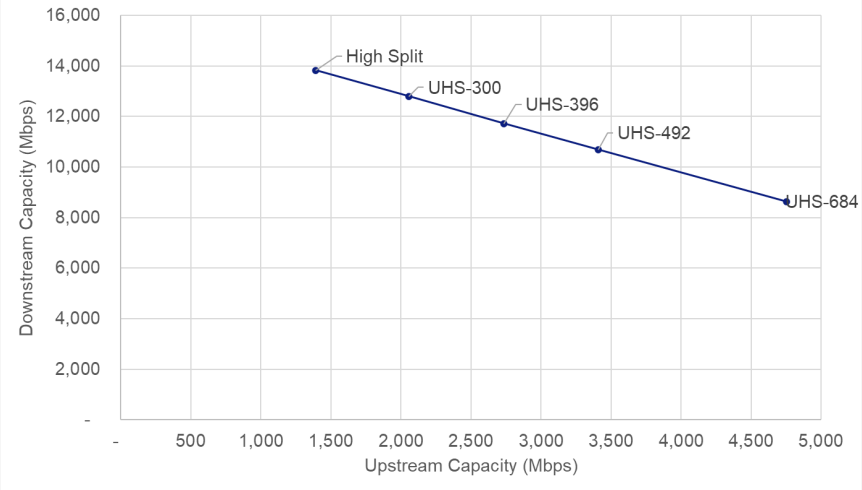


Figure 21 – High-split and UHS Capacities

The two strategies that make sense in terms of service offerings are to either maximize the downstream tier, as this is what subscribers tend to value, or maximize the symmetry of the tier. If an operator wants to offer 10 Gbps downstream and 1 Gbps upstream then high split, UHS-300, UHS-396 and UHS-492 have the capacity to support the offering. Which variant is most supportable and future-proof depends upon the operator's network starting point as well as their forecast for future growth. Here the Cloonan formula, referenced earlier can help as a guide.

$$\text{Serving Group Capacity} >= \text{Nsub}*\text{Tavg} + K*\text{Tmax_max}$$

If a K factor of 1 is used for simplicity, a downstream Tmax_max of 10 Gbps can be used, leaving Nsub and Tavg. If the network is expected to provide sufficient capacity for 10 years, then an Nsub and Tavg must be forecast for that time. Using an example of 100 for Nsub and 20 Mbps for Tavg the serving group capacity must be greater than or equal to 12 Gbps as calculated below.

$$\text{Serving Group Capacity} >= 100*20 + 1*10{,}000 = 12{,}000 \text{ Mbps}$$

Of the initial group of four variants, only high-split and UHS-300 meet this requirement. Doing the same exercise with the upstream, using a K factor of 1, an upstream Tmax_max of 1 Gbps and example values of 100 for Nsub and 2 Mbps for Tavg, the serving group capacity must be greater than or equal to 1,200 Mbps.

$$\text{Serving Group Capacity} >= 100*2 + 1*1{,}000 = 1{,}200 \text{ Mbps}$$

Both high-split and UHS-300 meet this requirement, assuming that no capacity is lost due to OTA interferers or similar.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

XGS-PON operators currently have 8 Gbps symmetric tiers in market, so if an operator wants to match the tier in the downstream while keeping the tier as symmetric as possible then UHS-684 or UHS-492 are likely to be most desirable. Using the same downstream numbers as above but substituting in 8 Gbps for the tier we arrive at a serving group capacity greater than or equal to 10 Gbps.

Serving Group Capacity >= 100*20 + 1*8,000 = 10,000 Mbps

In this case UHS-684 does not have sufficient capacity to meet this requirement, but UHS-492 does. To determine the maximum upstream, we can reorder the Cloonan formula to calculate the upstream Tmax_max as shown below.

Tmax_max <= (Serving Group Capacity – Nsub*Tavg)/K

Using the UHS-492 serving group capacity of 3,409 Mbps we arrive at a Tmax_max of 3,209 Mbps, making the most likely highest upstream service tier 3 Gbps.

Tmax_max <= (3,409 – 100*2)/1 = 3,209 Mbps

UHS-396 is also a potential middle ground, where an operator can offer multiple Gbps upstream and a higher downstream than what XGS-PON can offer, such as 10/2 Gbps.

8.5. FTTH Upgrade

An operator may also decide to upgrade their HFC network straight to FTTH, and indeed some operators have announced that intention publicly. This strategy makes sense if the operator believes an FTTH upgrade is not significantly more capital intensive than an HFC upgrade, or that network growth is great enough and sustained over time such that an HFC network upgrade will not delay an upgrade to FTTH long enough to justify the investment.

Making some assumptions, we can calculate under what traffic growth rates and FTTH upgrade cost conditions the total cost of ownership (TCO) of an HFC upgrade is more desirable than an FTTH upgrade and vice-versa. The analysis assumes a serving group with 200 households with 50% penetration, an UHS-396 HFC upgrade with a cost of $100-$300 (based on publicly released estimates), Tavg of 5 Mbps and Tmax_max of 10 Gbps.

FTTH Upgrade Cost per Household Passed

Traffic Growth Rate	$ 300	$ 400	$ 500	$ 600	$ 700	$ 800	$ 900	$ 1,000
5%	HFC	HFC	HFC	HFC	HFC	HFC	HFC	HFC
10%	HFC	HFC	HFC	HFC	HFC	HFC	HFC	HFC
15%	FTTH	HFC	HFC	HFC	HFC	HFC	HFC	HFC
23%	FTTH	FTTH	HFC	HFC	HFC	HFC	HFC	HFC
32%	FTTH	FTTH	FTTH	HFC	HFC	HFC	HFC	HFC
50%	FTTH	FTTH	FTTH	FTTH	FTTH	FTTH	HFC	HFC

Figure 22 shows the results, where purple cells represent growth and cost conditions where an HFC upgrade provides a lower and thus more desirable TCO, and orange cells represent conditions where an FTTH upgrade has a lower TCO. In this example we can see that the combination of low FTTH upgrade cost and high growth favours an FTTH upgrade, while high FTTH upgrade cost and low growth favours an HFC upgrade.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

FTTH Upgrade Cost per Household Passed

Traffic Growth Rate	$ 300	$ 400	$ 500	$ 600	$ 700	$ 800	$ 900	$ 1,000
5%	HFC	HFC	HFC	HFC	HFC	HFC	HFC	HFC
10%	HFC	HFC	HFC	HFC	HFC	HFC	HFC	HFC
15%	FTTH	HFC	HFC	HFC	HFC	HFC	HFC	HFC
23%	FTTH	FTTH	HFC	HFC	HFC	HFC	HFC	HFC
32%	FTTH	FTTH	FTTH	HFC	HFC	HFC	HFC	HFC
50%	FTTH	FTTH	FTTH	FTTH	FTTH	FTTH	HFC	HFC

Figure 22 – TCO Comparison Between FTTH and HFC Upgrades

There are many factors that determine the cost of an FTTH upgrade, but the most influential are whether the plant type is aerial or underground, ownership and regulatory rules around access to infrastructure, and labour costs having to do with installing fibre. While operator forecasts for traffic growth will vary, the trend is for slowing growth, as discussed earlier.

With generally high estimated costs for FTTH upgrades and low forecasted traffic growth, most operators would find themselves on the top right of the chart, making an HFC upgrade the optimal decision from a financial perspective. Within operator networks there are likely to be areas which can be upgraded at a lower cost, or traffic growth is higher than average due to demographic shifts or similar, and operators may decide to upgrade those areas directly to FTTH.

While from a financial perspective an HFC upgrade may represent the option with the best TCO, there are other reasons to choose an upgrade path. Operators may perceive FTTH as the only upgrade path that achieves technology leadership instead of asking whether subscriber needs can be met with an HFC network, and whether they will continue to be met over the expected lifespan of the network. Subscribers do not inherently ask for a technology solution; they ask for services. To the extent that they do ask for a specific technology, it tends to be driven by marketing activities.

There also exists confusion over what technology is most appropriate in new builds and what is capable of meeting the needs of subscribers. Thanks to the transition to all-IP services and improvements in CPE and CPE install practices, FTTH is the most future-proof technology, and at build time it is the most appropriate technology, assuming reasonable density of homes. This is why most HFC operators are building FTTH in new areas. This, in addition to the fact that many HFC operators are upgrading to FTTH add to the misleading messaging that FTTH is required to offer services that subscribers need, including traffic symmetry. Members of the US Senate recently urged the Federal Communications Commission (FCC) to change their broadband definition from 25 Mbps downstream by 3 Mbps upstream to 100 Mbps symmetric [11], a definition that would needlessly exclude HFC sub-split plants. Although an argument for 100 Mbps downstream can be made, it is difficult to envision what residential use cases require high-bandwidth symmetric services.

9. Conclusion

In order to offer higher-speed service tiers, cable operators will be obliged to upgrade their HFC networks. There are many upgrade paths and strategies to choose from, and each operator must decide what elements to prioritize. Strategies range from maintaining a sub-split network to minimize capital intensity to upgrading straight to FTTH. In between are strategies to upgrade to mid-split, high-split or

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

ultra-high-split using DOCSIS 4.0 technology. The choice of diplex frequencies has implications on what tiers can be offered and how closely network capacity matches the needs of subscribers.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

ADSL	Asymmetric DSL
bps	bits per second
CM	Cable Modem
CO	Central Office
COVID-19	Corona Virus Disease of 2019
CPE	Customer Premises Equipment
CWDM	Coarse WDM
DAA	Digital Access Architecture
DOCSIS	Data Over Cable Service Interface Specifications
DS	downstream
DSG	DOCSIS Set-top Gateway
DSL	Digital Subscriber Line
DSLAM	DSL Access Multiplexer
DWDM	Dense WDM
FCC	Federal Communications Commission
FDD	Frequency Division Duplex
FTTC	Fibre-to-the-curb
FTTH	Fibre-to-the-home
Gbps	Gigabit-per-second
GHz	Gigahertz
GPON	Gigabit PON
HD	High Definition
HFC	Hybrid Fibre-Coax
Hz	Hertz
IP	Internet Protocol
IPTV	IP Television
K	QoE Factor
Mbps	Megabit-per-second
MHz	Megahertz
MoCA	Multimedia over Coax Alliance
MPEG	Motion Pictures Expert Group
NHL	National Hockey League
Nsub	Number of Subscribers
OLT	Optical Line Terminal
ONU	Optical Network Unit
OTA	Over the Air
OTT	Over-the-top
PHY	Physical Layer
PON	Passive Optical Network
QAM	Quadrature Analog Modulation
QoE	Quality of Experience
SCTE	Society of Cable Telecommunications Engineers
SG	Serving Group
SNR	Signal-to-noise Ratio
Tavg	Average Traffic at Busy Hour
TCO	Total Cost of Ownership

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Tmax_max	Maximum Provisioned Traffic Rate
UHS	Ultra-high-split
US	upstream
VDSL	Very-high-speed DSL
WDM	Wavelength Division Multiplexing
XGS-PON	10 Gbps Symmetric PON
YoY	Year-over-year

Bibliography & References

[1] OECD Data - https://data.oecd.org/broadband/fixed-broadband-subscriptions.htm

[2] https://en.avm.de/guide/avm-explains-connection-types-dsl/

[3] https://seekingalpha.com/article/1922581-calix-the-local-telcos-favorite

[4] https://www.forbes.com/sites/kevinmurnane/2017/10/26/upgrading-to-4k-hdr-tv-how-far-you-sit-from-the-screen-is-critical/?sh=693155541801

[5] https://www.thebroadcastbridge.com/content/entry/2746/hevc-decoded

[6] *Maximizing Returns on the Path to DOCSIS 4.0*, Mike Darling; SCTE Cable-Tec 2021, SCTE

[7] https://www.bce.ca/investors/AR-2022/2022-bce-integrated-annual-report.pdf

[8] https://www.telus.com/en/about/investor-relations/reports/annual-reports/2022?intcmp=tcom_about_annual-reports_link_2022-annual-report

[9] *Predictions on the Evolution of Access Networks to the Year 2030 & Beyond,* Cloonan, Tom, Emmendorfer, Mike, Ulm, John, Al-Banna, Ayham and Chari, Santhana; SCTE Cable Tec 2014, SCTE

[10] *Network Capacity Options on the Path to 10G*, Karthik Sundaresan; SCTE Cable-Tec 2022, SCTE

[11] https://www.pcmag.com/news/senators-broadband-speed-minimum-should-be-100mbps-for-downloads-and-uploads

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Focusing (Stored) Energy Where It Matters Most

Optimizing the Impact of Energy Storage in the Access Network

A Technical Paper prepared for SCTE by

Toby Peck
Senior Director, Product Management
EnerSys

Creighton Brown
Engineering Manager, Energy Storage
EnerSys

Mark Vodopest
Broadband &Renewables Market Manager, Energy Storage
EnerSys

Greg Laughlin
Strategic Marketing Manager
EnerSys

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Evaluation Process 4
2.1. Load Profile 4
2.1.1. General Considerations 4
2.1.2. Cable Broadband Applications 4
2.2. Backup Runtime 5
2.2.1. General Considerations 5
2.2.2. Cable Broadband Applications 5
2.3. Cycling Frequency 6
2.3.1. General Considerations 6
2.3.2. Cable Broadband Applications 6
2.4. Site Limitations 6
2.4.1. General Considerations 6
2.4.2. Cable Broadband Applications 7
2.5. Environmental Resiliance 9
2.5.1. General Considerations 9
2.5.2. Cable Broadband Applications 9
2.6. Safety 9
2.6.1. General Considerations 9
2.6.2. Cable Broadband Considerations 9
2.7. Recyclability 11
3. Energy Storage System Technologies 11
3.1. Lead Acid 12
3.1.1. Description 12
3.1.2. Benefits 12
3.1.3. Limitations 13
3.2. Lithium-ion 13
3.2.1. Description 13
3.2.2. Benefits 13
3.2.3. Limitations 14
3.3. Hybrid Supercapacitor 15
3.3.1. Description 15
3.3.2. Benefits 15
3.3.3. Limitations 16
3.4. Nickel-Zinc 16
3.4.1. Description 16
3.4.2. Benefits 16
3.4.3. Limitations 17
3.5. Sodium-Ion 17
3.5.1. Description 17
3.5.2. Benefits 17
3.5.3. Limitations 18
3.6. Comparing Technologies 18
4. Determing the Best ESS Technology for any Given Site 18
4.1. Aligning Application Needs to Technology 19
4.2. Valuation of key considerations and costs 19
4.2.1. Key considerations: Costs 20
4.2.2. Key considerations: Benefits 20
5. Conclusions 22

Abbreviations 23

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References 23

List of Figures

Title **Page Number**

Figure 1 - Estimated distribution of North American power supply loads 5
Figure 2 - Example cable broadband powering sites. The variability in available space, loading, and local regulations makes a one-size-fits-all solution for ESS solutions nearly impossible. 8
Figure 3 – Hybrid supercapacitor 15
Figure 4 – Volumetric energy density of common energy storage chemistries 17
Figure 5 – Example Site Evaluation 19

List of Tables

Title **Page Number**

Table 1 – NFPA 855 Application 10
Table 2 – Chemistry Comparison 11
Table 3 – Battery chemistry analysis (5=Best) 18

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

We have become increasingly reliant on the hybrid fiber-coax (HFC) network for high-speed data, as much of the workforce has transitioned to a hybrid work environment between office and home. The line between business services and residential customers has been blurred – and the need for this ultra-reliable data source has become universal. We have charted a course to meet future bandwidth needs with the 10G initiative, and new mobility offerings are leveraging the HFC network to ensure connectivity while on the go. With all of this, it has never been more important to have reliable power to keep the network up and running.

Energy storage is fundamental to reliable powering; without it, every blip on the grid can cause a business transaction to fail or an important Zoom call to drop. There is an estimated 3.2 GWh of energy storage in the access network today that guards against outages for millions of customers. So how can we use this available energy most effectively? Is it being deployed and maintained efficiently to minimize operational costs? Can it be leveraged to lower utility impact? Is there intelligence that can be embedded to increase operational efficiency and plant reliability?

This paper will explore:

- Various energy storage technologies being reviewed in for use in the outside plant (OSP)
- Key considerations for deployment of OSP energy storage solutions
- A methodology for determining the best energy storage technology for any plant scenario

2. Evaluation Process

Choosing the best energy storage system (ESS) for a specific application can be challenging, but if you assess the right information during the design phase of a project, this task can be much less daunting. There are many key system requirements that need to be considered as part of the ESS such as required backup time, load profile, cycling frequency, environmental factors, cost, site limitations, local code requirements and safety. Making the final decision should be based on all these factors and not just one or two key characteristics. Purchasing the best value solution will ensure that the ESS will perform safely and optimize the operational benefits of the system. Below is a description of some of these key factors for choosing the right ESS and the role that they play in determining which solution is best for any unique application.

2.1. Load

2.1.1. General Considerations

The load is the amount of power that will be required from the ESS for a specified period. A load profile is given either as a constant current (constant amperage), constant power (constant watts), or a multi-step load, where loads turn on and off throughout the backup period. Understanding the unique charge/discharge characteristics of each battery will help you to make the best ESS selection.

2.1.2. Cable Broadband Applications

The traditional application for outdoor cable broadband backup power is for HFC nodes and amplifiers, which through distributed access architecture (DAA) are migrating to remote-phy devices (RPDs) and passive optical network (PON) optical line terminals (OLTs). The traditional load profile is a constant power load that only requires battery support during infrequent utility power outages. Figure **1** below

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

illustrates typical loads at an HFC power supply; while most HFC power supplies support 15A to 18A of power, many operators tend to load the supplies close to 50% capacity to allow for future additions.

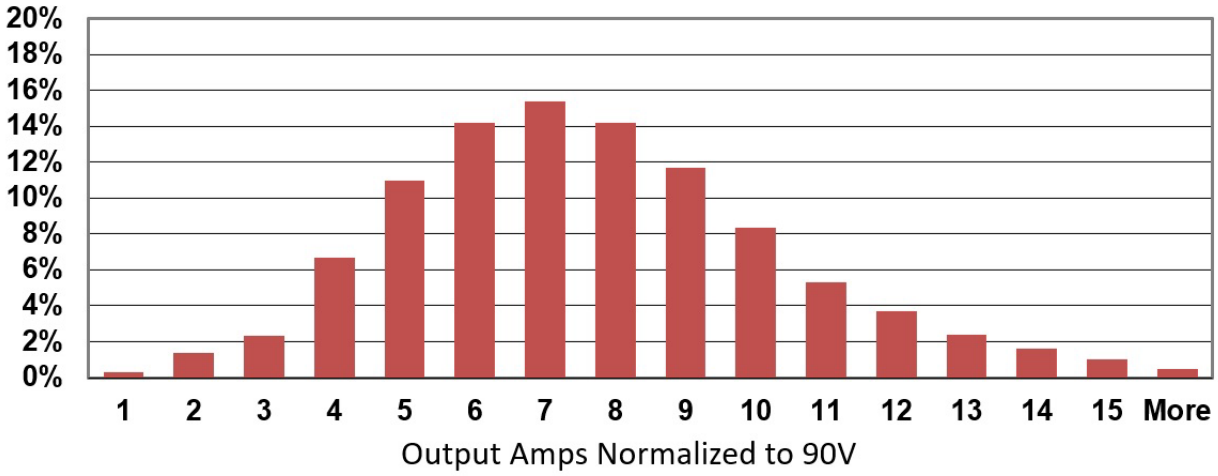


Figure 1 - Estimated distribution of North American power supply loads

2.2. Backup Runtime

2.2.1. General Considerations

Required backup time is an important attribute of the load profile and needs to be fully understood upfront. Batteries are often engineered with specific charge and discharge characteristics in mind. Certain chemistries perform better at different discharge rates and even within the same chemistry batteries can be designed for high instant power for short durations or high energy (capacity) low power for longer durations. For the purposes of this paper, the term battery can be defined by the Oxfor dictionary definition - a container consisting of one or more cells, in which chemical energy is converted into electricity and used as a source of power.

2.2.2. Cable Broadband Applications

Entire white papers can be written on strategies for determining required runtime at any unique OSP powering location, but to provide context for understanding the best ESS solution it is important to have a high-level understanding of what might drive runtime requirements.

The first question to be answered is how much downtime can you tolerate? The answer to this question may be dictated by competition, service agreements or mandates. Our societal reliance on network connectivity for work, life management and entertainment has made the cable network essential to the point where any downtime is unacceptable. Competitive pressure from other wireline, wireless and satellite options place power reliability as a differentiator for customer retention.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Government mandates for emergency services (911) – or in the case of California Public Utility Commission's (CPUC) 72-hour mandate for wildfire mitigation[i] – may also affect your backup power requirements.

The next consideration for estimating required backup time is understanding how long it will take to dispatch technicians to augment power with curbside generators until utility service is restored. It is important to factor not only the distance to reach a site, but also the quantity of sites and ability to serve during widespread outages in a given service area.

2.3. Cycling Frequency

2.3.1. General Considerations

Once a clear understanding of the load profile has been established, it needs to be determined how often this load will be applied to the ESS. Depending on the application, an ESS may be utilized daily, or in the case of standby backup power, may only be used a few times per year. It may not make financial sense to pay significantly more for a technology that offers thousands of cycles if those cycles will never be realized. In contrast, if the system is going to be cycled on a regular basis it may make sense to pay more up front for a battery system that cycles better to reduce the number of required battery changes over the life of the system.

2.3.2. Cable Broadband Applications

Decision factors with respect to cycling requirements have been traditionally placed around the stability of the utility grid serving each site. In areas where the utility grid is stable and reliable, the batteries spend most of their time in float mode waiting to be used, in which case battery cycling is not a concern. Alternately, areas prone to frequent (daily) outages may benefit from a storage solution designed for high cycles.

Energy mitigation may have some future application as well as any potential renewable applications, where battery power is intentionally invoked. A high-cycle energy storage solution would be more applicable for these energy strategies.

2.4. Site Limitations

2.4.1. General Considerations

As the two drivers connected to runtime, load demands and required backup times, continue to increase, designers are often challenged with meeting these requirements within a fixed space available for both the batteries and corresponding electronics. Some installations may have weight limitations, such as pole mounted systems or systems that are not installed on the base floor of a structure.

As an example, the California Public Utilities Commission [CPUC] increased the required backup time of critical communications infrastructure from 24 hours to 72 hours in fire prone areas. In many cases this extended runtime had to be met without expanding the existing available space for energy storage. In these scenarios, volumetric and gravimetric energy density of the ESS becomes a critical attribute that needs to be considered closely and may outweigh other attributes.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.4.2. Cable Broadband Applications

When considering the best energy storage for any OSP ESS, a key detail to understand is the physical space limitations of the specific site being built.

- Currently most sites have been designed specifically around the form-factor of three or six batteries in case size 27 (306 x 173 x 225mm) or case size 31 (330 x 173 x 240mm)
- Sites are often in easements in front of residences or on utility poles, so there is usually little or no space for additional battery cabinets to extend run times
- Many areas have height restrictions for cabinets making it impractical to add battery extensions to existing locations
- In locations where space does exist for larger installations, permitting can still be restrictive, and requirements vary between national, regional and local Authorities Having Jurisdiction (AHJs).
- When upgrading existing sites with established utility service connections, it is often a requirement to maintain that connection to reduce costs and minimize logistical issues with local electrical utilities.

2.4.2.1. Cabinet Space limitations (Typical)

There are numerous exceptions to these and thus there is no one-size-fits-all ESS solution for every site.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 2 - Example cable broadband powering sites. The variability in available space, loading, and local regulations makes a one-size-fits-all solution for ESS solutions nearly impossible.

2.4.2.2. Other Physical Restrictions

The access network has often been weaved in amidst its surroundings to stand out as little as possible. Network power systems in a particular location often existed prior to suburban development that now restricts their expansion or adaptation. Ground space around existing sites often has limited clearance to expand, especially in scenarios like the CPUC's 72-hour mandated backup time, which required immense increases in runtime with limited room for expansion. Pole attachments can be heavily regulated by the utilities that own them, and many local regulations exist around the space that an enclosure can occupy on a pole as well as the total weight of the system attached to a pole. It is important to understand these requirements as they can factor heavily in the process of deciding an energy storage solution for any given site.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.5. Environmental Resiliance

2.5.1. General Considerations

Outdoor ESS can be exposed to a wide array of environmental challenges including excessive heat, cold, heavy rains, wind, humidity, earthquake and more. Each battery chemistry handles these environmental concerns differently.

2.5.2. Cable Broadband Applications

The OSP broadband network is generally powered and backed up by energy storage systems in enclosures that have only passive thermal management. This means that in many geographies these systems will be subjected to a wide range of temperatures ranging from -40°C to + 60°C. Additionally, the majority of OSP enclosures are NEMA 3R rated, which provides a degree of protection against falling objects and precipitation but allows open airflow through the cabinet as a measure of passive cooling. This means that these systems are often subjected to moisture from humidity. Finally, as most of these systems are close to roads, they can be impacted by roadside vibration caused by larger vehicles.

2.6. Safety

2.6.1. General Considerations

Safety is perhaps the most important consideration when discussing energy storage technology. The harsh nature of outdoor installations, especially within proximity to the public, creates inherent risk. Any energy storage technology should be proven and certified safe to the highest possible standards before being deployed.

2.6.2. Cable Broadband Considerations

2.6.2.1. Standards and Codes

A key factor in deploying the best ESS for any application is the understanding of relevant safety standards. There are two main organizations who produce recommended fire codes, International Fire Code (IFC) and National Fire Protection Association (NFPA). Each of these organizations release recommended codes related to fire safety which are often adopted as requirements, either in part or in full, by fire marshals and AHJ's in their local building codes. Contained within chapters related to energy storage systems, IFC and NFPA855 documents often require adherence and certification to UL Solutions (UL – formerly Underwriters Laboratory) test standards and certifications to dictate fundamental system safety.

While limited in impact to most current broadband OSP applications, NFPA 855 2023 edition also dictates ground clearances and site protection required, which can have a drastic impact on the area required for deployment. Required compliance to NFPA 855 is defined by the ESS technology and the aggregate energy capacity. Table **1** below from NFPA defines where ESS must comply to this standard.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 – NFPA 855 Application[ii]

NFPA 855 – Application

ESS TECHNOLOGY	Aggregate CAPACITY[a]
BATTERY ESS	
Lead acid	70 KWh
Nickel cadmium	70 KWh
Lithium-Ion	20 KWh
Sodium	20 KWh
Flow batteries	20 KWh
Other battery technologies	10 KWh
Batteries in residential occupancies	1 KWh
CAPACITOR ESS	
Capacitors, all types	3 KWh
OTHER ESS	
All other ESS	70 KWh

2.6.2.2. UL Standards for Energy Storage Systems

In Energy storage markets, UL provides standards that are either pass/fail and carry a certification or listing or are merely to gather data to make educated decisions on installation parameters. The following sections summarize standards that most often impact energy storage.

2.6.2.2.1. UL1973

UL1973 is a series of tests that subject batteries (specifically at battery level) to abusive test situations (such as overcharge, drop test, crush test etc.) to ensure that the battery design is robust enough to safely handle challenging environmental and operational hazards. This standard was originally written with lithium chemistries in mind, but Annex H of the standard – with tests more applicable to lead acid batteries – has recently been released.[iii]

2.6.2.2.2. UL9540

UL9540 is a system level test that includes a UL1973 listed battery, along with other UL listed electronics and controls (inverters, breakers etc.). This is a safety standard to ensure that all system components interact correctly with each other to maximize safe operation of the system and is not a test of any individual component. It should be noted that per the scope of UL9540 Section 1.5: "systems using lead acid or Ni-Cad batteries that fall within the scope of UL 1778 and only serve an uninterruptible power system (UPS) application are outside the scope of this standard."[iv]

2.6.2.2.3. UL9540A

UL9540A is a "Test Method for evaluating thermal runaway and fire propagation in battery energy storage systems". The goal of this set of tests is to determine if a battery is capable of entering thermal runaway, and how it responds at a cell, module, unit, and installation-level once in thermal runaway.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Temperature, deflagration, propagation, and quantities of released gases are measured and recorded at each phase of the test. This data is then used to determine spacing requirements, fire protection and suppression requirements as well as other aspects of the installation.[v]

2.7. Recyclability

As our industry leads the way toward a more sustainable future, it is paramount to consider the impact of end-of-life disposal of any energy storage technology, as well as the assumed life cycle. While many technologies can be recycled at some level, market value of the recyclable material will determine whether recycling becomes an added cost.

3. Energy Storage System Technologies

Although there is continual research and development in potential ESS technologies and battery chemistries, we will only highlight technologies that are currently available or expected within the next 24 months, and applicable for use in the cable broadband market. This includes lead acid, lithium, hybrid supercapacitor, nickel-zinc and sodium-ion. Each battery chemistry has its own benefits which make it ideal for various applications. In Table 2 below, is a comparison of important characteristics to assess as you are considering available ESS options.

Table 2 – Chemistry Comparison[vi][vii]

	Lead Acid	Bi-Polar	LFP	NMC	NCA	Nickel Zinc	Sodium Ion	Hyb SuperCap
								
Full Chemistry Name	Lead Acid	Bi-Polar Lead Acid	Lithium Iron Phosphate	Nickel Manganese Cobalt	Nickel Cobalt Aluminum Oxide	Nickel Zinc	Sodium Ion	Lithium SuperCap
Applications	ESS/Telecom/ Datacenter UPS / Transportation/ Engine start	Under development	ESS/EV/Consumer Storage	EV, Ebike, Medical Devices, Industrial	EV, Consumer Electronics	UPS-High power discharge	EV, ESS	Telecom/Broadband, 36V, 48V
Nominal Voltage	2	2	3.2	3.6	3.6	1.7	3	3.6
Operational Voltage	1.67 to 2.27	1.67 to 2.27	2.5 to 3.65	3.0 to 4.2	3.0 to 4.2	1.2 -1.9	1.5 to 4.3	3.0 To 4.2
Cycle Life	200 to 2050 (50% DOD Carbon)	500+	3000-5000	1500-3000	500	500	3000	10000
Design Life	5-7 Years	5-7 Years	10 to 15 years	10 to 15 years	10 to 15 years	10 to 15 years	10 to 15 year Estimate	20+ years
Charge Current	.1C to Unlimited	.1C to Unlimited	1C	.7 to 1C	.7C	.45C		.5C
Discharge current	4C	4C	1C	1 - 2C	1C	.45C		1C
Energy Wh/KG	35 - 50	50-60	90 - 120	150 to 220	200 -260	62.5	160	126
Energy Wh/L	20 - 50	50-60	250	338-545	569	280	~200	~150
Main Benfits	Low cost, Low risk Thermal runaway	Increased power density, 15% lighter than trad. Lead Acid, lower cost	High Cycle life, Perceived safer	High Energy Density	High Energy Density	No Thermal Runaway	No Thermal runaway, Cheap BOM, Sodium cheap and plentiful	Very High Cycle life, No Thermal runaway
Challenges	High Weight, lower cycles @ high DOD	Under development	Lower Energy Density	Lower cycle life	Lower cycle life	Limited Application, limited charge/Discharge profile	New Tech, Manufacturing Challenges	High cost
Thermal Runaway Temp	NA	NA	270C	210C	150C	NA	NA	NA
Form Factor	Primarily prismatic	Prismatic	18650, 32650, Prismatic	18650, Prismatic	18650, Prismatic	Cylindrical	Prismatic, Cylindrical	Pouch
Cost/KWH	$75	~$50	$100	$150	$150	$150	$137	$500
Operat. Temp	-40C to 65C	-40C to 65C	-20 to 60 C (0C min for charging)	-20 to 55 C (0C min for charging)	-20 to 60 C (0C min for charging)	-20 to 50 c	-20 to 60c	-20 to 60 C (0C min for charging)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.1. Lead Acid

3.1.1. Description

Lead acid battery technology was first developed in 1859 and has been a commercially available battery for over 150 years. Although there are many different designs of lead acid batteries, the primary characteristics of positive and negative plates sandwiched around a separator material, which are stacked in parallel to provide a 2-volt cell are common amongst the different designs. These plates are either submerged in electrolyte or in the case of absorbent glass mat (AGM) design, the electrolyte is held close to the plates in a super absorbent fiber glass mat that also acts as the separator. Gel batteries use a silica-based gel to suspend the electrolyte. The most common forms of lead acid batteries include Vented Lead Acid (VLA or flooded lead acid), Valve Regulated Lead Acid (VRLA), Thin Plate Pure Lead (TPPL), Absorbent Glass Mat (AGM), and gel. Although many lead acid batteries today are made of 6 cells in series which provide a 12 V nominal rating, they can also be found in 2 V single cells, 4 V, 6 V, and 8v configurations.

3.1.2. Benefits

Advances in carbon infused AGM and pure lead technologies have improved the potential cycle life of lead acid batteries to reach as high as 2050 cycles at 50% depth of discharge in products that are already commercially available. These batteries also perform much better at a partial state of charge and thus are well suited for renewable or unstable grid scenarios where there is no guarantee that the batteries will be fully charged immediately after a discharge. The added carbon optimizes sulphate crystal size allowing the sulphate crystals to convert back into the electrolyte during the charging process, thus extending the life and kWh throughput of the battery.

Lead acid batteries in the USA are 99% recycled with over 95% of the components being recyclable. [viii] This has a staggering effect on the overall environmental impact of producing new batteries as a vast circular economy has been created. In addition, lead is a domestically available resource which is often mined as a byproduct of other mineral mining such as silver and zinc. Also, lead acid batteries do not require any rare minerals that may bring additional environmental impact from mining.

If properly maintained and protected lead acid batteries can perform well in a wide variety of environmental conditions. Although lead acid batteries lose some capacity at lower temperatures and have a shortened life at high temperatures, they perform safely at temperatures as low as -40C and as high as 65C. The ideal operational temperature, which is where most batteries are rated at, is between 20C and 25C.

With recent battery incidents in the news, regulators are revisiting the safety of ESS installations. Because lead acid batteries use an aqueous non-flammable electrolyte, they cannot enter a thermal runaway event in the same way that other technologies may. Under normal operation lead acid batteries function reliably and safely.

As batteries approach the end of their design life or if they sustain damage due to abuse, the internal resistance of the battery can increase, which in turn raises the required float current. As the float current increases, more heat is generated and much of the excess energy breaks the electrolyte into hydrogen and oxygen. The increase of hydrogen and oxygen exceeds the battery's ability to recombine the gases, which build in pressure until they are vented. This process is known as thermal walk away and in contrast to a lithium-ion battery in thermal runaway, this process is not self-sustaining and can be stopped at any time by disconnecting the incoming current to the battery. Proper ventilation and hydrogen alarms should be

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

considered when the project is being planned to ensure the system. Some battery manufacturers add a catalyst to lead acid batteries to aid in the hydrogen/oxygen recombination process. This helps to reduce electrolyte loss, vented gases, and prolongs the life of the battery.

3.1.3. Limitations

As a lead acid battery is discharged, the acid is absorbed from the electrolyte and sulphation of the lead plates begins to occur. The deeper the discharge, the more acid is absorbed allowing for increased sulphation throughout the plates. This sulphation process slowly degrades the capacity of the battery over each cycle. Due to the effects of sulphation in deep cycling applications over time, lead acid batteries perform best in constant float applications with in-frequent discharges. This makes lead acid batteries a great fit for standby reserve applications such as telecom, datacenter, broadband and switch gear where systems spend the majority of their working life on float.

3.2. Lithium-ion

3.2.1. Description

Although a much newer technology than lead acid, over the past 30 years lithium-ion batteries have taken over many markets due to their high volumetric and gravimetric energy density as well as excellent performance in cycling applications. This is especially true of portable and mobility devices such as cell phones, laptops, cordless hand tools, electric vehicles, and in more recent years ESS. Lithium-ion refers to a battery where lithium-ions are exchanged across a micro permeable separator. Many chemistry variations can be found within the category of lithium-ion batteries such as Lithium Iron Phosphate (LFP), Nickel Manganese Cobalt (NMC), Nickel Cobalt Aluminum Oxide (NCA). Each of these chemistries brings its own set of positive and negative attributes to the market, but there are some key elements that are common to all lithium-based chemistries. Some of the key differences between various lithium chemistries include cycle life, volatility, risk of thermal runaway, energy density, and cost of critical minerals used in the design of the battery. The scope of this paper will focus on the common attributes of lithium-based batteries and will refer to them as one chemistry.

3.2.2. Benefits

In general, lithium-ion batteries perform very well in cyclical applications with very minimal degradation or loss of capacity over time. At 80% depth of discharge, lithium-ion batteries are capable of between 2000 and 7000 cycles depending on chemistry and operating conditions.[ix]

One of the key differences between lithium-ion and many other chemistries is the requirement of a Battery Management System (BMS). The functionality of each BMS varies significantly from each application and each manufacturer, but at a minimum they monitor temperature, voltage and current. The main function of the BMS is to protect the battery from entering an unsafe state by disconnecting incoming and outgoing current if one of these monitored parameters goes beyond the specification limits. Some BMS also actively balance cells in a module to ensure the voltage remains consistent. A BMS used in a typical ESS will have additional functionality such as using past and present data to establish a state of health for the battery, recording or broadcasting data through Wi-Fi or wired connections, and alerting the user of an unsafe state. They also provide valuable data to the end user that can help understand usage and performance characteristics of the battery over time.

Applications that require a significant amount of energy in a limited space or have strict weight restrictions are perfect candidates for the use of lithium-ion batteries as the energy density is 2 to 3 times

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

higher than other chemistries. The data that is available thanks to the BMS also makes lithium-ion batteries a good solution for applications that require detailed data and remote monitoring.

3.2.3. Limitations

Although the BMS is a valuable tool in keeping lithium-ion batteries safe and providing data, it also presents more failure points to the system and adds cost to the overall ESS. BMS is a complex circuit board with many components and relies heavily on accurate data from temperature, voltage, and current sensors. If one of these sensors or components on the board fail or give a false reading, the BMS can disconnect the battery from the load, rendering it unusable. In addition, due to current limitations of the circuitry of the BMS, the BMS can become the limiting factor of how much instantaneous power can be provided by a battery.

Increasing public awareness of safety challenges associated with lithium-ion technologies has come to light over the past several years as issues have been reported. This includes issue with EV's, cell phones, grid-scale energy storage, E-bikes and more. Most currently available lithium-ion chemistries use an organic electrolyte and if a cell is damaged or abused it can enter into what is known as thermal runaway. UL9540A defines thermal runaway as: "The incident when an electrochemical cell increases its temperature through self-heating in an uncontrollable fashion. The thermal runaway progresses when the cells generation of heat is at a higher rate than the heat it can dissipate. This may lead to fire, explosion and gas evolution."[x] A thermal runaway event can be initiated from many sources including but not limited to: short circuit, puncture of cell, manufacturing defect, external heating, over charging, and discharging outside of specified parameters.

It is important to note that although lithium-ion batteries are capable of entering thermal runaway, modern design and manufacturing has significantly reduced this risk and that the number of issues as a percentage of batteries sold is very low. In addition, UL has created standards and test procedures to help ensure that batteries are as safe as possible. In choosing any lithium-based product, it is critical that all cells, modules, or batteries are sourced from a reliable manufacturer who adheres to strict quality and traceability guidelines and has undergone UL listing and testing as recommended in NFPA855.

Very cold climates can be particularly challenging for lithium-ion batteries, as they cannot be charged if the temperature drops below freezing without causing permanent damage at the cellular level. When the temperature drops below freezing the BMS will prevent the battery from charging until the temperature has returned to a safe level. One solution to overcome this challenge in cold climates is to add heat sources to maintain safe operating temperatures. Although this approach can be effective, depending on the severity of the climate, significant power could be required to maintain the temperature and it may take several hours to heat the batteries to a safe charging temperature after an extended outage.

Due to the difference in cell voltages and charge parameters between lithium-ion batteries and previously installed chemistries there may be some compatibility challenges between existing equipment and the requirements of the lithium-ion battery. In a retrofit scenario when swapping out lead-acid batteries for lithium, attention to detail should be given when considering voltage ranges, charge requirements, and BMS limitations of various chemistries under consideration.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.3. Hybrid Supercapacitor

3.3.1. Description

Hybrid supercapacitors are relatively new. They are considered electrostatic energy storage and combine the high cycling capability and high-power output of super capacitors with the power density of lithium-ion as illustrated in Figure **3**. Hybrid supercapacitors have primarily been targeted at 36V broadband and 48V telecom markets but have also seen some penetration into backup generator starting applications.

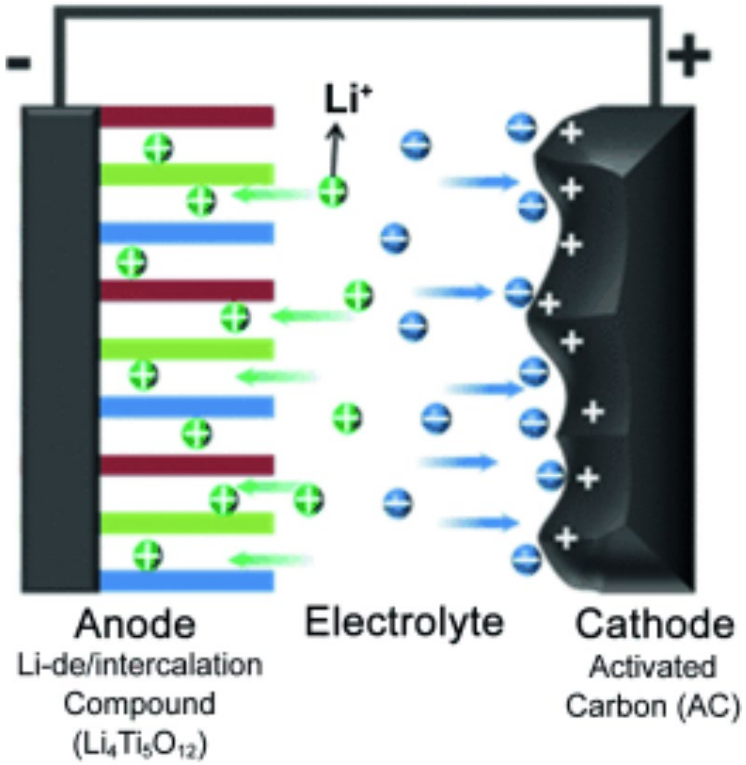


Figure 3 – Hybrid supercapacitor[xi]

3.3.2. Benefits

With a cycle life rating of up to 10,000 cycles, and up to a 20-year calendar life with minimal degradation, hybrid supercapacitors are ideal for use with unreliable grids or in off-grid applications. They can be recharged quickly (under 2 hours) and are capable of 1C continuous discharge – meaning the discharge current will discharge the entire battery in 1 hour. Although their energy density is lower than most lithium-ion chemistries, hybrid supercapacitors currently achieve a 100-150 Wh/kg, which is an improvement to most nickel-based and lead-based chemistries.[xii]

In a typical lithium-ion battery, during a thermal runaway event lithium-ion cathodes release oxygen which enhances or sustains the flames and makes it very difficult to extinguish a fire once it has started.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Conversely, hybrid supercapacitors use a lithium-doped carbon cathode which contains no oxygen, thus eliminating the risk of a self-sustained thermal runaway event.

3.3.3. Limitations

Currently, these batteries have an initial cost of up to 2 times that of lithium-ion. However, under certain use case scenarios where high throughput of kWh's or extreme cycling is important these batteries may offer an overall total cost of ownership (TCO) benefit over some of the other chemistries discussed, especially where longer-term TCO models of 20 years or more are used.

As with other batteries using lithium, hybrid supercapacitors have a limited operating and charging temperature range. They are unable to be charged in temperatures below 0°C and would require a heat source to recharge in colder climates.

3.4. Nickel-Zinc

3.4.1. Description

In recent years, nickel-zinc batteries have reemerged and have begun to gain ESS market share primarily in the datacenter industry. According to Zinc 5, Thomas Edison was awarded a US patent for a nickel-zinc battery in 1901, but he was never able to produce a commercially viable battery due to a very limited cycle life (Five, n.d.). The cycle life hurdle has been overcome in recent years allowing the technology to push forward. As a chemistry, nickel-zinc has many similarities with other nickel-based batteries such as nickel cadmium and nickel metal hydride which have been used for many years in various ESS applications. Nickel zinc batteries are valve-regulated, non-spillable batteries like many VRLA batteries.[xiii]

3.4.2. Benefits

Nickel-Zinc batteries use an alkaline, non-flammable liquid electrolyte and they have no propensity to enter thermal runaway. Depending on the manufacturer some nickel-zinc batteries offer a simplified BMS that can provide overcharge and overcurrent protection but are typically not as sophisticated as a BMS you would find on a lithium-ion module. This lack of a sophisticated BMS is primarily due to the inherently safe nature of the battery during operation.

Nickel-zinc batteries fall between lead acid and lithium-ion solutions on the energy density scale and is similar with other nickel-based chemistries as shown in Figure **4**. These batteries can provide significant power in a small footprint.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

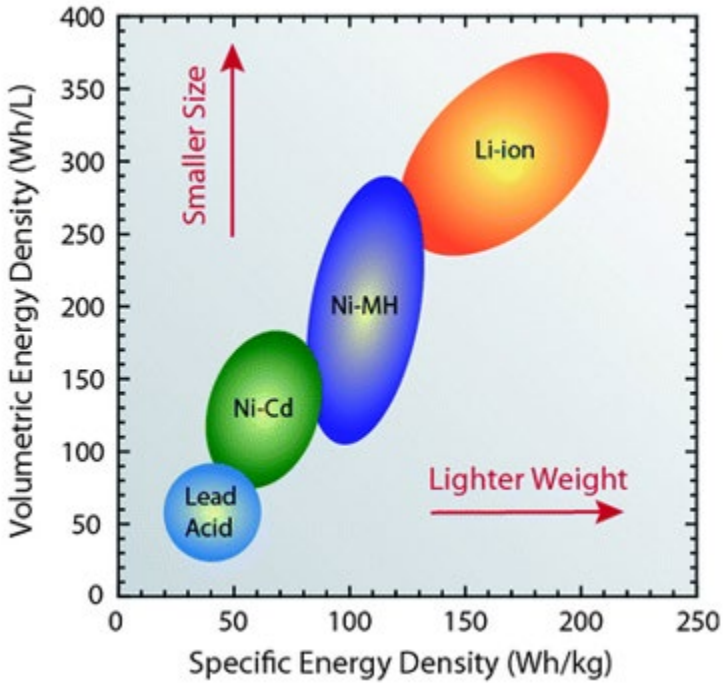


Figure 4 – Volumetric energy density of common energy storage chemistries[xiv]

3.4.3. Limitations

One important characteristic that is relatively unique to nickel-zinc chemistries that should be noted is the minimum charge and discharge profiles. In order to maintain optimal health and performance, nickel-zinc batteries can require 4-hour minimum charge and discharge rate. Lower charge and discharge rates may cause unwanted damage or premature aging of the cell, making it challenging to use this chemistry in applications requiring a multiple day discharge. However, the lower discharge rate makes it ideal for high-power short duration needs. In addition, this chemistry would not work well in applications with unreliable grids or when using intermittent sources of power such as renewables to charge the batteries.[xv]

3.5. Sodium-Ion

3.5.1. Description

Although this technology has been under development for many years, commercially viable sodium-ion batteries are very new to the market and have primarily been focused on the EV market. As the price and availability of lithium-ion has risen over the past few years, more focus has been put into working through the challenges with commercially producing a battery. At a chemistry level, sodium-ion batteries function much in the same way as lithium-based batteries but do not use lithium, cobalt, manganese or other rare minerals.

3.5.2. Benefits

Although slightly lower in energy density than current lithium-ion solutions, sodium-ion batteries use the low cost and abundantly available element of sodium as the as a key component. This has the potential to

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

significantly reduce costs and provide the ability to domestically source the raw materials instead of relying on delicate, and often unreliable, international supply chains. Safety is another key benefit when comparing sodium-ion to lithium-ion batteries. Sodium-ion batteries do not have any risk of entering thermal runaway.

3.5.3. Limitations

Sodium-ion batteries share many of the positive performance attributes of lithium-ion batteries while eliminating many of the concerns. As this is a new technology, initial deployments have yet to fully prove the validity of research-based claims of cycle life, safety, and overall performance of this chemistry. Sodium-ion will likely not completely displace lithium-ion batteries but the inherent safety of this chemistry along with the other key performance specifications make it a great candidate in the future of energy storage.

3.6. Comparing Technologies

Before discussing a decision process, it is useful to summarize how these technologies compare relative to each of our key considerations. As product cost and expected life are key considerations for analysis, we have added those into Table **3** below. As you can see, lead acid, lithium-ion and hybrid supercapacitor systems have distinct areas where they excel, while sodium-ion tends to be strong in all categories. Nickel-zinc does not seem to have any distinct advantages over other technologies.

Table 3 – Battery chemistry analysis (5=Best)

ESS Technology	Cost	Expected Life	Runtime		Temperature Resilience	Cycling
			By Volume	By Weight		
Lead Acid	5	2	2	1	5	1
Lithium-Ion	2	4	5	5	3	4
Hybrid Supercapacitor	1	5	2	3	3	5
Nickel-Zinc	2	2	3	2	2	2
Sodium-Ion	4	4	4	4	4	4

Notably missing from this chart are the safety and recyclability considerations discussed previously. These are vital inputs to the decision-making process, but they can be somewhat technology agnostic. While there is a proportional risk increase with any technology as the available energy increases, it is the management of that energy and thorough vetting via agency certification that determines how safe an ESS is. Second, within each technology there are significant variations in design and chemistry that impact their ability to be recycled. Both will still play a significant part in the decision-making process.

4. Determing the Best ESS Technology for any Given Site

Now that we have defined the key considerations for ESS deployment in the OSP and the benefits and limitations of common energy storage technologies, we can layout a process for deciding the best technology for any given site. By weighting the importance of each of the key considerations at a particular site and comparing those weightings to the relative strengths and weaknesses of deployable ESS technologies operators can be assured the best ESS is being deployed. While simple conceptually, the nuances in this process can drive significant gains in plant reliability. As we discuss this process, we

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

will do so with an understanding that every operator has their own methods of defining value for any decision and discuss valuation in general terms as this paper is not a study of valuation methods.

4.1. Aligning Application Needs to Technology

A basic initial method to align application needs to the best ESS technology is to create profile of the needs of a given site through the lens of the key ESS considerations. By ranking the importance of each key consideration one can quickly align site demands with technology advantages of a given ESS.

As an example: Figure **5** below evaluates a ground-mounted site in a location that generally sees very cold winters, is close to a service location, rarely has outages and has no additional value drivers for above standard backup time. As such, its profile might look something like the upper table in Figure 5.

ESS Technology	Cost	Expected Life	Runtime		Temperature Resilience	Cycling
			By Volume	By Weight		
Example Site #1	5	4	3	2	5	1

ESS Technology	Cost	Expected Life	Runtime		Temperature Resilience	Cycling
			By Volume	By Weight		
Lead Acid	5	2	2	1	5	1
Lithium-Ion	2	4	5	5	3	4
Hybrid Supercapacitor	1	5	2	3	3	5
Nickel-Zinc	2	2	3	2	2	2
Sodium-Ion	4	4	4	4	4	4

Figure 5 – Example Site Evaluation

By evaluating the site profile against the ESS Technology comparison, one would quickly see the best technology to align with site needs in this example is lead acid. The alignment of a need for cold temperature resilience with the wider functional operating range of lead-acid is the most critical synergy here. Cost is always a factor in any decision, however, with no other value drivers such as additional runtime regulations or critical loads being backed up, cost becomes more important relative to other considerations. This is a very targeted example of how to use a very simplistic approach, but often times there is no clear answer and it becomes necessary to use a more complex method.

4.2. Valuation of key considerations and costs

When alignments between technology and site needs are not as clear, leveraging a valuation method to determine the best path can be helpful. As with any operational decision, the ability to understand all key considerations and use data to assign them a value is paramount. A simple equation can be used to determine value added to the plant of any ESS technologies which is essentially a cost benefit analysis which determines which technology can add the greatest value.

$$\text{Value added} = \text{Benefits} - \text{Costs}$$

Within this “simple” equation are numerous complex nuances that any operator must determine for their own plant. First, what time period (T) should be used to determine my value added? This brings into play

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the life cycle of a product, its resiliency, plant impacts to customer churn rates, and, for public companies, the market expectation of return on capital investment. Recent sustainability initiatives have driven models with more long-term thinking, but each operator should use their own desired time period. Second, which of the key considerations should be considered a cost versus a benefit? For example, the safety of a system could be considered a cost or a benefit. Below we will lay out how key considerations can be thought of and even some scenarios where key considerations might eliminate an ESS from the decision process by default. Finally, what metrics does one use to value each of the potential benefits. This is something that should be decided by each operator, but we will discuss some finer points in determining value of each key consideration.

4.2.1. Key considerations: Costs

- **Product Costs –** This is probably the most intuitive part of the calculation. This includes the actual capital cost of the ESS hardware, the cost to install and permit and any assumed annual maintenance costs over the assumed life of the product. Assumed maintenance costs should be based on truck rolls required over the assumed time period following the manufacturer's recommended guidelines.

- **Safety –** While fundamentally a benefit, safety is more appropriately calculated as a cost. ESS's with well-developed safety systems that have been certified to high safety standards will have inherent cost built into their hardware, thus the potential long-term costs for safety should be close to zero. Energy Storage Systems that are not appropriately certified may have long term risks to employees or public safety which must be understood and quantified. Additionally, in many instances, certification to certain safety standards is required by Authorities Having Jurisdiction (AHJ) in order to deploy a technology in the plant.

- **Resilience** – A system's level of resilience to the harsh environmental conditions in the outside plant will impact its overall expected life and potentially its performance and runtime during its functional life. It is necessary to understand how resilience impacts these benefits and include this in the value-added calculations. In some cases, lack of resilience may preclude an ESS from being deployed in certain areas. For example, many ESSs functional temperature ranges only go to -20°C, so in colder areas where these temperatures occur regularly during winter these should be eliminated from consideration without a safe method of warming them. There are various environmentally controlled enclosures that can make deployment of these a possibility in extreme conditions, however, the incremental cost and maintenance for those enclosures must be considered as part of the valuation.

4.2.2. Key considerations: Benefits

- **Expected Life –** The benefit of long life is its ability to increase the duration of product replacement cycles and reduce Total Cost of Ownership (TCO.) Total capital cost of hardware used to calculate added value should be equal to the ESS hardware cost times the analyzed time period over the expected life. For example, if an ESS has a hardware cost of C_h, the analyzed time period is ten years, and the expected life is five years, the total hardware cost of the period is 10/5 x C_h or $2C_h$. While this should be based on the manufacturer's stated expected life, warranted life should be factored into this calculation

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

as well, as large gaps between warranted life and expected life for TCO calculations can leave operators with significant liability.

- **Runtime –** This is the primary benefit that can be quantified for an ESS. The amount of time that the system can run without needing to roll a truck to provide additional backup can greatly reduce operational costs and increase resiliency of the plant. In most cases having a plant that is adequately backed up provides value by reducing operational expense and improving customer satisfaction. Additionally, in some cases, being able to add greater runtime could open up new business opportunities with customers with higher reliability demand or could have significant cost reduction in sites that are more challenging to roll a truck to. While many times space is more of a premium and runtime by volume is more important, there are instances where runtime by weight can be key. A good example of this is where the best location for a site has no space to set a system on the ground and the pole owner has stringent weight restrictions to allow a cabinet to be mounted on a pole.

- **Cycling –** While many HFC powering sites rarely experience outages, in some cases where the grid is less reliable, frequent outages can mean increased operational costs from truck rolls or dissatisfied customers. In these instances, the ability for an ESS to handle more frequent outages without significant degradation to life or performance can have significant benefit. In addition to this, with the advent of time-of-use utility rate structures, there may be locations where ESS can be slightly oversized to provide significant reduction of utility expense, by cycling during peak rate periods.

- **Recyclability** – With many operators implementing sustainability initiatives to reduce their carbon footprint, the importance of using products with longer life which are composed of fully recyclable content cannot be overstated. Products which are recyclable at the end of their useful life have a significantly lower average emission factor and drastically reduce overall carbon footprint and hence impact to carbon neutral goals. Many energy storage products can be recycled, and while the return from recycling of products is rarely significant compared to the original capital investment, some require operators to pay to have them recycled creating additional liability at the end of their useful life. Additionally, paying to recycle spent products has a risk of motivating unsustainable behavior if thorough processes are not in place to ensure products are dealt with properly. When exploring an ESS solution, it is good to quantify any gain or liability from recycling at the end of the product's useful life.

By analyzing the value of each of these factors over time, one can determine the ESS technology that provides the most value in any scenario and determine which solution best meets the needs of any site.

While the valuation method is thorough, it involves a fair amount of characterization and calculation against each key consideration, for each technology. As more technologies become available this decision-making task can become increasingly daunting. One way to be more effective is to use the weighting and alignment method to reduce the decision to the two best aligned choices and use the valuation method to determine the best of the two ESS technologies for the application.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Conclusions

A fundamental element of the HFC network is the highly reliable power used to drive delivery of data and services to customers. Enabling this steadfast HFC power grid are the Energy Storage Systems that provide backup power when the utility grid fails to do so. As the network continues to evolve and becomes an increasingly more critical pipeline to connect us socially and economically, the make-up of the ESS holding that network up needs to evolve to meet ever-increasing demands for availability.

Because of this need for ESS to advance, innovative new solutions for storing energy in the plant are regularly being explored. These new solutions have continued to improve the reliability of the plant as it evolves around new challenges and regulations, validating the need for continued exploration and implementation of new solutions. The evolution of Lead-Acid, lithium-ion, Hybrid Supercapacitor, and in the near future Sodium Ion ESS technologies, have provided an array of solutions with various benefits and limitations which make available options to solve many of the challenges facing us today.

By continuing to build a library comparing the relative benefits and limitations of ESS technologies, we can use a simple weighting method to visualize how these technologies stack up with regard to cost, runtime, safety and other key considerations in the OSP. By using the weighting and alignment method to filter ESS choices for an application down to the best options, then using the valuation method to analyze those options in light of key considerations for OSP deployment, operators can make data-driven decisions on the most value-added energy storage solution possible. These methods can be expanded easily as new technologies and additional key considerations come to light, they provide a decision framework that allow our industry to continue to build the most reliable network possible.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AGM	Absorbent glass mat
AHJ	Authority Having Jurisdiction
CI	Critical infrastructure
CMTS	Cable modem termination system
CPUC	California Public Utility Commission
ESS	Energy storage system
HFC	Hybrid fiber-coax
IFC	International Fire Code
kWh	Kilowatt hour
NFPA	National Fire Protection Association
OSP	Outside plant
TCO	Total cost of ownership
TPPL	Thin plate pure lead
UL	UL Solutions (formerly Underwriters Laboratories)
UPS	Uninterruptable power supply
VLA	Vented lead acid
VRLA	Valve regulated lead acid

Bibliography & References

[i] (California Public Utilities Commission)
[ii] (NFPA, Standard for the Installation of Stationary Energy Storage Systems 2023 Edition, 2023)
[iii] (UL, ANSI/CAN/UL Batteries for Use in Stationary and Motive Auxiliary Power Applications, 2022)
[iv] (UL, Ansi/CAN/UL9540:2023 Standard for Safety Energy Storage Systems and Equipment, 2023)
[v] UL, ANSI/CAN/UL 9540A:2019 Standard for Safety Test method for Evaluating Thermal Runaway Fire Propagation in Battery Energy Storage Systems, 2019

[vi] (Battery University, 2021)
[vii] (Battery Design, 2023)
[viii] (International Battery Council, 2023)
[ix] (Battery University, 2021)
[x] UL, ANSI/CAN/UL 9540A:2019 Standard for Safety Test method for Evaluating Thermal Runaway Fire Propagation in Battery Energy Storage Systems, 2019
[xi] (Kim, Park, Cho, & Kim, 2013)
[xii] (ATX, 2023)
[xiii] (ZincFive, 2023)
[xiv] (epec engineered Technologies, 2023)
[xv] (ZincFive, 2023)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

From Manual to Automated

AI-Driven Network Engineering and Operations

A Technical Paper prepared for SCTE by

Nader Foroughi
CTO of Americas
Technetix Inc.
8490 Upland Drive, Englewood, CO
403 606 7427
nader.foroughi@technetix.com

Chris Beem
CTO Engineer
Technetix Inc.
Technetix B.V., Kazemat 5, 3905 NR, Veenendaal
chris.beem@technetix.com

Diego Royo Moros
CTO Engineer
Technetix Inc.
Technetix B.V., Kazemat 5, 3905 NR, Veenendaal
diego.royo@technetix.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Modern Day design Challenge 4
   - 2.1. HFC Statistics 5
   - 2.2. Histograms and Probability Density Functions 6
   - 2.3. Non-Linearities in HFC 8
   - 2.4. SNR/MER Convergence 9
3. AI-Driven Design 10
   - 3.1. Gradient Optimization 11
   - 3.2. Application 12
     - 3.2.1. Method 13
     - 3.2.2. Network Installation 14
     - 3.2.3 Amplifier Communication 14
     - 3.3 Use Cases 17
     - 3.3.1 Carrier Expansion and Reduction 17
     - 3.3.2 Split Change 20
   - 3.3. Power Optimization 23
4. Conclusion 26

Abbreviations 27

Bibliography & References 27

List of Figures

Title **Page Number**

Figure 1 – Various Teams Interfacing with Access Network 6
Figure 2 – Convolution of Two Normally Distributed Variables 8
Figure 3 – Cascade CCN vs Starting CCN 9
Figure 4 – Simplified Gradient Optimization 11
Figure 5 – Analyzed Plant Map 13
Figure 6 – Simulation Span 14
Figure 7 – PLC Implementation in HFC Amplifiers 16
Figure 8 – Channel Loading Examples for Various Traffic Patterns 16
Figure 9 – Insertion Loss of 20 Different Passive Devices from 100 Hz to 5 MHz 17
Figure 10 – Channel Loading Examples for Various Traffic Patterns 17
Figure 11 – System Operating at 204 MHz Split 19
Figure 12 – Carriers Added with No Optimization 19
Figure 13 – Optimized System Post Carrier Expansion 19
Figure 14 – Channel Loading Examples for Various Splits 20
Figure 15 – 1.8 GHz System Operating at 204 MHz Split 22
Figure 16 – The Same System Operating at 492 MHz Split – Post Optimization 22
Figure 17 – Example Amplifier CINR Curve 23
Figure 18 – Amplifier CINR and Required TCP – Full Channel Loading 24
Figure 19 – Amplifier CINR and Required TCP – Reduced Channel Loading 24

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 20 – Amplifer Cascade Analyzed for Power Saving Mode 25
Figure 21 – Amplifer Cascade High Power Mode Consumption 25
Figure 22 – Amplifer Cascade Low Power Mode Consumption 26

List of Tables

Title **Page Number**

Table 1 – Amplifier Telemetry Requirements 15
Table 2 – Amplifier Alignment – Channel Loading Changes 18
Table 3 – Amplifier Alignment – Split Change 20
Table 4 – Average CM Rx Level Before and After 21

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The hybrid fiber-coax (HFC) networks are rapidly changing as we approach mass deployments of next generation DOCSIS and spectrum expansion. Knowing that HFC networks amalgamate the strengths of optical fiber and coaxial cable technologies, we are faced with a new level of intricacy to their design and maintenance. As the frequencies utilized in these networks continue to climb, the intricacies associated with HFC networks have surged significantly. One of the key factors contributing to this intricacy is the complex interdependencies among the various departments engaged in the design and maintenance process of HFC networks. The alignment of different segments such as network planning, engineering, and operations has become pivotal for the seamless functioning of these networks.

As the demand for higher data speeds and bandwidth increases, the frequency spectrum used in HFC networks expands, leading to enhanced complexity in design. With higher frequencies, challenges such as signal attenuation, interference, and noise become more pronounced. Furthermore, different departments—such as engineering, operations, and maintenance—become increasingly interdependent due to the intricate nature of these networks. Ensuring efficient signal transmission, maintaining signal quality, and minimizing signal degradation are tasks that require a meticulous approach.

This paper delves into how network automation and software defined networks can alleviate a large majority of the concerns, both from plant upgrade and on-going maintenance perspectives. A central focus of this exploration is the transformative potential of artificial intelligence (AI) in reshaping the landscape of end-to-end design and maintenance within the realm of HFC networks. By harnessing the power of AI, it becomes possible to reimagine and revolutionize the conventional methodologies that govern the optimization of nodes, amplifiers, and the overall capacity management. The utilization of AI-driven algorithms and predictive analytics offers a promising avenue for addressing the challenges associated with HFC networks, paving the way for enhanced efficiency, reliability, and adaptability.

This paper interlaces the potential advantages of AI in HFC network design with its impacts across the diverse realm of network management. These benefits span from substantial cost savings via power and resource optimization to informed decision-making enabled by data-driven AI insights. Moreover, the infusion of AI can also increase the plant reliability of HFC networks, enhancing proactive maintenance and anomaly detection algorithms.

2. Modern Day Design Challenge

Historically, operators have predominantly used static parameters to formulate optimal network designs for specific regions. A case in point is the downstream design process, which traditionally unfolds through a series of steps. Initially, operators identify a worst-case or, at the very least, a reasonable worst-case scenario for span loss. Subsequently, they establish a corresponding worst-case or reasonable worst-case situation for drop loss. Following this, the optimal receive level of end devices such as modems and set-top boxes is determined. Finally, operators determine the maximum output power of amplifiers and nodes, a crucial criteria to not exceed. However, this conventional approach, while providing a baseline for network design, lacks the dynamic adaptability and precision that contemporary data-driven techniques can offer.

Furthermore, in the conventional approach, network designers often take the side of caution, designing the network infrastructure to accommodate peak demands and extreme usage scenarios. While this strategy ensures network reliability during spikes in usage, it leads to underutilization of resources during normal operation, resulting in inefficient capacity allocation. This over-provisioning of resources translates to increased costs in terms of both equipment and energy consumption, hindering cost-

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

effectiveness and sustainability. Moreover, traditional designs lack the agility to adapt to fluctuating usage patterns, leading to suboptimal performance during off-peak hours.

Contrastingly, emerging data-driven and AI-assisted methodologies for network design introduce a paradigm shift. Instead of relying solely on static worst-case assumptions, these advanced techniques can tap into real-time data streams and predictive algorithms. This empowers operators to tailor their designs more precisely to actual usage patterns and network conditions. The rigid sequence of steps in traditional design gives way to a dynamic and adaptive process, where network adjustments are made in response to fluctuating demand and environmental factors. By leveraging the insights gleaned from data analytics, operators can achieve a level of optimization that transcends the limitations of traditional static approaches. By analyzing historical usage patterns and current network conditions, AI models can optimize capacity allocation dynamically, ensuring that resources are allocated precisely where and when they are needed most. This adaptive approach not only maximizes the utilization of available bandwidth and infrastructure but also reduces the need for excessive over-provisioning. Consequently, this data-driven design methodology leads to substantial cost savings, as operators can efficiently allocate resources based on actual demand rather than hypothetical worst-case scenarios.

Data-driven and AI-assisted HFC design revolutionizes network planning by harnessing the power of near-real-time data and predictive analytics. By analyzing historical usage patterns and current network conditions,

Furthermore, the incorporation of AI and data-driven techniques in HFC design addresses another critical aspect: power reduction. Traditional designs tend to overlook energy efficiency, focusing primarily on network stability. This oversight leads to unnecessary power consumption, contributing to higher operational costs and a larger carbon footprint. In contrast, AI-driven designs take into account dynamic power management and predictive algorithms. By optimizing power distribution and consumption across the network, AI-based designs can significantly reduce energy expenditure without compromising network performance or reliability. This dual advantage of cost reduction and environmental sustainability positions data-driven design as a compelling alternative to traditional HFC approaches.

AI-driven design can also further enhance predictive maintenance capabilities, where machine learning algorithms continuously monitor network components and identify early signs of degradation or impending failures. This proactive approach enables network operators to perform targeted maintenance interventions, mitigating the risk of unplanned downtime and enhancing overall plant reliability. This reliability enhancement is crucial in maintaining customer trust and satisfaction, a factor that is often compromised in traditional design approaches.

2.1. HFC Statistics

Manual design and maintenance of HFC networks experiences significant challenges stemming from design, operations, and engineering domains bring heavily intertwined. The traditional approach to HFC network management relies on segregated teams responsible for different stages of the network lifecycle, including initial design, deployment, and ongoing maintenance. However, the high degree of interdependencies between these stages gives rise to a cascade of downfalls that hinder efficiency, scalability, and overall network performance.

The functional isolation of engineering, design and operations impedes the flow of critical information and insights that each domain possesses. Designers might lack real-world operational context, leading to impractical designs that struggle to adapt to the complex realities of network deployment. Conversely, operational teams might find themselves constrained by design choices that do not align with the practical

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

constraints they encounter in the field. This lack of synergy perpetuates inefficiencies, maintenance challenges, and suboptimal network performance.

Figure 1 outlines the interdependency of various departments involved in design and maintenance of a network.

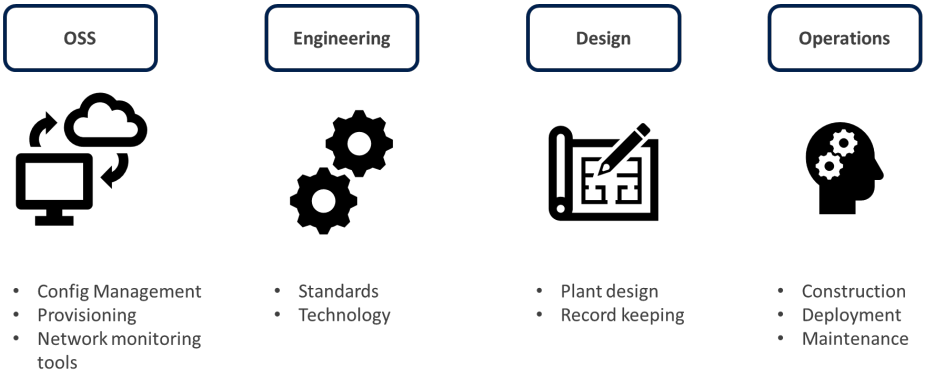


Figure 1 – Various Teams Interfacing with Access Network

Moreover, the manual approach encounters difficulties when confronted with the dynamic nature of modern telecommunication networks. HFC networks are subject to ever-evolving demands, technologies, and environmental factors. Manual design and maintenance processes struggle to keep up with rapid changes, resulting in delayed responses to network issues, prolonged downtimes, and ultimately, reduced customer satisfaction. The interplay between design, operations, and engineering requires adjustments and adaptive strategies that manual processes often struggle to accommodate.

In light of these challenges, the shift toward AI-driven solutions is gaining traction. AI has the potential to bridge the gaps between design, operations, and engineering by facilitating continuous data exchange, predictive analytics, and automated adjustments. By leveraging AI's ability to process large volumes of real-time data and optimize network parameters across the entire lifecycle, HFC networks can achieve enhanced efficiency, agility, and resilience, overcoming the pitfalls associated with the traditional manual approach.

In the following sections we will explore why some of the historical and manual design and deployment methodologies can result in suboptimal system performance.

2.2. Histograms and Probability Density Functions

In HFC networks, understanding the distribution of cable modem transmit levels and receive levels is crucial for optimizing network performance. As a matter of fact, modem and set top box levels, both in the upstream and downstream are the core of HFC design. As mentioned in the previous section, the historical approach entails selecting a worst-case scenario or reasonable worst-case scenario to design the access network. Knowing that many probability density functions (PDF) are Gaussian in nature, the historical method of designing the network can be described as designing for $\sim 3\,\sigma$ which can be deemed a worst-case scenario.

Gaussian distributions are commonly used to model these levels due to their prevalence in real-world scenarios. However, a noteworthy observation is that while individual cable modem transmit and receive levels might exhibit Gaussian distributions, the resulting signal-to-noise ratio (SNR) or modulation error

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

ratio (MER) might display left-skewed characteristics when considered in aggregate. This phenomenon arises due to the complex interplay of factors within the HFC environment.

Cable modem transmit levels are subject to variations caused by factors like attenuation, noise, and signal degradation. Similarly, receive levels are influenced by various factors including signal attenuation, noise amplification, and interference. The inherent variability in these parameters contributes to the Gaussian nature of their distributions. However, when these variations accumulate and affect the SNR or MER, a left-skewed distribution may emerge. This asymmetry is a result of the impact of unfavorable conditions that can lead to sudden drops in signal quality, producing extended tails on the left side of the distribution.

Mathematically speaking, we can also prove that the sum of two or more Gaussian distributed histograms does not necessarily result in a Gaussian distributed result and can exhibit skewness. Let us consider the sum of two Gaussian distributions.

Let X be a random variable following a Gaussian distribution with mean μ_1 and variance σ_1^2, and Y be another random variable following a Gaussian distribution with mean μ_2 and variance σ_2^2. The probability density function (PDF) of a Gaussian distribution is given by:

$$f(x) = (1 / (\sigma * \sqrt{(2\pi)})) * e^{(-((x-\mu)\text{^}2) / (2\sigma\text{^}2))}$$

Now, let us consider the sum Z = X + Y. The mean of Z is μz = μx + μy, and the variance of Z is $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$ due to the independence of X and Y.

The PDF of Z can be obtained by convolution of the PDFs of X and Y:

$$fz(z) = \int [f(x) * f(y)]\, dx$$

This convolution operation makes the result more complex, and in general, it does not yield a Gaussian distribution. When $\mu x \neq \mu y$ and/or $\sigma x \neq \sigma y$, the convolution may lead to skewness in the resulting distribution. The convolution of two Gaussian distributions may result in a distribution that is broader or skewed, rather than being perfectly Gaussian.

Another phenomenon that can be observed as a result of this is the fact that there is typically a low amount of correlation between transmit and receive levels in comparison to downstream and upstream SNR/MER.

In the context of Gaussian distributions, the correlation between X and Y can be expressed as:

$$\rho(X, Y) = \text{cov}(X, Y) / (\sigma_1 * \sigma_2)$$

Where:

- $\text{cov}(X, Y)$ is the covariance between X and Y,
- and $\sigma_1 * \sigma_2$ is the product of their standard deviations.

When the convolution of X and Y leads to a skewed or non-Gaussian distribution, the relationship between the original Gaussian PDFs and the resulting PDF becomes nonlinear. This nonlinearity can result in a lower covariance (cov(X, Y)) and consequently a lower correlation (ρ(X, Y)) between the original Gaussian variables and the resulting variable.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Graphically the derivation above can be shown as:

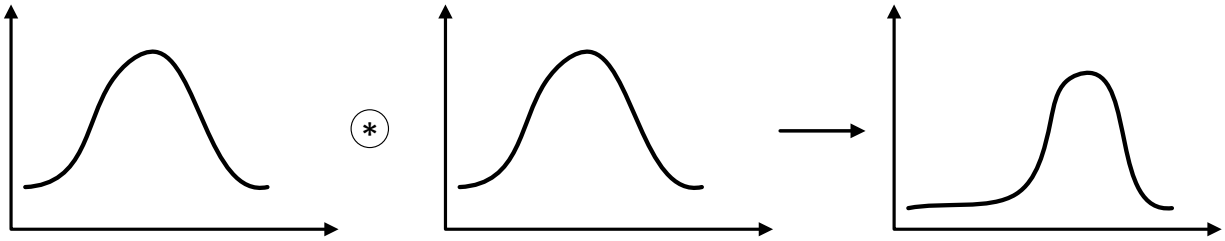


Figure 2 – Convolution of Two Normally Distributed Variables

From an RF perspective, we will explore this non-linearity in the section below.

2.3. Non-Linearities in HFC

In this section we will explain the phenomenon laid out in section 2.2, from a noise and distortion accumulation perspective.

We know that the carrier-to-noise ratio (CNR) of a single amplifier can be derived from the following formula:

$$C/N\,(dB) = \; C_i(dBmV) + 57.1 - NF(dB)$$

Where:

- C_i: input signal
- NF: Noise figure of the amplifier

The number 57.1 is the minimum thermal noise at ~16.7 °c expressed in dBmV/6.4MHz.

Note: the minimum thermal noise power for different temperatures can be calculated using the following formula:

$$n_p = kTB$$

where:

- n_p = noise power in watts
- k = Boltzmann's constant (1.34×10^{-23} joules/K)
- T = absolute temperature in K
- B = bandwidth of the measurement in Hz

Knowing the noise products accumulate in 10log fashion due to the non-coherent nature of noise, the overall cascade C/N for amplifiers operating at different output levels can be derived from the following equation:

$$C/N_{total}\,(dB) = \; -10log\left\{10^{\frac{-C/N_1}{10}} + \; 10^{\frac{-C/N_2}{10}} + \cdots + \; 10^{\frac{-C/Nn}{10}}\right\}$$

Where, C/N_x is the carrier to noise of each amplifier calculated independently.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Distortion products from an amplifier or series or amplifiers historically have been characterized by measuring composite second order (CSO) and composite triple beat (CTB) on analog carriers. These distortion products are harmonics of the primary signal. Today on the other hand, cable operators primarily use digital carriers. Digital carriers' distortion products do not appear similar to the analog carriers. Instead, they appear very similar to a raised noise. For this reason, composite intermodulation noise (CIN) is the best way to characterize the distortion performance of amplifiers today. The rate of accumulation of CIN products is dependent on whether the distortion products are coherent or non-coherent. The range of CIN accumulation can range between [10-20]log.

Carrier-to-composite noise (CCN) in this paper has been used as the primary method of determining signal quality. Although SNR and MER can be measured using meters and measurement equipment, there are inconsistencies in these types of measurements, especially when considering different measuring equipment. For this reason, we will consider carrier-to-composite noise (CCN) as the true measure of performance in a cascade. Summing noise and distortion products in a cascade, CCN can be derived from the following formula:

$$CCN(dB) = -10log\left\{10^{\frac{-CCN_i}{10}} + 10^{\frac{-CTN_{Total}}{10}} + 10^{\frac{-CIN_{Total}}{10}}\right\}$$

Where CCN_i is the starting CCN. In essence, the starting CCN (MER) of the modem in the upstream or the CCN (MER) of the node/remote PHY device (RPD) in the downstream.

2.4. SNR/MER Convergence

Based on the formulas laid out in the previous section, Figure 3 demonstrates how CCN can converge into a very predictable number based on the depth of the cascade.

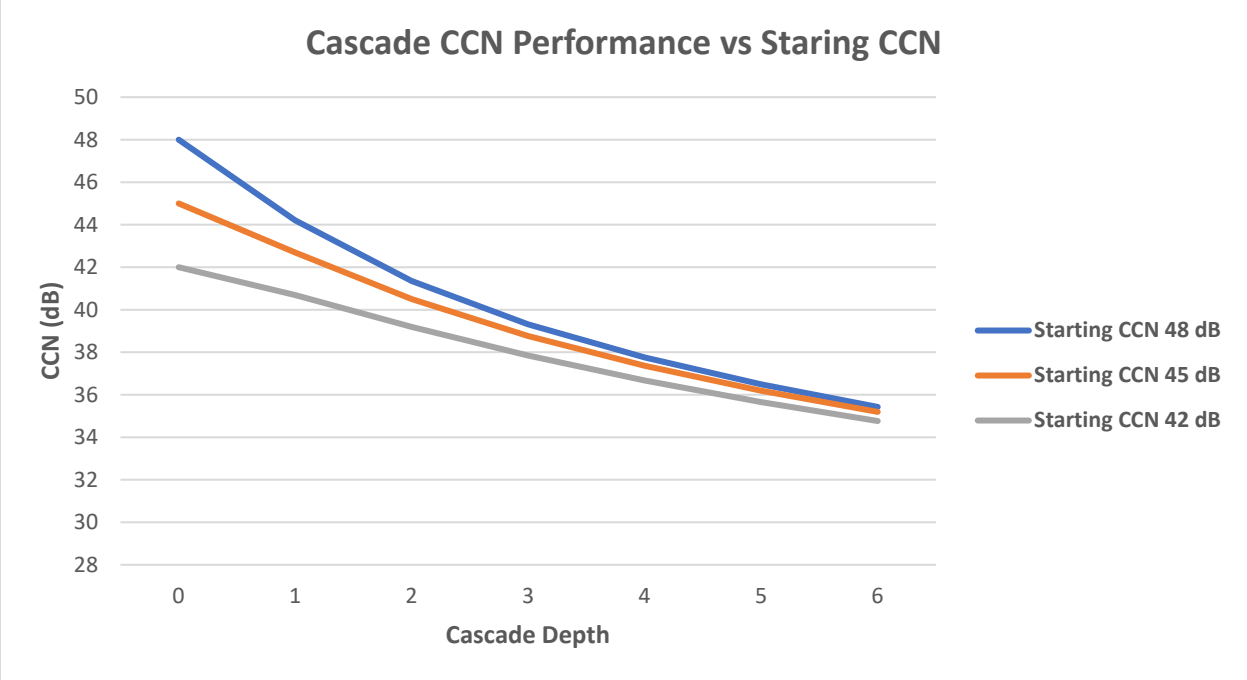


Figure 3 – Cascade CCN vs Starting CCN

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 3 serves as an illustration for what was discussed in section 2.2, the emergence of skewed SNR and MER distributions, despite the Gaussian nature of cable modem transmit and receive levels. This intriguing occurrence can be the main reason behind many complexities observed in the HFC plant.

As discussed in the previous section, the Gaussian distribution, characterized by its bell-shaped curve, often characterizes the individual transmit and receive levels of cable modems within HFC networks. However, the amalgamation of these individual Gaussian distributions into aggregate SNR and MER distributions reveals a fascinating skewness. The key insight that Figure 2 demonstrates is the manifestation of left-skewed tendencies, where the distribution's tail extends towards the lower values of SNR and MER.

This observation holds significance for HFC network design and management. Ensuring consistent and reliable network performance requires not only a comprehensive understanding of individual distributions but also an awareness of how these parameters interact and shape the overall quality of signal transmission. By considering the left-skewed tendencies in aggregated SNR/MER distributions, HFC network operators can implement strategies that address the challenges posed by such asymmetries. Advanced data-driven techniques and AI-driven algorithms can be employed to dynamically adjust network parameters and mitigate the impact of left-skewed distributions, leading to improved overall performance and end-user experiences in HFC systems.

This peculiar left-skewed behavior is attributable to the intricate dynamics at play within the HFC plant. Various factors, including attenuation, noise amplification, and distortions, contribute to the variations in both upstream and downstream signal quality. When these variations culminate and impact the SNR and MER, unfavorable conditions can lead to pronounced declines in signal quality. As a result, the left-skewed extension becomes pronounced, highlighting the critical role of understanding not only the characteristics of individual parameters but also their collective impact on the overall network performance.

In the context of HFC network management and optimization, the insights gleaned from Figure 2 are invaluable. By discerning the underlying causes of skewed SNR and MER distributions, network operators can tailor strategies to address and alleviate the asymmetries. Leveraging advanced analytical techniques and AI-driven algorithms, HFC systems can be dynamically adjusted to counteract the adverse effects of cumulative noise and distortions, ultimately resulting in enhanced signal quality and increased reliability in the access networks.

3. AI-Driven Design

So far in the paper we have explored the non-linear relationships between output power, signal quality and other factors. Network design has proven to be incredibly complex due to the sheer number of parameters that one has to take into consideration, especially taking the non-linear nature of these parameters. The integration of AI presents a revolutionary shift from traditional approaches. The limitations of static worst-case scenario planning have become evident, particularly when examining the results of the analyses in the previous sections.

AI-driven design offers a transformative solution to this challenge. By leveraging near-real-time data and predictive analytics, AI models can dynamically optimize network parameters. This data-driven methodology not only maximizes capacity utilization but also can mitigate the skewed SNR and MER distributions. Unlike traditional worst-case scenario planning, AI adapts to fluctuating network conditions, responding to variations in transmit and receive levels and thereby preventing the accumulation of noise and distortions leading to sub-optimal results.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The significance of AI-driven design extends beyond signal quality. Traditional methods involve manual intervention and static assumptions for capacity planning and channel loading, resulting in inefficiencies and underutilized resources. In contrast, AI optimizes node placement, amplifier output power, and capacity management, resulting in substantial cost savings and improved reliability.

The primary method explored in this paper is gradient optimization in any typical machine learning algorithm to optimize for an end objective. Due to the vast number of areas that the gradient optimization can be applied to, for the purpose of this paper, we have only considered the optimization of output power for signal quality and power consumption in the access networks.

3.1. Gradient Optimization

AI-driven HFC design can be framed as a gradient optimization problem, knowing that HFC design and optimization are differentiable functions in nature. In this context, the goal is to find the set of parameters that minimizes or maximizes an objective function, reflecting the desired outcome. The process involves iteratively adjusting network variables based on the calculated gradient of the objective function with respect to these parameters. This iterative refinement is guided by the general gradient optimization formula:

$$\theta_{k+1} = \theta_k - \alpha * \nabla J(\theta_k)$$

Where:

- θ: the parameters being optimized
- α: the learning rate determining the step size in each iteration
- $\nabla J(\theta_k)$: the gradient of the objective function J with respect to θ at the current iteration k.

The gradient $\nabla J(\theta_k)$ indicates the direction of steepest ascent (for maximization) or descent (for minimization) in the parameter space. By iteratively updating θ based on the gradient, the optimization process converges towards a solution that aligns with the desired network performance outcomes.

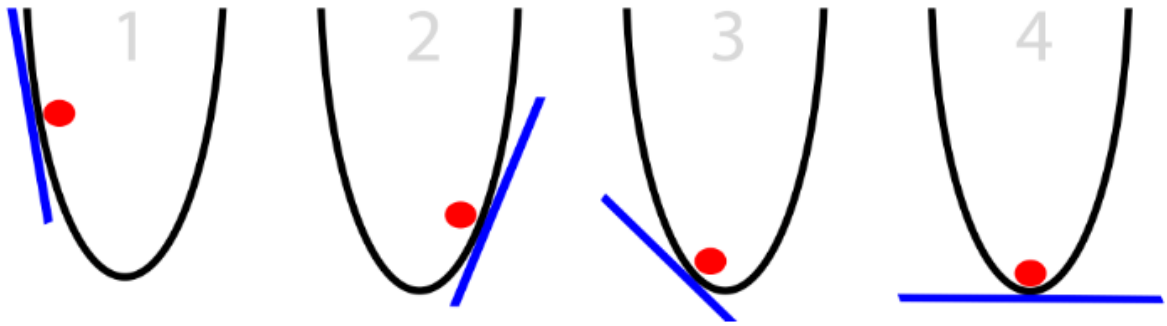


Figure 4 – Simplified Gradient Optimization

In the context of HFC network design, this gradient optimization paradigm enables AI algorithms to dynamically adjust variables such as cable modem transmit levels, amplifier settings, and even node placements. The objective function can encapsulate goals like maximizing data throughput, minimizing power consumption, and increasing signal quality/capacity. By iteratively fine-tuning network parameters based on the calculated gradients, AI-driven HFC design converges towards a configuration that

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

optimally balances diverse performance factors, resulting in a network that is efficient, reliable, and tailored to near-real-time demands.

3.2. Application

In the context of this paper, when the outside plant is established and calibrated to ensure adequate signal delivery to cable modems (CMs), an array of factors, including adjustments to outside amplifier settings, can induce fluctuations in CM signal levels. The impact of events such as variations in outside amplifier configurations can potentially disrupt signal consistency for CMs. However, the application of internal control techniques, such as temperature and gain control mechanisms inherent in digital amplifiers, effectively counteracts signal drift attributed to temperature fluctuations and passive component deviations.

Furthermore, network optimization plays a pivotal role in maintaining the stability of the entire network across amplifiers within the access network. This optimization process is designed to comprehensively address network stability concerns by accounting for the relationship of amplifiers and environmental factors. In the context of this paper, four distinct scenarios underscore the importance of network optimization as a control mechanism within the access network:

- Initial Deployment: During the initial network setup, network optimization techniques are employed to fine-tune parameters, ensuring an optimal signal distribution across CMs. This process enhances signal reliability and minimizes potential disparities in signal strength, thereby establishing a robust foundation for network performance.
- Carrier Expansion and Management: As operators expand their carrier offerings or activate/deactivate specific carriers, the dynamic nature of network optimization comes into play. This technique allows operators to precisely manage carrier configuration changes, maintaining equilibrium in signal distribution and preventing disruptions during transitions.
- Split Adjustment: When adjustments to the upstream and downstream splits are needed, network optimization techniques facilitate seamless adaptation. These methods ensure that the alterations are executed smoothly, without causing distortions in signal quality or network stability.
- Plant Element Changes: Plant element changes necessitate recalibration of the network to accommodate the altered load. Network optimization is crucial in this scenario, recalibrating parameters to ensure uniform signal strength and optimal network performance despite changes in the environment.

In order to demonstrate how the methodologies of this paper can be implemented, the design shown in Figure 5 has been taken into consideration.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

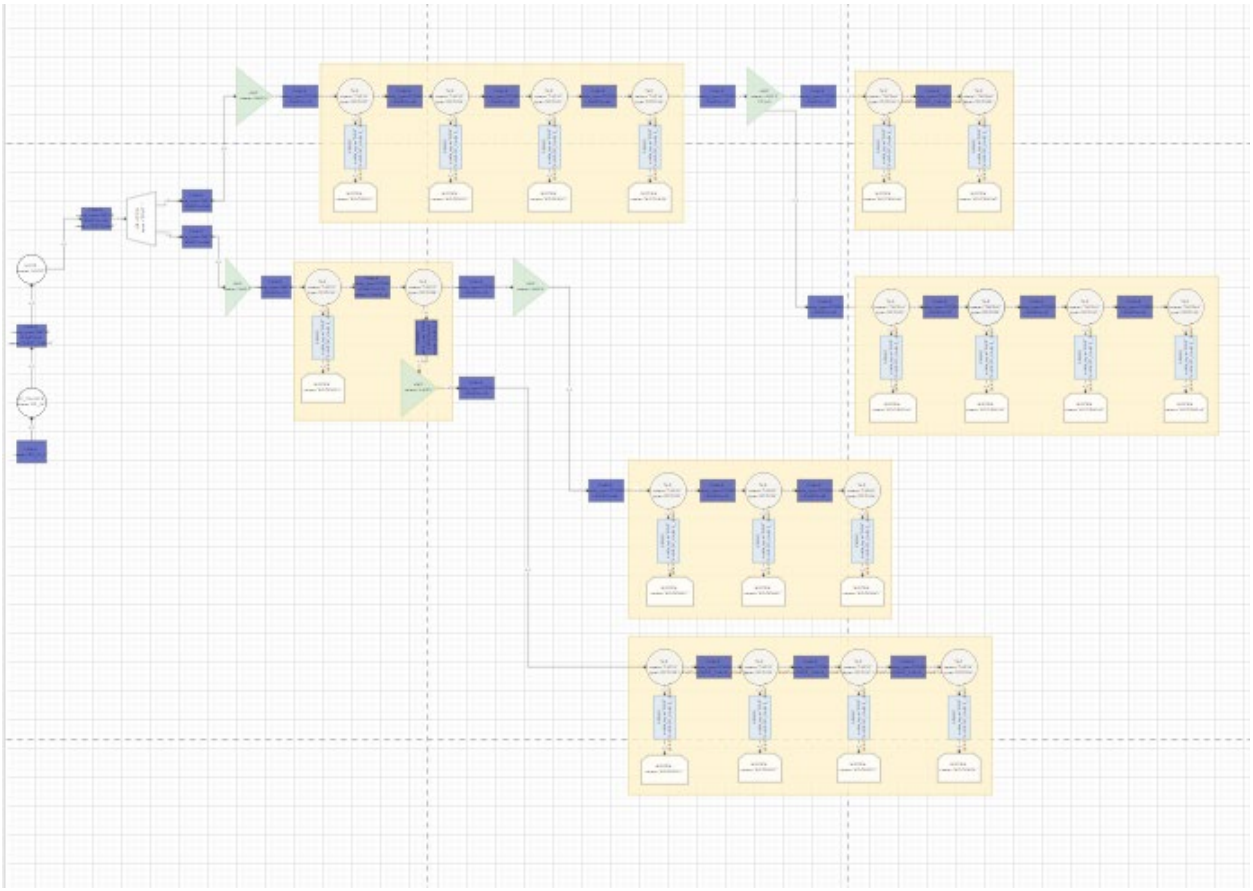


Figure 5 – Analyzed Plant Map

3.2.1. Method

Our network modeling involves the utilization of simulations that mirror real-world components like CMs, passives, amplifiers, and cables. These simulated elements emulate the behaviors and attributes of actual equipment. The internal configurations and settings of amplifiers in these simulations are reflective of their real-world counterparts, and passive losses are derived from empirical measurements. The network simulation process comprises two key components:

- Signal Propagation Simulation: This involves the propagation of signal levels through the network, emulating the real-world interactions between various components. This simulation provides insights into how the signal levels change as they traverse the network architecture. It sheds light on the dynamics of signal attenuation, amplification, and other factors that impact signal quality.
- Amplifier Settings Optimization: The second part of the simulation focuses on optimizing amplifier settings to achieve a specific objective. In this case, the optimization goal is to ensure that the majority of modems receive an optimal level of 0 dBmV. The amplifier settings are dynamically adjusted to achieve this objective, aiming for a flat distribution of signal levels across the network. This iterative process involves updating amplifier configurations based on real-time feedback to ensure that modems receive the desired signal level.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Each simulation scenario is accompanied by a comprehensive report detailing crucial parameters. These parameters encompass the amplifier output total composite power (TCP) levels, which provide insights into signal strength, and the convergence metric indicating the deviation of modem receive levels from the optimal 0 dBmV target. The network model also introduces variability by randomizing the drop cable lengths to CMs, reflecting the diverse conditions encountered in real-world HFC setups.

3.2.2. Network Installation

A classical network installation scenario shown in Figure 6 represents a worst-case instance where all amplifiers share identical output levels. Within this setup, amplifiers 1, 2, and 3 are each equipped with a 45 dB gain capability, sufficient to overcome the loss incurred across 150 meters of COAX9 cable connecting amplifier 1 to amplifier 2. However, the heightened gain of amplifier 2 results in an excessively potent signal transmission to amplifier 3, initiating a situation wherein the latter excessively pads the signal. Consequently, this higher output power degrades signal quality. This intricate chain of events not only leads to energy waste, but also impacts the signal reception quality of CMs such as CM1A, 1B, and 1C, potentially yielding higher-than-optimal receive levels.

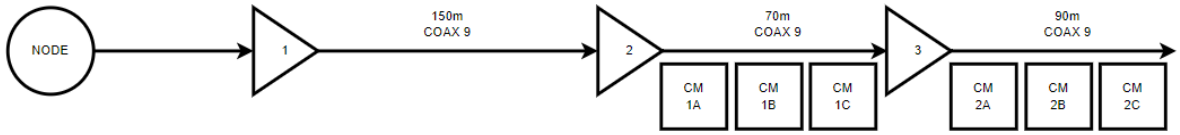


Figure 6 – Simulation Span

Unlike the uniform output level approach of the classical method, network alignment recommends a dynamic divergence in amplifier output levels. This strategic divergence holds the potential to drastically reduce energy consumption across the network while concurrently ensuring desirable signal levels for CMs. By distributing energy resources, network alignment establishes a harmonious balance wherein CMs receive optimal signal strength without the attendant wastefulness associated with uniform amplification. This adaptive strategy optimizes energy usage, eliminates energy waste, and enhances the quality of signal transmission to CMs, underscoring the promise of network automation in redefining efficient HFC network design and performance.

3.2.3 Amplifier Communication

In contrast to local amplifier control, network alignment involves the analysis of network data in a centralized location. This data encompasses measurements obtained from both digital amplifiers and CMs. The process of network alignment, as elaborated in the context of network optimization, requires the independent optimization of a specific service group or node, distinct from other service groups.

These dynamics extend across multiple service groups, particularly with regards to interdependencies shaped by external factors such as various plant conditions and operator-specific intricacies. The integration of such metadata holds the potential to enhance network optimization by considering these holistic influences. While details pertaining to service groups are likely to be shared efficiently within a cloud environment, this process remains detached from the communication that occurs between active elements in those individual service groups.

For the successful adoption of communication technology between amplifiers, certain prerequisites must be met. This includes the criteria of being cost-effective, energy-efficient, dependable, and easily

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

implementable. The incorporation of communication technology into the network architecture relies on its alignment with these essential attributes. These requirements are summarized in Table 1.

Table 1 – Amplifier Telemetry Requirements

communication technology objectives				
Cost	Power	Ease of implementation	Reliability and robustness	Performance
not a sizable part of amplifier itself **<$5**	Not noticeable in the amplifier bill of material <1 W	Integratable in new amplifiers, pluggable in existing amplifiers. No extra elements in headend	Independent of DOCSIS, communication under misaligned network conditions	Enough data throughput for access network details to be communicated to headend. Calculated to be ~3 kbs

The communication method of network optimization has been thoroughly researched to come up with a method that fits the objectives mentioned in Table 1. G3PLC powerline communications turned up as a cheap, easily implementable solution, that does not rely on in-band method of communication. G3PLC is a highly reliable mesh protocol for communication under extremely noisy conditions. It is an open international ITU standard and described in ITU-T G.9903, ITU-T G.9901 and IEEE 1901.2. based on narrowband power line communication (NB-PLC) orthogonal frequency division multiplexing (OFDM) technology based on state-of-the-art narrowband PLC standards.

The G3PLC standard specifies a link budget of at least 60 dB, the G3PLC community states that this in practice is >80 dB. Meaning between transmission and reception there can be 80 dB of loss before 50% of the messages are lost.

Figure 7 demonstrates a possible implementation method for PLC in HFC amplifiers.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

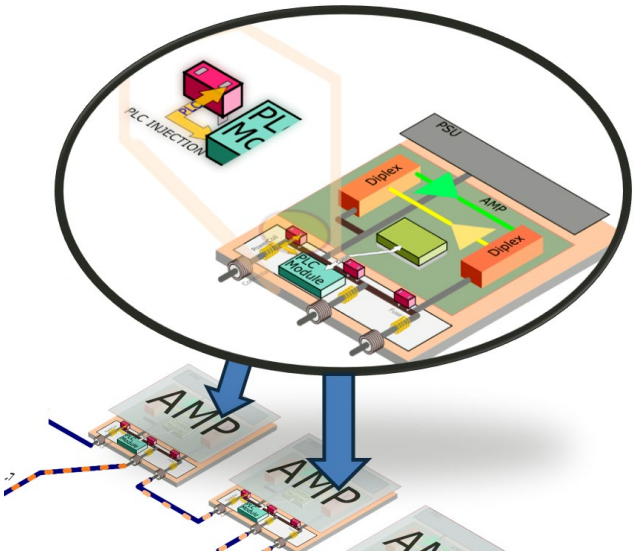


Figure 7 – PLC Implementation in HFC Amplifiers

Injection is done directly in the power passing section of the amplifier, thus bypassing all the RF circuitry, creating a communication path that is independent of existing spectrum plans. This is demonstrated in Figure 8.

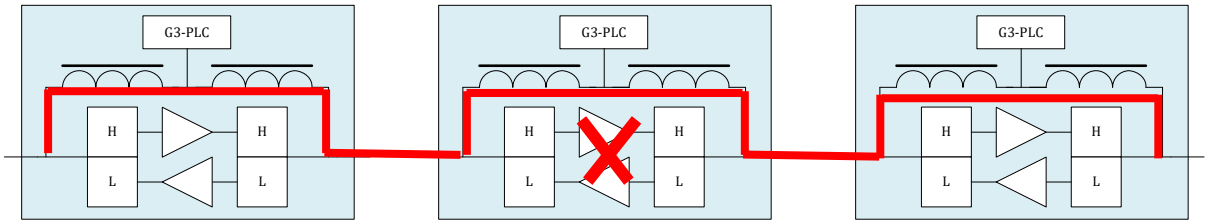


Figure 8 – Channel Loading Examples for Various Traffic Patterns

As shown in the middle amplifier in Figure 8, even if the amplifier themselves fail due to any reason, the communication path can still continue since only the power coils are used in the PLC communication path.

G3PLC uses OFDM modulation between 10 kHz and 490 kHz. Amplifiers can accommodate G3PLC modules so they can communicate to a centralized data polling and processing engine. Between each amplifier there can be many passives, which attenuate the signal. Often, specifications for loss below 500 kHz do not exist for passives. Thus, insertion loss has been measured on a many different passives (taps, splitters, cables). Additionally, PLC signal receive level and link quality were measured between PLC modules throughout a cascade of passives to determine signal degradation. The results are shown in Figure 9.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

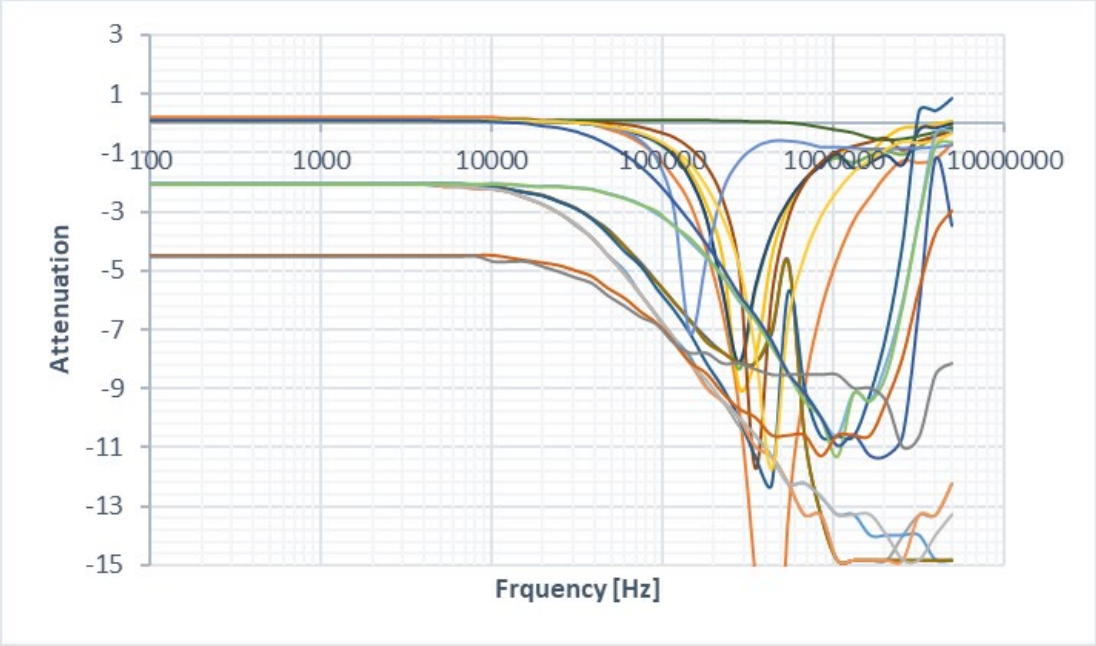


Figure 9 – Insertion Loss of 20 Different Passive Devices from 100 Hz to 5 MHz

3.3 Use Cases

The sections below discuss various applications and use cases of an AI-driven and optimized HFC system.

3.3.1 Carrier Expansion and Reduction

Operators may have the need to add or remove channels in response to fluctuations in bandwidth utilization, a prime example being the variation between low-demand periods like 2 AM and high-traffic times like 5 PM. However, this strategic maneuver, while effective, introduces a subtle realignment challenge to previously well-tuned networks. The arises from the modification of the spectrum plan, for which tilt and gain parameters of amplifiers had been calibrated. Consequently, with the expansion of the spectrum due to channel adjustments, the Rx (receive) levels shift to higher values. This transformation proves problematic when the receiving modems require a close-to-uniform input signal, as the spectrum's altered boundaries can potentially lead to an elevated Rx level, thus impacting network performance.

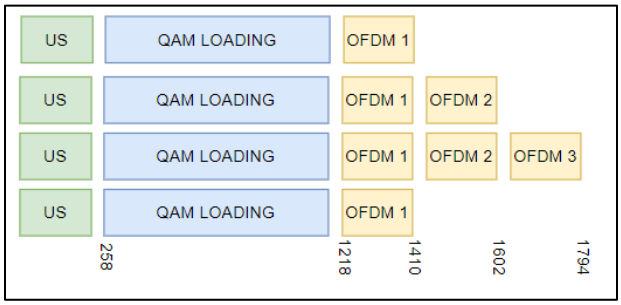


Figure 10 – Channel Loading Examples for Various Traffic Patterns

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 10 demonstrates various channel plans as traffic patterns change in the field. From top to bottom, it can be observed that more OFDM carriers can be deployed as traffic demand grows before reducing the channel loading to 1400 MHz when there is lower demand. This of course is an example of what could happen in the field and is very operator dependent. Table 2 outlines the total composite power (TCP) of each amplifier based on the channel loading. The TCP reduction is primarily caused by the reduction in channel loading, but the output level and tilt can be further enhanced by the optimization algorithm, rather than the blanket output levels.

It should be noted that in this example we are assuming that the amplifiers are optimized based on the span characteristics that they are exposed to. In reality, based on today's practices, these results will be much more exaggerated since every amplifier has the same rated output power regardless of the unique characteristics of each segment.

Table 2 – Amplifier Alignment – Channel Loading Changes

	AMP 1		AMP 2		AMP 3		AMP 4		AMP 5	
Amplifier TCP [dBmV]	**Before**	**After**	**Before**	**After**	**Before**	**After**	**Before**	**After**	**Before**	**After**
1 OFDM block	*61.8*		*46.3*		*58.9*		*57.8*		*53.3*	
1 to 2 OFDM blocks	*65.0*	*64.1*	*48.8*	*47.9*	*62.1*	*61.1*	*60.3*	*59.8*	*55.7*	*55.1*
2 to 3 OFDM blocks	*67.1*	*66.4*	*49.9*	*49.2*	*64.2*	*63.1*	*62.2*	*61.4*	*57.6*	*56.7*
3 to 1 OFDM blocks	*61.0*	*61.8*	*45.6*	*46.3*	*57.6*	*58.9*	*57.0*	*57.8*	*54.4*	*53.3*

The results of the system optimization have been demonstrated in Figure 11, Figure 12, and Figure 13.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

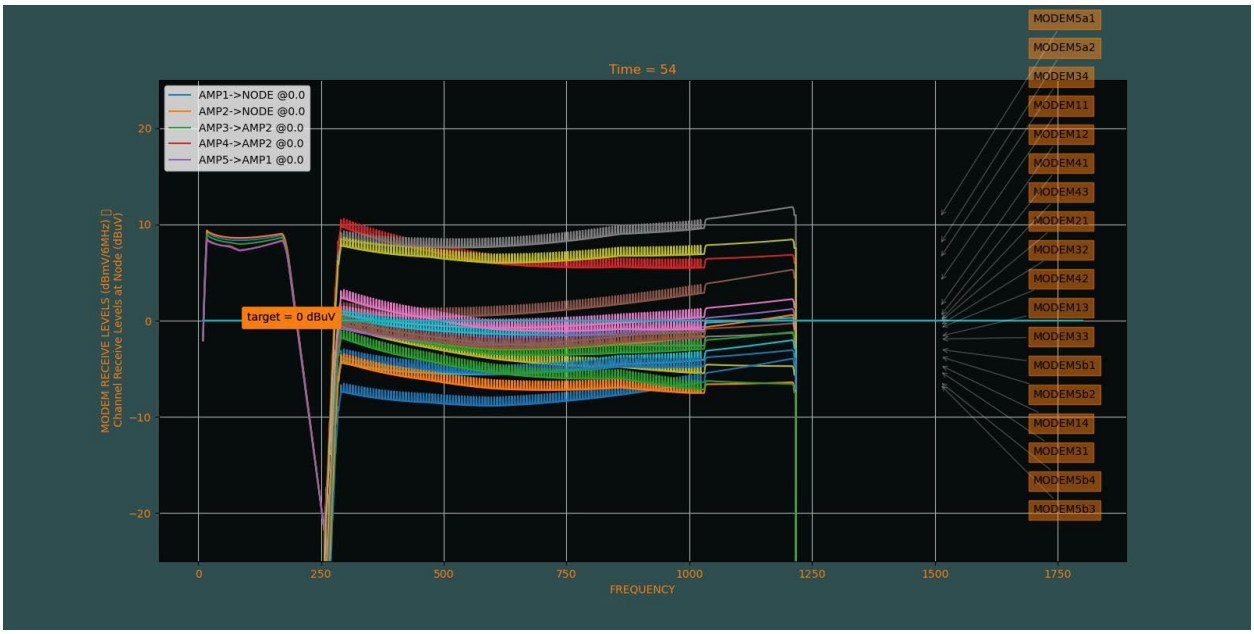


Figure 11 – System Operating at 204 MHz Split

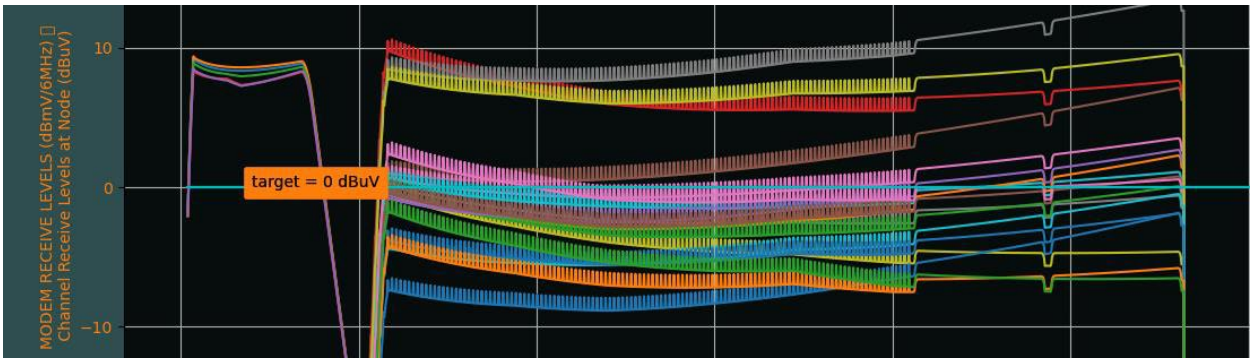


Figure 12 – Carriers Added with No Optimization

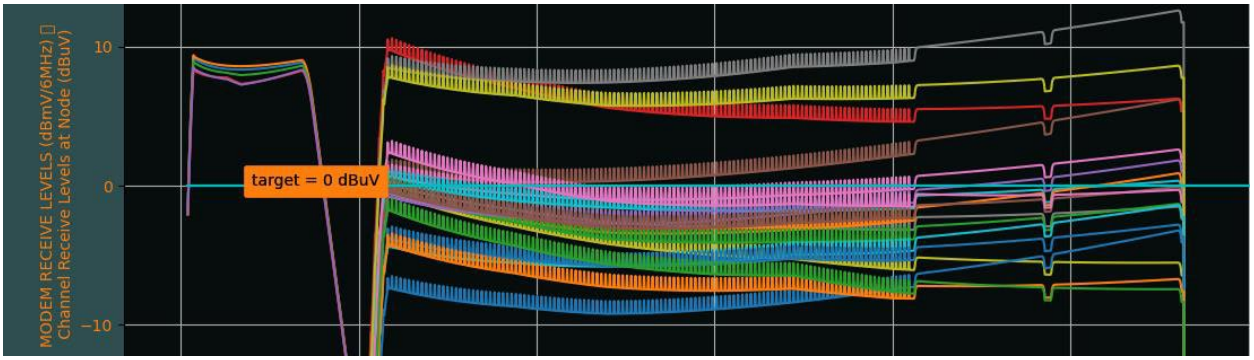


Figure 13 – Optimized System Post Carrier Expansion

It can be observed from Figure 11, Figure 12, and Figure 13 that the system can be further optimized based on the channel loading that each amplifier sees. As previously mentioned, this phenomenon is more exaggerated based on today's practices where every amplifier has the same output. In our example, even

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

with added or removed carries, we can observe how the system can be fine-tuned based on the unique characteristics of each span, rather than a blanket approach.

3.3.2 Split Change

This section explores network realignment following a split change. In the conventional static installation approach, the task of recalibrating each amplifier to accommodate the updated split configuration is requisite. Conversely, the subsequent scenario demonstrates the seamless automation of network alignment, rendering the process effortlessly orchestrated.

Figure 14 demonstrates the different splits that be deployed through time in a selected serving group. Table 3 demonstrates the difference in TCP before and after optimization for each split. This demonstrates why automation can further enhance deployment decisions since each segment and amplifier span needs to be optimized based on the end-to-end performance that is expected, rather than blanket output levels.

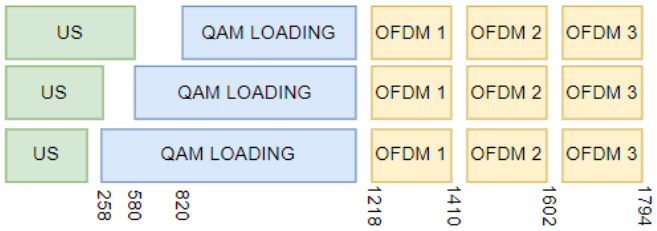


Figure 14 – Channel Loading Examples for Various Splits

Table 3 – Amplifier Alignment – Split Change

TCP [dBmV]	AMP 1		AMP 2		AMP 3		AMP 4		AMP 5	
	Before	After	Before	After	Before	After	Before	After	Before	After
204-258	*65.9*		*48.6*		*62.4*		*61.0*		*56.6*	
204-258 to 492-580	*63.7*	*66.4*	*46.6*	*48.8*	*59.3*	*63.0*	*57.6*	*61.4*	*53.1*	*56.6*
492-580 to 684-820	*64.4*	*66.4*	*47.2*	*48.7*	*60.3*	*63.1*	*58.23*	*61.3*	*53.2*	*56.4*

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 4 – Average CM Rx Level Before and After

Average modem Rx level [dBmV]	**Average modem Rx level Before Alignment**	**Average modem Rx level After Alignment**
204-258	0.4	
204-258 to 492-580	-10	+0.2
492-580 to 684-820	-5.1	+0.2

It is important to highlight that Table 3 shows a fundamental rationale underscoring the value automation offers to operators. With each split change, the task of realigning and optimizing the entire system becomes evident. Even in scenarios where the amplifiers feature switchable diplex filters, the process is not as straightforward as merely adjusting the split. Rather, a comprehensive optimization is necessitated, from the headend through the amplifiers and finally at the cable modem, to ensure optimal network performance. In contrast, system level optimization and automation can seamlessly overcome this challenge, while monitoring areas where performance is sub-optimal that may need manual intervention. This is done by optimizing output power and TCP based on the unique characteristics of each segment depending on the split that is deployed at the time.

Table 4 further drives the point by showing the convergence of Rx levels by the modems. In the blanket approach the ranges and receive levels can be sub-optimal in comparison to a fine-tuned and optimized system, as shown in the second column of the table.

Figure 15 and Figure 16 summarize the analyses in this section. Figure 15 shows a system operating at 204 MHz split while Figure 16 demonstrates the same system operating at 684 MHz split. It can be observed that with changes of splits, the results converge to the extrapolation of where each Rx level was prior to the split change, which demonstrates the high level of confidence in the optimization algorithm.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

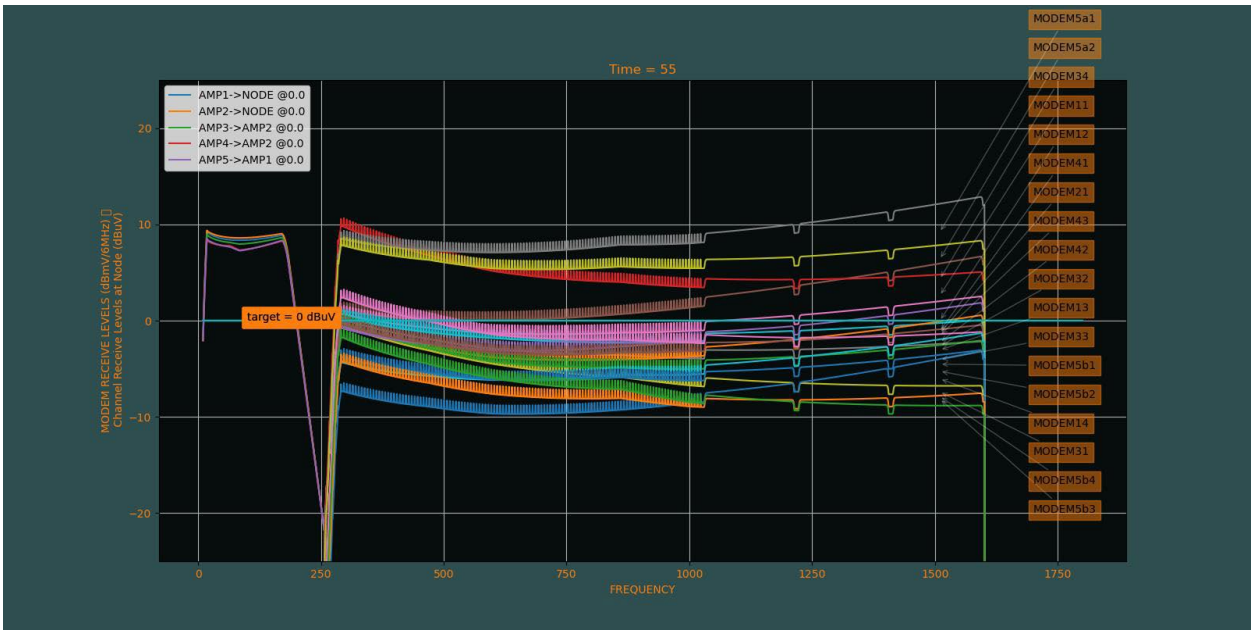


Figure 15 – 1.8 GHz System Operating at 204 MHz Split

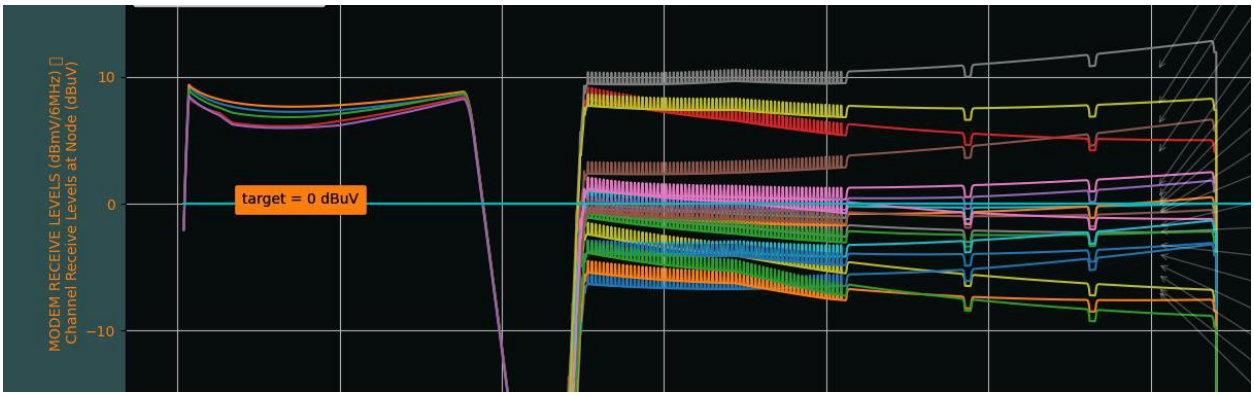


Figure 16 – The Same System Operating at 492 MHz Split – Post Optimization

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.3. Power Optimization

Amplifiers in a network have always operated utilizing a single output setting, in both forward and return. They are also constantly operating at full power to provide the maximum gain they can to feed signals to the subsequent receiving devices, whether amplifiers or modems. The amplifiers today are balanced at a certain voltage and current, which is referred to the bias point. This bias point ultimately defines the power consumption of the amplifier.

The performance of the amplifier is characterized by the carrier-to-interference-noise ratio (CINR) curve. In Figure 17, the y axis is CINR, a measure of signal quality. The x axis often represents the output power of the amplifier or the TCP. The more power the amplifier consumes, the wider this curve becomes. In other words, with more power the quality can become better.

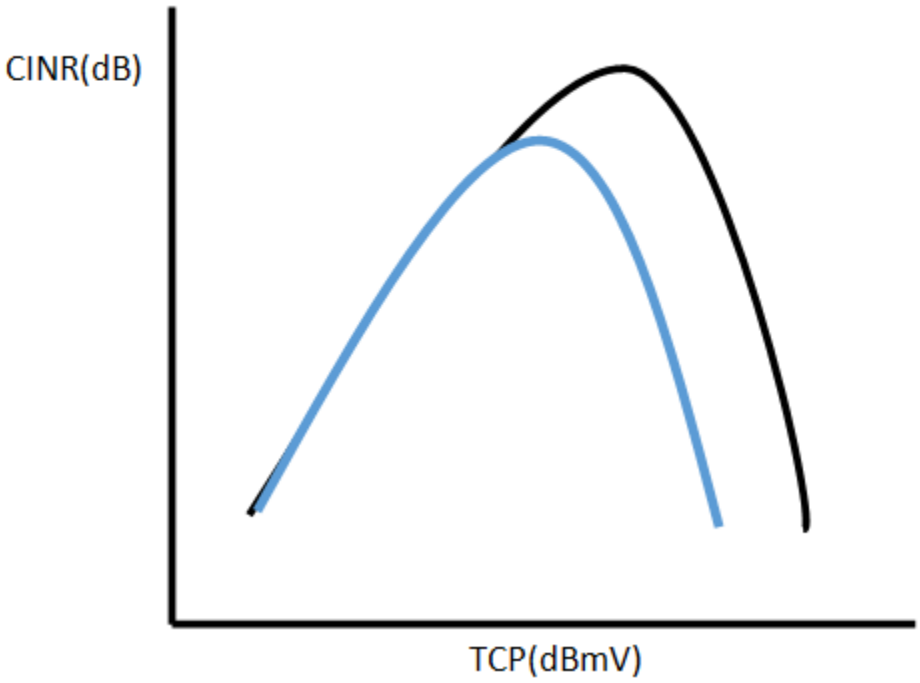


Figure 17 – Example Amplifier CINR Curve

In Figure 17, the blue line represents an amplifier operating with low power and the black line represents an amplifier operating with high power.

State of the art amplifiers have the possibility of changing the power consumption depending on the performance that is required by the amplifier chain. As an example, let us assume that a cable operator is willing to turn off the last OFDM block from 1 GHz to 1.2 GHz during off-peak hours. Figure 18 and Figure 19 demonstrate the required TCP for amplifiers in cascade, depending on the channel loading. The green line represents the TCP the amplifier needs to provide and the orange dashed line represents the required CINR performance. In our example, Figure 19 demonstrates a scenario where the last OFDM block is turned off and therefore the green line is shifted to the left. The amplifier can now operate in low power mode since it can be seen that the intersection of the green TCP threshold and the blue CINR curve are above the performance threshold shown by the orange dashed line.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

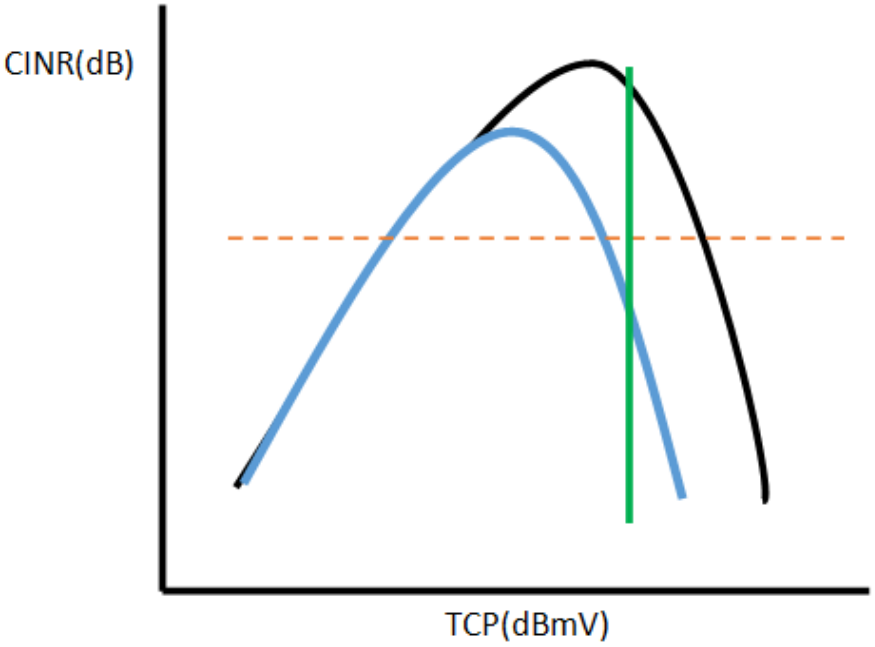


Figure 18 – Amplifier CINR and Required TCP – Full Channel Loading

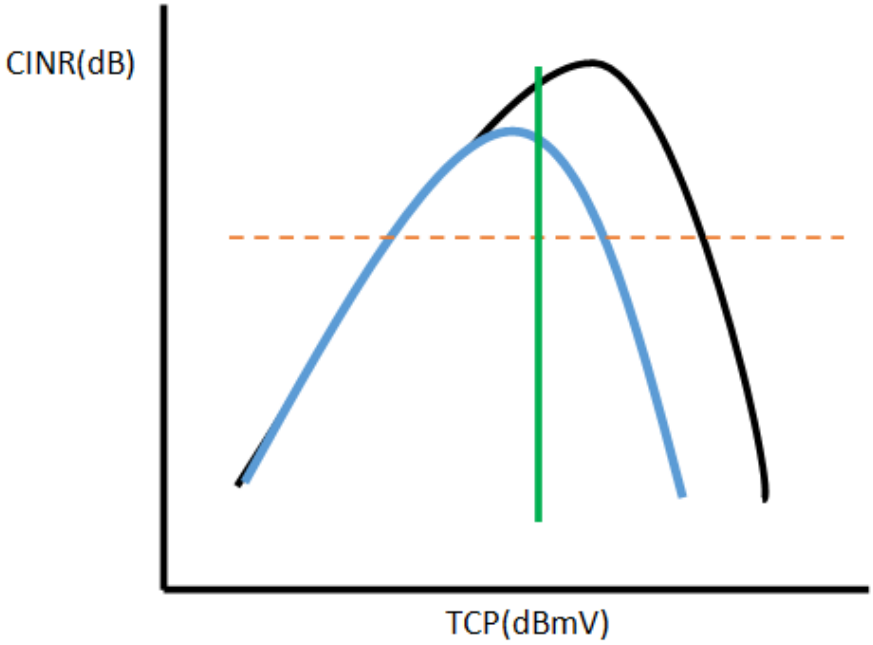


Figure 19 – Amplifier CINR and Required TCP – Reduced Channel Loading

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Let us analyze a real-life example. Figure 20 shows an example cascade of amplifiers.

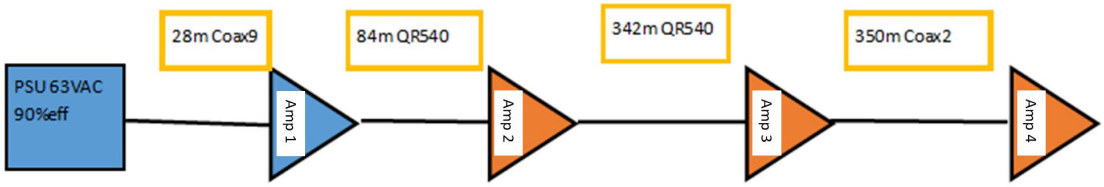


Figure 20 – Amplifer Cascade Analyzed for Power Saving Mode

Amplifier 1 is a three-output amplifier with the following power consumption:

- High power: 34 W
- Low power: 30 W

Amplifiers 2-4 are line extender amplifiers with the following power consumptions:

- High power: 19 W
- Low power: 17 W

The power consumption of this cascade at the mains in high power mode is 118.4 watts, as shown in Figure 21.



Figure 21 – Amplifer Cascade High Power Mode Consumption

In contrast Figure 22 demonstrates the same cascade's consumption at the mains in low power mode is 97.3 watts.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 22 – Amplifer Cascade Low Power Mode Consumption

A total saving of 20 W has been shown in this example cascade of amplifiers. Assuming that the node has four identical output legs, the saving could be 80 W when all amplifiers are operating in low power mode.

The theoretical power reduction should be 10 W (4 W in three output amplifiers and 2 W in each line extender). The remainder of the savings are from the power dissipated in the cable by I^2*R losses. Low power mode reduces the current in the network and therefore a reduction of the power wasted in the cables is realized.

4. Conclusion

The landscape of HFC networks is undergoing rapid transformation, particularly with the impending mass adoption of next-generation DOCSIS standards and spectrum expansion. These networks, which synergize optical fiber and coaxial cable technologies, require increasing intricacies in design and maintenance. Among the foremost challenges is the web of interdependencies that connect various departments engaged in the process of designing and maintaining HFC networks. The alignment of functions across network planning, engineering, and operations is now more crucial than ever to ensure seamless operation of the networks.

This paper navigated the potential of network automation and software-defined solutions to assuage a majority of these concerns, both in plant upgrades and ongoing maintenance. A central theme of this exploration is the transformative power of artificial intelligence in reshaping the entire landscape of end-to-end design and maintenance within the HFC network domain. The deployment of AI opens avenues for reimagining conventional methodologies governing the optimization of nodes and amplifiers, overall capacity management, and power optimization. Employing AI-driven algorithms and predictive analytics offers a revolutionary method for tackling the challenges to these areas, promising higher efficiency, reliability, and adaptability.

Linked with the potential benefits of AI in HFC network design are the various impacts that extend across various facets of network management. The advantages span from substantial cost savings stemming from optimized resource allocation, to enhanced decision-making through data-driven insights provided by AI models. Moreover, the integration of AI stands to enhance the plant reliability of HFC networks, as predictive maintenance and anomaly detection algorithms can proactively identify and mitigate potential issues.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AI	artificial intelligence
CCN	carrier-to-composite noise
CNR	Carrier-to-noise ratio
CIN	composite intermodulation noise
CINR	carrier-to-interference-noise ratio
CM	cable modem
CSO	composite second order
CTB	composite triple beat
dB	decibels
dBmV	decibels relative to one millivolt
DOCSIS	Data-Over-Cable Service Interface Specification
GHz	gigahertz
HFC	hybrid fiber-coax
MER	modulation error ratio
MHz	megahertz
NB-PLC	narrow band powerline communication
OFDM	orthogonal frequency division multiplexing
PDF	probability density function
PLC	powerline communication
QAM	quadrature amplitude modulation
RPD	remote PHY device
RF	radio frequency
Rx	receive
SCTE	Society of Cable Telecommunications Engineers
SNR	Signal-to-noise ratio
TCP	total composite power

Bibliography & References

Data-Over-Cable Service Interface Specification DOCSIS 4.0 – *Physical Layer Specification CM-SP-PHYv4.0*

Broadband Cable Access Networks – The HFC Plant, David Large and James Farmer

https://iamtrask.github.io/2015/07/27/python-network-part2/

http://neuralnetworksanddeeplearning.com/chap2.html

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Full Duplex DOCSIS (FDX) Amplifier Automatic Configuration

A Technical Paper prepared for SCTE by

Richard S. Prodan, Ph.D.
Engineering Fellow
Comcast Cable
Denver, Colorado
+1 (303) 372-2374
rich_prodan@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. FDX Amplifier Design 4
3. FDX Initial Amplifier Install and Setup 5
   3.1. Define Amp Configuration Files 5
   3.2. Initial Amp Setup to Enable SoC eCM Activation 6
      3.2.1. Initial Boot-up and Downstream Input Setup 6
      3.2.2. Legacy Upstream Input Setup 6
4. SoC Downstream Gain and Equalization Optimization 7
   4.1. Initial Downstream Output Setup for Mid-Split Operation 7
   4.2. Downstream Output Optimization using SoC Downstream Equalizer 9
5. SoC Upstream Gain and Equalization Optimization 11
   5.1. Initial Upstream Output Setup for FDX Operation 11
   5.2. Upstream Input Optimization using SoC Upstream Equalizer 12
6. Network Amplifier Topology Discovery 14
7. Final Downstream and Upstream Optimization of All Amps 18
8. Example Auto Setup of Cascaded FDX Amplifier Networks 18
   8.1. Single Family Unit (SFU) Model of a Node + 6 Amp Cascade 19
   8.2. Multiple Dwelling Unit (MDU) Model of a Node + 6 Amp Cascade 23
9. Conclusion 26

Abbreviations 27

Bibliography & References 27

List of Figures

Title **Page Number**

Figure 1 – FDX Amplifier Functional Diagram 5
Figure 2 – FDX Amplifier Configuration File Paths 5
Figure 3 – Initial Boot-up and Downstream Input Setup 6
Figure 4 – Legacy Upstream Input Setup 6
Figure 5 – Initial Downstream Output Setup for Mid-Split Operation 7
Figure 6 – Example of Initial Downstream Setup: 8
Figure 7 – Downstream Output Optimization using SoC Downstream Equalizer 9
Figure 8 – Example of Downstream Output Optimization Using SoC Downstream Equalizer : (a) Node/Amp Downstream Transmit Level, Inter-Amp Span Loss, and Amp Downstream Equalization; (b) Amp Downstrean Receive Level, Amp Downstream Equalization and Op Gain, and Node and Amp Downstream Transmit Levels 10
Figure 9 – Initial Upstream Output Setup for FDX Operation 11
Figure 10 – Example of Initial Upstream Output Setup for FDX Operation: (a) Inter-Amp Span Loss, and Amp Upstream Equalization; (b) Amp Upstream Receive Level, Amp Upstream Equalization and Op Gain, and Amp Upstream Transmit Level 12
Figure 11 – Example of Upstream Input Optimization Using SoC Upstream Equalizer: 14
Figure 12 – FDX Amplifier With Topology Discovery Mid-Split Upstream Signal Paths 15
Figure 13 – Example Tree and Branch Cable Network 16
Figure 14 – Matrix Describing the Tree and Branch Cable Network of Figure 13 17
Figure 15 – Set SoC Transmit Spectrum Generator to Check Amp/Node Receive Level 18

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 16 – Example Single Family Unit (SFU) Model for a Node + 6 Amp Cascade 19
Figure 17 – SFU Node Port Input/Output and Echo Interference Levels 20
Figure 18 – SFU Amp 1 Initial Downstream Setup 20
Figure 19 – SFU Amp 1 Final Optimized Downstream Setup 20
Figure 20 – SFU Amp 6 Initial Downstream Setup 21
Figure 21 – SFU Amp 6 Final Optimized Downstream Setup 21
Figure 22 – SFU Amp 1 Initial Upstream Setup 22
Figure 23 – SFU Amp 1 Final Optimized Upstream Setup 22
Figure 24 – Example Multiple Dwelling Unit (MDU) Model for a Node + 6 Amp Cascade 23
Figure 25 – MDU Node Port Input/Output and Echo Interference Levels 23
Figure 26 – MDU Amp 1 Initial Downstream Setup 24
Figure 27 – MDU Amp 1 Final Optimized Downstream Setup 24
Figure 28 – MDU Amp 6 Initial Downstream Setup 25
Figure 29 – MDU Amp 6 Final Optimized Downstream Setup 25
Figure 30 – MDU Amp 1 Initial Upstream Setup 26
Figure 31 – MDU Amp 1 Final Optimized Upstream Setup 26

List of Tables

Title **Page Number**

Table 1 – Network Topology Describing the Tree and Branch Cable Network of Figure 13 17

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Full Duplex Data-Over-Cable Service Interface Specifications (DOCSIS®)(FDX) technology has been realized in a custom system on a chip (SoC) successfully implemented in an operational FDX remote PHY (R-PHY) node design. A new implementation of this SoC is also being adapted for use in an FDX amplifier utilizing the same echo cancellation technology employed in the node.

FDX amps will begin to replace conventional mid-split amps and will support concurrent bidirectional communication within the same bandwidth. These amps are more complex to configure for FDX operation. Completely new and more complicated procedures for configuring these amplifiers are needed to install these amps. Additionally new procedures to maintain proper operation of these new amplifiers will be needed.

Embedded digital signal processing within DOCSIS 4.0 FDX amplifiers provides the ability to automatically configure the amp for FDX operation, replacing the manual alignment process in conventional amps today. Optimal results for cascades of amps can be obtained resulting in near perfect amp alignment that is otherwise not possible.

The heart of these FDX amps use digital signal processing (DSP) in a custom SoC that digitizes all inbound and outbound signals with flexible processing of the required amplifier gain, equalization, and echo cancellation.

Several test points and a wideband Fast Fourier Transform (FFT) processor for spectrum analysis are included in the SoC in addition to the required DSP for digital filtering, echo cancellation, and FFT signal processing. These digital test points can be processed by the wideband FFT for spectrum analysis. Such processed spectral data can be communicated through an embedded cable modem (eCM) in the SoC, as well as the loading of data as the result of processing this spectrum data for configuring proper amplifier operation. The alignment process for technicians can therefore be simplified through automation via a communication link between the amp and a configuration server using internal amp measurements that are described in this paper.

This paper discusses a method for automatically configuring an FDX amplifier for optimal operation using the spectral analysis capability and DOCSIS communication in the SoC. Additionally, a method for discovering the network topology of the amplifiers in a tree and branch network is derived. Such methodologies can also be used to initially set up and periodically check for continued proper operation of installed and configured amplifiers.

2. FDX Amplifier Design

A functional diagram of an FDX amplifier is shown in Figure 1. This amplifier functionally contains much of a conventional frequency division duplex (FDD) amplifier with the diplex filters replaced by directional couplers or splitters to enable simultaneous bidirectional spectral occupancy in the FDX band. The insertion of a SoC provides echo cancellation of the downstream echo into the upstream signal and DSP features such as digital equalization of upstream and downstream signals. A DOCSIS embedded cable modem (eCM) enables communication between the amp and a central controller (“Amp Central”) to provide network operational parameters and control amplifier setup and determination of network topology of the amplifier cascades within the tree and branch system architecture. The mid-split upstream path is separated via diplex filters and does not pass through the SoC to facilitate network connectivity of the eCM within the SoC.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

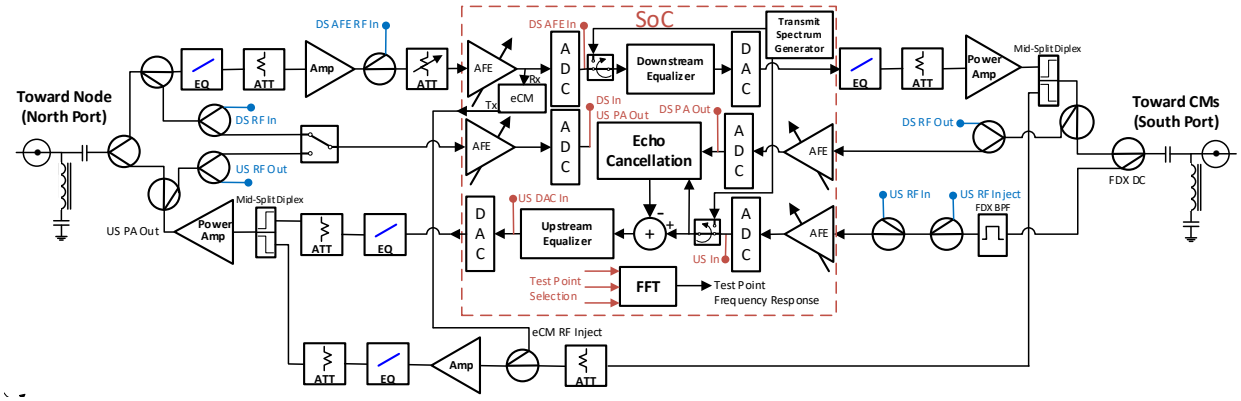


Figure 1 – FDX Amplifier Functional Diagram

An FFT for frequency response (FR) calculations of signal test points is provided within the SoC. Various test points within the SoC are identified in Figure 1. An FFT can be applied to any selected test point to calculate the test point signal FR. Such calculated FRs along with configuration file FRs of the paths between the amp ports and the SoC are used to determine automatic equalization DSP coefficients (in frequency and/or time domain) in the SoC to augment conventional gain and equalization within the amp signal paths.

3. FDX Initial Amplifier Install and Setup

3.1. Define Amp Configuration Files

Configuration file paths #1 through #7 are shown in Figure 2.

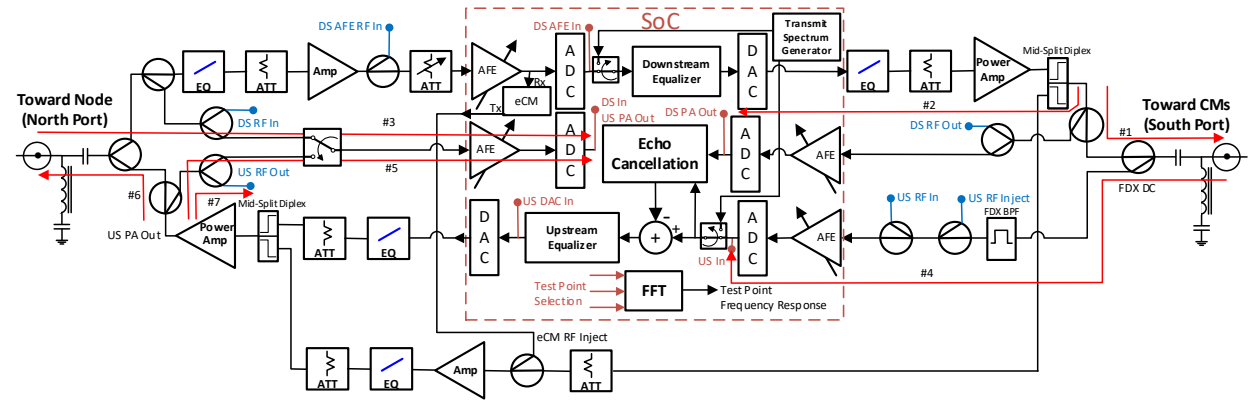


Figure 2 – FDX Amplifier Configuration File Paths

Every configuration file contains an FR (i.e., complex valued scattering parameter s21 with log magnitude < 0 dB) that was measured during the manufacturing process and subsequently stored in the amplifier's memory. By utilizing the configuration files, the levels at the amplifier ports are calculated based on the measurements of the SoC internal test point FR. These calculations, and the resultant calibration of the gains and equalization, both internal and external to the SoC, are discussed in subsequent sections.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.2. Initial Amp Setup to Enable SoC eCM Activation

3.2.1. Initial Boot-up and Downstream Input Setup

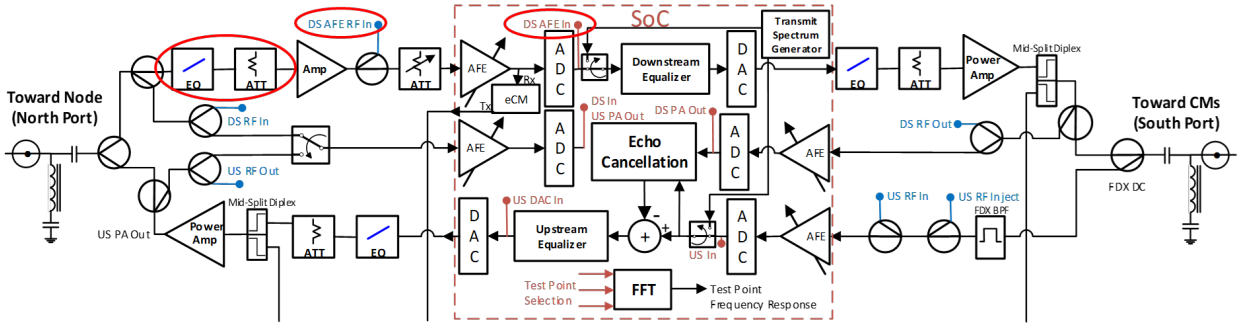


Figure 3 – Initial Boot-up and Downstream Input Setup

Assume that an existing amp is being replaced with the FDX amp shown in Figure 3. The following procedure will setup an initial downstream input to the FDX amp SoC and enable the eCM to establish connectivity with Amp Central as follows:

- Replace the amp housing (new) or RF lid (existing) with the AC power shunts initially removed, then reinstall after the replacement is complete
- Power-up and boot the RF board and SoC module to set the default state for the RF South Port digital-to-analog converter (DAC) attenuator and equalizer (EQ), and the SoC configuration default states for the North Port analog front end (AFE), South Port DAC, North Port DAC, South Port US and DS AFEs
- Set the ADU to manual
- Measure the spectrum and total composite power (TCP) at the **DS AFE RF In** test point with the (measurement) meter tool across test frequencies
- Calculate and set DS input attenuator (ATT) and equalizer (EQ) for approximately a flat level across frequency less than or equal to the TCP limit including the North AFE backoff attenuator (0 dB default) at the **DS AFE RF In** input for the best analog-to-digital converter (ADC) performance
- Measure and adjust the ADU pilot level attenuation in auto mode to match the manual mode

3.2.2. Legacy Upstream Input Setup

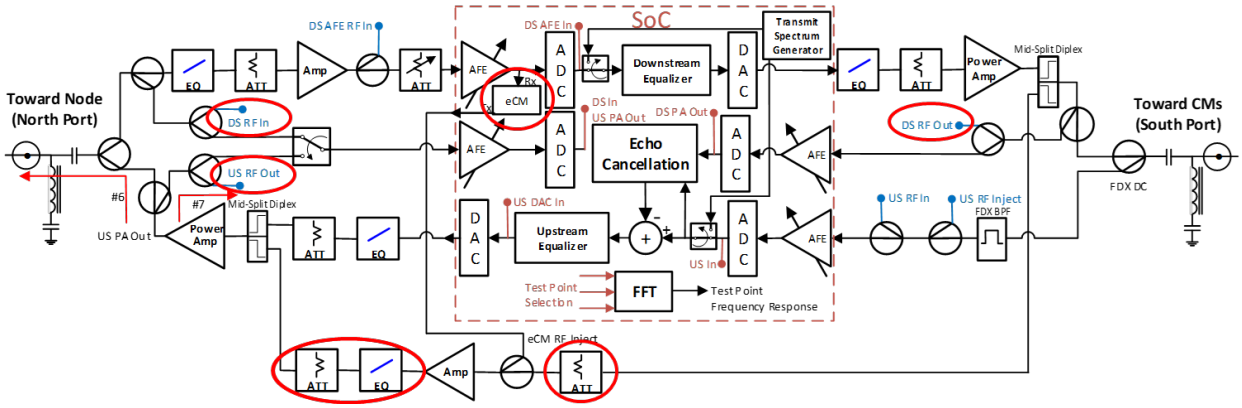


Figure 4 – Legacy Upstream Input Setup

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Setup the legacy mid-split upstream path of the FDX amp in Figure 4 as follows:

- Consult the design reference for return input attenuation value
- Set the default state for the legacy US pre-amp input attenuator, US power amp output attenuator, and US output EQ
  - Legacy upstream plug-ins could use previously installed amplifier values initially
- Connect the (measurement) meter tool eCM to the **DS RF In** (CM downstream input) and **DS RF Out** (CM upstream output) test points
- Check the **DS RF Out** test point legacy upstream FR with the (measurement) meter tool with the meter eCM injecting an upstream signal into **DS RF Out** and displaying the upstream spectrum on the meter
- Adjust the US power amp output attenuator and US output EQ to match the upstream FR design reference value at **DS RF Out**
  - **DS RF Out** FR = South Port upstream input FR + Config #1 FR – Config #2 FR[1]
  - Alternatively, define and store a single Config #1-2 FR = Config #1 FR – Config #2 FR

The SoC eCM then locks with a positive status indicator and establishes connectivity to Amp Central. Automatic calculation can proceed with this communication path established. Calculations and results can be handled in Amp Central or CPU processors in the amp or in the SoC (in addition to internal signal test point FFTs).

4. SoC Downstream Gain and Equalization Optimization

4.1. Initial Downstream Output Setup for Mid-Split Operation

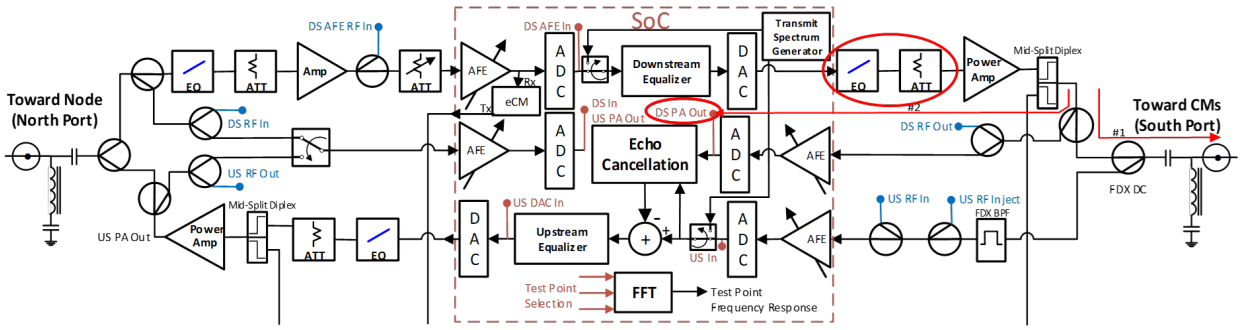


Figure 5 – Initial Downstream Output Setup for Mid-Split Operation

The eCM is online and can query Amp Central for downstream transmit output level and tilt as well as the upstream receive input level. Then, proceed to set the initial gain and tilt in the FDX amp downstream path shown in Figure 5, excluding the SoC, as follows:

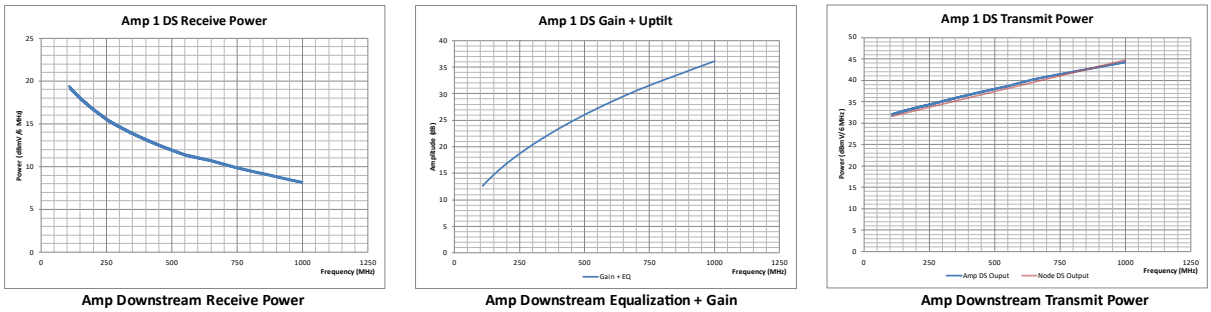
- Command the SoC downstream path gain and equalization to 0 dB flat across the downstream band
  - DS DAC input = **DS AFE In** (initially no influence by the SoC digital equalizer, just use conventional or electronic plug-ins)
- Measure the **SoC DS PA out** Frequency Response (FR) downstream output level and tilt values

[1] **DS RF Out** FR + Config #2 FR = South Port upstream input FR + Config #1 FR. So South Port upstream input FR = **DS RF Out** FR + Config #2 FR – Config #1 FR.

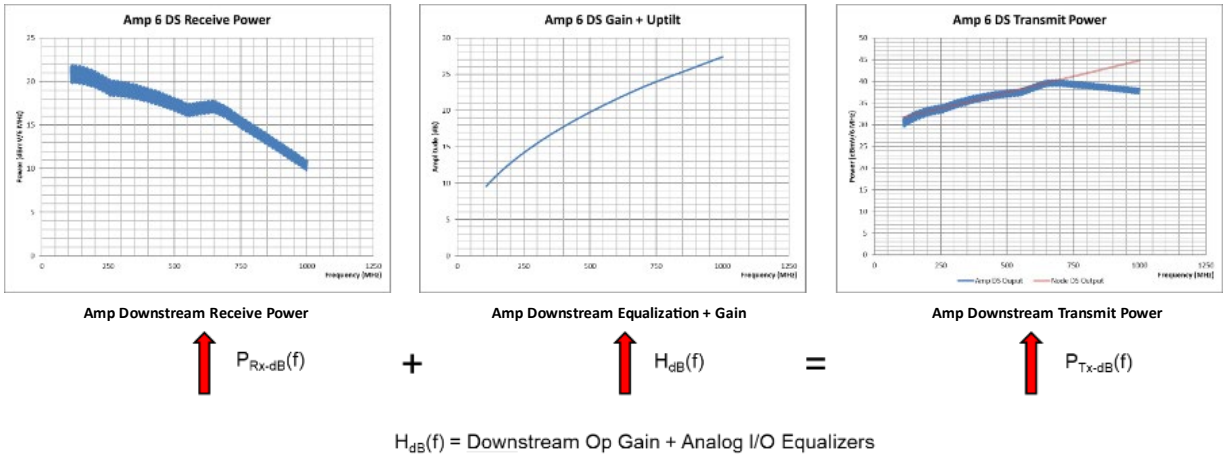
© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Calculate the South Port-to-**DS PA out** path loss FR = Config #2 FR – Config #1 FR[2]
  - Alternatively, use a single Config #2-1 FR = Config #2 FR – Config #1 FR = –Config #1-2 FR
- Calculate the South Port output FR = **DS PA Out** FR – South Port-to-**DS PA out** path loss FR
- Set the downstream output attenuator and equalizer to approximate the design output level and tilt
- Check the output level and tilt by remeasuring the **DS PA Out** FR and calculating the South Port output FR

The initial setup involves passing the signal through the SoC, which then establishes the downstream connection utilizing conventional attenuators and equalizers, similar to the approach followed in conventional (analog) amplifiers.



(a)



(b)

Figure 6 – Example of Initial Downstream Setup:
(a) Amp 1 Initial Downstream Setup, (b) Amp 6 Initial Downstream Setup

[2] The South Port output FR – Config #1 FR = Diplex common port FR, and the Diplex common port FR + Config #2 FR = **DS PA Out** FR. So the South Port output FR – Config #1 FR = **DS PA Out** FR – Config #2 FR, or the South Port output FR = **DS PA Out** FR – (Config #2 FR – Config #1 FR).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The following network simulations are calculated using the methodology in [1]. As shown in Figure 6, this approximates the design level and tilt design profile and passes any input linear distortion (amplitude variation) to the output signal. The following section removes these artifacts.

4.2. Downstream Output Optimization using SoC Downstream Equalizer

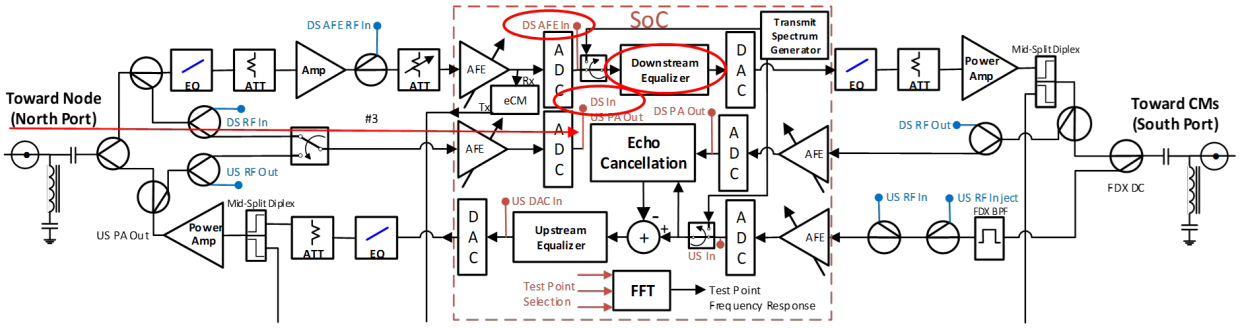


Figure 7 – Downstream Output Optimization using SoC Downstream Equalizer

In this section, the design level and tilt design profile are restored and any input linear distortion (amplitude variation) to the output signal is removed as follows:

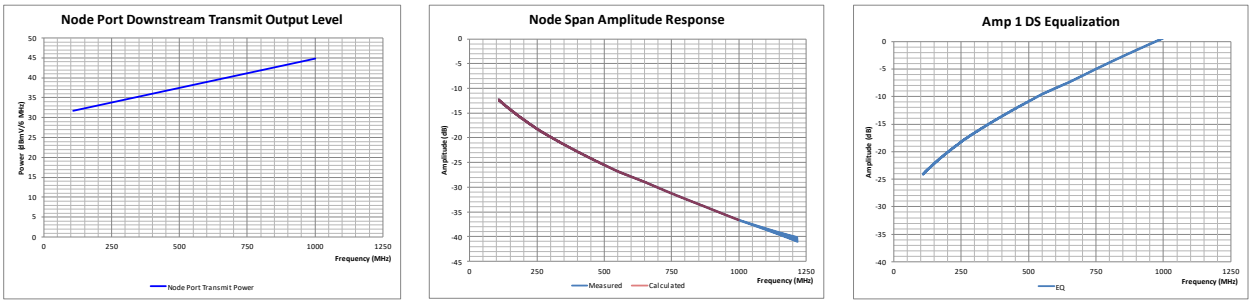
- Measure the **SoC DS in** FR and **SoC DS AFE in** FR
- Set the North Port-to-**DS in** FR = Config #3 FR
- Calculate the North Port DS input FR = **DS in** FR – North Port-to-**DS in** FR
- Calculate the Inter-Amp Span Loss FR = North Port DS input FR – Node/Amp output design FR
  - Node/Amp output design FR = Node/Amp downstream output design level and tilt FR
- Calculate the North Port-to-**DS AFE in** FR = **DS AFE in** – North Port DS input FR
- Calculate the SoC DS DAC Out-to-South Port FR = **DS AFE in** – South Port output FR
  - SoC Downstream Equalizer = 0 dB in FDX band → DS DAC input = **DS AFE in**
- Calculate DS Equalization + Gain FR = –Inter-Amp Span Loss FR (unity gain)
- Calculate DS Op Gain (> 0 dB) for DS Equalizer FR (≤ 0 dB)
  - DS Op Gain = |DS Equalizer + Gain FR @ maximum DS frequency|
- Calculate DS Equalization FR = –(DS Op Gain + Inter-Amp Span Loss FR)
- DS Equalization FR + DS Op Gain = North Port-to-**DS AFE in** FR + SoC Downstream Equalizer FR + DS DAC input-to-South Port FR
- Set SoC Downstream Equalizer FR = DS Equalization FR + DS Op Gain – North Port-to-**DS AFE in** FR – DS DAC input-to-South Port FR
- FDX downstream path configuration is complete

An example depicting the DS Equalization + Gain FR, the Op Gain, and DS Equalization FR calculation steps above that are needed for determining the SoC DS Equalizer FR is shown in Figure 8.
The node transmitted downstream output FR is shown in the left plot of Figure 8a. The amp received downstream input FR is shown in the leftmost plot of Figure 8b. The DS Equalization + Gain FR and DS Op Gain are shown in the center plot of Figure 8b. The output FR (i.e., the level and tilt) matches that of the negative inverse of the inter-amp span loss in the center plot of Figure 8a. The downstream equalization calculation without the op gain is shown in the right plot of Figure 8a. The amp downstream

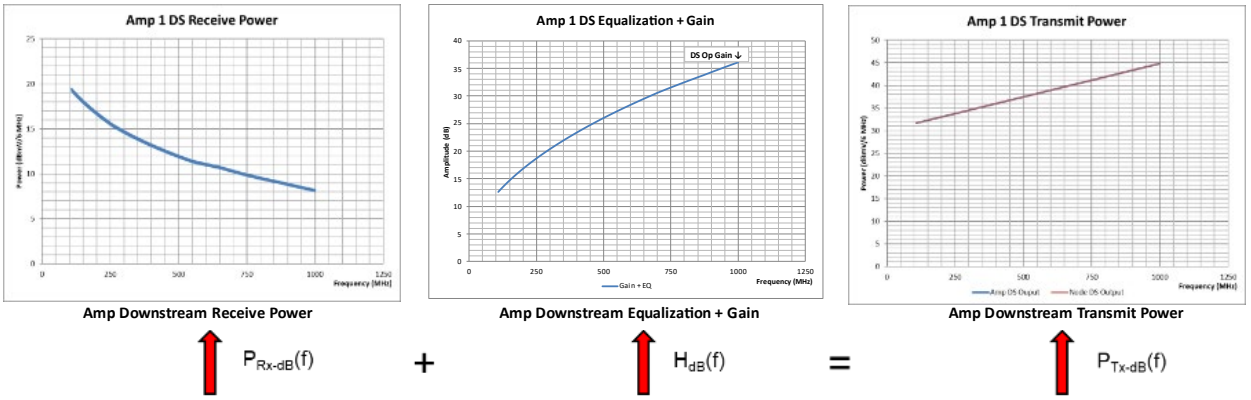
© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

output transmit FR is shown compared to the node downstream output transmit FR in the right plot of Figure 8b. The downstream output FR (i.e., the level and tilt) matches that of the node with a final 37 dB amp downstream operational gain.

This final setup with activation of the SoC downstream (digital) equalizer and gain setting achieves the design level and tilt design profile and removes any input linear distortion (amplitude variation) to the output signal. Examples demonstrating this result are shown in Section 8 Example Auto Setup of Cascaded FDX Amplifier Networks.



(a)



(b)

Figure 8 – Example of Downstream Output Optimization Using SoC Downstream Equalizer : (a) Node/Amp Downstream Transmit Level, Inter-Amp Span Loss, and Amp Downstream Equalization; (b) Amp Downstrean Receive Level, Amp Downstream Equalization and Op Gain, and Node and Amp Downstream Transmit Levels

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. SoC Upstream Gain and Equalization Optimization

The following sections describe a similar process for the upstream. This is possible because the upstream and downstream signals share the same spectrum and experience equivalent amp span loss from the amp North Port to the Node Port, thanks to the symmetry in path loss between the two directions.

5.1. Initial Upstream Output Setup for FDX Operation

The initial upstream output setup procedure of the FDX amp shown in Figure 9 is described in the following:

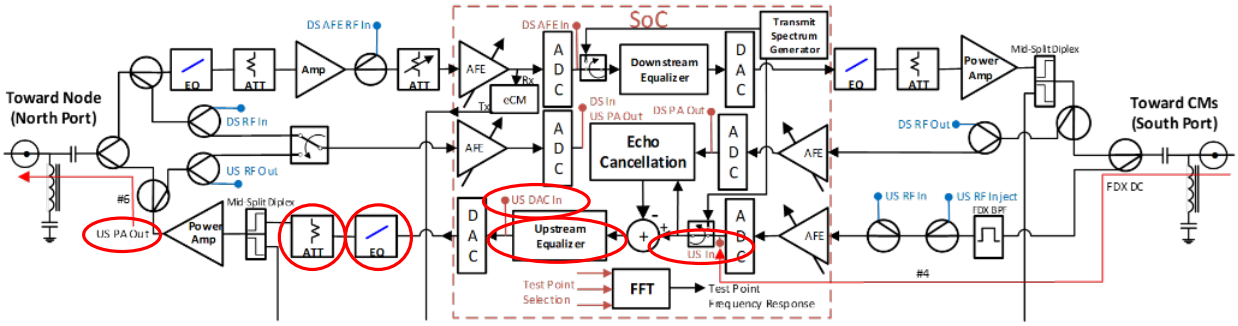
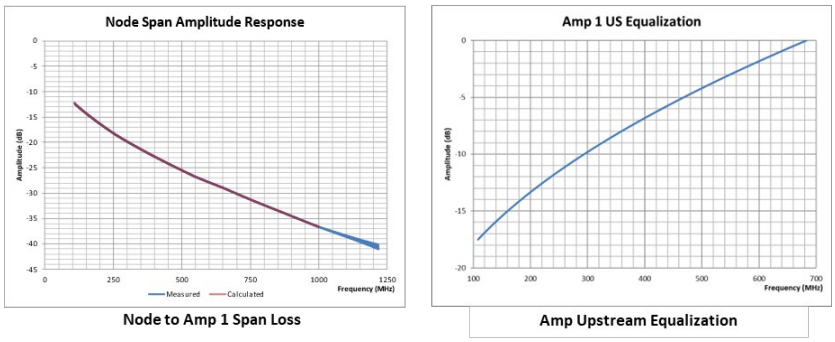


Figure 9 – Initial Upstream Output Setup for FDX Operation

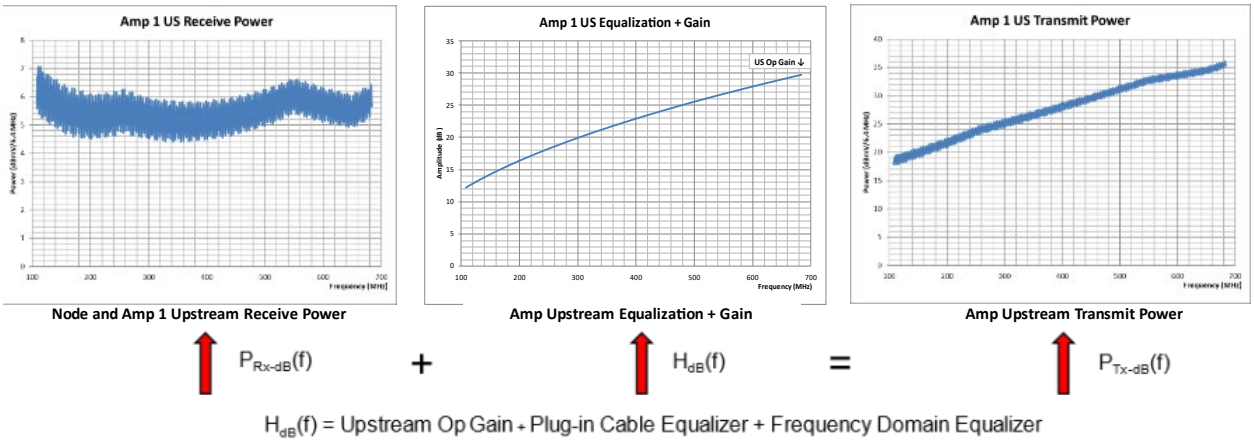
- Command the SoC upstream path gain to 0 dB flat across the upstream band
  - **US in** to **US DAC in**
  - Initially there is no influence by the SoC digital equalizer, just use conventional or electronic plug-ins
- Set the South Port-to-**SoC US in** FR = Config #4 FR
- Set the US PA out-to-North Port FR = Config #6 FR
- Calculate South Port to North Port Amp FR = –Inter-Amp Span Loss FR (symmetric insertion loss)
- Calculate US Equalization + Gain FR = –Inter-Amp Span Loss FR (unity gain)
- Calculate US Op Gain (> 0 dB) for US Equalizer FR (≤ 0 dB)
  - US Op Gain = |US Equalizer + Gain FR @ maximum US frequency|
- Calculate US Equalization FR = –(US Op Gain + Inter-Amp Span Loss FR)
- Calculate US Power Amp Gain = US Op Gain – South Port-to-SoC US in FR – US PA out-to-North Port FR and set upstream output attenuator
- Select plug-in or set electronic US EQ FR ≈ US Equalization FR (match cable EQ uptilt)

An example depicting the US Equalization + Gain FR, the Op Gain, and US Equalization FR calculation steps above that are needed for determining the SoC US Equalizer FR is shown in Figure 10. The US Equalization + Gain FR and US Op Gain are shown in the center plot of Figure 10b. The output FR (i.e., the level and tilt) approximately matches that of the negative inverse of the inter-amp span loss in the center plot of Figure 10a. The upstream equalization calculation without the op gain is shown in the right plot of Figure 10a. The amp upstream output transmit FR is shown in the right plot of Figure 10b. The amp received upstream input FR is shown in the left plot of Figure 10b. The node port upstream input level to be matched at the amp upstream input (South) port is 5 dBmV/6.4 MHz.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



(a)



(b)

Figure 10 – Example of Initial Upstream Output Setup for FDX Operation: (a) Inter-Amp Span Loss, and Amp Upstream Equalization; (b) Amp Upstream Receive Level, Amp Upstream Equalization and Op Gain, and Amp Upstream Transmit Level

This approximates the upstream input design level and passes any input linear distortion (amplitude variation) to the output signal. A cable modem transmitting towards the amp south port will transmit with pre-equalization to be received at the node port with a flat level without linear distortion. The upstream input FR (i.e., the level and tilt) approximates that of the node since the amp has not applied upstream equalization in the SoC to reduce the CM linear distortion due to pre-equalization. This is discussed in the next section.

5.2. Upstream Input Optimization using SoC Upstream Equalizer

Completion of the upstream input optimization is accomplished by setting the SoC upstream equalizer to correct the deviation of the US EQ FR from the ideal US Equalization FR calculated in the previous section as follows:

- SoC Upstream Equalizer FR + US EQ FR = US Equalization FR

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Set SoC Upstream Equalizer FR = US Equalization FR – US EQ FR
  - SoC corrects any deviation of US EQ FR from US Equalization FR
- FDX upstream path configuration is complete

Figure 11 shows the effect of the SoC upstream equalizer on the previous example. The US Equalization + Gain FR and US Op Gain are shown in the center plot of Figure 11b. The output FR (i.e., the level and tilt) now matches that of the negative inverse of the inter-amp span loss in the center plot of Figure 11a. The upstream equalization calculation without the op gain is shown in the right plot of Figure 11a. The amp upstream output transmit FR is shown in the right plot of Figure 11b. The amp received upstream input FR is shown in the left plot of Figure 11b. The node port upstream input level to be matched at the amp upstream input (South) port is 5 dBmV/6.4 MHz.

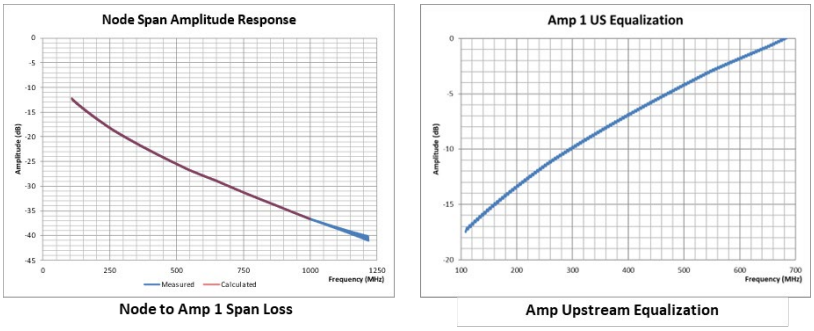
The final upstream equalization calculation is shown in Figure 11a. The amp upstream input FR is shown in the leftmost plot of Figure 11b. The initial US Equalization + Gain FR is shown in the center plot. The amp upstream output transmit FR is shown in the rightmost plot. The upstream input FR (i.e., the level) is high due to a low initial 0 dB amp gain.

The final upstream equalization calculation is shown in Figure 11b. The final US Equalization + Gain FR is shown in the center plot. The amp upstream output transmit FR is shown in the rightmost plot. The input FR (i.e., the receive level) now matches the node port upstream input level with a final 30 dB amp upstream operational gain.

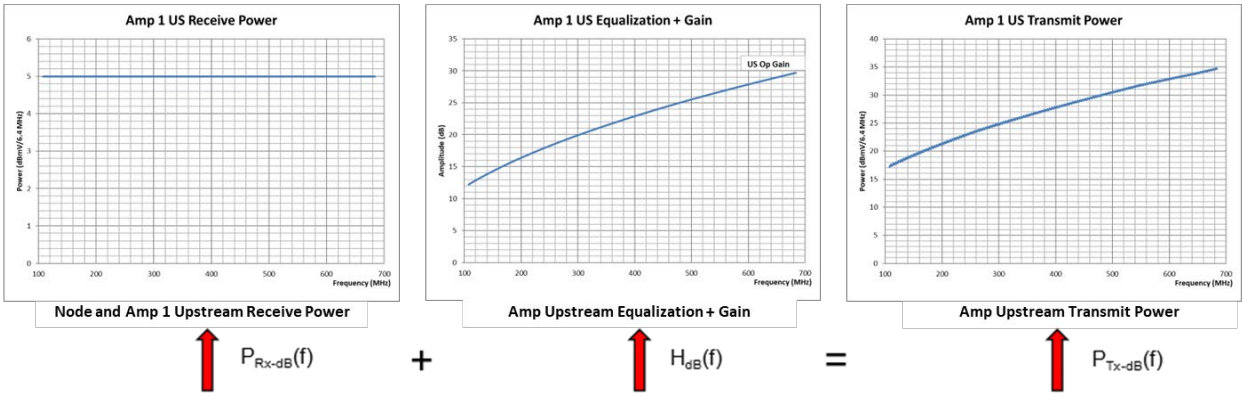
This final setup with activation of the SoC upstream (digital) equalizer and final gain setting achieves the node upstream input design level and removes any input linear distortion (amplitude variation) to the input signal. Examples demonstrating this result are shown in Section 8 Example Auto Setup of Cascaded FDX Amplifier Networks.

Proceed to the next Section 6 Network Amplifier Topology Discovery after all amps in the node leg have been configured before the final optimization in Section 7 Final Downstream and Upstream Optimization of All Amps.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



(a)



(b)

Figure 11 – Example of Upstream Input Optimization Using SoC Upstream Equalizer: (a) Inter-Amp Span Loss, and Amp Upstream Equalization; (b) Amp Upstream Receive Level, Amp Upstream Equalization and Op Gain, and Amp Upstream Transmit Level

6. Network Amplifier Topology Discovery

A functional diagram of an FDX amplifier with topology discovery paths is shown in Figure 12. Using the embedded cable modem in each FDX amplifier, as depicted in Figure 12, the network topology can be determined by transmitting individually, one at a time, in any order, through the mid-split upstream. During each transmission, all amplifiers that detect the signal at the SoC US PA out test point are connected in the upstream path, forming a link between the transmitting amplifier and the node. Note that the transmitting amp will also receive its own injected transmission into the mid-split upstream path.

An example tree and branch network diagram of a cable plant depicting amplifier cascade connectivity is shown in Figure 13. A single port is connected to 25 amplifiers with a maximum cascade depth of 6 (Node + 6). Each amp has a single downstream input (North Port). An amp can have multiple downstream outputs (South Ports), which are also multiple upstream inputs. Splitters/directional couplers can also combine paths that are external to the amplifier port(s) but are considered direct connections to the amplifier. Either a direct or a combined path between amps is a network connection.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Such topology can be represented as a linear graph. Linear graphs are commonly used to describe electrical network topology with associated cut-set (Kirchoff Current Law) or loop-set (Kirchoff Voltage Law) matrices. Network elements (e.g., R, L, and C) are represented as nodes with connections between nodes represented as edges. A non-zero matrix element indicates an edge connection between nodes.

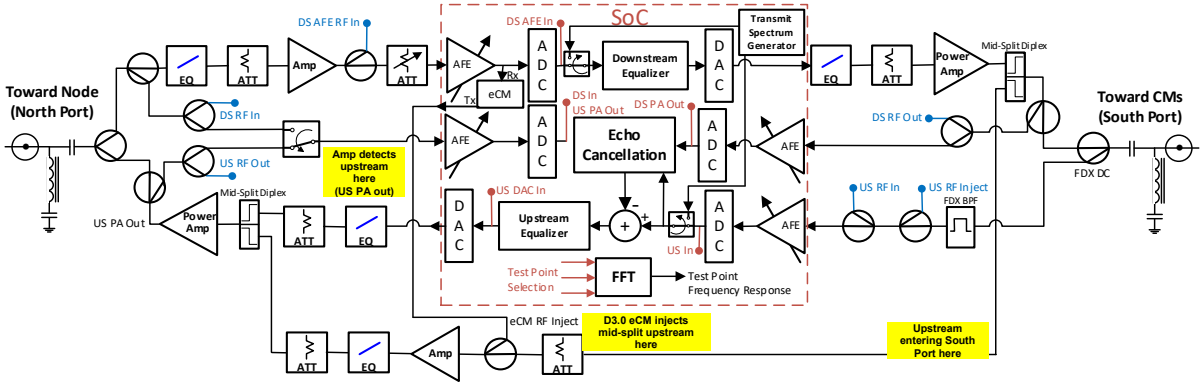


Figure 12 – FDX Amplifier With Topology Discovery Mid-Split Upstream Signal Paths

For determining network topology of amplifier connections, a similar principle can be applied. Each amp in the cable network is a "node", and each input or output connection between amps is an "edge". Network topology is represented as a matrix of unidirectionally edge connected nodes from each end-of-line amplifier toward the FDX node.

A linear graph matrix describing such edge connections of amp nodes between each transmitting amp and all receiving amp nodes in the path toward the cable system fiber node for the network of Figure 13 is shown in Figure 14. A non-zero row element in the matrix denotes an edge connecting amp in the path between, and including, the transmitting amp and the root amp connected directly to the fiber node. Each "1" cell represents a mid-split upstream transmission in a row amp that is received by that column amp. For example, node row 14 indicated that this amp transmission can be heard by receiving column node amps 13 and 1 (as well as itself).

Some properties of this matrix can be used to determine the network topology of all amp cascades from an end-of-line amp to the fiber node. The row sum is equal to the number of amps that received the transmission of that row amp and represents the cascade depth of the transmitting amp in a network branch. The column sum is equal to the number of amps whose transmissions were received by that column amp (including itself). For example:

- The column of amp 14 did not receive any other amp transmission except itself (a single "1" in column 14) and is therefore at the end-of-line of a network branch.
- The column of amp 13 received a transmission only from itself and amp 14 (rows 13 and 14).
- The column of amp 1 received transmissions from every amp in the network, i.e., it is the root amp of the network tree (amp cascade level 1) connected to the fiber node.

A branch starts at an end-of-line amp where the # of edges in the column sum is equal to 1, i.e., there are no amps downstream of this amp. Each amp in the branch (cascade) from that end-of-line amp to the root amp (connected to the node) has a non-zero cell in the row of that end-of-line amp. To determine network topology, sort each non-zero cell amp in the end-of-line amp row in decreasing order (from the end-of-line amp to the root amp) according to the row sum (amp level) of each non-zero amp in the row.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

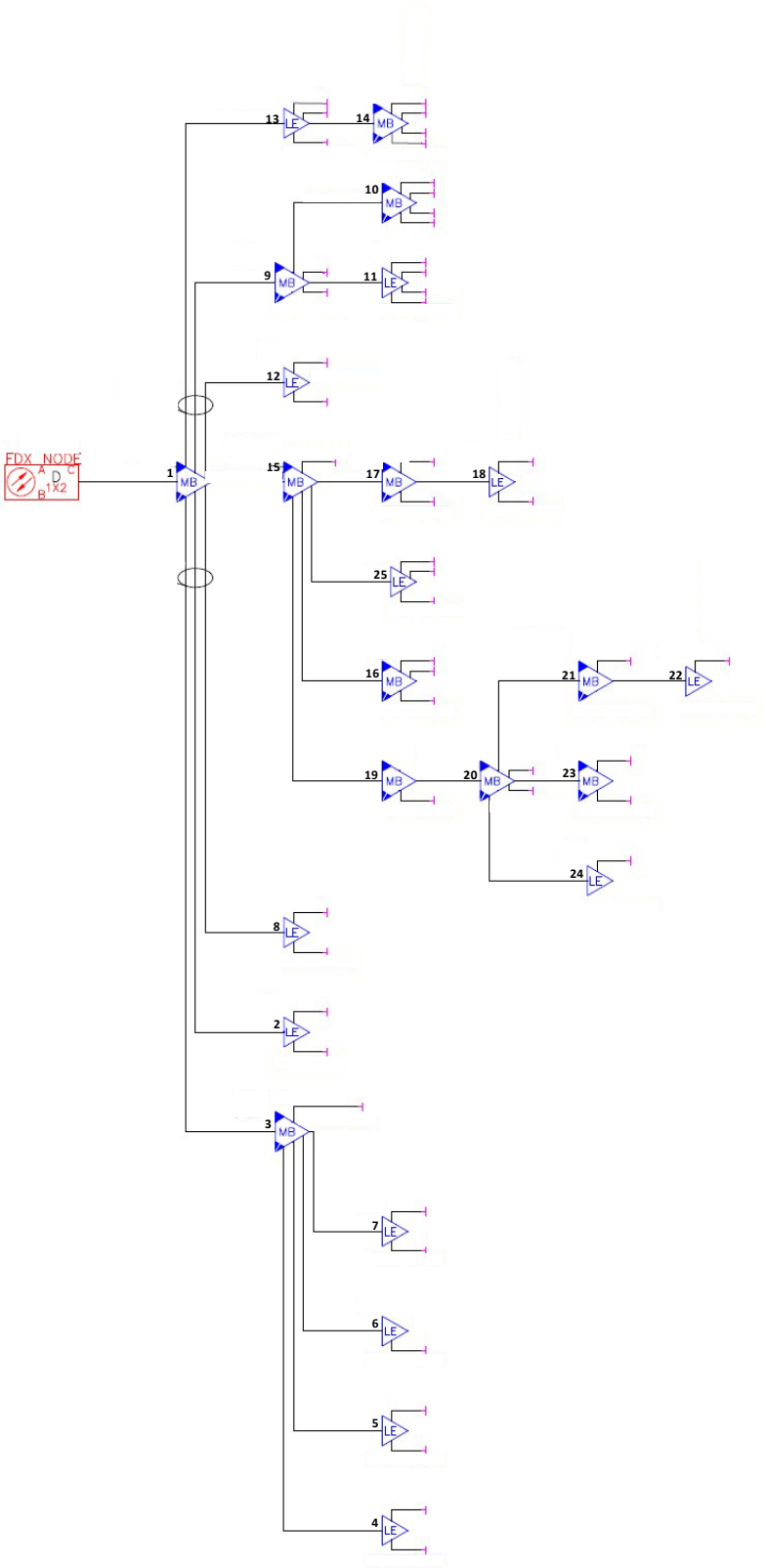


Figure 13 – Example Tree and Branch Cable Network

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

For example:

- amp 14 column total (# of edges) = 1 and is at a branch end-of-line
- row 14 connects with edges to columns 13 and 1
- row 14 sum (amp cascade level) = 3
- row 13 sum (amp cascade level) = 2
- row 1 sum (amp cascade level) = 1
- Therefore, the path of this branch from the end-of-line (amp level 3) to the root (amp level 1) is 14, 13, 1 as shown in Table 1

	# of edges	25	1	5	1	1	1	1	1	3	1	1	1	2	1	11	1	2	1	6	5	2	1	1	1	1
Amp Level		Receive																								
	nodes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	Transmit 1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	3	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	4	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	5	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	6	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	7	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	8	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	9	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	10	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	11	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	12	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	13	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
3	14	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
2	15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3	16	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
3	17	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
4	18	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0
3	19	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
4	20	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0
5	21	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	0
6	22	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	0	0	0
5	23	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1	0	0
5	24	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	1	0
3	25	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1

Figure 14 – Matrix Describing the Tree and Branch Cable Network of Figure 13

Table 1 – Network Topology Describing the Tree and Branch Cable Network of Figure 13

Network Topology for Node +6

End of Line Amps:		2		4	5	6	7	8		10	11	12		14		16		18				22	23	24	25
Toward Node ↓		1		3	3	3	3	1		9	9	1		13		15		17				21	20	20	15
				1	1	1	1			1	1			1		1		15				20	19	19	1
																		1				19	15	15	
																						15	1	1	
																						1			

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

7. Final Downstream and Upstream Optimization of All Amps

After completing Network Amplifier Topology Discovery, repeat previous Downstream and Upstream Optimizations successively from ALL level 1 to level x amps. Each amp at level n initially relies on the downstream output profile of amp n-1 matching that of the node. This is only true if FDX amps are configured successively from the node toward the end of line. Configuring amps in any other order will measure conventional amplifier downstream output profiles which may differ from that of the node resulting in some deviation from the node downstream output level profile. A second pass configuration in order from the node toward the end of line (N+1 to N+x) corrects any such deviations resulting in each amplifier downstream output matching that of the node.

Each amplifier upstream gain and FR can now be precisely adjusted by using a transmit spectrum generator signal from each end-of-line amplifier as follows:

- Check Amp/Node upstream receive level with SoC Transmit Spectrum Generator (Figure 15)
  - Query Amp Central for Node/Amp Port FDX upstream input level (constant value over frequency)
  - Set Transmit Spectrum Generator level successively at each end of line amp = FDX upstream input level + <South Port-to-SoC US in FR>

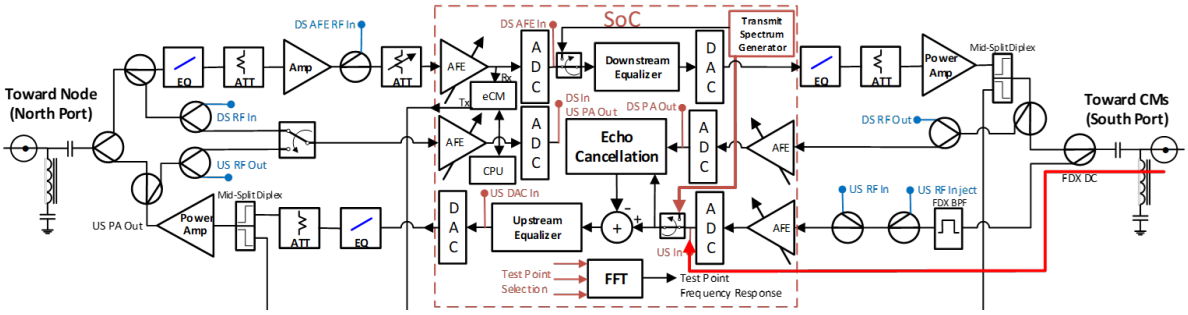


Figure 15 – Set SoC Transmit Spectrum Generator to Check Amp/Node Receive Level

- Check upstream alignment (and redo if needed) from the end of line toward the node
  - Back to front from amp cascade level x to 1 for a N + x system after network amp topology discovery
  - Turn on the Transmit Spectrum Generator signal in the FDX upstream band on the end of line amp x
  - Successively check each amp upstream receive level at amp x-1,x-2,…,1 at SoC US in test point in each cascade path from each end of line amp x
  - Check for each amp received upstream input level at SoC US in FR = Node/Amp Port FDX upstream input level FR, where the Node/Amp Port FDX upstream input level FR should be a constant value over frequency[3]

8. Example Auto Setup of Cascaded FDX Amplifier Networks

The methodology described for automatic configuration of FDX amplifiers, designed with DSP in the SoC shown in Figure 1, is applied to two examples of amplifier cascades within cable systems. Firstly, to a suburban system of single-family units with distributed taps separated by about 100 feet with amps

[3] Since the Transmit Signal Generator injects a constant level across the FDX band at the SoC US In test point, the Config #4 FR from the South Port to the SoC US In test point is not shaping the test signal FR. So the upstream input of the receiving amps South Port north of the transmitting amp will be the Node/Amp Port FDX upstream input level FR - Config #4 FR. Then the successive SoC US In test point level FRs will be the Node/Amp Port FDX upstream input FR - Config #4 FR + Config #4 FR = Node/Amp Port FDX upstream input FR.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

spaced about 800 feet in a 6-amp cascade. Secondly, to an urban multiple dwelling unit, multi-story building with 4 or more back-to-back tap clusters separated by about 45 feet with 8 taps per amplifier at spacing of about 50 feet for each group of 6 floors.

These two types of systems have significantly different frequency response characteristics that present different challenges to amplifier setup and equalization due to the different mix of cable lengths and tap spacing. Nevertheless, it will be demonstrated that the aforementioned procedures can achieve nearly identical input and output levels across the cascade of amplifiers in a uniform manner, while significantly reducing the accumulation of linear distortion that is typically observed in conventional amplifier cascades.

8.1. Single Family Unit (SFU) Model of a Node + 6 Amp Cascade

An example of a Node + 6 SFU system is shown in Figure 16. The node port input/output and echo interference levels, upstream received signal-to-noise ratio (SNR), and resultant bit-loading is shown in Figure 17.

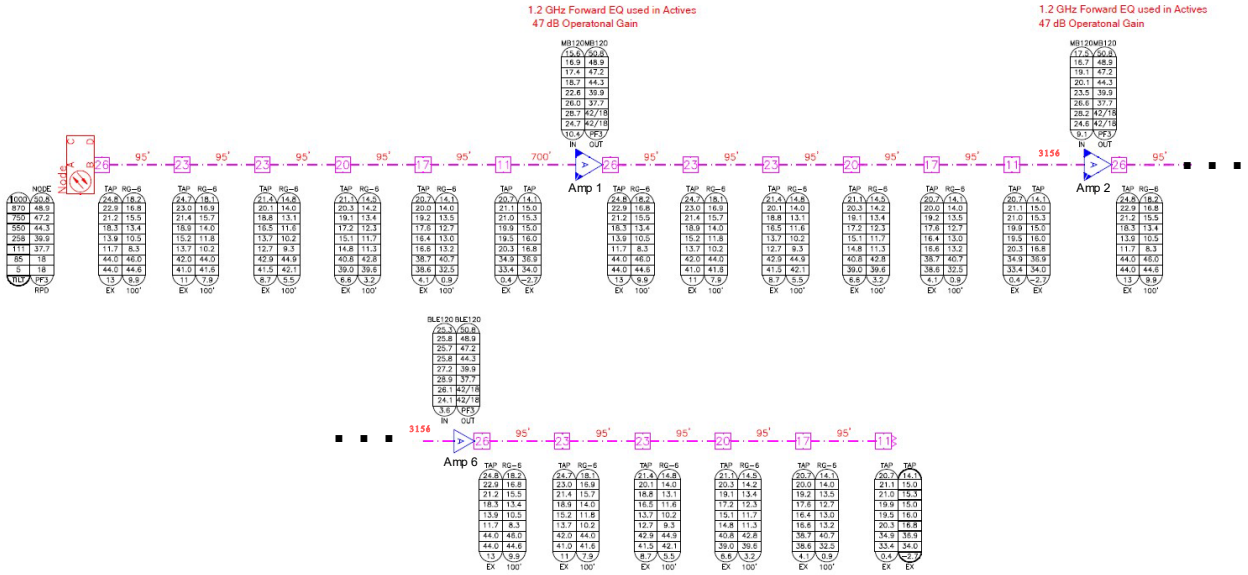
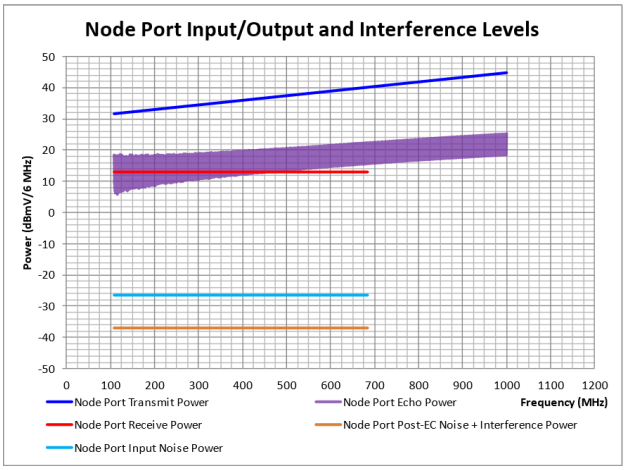


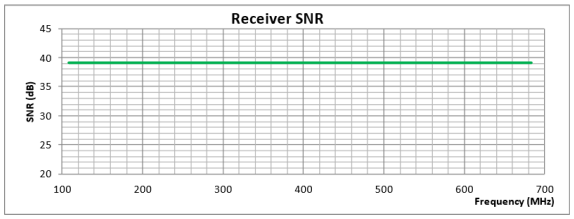
Figure 16 – Example Single Family Unit (SFU) Model for a Node + 6 Amp Cascade

The procedure of Section 4.1 for initial downstream setup for mid-split operation of the amplifiers was applied to Amp 1 of the SFU system of Figure 16. This initially set up the attenuators and equalizers with the SoC passthrough of the downstream signal. The result in Figure 18 for the Amp 1 downstream gain, plus uptilt on the downstream received input, yields the downstream transmitted output shown. Note that the amp output slightly deviates from the node transmitted signal frequency response.

Next, the procedure of Section 4.2 for the final optimized downstream setup of the amplifiers using the digital equalizer in the SoC was applied. The result in Figure 19 for the Amp 1 optimized downstream gain, plus uptilt on the downstream received input, yields the downstream transmitted output shown. Note that the amp output deviation from the node transmitted signal frequency response has been eliminated by the SoC digital equalizer action.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.





Node and Amp Upstream Input Level: 13 dBmV/6.4 MHz

3.1 Spec	3.1 PMA
Average Bit-Loading Post-EC Echo + D4.0 US Rx + CNR 108 - 684 MHz	Average Bit-Loading Post-EC Echo + D3.1 US Rx + CNR 108 - 684 MHz
9 bits/subc	11 bits/subc

Figure 17 – SFU Node Port Input/Output and Echo Interference Levels

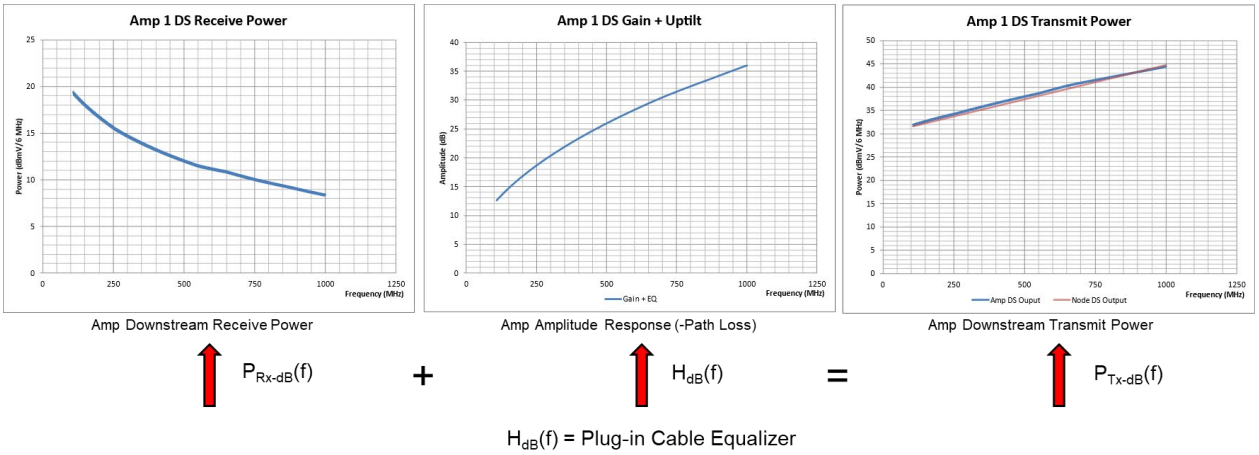


Figure 18 – SFU Amp 1 Initial Downstream Setup

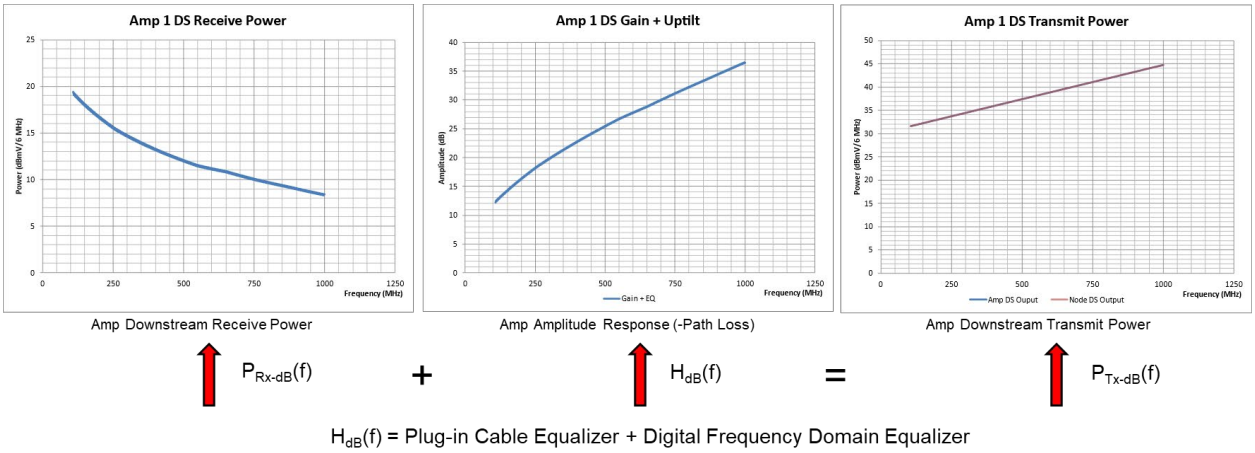


Figure 19 – SFU Amp 1 Final Optimized Downstream Setup

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The procedure of Section 4.1 for initial downstream setup for mid-split operation of the amplifiers was applied to Amp 6 of the SFU system of Figure 16. This initially set up the attenuators and equalizers with the SoC passthrough of the downstream signal. The result in Figure 20 for the Amp 6 downstream gain, plus uptilt on the downstream received input, yields the downstream transmitted output shown. Note that the amp output significantly deviates from the node transmitted signal frequency response. High frequency roll-off and linear distortion (amplitude ripple) due to signal reflections increases as the downstream signal traverses through the amp cascade.

Next, the procedure of Section 4.2 for final optimized downstream setup of the amplifiers using the digital equalizer in the SoC was applied. The result in Figure 21 for the Amp 6 optimized downstream gain plus uptilt on the downstream received input yields the downstream transmitted output shown. Note that the amp output deviation from the node transmitted signal frequency response has been eliminated by the SoC digital equalizer action. This is true for each amp in the cascade.

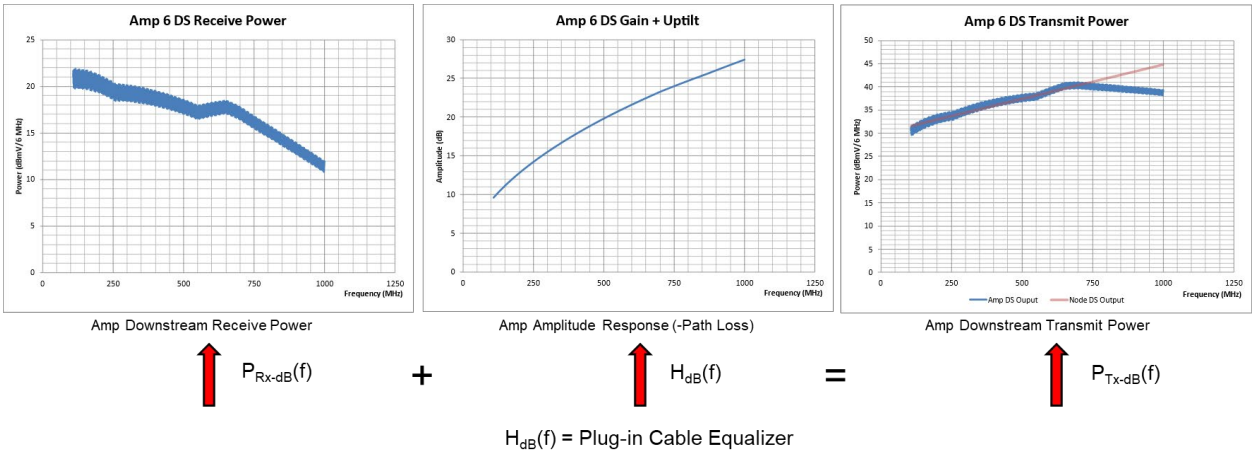


Figure 20 – SFU Amp 6 Initial Downstream Setup

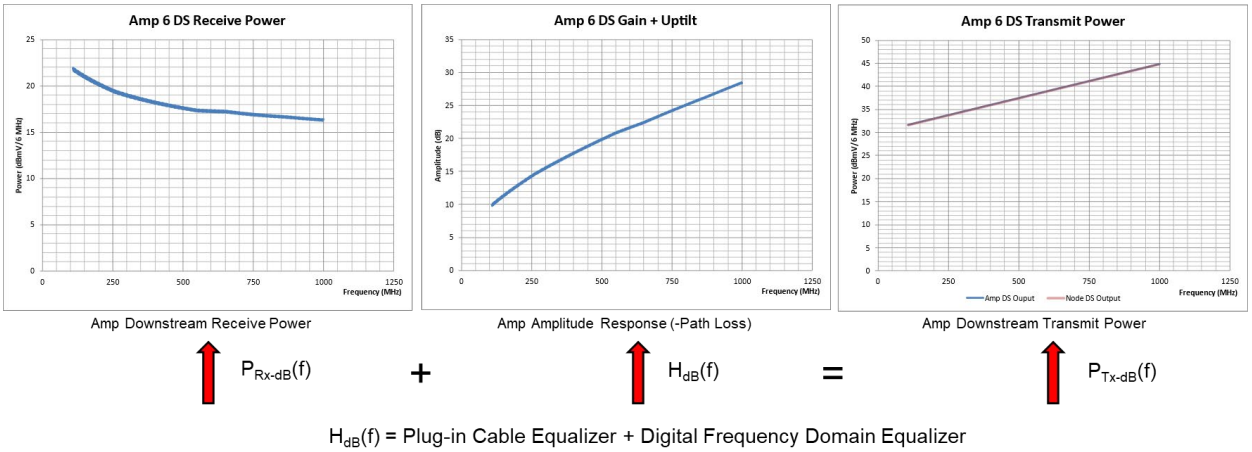


Figure 21 – SFU Amp 6 Final Optimized Downstream Setup

A similar procedure of Section 5.1 for initial upstream setup for FDX operation of the amplifiers was applied to Amp 1 of the SFU system of Figure 16. This initially set up the attenuators and equalizers with the SoC passthrough of the upstream signal. Note that in the upstream direction, Amp 1 is at the end of

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the upstream signal cascade. Hence, this amp will have the maximum level deviation and linear distortion in the upstream. The result in Figure 22 for the Amp 1 upstream gain, plus uptilt on the upstream received input, yields the upstream transmitted output shown. Note that the amp input significantly deviates from the node received signal frequency response (flat level). Linear distortion (amplitude ripple) due to signal reflections increases as the upstream signal traverses through the amp cascade.

Next, the procedure of Section 5.2 for the final optimized upstream setup of the amplifiers using the digital equalizer in the SoC was applied. The result in Figure 23 for the Amp 1 optimized upstream gain, plus uptilt on the upstream received input, yields the upstream transmitted output shown. Note that the amp input deviation from the node received signal frequency response (flat level) has been eliminated by the SoC digital equalizer action. This is true for each amp in the cascade.

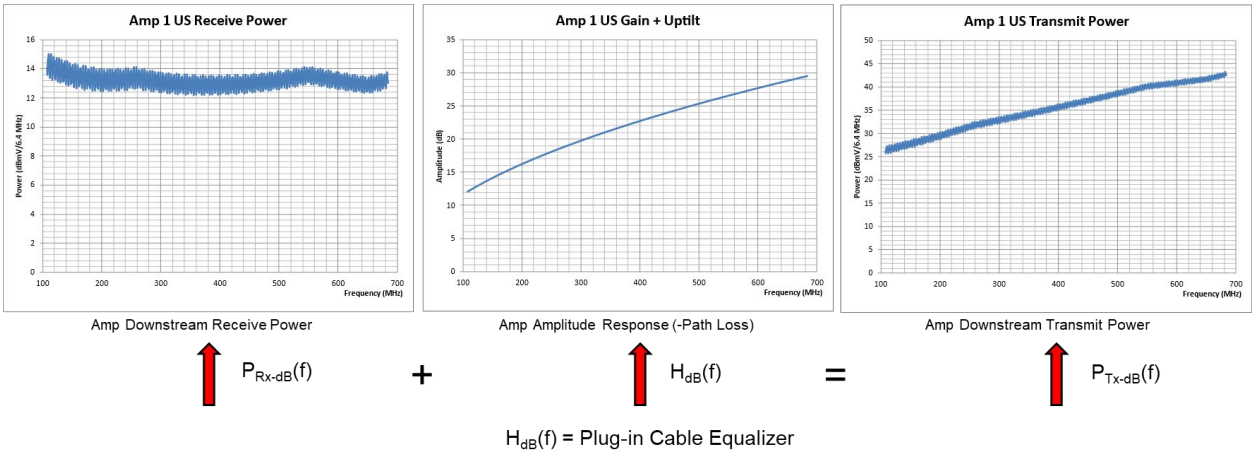


Figure 22 – SFU Amp 1 Initial Upstream Setup

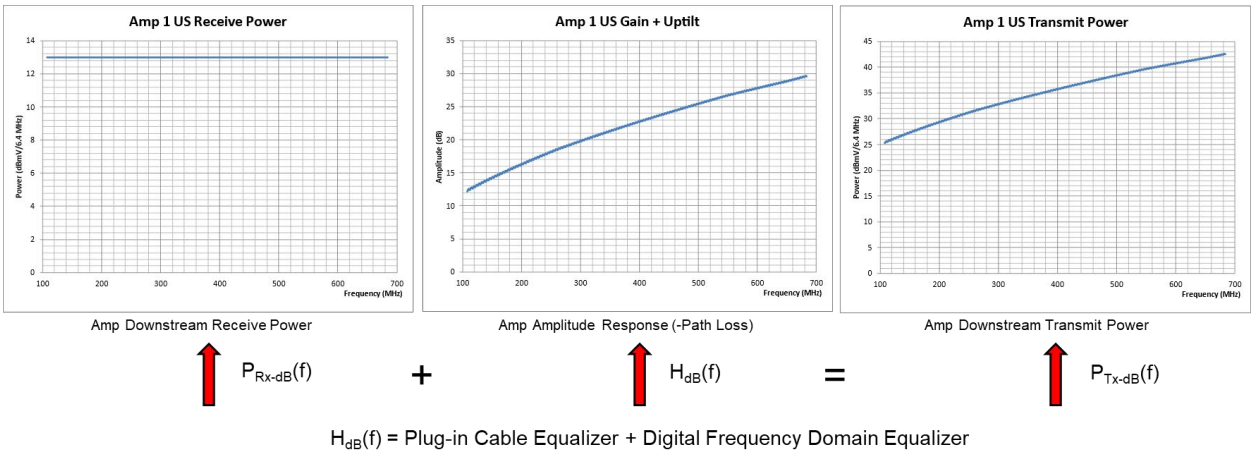


Figure 23 – SFU Amp 1 Final Optimized Upstream Setup

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

8.2. Multiple Dwelling Unit (MDU) Model of a Node + 6 Amp Cascade

An example of a Node + 6 MDU system is shown in Figure 24.

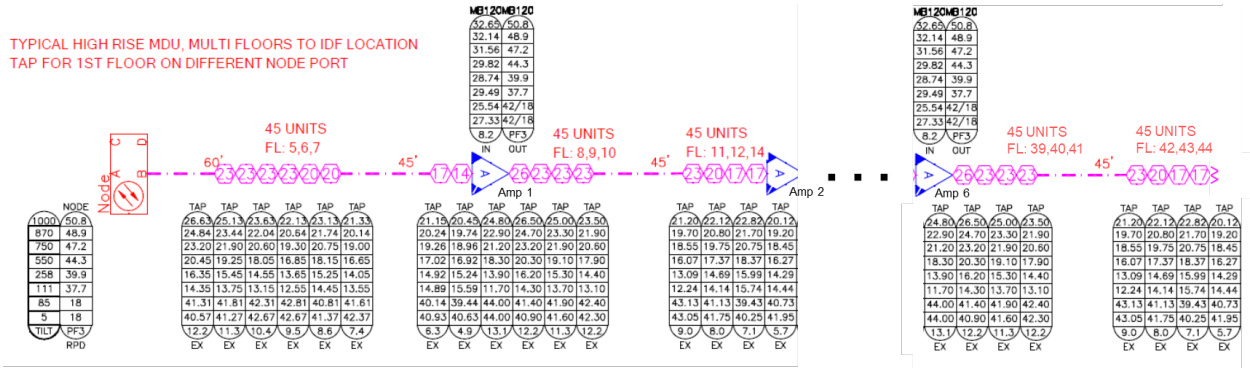


Figure 24 – Example Multiple Dwelling Unit (MDU) Model for a Node + 6 Amp Cascade

The node port input/output and echo interference levels, the upstream received SNR, and resultant bit-loading are shown in Figure 25.

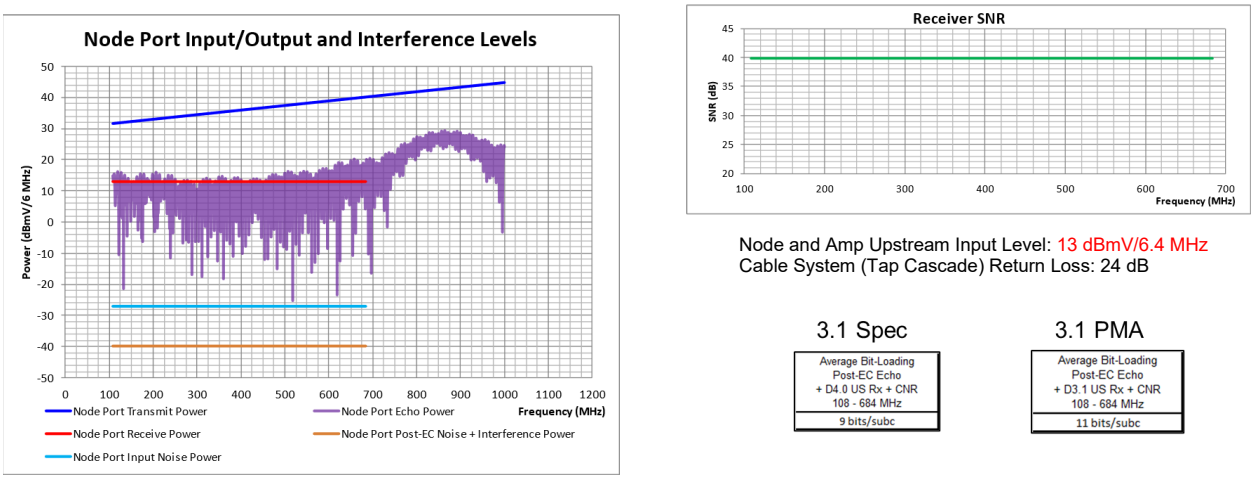


Figure 25 – MDU Node Port Input/Output and Echo Interference Levels

The procedure outlined in Section 4.1, which pertains to the initial downstream setup for mid-split operation of the amplifiers, was implemented specifically on Amp 1 of the MDU system depicted in Figure 24. This initially set up the attenuators and equalizers with the SoC passthrough of the downstream signal. The result in Figure 26 for the Amp 1 downstream gain, plus uptilt on the downstream received input, yields the downstream transmitted output shown. Note that the amp output slightly deviates from the node transmitted signal frequency response.

Next, the procedure of Section 4.2 for final optimized downstream setup of the amplifiers using the digital equalizer in the SoC was applied. The result in Figure 27 for the Amp 1 optimized downstream gain, plus uptilt on the downstream received, input yields the downstream transmitted output shown. Note that the amp output deviation from the node transmitted signal frequency response has been eliminated by the SoC digital equalizer action.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

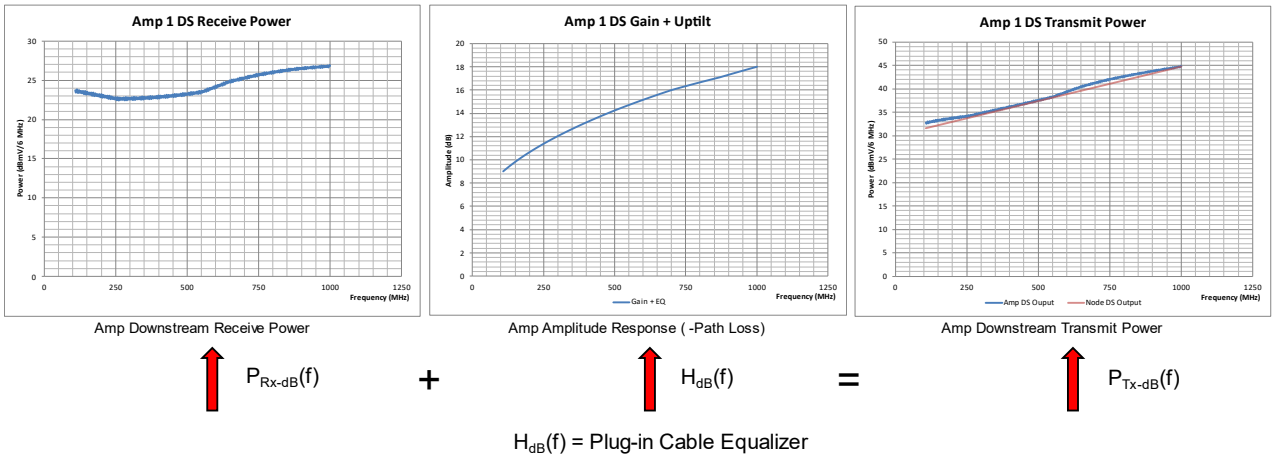


Figure 26 – MDU Amp 1 Initial Downstream Setup

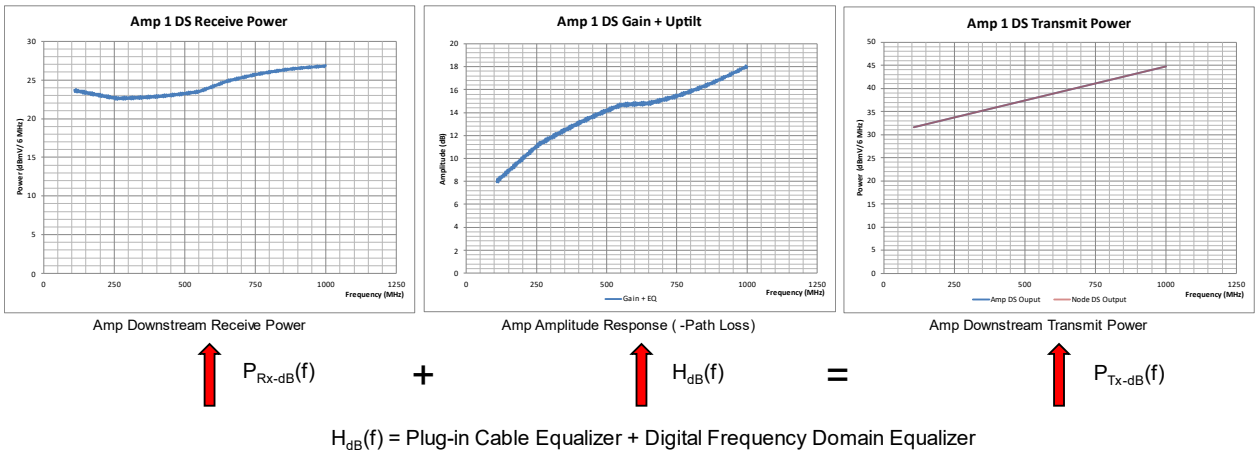


Figure 27 – MDU Amp 1 Final Optimized Downstream Setup

The procedure of Section 4.1 for initial downstream setup for mid-split operation of the amplifiers was applied to Amp 6 of the MDU system of Figure 24. This initially set up the attenuators and equalizers with the SoC passthrough of the downstream signal. The result in Figure 28 for the Amp 6 downstream gain, plus uptilt on the downstream received input, yields the downstream transmitted output shown. Note that the amp output significantly deviates from the node transmitted signal frequency response. High frequency roll-off and linear distortion (amplitude ripple) due to signal reflections increases as the downstream signal traverses through the amp cascade.

Next, the procedure of Section 4.2 for final optimized downstream setup of the amplifiers using the digital equalizer in the SoC was applied. The result in Figure 29 for the Amp 6 optimized downstream gain, plus uptilt on the downstream received input, yields the downstream transmitted output shown. Note that the amp output deviation from the node transmitted signal frequency response has been eliminated by the SoC digital equalizer action. This is true for each amp in the cascade.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

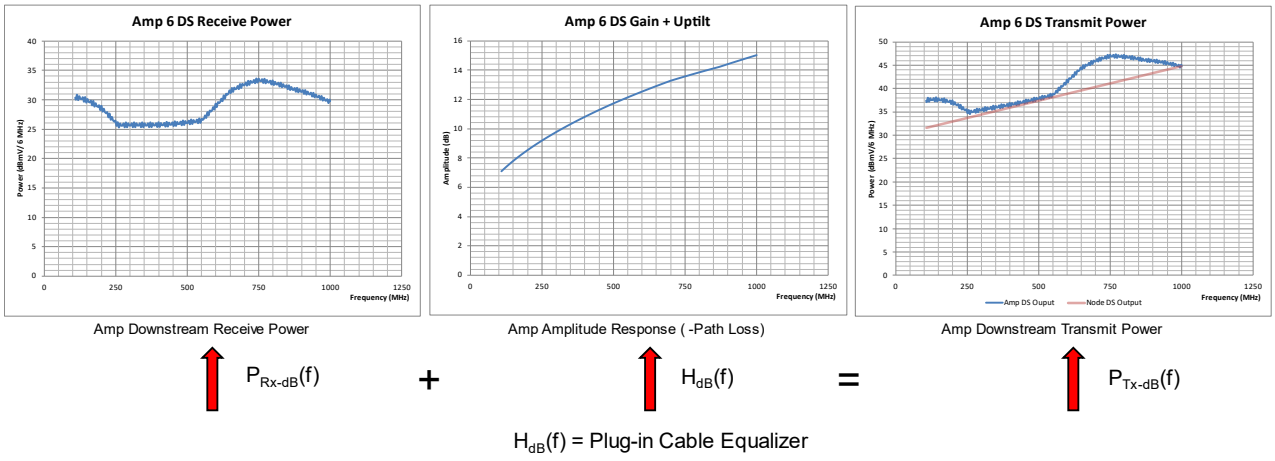


Figure 28 – MDU Amp 6 Initial Downstream Setup

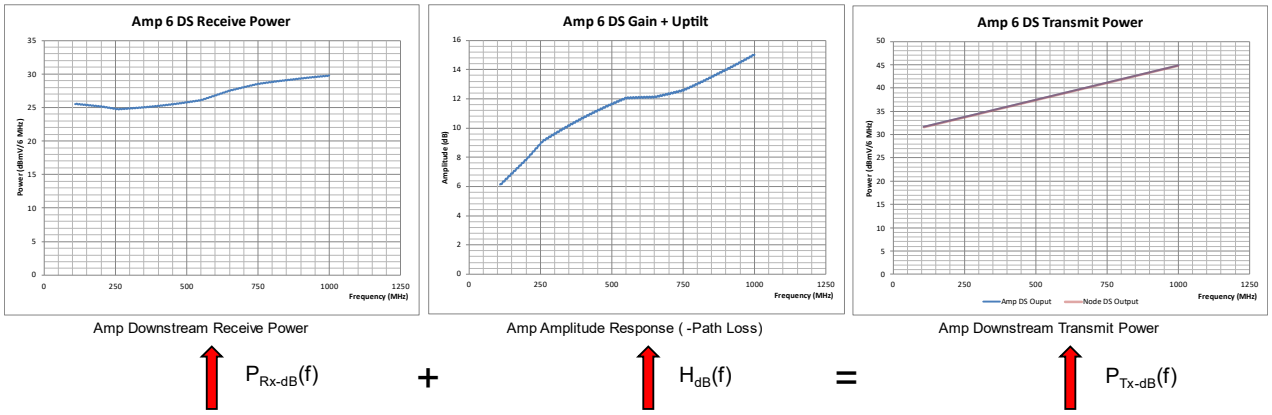


Figure 29 – MDU Amp 6 Final Optimized Downstream Setup

A similar procedure of Section 5.1 for initial upstream setup for FDX operation of the amplifiers was applied to Amp 1 of the MDU system of Figure 24. This initially set up the attenuators and equalizers with the SoC passthrough of the upstream signal. Note that in the upstream direction, Amp 1 is at the end of the upstream signal cascade. Hence, this amp will have the maximum level deviation and linear distortion in the upstream. The result in Figure 30 for the Amp 1 upstream gain, plus uptilt on the upstream received input, yields the upstream transmitted output shown. Note that the amp input significantly deviates from the node received signal frequency response (flat level). Linear distortion (amplitude ripple) due to signal reflections increases as the upstream signal traverses through the amp cascade.

Next, the procedure of Section 5.2 for final optimized upstream setup of the amplifiers using the digital equalizer in the SoC was applied. The result in Figure 31 for the Amp 1 optimized upstream gain, plus uptilt on the upstream received input, yields the upstream transmitted output shown. Note that the amp input deviation from the node received signal frequency response (flat level) has been eliminated by the SoC digital equalizer action. This is true for each amp in the cascade.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

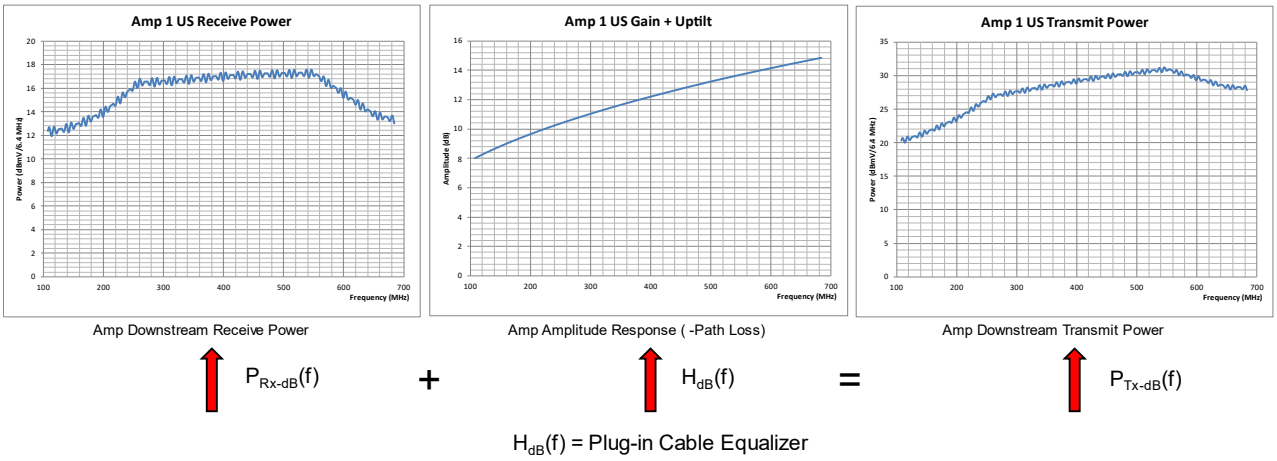


Figure 30 – MDU Amp 1 Initial Upstream Setup

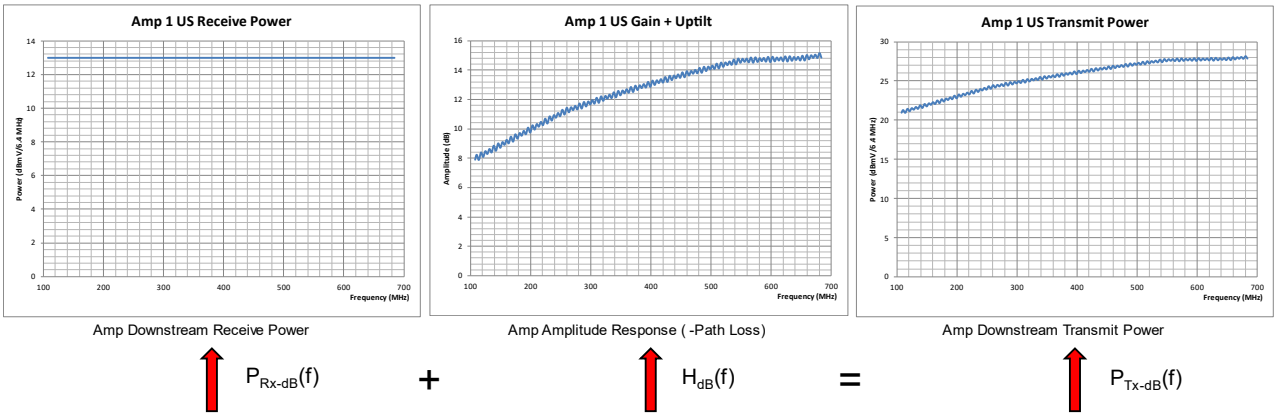


Figure 31 – MDU Amp 1 Final Optimized Upstream Setup

9. Conclusion

Optimized amplifier gain and digital equalization demonstrate superior signal regeneration compared to analog plug-in cable equalizers. ADC, DAC, and FFT processing of signals at input and output test point interfaces included in the DOCSIS 4.0 FDX Smart Amplifier SoC allow frequency domain calculation of amplifier gain and DSP-based equalization.

A multiple step algorithm for automatically configuring FDX amplifier downstream and upstream signals was described, and the resulting performance demonstrated. Optimized digital equalization can restore node output transmission levels, tilt, and input receive levels in cascaded FDX amplifiers. Both example networks discussed demonstrate optimized digital equalization that can restore node transmission levels, tilt, and input receive level in cascaded FDX amplifiers.

A method was described for determining network amp topology using a linear graph matrix with elements determined by mid-split upstream transmission and reception of amp embedded cable modems. This method determines the amp cascade sequence order that could be used to order the automatic amp

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

configuration process from the node to the end-of-line amp. Such methodologies can also be used to initially set up and periodically check for continued proper operation of installed and configured amplifiers.

Abbreviations

ADC	analog-to-digital converter
ADU	automatic drive unit
AFE	analog front end
DAC	digital-to-analog converter
dB	decibel
dBmV	decibel millivolt
DOCSIS	Data-Over-Cable Service Interface Specifications
DS	downstream
DSP	digital signal processing
eCM	embedded cable modem
EQ	equalizer
FDD	frequency division duplex
FDX	full duplex DOCSIS
FFT	Fast Fourier Transform
FR	frequency response
MDU	multiple dwelling unit
MHz	megahertz
N+x	node plus x amplifiers
PA	power amp
R-PHY	remote PHY
SCTE	Society of Cable Telecommunications Engineers
SFU	single family unit
SNR	signal-to-noise ratio
SoC	system on a chip
TCP	total composite power
US	upstream

Bibliography & References

[1] R.S. Prodan, *Full Duplex DOCSIS PHY Layer Design and Analysis for the Fiber Deep Architecture*, SCTE 2017 Cable-Tec Expo

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Future Of Cryptography

Understanding Quantum-Safe Timelines and Deployments

A Technical Paper prepared for SCTE by

Massimiliano Pala
Director, PKI Architectures and Principal Architect
CableLabs Inc.
858 Coal Creek Cir, Louisville, CO 80027
+1 (303) 661-3331
m.pala@cablelabs.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
   - 1.1. Quantum and the Broadband Industry ... 3
   - 1.2. Encryption and User-Level Privacy under Quantum ... 3
     - 1.2.1. Grover's Estimation and Entropy Requirements ... 4
     - 1.2.2. Quantum Technologies for Cryptography ... 5
   - 1.3. A Crypto-Agility Example for the Broadband Industry: The CARAF Framework ... 6
     - 1.3.1. Different Types of Migration Strategies ... 6
2. The Future of Cryptography Project ... 7
   - 2.1. Quantum, Security, and Timelines ... 8
     - 2.1.1. Deployment Timelines and the Mosca Framework ... 9
     - 2.1.2. Migration Timelines: An Example ... 10
3. DOCSIS and Quantum-Safety ... 12
   - 3.1. The BPI+ V2 Authentication Protocol ... 12
   - 3.2. The Simulation Framework ... 13
     - 3.2.1. The Framework in a Nutshell ... 14
   - 3.1. Current Framework Limitations ... 15
   - 3.2. The Methodology ... 16
   - 3.3. The Results ... 16
     - 3.3.1. Considerations on Message Sizes ... 17
     - 3.3.1. Considerations on Execution Complexity ... 17
4. Conclusions and Future Work ... 18

Abbreviations ... 19

Bibliography & References ... 20

List of Figures

Title **Page Number**

Figure 1 - Expected Running Time (Computational Cycles) for Grover's Algorithm. ... 5
Figure 2 - The Mosca Framework ... 9
Figure 3 - The Mosca-Pala framework ... 9
Figure 4 - Example Timeline for QS migrations based on the Mosca-Pala framework. ... 11
Figure 5 - BPI+ V2 authorization process (Auth Req and Auth Reply only) ... 13

List of Tables

Title **Page Number**

Table 1 - Size of Digital Certificates that use Hybrid Technologies (Explicit Composite) ... 7
Table 2 - Message sizes for BPI+ V1, BPI+ V2, and BPI+ V2 QS. ... 16

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The digital era, recently fueled by amazing advancements and synergies in computer science and information technology, has placed an unprecedented reliance on cryptography for securing data. However, the very cryptography that provides security and resilience across the world now stands on the cusp of a significant paradigm shift. The advent of quantum computers, a technological achievement in itself, is indeed poised to challenge the very bedrock of contemporary cryptographic systems, thus urging for fundamental shifts in our algorithms adoptions. In fact, quantum computers, with their fundamental differences in their computational capabilities from classical computers, have the potential to threaten current encryption and authentication systems. Traditional encryption schemes like RSA [Rsa16] or ECDSA [Ec05] rely heavily on mathematical problems such as factoring large integers and discrete logarithm problems, tasks that are computationally infeasible for classical computers but can be swiftly dealt with by a Cryptographic Relevant Quantum Computer (CRQC).

In the face of these imminent threats to digital security infrastructure, the security community must respond urgently with an adaptable and future-proof strategy: crypto-agility. This strategy, which emphasizes the ability to shift from one cryptographic system or algorithm to another, is quickly becoming a necessity rather than an option.

This paper aims to delve into the impending risks posed by quantum computing to present-day cryptography, scrutinize the challenges in crypto-analysis, and explore the need for the adoption of crypto-agility processes within digital security paradigms. In this work we focus on providing important considerations on practical timelines for quantum-safe migrations for the broadband community together with a look at *Future of Cryptography* project that aims at providing guidelines for how to build a safer and secure future in the quantum age.

1.1. Quantum and the Broadband Industry

Quantum computers pose a particular danger to broadband network operators and indeed all sectors relying on authentication and encryption for data security. The primary reason is that quantum computers are exceptionally good at solving the factoring problem — i.e., factoring a large number into its prime components. Traditionally, this fundamental problem has provided the backbone of many encryption techniques, including RSA (Rivest-Shamir-Adleman) encryption, which is one of the most widely authentication scheme used in the broadband industry.

However, a fault-tolerant quantum computer with enough qubits can solve these factoring problems far more quickly and efficiently than a classical computer, potentially breaking the security of device identities and trusted authorities (i.e., digital certificates) almost effortlessly. Such quantum systems are referred to as Cryptographic Relevant Quantum Computers or CRQCs.

If and when these computational devices become practical, they could disrupt the authentication methods currently used by broadband network operators to transfer authorization and encryption keys, making it essential for these organizations to start preparing alternative encryption methodologies. This implies that securely transmitted data today (or stored in the Company's databases) could be captured and stored for decryption later, especially if the process leverages quantum-vulnerable key exchange mechanisms.

1.2. Encryption and User-Level Privacy under Quantum

In the broadband industry, encryption is used both to protect corporate data and Access Network communications. For example, when considering user-level privacy, DOCSIS [SECv4.0] uses BPI+

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

[BPI+08] to provide privacy of communication between the CM and the CMTS. In this context, it is therefore important to understand the practical impact of Quantum Computing and Grover's Algorithm [Gro96] on symmetric encryption, particularly focusing on the Advanced Encryption Standard (AES).

In the quantum computing world, Grover's Algorithm, developed in 1996, theoretically allows for a very efficient search for unstructured data (black-box function) by leveraging the possibility to construct all possible output in a single operation via the implementation of an Oracle function and the use of quantum-mechanical properties such as superposition and entanglement. The algorithm, however, does not offer the same significant speed-up as Shor's quantum algorithm for factoring large integers.

This means that cracking a symmetric cipher like AES-128 with a quantum computer using Grover's Algorithm would still require significant computational resources, involving serial computing operations in a quantum setting that is currently thought to be a quite difficult task (i.e., more than 2^{64+} operations). Thus, switching from AES-128 to AES-256 due to quantum computing threats might need to happen a bit further in the future than originally expected, given today's understanding of quantum-computing and its limitations. Even institutions like NIST are still convinced that AES-128 will likely remain secure against quantum threats for several more decades because of this hardware dependency and lack of parallelizable quantum algorithms.

Nevertheless, many different environments are preparing to switch to higher entropy encryption systems (or enabling the possibility to do so), should the assumptions behind the implementation of the Grover's algorithms should not hold in the future [3GPP23].

1.2.1. Grover's Estimation and Entropy Requirements

Although the use of AES-128 can still be considered secure for a decade or more, a recent paper from ATIS [Deakin23] quantifies the effect of poor entropy on AES encryption. The paper shows how the run time of Grover's algorithm on a quantum computer can be significantly reduced unless the source of entropy is, in fact, truly random. Furthermore, as quantum computer gate fidelity improves over time, also the number of logical qubits required to run Grover's algorithm is further reduced – this combined effect with poor entropy could make symmetric key encryption less resistant to quantum attack than many expect. Using the equations 9 and 11 from ATIS' paper, we can provide an evaluation of the relative security of AES assuming keys are produced from low entropy sources.

Let's assume that the search space is n = 256, thus requiring the size of the key to be 256 bits.

According to [Deakin23], Grover's run time, namely how long the algorithm is expected to run given a physical key of n-bits produced from an h-bit entropy source, can be approximated by the time needed to run the Oracle operator followed by the execution of the diffusion operator for $\sqrt{N}$ (where N is equal to 2^{256}). Therefore, the runtime can be expressed as:

$$Runtime\ (n,h) = 2^{h/2}(n^2 + 4n + c_n) + 2n$$

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Where ‘h’ represents the number of entropy bits where the value ranges between 32 to 256 bits.

Figure 1 from the Deakin’s paper shows how important the use of a high source of entropy is when evaluating the expected computational cycles of Grover as the exponential nature of the runtime curve manifests itself after the entropy exceeds that of approximately 220 bits.

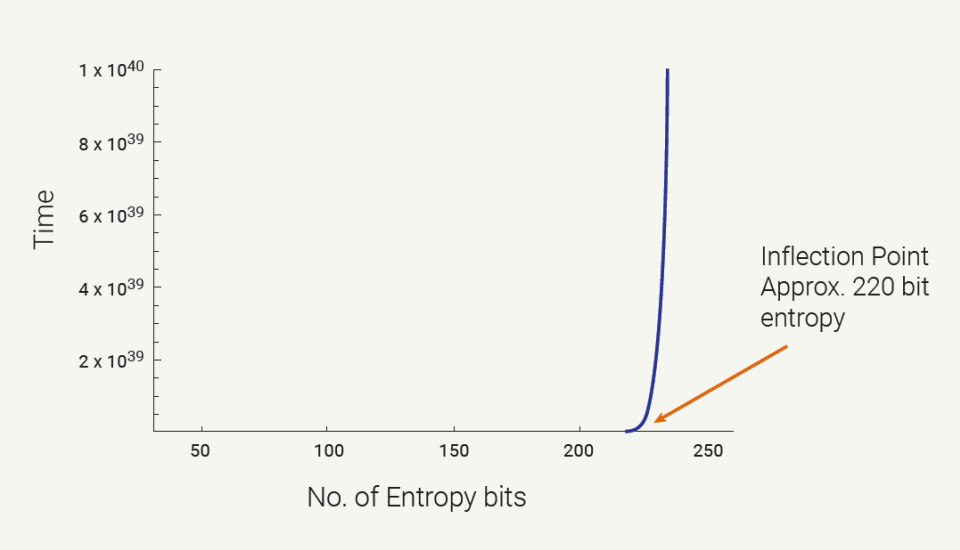


Figure 1 - Expected Running Time (Computational Cycles) for Grover's Algorithm.

It is important to highlight how this analysis not only shows how important it is to have larger keys to protect against quantum computers, but also how paramount it is to ensure that it is produced from a high entropy source. In other words, although we currently estimate the current generation of encryption algorithms (e.g., AES-128) to still be secure in the context of the broadband industry for a couple of decades, this work shows how important it will be to rely on true source of randomness to generate symmetric keys with the required number of entropy bits (e.g., how truly “random” the bits generated for the symmetric key are) and not just doubling the size of the key without relying on true random sources.

1.2.2. Quantum Technologies for Cryptography

Quantum technologies comprise many different fields, not only quantum computing. In particular, the better understanding of quantum-mechanics processes can be used to build new quantum-based devices (e.g., quantum-based sensing) that can be used, for example, to provide functions that are provable secure. Specifically, when it comes to security, the use of quantum technologies can help in solving one of the most fundamental and hard problems: the generation of random bits. Although conventional hardware random number generators continue to be robust and secure, Quantum Random Number Generators (QRNGs) are today available on the market that provide verifiable sources of randomness that can better address advanced entropy/randomness requirements.

In conclusion, it is important to underline that the source and quality of entropy will have a significant impact on the security of symmetric key cryptography to be quantum resistant. Specifically, as we have seen in the previous section, unless a source of entropy used for 256-bit symmetric key AES is truly

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

random across at least 225 bits, it may have little effect in terms of being any more secure against the quantum-threat.

1.3. A Crypto-Agility Example for the Broadband Industry: The CARAF Framework

When it comes to the deployment of new cryptography, the fundamental step that is crucial for the success of the migration plan is the availability of a clear map of what is available today, a detailed prioritization plan, and a cost-effective action plan that both fits the plans from the ecosystem and the vendor's community.

In order to foster good processes across a single organization and/or a whole industry sectors, many companies will have to identify processes that can be used for a successful multi-year effort such as quantum-safe migration. To help with the internal organization and the coordination across industry-sectors and supply-chains, some frameworks that focus on crypto-agility have been developed. The CARAF framework [Car20] is one of these framework that outlines strategies to develop or maintain crypto-agility within an organization.

CARAF stands for *Cryptographic Agility Risk Assessment and Management Framework* and, in summary, is a tool to help managing the crypto assets of an organization by understanding where and how cryptography is used within the organization, conducting a risk assessment based on the criticality of operations and business functions, prioritizing risk mitigation activities, and ongoing management of cryptographic systems.

The core of CARAF, and other similar frameworks, is aimed at ensuring that organizations are prepared for any cryptographic changes via five main areas of interest/work:

1. **Creation of a Crypto Application Inventory.** Keeping track of all cryptographic algorithms and where they are used is the first step for any organization. This step of the process includes noting any hardware, software, and systems that relies on any element of cryptography.
2. **Categorization.** Categorize relevant assets (e.g., hardware, software, etc.) based on operational and business criticality, i.e., how important they are to ongoing operations and business functions.
3. **Assessment.** Assess the risks associated with each category. A higher risk might be assigned to systems utilizing outdated or weak cryptographic algorithms, those with no available alternate algorithms, or systems which are integral to important business functions.
4. **Prioritization.** Prioritize risk mitigation strategies based on the risk assessment findings and business needs. Although it is important to address high risk areas first, there might be other important considerations factors that might change the risk-based priorities.
5. **Management.** Implement changes, monitor effectiveness, and ensure to regularly repeat the crypto-agility process to keep your systems secure and up-to-date.

It is important that Companies start exploring the availability of such tools and start defining appropriate process for the specific organization, possibly in harmony with the larger ecosystem(s) the Company is a member of.

1.3.1. Different Types of Migration Strategies

As discussed in the previous section, the migration towards quantum-resistant cryptography is a critical step considering the potential threats posed by quantum computing. To effectively transition, a risk-informed strategy that involves a Quantum Risk Assessment (QRA) is necessary to identify and prioritize

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

which assets will be affected. The options to migrate systems from quantum vulnerability to quantum resistance include a one-time migration, crypto agility by design, or a hybrid solution.

A one-time migration involves the singular migration of prioritized technologies to standardized post-quantum cryptography algorithms. It ranges from hardware replacement to a software patch, with constant monitoring of the quantum landscape to ensure the selected method remains quantum resistant.

Crypto Agility by design, instead, refers to the ability to quickly change cryptographic algorithms without requiring long downtimes or extensive code revisions. Products should be designed to swap in different cryptographic algorithms, including PQC, and they should be able to migrate quickly to NIST's PQC algorithms once they are available. When it comes to the practical aspects of updating credentials, especially device ones, there are many challenges that need to be considered. Specifically, although protocols exist that can provide some level of automation (provided that the software has been updated to handle the new algorithms), the possibility for infrastructure-deployed devices to directly reach certificate-renewal services might be limited (e.g., no direct access outside the operator's network) and additional considerations are warranted about the deployment of hybrid (multi-algorithm) identities.

The hybrid approach involves the simultaneous use of PQC and classical algorithms for comprehensive security. This approach acknowledges that vulnerabilities may still be discovered in quantum-resistant cryptography, hence the need to maintain the use of classical algorithms. A downside of the hybrid approach is that it does increase the overhead from cryptography and introduce new challenges related to certificate revocation. On the increased size of cryptography, it is important to notice how the addition of traditional keys and signatures to quantum-safe ones only increases the over-all size of the certificate to be less than 10%. Table 1 provides examples of one type of hybrid technology, i.e., Composite Crypto [Pala23], and the size of the average hybrid certificate when compared to the equivalent non-hybrid version.

Table 1 - Size of Digital Certificates that use Hybrid Technologies (Explicit Composite)

Algorithm	Sec. Level	PubKey Size	Sig Size	Cert Size
id-Dilithium3-RSA-SHA256	192 (112)	2275 (+320)	3564 (+270)	6103 (+600)
id-Dilithium5-ED448	192 (224)	2690 (+97)	3294 (+126)	7660 (+208)
id-Falcon512-ED25519	128 (128)	964 (+66)	734 (+79)	1932 (+136)
id-Falcon512-P256-SHA256	128 (128)	1101 (+203)	738 (+83)	1993 (+197)

All sizes are all expressed in number of bytes and the difference with the non-hybrid version of the certificate is reported in parenthesis. Other mechanism to provide hybrid approaches might focus on the use of multiple certificates (or identities) which, usually, require changes in protocols and/or application logic for their deployments.

Despite these challenges, the migration towards quantum-resistant cryptography is a necessary step to ensure robust security and different industry sectors will need to decide which migration strategy better fits the specific ecosystem and individual companies.

2. The Future of Cryptography Project

To help with the definition of community-wide education, experimentation, and guidelines production, CableLabs and its Members have started a new project named Future of Cryptography. This initiative

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

stems from the previous work done under the Extensible and Programmable Infrastructure for Cryptography and Access (EPIC Access) project, led by Steve Goeringer, which identified potential impacts of crypto failures on the industry and the need for a cost-effective plan for deploying crypto-agility processes.

The Future of Cryptography is detailed as a three-phased project involving education, experimentation, and standardization guidelines.

The first activities for the project are focused on collaboratively building an understanding of threats and solutions related to the deployment of quantum-safe cryptography by leveraging the input from experts across the community. During the second phase, the project aims at experiment with different methodologies for migration to post-quantum solutions, that will result in understanding the practical impact of the new cryptography. In phase three, the project's final efforts will be focused on summarizing the results of the experiments and provide the community with comprehensive guidelines that are meant to be leveraged for standardizing the use of the new cryptography within the broadband community and specifications.

Ultimately, the Future of Cryptography project is an example of a crucial activity that will provide the groundwork towards creating crypto-agile processes across the industry. However, to better understand the length of the timelines that lie in front of us, we provide, in the next few sections, considerations on the processes that need to happen *after* the new algorithm's standardization has been finalized and *before* the industry is ready to start the deployment of the new algorithms.

2.1. Quantum, Security, and Timelines

Today, we start building our Trust by embedding long-term keys in the form of Trust Anchors that are usually distributed via Self-Signed X.509 certificates (also called Root CAs) [X509]. Because the secure distribution of Trust Anchors (TA) is one of the most difficult problems to be solved in security, TAs are usually required to be somewhat long lived across multiple years or even decades. It is not uncommon to have TAs that are meant to provide the root of trust for specific ecosystems (i.e., the WebPKI, DOCSIS, Matter, etc.) for 30 or 50 years. Following similar considerations for the intermediates and edges of the Trust Infrastructure (i.e., Intermediate CAs and End-Entity certificates), it is easy to understand how updating or replacing such identities in the field requires large investments and, in many cases, hardware replacements.

When it comes to the distribution of the authorization key (or session key) in protocols, the situation is quite different. Specifically, at least today, the key exchange process is usually ephemeral, and the public keys used in the process are never persisted across sessions (e.g., ECDHE). This means that switching between different key exchange mechanisms (or key encapsulation ones) does not require global support for the algorithms (such as the public key used in the Root of a PKI), only local support between the entities participating in the communication – thus removing the need for algorithmic long-term commitment. Similarly, for Key Encapsulation Mechanisms or KEMs [Dent02], algorithms are expected to use ephemeral keys (i.e., negotiated at a local level).

In other words, while new algorithms for key exchange/key encapsulation mechanisms can be *deployed and negotiated locally*, planning for new algorithms for Trusted Identities is a *global issue* (in the sense of the entire ecosystem that relies on those identities).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.1.1. Deployment Timelines and the Mosca Framework

Let's now look at some back-of-the-envelope estimations for when quantum-safe solutions (not just algorithms) can be reliably deployed across the broadband industry and, ultimately, across the world.

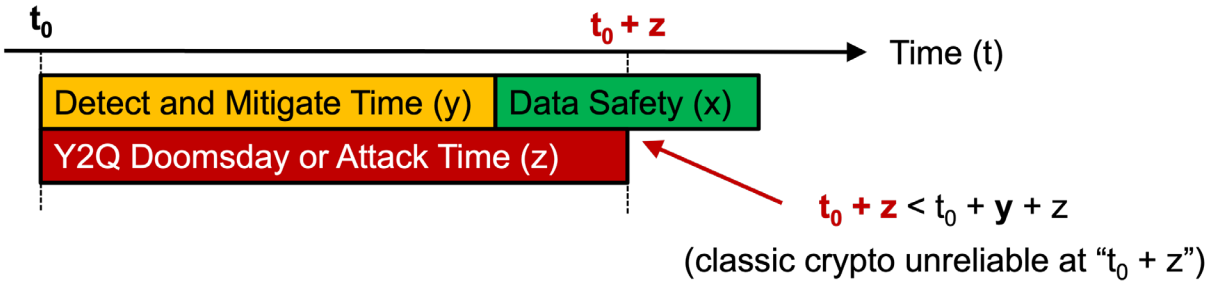


Figure 2 - The Mosca Framework

In Figure 2, we use the Mosca Framework [Mosca15] where t_0 is the time when the deployment of quantum-safe solutions starts. In this framework, the time needed to complete the migration to quantum-safe solutions is indicated by the variable y, while the time until a CRQC is available is indicated with z.

Given the need for multi-year processes for migrating all systems to quantum-safety, we can estimate a base number between 5 and 10 years for the value of y. Additionally, if there are data safety requirements, an additional period, indicated with the variable x (e.g., another 3 to 5 years), to achieve full data safety (i.e., the green bar in Figure 1). When it comes to the estimation of z, we can use a large fork to encompass many different positions about the engineering feasibility and timelines for CRQCs. For our discussion, we will use a value of z between 7 and 20 years.

With these numbers in mind, if we already had all the algorithms standardized and integrated with existing protocols today, if all the TAs had already been upgraded and available for integration with products and network infrastructures, if all vendors had deploy-able and interoperable solutions that can be bought on the market…. That is when we could say that today is t0, however, it is clear that we are quite far from that today.

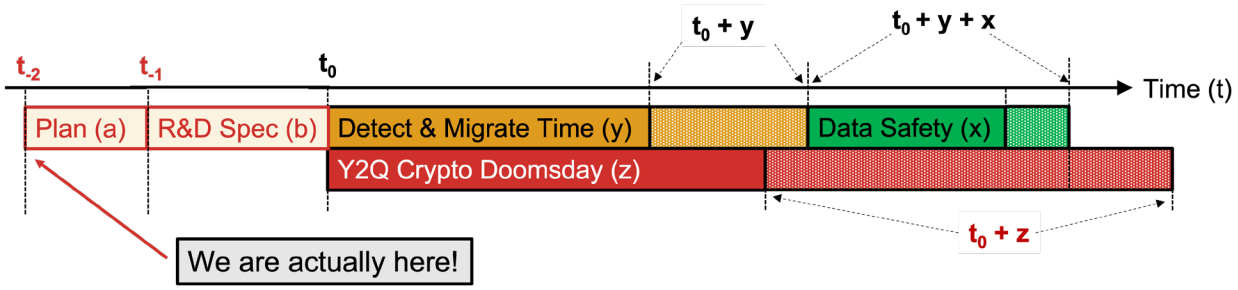


Figure 3 - The Mosca-Pala framework

In Figure 2 we propose an updated formulation of the Mosca framework, the Mosca-Pala framework, where two additional variables are introduced.

We use the variable 'b' to indicate the time needed for the different ecosystems (e.g., Broadband Industry, IoT, Internet Routing, DNS, etc.) to integrate the new algorithms in their protocols. We estimate this

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

process to take between 3 and 5 years (we are very optimistic!). This time 'b' includes the development of new specifications and products for the market, where embedded or dedicated hardware needs to be upgraded (e.g., System on a Chip or SoC).

Even if we had the algorithms and implementations already standardized and available today, there is another variable that needs to be taken in considerations: the need for ecosystems to experiment and agree on which solutions to adopt (e.g., straightforward replacement? Hybrid deployments? etc.). We call this variable 'a' and we estimate its value to be between 2 and 5 years for most environments.

We can finally put everything together and try to estimate how much time will be needed to complete the migration to quantum-safe solutions in different industry sectors such as communications:

$$\text{Migration Time} = a + b + y + x$$

Which gives us a low and high boundary for our estimations to be the following Time To Quantum-Safety or TTQS:

$$TTQS_{MIN} = 2 + 3 + 5 + 3 = 13 \text{ (years)}$$

$$TTQS_{MAX} = 5 + 5 + 10 + 5 = 25 \text{ (years)}$$

Therefore, even with the most optimistic estimates, we are looking at a timeline of over a decade, at the minimum, to complete the migration to quantum-safe solutions, and, potentially, as long as 25 years given the complexities and challenges involved.

Action to prepare for quantum-safe solutions must be taken well in advance of the realization of practical quantum computing and given these estimations for migration timelines and the range of estimates for when a CRQC might become available, it is critical that organizations begin to evaluate and implement quantum-safe solutions now to ensure a smooth and successful transition in the future.

A typical migration timeline for the broadband industry is provided in Figure 4 where we highlight the two new variables of the Mosca-Pala framework: the Planning Time ('a') and the Specifications/Products Updating Time ('b'). During this phase of the migration, we envision the need for ecosystems to evaluate and experiment with tools that provide practical information about risk changes and operational needs. For example, even while algorithms and their parameters are being standardized to address first deployments and risk-mitigating solution (horizontal or vertical algorithm-agility), the specific ecosystem(s) may start collecting, pending the availability of easy-to-use tools and implementations for prototypes and simulations, data on the use of different algorithms, their performances, and experienced bottlenecks in relation to the specific protocol where the new cryptography is being tested. Other factors should also be evaluated such as what lifetime for the device/software is envisioned and the availability of hybrid solutions.

2.1.2. Migration Timelines: An Example

In this section we propose imaginary timeline for migrating an ecosystem to use quantum-safe solutions. For this scenario, let's imagine that the specific industry sector or ecosystem starts paying attention to the quantum-threat and how to address it at time T_{INIT} which can be easily expressed as:

$$T_{INIT} = t_0 - a - b.$$

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 4 - Example Timeline for QS migrations based on the Mosca-Pala framework.

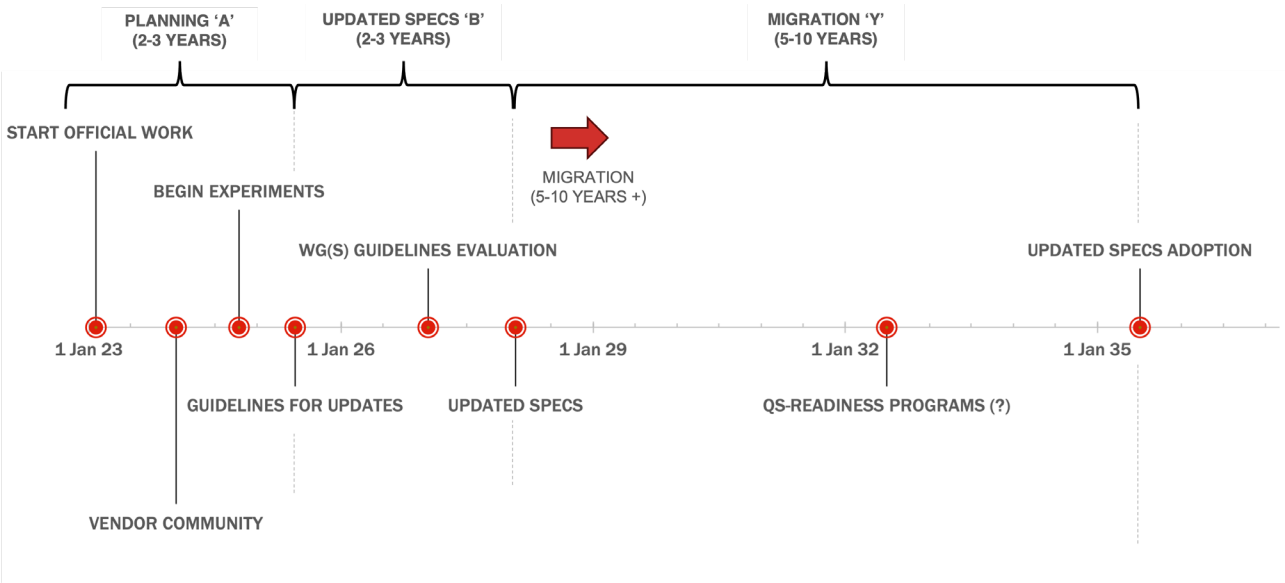


Figure 4 depicts the envisioned scenario where we can imagine that the ecosystem was ready to start its planning phase at the beginning of 2023.

During the initial phase of the migration effort, after involving the vendors' community, the initial set of experimental results should be shared across the whole community to evaluate risks, performances, and costs of each individual solution. On top of that, each entity in the ecosystem should list and prioritize assets updates across the board also in accordance with the envisioned timelines from the vendors' community. If everything goes well and specific requirements and solutions have been identified, we are now around 2026.

At this point, after agreeing on an interoperable approach that might involve the adoption of multiple algorithms (useful to guarantee the implementation of crypto-agility controls that go beyond the single algorithm replacement), the community needs to update the specifications of protocols and devices deployed across their networks. This second phase of the migration must make use of the results produced by the community in the first phase to inform how the specifications must change, when to adopt the new standards, and when devices and certification programs will be available to support such solutions. An important part of this second phase is also defining the characteristics of the Trust Infrastructures (or PKIs) and associated processes to provide a quantum-safe migration paths that different protocols, working-groups, and corporate environments will integrate. We envision this phase to take another three (3) to five (5) years to complete.

Last but not least is the deployment phase. This is the phase that individual operators have absolute control over. In fact, although the specifications and products might be ready for adoption at this stage, cost-related considerations and risk management considerations might drive the effective migration planning. During this phase, it is important to understand that coordination and planning must occur at all levels: not only across the supply-chain for corporate systems, but also explore the dependencies, requirements, and limitations of the access network. In our scenario, for example, even if the new specifications and some products are available for purchasing and deployment before, not all vendors might be ready. Delays in the actual migration might come from the lack of solutions (or upgrade path) from specific vendors.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. DOCSIS and Quantum-Safety

Several technical solutions have been proposed in the literature that investigate the possibility for delivering quantum-safe solutions not only for the newer versions of DOCSIS (4.0+), but also for earlier versions by leveraging a new registration process [Pala20, Pala21]. In this work, we took a step further and built a tool to provide practical measurements and initial comparison between the use of traditional vs. quantum-safe algorithms in BPI+.

In the following sections we share the results of the experiments we ran aimed at the impact of the new quantum-safe algorithms with BPI+ V2 [SECv4.0] authentication protocol. This is as an example of the activities needed, within the different eco-systems, to understand the impact of enabling the use of quantum-safe cryptography in networks and applications.

3.1. The BPI+ V2 Authentication Protocol

Before we provide a description of the simulation framework and the collected results, it is important to review the core design of the BPI+ V2 protocol and highlight the characteristics that could make it a candidate for drop-in replacement with quantum-safe cryptography.

BPI+ V2 design introduced several enhancements (when compared to BPI+ V1) that removed, on paper, the limitations for the deployment of quantum-safe cryptography, such as:

- The use of digital signatures with trusted X.509 certificates to authenticate BPKM messages
- The introduction of ECDHE key-exchange mechanisms for the derivation of authorization keys in alignment with TLSv1.3 [RFC8446]
- The use of fragmentation-enabled messages to extend the max supported message size (MMM V5)

The core of the BPI+ V2 messages (i.e., the Auth Request and Auth Reply ones) is depicted in Figure 5 where the Auth Info message is omitted for clarity.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

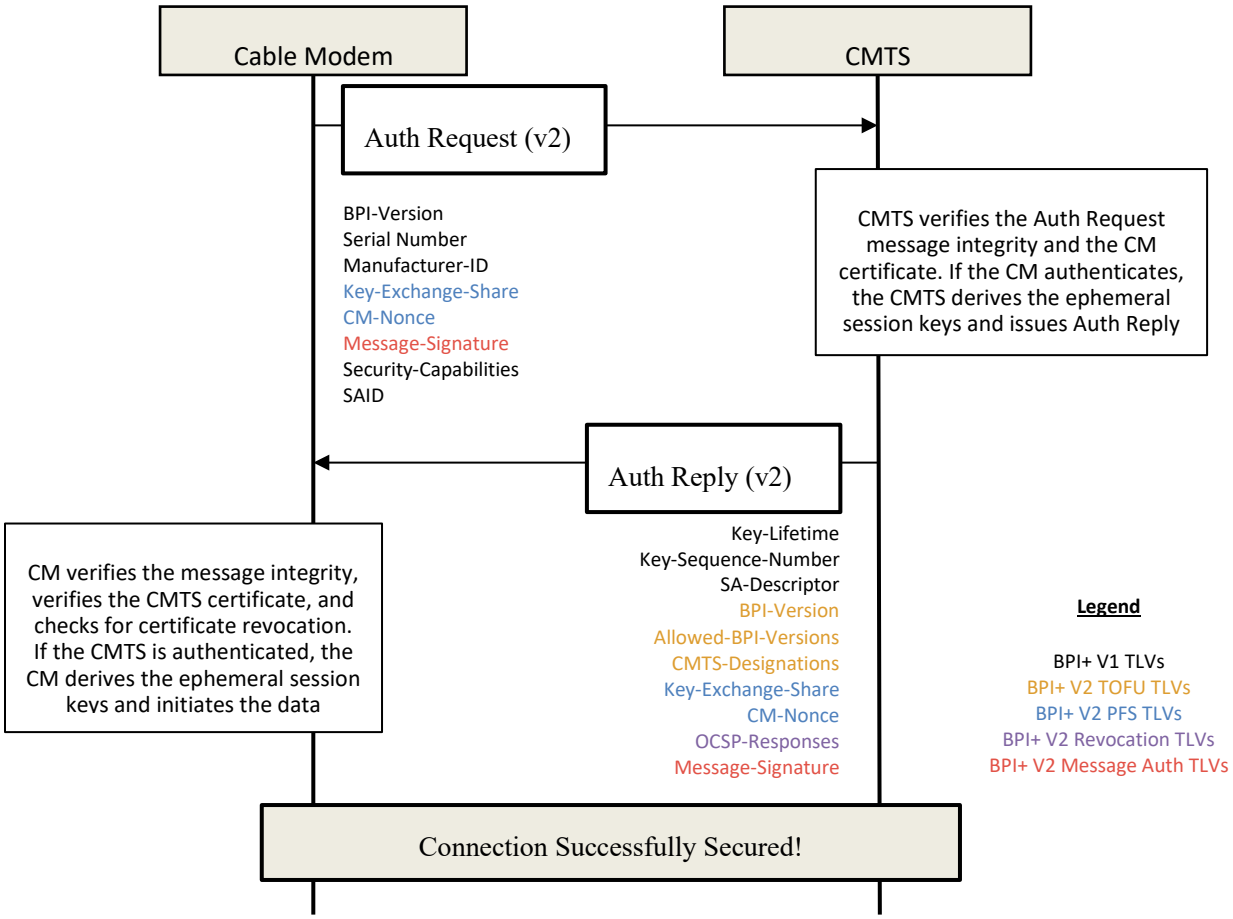


Figure 5 - BPI+ V2 authorization process (Auth Req and Auth Reply only)

3.2. The Simulation Framework

To better understand the impact of quantum-safe algorithms on core broadband algorithms, CableLabs built an experimental simulator for BPI+ V1 and V2, namely *LibBPLUS* framework.

Although the framework supports both versions of the BPI+ protocol, given the limitations of the type of MAC message used in the implementation of BPI+ V1 (i.e., MMM V1 that support maximum sizes smaller than 2000 bytes), BPI+ V1 is not suitable for drop-in replacement(s). Moreover, the explicit dependency on the RSA algorithm for delivering the authorization key to the CM (i.e., the lack of authenticated key exchange mechanisms) makes it impossible to use device certificates with quantum-safe public keys.

However, when it comes to BPI+ V2, all those limitations have been removed as described in the previous Section and the rest of the paper focuses on latest version of the protocol (i.e., BPI+ V2).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.2.1. The Framework in a Nutshell

The LibBPLUS framework has been developed to be easily extensible and easy to integrate in existing and future systems. Its main purpose is to provide an implementation of the BPI+ authentication protocol (both V1 and V2) that can be used to test different cryptographic and communication parameters, even experimental ones (i.e., BPI+ V1 and BPI+ V2 messages, authentication certificates, post-quantum cryptography, etc.). It is important to highlight that this work, LibBPLUS, is not meant for production use – the code is not optimized and is meant to only provide early access to a reference implementation, even when devices that support new versions of the protocol are not available yet.

The codebase comes with a shared library, a message generation and parsing tool, and a TCP-based simulator that implements both client and server side. The project depends on few open-source packages to be installed (SDKs are needed for compilation) such as LibPKI [LibPKI], LibOQS (optional, only needed for QS-support) [LibOQS], and finally OpenSSL-OQS (1.1.1) [OSSL-OQS] or OpenSSL (1.1.1) [OSSL].

The LibBPLUS library has been architected with the developer and researcher in mind by providing different abstraction layers and easy integration options. The core piece of the LibBPLUS project is distributed as a shared library that can be easily linked in C/C++ applications and/or Python/Java/Node/etc. wrappers. The library is organized in a set of abstraction layers that implement different layers of functionalities: from a very high-level abstraction layer that provides easy-to-use message creation functions, to the lower-level BPKM and TLV libraries that allow for customization of the messages (even outside the standardized parameters).

The abstraction layers are organized as follows, from the higher level to the lowest:

- **BPLUS_CTX** - A complete set of support tools for loading and using X.509 certificates, Keys, Configurations, Server Configurations, etc. It also provides the support layer for the authorization and Traffic Encryption Key or TEK state machines.
- **BPLUS_API** – A high-level set of functionalities to generate and manage BPKM messages. The API provides message-specific functionalities such as generating the public key for the key exchange mechanism or deriving the authorization key from Auth Request and Auth Reply messages.
- **BPLUS_BPKM** - Provides support for BPKM message generation, encoding, and decoding. Specifically, the BPKM sub-library provides the definition of BPKM messages (grouped by BPI+ version) and associated management functions.
- **BPLUS_TLV** - Provides low-level layer for TLV generation, encoding, and decoding. The TLV sub-library provides the definition of TLVs (also grouped by BPI+ version) and associated management functions.

Additionally, the LibBPLUS project comes with a fully fledge network simulator and message generator that is provided via a separate Command Line Interface (CLI) tool called ‘bplus’:

- **BPLUS Tool** - A tool that leverages the BPLUS library to provide a threefold implementation: a message generator and parser, a BPI+ Auth state machine for server-side implementations (TCP server), and a BPI+ Auth state machine for client-side implementations (TCP client).

The package comes with extensive tests (src/tests) and tools (src/tools) that show the usage of the library in different scenarios. Additionally, extensive documentation for the supported APIs for the different

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

abstraction layers is provided directly in the header files that modern IDEs such as VS Code made readily available to the developer[1].

3.1. Current Framework Limitations

As we hinted in the introduction to the section, the LibBPLUS framework is highly experimental and comes with some limitations that we are aware of and plan to remove in the next generation.

On the message generation and parsing front via the CLI application, the tool is still limited to be able to generate only few types of messages such as Auth Info, but it does not currently support the generation of all types of BPI+ messages.

On the simulation front, although the core of the simulator is designed to support multiple concurrent connections, the implementation is currently limited to handling a single individual authentication and encryption key delivery. After the Authorization State Machine (ASM) reaches the "Authorized" state and the TEK state machine reaches the "Operational" state, the simulation is terminated on both sides.

When it comes to implementation of quantum-safe testing, the dependency on the availability of implementations or, more precisely, the lack thereof, has proven to be a challenging problem to solve. In particular, the enhancements that we needed to provide for the crypto libraries can be summarized as follows:

- **Provide an implementation for hash-n-sign paradigm (Quantum Safe)**. This enhancement allows the crypto layer to be able to use the traditional model we are used to when it comes to signing. In particular, the provided enhancements allow to sign the CMS structure that is used for authenticating the messages (i.e., detached CMS signature) by pre-hashing the content and then signing the hash value (instead of signing the entire message directly). This was needed to get around OpenSSL's requirements to use the hash-n-sign paradigm for signing CMS structures.

- **Provide an implementation for Hybrid certificates (Composite Crypto).** This enhancement implements the Composite Crypto technology developed at CableLabs to support the testing of hybrid solutions. The availability of such technology is aimed at studying the impact of using multiple algorithms to lower the risk coming from the adoption of new cryptography without the need for changing the application logic to use, for example, multiple certificates.

On the front of Key Encapsulation Mechanisms (KEMs), there is still much work to do.

Since the OpenSSL-OQS wrapper does not support the generation and use of key pairs that use the Kyber algorithm (currently the only one selected for quantum-safe KEM standardization), the simulator does not currently have the option of replacing traditional key exchange mechanisms with quantum-resistant KEMs. This limitation is to be kept in mind during the discussion of the collected results that do not include the use of quantum-safe KEMs, even in the quantum-safe tests.

Finally, the simulator does not provide support for the encapsulation of BPKM messages (i.e, the BPI+ ones) inside MAC Management Messages. This means that the simulator does not create full MMM messages where BPKM ones are embedded as the payload of an MMM, but, instead, uses the TCP layer to transfer BPKM messages directly across the client and the server. Consequently, even for large BPI+ V2 BPKM messages the simulator does not have to fragment and then reconstruct messages that are

[1] The GITHUB repository is not currently publicly available. We are working to understand the implications (and required processes and improvements) for releasing the code to the community.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

larger than a single DOCSIS frame (e.g., 2000 bytes) since it relies on the fragmentation support built into the TCP/IP layer.

3.2. The Methodology

To collect the experimental data, we ran several simulations where we used different types of certificates and provide a comparison among the collected results. All experiments have been run on a MacBook Pro laptop (2.3 GHz 8-core Intel Core i9) with 16Gb of RAM. In this first phase of experimentation, we focused our metrics on two dimensions only: size of the authentication trace and time to execute the authorization process.

First, we ran BPI+ V1 and BPI+ V2 simulations to get baseline values that are used to build the comparison matrix. After that, we configured the simulator to use certificates from a quantum-safe PKI where we used CRYSTALS-Dilithium keys at level 2 (the lowest security level provided by this algorithm) for the whole chain [Dil17].

During the simulation, every message received or sent is automatically captured by the framework and saved both in binary and textual format on local storage for later analysis – although these operations added extra execution time that we needed to be aware of, it did not have an impact on the overall set of measurements. The framework also saves the contents of the `Message-Signature` TLVs of authenticated messages in separate files. Additionally, extensive logs are also being recorded that provide additional information on the execution of the protocol such as the selected key-exchange field (e.g., P256, X448, etc.) or possible issues with the validation of signatures and certificates.

Table 2 - Message sizes for BPI+ V1, BPI+ V2, and BPI+ V2 QS.

Message Type	BPI+ Version	Certificate Type	Message Size	Signature TLV
Auth Info	BPI+ V1	RSA	1,067	n/a
Auth Request	BPI+ V1	RSA	1,400	n/a
Auth Reply	BPI+ V1	RSA	291	n/a
Auth Info	BPI+ V2	RSA	1,067	n/a
Auth Request	BPI+ V2	RSA	1,703	1,564*
Auth Reply	BPI+ V2	RSA	2,809	2,689**
Auth Info	BPI+ V2	Dilithium (level 2)	4,242	n/a
Auth Request	BPI+ V2	Dilithium (level 2)	7,053	6,914*
Auth Reply	BPI+ V2	Dilithium (level 2)	11,318	11,198**

3.3. The Results

The results are summarized in Table 2 where we provide the results from the BPI+ V1 runs with traditional cryptography (RSA PKI) and BPI+ V2 with traditional cryptography (RSA PKI) and quantum-safe cryptography (CRYSTALS-Dilithium Level 2) for certificates and signatures. As we detailed in Section 3.1, the key exchange mechanism has not been updated to quantum-safe variations because of the limitations of the OpenSSL-OQS library.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.3.1. Considerations on Message Sizes

Since the size of crypto is one of the important factors for drop-in replacement of cryptography, let's first analyze the size of the authorization traces in Table 2 where different types of messages together with their sizes and signatures' sizes are reported. The signature size is the size of the encoded CMS data structure the implements the SigneData type (i.e., detached signature). In the table, we use the '*' (asterisk) symbol to remind the reader that the CMS structure also includes the signer's certificate in the `certs` field of the `SignedData` structure. Similarly, the '**' (double asterisk) symbol is used to remind the reader that the CMS structure that authenticates CMTS messages (e.g., the Auth Reply) not only includes the signer's certificate, but it also includes the CA one (i.e., the CMTS certificate's CA).

As expected, *BPI+ V1 messages with traditional cryptography* (i.e., RSA certificates) are the smallest ones. This is due to the fact that BPI+ V1 does not provide any direct authentication and, therefore, no signatures are generated and/or transferred across the wire.

For *BPI+ V2 authorization process with traditional cryptography* (i.e., RSA certificates), while the Auth Info message size is not changed from BPI+ V1 as it only carries the CA certificate, the Auth Request message size, instead, is increased roughly by the size of the message signature in the order of ~525 bytes. The size of the Auth Reply is also increased considerably. In particular, since BPI+ V2 provides mutual authentication, the Auth Reply message must include not only the CMTS' certificate and the CMTS' message signature, it also must include the CMTS certificate's issuing CA certificate. Therefore, while the size of Auth Request messages can stay below the 1976 limit for a non-fragmented MMM payload, the size of the Auth Reply goes beyond that limit having to include two certificates (i.e., the signer's and the CA's) in the CMS structure.

Ultimately, to test the possibility for drop-in replacement for BPI+ V2, we changed the configuration of the simulator to use X.509 certificates not from a traditional RSA infrastructure, but *from a quantum-safe PKI*. The new infrastructure has been setup as defined in the DOCSIS® PKI (the "new" PKI or 2nd Generation) [ClTi23] and DOCSIS 4.0 SEC specifications (i.e., the certificates' "profile"), except for the public-keys and signatures algorithm. In this case, we used the CRYSTALS-Dilithium algorithm at level 2 for the keys for all levels of the hierarchy (i.e., Root CA, Intermediate CA, and End-Entity). The collected results confirm the expected increase for all the messages in the Authorization process by an approximate factor of four (4). In fact, the Auth Info message is almost four times the size of the same message with traditional cryptography – the increase comes exclusively from the increased size of the CA certificate. The Auth Request message size, instead, is increased four fold because of the increased sizes of both the signature and the CM certificate. Similarly, the Auth Reply message is increased roughly four times because of the CA certificate's increased size together with signature size.

Putting everything together, we can consistently estimate that, in this configuration, quantum-safe BPI+ V2 authorization processes have a four-times increase (4x) in the size of transferred messages when compared to the same protocol using traditional cryptography.

3.3.1. Considerations on Execution Complexity

Although the simulator's is not optimized for speed, since it is in its first release, we can try to provide some simple comparison of the execution times for the BPI+ protocol when using the different configurations.

Before we provide any considerations on the results, though, it is important to remember the limitations of the simulator. Specifically, in the executed simulations we connected the two ends of the simulator via the TCP/IP interface on the localhost. This eliminates the need to account for the time needed to transfer

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the data over the wire and, therefore, should reduce the variability in reported measurements. Moreover, it is important to consider that both sides of the simulation (client and server) have been executed on the same host where measurements also have been captured via the simulator directly.

Let's first start our analysis by looking at the BPI+ execution with traditional cryptography. In this configuration, while the simulated BPI+ V1 (RSA PKI) takes approximately 50ms to transfer the Authorization Key from the CM to the CMTS, BPI+ V2 (RSA PKI) doubles that time. This is consistent with the use of signatures for both the CM and CMTS because of the increase in the number of cryptographic operations.

Among all the execution times for the different versions of BPI+ with different configurations, the results for the quantum-safe version of BPI+ V2 proved to be the most surprising. In fact, despite the increased size of cryptography, the run time for completing the delivery of the Authorization Key between the CM and CMTS (via ECDHE by using X448) is constantly shorter when compared to using traditional cryptography. Specifically, with the selected configuration, the overall authorization process execution time is reduced to almost half when compared with BPI+ V1. Similarly, when the quantum-safe BPI+ V2 execution time is compared with the traditional one, the quantum-safe one appears to be less than a fourth of the runtime needed to complete BPI+ V2 with RSA certificates (i.e., 20-25ms vs. 100-120ms).

Although interesting, these initial results need to be considered in the context of a simulation and its limitations. More work is envisioned to better understand these initial results.

4. Conclusions and Future Work

In this paper we looked at the *Future of Cryptography* for the broadband industry with the intent of providing important considerations on timelines and required processes for migrating the cryptographic algorithms we use today to quantum-safe solutions.

While in the first part of the paper we introduced the quantum-threat and looked at some practical considerations on the impact of quantum computers over encryption, in the second part of the paper we proposed the Mosca-Pala framework and use it to reason about migration timelines and required steps to execute them. The provided back-of-the-envelope estimations suggest that the industry might not have time to wait to be able to update and deliver quantum-safety across the corporate, access, and converged networks in the next decade (e.g., 10-15 years).

The third and last part of the paper focuses on providing considerations for quantum-safety and DOCSIS. For this work, we built a BPI+ simulator that we then used to simulate the authorization process and compare the results among traditional and quantum-safe implementations. While BPI+ V1 is inherently tied to traditional cryptography (RSA) and cannot be used with quantum-safe certificates (i.e., device identities), the removal of algorithmic dependency between the certificate and the key exchange mechanism makes BPI+ V2 a good candidate for drop-in replacement. In fact, the collected results on the simulation show how BPI+ V2 supports the use of quantum-safe cryptography, thus making BPI+ V2 and DOCSIS 4.0 suitable for algorithm-replacement. On top of that, our experiments show how the use of quantum-safe cryptography could be quite efficient when compared to the RSA algorithm, nevertheless the increased size of crypto-objects.

As described throughout our paper, the work is far from being completed though.

First, we need to better understand the efficiency of quantum-safe algorithms and update the simulator to be able to use quantum-safe KEMs (still not implemented). Once these updates are in place, we also need to investigate the use of algorithms other than CRYSTALS-Dilithium such as Falcon [Fa17], together

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

with the impact of using hybrid technologies such as Composite Crypto or the use of multiple certificates. Last but not least, we need to improve the measurement tools to better understand how the crypto-load might change between traditional and quantum-safe solutions for the different types of protocols, especially on the server-side (i.e., the CMTS, AAA, etc.).

Abbreviations

AES	Advanced Encryption Standard
BPI	Baseline Privacy Interface
BPI+	Baseline Privacy Interface Plus
BPKM	Baseline Privacy Key Management
CA	Certificate Authority
CRL	certificate revocation list
CRQC	Cryptographic Relevant Quantum Computer
DER	Distinguished Encoding Rules
DN	X.500 Distinguished Name
DOCSIS	Data Over Cable Service Interface Specifications
EC	Elliptic-Curves
ECC	Elliptic-Curves Cryptography
ECDH	Elliptic-Curves Diffie-Hellman
ECDSA	Elliptic-Curves Digital Signing Algorithm
EE	end entity
DH	Diffie-Hellman
HSM	hardware security module
IETF	Internet Engineering Task Force Standards Organization
KEM	key encapsulation mechanism
KEX	key exchange (algorithm)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

NIST	National Institute of Standards and Technologies
P256	NIST EC Curve at 128bit of security
PKC	public-key cryptography
PKI	public-key infrastructure
QC	quantum computing
QS	Quantum Safe
R-PHY	Remote RF Layer (PHY)
R-MACPHY	Remote Media Access Control and RF Layer (PHY)
RSA	Rivest-Shamir-Adleman (cryptosystem)
SHA-256	Secure Hash Algorithm 2 (256-bit)
SHA-3	Secure Hash Algorithm 3
SCTE	Society of Cable Telecommunications Engineers
SE	Secure Element
SSD	Secure Software Download
TEK	Traffic Encryption Key
TLS	Transport Layer Security
TTQS	Time To Quantum Safety
X448	EC Curve for ECDHE that provides 192 bits of classical security
X25519	EC Curve for ECDHE that provides 128 bits of classical security

Bibliography & References

[Ec05] American National Standards Institute, *Public Key Cryptography for the Financial Services Industry: the Elliptic Curve Digital Signature Algorithm (ECDSA)*, ANSI X9.62, November 2005.

[Rsa16] The Internet Engineering Task Force (IETF) – IETF RFC 8017. PKCS #1: RSA Cryptography Specifications Version 2.2, edited by K. Moriarty et al., November 2016. Also available at https://datatracker.ietf.org/doc/rfc8017/

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

[Gro96] Lov K. Grover. *A Fast Quantum Mechanical Algorithm for Database Search.* In Gary L. Miller, editor, Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing, Philadelphia, Pennsylvania, USA, May 22–24, 1996, pages 212–219. ACM, 1996.

[3GPP23] GSM Association, Post Quantum Telco Network Impact Assessment Whitepaper, Version 1.0, February 2023. Available at: https://www.gsma.com/newsroom/wp-content/uploads/PQ.1-Post-Quantum-Telco-Network-Impact-Assessment-Whitepaper-Version1.0.pdf

[Deakin23] I. Deakin, *Implications of Entropy on Symmetric Key Encryption*. February 2023. ATIS. Available online at: https://www.atis.org/?smd_process_download=1&download_id=1783427

[Car20] Chujiao Ma and others, *CARAF: Crypto Agility Risk Assessment Framework*, Journal of Cybersecurity, Volume 7, Issue 1, 2021. Available online at: https://doi.org/10.1093/cybsec/tyab013

[Pala23] The Internet Engineering Task Force (IETF) – I-D draft-ounsworth-pq-composite-sigs-09 - Composite Keys and Signatures For Use In Internet PKI, edited by M. Ounsworth and M. Pala, Apr 2023. Also available at https://datatracker.ietf.org/doc/draft-ounsworth-pq-composite-sigs/

[X509] ITU-T Recommendation X.509 (2005) | ISO/IEC 9594-8:2005, *Information Technology - Open Systems Interconnection – The Directory: Public-key and attribute certificate frameworks*.

[Dent02] Alexander W. Dent, *A Designer's Guide to KEMs*, Cryptology ePrint Archive, 2002. Available online at: https://eprint.iacr.org/2002/174

[Mosca15] Michele Mosca, Cybersecurity in an era with quantum computers: will we be ready?, Cryptology ePrint Archive, 2002. Available online at: https://eprint.iacr.org/2015/1075.pdf

[BPI+08] Data-Over-Cable Service Interface Specifications, Baseline Privacy Plus Interface Specification, CM-SP-BPI+-C01-081104, November 4, 2008.

[SECv4.0] DOCSIS 4.0 Security Specification, CM-SP-SECv4.0-I04-220328, March 28, 2022, Cable Television Laboratories, Inc.

[Pala20] Massimiliano Pala. DOCSIS® PKI: A Proposal for a Next-Generation Quantum-Resistant Infrastructure. SCTE Cable-Tech Expo, October 2020.

[Pala21] Massimiliano Pala. Enabling Encryption and Algorithm Revocation for Post-Quantum DOCSIS Certificates. SCTE Cable-Tech Expo, September 2021.

[RFC8446] The Internet Engineering Task Force (IETF) – IETF RFC 8446. *The Transport Layer Security (TLS) Protocol Version 1.3*, edited by E. Rescorla, Aug 2018. Available online at: https://datatracker.ietf.org/doc/html/rfc8446

[LibPKI] The LibPKI project. https://www.openca.org/projects/libpki

[OSSL] The OpenSSL project. https://www.openssl.org

[LibOQS] The Open Quantum Safe project. https://www.liboqs.org

[OSSL-OQS] The OpenSSL-OQS project. https://www.openssl.org

[OCAOD] The OpenCA OCSPD project. https://www.openca.org

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

[RFC5652] IETF RFC 5652, R. Housley, *Cryptographic Message Syntax (CMS)*, September 2009.

[Di17] Dilithium-Crystals, *Dilithium digital signature scheme.* Available online at: https://pq-crystals.org/dilithium/

[Fa17] Falcon, *Fast Fourier Lattice-based Compact Signatures over NTRU*. Available online at: https://falcon-sign.info

[ClTi23] CableLabs C-PKI-TI-V1.4, Trust Infrastructure Document (Certificate Templates), November 2022. Available online at: https://www-res.cablelabs.com/wp-content/uploads/2022/12/13105005/CableLabs-Trust-Infrastructure-C-PKI-TI-V1-4-2022-11-15-WEBSITE.pdf

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

FWA, The Promise of Broadband Everywhere Lessons Learned & Best Practices

A Technical Paper prepared for SCTE by

Ahmed Jafri
Manager Wireless Access Network
Rogers Communications Inc.
8200 Dixie Rd. Brampton ON
Ahmed.Jafri@rci.rogers.com

Osama Bazan
Principal RAN Designer
Rogers Communications Inc.

Rizwan Butt
Manager Insights & Analytics
Rogers Communications Inc.

Mahmoud ElMeligui
Principal Designer
Rogers Communications Inc.

Carlos Moco
Manager RAN Optimisation
Rogers Communications Inc.

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. State of FWA in Canada 5
3. Best Practices & Processes for FWA 6
   3.1. Serviceability and Coverage Qualification 6
   3.2. Capacity Management 8
      3.2.1. Estimating User Capacity Requirements 8
      3.2.2. Available Capacity 9
   3.3. Serviceability Recommendations 13
   3.4. Device Considerations 14
   3.5. Fair Usage Policy 16
4. Network Management Features for FWA 16
   4.1. FWA Recognition in the RAN 16
      4.1.1. UE-level Identification 17
      4.1.2. Bearer/Flow-level Identification 17
   4.2. Carrier Management 18
      4.2.1. Limiting FWA Utilization of Coverage Layers 18
      4.2.2. Prioritizing FWA utilization of the spectrum licensed only for Fixed Wireless Services 19
   4.3. Radio Resource Management 19
      4.3.1. FWA-Mobile Relative Priority Scheduling 19
      4.3.2. Differentiated Admission Control 21
      4.3.3. Differentiated Uplink Power Control 21
   4.4. Revisiting the Battery Power Saving Features 22
      4.4.1. Discontinuous Reception in Connected Mode (C-DRX) 22
      4.4.2. User Inactivity Timer 22
5. Case Studies 22
   5.1. Case Study 1 – Consequences of Low Band Utilization 22
   5.2. Case Study 2 – Massive MIMO 24
6. FWA: Lessons Learned, Prospects & Future 25
   6.1. FWA Recommendations 25
   6.2. Urbanization of FWA 27
   6.3. Future Developments in FWA 28
7. Conclusion 28

Abbreviations 29

Bibliography & References 30

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Figures

Title **Page Number**

Figure 1 – Underserved Households across Canada 5
Figure 2 - Rural FWA Coverage 7
Figure 3 - Impact of Cell Loading on SINR for 8x8 and 32x32 radios 9
Figure 4 - Impact of Spectral Efficiency on Capacity 11
Figure 5 - High Traffic High Efficiency Site 11
Figure 6 - Low Traffic Low Efficiency Site 12
Figure 7 - Spectral Efficience & Traffic 12
Figure 8 - Impact of user distribution across a cell on capacity 13
Figure 9 - Indoor & Outdoor CPE Setup 15
Figure 10 - Relation between High Band RSRP and distance from site 16
Figure 11 - FWA users sharing the capacity layers with mobile users 18
Figure 12 - FWA traffic prioitrized on FWA-exclusive 5G NSA carrier with no mobile access 19
Figure 13 - FWA vs Mobile: Data Volume and Throughput 20
Figure 14 - FWA vs Mobile: Number of Users and Resource Block Utilization % 21
Figure 15 - Aggregated Bandwidth Based on Serving Band 23
Figure 16 - FWA Network Level Statistics 23
Figure 17 - Average Spectral Efficiency of Massive MIMO vs Non-Massive MIMO 24
Figure 18 - MU-MIMO Usage: All FWA Nodes vs Mobile Nodes 25
Figure 19 - MU-MIMO Usage Trend for FWA Serving Nodes 25

List of Tables

Title **Page Number**

Table 1 - FWA Link Budget Assumptions & Considerations 6
Table 2 - Indoor vs Outdoor CPE Comparison 14

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

In today's digital age, access to high-speed internet has become critical for individuals, businesses, and communities to thrive and participate fully in the modern economy. However, despite significant technological progress and infrastructure development, the digital divide persists, leaving a considerable number of households underserved or unserved by broadband connectivity. This paper explores the potential of Fixed Wireless Access (FWA) as a solution on 4G/5G networks to bridge this divide, focusing on the Canadian context while acknowledging the global nature of the issue.

Canada, a developed country with an estimated population of 16.2 million households, faces challenges in providing universal broadband access. According to the Canadian Radio-television and Telecommunications Commission (CRTC), approximately 91.4% of households have access to broadband speeds of at least 50 Mbps download and 10 Mbps upload. However, this leaves around 1.4 million homes as underserved or unserved, with rural communities being particularly affected, where only 62% have access to these reasonable broadband speeds. This persistent digital divide highlights the multifaceted challenges associated with socioeconomic, political, and geographic factors that shape broadband accessibility.

Fixed Wireless Access is not a novel concept, having existed in various technological forms where the last mile connection provided to users is wireless. However, as the demand for data consumption on broadband connections continues to increase, delivering a satisfactory customer experience over wireless networks has become increasingly challenging. The convergence of mobile and fixed wireless broadband services over 4G networks began with the expansion of 4G cellular networks. But bridging the digital divide requires ensuring FWA can deliver an experience comparable to fixed broadband. While 4G based FWA has proven to be promising, it has fallen short particularly in terms of being able to provide comparable throughput speeds and adequate capacity.

However, 5G networks have emerged as a more robust and appropriate solution for applications like FWA while catering to the growing needs of mobile traffic. This is mainly due to the advanced features and capabilities 5G brings such as access to higher bandwidths and frequency spectrum, more efficient radio interface, enhanced radio resource management, higher order Multi-User Multiple Input Multiple Output (MU-MIMO), Beamforming, Dual Connectivity and Carrier Aggregation (CA) for peak performance and coverage extension, higher order modulation, and more capable customer premises equipment (CPE).

It is essential to recognize that FWA users typically consume significantly more data compared to mobile users, often by a factor of 40 to 50 times. Additionally, seasonal traffic variations introduce complexities that necessitates careful capacity planning and resource management. This requires operators offering FWA services over their 4G/5G network to not only rely on the enhanced features now available with 5G, but also to adopt processes and best practices that will protect their network, improve customer experience, provide sustainability, and ensure success.

This technical paper aims to cover the features, processes and best practices utilized by Rogers to accommodate FWA as an application over its 4G/5G Mobile Broadband (MBB) network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2. State of FWA in Canada

Wireless coverage for 4G extends to 99.4% of the population. Technically FWA can be made available to them; however, not all areas with 4G coverage are able to provide or support speeds of 50 Mbps on the downlink and 10 Mbps on the uplink. It is important to note, while many communities within Canada are served, the high cost of these services makes broadband connectivity inaccessible to many households. Due to these socio-economic factors, it is necessary for service providers to build efficient and cost-effective networks to make high speed internet connections readily available and accessible.

Figure 1 map highlights areas across Canada that have approximately 1.4 million underserved households. As it can be deduced, to capture this FWA opportunity, a very large area needs to be addressed, which makes building efficient networks ever more challenging.

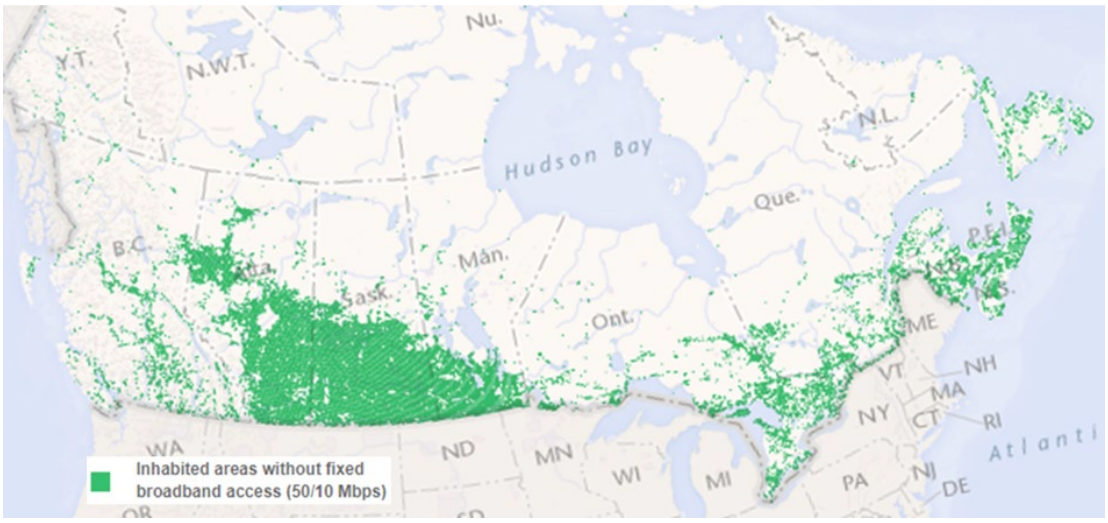


Figure 1 – Underserved Households across Canada

Low population density appears to be a big challenge for Canada. To put this into perspective, Canada only has four 4 people per square km, while USA has 34 people per square km.

Therefore, to bridge this digital divide, the government has initiated broadband fund programs to subsidize the build of broadband connectivity to these underserved communities. This primarily includes fiber build but also considers FWA as a viable alternative option.

The success of FWA relies on the support of clear and transparent regulatory policies governing spectrum licensing and usage. The Canadian Radio-television and Telecommunications Commission (CRTC) plays a crucial role in promoting fair competition and universal access, encouraging operators to invest in FWA deployments. Therefore, FWA deployment has gained momentum across Canada with various operators utilizing their existing 4G/5G mobile networks.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Best Practices & Processes for FWA

This section will cover best practices and recommended processes based on Rogers' learnings and experience from onboarding FWA customers to its 4G/5G MBB network.

3.1. Serviceability and Coverage Qualification

Typical network strategy is to offload heavy traffic on to higher capacity or larger bandwidth channels. 3.5GHz (n78) being the primary Frequency Range 1 (FR1) 5G high bandwidth band available in Canada, FWA traffic is preferred to be steered towards this band, if available. Therefore, when it comes to a serviceability check for customers, it is desirable to base it on the coverage of this 5G band or other 4G high bands.

The CRTC defines households as underserved if they do not have access to broadband service offerings of at least 50 Mbps on the downlink (DL) and 10 Mbps on the uplink (UL). To provide the mandated service levels, appropriate coverage thresholds should be considered depending on the operator's network configuration. A detailed Link Budget analysis is required that takes into consideration the spectrum available to the operator, Radio hardware capabilities, and end user equipment specifications. The table below shows an example of a Link Budget analysis for a given scenario with the given assumptions.

Table 1 - FWA Link Budget Assumptions & Considerations

Table Heading	Scenario 1	Scenario 2
Technology	NR TDD	
Frequency	3.5 GHz	
Band	N78	
DL/UL Ratio (for TDD)	70/30	
Implementation (Mobile/FWA)	FWA	
Radio	32T32R (Massive MIMO)	
Power	200 W	
Antenna Gain	N/A	
EIRP	70 dBm	
UE Antenna Gain	16 dBi	
UE TX Power	26 dBm	
DL MIMO (SU)	4x4	
UL MIMO	1x4	
Modulation Supported (DL/UL)	256 QAM/64 QAM	
Bandwidth	60 MHz	
Morphology	Rural	
Site Height	40 m	
UE Height	4 m	
Environment (Indoor/Outdoor)	Outdoor	
LOS/NLOS	NLOS	
Coverage Probability	85%	
Cell Loading	80%	
Cell Edge DL Throughput Target	50 Mbps	25 Mbps
Cell Edge UL Throughput Target	10 Mbps	5 Mbps
Limiting Channel	PUSCH	PUSCH
MAPL	155.4 dB	159.6 dB

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table Heading	Scenario 1	Scenario 2
Min RSRP Threshold (Cell Edge)	-101.3 dBm	-105.5 dBm
- 0 dB log normal fading (LNF) margin was applied since receiver is static - Considering high cell ranges and directional antennas on the CPE, UL interference margin considered was <0.5 dB - Propagation loss for a RMa (Rural Macro) environment considered - CPE considered to be Power Class 2 (PC 2) for N78		

For the given cell edge throughput targets, it can be observed that the UL data channel (PUSCH) is the limiting channel while there is still substantial headroom available on DL data channels. It is critical for operators to calculate their link budgets based on the site configuration and the equipment deployed as well as align with vendor link curves for radio hardware used. The minimum cell edge RSRP threshold is then calculated for desired service levels, which operators can then use as a reference or as an input to their preferred RF propagation modeling tools. It is recommended to use industry standard tools that are capable to model exact network and site configurations and use up to date terrain, clutter, building and foliage data. This will allow operators to precisely identify serviceable areas and accurately plan for their FWA deployment.

Rogers has worked extensively with its vendor partners to have tuned propagation models and then calibrated those models for FWA deployments, accommodating higher CPE heights and antenna gains. Field measurement validation has shown a correlation of > 0.7 between measured and predicted values when within 3 km range of the site. Beyond that accurate alignment of CPE towards the best server becomes more challenging considering the environmental and clutter unknowns contributing to the variations. As per industry standards, propagation models when tuned are converged to have a mean error of < 2 dB and a standard deviation of < 7 dB. Therefore, adjusting propagation models and tool settings should be a continuous improvement process undertaken by operators along with continuous investment in highly accurate and up to date geo data sets. Figure 4.12 shows an example of a propagation analysis done for a cluster of sites planned in a rural area for FWA.

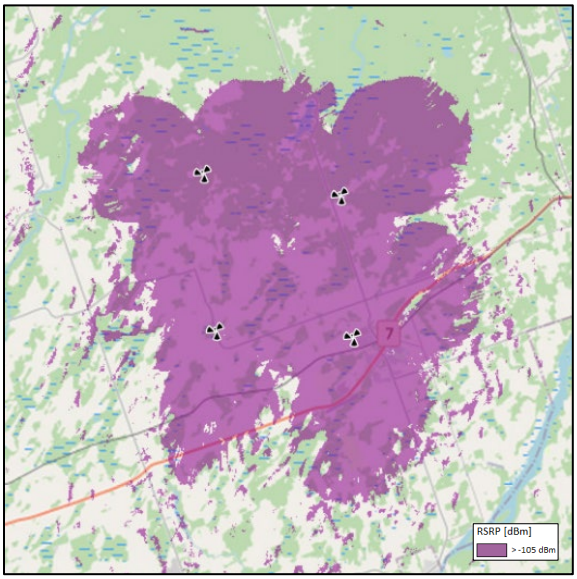


Figure 2 - Rural FWA Coverage

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.2. Capacity Management

As service providers embark on offering FWA services over their cellular network, capacity planning becomes an even more critical component of their cellular network strategy. Careful capacity planning ensures that the network is designed and optimized to manage the growing demands of FWA subscribers, provides subscribers with a high-quality experience, and allows operators to make informed decisions regarding network expansion and upgrades.

By analyzing historical data, traffic patterns, seasonal variations, and user behavior, operators can identify peak usage times, areas with high demand, and any potential bottlenecks that could lead to congestion and poor experience. Similarly, it is critical to forecast factors like subscriber growth, market penetration, year over year increase in data usage, and emerging applications so that operators can make provisions for expanding network capacity accordingly. Effective planning also helps with cost optimization which allows operators to avoid over provisioning that may cause unnecessary capital investment or under-provisioning that would lead to performance issues and dissatisfied customers.

The strategy for each operator may differ depending on the type or grade of services offered, dedicated, or shared spectrum available, network sharing with other services such as mobile, expected market penetration and product offerings. This section will focus on how to estimate capacity requirements and available capacity on a network that serves both mobile and FWA users.

3.2.1. Estimating User Capacity Requirements

There are several frameworks available to estimate user requirements. Utilizing usage statistics has proven to be effective and widely adopted, which will be discussed in this section. The following details are required:

- Average monthly data usage $\mathbf{D_{Avg}}$. This can be further categorized into the type of plans offered to customers. E.g., typical monthly usage observed in a market on a 25DL/5UL plan is ~500 GB/Month
- Average duration of daily peak usage times $\mathbf{T_{Peak}}$. E.g., high traffic or peak data consumption is observed on the network between 6 PM to 10 PM [4 hours]
- Percentage of traffic during peak hours $\mathbf{P_{\%}}$. An 80-20 rule can be applied here which can be interpreted as 80% of the traffic is observed during 20% of the time

With the above variables known, average user consumption during peak hours $\mathbf{U_{Avg}}$ can be determined as follows:

$$\boldsymbol{U_{Avg.}} = \left[\frac{\left(\boldsymbol{D_{Avg.}} \times \boldsymbol{P_{\%}}\right)}{\left(\boldsymbol{T_{Peak}} \times \mathbf{30}_{\boldsymbol{Days}} \times \mathbf{60}_{\boldsymbol{min}} \times \mathbf{60}_{\boldsymbol{sec}}\right)}\right] \times \mathbf{1000}_{\boldsymbol{MB}} \times \mathbf{8}_{\boldsymbol{bits}}$$

e.g.

$$U_{Avg.} = \left[\frac{(500GB \times 80\%)}{(4.8 \times 30_{Days} \times 60_{min} \times 60_{sec})}\right] \times 1000_{MB} \times 8_{bits} = 6.2Mbps$$

Therefore, the capacity on the network is not necessarily determined by the plans or bandwidth offered, but by the expected consumption of network resources. This makes it important to introduce policies that would limit or influence the usage pattern of customers.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.2.2. Available Capacity

Available capacity can be calculated on cell level that can then be aggregated to sector level, site level or market level capacity. There are several different approaches that can be adopted. However, utilizing channel spectral efficiency simplifies many aspects of capacity calculation and will be the one discussed here. One of the main assumptions associated with it is that any new customers added within the cell coverage range will conform to the same statistical distribution as existing users or traffic. They will experience radio conditions that are reflective of the spectral efficiency used, which can either be a network average, cell average, cell average during busy hours or spectral efficiency with any other criteria. This approach blends in all the radio parameters, such as Signal to Interference Noise Ratio (SINR), Rank Indicator (RI), Channel Quality Indicator (CQI), Modulation Coding Scheme (MCS), Block Error Rate (BLER), that have a direct impact on achievable throughput. Therefore, there are limitations of using spectral efficiency which will be examined later in this section as well as possible ways to address them.

Realistically, however, with each added user, there is increased interference within the cell and on adjacent cells, which should be factored in separately. The link budget analysis done initially accounts for the impact of a given cell load (80% in the above example) and reflects that in the required RSRP for the target throughput desired. For further accuracy, actual cell loading can be taken into consideration and align it with vendors' product capabilities of mitigating interference, which could either be through interference cancellation soft features or hardware capabilities such as beamforming. Figure 3 shows the impact of cell loading on SINR across the cell for two different 3.5 GHz radios. As the loading increases, the delta of achievable good SINR (>20 dB) increases between the non-beamforming 8x8 radio and the massive MIMO beamforming 32x32 radio. The impact of massive MIMO on spectral efficiency is further discussed in detail in case study 2 in section 5.2

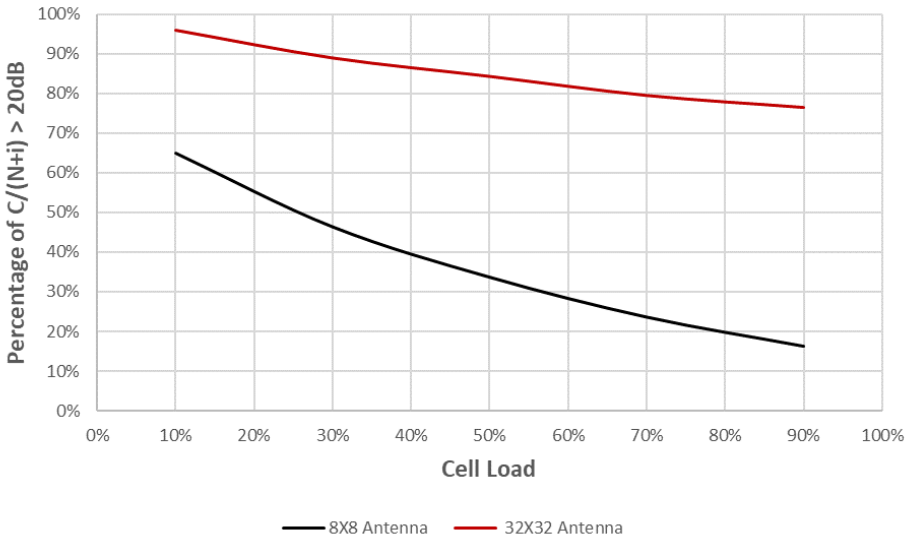


Figure 3 - Impact of Cell Loading on SINR for 8x8 and 32x32 radios

Existing usage, which may either be mobile usage or FWA usage, is accounted for by taking current average Physical Resource Block (PRB) utilization during peak hours. Only the remaining resources are then considered as available capacity. Maximum allowed PRB utilization threshold sets the upper limit

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

before a sector is considered as congested. Operators can decide to relax or tighten this limit depending on the market type and the amount of headroom they prefer to keep.

To efficiently serve a large customer base, it is common and economically feasible for operators to oversubscribe. Oversubscription ratio is a term used to describe the practice of selling more network resources such as bandwidth or data capacity to customers than what is available on the network. Not all users use their capacity simultaneously, therefore a concurrency factor or appropriate oversubscription ratios can be determined by analyzing historical data and customer usage patterns. Oversubscription should be carefully managed to avoid potential issues of network congestion and poor user experience during peak hours. The objective is to ensure maximum number of users to experience satisfactory performance without any noticeable degradation.

In the equation below, we aggregate the capacity to sector level.

∴ for x number of channels on sector S:

C_S = Sector Capacity (Mbps)

P_t = Max PRB threshold considered before congestion (%)

P_k = Average PRB utilization for channel k during peak hours (%)

B_k = Channel Bandwidth of channel k (MHz)

S_k = Average Spectral Efficiency of channel k (bits/sec/Hz)

R_{TDD} = Downlink Ratio for TDD channels only

I = Interference factor due to added traffic {I < 1}

U_{Avg} = Average usage during busy period (Mbps)

OSR = Oversubscription ratio {OSR ≥1}

N = Number of users that can be supported on sector S with U_{Avg} usage

$$C_S = \sum_{k=1}^{x} \left(\left(\frac{P_t - P_k}{100} \right) B_k S_k R_{TDD} I \right) OSR$$

$$N = C_S \div U_{Avg.}$$

Figure 4 shows the impact spectral efficiency has on the number of users that can be supported on a 20 MHz channel as the usage pattern changes. An OSR of 2.5:1 is considered, and the cell is assumed to be 50% loaded with 30% resources still available before congestion occurs.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

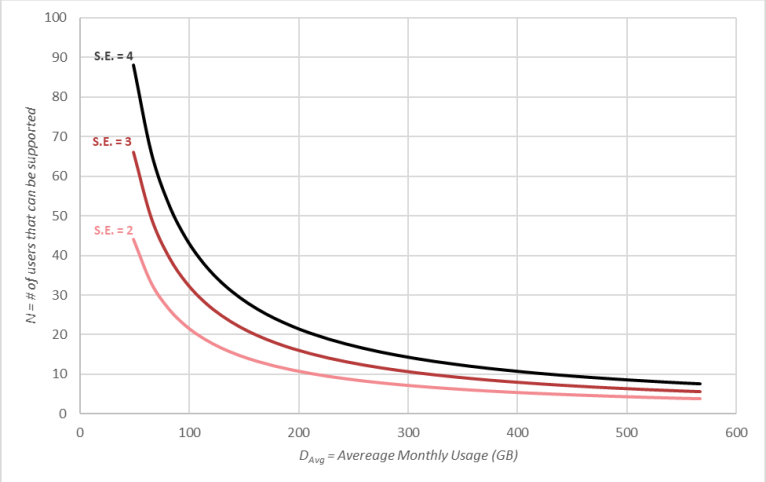


Figure 4 - Impact of Spectral Efficiency on Capacity

3.2.2.1. Limitations of Spectral Efficiency

While spectral efficiency is a good indicator of how well network resources are utilized, it can sometimes be misleading if there is no demand on the network. Evaluation was performed for congested cases and cases with low traffic demand.

The graph in Figure 5 shows a typical case where high data consumption during busy hours drives congestion and interference, resulting in reduced spectral efficiency. As utilization decreases, the interference reduces, and the spectral efficiency improves.

The site shown here in Figure 5 has 10MHz of high band LTE deployed on three sectors and is serving ~60 FWA users that account for almost 70% of the total traffic, the remaining being mobile traffic. Average distance of FWA users from the site is approximately 3.5 km with 20% users at cell edge beyond 6 km from the site. Weekly FWA download volume incurred is around 4.5 TB

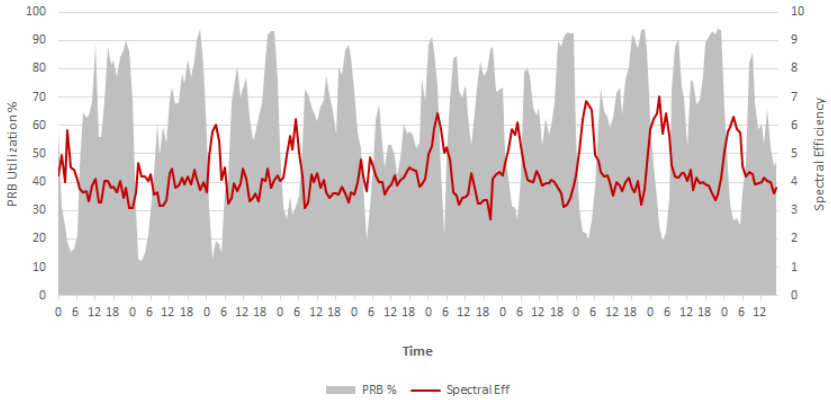


Figure 5 - High Traffic High Efficiency Site

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The graph in Figure 6 ,however, shows a case which is less loaded but also with lower than expected, spectral efficiency. This is because most of the FWA traffic is generated from cell edge. As the traffic reduces, there is not much impact on spectral efficiency because there is not enough demand to efficiently utilize all resources. In these cases, spectral efficiency achievable during non-busy hours could potentially be much higher.

The site shown here in Figure 6 has 10 MHz of high band LTE deployed on three sectors and is serving ~25 FWA users that account for almost 36% of the total traffic, the remaining being mobile traffic. Average distance of FWA users from the site is approximately 5 km with 30% users at cell edge beyond 10 km from the site. Weekly FWA download volume incurred is around 2 TB

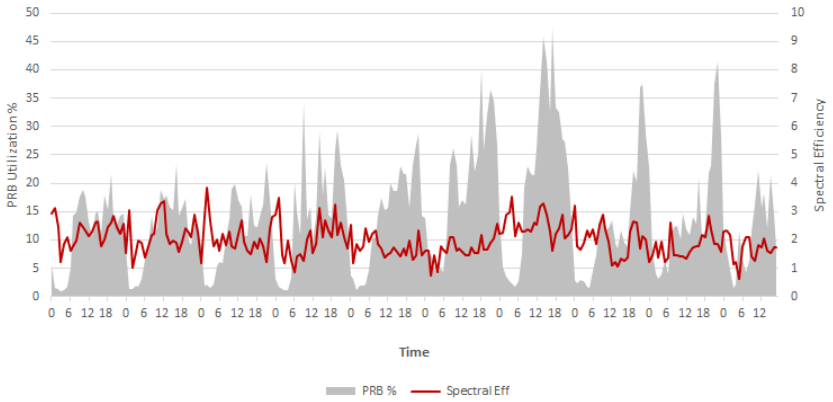


Figure 6 - Low Traffic Low Efficiency Site

Figure 7 shows how variation in demand and average RF conditions impacts spectral efficiency and could therefore impact capacity analysis. It is important for operators to consider these cases and determine whether the spectral efficiency considered is a true representation or not.

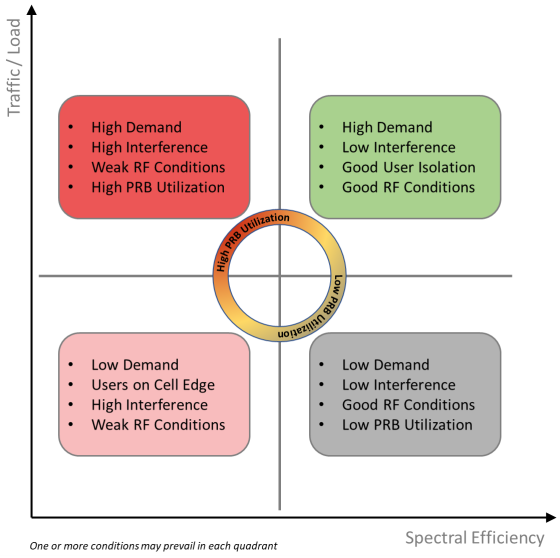


Figure 7 - Spectral Efficience & Traffic

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.3. Serviceability Recommendations

To deem a customer as serviceable, both coverage in the form of RF conditions and capacity play an integral role. RF conditions dictate how efficiently network resources can be utilized and that efficiency dictates how much capacity we obtain.

The following factors have significant influence on serviceability:

- **Technology deployed:** 5G in comparison to LTE can have up to 6% more resources blocks available for the same bandwidth at 15khz sub carrier spacing (SCS). 5G also allows for higher bandwidths and more flexibility on the air interface in terms of higher SCS and shorter timeslots.
- **Radio conditions:** Ability to achieve higher order modulation such as 256 QAM
- **Spectrum:** Available bandwidth and number of channels
- **CPE capability:** User Equipment (UE) category and type, e.g., equipped with high gain directional antenna, modulation and MIMO order supported, indoor in comparison to outdoor
- **Radio/Base Station capability:** MIMO streams or number of DL & UL streams supported. MU-MIMO capability increases spectrum efficiency tremendously, however, to achieve the right conditions to utilize MU-MIMO is very challenging
- **Data consumption / Traffic pattern:** Busy hour usage or data volume observed on the channels being considered
- **User distribution:** How are users spread across the coverage of the cell

All these factors significantly impact cell capacity and coverage reach. While it is possible to get a reasonable view of how many users can be supported on a given cell, the actual numbers will always vary depending on the impact on resource utilization with the addition of each user.

Figure 8 illustrates the concept of how the calculated capacity of X users with a given spectral efficiency can be impacted if users are added in favorable or unfavorable conditions.

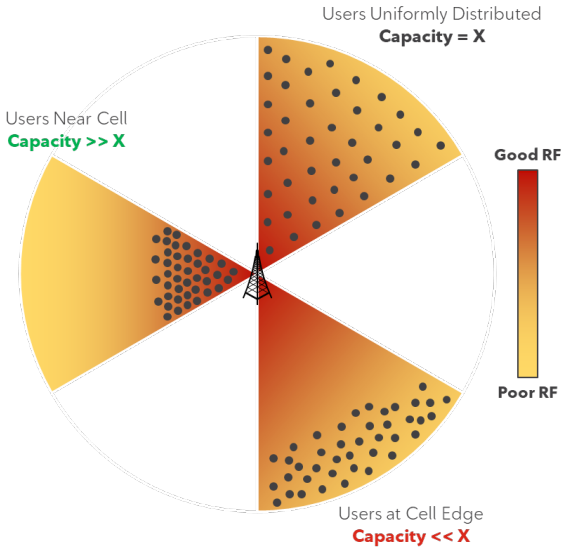


Figure 8 - Impact of user distribution across a cell on capacity

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Therefore, when adding FWA users to a network, it is critical to target users that are:

- Located near the site, preferably with LOS (line of sight)
- Within good RF conditions of high band spectrum
- Using an outdoor CPE with a high gain directional antenna installed at a good height
- Served by cells that have high channel bandwidth and use advanced radios capable of MU-MIMO and beamforming
- Offered plans that control user behavior to avoid misuse or abuse of network resources. E.g., limit plans to advertised throughput speeds such as 50DL/10UL or 25DL/5UL. Or impose data volume caps or enforce throttling after a certain amount of monthly consumption.

3.4. Device Considerations

Device or CPE selection is critical as its capabilities determine the efficiency with which network resources can be utilized. With the device landscape maturing for 5G, it is recommended for operators to offer devices that support:

- 4G & 5G, including Dual Connectivity for supporting 5G-NSA (EN-DC)
- Carrier aggregation with 3, 4 or 5 component carriers depending on technology & network mode
- Higher order MIMO, supporting up to 4 downlink MIMO layers. Devices are also beginning to support up to 2 uplink MIMO layers
- 256 QAM on both downlink and uplink
- High channel bandwidth (individual and aggregated)
- 26 dBm maximum output power, power class 2 support on higher bands for better uplink performance
- Device management and performance measurement via TR-069, TR-143, TR-390
- External high gain directional antenna for better reception

However, service providers are often challenged with the kind of CPE that can be offered to their target markets. Primarily it is a choice between an indoor and outdoor device, and there are various factors and CPE attributes that operators need to evaluate to make that decision. These factors also lead to compromises on the features listed above. Table 2 - Indoor vs Outdoor CPE Comparison covers some of them.

Table 2 - Indoor vs Outdoor CPE Comparison

Attribute	Indoor CPE	Outdoor CPE
Form Factor	Typically, the size of a regular indoor wireless router	Bulky, like an outdoor access point, depends on antenna size
Installation	Simple, quicker, and easier to install. Do it yourself (DIY) job for customer	Complex, needs to be mounted securely and aligned towards the best server, usually requires a professional install
Coverage & Range	Limited coverage and signal reception due to in-building penetration losses, higher interference from other close by devices and lower antenna gains	Extended coverage. Outdoor placement allows for direct line of sight with the base station, which leads to better signal reception. Directional high gain antenna also leads to better isolation from other devices for lower interference and improved signal quality

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Attribute	Indoor CPE	Outdoor CPE
Performance	Peak throughput performance is impacted because it is usually equipped with smaller lower gain antennas which causes weaker signal reception. Its small factor also often results in reduced capabilities such as number of downlink MIMO layers supported.	With the use of large higher gain antennas and better signal reception, peak throughputs are easily achievable.

Outdoor CPEs, professionally installed, are generally preferred for their better signal reception, extended coverage, and higher performance particularly in rural or challenging deployment scenarios. They help with improving spectral efficiency and put less stress on network capacity. However, indoor devices are still offered for better market penetration and customer acceptance because of their quick and simple DIY installation, aesthetic designs, and flexibility to place them anywhere where there is coverage. Figure 9 gives a general FWA setup for indoor and outdoor CPEs.

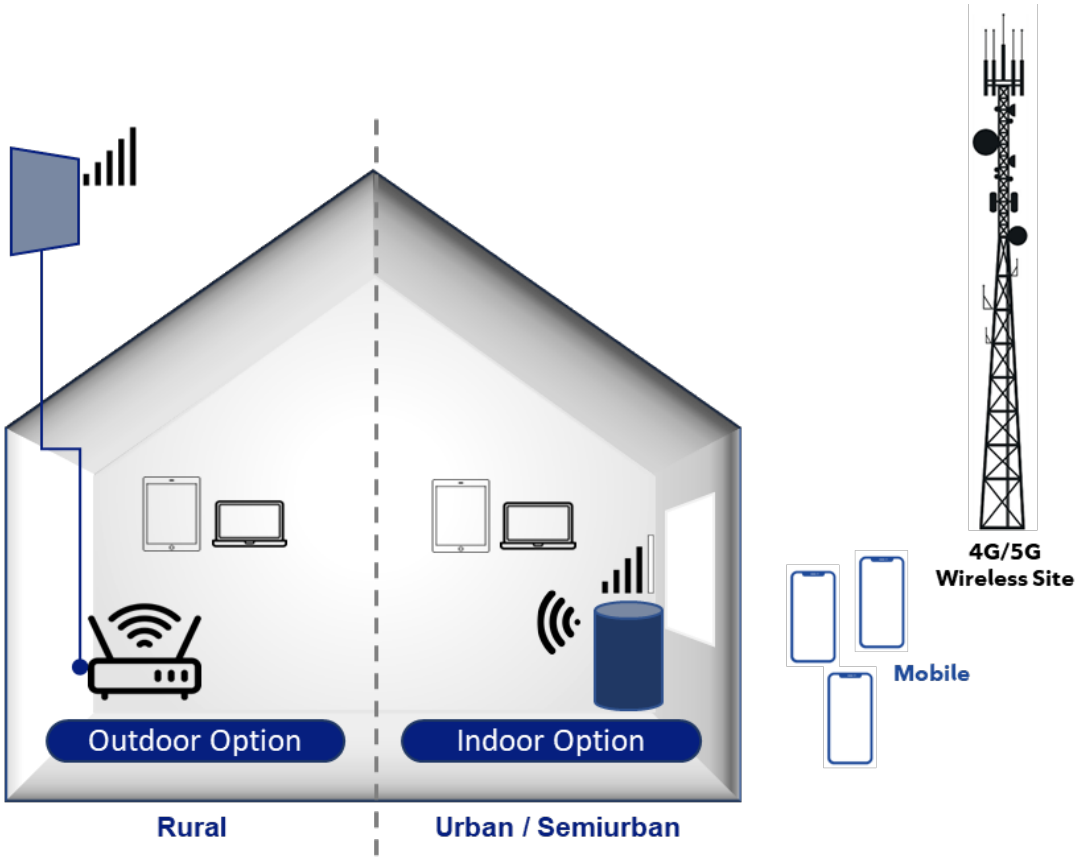


Figure 9 - Indoor & Outdoor CPE Setup

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The graph in Figure 10 shows LTE high band RSRP variations observed against distance of CPE from the base station. Outdoor CPE considers an external antenna with a gain of 16 dBi. A trend is extrapolated for indoor CPE on the same graph that assumes an additional loss of 18 dB (10 dB inbuilding penetration loss and 8 dB reduced antenna gain). Assuming a cut-off at -105 dBm for target service levels, it can be deduced that Indoor CPEs reduce coverage from 8 km to approximately 1 km

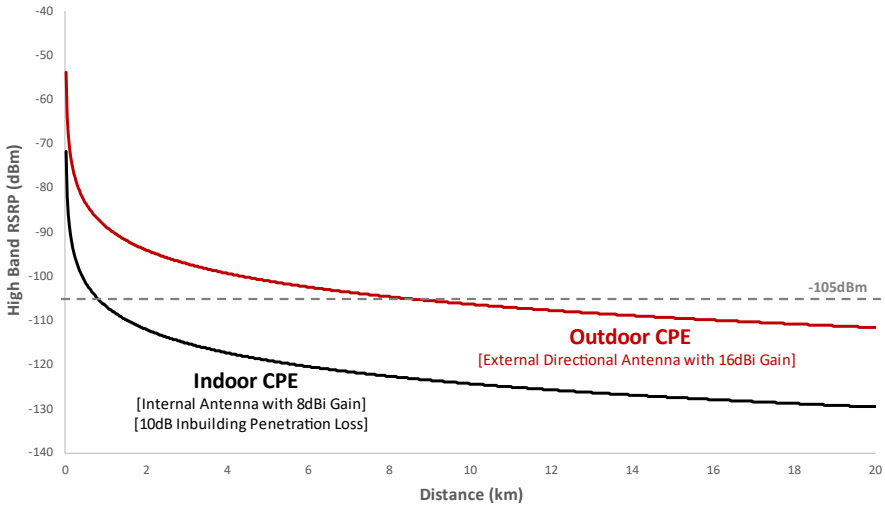


Figure 10 - Relation between High Band RSRP and distance from site

3.5. Fair Usage Policy

Across broadband networks, it has been observed that there are users with extremely high data consumption. With limited spectrum resources available on the wireless air interface, this would create capacity problems and degrade the experience for the rest of the users. Therefore, implementation of proper fair usage policies is recommended, which may be in the form of introducing:

- Data bucket plans where the users are only allowed a certain volume of data each month. Any overage after that may be charged with a higher rate
- Throttling speeds to a lower bandwidth once users surpass a certain usage each month

4. Network Management Features for FWA

As previously stated, a practical FWA solution is to deploy FWA service over 4G/5G networks. Utilizing an existing mobile network infrastructure, FWA deployment is relatively quick and cost-efficient. Despite the benefits, co-existence of FWA users and mobile users, each with their own distinct requirements brings up new challenges. It is vital to optimize the radio access network (RAN) features and parameters to ensure efficient sharing of resources between FWA and mobile users.

4.1. FWA Recognition in the RAN

As part of 4G/5G security architecture, the identity of users is not visible by design within the access network. Temporary identities are used instead with the mapping to the users' permanent identities known only by the core network. For the RAN to apply customized treatment, FWA users and FWA traffic must be identified. Within 3GPP specifications framework, there are several techniques that can be implemented in the core network to assist the RAN in identifying FWA traffic.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.1.1. UE-level Identification

To support FWA user management in the RAN, the core network can assign a planned UE-specific ID to FWA users and signal the value to the RAN every time the FWA user connects to the network.

4.1.1.1. Subscriber Profile ID (SPID)

The SPID, also known as RAT/frequency selection priority (RFSP) index, is typically assigned to specific subscriptions such as FWA. When received by the RAN, it is mapped by the radio base station to locally defined configurations designed to apply specific FWA strategies.

4.1.1.2. Network Slicing

One of the fundamental features of 5G is the concept of network slicing. A network slice is an end-to-end logical network consisting of a subset of network resources configured by the operator to serve a specific business requirement (e.g., MBB, FWA, WPN). The network slice is identified by the user, RAN and core through a single network slice selection assistance information (S-NSSAI) value. With the RAN, network slices are mapped to logically separated radio resources designed to ensure the requirements of each network slice.

4.1.2. Bearer/Flow-level Identification

In 4G/5G networks, the most fundamental unit for Quality of Service (QoS) is a single end-to-end data stream that is called “bearer” in 4G and “QoS flow” in 5G. For customized FWA QoS treatment, FWA bearers/flows should be identified and configured with distinct QoS parameters compared to mobile bearers/flows. Even within FWA bearers/flows, there might be a requirement to distinguish between them. For instance, each FWA CPE may have two different bearers/flows: one for user data traffic and the other for CPE remote management.

Among the most important QoS parameters linked to each bearer/flow are the QoS class identifier (QCI) for 4G, the 5G QoS identifier (5QI) for 5G and the allocation and retention priority (ARP) used for both 4G bearers and 5G QoS flows

4.1.2.1. QCI/5QI

The QCI/5QI is an identifier that combines multiple QoS characteristics such as packet delay budget, packet error loss rate, priority and whether the bit rate is guaranteed or not. In addition to standardized QCI/5QI values, the operator can configure operator defined QCI/5QI values to meet service requirements. Within the RAN, the configured QoS (based on the QCI/5QI value) is mapped to internal RAN features and parameters to satisfy the required QoS of each bearer/flow.

4.1.2.2. ARP

ARP parameters dictate the desired QoS level of a bearer/flow during admission and preemption decisions. ARP parameters include a priority level, preemption capability and preemption vulnerability. At the time of network congestion, ARP configuration determines if a new bearer/flow with a higher ARP priority can be admitted at the expense of preempting an existing bearer with lower ARP priority. To protect mobile subscribers during overload periods (e.g., special events), operators may assign FWA bearers/flows with ARP priority lower than that of the mobile bearers/flows

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2. Carrier Management

Cellular networks are typically deployed using overlaid multi-layer frequency bands. Depending on the operator's spectrum assets, the number of layers may vary. However, the layers are traditionally categorized into either capacity layers or coverage layers. A capacity layer is usually a higher frequency carrier with large bandwidth and advanced MIMO capabilities, but with limited wireless coverage. On the other hand, a coverage layer is a low frequency carrier (e.g., sub-1GHz) that can provide wide-area and deep indoor coverage; however, these layers typically have smaller bandwidths and basic MIMO capabilities and hence limited capacity.

To improve radio resource utilization and spectral efficiency, several carrier management techniques can be utilized within the RAN to ensure the subscribers are being served by the best carrier(s).

4.2.1. Limiting FWA Utilization of Coverage Layers

One of the key characteristics of FWA service is the stationary CPE. Before selling the service, the operator must perform a serviceability check and during the installation, especially for outdoor CPEs, a professional installer ensures the CPE is being served by the planned carrier. It is highly recommended to have FWA served primarily by the capacity layers as shown in Figure 11. The benefit is twofold: First, capacity layers are more suited to accommodate the high traffic demand of FWA. Second, the scarce resources on the coverage layer can be available to mobile users in weak coverage.

Utilizing SPID, the operator can configure customized FWA settings for both idle mode reselections and connected mode handovers that are different from default settings used by mobile users.

For idle mode control, the RAN sends dedicated camping priorities for the capacity layers (excluding the coverage layers) to the FWA user every time the connection is released. These dedicated priorities are used by FWA users instead of the broadcasted priorities used by all other users. If the dedicated priorities are valid, the FWA user (re)selects the best carrier among the capacity layers only.

In connected mode, the RAN is in control of steering the users between different carriers. In general, handovers between layers are triggered for various reasons e.g., RF based session continuity, load balancing, and priority based layer control. For FWA users, it is preferred to keep them on capacity layers. This can be achieved by restricting handover from capacity layers to coverage layers. However, even then if an FWA user lands on a coverage layer, either due to poor high band coverage or as the only available carrier during maintenance, RAN features can be configured to trigger the FWA user to handover back to capacity layers once they are available.

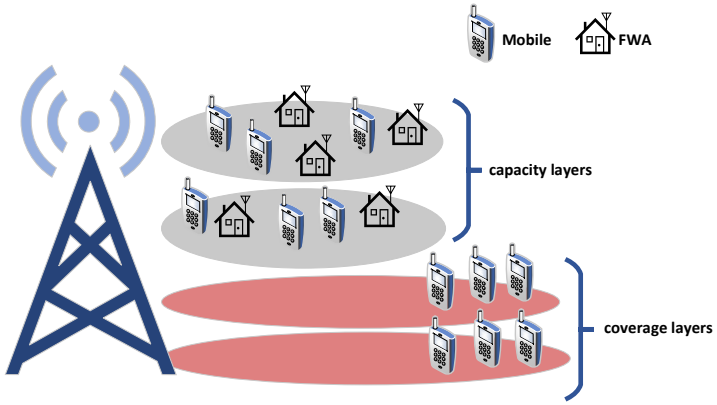


Figure 11 - FWA users sharing the capacity layers with mobile users

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2.2. Prioritizing FWA utilization of the spectrum licensed only for Fixed Wireless Services

When the spectrum is auctioned, the license for spectrum usage may come with some restrictions. For instance, the 3.5 GHz band in Canada has been traditionally licensed for the exclusive use of fixed wireless services. Due to the importance of this band for 5G and following global trends, the Canadian government re-issued new licenses for flexible usage that allows for both mobile and fixed services. However, incumbent licensees are protected from having to transition to the new spectrum allocations for up to a 3-year transitional period in some areas.

When the operator has regulatory and/or business requirements to restrict the access of specific carriers to FWA usage only, RAN should provide a solution to support such requirements. In 5G non-standalone (NSA) networks, the access control on 5G carrier usage is performed on the 4G anchor side. In the scenario where a specific 5G carrier can only by used by FWA, 4G RAN must be configured such that 5G-capable mobile users cannot be anchored towards the restricted carrier, but can be anchored to use other 5G carriers as shown in Figure 12. To maximize radio resource utilization, 4G RAN should prioritize the FWA usage of the 5G restricted carrier and hence freeing enough resources for mobile on 4G and other 5G carriers. To further support this strategy, aggressive traffic steering mechanisms with 4G anchor carriers might be needed.

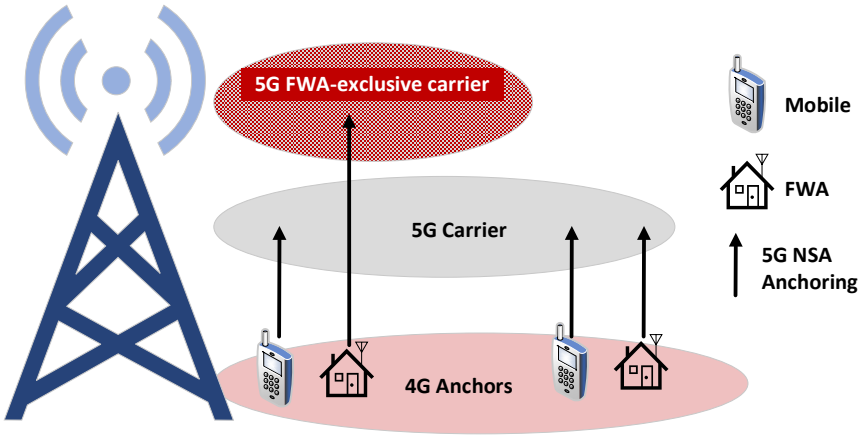


Figure 12 - FWA traffic prioitrized on FWA-exclusive 5G NSA carrier with no mobile access

4.3. Radio Resource Management

In wireless networks (point-to-multi-point communication), the radio resources are shared among all users. Radio resource management is a set of management techniques deployed in the RAN to control how the scarce radio resources are used to serve users as efficiently as possible. Radio resource management algorithms include admission control, scheduling, power control, interference management, modulation and coding schemes, MIMO, and beamforming.

4.3.1. FWA-Mobile Relative Priority Scheduling

In 4G/5G networks, the available radio resources are divided into small frequency-time units called resource blocks. The RAN node acts as a scheduler that assigns different chunks of resource blocks to

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

different users for a data transmission time interval (TTI) of less than or equal to 1 millisecond. The scheduler needs to take a resource allocation decision every TTI taking into consideration the traffic load, the required QoS of each bearer/flow and the available resources. The scheduler's objective is to optimize network capacity and user experience. Generally, there are different scheduling algorithms such as round robin, maximum carrier-to-interference, or proportional fairness. In practice, the proportional fair scheduling is the most common algorithm that attempts to prioritize users with good radio conditions without starving users with poor radio conditions.

The deployment of FWA on a 4G/5G network is typically motivated by the fact that the radio resources are under-utilized on average. If the resources are available to satisfy the demands of both mobile and FWA users, the scheduling decision is somehow straightforward. However, at the TTI level, it is common to have high resource contention especially in the presence of bursty applications. In these scenarios, it may be preferable for the operator to configure the scheduler to protect the experience of mobile users after FWA deployment.

In addition to assigning different QCI/5QI for mobile users and FWA users, it is recommended to configure relative weights to protect the throughput of mobile users when there is high demand on resources from both mobile and FWA. The ratio between the FWA and mobile relative weights will depend on the level of desired relative experience during high resource contention. During the periods with no contention, full resource grants will be given to FWA users.

Figure 13 and Figure 14 show the statistics on a typical FWA site. While FWA users represent less than 10% of the total users, they consume more than 85% of the data traffic. Thanks to the relative priority scheduling, the throughput of mobile users is not impacted by the introduction of FWA service.

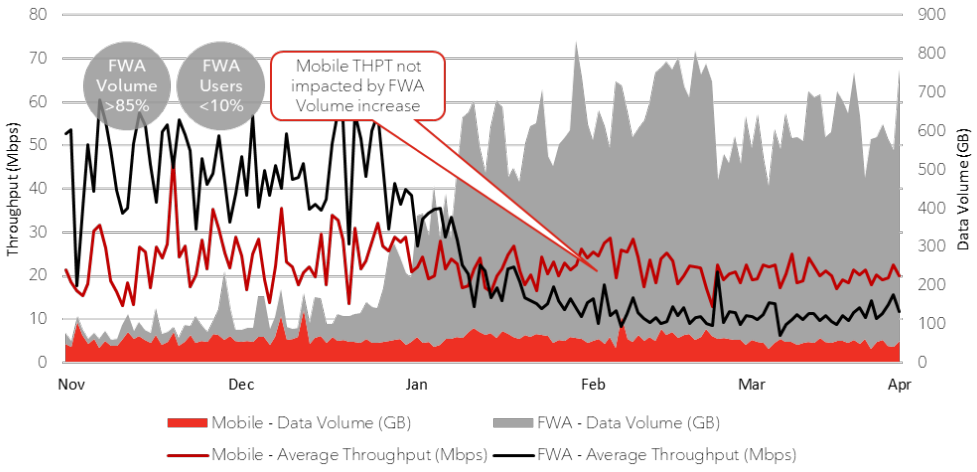


Figure 13 - FWA vs Mobile: Data Volume and Throughput

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

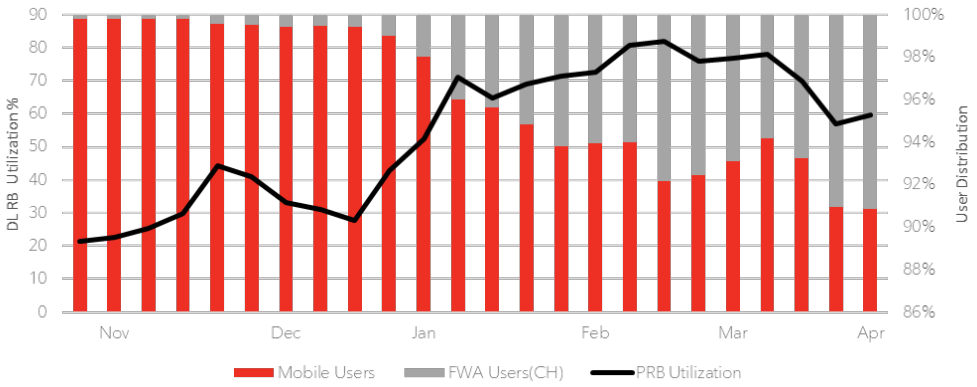


Figure 14 - FWA vs Mobile: Number of Users and Resource Block Utilization %

4.3.2. Differentiated Admission Control

Admission control is a key functionality that is used by the RAN node to decide whether there are enough resources to allow a new call or session to be adequately served without impacting the QoS of existing connections. When different services, such as mobile and FWA, are deployed on the same network, it is recommended to utilize differentiated admission control by assigning different ARP values. To protect mobile subscribers, FWA bearers/flows can be configured with lower ARP priority values.

Differentiated admission control could be achieved using differentiated blocking and/or preemption. During the periods of system resource shortage, the RAN prioritizes the admissions of mobile users with higher ARP priority over FWA users with lower ARP priority. Moreover, when preemption of existing sessions is necessary to free resources for incoming requests, the RAN prioritizes the preemption of FWA users with lower ARP priority over mobile users with higher ARP priority. The differentiated admission control preserves the accessibility and mobility performance of legacy mobile users after the introduction of FWA service in the network.

4.3.3. Differentiated Uplink Power Control

One of the mechanisms of radio resource management is the uplink power control. The RAN node controls how much power the device should use during uplink transmissions. This is typically dictated by specifying an uplink target power spectral density received by the base station known as P0. P0 is a parameter that is optimized to balance between the maximum uplink throughput and the overall network capacity. A higher P0 leads to higher signal to interference and noise ratio (SINR) resulting in higher uplink throughput, especially for users located close to the cell center. However, the trade-off of setting a high P0 value is the increase in the co-channel interference due to high power transmission by cell edge users.

If FWA devices are pre-planned to be in good/acceptable RF conditions, they might benefit from a P0 value, higher than that of the mobile users, to improve the reception of uplink transmissions from FWA, resulting in improvement in uplink throughput as well as downlink throughput due to more robust acknowledgement of downlink transmissions. Since the number of FWA users per cell are usually limited and their locations are typically near cell center, the increase in the co-channel interference has negligible impact on the SINR of mobile users on neighboring sites.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.4. Revisiting the Battery Power Saving Features

Battery life is one of the major challenges facing mobile devices. It is a common practice that the RAN in 4G/5G networks deploy features that aim to provide battery power savings even if it comes at the expense of minor degradation in performance. On the other hand, FWA devices are usually not power-constrained and hence the settings of the power saving features should be customized in an FWA-aware RAN to alleviate any unnecessary performance degradation to FWA devices.

4.4.1. Discontinuous Reception in Connected Mode (C-DRX)

C-DRX is a mechanism that allows the 4G/5G devices to stop monitoring the downlink channel during certain reoccurring time periods, similar to entering a temporary sleep mode. For battery-operated devices, C-DRX can extend the battery life as the device has the chance to turn off some RF circuits and reduce power consumption. The trade-off is an increase in the downlink latency and a slight degradation in the user throughput which is a cost that most mobile users are willing to pay to extend their battery life.

Since FWA devices are not power constrained, there is no benefit of utilizing C-DRX and hence it is recommended that operators configure the RAN to disable C-DRX feature for FWA devices.

4.4.2. User Inactivity Timer

When the RAN detects that a connected user does not transmit or receive any data for a pre-configured duration, the UE is released to idle mode. The optimization of the user inactivity timer for mobile users is to balance between the benefits of battery life extension when the user is in idle mode and the increase in signaling traffic for frequent changes between idle and connected modes. For FWA users, a customized user inactivity timer is recommended since there are no concerns regarding power consumption.

5. Case Studies

This section will cover two case studies evaluated on the Rogers network and how their results influenced decisions on optimizing the performance of the fixed wireless access product provided to customers.

5.1. Case Study 1 – Consequences of Low Band Utilization

Fixed Wireless Access is generally an alternative to cable or fiber connectivity to the home, hence the requirements of the service are different to what a mobile UE would need. Having the FWA user served by low bands that generally have lower capacity introduces worse performance to the end user for a few reasons:

- Low bands generally serve users in poor radio conditions, hence the users utilize more resources. Adding FWA users to the pool of users on low band could lead to congestion and a negative impact on all users served by the cell.
- The lower capacity on low bands generally leads to lower end user throughputs even in the absence of congestion.
- Low bands are sometimes limited to lower MIMO configuration (using 2x2 MIMO instead of 4x4 MIMO) this essentially reduces the peak throughput by half.

These limitations on low band were observed from network and device statistics as well as from customer complaints. Due to these drawbacks, settings were introduced as discussed in section 4, to make sure FWA UEs are kept on high bands longer which resulted in improved overall experience.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In Figure 15, data collected from the CPE show that when the FWA UE is served by 700 MHz (Band 12) as the primary cell, the LTE and NR aggregated spectrum bandwidth available to the user is more than 50 MHz for just 6% of the time from the collected samples. While when the UE is served by 2600 MHz (Band 7) as the primary cell, around 70% of the collected samples show an aggregated bandwidth larger than50 MHz.

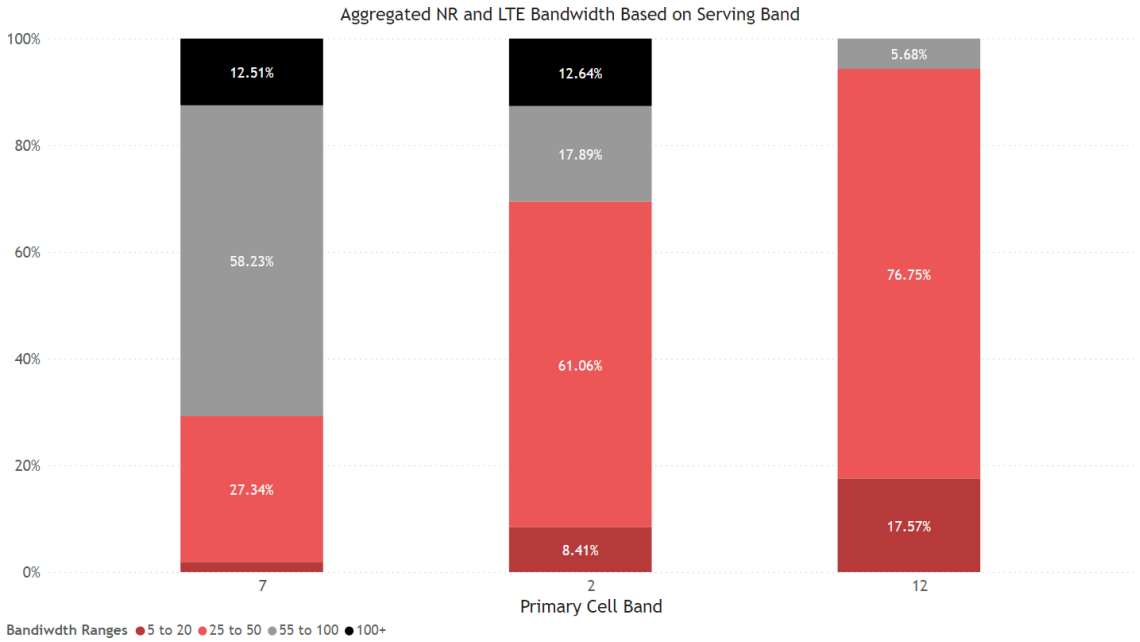


Figure 15 - Aggregated Bandwidth Based on Serving Band

From a network statistics point of view, Figure 16 shows the improvement in throughput attained on LTE and NR as a result of changes implemented to keep FWA users on high LTE bands for longer. The changes resulted in around 12% average throughput improvement for FWA users after reducing the percentage of traffic carried on low bands by 8%.

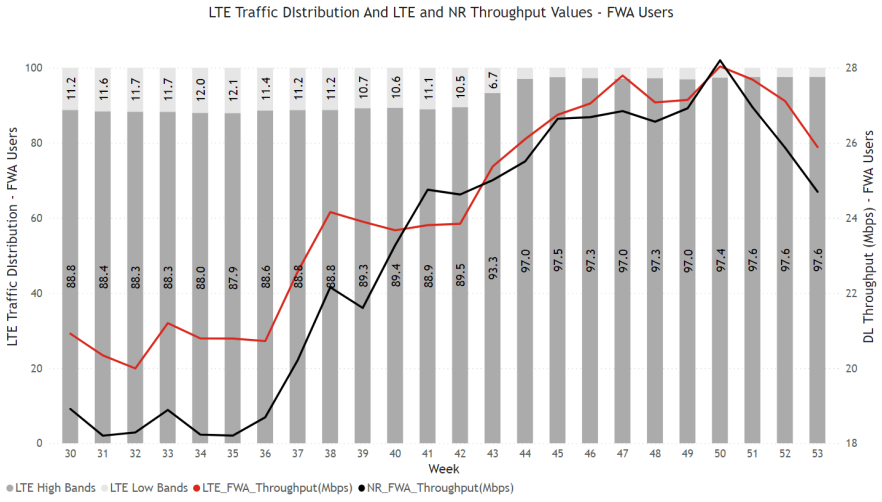


Figure 16 - FWA Network Level Statistics

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.2. Case Study 2 – Massive MIMO

Massive MIMO is a technique that utilizes an antenna array with large number of antenna elements and large number of transmitter and receiver chains [TRXs], typically significantly more than 8 [ref: Ericsson Massive MIMO Handbook]. Massive MIMO allows for improved air interface performance by enabling advanced beamforming, which can improve antenna directivity. In addition, massive MIMO facilitates the use of Multi-User MIMO (MU-MIMO) which can expand the capacity of the air interface by allowing the RAN node to schedule multiple users to utilize the same time and frequency resources, provided the users are sufficiently spatially separated. Massive MIMO is mainly used with an active antenna system, where the radio and the antenna are integrated into a single unit.

Massive MIMO active antenna systems are mainly used with NR TDD bands that are generally in frequency ranges greater than 2.5 GHz. As the frequency increases the optimal size of each antenna element in the array reduces allowing for the use of larger arrays while maintaining or even reducing the overall size of the antenna. In addition, for TDD, having a reciprocal channel allows for more accurate channel estimation and therefore, further improves gains from the use of massive MIMO antennas. Network statistics show that with massive MIMO antennas, a higher average spectral efficiency is achieved for FWA users. Figure 17 shows the median spectral efficiency for FWA users estimated from network statistics is 22% higher for Massive MIMO deployments compared to non-Massive MIMO deployments.

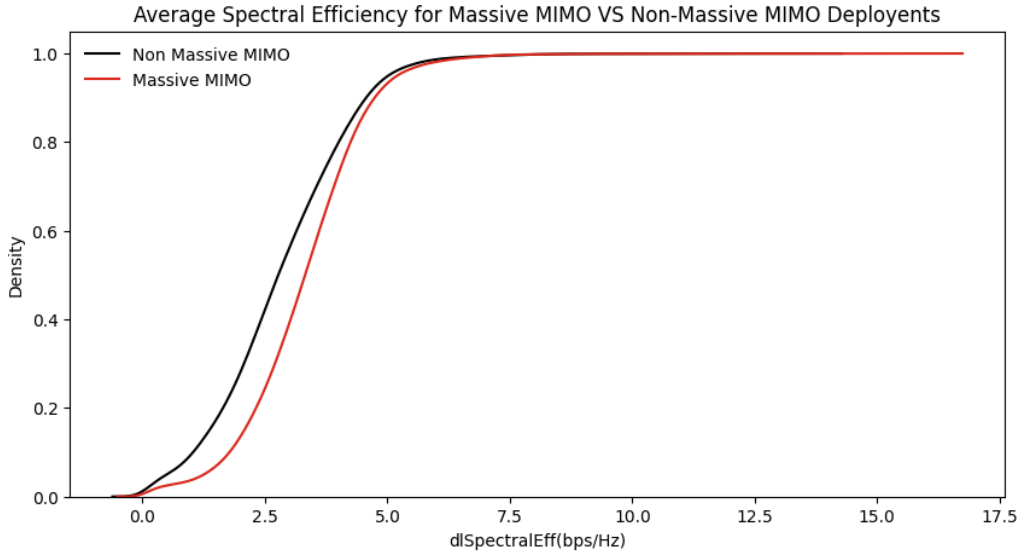


Figure 17 - Average Spectral Efficiency of Massive MIMO vs Non-Massive MIMO

MU-MIMO is expected to be more efficient for FWA users compared to mobile users since they have a higher data consumption rate and remain static. If there are more than two FWA users served by a particular cell and are sufficiently isolated to satisfy MU-MIMO conditions, those UEs will almost always satisfy the conditions. To observe this, network KPIs were analyzed to compare a set of nodes that predominantly serve FWA users and other nodes that serve mobile users and evaluate the difference in their cell capacity and MU-MIMO usage. Figure 18 shows sites predominantly serving FWA users utilize MU-MIMO 7% of the time while for sites with Mobile users this value is under 1%.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

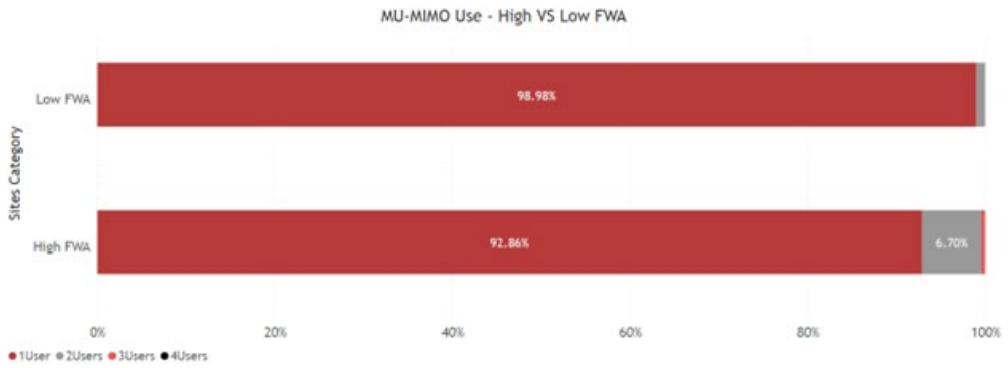


Figure 18 - MU-MIMO Usage: All FWA Nodes vs Mobile Nodes

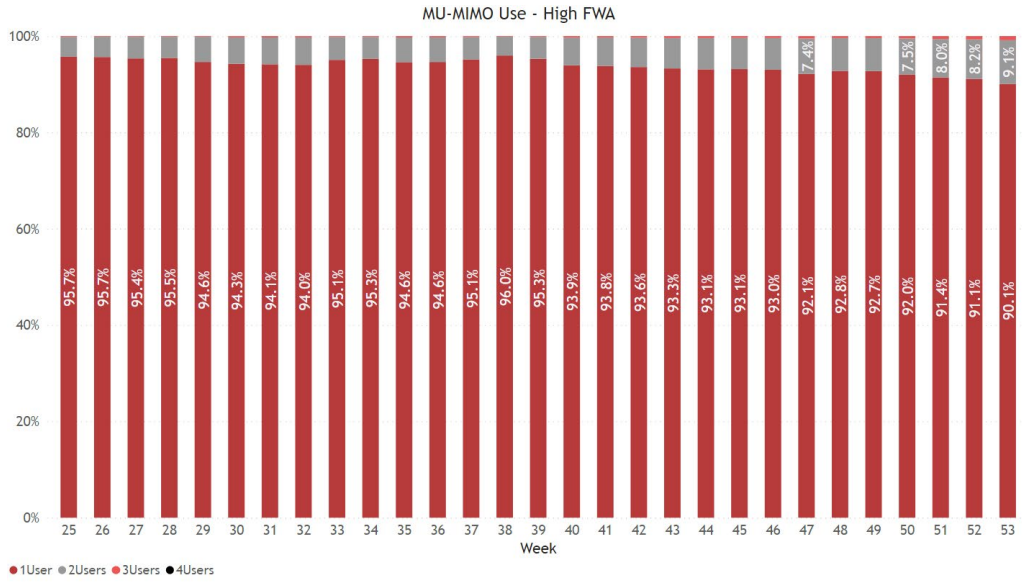


Figure 19 - MU-MIMO Usage Trend for FWA Serving Nodes

6. FWA: Lessons Learned, Prospects & Future

This section serves to summarize the best practices discussed in this paper and translate them into practical approaches and lessons learned. It then evaluates how FWA is also a valid use case for the urban environment and what is upcoming in context of the Canadian market.

6.1. FWA Recommendations

As operators continue to onboard more customers onto their 4G/5G networks and they tackle through issues, more clarity is achieved in terms of customer behavior and experience, traffic trends, installation issues, CPE issues, network management and impact of network features. Some of the key challenges broadly categorized with recommendations on how to address them are listed below.

- **Coverage Qualification:** To ensure customers are only onboarded where there is sufficient coverage available that can provide the desired service levels. This will avoid adding users on cell

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

edge and/or low band, that will have greater impact on network resources and provide poor customer experience. It will also prevent failed truck rolls or technician visits to remote locations where there is no coverage available. Coverage qualification also helps with determining the dominant best server at customer locations and can guide field technicians to position and orient the CPE for best results. The following steps are recommended:

- Conduct proper link budget analysis as per network configuration to determine required RSRP thresholds. This will define the minimum signal strength customers will need to have at their locations
- Use industry standard propagation modelling tools to determine serviceable areas and best server
- Continuously invest in updating & tuning propagation models and in latest geo datasets

- **Capacity Evaluation:** To ensure customers are only onboarded where there is sufficient capacity or headroom available on best server cells. Many operators choose to calculate an absolute number that can be supported on each of their sites based on the most common configuration found on their networks and use that as a reference. While this may be sufficient for high level planning, it is not the preferred method for actual customer onboarding especially when the network is being shared with other services such as mobile which may be more significant for the operator. Operators are recommended to be more surgical about it and define a framework around capacity evaluation at cell level using methodologies like the one discussed in this paper and incorporate it into their serviceability process. To accommodate that, some best practices include:
  - Continuously monitoring the network for congestion
  - Define appropriate congestion thresholds, could be based on PRB utilization or any other performance KPI. Different markets (urban vs rural) can be addressed differently
  - Prioritize network build or upgrades where congestion is observed or fast approaching
  - Analyze traffic trends for seasonal variations and year over year growth. It is recommended to account for those when calculating available headroom to avoid future congestion and customer churn

- **Maximizing Spectral Efficiency**: This is essential for efficient utilization of network resources. While coverage and capacity checks during serviceability evaluation also helps with improving spectral efficiency, some network strategies can help further improve this as seen with the case studies discussed in this paper.
  - Use of Massive MIMO radios capable of MU-MIMO and beamforming. It helps with reducing interference and re-using resources amongst users sufficiently isolated within the same cell coverage
  - Using outdoor CPEs with high gain directional antenna, especially for cell edge users. This helps with reducing interference and improves uplink performance, prevents excessive low band utilization, and gives better coverage reach and experience to cell edge users

- **Network Management Features**: Many operators are offering FWA services on infrastructure, spectrum and network assets that are shared with mobile users. Depending on the operator's strategy and primary business, traffic balancing and prioritization becomes critical. Operators may choose to implement network features that prioritizes Mobile traffic to ensure FWA usage does not adversely impact critical mobile services. Some of the practices that have proved valuable are:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

  - Steering FWA traffic to high band capacity layers, and away from low bands that are generally reserved for cell edge mobile users and are also limited to lower order MIMO configurations
  - Relative priority scheduling and differentiated admission control for mobile users over FWA users that protects the mobile experience as congestion on the cell increases due to FWA traffic
  - Implement fair usage policies such as data caps, throttling or limited bandwidth plans to protect network from heavy users

  It is important to note, not all features that work for Mobile use cases will work for FWA users as well. Operators need to work out customized settings for their FWA setup. E.g., FWA users can have higher differentiated uplink power control because they are less likely to cause adjacent channel interference or using battery saving features such as C-DRX and inactivity timer may not be useful for FWA users because these devices are not power constrained.

- **Device Selection & Installation**: Operators tend to select devices that are more acceptable by customers and have an attractive form factor. However, that often results in reduced capabilities of devices. Some important factors to consider related to FWA devices are:
  - Select CPEs capable of supporting latest 4G/5G features and enhancements
  - Deploy managed devices as it provides visibility to customer experience
  - Prefer outdoor CPEs over indoor for better spectral efficiency
  - Consider professional installs of outdoor CPEs that are securely mounted at sufficient heights to achieve line of sight and are oriented accurately towards the base station of the best server. This allows operators to control which part of the network (cells) traffic is added to. Improper or sub-standard installation may defeat the purpose of improving spectral efficiency

6.2. Urbanization of FWA

While the focus of developing FWA services has been on rural areas to serve unserved or underserved communities, 5G opens-up the potential for increased adoption of FWA in urban areas as well, where the demand for high-speed internet continues to grow.

High population density leads to a greater demand for broadband connectivity in a confined area, which can be efficiently addressed by FWA, especially in areas where wired infrastructure is either unavailable due to cost and implementation challenges or is outdated and difficult to maintain. Fewer physical components involved in FWA enable its rapid deployment, making it an attractive option for meeting immediate connectivity demands. Since last-mile connectivity to users is wireless, it causes minimal disruption to existing urban infrastructure.

Despite that, for FWA to be successful in urban markets and be competitive to the cable broadband offerings, its success largely depends on:

- Providing throughput speeds comparable to cable broadband
- Provide reliable indoor coverage
- Use convenient indoor devices as physically mounting outdoor CPEs may not be possible for multi dwelling units or urban environments
- Offer competitive pricing

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

As technology continues to evolve and 5G networks become more widespread, FWA is likely to play an increasingly important role in meeting the digital demands of our expanding urban centers.

6.3. Future Developments in FWA

The future of FWA appears to be promising, with several developments and prospects on the horizon. As 5G deployment gains momentum with operators in Canada, FWA stands to benefit from its expanded coverage and greater network efficiency, extending its reach to more underserved communities and communities previously only served with 4G.

A key feature introduced with 5G is network slicing that particularly benefits the FWA use case. It allows providers to partition their network into virtual end to end slices, each of which can be tailored to specific applications and user needs. This ensures optimized performance and quality of service for FWA and other services sharing network resources. Operators have been trialing this feature and are now expected to start implementing it.

The forthcoming 3800MHz auction in Canada also holds the promise of making additional capacity bands available, presenting a valuable opportunity to address FWA congestion concerns and empowering operators to improve the quality of services offered to FWA consumers.

Next on the horizon are prospects of using millimeter-wave (mm-Wave) technology for FWA. Its immense capacity, ultra-high speeds, and minimal latency, achieved by using large channel bandwidths, holds tremendous promise for FWA applications. However, the adoption of mm-Wave does come with its challenges, particularly in terms of coverage and propagation. In building penetration, foliage or propagation losses are extensive which makes it infeasible for FWA in non-line of sight conditions. The mm-Wave spectrum in Canada is yet to undergo auction, and while some operators are conducting trials, widespread implementation remains to be seen globally.

7. Conclusion

FWA isn't a threat to fiber or vice vera, rather the two complement each other. In context of bridging the digital divide and depending on an operators' broadband strategy, some key FWA deployment scenarios that emerge are:

- Quick turnaround to reach underserved communities via FWA, while fiber build follows. 5G infrastructure developed for FWA will gradually progress to cater mobile growth in the area once the fiber infrastructure is developed
- Customers in served communities that can compromise on performance and are on a budget. Key opportunity for operators is to utilize unused wireless capacity during non-busy hours
- Businesses that demand backup connections or an alternate to a wired connection
- Address residential customers in urban areas where wired connection is not available or outdated

With the ever-increasing demand of broadband connectivity and finite resources on the air interface that are usually shared with other services such as mobile, service providers continue to face challenges of providing fast, reliable, and consistent speeds to their customers along with sufficient capacity during busy hours. The paper has tried to identify best practices and recommendations that would help operators continue successfully on this journey and truly help bridge the digital divide in our communities.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

FWA	Fixed Wireless Access
MBB	Mobile Broadband
CRTC	Canadian Radio-television and Telecommunications Commission
LTE	Long Term Evolution
CA	Carrier Aggregation
CPE	Customer Premises Equipment
MIMO	Multiple Input Multiple Output
MU-MIMO	Multi-User Multiple Input Multiple Output
SCTE	Society of Cable Telecommunications Engineers
SPID	Subscriber profile ID
RAN	Radio access network
RFSP	RAT/frequency selection priority
RAT	Radio access technology
DL	Downlink
UL	Uplink
TDD	Time Division Duplexing
LNF	Log Normal Fade Margin
RMa	Rural Macro
S-NSSAI	Single network slice selection assistance information
mmWave	Millimeter-wave
QoS	Quality of service
PC-2	Power Class 2
PUSCH	Physical Uplink Shared Channel
SINR	Signal to Interference and Noise Ratio
CQI	Channel Quality Indicator
MCS	Modulation Coding Scheme
BLER	Block Error Rate
QCI	QoS class identifier
5QI	5G QoS identifier
ARP	Allocation and retention priority
NSA	Non-standalone
TTI	Transmission Time Interval
SINR	Signal to Interference and Noise Ratio
C-DRX	Discontinuous Reception in Connected Mode

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

Massive MIMO Handbook, 2nd Edition; Ericsson AB; 2023

Fixed Wireless Access Handbook, 2023 Ed.; Ericsson AB

Fixed Wireless Access With 5G Networks; A 5G Americas White Paper

Statistics Canada. Table 98-10-0001-01 Population and dwelling counts: Canada, provinces and territories

Canadian Radio-television and Telecommunications Commission
https://crtc.gc.ca/eng/internet/internet.htm

A. Lappalainen and C. Rosenberg, "Can 5G Fixed Broadband Bridge the Rural Digital Divide?," in IEEE Communications Standards Magazine, vol. 6, no. 2, pp. 79-84, June 2022, doi: 10.1109/MCOMSTD.0001.2100092.

Carto Vista https://crtc.gc.ca/cartovista/internetcanada-en/

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Graph Algorithms and Real-Time Telemetry for Intelligent Plant Operations

A Technical Paper prepared for SCTE by

Bob Lutz
Senior Engineer
Comcast
1800 Arch St, Philadelphia, PA
Bob_Lutz@comcast.com

Matthew Stehman
Principal Engineer
Comcast
1800 Arch St, Philadelphia, PA
Matthew_Stehman@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Problem Statement and Use Cases ... 3
   - 2.1. Legacy Plant Equipment Detection and Isolation ... 3
   - 2.2. Degraded Modulation Error Ratio ... 4
   - 2.3. Downstream Full Band Spectrum Clustering ... 5
3. Setup ... 5
   - 3.1. Network As a Graph ... 5
   - 3.2. Telemetry Summary ... 6
4. Algorithms ... 7
   - 4.1. Binary Attribute Clustering: Combinatorial Approach ... 7
   - 4.2. Continuous Attribute Clustering: Statistical Approach ... 9
   - 4.3. Cluster Analysis ... 11
     - 4.3.1. Root Cause Analysis ... 11
     - 4.3.2. Multi-Cluster Correlation ... 13
5. Algorithm Performance ... 14
   - 5.1. Graph Algorithms vs. Geospatial Clustering ... 14
6. Conclusions and Next Steps ... 16
7. Acknowledgments ... 16

Abbreviations ... 17

Bibliography & References ... 18

List of Figures

Title **Page Number**

Figure 1 – Downstream SC-QAM MER distributions. A.) Majority of devices with good MER B.) Majority of devices with low MER ... 4

Figure 2 – Common impairment patterns on full band capture of downstream receive power at the modem ... 5

Figure 3 – Graph database schema example for virtual CMTS ... 6

Figure 4 – Precision-based crawl in an access network graph with terminal and target highlighted ... 8

Figure 5 – Device clusters and corresponding root cause vertices ... 9

Figure 6 – Clustering impaired downstream SC-QAM MER with network topology and graph algorithms ... 11

Figure 7 – Cluster of device spectra with similar water wave patterns and corresponding root cause vertex ... 12

Figure 8 – Performance of baseline and geospatial clustering vs. graph-based clustering ... 15

List of Tables

Title **Page Number**

Table 1 – Average performance metrics and lift of baseline and geospatial clustering vs. graph-based clustering ... 16

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Real-time telemetry analysis has always been a core requirement for cable plant operators to detect customer-impacting events and dispatch the appropriate field teams. In the age of 10G networks, the cable plant's margin for error is slimmer than ever to offer state-of-the-art technology to potential subscribers. Minimizing the time field technicians need to troubleshoot issues will be key to maintaining a high cadence for 10G node deployments and conversions. To this end, telemetry-based alerts should not only identify problems in the plant but should recommend potential solutions as well.

We present an approach for combining node-level telemetry data and graph algorithms to help technicians resolve plant issues more efficiently and reduce mean time to repair. This approach has been successfully applied to use cases on the road to 10G, including distributed access architecture (DAA)-based 2G service deployments, streamlining demand maintenance with reduced truck rolls, and improving proactive maintenance by detecting likely network impairments. The algorithms involved have been field tested, incorporating technician feedback, and are now being integrated into production operations. This paper will highlight the importance and impact of combining network telemetry with plant topology to support the continued rollout of 10G.

2. Problem Statement and Use Cases

For multiple systems operators (MSOs), processes to efficiently identify and isolate network impairments are key to running a high-performance network and ensuring that customers receive the level of service they expect. Today's 10G networks can produce vast amounts of real-time telemetry from customer premise equipment (CPE). This paper will discuss approaches for combining CPE telemetry with a graph database of the hybrid fiber-coaxial (HFC) network to identify and resolve multi-home network impairments.

When multiple customer devices on a node show a degraded level of service, a decision needs to be made on what type of resource is needed to resolve the issue. Some scenarios require a network technician to address impairments in the HFC plant, while others require an in-home technician for issues inside the customer premise. From an operations standpoint, each ticket generated for field teams should ideally address as many customer issues as possible. This will reduce the overall number of jobs while maintaining or even increasing the number of customer issues resolved. Thus it is critical to be able to identify network impairments that are impacting multiple customers or homes at once. The following sections will discuss use cases for multi-home network impairment detection algorithms that are being used daily as the development of the 10G network continues.

2.1. Legacy Plant Equipment Detection and Isolation

The requirements of the network are changing as analog architectures are converted to digital on the road to 10G. As laid out in Harb et al. (2023) and Sundaresan (2022), the path to 10G requires potentially redesigning the spectrum and channel layouts to increase speeds. If legacy analog components in the HFC plant are not designed compatibly with the updated spectrum plans, they can cause issues with CPE bonding on these new channels.

Two areas of the spectrum that demand particular attention are:

1. 45–200 MHz with transitions from sub-split to mid/high-split designs and
2. > 750 MHz with the expansion of the existing downstream frequencies to include more channels.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In the first case, upstream frequencies are expanding into what were previously downstream frequencies. If certain components of the legacy network attenuate signals in this frequency range, customer devices will be unable to bond on these new channels when the digital network is turned up. In the second case, physical limitations of network equipment, such as amplifiers causing signal power roll-off, can prevent signals at high frequencies from reaching customers.

Typically, field teams perform a network sweep for incompatible legacy equipment when preparing for digital conversion. However, some network elements can be missed—especially if they are undocumented. If this happens, all customers downstream of incompatible network equipment will be unable to bond on any channels in the new frequency ranges. These bonding issues can occur immediately after the digital cutover or introduction of new channels.

In Section 4.1, we present an approach to identifying these issues and directing field teams to specific network elements of interest. This ability to isolate incompatible network devices enables efficient resolutions that minimize negative customer impact.

2.2. Degraded Modulation Error Ratio

Customers might continue to experience degraded signal quality, meaning packet loss and underperforming speeds, even after the HFC plant is upgraded to 10G-compatible equipment. To identify customers with degraded signal quality, we can look for CPE with low Modulation Error Ratio (MER) on certain channels. Figure 1 shows example device populations with high and low MERs for downstream single-carrier quadrature amplitude modulation (SC-QAM) channels.

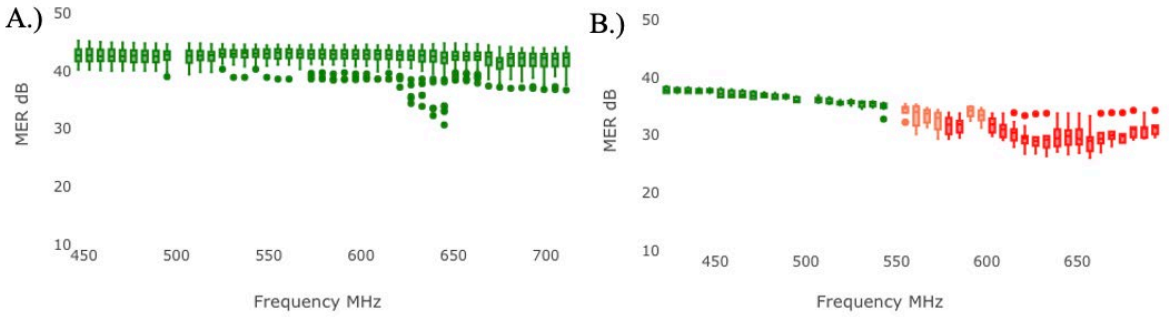


Figure 1 – Downstream SC-QAM MER distributions. A.) Majority of devices with good MER B.) Majority of devices with low MER

As in the previous use case, our goal is to identify impairments affecting multiple homes and direct field teams to root causes of the impairments. One challenge in this use case is that a radio frequency (RF) antenna in the vicinity of a node can cause CPE in separate legs of the plant topology to share similar MER characteristics. If this happens, a collection of in-home issues degrading MER, such as loose connections allowing ingress, can be mistaken for a shared multi-home issue when viewing telemetry alone. Another challenge is that devices only report MER values for channels on which they are bonded, but the CMTS can allow different CPE to bond on different channels. Thus, even if multiple devices are impacted by a single network impairment, their MER values might be incomparable if the devices are bonded on different channels.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In Section 4.2, we present a hybrid statistical and graph-based approach that overcomes these challenges by combining MER data with the plant topology. As in the previous use case, our algorithm identifies a particular network element as the likely root cause of each detected impairment.

2.3. Downstream Full Band Spectrum Clustering

Full band capture (FBC) data can reveal network impairments even when channel bonding and MER values are nominal. As discussed in Dugan et al. (2022), certain patterns in the FBC waveform indicate known network impairment types. Some of these patterns are illustrated in Figure 2 for FBC of downstream receive power at the modem.

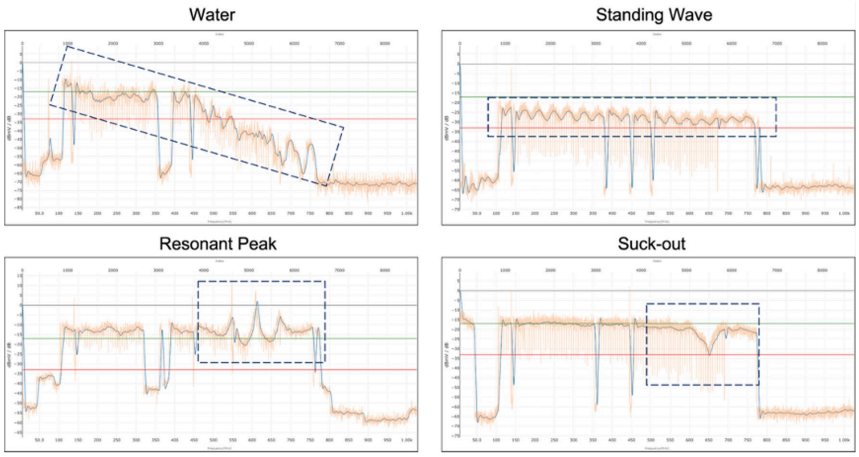


Figure 2 – Common impairment patterns on full band capture of downstream receive power at the modem

In this use case, output from pre-existing classification and clustering models processing FBC data is combined with a graph topology view of the network to identify network elements as root causes of multi-home impairments. These root cause elements are combined with results from the previous use case to find correlation between MER impairments and spectrum impairments, bundling demand maintenance (DM) events with proactive maintenance (PM) events. Our approach to this problem is presented in Section 4.3.

3. Setup

3.1. Network As a Graph

Traditional approaches to batch telemetry analysis are generally either implicit (e.g. inferring relationships between network elements based on latitude/longitude) or manual (e.g. performing root cause analysis by visual inspection of plant maps). Graph-structured data helps us outperform these baseline approaches by codifying explicit connections between network elements and allowing for full automation.

The algorithms in this paper operate on data from routing of cable infrastructure (ROCI), a graph database representing the access network. Vertices in the database represent the logical and physical entities that make up the network, from the cable modem termination system (CMTS) down to CPE. Edges represent either physical connections (e.g. by coaxial cable) or logical relationships (e.g. customer account to street address). Vertices and edges can have attributes representing properties of the given network element or

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

relationship, such as IP address, cable length, or latitude/longitude. To support real-time analysis, the data in ROCI is refreshed automatically as the topology of the access network evolves, by e.g. the addition of new customers, the conversion of analog nodes to digital, or the redrawing of plant maps. For more detail on ROCI, see Narayanaswamy et al. (2021).

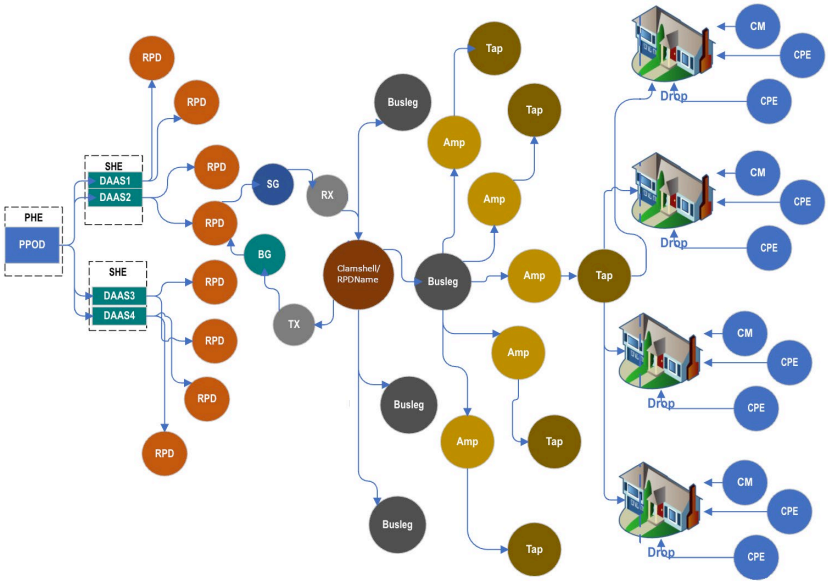


Figure 3 – Graph database schema example for virtual CMTS

3.2. Telemetry Summary

Addressing the use cases in Section 2 requires up-to-date telemetry from millions of customer devices. Fortunately, DOCSIS management information bases (MIBs) allow vast amounts of telemetry to be polled from CPE in near-real time. We briefly summarize the telemetry relevant to our use cases.

Most CPE have onboard spectrum analyzers and support the ability to report a FBC of downstream receive power across the entire frequency range. This view of receive power versus frequency gives a detailed view of the signals reaching the CPE and can be processed in a variety of ways based on project needs. Frequency-specific variations in the FBC indicate impairments related to amplification and attenuation of the signal power levels as the signals traverse the network. The FBC data is used heavily to detect HFC plant-related impairments as well as determining if certain frequencies can reach given CPE.

Even if the CPE receives signal power at a sufficiently high level, it may not be able to decode the DOCSIS code words if the signal quality is poor. In these cases, the CPE will drop packets, causing slower speeds and degraded customer experience. Channel-level MER can be used to identify these types of impairments, where the signal level is high, but noise corrupts the signal before it reaches the CPE. Processing MER data at the channel level is key, as it allows for clustering of devices based on frequency-specific patterns. This is not possible with aggregated MER values.

DOCSIS overcomes some amount of noise in a signal by including redundancy in its code words to help account for uncertainty in the packet. However, if the MER is too low and packets are lost, then the noise will impact the customer. Packet error rate (PER) can be used to measure customer experience in this situation. As will be seen in Section 5, MER and PER can be used together to efficiently identify customer-impacting, multi-home network impairments.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. Algorithms

We present a suite of algorithms that address the use cases from Section 2 by combining device telemetry with the network topology to identify multi-home network impairments. Broadly speaking, our goal is to group together CPE or homes whose degraded service is likely due to the same underlying issue and identify the network element that is the most likely root cause.

We first establish some terminology. A *path* in a graph is a sequence $(v_0, v_1, \dots, v_n)$ of vertices such that there is an edge between v_i and v_{i+1} for all i. A graph is a *tree* if, for each pair of vertices u and v, there is exactly one path from u to v. The *leaves* of a tree are the vertices incident to exactly one edge. When designing algorithms for the access network, the key data structure is a *rooted tree*, i.e. a tree in which a single vertex is designated the *root*. In this section, the root of our graphs is always the RF node, and the leaves are CPE. A *crawl* in a rooted tree is a path $(v_0, v_1, \dots, v_n)$ where v_0 is a leaf and v_n is a root.

Rooted trees enjoy a natural notion of hierarchy, where the root is considered the unique common ancestor of all other vertices. Algorithms on rooted trees can leverage this structure in contexts where vertices inherit behavior from their ancestors. For example, a misconfigured amplifier can pass a resonant peak to all downstream CPE. By identifying the relevant set of CPE and working backwards, the amplifier can be identified.

Our setup presents two main challenges. First, trees are the sparsest connected graphs, making them poor candidates for algorithms that rely on interconnectedness to simulate message passing or information spread. Many out-of-the-box solutions to problems like clustering and classification fall into this class of algorithms. Second, most approaches to analyzing graphs with vertex attributes assume that *all* vertices of the graph are attributed. But that is not the case here; while we have telemetry for the CPE, we have none for the internal vertices. In this way, access network graphs can be considered discrete "sensor networks," where the sensors are the CPE.

4.1. Binary Attribute Clustering: Combinatorial Approach

We start with the simplest possible case, when the device telemetry consists of a single binary variable, and apply our technique to the use case from Section 2.1. In practice, this variable could be truly binary (e.g. online vs. offline) or could express whether a continuous variable meets a certain threshold (e.g. whether signal-to-noise ratio is above or below 35 dB). For uniformity, we will call the values of this binary variable *impaired* and *unimpaired*. We will use the terms "CPE" and "device" interchangeably, so "device" necessarily means a customer device.

Let $(v_0, v_1, \dots, v_n)$ be a crawl in an access network graph, so that v_0 is a device, i.e. a leaf, and v_n is the RF node, i.e. the root. At each step of the crawl, we take a quantitative measurement $m(i)$ of the devices downstream of v_i. For example, let D_i denote the set of all devices downstream of v_i, and let I_i denote the set of devices in D_i that are impaired. We could take $m(i)$ to be the *precision*, i.e. the fraction of devices downstream of v_i that are impaired:

$$p(i) = \frac{|I_i|}{|D_i|}.$$

Alternatively, we could take $m(i)$ to be the *recall*, i.e. the fraction of all impaired devices that are downstream of v_i:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

$$r(i) = \frac{|I_i|}{|I_n|}.$$

For further discussion on these measurements, see Section 6.1 of Dugan et al. (2022).

We fix a value M to act as an inclusive lower threshold for the measurement $m(i)$. Simple modifications can be made for upper or exclusive thresholds. Typically, we will assume that the threshold is satisfied at the starting device, i.e. $m(0) \geq M$. The *terminal* is the last vertex in the crawl whose measurement satisfies the threshold. In other words, the terminal is v_{s-1}, where s is the smallest value of i for which $m(i) < M$, or $s = n + 1$ if no such value exists. The *target* is the first vertex in the crawl whose set of downstream impaired devices is the same as the terminal's. In other words, the target is v_t, where t is the smallest index such that $|I_t| = |I_{s-1}|$.

An example crawl is illustrated in Figure 4. Each vertex v_i of the crawl is labeled by its index i. There are 11 impaired devices, colored blue, and 4 unimpaired devices, colored gray. We take $m(i) = p(i)$ to be the precision and $M = 1$. The terminal, colored red, is the last vertex in the crawl for which all downstream devices are impaired. The target is colored green. The remaining vertices of the crawl are colored orange.

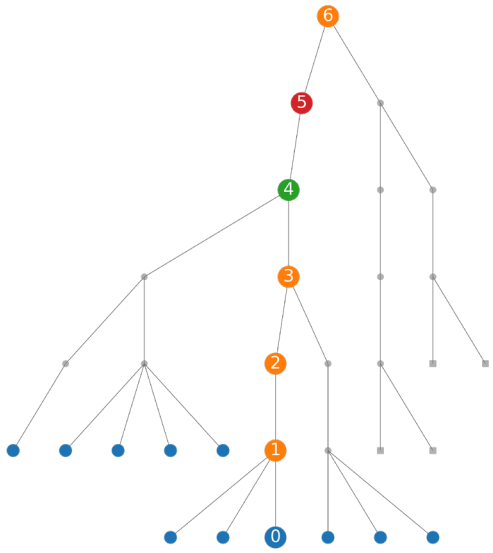


Figure 4 – Precision-based crawl in an access network graph with terminal and target highlighted

To cluster the impaired devices on the node, we perform a crawl starting at each impaired device. In this way, each impaired device is associated with a target. Sometimes the set of all targets is an *antichain*, in the sense that no target is an ancestor of any other target. In this case, we can partition the impaired devices into disjoint clusters by grouping together all devices with the same target. The clusters can be interpreted as separate underlying network impairments, each equipped with a corresponding *root cause vertex*, i.e. the target itself, that is interpreted as the source of the cluster's impairment. The validity of these interpretations depends on data quality and the choices of measurement and threshold.

The targets do not always form an antichain, however. If some targets are ancestors of other targets, measures must be taken to prevent the clusters from overlapping. One such measure is to discard any

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

targets that are descendants of another target in the set. The resulting "minimal" set of targets will guarantee a partition into disjoint clusters. Another remedy is to order the targets (or, equivalently, the impaired devices), and to define each cluster as the set of impaired devices downstream of the target that do not belong to any previous cluster. These "exclusionary" clusters will depend on the ordering; choosing an appropriate ordering is a potentially subtle problem.

This procedure can be used to detect incompatible legacy equipment as described in Section 2.1. Here a customer device is considered unimpaired if it is bonded on any new channels and impaired otherwise. Since all devices downstream of incompatible legacy network elements are guaranteed to be impaired, we can take $m(i) = p(i)$ to be the precision and $M = 1$. Each crawl will then terminate at the last step for which every downstream device is impaired.

The terminals always form an antichain in the case $M = 1$. They do not necessarily form an antichain, however, if $M < 1$. This threshold could be a more appropriate choice if the impairment in question were not guaranteed to impact the telemetry of every single downstream device. An example result for the case $M = 1$ is shown in Figure 5, where the impaired devices and root cause vertices are colored according to their corresponding cluster (red, green, orange or blue).

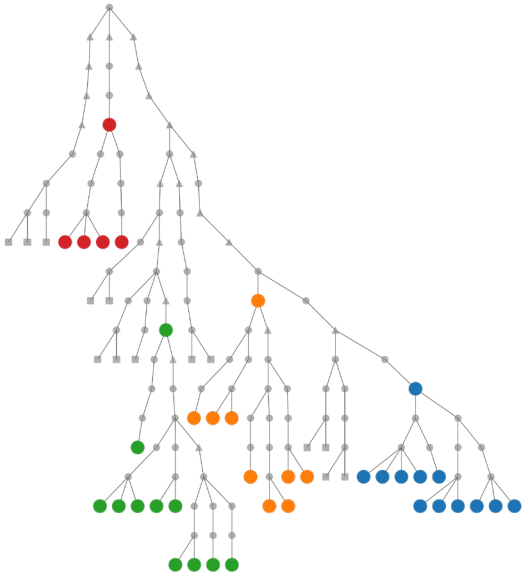


Figure 5 – Device clusters and corresponding root cause vertices

In practice, additional context is often needed to generate meaningful tickets for field teams. For example, how many homes are impacted by a given cluster? Is the root cause a cable drop in a multi-dwelling unit (MDU)? Is it an "end-of-line" element? Answers to these questions can help decide how to prioritize different network events and which personnel to dispatch. The algorithm's ability to provide these answers depends on what data is available in the graph database.

4.2. Continuous Attribute Clustering: Statistical Approach

We now consider the case where the device telemetry takes continuous values in order to address the use case from Section 2.2. We will continue to use the notation from Section 4.1, where $(v_0, v_1, \dots, v_n)$ is a crawl in an access network graph and $m(i)$ is a measurement at each step i of the crawl. It is convenient to think of this process from a statistical point of view, where at each vertex v_i the attribute values of the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

downstream devices constitute a sample from a probability distribution, and the measurement $m(i)$ is a descriptive statistic of this sample. With binary or categorical attributes, the distribution is discrete, and the statistic is enumerative. With continuous attributes, the distribution is continuous, and the statistic can take more familiar forms like the mean, standard deviation, etc.

In addition to allowing continuous attributes, we will generalize the previous setup in several ways. First, instead of a single measurement $m(i)$, we consider a family of measurements $m_k(i)$. For example, $m_1(i)$ could be the maximum attribute value across all devices downstream of v_i, and $m_2(i)$ could be the median. Second, instead of a fixed threshold M_k for each measurement, we define dynamic thresholds $M_k(i)$ that can depend on the current step i or on previous steps. For example, $M_1(i)$ could be the 50th percentile of the attribute values of all devices downstream of v_i.

Depending on the problem, we will have different criteria for where crawls begin and end. For where to begin the crawls, there is no longer a built-in notion of "impaired" devices as in the binary case, so we must define a set of impaired devices based on the problem statement and telemetry. As for terminals, there are many ways to generalize the termination criterion $m(i) < M$ to the multi-measurement case. Typically, a crawl will terminate when $m_k(i) < M_k(i)$ for *any* k. As before, simple modifications can be made for upper or exclusive thresholds $M_k(i)$. Following the same clustering process as in Section 4.1, we obtain disjoint clusters of the impaired devices with a root cause vertex associated to each cluster.

We can use this setup to detect pockets of customer devices with degraded MER as discussed in Section 2.2. Because MER is measured on multiple channels, the telemetry in this example is vector-valued, and the relevant probability distributions are multivariate. For simplicity, we will assume that all devices are bonded on the same channels. In reality, however, devices are commonly bonded on different channels, so the telemetry vectors might contain null values. The null values can be either ignored or imputed.

In this example, a device is impaired if it is bonded on any sufficiently "degraded" channel, where the definition of "degraded" depends on context. We set the following measurements and thresholds:

- $m_1(i)$ the maximum value of $f(d)$ as d ranges over all devices downstream of v_i, where $f(d)$ is the minimum MER value for device d across all degraded channels
- $m_2(i)$ the median of $g(c)$ as c ranges over all degraded channels, where $g(c)$ is the standard deviation of MER values on channel c across all devices downstream of v_i
- $m_3(i)$ the 80th percentile of $h(d)$ as d ranges over all devices downstream of v_i, where $h(d)$ is the minimum MER value of device d across all degraded channels
- $M_1(i)$ a constant value; MER above this value is considered "very good"
- $M_2(i)$ a constant value; MERs differing by more than this value are considered "dissimilar"
- $M_3(i)$ a certain percentile of MER values across all devices downstream of v_{i-1} (the *previous* step in the crawl) and all degraded channels

The crawls terminate when $m_k(i) > M_k(i)$ for any i. For $k = 1$, the termination condition checks whether the MER of downstream devices is too high for the sample to be considered impaired; for $k = 2$ and $k = 3$, the conditions detect whether the current step has added outliers to the previous MER sample. In this example, we take the "minimal" approach to clustering described in Section 4.1. Devices with healthy MER will be clustered by themselves, due to the crawl termination conditions; all singleton clusters and unimpaired devices can be merged to form a cluster of "normal" devices. This is depicted in Figure 6, where the lone impaired cluster is orange, and the merged "normal" cluster is blue.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

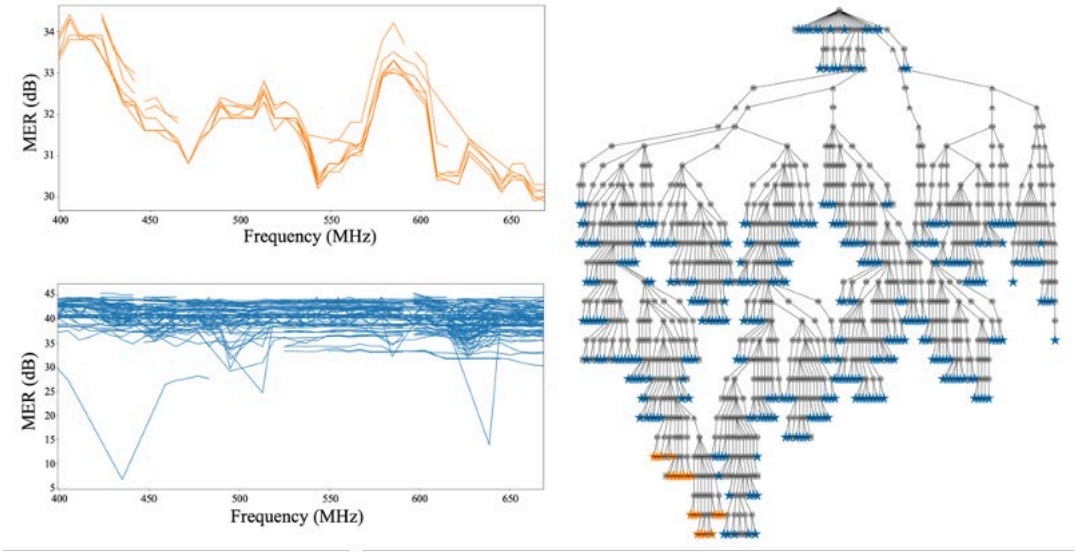


Figure 6 – Clustering impaired downstream SC-QAM MER with network topology and graph algorithms

4.3. Cluster Analysis

So far, we have described graph algorithms that use telemetry and graph structure *simultaneously* to perform clustering and root cause analysis. But this is not always the preferred approach. For some problems, it makes sense to operate on telemetry alone, or on the graph structure alone, and to combine the results post hoc. Alternatively, we might wish to leverage existing solutions, such as legacy clustering or classification algorithms that do not make use of topology, and enrich them with graph data. In this section, we apply these approaches to the use cases of Sections 2.2 and 2.3.

4.3.1. Root Cause Analysis

In some cases, we are given existing clusters of impaired devices, and our goal is to identify the network element that most likely caused each impairment cluster. This is essentially the reverse of the problems in Sections 4.1 and 4.2, where we identified a root cause vertex *first*, and the clusters are obtained as a corollary. Hence a different approach is needed.

For example, suppose that we want to perform root cause analysis for impairment patterns in FBC data, as described in Section 2.3. In this scenario, devices are classified by impairment type (e.g. resonant peak), and devices with the same impairment type(s) are clustered together if their impairment patterns exhibit similar characteristics (e.g. resonant peaks with the same resonant frequency). This clustering step is needed to distinguish separate impairments of the same type occurring simultaneously. We can perform root cause analysis on these clusters by viewing them in the appropriate access network graph. In addition to providing key diagnostic context, this process also allows us to filter out clusters for which the clustering algorithm has underperformed, or the underlying data quality is poor. In this way, the telemetry and topology can act as checks and balances for one another, instead of as potentially noisy simultaneous input.

Returning to the general setting, let $C_1, C_2, \ldots, C_n$ denote device clusters. These clusters need not be disjoint. In the case of spectrum impairments, for example, the clusters might overlap if some devices are

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

subject to multiple underlying network impairments with distinct patterns; ideally, we would see one cluster for each impairment pattern and another cluster comprising all devices with "normal" spectra.

Naively, we could take the root cause vertex of the cluster C_k to be the *lowest common ancestor* of all devices in C_k. However, this approach gives disproportionate influence to "topological outliers," i.e. devices in C_k that are topologically distant from the other devices in the cluster. We need an approach that can discriminate some topological outliers, since they are common in practice.

We will focus on a particular cluster C_k and consider a device "impaired" if it belongs to C_k or "unimpaired" otherwise. Recall the definitions of *precision* $p(i)$ and *recall* $r(i)$ from Section 4.1. Instead of taking these measurements at a step i of a crawl, we now take them at any vertex v in the graph. Thus, for example, $p(v)$ is the fraction of devices downstream of v that are impaired. We can define a corresponding *F-score* at each vertex as follows:

$$F_\beta(v) = \frac{(1+\beta^2)\cdot p(v)\cdot r(v)}{\beta^2\cdot p(v) + r(v)},$$

where β is a positive parameter. This score can be interpreted as a weighted harmonic mean of precision and recall, where recall is considered roughly β times as important as precision. It is generally high if both precision and recall are high, and low otherwise. The usual F_1 score is the special case $\beta = 1$.

The root cause vertex of cluster C_k is taken to be the vertex v that maximizes $F_\beta(v)$. Higher values of β will give more importance to topological outliers, resulting in root cause vertices closer to the RF node. Lower values of β will give less importance to topological outliers, resulting in root cause vertices closer to impaired devices. Choosing an appropriate value of β depends on the problem statement, the quality of the clusters, and the nature of the impairments in question. Figure 7 shows a cluster of devices exhibiting similar water wave patterns with a single topological outlier and the associated root cause vertex.

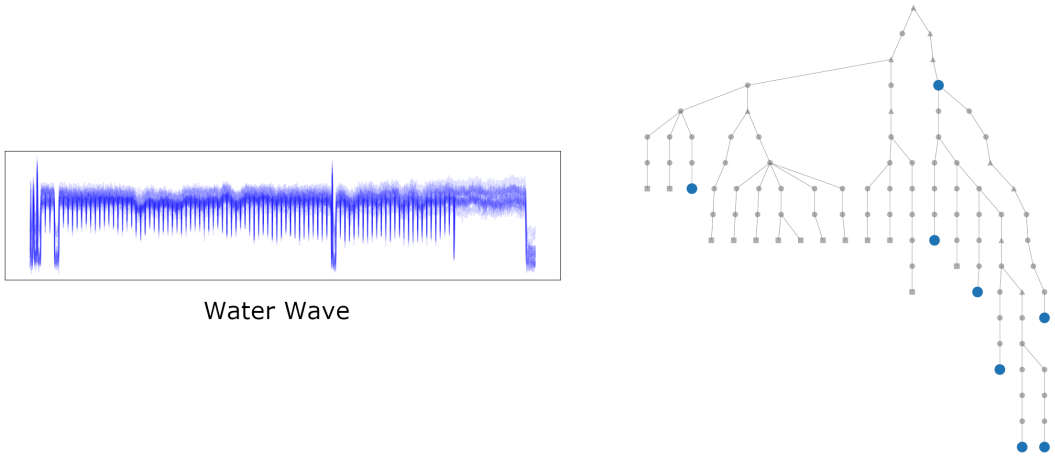


Figure 7 – Cluster of device spectra with similar water wave patterns and corresponding root cause vertex

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

It is not always possible to adequately characterize a device cluster with a root cause vertex. Clusters can exhibit sparseness, bifurcation, or otherwise low topological correlation when viewed within the network topology. Typically, these phenomena point to one or more of the following issues:

1. Poor clustering performance
2. Inaccurate data, either in the telemetry or in the graph database
3. Indistinguishable telemetry characteristics across distinct impairments.

An example of Item 3 in the case of FBC could be two impairment patterns occurring on the same frequencies, such as a water wave superimposed with a standing wave, or two water waves overlapping. This poses a separate issue from poor clustering performance, since even an ideal clustering algorithm could not necessarily distinguish devices subject to the different underlying impairments without additional data.

To detect clusters with low topological correlation, we enforce lower thresholds for precision and recall. Specifically, we set numbers P and R and check whether $p(v) \geq P$ and $r(v) \geq R$, where v is the candidate root cause vertex described above. If either condition fails, the cluster is rejected as having low topological correlation and receives no root cause. The cluster can then be reported and evaluated to determine the influence of Items 1–3 above.

4.3.2. Multi-Cluster Correlation

When clustering impaired devices based on multiple categories of telemetry, different clusters can point to related network impairments, or even the same underlying impairment. For example, a cluster of devices with degraded MER and a cluster of devices exhibiting water wave could point to the same issue if the water wave is *causing* the low MER. Thus we need a way to determine if clusters and root cause vertices from different sources are correlated, independent of the individual algorithms.

Let K and L denote two clusters in the same access network graph, potentially from different algorithms, with associated root cause vectors u and v, respectively. A classical measure of correlation between clusters K and L is the *Jaccard index*:

$$\frac{|K \cap L|}{|K \cup L|},$$

Where ∩ denotes intersection and ∪ denotes union. This measure is appropriate when the clusters are generally "comprehensive," in the sense that they contain most of the devices they should, and few they should not. In practice, however, clustering results are often sensitive, noisy and non-deterministic. This can be due to inconsitent data quality, the behavior of the algorithm, or both.

To address these limitations, we focus on root cause vertices instead of the clusters themselves. Ultimately, we want to know if a technician can solve multiple issues by visiting a single network element. The particular devices implicated in a cluster often have no bearing on this result, especially taking into account the performance of the clustering algorithm. For each vertex w, let $S(w)$ denote the set of all CPE downstream of w. We define the *Jaccard index* of root cause vertices u and v as

$$J(u, v) = \frac{|S(u) \cap S(v)|}{|S(u) \cup S(v)|},$$

In other words, $J(u, v)$ is the number of common downstream devices between u and v divided by the total number of downstream devices. Higher values indicate higher correlation between the root cause

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

vertices. Identical root cause vertices will have Jaccard index 1, while root cause vertices from distinct branches of the graph will have Jaccard index 0. A Jaccard index strictly between 0 and 1 occurs when one root cause vertex is an ancestor of the other. This definition can be adapted to arbitrarily many sets of clusters.

The Jaccard index of root cause vertices can be used to measure correlation between degraded MER (Section 2.2) and impairment patterns seen on FBC (Section 2.3). Here the root cause vertices come from the algorithms in Sections 4.2 and 4.3.1, respectively. Root cause vertex pairs with sufficiently high Jaccard index are interpreted as expressions of related or identical underlying network impairments. This effectively correlates demand maintenance (DM) events with proactive maintenance (PM) events.

5. Algorithm Performance

The algorithms in Section 4 perform well when applied to the use cases in Section 2. The precision-based algorithm in Section 4.1 has been used to identify legacy equipment preventing mid-split enablement on the path to 10G (Section 2.1). In this application, the root cause vertex identified by the algorithm is within 300 feet of the actual legacy equipment in 95% of cases. A similar approach has also been used to identify root causes of severe roll-off in a region of spectrum intended for orthogonal frequency-division multiplexing (OFDM) expansion; see Harb et al. for further discussion (2023).

The statistical algorithm in Section 4.2 has been used to identify multi-home MER impairments and their root causes (Section 2.2). The results of the algorithm are overlayed with PER telemetry to determine whether the degraded MER is impacting customers. A relevant network impairment is found at the predicted root cause vertex in 85% of customer-impacting cases. To enhance these results, the root cause analysis and cluster correlation algorithms in Section 4.3 are used to identify spectrum impairments (Section 2.3) that are likely related to or even causing multi-home MER issues.

5.1. Graph Algorithms vs. Geospatial Clustering

Network topology data is sometimes unavailable or of insufficient quality for meaningful analysis. When this happens, a non-graph-based fallback approach is necessary. There is a question of how such an approach will perform when compared to the graph algorithms in Section 4. To answer this question, we perform clustering on CPE downstream of 2,071 physical layer (PHY) devices (remote PHY devices (RPDs)) with high-quality network topology data and a single binary attribute (“impaired” vs. “unimpaired”). We cluster the impaired customer devices on each RPD using three approaches:

1. (Baseline) All impaired devices assigned to the same cluster
2. (Geospatial) Density-based spatial clustering of applications with noise (DBSCAN) on latitude/longitude alone
3. (“Ground truth”) Clustering with graph data, precision-based approach described in Section 4.1.

We assess the performance of the baseline and geospatial approaches, treating the results of the graph-based clustering as the ground truth. Let n be one of the RPDs analyzed. Define S to be the set of all pairs of impaired devices on RPD n such that both devices belong to the same cluster, according to the graph-based approach. Let T be defined similarly but with one of the non-graph-based approaches instead. The *precision* of the non-graph-based approach for node n can be defined as

$$p = \frac{|S \cap T|}{|T|}.$$

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In words, p is the fraction of device pairs sharing the same non-graph-based cluster that also share the same graph-based cluster. The *recall* is then

$$r = \frac{|S \cap T|}{|S|}.$$

In words, r is the fraction of device pairs sharing the same graph-based cluster that also share the same non-graph-based cluster. The results are illustrated in Figure 8, where we plot histograms and cumulative distribution functions (CDFs) of precision, recall and F_1 score across all 2,112 nodes. Here the F_1 score is the usual harmonic mean of p and r.

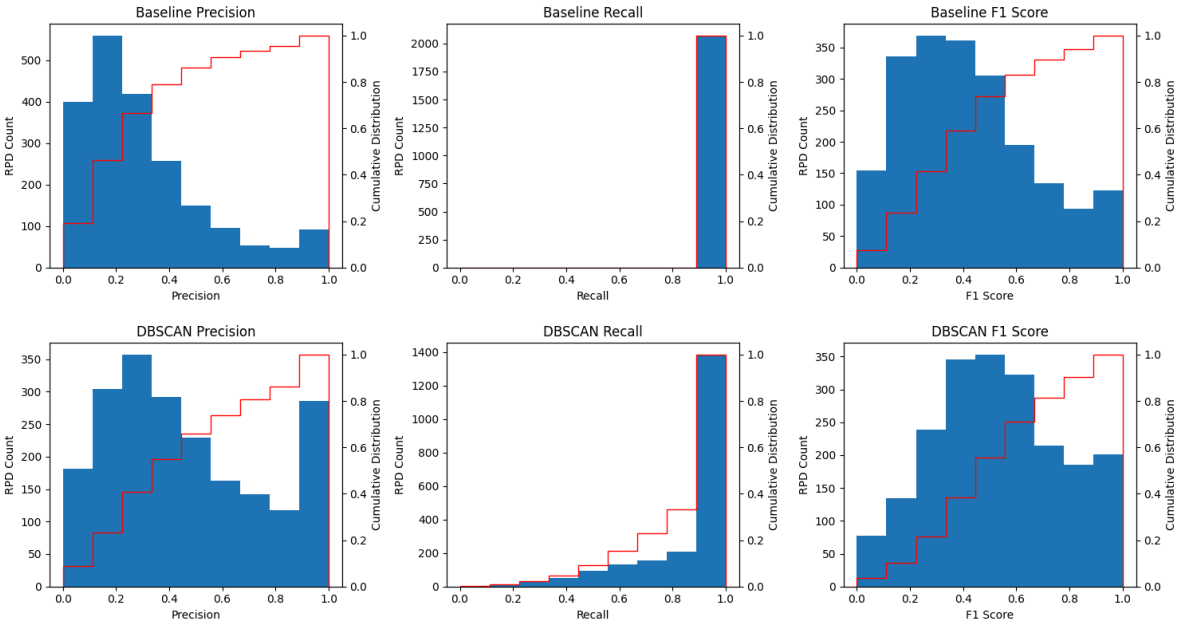


Figure 8 – Performance of baseline and geospatial clustering vs. graph-based clustering

A numerical summary is given in Table 1. We compute the mean and median of precision, recall and F_1 score for the baseline and geospatial approaches across all RPDs, and then for each metric compute the mean and median *difference* between the two approaches across all RPDs. While the baseline approach has perfect recall by design, the geospatial approach outperforms the baseline in both precision and F_1 score, suggesting that it improves on the baseline overall. However, the average F_1 score for the geospatial approach is 0.54, which is still low. This suggests that the geospatial approach does not perform similarly overall to the graph-based clustering.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 – Average performance metrics and lift of baseline and geospatial clustering vs. graph-based clustering

	Baseline	Geospatial	Geo minus baseline
Precision (mean)	0.30	0.47	0.16
Precision (median)	0.24	0.40	0.11
Recall (mean)	1.00	0.87	−0.13
Recall (median)	1.00	0.98	−0.02
F_1 score (mean)	0.42	0.53	0.11
F_1 score (median)	0.38	0.52	0.10

6. Conclusions and Next Steps

We presented a suite of approaches to combining node-level telemetry with graph algorithms to identify and describe multi-home network impairments. Applying similar algorithms to additional categories of telemetry will help to further reduce the number of jobs generated for field teams while increasing the number of customer issues resolved. Possible candidates for such telemetry include MER for OFDM channels, upstream forward error correction (FEC) rate, and downstream PER.

With more telemetry comes more complexity, however. The approaches in this paper treat each category of telemetry separately, comparing and combining the results at the end (e.g. by cluster correlation in Section 4.3.2). As more telemetry is included in our analyses, it will be critical to evaluate this divide-and-conquer approach against more holistic approaches that attempt to perform clustering and root cause analysis on multiple attributes at once. The goal, ultimately, is a comprehensive data pipeline that ingests all pertinent customer device telemetry and generates jobs that are optimized for customer impact.

7. Acknowledgments

The authors thank Andy Martushev, Doug Sitkin, and Nathaniel Lee for operationalizing the algorithms in this paper and providing critical technical input. The authors also thank Ramya Narayanaswamy for helpful comments on a draft of this paper.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

BG	bonding group
CDF	cumulative distribution function
CMTS	cable modem termination system
CPE	customer premise equipment
DAA	distributed access architecture
DAAS	DAA switch
DBSCAN	density-based spatial clustering of applications with noise
DM	demand maintenance
FBC	full band capture
FEC	forward error correction
HFC	hybrid fiber-coaxial
MDU	multi-dwelling unit
MER	modulation error ratio
MIB	management information base
MSO	multiple systems operator
OFDM	orthogonal frequency-division multiplexing
PER	packet error rate
PHE	primary headend
PHY	physical layer
PM	proactive maintenance
PPOD	physical point of deployment
RF	radio frequency
ROCI	routing of cable infrastructure
RPD	remote PHY device
RX	receive
SC-QAM	single-carrier quadrature amplitude modulation
SG	service group
SHE	secondary headend
TX	transmit

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

Dugan, K., Evans, J. & Harb, M. (2022). A Deep Learning Approach for Detecting RF Spectrum Impairments and Conducting Root Cause Analysis. *SCTE 2022 Fall Technical Forum*.

Harb, M., Shen, W., Stehman, M., Walavalkar, S. & Rice, D. (2023). Accounting for Every MHz of Bandwidth: Data & Algorithms for Artifact Discovery and Close-Packing of QAMs in Support of Spectrum Activation. *SCTE 2023 Fall Technical Forum*

Narayanaswamy, R., Subramanya, K., Prodan, R. & Wolcott, L. (2021). When Physical Layer Simulation Gets Real. *SCTE 2021 Fall Technical Forum*.

Prodan, R. S. (2020). Optimizing the 10G Transition to Full-Duplex DOCSIS® 4.0. *SCTE 2020 Fall Technical Forum*.

Sundaresan, K. (2022). Network Capacity Options on the Path to 10G. *SCTE 2022 Fall Technical Forum*.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Hitting the Target: Measuring the Success of Online Learning

An Operational Practice prepared for SCTE by

Matthew Gress
Instructional Designer
Charter Communications
6399 S Fiddlers Green Circle
Greenwood Village, CO 80111
303-323-1327
Matthew.gress@charter.com

Sara Williamson
Manager, Curriculum Development
Charter Communications
6399 S Fiddlers Green Cir
Greenwood Village, CO 80111
303-323-1479
Sara.williamson@charter.com

Abbie O'Dell
Senior Director, Learning Services: Field Operations
Charter Communications
6399 S Fiddlers Green Cir
Greenwood Village, CO 80111
720-482-4205
Abbie.odell@charter.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Literature Review ... 3
   - 2.1. Learner Attitudes Toward Online Learning ... 3
   - 2.2. Role of Technology in Online Learning ... 3
   - 2.3. Self Directed Learning ... 4
   - 2.4. Impact of Online Learning on Performance ... 4
   - 2.5. Conclusion ... 4
3. Research Methods ... 5
   - 3.1. Participants ... 5
   - 3.2. Data Collection ... 5
   - 3.3. Data Analysis ... 6
4. Results ... 6
   - 4.1. Post-Training Evaluation Survey Results ... 6
   - 4.2. Performance Results ... 7
5. Discussion ... 8
   - 5.1. Learner Perceptions ... 8
   - 5.2. Business Performance ... 9
   - 5.3. Study Limitations ... 9
6. Conclusion ... 10

Abbreviations ... 11

Bibliography & References ... 11

List of Figures

Title **Page Number**

Figure 1 – Proactive Maintenance Success Rate ... 8

List of Tables

Title **Page Number**

Table 1 – Evaluation Results – Question 1 ... 6
Table 2 – Evaluation Results – Question 2 ... 7
Table 3 – Evaluation Results – Question 3 ... 7
Table 4 – Evaluation Results – Question 4 ... 7

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

As digital technologies continue to reshape various industries, the corporate world is witnessing an unprecedented shift in its approach toward training and development. Online learning has emerged as a pivotal tool in this context, offering flexibility, scalability and a broad range of learning experiences. The potential benefits of online learning for businesses, such as workforce upskilling, talent retention and increased productivity, are significant. The question of effectiveness remains critical.

In the telecommunications learning and development environment, many courses are geared toward preparing the learner for hands-on work. These courses have historically been conducted through in-person classroom courses. With the advent of online learning, the training medium has shifted, which leads learners and learning and development experts to wonder, "Is this effective?"

This study aims to identify the effectiveness of online learning. In the literature review, we discuss how the learner's attitude affects online learning outcomes, the role that technology plays in online learning, the effects of self-directed learning and the impact that online learning has on learning performance. Following the literature review, the training course series geared toward field technicians will be scrutinized. From these observations, we hope to measure the effectiveness of online learning in the telecommunications industry.

2. Literature Review

2.1. Learner Attitudes Toward Online Learning

Adult learners are unique in their approach to learning. Their previous life and work experiences inform their ability to understand and assimilate new information (Driscoll, 1998). Therefore, understanding these attitudes toward online learning is crucial in creating successful learning programs. As online education is a relatively new and rapidly evolving field, learners might not be comfortable with it yet. Interestingly, despite the advantages of convenience and accessibility, many still express preferences for the personal interaction and immediacy of feedback found in traditional face-to-face courses (Kuong, 2015). However, once they overcome the initial technological hurdles, learners generally find online platforms user-friendly and easy to navigate (Almahasees et al., 2021). To harness the full potential of online learning, an attitude shift toward positivity and readiness is needed (Hergüner et al., 2020).

2.2. Role of Technology in Online Learning

Technology plays a pivotal role in online learning, acting as the bridge between content and learner. It is not just about technology though; successful online learning also requires time investment and support from leadership or peers (Noosegard & Ørngreen, 2015). Learning management systems (LMS) like Blackboard or Cornerstone have revolutionized online learning settings, offering a centralized hub for learning materials and a powerful source of analytics. However, the digital divide can pose a challenge, as access to stable internet is not universal. This can impact the efficacy of online learning, especially in multimedia-rich environments (Seladorai & Mohamed, 2021). It underscores the necessity of strong infrastructure to ensure every learner can access the content (Gaebel et al., 2013). With the increasing digitization of workforce training, it is clear the future of employee development is intertwined with online learning (Chen, 2008).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.3. Self Directed Learning

The nature of online learning often necessitates a self-directed approach, where learners are responsible for their own learning journey. This sense of autonomy can be empowering for adult learners, providing control over their learning process (Knowles et al., 1998). However, the tendency for learners to choose safe, comfortable paths could inhibit their growth, posing an interesting conundrum for content designers (Knowles et al., 1998). It is also worth noting that not all adults may be fully equipped for self-directed learning and, in such cases, instructional scaffolding is essential to guide learners effectively (Jumaat & Tasir, 2014).

Online learning is essential for learning and development organizations because it opens up learning outside of the traditional classroom, which allows for self-directed learning. Online learning is an ideal platform that can be used both in the classroom as well as outside the classroom, giving autonomy to learners to seek their knowledge in a way that resonates with them. “Although this time in our work lives may be a period of stability – at least in comparison to the last few years – there is still an ongoing workplace change that organizations must adapt to – the increases in employee agency. This means organizations need the help of software and services designed and adapted for addressing The Great Resignation” (Stevenson et al. 2022).

2.4. Impact of Online Learning on Performance

The impact of online learning on performance is multi-faceted. It is not just about imparting knowledge, but how it can be applied in a practical context. Factors such as the quality of learning materials, the learner's motivation, the robustness of the technological infrastructure and the relevance of the content to the learner's job all play a role in the efficacy of online learning. Therefore, a successful e-learning initiative must be closely aligned with organizational objectives and learner interests, and should foster a supportive learning environment (Chen, 2008). Business objectives can be measured from online training through on-the-job training (OJT) effectiveness. OJT effectiveness items can include overall task achievement, job accomplishing satisfaction and job performance (Cheng & Chen, 2008). If the impact of OJT effectiveness is measurable, a measure of the training effectiveness can be completed too. This leads us to the notion that a majority of studies report online programs as effective or even superior to traditional classroom learning, signifying the transformative potential of online learning in the workplace (Noosegard & Ørngreen, 2015). OJT effectiveness helps to measure the success of online learning.

2.5. Conclusion

Existing research provides insights into how learner attitudes affect online learning outcomes, the role that technology plays in online learning, the effects of self-directed learning and the impact that online learning has on learning performance. Core adult learning theories on self-direction support the effectiveness of online learning methodologies, and more recent research highlights potential challenges for learners. There is still a gap in the collective body of work, specifically related to the opportunities to clearly align quantitative performance data back to online learning experiences in the workplace.

This quantitative study examines the effectiveness of targeted online learning materials within the workplace. Through an investigation of the learner attitudes, technology, self-directed learning and the success of online learning, we will seek to understand how online learning materials impact learning and business outcomes. To do this, we pose following questions:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Do short, targeted online modules and online learning materials impact learner experience and business outcomes?
- What are the attitudes of online learners?
- What is the role of technology in online learning effectiveness?
- How does online learning impact performance in the workplace?

3. Research Methods

As detailed in the literature review, there is a gap in current research that directly addresses aligning quantitative performance measures back to online learning programs. To investigate a relationship between performance and online learning, a quantitative method that seeks to understand performance measures of a sample group was used. This quantitative method provides the opportunity to both understand learner perceptions and attitudes of online experiences and to subsequently identify possible relationships to performance. Quantitative data provides a more concrete picture of impacts to job performance, helping to validate the investments of learning and development organizations into their online learning programs.

3.1. Participants

The participants whose performance and attitudes were investigated as part of the study are frontline field technicians in a large telecommunications organization who participated in a series of online training courses on proactive maintenance. These modules covered topics including core technical concepts such as modulation error ratio (MER), downstream uncorrectables and signal-to-noise ratio (SNR), as well as instructions on how to use internal telemetry tools for problem diagnosis and troubleshooting steps to resolve such issues.

3.2. Data Collection

The internal research of the proactive maintenance series covered four major topics. The first topic was learner attitudes. Ten proactive maintenance online learning courses were identified as the main courses in question and their data was assembled. At the time of data collection, the minimum number of completions for all e-learns was 10,400 and the maximum was 12,800, with a mean of 11,500. The users or learners who completed these courses are field technicians with varying degrees of experience. 1,039 user ratings from the LMS were also obtained. A report that generated multiple choice evaluations from users was also compiled and included 82 multiple choice evaluations between all proactive maintenance courses. The multiple-choice evaluation statements were as follows:

- Select the statement that best describes examples and activities in the course.
- Select the statement that best describes the use of technology in this course.
- Select the statement that best reflects your ability to apply course information.
- Select the statement that is most applicable.

A retroactive observation of the technology used in this online learning series was also completed. This observation included information covering the implementation of the series, the technology used to design the series and the technology used by the learners. Information about the online learning series courses was also collected. This included the naming scheme of the lessons, the duration of the entire series and the expected duration of each course.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The final collection of data came from the learning and business outcomes. For the learning outcomes, data retrieved identified the changes to trouble calls and their resolution. Business outcome information was also provided by the proactive maintenance department and included information covering success rate, ticket completion and "refer to maintenance" work orders. The final data point was the metrics derived from completions of proactive maintenance courses and the total assignments.

3.3. Data Analysis

Post-training learner evaluation data were exported from the LMS into an Excel spreadsheet, and all personally identifiable user data was removed. The star ratings were organized and aggregated as an average overall score by individual course and by overall program (all courses). The post-training survey data was organized by question and responses into table format and expressed as a percentage of positive responses.

Performance data were gathered from internal work order data systems and organized into an aggregate score known as "proactive maintenance success." Proactive maintenance success was identified as all impairments being resolved and compliant telemetry on the account for 48 hours post truck roll. These data were expressed as a graph showing overall company performance for all employees (see Figure 1).

4. Results

4.1. Post-Training Evaluation Survey Results

To examine this online learning program's effectiveness, the course evaluations were first considered. A cumulative average rating of 4.58 stars out of five stars was recorded from the ten e-learns. User-generated responses and evaluations from the entire training series were positive overall. Of the 10 courses identified for this study, the lowest star rating for a course was 4.4 stars and the highest rating was 4.7 stars. The average between all courses was 4.58 stars.

Users were presented questions that asked them to characterize their impressions of the course. Question 1 asked learners to consider the examples and activities in the course and rate whether they helped the learner understand the content. The overall score for Question 1 was 81 percent positive, with the majority of learners indicating they felt the examples and activities helped (See Table 1).

Table 1 – Evaluation Results – Question 1

Select the statement that best describes examples and activities in the course.	58
The examples or activities in the course did not help my understanding of the content.	11
The examples or activities in the course somewhat helped me to understand the content.	12
The examples or activities in the course helped me to understand the content.	12
The examples or activities in the course significantly helped me to understand the content.	23

Question 2 asked learners to consider the use of technology in the course and their overall impressions of whether technology enhanced or interfered with their ability to learn. The score for this question was 63 percent positive, with 20 of the total 54 respondents indicating the use of technology "did not enhance" or "interfered with" their learning (see Table 2).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 2 – Evaluation Results – Question 2

Select the statement that best describes the use of technology in this course (if applicable).	54
Technology used in the course interfered with my ability to learn.	10
Technology used in the course did not enhance my ability to learn.	10
Technology used in the course enhanced my ability to learn.	13
Technology used in this course greatly enhanced my ability to learn.	21

Question 3 asked learners to consider their confidence level relative to their ability to apply what they learned back on the job. The majority of learners were favorable on this question, with an overall positive score of 62.5 percent. Despite this positive response overall, it is noteworthy that 21 learners of 56 total respondents still rated it as "not at all" or "somewhat" confident (see Table 3).

Table 3 – Evaluation Results – Question 3

Select the statement that best reflects your ability to apply course information.	56
I am not at all confident in my ability to apply the course content to my job.	10
I am somewhat confident in my ability to apply content to my job.	11
I am confident in my ability to apply course content to my job.	10
I am very confident in my ability to apply course content to my job.	25

The final evaluation question asked learners to consider whether the overall information was useful and could be applied back on the job. 25 of the 36 total respondents to this question indicated that the content was useful, for an overall score of 69.4 percent (see Table 4).

Table 4 – Evaluation Results – Question 4

Select the statement that is most applicable.	36
This course did not provide information which I can apply in my role.	11
This course provided useful information which I can apply in my role.	25

4.2. Performance Results

Course attendees' performance was measured relative to an overall proactive maintenance success score. Proactive maintenance success is identified as all impairments being resolved and compliant telemetry on the account for 48 hours post truck roll.

Before the introduction of the e-learns, proactive maintenance success rates hovered around 80.96 percent Following the introduction of the online learning series, an increase in success rates can be clearly seen. From July of 2022 to January of 2023, success rates climbed to a peak of 88.40 percent and the average success rate was 87.83 percent.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

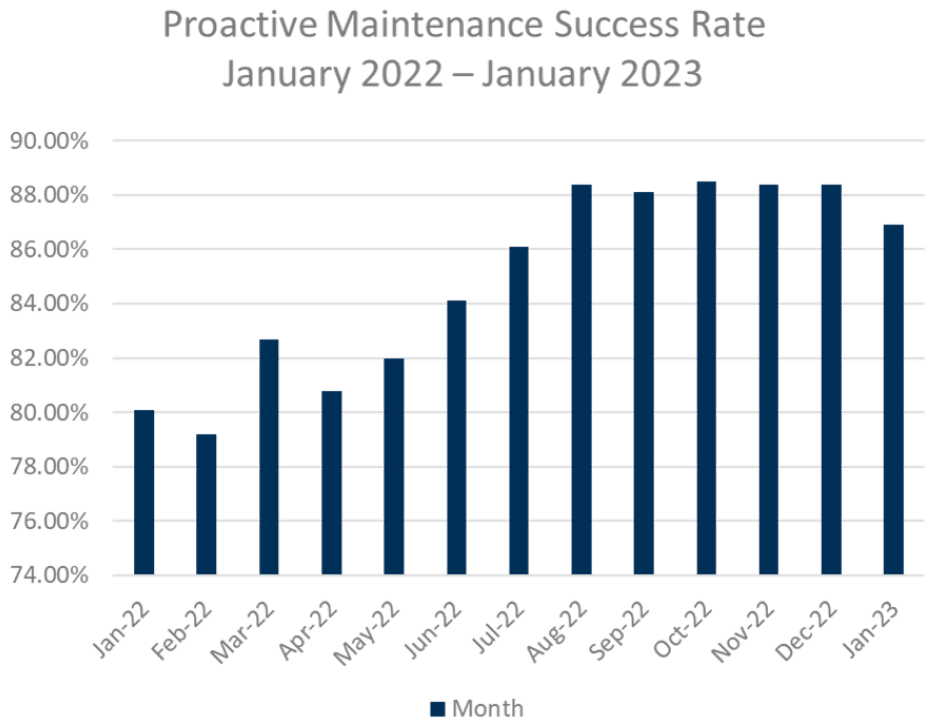


Figure 1 – Proactive Maintenance Success Rate

5. Discussion

This research sought to examine the effectiveness of online learning and was completed in the context of telecommunications training, focusing specifically on one e-learn series. We aimed to evaluate learner attitudes toward the series, the role that technology played, the effects of self-directed learning and the impact of online learning on job performance and business outcomes.

5.1. Learner Perceptions

Learners' overall star ratings of the courses were overwhelmingly positive (4.58 out of 5 stars), indicating that the perception of these online courses was generally positive for the majority of learners, aligning with the research (Almahasees et al., 2021). While these star ratings do not provide the context needed to fully identify the impact of online learning effectiveness, they do shed light on the learner attitudes. These e-learns were well-received by the thousands of learners that completed the courses.

The post-course evaluations provide additional insights into learner attitudes through multiple-choice evaluation statements. While the examples and activities included in the training series are different from face-to-face activities or hands-on examples that would be presented in facilitated courses, learners still feel that the activities and examples helped them to better understand the content. Of the 58 users that responded to this survey statement, 47 responded positively. This is an 81 percent positivity rate for the examples and activities. The proactive maintenance online series included many examples of real-world telemetry that users would see on their signal level meters when on a proactive maintenance trouble call.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Similar to what Driscoll (1998) identified, learners were able to make the connections between their own equipment and the new processes.

Considering the topic of learner experience, our research found that 63 percent of learners (34/54) reported the technology in the training series enhanced their learning experience. This aligns with the literature suggesting technologically mediated instruction can provide a more flexible self-paced learning environment (Noosegard & Ørngreen, 2015). This could also be attributed to the robust LMS that the programs reside in. The LMS provides an all-in-one experience for the user so they can access the online learning courses at any time and any place. Also, due to the way these courses were created and the understanding that learners may have limited access to strong internet connections, learners can still access these courses from anywhere with minimal internet connection. This allows users to self-direct their learning experience even more, no matter where they are or what device they choose to access the training with.

A sizable portion of learners, 37 percent, felt that technology did not enhance their learning or interfered with their ability to learn. This could be attributed to potential barriers such as digital literacy. These results underline the importance of intuitive, user-friendly online course design.

The findings from Question 3, related to ability to apply the learnings, suggest a strong correlation between online course participation and confidence in job-related application. 82.1 percent (46/56) of participants expressed confidence in their ability to apply the course information to their job. This supports the argument that effective online learning can provide pragmatic skills transferable to the work environment (Chen, 2008). Training's effectiveness can be measured by this ability to apply learned concepts, which can lead to enhanced performance.

The final survey question covered the application of course information to their job role. The relevance of learning material to an individual's role is a crucial aspect of effective training. Of the participants who responded, 69.4 percent (25/36) affirmed that the training series provided them with information applicable to their job. This finding supports the literature assertion that aligns between the learning content and real-world job roles and the enhancement of the training's effectiveness (Chen, 2008). However, the data also implies that 30.6 percent (11/36) of responding learners did not find the course content to be directly applicable to their roles. This gap in relevancy for learners points toward an opportunity for further refinement of the proactive maintenance series content. Further refinement may lead the content of online learning to be more relevant to the diverse group of learners.

5.2. Business Performance

Our study found a significant increase in success rates following the June 2022 release of the training series. Before the introduction of the e-learns, proactive maintenance success rates hovered around 80.96 percent. Using the January 2022 and January 2023 data points, we see an approximate 6.9 percent increase in success rates. Most notably, this growth in success rates was not a temporary boost, but rather showed sustained positive changes after the introduction of the training. These findings are in line with our research that suggests a strong correlation between effective online training and improved performance metrics (Cheng & Chen, 2008).

5.3. Study Limitations

This study, while providing valuable insights, is marked by several limitations that we must consider when interpreting the results. Limitations are expected in studies, especially when working in an evolving

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

business environment and these limitations do not discredit the evidence that we uncovered in the research (Kirkpatrick, 1996).

The first limitation is the lack of a fully controlled research group. While this training series is geared toward field technicians, it is not exclusive to field technicians. This may introduce minor variability and inconsistency in the data.

The second limitation is the limited evaluation options. The evaluations used in this research are pre-set and were not customized for these courses. Without the ability to set custom evaluations, the study may have missed information pertaining to the attitudes of the learners. Tailored evaluations could provide more precise insights into how we measure and enhance the success of online learning.

The final limitation is the variables that coincide with the introduction of the training series. Other variables may have affected the business outcomes, such as the enhancements to tools used by technicians and the reprioritization of proactive maintenance severity levels. These factors could change the relationship between the online training and the observed increase in proactive maintenance success rates.

These limitations highlight areas that could be addressed in future research, potentially leading to a better understanding of the dynamics of online learning. It is essential that readers consider these limitations when interpreting the conclusions drawn from this study.

6. Conclusion

This research explored the effectiveness of online learning in the context of the telecommunications industry, focusing on the proactive maintenance e-learn series. The investigation covered a diverse range of factors, including learner attitudes, technology's role, self-directed learning and the impact on business outcomes.

The findings indicate that a majority of the learners found the use of technology as an enhancement to their learning experience and felt confident in applying the course content to their jobs. Further evaluations prove that learner attitudes are overwhelmingly positive when completing online learning. These positive responses, combined with the demonstrated increase in business success rates, provide robust evidence supporting the effectiveness of online learning.

This study also identified areas for potential improvement. A sizable portion of learners did not see technology as an enhancer or did not find the course content directly applicable. Addressing these areas could further boost the success of future online learning initiatives.

In conclusion, while online learning shows significant promise as an effective medium for training in the telecommunications industry, there is still room for enhancement and a continued need for investigation. With further research and improvements to our online learning materials, we can further optimize the impact of online learning on the development of our workforce and performance of our businesses.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

LMS	Learning Management System
OJT	On-the-job training
MER	Modulation error ratio
SNR	Signal to noise ratio

Bibliography & References

Almahasees, Z., Mohsen, K., & Amin, M. O. (2021). Faculty's and Students' perceptions of online learning during COVID-19. Frontiers in Education, 6. https://doi.org/10.3389/feduc.2021.638470

Bowles, M. (2003). Learning to E-learn Project: Rediscovering the Benefits of E-learning. Malaysian Online Journal of Instructional Technology

Chen, E. T. (2008). Successful E-learning in corporations. Communications of the IIMA, 8(2). https://doi.org/10.58729/1941-6687.1080

Cheng, W. T., & Chen, C. (2015). The impact of E-learning on workplace on-the-job training. International Journal of E-Education, e-Business, e-Management and E-learning, 212–228. https://doi.org/10.17706/ijeeee.2015.5.4.212-228

Driscoll, M. (1998). Web-based training: Using technology to design adult learning experiences. Jossey-Bass/Pfeiffer.

Gaebel, M., Kupriyanova, V., Morais, R., & Colucci, E. (2014). E-learning in European Higher Education Institutions. https://www.eunis.org/wp-content/uploads/2015/02/E-learning-survey2.pdf

Hergüner, g., Buğra, S., Hergüner, s., & Dönmez, a. (N.D.). The effect of online learning attitudes of university students on ... - ed. Https://files.Eric.Ed.Gov/fulltext/EJ1272871.Pdf

Jumaat, N. F., & Tasir, Z. (2014). Instructional scaffolding in online learning environment: A meta-analysis. 2014 International Conference on Teaching and Learning in Computing and Engineering. https://doi.org/10.1109/latice.2014.22

Kirkpatrick, D. L. (1996). Evaluating training programs the four levels. Berrett-Koehler.

Knowles, M. S., Holton, E. F., Swanson, R. A., & Robinson, P. A. (1998). The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development (5th ed.). Butterworth-Heinemann.

Kuong, H. C. (2015). Enhancing online learning experience: From learners' perspective. Procedia - Social and Behavioral Sciences, 191, 1002–1005. https://doi.org/10.1016/j.sbspro.2015.04.403

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Seladorai, D., & Mohamed, M. (2021). Digital learning among postgraduate students in the times of covid-19: A literature review. Creative Education, 12(07), 1494–1502. https://doi.org/10.4236/ce.2021.127114

Stevenson, C., Cooke, M., Rochelle, M., & Cooke, R. (2022). Challenges of Onboarding: The Ups and Downs of Onboarding in Remote or Hybrid Work. (E. Bui & R. Pachter, Eds.). Brandon Hall Group & NovoEd.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

How ATSSS and Trusted Wi-Fi Access Will Enable the Converged Mobile User Experience

A Technical Paper prepared for SCTE by

Christopher Burke
Principal Engineer III
Charter Communications
6360 S Fiddler's Green Circle, Greenwood Village, CO
720-699-6430
christopher.burke@charter.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Access Traffic Steering, Switching, and Splitting ... 4
   2.1. Wi-Fi Connectivity ... 7
3. Trusted Wi-Fi Access ... 7
4. What About Untrusted Wi-Fi Access? ... 10
5. Putting 5G/Wi-Fi Convergence to Use ... 12
   5.1. Converged Experience #1: Mobility Coverage and Network Resiliency ... 13
   5.2. Converged Experience #2: Mixed Private Wireless ... 14
   5.3. Converged Experience #3: Improved Full MVNO Offload ... 14
   5.4. Converged Experience #4: Wi-Fi Mobile Operators ... 15
6. Conclusion ... 15

Abbreviations ... 16

Bibliography & References ... 17

List of Figures

Title **Page Number**

Figure 1 - Convergence Concept ... 3
Figure 2 - ATSSS Concept ... 4
Figure 3 - ATSSS architecture as described in 3GPP TS 23.501 ... 5
Figure 4 - UE ATSSS Steering Functionalities as described in 3GPP TS 23.501 ... 6
Figure 5 - ATSSS Architecture w/ Steering, Policy, and Performance Details ... 7
Figure 6 - Trusted Non-3GPP Access as described in 3GPP TS 23.501 ... 8
Figure 7 - Example Trusted Wi-Fi Access Architecture Using Existing Wi-Fi System ... 8
Figure 8 - Trusted non-3GPP Access Authentication and Registration as described in 3GPP TS 23.502 ... 9
Figure 9 - Trusted non-3GPP Access PDU Session Establishment as described by 3GPP TS 23.502 ... 10
Figure 10 - 4G non-3GPP Architecture as described by 3GPP TS 23.402 ... 11
Figure 11 - Untrusted non-3GPP Access as described in 3GPP TS 23.501 ... 11
Figure 12 - NSWO as described in 3GPP TS 23.501 ... 12
Figure 13 - ATSSS and Trusted Wi-Fi Combined Architecture ... 13
Figure 14 - 5G/Wi-Fi Convergence User Transitions ... 13
Figure 15 - Full MVNO ATSSS Architecture ... 14
Figure 16 - Example Home Routed Roaming Architecture with Trusted Wi-Fi Access ... 15

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Mobile devices today can connect to services over cellular and Wi-Fi networks; two completely different technologies with their own designs and sets of standards. Additionally, mobile services are often combined with other services, such as video or internet, and charged to the customer on a single bill. Does this provide a converged user experience? Not really.

While bundling of services is convenient to the customer, it doesn't really converge the delivery of these services. And a device that can connect to multiple networks is great, but if the networks are completely separate, with different authentication methods, policy enforcement, charging rates, and data paths to services, then these networks merely coexist and are not converged.

The convergence vision is to deliver ubiquitous connectivity to customers anywhere and on any device, delivered on high capacity and low latency networks. This means that customers would carry their applications, services, policies, and identity with them wherever they go – inside or outside of home, on a fixed or mobile device.

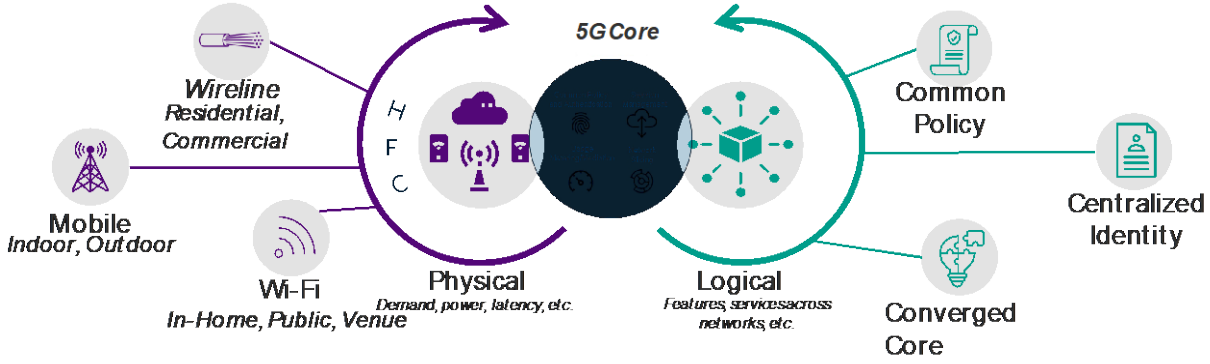


Figure 1 - Convergence Concept

The emergence of 5G networks provides the ability to deliver on this experience, with standards detailing the early beginnings of the convergence vision. While 5G convergence looks at several different convergence opportunities, one of the first areas to explore is mobile 5G/Wi-Fi convergence because it can offer significant benefits such as:

- Mobility: The ability to seamlessly combine Wi-Fi with cellular connectivity is extremely valuable, especially for use cases like moving through indoor/outdoor campuses.
- Resiliency: Combining 5G and Wi-Fi to provide better coverage in any space and making use of both simultaneously.
- Security: Benefits from the ability to authenticate all devices via 5G cores, even as some operate over non-cellular networks.
- Differentiation: Many cable operators with extensive Wi-Fi deployments now act as mobile virtual network operators (MVNOs). New public/private solutions that tightly integrate 5G with Wi-Fi could open compelling new business.

A converged mobile experience is about having a singular experience that follows you wherever and however you connect, utilizing the same authentication credentials, subscriber policies, charging rules, and gateways. This paper explores how two 5G mobile technologies, access traffic steering, switching,

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

and splitting (ATSSS) and trusted Wi-Fi access, can work together to provide the first converged mobile user experience.

2. Access Traffic Steering, Switching, and Splitting

Beginning with 3GPP Release 16, 5G introduces the access traffic steering, switching, and splitting (ATSSS) feature, which allows the service provider to steer, switch, and split traffic across 3GPP (5G cellular) and non-3GPP (Wi-Fi) access networks.

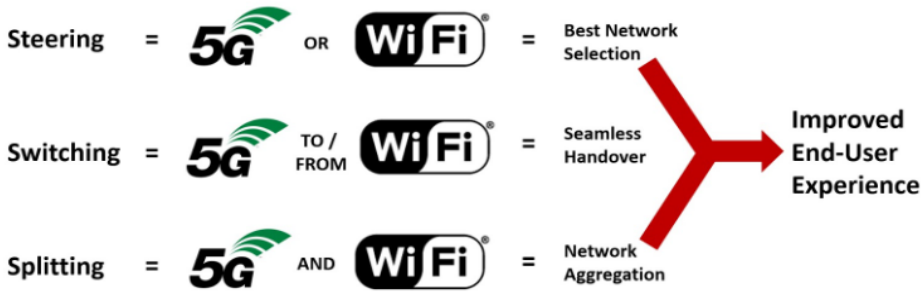


Figure 2 - ATSSS Concept

As a key technology in 5G/Wi-Fi convergence, the implementation of ATSSS by service providers can improve the overall quality of experience (QoE) through:

- Use of standards-based common authentication and policy control to select the best available access network.
- The ability to facilitate seamless transition between cellular and Wi-Fi networks.
  - Customers don't need to disable Wi-Fi due to dead zones.
  - Improved performance for sensitive applications (gaming, banking, OTT voice).
- 5G cellular and Wi-Fi data aggregation, which will deliver speeds that combine the channels from both access networks to be used simultaneously.

ATSSS provides simultaneous connectivity across 3GPP networks and non-3GPP networks to a common user plane function (UPF) within the 5G core, allowing for common authentication and policy control and the smoothing of transitions between access types.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

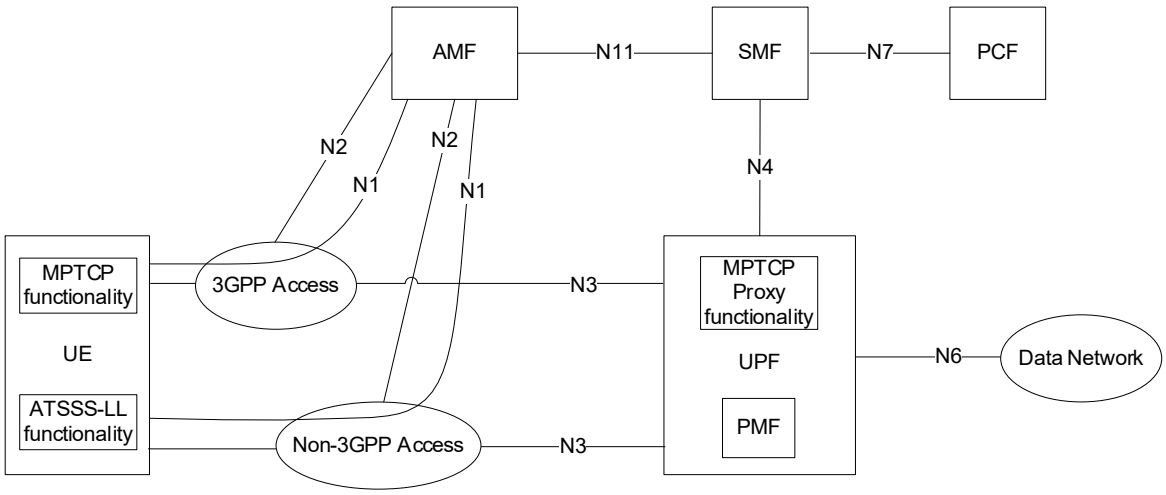


Figure 3 - ATSSS architecture as described in 3GPP TS 23.501

To enable ATSSS, the user equipment (UE) requests a multi-access packet data unit (MA-PDU) session, which may have user plane resources on two access networks: one 3GPP access network and one non-3GPP access network, with independent N3 tunnels between each access network and a single UPF in the 5G core.

An ATSSS capable UE specifies support for one or more steering functionalities:

- High Layer Steering: Steering functionality that operates above the IP layer.
  - The initial support for high layer steering is multi-path TCP (MPTCP), which can be applied to steer, switch, and split TCP traffic. The MPTCP functionality in the UE communicates with an associated MPTCP proxy functionality in the UPF, using MPTCP protocol over the 3GPP and/or the non-3GPP user plane.
  - In 3GPP Release 18, high layer steering will include multi-path QUIC (MPQUIC), which can be applied to steer, switch, and split UDP traffic. This is not covered in this paper but operates functionally the same as MPTCP.
- Low Layer Steering: Steering functionality that operates below the IP layer.
  - ATSSS-LL steering functionality can be applied to steer, switch and split all types of traffic, including TCP traffic, UDP traffic, Ethernet traffic, etc.

The UE may indicate support for ATSSS-LL only, or support for both ATSSS-LL and MPTCP. If using MPTCP, the UE must be assigned three IP addresses; a MA-PDU session IP address and two link-specific multipath IP addresses.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

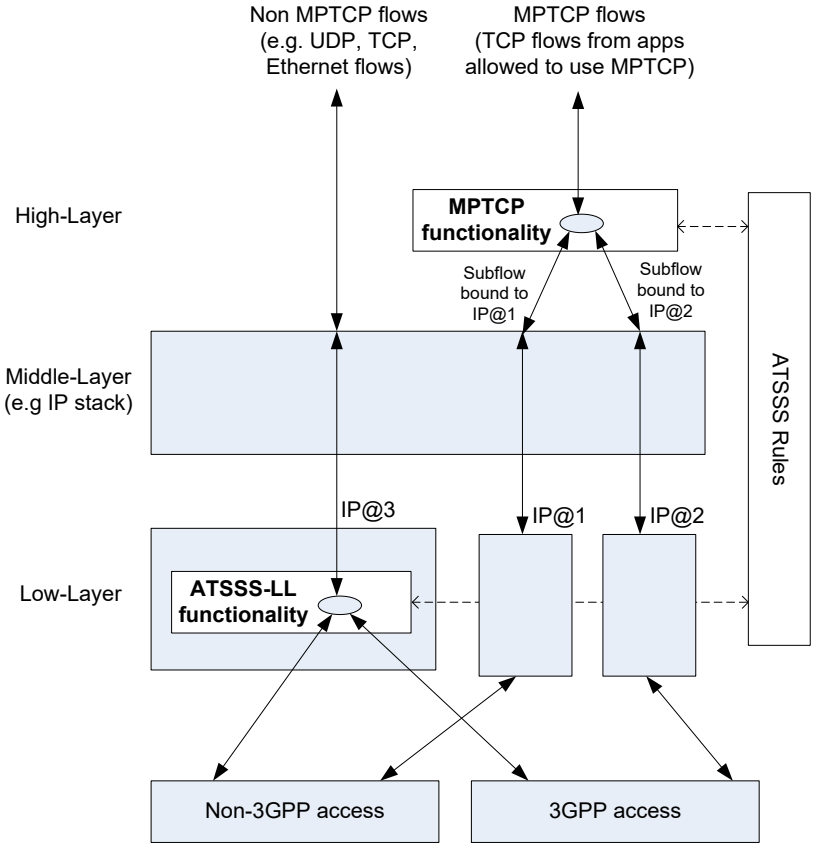


Figure 4 - UE ATSSS Steering Functionalities as described in 3GPP TS 23.501

Once the MA-PDU session is established, the UE uses network provided policy rules (ATSSS rules) and network measurements to decide how to distribute the uplink traffic across the two access networks. Likewise, the UPF uses network provided policy rules (MAR) and network measurements to decide how to distribute the downlink traffic across the two access networks. Network performance measurements are exchanged between UE and UPF performance measurement functions (PMF) using the PMF protocol (PMFP) to assist with determining the latency, packet loss, and access network availability. Traffic is routed in any combination of active/standby, smallest delay, load balancing, or priority, based as determined by policy and network conditions. The UE and UPF may make uplink and downlink routing decisions, respectively, independent of each other.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

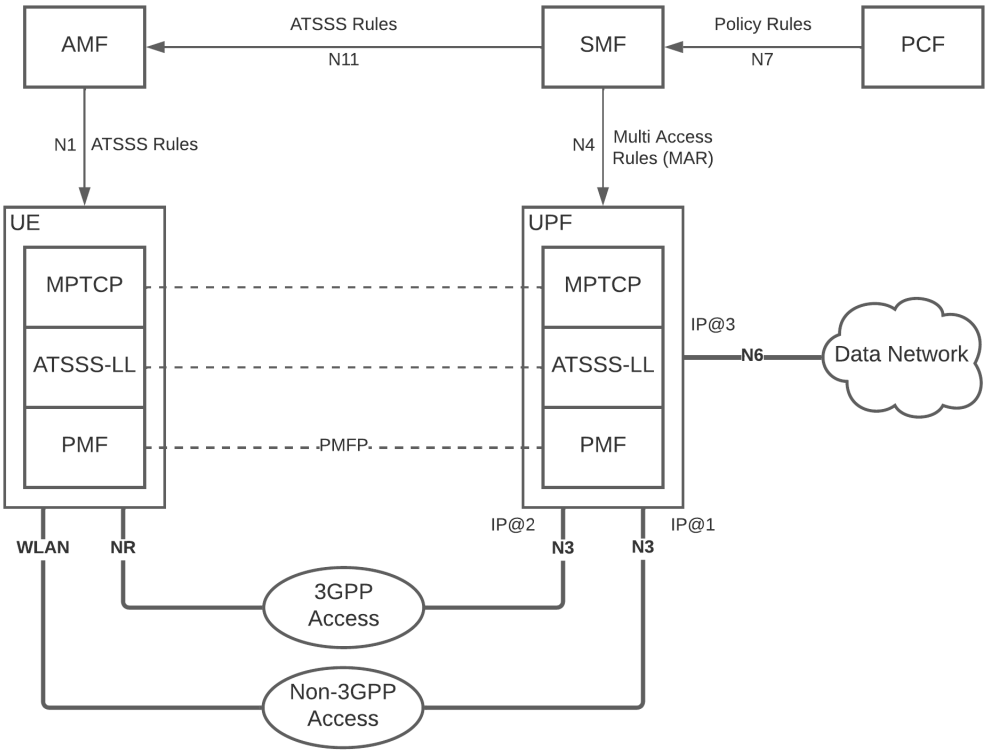


Figure 5 - ATSSS Architecture w/ Steering, Policy, and Performance Details

2.1. Wi-Fi Connectivity

3GPP access is the 5G radio access network (RAN), which connects natively to the 5G core and is well defined in 3GPP standards, making non-3GPP access the key component of ATSSS. ATSSS can technically work with any non-3GPP access (WLAN, wired, etc.), but for mobile device 5G/Wi-Fi convergence, we will focus on Wi-Fi connectivity specifically.

3. Trusted Wi-Fi Access

Beginning with 3GPP Release 16, 5G introduces the trusted non-3GPP access feature, which allows Wi-Fi access for mobile devices to behave similarly to the 5G RAN regarding network discoverability, authentication, registration, and PDU session establishment.

Trusted non-3GPP access used for mobile device Wi-Fi access is a key component in 5G/Wi-Fi convergence because it enables the mobile device to:

- Select the Wi-Fi access based on PLMN ID.
- Gain authenticated access to the Wi-Fi network using the device's SIM credentials.
- Register with the 5G core.
- Retain mobile data policies and services while connected to Wi-Fi.

Trusted non-3GPP access introduces the concept of a trusted network access network (TNAN), which consists of the trusted non-3GPP gateway function (TNGF) and the trusted non-3GPP access point.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

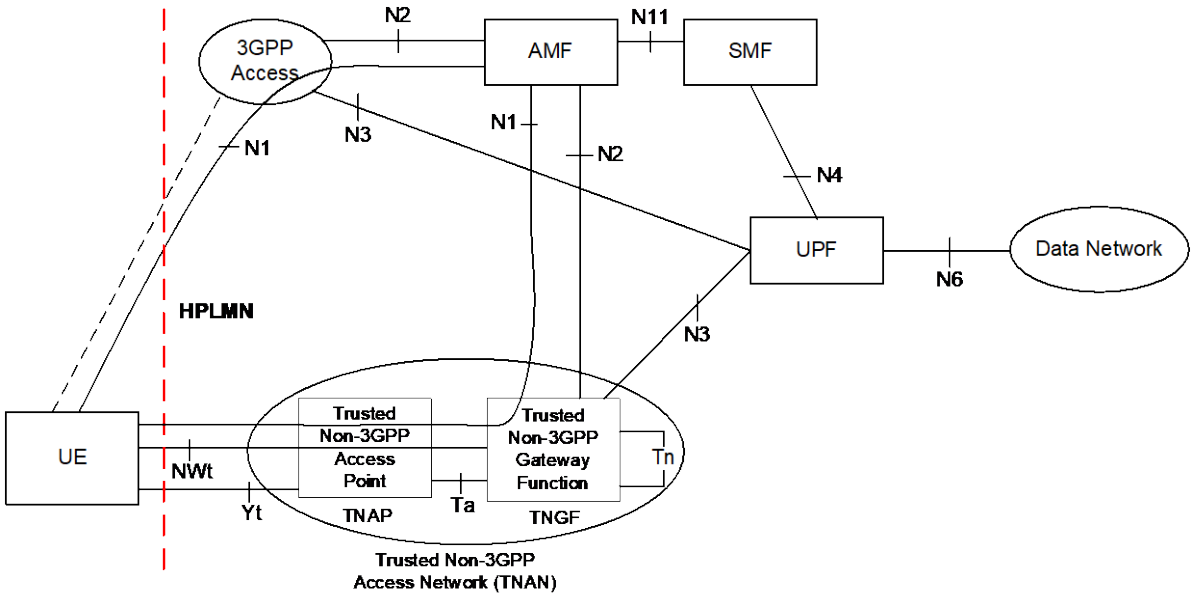


Figure 6 - Trusted Non-3GPP Access as described in 3GPP TS 23.501

The TNGF provides interworking for the UE to connect with and interface with the 5G core for 5G authentication and registration over a standard N2 interface to the AMF. The 5G user plane is handled over a standard N3 interface to a UPF.

With trusted Wi-Fi access, the TNAP is a Hotspot 2.0 compliant Wi-Fi system. The Ta interface between the TNAP and the TNGF is a AAA interface that facilitates authentication procedures between the Wi-Fi system and the 5G core using either RADIUS or diameter messaging. Any existing Wi-Fi system that supports the Hotspot 2.0 standard already supports AAA authentication; therefore, the entire Wi-Fi system effectively becomes the TNAP.

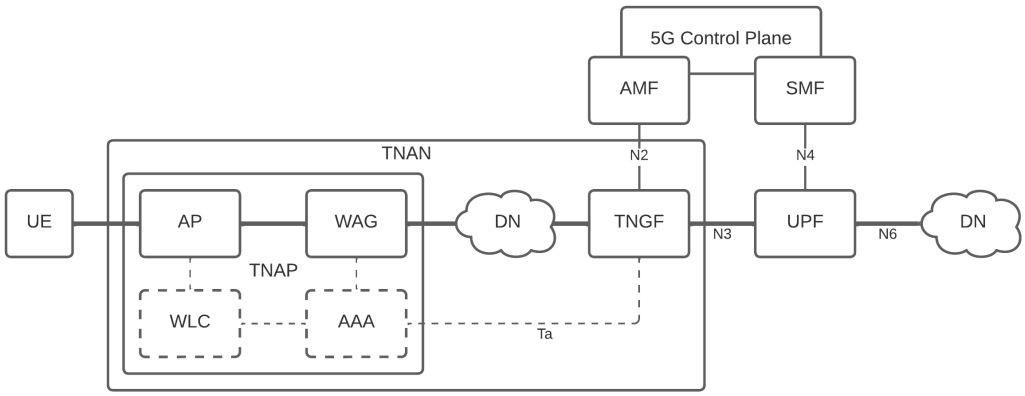


Figure 7 - Example Trusted Wi-Fi Access Architecture Using Existing Wi-Fi System

In the trusted Wi-Fi access system, the Wi-Fi Hotspot 2.0 access point (AP), using the access network query protocol (ANQP), advertises the public land mobile network identifier (PLMN ID) and an indicator that it supports 5G connectivity. The mobile device requests the "3GPP cellular network" information

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

over ANQP and selects the highest priority service set identifier (SSID) of a WLAN that contains a suitable PLMN ID with 5G connectivity.

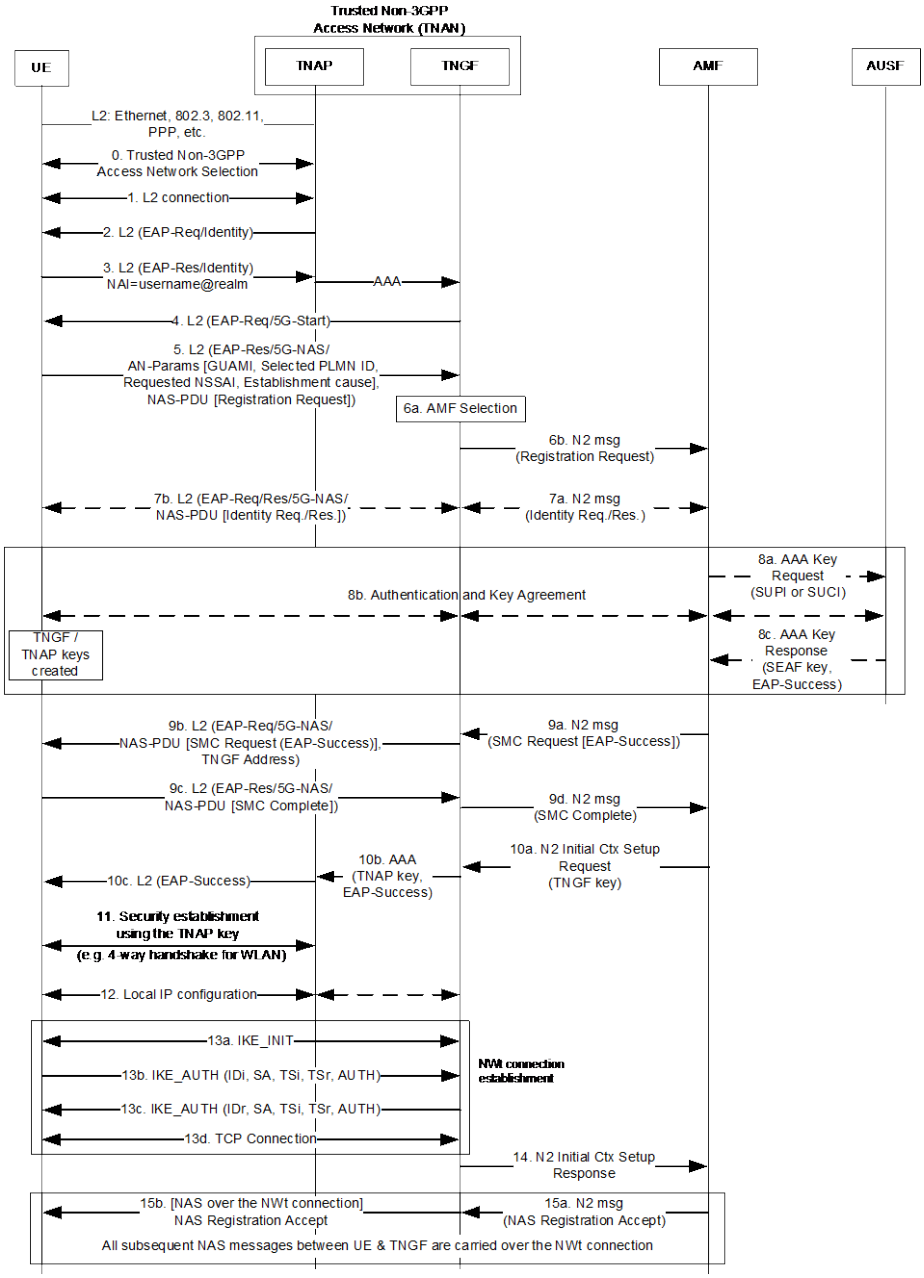


Figure 8 - Trusted non-3GPP Access Authentication and Registration as described in 3GPP TS 23.502

The UE then uses standard EAP authentication procedures to authenticate with the 5G core through the trusted network gateway function (TNGF). The TNGF has an AAA interface (Ta) that supports the exchanging of EAP messages with the Wi-Fi authentication systems using either RADIUS or diameter messaging. The EAP payload is set to “EAP-5G” which is an extended EAP type method that allows for the full exchange of 5G NAS messages between the UE and the 5G core.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Once authenticated, the mobile device is granted access to the WLAN and receives an IP address. The UE then establishes an IPSec tunnel directly to the TNGF (NWt) to complete 5G registration and establish PDU sessions using 5G NAS. The TNGF creates IPSec child security associations to the UE for each PDU session.

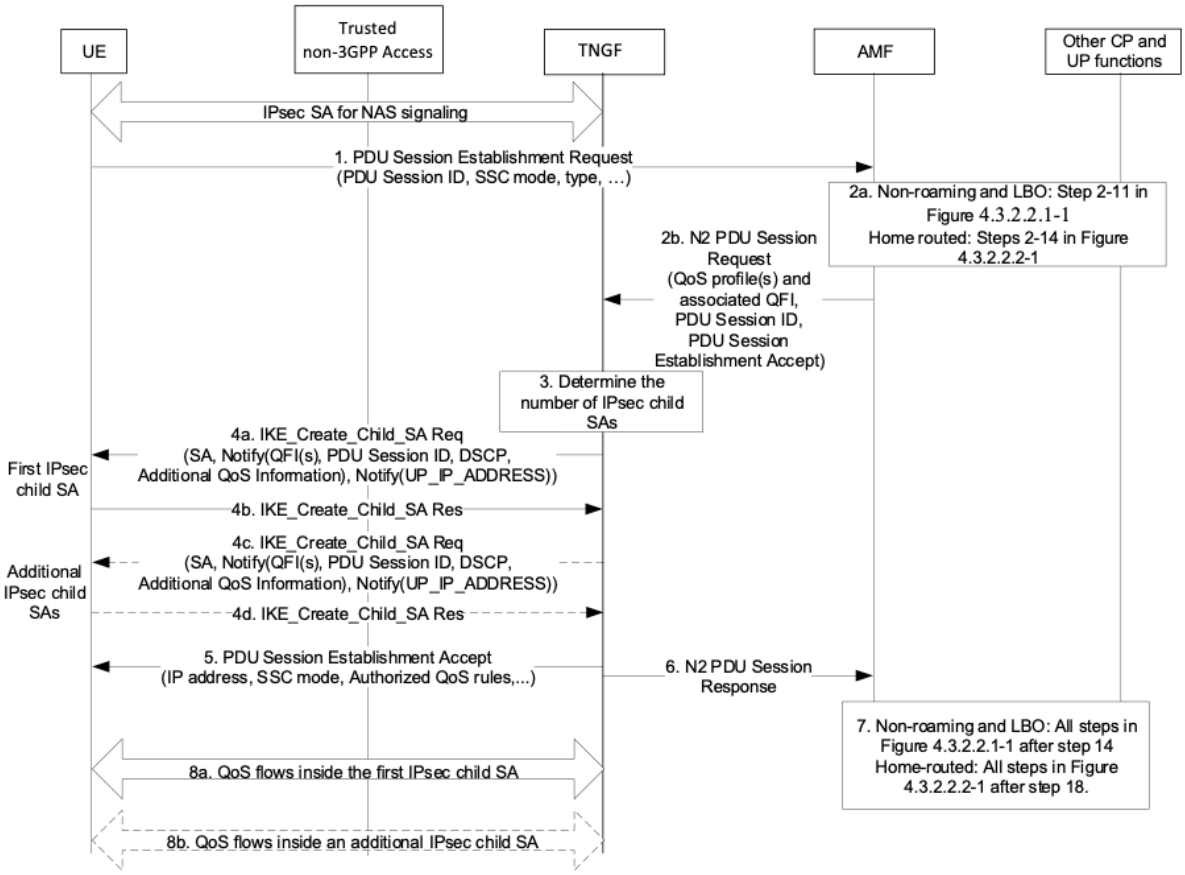


Figure 9 - Trusted non-3GPP Access PDU Session Establishment as described by 3GPP TS 23.502

The complete trusted Wi-Fi access flow enables the mobile device to authenticate, register, and establish data sessions to the 5G core over Wi-Fi access as if it were connecting over cellular, using the exact same credentials, 5G NAS control plane, and 5G user plane.

4. What About Untrusted Wi-Fi Access?

The goal of convergence is to bring together disparate networks and services to provide a new user experience. Part of the requirements of convergence is the seamless nature of the converged experience. 5G untrusted non-3GPP access does not define how the user connects to a Wi-Fi network and does not define a method or architecture to support authentication with the 5G core. This is different from 4G.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

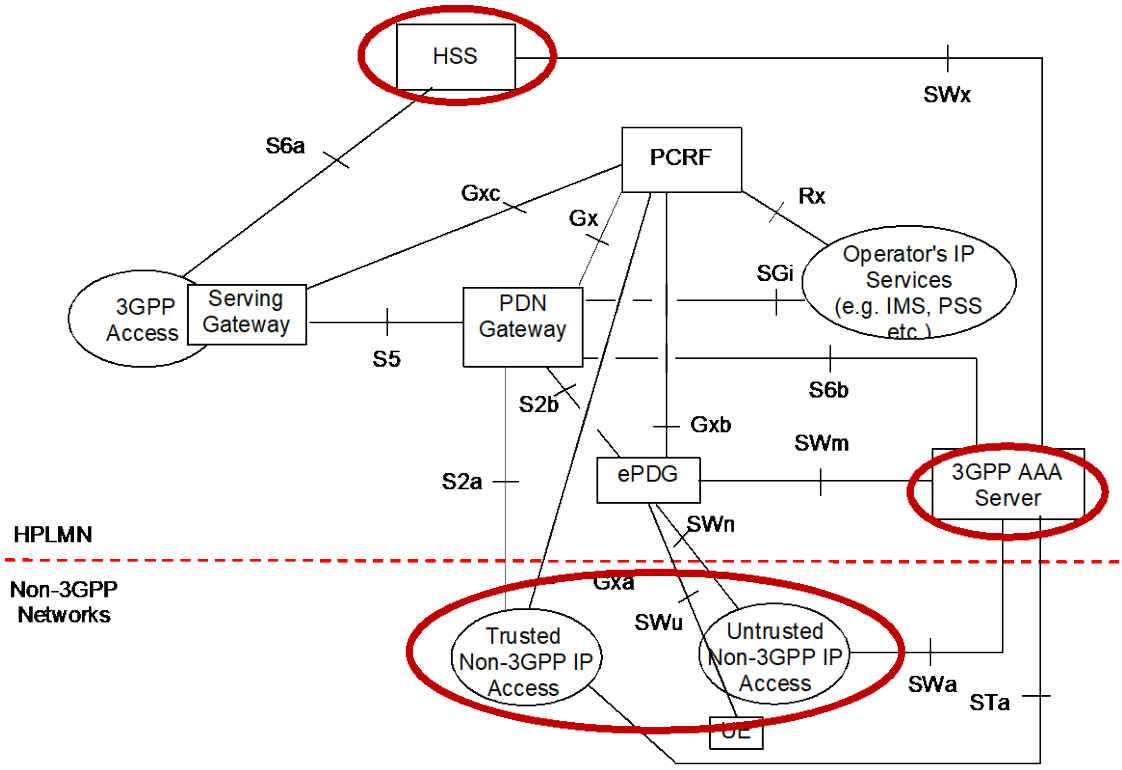


Figure 10 - 4G non-3GPP Architecture as described by 3GPP TS 23.402

If we look at 4G architectures, both trusted and untrusted non-3GPP access have interfaces to a 3GPP AAA server (STa and SWa, respectively) to facilitate authentication to the HSS. However, in 5G, the untrusted non-3GPP access contains no such interface. Only trusted non-3GPP access provides an authentication interface (Ta) between the Wi-Fi network and the 5G core.

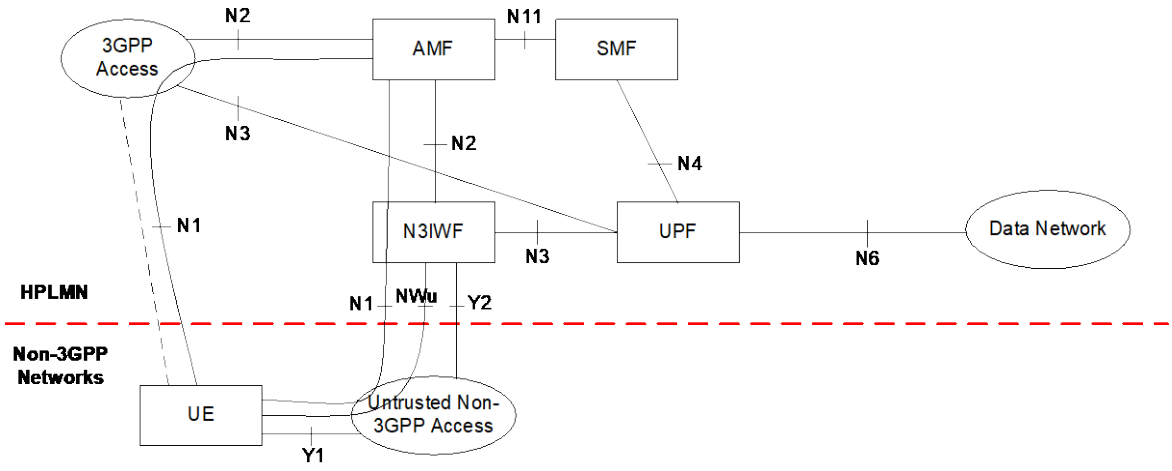


Figure 11 - Untrusted non-3GPP Access as described in 3GPP TS 23.501

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In release 17, 3GPP defined non-seamless WLAN offload (NSWO), which allows for authentication to the 5G core without automatically registering or establishing PDU sessions with the 5G core, very much like a 3GPP AAA did in 4G.

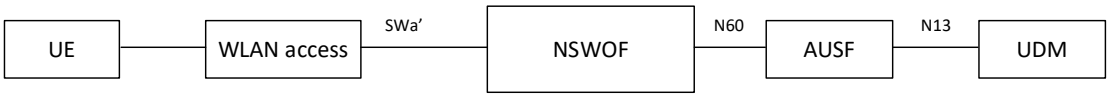


Figure 12 - NSWO as described in 3GPP TS 23.501

Theoretically, NSWO could be combined with untrusted non-3GPP access to have a behavior more closely related to the 4G experience. However, NSWO with untrusted Wi-Fi access has several drawbacks.

1. NSWO and untrusted non-3GPP access each require an authentication. This means there are two authentications: one to gain access to the Wi-Fi network and a second to allow for untrusted access. This is like the current 4G experience.
2. NSWO requires EAP-AKA' to be used for authentication. Most 5G networks are using 5G-AKA as the primary authentication method. This means the authentication method for the NSWO is potentially different from the authentication method for the untrusted access. Additionally, the authentication method is specified in the subscriber record in the UDR as an either/or selection. It is unclear currently how to configure the subscriber to authenticate using one method for NSWO and another method for cellular and untrusted access.
3. The requirements for NSWO to work over Wi-Fi are identical to the requirements for trusted Wi-Fi access. NSWO requires PLMN lists, and their ability to support NSWO must be specified as an option in the ANQP by the Wi-Fi AP, just like trusted Wi-Fi. The UE must be able to have logic to select the correct PLMN and construct a NSWO specific NAI for authentication, just like trusted Wi-Fi. The actual authentication exchange is EAP between the NSWO server and the UE, just as it is between the TNGF and the UE.
4. NSWO with untrusted access requires two new platforms to be deployed (NSWOF and N3IWF) to perform the same functionality of the TNGF.
5. NSWO is designed for legacy Wi-Fi offload behavior only. However, trusted Wi-Fi access can also be used for offload as policy rules, such as UE route selection policy (URSP), can be configured to route traffic locally instead of through a PDU session back to the 5G core.

Essentially, using NSWO and untrusted Wi-Fi access is more complicated, less efficient, and requires the same amount of effort to implement without providing any additional benefit when compared to trusted Wi-Fi access.

Untrusted access still has a place for connecting when the WLAN does not support Hotspot 2.0, or the trusted WLAN cannot be used in the user's current location. 3GPP standards for trusted access have untrusted access as the fallback option when trusted networks are not available. In these cases, untrusted networks may be used, requiring manual Wi-Fi selection and authentication, to connect back to the 5G core to leverage the ATSSS technology.

5. Putting 5G/Wi-Fi Convergence to Use

Trusted Wi-Fi access provides seamless authentication and connectivity of a mobile device to the 5G core over a compatible Hotspot 2.0 Wi-Fi system. ATSSS provides the ability to steer, switch, and split 5G

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

user traffic between the UE and 5G core over both NR and Wi-Fi in real time. When these two technologies are combined, the first stages of the converged experience may be realized.

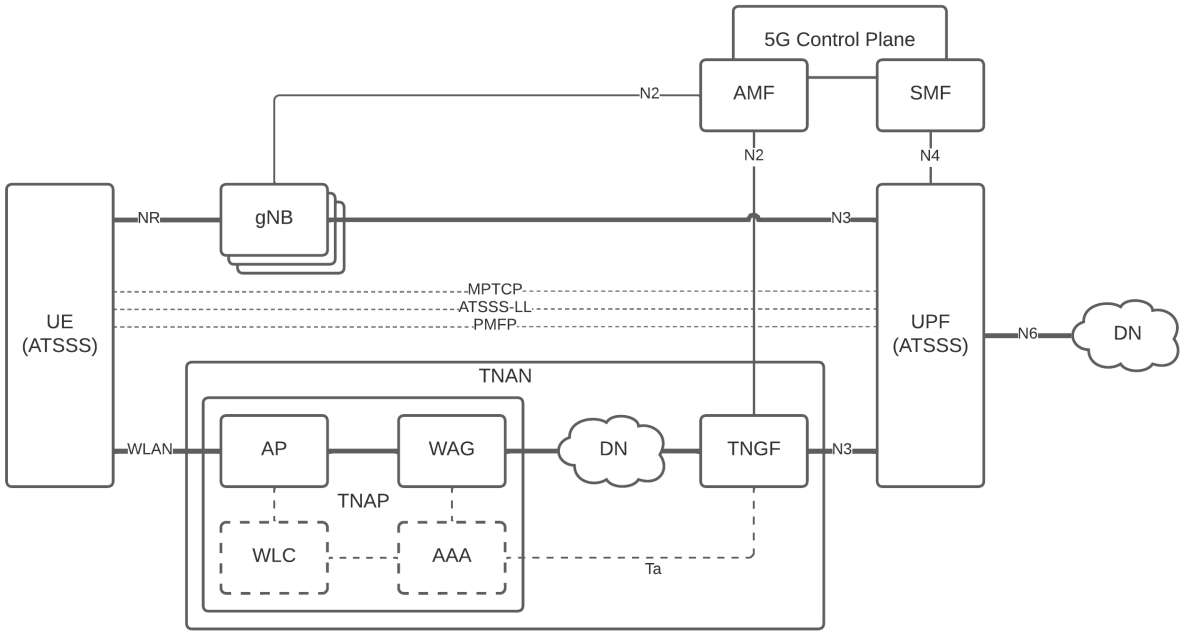


Figure 13 - ATSSS and Trusted Wi-Fi Combined Architecture

5.1. Converged Experience #1: Mobility Coverage and Network Resiliency

Since trusted Wi-Fi access treats access to Wi-Fi networks like a cellular RAN, a mobile operator who provides both 5G NR and Wi-Fi access can take advantage of the seamless authentication, registration, and PDU session establishment of trusted Wi-Fi access and the traffic switching capabilities of ATSSS to extend coverage to indoor areas or other obscure areas where cellular coverage doesn't reach. This creates a seamless experience where the user experiences no interruptions to services and maintains a consistent user policy across the two accesses, which is important when considering parental control policies.

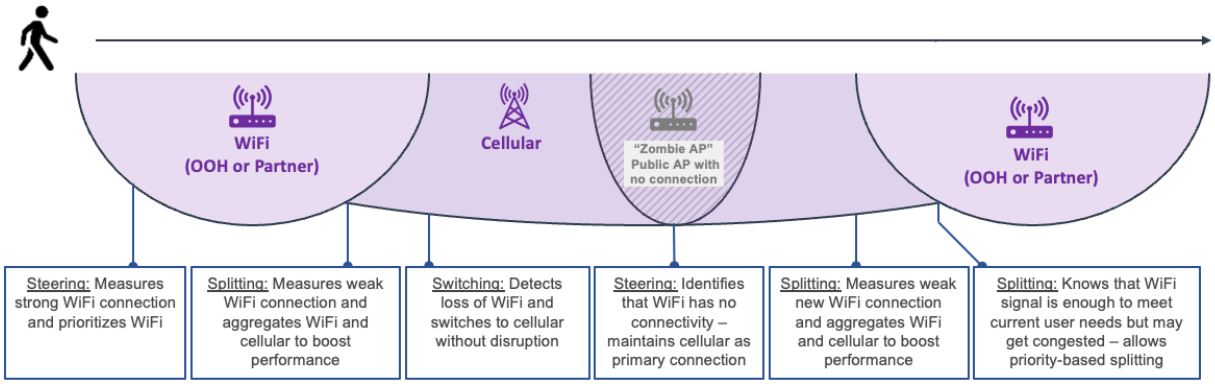


Figure 14 - 5G/Wi-Fi Convergence User Transitions

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Additionally, in areas where both cellular and Wi-Fi are available, the splitting capabilities of ATSSS can dynamically route traffic over the best possible access network. ATSSS features can improve the switching experience as the user transitions between access networks, and the steering functionality of ATSSS ensures users never suffer from a "zombie" Wi-Fi AP.

5.2. Converged Experience #2: Mixed Private Wireless

In private wireless, a mixed CBRS and Wi-Fi deployment on a university, hospital, or factory campus can be leveraged to serve a multitude of deployment specific use cases. For many of the same reasons a mobile operator would deploy 5G/Wi-Fi convergence, private wireless deployments would benefit greatly from such a converged architecture and are more likely to have much more overlap of unlicensed access networks. This can be especially powerful for users who use a dual-SIM, dual-standby (DSDS) device to access the private network. If the private wireless network is hosted or managed by a MNO, the MNO may design a deal with the private network, allowing MNO customers to connect over the private network assets and back to the MNO core.

5.3. Converged Experience #3: Improved Full MVNO Offload

Mobile virtual network operators (MVNOs) who provide large out of home Wi-Fi systems, such as a major cable operator, may be able to take advantage of 5G/Wi-Fi convergence to improve their service offering. This would require the MVNO to operate a full MVNO model where the cellular RAN is provided by the MNO partner, and the 5G core and all end services are provided by the MVNO.

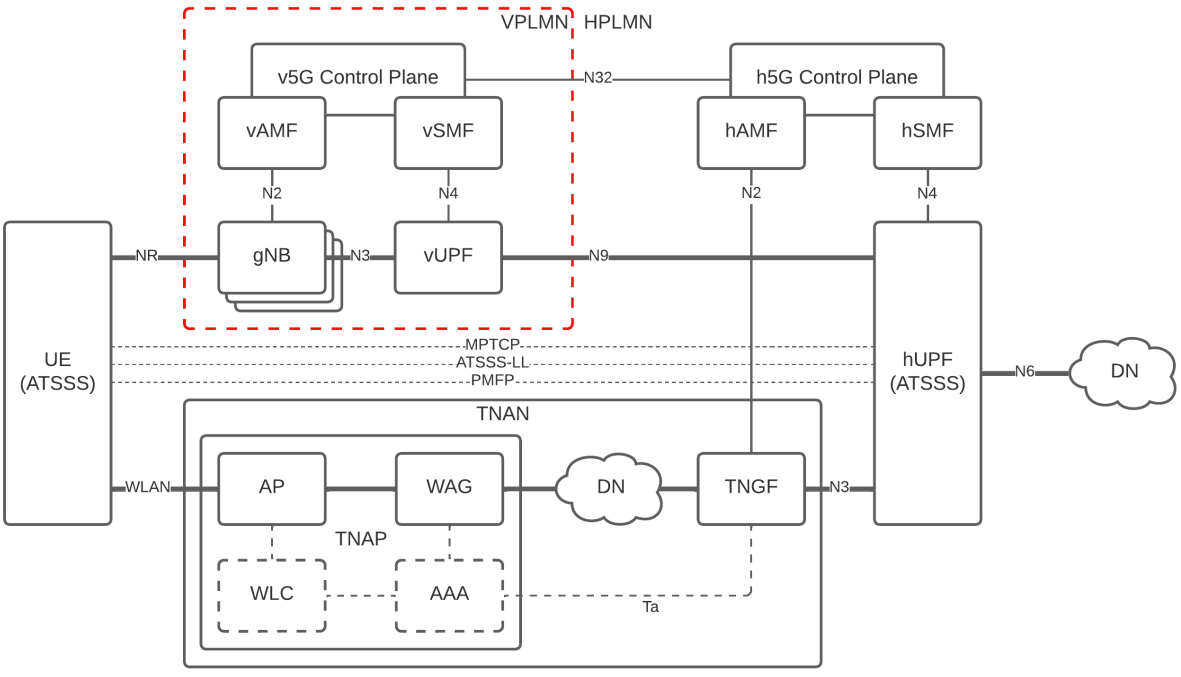


Figure 15 - Full MVNO ATSSS Architecture

Because the cellular traffic is routed to the MVNO's 5G core, the implementation of trusted Wi-Fi access connecting back to the common 5G core enables a seamless offload experience for the users. ATSSS features can improve the switching experience as the user transitions between networks, and the steering

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

functionality of ATSSS ensures users never suffer from a "zombie" Wi-Fi AP. Users are less likely to turn off Wi-Fi because they will never detect a bad connection, which will improve offload metrics as they pass into working Wi-Fi APs.

5.4. Converged Experience #4: Wi-Fi Mobile Operators

Because trusted Wi-Fi access treats the Wi-Fi network as a RAN to the 5G core, it is now possible to configure a fully secure mobile service that works over Wi-Fi. This can create new entrants to mobile network operators, especially if paired with CBRS. Not only could Wi-Fi only, or Wi-Fi preferred, MNOs begin to offer service directly, but their networks can be easily monetized to support new roaming relationships between traditional operators and these new Wi-Fi preferred operators.

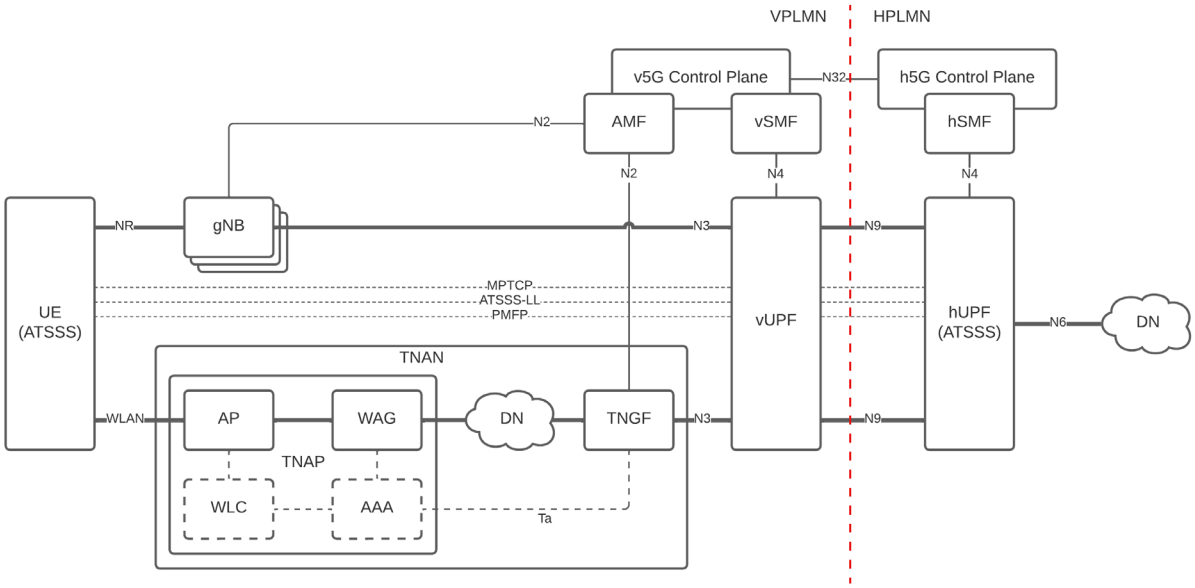


Figure 16 - Example Home Routed Roaming Architecture with Trusted Wi-Fi Access

6. Conclusion

5G networks provide a lot of opportunity for convergence, bringing together networks and services to enhance the user experience. The first opportunity to demonstrate the converged user experience in 5G is to combine cellular and Wi-Fi access networks to provide seamless, integrated connectivity of a mobile device to a common 5G core. 5G trusted Wi-Fi access connectivity provides seamless authentication, registration, and PDU session establishment, similar to cellular access, allowing the use of Wi-Fi networks as another RAN for the 5G core. When combined with ATSSS, a user's device can seamlessly move between the two access networks, load-share traffic, and avoid service disruptions. This converged network also opens new opportunities for private networks, MVNOs, and Wi-Fi network providers, changing the landscape of mobile network deployment options.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

3GPP	3rd Generation Partnership Project
4G	4th generation
5G	5th generation
5G-AKA	5G authentication and key agreement
AAA	authentication, authorization, accounting
AMF	access mobility management function
ANQP	access network query protocol
AP	access point
ATSSS	access traffic steering switching splitting
ATSSS-LL	ATSSS lower layer
AUSF	authentication server function
EAP	extensible authentication protocol
EAP-AKA’	extensible authentication protocol for authentication and key agreement prime
MA-PDU	multi-access packet data unit
MAR	multi-access rules
MNO	mobile network operator
MPQUIC	multi-path QUIC
MPTCP	multi-path TCP
MVNO	mobile virtual network operator
N3IWF	non-3GPP interworking function
NAS	non-access stratum
NSWO	non-seamless WLAN offload
NSWOF	non-seamless WLAN offload function
PCF	policy control function
PDU	packet data unit
PLMN ID	public land mobile network identifier
PMF	performance measurement function
PMFP	performance measurement function protocol
QoE	quality of experience
QUIC	quick UDP internet connections
RAN	radio access network
SCTE	Society of Cable Telecommunications Engineers
SMF	session management function
SSID	service set identifier
TCP	transmission control protocol
TNAN	trusted non-3GPP access network
TNAP	trusted non-3GPP access point
TNGF	trusted non-3GPP gateway function
UDP	user datagram protocol
UE	user equipment
UPF	user plane function
URSP	UE route selection policy
WAG	wireless access gateway
WLAN	wireless local area network
WLC	wireless LAN controller

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

3GPP TS 23.402: *Architecture enhancements for non-3GPP accesses*; 3rd Generation Partnership Project

3GPP TS 23.501: *System architecture for the 5G System (5GS)*; 3rd Generation Partnership Project

3GPP TS 23.502: *Procedures for the 5G System (5GS)*; 3rd Generation Partnership Project

3GPP TS 23.503: *Policy and charging control framework for the 5G System (5GS)*; 3rd Generation Partnership Project

3GPP TS 24.193: *Access Traffic Steering, Switching and Splitting (ATSSS)*, 3rd Generation Partnership Project

3GPP TS 24.502: *Access to the 3GPP 5G Core Network (5GCN) via Non-3GPP Access Networks (N3AN)*; 3rd Generation Partnership Project

3GPP TS 24.526: *User Equipment (UE) policies for 5G System (5GS)*; 3rd Generation Partnership Project

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

How to Cook a Duck: The Importance of M-Learning

An Operational Practice prepared for SCTE by

Fernando Durman
Technological Integration Manager
Telecom SA
Hornos 690, Ciudad de Buenos Aires. Argentina
54-11-62049804
fdurman@teco.com.ar

Damián Aguilar Cogan
Industries Outreach & Technical Training Management
Telecom SA
Hornos 690, Ciudad de Buenos Aires. Argentina
54-11-69546556
daguilar@teco.com.ar

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Context 4
2.1. The Arrival of Industry 4.0 4
2.1.1. Labor Market Impact 5
2.1.2. New Ways of Learning 5
2.2. Telecom in this market. 5
2.3. Customer Centrality 6
2.3.1. Importance of Customer Loyalty 6
2.3.2. NPS as a Loyalty Indicator 7
2.4. OKR Based Learning 7
3. Theoretical Framework 8
3.1. ICT Tools in the Educational Process 8
3.2. Mobile-Learning as part of Blended-Learning 9
3.3. The appearance of WhatsApp 9
3.4. Relationship between WhatsApp and M- learning 10
4. Study proposal 10
4.1. Justification 11
4.2. Hypothesis 11
4.3. Goals 11
4.4. Scope 11
4.5. Investigation methodology 12
4.6. Investigation 12
4.6.1. Demographic analysis of the sample 12
4.6.2. Survey on learning modalities 13
4.6.3. Launch of the m- learning initiative 15
4.6.4. Case 1: e-DOC - Substitution of Paper by Digital Systems 15
4.6.5. Case 2: Recovery of Power Supplies from CPEs 17
4.6.6. Case 3: Good FTTH Practices 18
5. Conclusion 20
6. Future steps 20

Abbreviations 21

Bibliography 22

List of Figures

Title **Page Number**

Figure 1: Percentages of staffing according to generation of belonging. 13
Figure 2: Percentage of type of learning device used by generation 13
Figure 3: Percentage of Learning modality used in the last year according to the response of field technicians 14
Figure 4: Learning modality recognized as valid by technicians 14
Figure 5: e-DOC. Communication board by WhatsApp 16
Figure 6: Impact of communication on the weekly evolution of adoption of the e-DOC App 16
Figure 7: Impact of communication on the monthly evolution of adoption of the e-DOC App 17
Figure 8: Recovery of power supplies. Communication board by WhatsApp 17

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 9: Impact of communication on the monthly evolution of the recovery of CPEs power supplies. ... 18
Figure 10: Good connectorization practices. Communication board by WhatsApp 19
Figure 11: Impact of communication on the bimonthly evolution of the failure rate of fiber connectors..... 19

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Mobile technology has transformed many aspects of our lives, including the way we learn and perform everyday tasks. An example of this statement occurs in daily life in the cooking process: in this case due to the need to prepare a duck. Through mobile devices, we have access to a wide range of educational resources that not only teach us how to cook but also improve our culinary skills and transform the kitchen experience.

Mobile learning allows us to access online cooking tutorials, detailed recipes, how-to videos, apps and interactive websites to become indispensable companions providing instant access to tips and tricks while preparing ingredients. They also offer commentary spaces for other amateur cooks, creating a collaborative learning environment.

Thus, from preparing a simple duck to creating gourmet dishes, the availability of educational resources through mobile devices has redefined the way we approach cooking and gives us the opportunity to be continuous learners in this culinary art.

Similarly, m-learning (mobile learning) has a significant impact on the work of a field technician, who are constantly on the move, performing maintenance, repair, and installation tasks at different locations.

Access to real-time information allows technicians to access manuals, repair guides, video tutorials and other resources directly on their mobile devices while on the job site, thus solving problems on the spot, avoiding delays and improving efficiency.

In complex or unfamiliar situations, technicians can search for solutions online, consult discussion forums, and access practical communities that can offer guidance and advice.

Instant communication is another very useful tool. Key to reinforce concepts in the deployment of new technology and services, provide the solution to unforeseen situations that require rapid and massive action, help as a reminder of new work processes, and all other information of a massive nature that requires rapid or planned dissemination

Lastly and very importantly, it keeps continuous training alive by offering updates with the latest technologies and procedures through online courses, webinars and specific training materials.

In short, mobile learning on the job offers a field technician an efficient and effective way to access information, improve skills, and communicate effectively, all while on the go to get the job done. This combination of technology and continuous learning can improve the quality of work and the satisfaction of both technicians and their customers.

In this paper we will cover all the actions carried on Telecom Argentina with this purpose, that include include the deployment of different tools and the creation of different contents/activities with the support of different frameworks and studies.

2. Context

2.1. The Arrival of Industry 4.0

The arrival of the fourth industrial revolution or Industry 4.0 brought technologies focused on intelligent environments, that is, processes and things with a certain autonomy. This trend is reflected in the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

automation, digitization and massive use of information from each component of that universe. IOT, Big Data, Cloud Computing, Cybersecurity are the new paradigms (Baygin, Yetis, Karakose, Akin , 2016).

Undoubtedly this impacts society, industries, organizations, business, culture and education. The new professions and workers need, in addition to incorporating new disciplines and practices, they must have a broader and more global vision, in a digital learning environment (Cortés, Landeta, Chacón, 2017).

The use of ICT (Information and Communication Technology) tools will be key in the management of information and communication. Billions of different digital devices access Internet-based networks, as key drivers in cost reduction and customer compliance (Joyanes, 2017).

2.1.1. Labor Market Impact

Digital transformation will demand a greater number of skilled workers. It is estimated that by 2030 this demand will considerably exceed supply, causing a scenario of global talent shortage to face the evolution of Industry 4.0 (Ferry, 2018).

Future organizations will need to attract and retain staff, while improving their skills and knowledge, within a framework of motivation and commitment.

This scenario requires companies to evolve taking advantage of technology and promoting the autonomy of people, generating collective knowledge through learning ecosystems that promote a culture of constant, social and collaborative development, as well as experiential.

New workers require experience in multidisciplinary teamwork and communication skills, not only with other humans but also with machines. They must be prepared to learn new things and with the conviction of being the best prepared for the changes that are coming (Kozák, Ruzicky, 2018)

2.1.2. New Ways of Learning

The new learning ecosystems must focus on people and not on disciplines, giving importance to autonomous learning and the practice of "learning to learn", highlighting the ubiquity of learning, at any time and place, according to the needs of the people, and with special importance of innovation and methodological variety, to make it more adaptable to the requirements of each individual and group.

"Knowledge" is valuable, and it must flow in the organization, which is why new learning modalities are necessary beyond 70-20-10, such as virtual learning, M- learning, collaboration or co-creation, adjusted to the new organizational environments marked by ubiquity (DESI, 2018)

The teaching model for these new skills must be adjusted to the learning styles of each one, their independence of time and location, individual needs and contents that represent permanent or lifelong learning. A problem-oriented model, in a way that is connected to a reality based on the resolution of a particular case.

A challenge that implies making mentality changes for both teaching and learning, (Kozák, Ruzicky, 2018)

2.2. Telecom in this market.

As far as the future is concerned, Telecom's focus is to continue consolidating itself as an ecosystem of digital solutions and platforms, leveraged on connectivity as its competitive advantage. The company's

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

compromise in expanding and boosting its fixed and mobile networks is unwavering, but with the goal of going beyond connectivity, accompanying the challenges of the digital economy and developing new business products and services related to IoT, fintech, cybersecurity, entertainment, and smart home, among others.

In this evolution from a telecommunications company to a Tech-Co with regional reach, the development of digital talent is key. The company's evolution comes with a cultural transformation among its collaborators, implementing agile and collaborative methodologies and developing cross-functional work practices focused on enhancing customer experience.

Telecom continues to drive the evolution of its industry and boosting the growth of the digital economy, generating value for its customers by incorporating the most innovative technologies in the market, such as the 5G standard, and expanding the regional footprint of its services (CNV, 2022).

2.3. Customer Centrality

The technological and digital revolution is modifying the ways of living, relating and working. This phenomenon is manifested not only in the speed at which the changes take place, but also in the global scope, depth and breadth of the transformations in the models of production and governance management of the global economic system. The telecommunications industry is at the center of this revolution, offering people the opportunity to easily connect, access information, and handle enormous amounts of data storage and processing (Camps, 2016).

Paradoxically, this digitization of economic activity and the appearance of new business models did not generate a significant increase in value for telecommunications companies. The appearance of new actors, called over the top (OTTs) and digital native companies got ahead and were able to identify and understand the new consumer demands and transform them into successful business models (Cullem, 2019).

Thus, telecommunications companies are forced to reinvent their relationship with the customer, migrating from the classic access-based management model to the customer-based model. To do this, they must strengthen the customer experience by incorporating service quality as a key element of differentiation, complementing the traditional SLA metrics (Theoretical vs Real Quality) with others that incorporate the measurement of perceived service as a variable for monitoring and managing the experience. of the client (Martinez Arroyo, 2018).

2.3.1. Importance of Customer Loyalty

The competitive context leads companies to multiple pricing, product and technology strategies, however, only those that focus on customer satisfaction achieve a sustainable differentiator to boost their profits and growth.

In turn, consumers have a variety of product, channel, and service options and are beginning to identify companies by the overall experience they receive rather than by one particular factor. This relationship is manifested through loyalty to the company. (Gasull, 2019).

This loyalty is a sales and growth enhancer, since those more loyal customers stay with the company longer and are willing to pay more for an experience that they value more. Additionally, they become promoters of the brand and promote the acquisition of new customers.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.3.2. NPS as a Loyalty Indicator

The NPS (Net Promoter Score) is an indicator created by Frederick F. Reichheld, Bain & Company and Satmetrix to measure the level of customer loyalty to a brand or company, through its probability of recommendation. After asking customers how likely they would recommend a certain brand or company, they are categorized as promoters, passives, or detractors.

For this, organizations conduct customer surveys in order to obtain the necessary information. The main question posed by Reichheld (2003), asks the client to say in a range of 0 to 10 how willing he is to recommend the organization. Customers who respond between 0 and 6 are detractors of the company (dissatisfied customers who can generate a series of criticisms that would affect the growth of the company), those who respond between 7 and 8 are Neutral customers (customers who are satisfied, but are vulnerable to the activities of the competition or opinions of acquaintances), and those who give an answer within 9 and 10 are promoter clients and are in charge of recommending the organization and allowing it to have greater growth.

To calculate the NPS, the percentage of detractors must be subtracted from the percentage of promoters of the sample, as shown below:

$$NPS = \%Promoters - \%Detractors$$

The Net Promoter Score has become a key indicator of customer loyalty and used as a powerful competitive benchmark.

It stands out from this indicator that it is easy to obtain, favoring its adoption, that it is an easy-to-interpret metric, that it offers the possibility of easily benchmarking with companies and sectors, simply by sharing the NPS metrics of each organization and that it allows the approach to consumers and relate them to the indicator (Balan, 2012).

If we apply this perspective to the telecommunications industry, which has been converging in recent years through intense merger and acquisition activity and which, many times, makes them divert from the focus on the client, by monitoring the Net Promoter Score, these Companies could manage change more effectively, keeping their most valuable customers and attracting new ones (Deloitte, 2018).

2.4. OKR Based Learning

Learning based on OKR (Objectives and Key Results) has several advantages and benefits that can positively impact personal growth, professional development and organizational success.

The OKRs help align an organization towards the strategic objectives it wants to achieve. OKR goals should be qualitative and even inspirational, limited in time, and actionable (Paine, 2011).

A training plan based on OKR will help us clearly define the objectives of a training and the results that indicate that these objectives have been met.

In this way, the effectiveness of the training will be increased, the efficiency of the application of the courses in the Organization will be improved and, fundamentally, the speed of cultural change will be increased.

This focus on measurable results will help focus learning on tangible, measurable results, allowing people to visualize their progress and success. In addition, measurement and feedback provide a framework for

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

measuring progress and results. This facilitates the identification of areas for improvement and constant feedback to adjust the learning approach based on the results obtained (Wodtke, 2021).

An example of this might be a Primary Objective of Raising the technical quality of the work of field technicians to provide exceptional service to customers.

The key results are to evaluate technical skills, approval of certifications, efficiency in solving problems, customer satisfaction.

3. Theoretical Framework

3.1. ICT Tools in the Educational Process

In teaching practice, the importance of integrating information and communication technologies (ICT) in teaching-learning processes is established, particularly the incorporation of mobile devices, given that both globally and nationally their use has increased among people due to its portability, immediacy, connectivity and adaptability. These technologies have generated a paradigm shift in education in general and in distance education because they increase the possibilities of interacting with group members, improve communication and blur the barrier that separates students and teachers (Cantillo Valero, Roura Redondo and Sánchez Palacín, 2012).

The speed of technological advances and their global reach give rise to adaptive and proactive responses in organizations and individuals, based on the innovation-acceptance-consolidation-obsolescence flows, whose knowledge management strategies are not alien to them and have an impact by modifying Teaching and Learning processes (Illanas & Llorens, 2011).

This implies reconsidering alternatives that go beyond educational challenges since the solutions are both the cause and origin of new changes in the concept of knowledge management (Schlemmer, Saccol & Garrido, 2007).

On the other hand, the influences that the irruption of networks and social media produce in the daily habits of users, causes a greater demand for personalization of learning. The demand for absolute connectivity with other peers is increasing, unlimited access to resources and sources of information with total flexibility in the way, place and time of access.

An increasingly natural and necessary coexistence of formal and informal learning flows.

Thus, the educational ecosystem must provide students in the digital age with a series of skills that prepare them for a life in continuous technological evolution, among which it would be worth highlighting the skills to listen, search for various information in technological ecosystems, interact with other peers. constantly manage their digital identity and/or actively produce content, among others (Schlemmer, Saccol & Garrido, 2007).

Technology allows the e-learning system to bring the teacher's classroom closer to the student for their personal interaction, in an environment of flexibility and autonomy but that maintains the focus on the follow-up, guidance and monitoring of the learning process (Rosini 2013).

However, the introduction of new IT developments in education does not always translate into innovation (Oliveira, PC de, Cunha, CJC de A., & Nakayama, MK 2016) because their use is not a guarantee of a new form of education. This shows that the new pedagogical projects based on e-learning must be

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

designed differently than the traditional face-to-face ones, while the former seeks to strengthen the student-student relationship.

3.2. Mobile-Learning as part of Blended-Learning

Internet access from mobile technology allows personalized and localized learning beyond formal spaces (Pereira, Fillol and Moura, 2019). This learning is known as mobile learning. Given the great expansion of mobile technology, the best tools for learning are available to a large part of the population so that they can carry them in their pockets.

The essence of m- learning is "accessing information and knowledge anytime, anywhere, from devices that learners often take with them everywhere and find friendly and personal". (Gikas and Grant, 2013).

For this, it is necessary to digitally empower the apprentice to achieve the relevant skills for their learning. Without a doubt, more and more companies are betting on online training courses and that require their workers to take them.

Smartphones and Tablets are two of the products that have experienced the greatest expansion in recent years. These portable mobile devices have Internet access and a relatively high autonomy of use/charge, with the possibility of downloading numerous applications and that have functionalities of various kinds to carry out any task.

Adult learning adjusts skills and knowledge to fit the future of work, but its design needs serious rethinking (World Bank, 2019). Indeed, e-learning with the 70-20-10 model and the mobile learning emerge as an effective possibility to become the training channel.

Companies consider e-learning as a budgetary and productivity solution to business needs and changes in technology, which encourages them to take online training measures compared to more traditional solutions, which includes training through social networks and mobile devices, the inclusion of learning analytics and learning in the cloud. Companies, like other civil society organizations, undertake education projects beyond the educational system, in what is called non-formal education. This training process allows training throughout life and it is in this space that training in technology is being made possible as a transforming tool (Guzmán-Flores and García-Redondo, 2016).

3.3. The appearance of WhatsApp

WhatsApp is a free application and social network that is currently owned by Facebook that allows users to send messages, make calls and video calls through the Internet connection on their mobile device. It is compatible with Android, iPhone, Mac, Windows and Windows Phone and has multiple features that have made it a highly demanded platform (Clement, 2019).

Access from multiple devices such as mobile phones, tablets and computers, highlights greater flexibility in communication, with an intuitive and easy-to-use interface, for different types of users close to or far from technology (Wali, Winters, Oliver, 2016). It is currently used in more than 180 countries by over two billion people and is widely used in Latin America.

In 2023, WhatsApp became the preferred social network for Argentine Internet users, with close to 93% of Internet users interacting on the platform. Instagram and Facebook followed as the most popular social networks in Argentina, both with a participation of more than 85% of users (Statista, 2023).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Its irruption has to do with the wide benefits that its use brought. Starting with real-time and instant communication that facilitates fast and efficient interaction of different messaging formats (multimedia, text, images, videos, voice recordings, and documents), providing a richer form of interaction. This is free of charge through an internet connection from different geographical locations.

The possibility of creating broadcast groups facilitates contact with multiple people at the same time in a context of security and privacy since end-to-end encryption means that messages are protected and only the parties involved can read them.

It is these virtues that install WhatsApp in educational environments, as a complement to communication processes, collaborative learning and facilitating the transfer of knowledge through ICT to learn at any time and in any place (Suárez Ramírez, 2018),

Therefore, it is not only about using a mobile phone as an end, but about a means that the student (usually) already has to enhance their learning and that allows them to acquire knowledge and communicate with their peers and the teacher. (Khaddage, Muller, Flintoff, 2016))

3.4. Relationship between WhatsApp and M- learning

The relationship between m- learning and WhatsApp lies in how WhatsApp can be used as a tool in the context of mobile learning.

The first point of relationship is the communication and collaboration that both educators and students have on WhatsApp by sharing announcements, reading material, presentations or links to online learning resources, homework, and even providing assistance through text messages, images, and videos. Interaction that can occur in real time or asynchronously, helping participation and fostering collaborative learning.

Many times, the app acts as a massive and instantaneous channel of information that meets this need. In other cases, as a reinforcement of new concepts or others that need to be updated or kept current.

A widely used functionality of the app is the one that transforms it into a reminder of deadlines for tasks, exams or events related to learning as an organizational tool.

At the same time, it is also becoming a social- learning tool, in which students can collaborate on projects and teamwork through WhatsApp, sharing ideas, resources and progress updates (Younus & Qureshi, 2019).

In short, the intersection of e-learning and m- learning is messaging-based learning.

It is important to mention that although WhatsApp can be a useful tool in mobile learning, there are also privacy and security considerations that need to be addressed, especially when it comes to communication between educators and students. In addition, the effectiveness of using WhatsApp in an educational setting will depend on factors such as group dynamics, the culture of use of the platform, and the preferences of the participants (Ally, Prieto-Blázquez, 2014).

4. Study proposal

In previous points, WhatsApp is highlighted as a key tool of m-learning within the framework of a Blended ecosystem learning.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

This argument is supported by the massive use of mobile devices and the messaging app, real-time communication between educators and students enriched by content in multiple formats, personalized and independent of time and place.

In a context of high dynamism and change, WhatsApp is presented as an agile communication channel and personalized information, easy to use and read, and crossing the barriers of time and distance for learning. WhatsApp as a key m-learning tool within a blended learning ecosystem will be explained through the demographic analysis of the field technical staff.

This work will seek to demonstrate through real use cases the improvement in operational and service indicators from mass communications to the field technical staff by WhatsApp.

To do so, we pose the following questions:

- Are operational and service indicators impacted by the use of WhatsApp as a communication tool during technical training of field employees?

4.1. Justification

The study is framed in a VUCA (Volatile, Uncertain, Complex and Ambiguous) context in which the need to adapt to the changing and challenging circumstances that characterize said environment appears.

Thus, the learning model through WhatsApp is offered as an effective strategy to address learning challenges in a VUCA environment, providing flexibility, access to information and effective communication. At the same time, it becomes a continuous training tool that adapts to different learning styles and facilitates quick decision-making.

4.2. Hypothesis

The use of WhatsApp as a training tool for the field technician improves operational technical indicators. Through reports in pre and post learning instances it is possible to demonstrate this correlation.

4.3. Goals

Demonstrate improvement in operational and service indicators through training initiatives via WhatsApp.

Formulate a rapid deployment model for the real-time management of these metrics that allow configuring a learning ecosystem for continuous improvement.

Reflect on the implications and feasibility of applying this work in other organizations and what conditions should exist in them for its implementation.

4.4. Scope

The study reaches the field technical staff within the framework of the learning ecosystem that trains them with the necessary knowledge to carry out their task.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.5. Investigation methodology

A mixed qualitative -quantitative approach will be applied. Quantitative because it will be oriented to the identification of the communications that had an impact on certain operating parameters. Qualitative because it will use data collection and analysis to answer the research questions without following a rigid and sequential process within a context in which the metrics analyzed are affected by other variables not included in this research.

The research method will be Explanatory since it will be oriented to give an answer on how the training cause impacts on technical indicators, explaining how and under what conditions it occurs, and Correlational because in the study of these cause-effect relationships it will not maintain rigorous control of all variables.

It is non-experimental because it is a study that will be carried out without the deliberate manipulation of variables and in which the phenomena are only observed in their natural environment and later analyzed. Finally, it will be transversal since its purpose is to collect data at a single moment, whose purpose will be to describe variables and analyze their incidence and interrelationship at a given moment.

4.6. Investigation

4.6.1. Demographic analysis of the sample

The relationship between generations and the use of technology is a complex topic that reflects how different groups of people interact, adopt and adapt to technological tools based on their experiences and historical contexts.

The study mapped the field technical staff according to the well-known classification in Baby Boomers, Generation X, Millennials and Generation Z, validating that there are different approaches and diverse attitudes towards online learning (e-learning) due to their experiences and levels. of familiarity with technology (Chitiyo, Nyemba, 2023)

Baby Boomer Generation (born between 1946 and 1964): grew up in a pre-digital age, however, those who use technology adapt to e-learning if they receive the right support, more structured formats and more traditional learning approaches, such as webinars or online courses with a clear structure.

Generation X (born between 1965 and 1980): They are more familiar with technology than baby boomers and are prone to adopting the digital model of learning. They appreciate the flexibility of online learning and interactive formats.

Generation Y or Millennials (born between 1981 and 1996): they are digital natives and have a high affinity for e-learning. They value accessibility, interactivity and personalization in their learning experiences and tend to prefer more visual and multimedia formats.

Generation Z (born after 1997): They are very adept at digital formats, accustomed to technology from an early age. It is natural for them to learn through digital platforms, online videos, educational games and multimedia resources, they have a greater disposition for self-education and autonomous exploration online.

The graph shows that more than 80% is generation Y or Z, which allow digital learning initiatives naturally, 13% of generation X will accompany the wave of digital proposals and only the remaining 5% should be supported.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

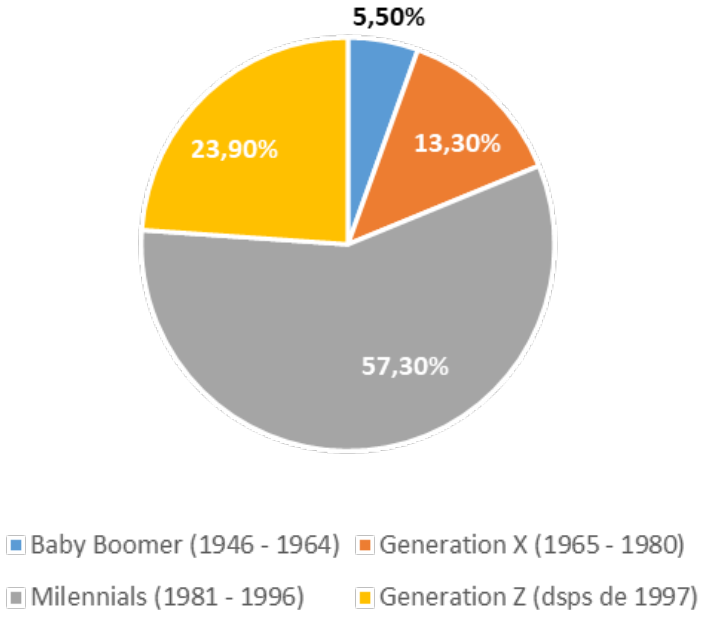


Figure 1: Percentages of staffing according to generation of belonging.

4.6.2. Survey on learning modalities

The study surveyed the technical staff (a sample of 857 responses) about the device used to consume content. The graph shows the use of the smartphone as the most widely used device, exceeding 80% in all generations, with the exception of Baby Boomers who incorporate laptops or desktop PCs to a greater extent.

The high usage of smartphones requires instructional designers to design content in responsive mode and viewable on smartphone screens.

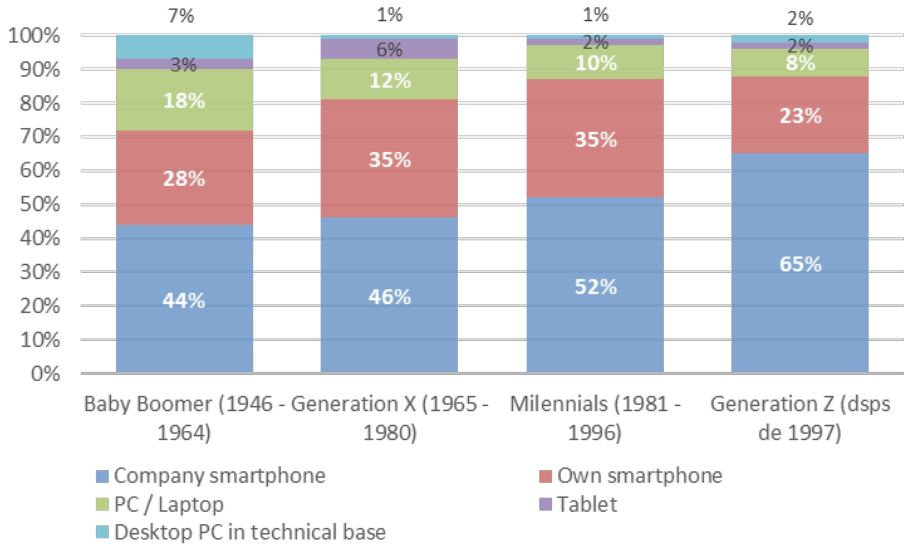


Figure 2: Percentage of type of learning device used by generation

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The expected data that the younger generations have greater adoption of cell phone use is confirmed.

The other important data surveyed is on which platform they received training in the last 12 months. We found significant adoption of virtual platforms (knowledge base, e-learning) reaching 75%.

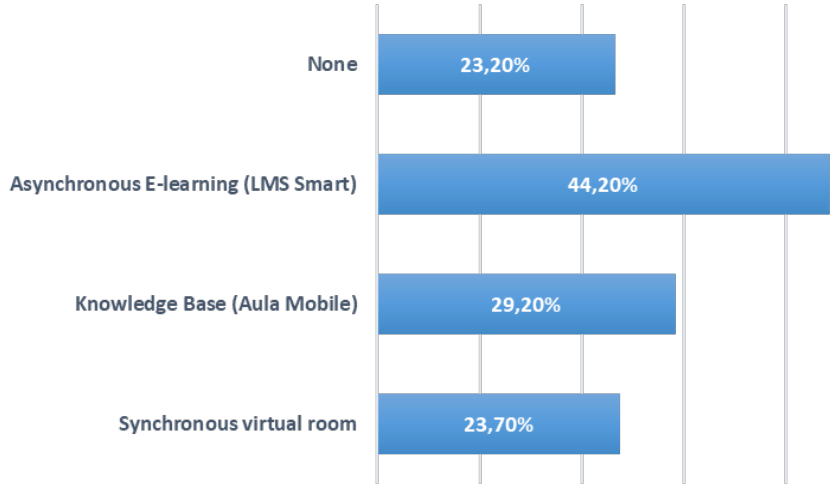


Figure 3: Percentage of Learning modality used in the last year according to the response of field technicians

The striking number that indicates 23% without training when crossed with the installed base of personnel that received training, an inconsistency appears because all the personnel received at least one training. This is because learning on virtual platforms is not perceived as formal training.

This is confirmed when the sample is asked about the valid means of training (choosing one invalidates the rest, placing them only as informative).

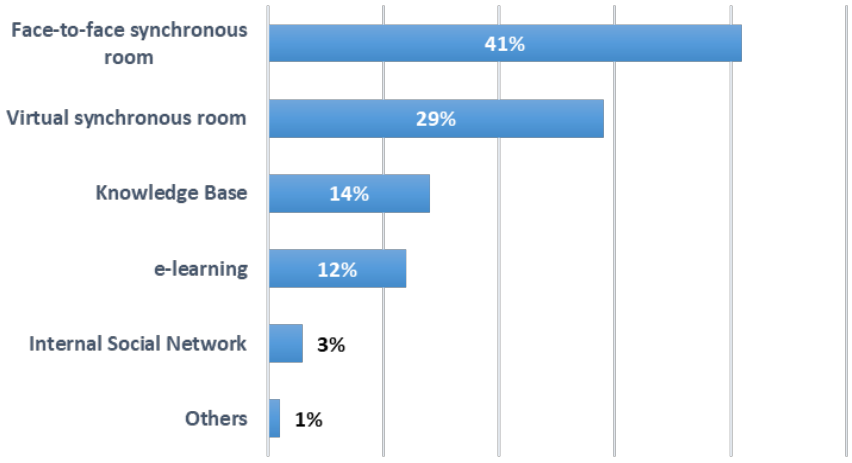


Figure 4: Learning modality recognized as valid by technicians

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The endowment finds as valid the synchronous format with an instructor, whether face-to-face or virtual. For this reason, when answering about training received in the last year, it does not recognize those that were virtual as such.

4.6.3. Launch of the m- learning initiative

The generational mapping of the staff shows a high affinity with the use of technology as a means of learning. When the surveys regarding the true adoption of digital learning tools are added, it is concluded that fertile ground is available to launch the m-learning initiative through the WhatsApp messaging tool on the smartphones of field technicians. The theoretical framework developed highlights the good qualities of the model, ideal to face the context described at the beginning.

Thus, the challenge is to develop a format that is perceived by the field technician as a quality training modality within the learning ecosystem.

However, this path began to be followed several years ago on an experimental basis and this study confirms it theoretically and practically as a virtuous path.

Since 2012, Telecom has been using instant messaging groups as a means of information and informal training with its staff. The demand of the operation is clear: to be close to the field technicians with daily learning actions through a fast and easy-to-use channel during the execution of service tasks.

The first stage of deployment focused on the staff of 300 supervisors who at that time orchestrated the more than +3000 regional operations technicians with whom. The same WhatsApp mass communication tool was used, which is still used today. Technical supervisors are a fundamental link in the training of the technical staff, since they not only promote the consumption of training, but they are also the ones who must allocate the necessary time for the training of their teams.

In parallel, several focus groups were carried out with the field technicians who concluded that the best way to deliver knowledge was to achieve a direct link. As a result, it was decided to extend the scope of this messaging via WhatsApp to the more than 6,000 technicians including the operations staff, both residential, external plant and internal network, for HFC, FTTH, Mobile and Environment technologies.

The communications generate a space for consultation and learning supported by a Smart LMS (Learning Management System) on SAP and on an Aula Mobile knowledge base on Oracle. Given that the interactions take place during the field technicians' workday itself, the communication must be clear and graphically attractive and synthetic in order to capture the attention of the recipients.

4.6.4. Case 1: e-DOC - Substitution of Paper by Digital Systems

This communication accompanied the initiative to save paper through an e-DOC digital documentation app. In this way, the work orders in which the client confirmed the work carried out by means of a paper signature, the replacement of the loan book of the installed customer premises equipment (CPE) and the book of equipment withdrawal remittances, are replaced.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

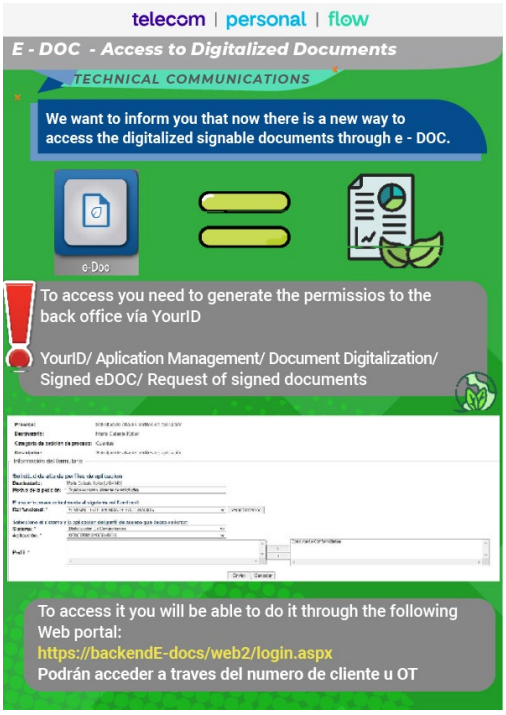


Figure 5: e-DOC. Communication board by WhatsApp

The communication highlights the benefits of the initiative and offers a video tutorial for learning how to use it through a link to the knowledge base.

To measure the impact, we take the connection of unique users to the app. During the first two weeks of April 2023, we observed a group of technicians connecting as MVPs as their first deployment. In week 3 and week 4 of April, the communications of the figure were sent with the consequent increase in usability.

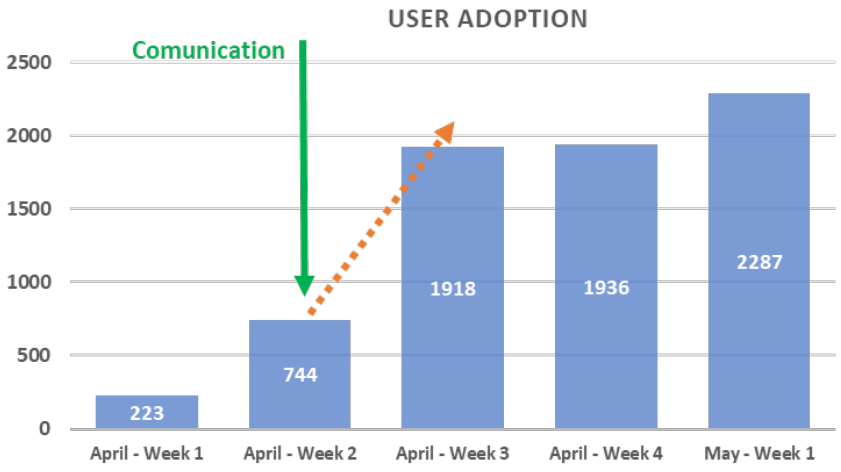


Figure 6: Impact of communication on the weekly evolution of adoption of the e-DOC App

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

On June -23 a reinforcement was carried out that operated as a final push for the total adoption of e-DOC.

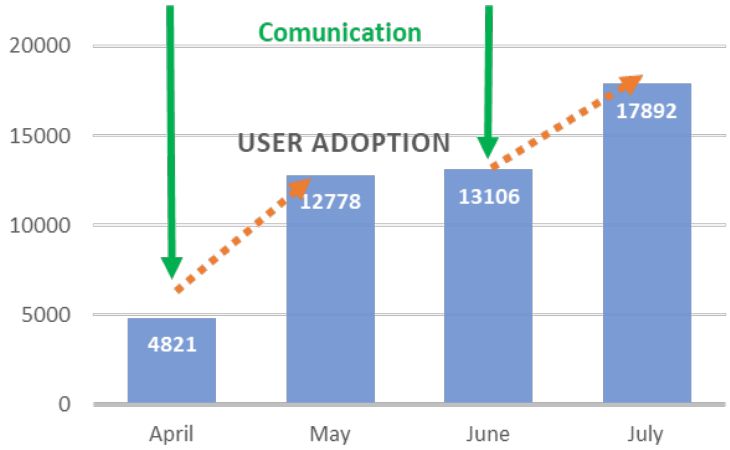


Figure 7: Impact of communication on the monthly evolution of adoption of the e-DOC App

4.6.5. Case 2: Recovery of Power Supplies from CPEs

This example is the result of joint work with the Legal and Supply area. An investigation found a large resale business for CPE power supplies. To reinforce logistics controls and raise awareness about the importance of caring for the company's assets, recovery control boards and training communications and information about the process were defined.



Figure 8: Recovery of power supplies. Communication board by WhatsApp

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

All field technicians received a mass-notification via WhatsApp to reinforce the importance of recovering power supplies (see Figure 8).

The metric used to measure the impact of the initiative is the number of power supplies entering the Repair Depot.

The first communication of November-22 had an immediate effect on the increase of recovered sources. In the fifth and sixth month after the start of the initiative, a relaxation in the process was noted, reinforcing itself in May-23 with a new communication that clearly shows an improvement in the following month (see Figure 9).

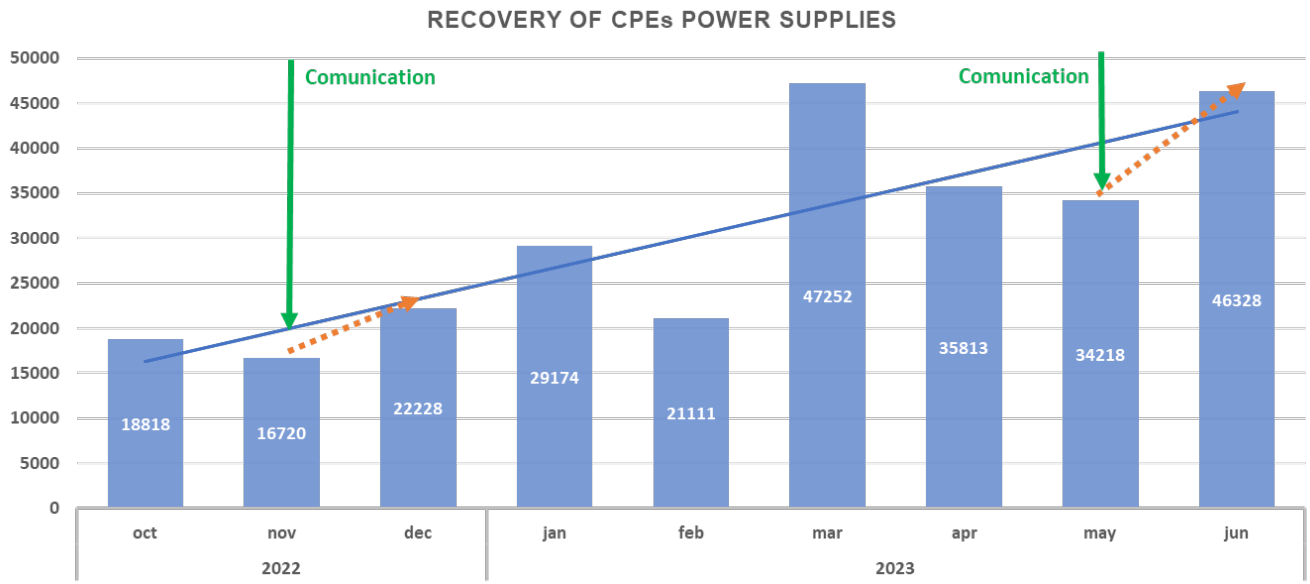


Figure 9: Impact of communication on the monthly evolution of the recovery of CPEs power supplies.

4.6.6. Case 3: Good FTTH Practices

Today, new deployments of access networks focus on technology. For years, the company evolved on HFC networks, so the fiber network is presented as a new technology that requires a learning curve and adoption of the best practices of the field technical staff.

In this context, permanent training and continuous learning are two basic needs to contain.

An analysis of the issues that most affect the service finds it to be one of the biggest failure points in the connectorization. Sensitivity to cleanliness and neatness and respect for fiber haircut measures are the focus of attention and training.

Another not minor point is the complexity and care in the use of tools and handling of materials, especially when we have various types of cables and brands of connectors, unlike coax, which after many years was able to work with a universal connector. axial compression.

The initiative in this case is a communication on the care of the connectorized, with a link to training documents stored in the knowledge base and a mailbox for consultation.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

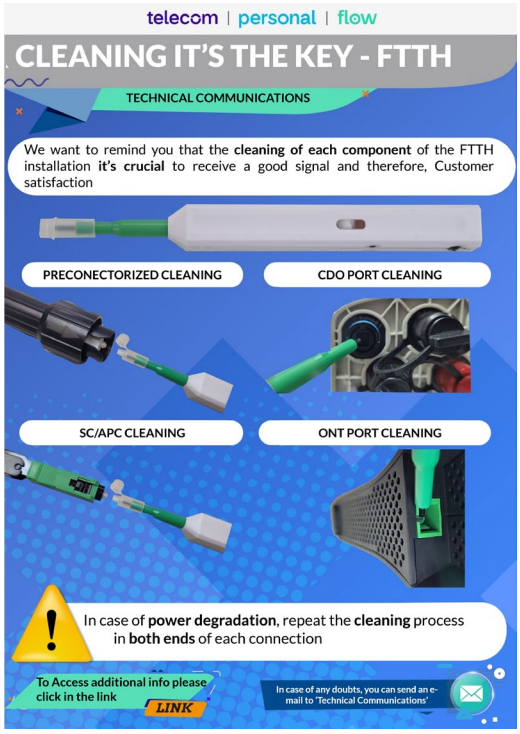


Figure 10: Good connectorization practices. Communication board by WhatsApp

The communication impact metric is the proportion of the number of connector failures within the universe of affected failures. Since the inspected samples can have a lot of variation from month to month, the analysis is carried out every two months.

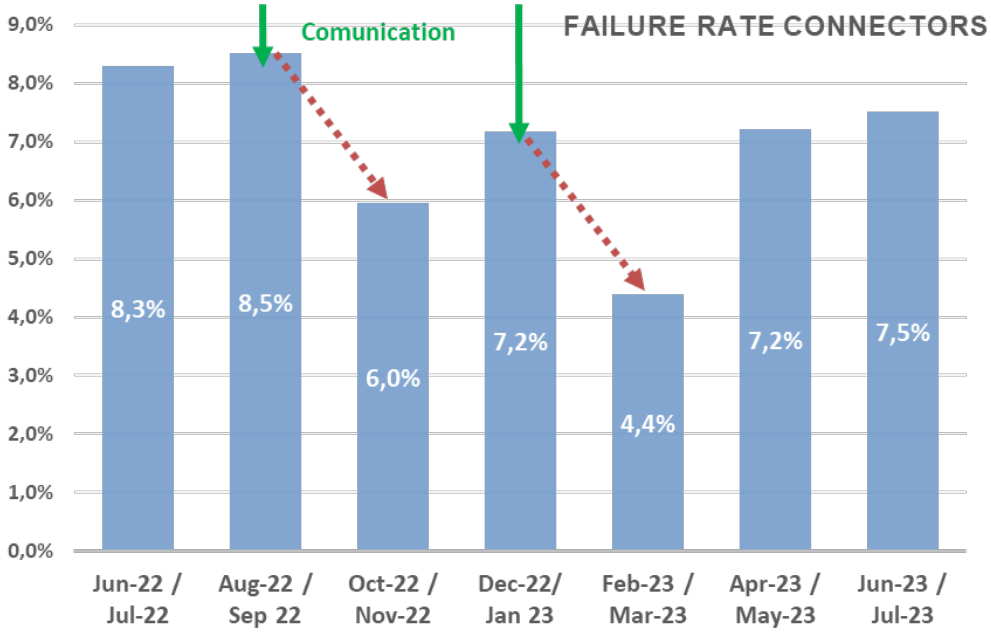


Figure 11: Impact of communication on the bimonthly evolution of the failure rate of fiber connectors.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 11 clearly shows the benefits of these communications through the rapid improvement in connector failure rate.

However, an interesting phenomenon is also displayed. As of each passing two-month period, there is a degradation of this indicator, which shows the need to be in permanent contact with the field coaching staff to provide continuous training support.

This may be due to various factors such as third-party personnel turnover, Ebbinghaus[1] forgetfulness curve, lack of application in the task of connectorizing by rotation in the performance of other tasks, etc.

5. Conclusion

This study has explored from different angles the impact of m- learning through the Whatsapp platform on the improvement of company indicators. The results obtained offer a solid and well-argued look at the positive impact on these indicators.

The accessibility of educational content for field technicians, by allowing them to conveniently learn anytime, anywhere while on the job, translates into a better understanding of technical best practices, new operational processes, and adoption of digital tools. On the other hand, the universality of the use of WhatsApp offers a flexibility of formats and contents that can face not only aspects of technical learning, but also address and strengthen cultural principles, stimulate digital transformation, disseminate business information and establish sudden communications in situations short-term.

Effectively trained employees are better prepared to address new challenges and take advantage of opportunities, which is ultimately reflected in improved business indicators.

On the other hand, although it was not verified by this work, m- learning through WhatsApp can be implemented with relatively low costs compared to other traditional training modes. This economic advantage is particularly significant for companies looking to optimize their resources and maximize the return on investment in training and personnel development. This effectiveness can be measured in an increase in labor productivity and efficiency.

In summary, as anticipated by the theoretical framework, the verification of the research hypotheses supports the idea that m- learning through WhatsApp can have a positive impact on business indicators.

Flexibility and economic efficiency are key aspects of this statement.

6. Future steps

The challenge is not to confuse m- learning as an end but as part of a blended learning that seeks to enhance learning in a learning ecosystem.

It is important to take advantage of and evolve the WhatsApp space as a source of constant feedback from technicians on the effectiveness of the blended program. learning and use this information as part of a plan for continuous improvement in content and structure. That is, to collaborate in a b- learning design with face-to-face, online (e-learning) and mobile (m- learning) learning modules.

[1] "Research consistently demonstrates that memory degrades over time (Ebbinghaus 1895) and this may be one of the factors impacting the degradation here."

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Developing this WhatsApp universe would allow technicians to participate in discussions and get involved with learning, share best practices, foster collaboration, and thus begin to promote social-learning.

Although the study shows progress in the monitoring of business indicators, it is essential to also design tools and dashboards to monitor the learning progress of technicians. In this way, establish certifications or recognitions for field technicians as a motivation tool and also to validate the quality of their learning.

Blended strategy learning that includes m- learning through WhatsApp will be highly effective in training field technicians. The combination of learning modalities will provide a more complete and adaptable experience, allowing technicians to learn effectively and apply their knowledge in real situations.

Abbreviations

APP	Application
CDO	Optical Distribution Box
CPE	Customer Premises Equipment
FTTH	Fiber to the Home
HFC	Hybrid Fiber-Coax
ICT	Information and Communication Technology
IOT	Internet of Things
KPI	Key Performance Indicator
LMS	Learning Management System
MVP	Minimum Viable Product
NPS	Net Promoter Score
OKR	Objectives and Key Results
ONT	Optical Network Terminal
OTT	Over The Top
SC-APC	Standard Connector - Angled Physical Contact
VUCA	Volatility, Uncertainty, Complexity, Ambiguity

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography

Ally, M., & Prieto-Blázquez, J. (2014). *"What is the future of mobile learning in education?".* Education and Information Technologies

Bălan, C. (2012). *"Net promoter score: Key metric of customer loyalty*". Quality - Access to Success, Durban University of Technology

Baygin, M. Yetis, H. Karakose, M. Akin, E. (2016). *"An effect analysis of industry 4.0 to higher education"*. DOI:10.1109/ITHET.2016.7760744.

Camps, C. (2016). *"Telecommunications operators in the new digital era"*. Retrieved from Management Solutions website: https://www.managementsolutions.com/sites/default/files/publicaciones/esp/telecomunicaciones-era-digital.pdf

Cetinkaya, L. (2017). *"The Impact of Whatsapp Use on Success in Education Process". The International Review of Research in Open and Distributed Learning*, *18*(7). https://doi.org/10.19173/irrodl.v18i7.3279

Clement, J. (2019). *"Mobile App Usage—Statistics & Facts"*. Statista. https://www.statista.com/topics/1002/mobile-app-usage

Chitiyo, R. Nyemba, F. (2023). *"A Theoretical Perspective of Inequities in Online Learning/Education Based on Generational Differences"*. IGI Global

Deloitte. (2018). "*Deloitte NPS Technology Industry. Why customer loyalty matters?".* Retrieved from https://www2.deloitte.com/ar/es/industrias/technology-media-and-telecommunications.html.

European Commission (2018), *"The Digital Economy and Society Index (DESI)"*. Retrieved from https://digital-strategy.ec.europa.eu/en/library/digital-economy-and-society-index-2018-report

Gasull, E. (2019). *Customer 2020: Are you ready for the future or are you still living in the past?*. Accenture Strategy.

Gikas, J. Grant, M. (2013). *"Mobile Computing Devices in Higher Education: Student Perspectives on Learning with Cellphones, Smartphones & Social Media".* The Internet and Higher Education. 19.

Guadamuz-Villalobos, J. (2020). *"First steps of mobile learning in Costa Rica: use of WhatsApp as a means of communication in the classroom"*. Revista Electrónica Educare, vol. 24, núm. 2

Guzmán-Flores, T. y García-Redondo, O. (2016). *""Application of the mainstreaming methodology to design a virtual education model for government institutions that provide non-formal education"".* Fundación Dialnet, Universidad de Rioja.

Illanas, A. & Llorens, F. (2011). *"Web 2.0 challenges facing the eeeS".* Universidad y Desarrollo Social de la Web, Washington DC. Editandum

Instituto Nacional de Estadística y Censos (2023). *"Access and use of information and communication technologies"* https://www.indec.gob.ar/indec/web/Nivel3-Tema-4-26.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Joyanes, L (2017). *"Industry 4.0 - The fourth industrial revolution*". Alfaomega.

Khaddage, F. & Müller, W. & Flintoff, K. (2016). *"Advancing Mobile Learning in Formal And Informal Settings via Mobile App Technology: Where to From Here, and How?"*. Educational Technology & Society. 19.

Kozák, Š. Ružický, E. Štefanovič, J and. Schindler, F (2018). *"Research and education for industry 4.0: Present development"*. Cybernetics & Informatics (K&I), Lazy pod Makytou, Slovakia.

Korn Ferry Institute. (2018) *"Future of Work: The Global Talent Crunch"*. https://www.kornferry.com/insights/featured-topics/future-of-work

Kukulska-Hulme, A. Traxler, J. (2005). "*Mobile Learning: A Handbook for Educators and Trainers*". Routledge

Martinez Arroyo, J. (2018). *"Reinvent yourself: the great challenge of Customer Experience. Association for the development of customer experience"*. BCG

Oliveira, P. C. de, Cunha, C. J. C. de A., & Nakayama, M. K. (2016). *"Learning management systems (lms) and e-learning management: an integrative review and research agenda"*. JISTEM - Journal of Information Systems and Technology Management, 13(2),

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO]. 2013. "*UNESCO Guidelines for Mobile Learning Policies. Paris: UNESCO*". http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/ED/ICT/images/114_13_ED_UNESCO_Policy_Guidelines_for_Mobile_Learning_S.pdf

Paine, K. D. (2011). *"Measure What Matters (1st ed.)"*. Retrieved from https://www.perlego.com/book/1014890/measure-what-matters-online-tools-for-understanding-customers-social-media-engagement-and-key-relationships-pdf

Pereira, S. Fillol, J. Moura, P. (2019). *"Young people learning from digital media outside of school: The informal meets the formal"*. Comunicar, 58, https://doi.org/10.3916/C58-2019-04

Reichheld F. (2003). *"The One Number You Need to grow"*. Harvard Business Review, December 2003, 9.

Rosini, A.M. (2013). *"As novas tecnologias da informação e a educação a distancia"*. Cengage Learning, São Paulo.

Schlemmer, E. Saccol, A. Garrido, S. (2007). *"Um modelo sistêmico de avaliação de softwares para educação a distância como apoio à gestão de EaD"*. Revista de Gestão USP, 14(1).

Statista (2023). *"Percentage of users in selected social networks in Argentina in 2023"*. https://es.statista.com/estadisticas/1218938/argentina-porcentaje-de-usuarios-por-red-social

Suárez Ramírez, J. (2018). *"WhatsApp: Uses, emotional dimension, benefits and harms in the life of university students"*. V Jornadas Iberoamericanas de Innovación Educativa en el Ámbito de las TIC y las TAC (InnoEducaTIC 2018). http://hdl.handle.net/10553/52679

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Trejos Buriticá, O. I. (2018). *"WhatsApp as a support tool for the teaching and learning process of computer programming"*. Educación y Ciudad, (35), https://doi.org/10.36737/01230425.v0.n35.2018.1970

Valero, C. Roura Redondo, M. Sánchez Palacín, A. (2012). *"Current trends in the use of mobile devices in education"*. La educación. No. 147.

Wali, E. Winters, N. Oliver, K. (2016). *"WhatsApp: A tool for sharing, communication, and evaluation in libraries and research support services"*. Journal of the European Association for Health Information and Libraries, 12(1).

Web Inversores Telecom (2022) *"Integrated Annual Report Year 2022"* https://inversores.telecom.com.ar/es/reporte-integrado.html

Wodtke, C (2021). *"Radical focus : achieving your most important goals with objectives and key results"*. Cucina Media LLC.

World Bank (2019). *"World Development Report 2019: The Changing Nature of Work*". The World Bank Group. https://www.worldbank.org/en/publication/wdr2019

Ynzunza Cortés, C. Izar Landeta, J. Bocarando Chacón, J. (2017) *"The Industry 4.0 Environment: Implications and Future Perspectives"*. Conciencia Tecnológica, no. 54.

Younus, A. Qureshi, A. (2019). *"WhatsApp as an innovative teaching tool: A cross-sectional study"*. Anatomical Sciences Education, 12(2),

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Hybrid Cloud to Power Private 5G Networks

Industrial Worker Safety Use Case

A Technical Paper prepared for SCTE by

Mohamed Daoud
Director of Engineering, Emerging Technology
Charter Communications
6399 S. Fiddlers Green Circle, Greenwood Village, CO 80111
720-699-5077
Mohamed.Daoud@charter.com

Ammar Latif
Principal Solutions Architect- Telco
Amazon Web Services (AWS)
2121 7th Avenue, Seattle, WA 98121
732-983-2708
latammar@amazon.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Private 5G overview 4
3. AWS Edge Service For P5G Infrastructure 6
4. Charter's Private 5G Lab Setup 8
   4.1. AWS Outposts 9
   4.2. Charter Lab AWS Outpost Configuration 10
   4.3. Charter Private Wireless Lab Setup 15
5. Lab Private Wireless Use Case 17
   5.1. Worker Safety Use Case 18
   5.2. Use Case Computer Vision Model 19
      5.2.1. Annotating Images using AWS Mechanical Turk 20
      5.2.2. Training Machine Learning Model on EC2 21
   5.3. Private 5G Network and Uplink Capacity 23
      5.3.1. Camera Configuration 24
      5.3.2. Private 5G Uplink Capacity, Camera FPS, and Resolution 24
      5.3.3. The Video Streaming Choice for the Lab Use Case 25
   5.4. Private 5G Use case Latency 26
6. Conclusion 28

Abbreviations 29

Bibliography & References 29

List of Figures

Title **Page Number**

Figure 1: 3GPP NPN Deployment models 5
Figure 2: P5G Cloud Deployment Scenarios 7
Figure 3: AWS Edge Services 7
Figure 4: AWS Outpost Ordering and Tracking via the Console 9
Figure 5: Outpost Management via AWS Console 10
Figure 6: GPU-based Instances on Outpost 11
Figure 7: Outpost Original Instance Capacity before Slotting 12
Figure 8: Example of g4dn.metal Slotting Options 13
Figure 9: AWS Outpost in Charter Lab 14
Figure 10: AWS Outpost Inside View Showing Installed RUs 15
Figure 11: Charter's Lab Private Wireless on AWS Outpost Architecture 16
Figure 12: Charter Lab Use Case Architecture Deployed on AWS Outpost 19
Figure 13: AWS Mechanical Turk Labeling Task 21
Figure 14; AWS Mechanical Turk - Charter's Worker Safety Task 21
Figure 15: Training Loss Curve 22
Figure 16: Charter Developed Model Inference 23
Figure 18: Resolution vs. Average Bitrate 25
Figure 19: Uplink Bitrate for Multiple Cameras 26

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Tables

Title	Page Number
Table 1: Charter's AWS Outpost Compute Configuration	10
Table 2: Uplink Bit Rate vs. FPS and Resolution	24
Table 3: Uplink Bitrate for Multiple Cameras	26
Table 4: Latency Causing Delays in the Worker Safety Use Case	27

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Cloud adoption among enterprises continues its strong momentum and many enterprises depend on the scale of the public cloud for their digital transformation journey. These enterprises have adopted a development framework and workloads that utilize public cloud scale and an expansive set of services. The developers utilize as well as innovate cloud services such as databases, analytics, IoT, and artificial intelligence/machine learning (AI/ML) that are available to them with the rich set of cloud APIs. This affords the enterprises the ability to innovate faster without the undifferentiated heavy lifting on managing the underlying infrastructure.

Cloud edge computing allows workloads to benefit from such cloud innovation and provides a consistent experience to enterprise workloads that require on-prem deployment. These applications usually have low latency requirements for interaction between its components as well as handling massive amounts of data processing and storage.

For telco use cases, private 5G (P5G) is an interesting use case for cloud edge computing due to its distributed architecture with a disaggregated set of telco network functions and applications that need to run on-premises due to latency constraints (e.g., gNB and UPF), local data storage (e.g., UDM), or local data processing needs such as applications that require AI/ML capabilities.

In this paper, we provide an overview of P5G and relevant hybrid cloud architectures. We review Amazon Web Services (AWS) edge offerings relevant to the P5G space. We then dive into Charter labs undertaking of building an industrial worker safety application that utilizes a hybrid cloud deployment of P5G. The paper describes the architectural choices for such hybrid cloud deployment as well as the use of cloud AI/ML tools as part of the application development and delivery.

2. Private 5G overview

A P5G network is a cellular-based network that is designed and operated for a specific organization's private use, rather than being accessible to the public. These networks provide dedicated connectivity and services tailored to the organization's requirements. 3GPP defines private 5G networks as 5G non-public networks (NPN). These private networks are particularly beneficial for industries that require high reliability, low latency, and secure connectivity, such as manufacturing, logistics, healthcare, utilities, and smart cities. P5G can support various applications, including industrial automation, remote monitoring, augmented reality, robotics, and mission-critical communications. Here are some key characteristics and benefits of P5G technology:

1. Dedicated Network: Providing organizations with dedicated cellular network infrastructure. This results in greater control and customization over network resources and enhanced performance while benefiting from 5G technology.
2. Low Latency: Offering low-latency connectivity with dedicated deployments, which is critical for applications that require low latency real-time communication, such as industrial automation, autonomous vehicles, and remote surgery.
3. Enhanced Security: Offering increased security measures compared to public 5G networks. This allows organizations to implement their own security protocols, encryption methods and access controls, providing a higher level of data protection.
4. IoT Enablement: The ability to support a large number of Internet of Things (IoT) devices, which allows for the deployment of advanced applications and services that rely on machine-to-machine communications at scale.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Industry-Specific Use Cases: P5G technology can be particularly beneficial for industries such as manufacturing, logistics, healthcare, utilities and transportation, where reliable, low-latency and secure connectivity is crucial.

To address the need for P5G networks, 3GPP has introduced specific features and enhancements to support private networks. The key elements of a 3GPP private network include:

1. Flexible Spectrum Allocation: In P5G architecture, private network operators can either use licensed spectrum dedicated to them or utilize shared/unlicensed spectrum for their network deployment.

2. Network Slicing: Enables sharing a single physical network infrastructure into multiple virtual networks (slices), allowing each private network to have its own isolated and customized resources, performance characteristics and services.

3. Security and Isolation: Private networks have enhanced security measures to protect sensitive data and ensure isolation from public networks. This is done through utilizing encryption, authentication and access control mechanisms.

4. Quality of Service (QoS): As P5G networks can use partially or completely separate infrastructure, this architecture allows organizations to define and prioritize certain types of traffic or applications, ensuring the desired quality of service for their specific use cases.

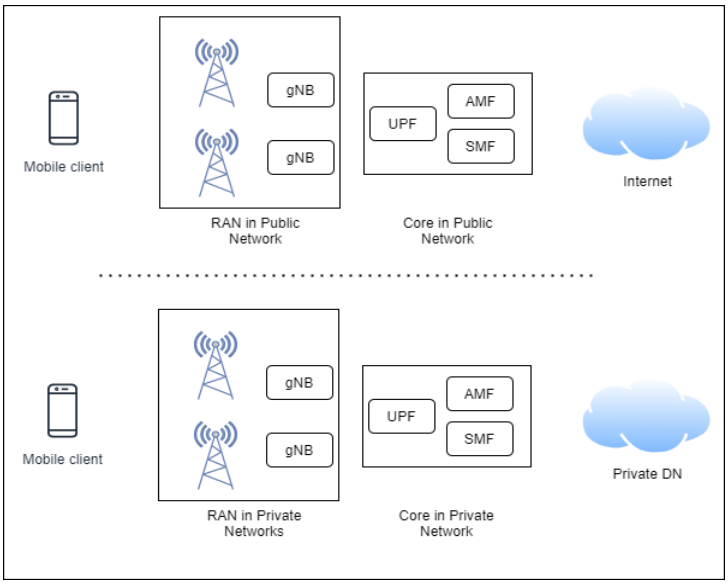


Figure 1: 3GPP NPN Deployment models

3GPP defines two main categories for deploying P5G network:

a) Public Network Integrated NPN (PNI-NPN) where the P5G network relies on some of the network functions that are provided by Public Land Mobile Networks (PLMN). This option is typically provided by PLMN with some of the equipment residing within the targeted enterprise facility.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

b) Standalone NPN (SNPS) which by its name is a standalone deployment that does not rely on network functions provided by PMLN. In this option, organizations typically work with network equipment providers, system integrators, and licensed spectrum holders. The solution requires deployment of its own base stations, antennas and network management systems, ensuring dedicated coverage within the targeted premises or desired coverage areas.

SNPNs enable the enterprise customer or the service provider to retain full control of the P5G network. On the other hand, PNI-NPNs can help with introducing P5G networks to enterprises with lower barriers to entry due to reduced initial investment and streamlined ongoing operations that are provided by PLMN. It's important to note that while 3GPP standards provide guidance on architectural options for building private networks, the specific implementation and deployment details may vary depending on the operator, organization and local regulatory requirements.

3. AWS Edge Service For P5G Infrastructure

P5G network requirements vary by industry and use case and thus the architecture for edge placements of network functions vary. For example, use cases that have higher data security or data residency requirements normally would build exclusive UPF to ensure the data flow never gets out of the local network within the targeted zone. Another consideration is latency where industries with lower latency requirements or higher reliability requirements prefer to put user plane function (UPF) in addition to access and mobility management function (AMF) and session management function (SMF) in the targeted zone. On the other hand, deploying NFs in AWS Region offers more versatility of services available to the deployment that are fully managed by the cloud provider.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

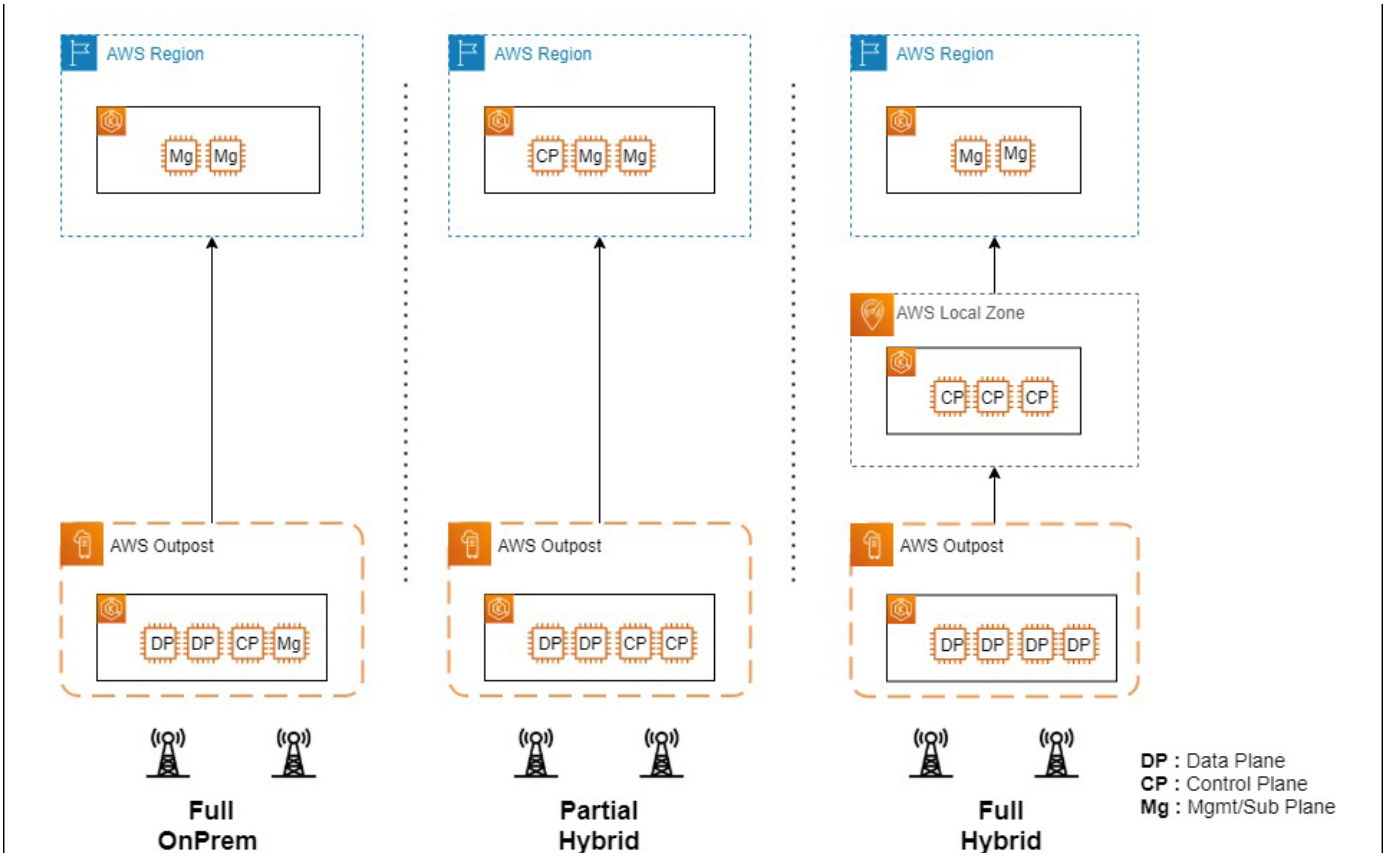


Figure 2: P5G Cloud Deployment Scenarios

AWS offers several edge services that can be utilized by telecom companies to enhance their services and improve the user experience. Here are a few AWS edge services relevant to the telecom industry:

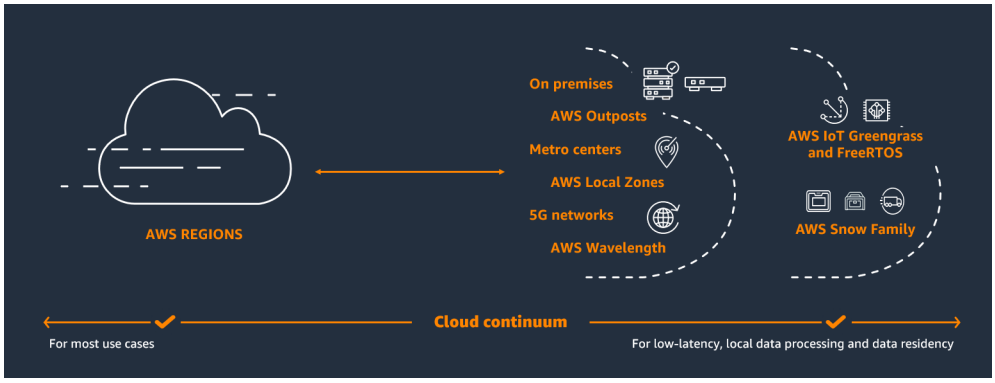


Figure 3: AWS Edge Services

1. AWS Outposts Family: AWS Outposts is a fully managed service by AWS that extends AWS infrastructure, services, APIs and tools to customer on-premises data centers or co-location facilities. It enables organizations to run AWS services locally, thereby providing a consistent hybrid experience with a unified management console. With AWS Outposts, customers can enjoy the same APIs, control plane, and management tools as in the AWS cloud while keeping their data and applications on-premises. This is particularly useful for organizations that have specific

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

data residency requirements, low-latency application requirements, or legacy applications that cannot easily be migrated to the cloud.

AWS Outposts brings AWS infrastructure and services to on-premises data centers or edge locations. This can be beneficial for telecom companies that require low-latency access to AWS services for specific applications or data processing tasks at the edge.

2. AWS Local Zones: AWS Local Zones are an extension of AWS infrastructure designed to bring AWS services closer to end-users and data centers in specific geographic areas. AWS Local Zones provide low-latency access to select AWS services for applications and workloads that require single-digit millisecond latency. AWS Local Zones are located in metropolitan areas separate from AWS Regions and are connected to the parent AWS Region via a high-bandwidth, low-latency network link. They are typically established to address specific needs, such as providing services to customers in a particular city or region with high local demand or latency-sensitive applications.

   By leveraging AWS Local Zones, telecom companies can deploy applications closer to end-users and process data in a local environment, reducing the latency between the application and its users. This can be beneficial for applications that require real-time responsiveness, such as gaming, media streaming, and financial services.

3. AWS Wavelength: Wavelength brings AWS compute and storage services to the edge of telecom networks, specifically at the edge of 5G networks. It allows developers to build applications that can run directly on Wavelength zones, reducing latency and enabling ultra-low latency use cases like real-time gaming, machine learning inference, and interactive video streaming.

4. AWS Snow Family: The AWS Snow Family helps customers that need to run operations in austere, non-data center environments, and in locations where there's lack of consistent network connectivity. The Snow Family, comprising AWS Snowcone and AWS Snowball, offers a number of physical devices and capacity points, most with built-in computing capabilities. These services enable customers to access the storage and compute power of the AWS Cloud locally and cost effectively in places where connecting to the internet might not be an option.

These services are designed to help telecom companies leverage the power of edge computing, improve performance, reduce latency and enhance the overall user experience for their customers.

4. Charter's Private 5G Lab Setup

In Charter Communications' lab, the team explored diverse P5G network architectures tailored to suit various enterprise use cases. Among the architectures investigated, one notable implementation involved constructing a P5G network for remote location scenarios utilizing AWS Snowcone. This setup incorporated a 5G standalone (SA) core network alongside 5G small cells, offering a robust and versatile solution.

Additionally, the team delved into a hybrid approach that employed AWS Snowcone to host the UPF. In contrast, the SMF, AMF, and Unified Data Management (UDM) components were hosted in the AWS region. Although this hybrid scenario presented intriguing possibilities, the focal point of this paper revolves around Charter's lab architecture, centered explicitly on a P5G network built using the AWS outpost.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

By focusing on the AWS Outposts-based architecture, the team aimed to demonstrate this particular setup's unique capabilities and advantages. Through careful design and implementation, Charter's lab succeeded in creating a P5G network that leveraged AWS Outpost infrastructure, showcasing the potential for efficient and reliable connectivity within enterprise environments.

4.1. AWS Outposts

AWS Outposts represents a comprehensive suite of fully-managed solutions designed to seamlessly bring AWS infrastructure and services to on-premises or edge locations, ensuring a unified hybrid experience. These innovative AWS Outposts solutions empower users to extend and operate native AWS services directly on their premises, offering unrivaled flexibility and control. The range of AWS Outposts form factors caters to diverse requirements, spanning from compact 1U and 2U servers to expansive 42U racks and multi-rack deployments.

With the integration of AWS Outposts, operators and enterprises can deploy select AWS services locally while retaining seamless connectivity to an extensive array of services available within their local AWS Region. This empowers them to run applications and workloads on-premises effortlessly, leveraging the familiar ecosystem of AWS services, tools, and APIs. AWS Outposts are optimized to support workloads and devices necessitating low-latency access to on-premises systems, local data processing, data residency compliance, and application migration with local system interdependencies.

Within the Charter team, a full AWS outpost rack boasting 42U capacity has been procured, comprising a balanced mix of CPU (Central Processing Unit) and GPU (Graphics Processing Unit) resources. The entire ordering process for the outpost was easily executed through the user-friendly AWS console, ensuring transparency and enabling real-time tracking at each acquisition stage.

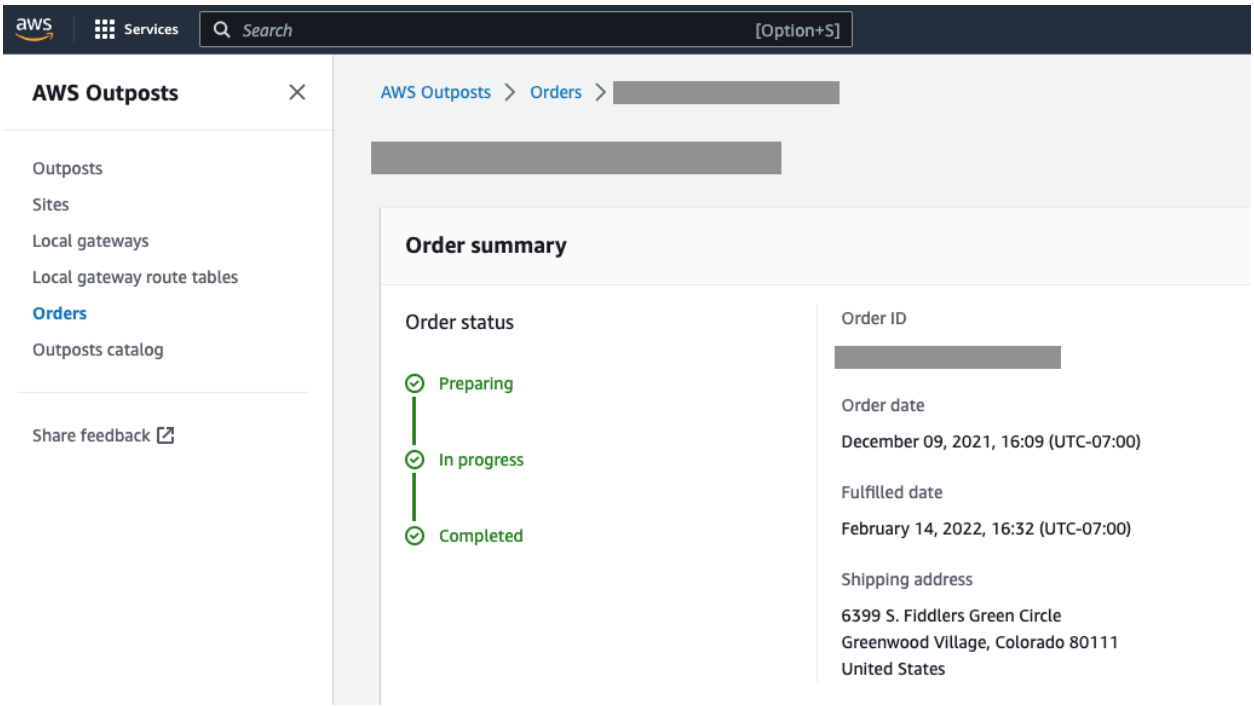


Figure 4: AWS Outpost Ordering and Tracking via the Console

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2. Charter Lab AWS Outpost Configuration

The AWS Outpost deployed within the Charter lab establishes a secure connection with the AWS US West region through a robust site-to-site VPN (Virtual Private Network). This enables the extension of the Virtual Private Cloud (VPC) from the region directly to the Outpost's on-premises infrastructure, ensuring seamless connectivity between workloads in both environments. While the AWS Outpost offers the alternative of connecting back to the region via AWS Direct Connect, the Charter lab team deliberately chose the VPN option to meet their specific requirements.

Following the successful arrival and configuration of the Outpost, it seamlessly integrated into the AWS ecosystem, accessible and manageable through the same familiar AWS console interface utilized for other resources within the AWS region. This cohesive experience enhances ease of use and ensures a consistent operational environment for the Charter lab team.

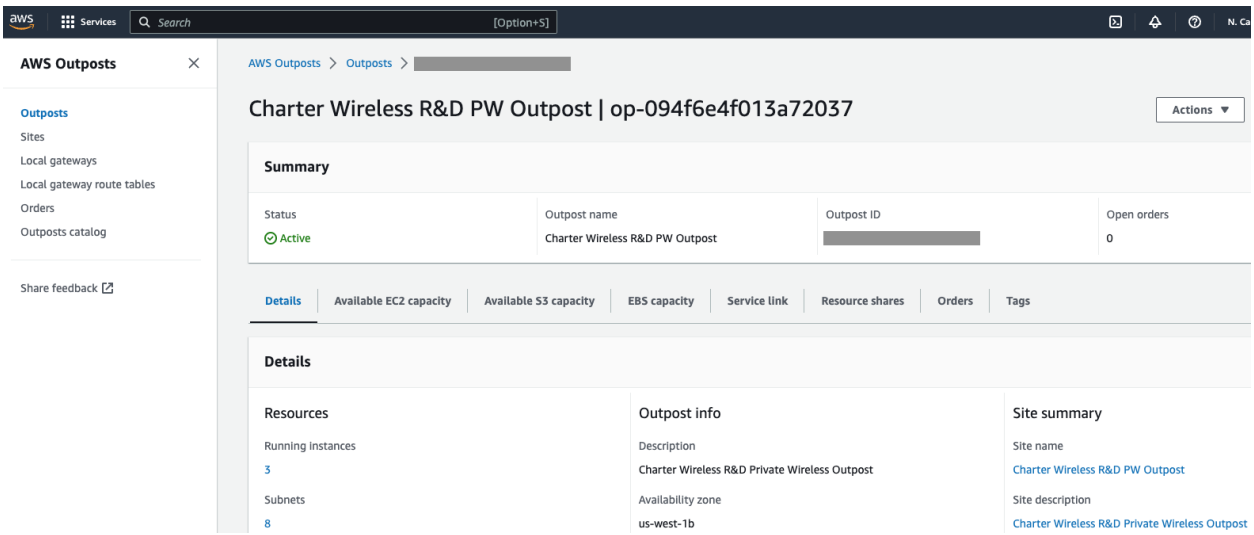


Figure 5: Outpost Management via AWS Console

The outpost in Charter lab has GPUs in the form of four G4dn.2xlarge and four G4dn.4xlarge and CPUs in the form of six M5.4xlarge instances.

Table 1: Charter's AWS Outpost Compute Configuration

Instance Type	Quantity
G4dn.2xlarge	4
G4dn.4xlarge	4
M5.4xlarge	6

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The distinction between G4dn.2xlarge and G4dn.4xlarge instances lies in their vCPU count and CPU memory capacity. Both instance types, powered by NVIDIA T4 GPUs and custom Intel Cascade Lake CPUs, offer exceptional performance. Notably, the GPUs are consistent across both types, featuring a single T4 GPU equipped with 16 GB of GPU memory.

Instance	GPUs	vCPU	Memory (GiB)	GPU Memory (GiB)	Instance Storage (GB)	Network Performance (Gbps)***	EBS Bandwidth (Gbps)
g4dn.xlarge	1	4	16	16	1 x 125 NVMe SSD	Up to 25	Up to 3.5
g4dn.2xlarge	1	8	32	16	1 x 225 NVMe SSD	Up to 25	Up to 3.5
g4dn.4xlarge	1	16	64	16	1 x 225 NVMe SSD	Up to 25	4.75
g4dn.8xlarge	1	32	128	16	1 x 900 NVMe SSD	50	9.5
g4dn.16xlarge	1	64	256	16	1 x 900 NVMe SSD	50	9.5
g4dn.12xlarge	4	48	192	64	1 x 900 NVMe SSD	50	9.5
g4dn.metal	8	96	384	128	2 x 900 NVMe SSD	100	19

All instances have the following specs:

- 2.5 GHz Cascade Lake 24C processors

Figure 6: GPU-based Instances on Outpost

The configuration mentioned in Table 1 is what Charter decided to have. However, the Outpost compute instances are configurable via a slotting mechanism. Once slotted, the configuration can't be changed unless all EC 2 instances are terminated.

The original Outpost configuration is two g4dn.12xlarge and one m5.24xlarge, as shown in Figure 7 but the Charter team decided to slot it as in Table 1 to run a 5G SA core and enterprise application simultaneously on the outpost.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

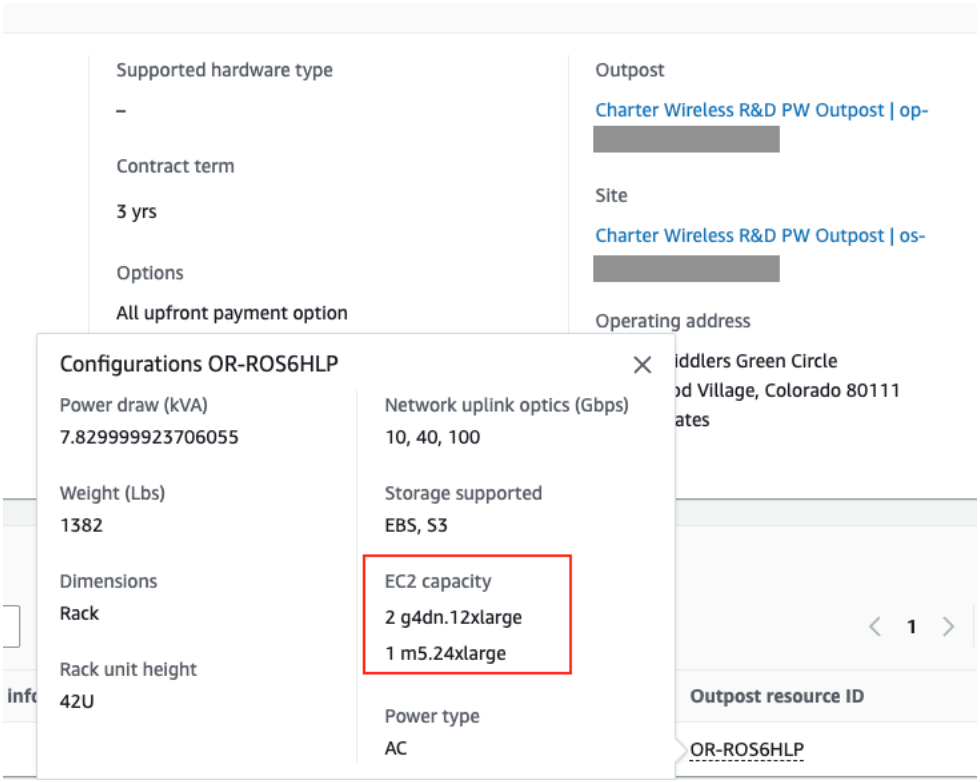


Figure 7: Outpost Original Instance Capacity before Slotting

Figure 8 illustrates various g4dn slotting options, each offering distinct configurations in terms of CPU, GPU, and memory utilization. The Charter team engaged in several productive discussions with the AWS team to carefully determine the optimal slotting requirements for the project. These discussions took into account the specific use cases, computing needs, capacity requirements, as well as potential future workloads anticipated to run on the outpost. Through collaborative deliberation, the team arrived at a well-informed decision on the most suitable slotting configuration.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

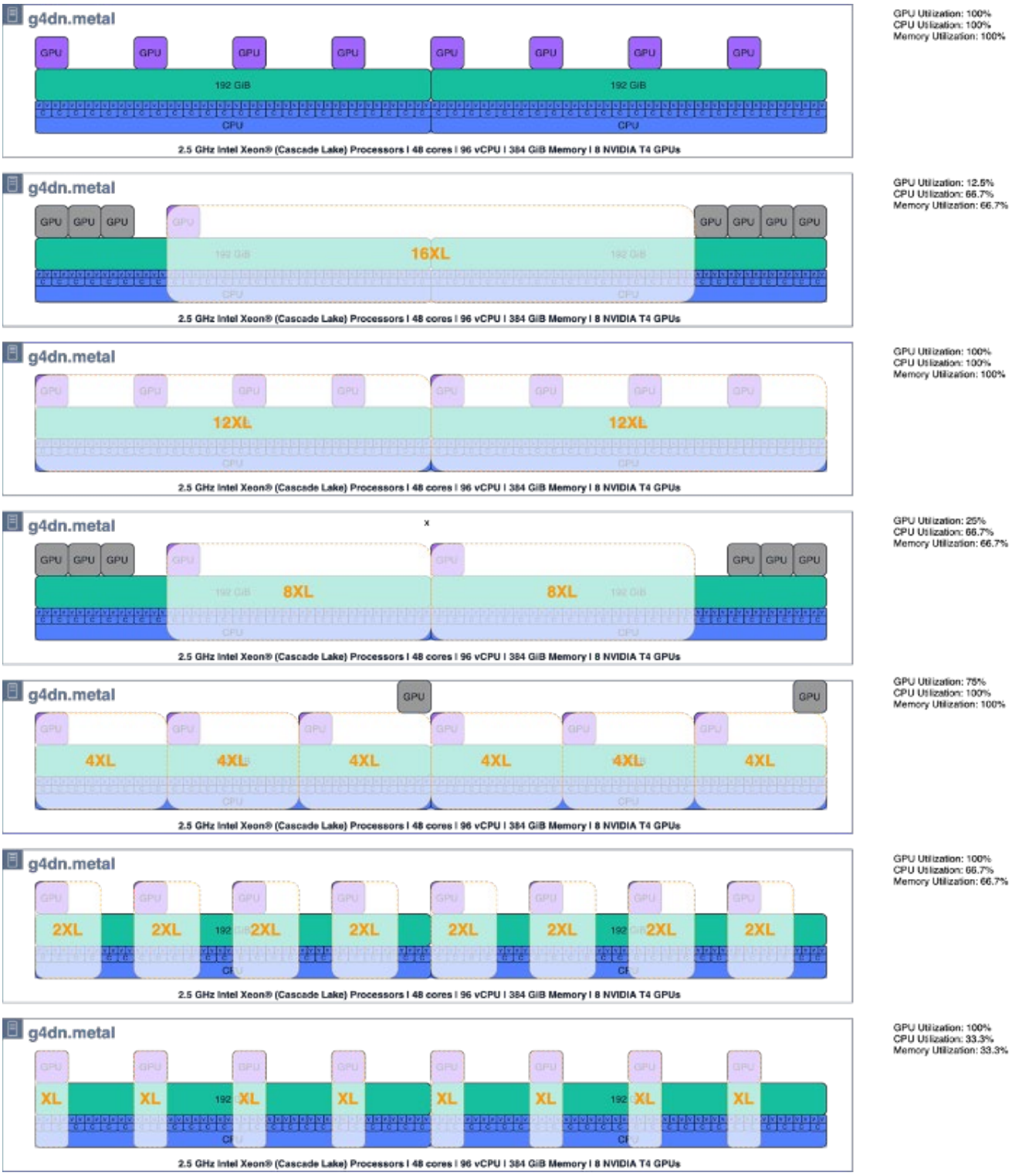


Figure 8: Example of g4dn.metal Slotting Options

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Considering future scalability, the Charter team deliberately selected an AWS Outpost configuration allowing for expansion. As depicted in Figures 9 and 10, the Charter AWS Outpost is designed with ample physical space to accommodate additional compute and storage resources as required. This strategic decision ensures that the infrastructure can quickly scale and adapt to evolving demands, providing the flexibility needed for future growth and optimization.



Figure 9: AWS Outpost in Charter Lab

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

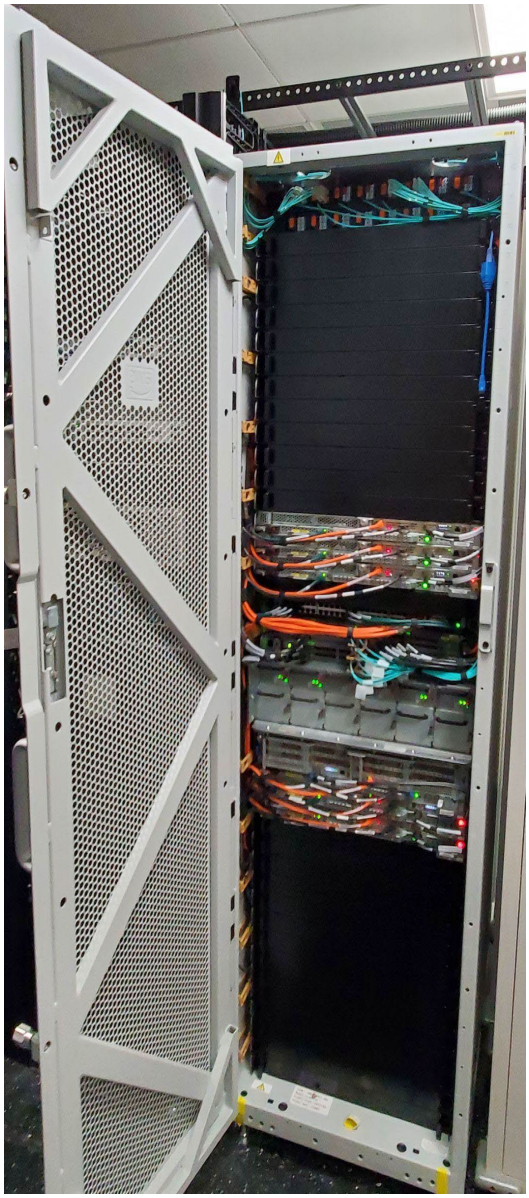


Figure 10: AWS Outpost Inside View Showing Installed RUs

4.3. Charter Private Wireless Lab Setup

In the Charter lab, a robust 5G system has been successfully implemented by leveraging a combination of on-premises servers and cloud infrastructure provided by AWS Outpost. To ensure efficient deployment of specific components, three Linux-based hosts have been utilized: one for the 5G core, another for the Radio Security Gateway (R-SGW), and the third for the OAM NAT (Network Address Translation).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The 5G core and R-SGW have been deployed on-premises to optimize performance and reduce latency. This strategic placement near the 5G small cell enables efficient security processing, resulting in lower latency during data transmission. Conversely, the OAM component responsible for network management and monitoring has been deployed in the AWS cloud region (us-west-1). This cloud-based approach allows for centralized management and monitoring of the 5G network infrastructure, eliminating the need for physical presence at each network site. Leveraging the cloud-based OAM offers advantages such as streamlined management and effective control over the 5G networks.

The Charter team has implemented a site-to-site VPN to establish secure communication and seamless data transfer between the on-premises servers and cloud resources. This VPN ensures the confidentiality and integrity of the communication between the on-premises infrastructure and the cloud-based resources.

For reliable communication between the 5G core and the 5G small cell, a high-speed 1 Gbps Ethernet connection has been employed. This connection utilizes an Ethernet cable connected to a gigabit switch, facilitating efficient and robust communication between these crucial components of the 5G system.

The Charter team chose 5G Citizen Broadband Radio Service (CBRS) small cells for the lab setup. CBRS is the 150 MHz of spectrum in the 3550 MHz to 3700 MHz range (3.5 GHz to 3.7 GHz) that the United States Federal Communications Commission (FCC) has designated for sharing among three tiers of users: incumbent users, Priority Access License (PAL) users, and General Authorized Access (GAA) users.

Charter Communications owns a significant amount of PAL across major U.S. markets.

Collaborating with the AWS team, Charter has deployed a 5G SA core supplied by a major 5G core provider and two small cells from two different vendors capable of transmitting on n48 band. This combination of cutting-edge technologies ensures the successful operation of the 5G system, meeting the requirements of Charter's implementation.

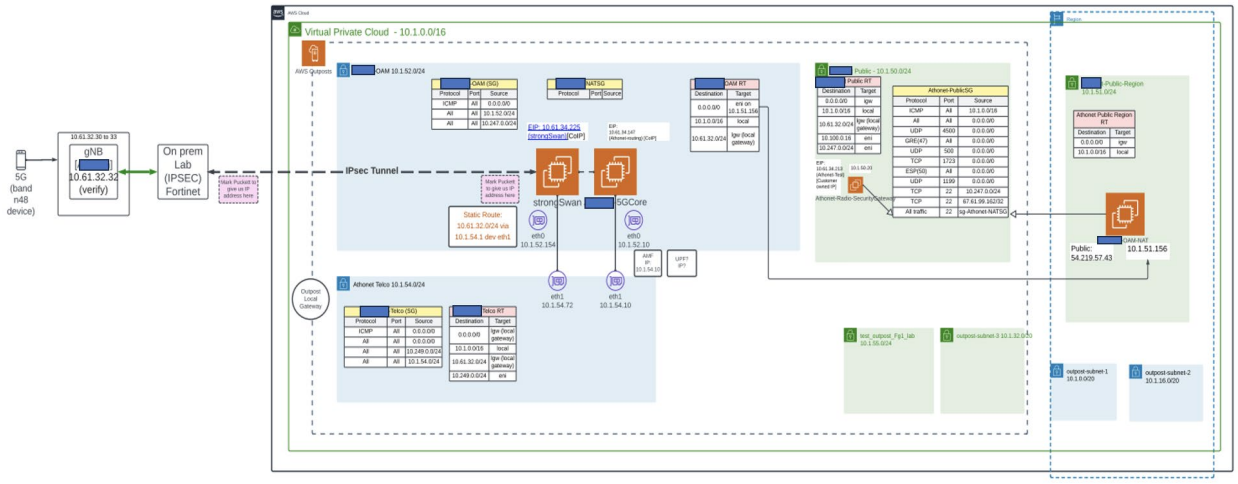


Figure 11: Charter's Lab Private Wireless on AWS Outpost Architecture

Implementing the 5G Core network in the Charter lab demonstrates a well-designed architecture that effectively utilizes AWS Outpost, as depicted in Figure 11. The core network is deployed on-premises

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

using the "m5.4xlarge" server instance provided by AWS. This powerful server instance ensures optimal performance and resource allocation for the 5G Core software.

In the us-west-1 region, a dedicated "t3.micro" instance operates the Operations, Administration, and Maintenance (OAM) service. This instance is vital in comprehensive management, monitoring and fault detection within the 5G Core network. By leveraging this cloud-based OAM instance, Charter can efficiently manage and maintain the network infrastructure remotely.

The "m5.4xlarge" servers are intelligently configured with virtualized network functions specifically designed for the 5G Core. These servers connect to the local network through a Layer 3 (L3) switch, directly linking with the 5G small cell. The 5G small cell supports Band n48, allowing 5G phones to establish connections and benefit from the high-speed network.

This carefully planned setup enables the deployment of a robust 5G Core network on-premises, providing greater control and minimizing latency throughout the network infrastructure. The powerful "m5.4xlarge" server instance ensures smooth operation of the 5G Core software. In contrast, the OAM instance in the us-west-1 region enables efficient management and maintenance of the network infrastructure from a remote location.

The 5G SA core setup was easy and quick since the 5G core was based on containers and predefined templates. Using the AWS console and with just a few clicks, the 5G SA core was instantiated and configured successfully. This easy and quick 5G Core setup results from AWS and telecom vendors' work to ensure compatibility and a hassle-free setup.

After standing up the 5G core and connecting the 5G small cells, Charter and AWS collaborated closely to conduct comprehensive testing of the lab setup, ensuring that devices could successfully attach and achieve the maximum throughput, validating the reliability and performance of the implemented 5G Core network.

5. Lab Private Wireless Use Case

The Charter team has successfully leveraged the AWS Outpost and a P5G network within their lab to implement an enterprise use case, showcasing an end-to-end network and enterprise application integration on the same hardware. Specifically, they have utilized a portion of the outpost compute as Multi-Access Edge Computing (MEC), enhancing the capabilities of their infrastructure.

The highlighted enterprise use case focuses on worker safety, featuring an innovative application that detects non-compliant workers and takes immediate action to ensure their safety. In unsafe scenarios, the application promptly intervenes by stopping the machine being operated by the worker.

This integration of AWS Outpost, P5G network and the worker safety application demonstrates the Charter team's commitment to harnessing cutting-edge technologies to create a secure and efficient working environment. By utilizing MEC functionality, they have effectively brought compute capabilities closer to the edge, enabling swift and intelligent decision-making at the point of action, thereby prioritizing worker safety and well-being.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.1. Worker Safety Use Case

The concept of MEC has revolutionized the way traffic and services are handled by moving them from the cloud to the network's edge. This shift brings significant improvements in terms of latency and security. MEC is advantageous for applications such as Data and Video Analytics, Augmented Reality, and various Internet of Things (IoT) use cases. Recognizing the strengths of MEC, the idea emerged to host a low-latency safety application on this platform, specifically designed to monitor and prevent accidents in a factory environment involving a drill press.

Disclaimer: The following example involving Personal Protective Equipment (PPE) and Mobile Edge Computing (MEC) is purely for illustrative purposes to showcase the potential benefits of MEC in a private network edge scenario. It is not intended to represent or comply with real-life safety standards, safety codes, or regulations. For accurate and up-to-date information on safety standards and regulations related to Personal Protective Equipment or any safety-related matters, it is essential to consult and adhere to the guidelines provided by relevant safety administrations and authorities. This example should not be used as a reference for establishing safety protocols or procedures. Always prioritize safety by following official safety guidelines and seeking advice from appropriate experts and regulatory bodies.

The lab worker safety application developed for this purpose focuses on ensuring compliance with Personal Protective Equipment (PPE) requirements, including helmets, safety vests, gloves and goggles. By verifying the presence of the necessary PPE, the application categorizes workers as either "green" or "red" based on compliance. Its primary objective is to prevent accidents caused by the tips of drill press bits. The application incorporates a computer vision model and algorithm responsible for recognizing and localizing detections to achieve this.

The architecture of the overall worker safety application is depicted in Figure 12, highlighting the interconnected components and their functionalities.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

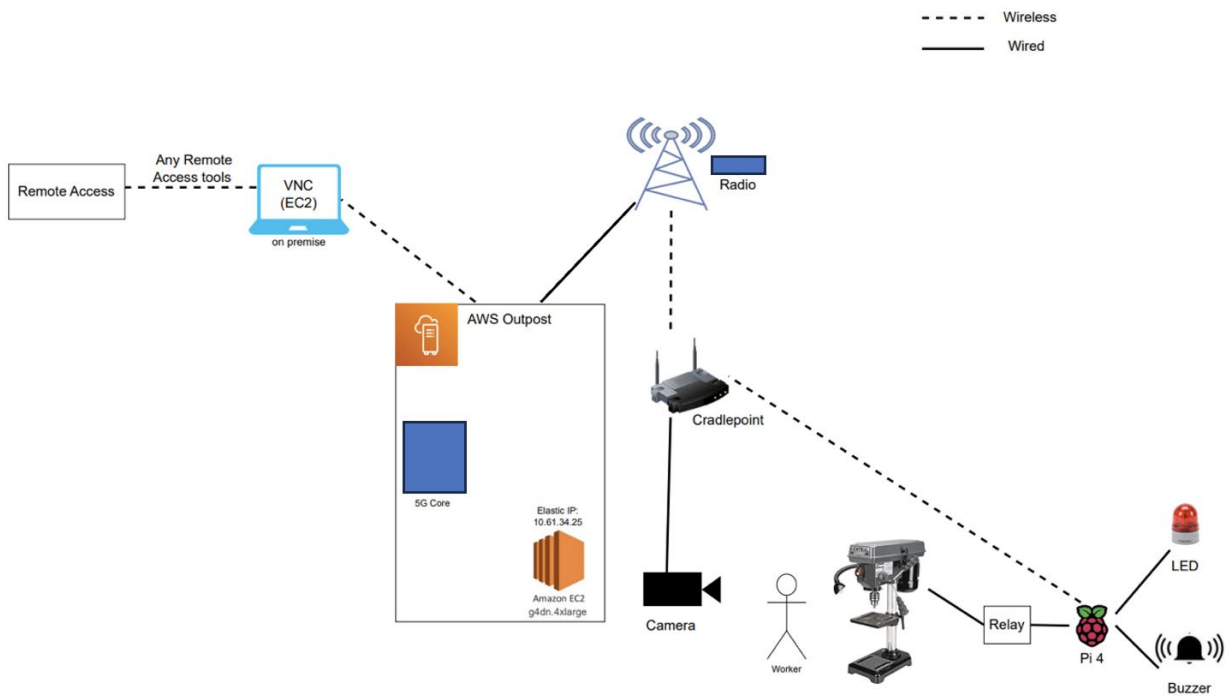


Figure 12: Charter Lab Use Case Architecture Deployed on AWS Outpost

To implement this application, the setup utilizes AWS Outpost as the MEC platform, hosting the 5G SA core, 5G radio components, a Cradlepoint router, a camera, a drill press machine, a Raspberry Pi, a relay switch, LEDs and a buzzer.

This configuration creates a Radio Access Network (RAN) ecosystem, with the 5G core hosted on the AWS Outpost and the radios strategically placed throughout the workplace, establishing connections with routers and integrating all the components. The camera, which operates using Power over Ethernet (PoE), is positioned to focus on the drill press. The camera leverages the Real-Time Streaming Protocol (RTSP) to transmit camera frames over the network.

A GPU processes each frame captured by the camera within the AWS Outpost, and the resulting output responses are sent back over the radio to the corresponding Raspberry Pi. The Raspberry Pi is wirelessly connected to the router and controls the relay switch, which is responsible for turning the drill press on/off and managing the LEDs and buzzer.

Integrating these components in the MEC architecture ensures real-time safety compliance monitoring and swift preventive action, effectively reducing the risk of accidents in the factory environment.

5.2. Use Case Computer Vision Model

Implementing a computer vision model is essential to detect workers, PPE, and machines accurately. The chosen model for this purpose is YOLOv4, a well-known and widely used deep neural network computer vision model. YOLO, which stands for You Only Look Once, has gained popularity as an open-source solution.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The development cycle of the computer vision model for object recognition tasks follows a sequential process comprising several crucial steps: data collection, data annotation, model training, model validation and model deployment.

The data collection process involves creating a custom dataset by capturing images specific to the application scenario and utilizing publicly available datasets online. We employed both approaches to develop our model, resulting in the collection of approximately 10,000 images. Once the data is gathered, the next step is annotation.

These 10,000 images needed to be annotated to be fed into the machine learning model to train it. Annotating is the practice of assigning labels to an image or parts of it. In this case, annotating was especially hard as each image included several objects to annotate, like a vest, helmet, gloves, worker, and more. Annotating the 10,000 images manually would have taken hundreds of hours to do. The Charter team had to find a solution for that.

5.2.1. Annotating Images using AWS Mechanical Turk

The Charter team initiated the annotation process by meticulously labeling a subset of images from the dataset, gaining valuable insights into the requirements, time consumption and obtaining examples of annotated images.

To ensure accurate and efficient annotation of the entire 10,000-image dataset, the team leveraged AWS SageMaker's Ground Truth, a cloud service that simplifies data labeling and provides streamlined annotation jobs. Ground Truth facilitated the utilization of human annotators through various means such as Amazon Mechanical Turk, third-party vendors, or an in-house workforce. Amazon Mechanical Turk functions as a crowdsourcing marketplace, enabling the delegation of tasks to a distributed global workforce. These tasks range from simple data validation and research to more subjective assignments like survey participation and content moderation. AWS Mechanical Turk empowers businesses to tap into a diverse workforce's collective intelligence, skills, and insights, enhancing operational efficiency, data collection, and analysis and expediting machine learning development.

By combining the manual annotation efforts of the Charter team with the assistance of AWS SageMaker's Ground Truth service, the dataset received comprehensive annotation, providing the crucial training data required for the YOLOv4 model. This approach ensures accurate and reliable object detection and classification within the computer vision system for workers, PPE, and machinery.

Remarkably, the annotation of the 10,000 images was completed in under 24 hours, delivering highly satisfactory results.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

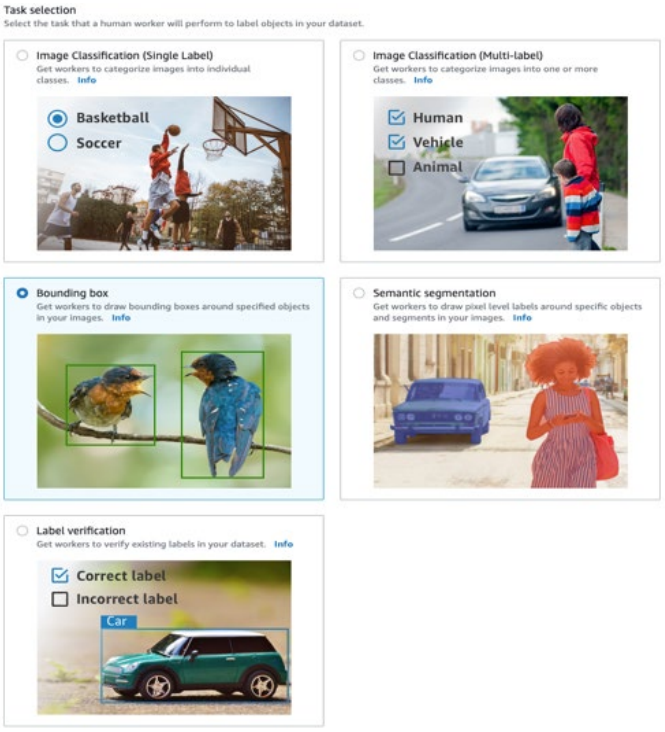


Figure 13: AWS Mechanical Turk Labeling Task

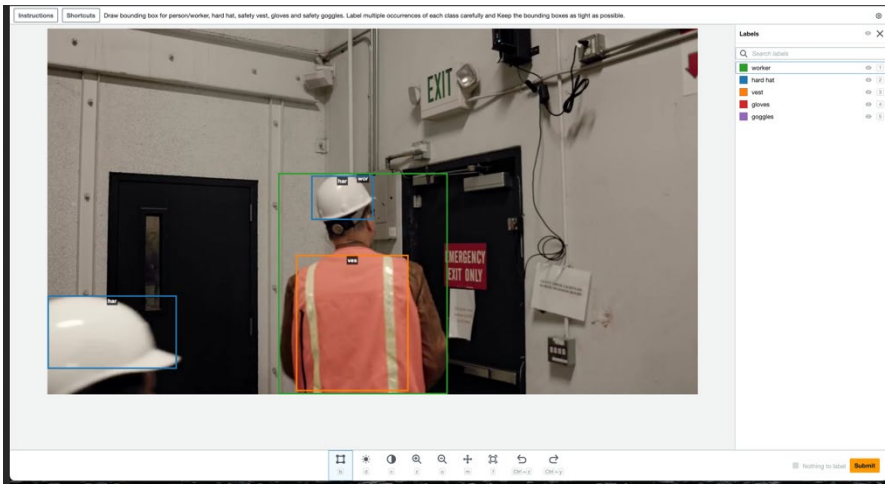


Figure 14; AWS Mechanical Turk - Charter's Worker Safety Task

5.2.2. Training Machine Learning Model on EC2

The computer vision model employed for this project is highly complex, comprising more than 60 million parameters. If the training of this model were to be carried out on a conventional computer, it would take

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

several months to complete. The model training was executed on an EC2 instance within the AWS regions to expedite the training process.

Specifically, the training took place on a p3.16xlarge instance, utilizing the Deep Learning AMI. The p3.16xlarge instance boasts eight Tesla V100 GPUs, collectively offering a GPU memory capacity of 128 GB. Leveraging the computational power of this setup, the training task was significantly accelerated. Instead of taking months, it was completed within hours.

The training process effectively utilized the CUDA platform, which was pre-installed on the AMI, further optimizing the training efficiency. The model's input picture size was set at 416 x 416, and 85% of the dataset was randomly selected for training. Important training parameters, such as the batch size and number of epochs, were determined as 64 and 25000, respectively. The model training yielded an average loss estimate of 1.5387, indicating a satisfactory training outcome.

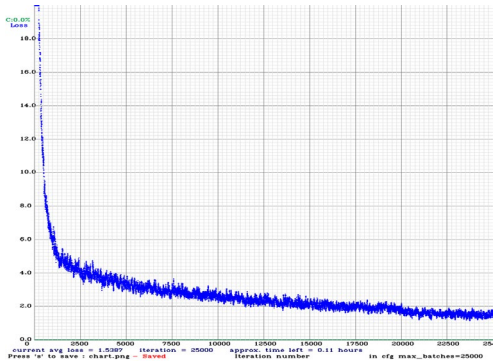


Figure 15: Training Loss Curve

To ensure the reliability of the trained model, it underwent validation using the remaining 15% of images from the dataset. The model's performance was evaluated using the mean average precision (mAP) metric, which comprehensively assesses its behavior.

For the specific trained model with an input size of 416x416, benchmarking revealed an impressive frames-per-second (FPS) rate of 72 FPS. Additionally, the mAP at an Intersection over Union (IoU) threshold of 0.75 was measured at 0.7238 or 72.38%. These results demonstrated the model's ability to detect and effectively classify objects within the images.

During the benchmarking process, it was observed that altering the input size impacted the model's performance. Increasing the input size resulted in improved mAP scores but reduced the frame rate, highlighting the trade-off between accuracy and computational efficiency.

To facilitate the training model's integration and deployment, the binary weight file was converted to the ONNX framework, streamlining its conversion to the TensorRT framework. TensorRT is a model acceleration framework that optimizes weights to a lower resolution format while preserving the model's behavior. As a result, the accelerated model achieved an impressive 96 FPS frame rate.

The inference of the accelerated model was conducted on an AWS Outpost g4dn.4xlarge instance, utilizing the Deep Learning AMI. This particular instance is equipped with a single T4 GPU and 64 GB of GPU memory, providing sufficient computational resources for efficient model inference.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In the Charter lab, the trained model was successfully implemented and executed using the AWS Outpost. The provided pictures in Figure 16 are samples of the model's inference, demonstrating its ability to detect objects within the given images accurately.



Figure 16: Charter Developed Model Inference

5.3. Private 5G Network and Uplink Capacity

It was necessary to explore solutions enabling efficient video uplink to accommodate the demands of multiple cameras in enterprise use cases. Two approaches were considered for addressing this challenge.

The first approach focuses on optimizing the 5G network configuration. One way to achieve this is by modifying the Time Division Duplex (TDD) frame configuration, allowing for the allocation of more uplink slots. This enables an increased capacity for camera uplink transmissions. Additionally, the uplink carrier aggregation feature can be employed to augment the available bandwidth for uplink purposes. Leveraging these network-level adjustments can enhance the capacity for handling multiple camera streams.

The second approach involves optimizing the video feed itself. Various techniques can be employed to achieve this goal. One such technique is utilizing advanced encoding methods to reduce the size of the video data without compromising its quality. Lower frame rates and resolutions can also be employed to minimize the bandwidth requirements of each camera stream. Implementing these optimizations makes it feasible to support multiple camera feeds simultaneously on the 5G private wireless network.

In the subsequent section, we will delve deeper into the specifics of optimizing the video camera feeds, exploring methods and strategies that enable the efficient operation of multiple cameras within the 5G private wireless environment.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.3.1. Camera Configuration

Typically, cellular networks prioritize downlink traffic and possess a restricted capacity for uplink transmission. This inherent limitation significantly impacts the number of cameras that can be accommodated within the network and their resolution and frame rate configurations. The encoding algorithm employed by the cameras directly influences the instantaneous upload bitrate required for transmitting the video data over the network.

Over time, encoding techniques have evolved, becoming more dynamic in nature. The H.264 encoding standard was employed in the cameras utilized for testing. This particular encoding standard offers distinct advantages over other encoding types, especially when dealing with dynamic field-of-view scenarios. The utilization of H.264 ensures efficient compression of video data while maintaining reasonable image quality, thereby optimizing the utilization of the limited uplink capacity in cellular networks.

5.3.2. Private 5G Uplink Capacity, Camera FPS, and Resolution

In this scenario, a crucial aspect was finding the right balance between uplink bitrate/throughput, measured in megabits per second (Mbps), and the camera feed's desired FPS and resolution. Adjustments to the FPS and resolution were made within the camera's stream settings to achieve the desired outcome.

To assess the uplink throughput, the router was configured to estimate the data transmission rate over a 30-minute period, gathering 100 samples per hour. However, it should be noted that certain tests could not be conducted due to the cameras used for testing lacking support for 60 FPS at higher resolutions.

It's important to recognize that higher resolutions necessitate a greater uplink bitrate. Consequently, a delicate balance needed to be struck between the FPS settings that would ensure accurate operation for the specific use case and the inherent limitations of the uplink capacity.

By carefully considering these factors, an optimal configuration was sought to enable effective camera performance while working within the constraints of the available uplink capacity.

Table 2: Uplink Bit Rate vs. FPS and Resolution

FPS	*15*			*30*			*60*		
Resolution	*Uplink Bitrate (Mbps)*			*Uplink Bitrate (Mbps)*			*Uplink Bitrate (Mbps)*		
	Min	*Avg*	*Max*	*Min*	*Avg*	*Max*	*Min*	*Avg*	*Max*
640x480	0.37	0.59	4.2	0.29	0.97	1.34	1.20	1.62	2.30
1280x720	0.62	1.03	1.63	1.52	1.86	2.54	3.24	5.03	6.93
1920x1080	2.29	2.96	3.85	3.30	4.13	6.64	NA	NA	NA
2992x1680	5.15	6.96	9.32	3.89	8.12	11.13	NA	NA	NA

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

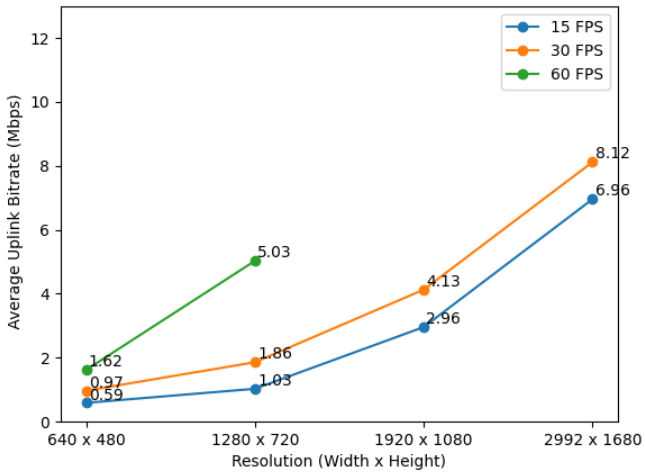


Figure 17: Resolution vs. Average Bitrate

5.3.3. The Video Streaming Choice for the Lab Use Case

To ensure the scalability of the worker safety solution, careful consideration is given to the tradeoff between the resolution and frame rate of the video stream. This decision enables the connection of multiple cameras to a single CBRS 5G router, facilitating a comprehensive monitoring setup.

When analyzing the uplink bitrate usage, it becomes evident that lower-resolution streams consume less bandwidth than higher-resolution streams. Conversely, the uplink bitrate increases with higher frame rates. While lower-resolution images may exhibit graininess and pixel loss, they tend to be less effective in challenging lighting conditions.

Since this application pertains to safety, lower frame rates are a disadvantage as they capture fewer frames per second, resulting in less information. A higher frame rate is preferred to maximize the potential for incident detection and prevention. According to latency data, a frame capture latency of approximately 9ms is achieved at 60 FPS, allowing for more frequent updates and providing more information to the algorithm. This advantage positions higher frame rates as capable of preempting unfortunate incidents more effectively than lower frame rates.

Considering the camera capture latency and uplink bitrate considerations, the application was run at a resolution of 1280x720 and a frame rate of 60 FPS. This configuration strikes a balance between capturing crucial details and minimizing bandwidth requirements.

A test was conducted using up to three cameras to assess the solution's scalability, examining the system's ability to handle multiple camera inputs simultaneously.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 3: Uplink Bitrate for Multiple Cameras

Number of Cameras	Uplink Bitrate (Mbps)		
	Min	Avg	Max
2	6.48	7.37	7.76
3	9.04	11.7	14.61

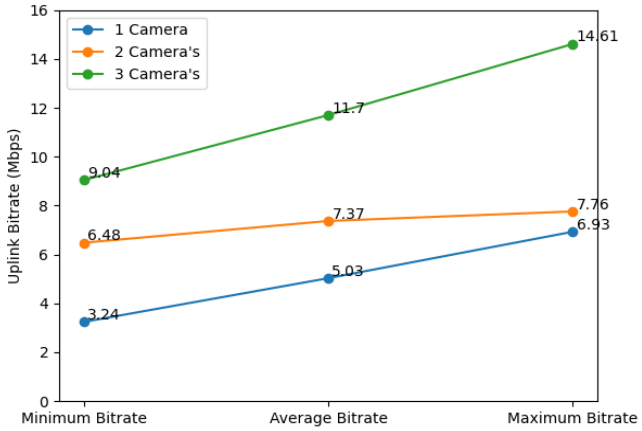


Figure 18: Uplink Bitrate for Multiple Cameras

5.4. Private 5G Use case Latency

The end-to-end latency measured for the use case was below 50 msec. This 50 msec is measured once the camera captures the first video frame until the drill press stops. All delays in the system are mentioned in Table 4.

Having both the network and use case deployed on edge helped reduce the latency significantly. The Charter team believes that further optimization could be done to reduce the overall latency further.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 4: Latency Causing Delays in the Worker Safety Use Case

System Delay	Definition
Camera Fetch	Time between the EC2 request a single frame from the camera till the single frame is received by the EC2
Processing	Time taken by the EC2 to run a single frame through the neural network model (YOLOv4) detect PPE objects
Code Algorithm	Time taken by EC2 to determine if the worker is compliant or not and what kind of non-compliance is happening
Code Processing per Frame	Time taken to run the overall algorithm for a single frame. Code Processing per Frame = Camera Fetch + Processing + Code algorithm.
Downlink	Time to send a command from outpost EC2 to RPi via wired, cellular, and WIFI links and RPi taking action
Relay trigger time	Time taken by relay to turn on/off
LED light change	Time taken by LED to change colors
RPi	Delay added in raspberry Pi to properly receive payloads over wireless link. At the current rate of 60FPS the RPi was unstable so we added a delay/buffer in the RPi to prevent crashes
Buzzer	Delay added for proper functioning of buzzer, without adding this delay the buzzer was out of sync with the LED and Relay
Overall Delay per frame	Time taken to receive a frame and translate the frame to give output at the RPi (end to end delay)

Finally, the worker safety use case and Private 5G network functions ran on the same hardware AWS outpost and the team demonstrated how to quickly and efficiently spin up a private network along with a useful enterprise use case.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Conclusion

Cloud edge computing allows telcos to provide their customers private 5G services while taking advantage of all services/offerings provided by public cloud. The Charter emerging technologies team in partnership with AWS telco team successfully implemented a P5G network hosted in AWS Outpost. The team deployed a worker safety application in this P5G environment. This project is a showcase for establishing a fully functional end-to-end 5G enterprise network, incorporating a practical video analytics application, all within a single lab hardware setup. The potential for expansion and further development is evident.

Looking ahead, the future work will explore the possibilities of deploying multiple worker safety applications across various locations, all powered by a centralized AWS outpost within the Charter network. This approach entails consolidating computational resources from individual locations into a central hub, enabling multiple enterprises to deploy their applications without the need for owning individual outposts. This centralized approach promises increased efficiency and cost-effectiveness for enterprises, streamlining their operations and fostering collaboration within the network.

Finally, CBRS proved very useful for private networks and can be leveraged to provide enterprise, government, and business-oriented 5G networks and low latency use cases.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

3GPP	The 3rd Generation Partnership Project
SA	Standalone
P5G	Private 5G
PLMN	Public Land Mobile Networks
NPN	Non-Public Network
PNI-NPN	Public Network Integrated NPN
SNPS	Standalone NPN
UPF	User Plane Function
SMF	Session Management Function
AMF	Access and Mobility Management Function
UDM	Unified Data Management
gNB	Next generation NodeB
CPU	Central Processing Unit
GPU	Graphics Processing Unit
VPC	Virtual Private Cloud
VPN	Virtual Private Network
OAM	Operations, Administration, and Maintenance
NAT	Network Address Translation
L3	Layer 3
MEC	Multi-Access Edge Computing
IoT	Internet of Things
PPE	Personal Protective Equipment
RAN	Radio Access Network
PoE	Power over Ethernet
RTSP	the Real-Time Streaming Protocol
YOLO	You Only Look Once
mAP	Mean Average Precision
FPS	frames-per-second
IoU	Intersection over Union
Mbps	Megabits per second
CBRS	Citizen Broadband Radio Service
FCC	Federal Communications Commission
PAL	Priority Access License
GAA	General Authorized Access

Bibliography & References

Non-Public Networks (NPN). Dongwook Kim, 3GPP MCC https://www.3gpp.org/technologies/npn

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Fully Automated Athonet 4G/5G Core Management and Orchestration on AWS *https://aws.amazon.com/blogs/apn/fully-automated-athonet-4g-5g-core-management-and-orchestration-on-aws/*

AWS for Edge https://aws.amazon.com/edge/

AWS Outposts Family Overview https://aws.amazon.com/outposts/

AWS Local Zones Overview https://aws.amazon.com/about-aws/global-infrastructure/localzones/

AWS Wavelength Overview https://aws.amazon.com/wavelength/

AWS Snow Family Overview https://aws.amazon.com/snow/

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Hyperscale Virtual Services Gateway

A Technical Paper prepared for SCTE by

Carl Klatsky
Senior Principal Engineer
Comcast
1800 Arch St., Philadelphia, PA 19103
+1 215 286 8256
carl_klatsky@comcast.com

DeFu Li
Distinguish Engineer
Comcast
1800 Arch St., Philadelphia, PA 19103
+1 267 260 3704
defu_li@comcast.com

Jason Combs
Senior Principal Engineer
Comcast
Jason_combs@comcast.com

Anton Grichina
Senior Software Engineer
Harmonic
anton.grichina@harmonicinc.com

Adam Levy
Director System Architecture
Harmonic
19 Alon Hatavor St., P.O.B 3600
Caesarea Industrial Park 3088900
Israel
adam.levy@harmonicinc.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. System Overview ... 3
   - 2.1. Virtual Cable Modem Termination System ... 3
   - 2.2. Virtual Services Gateway ... 4
   - 2.3. SmartNIC ... 5
3. Challenges & Evolution ... 6
   - 3.1. VCMTS & VSG Challenges Requiring SmartNIC Capabiltities ... 6
   - 3.2. Evolving the VCMTS ... 8
   - 3.3. Evolving the VSG ... 9
4. Integrated VCMTS-VSG Solution ... 9
   - 4.1. Integrated System ... 9
   - 4.2. Ability to Scale the System ... 11
   - 4.3. Future Plans for VSG Containerization ... 12
5. Conclusion ... 13

Abbreviations ... 14

Bibliography & References ... 15

List of Figures

Title **Page Number**

Figure 1 - MHAv2 ... 3
Figure 2 - VSG High Level ... 4
Figure 3 - VSG Reporting Infrastructure ... 5
Figure 4 - An Example of VCMTS CPU Core Allocation Model ... 6
Figure 5 - VSG VLAN Interconnection ... 7
Figure 6 - VSG NIC Internal View ... 8
Figure 7 - Traffic Steering in VCMTS ... 10
Figure 8 - VSG Replication & Encapsulation Rules in VCMTS ... 11
Figure 9 - Baremetal VSG Deployment View ... 12
Figure 10 - Containerized VSG & VCMTS Deployment View ... 13

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The Virtual Cable Modem Termination System (VCMTS) and Virtual Services Gateway (VSG) are newer network elements that have quickly become vital to running a large MSO network. The VCMTS disaggregates the traditional Cable Modem Termination System (CMTS) functions across a generalized computing structure. The VSG is in the logical path of the customer data packets enabling usage and quality analysis use cases. The next step is to integrate these two virtualized network elements.

The authors will outline an approach to building a hyper scale VSG that can scale in proportion to future needs. This approach uses a Smart Network Interface Card (NIC) with programmable network function accelerators for the platform's physical connectivity along with a control plane that utilizes the Linux SwitchDev model and Linux Traffic Control (TC) flows to direct traffic between the VCMTS and VSG systems. The approach allows both elements to rapidly scale up by simply increasing the capacity of the switching fabric and adding more compute nodes.

2. System Overview

2.1. Virtual Cable Modem Termination System

VCMTS is a collection of software applications implementing the Converged Cable Access Platform (CCAP) core functions according to the Distribute Access Architecture (DAA) specification, also known as the Modular Head-End Architecture version 2 (MHAv2) [1].

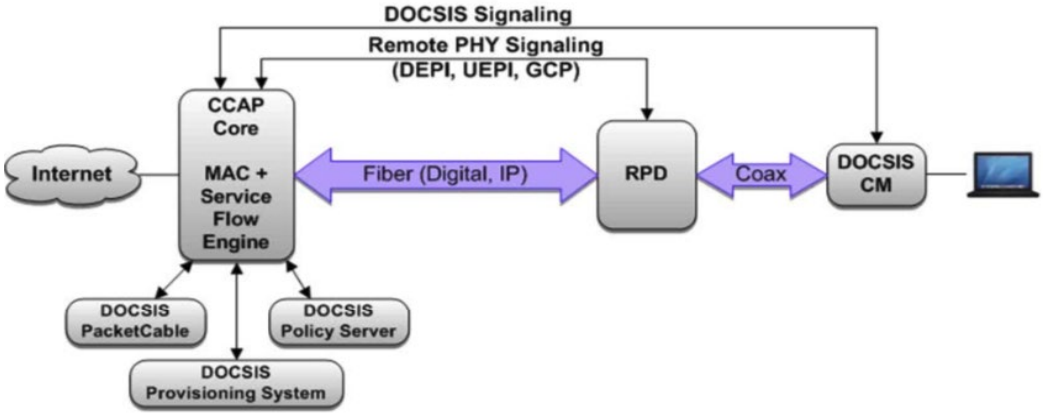


Figure 1 - MHAv2

This suite of software applications is built and deployed on the Edge Cloud Platform. As a software CMTS, all packet processing, encapsulation, encryption, and Data Over Cable System Interface Specification (DOCSIS) protocols are implemented with software running over Commercial Off-The-Shelf (COTS) servers. This approach provides significant power savings, cost reduction, flexible scalability, and performance increases as each generation of Central Processing Units (CPU) becomes more powerful, with very little to no additional development costs. The VCMTS architecture has maintained full independence from hardware from the beginning, while also taking advantage of. opportunities for offloading expensive operations onto the hardware components it runs over.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.2. Virtual Services Gateway

The Virtual Services Gateway (VSG) provides a new platform where next-generation network services are deployed close to the customer edge. Automation and flexible deployment best practices are used to simplify the creation of new services and augment the capabilities of existing network services. Current state-of-the-art Software-Defined Networking (SDN) and Network-Function-Virtualization (NFV) technologies are used to enable agile implementation and orchestrated management and monitoring.

The VSG is a software application capable of running on general-purpose compute nodes, either virtual or bare metal. It can share platforms with the VCMTS or other virtual elements deployed at the network edge. Initially placed in line with customer traffic, the performance of the VSG is critical to improving the customer experience. It is a low-latency flow-through platform that can forward traffic at a line rate, ensuring the introduction of the VSG does not cause a performance bottleneck.

The VSG is introduced between the CMTS/VCMTS and Residential Edge Router (RUR) in Headend or Hub sites where there is access to customer traffic near the customer's premises.

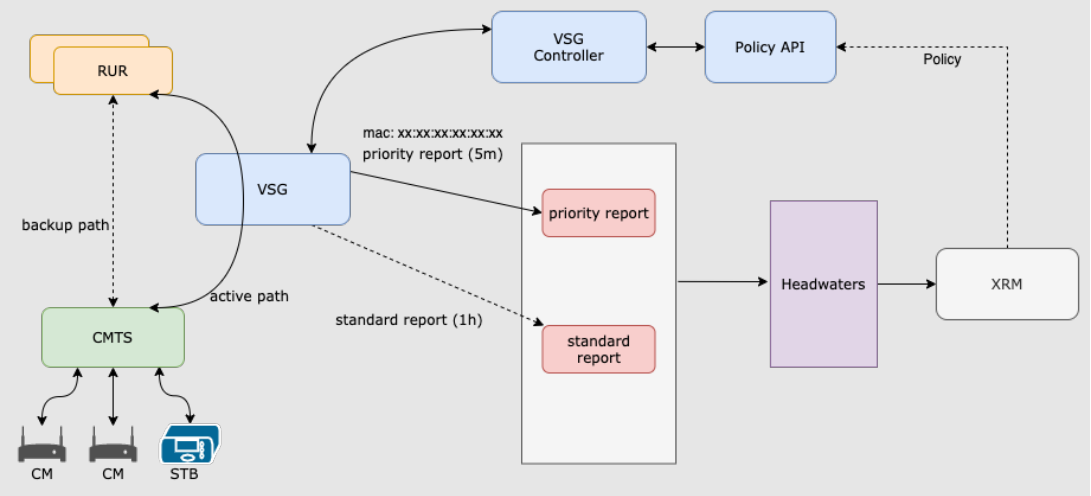


Figure 2 - VSG High Level

Potential use cases for the VSG are Usage Based Billing (UBB) and customer Quality of Experience (QoE) measurements.

For the UBB use case, the VSG platform needs to collect the usage byte information for each customer connected to the network. The VSG platform can be deployed in head-ends where it is able to collect and aggregate upstream and downstream byte usage per IP family per DSCP in regular intervals for each cable modem device. A mediation layer is needed to consume the byte usage, apply any needed business rules, and publish the data to back-office systems such as the Xfinity Resource Manager (XRM). This same data set can be used for capacity planning, usage trend analysis, and other analysis.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

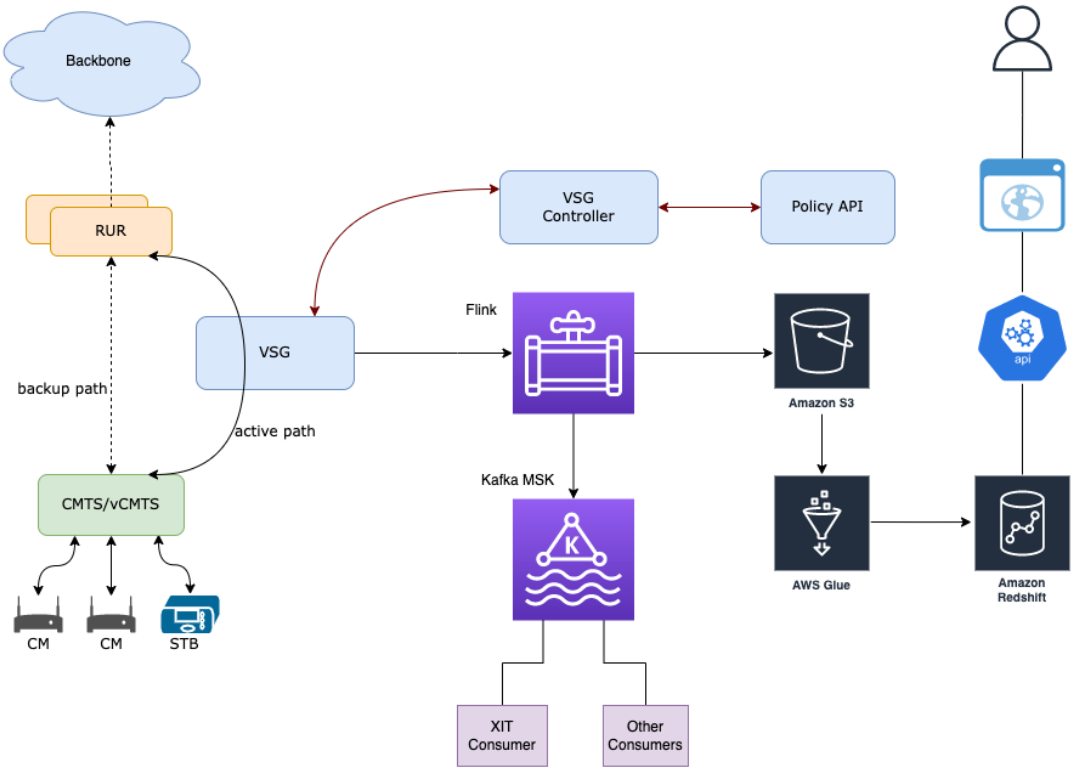


Figure 3 - VSG Reporting Infrastructure

The VSG Quality of Experience use case is to collect and derive the QoE of residential High-Speed Data (HSD) services provided by the MSO. The VSG is intended to be an IP flow-aware system. As such, it is able to analyze key Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) performance and network latency statistics. Statistics are able to be captured for each IP flow, aggregated on a per-customer basis, and streamed to back-office systems. The VSG back-office components are able to tie the collected performance data with other meta-data pulled from various available data sources to calculate the *Quality of Experience* for the HSD customer.

QoE is calculated based on multiple network performance parameters, including latency, bandwidth consumption, and packet retransmission. Thresholds are identified from research and data analysis over a period of time.

2.3. SmartNIC

In recent years, the industry term "SmartNIC" has been coined to describe a NIC (Network Interface Card) that includes advanced switching and/or processing offload capabilities. Some SmartNICs even include a CPU as an on-board component. The term SmartNIC is not defined by any standards body such as the Institute of Electrical and Electronics Engineers (IEEE) or the Association for Computing Machinery (ACM), and as such, has differing definitions as presented by the various NIC vendors. With no clear standards-based definition, the selection of a suitable NIC for various networking applications is quite a challenge.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Within the context of this paper, the authors use this definition of a SmartNIC: Any NIC with network switching offload capability to modify, copy, redirect, or forward packets in hardware, regardless of the presence of additional on-board CPU.

3. Challenges & Evolution

3.1. VCMTS & VSG Challenges Requiring SmartNIC Capabiltities

In running the Edge Cloud Platform, the primary business challenge comes in the form of ever-increasing bandwidth and scale.

Service Group (SG) bandwidth capacity continues to increase as part of the ongoing Radio Frequency (RF) plant and access technology upgrades. For example, going from 96 MHz to 192 MHz full band Orthogonal Frequency Division Multiplexing (OFDM) channels, Low-split to Mid-split RF plant, Full-Duplex DOCSIS (FDX), node splits, and so on. These are driven by subscriber growth, network demand increases, and new service offerings for the multi-gigabit symmetric speed-tiers. The VCMTS workloads running on the servers need to scale accordingly.

To attempt to keep up with this growth, each generation of the server rack built for the Edge Cloud Platform has a specific subscriber scale and network bandwidth capacity target. When transitioning to a newer generation server rack build, the number of subscribers is typically doubled, and the network bandwidth supported is more than doubled. For example, in the most recent generation, the change has been from 10 Gigabit per second (Gbps) ethernet connectivity per compute host to 100 Gbps ethernet connectivity—to allow for both the multiplexing of high bandwidth services to individual customers as well as continued general usage growth.

This increase in both subscriber scale and bandwidth utilization presents a problem for maintaining the VSG function. Prior to this generation, standalone bare metal servers were able to scale vertically to support 400 Gbps and thereby keep up with demand. However, this is proving to be unsustainable with the latest generation of the Edge Cloud Platform; this presents a clear need to be able to steer traffic more granularly to multiple VSG servers. Existing NIC solutions have many shortcomings in network traffic steering functions which need to be addressed. Figure 4 shows an example of the host CPU core allocation scheme for the deployment of VCMTS pods. The VCMTS Data Plane (DP) has dedicated CPU cores, while the VCMTS Control Plane shares the remaining CPU cores. The data and control plane traffic relies on the NIC for steering to their targeted Virtual Functions (VF) or Physical Functions (PFs) to reach the intended network interfaces of the VCMTS pods.

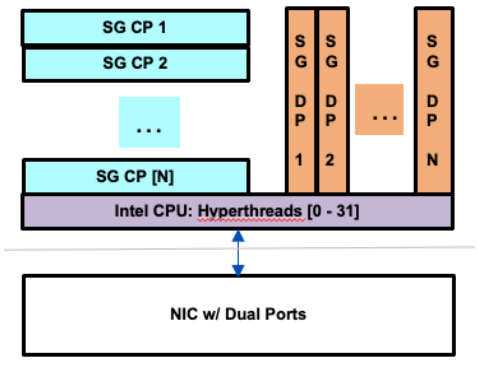


Figure 4 - An Example of VCMTS CPU Core Allocation Model

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The proposed solution, detailed in a later section, mirrors and tunnels the traffic from the VCMTS to the VSG utilizing the more capable SwitchDev model. A SmartNIC with programmable network function accelerators provides this ability to address the above challenges without sacrificing VCMTS throughput performance.

In the VSG system, a SmartNIC is needed to perform the Virtual Local Area Network (VLAN) translation and packet copy functions. As seen in Figure 5 - VSG VLAN Interconnection and Figure 6 - VSG NIC Internal View, the VSG is physically wired to the redundant routers with one 100 Gbps link to each router. Each link supports both upstream and downstream traffic. For the VSG to properly analyze the subscriber QoE, it is critical that the VSG application has some parameter to indicate upstream traffic from downstream traffic; this is achieved by VLAN ID or "VLAN tags" to distinguish the directionality of the traffic as it traverses the VSG. Traffic to and from the VCMTS is marked with a unique VLAN tag. Using the router's patch panel function, the VLAN tagged packets are re-directed through the router between the VCMTS port facing and the VSG-facing port. At the VSG, the VSG NIC makes a copy of the packet to send to the VSG application and changes the VLAN tag of the original packet for transmission back to the RUR and onwards upstream to the Aggregation Routers (AR). From the perspective of the router, it is only aware of the VLAN tag value after it has been modified by the VSG. The router has a layer three sub-interface corresponding to this VLAN tag value and proceeds to route the traffic onto the network as per standard layer three routing functions. The same process occurs in reverse for downstream traffic.

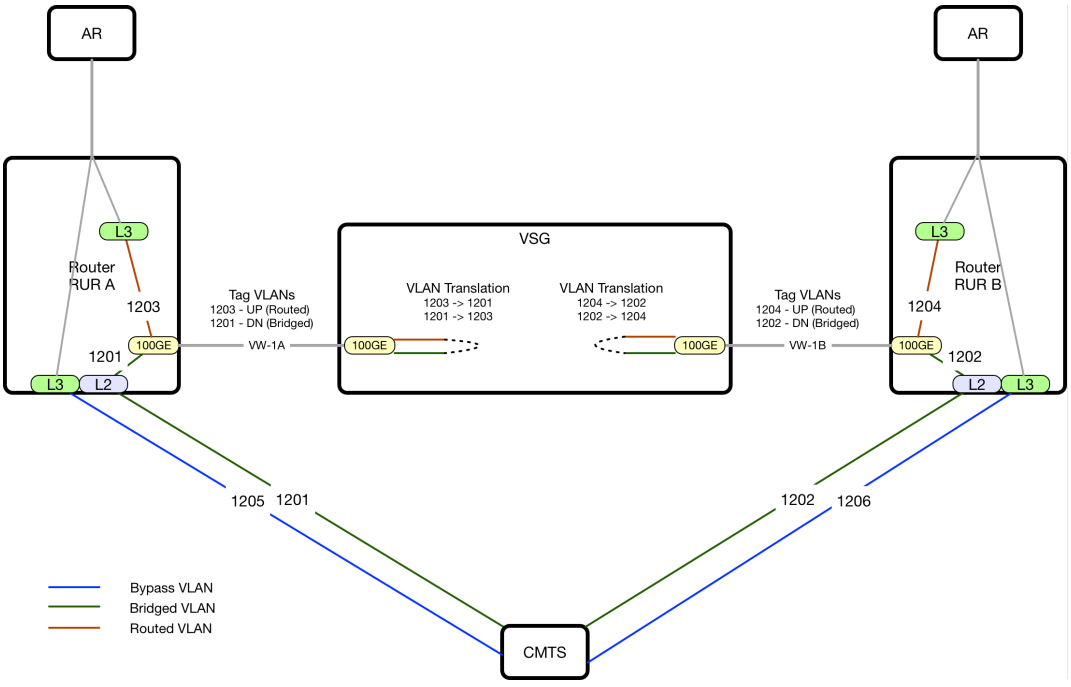


Figure 5 - VSG VLAN Interconnection

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

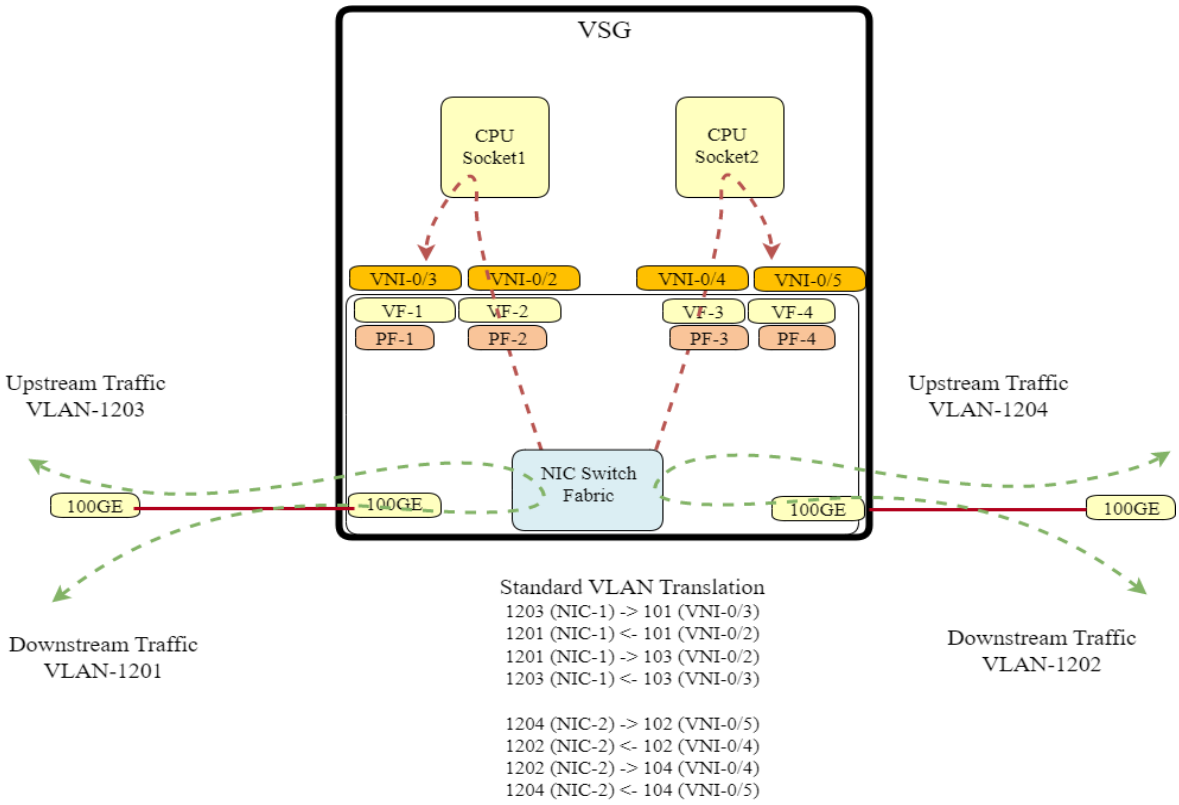


Figure 6 - VSG NIC Internal View

3.2. Evolving the VCMTS

As business needs evolve, the VCMTS needs to evolve to meet those needs. New capabilities are needed, such as increased throughput and improved packet processing offload.

A true SmartNIC with refined traffic steering and increased performance is required—one that supports a higher number of virtualized functions and more elaborate traffic management mechanisms.

Single Root-Input / Output Virtualization (SR-IOV) is an industry standard for NIC virtualization; it allows a physical device to be used simultaneously from different applications or virtual machines. Each user gets one or multiple VFs which are logical Peripheral Component Interconnect express (PCIe) devices with no performance overhead. SR-IOV is standardized and supported by different hardware vendors. SR-IOV can be utilized in a VCMTS.

An advanced traffic management technique is also needed to efficiently re-direct packets through the NIC. SwitchDev mode provides the needed traffic management technique. SwitchDev is an open-source infrastructure construct in the Linux kernel that controls how a SmartNIC processes incoming and outgoing packets. It standardizes a configuration mechanism for implementing SR-IOV. Further, it enables using the NIC as an e-switch and embedded switch within the NIC.

While with SR-IOV, each application is connected directly to the network via a VF, with SwitchDev the VF is connected to the eSwitch in the NIC. The SwitchDev mode enables more advanced schemes of communication between applications over VFs. The NIC becomes a configurable switch with extended functionality. It is possible to forward packets between “ports” according to a defined configuration. The

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

NIC utilizes two forwarding layers from the port to the e-switch and from the e-switch to the VFs. This additional e-switch layer makes the system more complex, but it also allows for greater flexibility. It also forces multi-tenant security by enforcing the e-switch steering to be configured from the host.

One example of a host-based infrastructure which uses SwitchDev is Open vSwitch (OVS). OVS is a mature virtual switching software widely used by telecom operators. OVS provides control plane and data plane functionality, and with SwitchDev integration, its data plane could be offloaded to the SmartNIC, which tremendously increases throughput while minimizing CPU load.

A SmartNIC is able to support OVS offloading via SwitchDev. The dynamic nature of OVS – learning from packets sent to the software – is superfluous for the VCMTS application as it already has all the information about all connected hosts. However, OVS provides a standard implementation and consistent behavior, and is therefore worth considering.

A SmartNIC enables many new opportunities, and a VCMTS implementation with a SmartNIC should try to balance between hardware offload opportunities and reduced complexity for best-in-class product stability and performance.

3.3. Evolving the VSG

In a typical deployment, a VSG needs the capability to perform 100 Gbps line rate packet copy and VLAN translation functions. A suitable VSG NIC did not include any additional onboard CPU for generalized compute functions, but as originally envisioned, the VSG does not need that functionality. The needed VLAN translation is shown above in Figure 5 - VSG VLAN Interconnection. The NIC needs to change the VLAN ID in all received packets based on the directionality of the packet. The NIC also needs to provide a packet copy function to send a copy of all received packets to the VSG software application.

To integrate a VSG more closely with the VCMTS and to do so at scale, the required packet copy and VLAN translation functionality will need to split across the two systems. The VCMTS SmartNIC will need to perform the packet copy function, plus an additional function to tunnel the copied packets to the VSG using the Virtual eXtensible LAN (VXLAN) [2] tunneling protocol. The VSG will then need to receive and decapsulate the tunneled packets, translate the VLAN ID as needed, and forward the packet to the VSG software application. A SmartNIC with hardware offload support for the VXLAN protocol decapsulation function is needed. This SmartNIC also needs to provide a line-rate VLAN translation function with hardware offload support.

4. Integrated VCMTS-VSG Solution

4.1. Integrated System

The core of the VCMTS is a collection of software components deployed in a highly available cluster. Kubernetes is an integral part of the system, playing a key role in the management, monitoring, and provisioning of various applications. Kubernetes isolates components from each other into multiple pod instances, enabling horizontal scaling.

Packet processing applications based on the DPDK library, such as the VCMTS, must have direct access to the network to minimize latency and increase throughput without intermediate software components. The SmartNIC provides two mechanisms for network direct access.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The first mechanism is SR-IOV, which allows the splitting of a single physical PCIe device into multiple virtual PCIe devices. Each packet processing application pod is allocated with one or more VFs, used as a networking interface.

The second network direct access mechanism provided by the SmartNIC is traffic steering, the ability to maintain steering rules that determine which traffic will be forwarded to each VF.

NIC traffic steering is either vendor-specific or based on standard APIs, such as rte_flow, SwitchDev, or OVS. SwitchDev is an infrastructure construct provided by recent Linux kernel versions which enables traffic steering configuration via the "tc" utility. The result of "tc" commands are the configured traffic steering rules. A traffic steering rule consists of a matcher, which matches specific packets and actions, which defines what the NIC will do with the matched packets. There are two classes of steering rules: receive and transmit rules. Receive rules define which packets will be sent from the NIC to a specific VF; transmit rules define how to process packets sent from the VF to the NIC.

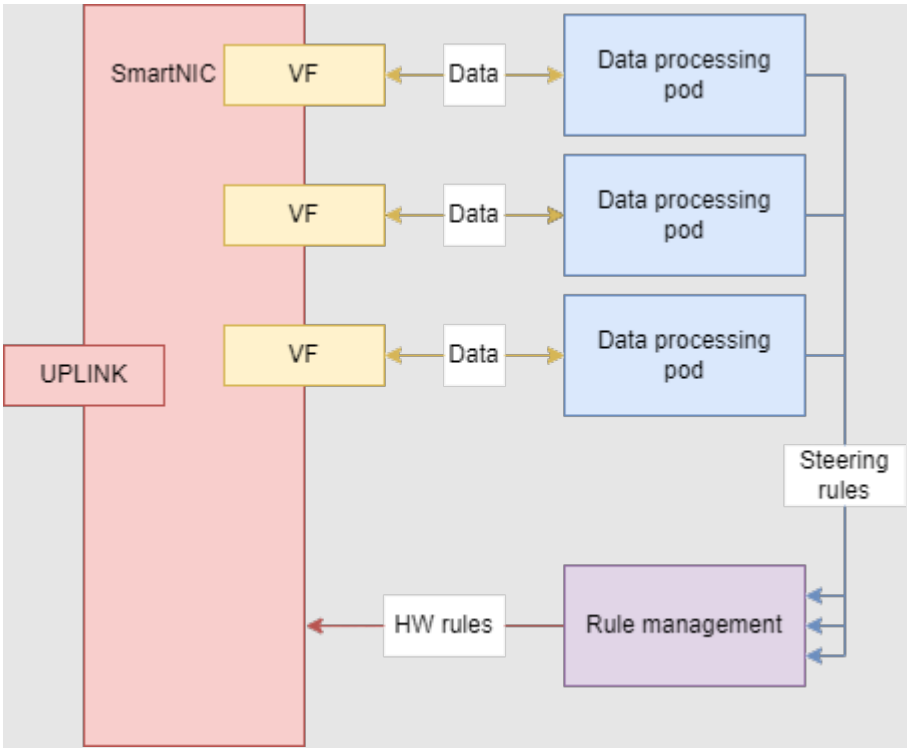


Figure 7 - Traffic Steering in VCMTS

With the development of the next-generation Edge Cloud Platform architecture solution, one of the new requirements has been to support any scale. Instead of relying on the VSG server being in a specific network location and mirroring specific ports, the VCMTS now takes over the mirroring of all traffic towards the VSG server instances, thus reducing the dependency on the network topology almost completely.

The VSG application requires all downstream and upstream traffic to be copied, encapsulated, and sent to the VSG server. Steering rules are used to separate, mirror, and encapsulate traffic into a VXLAN tunnel toward the VSG server.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

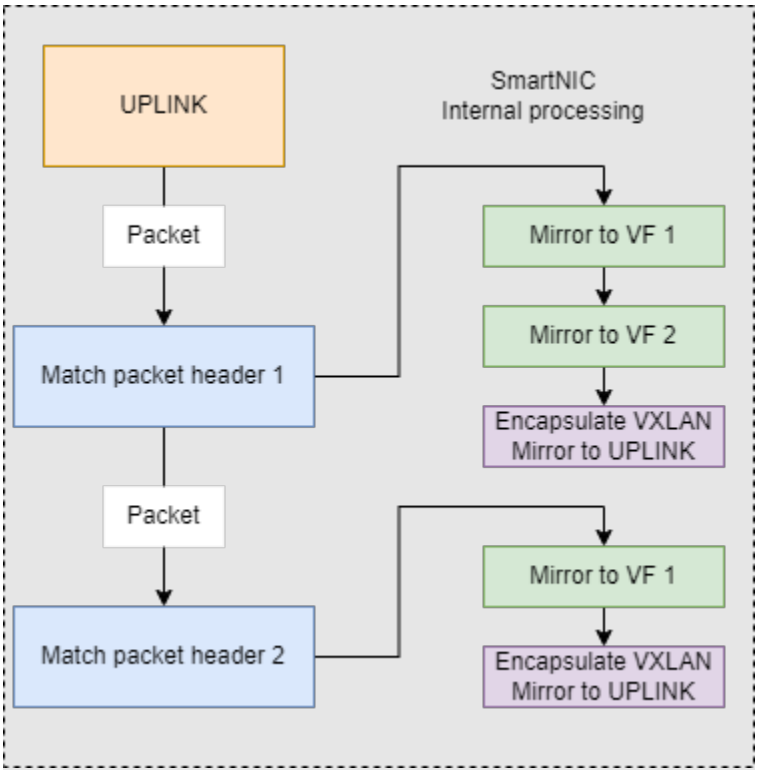


Figure 8 - VSG Replication & Encapsulation Rules in VCMTS

The mirroring and encapsulation could be implemented with software on the CPU. The memory and computing required for mirroring and encapsulating the entire VCMTS traffic is extremely costly, even when run on modern CPUs, with a potentially considerable performance hit.

With SmartNIC offloading enabled, there is no CPU performance penalty. The potential performance improvement is a clear driver towards implementing packet processing applications with SmartNIC HW offloading.

Potential disadvantages with SmartNIC usage are few but should be mentioned:

- Encountering vendor-specific behavior is probable, even when using standard APIs or open-source implementations. Configuring the same feature with identical API calls may result in different behavior across multiple NIC vendor implementations.
- The ability to debug issues is limited, a natural consequence of using hardware offloads. Debugging hardware is harder than debugging a software-only implementation.

4.2. Ability to Scale the System

The deployment philosophy is to “build it once and deploy it at any scale.” Sizing and scaling are part of the deployment plan to match the number of VCMTS pods that can be served by a single VSG bare metal server. If more VSG capacity is needed, we simply add more VSG bare metal servers and connect it to the switch fabric.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

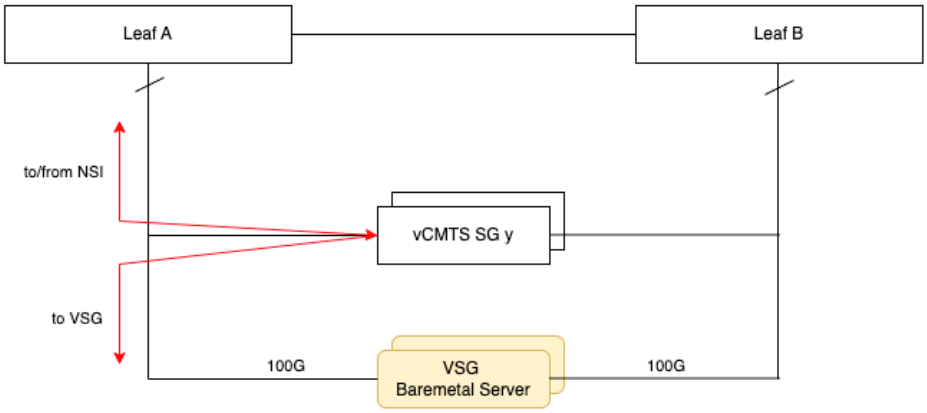


Figure 9 - Baremetal VSG Deployment View

The VCMTS pod to the VSG bare metal server mapping can be dynamically provisioned by the service orchestration layer. Together, the VCMTS pods and the steering operator are able to manage the SmartNIC traffic steering rules and mirroring functions. The ability to steer and mirror VCMTS traffic is, in essence, the ability to partition the service; the system scale is no longer bound by a single VSG bare metal server's limitation. Therefore, the system scales horizontally.

4.3. Future Plans for VSG Containerization

The Edge Cloud Platform can have a tremendous amount of computing and network resources. The next level of sophistication is to build the system such that it is elastic and can auto-scale. These aspects enable the systems to automatically adjust the provisioned network and compute resources to match the current demand.

Future ideas for integrating the VSG with the VCMTS may involve "containerizing" the VSG application to run on the Edge Cloud Platform. The VSG application function may be disaggregated into its constituent functions, potentially with each function running in its own container. There are several possibilities on how to distribute the VSG container workload across the control and data compute nodes of the edge cloud. The key requirements for this solution are:

- The ability to create and destroy the VSG application containers automatically and/or on request, scaling the system as demand grows and subsides.
- The ability to copy and forward data plane packets to the appropriate VSG application container
- The ability to copy and forward VSG control plane packets amongst the VSG application containers

One potential solution is shown in Figure 10 - Containerized VSG & VCMTS Deployment View. In this solution, the VSG pod supports all the needed VSG application containers running using available compute from a VCMTS "worker" node. The VSG pod supports both the VCMTS pod co-resident on the same compute and a VCMTS pod running on separate compute nodes. A variation on this solution is to only copy the packet headers instead of the full packet. This potential solution will reduce the overall link input/output bandwidth needed to support the packet copy function.

The key takeaway for these solutions is the capability of the SmartNIC to modify, copy, redirect, or forward packets both across and within the supported compute nodes. The SmartNIC in use also provides

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

packet encryption using hardware offload, so it is possible to encrypt the VXLAN tunneled traffic between the VCMTS and VSG if desired.

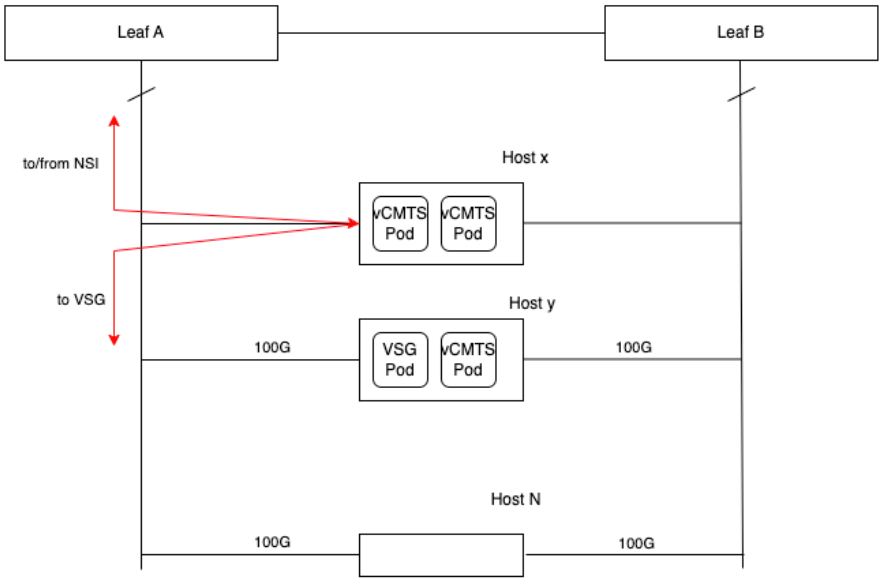


Figure 10 - Containerized VSG & VCMTS Deployment View

5. Conclusion

The VCMTS and the VSG platforms may be developed to achieve new functionality to meet their respective goals. They are able to scale to a point where the next step must be taken together to continue to grow, which is the creation of an Edge Cloud Platform. This platform is modeled after a traditional elastic cloud infrastructure but modified to the unique requirements of the edge and access networks. As such, it must be capable of hyper-scaling automatically and horizontally while maintaining the performance and demands of all the services provided by its individual components.

In this paper, the authors have laid out:

- The function and architecture of the vCMTS and VSG systems
- Introduced the SmartNIC and its capabilities
- Describe challenges that may require SmartNIC functionality
- A proposed architecture for implementing a horizontally scaled solution
- Additional future improvements to move toward and improve the Edge Cloud Platform

The edge cloud is the future of the access network. The authors believe the potential solution documented in this paper is not only the right solution for hyper-scaling the VSG but has defined a roadmap for incorporating other edge and access services into that same platform seamlessly and indefinitely scalable.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

ACM	Association for Compute Machinery
AR	Aggregation Router
COTS	Commercial-Off-The-Shelf
CCAP	Converged Cable Access Platform
CPU	Central Processing Unit
DAA	Distributed Access Architecture
DOCSIS	Data Over Cable System Interface Specification
DP	Data Plane
FDX	Full Duplex DOCSIS
Gbps	Gigabit per second
HSD	High Speed-Data
IEEE	Institute of Electrical and Electronics Engineers
LAN	Local Area Network
M-CMTS	Modular-Cable Modem Termination System
MHA	Modular Head-end Architecture
NFV	Network Function Virtualization
NIC	Network Interface Card
OFDM	Orthogonal Frequency Division Multiplexing
OVS	Open vSwitch
PCIe	Peripheral Component Interconnect express
PF	Physical Function
QAT	Quick Assist Technology
QOE	Quality of Experience
RF	Radio Frequency
RUR	Residential U-ring Router
SDN	Software Defined Network
SG	Service Group
SR-IOV	Single Root - Input / Output Virtualization
TC	Traffic Control
TCP	Transmission Control Protocol
UBB	Usage Based Billing
UDP	User Datagram Protocol
VCMTS	Virtual Cable Modem Termination System
VF	Virtual Function
VSG	Virtual Services Gateway
VXLAN	Virtual eXtensible LAN
XRM	Xfinity Resource Manager

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

[1] Data-Over-Cable Service Interface Specifications MHAv2 Remote Downstream External PHY Interface Specification, CM-SP-R-DEPI-I16-210804, August 4, 2021, Cable Television Laboratories, Inc.

[2] IETF RFC 7348 https://datatracker.ietf.org/doc/html/rfc7348

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Implementing Nanogrid Into the Critical Facility

Improving Reliability, Efficiency, and Future Adaptability

A Technical Paper prepared for SCTE by

Rahul Khandekar, Ph.D.
Director, Research and Advanced Development
EnerSys

Ron Slutter
Senior Technical Support
EnerSys

Francisco Paz, Ph.D.
Lead Systems Engineer
EnerSys

Dan Cooper
Sr. Technology Executive
Greenside LLC

Greg Laughlin
Strategic Marketing Manager
EnerSys

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
   - 1.1. Critical Facility Power Elements ... 4
   - 1.2. Figure 1Utility Power Outage ... 5
2. Reliability Weaknesses with Today's Critical Facility Architecture ... 5
   - 2.1. Generator Failure ... 5
   - 2.2. Automatic Transfer Switch Failure ... 6
   - 2.3. Loss of HVAC ... 6
   - 2.4. Battery Reliability ... 6
3. Evolving Critical Facility Changes Will Add to Reliability Concerns ... 6
   - 3.1. Increasing Energy Demands ... 6
   - 3.2. Higher Loads Make HVAC Critical ... 6
   - 3.3. Utility Grid Strains Leading to Increased Reliance on the ATS ... 7
4. Implementing Nanogrid to Address Power Reliability ... 7
   - 4.1. Transition #1: Distributed Energy Resource Connectivity ... 8
   - 4.2. Transition #2: Protected 380VDC bus ... 9
   - 4.3. Transition #3 / Final Proposed Architecture: Protected DC Bus ... 9
5. Additional Benefits in Efficiency and Adaptability ... 11
   - 5.1. Efficiency Gains by Using High Voltage DC Distribution ... 11
   - 5.2. Distributed Energy Resources Adaptability ... 11
6. Conclusions ... 12

Abbreviations ... 13

Bibliography & References ... 14

List of Figures

Title **Page Number**

Figure 1 - Example of Present Powering Architecture in a Critical Facility ... 4
Figure 2 – Critical Facility Outage Time Diagram ... 5
Figure 3 – Average Duration Total Annual Utility Power Interruptions, United States (2013-2020)[ii] ... 7
Figure 4 - Transition Architecture #1 ... 8
Figure 5 - Transition Architecture #2 ... 9
Figure 6 - Transition #3 / Final Proposed Architecture ... 10
Figure 7 - Critical Facility Outage Time Diagram with the Proposed Architecture ... 11

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

As cable broadband networks evolve to meet capacity and performance expectations – pushing legacy functions outward toward the edge – critical facilities are challenged to maintain reliability, forcing operators to revisit their power architectures. The ever-growing reliance on high-speed network connectivity are driving the transition to distributed access architectures, with Remote-PHY Devices (RPDs) in edge nodes taking on much of the work that had been traditionally performed in critical facilities. High-density virtual CMTS (vCMTS) and edge gateway server racks are replacing legacy equipment in the critical facility, increasing network capacity and performance, but at the cost of increased energy consumption and associated thermal loading. Heating Ventilation and Air Conditioning (HVAC) systems, not traditionally backed up by batteries, may need to be reconsidered as a critical load. As described in SCTE 184, "Modern equipment densities can overheat an environment in less than an hour without adequate cooling making any additional installed capacity unusable."[i]

At the same time, external threats to the utility grid – extreme weather events, electric vehicles, heat pumps and air conditioning units – are leading to more frequent and longer utility outages, adding to grid instability[ii], and placing more reliance on the effectiveness of the power backup system. This combination of increasing grid instability, reliability expectations and added thermal load may drive operators to revisit their existing critical facility (CF) power architectures.

In addition to increasing reliability, operators are focused on meeting their corporate sustainability goals toward reducing overall energy consumption and reliance on fossil fuel sources[iii,iv]. Improving power efficiencies in the critical facilities and introducing distributed energy resources can help these efforts[v]. A nanogrid architecture can allow the use of distributed energy resources (DERs) such as wind or solar to supplement utility power, allowing operators to realize additional cost benefits by avoiding peak usage utility rates or selling power back to the grid[vi].

In this paper we will:

1) Walk through a complete critical facility outage with today's architectures.
2) Identify threats to critical facility reliability.
3) Look at how changes to critical facilities will lead to additional reliability challenges.
4) Explore nanogrid architectural steps that could be implemented to increase reliability.
5) List additional benefits from moving to the proposed nanogrid architecture.
6) Anatomy of a Critical Facility Blackout.

Figure 1 shows a diagram of a typical high-level medium sized cable broadband critical facility (CF). Utility power supports the entire facility through the Essential AC bus, with on-site generators (GEN 1 and GEN 2) to maintain operation during utility outages. Figure 1 shows two Automatic Transfer Switches (ATSs): the Utility-to-Generator (UTG) and Generator-to-Generator (GTG). The UTG detects the loss of the main power source (the utility) and automatically switches to power from the generators. The GTG selects which generator to connect. In the event of a prolonged outage, operators can bring a portable generator (PORT GEN) and, using the Manual Transfer Switch (MTS), start supplying energy to the CF.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

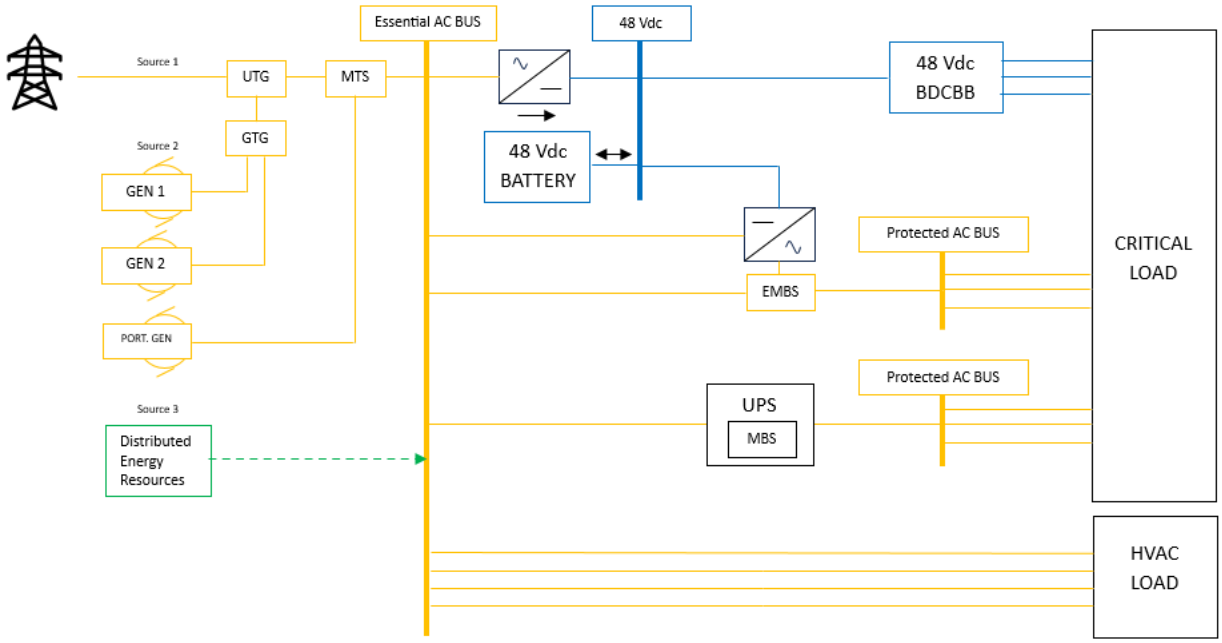


Figure 1 - Example of Present Powering Architecture in a Critical Facility

1.1. Critical Facility Power Elements

The Essential AC bus is the common point from where the different loads throughout the critical facility are fed, with various power conversions and battery backup. The Essential AC bus also supplies the Heating, Ventilation, and Air Conditioning (HVAC) load, which oversees keeping the operating temperature of the CF. In this architecture, the HVAC system is not supported by batteries. If both the utility and the generators are lost, the HVAC system ceases to work.

Critical loads (CL) include all IT equipment such as servers, routers and CMTS racks whose operation is deemed absolutely necessary for maintaining network operations. In the present architecture, the CL must be supported for a typical period of 2 to 4 hours when utility power and the generators fail. To support the CL, there is a 48VDC bus supported by battery energy storage. A unidirectional AC to DC converter (rectifier) charges the batteries and provides power to the CLs.

Most CLs accept 48VDC as input power and can be fed directly from the 48VDC bus through a Battery Distribution Circuit Breaker Bay (BDCBB). This provides very high source reliability and continuity.

A portion of CLs must run from AC. To supply these loads and support them in case of outages, the CF has a Protected AC Bus that is supplied from the 48VDC bus through a unidirectional DC to AC converter (inverter). During an outage, the inverter supplies power to the load until the batteries are exhausted. The connection includes an External Maintenance Bypass Switch (EMBS) which allows the inverter to be disconnected for maintenance.

Finally, the CF designers may want to include Distributed Energy Resources (DERs) such as solar PV or wind to help achieve the company's sustainability goals, lower the energy bills, and in some cases extend the runtime of the batteries. The Essential AC Bus also provides the point of common coupling for the DERs. However, integrating diverse DERs into the AC bus is more complex than integrating them using DC, due to the need to match the AC frequency, phase, and voltage[vi].

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1.2. Figure 1Utility Power Outage

Figure 2 shows a time diagram of a utility power outage followed by a loss of the generator backup. At the beginning, the HVAC and CL are running from the Utility AC Grid. At instance (1), the CF loses power from the Utility, the ATS (the UTG) connects to the Generator, and the Generators start supporting the HVAC and the CL. During this period, the HVAC regulates the temperature at a safe level. After some time, at instance (2), the Generator is lost, and the CL must run from the battery storage. This can last for between 2 to 4 hours, until the batteries are exhausted in instance (3). However, in this condition, the HVAC is not working, and the temperature starts to increase since cooling is only due to the natural air flow in the CF. Depending on the power density of the CF, within a short period of time it will reach the Critical Temperature, at which point the CL will be dropped.

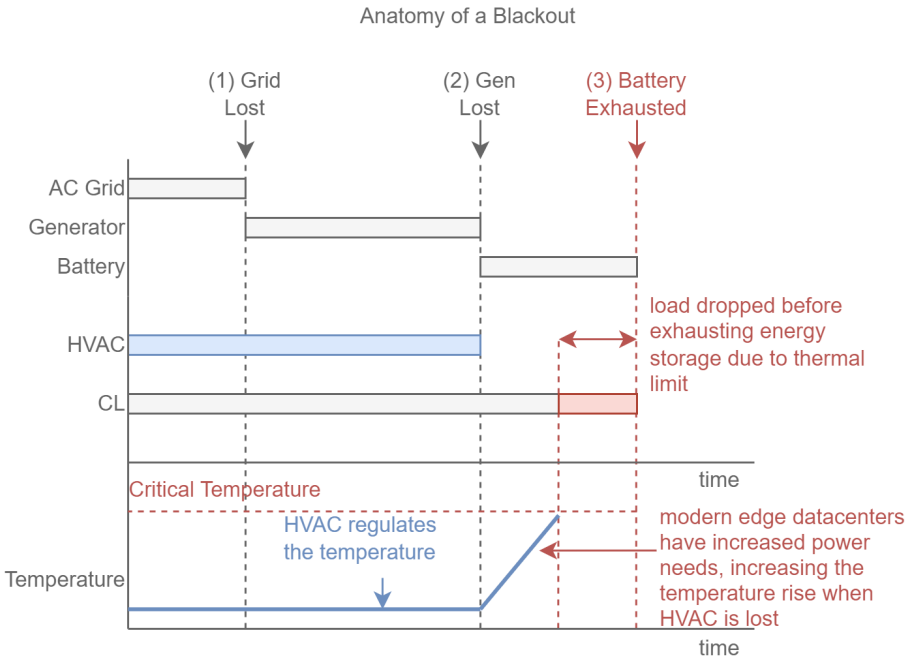


Figure 2 – Critical Facility Outage Time Diagram

2. Reliability Weaknesses with Today's Critical Facility Architecture

The current CF architecture shown in Figure 1 has been serving well for years, but new emerging technologies are pushing the limits. The current architecture can have several critical risk factors that compromise its reliability: the generator, the ATSs (GTG and UTG), the HVAC, and the batteries.

2.1. Generator Failure

The generator is made of thousands of parts and pieces, all required to function properly to provide backup power. As such, the generator reliability can be a risk factor.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.2. Automatic Transfer Switch Failure

As shown in Figure 1, the Utility-to -Generator (UTG) detects the loss of the Utility Power and automatically connects to the generators (GEN 1 or GEN 2). If the UTG fails and does not perform its function, there is no redundant component that can connect to the generators. At this point, the CL starts drawing power from the 48VDC battery bus. However, the HVAC, not supported by the 48VDC battery bus, will not be running. This greatly increases the chance of either exhausting the batteries or reaching critical temperature before the utility power is restored or the portable generator is connected. The inverse problem can also happen: the UTG can remain in the GEN position, even when the Utility outage is resolved, causing the system to exhaust the fuel and batteries, resulting in dropping the CL.

A similar problem can be observed with the Generator-to-Generator (GTG) ATS. The GTG may make the wrong decision (connecting to GEN 1 when it should connect to GEN 2), or it may fail to perform the action. Transfer Switches are complex devices with mechanical, electrical, and software components. They have a relatively high failure rate.

2.3. Loss of HVAC

Figure 1 indicates how the HVAC is connected to the Essential AC Bus. In this architecture, energy flows into the Essential AC Bus only from the Utility, GEN1, GEN2, or PORT GEN. This is important since it implies that the HVAC system is not supported by the batteries. To support the HVAC system from the batteries, the AC/DC battery charger should be bidirectional and there should be many more batteries.

A loss of the HVAC system does not immediately trigger a failure based on temperature. Forced air circulation can delay temperature increase long enough to connect the PORT GEN. However, as power density increases, loss of the HVAC system will result in the failure of the CL.

2.4. Battery Reliability

Battery reliability is high when batteries are properly maintained and in a controlled temperature environment.

3. Evolving Critical Facility Changes Will Add to Reliability Concerns

3.1. Increasing Energy Demands

As an example, the CL energy requirements per rack for a medium facility may grow by 40% in the upcoming years. Since the HVAC continues to be sized at a one-to-one ratio of cooling to power, the facility will need 40% additional HVAC loads.

3.2. Higher Loads Make HVAC Critical

Modern edge datacenters are increasing in power and power density. In this condition, there is more heat to be removed from the CF. Current load densities will raise the temperature above the critical temperature threshold in less than one hour without support from the HVAC system[vii]. The CL will be dropped before the Battery Energy Storage System (BESS) and DERs are exhausted. Given this scenario, it becomes clear that the HVAC needs to become part of the CL. The advanced infrastructure will need to support the HVAC load. Supporting the HVAC load requires higher DC voltages to be able to distribute available power from the BESS to the HVAC.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.3. Utility Grid Strains Leading to Increased Reliance on the ATS

The CFs will have to perform under significantly more stress due to the increase in Utility power outages and at significantly higher reliability levels. Data from the U.S. Energy Information Administration (shown in Figure 3) shows that utility outages are increasing. Moreover, this increase is driven by major events (snowstorms, hurricanes, and wildfires), which are expected to keep increasing in frequency and severity. Each outage event is one chance for the ATS to fail, increasing the probability of a failure.

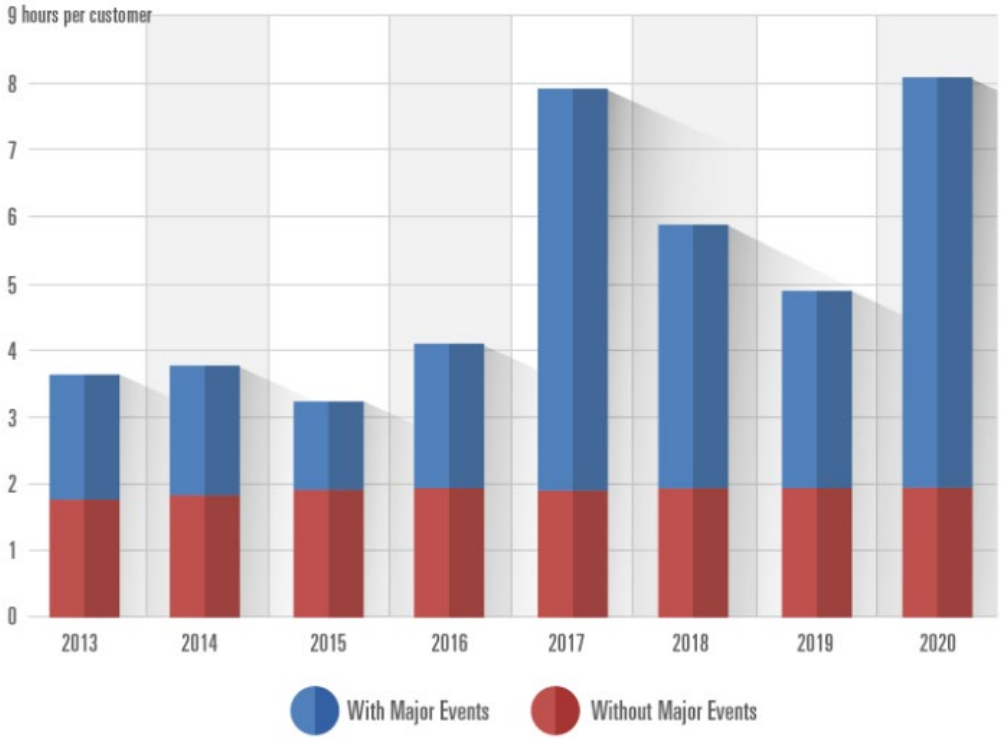


Figure 3 – Average Duration Total Annual Utility Power Interruptions, United States (2013-2020)[ii]

Other industry trends such as autonomous vehicles, IoT, smart city applications, and industry 4.0 solutions push the reliance on the internet infrastructure. Future applications that rely on an underlying, high bandwidth, low latency data network will impose further restrictions on the reliability of the CF.

4. Implementing Nanogrid to Address Power Reliability

This section outlines a new proposed CF architecture to overcome the reliability challenges, increase efficiency, and simplify integration with DER. The proposed architecture is based on a DC nanogrid concept using a higher DC voltage essential bus. This section also presents three transition architectures to illustrate the possible evolution from the current architecture to the final proposed architecture. A full migration to the final architecture would be required to benefit from the proposed reliability increases.

In the context of this paper, a nanogrid utilizes the common concepts found in a microgrid but for a building-size microgrid.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.1. Transition #1: Distributed Energy Resource Connectivity

Figure 4 shows a diagram of the Transition Architecture #1. This architecture shows several significant changes with respect to the architecture in Figure 1. Namely:

1. The AC to DC converter linking the Essential AC bus and the 48VDC bus is now bidirectional. This allows the DERs to be connected to the 48VDC bus and then transfer their energy to the essential AC bus.
2. There is now also a high voltage DC bus (380V in this paper) supported by 380V batteries. This bus can support some native 380VDC CLs which are now emerging. The 380VDC bus is linked to the 48Vdc bus through a bidirectional DC to DC converter, which allows it to run from the 48VDC battery backup and even support it. The loads are directly connected using a 380VDC BDCBB.
3. The DERs can now be connected either to the Essential AC bus, the 48VDC bus, or the 380VDC bus. Integrating DER at the 380VDC level is straightforward as it does not require synchronization and uses less power conversion stages (e.g., typical applications of PV use 600VDC or 1000VDC arrays, which are allowed by the NEC[viii], integrating at the 380VDC level requires just a simple DC to DC converter).

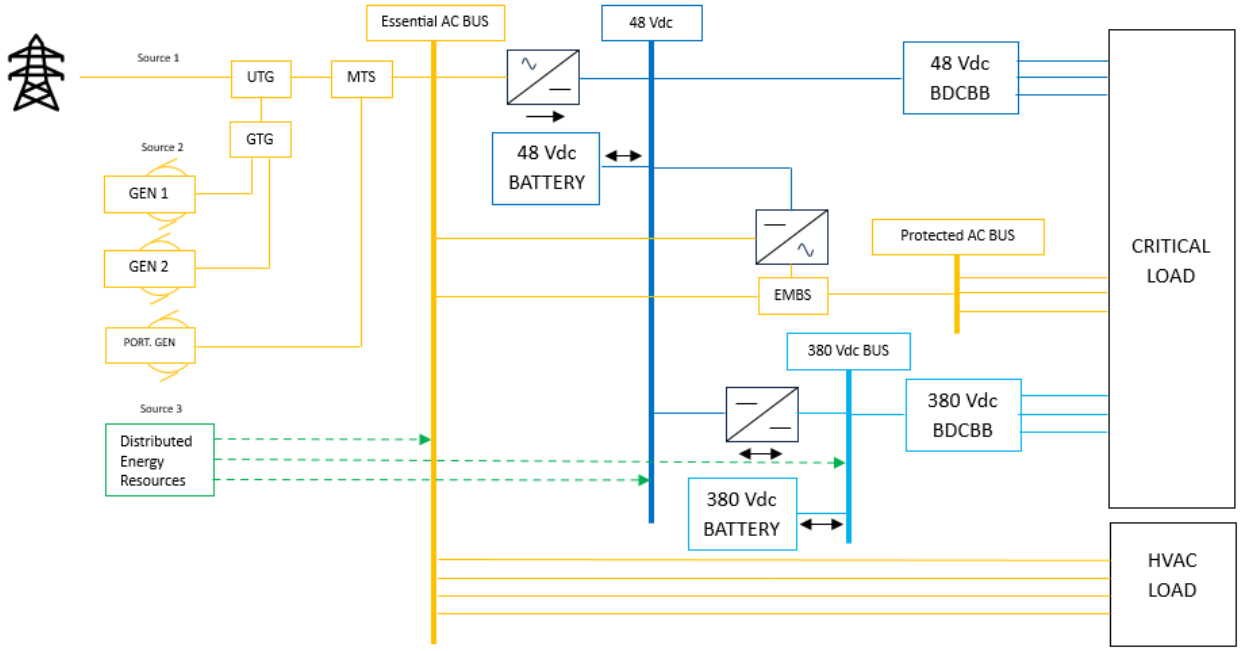


Figure 4 - Transition Architecture #1

Although the Transition architecture #1 provides improvements in efficiency and allows better integration of DERs, it still has the same critical reliability problems: the HVAC is not supported by the batteries and the ATS is a critical risk factor. Furthermore, integrating DERs at the 380VDC level can result in lower efficiency if the energy must be transferred to the Essential AC bus for utilization.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2. Transition #2: Protected 380VDC bus

Figure 5 shows a diagram of Transitional architecture #2. This architecture shows a significant change with respect to Transitional Architecture #1 in Figure 4: The high voltage DC bus (380VDC in this case) is fed from the Essential AC bus, instead of the 48VDC bus, using a bidirectional AC to DC converter. This arrangement reduces the number of power conversions needed to feed the 380V load and charge the batteries, increasing the efficiency of the system. Furthermore, it reduces the number of power conversion stages needed to feed the energy from the DERs into the Essential AC bus.

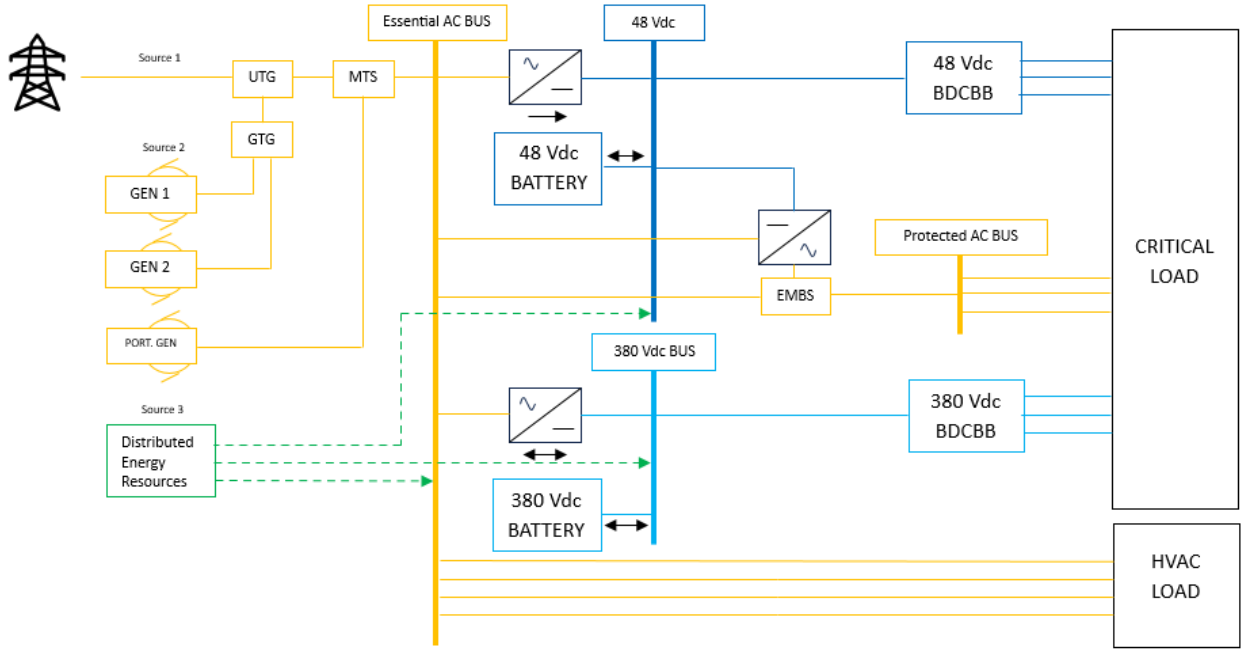


Figure 5 - Transition Architecture #2

Although the Transition Architecture #2 provides improvements in efficiency and integration of DERs compared with Transition Architecture #1, it still has the same critical reliability problems: the HVAC is not supported by the batteries due to capacity concerns and the ATS is a critical risk factor.

4.3. Transition #3 / Final Proposed Architecture: Protected DC Bus

Figure 6 shows a diagram of the proposed advanced nanogrid CF architecture. Compared with the architectures in Figure 1, Figure 4, and Figure 5, the proposed architecture includes the following changes:

1. This advanced nanogrid architecture implements the creation of a protected 380VDC bus, where all the sources and loads are connected. This replaces the essential AC bus.
2. Multiple energy sources:
    a. Source 1: The utility connection will be transposed using bidirectional rectification. It will be modular and use redundant components like the rectifier stages in Figure 1. A contactor could be provided which could provide islanding of the electrical system if desired.
    b. Source 2: The Generator(s) (GEN1, GEN2, and PORT GEN, if available) connect to the system using a unidirectional AC to DC converter (a simple rectifier). It will be modular

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

and use redundant components like the rectifier stages in Figure 1. The output of the rectifier is DC, eliminating the need for AC synchronization.

c. Source 3: Utilizing a Battery Energy Storage System (BESS) which supports the protected 380VDC bus, potentially using a DC-to-DC converter. It will be modular and use redundant components like the rectifier stages in Figure 1.

d. Source 4: The DERs are connected directly to the protected 380VDC bus.

3. The HVAC load is fed from the protected 380VDC bus using a DC to AC inverter system, and therefore is supported by multiple sources of energy. Alternatively, the HVAC could use 380VDC directly as equipment becomes available. This allows the cooling function to continue.
4. An ATS is no longer required; the generators are attached to the protected 380VDC bus by using rectification.

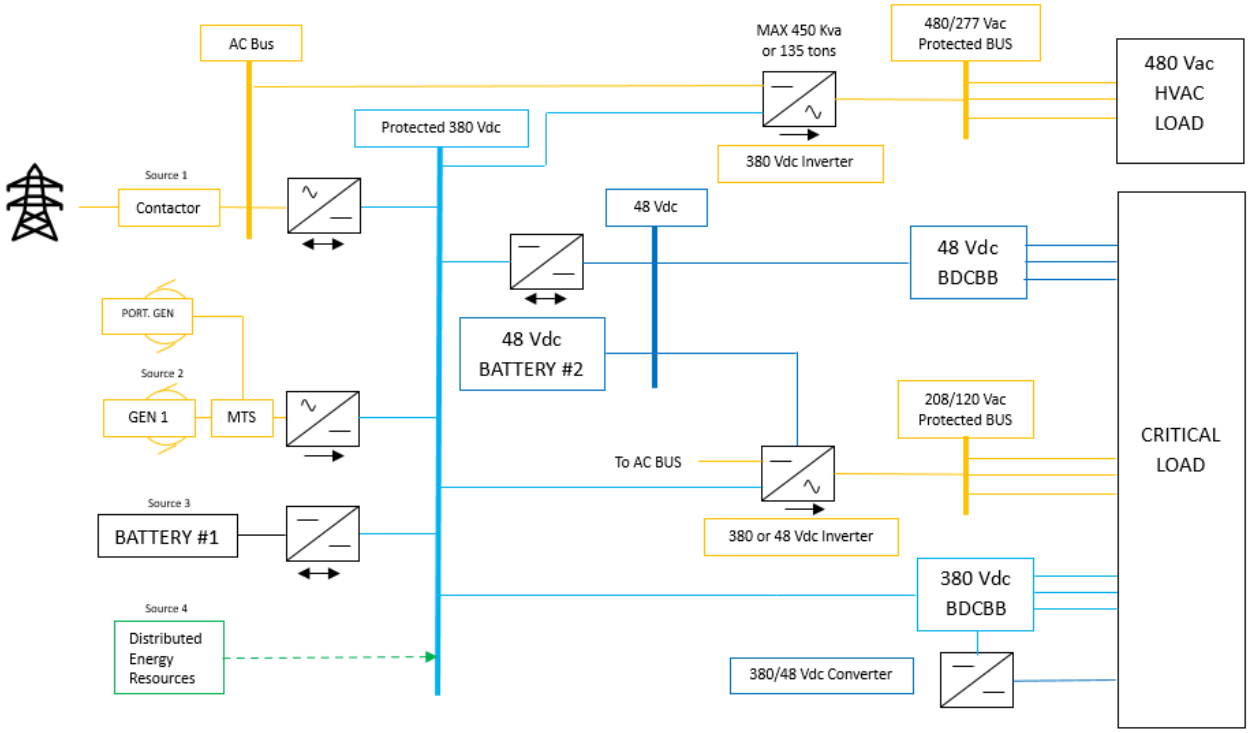


Figure 6 - Transition #3 / Final Proposed Architecture

The proposed nanogrid architecture addresses the major critical reliability factors – supporting the CL while providing the required cooling.

Figure 7 shows an updated time diagram of a utility blackout. Using the proposed architecture, the HVAC system is supported by the BESS and DERs. If the generator is lost in (2), the temperature of the CF is still controlled by HVAC system. This allows the facility to use the entirety of the BESS and DER. If DERs are available, they can help support the load and batteries during the blackout to extend the runtime of the CF.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

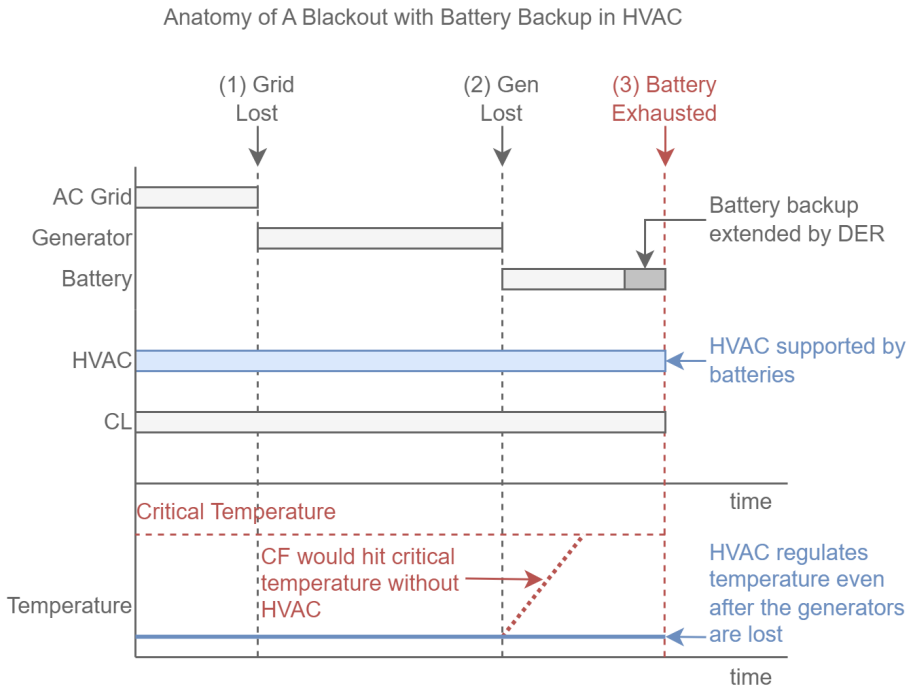


Figure 7 - Critical Facility Outage Time Diagram with the Proposed Architecture

5. Additional Benefits in Efficiency and Adaptability

The proposed nanogrid architecture can directly mitigate the major critical reliability risks associated with CFs; however, this is not the only benefit. Using the 380VDC bus architecture provides reduced losses in power transmission and simplifies integration of DERs.

5.1. Efficiency Gains by Using High Voltage DC Distribution

Using a protected 380VDC bus, allows the power distribution to be done at higher voltage (380VDC compared to 48VDC). Using 380VDC has the significant advantage of reducing losses in the DC distribution.

Using a 380VDC distribution can significantly reduce the number of conductors needed to route power to the loads. For example, an arrangement of cables (72 750mcm cables) in a tray can carry 359kW at 48VDC or 2.5MW at 380VDC, almost 7 times more power. Furthermore, a smaller arrangement of cables (12 750mcm) can carry over 85% more power at 380VDC than the 72-cable arrangement at 48VDC[ix]. This is an 83% reduction in conductor costs while increasing power delivered by 85%.

5.2. Distributed Energy Resources Adaptability

Integrating DERs at the 380VDC level is significantly easier than integrating using an AC bus and more efficient than integrating at a 48VDC level. This improved integration of DERs can allow operators to achieve their sustainability goals by using more renewable energy sources.

Using the bidirectional AC-to-DC grid interface, the CF can interface the grid and provide power from the renewables to support it. Typical applications may include frequency support and reactive power compensation.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Conclusions

This paper presented an analysis of existing critical facilities for cable broadband applications from the point of view of energy consumption and reliability. It describes the current architecture and impacts on the facility performance during an outage. Two critical areas for improved reliability were identified:

1) Adding battery backup to support the HVAC load.
2) Reducing the reliance on the ATS.

This paper proposes using a high voltage DC nanogrid architecture for the critical facility power system. Replacing the essential AC bus with a high voltage protected DC bus (380VDC in this example) not only allows more efficient power distribution, but also simplifies integration of the battery energy storage system and distributed energy resources.

Incorporating DERs can also help operators meet their sustainability goals by reducing utility consumption and implementing other operating modes, such as peak shaving and demand side management.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

BDCBB	Battery Distribution Circuit Breaker Bay
BDFB	Battery Distribution Fuse Bay
BESS	Battery Energy Storage System
CF	Critical Facility
CI	Critical Infrastructure
CL	Critical Load
CMTS	Cable Modem Termination System
DCDC	DC to DC voltage transformation
DER	Distributed Energy Resources
GTG	Generator-to-Generator
HVAC	Heating Ventilation and Air Conditioning
MTS	Manual Transfer Switch
PORT GEN	Portable Generator
PFC	Power Factor Correction
RPD	Remote-PHY Device
UPS	Uninterruptible Power Supply.
UTG	Utility to Generator

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

[i] Society of Cable Telecommunications Engineers, Inc. (SCTE), "SCTE 184 2022: SCTE Energy Management Design, Construction and Operational Practices for Cable Facilities," Exton, PA, USA, Oct. 2022.
[ii] U.S. energy Information Administration, 2021, Today in Energy, Viewed 8/1/2022, **www.eia.gov/todayinenergy/U.S.-electricity-customers-experienced-eight-hours-of-power-interruptions-in-2020**
[iii]Climate Impact Partners, FORTUNE GLOBAL 500 CLIMATE COMMITMENTS, accessed Aug. 2023, https://www.climateimpact.com/news-insights/fortune-global-500-climate-commitments/
[iv] Comcast, *Carbon Neutral by 2035,*Accessed Aug. 2023, https://corporate.comcast.com/impact/environment
[v] V. Vossos, R. Heard, E. Mannarino, K. Johnson, D. Gerber, M. Khuttar, G. Kloss, B. Nordman, R. Brown (prepared for California Energy Commission), "Direct Current as an Integrating and Enabling Platform for Zero-Net Energy Buildings", April 2019, CEC-500-2019-038
[vi] D. Boroyevich, I. Cvetković, D. Dong, R. Burgos, F. Wang and F. Lee, "Future electronic power distribution systems a contemplative view," 2010 12th International Conference on Optimization of Electrical and Electronic Equipment, Brasov, Romania, 2010, pp. 1369-1380, doi: 10.1109/OPTIM.2010.5510477.
[vii] Society of Cable Telecommunications Engineers, Inc. (SCTE), "SCTE 184 2022: SCTE Energy Management Design, Construction and Operational Practices for Cable Facilities," Exton, PA, USA, Oct. 2022.
[viii] NFPA 70 National Electrical Code (NEC), 2023 Edition, Article 690.7.
[ix] R. Ambriz and M. Kania, "A service provider's decision to move from 48V to 380V powering: The problem statement, technical assessment, financial analysis and practical implementation plan," 2016 IEEE International Telecommunications Energy Conference (INTELEC), Austin, TX, USA, 2016, pp. 1-7, doi: 10.1109/INTLEC.2016.7749117.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Intelligent Outside Plant Power Operations with Machine Learning

A Technical Paper prepared for SCTE by

Matthew Stehman
Principal Machine Learning Engineer
Comcast
1800 Arch St, Philadelphia, PA
Matthew_Stehman@comcast.com

Chris D'Andrea
Data Scientist
Comcast
1800 Arch St, Philadelphia, PA
Christopher_D'andrea@comcast.com

Ilana Weinstein
Data Scientist
Comcast
1800 Arch St, Philadelphia, PA
ilana_weinstein@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. The Life Cycle of a Commercial Power Outage ... 3
   - 2.1. Outage Detection ... 3
   - 2.2. Workforce Planning and Maintenance Activity ... 4
   - 2.3. Outage Resolution and Closure ... 5
3. Machine Learning on Historical Outages ... 6
   - 3.1. Creating a Historical Outage Data Base ... 6
   - 3.2. Power Supply Battery Run Time Modeling ... 7
   - 3.3. Customer Impact Modeling ... 8
4. CPO Dynamics and Long Term Planning ... 10
   - 4.1. Learning Patterns from CPOs ... 10
   - 4.2. AI Enhanced CPO Operations and Planning ... 11
5. Conclusion ... 12

Abbreviations ... 13

Bibliography & References ... 13

List of Figures

Title **Page Number**

Figure 1 - Detecting a CPO ... 4
Figure 2 - CPO impact assesments ... 5
Figure 3 - Outage geospatial example ... 7
Figure 4 - Battery run time model performance: A.) Distributions of model prediction errors B.) Raw predictions of each model (colored dots) compared to the actual run times (black dots) ... 8
Figure 5 - Customer correlation to power supply status: A.) customers online while PS is running on battery B.) customers offline when PS batteries deplete. ... 10
Figure 6 - Timeline of a CPO ... 10
Figure 7 - AI enhanced CPO resolution and OSP power asset management ... 11

List of Tables

Title **Page Number**

Table 1 - Actions for various outage states ... 9

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

As cable providers continue to develop and deploy 10G technologies, customer experience expectations are also growing in stride. These more reliable networks and faster speeds mean that even the smallest outages have a high impact on daily life. In the realm of outside plant power, the power supplies are responsible for powering the plant as well as providing back up battery power and keeping the cable plant online during commercial power outages (CPOs). Artificial intelligence (AI) can be integrated with real-time telemetry to plan resource deployments based on the expected performance of the batteries combined with predictions on customer impacts during the outage. While CPOs are likely unavoidable for multiple systems operators (MSOs), advanced techniques can be employed to mitigate potential connectivity outages during such events.

The authors will discuss Comcast's approach to minimizing customer impact during CPOs using Artificial Intelligence and Machine Learning (ML). Big Data tools and pipelines are employed to gather power supply telemetry and customer connectivity data for historical CPOs. Predictive models are built to learn from these historical events and can provide insights during active CPOs. These predictive models can then be integrated into field operations platforms to help plan work and even proactively message customers. This paper will illustrate the power of using historical power outage data with AI/ML frameworks to support the connectivity expectations of a 10G network.

2. The Life Cycle of a Commercial Power Outage

The feeling during the first few minutes of a power outage is all too familiar to most people; lights go out, a scramble to find a light source begins (candle, flashlight or more likely a smart phone) and then the discussion of "how long will this one last." During commercial power outages, impacted customers are focused on what the power company is doing to bring power back up, but few know how stressful these outages can be for MSOs as well. MSOs generally would like to avoid the scenario where customers find their cable services unavailable after commercial power is restored.

While MSOs are typically in a reactive state of operations during CPOs, smart data collection and predictive models can significantly help with more targeted operations and clearer communications to subscribers. The following sections will discuss the different stages of a CPO and how Comcast is using data-driven approaches to build AI/ML-driven systems to learn from real outage data.

2.1. Outage Detection

The first step in resolving a CPO is knowing there is one. The detection of a CPO starts with data over cable service interface specifications (DOCSIS) customer premise equipment (CPE) registration data. Typically, during a CPO, a cluster of modems will become unreachable around the same moment in time. The CPE polling systems will label these devices with offline statuses, and they typically remain offline together until services are restored. Once a group of devices is determined to be offline, a clustering algorithm is invoked to determine if they are part of the same outage. The next step is to rule out network plant damage, such as a fiber cut, or something related to a hybrid fiber coax (HFC) outage. Comcast's HFC network is built upon a vast Outside Plant (OSP) power network with real-time telemetry. If MSOs have visibility into CPE connectivity and OSP power supply statuses, the two can be correlated together to see if the group of offline modems have any corresponding power supply (PS) that also lost input power around the same time.

It is important to note that the OSP power supplies typically have battery backup systems to keep the HFC plant operational for a prescribed amount of time during typical CPOs. More information on the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

types of HFC-powering options during CPOs can be found in Peck and Frankhouser (2021). In these cases, there are two types of scenarios for CPE status: customers that do not have an external power source will register as offline, and customers that do have an external power source (home generator or local backup batteries) will remain online as long as the HFC plant is online or until their power source depletes. In these cases, the combination of scenarios needs to be considered as not all customers experiencing a power outage will appear offline immediately.

All previous steps can be automated using internal data sources. The final verification of a CPO can be made by cross-checking with the local power provider in the area of the outage. This can be done with automation and manual engagements depending on the circumstances. The final step can take some time as there may be different lag times between the various systems that need to communicate. It is thus the case that the automation should first create the outage as soon as the confidence level is high enough, and then when the CPO classification is made, the outage is updated, and a potentially different workflow is engaged.

A schematic representation of this process is described in Figure 1. The next section will discuss what actions are typically taken once the CPO is declared.

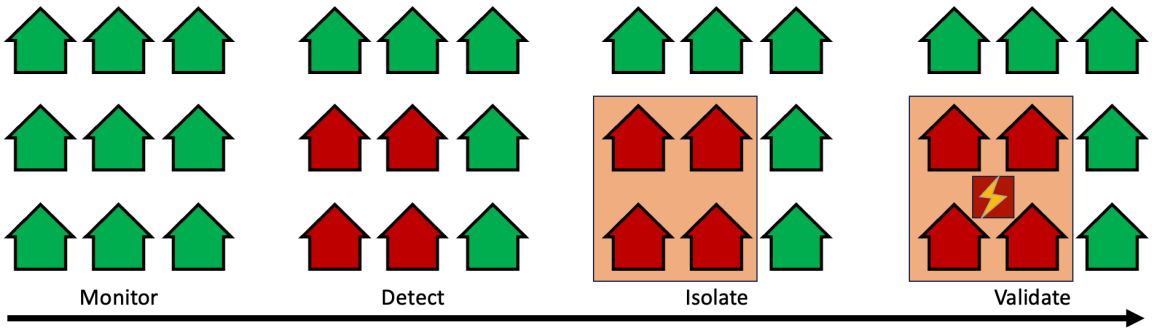


Figure 1 - Detecting a CPO

2.2. Workforce Planning and Maintenance Activity

Once the outage is declared the field teams take over and determine the necessary courses of action. An evaluation of the outage is usually the first task. This typically involves answering questions around: how many customers are offline, how many nodes are impacted, what's the status of our power supplies, is there an Estimated Time to Repair (ETR) from the power provider, is it weather related or something else. In this phase, teams are triaging the situation and getting as much information gathered as possible to determine what operations will be needed to resolve any issues. This data is extremely useful in cases where the demands of the outages exceed the available resources. In such cases, workforce planning efforts may be needed to determine the best sequence of operations based on a variety of factors.

Customer impact is at the forefront of the evaluation. Who is impacted and who may become impacted, may drive the maintenance activities in compliance with the Telecommunications Service Priority (TSP) program. To answer the first part of the question, who is impacted, the status of the CPE and power supplies give an initial estimate of the blast radius of the outage. To answer the second question, who may become impacted, there are two factors. The first is the progression of the outage, which is likely due to factors outside of the MSO's control. The second aspect is related to the power supply batteries mentioned in Section 2.1. During the loss of commercial input power, the OSP power supplies can run on battery to keep the HFC plant operational. However, the batteries have a finite run time. If the duration of

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the CPO exceeds the run time of the backup batteries the HFC plant will go down, resulting in customers being disconnected.

The potential impact to currently online customers' needs to be considered when planning resources and operational activities. Comcast employs a machine learning battery run time model to predict the run time of the power supply batteries at the beginning of every CPO. This run time prediction is then used to determine the need to roll out backup generators to support the HFC plant during outages that exceed the run time capacity of the power supply batteries. Since the run time prediction is made at the beginning of the outage, these generator decisions can be discussed very early in the planning process. Section 3.2 will discuss the run time prediction model in more detail. Figure 2 shows a schematic of the general information that should be considered when planning resources and potential actions to address a CPO.

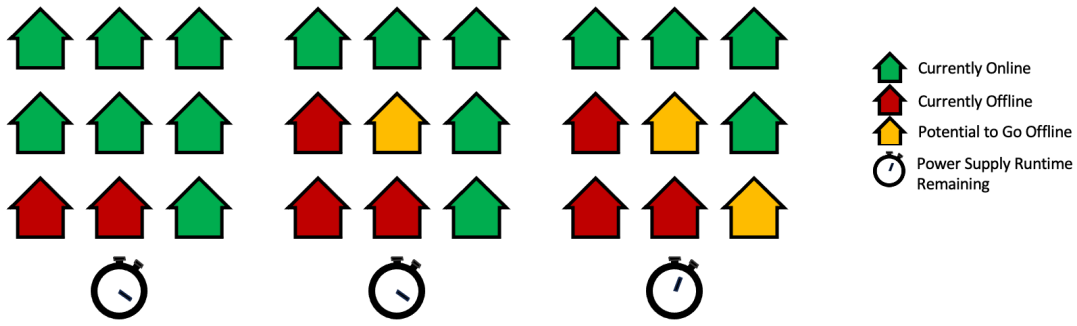


Figure 2 - CPO impact assesments

Another key aspect of the CPO action plan is customer communications. As shown in Figure 2, customers can be in a variety of states during CPOs and a message from their cable provider acknowledging the situation and explaining the next steps can go a long way. Pertinent information for these communications includes current status, possible future statuses, and timelines where available. These items can be derived by combining real-time telemetry with AI/ML systems that learn from historical outage data.

2.3. Outage Resolution and Closure

Once the outage begins to resolve and customers begin to come back online, procedures can be taken to start clearing the outages. Typically, this begins to show via telemetry in the form of the outage footprint shrinking as more customers begin connecting to the HFC plant and their DOCSIS registration statuses are online. Depending on the nature of the outage and the amount of time it takes all customers to come back online, a physical plant inspection may be needed. When customers are back online, the outage will be closed, and this lifecycle will begin again in the next outage.

When the outage is closed, some amount of data will need to be saved for documentation and outage history tracking (depending on the details of the outage, reports may be filed with the Federal Communications Commission and some state regulatory agencies). In addition, if machine learning techniques are to be employed to learn from historical data, as is done at Comcast, then some thought should be put into what data should be stored to compile a training data set for building predictive ML/AI models to aid future outage operations and resolutions. As overviewed in this section, MSOs make many customer-impacting decisions in real-time during CPOs and having an AI/ML framework to help direct these decisions could be beneficial. Section 3 will discuss more on the data storage and ML/AI engagements at Comcast.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Machine Learning on Historical Outages

The ability to leverage predictive models trained on historical CPOs can lead to more efficient resolutions of future CPOs. The various steps in the outage life cycle require different levels of validation and workforce planning based on dynamically changing data as the outage progresses. It is in these cases that machine learning can provide more confidence in validating different scenarios as it continues to learn and evolve from previous outages. As discussed in Cruickshank (2021), MSOs often have high fidelity data around CPOs and thus there is great potential to apply AI/ML techniques.

This section discusses what data related to each outage is stored and how that enables Comcast to run models that both allow for more effective action during subsequent outages as well as shows what steps can be taken prior to future outages that would minimize their impact on customers.

3.1. Creating a Historical Outage Data Base

When an outage concludes, a data structure is stored that contains information necessary to reconstruct the event, so that AI/ML models can be subsequently trained to learn and predict aspects of future outages. In what follows, we refer to this data structure as an 'outage object,' distinct from an 'outage event,' which is the clustering of offline customers at a given time.

At its highest level, an outage object stores summary information: the classification of the outage (CPO or HFC), the starting and closing time of the outage; the physical boundary enclosing all locations that experienced an offline event during the outage; the list of nodes and power supplies impacted; and the number of customers who experienced a service disruption. This information is useful for characterizing and comparing the overall impact of an outage. Figure 3 shows a snapshot of outages from a geospatial point of view for context about the meta data stored.

The outage classification is determined based on several factors. For example, in Figure 3, Outage A would be classified as a CPO due to the involvement of PSs. If Outage B happened at a similar time to Outage A, then it could be classified as a CPO based on proximity to Outage A, even though it has no PSs directly contributing. Finally, if Outage C happened in isolation at a separate time, it would be hard to classify it as a CPO without more context. A potential method for having a confidence factor when classifying CPOs with no PSs or nearby CPOs using historical data is mentioned in Section **Error! Reference source not found.**. The outages can evolve over time as the situation changes. For example, power providers bringing power back block-by-block. In these cases, an outage object should contain all the various states of a given outage, from start to resolution.

The next level of an outage object is a time series of outage events. For each observed timestamp and outage identifier, the outage object contains a list of the nodes and power supplies impacted; the number of customers offline; and the geographic boundary of the event. The boundary is derived from the locations of the offline customers and power supplies. This data can be used for further modeling and analysis efforts after the fact.

The final level of an outage object contains records for the components of the outages indicating the start and end time of their outage; their location (latitude/longitude); the node they are serviced by or power supply; and their status (offline for customers; offline or on standby for power supplies). Each record also stores the outage identification(s) the record is associated with.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 3 - Outage geospatial example

More detailed information is available from the power supplies due to their continued telemetry, sampled every minute even while operating on standby. This includes detailed information about the power supply itself, such as the quantity of batteries and their make/model; the temperature; and the output voltage and power provided to the active devices in the OSP. Comcast also stores the individual battery voltages and the voltages over each string of batteries, allowing for critical insights into the operation and health of the batteries during their discharge. The PS telemetry that is stored for these historical outages are then used to train ML models that can predict battery run times during future outages as shown in Section 3.2

3.2. Power Supply Battery Run Time Modeling

By harnessing the rich battery run time telemetry from historical CPOs and leveraging the extracted historical performance data for each power supply, an advanced battery run time prediction model has been put into production. The model combines real-time telemetry with the aforementioned historical performance features to predict how long each individual power supply's batteries may be able to power the HFC plant until the batteries deplete and reach end of discharge (EOD). The current model is a Random Forest (RF) Regression model that was chosen as RFs are known to be capable of capturing complex relationships to outliers resulting in higher accuracy. Moreover, this model offers valuable insights through feature importance explanations and provides a confidence score in the prediction, making it useful in the field. This concept was first introduced in Ohnmacht and Stehman (2022) and the model accuracy has since improved due to the accumulation of more training data and enhancements to the features. The model in use is trained, validated, and tested on ~40K outages spanning over 10 months of data, resulting in a model that is informed by a wide distribution of possible run time events, empowering it with more predictive ability.

Figure 4a shows the current model performance against the first iteration of the model and a state-of-the-art physics-based model, described in Lin and Nispel (2022), on the test set of ~6k events, excluding events from the training set. It is evident that the new model performs significantly better overall but also on shorter run times, centering the raw prediction error closer to zero. Given the model's knowledge of historical events, it is outperforming the preceding model by 15% and the baseline physics-based model by 50%. The latest model was specifically tuned to make more accurate predictions for shorter run time situations. This accuracy improvement is seen in Figure 4b as the ML 2023 model is able to capture the trend of the shorter run times while previous models cannot. This was accomplished by adding a new set

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

of features to the model that indicate if the power supply's batteries are not in a nominal state (many recent discharges, loss of commercial power while charging, etc.). This addition allows the model to make predictions that consider the current state of the batteries at run time.

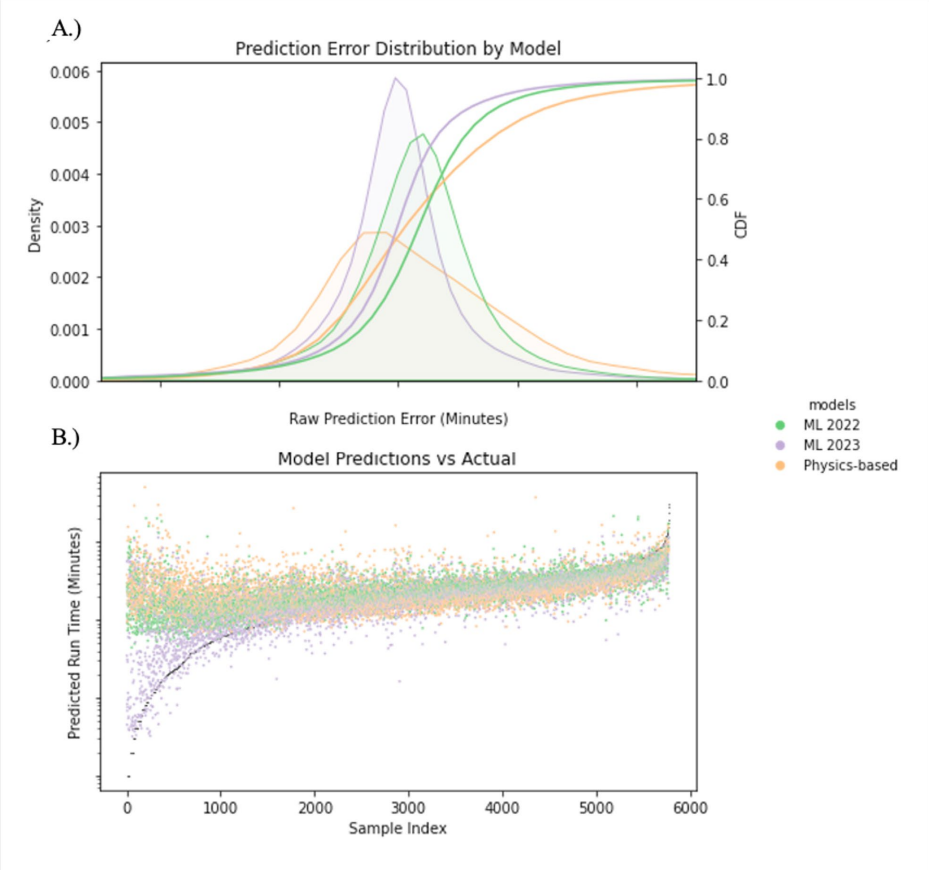


Figure 4 - Battery run time model performance: A.) Distributions of model prediction errors B.) Raw predictions of each model (colored dots) compared to the actual run times (black dots)

As with any production ML model, it is vital to have a model maintenance plan to avoid any sort of input data or model performance drift. From the most recent model, it is apparent that adding more events can improve accuracy. The current plan is to re-train this model on a quarterly basis, incorporating the most up-to-date distribution of possible outage scenarios while also exploring new features and methodologies for model improvement. There is also an internal view to monitor model performance that can eventually be incorporated into a pipeline for automatic model re-training.

3.3. Customer Impact Modeling

Once a historical database of validated CPOs is stood up, AI/ML algorithms can be deployed to find correlations and make predictions about future outcomes for similar outages. It is well known that the HFC plant and commercial power grid don't always overlap nicely. This discord can lead to various scenarios with customer connection status not aligning with the HFC plant status. The path to resolution may vary in these different scenarios and being able to predict beforehand how the customer connectivity in each outage scenario will unfold poses a huge head start in a path to resolution. See Table 1 for a breakdown of these scenarios, the most probable actions and potential customer communications.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 - Actions for various outage states

Commercial Power Status	HFC Status	Customer Status	MSO Action	Customer Communication
In Outage	Online	Online	Monitor outage progression. Plan generator deployment around PS run time model prediction	Notify of current outage and that they may experience service disruptions if outage last longer than run time of batteries
		Offline	Monitor customer. Likely lost power and cannot connect to service	Notify of current outage and that services are available. Include estimated time to power restoration
	Offline	Offline	Deploy generator and bring plant online ASAP	Notify of current outage and expected time to repair
Nearby Outage	Online	Online	Monitor nearby outage progression	Notify of nearby outage and potential disruptions

From Table 1 there are a few interesting scenarios relating to customer experience. The first case is where some customers remain online from a DOCSIS standpoint even though the PS supporting their branch of the HFC plant is running on battery, indicating a loss of commercial input power. In this scenario, either the customer is on a different commercial grid than the HFC plant, or they have an alternative power source. These customers may not even be aware that there is an outage in their area and that their cable data services may be impacted. Being able to use advanced techniques to identify these cases and message customers about an outage and potential interruptions can go a long way in customer satisfaction.

The other scenario is when the commercial grid directly overlaps with the HFC plant. In these outages, most customers will show as offline while the PS is running on battery. This can completely change both operations and customer communications. Firstly, if all customers on a node are offline, even when the power supply is running on battery, that indicates that the customers share the same commercial power grid as the HFC plant. In these cases, if customers lose power, they might not know that cable data services are still available to them even though they lost power. Thus, in these scenarios, messaging a customer to let them know that cable services are still available could completely change their outage experience. Furthermore, if the AI model indicates areas where this scenario is pervasive over time, a proactive notification can be sent to customers indicating that an addition of a backup battery to their cable modem will keep them connected even during CPOs.

Comcast is developing these AI systems to learn trends about our HFC and CPO relationships using all the data and tools mentioned above. Figure 5 shows a plant topology view from a small example where correlation scores are given to customer locations on whether they share the same power grid as the HFC plant. In this example an entire busleg remained online while the PS was running on battery and went offline when the PS batteries depleted. Having a model that can learn from these scenarios and make predictions about the various states of customer connectivity during CPOs will significantly aid in outage resolution going forward.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

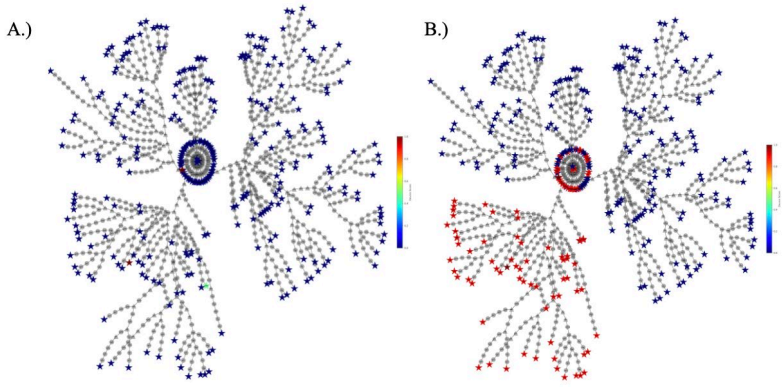


Figure 5 - Customer correlation to power supply status: A.) customers online while PS is running on battery B.) customers offline when PS batteries deplete.

4. CPO Dynamics and Long Term Planning

In this section, Comcast's approach to rethinking the CPO life cycle will be presented, highlighting the integration of AI and ML models. These models not only assist in resolving current CPOs but also play a crucial role in long-range planning to mitigate future outages. First in Section **Error! Reference source not found.**, an example CPO timeline is presented and discussed to showcase the power of building predictive models from historical outages. Then Section 4.2 will discuss the operational improvements that can be made with this type of approach.

4.1. Learning Patterns from CPOs

There are certain patterns that emerge when analyzing CPOs and understanding how to analyze these patterns can go a long way in being able to better action around future CPOs. Figure 6 shows a representative illustration of a typical CPO. This is a time series view of the typical components involved in a CPO i.e., counts of subscribers offline; power supplies running on battery; and power supplies whose batteries have reached EOD.

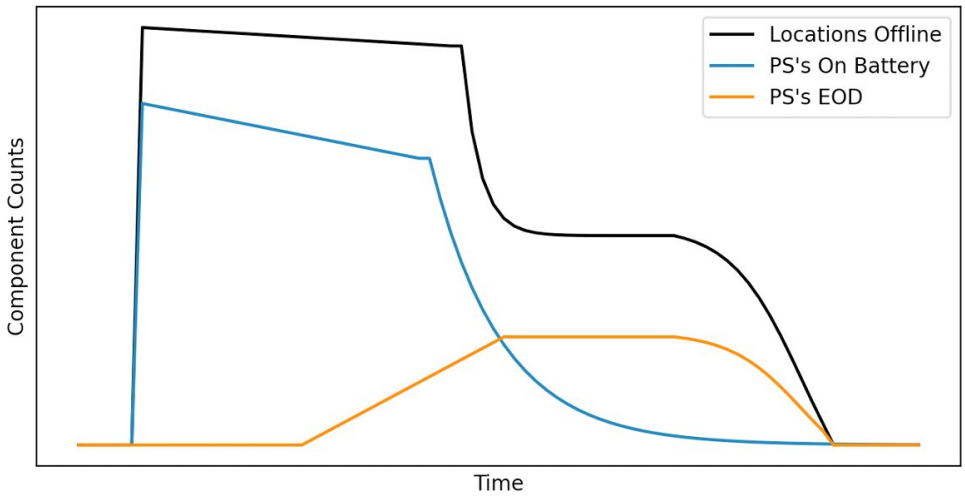


Figure 6 - Timeline of a CPO

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

There are a few features to note in this figure that will be relevant to modeling CPOs. The first feature being the initial impulse where subscribers and PSs lose commercial power almost instantly. As the power supplies run on battery, there tends to be a linear increase in the number of power supplies reaching EOD vs time. The number of PS's and customers that regain commercial power increases with time as the outage begins to resolve. It is easy to see that the offline location counts closely resembles the sum of the time series of the PS's on Battery and PS's EOD curves. This highlights the fact that some customer locations share the same commercial power as the power supplies and some locations have a different power source than the power supplies. For this reason, being able to model the relationship between customers, HFC plant and commercial power grid would allow MSOs to be more targeted and efficient when managing future CPOs.

Some important aspects worth modeling from this outage will be discussed as they can lead to increased confidence in identifying a potential future outage in the same area of service. Modeling which customers lose service at the same time will give an indication of which customer's share the same commercial power grid. In a similar fashion, knowing which customers regain service at the same time can help understand trends in resolving outages. Correlating customers to power supply status's is also of interest. Keeping track of which customers still have service when their power supply is running on battery can indicate that they have an alternative source of power. Analyzing which customers loose service when a PS runs to EOD can help inform power to HFC plant mappings and help in impact assessments of future EOD events, as mentioned in Section 2.3. Finally monitoring how long the Power Supply batteries last before reaching EOD can help in for future predictions of run time for subsequent outages, as discussed in Section 3.2.

In the event of a subsequent outage in the same area of service having models around the data discussed in this section can aid significantly in identifying and resolving the outage in a CPO standpoint. Understanding which customers tend to lose service together during CPOs will help add a level of confidence when declaring a future CPO. Having a model of which customers tend to regain service together can help when declaring an CPO closed and even indicate that the CPO is closed but an HFC outage remains. And finally having a baseline for how long a PS's battery lasted in a previous outage will be helpful for predicting its runtime in a future outage.

4.2. AI Enhanced CPO Operations and Planning

The previous section introduced several opportunities for AI/ML systems to add confidence in CPO identification and isolation as well as to help drive efficient field operations during outages. Figure 7**Error! Reference source not found.** shows a diagram of the overarching effort to have AI platforms not only impact real-time operations but also drive long-range planning.

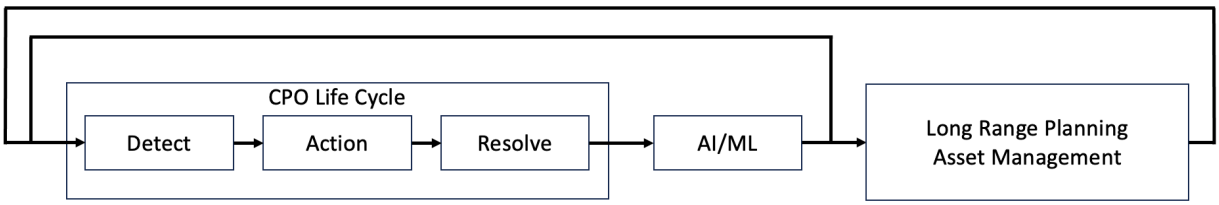


Figure 7 - AI enhanced CPO resolution and OSP power asset management

As showcased earlier in this section, while each CPO has its own features, CPOs don't tend to be random, and thus building models from historical CPOs can be useful for long-term planning. A few AI/ML initiatives that Comcast believes to be significant enhancements to the current state of the art are discussed in the remainder of this section.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

From an operations standpoint, making targeted and calculated field decisions during CPOs is needed to mitigate customer offline time. Predictive models can aid in identifying CPOs and further distinguish between CPO and HFC outages. By modeling which customers typically go offline together in CPOs, a confidence score can be derived to help with the validation stage of future CPO lifecycles. Also understanding which customers share commercial power with the HFC plant allows for accurate proactive messaging and better planning in real time. This can be achieved by modeling which customers are typically online while being located inside an outage footprint.

With a rich historical database of CPOs, commercial power grid reliability and modeling could be leveraged for long-term planning. Building predictive models for commercial power grid performance and reliability can aid long-range planning on several fronts including hardening OSP power supply architectures and OSP asset placement. In the first use case, being able to model the commercial power grid from a reliability and geographic footprint standpoint would allow MSOs to set location-specific run time requirements for different regions of the HFC plant. The current PS backup battery systems may be over designed in some cases and under designed in others. Tailoring the PS run time designs to each location could help optimize OSP power supply efficiency. This information can then be made available to consumers through partnerships with CableLabs and platforms like Grid Metrics to continue improve customer experience across the cable and commercial power domains.

Another interesting use case is the potential to relocate OSP power assets to more reliable areas of the commercial power grid. This initiative has potential to offer savings on multiple fronts. Firstly, if the OSP power supplies can be moved to different and more reliable portions of the commercial power grid than the customers it serves, there is potential to have less HFC related outages and thus lower customer downtimes. From an energy usage standpoint, smart placement of OSP PSs can lead to less run time on battery, resulting in fewer re-charge cycles and further reducing the energy footprint of the OSP. Optimal placement of power supplies can also increase the service life of the batteries, resulting in a reduction in truck rolls and capital expenses.

5. Conclusion

This paper showcases how DOCSIS connectivity data and OSP PS telemetry can lead to advanced CPO detection, actioning, and resolution methodologies. Applying AI/ML models to a historical database of CPOs can lead to more efficient field operations during CPOs and aid in long-term planning efforts to further optimize the OSP power network. Comcast is currently running a ML PS battery run time model in production that significantly outperforms the state-of-the-art physics-based models. Comcast is also developing a novel CPO outage detection algorithm and creating a historical data base of CPOs for future ML/AI models. By applying novel AI/ML approaches in the commercial power domain, Comcast hopes to improve customer experience during CPOs by having more targeted, proactive customer communications and running a more efficient OSP power network, both physically and operationally.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AI	Artificial intelligence
CPE	Customer premise equipment
CPO	Commercial power outage
DOCSIS	Data over cable service interface specifications
EOD	End of discharge
ETR	Estimated time to repair
HFC	Hybrid fiber coax
ML	Machine learning
MSO	Multiple systems operator
OSP	Outside plant
PS	Power supply
RF	Random forest
TSP	Telecommunications service priority

Bibliography & References

Ohnmacht, S., & Stehman, M. (2022). *Machine Learning and Telemetry Improves Outside Plant Power Resiliency for More Reliable Networks*. SCTE Cable-Tec Expo 2022, Philadelphia, Pa.

Cruickshank, R. (2021). *What's Smart About Smart Power?*. SCTE Cable-Tec Expo 2021, Virtual.

Lin, K., & Nispel, M. (2022). *A New Model for Plant Power and Health Estimation*. SCTE Cable-Tec Expo 2022, Philadelphia, Pa.

Peck, T., & Frankhouser, J. (2021). *Ensuring HFC Network Resiliency During Extended Utility Outages*. SCTE Cable-Tec Expo 2021, Virtual.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Internet Service Providers and Cybersecurity

Device Identification and Security at the Edge

An Operational Practice prepared for SCTE by

Scott Dotto
Senior Product Manager
Charter Communications
6360 S Fiddlers Green Circle
(303) 721-3758
Scott.Dotto@Charter.com

Justinas Bisikirskas, CUJO AI

Kyle Johnson, Charter Communications

Leonardas Marozas, CUJO AI

Tyson Vinson, Charter Communications

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

Table of Contents 2

List of Figures 2
1. Introduction 3
2. Cybersecurity and ISPs Landscape 3
3. Tools and Techniques to Gather Data 4
   - 3.1. Connected Device Identification 4
   - 3.2. Connected Device Security 5
4. Data Analysis 7
   - 4.1. Connected Device Overview 7
   - 4.2. Popular Brands and Models 11
5. Case Studies (Examples) 12
   - 5.1. Hikvision IP Cameras & DVRs 12
   - 5.2. Space Monkey NAS Devices 13
6. What Actions Can ISPs Take? 13
7. Conclusion 15

Abbreviations 16

Bibliography & References 16

List of Figures

Title **Page Number**

Figure 1: Internet Access Layers 4
Figure 2: Device Identification 5
Figure 3: Cybersecurity Threats by Type 7
Figure 4: Device Penetration by Category in Charter Homes 8
Figure 5: Percentage of Threats by Device Type 10
Figure 6: Brand “Threat Index” 11
Figure 7: Percentage of Threats by IP Camera Brand 12
Figure 8: Percent of Cybersecurity Events and Number of Devices by Brand Among NAS Devices 13
Figure 9: ISP Opportunities 15

List of Tables

Title **Page Number**

Table 1: Cybersecurity Event Type Definition 7

1. Introduction

Cybercrime is a pervasive and increasing risk to businesses and consumers alike. This increasing threat is driven by the expanding influence of digital lives, the growing reliance on technology to deliver essential services and the rise in connected devices, which are projected to reach an estimated 15 billion by the end of 2023. According to a report by Cybersecurity Ventures, the global cost of cybercrime is predicted to reach a staggering $10.5 trillion annually by 2025 (about $1,300 per person on average).[1] In the first half of 2022, an estimated 53 million Americans were impacted by cybercrime, costing US households billions of dollars.[2] Such a massive problem demands effective solutions to safeguard customers' devices, data and financial security.

Internet service providers (ISPs) are in a unique position to improve the cybersecurity landscape. To serve customers, ISPs must know certain traffic characteristics such as IP source and destination. These observations offer ISPs valuable insight into the capabilities of devices on the network and whether traffic from those devices may be suspicious or malicious. Further analysis may uncover vulnerabilities in customer devices and other cybersecurity risks.

The knowledge gained from this analysis may be used to develop new security services and techniques. Sharing deidentified and aggregated information (with a commitment to protecting personal information and restrictions on how that information can be used) with relevant parties—manufacturers of offending devices, other ISPs, or regulatory and standards bodies—can lead to better outcomes in securing customers' data, identities, devices and, consequently, the internet as a whole. This deidentified and aggregated data can be directly used by other parties to make informed decisions to drive industry standards and respond promptly and efficiently to an ever-evolving threat. Additionally, more specific information can be shared with impacted customers so they may take action to protect themselves.

ISPs have a critical role in enhancing cybersecurity and securing the internet. ISPs' access to traffic data empowers them to take proactive steps in protecting their customers and the internet ecosystem at large. By leveraging these insights and collaborating with relevant stakeholders, ISPs can meaningfully contribute to mitigating cybersecurity threats to customers.

2. Cybersecurity and ISPs Landscape

ISPs are a major component in the infrastructure of the internet ecosystem, facilitating the smooth flow of data between end-users and various online resources. Protecting this critical infrastructure, as well as customer security, is standard practice and a high priority. ISPs employ a variety of techniques for security, including but not limited to:

1. **Traffic Filters:** These filters close off commonly abused & attacked services, such as open DNS resolvers and SMTP servers.
2. **Distributed Denial of Service (DDoS) Scrubbing**: DDoS scrubbing helps remove high-volume attack traffic used to impact the availability of internet access.
3. **"Walled Gardens" or Enclosed Network Environments:** Isolates a compromised customer to prevent spreading malware and to provide a call-to-action for the customer to contact the ISP for further assistance.
4. **Endpoint protection software:** Customers can protect their laptops and mobile devices via software provided by the ISP

These techniques are robust and capable; however, ISPs have a responsibility and opportunity to protect customers at all relevant network layers and on all devices. Techniques available at the core network can be reinforced by data and advanced learning models deployed at the edge. These methods enhance the

ability to apply policy to traffic at a more granular level—the device or application. Of particular importance is providing customers with protection for the individual devices they connect to their home WiFi. This enables an ISP to protect multiple households or groups of users and against threats that target individual devices or specific applications within a household.

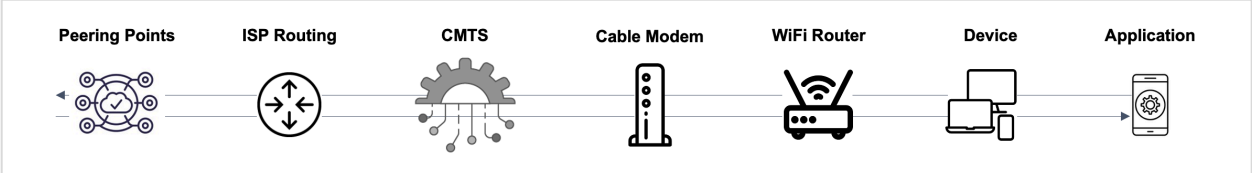


Figure 1: Internet Access Layers

ISPs have a responsibility and the unique capability to protect customers at all relevant points of service delivery. An agent on the WiFi router bolsters cybersecurity for customers beyond the traditional protection at peering and core network.

3. Tools and Techniques to Gather Data

3.1. Connected Device Identification

At the edge of their network, ISPs may deploy capabilities on customer premises equipment. One such example is Charter Communications, which employs a solution developed by CUJO AI, a company focused on cybersecurity and advanced learning models. This solution is integrated onto the WiFi router platform.

By deploying an agent on the WiFi router, it can capture unique device identifiers from network metadata, which is then analyzed by a classification engine to categorize each device by type, brand, model and the device's specific capabilities. For instance, the engine may determine that a device is a streaming device and whether it supports 4K resolution, among other characteristics.

The classification engine also assigns a unique device identity, ensuring consistency in measuring data over time for many essential use cases, including quality of experience (QoE) metrics, improved customer support and security performance over the device's lifespan. Recent privacy measures, such as private MAC and random MAC, may obfuscate the data to uniquely identify traffic overtime.

Improvements in privacy for online activities also do not necessarily equate to an improved security profile for devices. While many innovative solutions have been introduced to support consumer privacy, they must also aim to ensure that consumers remain protected while online. Appropriate security measures play a vital role in safeguarding devices from potential threats and attacks.

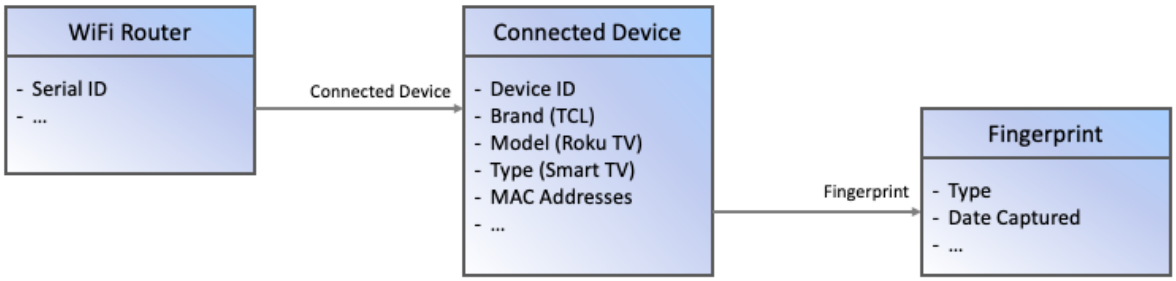


Figure 2: Device Identification

Device identification solution gives both details on the type of device and unique fingerprints.

3.2. Connected Device Security

Each device, depending on its profile, communicates differently on the network making it impractical to have a single cybersecurity solution that fits all. For instance, devices which support web browsing, like smartphones, require safe-browsing security to protect consumers from rogue websites or malicious URLs. On the other hand, IoT devices necessitate a different type of protection, such as blocking unauthorized access from malicious IP addresses and detecting malicious device behavior. Security solutions deployed at the edge should be multi-layered, tailoring appropriate security services specifically to each device based on its profile.

To ensure security for all connected devices, real-time network metadata is captured for processing by a low latency AI engine. This engine investigates and categorizes cybersecurity events into four distinct categories: secure traffic, secure browsing, outbound denial of service (DoS) prevention and smart device protection (remote access threats).

This paper focuses on secure traffic cybersecurity events, which are the most common threats to single-purpose devices such as smart home IoT devices (*single-purpose device* implies a device that lacks a browser and typically requires minimal interaction with the user to perform its purpose). Unlike secure browsing threats, secure traffic better reflects the devices that pose a higher risk to cybersecurity, as it factors out threats blocked due to user behavior (e.g., clicking on an insecure link through a browser on a mobile phone). By focusing on secure traffic for single-purpose devices, the analysis excludes the cybersecurity events stemming from an individual's unsafe browsing behavior and allows for a better comparison of the devices themselves.

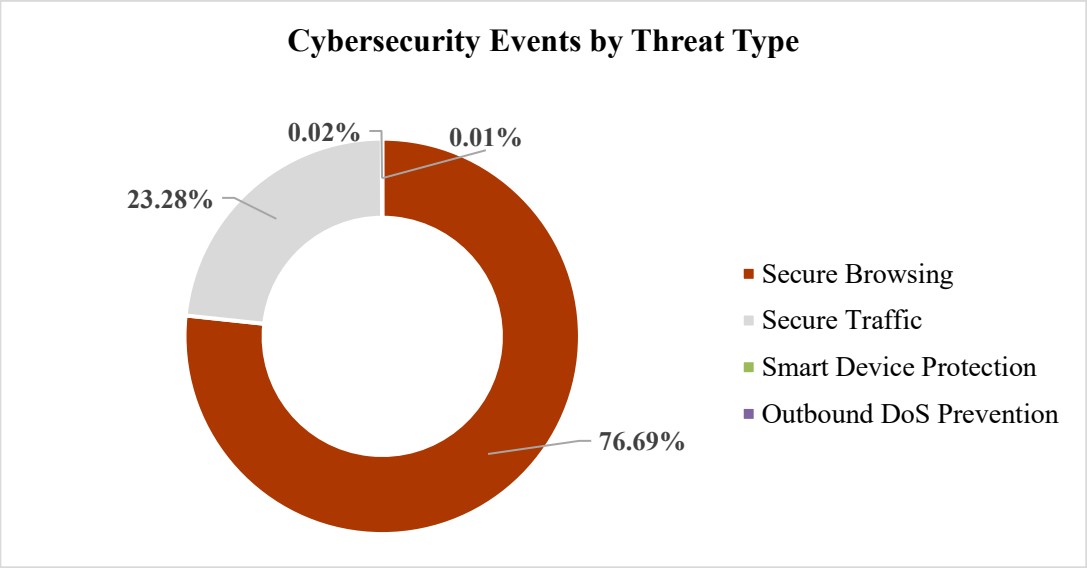

Cybersecurity Events by Threat Type
0.02%
0.01%
23.28%
76.69%
Secure Browsing
Secure Traffic
Smart Device Protection
Outbound DoS Prevention

Figure 3: Cybersecurity Threats by Type

Secure browsing and secure traffic account for the vast majority of cybersecurity events (99.97%)[3].

Table 1: Cybersecurity Event Type Definition

Secure Browsing protects users from accessing websites that are known to be malicious or suspicious
Secure Traffic monitors network traffic at the IP layer; if traffic (inbound or outbound) is detected from a known, malicious IP residing in the reputation database, the connection is blocked
Outbound DoS Prevention prevents customers' devices from participating in a DoS attack; attack traffic to the victim IP address is blocked, while other traffic is allowed
Smart Device Protection detects and blocks IoT communication that deviates from normal communication patterns

When a cybersecurity event is detected, the specific malicious traffic to or from the device is automatically blocked in real time, preventing any negative impact to the customer and their devices. Furthermore, cybersecurity events are recorded for further analysis.

The device identification dataset can be combined with the cybersecurity threat dataset. This integration allows for in-depth exploration of the types of devices, brands, and models that over-index on cybersecurity events—devices that tend to attract unwanted attention from cybercriminals. Identifying these devices constitutes the first step toward better protecting internet customers.

4. Data Analysis

4.1. Connected Device Overview

Unsurprisingly, the most prevalent devices in customers' homes are phones and computers. While many of these devices are subject to cybersecurity threats, most of these threats are from either insecure browsing behavior or exposing ports to the internet, rather than other threat categories to which all devices are susceptible. In fact, most threats (approximately 80%)[3] are related to unsafe browsing and infections directing users to insecure and malicious content. Since cybersecurity events may arise from either user behavior or characteristics of the device, when analyzing the security of a device it is critical to control for user behavior. For devices with multiple, divergent use cases and especially those with brower support (such as phones and computes), it may be difficult to draw strong conclusions about the device.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

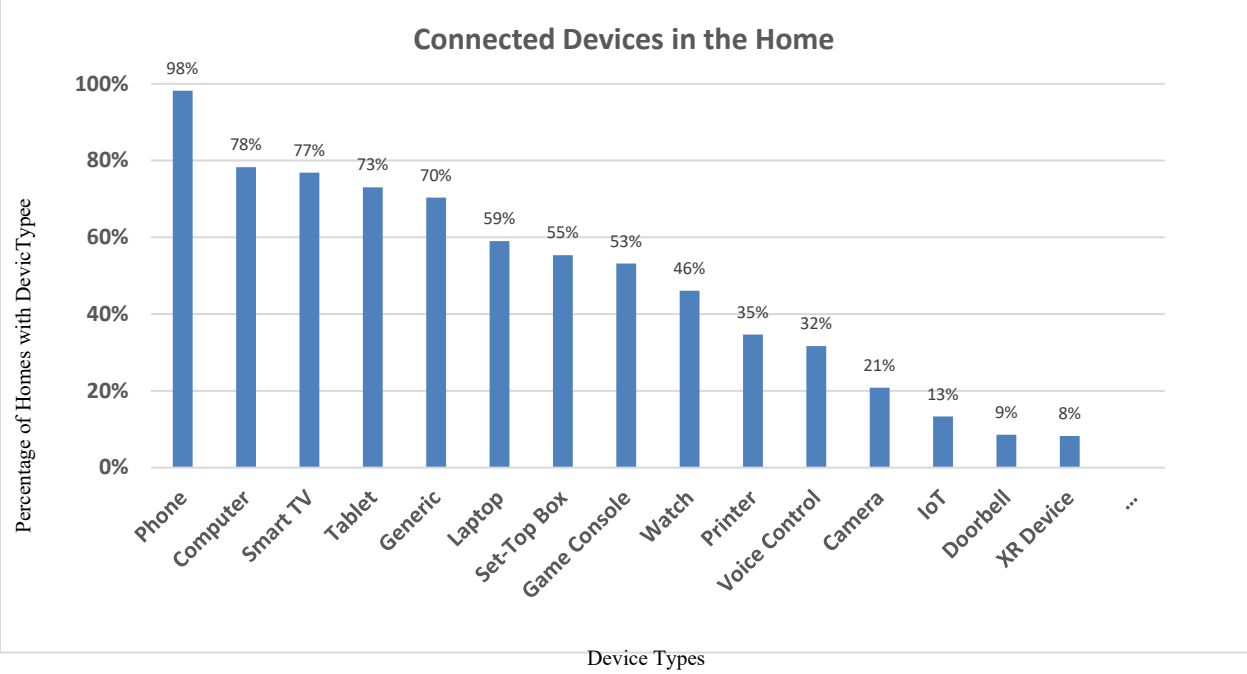


Figure 4: Device Penetration by Category in Charter Homes

Almost all homes have mobile phones (98% of homes), computers (78%) and smart TVs (73%). The prevalence of IoT devices such as security cameras, doorbells, thermostats, smart lights or plugs is also increasing. This chart omits less prevalent device types[3].

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Conversely, for single-purpose devices lacking a browser, the volume of cybersecurity incidents may lead to clear insights on which devices tend to have more vulnerabilities or invite more attacks. Certain IoT device types (e.g., IP cameras, DVRs and doorbells) and network attached storage (NAS) devices tend to over-index on cybersecurity threats—these device types, when controlling for the number of devices represented within each category, consistently have more cybersecurity attempted actions per device. For these devices, the most common blocked security threat is secure traffic, a connection attempt from known, malicious IP addresses.

Over 99% of connection attempts arise from recognized, malicious IP addresses, while less than 1% of threats originate from the infected device, initiating communication with known botnet command and control centers or other malicious IPs. Known, malicious scanners can quickly identify new assets with open ports on the Internet, catalog them and attempt exploitation within seconds.

As Figure 5 below illustrates, IP cameras account for an outsized percentage of all threats observed while only constituting a small percentage of the devices (0.66% of total device count). IP cameras are a distinct category and differentiated from other cameras by sending digital data to be stored in the network. Similarly, NAS devices also over-index relative to other device types in cybersecurity events, accounting for 36% of Secure Traffic events while only accounting for 0.03% of all devices[3].

NAS devices and IP cameras are enticing targets for several reasons:

1. **Weak or Absent Security Features:** Many IoT devices, including NAS devices and IP cameras, lack essential built-in security features, making them an easy target.
2. **Lack of Regular Updates:** Despite the regular appearance of new vulnerabilities and subsequent exploits of NAS devices and IP cameras, they are not consistently updated. Both vendors and users share responsibility—vendors to deliver updates to patch known vulnerabilities and users to ensure devices are operating with the latest firmware.
3. **Value of Data or Function:** Devices that store valuable data, such as NAS devices (or cryptocurrency miners), or perform certain functions, such as IP cameras, are attractive targets due to potential gains for attackers in terms of data or information.
4. **Increasing Interconnectivity:** Many NAS devices and IP cameras expose ports to the internet, either via UpnP or by asking the user to manually configure forwarding on their routers.
5. **High Data Rates:** NAS devices and IP cameras, as opposed to many other IoT devices, transmit high-bandwidth data, enabling obfuscation of DDoS traffic by interspersing with higher-volume legimate traffic.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

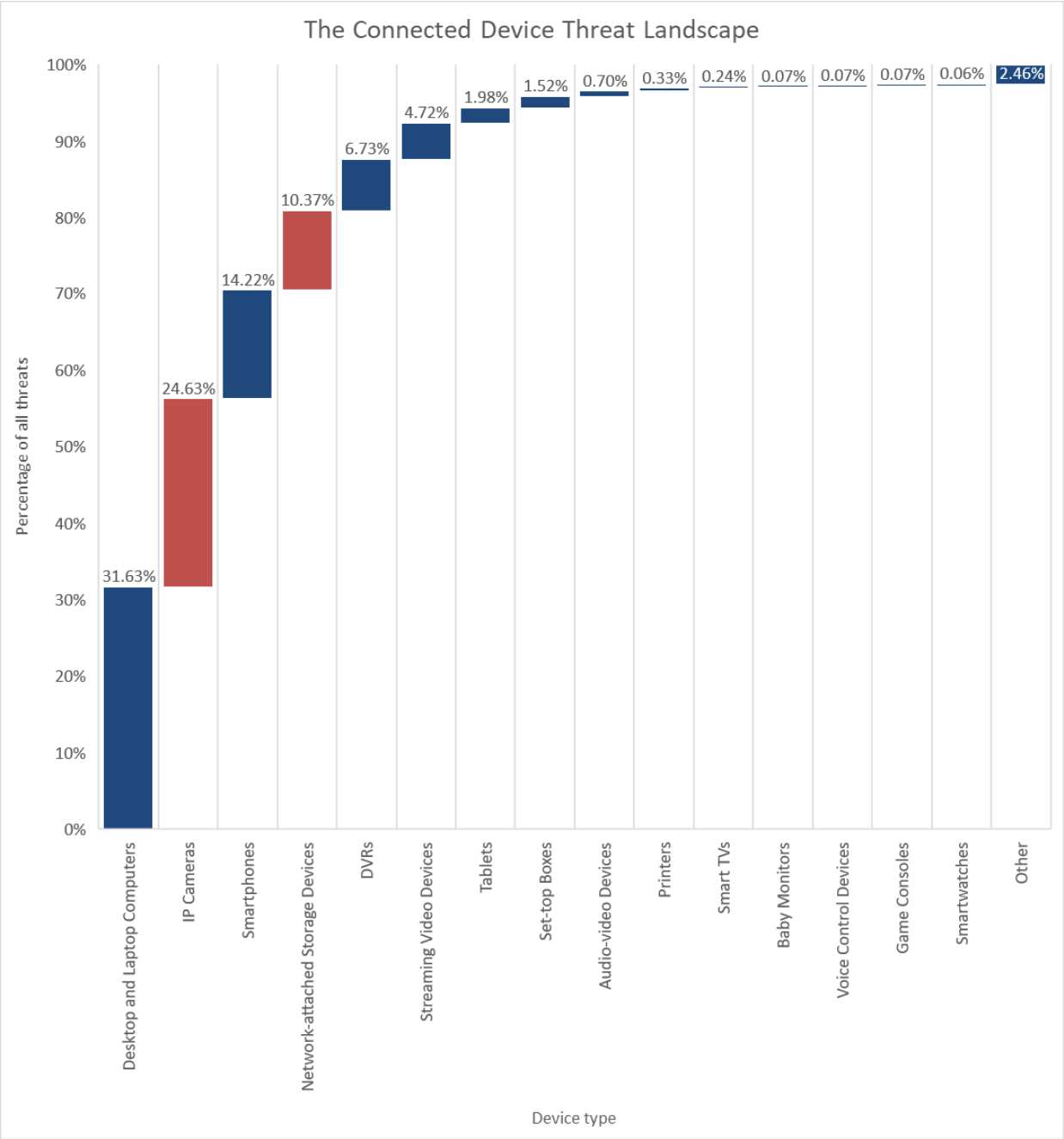


Figure 5: Percentage of Threats by Device Type

A relatively small number of device types account for most cybersecurity threat events[3].

While these device types often carry greater risk, recommending customers completely remove such devices from their homes is not a practical stance. Gaining insight into which specific devices within these categories pose the highest risk can offer more alternatives for effectively mitigating these threats.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2. Popular Brands and Models

This analysis can also be extended further to device brands and models. Comparing brands and models within a specific category better illustrates which brands and models over-index in cybersecurity events, given certain device types may just be inherently riskier to provide the services they are designed for. Devices from those over-indexing brands pose a higher risk to customers compared to their peer devices.

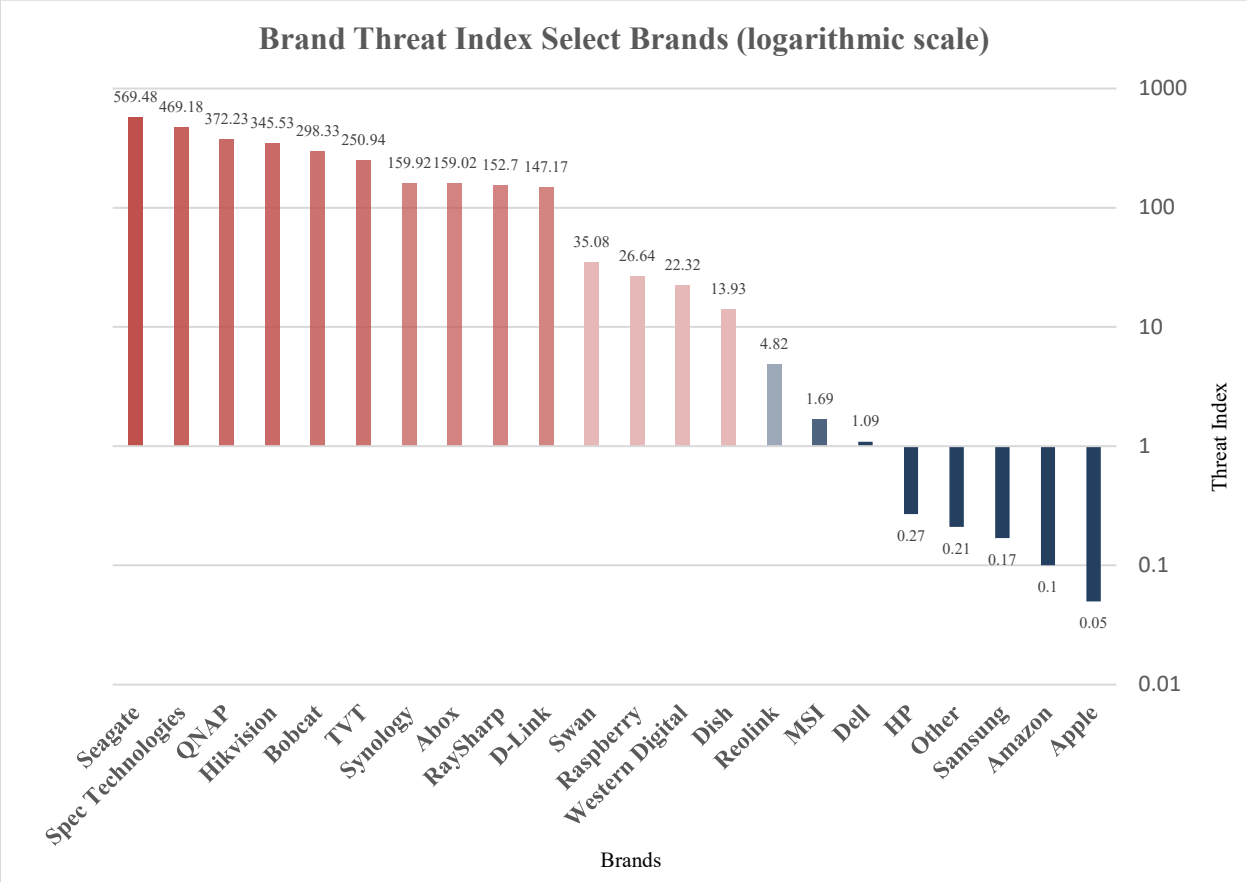


Figure 6: Brand "Threat Index"

The *threat index* represents the ratio of average threats per device in a category. A group of devices with a threat index above or below 1.00 experience more or fewer than average threats, respectively. For example, a device model with a threat index of 400 experiences 10 times more threats on average than a device with a threat index of 40[3].

The data indicates that well-recognized brands generally exhibit strong performance in terms of the number of threats per device. A significant portion of the brands that over-index in security threats are those with a smaller presence in the market. Notably, these brands also tend to correspond to devices within the categories previously identified as having a higher incidence of cybersecurity events, including NAS devices, IP cameras, doorbells (mainly equipped with cameras) and cryptocurrency miners.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Additionally, the data shows that a large preponderance of cybersecurity events come from a small number of device brands. Small, targeted efforts to correct issues with a few brands or models could potentially yield large reductions in cybersecurity events for customers.

Exploring more deeply how specific brands or models perform within their respective device type categories reveals more clearly which devices tend to pose the highest risks. Further investigation may provide insights into the underlying reasons for this trend. When factoring in the number of devices represented by each brand or model, a select few become more prominent.

5. Case Studies (Examples)

5.1. Hikvision IP Cameras & DVRs

As mentioned earlier, IP cameras generally have a higher incidence of cybersecurity events relative to other device types when controlling for number of devices represented. When examining specific brands within the IP camera device category, the disparity becomes even more pronounced. Despite constituting only 3.7% of the IP camera market share, Hikvision devices contribute to 58% of the documented cybersecurity threats[3]. Although another brand, ClareVision, fares even worse when factoring in the device count, their limited presence in the market mitigates the substantial threat to customers.

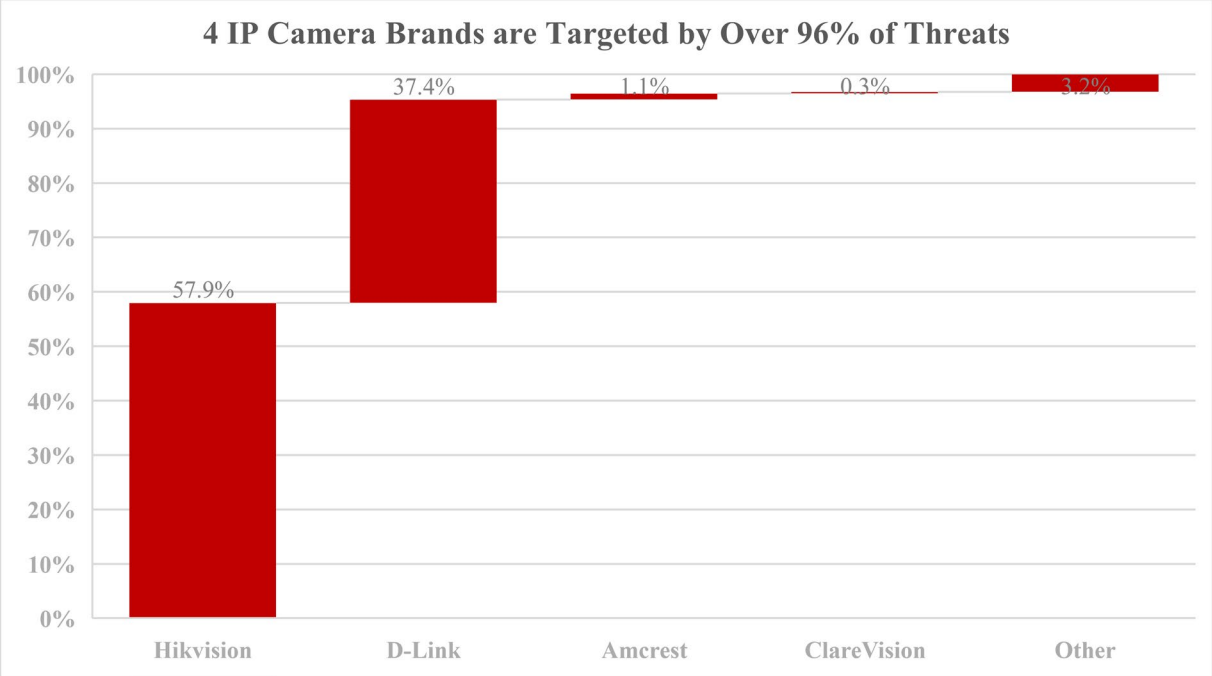


Figure 7: Percentage of Threats by IP Camera Brand

Hikvision accounts for over half of the threats against IP cameras, with D-Link accounting for another 37%. Combined they account for 95% of threats[3].

The notable prominence of Hikvision is even more intriguing in light of its ban by the Federal Communications Commission (FCC) in November 2022 due to its "unacceptable risk to national security."[4] Specifically, the ban blocks import and sale of new video surveillance and telecommunications equipment from Hikvision in the US "for the purpose of public safety, security of government facilities,

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

physical security surveillance of critical infrastructure, and other national security purposes."[4] This ban is not enforceable for all consumer-grade equipment, which may still find its way to customers in the US market. Hikvision devices potentially introduce an undisclosed risk to the consumers who possess them, originating from a company deemed a national security risk by the US government.

5.2. Space Monkey NAS Devices

Space Monkey was a brand of NAS devices offering a unique solution that combined the concepts of cloud storage and an in-home device. This approach provided a distributed and decentralized storage solution. Customers' data was not exclusively stored on their NAS device at home but was duplicated over Space Monkey's extensive device footprint. Frequently used files were locally cached for quick access[5].

In 2014, Vivint acquired Space Monkey but later chose to discontinue support for that business line[5] and the devices no longer receive firmware updates. These devices are now susceptible to cybercrime through known vulnerabilities, making them an appealing target.

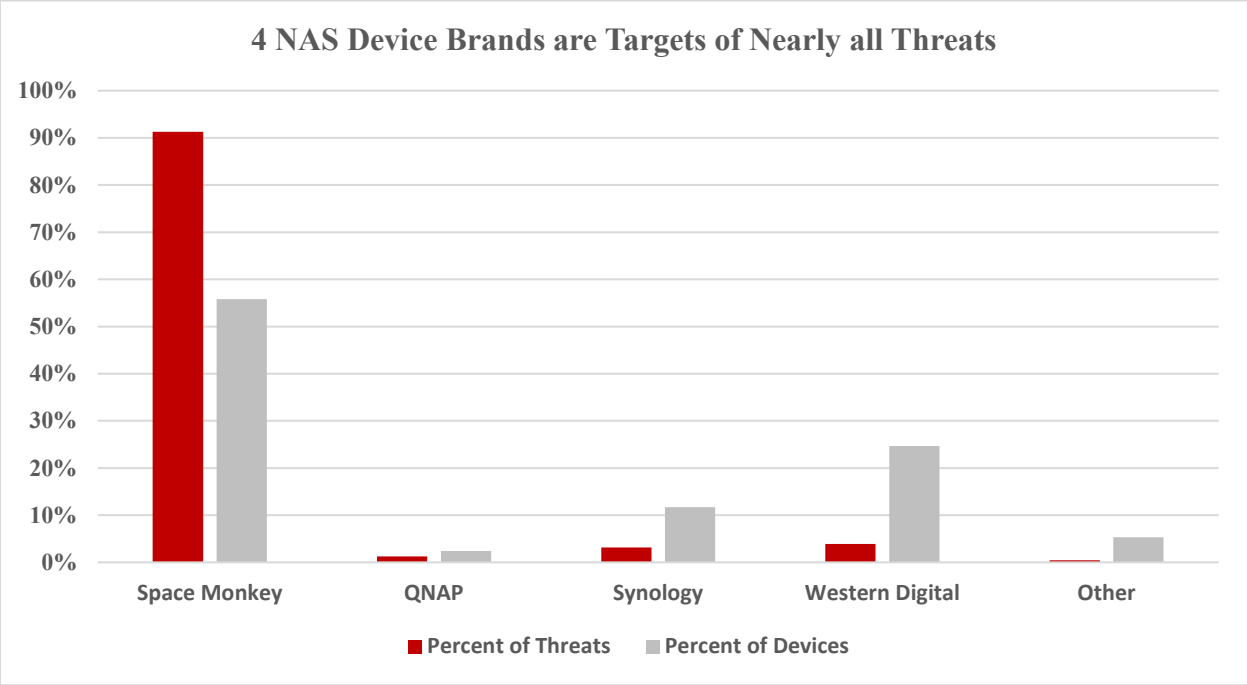


Figure 8: Percent of Cybersecurity Events and Number of Devices by Brand Among NAS Devices

Space Monkey devices account for over 90% of cybersecurity events observed across all NAS devices, while accounting for 56% of NAS devices observed in market[3].

6. What Actions Can ISPs Take?

As the above examples illustrate, a large percentage of the threats can be attributed to relatively few brands. That is helpful in considering a response strategy given a small, specific action may yield great results. Better securing a few devices may drastically reduce the number of cybersecurity events on an ISP's network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

This data provides insights into various measures to enhance network and customer security. First, for devices that might pose significant risks or are plagued by difficult-to-address vulnerabilities (such as those arising from limited or absent firmware support, as seen in the case of discontinued brands), the ISP could share this information with the customer. Often, customers may not be aware of the risks associated with using certain devices. In situations where devices are well supported and established, sharing deidentified and aggregated information directly with the brand/manufacturer could provide them with valuable insights to mitigate specific security risks and enhance their product line.

In the cases of both Hikvision and Space Monkey, notifying customers may be the most effective strategy for mitigating ongoing cybersecurity threats. Hikvision might not have the incentive to patch devices in a market that bans their products, and Space Monkey is no longer a supported device, thus lacking available updates. In other scenarios, conveying these insights to device manufacturers directly might equip them with the necessary information to rectify vulnerabilities, allowing their devices to perform more closely with industry peers.

This deidentified and aggregated information could also be reinforced by sharing it with other ISPs possessing similar capabilities. By exchanging additional data, ISPs could potentially verify or dismiss specific concerns. Cumulatively, there might be ample data and coverage to facilitate broader dissemination, reaching standards bodies, regulatory entities or open-source cybersecurity threat-sharing initiatives like Malware Information Sharing Platform (MISP) Threat Sharing. This data could contribute to aiding other organizations by offering information on emerging threats.

Another instance of such interorganizational collaborations could involve consortiums of incident response teams, such as Forum of Incident Response and Security Team (FIRST[6]). These platforms would enable partners and other organizations to collectively engage in threat research, enhancing their incident response capabilities. Furthermore, they could provide highly-valuable data for Special Interest Group (SIG) chapters, which could then elevate compliance and standards pertaining to data handling and protection based on the gathered data. Additionally, sharing information with Information Sharing and Analysis Centers (ISAC) or Information Sharing Analysis Orgranizations(ISAO), such as the Comm-ISAC may provide another path to explore threat information and best practices through exisiting entities. Many ISPs are already members of Comm-ISAC, and increased membership and participation may aid in collective goals regarding cybersecurity.

Collaboration with standards bodies or academic institutions with this data who could perform more extensive research would benefit internet security. The National Institute for Standards and Technology (NIST), for example, publishes device standards for consumer-grade IoT products, which may benefit from these insights. In addition to end devices, this data may guide other initiatives aimed at securing gateways and the traffic they route such as the “Gateway Device Security Best Common Practices” published by CableLabs[7].

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

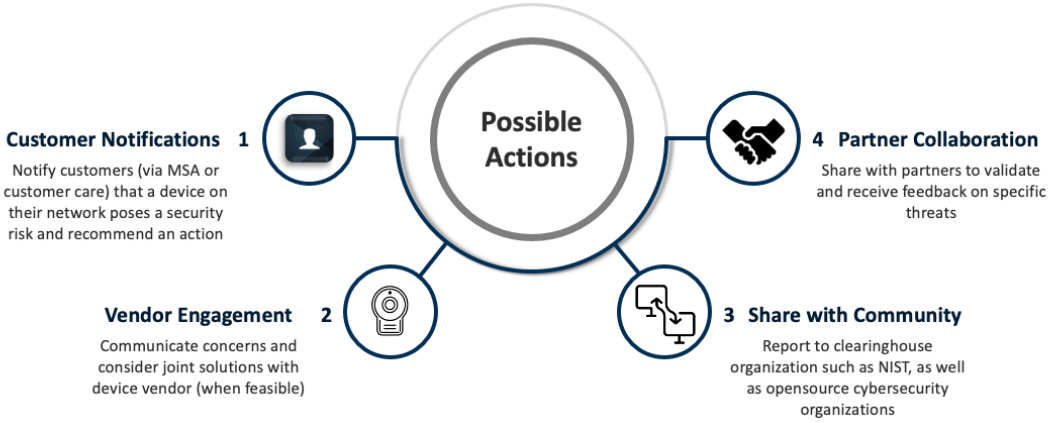


Figure 9: ISP Opportunities

7. Conclusion

The escalating threat of cybercrime poses a widespread and growing danger to everyone, and this challenge necessitates effective solutions to safeguard customers' devices, information and financial assets. ISPs have a distinct opportunity to play a pivotal role in bolstering the cybersecurity landscape.

ISPs should be building and deploying capabilities to profile single-purpose devices on their networks with respect to cybersecurity. And when armed with that information, ISPs should engage in discourse with all relevant parties to determine how best to fix uncovered vulnerabilities in end devices.

ISPs wield a critical function in elevating cybersecurity and safeguarding the integrity of the internet. Their access to traffic data empowers them to take proactive measures in shielding their customers and the broader online ecosystem. By harnessing these insights and collaborating closely with stakeholders, ISPs stand poised to make substantial strides in protecting customers from the ever-changing landscape of cybersecurity threats.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AI	artificial intelligence
CMTS	Cable Modem Termination System
DoS	Denial of Service
DDoS	Distributed Denial of Service
DHCP	Dynamic Host Configuration Protocol
DNS	Domain Name System
FCC	Federal Communications Commission
FIRST	Forum of Incident Response and Security Team
HTTP	Hypertext Transfer Protocol
IoT	Internet of Things
IP	Internet Protocol
IR	internal report
ISP	Internet Service Provider
MAC	Media Access Control
MISP	Malware Information Sharing Platform and Threat Sharing
NAS	network attached storage
NIST	National Institute of Standards and Technology
OSI	Open Systems Interconnection
SIG	Special Interest Group
SMTP	Simple Mail Transfer Protocol
SSDP	Simple Service Discovery Protocol
Tbps	terabits per second
TCP	Transmission Control Protocol
UPnP	Universal Plug and Play
URL	Uniform Resource Locator

Bibliography & References

1. Forbes, “10.5 Trillion Reasons Why We Need a United Response to Cyber Risk”; https://www.forbes.com/sites/forbestechcouncil/2023/02/22/105-trillion-reasons-why-we-need-a-united-response-to-cyber-risk/?sh=3b14abce3b0c
2. AAG, “The Latest 2023 Cyber Crime Statistics (updated July 2023)”; https://aag-it.com/the-latest-cyber-crime-statistics/#:~:text=Headline%20Cyber%20Crime%20Statistics&text=1%20in%202%20American%20internet,the%20first%20half%20of%202022.
3. Charter Advanced WiFi Platform Data, June 2023
4. FCC, “FCC Bans Authorizations for Devices That Pose National Security Threat”; https://www.fcc.gov/document/fcc-bans-authorizations-devices-pose-national-security-threat
5. Wikipedia, “Space Monkey (company); https://en.wikipedia.org/wiki/Space_Monkey_(company)
6. https://www.first.org/
7. CableLabs, ““Gateway Device Security Best Common Practices”; https://www.cablelabs.com/specifications/CL-GL-GDS-BCP

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Leveraging DAA Investments to Accelerate FTTP

Considerations for FTTP that are Specific to Cable Operators

A Technical Paper prepared for SCTE by

Brian Soloducha, P.Eng
Principal Network Architect
Rogers Communications Inc
2400 32 Ave NE Calgary AB Canada
403-716-6041
brian.soloducha@sjrb.ca

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Background – Legacy FTTP deployments at a Cable Operator 4
   - 2.1. Limitations of Backbone Fiber Infrastructure 5
   - 2.2. Limitations of Video Service Delivery 8
3. Evolution of the Access Network and Hubsite Infrastructure 9
   - 3.1. Sharing of Backbone Fiber within an Analog Node Service Area 9
   - 3.2. Deploying Distributed Access Architecture 10
     - 3.2.1. Upgraded 40ch passive DWDM in the Access Network 10
     - 3.2.1. Simplified Hubsite Configurations 14
4. Remote OLTs and Distributed FTTP 15
   - 4.1. Defining Remote OLTs 15
   - 4.2. Defining Distributed FTTP 16
   - 4.3. Benefits 17
   - 4.4. Drawbacks 20
   - 4.5. Considerations for Distributed FTTP specific to Cable Operators 20
   - 4.6. Future PON Protocols 21
     - 4.6.1. 25G PON (25G EPON and 25GS-PON MSA) 22
     - 4.6.2. 50G PON (50G EPON and an ITU-T variant) 24
     - 4.6.3. 100G PON (CableLabs Coherent PON Working Group) 25
5. Conclusion 26

Abbreviations 27

Bibliography & References 28

List of Figures

Title **Page Number**

Figure 1 - Backbone Fiber Design Parameters 5
Figure 2 - HFC Node Service Areas 6
Figure 3 - FTTP Node Service Areas 6
Figure 4 - FTC with Active Drop Configuration 7
Figure 5 - Hubsite Configuration for Legacy FTTP and Legacy HFC Architectures 8
Figure 6 - Deploying Passive DWDM (MWL gen 1) into Node Service Areas 9
Figure 7 - 4ch/8ch Passive DWDM (MWL gen 1) 11
Figure 8 - 5ch/40ch Passive DWDM (MWL gen 2) 12
Figure 9 - Deploying Passive DWDM (MWL gen 2) into Node Service Areas 13
Figure 10 - Backbone Fiber Cable with Passive DWDM at different Node Service Areas 13
Figure 11 - Hubsite Configuration for currently-supported architectures 14
Figure 12 - Hubsite Configuration for DAA and Distributed FTTP 15
Figure 13 - Remote OLT Node Enclosure 15
Figure 14 - Remote OLTs sharing Backbone Fiber through Passive DWDM 17
Figure 15 - New RPDs and Remote OLTs into existing Node Service Areas 18
Figure 16 - FTC with Hot Drop Configuration 19

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Tables

Title **Page Number**

Table 1 – Defining Remote OLT Characteristics 16
Table 2 – Expected DOCSIS 4.0 Modulation Throughput 23
Table 3 – Expected DOCSIS 4.0 Downstream Spectrum Usage 24
Table 4 – Expected DOCSIS 4.0 RPD Backhaul Requirements 24

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Cable Operators know how to build and operate high-quality Hybrid Fiber-Coax (HFC) networks for Residential and Business customers. Recent developments in the last few years have focused energies on new Passive Optical Networks (PON) for serving new customers in Greenfield builds with Fiber-to-the-Premise (FTTP) architectures and technologies. With network expansion initiatives in both the United States and Canada, Cable Operators need to be flexible with strategies for growing their networks and increasing revenues from new Commercial and Residential customers.

This Technical Paper discusses a physical architecture for FTTP that will look familiar to those Cable Operators that have already adopted Distributed Access Architecture (DAA) for their HFC deployments. Starting with a reminder of legacy HFC architecture and critical Hubsite components, the paper then discusses legacy PON architecture and Hubsite components, focusing on key limitations that have previously prevented widespread FTTP deployments at Cable Operators. Topics under discussion will progress to recent developments that overcome these limitations. The paper will conclude by looking into the future at problems that are already foreseeable and that will need to be addressed in due time.

This Technical Paper is targeted for an audience curious about starting (or restarting) an FTTP program, and for Cable Operators who have deployed DAA for HFC and are looking to leverage their investment towards PON architectures in Greenfield builds. For those Operators who have already deployed Remote OLTs, the later sections that are forward-looking may be of interest.

2. Background – Legacy FTTP deployments at a Cable Operator

Back in the early 2010s in Western Canada, the regional Cable Operator trialed FTTP builds for greenfield deployments to learn both the strengths and weaknesses of the architecture, technology, and operationalization thereof, that a strong and ambitious competitor Telco was deploying. Residential households-passed and Business drops exceeded 10,000 as an experiment to determine if and how FTTP could be successfully adopted at scale by a Cable Operator. Two major limitations of these legacy PON deployments were directly experienced in Western Canada that are specific to Cable Operators. These limitations were not relevant to competitors that did not operate legacy HFC networks. Other Cable Operators may have experienced other challenges above & beyond those observed in Western Canada.

The initial years of this experimental deployment program featured a 1G EPON protocol providing data / Internet services, with Radio-Frequency-over-Glass (RFoG) technology overlays being used to provide Video and Voice services. This allowed for the use of a DOCSIS Provisioning of EPON (DPoE) back-office provisioning engine. In later years, the EPON technology was slowly deprecated and migrated over to GPON protocols in some instances.

The physical architecture of this broad experiment utilized Centralized Optical Line Terminals (OLT) with the OLT located in Hubsites that fed the Optical Distribution Network (ODN). A split ratio of 1:32 with a distance limit of 20.0 km was achieved with Central Splitting at a single location per ODN. This location was called a Fiber Transition Cabinet (FTC), although other literature for the industry may call this a “Fiber Distribution Hub” (FDH) or “Local Convergence Point” (LCP). The idea was to use existing Backbone Fiber from a Hubsite to a Node Service Area and place the FTC/FDH for each ODN inside the Node Service Area.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.1. Limitations of Backbone Fiber Infrastructure

The program for FTTP was designed to use existing Backbone Fiber cables from Hubsites out to new Node Service Areas over the Access Network Infrastructure that was in-place at that time.

The existing legacy HFC network was designed with the intended following parameters:

- Node Service Areas would typically be sized at approximately 500 Households-passed (HHP)
- One analog RF node per Node Service Area
- 2 fibers (one forward and one return) per analog RF node
- Backbone Fiber Cabling sized at 288 fibers per cable
    - non-ribbonized to reduce Mean-Time-to-Repair (MTTR) in the event of fiber cuts
- Backbone Fiber allocated as follows:
    - 48f reserved for Hubsite to Hubsite traffic such as Broadcast Video and Metro Dense-Wavelength-Division-Multiplexing (DWDM)
    - 240f reserved for Node Service Areas
- Fiber for Node Service Areas were allocated as follows:
    - 2f used by the current analog RF node
    - 2f reserved for future node segmentation activities
    - 2f reserved for Business services

Not all deployments meet these intended parameters; recent deployments have higher rates of compliance.

Given that 6f are allocated per Node Service Area out of a total of 240f reserved for Node Service Areas on the Backbone Fiber, this allows for 40 Node Service Areas to be served (unprotected, single-ended) out of a Hubsite per Backbone Fiber cable. In Metro areas, the Backbone Fiber cables are typically built from one Hubsite to another Hubsite, so another 40 Node Service Areas are reverse-fed from the far-end Hubsite that the Backbone Fiber cable connects to.

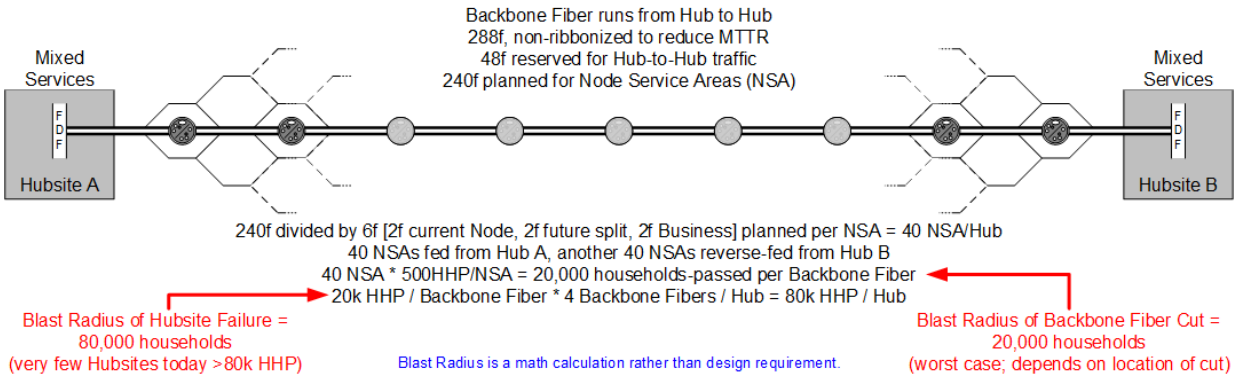


Figure 1 - Backbone Fiber Design Parameters

Western Canada typically has smaller Hubsites than what might be common in Central Canada or the United States, with typical Hubsites serving approximately 50,000 HHP and connecting 3 or 4 Backbone Fiber cables. Vaults for separate entrances for each Backbone Fiber cable would be ideal.

The math calculations for Backbone Fiber then result in a 'blast radius' of up to 20,000 HHP for a Backbone Fiber cable cut, or up to 80,000 HHP in the event of a complete failure at a larger Hubsite (power outage & generator failure, forest fire, catastrophic vehicle crash, etc). This concept of a blast

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

radius is not a design requirement; rather, it is just a resultant of the sizing parameters used elsewhere in the design. The (in)frequency of these failure events has not yet warranted a program to build protection into the Node backhaul connectivity. While some Cable Operators have built protection paths to Nodes out of the same Hubsite for Layer 1 protection; and other Cable Operators are looking at Layer 2 or Layer 3 protection out of the far-end Hubsites (when Backbone Fiber is built from one Hubsite to another), protection paths for Node Service Area connectivity has not been prioritized in Western Canada.

This architecture for the HFC network was deployed in field trials in the late 1980s. Design parameters and fiber cable sizing (288f total, 240f reserved for Node Service Areas) were finalized in the late 1990s.

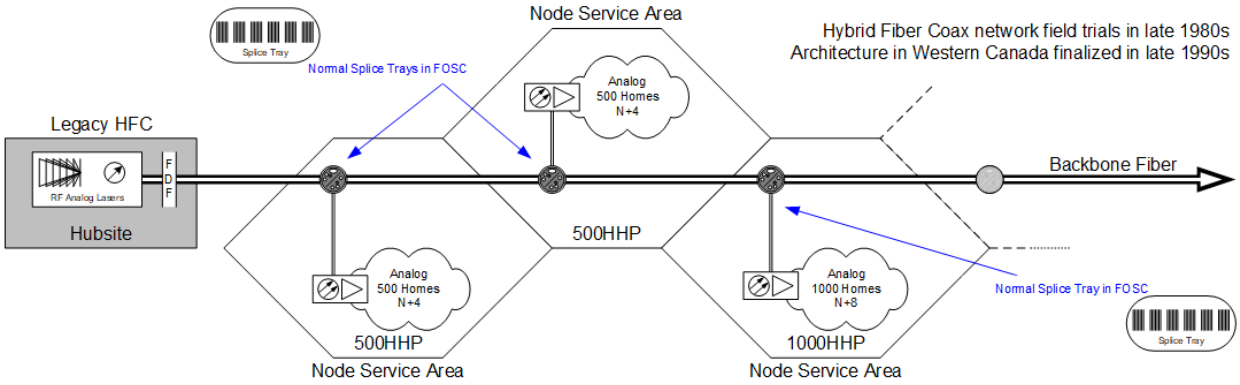


Figure 2 - HFC Node Service Areas

As the experimental FTTP deployment program kicked-off in approximately 2010, it became apparent that the consumption of Backbone Fiber for this FTTP architecture vastly exceeded the planned parameters for fibers assigned per Node Service Area. If all customer drops were connected to 1:32 splitters at the FDH, then a maximum of only 7680 drops could be served off of that Backbone Fiber cable, instead of the planned 20,000 HHP. Further, there would be no room for growth, no back-up fibers for new technologies, and no fibers available for dedicated Business services.

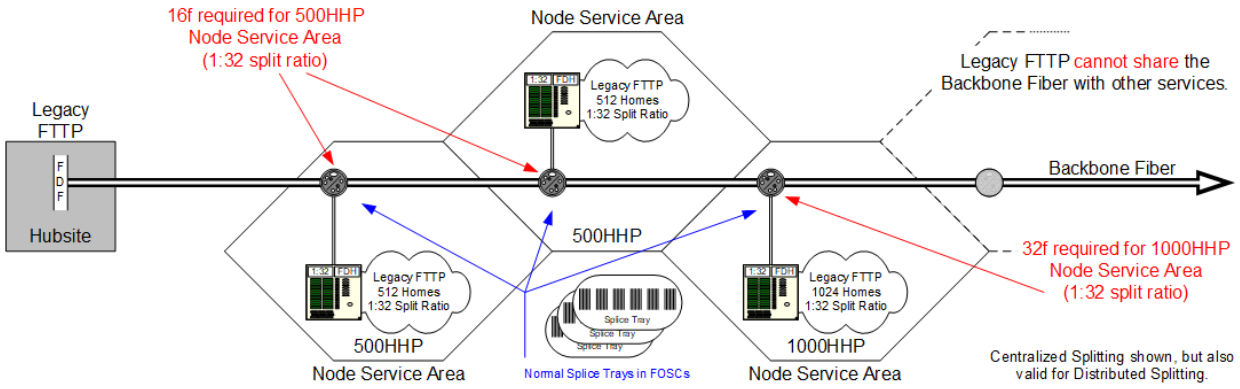


Figure 3 - FTTP Node Service Areas

An attempt was made to minimize unnecessary consumption by using an “Active Drop” configuration at the FDH; rather than connecting every drop to a 1:32 splitter, only active customers would be connected, thereby saving Backbone Fiber usage back to the Hubsite. This configuration also helps to limit the number of PON ports required at the Centralized OLT.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

This proved out to be an effective way to minimize the Backbone Fiber usage, but had three horrendous side effects:

1. A truck roll became mandatory for every single customer add AND disconnect.
2. The documentation became difficult to keep up-to-date.
3. The FDH became difficult to properly maintain and route fiber connections over time, making subsequent modifications (customer additions and removals) progressively more difficult.

With the Active Drop configuration, some FDHs became so unwieldly as to be considered “rats’ nests”.

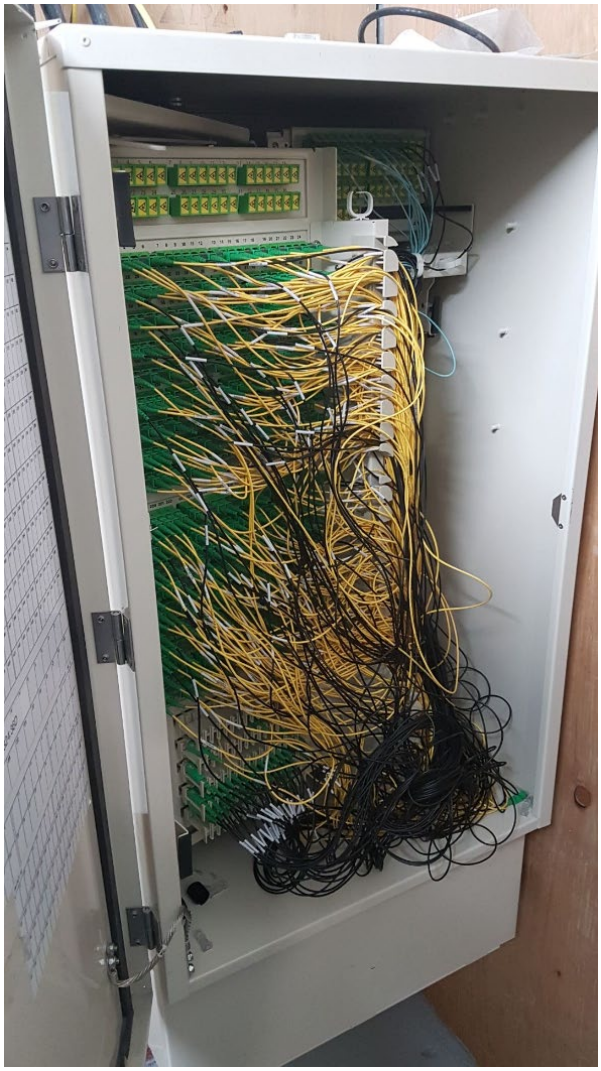


Figure 4 - FTC with Active Drop Configuration

In addition to Backbone Fiber usage, another problem is starting to crop up in older FTTP builds in various parts of both the United States and Canada, namely OLT PON port availability. Some Hubsites and Central Offices do not have rackspace or power to add additional OLT PON ports, and depending on planned penetration rates, there may not be OLT PON ports or Backbone Fiber available to light up new 1:32 power splitters regardless of whether the customer location has a fiber drop. In these cases, the additional revenue from a single new customer would not justify the costs for either a Backbone Fiber

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

build or a new OLT PON port activation, and hence that customer must wait until another customer is disconnected. This does not lead to happy customers.

Backbone Fiber networks used for deploying HFC in Western Canada were largely sized and built decades ago. It is on this infrastructure that the widescale experimental FTTP deployment program was attempted in Western Canada, and where the severity of the Backbone Fiber usage requirement became apparent, despite trying to mitigate the challenge through the use of an Active Drop configuration. This limitation was one of the two key reasons why the experimental FTTP deployment program was shuttered for a decade.

2.2. Limitations of Video Service Delivery

There is a multitude of equipment required in Hubsites for Cable Operators to serve Voice, Video, and Internet product solutions to customers over legacy HFC and legacy FTTP architectures. Internet service is easy over the data path provided by 1G EPON and GPON signals, and Voice services can be added with Session Initiation Protocol (SIP). Unfortunately, Video services requires either an IP Video solution or an RFoG overlay solution.

Back when the experimental FTTP deployment program kicked off, IP Video solutions were complex, costly, and seen as unnecessary by most Cable Operators who already had working Video solutions for their HFC networks. As such, an RFoG overlay solution was chosen for this experimental FTTP program.

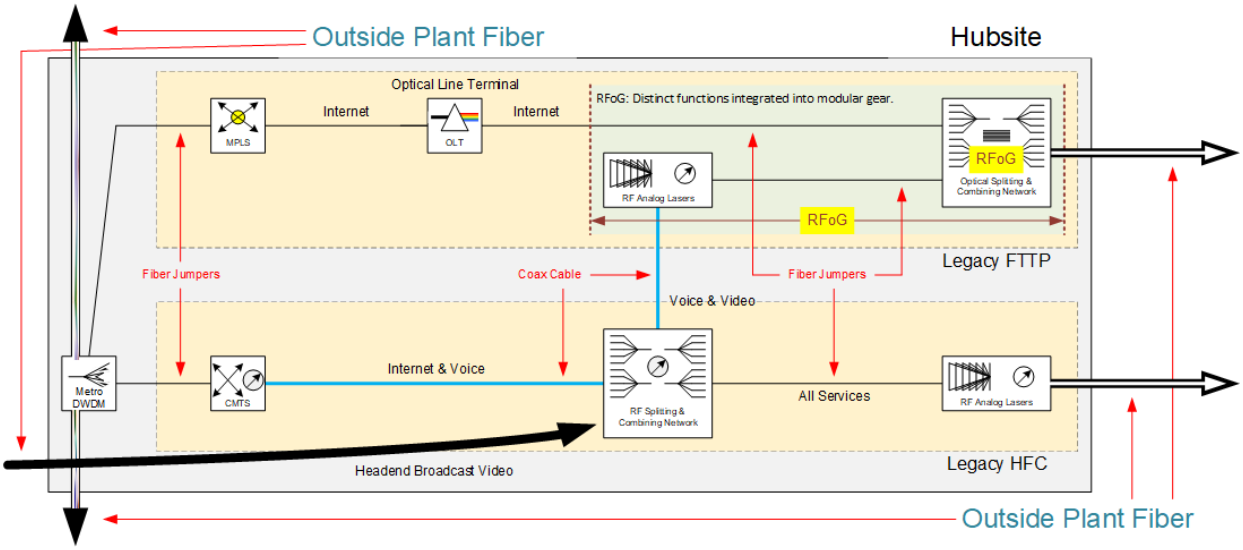


Figure 5 - Hubsite Configuration for Legacy FTTP and Legacy HFC Architectures

Service Delivery for Legacy FTTP was accomplished by combining Voice & Video signals from the Cable Modem Termination Station (CMTS) and Broadcast Video feeds, together with the Data / Internet feed from the OLT. This combining for legacy FTTP occurred at RFoG equipment in the Hubsite, and additional Customer Premise Equipment (CPE) was required to separate out the signals at the customer location. Additional challenges with RFoG, such as Optical Beat Interference, became apparent over time.

As more Cable Operators became aware of the requirements and operational issues of RFoG, demand tapered off while costs for RFoG equipment remained high with a limited number of suppliers in the market.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

This requirement for an RFoG overlay for a Video product proved to be the other key limitation as to why the experimental program for FTTP deployment was shuttered.

3. Evolution of the Access Network and Hubsite Infrastructure

There have been several changes in the past decade worthy of discussion that alleviate some of the limitations for FTTP deployments that are relevant to Cable Operators.

Many Cable Operators have now pivoted, or are in the process of pivoting, to an IP Video solution as part of larger operational considerations. A discussion of this general trend is beyond the scope of this paper. This change negates the need for an RFoG overlay solution for Video services and therefore removes entirely the need for costly RFoG equipment in both the Hubsite and the customer premise.

Backbone Fiber utilization trends and the implications of changing to DAA for HFC are worthy of a more detailed discussion.

3.1. Sharing of Backbone Fiber within an Analog Node Service Area

Fiber congestion mitigation projects are still undertaken in small batches on a regular basis, but large scale Backbone Fiber build programs spanning many months or even years for permitting and construction are to be avoided or deferred as much as possible. As traffic grew on legacy analog RF nodes, and as new neighborhoods were built and new Business services deployed, alternative methods were investigated for growing capacity without requiring large-scale Backbone Fiber overbuilds.

In the mid-2010s, passive DWDM filters and lasers for the analog RF nodes were introduced into the network to allow for segmentation of serving groups within the same Node Enclosure. This allowed for additional capacity within the Node Service Area while still using the same amount of Backbone Fiber for that pocket. This passive DWDM architecture was internally called MWL (Multi-Wavelength) and deployed as-needed in a Pay-As-You-Grow fashion in both 4-channel and 8-channel configurations. This first generation of passive DWDM was only intended and used for connecting analog RF nodes within a common node housing. [For Business customers and possibly a few rural applications, Coarse-Wavelength-Division-Multiplexing (CWDM) has alternatively been used.]

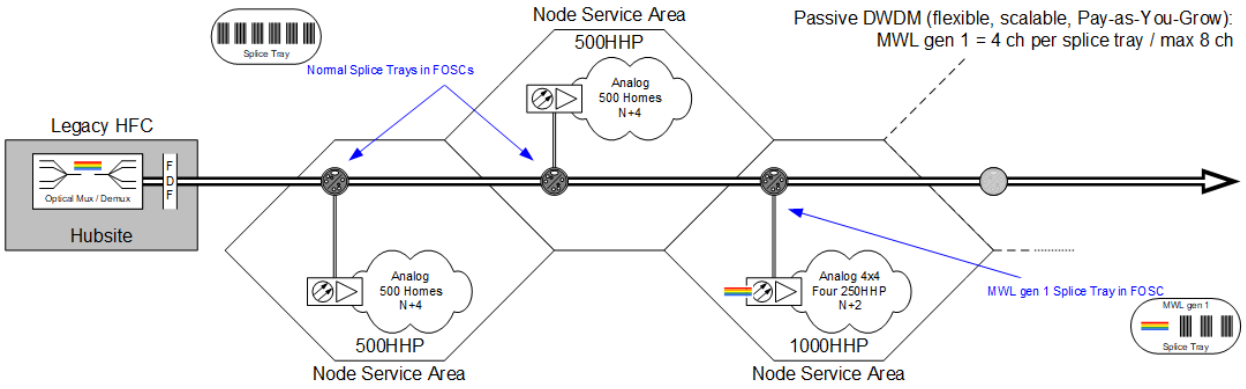


Figure 6 - Deploying Passive DWDM (MWL gen 1) into Node Service Areas

Of all metro Node Service Areas in Western Canada, approximately 20% have 4-channel passive DWDM configurations, and 5% have 8-channel passive DWDM configurations. Typically, a 4-channel passive

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

DWDM mux was placed into a splice tray in the FOSC; however, a 4-channel mux could also be placed inside the Node Enclosure (rare).

The deployment of passive DWDM at existing Node Service Areas allowed for relatively easy segmentation without the need for re-spacing of amplifiers in the coax portion of the network. This work separated one Node in one Node Enclosure into multiple nodes in the same Node Enclosure, thereby lowering the number of customers on each coax feed from the Node Enclosure and hence lowering total utilization and congestion by reducing the denominator (total coax capacity / HHP).

This approach allowed the same pair of fibers from the Backbone Fiber cable into a Node Service Area to be shared amongst multiple Nodes for that Node Service Area. This was an important technology breakthrough and operational mode for staying ahead of congestion on the HFC network in Western Canada.

Another important technology breakthrough and operational mode vital to staying ahead of congestion on the HFC networks in Western Canada was the conversion of the entire network to an 85 MHz return. A discussion of that mid-split upgrade program is beyond the scope of this paper.

3.2. Deploying Distributed Access Architecture

Approximately 5 years ago, a program was kicked-off in Western Canada to evaluate and deploy Distributed Access Architecture (DAA) for HFC. There have been many Technical Papers in recent years relating to DAA, so a full summary of the benefits and risks of DAA is unwarranted, but a short summary for both the Outside Plant and for the Hubsite as DAA relates to FTTP for Cable Operators is as follows:

In the Outside Plant, DAA allows for:

1. 10G Ethernet backhaul, allowing for
2. I-Temp DWDM full-band tunable pluggable optics, allowing for
3. Full-band passive DWDM filters and muxes

In the Hubsite, DAA allows for:

1. Converged Interconnect Network (CIN) routers, allowing for
2. Simplified Hubsite Configurations

These changes are key to a new physical architecture that will allow Cable Operators to deploy FTTP in greenfield Node Service Areas, or for in-fill demands from new or unexpected Multiple Dwelling Unit (MDU) buildings such as apartment buildings, condo highrise towers, row townhouses, etc.

3.2.1. Upgraded 40ch passive DWDM in the Access Network

The move to DAA has led to a flourishing of 10G DWDM optical links powered by Full-band Tunable DWDM pluggable optics. These optics are hosted by Remote PHY Devices (RPDs) in the Access Network and by CIN routers in Hubsites. Tunable DWDM pluggable optics in the SFP+ form-factor have now been used in Western Canada since late 2019. Some Cable Operators may choose to use fixed-wavelength DWDM optics; that conversation is beyond the scope of this paper.

The first generation of passive DWDM filters and muxes deployed in Western Canada were 4-channel and 8-channel configurations. These were based on the fixed-wavelength DWDM transceivers suggested by the analog RF node vendors and based on supply availability.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

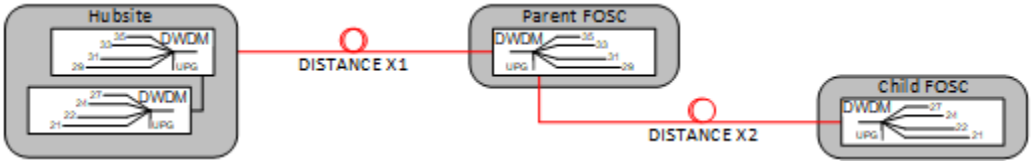


Figure 7 - 4ch/8ch Passive DWDM (MWL gen 1)

With the adoption of DAA and tunable optics, a new generation of passive DWDM filters and muxes were warranted. While the tunable optics were capable of 96-channels across a 50GHz-spaced grid, this was deemed excessive for foreseeable demands.

A next-generation passive DWDM system was developed:

- 100 GHz-spaced, 40-channel total capacity
- 10-channel initial capacity in Hubsites, 5-channel initial capacity in FOSCs in the Outside Plant
- Flexible, Scalable, Pay-as-You-Grow
- Allows for 8 total FOSCs in a Node Service Area
- Allows for 16 channels reserved for HFC (allowing for N+0 in 500 HHP Node Service Areas)
- Allows for 24 channels for other services (Business waves, Small Cells / Wireless, Remote OLTs, etc), and more if N+0 is not a required HFC configuration
- Adds channel groups in parallel

A key feature of this next-generation system is that channel groups are added in parallel rather than cascaded in series; this design minimizes total insertion loss at higher channel counts when compared to a cascaded design.

This next-generation system is called MWL gen 2 to maintain internal consistency for users in Western Canada, and features a few mandatory components in the Hubsite and one mandatory component in the Outside Plant. Surprisingly, due to the optical transmit and receive specifications of the tunable optics and the low total insertion losses achievable per channel, attenuators (Global Pads and Group Pads in the diagram below) are usually required for these wavelengths. [Short range tunable optics with looser transmit and receive specifications did not make sense from a supply chain perspective.]

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

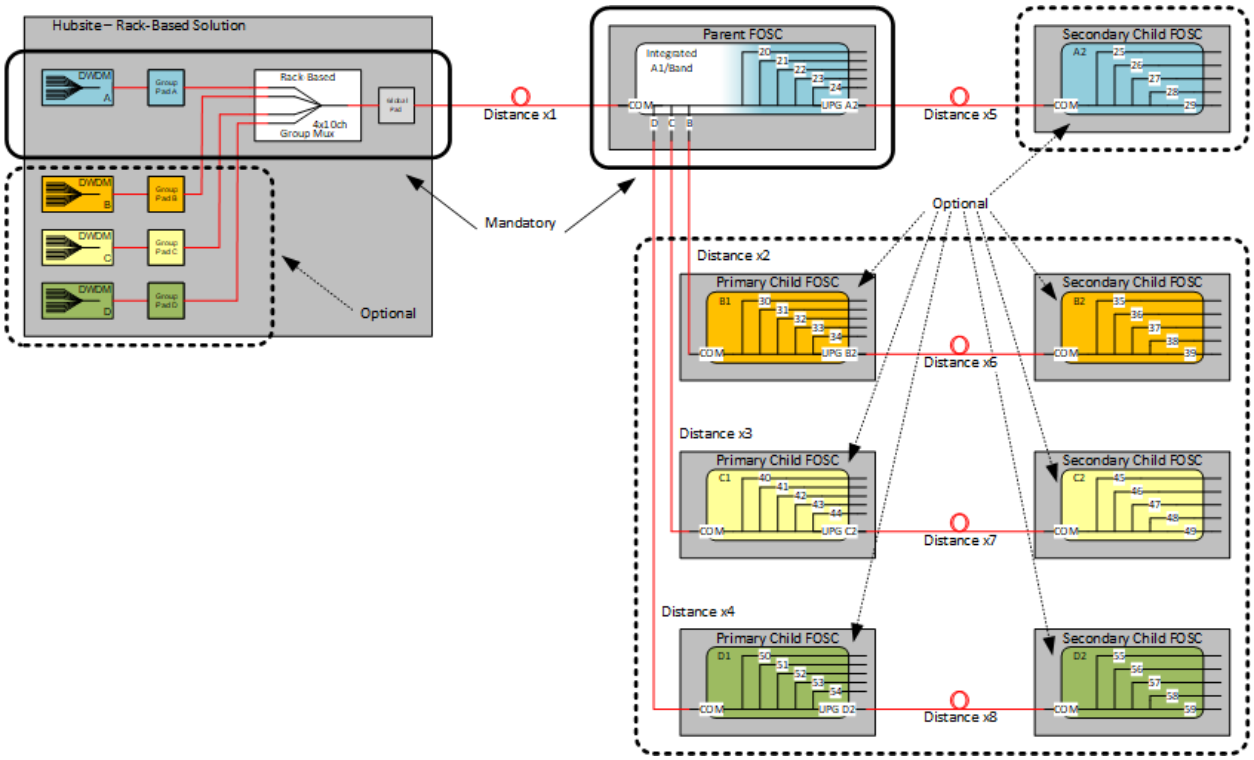


Figure 8 - 5ch/40ch Passive DWDM (MWL gen 2)

The equipment qualified and selected for the MWL gen 2 passive DWDM was tested with 10G wavelengths at all channels, and with higher bitrate signals of 100G, 200G, and 400G. There was some clipping observed on an Optical Spectrum Analyzer (OSA) for the 400G wave, but the Forward Error Correction (FEC) was sufficient for a clean Bit Error Rate (BER) Test of 72 hours over 50 km (lab spools of fiber). Therefore, if utilization of the MWL gen 2 passive DWDM systems start to congest with a plethora of 10G waves to a common Node Service Area, it will be possible to upgrade to higher bitrate wavelengths by using Point-to-Point Coherent waves with a Coherent Termination Device in the field to aggregate those smaller 10G demands. This is expected to future-proof the passive DWDM architecture.

With the introduction of this MWL gen 2 passive DWDM system, there is now a clear delineation for how Node Service Areas are to be served:

- Older analog RF nodes that have not yet been segmented will remain on non-shared fiber pairs.
- Analog RF nodes deployed after ~2015 will remain on MWL gen 1 passive DWDM systems.
- All Remote PHY Devices for DAA, all Remote OLTs for FTTP, all waves for Business services, all Small Cell & Wireless Backhaul waves, and all other wavelengths are to use the MWL gen 2 passive DWDM systems to share the Backbone Fiber.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

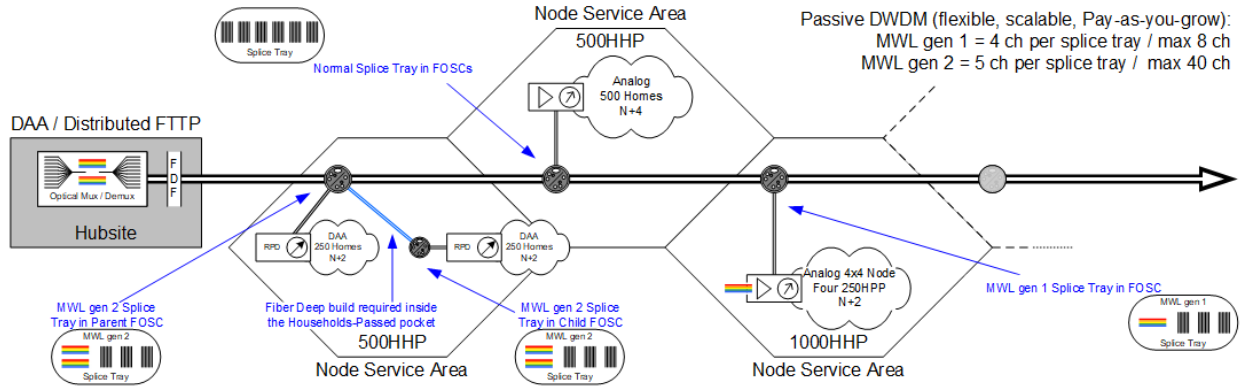


Figure 9 - Deploying Passive DWDM (MWL gen 2) into Node Service Areas

In this manner, DAA nodes can be deployed in Node Service Areas using the existing Backbone Fiber cables. Fiber can be built deeper into Node Service Areas and additional RPDs can be placed to allow for segmenting down from N+4 configurations (or higher) to N+2 configurations (or lower, if desired).

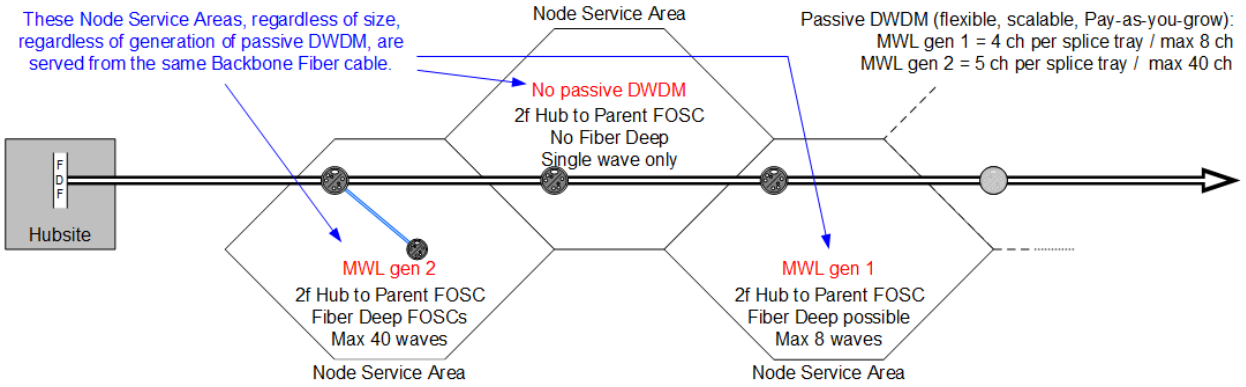


Figure 10 - Backbone Fiber Cable with Passive DWDM at different Node Service Areas

The important aspect under consideration is that the limitation of Backbone Fiber cable utilization remains; these are 288f cables and only 240f can be used for Node Service Areas. Adding passive DWDM filters and muxes allows for more efficient utilization of the fiber already in the ground.

There is no current plan to insert passive DWDM onto existing paths that do not have it already, however, all migrations from analog RF nodes to RPDs are required to include passive DWDM on the Backbone Fiber as part of that conversion process.

By pivoting from analog nodes with very specific fixed DWDM wavelengths over to RPDs with tunable optics, the move to DAA allows Cable Operators to save time, effort, and money. While narrowband DWDM optics and the additional filters add costs when compared to wideband 'grey' optics, this approach allows Cable Operators to defer, or maybe avoid altogether, costly Backbone Fiber overbuild programs for the Access Network.

There are many operational benefits for tunable DWDM optics that outweigh the higher cost when compared to fixed-wavelength DWDM optics, but that discussion is beyond the scope of this paper.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.2.1. Simplified Hubsite Configurations

The move to DAA will eventually lead to simplified Hubsite configurations.

One of the many promises of DAA is a significant reduction in Hubsite rackspace and power required if all service delivery is moved to a DAA configuration. This is both very true and highly unrealistic in the next 5 to 10 years. Changing Hubsite configurations (and analog nodes for RPDs in the Outside Plant) for existing customers is a lot of work that does not produce new revenue.

Unless there are pressing reasons for this DAA conversion for existing customers, such as a need to vacate the Hubsite due to a sale of the facility or being unable to renew a lease, or a need to upgrade to DOCSIS 4.0 for congestion or competitive reasons, or a need to remove End-of-Life / End-of-Support equipment, then the most likely operational mode for Cable Operators will be to Cap & Grow. Existing service delivery would be capped on existing platforms (CMTS, Centralized OLTs), and then new service delivery would be brought up on infrastructure used for DAA. This would include new RPDs (DOCSIS 4.0 upgrades) and Remote OLTs (new Distributed FTTP deployments in Greenfield SFUs, MDUs, and commercial / business parks), as well as Business and Small Cell / Wireless wavelengths (depending on the functionality of the CIN routers).

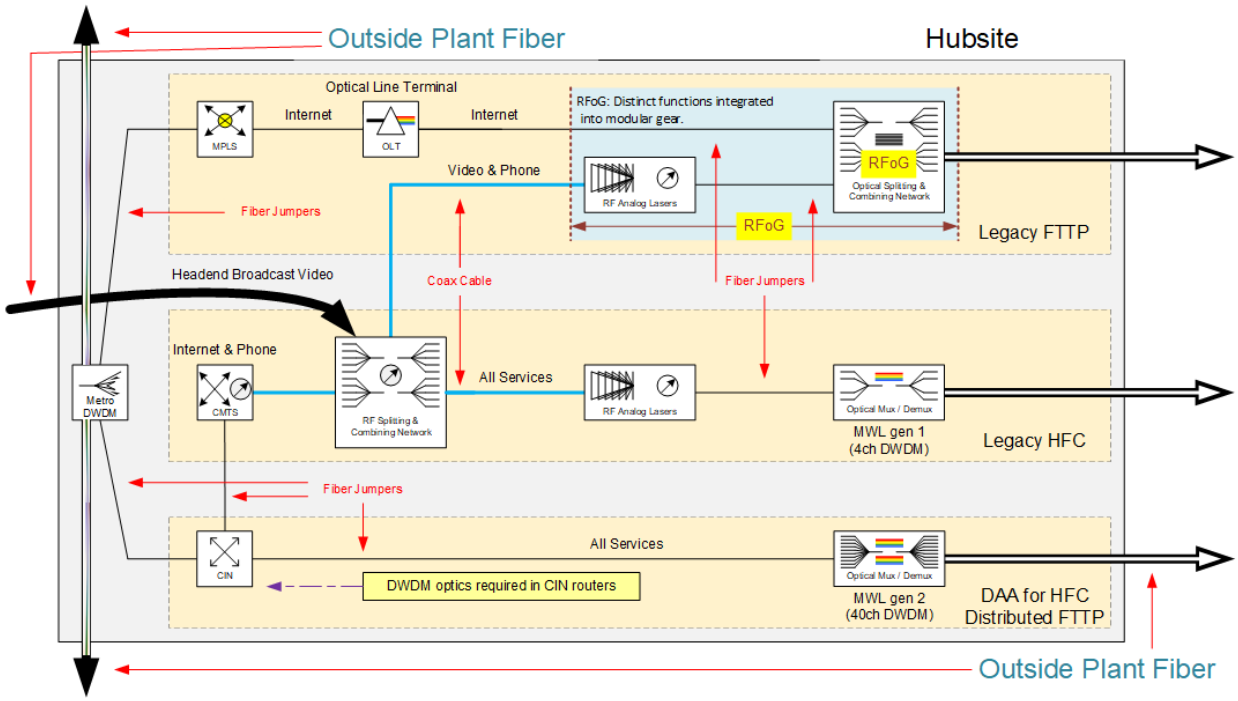


Figure 11 - Hubsite Configuration for currently-supported architectures

Hubsites will still be pressed for rackspace and power as this Cap & Grow model is implemented. CMTS links through Metro DWDM transponders back to Core Routers can be moved over to the CIN routers to alleviate some transponder costs; this might also be possible for Business Services routers (MPLS in the diagram above) depending on functionality.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

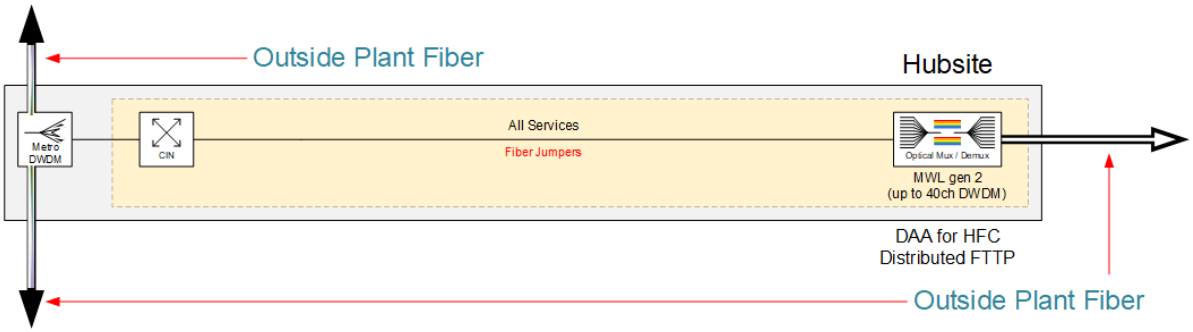


Figure 12 - Hubsite Configuration for DAA and Distributed FTTP

The end goal once all legacy HFC has been switched to DAA, and if Central OLTs can be replaced with Remote OLTs, is a simplified Hubsite configuration. This will take many years for most Hubsites. Future PON protocols may warrant new Centralized OLT equipment; consolidation & DAA conversion activities to make rackspace and power available at existing Hubsites would be required in Western Canada and likely for many Cable Operators.

4. Remote OLTs and Distributed FTTP

The terms Remote OLT and Distributed FTTP have been lightly-used previously in this Technical Paper. This discussion will now define these terms and proceed with discussing their merits and drawbacks.

4.1. Defining Remote OLTs

Remote OLTs are the functional equivalent for FTTP as a Remote PHY Device is for DAA. These Remote OLT devices have a small number of PON ports that terminate the ODNs for FTTP networks, have small number of Ethernet backhaul ports for connectivity back to a CIN router, and are environmentally-sealed and temperature-hardened with no moving parts (no fans). Similar to an RPD, a Remote OLT will have fins for heat dissipation, a coax seizure mechanism for power from an RF-isolated coax plant, and a fiber entrance port for a Node Service Cable. A Remote OLT will look like an RPD, but the inside guts will be different.



Figure 13 - Remote OLT Node Enclosure

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Perhaps controversially, there is a need to differentiate Remote OLTs from remotely-placed OLTs.

Table 1 – Defining Remote OLT Characteristics

Remote OLT	Remotely-Placed OLT
• Sealed Node Enclosure (sometimes referred to as 'clamshell') • 90 Vac from RF-isolated coax • No moving parts • Small number of PON and backhaul ports • Temperature hardened	• Rackmount in controlled environment (Hubsite, Environment Cabinet, etc) • -48 Vdc or 120 Vac power • Modular Chassis with different cards • # PON ports based on modules in slots • Requires fans

A rack-mount chassis-based OLT where the number of PON ports is expandable by adding modules to the chassis, where -48Vdc or 120 Vac powering is required, and where the temperature is both controlled in an environment cabinet and moderated by fans is considered to be a remotely-placed OLT.

Over the past 2 years, a total of four different Remote OLTs were put through evaluation and qualification activities in Western Canada, including 10G EPON and XGS-PON variants. On the physical specification requirements side, testing was done to ensure compliance with temperature conditions (-40 C to +60 C), powering needs (compatibility with 90 Vac coax power supplies), and Access Network compatibility (must utilize tunable DWDM optics to share the Backbone Fiber over MWL gen 2 passive DWDM).

Similar to RPDs that were evaluated, the Remote OLTs evaluated required pluggable optics to be rated for Industrial Temperature use (I-Temp, -40C to +85 C), and suffered from reboots during minor power bumps (such as may be experienced when a Plant Technician switches from utility-feed to battery at the coax power supply during routine maintenance activities).These reboots for Remote OLTs, similar to RPD behavior and performance, result in a loss of customer service for a duration measured in minutes as opposed to durations measured in seconds for power bumps on analog RF nodes.

Cable Operators looking at Remote OLTs will also need to consider Node Service Cable requirements.

Some of the Remote OLTs tested required a laptop for initial System Line-up and Test (SLAT). Unfortunately, this appears to be a holdover for systems designed for Telcos and then ported over to a form-factor suitable for Cable Operators. An automated, zero-touch provisioning system for initial deployments is preferable, and vendors for these Remote OLTs should be encouraged to adapt their products accordingly.

All four Remote OLTs that were tested used less power than typical RPDs; Cable Operators may want to look at smaller power supplies for deployments where only a single Remote OLT is required.

4.2. Defining Distributed FTTP

For this discussion, the physical architecture when discussing Distributed Access Architecture consists of:

- CIN routers in Hubsites
- Passive DWDM and tunable optics
- Remote PHY Devices terminating the coax portion of HFC service delivery

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Distributed FTTP is then defined for this discussion as the physical architecture consisting of:

- CIN routers in Hubsites
- Passive DWDM and tunable optics
- Remote OLTs terminating the ODN portion of PON service delivery

When talking about Distributed FTTP, this explicitly means a physical architecture that involves a Remote OLT as defined above.

4.3. Benefits

Moving the PHY layer closer to the customer has benefits with both DAA and Distributed FTTP.

For DAA, RPDs close to customers results in high modulation rates and hence more capacity on the plant for those customers.

For Distributed FTTP, Remote OLTs close to customers results in lower optical attenuation due to fiber distance, and as a result there is more optical power available; this allows for higher split ratios.

Legacy FTTP with Centralized OLTs standardized on 1:32 split ratios due to distances from Hubsites & Central Offices (COs) being relatively far. This was not too problematic for operators running 1G EPON or GPON because there was a limited amount of bandwidth to be shared between customers. With newer protocols like 10G EPON and XGS-PON, the bandwidth ceased to be a limiting factor, and hence higher split ratios become desirable as this is a more efficient use of available PON ports and fiber required from the PON port into the ODN. For Centralized OLTs, because of the distance to the customers, higher split ratios are not always possible, and therefore the Telco or Cable Operator is stuck with inefficient use of PON ports, and more importantly, inefficient use of the Backbone Fiber used for that PON port.

Remote OLTs have two strong benefits; by being close to the customers, higher split ratios result in a lower number of PON ports required for the same number of Households-passed. More importantly, by being able to utilize tunable DWDM optics for the backhaul, a Remote OLT is able to share fiber into a Node Service Area from the Backbone Fiber cable, and therefore this allows the operator to avoid or defer costly and intensive Backbone Fiber overbuild construction activities.

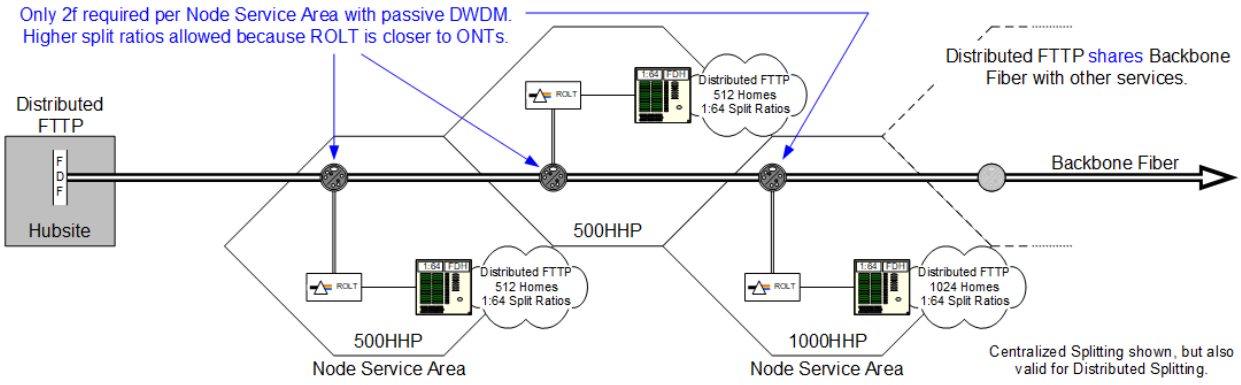


Figure 14 - Remote OLTs sharing Backbone Fiber through Passive DWDM

This also allows for unexpected growth within a Node Service Area without breaking the design parameters for the number of Backbone Fibers assigned to that Node Service Area, such as when a Node Service Area with RPDs is suddenly targeted by a developer for a new in-fill MDU.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

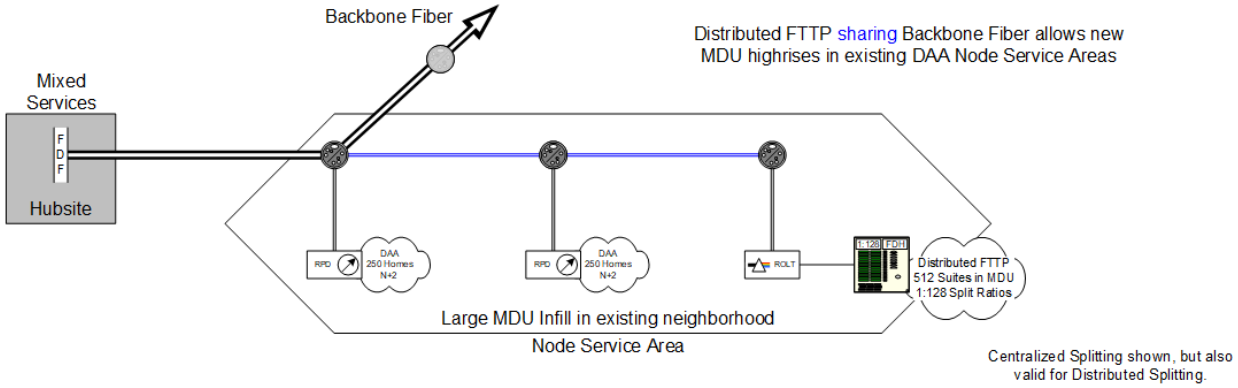


Figure 15 - New RPDs and Remote OLTs into existing Node Service Areas

This idea of adding services to an existing Node Service Area also applies to Rural Network Expansion opportunities where the expansion for that rural network would be single-ended. If the Node Service Area has passive DWDM but few active wavelengths, it would be possible to serve a low number of new Remote OLTs from that Node Service Area directly without consuming new fibers on the existing Backbone Fiber cable.

Distributed FTTP with Remote OLTs also creates an interesting dynamic where the linkage between Backbone Fiber utilization and PON port activation is severed. With Centralized OLTs, there is a 1:1 linkage between Backbone Fibers used per PON port activated. With Remote OLTs, there is no correlation between Backbone Fiber utilization and PON port activation; with a 40-channel passive DWDM system, the Backbone Fiber utilization now becomes dependent on how many channels an operator is assigning to a Remote OLT (for both capacity and protection), and how many PON ports need to be activated inside that Node Service Area, and further by the number of PON ports per Remote OLT. [All Remote OLTs tested in Western Canada had at least four PON ports per device, and some had more.]

With this linkage intentionally severed, additional PON ports can be lit without concern over Backbone Fiber usage. This allows operators to be intentionally inefficient in their PON port utilization. While this would require additional capital costs, it can be of great operational benefit. With Centralized OLTs, operators use Active Drop configurations to squeeze every available port out of a lit 1:32 splitter. As discussed above, this now requires a truck roll for every customer add or disconnect. By being intentionally inefficient on PON port utilizations, operators can instead adopt a “Hot Drop” configuration, whereby all drops are built, tested, and connected during the construction phase of the ODN’s lifecycle.

This greatly simplifies operations and maintenance at a central splitting location, reducing the need to visit the location and move fiber jumper connections for every single customer addition or disconnect.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 16 - FTC with Hot Drop Configuration

Hot Drop configurations also allows an operator to plan for extra drops to be available and spliced during the initial construction and build, useful for common scenarios experienced in Western Canada:

1. Basement suites rented to tenants in SFU houses & neighborhoods.
2. Professional Work-from-Home accounts requiring billing that is separate from a family account at the same address.
3. Convenience stores requiring Point-of-Sale accounts while co-located bank Automatic Teller Machines require separate accounts.

If limits are applied to how split ratios and distances are managed, these scenarios could also be accommodated with a 1:2 splitter at the side of the building. An operator would have to measure optical receive power levels prior to intentionally inserting additional optical loss into the path of an existing customer to add a 1:2 splitter; for this reason, having margin for additional drops pre-spliced and tested back to the central splitting location is ideal.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.4. Drawbacks

One of the drawbacks for Remote OLTs is that it is an active device requiring power and maintenance in the Outside Plant. For a Cable Operator with decades of experience placing actives in the field, this should be possible. For Telcos that have decommissioned & removed DSLAMs out in the field, this might be an untenable position. This creates uncertainty for vendors whereby Centralized OLTs are a known market (Telcos), but Remote OLTs have unknown potentials because they are an entirely new market for both products and customers (Cable Operators).

Another drawback of Remote OLTs is that the product cycle with the latest generation of protocols has lagged development of Centralized OLT products with the same protocols by several years. This is a function of temperature-hardening, form-factor, and the effort required to produce suitable prototypes, amongst other reasons. It is hoped that CableLabs' efforts on a next-generation Coherent PON (CPON) protocol will shorten or remove this gap that exists between Telcos able to deploy FTTP with Centralized OLTs and Cable Operators who require Backbone Fiber to be shared amongst services.

4.5. Considerations for Distributed FTTP specific to Cable Operators

Having briefly touched on the difference between Telcos and Cable Operators above, a focused discussion is warranted. The DOCSIS protocols have been wildly successful for Cable Operators providing Internet services to customers over the past two decades. Telcos with legacy twisted-pair networks were faced with a horrible reality and eventually acknowledged that they either needed to build expensive FTTP networks or they would be left out of (or even have to abandon) the Residential market for Internet, and potentially the market for Residential Voice, while leaving the potential market for Residential Video untapped.

In the United States, operators started ambitious FTTP programs like AT&T U-Verse, Verizon Fios, and Google Fiber, but these programs delivered a product roughly comparable to what was available with DOCSIS at Cable Operators at the time, and these FTTP programs scaled back as costs escalated.

In Canada, Telcos also started ambitious FTTP programs, but continued building as the costs per household-passed increased, partially-enabled by a falling interest rate environment that allowed for a debt-fueled spending binge.

With the protocols available at the time, companies that chose to build FTTP networks to salvage Residential markets knew that Centralized OLTs were required, and probably planned for 1:32 split ratios. They therefore knew that large Backbone Fiber constructions programs would be required everywhere in their serving areas.

In Western Canada, the regional Cable Operator's experimental FTTP deployment program was halted in 2011 when the two monumental limitations, both the lack of an IP Video solution and FTTP-related Backbone Fiber consumption, proved to be insurmountable with the technology solutions available at that time. There was still sunshine among the clouds; DOCSIS was a great service delivery method with a known architecture supported by sufficient infrastructure resources in both the Hubsite and the Outside Plant, and the protocols being used by the competitor Telcos did not provide a superior downstream for their Internet product.

In the years since the suspension of the experimental FTTP deployment program, success-based growth in network traffic required upgrades and augments to the infrastructure in Hubsites, where rackspace and power usage has now approached upper limits of what the Hubsites can support. The pivot to DAA has allowed for a cap & grow operational mode that has kept these constraints manageable so far, but these

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Hubsite constraints are still large enough to prevent a widescale deployment of Centralized OLTs across the network.

In the Hubsite, DAA requires approximately 1 port in a CIN router per 250 HHP, and Distributed FTTP appears to have similar requirements for unprotected links for Remote OLT backhaul connectivity.

Growth in upstream traffic has been accommodated by a 3-year mid-split (85 MHz) return program that was fortuitously completed just prior to the COVID-19 pandemic lockdowns.

Considering Brownfield overbuilds, recent deployments of DOCSIS 3.1 and the mid-split program, together with DOCSIS 4.0 coming soon, make it hard to justify overbuilding existing HFC plants with FTTP. Additionally, existing HFC costs points are currently much lower than FTTP on a per HHP basis. HFC is still a great architecture that has a long life ahead of it and that will still meet the needs of most customers for many years with good economics.

Considering fresh Greenfield builds, today's FTTP protocols of 10G EPON and XGS-PON provide adequate capacities, in both the downstream and upstream, while being capable of serving a higher number of customers per PON port than what was possible in the past. Developers are starting to prefer FTTP over HFC for new builds, and customers are generating more upstream traffic, enough to consider symmetrical services in their decision-making process. [Price still appears to be the primary consideration where customers have a choice between ISPs that use different technologies for their service delivery.]

With these thoughts in mind, the experimental FTTP deployment program was restarted in Western Canada a few years ago. The two prohibitive limitations had been resolved:

- An IP Video solution is now in place that allows for a Residential triple-play Voice, Video and Internet product offering.
- Remote OLT products are now available that allow for the use of passive DWDM and tunable optics so as to allow the Backbone Fiber to be shared amongst a variety of users.

There are now revenue-generating customers on multiple Distributed FTTP networks in Western Canada.

FTTP deployments are no longer generally-problematic for Cable Operators, but constraints in the existing Backbone Fiber might require an architecture like Distributed FTTP at other operators like it is required in Western Canada. Ultimately, the Backbone Fiber in Western Canada must be shared amongst many services (RPDs, Remote OLTs, Business services, Small Cell / Wireless services, etc), and it is anticipated that other Cable Operators face similar constraints.

To summarize, for the existing Access Network Infrastructure in Western Canada, Remote OLTs are absolutely required to enable a standardized architecture that simplifies planning, construction, deployment, acceptance testing, and service delivery for new FTTP customers in Greenfield SFU, MDU, Commercial Business Park, and Rural Network Expansion builds. Centralized OLTs consume too much Backbone Fiber, and there are not enough resources available in Hubsites, whether considering rackspace or power, to assume that new equipment can be added at existing Hubsite facilities.

4.6. Future PON Protocols

There are many developments with regards to future PON protocols, but the following discussion is meant to narrow the focus to those details specifically relevant to Distributed FTTP with Remote OLTs.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.6.1. 25G PON (25G EPON and 25GS-PON MSA)

The 25GS-PON Multi-Source Agreement (MSA) is an initiative spearheaded by Nokia and AT&T to create a market ecosystem for a PON protocol with a higher capacity than today's XGS-PON protocol. To summarize, the PHY layer essentially copies what the IEEE has specified for 25G EPON, while extending the Transmission Convergence (TC) layer used for XGS-PON.

As of July 2023, there were 54 member companies officially participating in the initiative, but notably absent were large suppliers Huawei and ZTE, and the large operator Verizon. Speculation as to why they have not signed on as of the writing of this Technical Paper is beyond the scope of this paper.

There are two specific benefits of 25GS-PON worth discussing in the context of Distributed FTTP when compared to 10G EPON or XGS-PON protocols:

1. 25GS-PON provides a higher aggregate data capacity over an ODN.
2. 25GS-PON provides a higher data capacity to a single subscriber.

For the first specific benefit of 25GS-PON to provide higher aggregate data capacities, the protocol should provide approximately 21 Gbps over the shared ODN as compared to a data rate of approximately 8.8 Gbps over an ODN running 10G EPON or XGS-PON protocols.

This first specific benefit of 25GS-PON is necessary for legacy architectures with Centralized OLTs because each PON port consumes a single strand of Backbone Fiber, and any constraints on Hubsite rackspace or power, or Backbone Fiber availability, will shackle the ability of the operator to add new revenue-generating customers if there are constraints or congestion anywhere in that network.

This first specific benefit of 25GS-PON is completely negated for Distributed FTTP architectures since lighting up new PON ports on Remote OLTs does not require new assignments of Backbone Fiber. In a Distributed FTTP network, a congested ODN with a 1:128 split ratio can be easily segmented into two ODNs with 1:64 split ratios, or four ODNs with 1:32 split ratios. Another alternative would be to 'cherry-pick' high-traffic users off the congested ODN and onto a new ODN with a smaller split ratio.

Operators with Distributed FTTP architectures do not require 25GS-PON to provide higher aggregate data capacities to customers.

In Western Canada, the splitters at the central splitting location are deliberately planned with 1:2 and 1:4 splitters cascading into 1:32 splitters so that segmentation activities can be easily accomplished in future years. This adds a small amount of additional insertion loss from the additional connectors, but that loss is easily overcome because the Remote OLTs are so close to the customers.

For the second specific benefit of 25GS-PON to provide higher data capacity to a single subscriber, the customer premise equipment would need to support higher flows for 25G Ethernet, probably through a pluggable SFP28 module. If customers have high traffic loads that consistently exceed the capacity of a 10G EPON or XGS-PON network, then a dedicated wavelength service would be more appropriate for them over a shared-media PON service.

For operators with Centralized OLTs, or for operators with distributed splitting, it is difficult to provide a dedicated wavelength service as separate fibers and splicing activities are required. For operators running a Distributed FTTP architecture with central splitting and with passive DWDM between the Hubsite and the Node Service Area, adding a dedicated wavelength bypass service would still require splicing at the central splitting location and at the FOSC(s) with the passive DWDM breakout, but no new Backbone

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Fibers would be required. A router port with DWDM optics in the Hubsite would be required to terminate the wavelength.

For those customers with a high peak burst traffic approaching 20 Gbps, but with much lower day-to-day traffic, a shared-media PON service running 25G EPON or 25GS-PON protocols could be a workable solution. Operators with Distributed FTTP architectures should evaluate Remote OLTs capable of these 25G PON protocols for these customers with specialized needs of this burst traffic that exceeds the capabilities of today's 10G EPON and XGS-PON protocols. Some of the participants in the 25GS-PON MSA are working on 25GS-PON Remote OLTs, and it is believed that there are some 25G EPON Remote OLT products also available or in development.

If operators do not have customers with high peak burst traffic patterns, AND these operators have Distributed FTTP architecture, AND future generations of PON protocols develop as expected (CableLabs Coherent PON Working Group, for example), then the decision as to whether to deploy 25G PON protocols (25G EPON or 25GS-PON) would come down to economics and cost points.

Another thing to consider for 25G PON specific to Remote OLTs is the Ethernet backhaul component. Development is ongoing for pluggable 25G optics that meet the requirements for devices in the field:

- Industrial Temperature (-40 C to +85 C)
- Distances preferably up to 80 km (optical power and dispersion tolerance)
- DWDM full-band tunable to share Backbone Fiber

Products that meet these requirements hopefully will be generally available in 2024.

If two 10G Ethernet pluggable optics are used, then the full capacity of the 25G protocol would not be available back to the router, and only a single 25G PON port would be available at the Remote OLT.

For operators needing multiple 25G PON ports activated at single Remote OLT, then 100G DWDM wavelengths could be used, like those envisioned and standardized by the CableLabs Coherent Optics P2P Working Group, published specification P2PCO-SP-PHYv1.0-I03.

An interesting dynamic for Cable Operators thinking about deploying Remote OLTs with 25G PON is that Remote PHY Devices for DOCSIS 4.0 might have similar backhaul requirements, if multiple serving groups are fed out of the same Node Enclosure, depending on the achievable throughput.

Table 2 – Expected DOCSIS 4.0 Modulation Throughput

Modulation Rate (QAM)	Theoretical Throughput (bits/s/Hz)	Achievable Throughput (bits/s/Hz)
256	8	6
1024	10	7.5
2048	11	
4096	12	9
8192		
16384		10.5

The DOCSIS 4.0 spectrum available for downstream (DS) is dependent on where the split for upstream (US) return occurs.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 3 – Expected DOCSIS 4.0 Downstream Spectrum Usage

DS Lower Band Edge	54 MHz	87.5 MHz	108 MHz	258 MHz	372 MHz	492 MHz	606 MHz	834 MHz
192 MHz DS blocks	9	9	9	8	8	7	6	5
Total DS Spectrum (MHz)	1728	1728	1728	1536	1536	1344	1152	960

Noting that achievable throughput varies and is lower than the theoretical throughput, an estimate for the downstream traffic possible on a DOCSIS 4.0 network can be made based on the number of 192 MHz downstream blocks are possible (which is a function of the where the upstream split is, as per the DOCSIS 4.0 specifications).

Table 4 – Expected DOCSIS 4.0 RPD Backhaul Requirements

192 MHz Blocks	Achievable QAM	Achievable Throughput (bits/s/Hz)	Usable Downstream Spectrum (MHz)	Capacity Calculation (Gbps), per Serving Group
1	4096	9	192	1.728
2	4096	9	384	3.456
3	4096	9	576	5.184
4	4096	9	768	6.912
5	4096	9	960	8.64
6	4096	9	1152	10.368
7	4096	9	1344	12.096
8	4096	9	1536	13.824
9	4096	9	1728	15.552

It is expected that initial iterations of DOCSIS 4.0 RPDs will only be capable of 1 serving group, and therefore only two 10G Ethernet backhaul connections combined in a Link Aggregation Group (LAG) would be required. Even at 12 bits/s/Hz and 1536 MHz of downstream spectrum (upstream 5 MHz to 300 MHz, downstream 372 MHz to 1794 MHz), a single serving group would require 18.432 Gbps and thus be satisfied by a 20 Gbps LAG for backhaul.

Regardless of whether an operator is looking at 25G PON in Remote OLT, or DOCSIS 4.0 in an RPD, once the number of PON ports or serving groups grows beyond one per Node Enclosure, the backhaul will start to get non-trivial and will require some planning and forethought. This is not expected for a number of years, but will need to be a consideration for how Remote OLTs and Remote PHY Devices are designed in the future.

4.6.2. 50G PON (50G EPON and an ITU-T variant)

For Cable Operators with Distributed FTTP, 50G PON variants are not interesting. These 50G PON variants address the same challenges as 25G PON, namely higher aggregate data and higher data to a singular customer. For higher singular data flow requirements, Distributed FTTP with central splitting allows problematic traffic flows, whether 25G or 50G, to be moved to separate ODNs or dedicated

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

wavelength services without additional Backbone Fiber, as discussed above. For aggregate data flow requirements, Distributed FTTP with Remote OLTs allows for easy segmentation of congested ODNs.

4.6.3. 100G PON (CableLabs Coherent PON Working Group)

In Western Canada, any PON protocol that cannot easily share Backbone Fiber must be relegated to Remote OLTs; building more Backbone Fiber is not an acceptable pre-requisite for connecting new FTTP customers.

If a PON protocol was made available that could easily share the Backbone Fiber by being tunable across a wide range of the DWDM spectrum currently used to share this Backbone Fiber, and if the optical link budget was sufficient enough to overcome distances and insertion losses typical for these passive DWDM deployments, while still providing a high enough split ratio to be considered efficient, then a Centralized OLT solution would be acceptable.

Remote OLTs are not necessarily desirable because they become another active in the field that requires planning, permitting, construction, deployment, and on-going maintenance. Remote OLTs are not being deployed as a function of choice; rather, Remote OLTs are being deployed because Centralized OLTs do not work with the 10G PON protocols today for the requirements that Cable Operators face, namely using existing Backbone Fiber sized and built decades ago.

This presents a gap in PON protocol developments. The market adoption of XGS-PON vs NG-PON2 at Telcos leads to the assumption that future development at the ITU-T will not prioritize sharing of Backbone Fibers. CableLabs' Coherent PON Working Group is working on a specification for a PON protocol for a next-generation 100G coherent PON (CPON). For deployment in Western Canada, CPON will need to be able to share the Backbone Fiber with other DWDM wavelength services.

When considering the differences between Centralized OLTs and Remote OLTs, the backhaul portion for Remote OLTs at 10 Gbps is similar to RPDs running DOCSIS 3.1, and is a known quantity supportable by the CIN routers used for the initial DAA deployments. However, at 100 Gbps DWDM backhaul, the costs are much higher for that 100 Gbps backhaul link with Remote OLTs. This is because 100G router ports and 100G tunable DWDM optics will be required, which are currently generally more expensive for Cable Operators than 10G ports.

This 100G cost premium becomes a Day 1 cost incurred at the initial deployment of the Remote OLT regardless of the number of planned customers. If the tunability (required for sharing the Backbone Fiber) and the optical link budget (required for high split ratios and long distances) can be obtained through design of the OLT and ONTs with sufficient specification and forethought in the PON protocol, then the economics place the cost burden on a success-based ONT deployment model rather than a Day 1 backhaul requirement equipment investment burden. This allows for a Pay-as-You-Grow model to be applied for deploying 100G PON overlays on top of existing 10G PON deployments. This is desirable for upgrading individual existing customers on existing ODNs, and will extend the product lifecycle for the 10G FTTP networks.

The scenario for minimizing 100G PON ONT costs with limited tunability and optical link budget specifications will tend to force deployment models at Cable Operators towards Distributed FTTP with Remote OLTs and therefore lead to higher Day 1 deployment costs. This might be acceptable in larger Greenfield builds but will be problematic when trying to proceed with success-based upgrades of existing 10G PON deployments.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

For the upgrade scenario where a CPON overlay is being used to cherry-pick problematic traffic flows off existing 10G PON ODNs, Centralized CPON OLT with tunability and high optical link budget specifications split at a Hubsite feeding multiple Node Service Areas should be dramatically simpler, and hopefully cheaper, than building new Remote OLTs at each of the Node Service Areas where this traffic grooming effort is required. An operator could even proactively deploy CPON signals into Node Service Areas with existing 10G FTTP ODNs and migrate the highest 2 or 3 customers per ODN from 10G PON to CPON and thus alleviate congestion before other customers even start to notice.

This ability to cherry-pick customers does require a central splitting architecture; a complete comparison of different splitting architectures is beyond the scope of this paper, but central splitting is used in Western Canada.

As previously discussed in this paper, an experimental FTTP deployment program was halted a decade ago because of unacceptable Backbone Fiber overbuild requirements; the CableLabs CPON protocol must not shackle Cable Operators in the future with a similar requirement. A low cost ONT for Cable Operators that forces Backbone Fiber overbuilds will result in a much-higher total overall deployment cost that will ultimately slow, or even halt, 100G PON deployments at Cable Operators in the future.

It is therefore recommended that the CPON protocol development efforts focus on PHY-layer specifications that meet the tunability and optical link budget specifications that Cable Operators need rather than having a narrow focus on ONT material costs for a total overall lowest-cost ONT.

As a cautionary note, deployments of equipment in Hubsites always require rackspace and power; as such, while this is still many years away, Cable Operators should proceed with DAA conversions so that rackspace and power can be reclaimed in sufficient quantities and timeframes so as to allow for new Centralized OLT deployments once future generations of PON protocols are available (such as CPON). Remote OLTs can be used for 10G PON deployments, and probably 25G PON deployments, but it is unclear whether this will be desirable, let alone possible, for 100G PON protocols and beyond.

5. Conclusion

It is possible for operators to deploy FTTP architectures for Greenfield MDU, SFU, Commercial Business Park, and Rural Network Expansion builds. For Cable Operators who have migrated to an IP Video solution and who have Backbone Fiber cables that were sized & built decades ago, existing DAA equipment can be leveraged to accelerate the deployment of FTTP.

Distributed FTTP architecture utilizes new Remote OLTs for PON port activations, and together with existing DAA equipment in both the Hubsite (CIN routers) and Outside Plant (passive DWDM), allows Cable Operators to light up new ODNs while decoupling the linkage between PON port utilization and Backbone Fiber utilization. Cable Operators can now deploy FTTP for customers without worrying about exhausting scarce Backbone Fiber resources, which was an insurmountable constraint for those operators trying to deploy legacy FTTP architectures.

Care should be taken that future PON protocols allow Cable Operators to revert to FTTP architectures with Centralized OLTs, rather than forcing Cable Operators to adopt Distributed FTTP architectures with Remote OLTs. This will allow Cable Operators to choose architectures and parameters best suited to their needs, rather than be forced into a specific architecture with limited product availability. This will require future PON protocols to be tunable across multiple wavelengths and to have sufficient optical link budget specifications to overcome passive DWDM insertion losses, fiber attenuation loss from distance, and optical power loss from power splitters inside the ODN.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

BER	bit error rate
bps	bits per second
CableLabs	Cable Television Laboratories
CIN	Converged Interconnect Network
CO	central office
CPON	coherent passive optical network
DAA	Distributed Access Architecture
DOCSIS	Data-Over-Cable Service Interface Specifications
DPoE	DOCSIS provisioning of EPON
DS	downstream
DWDM	dense wavelength division multiplexing
EPON	Ethernet passive optical network
FDH	fiber distribution hub
FEC	forward error correction
FOSC	fiber optic splice cabinet
FTC	fiber transition cabinet
FTTP	fiber-to-the-premises
GPON	gigabit passive optical network
HFC	hybrid fiber-coaxial
HHP	households-passed
Hz	hertz
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet Protocol
ITU-T	Telecommunication Standardization Sector of International Telecommunication Union
LAG	link aggregation group
LCP	local convergence point
MDU	multiple dwelling unit
MPLS	multiple protocol label switching
MSA	multi-source agreement
MTTR	mean-time-to-repair
MWL	multiple wavelength
ODN	optical distribution network
OLT	optical line terminal
ONT	optical network terminal
OSA	optical spectrum analyzer
PHY	physical layer
PON	passive optical network
RFoG	Radio-Frequency-over-Glass
RPD	remote PHY device
SFP	small formfactor pluggable
SFU	single-family unit
SIP	Session Initiation Protocol
US	upstream
XGS-PON	10-Gigabit-capable symmetric passive optical network

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

CableLabs CM-SP-PHYv4.0-I06-221019: *Data-Over-Cable Service Interface Specifications DOCSIS 4.0 – Physical Layer Specification*

CableLabs CPON-SP-ARCH-I01-230503: *Coherent Passive Optical Networks 100 Gbps Single-Wavelength PON – Coherent PON Architecture Specification*

CableLabs P2PCO-SP-PHYv1.0-I03-200501: *Point-to-Point Coherent Optics – P2P Coherent Optics Physical Layer 1.0 Specification*

ITU-T G.9807.1 Amendment 1: *10-Gigabit-capable symmetric passive optical network*

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Leveraging DOCSIS® 4.0 CM Backward Compatibility on DOCSIS 3.1 Networks

For an Increase in Capacity and Downstream Speed Tiers

A Technical Paper prepared for SCTE by

Karthik Sundaresan
Distinguished Technologist
Cable Television Laboratories Inc.
858 Coal Creek Circle, Louisville, CO USA
+1 303.661.9100
k.sundaresan@cablelabs.com

Doug Jones
Principal Architect
Cable Television Laboratories Inc.
858 Coal Creek Circle, Louisville, CO USA
+1 303.661.9100
d.jones@cablelabs.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
1.1. History 4
2. DOCSIS Evolution and Capabilities 5
2.1. DOCSIS 3.0 Technology 5
2.2. DOCSIS 3.1 Technology 5
2.3. DOCSIS 4.0 Technology 6
2.4. DOCSIS 3.1 and DOCSIS 4.0 Device Channel Capabilities and Peak Speeds 7
2.5. Transition from DOCSIS 3.1 Networks to DOCSIS 4.0 Networks 8
2.6. Updating the DOCSIS 3.1 CMTS 9
2.6.1. Additional OFDM channels with channel bonding 9
2.6.2. Support DOCSIS 4.0 CM registration 10
2.6.3. Other Features 10
3. Cable Modem Registration 11
3.1. DOCSIS 4.0 CM Capabilities 11
3.2. Registration on a DOCSIS 3.1 CMTS and CMTS Band Edge Capabilities 12
3.3. DOCSIS 4.0 FDD CM on a D3.1 CMTS 13
3.4. DOCSIS 4.0 FDX CM on a D3.1 CMTS 14
4. Results from Lab Testing 15
4.1. Equipment 16
4.2. Speeds obtained with D3.1 CMTS and D4.0 CMs 16
4.2.1. Downstream 32 SC-QAM + 4 OFDM channels 16
4.2.2. Downstream of 5 OFDM Channels 18
4.2.3. Upstream with 2 OFDMA Channels 19
4.2.4. Fastest DOCSIS 3.1 Configuration 19
4.2.5. Speeds including DOCSIS 3.0 and Earlier Modems 20
5. Conclusions 20

Abbreviations 21

Bibliography & References 22

List of Figures

Title **Page Number**

Figure 1 – Simplified DOCSIS 4.0 CM Registration 11
Figure 2 - D4.0 FDD CM, D3.1 CM w High-Split DiplexFilter, on High-Split D3.1 CMTS 14
Figure 3 - D4.0 FDD CM, D3.1 CM w Mid-Split Diplex Filter, on Mid-split D3.1 CMTS 14
Figure 4 - D4.0 FDX CM, D3.1 CM w Mid-Split Diplex Filter, on a Mid-split D3.1 CMTS 15
Figure 5 - D4.0 FDX CM, D3.1 CM w High-Split Diplex Filter, on High-Split D3.1 CMTS 15
Figure 6 – Downstream with 32 SC-QAM and 4 OFDM Channels 17
Figure 7 – Close-up of First Seven QAM Channels 17
Figure 8 – Downstream with 5 OFDM Channels 18
Figure 9 – Upstream with 2 OFDMA Channels 19

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Tables

Title **Page Number**

Table 1 – DOCSIS Equipment minimum channel capabilities 7
Table 2 – CM Peak Downstream Capacities with enhanced DOCSIS 3.1 CMTS 7
Table 3 – DOCSIS 4.0 CM Capabilities 12
Table 4 – DOCSIS 3.1 CMTS MDD Diplexer Band Edge Value 13

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Every generation of DOCSIS® technology includes requirements for backward compatibility with the previous generation to enable seamless transitions. With a DOCSIS 4.0 cable modems (CM), there is opportunity to further exploit the capabilities of a currently deployed DOCSIS 3.1 cable modem termination system (CMTS).

Specification work was completed clarifying the behavior of DOCSIS 4.0 CMs with DOCSIS 3.1 CMTSs. This paper examines these possibilities, it discusses how a DOCSIS 3.1 CMTS negotiates capabilities with a DOCSIS 4.0 CM, how a DOCSIS 4.0 CM, either FDD (frequency division duplex, (also known colloquially as Extended Spectrum DOCSIS), and FDX (full duplex) will operate with various diplex filters on the DOCSIS 3.1 network, and how a DOCSIS 4.0 CM can be used to increase capacity on an existing HFC network.

With recent product upgrades, a DOCSIS 3.1 CMTS can now support up to five (or more) OFDM channels; however, currently available DOCSIS 3.1 CMs only support two OFDM channels that have a peak downstream rate of about 5 Gbps. Operating a DOCSIS 4.0 CM on a DOCSIS 3.1 CMTS (with additional OFDM channels), can support a downstream peak capacity of up to 8.8 Gbps—a 75% increase relative to currently deployed DOCSIS 3.1 CMs. With higher capacities, operators can provide enhanced services and faster downstream speed tiers even before implementing DOCSIS 4.0 network topology changes and installing DOCSIS 4.0 CMTSs.

For DOCSIS 4.0 CMs to become operational, a DOCSIS 3.1 CMTS may need a software update. This update enables a DOCSIS 3.1 CMTS to gracefully interpret and negotiate DOCSIS 4.0 modem capabilities, thereby enabling DOCSIS 4.0 CMs to be operated on a DOCSIS 3.1 CMTS. This paper touches on other dependencies such as support by the CMTS for bigger certificate sizes, and bonding on more than two OFDM channels, support for OUDP testing and overlapping channels. We will also report on lab testing results that we plan to conduct.

1.1. History

From a historical perspective, DOCSIS 3.0 technology had several iterations of CM silicon, each supporting an increasing number of bonded single-carrier quadrature amplitude modulation (QAM) (SC-QAM) channels. The DOCSIS 3.1 specifications (which introduced orthogonal frequency division multiplexing (OFDM) technology) were designed the same way, mandating that a CM support a minimum of two OFDM channels (in addition to 32 SC-QAM channels) while allowing a CM to support more than two OFDM channels. Currently deployed DOCSIS 3.1 CMs, however, only support two OFDM channels in the downstream direction, and DOCSIS 3.1 CMs with additional OFDM channels are no longer anticipated, as the industry's energy and focus are now directed toward DOCSIS 4.0 CM designs.

In contrast, while early DOCSIS 3.1 CMTS solutions supported only two OFDM channels, updated designs now support up to five OFDM channels. The additional OFDM channels on these more capable DOCSIS 3.1 CMTSs can be used by a DOCSIS 4.0 CM to provide much higher downstream peak capacity. As cable operators enhance the available spectrum in their HFC networks—such as moving the upper band edge up to 1.0 GHz or 1.2 GHz—spectrum is opened to support those additional OFDM channels. (This outside plant work is typically combined with efforts to increase the upstream bandwidth by transitioning from a sub-split to either a mid-split or high-split HFC network.)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2. DOCSIS Evolution and Capabilities

2.1. DOCSIS 3.0 Technology

DOCSIS 3.0 Technology was introduced in 2006~2008 and included a new feature called channel bonding which led to a substantial increase in broadband capacity. Prior to the DOCSIS 3.0 specifications, modems used a single downstream QAM channel and a single upstream QAM channel.

Downstream Channel Bonding enabled functionality at a CM to receive data simultaneously on multiple receive channels. Downstream Channel Bonding refers to the ability (at the MAC layer) to schedule packets for a single service flow across those multiple channels. Upstream Channel Bonding enabled a CM to transmit simultaneously on multiple transmit channels. This allowed the CMTS to schedule the traffic for a single upstream service flow across those multiple channels.

Channel Bonding offers significant increases in the peak data rate that can be provided to a single CM, in both directions. Other enhancements in the upstream request-grant process like Continuous Concatenation and Fragmentation (CCF) also improved the efficiency of the upstream link.

DOCSIS 3.0 technology also introduced support for the Internet Protocol version 6 (IPv6). CMs can be provisioned with an IPv4 management address, an IPv6 management address, or both. Further, CMs can provide transparent IPv6 connectivity to devices behind the cable modem (CPEs), with full support for Quality of Service and filtering. DOCSIS 3.0 technology also introduced support for delivery of Source-Specific IP Multicast streams to CPEs.

DOCSIS 3.0 technology started off with minimum requirements of 4 Quadrature Amplitude Modulation (QAM) channels in the downstream and 4 QAM channels in the upstream. After starting with 4 QAM channels and enabling channel bonding, future iterations of DOCSIS 3.0 products, CMs and CMTS, enabled more channels in each direction. In the downstream the number of channels (bonded together) grew from 4 to 8 to 16 to 24 to 32 channels (32 channels provided the means for up to 192 MHz of downstream DOCSIS spectrum). In the upstream direction the number of channels grew from 4 to 8 SQAM upstream channels.

At that time most of the cable spectrum was used for video programming, and DOCSIS 3.0 technology ushered in a new era of dedicating more spectrum to broadband and the beginning of using less spectrum for the traditional video offering. This also led to the introduction of MPEG-4 video coding technology, however, the driver was to allocate more spectrum to broadband.

2.2. DOCSIS 3.1 Technology

DOCSIS 3.1 technology was introduced in 2013~2015 and added Orthogonal Frequency Division Multiplexing (OFDM) downstream signals and Orthogonal Frequency Division Multiple Access (OFDMA) upstream signals to achieve robust operation and provide more efficient use of the spectrum than previous DOCSIS versions. To reach the higher service tiers in the upstream direction, plant changes on the upstream/downstream spectrum split were also created. The DOCSIS 3.1technology allowed for a system to have options of several split configurations that can be chosen by an operator based on traffic demand, services offered and the capability of the cable plant. The upstream split options available were 42 MHz, 85 MHz, and 204 MHz of spectrum. In the downstream direction, HFC network spectrum was designed to start at 108 MHz or 258 MHz and extend up to 1.2 GHz of spectrum.

The OFDM downstream multicarrier system is composed of many subcarriers that have either 25 kHz or 50 kHz spacing. These subcarriers are grouped into independently configurable OFDM channels each

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

occupying a spectrum of up to 192 MHz in the downstream, totaling 7680 25 kHz subcarriers or 3840 50 kHz subcarriers; of which up to 7600 (25 kHz) or 3800 (50 kHz) active subcarriers span 190 MHz of spectrum. The OFDMA upstream multicarrier system is also composed of either 25 kHz or 50 kHz subcarriers. In the upstream, the subcarriers are grouped into independently configurable OFDMA channels each of up to 95 MHz encompassed spectrum, totaling 3800 25 kHz spaced subcarriers or 1900 50 kHz spaced subcarriers or 1920 50 kHz spaced subcarriers. Many parameters of these channels can be independently configured thereby optimizing configuration based on channel conditions.

In DOCSIS 3.1, the OFDM/OFDMA technology allows different subcarriers to use different modulation orders. This is referred to as variable bit-loading on the channel. A downstream profile will define the modulation order (i.e., bit-loading) on each carrier. To account for varying downstream plant conditions across different devices, an operator can define multiple downstream profiles, where each profile can be tuned to account for specific plant conditions. By optimizing the downstream profiles, this will allow a downstream channel to be able to operate with lower SNR margin, potentially allowing a channel to operate at an overall higher throughput. For OFDMA Channels, a minislot is a set of symbols and subcarriers and different modulations across minislots are allowed. This enables IUCs that allow different modems to transmit with different modulations in the upstream.

The DOCSIS 3.0 channels are referred to as single carrier QAM (SC-QAM) systems in contrast to the multicarrier DOCSIS 3.1 OFDM/OFDMA system.

DOCSIS 3.1 technology allowed the continued use of channel bonding between SC-QAM and OFDM/OFDMA channels. The result was an increase in the spectrum that could be allocated to broadband. DOCSIS 3.1 technology required support for a minimum of 32 SC-QAM channels and 2 OFDM channels in the downstream, and support or 8 SC-QAM channels and 2 OFDMA channels in the upstream. This provided the means for up to 576 MHz of downstream DOCSIS spectrum and now the move was away from legacy video technology and the pivot toward IPTV service.

2.3. DOCSIS 4.0 Technology

DOCSIS 4.0 technology is expected to be introduced in the 2024 timeframe and builds on the proven DOCSIS 3.1 OFDM and OFDMA technology, providing even more capacity. DOCSIS 4.0 technology leverages the existing DOCSIS3.1 Media Access Control (MAC) and Physical (PHY) layers. It includes backward compatibility for the existing PHY layers to enable a seamless migration to the new technology. DOCSIS 4.0 technology is enabled by two key modes of operation: frequency division duplex and full duplex modes of operation.

Full Duplex (FDX) DOCSIS technology is a new set of features that significantly increases upstream capacity and allows for the same spectrum to be used as downstream or upstream. FDX allows for simultaneous upstream and downstream transmission and reception within the same cable system spectrum. The FDX technology increases upstream capacity without significant loss of downstream capacity as compared to DOCSIS 3.1 technology. The FDX node transmits downstream and receives upstream at the same time on the same frequencies. FDX transmissions from node to modem and modem to node are overlapped both in time and frequency where new echo cancellation technology in the node allows this simultaneous bidirectional communication. While an FDX CMTS will simultaneously receive and transmit in the same FDX spectrum, FDX CMs can only receive or transmit at a time in the same FDX spectrum.

Frequency division duplex (FDD) DOCSIS technology is the traditional frequency division duplex mode of operation that has been operating for years and now supports significant increases in upstream and downstream capacity. FDD supports legacy high-split (204 MHz) and provides extended upstream splits

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

up to 684 MHz in an operational band plan which is referred to as ultra-high split (UHS). There are 4 new UHS options available including 300 MHz, 396 MHz, 492 MHz, or 684 MHz for the upstream upper band edge. FDD also increases the downstream bandwidth by extending the DOCSIS downstream upper band edge to 1794 MHz of spectrum.

DOCSIS 4.0 technology will provide a means for up to 1,536 MHz of downstream spectrum, almost tripling what was available with DOCSIS 3.1 technology. DOCSIS 4.0 technology envisions all services being provided over broadband and in fact, some operators are already moving in this direction.

2.4. DOCSIS 3.1 and DOCSIS 4.0 Device Channel Capabilities and Peak Speeds

DOCSIS 3.1 technology and DOCSIS 4.0 technology are very similar. DOCSIS 3.1 specifications laid the groundwork for including OFDM and OFDMA technology. DOCSIS 4.0 technology provides methods to get more of these channels on the coaxial cable network to provide even higher speeds.

Table 1 – DOCSIS Equipment minimum channel capabilities

DOCSIS Version	Device Type	Downstream OFDM Channels	Upstream OFDMA channels
DOCSIS 3.1	CM	2	2
	CMTS	2	2
DOCSIS 4.0	CM	5	7
	CMTS	6	8

Table 1 presents the minimum number of downstream OFDM channels supported on DOCSIS-compliant equipment. As noted above, currently deployed DOCSIS 3.1 CMs support two downstream OFDM channels, whereas several upgraded DOCSIS 3.1 CMTSs now support more than two OFDM channels. A DOCSIS 4.0 CM supports at least five downstream OFDM channels. A DOCSIS 4.0 CMTS will support 6 OFDM downstream channels.

Also on a related note, the upstream OFDMA channel capability for CMs increases from 2 (for DOCSIS 3.1 CMs) to 7 (for DOCSIS 4.0 CMs) and on the CMTS side increases from 2 to 8. The focus of this paper is increased capacity on DOCSIS 3.1 networks, so the increase in upstream channels is not as useful on a DOCSIS 3.1 network, given 2 OFDMA channels are enough to cover a mid-split or high-split plant.

Table 2 – CM Peak Downstream Capacities with enhanced DOCSIS 3.1 CMTS

CMTS	Number of downstream channels on Service Group		Spectrum allocated to DOCSIS Downstream (MHz)	CM	Capacity of Service group/ Peak Downstream Speeds at CM
	OFDM	SCQAM			
D3.1	2	32	576	D3.1	5.0 Gbps
D3.1	3	32	768	D4.0	6.9 Gbps
D3.1	4	32	960	D4.0	8.8 Gbps
D3.1	5	0	960	D4.0	9.5 Gbps

Table 2 shows how device capabilities relate to capacity of the downstream with a mixture of DOCSIS 3.1 and DOCSIS 4.0 CMs operating on a DOCSIS 3.1 CMTS. These speeds indicate the downstream capacity of the plant and also equate to the peak downstream speeds possible at a CM, assuming the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

bandwidth is not shared with any other CMs on the plant. The speeds in Table 2, are 'goodput' speeds available for the user data traffic, i.e., after considering the DOCSIS overhead (including FEC, MAC management and other PHY layer overhead). For an understanding of how to calculate speeds on a OFDM and OFDMA channel and DOCSIS 3.1 or 4.0 networks please refer to [D31 Capacity], [10G Capacity]. The speeds in Table 2 are not the product speeds or service tiers, that an MSO would offer, this is the total capacity of the plant, typically operators offer speed tiers well within the capacity of the network or the peak speed enabled by a CM. See [Capacity Model] for further analysis on network capacity planning. Some of the scenarios in lab testing that achieved some of these speeds are described in Section 4.2.

In the current baseline scenario, a DOCSIS 3.1 CM supports a peak of up to 5 Gbps downstream speed (using both OFDM and SC-QAM channels). In the past, D3.1 CMTS supported only two OFDM channels in a MAC domain. These CMTS are typically deployed in a mix of sub-split, mid-split (more prevalent) and high-split spectrum plans, with the downstream extending up to 1.2 GHz , all of which are the design parameters for a DOCSIS 3.1 network. A DOCSIS 4.0 network on the other hand increases the upstream by enabling ultra-high-split spectrum for the upstream and extending the downstream , or enabling full duplex operations, see section 2.3 and 2.5.

Now recently, there have been new product upgrades to existing DOCSIS 3.1 CMTS for use on the existing DOCSIS 3.1 Networks. DOCSIS 3.1 CMTSs have added support for additional OFDM channels in the downstream, within the same MAC domain. This upgraded DOCSIS 3.1 CMTS can support 3, or 4 or 5 OFDM channels in the downstream. To clarify, this is not a DOCSIS 4.0 CMTS and it doesn't understand the any new MAC management messages from DOCSIS 4.0 technology, neither does it support any of the new DOCSIS 4.0 features such as ultra-high splits, extended downstream, or full duplex operations.

A DOCSIS 4.0 CM which supports five OFDM channels can make use of these new OFDM channels on a DOCSIS 3.1 CMTS and a DOCSIS 3.1 network. A DOCSIS 3.1 CMTS that supports four OFDM channels paired with a DOCSIS 4.0 CM can achieve up to 8.8 Gbps downstream speed. Also keep in mind that to enable these additional OFDM channels, an operator does need to plan their spectrum assignment to various services and may need to take away bandwidth from say QAM video to enable spectrum for data. A DOCSIS 3.1 CMTS that supports five OFDM channels paired with a DOCSIS 4.0 CM can achieve up to 9.5 Gbps downstream speed. In this case all the bandwidth on the plant is allocated to DOCSIS spectrum, meaning there is no spectrum for QAM video , assumption here is video services have moved to be delivered over IP.

The actual speeds achieved will depend on the specific OFDM channel parameters and the signal to noise ratio of the fiber node or plant. When the channel parameters are tuned for robust operation, the speed could be different. In initial lab tests we proved out 8.8 Gbps speeds in a lab network tests, on a simple coax network, see notes in section below.

2.5. Transition from DOCSIS 3.1 Networks to DOCSIS 4.0 Networks

DOCSIS 3.1 technology was straight-forward to introduce onto the existing HFC network. Initially, spectrum had to be cleared for the first downstream OFDM channel and it was possible to get started with half of an OFDM channel, 96 MHz, instead of a full 192 MHz of spectrum. And the upstream was initially not changed and many operators in North America remained at the usual sub-split (42 MHz) configuration. Over time many operators made the upstream transition to mid-split (85 MHz) and some to high-split (204 MHz) configuration and introduced OFDMA channels in that new upstream spectrum.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

DOCSIS 4.0 technology can be more fully utilized with the evolution of the HFC network. The existing coaxial plant can carry the additional DOCSIS 4.0 spectrum once the actives and passives on the network need to be changed to enable all the spectrum capabilities. This equipment is available, and some operators have already begun evolving the network to support DOCSIS 4.0 technology.

The HFC network evolution will provide more spectrum on the coaxial cable that will be allocated to broadband. This new spectrum will be used with OFDM and OFDMA technology, which will rapidly become the new normal for providing broadband service, as compared to the SC-QAM technology which had been used through DOCSIS 3.0 technology. OFDM and OFDMA channels bring in new operational and PNM activities which operators are quickly getting familiar with. DOCSIS 4.0 FDD technology will be based on the access network upgrades for existing actives and passives up to 1.8 GHz along with a choice of diplex ultra-high split configurations. DOCSIS 4.0 FDX technology will be based on upgrades to introduce new Echo Cancellation functions into the RPHY nodes and amplifier platforms.

A DOCSIS 4.0 network uses a Distributed Access Architecture (DAA) which provides better end-of-line performance as compared to a traditional integrated headend CMTS. See [10G Playbook] for further details. The DAA technologies place part of the CMTS in the fiber node and are known as either Remote MACPHY or Remote PHY technology, see CableLabs specifications for both. DOCSIS 4.0 technology will need an upgrade in the network to migrate to 1.8 GHz DOCSIS 4.0 FDD DAA nodes or DOCSIS 4.0 FDX nodes (with Echo cancellation)

DOCSIS 4.0 technology drives additional changes including the method for signal leakage detection, legacy video services, CableCARD™ technology, and other technologies that are beyond the scope of this paper. While the DOCSIS 4.0 network upgrades are ongoing, updating the current DOCSIS 3.1 CMTS, Section 2.6, enables operators to seed the network with DOCSIS 4.0 CMs.

2.6. Updating the DOCSIS 3.1 CMTS

The DOCSIS 3.1 specifications introduced OFDM and OFDMA technology to the DOCSIS ecosystem. The DOCSIS 4.0 specifications added support for more OFDMA and OFDMA channels, and so the DOCSIS 4.0 CM can support five OFDM channels as compared to a DOCSIS 3.1 CM supporting two OFDM channels, see Table 1.

2.6.1. Additional OFDM channels with channel bonding

DOCSIS 3.1 CMTS equipment was originally designed to provide two OFDM channels on each service group, to match the DOCSIS 3.1 CM. With the advent of DOCSIS 4.0 CMs, it is possible to modify the DOCSIS 3.1 CMTS to provide more than two OFDM channels per service group, even up to five OFDM channels per service group. Generally, this can be done without new CMTS equipment; however, check with the CMTS supplier if this new operation impacts the number of service groups that can be served with this increased number of OFDM channels.

The DOCSIS 3.1 CMTS will also need to bond the additional OFDM channels with the existing downstream OFDM channels and the SC-QAM channels, thereby providing the additional downstream capacity for the DOCSIS 4.0 CMs. The CMTSs already supports different bonding groups, which will continue to be needed, to allow DOCSIS 3.0 , 3.1 and the new 4.0 CMs to operate seamlessly.

With the addition of up to 5 OFDM channels on the downstream, there are now new options for a DOCSIS 3.1 CM, specifically which two OFDM channels to use. Hence the DOCSIS 3.1 CMTS may want to introduce both downstream bonding groups for DOCSIS 3.1 CMs and a method for downstream load balancing for DOCSIS 3.1 CMs. A DOCSIS 4.0 CM will use all the OFDM channels available,

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

however, a DOCSIS 3.1 CM can only use two of the OFDM channels and there should be a method to get the DOCSIS 3.1 CMs on certain OFDM channels.

On the upstream, the DOCSIS 3.1 CMTS already supports mid-split or high-split configuration, and this includes the two OFDMA channels available per service group. Based on the available spectrum for upstream in a high split plant, DOCSIS 4.0 CMs should be able to use all the upstream spectrum without adding any more OFDMA channels per service group on the CMTS. That is, the DOCSIS 3.1 CMTS supports two OFDMA channels per service group and will continue to support those two OFDMA channels in a high-split configuration when DOCSIS 4.0 CMs are introduced. A DOCSIS 4.0 CM is required to operate on a high-split network.

2.6.2. Support DOCSIS 4.0 CM registration

2.6.2.1. CM Capabilities

Backwards compatibility for CMs and CMTSs has been a long time goal of DOCSIS protocol design and specification. The DOCSIS 3.1 CMTS needs to gracefully interpret and negotiate DOCSIS 4.0 modem capabilities, thereby enabling DOCSIS 4.0 CMs to operate on a DOCSIS 3.1 CMTS, albeit only with the DOCSIS 3.1 features.

Though a DOCSIS 3.1 CMTS should have supported the newer DOCSIS 4.0 CMs to come online and operate as DOCSIS 3.1 CMs, as per the DOCSIS specifications, the reality of DOCSIS 3.1 CMTS software implementations was that these CMs were initially unable to register. This was about a year ago, and since then the quick software fixes to DOCSIS 3.1 CMTSs already enable the recognition of the DOCSIS 4.0 CMs and bring them online. DOCSIS 4.0 CMs are backward compatible with DOCSIS 3.1 CMTS in an interoperable fashion. A DOCSIS 3.1 CMTS negotiates the best possible service offering for a DOCSIS 4.0 CM, considering several factors, including the number of OFDM channels, diplex filter, etc., supported by the CM.

Operators need to ensure that their DOCSIS 3.1 CMTS has the appropriate software updates to recognize and register DOCSIS 4.0 CMs on the DOCSIS 3.1 network. Registration of DOCSIS 4.0 CMs on DOCSIS 3.1 CMTSs has now been proven in the lab in the latest CableLabs DOCSIS 4.0 Interoperability events.

2.6.2.2. Larger Auth-Req

There are other dependencies such as support by the CMTS for bigger certificate sizes. A DOCSIS 3.1 CMTS needs to allow support for DOCSIS 4.0-style CM certificate and the associated larger Auth-REQ (authorization) message for BPI+ V1. Without this support, an operator cannot turn on BPI+ security when running a DOCSIS 4.0 CM on a DOCSIS 3.1 CMTS. A DOCSIS 3.1 CMTS is not required to support BPI+ V2 though does need to support the new DOCSIS 4.0 device certificate and the associated larger Auth-REQ message; both features have been proven in the latest CableLabs DOCSIS 4.0 Interoperability events.

2.6.3. Other Features

There are other features that would be nice to have on a DOCSIS 3.1 CMTS, but not necessary to bring up DOCSIS 4.0 CMs on the CMTS. OUDP Testing (OFDMA Upstream Data Profile) is a part of the DOCSIS 3.1 technology with CM vendors including support for it. In high-split networks, OUDP used as a signal to detect RF signal leakage. The DOCSIS 3.1 CMTS will need to support a mechanism to trigger the CMTS to command high-split CMs to transmit OUDP bursts to enable RF leakage testing.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Overlapping OFDMA channels (OOC) is a CMTS feature that allows operators to create a logical upstream channel that overlaps a physical one. The CMTS essentially creates logical channels such that it looks like there are two OFDMA channels: a "base" channel and an "overlap" channel. This can then be used to support mid-split CMs on a high-split CMTS without capacity loss. DOCSIS3.1 mid-split CM cannot use a partial OFDMA channel, if say a channel goes above the 85 MHz in the spectrum. Without OOC, operators will have to discard at least 23 MHz of upstream spectrum (between 85 to 108 MHz) when operating in a high split plant (204 MHz). A DOCSIS 4.0 FDX CM when booting on a DOCSIS 3.1 CMTS will boot as a mid-split CM and so the benefits extend to those CMs in the plant as well.

3. Cable Modem Registration

DOCSIS systems follow a common registration process where the CM reports capabilities that are negotiated with the CMTS, with the CMTS deciding how a CM uses the network (based on operator policy). Figure 1 shows a simplified registration process where the CM presents its capabilities to the CMTS, the CMTS decides which services and channels that CM can have access, and then provides a service configuration back to the CM.

During registration, a DOCSIS 4.0 CM must identify its version as "DOCSIS 4.0" to the CMTS. This will allow operators to both track DOCSIS 4.0 CMs on their networks and have their DOCSIS 3.1 CMTS enforce policy regarding how those DOCSIS 4.0 CMs become operational on the DOCSIS 3.1 network. As described in Section 2.6.2, a DOCSIS CMTS may need a software fix to recognize the DOCSIS 4.0 CMs correctly.

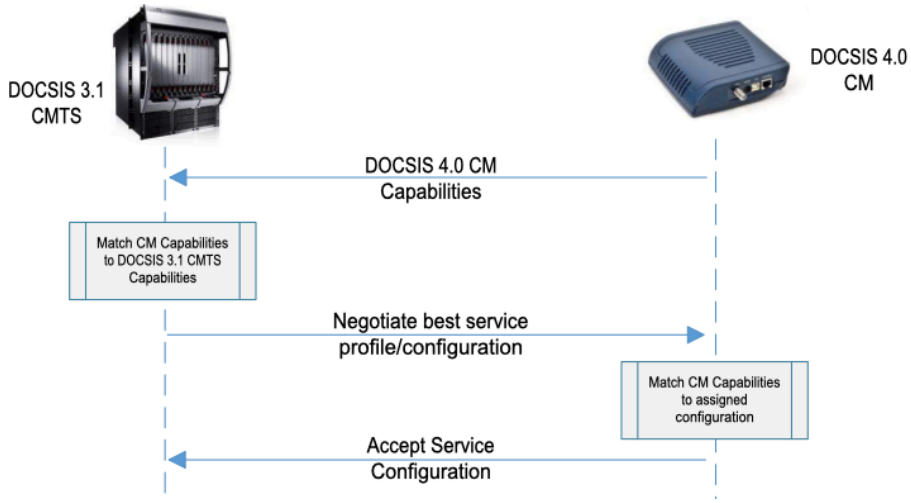


Figure 1 – Simplified DOCSIS 4.0 CM Registration

3.1. DOCSIS 4.0 CM Capabilities

Every CM reports its capabilities to the CMTS. Table 3 shows the relevant DOCSIS 4.0 CM capabilities, such as the new ultra-high-split DOCSIS 4.0 diplex filters and lists the new values that the DOCSIS 4.0 CM will indicate for some of the legacy DOCSIS 3.1 CM capabilities.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 3 – DOCSIS 4.0 CM Capabilities

Description	Capability Sub-TLV	CM Capability	D4.0 FDX CM	D4.0 FDD CM
Version	5.2	DOCSIS Version	v4.0	v4.0
Channel Support	5.49	OFDM Multiple Receive Channel Support	5	5
	5.50	OFDMA Multiple Transmit Channel Support	7	7
Legacy Band Edge Support	5.62	Diplexer Upstream Upper Band Edge Options	85	204
	5.60	Diplexer Downstream Lower Band Edge Options	108	258
	5.61	Diplexer Downstream Upper Band Edge Options	1218	1794
New D4.0 Capabilities	5.63	Adv. Band Plan Capability	FDX	FDD
	5.84	Adv. Diplexer Upstream Upper Band Edge Options List	-	204/300/396/ 492/684
	5.82	Adv. Diplexer Downstream Lower Band Edge Options List	-	258/372/492/ 606/834
	5.83	Adv. Diplexer Downstream Upper Band Edge Options List	-	1794

A DOCSIS 3.1 CMTS needs to support parsing of these new values from the legacy CM capability TLVs (Type/Length/Value) and to gracefully ignore the new DOCSIS 4.0 capabilities that it does not support. Based on the capabilities indicated by the CM for the diplexer band edges (TLVs 5.60, 5.61, 5.62) and matching them to the configuration of the DOCSIS 3.1 network (see description of MDD , Section 3.2), the CMTS will allocate appropriate channels to the CM.

In a DOCSIS system, the modem is required to report its capabilities to the CMTS. It is then the role of the CMTS to decide how to allow that modem to come online. Hence, a DOCSIS 3.1 CMTS needs to be upgraded to understand and gracefully handle the values in the capabilities of a DOCSIS 4.0 CM, at least for the relevant DOCSIS 3.1 CM capabilities.

3.2. Registration on a DOCSIS 3.1 CMTS and CMTS Band Edge Capabilities

A significant factor in the process of a DOCSIS 4.0 CM registering on a DOCSIS 3.1 CMTS relates to the various upstream and downstream bands that can be used (along with the associated diplex filter settings) and is different for each type of DOCSIS 4.0 CM.

The MAC Domain Descriptor (MDD) message is used to convey various DOCSIS system settings from the CMTS to all CMs. The MDD Diplexer Band Edge Sub-TLV (21) indicates the diplexer upstream and downstream band edges to which the HFC network is configured. The Table 4 shows the values of this TLV on different HFC networks.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 4 – DOCSIS 3.1 CMTS MDD Diplexer Band Edge Value

MDD Sub-TLV	CMTS MDD Sub-TLV Name	D3.1 Sub-Split	D3.1 Mid-Split	D3.1 High-Split
17	CMTS DOCSIS Version	v3.1	v3.1	v3.1
21.1	Diplexer Upstream Upper Band Edge	42	85	204
21.2	Diplexer Downstream Lower Band Edge	108	108	258
21.3	Diplexer Downstream Upper Band Edge	1218	1218	1218

DOCSIS 3.1 CMs can have various diplex filters and DOCSIS 4.0 CMs support additional diplex filters. A CMTS will consider the combinations of CM diplex filters and allocate the right channels to each CM.

Upon detecting the broadcast MDD message and the band edge settings, a DOCSIS 3.1 CM and a DOCSIS 4.0 CM will reset its diplex filter to best match the MDD message, and then the CM will reboot and attempt to register with the CMTS. Below are some of the rules that describe the expected behavior.

- The CM sets its diplexer upstream upper band edge to the highest upstream frequency that it supports within the range reported in the Diplexer Upstream Upper Band Edge MDD sub-TLV.
    - If the supported CM diplexer upstream upper band edge frequencies are greater than the range reported in the MDD message, the CM sets its diplexer upstream upper band edge to the lowest upstream frequency that it supports.
- The CM sets its diplexer downstream lower band edge to the lowest downstream frequency that it supports within the range reported in the Diplexer Downstream Lower Band Edge MDD sub-TLV.
    - If the supported CM diplexer downstream lower band edge frequencies are less than the range reported in the MDD message, the CM sets its diplexer downstream lower band edge to the highest downstream frequency that it supports.
- The CM sets its diplexer downstream upper band edge to the highest downstream frequency that it supports within the range reported in the Diplexer Downstream Upper Band Edge MDD sub-TLV.
    - If the supported CM diplexer downstream upper band edge frequencies are greater than range reported in the MDD message, the CM sets its diplexer downstream upper band edge to the lowest downstream frequency that it supports.

For a DOCSIS 4.0 CM, as shown in Table 3, the legacy band edge support CM capabilities, describing the diplexer band edges have one set of values for the DOCSIS 4.0 FDD CM (204-258MHz) , and one set of values for the DOCSIS 4.0 FDX CM (85-108MHz). So, per the MDD message diplexer setting rules above, these DOCSIS 4.0 CMs will always use these diplexer values and boot up with those diplexer settings.

As a related note, if a CMTS calculates that any of the supported CM diplexer band edge frequencies (as indicated by the CM capabilities) are outside the frequency range reported in the Diplexer Band Edge MDD TLV, the CMTS can allow or disallow the CM to register depending on operator policy. [MULPIv3.1]

3.3. DOCSIS 4.0 FDD CM on a D3.1 CMTS

A DOCSIS 4.0 FDD CM is required to support at least two diplex filters, including both high-split and one of the new ultra-high splits. Thus, on a DOCSIS 3.1 network, the FDD CM will use a high-split diplex filter because that is the filter that most closely aligns with the capabilities of a DOCSIS 3.1 network. An FDD CM can also operate using its high-split diplex filter on any of the following: a sub-, European-, mid-, or high-split DOCSIS 3.1 network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

For a high-split DOCSIS 3.1 network, the downstream begins at 258 MHz. With its high-split diplexer an FDD CM can only use downstream channels above 258 MHz in the spectrum. If the DOCSIS 3.1 CMTS supports more than two OFDM channels, this DOCSIS 4.0 CM can take advantage of these channels both for more capacity and higher speeds. On the upstream, an FDD CM will get the most upstream speeds out of a high-split network. A sample spectrum plan of a DOCSIS 3.1 high-split CMTS is shown in Figure 2.

A DOCSIS 4.0 FDD CM will reconfigure its diplexer, from one of the Ultra High splits that it supports and will use the 204-258 MHz diplexer when it boots up on a DOCSIS 3.1 network.

On a DOCSIS 3.1 high-split network, with the channel configuration as shown in Figure 2, the main difference between the DOCSIS 3.1 high-split CM and the DOCSIS 4.0 FDD CM is that the DOCSIS 4.0 CM can use the additional 2 OFDM channels to enable higher bandwidth service tiers.

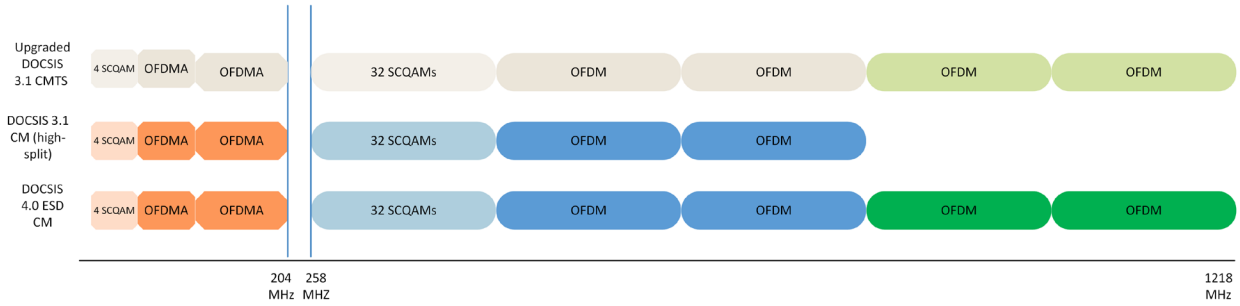


Figure 2 - D4.0 FDD CM, D3.1 CM w High-Split DiplexFilter, on High-Split D3.1 CMTS

The FDD CM can also operate on a sub- or mid-split HFC network and use the available upstream channels, with the caveat that because that FDD CM will still be using a high-split diplex filter, so it can only use DOCSIS downstream channels above 258 MHz in the spectrum. This is shown in Figure 3.

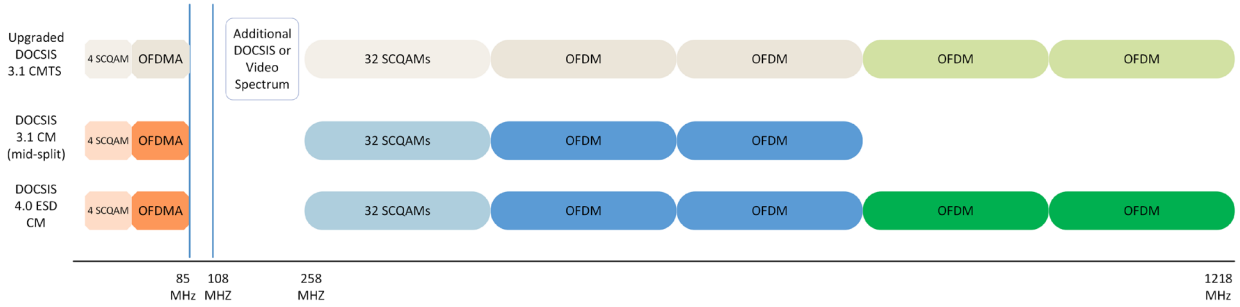


Figure 3 - D4.0 FDD CM, D3.1 CM w Mid-Split Diplex Filter, on Mid-split D3.1 CMTS

3.4. DOCSIS 4.0 FDX CM on a D3.1 CMTS

A DOCSIS 4.0 FDX CM is required to support two modes of operation; one in FDX mode and the other as a mid-split CM. Thus, on a DOCSIS 3.1 network, the FDX CM will use a mid-split diplex filter as that is the filter that most closely aligns with the capabilities of a DOCSIS 3.1 network. An FDX CM can also operate using its mid-split diplex filter on any of the following: a sub-, European-, mid-, or high-split DOCSIS 3.1 network.

A DOCSIS 3.1 CMTS does not support FDX technology. On a DOCSIS 3.1 network, a DOCSIS 4.0 FDX CM is expected to come online as a mid-split CM with no transition band from 684 MHz to 804

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

MHz in the spectrum. Figure 4 shows that an FDX CM will register as mid-split and use upstream spectrum up to 85 MHz and downstream spectrum starting at 108 MHz up to the limit of where the HFC network is operated.

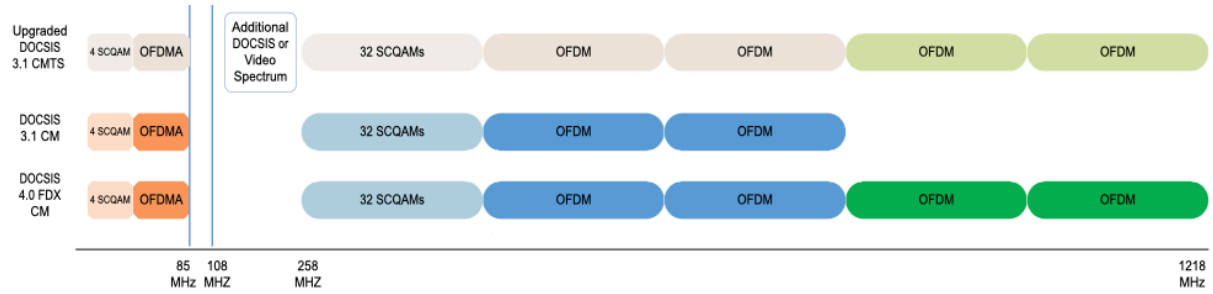


Figure 4 - D4.0 FDX CM, D3.1 CM w Mid-Split Diplex Filter, on a Mid-split D3.1 CMTS

An FDX CM can be deployed on any DOCSIS 3.1 network and is best suited for a mid-split network. For a mid-split DOCSIS 3.1 network, the downstream begins at 108 MHz in the spectrum. With its mid-split diplexer an FDX CM can use downstream channels above 108 MHz in the spectrum. If the DOCSIS 3.1 CMTS supports more than two OFDM channels, this DOCSIS 4.0 CM can take advantage of these channels both for more capacity and higher speeds. On the upstream, this DOCSIS 4.0 FDX CM will be able to use all the upstream capacity in the mid-split network as also shown in Figure 4.

As shown in Figure 5, the FDX CM will become operational on any DOCSIS 3.1 network, including sub-split or high-split networks, and will only be able to use the OFDMA channels that are below 85 MHz and the OFDM channels that are above 108 MHz in the spectrum.

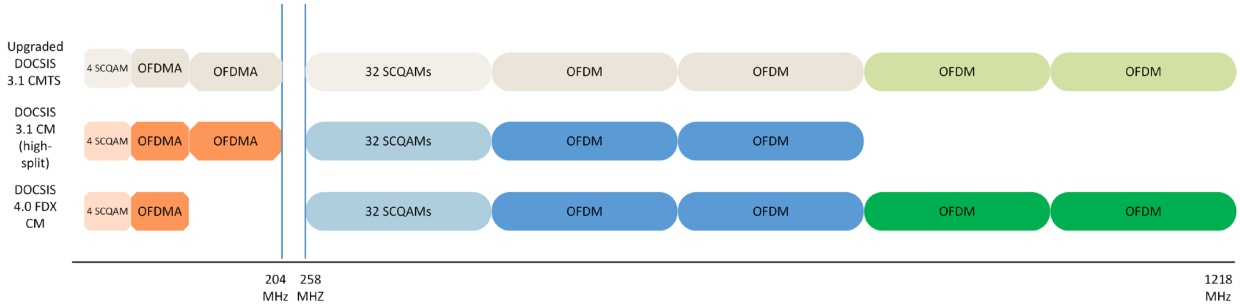


Figure 5 - D4.0 FDX CM, D3.1 CM w High-Split Diplex Filter, on High-Split D3.1 CMTS

4. Results from Lab Testing

CableLabs did quite a bit of lab work with the latest DOCSIS equipment this year. This included DOCSIS CMTS setup and bring up of DOCSIS 4.0 CMs. The results in this section are for a DOCSIS 3.1 CMTS with the cable spectrum filled with only DOCSIS channels. There are no video QAMs or other signals on the coaxial cable. This work was done in a lab with passive coaxial cable, that is, no amplifiers.

The speeds listed below are measured throughputs for the downstream and upstream channels. The downstream capacity exceeds that of a single DOCSIS 3.1 CM, hence, multiple high-split DOCSIS 3.1 CMs would be needed to utilize this capacity. In the case of a DOCSIS 4.0 CM, it can utilize all the capacity described in this section.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.1. Equipment

CableLabs has been working with high-split DOCSIS 3.1 CMTS equipment from multiple suppliers and in multiple form-factors.

Our lab network included the following CMTS setups:

- Integrated CMTS
- Integrated CMTS acting as a CCAP-Core hosting a Remote PHY Device (RPD)
- Virtual CCAP-Core hosting an RPD
- Remote MACPHY Device (RMD)

All these CMTS equipment are the latest generation (hardware and software) of equipment that can support a high-split network where the return path runs up to 204 MHz and the forward path runs from 258 MHz to 1218 MHz for a total of 960 MHz downstream spectrum which can all be filled with DOCSIS channels.

This is way more downstream than was available in either 750 MHz or 860 MHz networks and even more downstream spectrum than is available on a sub-split 1 GHz plant. And now, all this spectrum can be allocated to DOCSIS services with the latest generation of DOCSIS 3.1 CMTS.

For example, that downstream 960 MHz can be filled with downstream channels in a few ways:

- 32 x 6-MHz SC-QAM channels (total of 192 MHz) and 4 x 192 MHz OFDM channels (total of 768 MHz). The use of downstream SC-QAM channels allows DOCSIS 3.0 CMs to operate.
- 5 x 192 MHz-OFDM channels. With no SC-QAM channels, this network setup only supports DOCSIS 3.1 CMs and DOCSIS 4.0 CMs.

The latest generation DOCSIS 3.1 CMTS (with support for an increased number of OFDM channels) paired with DOCSIS 4.0 CMs will provide a useful increase in speed and capacity. DOCSIS 3.1 and earlier CMs can still be supported on the same network.

4.2. Speeds obtained with D3.1 CMTS and D4.0 CMs

4.2.1. Downstream 32 SC-QAM + 4 OFDM channels

The first network that we built had a DOCSIS 3.1 downstream spectrum that spans from 258 MHz to 1218 MHz, with 32 SC-QAM channels, each 6 MHz for a total of 192 MHz, and 4 OFDM channels, each 192 MHz wide. These channels fill 960 MHz from 258 MHz to 1218 MHz in the spectrum. Figure 6 shows a spectrum capture of the SC-QAM channels from 258 MHz to 450 MHz, and the spectrum from 450 MHz to 1218 MHz being occupied by 4 OFDM channels.

A DOCSIS 4.0 CM can successfully register on this system and use all the downstream channels. Both the DOCSIS 4.0 CMs (both FDD and FDX) can successfully register on the system and were able to use the 4 OFDM channels along with the 32 SC-QAMs. This DOCSIS 4.0 CM was successfully able to

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

receive traffic which the CMTS had bonded on the downstream across all of the SC-QAM and OFDM channels.

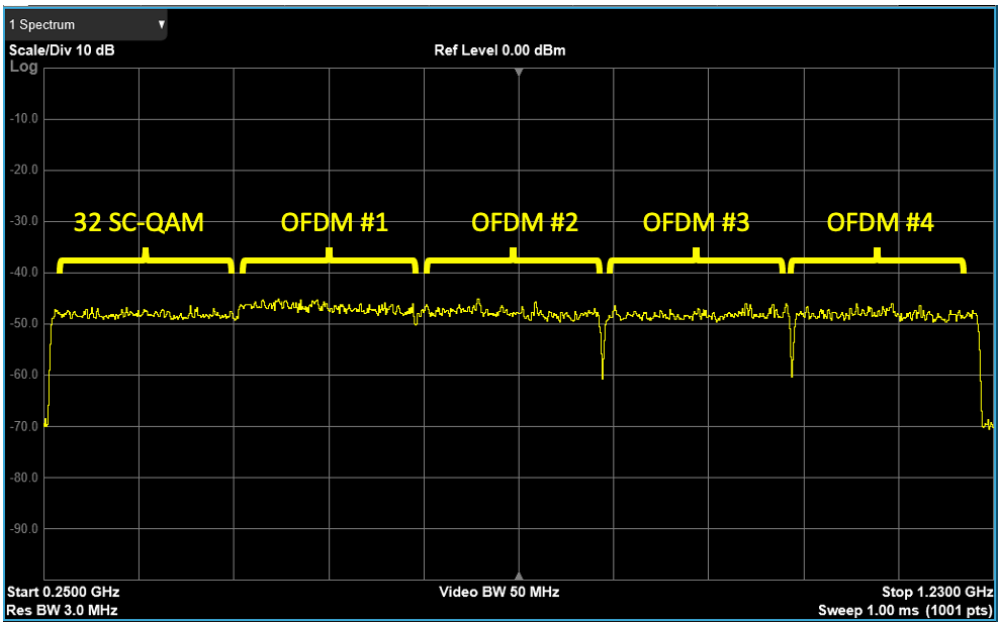


Figure 6 – Downstream with 32 SC-QAM and 4 OFDM Channels

Figure 7 shows a close-up of the first seven SC-QAM channels from 258 MHz to 300 MHz, illustrating that each is 6 MHz wide.

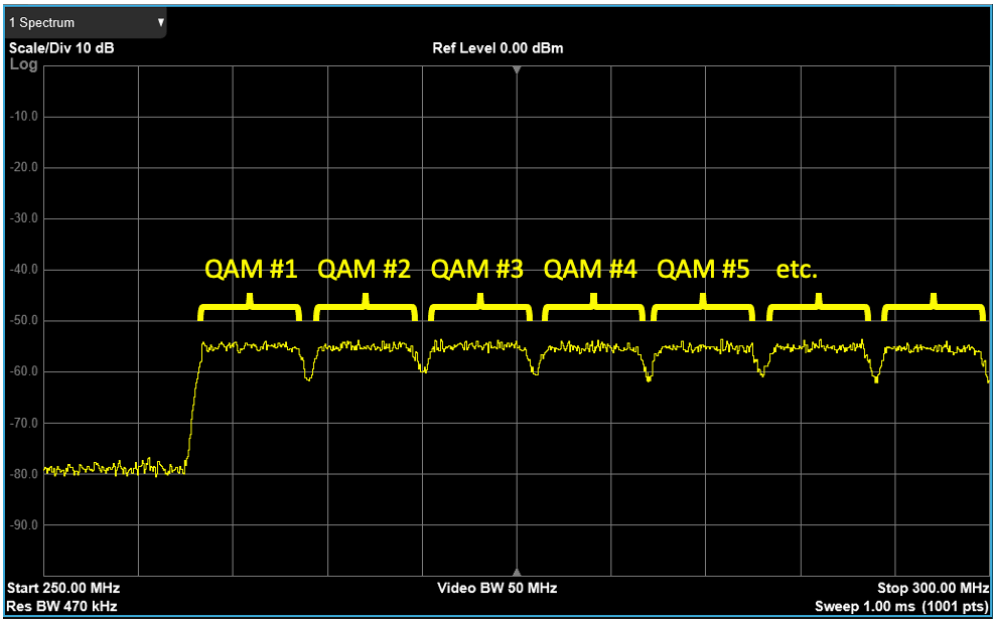


Figure 7 – Close-up of First Seven QAM Channels

Each downstream 6 MHz SC-QAM channel, using a modulation of 256 QAM, supports about 35 Mbps of throughput; hence, 32 of these channels provides about 1.1 Gbps of throughput. In Europe, a similar setup would use 8 MHz SC-QAM channels and 24 channels would fit within that 192 MHz of spectrum.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

A 192 MHz OFDM channel (as configured here) can support approximately 1.76 Gbps of throughput and theoretically up to 1.9 Gbps under the best conditions. In the lab, these OFDM channels were set up for throughput which can vary with different choices of cyclic prefix and other OFDM parameters. Add this all up, and this configuration provides ~ 8.1 Gbps of capacity downstream. This speed was achieved using a traffic generator sending traffic from the NSI of the CMTS to the CPE side of the DOCSIS 4.0 CM. This was achieved on multiple CMTS setups, including one test CMTS platform. Theoretically, with different OFDM parameters the rates could go up to 8.5 Gbps, see [D31 capacity], [10G Capacity].

4.2.2. Downstream of 5 OFDM Channels

The next cable network that we built had a DOCSIS 3.1 downstream spectrum that spans from 258 MHz to 1218 MHz, with simply 5 OFDM channels, each 192 MHz wide. These channels fill 960 MHz from 258 MHz to 1218 MHz in the spectrum. Figure 8 shows a spectrum capture of the spectrum from 258 MHz to 1218 MHz being occupied by 5 OFDM channels.

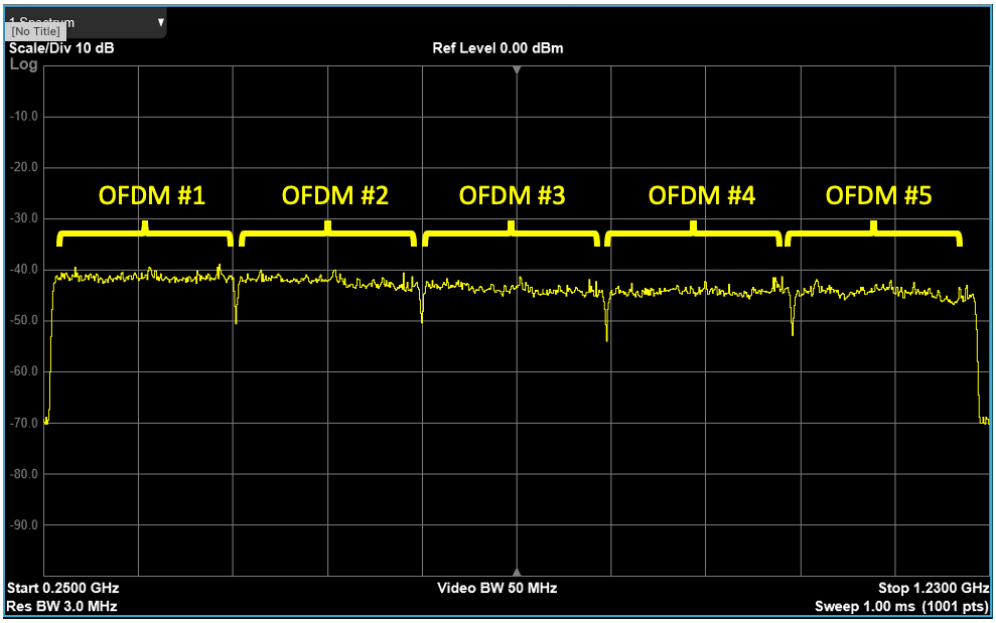


Figure 8 – Downstream with 5 OFDM Channels

A 192 MHz OFDM channel can support up to ~1.76 Gbps of throughput, and five of these channels provides about 8.8 Gbps of capacity on the downstream. This speed was achieved using a traffic generator sending traffic from the NSI of the CMTS to the CPE side of the DOCSIS 4.0 CM. The DOCSIS 4.0 CM was able to bond across all 5 OFDM channels. Not all CMTS setups supported this configuration, but the proof of concept was achieved at least on one CMTS setup.

A 192 MHz OFDM channel can theoretically support up to 1.9 Gbps under the best conditions and configuration (this would amount to 9.5 Gbps), see [D31 capacity] and [10G Capacity]. It is possible in the lab to tweak the channel settings to achieve upwards of 8.8 Gbps. In the lab, the current effort has not been able to attain 9+ Gbps although this will be a part of future testing and refining the channel parameters. The lab is an easier environment than an outside plant and with real-world settings, a real network should support peak speeds in the range of 8.5-9.5 Gbps in many deployments.

Downstream SC-QAM channels are required for DOCSIS 3.0 and earlier CMs; hence, only DOCSIS 3.1 CMs and DOCSIS 4.0 CMs would work on this system since there are no downstream SC-QAMs. An

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

operator could choose to include fewer than 32 SC-QAMs and then do a partial fifth OFDM channel and get a total downstream capacity over 8 Gbps while still supporting DOCSIS 3.0 and earlier CMs.

4.2.3. Upstream with 2 OFDMA Channels

The next cable network that we built had a DOCSIS 3.1 upstream spectrum that spans from 12 MHz to 204 MHz, with simply 2 OFDMA channels, each 96 MHz wide. This enables us to reach the maximum upstream speeds possible on a DOCSIS 3.1 upstream plant. Only DOCSIS 3.1 and DOCSIS 4.0 CMs would operate on this plant as DOCSIS 3.0 CMs and earlier CMs need upstream SC-QAM channels.

The two OFDMA channels, each 96 MHz wide, span from 12 MHz to 108 MHz, and 108 MHz to 204 MHz, respectively. This is clearly an increase from a traditional DOCSIS upstream (sub-split) of 4 SC-QAM channels that provide about 25.6 MHz of DOCSIS spectrum. The DOCSIS 4.0 CMs we tested successfully registered in this setup.

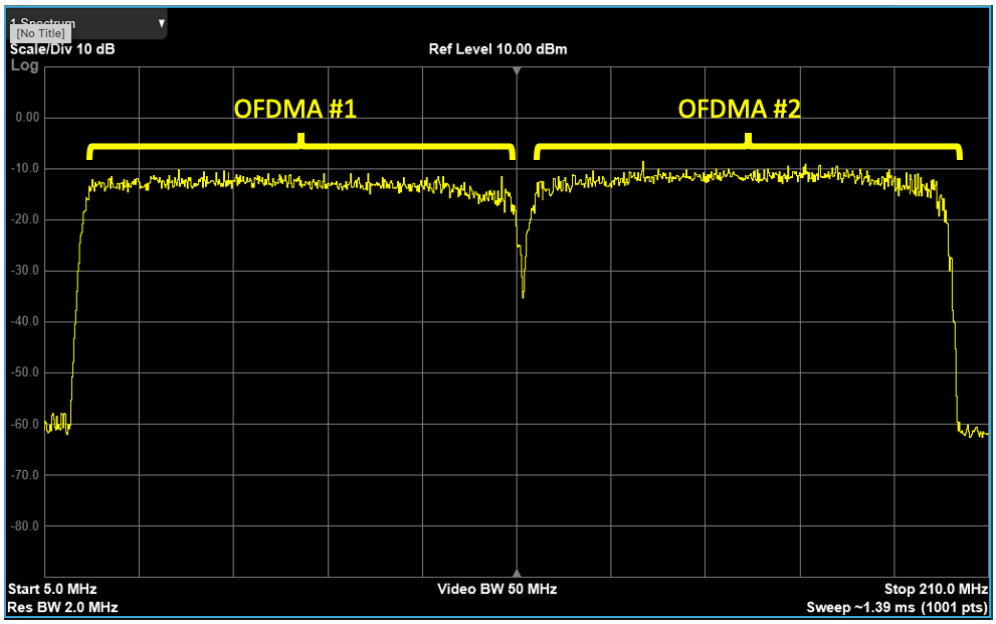


Figure 9 – Upstream with 2 OFDMA Channels

A 96 MHz OFDM channel can support around 0.8 Gbps (800 Mbps) of throughput on many cable plants. Theoretically one could tweak the parameters further to get a bit more throughput. In the lab, these OFDMA channels were set up for throughput which can vary with different choices of cyclic prefix and other OFDMA parameters. Two of these channels provided 1.6 Gbps of capacity upstream. This speed was achieved using a traffic generator sending traffic from CPE side of the DOCSIS 4.0 CM to the NSI of the CMTS. This is like speeds we obtain with DOCSIS 3.1 high-split CMs on a high-split CMTS.

4.2.4. Fastest DOCSIS 3.1 Configuration

The fastest DOCSIS 3.1 channel configuration is obtained with all OFDM and OFDMA channels. Specifically on a high-split upstream with two OFDMA channels and 960 MHz downstream with five OFDM channels. This configuration/setup provided up to 1.6 Gbps upstream and 8.8 Gbps downstream (and potentially up to 9.5 Gbps). We successfully got DOCSIS 4.0 CMs to register in this configuration as well with a high-split CMTS configuration but did not get a chance to run traffic throughput tests. This spectrum plan is quite interesting with only 7 OFDM/A channels and no SC-QAMs in the upstream or

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

downstream. Also there cannot be any traditional video service on the network (no available spectrum), and only IPTV service can be supported on such a plant.

4.2.5. Speeds including DOCSIS 3.0 and Earlier Modems

A real-world requirement for most operators is to support DOCSIS 3.0 and earlier CMs, which still live on the plant. To support this, SC-QAM channels would need to be added in both the upstream and downstream. For a given amount of spectrum, a SC-QAM is less spectrally efficient, i.e., provides less speed than using that spectrum for OFDM.

Typically, in the downstream up to thirty-two SC-QAMs would be included although fewer are possible, depending on how many DOCSIS 3.0 CMs are on the plant and the speed tiers that need to be supported. Regardless, a downstream of 8 Gbps or better can still be achieved. In the upstream, typically 25.6 MHz is reserved for four upstream SC-QAMs and then the lower OFDMA channel would run from about 40 MHz to 108 MHz (assuming the FM band is used). This combination of channels would provide up to 1.5 Gbps.

Hence a DOCSIS configuration to support DOCSIS 3.0 and earlier CMs as well as DOCSIS 3.1 and DOCSIS 4.0 CMs, the system can still provide about 8 Gbps in the downstream and about 1.5 Gbps in the upstream.

5. Conclusions

With recent product upgrades, a DOCSIS 3.1 CMTS can now support four or even five OFDM channels on the same cable plant. This is still a DOCSIS 3.1 CMTS operating on a DOCSIS 3.1 sub/mid/high-split network. Currently available DOCSIS 3.1 CMs only support two OFDM channels that, along with 32 SC-QAMs, have a peak downstream rate of about 5 Gbps.

The new DOCSIS 4.0 CM can use those additional downstream OFDM channels on the DOCSIS 3.1 CMTS, yielding a total capacity and peak downstream rate of up to 8.8 Gbps with 4 OFDM channels + 32 SC-QAMs, and potentially up to 9.5 Gbps with 5 OFDM channels. These increases network capacities can enable operators to offer faster downstream speed tiers than they could with just DOCSIS 3.1 CMs.

DOCSIS specifications have always been designed with backwards compatibility in mind. DOCSIS 4.0 CMs can become operational on a DOCSIS 3.1 CMTS. (Software bugs which prevented this last year have been fixed in CMTS implementations). With these fixes and upgrades to bond more than 2 OFDM channels, DOCSIS 3.1 CMTS can now interpret and negotiate DOCSIS 4.0 modem capabilities and register these CMs on the network allowing them to use the additional downstream capacity.

DOCSIS 4.0 CMs can operate seamlessly on a DOCSIS 3.1 CMTS. In the testing of various CMTS configurations (Integrated, RPHY, RMACPHY, etc.), we successfully brought up DOCSIS 4.0 CMs to register on multiple D3.1 CMTS. These CMs with support for 5 OFDM channels can use the additional CMTS downstream channels and bond data traffic across all of them to prove the ~8.8 to 9.5 Gbps peak downstream data rates.

Enabling DOCSIS 4.0 CMs on the DOCSIS 3.1 network today provides operators with a cost-effective means of providing enhanced services during the transition to DOCSIS 4.0 technology. As seen in lab testing with the example scenarios in this paper, cable operators have many useful and cost-effective methods to increase the service speeds to their customers.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The combination of DOCSIS 3.1 CMTS and DOCSIS 4.0 CM technologies give operators plenty of options in both the upstream and downstream network design. The speeds provided by this combination of technologies will be able to stand up to market pressures until the full network evolution for the DOCSIS 4.0 network are complete over the next many years. A DOCSIS 3.1 CMTS can now provide up to 9.5 Gbps of downstream capacity and almost 1.7 Gbps upstream capacity, enabling high service tier speeds to the customers. D4.0 FDD and FDX CMs can seamlessly make use of this upgraded capability in the DOCSIS 3.1 network.

Abbreviations

BPI +	Baseline Privacy Interface Plus
BPKM	Baseline Privacy Key Management
CM	cable modem
CMTS	cable modem termination system
DAA	distributed access architecture
DOCSIS	data over cable service interface specification
ESD	extended spectrum DOCSIS 4.0 technology
FDD	frequency division duplex
FEC	Forward Error Correction (code)
FDX	full duplex DOCSIS 4.0 technology
FM	frequency modulation
Gbps	gigabits per second
GHz	gigahertz
HFC	hybrid fiber-coax
IPTV	Internet protocol television
MAC	media access control
MACPHY	media access control (MAC) and physical (PHY) layers
MDD	MAC domain descriptor
MHz	megahertz
MPEG	motion pictures experts group
OFDM	orthogonal frequency-division multiplexing
OFDMA	orthogonal frequency-division multiple access
OOC	overlapping OFDMA channels
PHY	physical layer
PNM	proactive network maintenance
QAM	quadrature amplitude modulation
RMD	Remote MACPHY device
RPD	Remote PHY device
SC-QAM	single carrier quadrature amplitude modulation
SCTE	Society of Cable Telecommunications Engineers
TLV	type/length/value

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

[D40MULPI] *DOCSIS 4.0 MAC and Upper Layer Protocols Interface Specification, SP-CM-MULPIv4.0-I07-230503, 2023*.

[D40PHY] *DOCSIS 4.0 Physical Layer Specification, SP-CM-PHYv4.0-I06-221019, 2022*.

[D31 Capacity] *Accurately Estimating D3.1 Channel Capacity, Karthik Sundaresan. SCTE 2017*

[10G Capacity] *Network Capacity Options on the Path to 10G, Karthik Sundaresan. SCTE 2022*

[10G Playbook] *Developing the DOCSIS 4.0 Playbook for the Season Of 10G, Robert Howald, John Williams. SCTE 2021*

[Capacity Model] *Broadband Capacity Growth Models. Will the end of Exponential Growth eliminate the need for DOCSIS 4.0? John Ulm, Zoran Maricevic, Ram Ranganathan, SCTE 2022*

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Leveraging Lidar Technology in Telecommunications Construction

A Technical Paper prepared for SCTE by

Camille Westlake
Program Manager of Commercial Innovation
NV5 Geospatial Inc.
421 SW 6th Ave, Suite 800, Portland, OR. 97204
503.505.5100
Camille.Westlake@nv5.com

Steve Lucente
Director, Business Development - Telecom
NV5 Geospatial Inc.
421 SW 6th Ave, Suite 800, Portland, OR. 97204
305.926.9179
Steve.Lucente@nv5.com

Dan Wagner, NV5 Geospatial Inc.

Arley Oddo, NV5 Geospatial Inc.

Conrad Hilley, NV5 Geospatial Inc.

Chris Andersen, NV5 Geospatial Inc.

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Methods 5
2.1. Overview 5
2.2. Data Acquisition 6
2.2.1. Mission Planning 6
2.2.2. Survey Control 6
2.2.3. In-Situ Measurements 6
2.3. Lidar & Imagery Survey 7
2.4. Data Processing 9
2.4.1. Lidar Processing 9
2.4.2. Imagery Processing 10
2.4.3. Primary Asset Extraction via Pole & Span Modelling 10
2.4.4. Wire (Span) Measurements 13
2.4.1. Clearance Violations 15
2.4.2. SPIDA®calc Integration 15
3. Results 18
4. Conclusions 19

Abbreviations 20

Bibliography & References 20

List of Figures

Title **Page Number**

Figure 1 – Pole Modelling Workflow 5
Figure 2 – In-Situ Measurements 6
Figure 3 – Mobile Mapping System 7
Figure 4 – Mobile Lidar Point Cloud Data 8
Figure 5 – Spherical Imagery Data 9
Figure 6 – Pole Asset Extraction from Mobile Lidar Data 10
Figure 7 –Drip Loop (Red Dot) 10
Figure 8 – Pole Measurements 11
Figure 9 – Pole Measurements Inputs and Outputs 12
Figure 10 – Wire Measurements 13
Figure 11 – Wire Measurement Inputs and Outputs 14
Figure 12 – User-Defined Violation Criteria 15
Figure 13 – Pole Models in Pole Loading Software (SPIDA®calc) 16
Figure 14 – Image Recognition Workflow 16
Figure 15 – Image Recognition Results View 17
Figure 16 – Image Recognition QC View 17

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Tables

Title **Page Number**

Table 1 – Field Measurement List 7
Table 2 – Mobile Mapping Sensor Specifications 8
Table 3 – Pole Measurements Common Attachment Points 11
Table 4 – Pole Measurement Inputs and Outputs 12
Table 5 – Wire Measurements 13
Table 6 – Wire Measurement Inputs and Outputs 14
Table 7 – Variance of lidar-derived measurements vs. in-situ ground truth measurements 18

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The telecommunications industry is experiencing rapid growth and technological advancements with higher demand for data. The 5G rollout is driving that growth and the demand is putting pressure to build the infrastructure. 5G can use existing 4G/3G cell sites but has a lower cell radius of 100m, which means it will require the deployment of many more cell sites. The best locations for these existing cell sites are on existing infrastructure such as utility poles or buildings. Utility poles are ideal locations because the 5G cell sites use more energy (around 11.5 kW vs. 6.8 kW of 4G) and need a backhaul network (i.e., fiber connection).[1]

The struggle is working within the utility industry regulations to get additional fiber and antennas placed in a timeline needed for the 5G rollout. Telecommunication companies are finding their applications to deploy cell sites taking much longer than the shot clock governance of 90 days for reviewing the application. Some have reported that seventy percent (70%) of their applications have exceeded the 90-day shot clock and forty-seven percent (47%) have exceeded 150 days.[2]

Utilities want reassurance that the pole passes the make-ready assessment and can support the added infrastructure 5G needs. The traditional boots-on-the-ground method of measuring attachment points is time-consuming and cannot keep up with the needs of the 5G rollout. An alternative solution is a Mobile Mapper Unit which consists of a mobile lidar (light detection and ranging) system and a spherical camera. The lidar unit scans the pole and wires generating 3D point cloud data that can be analyzed by computer algorithms to extract the attachment locations on the pole, while image recognition can be applied to the spherical imagery to identify smaller equipment and pole tags.

Many telecommunication companies have large fiber densification projects in the largest cities. As a result of these projects, they will need to attach fiber cables to millions of electric utility poles. This will more than double the number of new attachments that would normally go on a pole. Time is of the essence for these projects and utilizing the normal utility attachment/inspection process will not meet the required timelines. However, utilities need to ensure that their facilities are being managed correctly and that all rules and regulations are being followed when it comes to pole attachments. Telecommunication companies and utilities are looking for ways to meet their timelines, while still achieving regulatory and safety compliance.

A large utility company, referred to as Utility, approached NV5 Geospatial (NV5G) to collaborate on investigating whether remote sensing can be utilized to:

- Shorten the asset attachment inspection timeline;
- Ensure that the Utility is capturing the information necessary to determine if telecommunication companies can attach to its poles while still adhering to government regulations (e.g. National Electric Safety Code (NESC) clearances), and;
- Create as-built records of change after construction.

Through this project, an accurate, efficient, and cost-effective remote sensing-based methodology was developed that produced an average ninety-four percent (94%) match with in-situ measurements and allowed for 40 miles of pole lines to be collected in a day. This white paper presents an overview of the remote sensing methodologies developed and a comparative quantitative analysis of conventional survey methods vs. a remote sensing based option.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2. Methods

2.1. Overview

A remote sensing mobile mapping system was used to collect lidar and spherical imagery data, which were then processed via automated routines, pole modeling tools, and image recognition to identify and extract utility poles, wires, and various pole attachments, as well as calculate wire measurements and identify clearance detections. This process also generated output data that could be run through a 3rd party application to generate pole-loading models.

The following workflow overviews the different phases from data collection to data processing and attachment, clearance, and loading analyses; highlighting areas of manual and automated data processing along with optional phases that can be applied to improve accuracy and provide more detailed results. Each phase is further detailed below, and the results comparing in-situ to remote-sensing-derived measurements are presented.

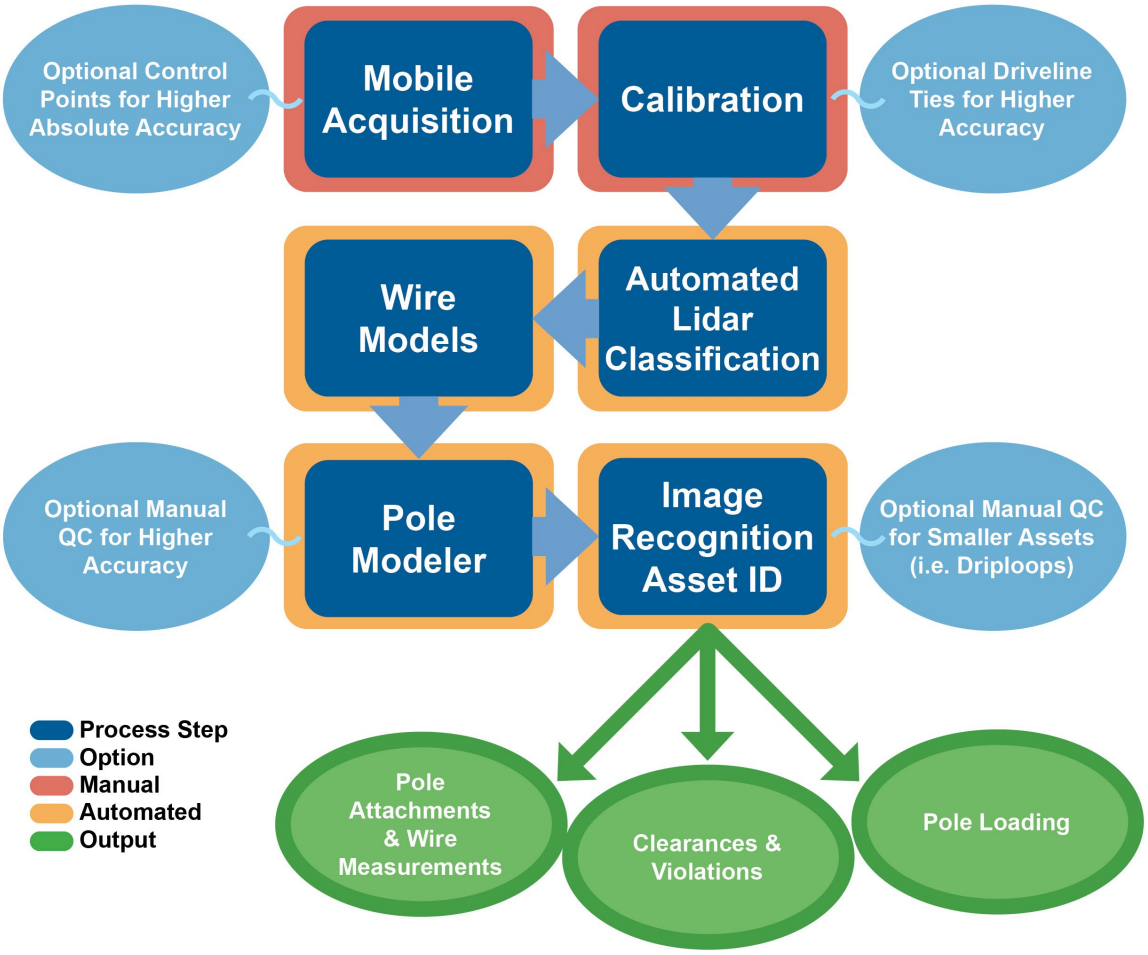


Figure 1 – Pole Modelling Workflow

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.2. Data Acquisition

2.2.1. Mission Planning

To maximize asset detection by the lidar and imagery systems, an acquisition route must be planned. Upon receipt of pole locations or a survey area of interest (AOI) from the utility client, the mission planning team creates a drive-route utilizing an in-house or 3rd party routing application. Lane prioritization (maximize feature visibility) and additional factors such as weather (precipitation), solar windows (sun angle for imagery), traffic volume (time of day dependent), and, roadway accessibility (construction, blockages, one-way/bi-directional/divided) and conditions (paved, gravel, etc.) are taken into account. This process ensures complete route coverage in the most time and cost-efficient manner possible.

2.2.2. Survey Control

NV5G survey teams utilize industry standard equipment (Trimble GNSS dual-frequency L1-L2 receivers) and survey methodologies (static, real-time kinematic (RTK) or precise point positioning (PPP)) for survey and ground control. Absolute vertical and horizontal accuracies in the range of a few centimeters are possible. Additional ground control points are collected for both calibration support and positional (vertical and horizontal) accuracy assessments, with typically a handful of targets being collected every mile.

For some applications, such as the current one of asset attachment inspections and clearance analyses, this level of absolute accuracy is not required as the primary focus is not on having high accuracy latitude and longitude coordinates for features, but rather that features are correctly situated relative to one another so that measurements taken between them are accurate. For mobile systems relative accuracies in the millimeter range are possible, allowing for highly accurate clearance measurements.

2.2.3. In-Situ Measurements

Field crews were deployed to carry out in-situ measurements in two phases. One phase involved measurements taken by field crews as directed by a utility engineer. The second phase involved measurements collected by field crews following conventional survey methods. These later measurements were collected twice to analyze and control for inconsistent field measurements. The engineer measurements were utilized as truth control values for comparison against both conventional survey and remote-sensing-derived measurements.

Field measurements were made using a measured telescoping fiberglass pole with a cable hook on the end. Once hooked over the wire or placed adjacent to an asset, the measurement was read off the telescoping pole and recorded.

A total of 11 different measurements (Table 1 – Field Measurement List) were taken at each of the 100 study poles.

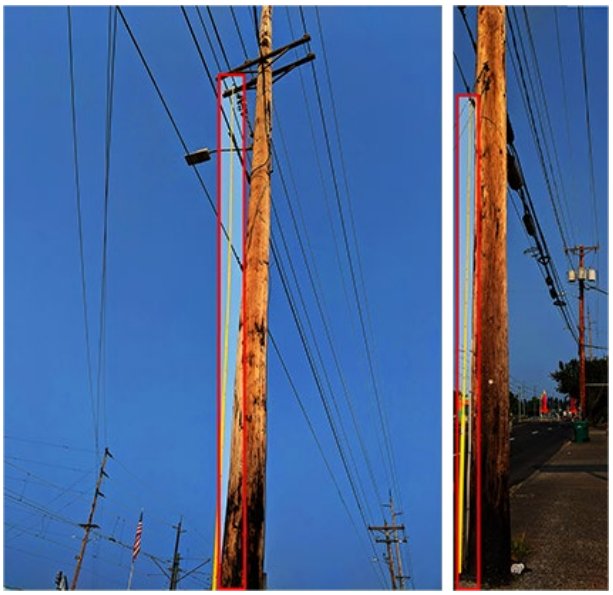


Figure 2 – In-Situ Measurements

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 – Field Measurement List

Highest Comm	Streetlight	Comm/Electric
Neutral	SL Drip	Comm/Street Light
Secondary	Electrical 4” Riser	Comm/SL D Loop
Secondary Drip	Electrical 6” Riser	

2.3. Lidar & Imagery Survey

High-density lidar and georeferenced panorama spherical imagery were collected using Riegl’s VMX-2HA lidar sensor and a LADYBUG-5 spherical imagery sensor (Table 2 – Mobile Mapping Sensor Specifications) mounted onto a truck. The VMX-2HA system is comprised of two (2) rear-facing VUX-1HA lidar sensors which collect point data of roadside features up to 475 meters away. The spherical imagery was included to collect images at fixed-distance intervals. All sensor inputs, monitored in real-time by operators, are recorded and time-stamped.

Additionally, Riegl’s RiACQUIRE software is utilized to verify successful data collection by monitoring real-time sensor status, route coverage, and data condition.



Figure 3 – Mobile Mapping System

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 2 – Mobile Mapping Sensor Specifications

Sensor	Specifications	
Lidar RIEGL VMX-2HA	Pulse Repetition Rate (PRR) (peak)	3.6 MHz
	Range	Min = 1m, Max = 475m
	Accuracy/Precision	5 mm / 3 mm
	Rotation	360°
Spherical Imagery LADYBUG-5	Number of Lenses	6
	Maximum Resolution	2048x1224 pixels (30 MB)
	Spectral Bands	3 (RGB)
IMU/GNSS	Type	Integrated
	Absolute Accuracy (typical)	xy = 2cm, z = 3cm

*inertial measurement unit (IMU)
ˆpulse repetition frequency (PRF)
˜red, green, and blue (RGB)
ɛglobal positioning system (GPS)
¨global navigation satellite system (GLONASS)

While mobile scanners can collect data at a point density of thousands per square meter and ideal densities are project and application-dependent, for reliable capture of pole and attachment features densities in the range of 750 PPSM (points per meter squared) are sufficient and more cost effective. (Figure 4 –Mobile Lidar Point Cloud Data).

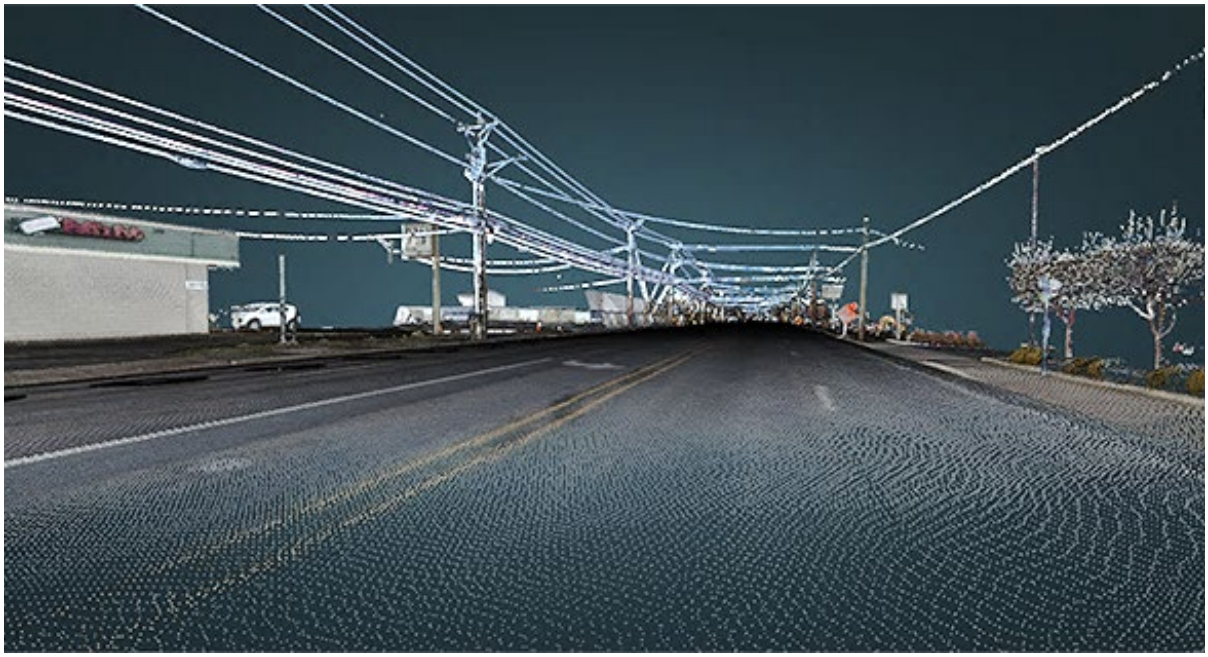


Figure 4 – Mobile Lidar Point Cloud Data

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

To aid in the identification of assets not readily identifiable with lidar data alone, spherical imagery is also collected. (Figure 5 – Spherical Imagery Data).



Figure 5 – Spherical Imagery Data

2.4. Data Processing

2.4.1. Lidar Processing

Post-acquisition processing of acquired lidar data begins with georeferencing and spatial refinement of the vehicle drive path known as the trajectory. Data from the onboard GNSS positioning system is differentially corrected using nearby reference base stations or PPP methods. This corrected GNSS position data is then combined with data from the onboard IMU to produce a smoothed best estimate of trajectory (SBET), data that captures the sensor position throughout the acquisition drive. With the vehicle and sensor position provided within the SBET, georeferenced lidar point clouds can be created.

To improve the absolute spatial accuracy of the derived point cloud, ground survey points placed on identifiable features can be used to make corrections to the point cloud position or further refine the SBET.

In areas where drive paths overlap, automated methods can be used to improve the alignment of the overlapping data, increasing the relative accuracy of the resulting point clouds.

2.4.1.1. Automated Point Cloud Classification

Following calibration, each point in the lidar point cloud is classified as real-world features (e.g. ground, poles, wires, etc.) using a series of advanced automated and manual routines. First, a ground surface model is developed using automated ground detection routines. The results are then refined using a statistical surface algorithm with constraints based on geometric relationships between points and lidar return characteristics. The resulting ground classification can then be used to generate a bare-earth digital elevation model (DEM). The DEM is quality-controlled to identify any ground classification issues that can quickly be resolved manually.

After ground points have been classified, above-ground features (e.g., buildings, wires type, poles, vegetation, fences, etc.) are automatically classified using a machine learning model specifically trained to classify utility infrastructure. Classification models used by NV5G leverage features of the lidar point cloud (e.g., height above ground, return type, return intensity, geometric properties, etc.). Further

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

refinements of the wire types (e.g. Primary and Secondary electrical wires) can be completed via manual methods. As necessary, quality control is completed, and any misclassifications are corrected to ensure required classification accuracies are met. Using automated routines alone, classification accuracy is 95% or better. Typically, the automated classification results are sufficient to extract 3D vector models of utility infrastructure automatically.

2.4.2. Imagery Processing

The Ladybug 5 camera generates video stream files (.prg) that are turned into panoramic 3-band (RGB) JPEG (Joint Photographic Experts Group) files using the Ladybug CapPro software. Image position and orientation are calculated by linking the time of image capture, the corresponding vehicle position and attitude, and the SBET data (the same utilized in 2.4.1 Lidar Processing) in POSPac Mobile Mapping Suite (MMS) and outputting an initial Exterior Orientations (EO) file. The EO file is combined with the JPEG images in a proprietary georeferencing tool, tagging them with geographic locations. Image georeferencing is quality controlled (QC) to ensure processing is performed as per protocols.

2.4.3. Primary Asset Extraction via Pole & Span Modelling

Once the lidar data has been classified and wires are extracted, Pole Modeler is used to extract attachments on the pole (Figure 6 – Pole Asset Extraction from Mobile Lidar Data), calculate mid-span and pole clearances, and generate models for pole loading analyses. These tools are run in an automated fashion with no initial manual QC. With added manual QC, accuracies can be improved further by removing errors produced during automation, such as missing or misplaced attachments (± 6 inches).

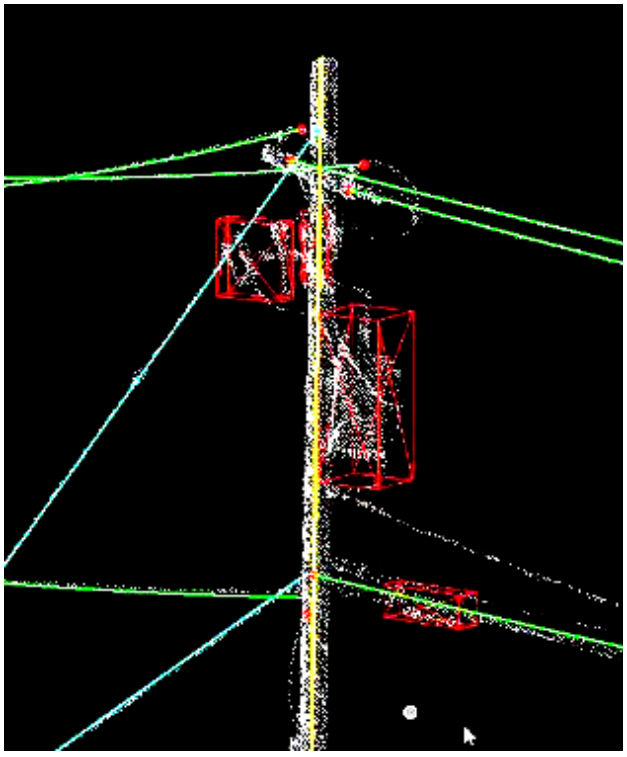


Figure 7 – Pole Asset Extraction from Mobile Lidar Data



Figure 6 – Drip Loop (Red Dot)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Pole Modeler first isolates the pole and calculates height, lean, lean angle, and diameter. Other features on the pole, such as cross-arms, larger equipment (e.g. streetlights, transformers), guy wires, service drops, and other attachments discernible within the lidar point cloud are also extracted. Next, if required, manual QC can be completed to adjust misplaced attachment points and capture smaller features that may have be missed (e.g. drip loops) (Figure 7 – Drip Loop) due to a lack of lidar point returns from the narrow cables. An analyst can quickly snap offset attachment points that are misaligned and add attachment points that were not identified by the automated routines.

2.4.3.1. Attachment Measurements

Once the necessary attachment assets have been identified, several measurements (Table 3 – Pole Measurements Common Attachment Points) between them can be completed (Figure 8 – Pole Measurements) automatically and used for clearance violation pole loading analyses.

Table 3 – Pole Measurements Common Attachment Points

Electric		Communication
Primary	Streetlight	Mainline
Secondary	Streetlight Drip Loop	Service Drop
Neutral	Riser	Crossarm
Transmission	Weather Head	Riser
Service Drop	Pole to Pole Guy	
Drip Loop	Down Guy	
Pole Top		

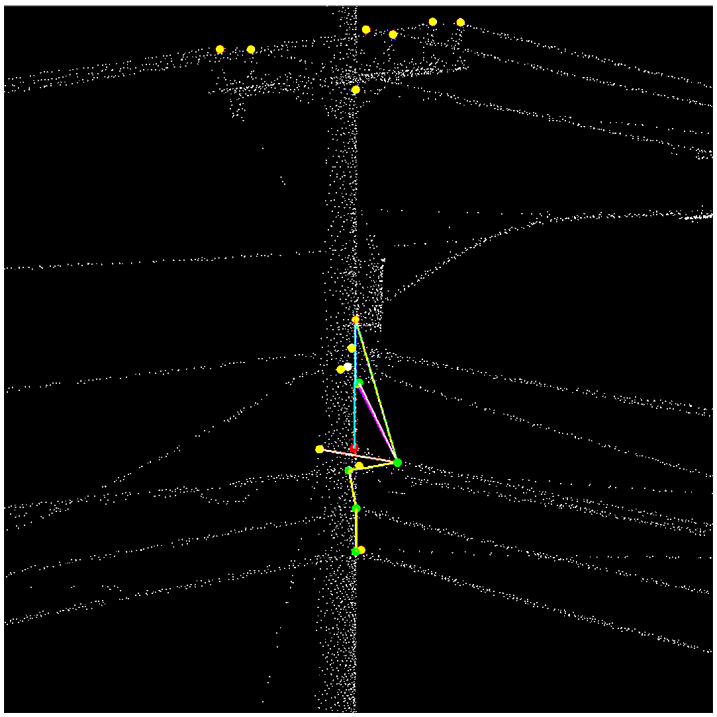


Figure 8 – Pole Measurements

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Several different inputs (Table 4 – Pole Measurement Inputs and Outputs) are required within Pole Modeler to calculate pole attachment measurements. Different parameters (e.g. units, attachment radius) can be adjusted based on project-specific requirements (Figure 9 – Pole Measurement Inputs and Outputs).

Table 4 – Pole Measurement Inputs and Outputs

Inputs	Outputs
Lidar Ground Model	Measurements between Attachment Points
Modeled Attachment Points	Heights Above Ground
Pole Locations and Names	Minimum Clearances
	Clearance Violations

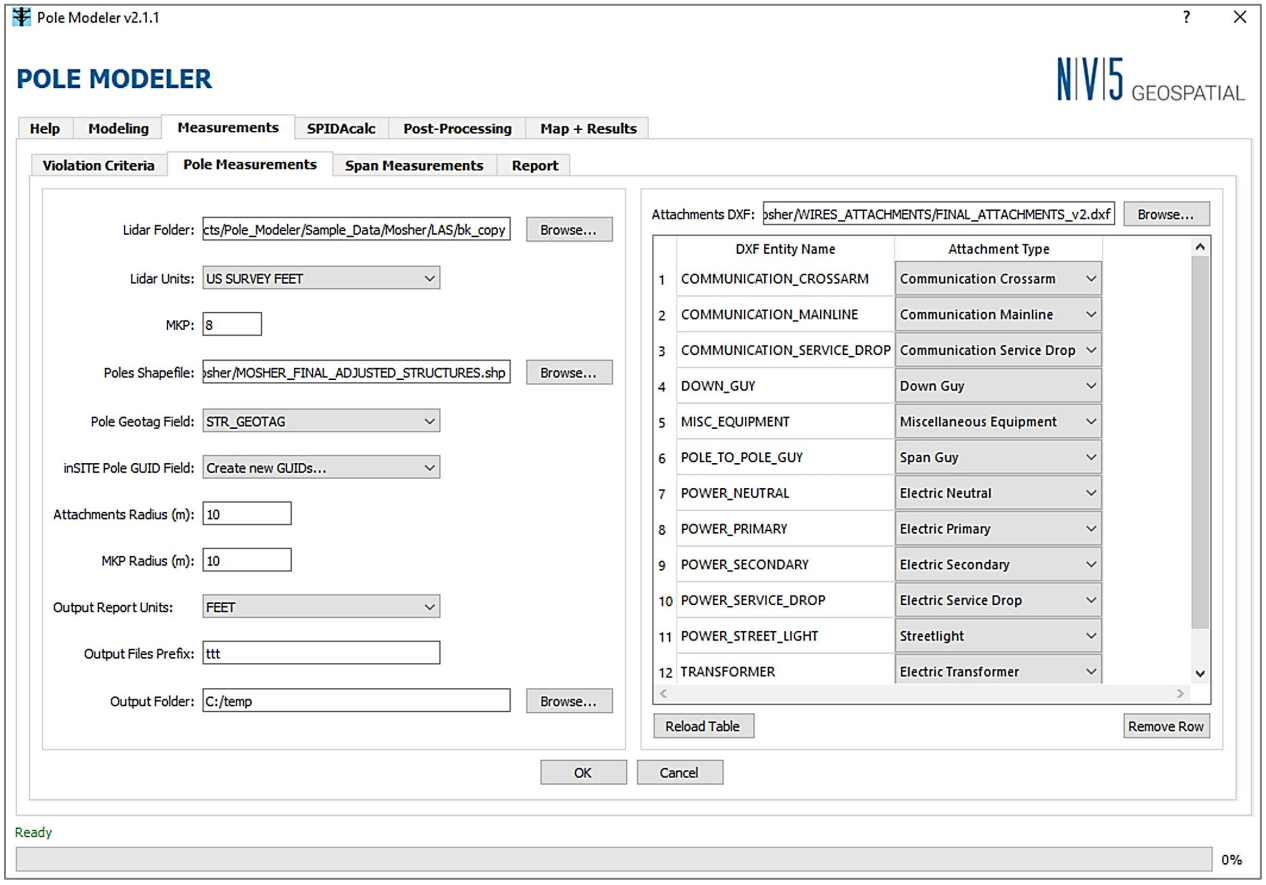


Figure 9 – Pole Measurements Inputs and Outputs

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.4.4. Wire (Span) Measurements

Wire (span) measurements (Table 5 – Wire Measurements) are calculated automatically within Pole Modeler from wire-to-wire or wire-to-ground between them can be completed (Figure 10 – Wire Measurements) and then utilized for clearance violation and pole loading analyses.

Table 5 – Wire Measurements

Wire Measurements
Communication Mainline
Communication Service Drop
Transmission
Primary
Secondary
Neutral
Service Drop
Pole to Pole Guy

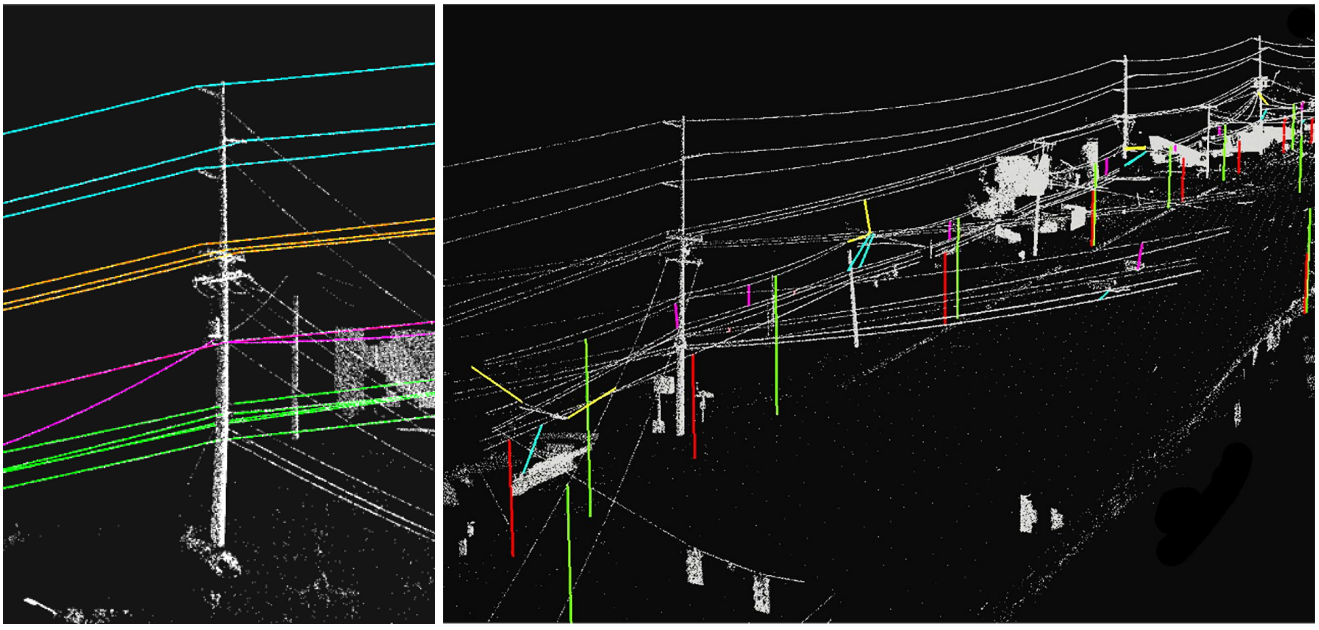


Figure 10 – Wire Measurements

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

As with pole measurements, several different inputs (Table 6 – Wire Measurement Inputs and Outputs) are required within Pole Modeler to calculate required wire distances. Different parameters (e.g. input and output units, surface type) can be adjusted based on project-specific requirements (Figure 11 – Wire Measurement Inputs and Outputs).

Table 6 – Wire Measurement Inputs and Outputs

Inputs	Outputs
Lidar Ground Model	Measurements between Span Wires
Hard Surface Points (optional)	Heights Above Ground
Span Wires	Minimum Clearances
Pole Locations and Names	Clearance Violations
Span Locations and Names	

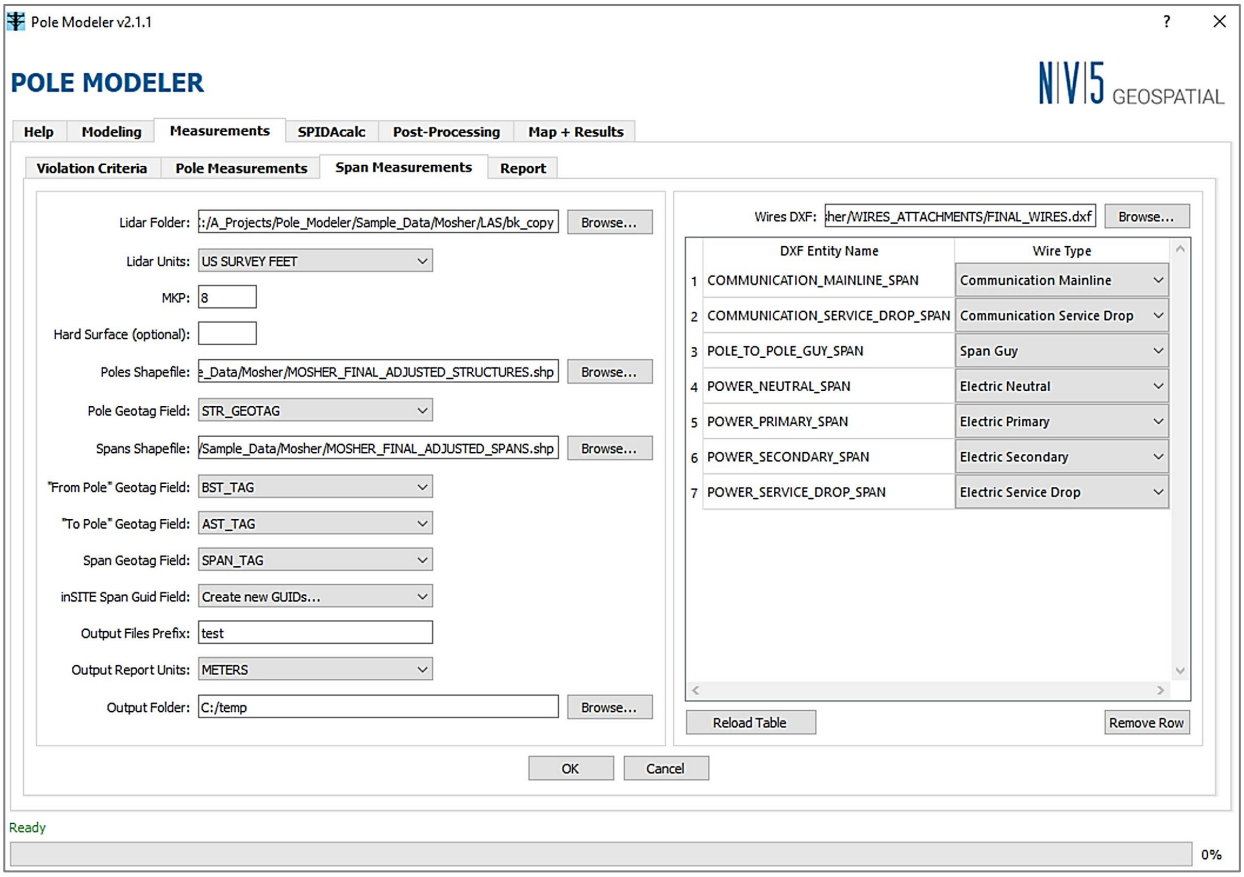


Figure 11 – Wire Measurement Inputs and Outputs

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.4.1. Clearance Violations

Once all the measurements and clearances have been calculated, analysis can be completed to determine if any violations are present. The clearance tool defaults to NESC regulations, but also allows customized clearance configurations (Figure 12 – User Defined Violation Criteria).

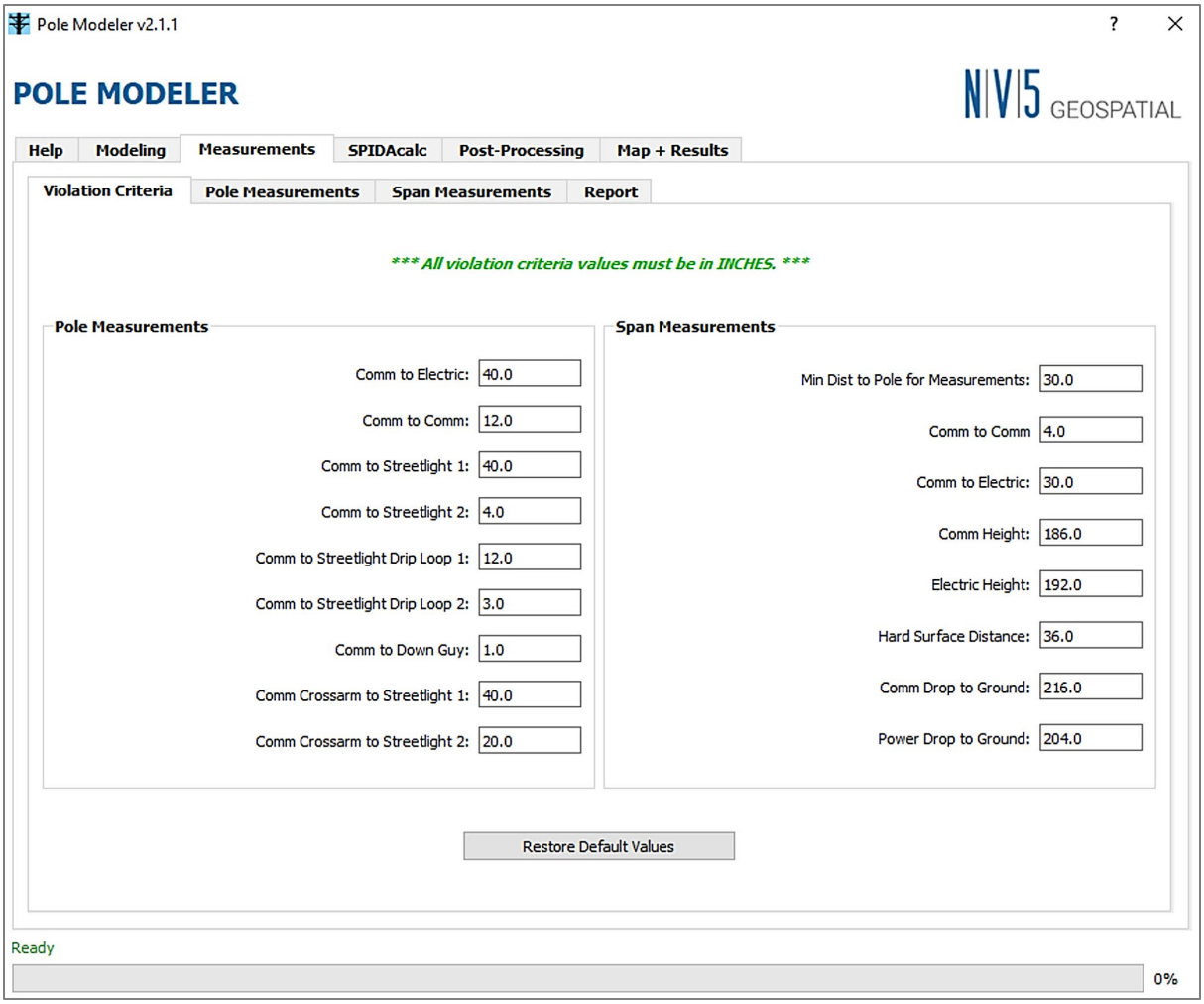


Figure 12 – User-Defined Violation Criteria

2.4.2. SPIDA®calc Integration

Two options are available that can be used to help generate SPIDA®calc files from the modeled asset polylines and their attachment points. The first tool ingests the modeled assets' DXF (drawing interchange format) files, takes dozens of measurements for each asset, assigns their positions on the pole, and connects the wires from span to span. It generates an Excel database with all the information required to build a SPIDA®calc model. The Excel file can easily be edited as part of a quality assurance (QA) or refinement process. For example, the insulator or wire type can be changed, insulator positions can be refined, wires can be connected to one another, etc. Once the Excel database has been deemed acceptable, it is processed by a second tool, which translates that data into a SPIDA®calc JSON (JavaScript object notation) file. During this step, a custom client file can optionally be incorporated into the SPIDA®calc file. It can then be opened and viewed directly within SPIDA®calc (Figure 13 – Pole Models in Pole Loading Software (SPIDA®calc).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

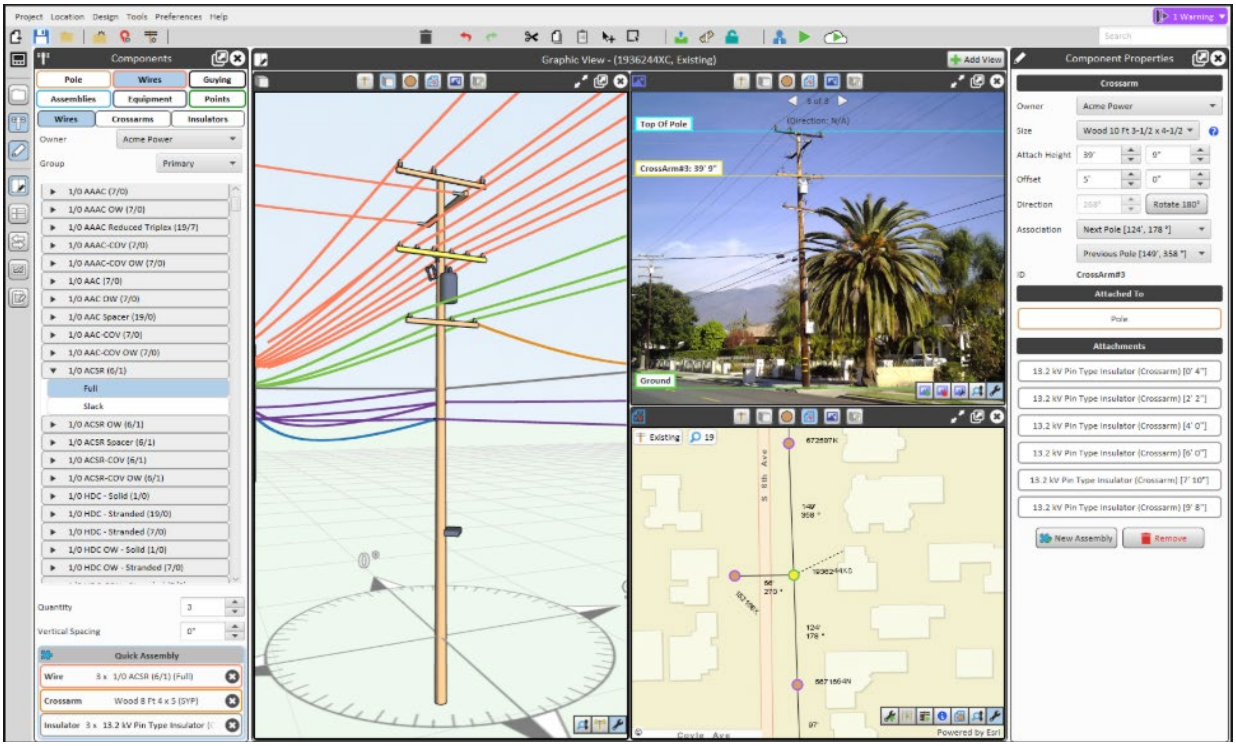


Figure 13 – Pole Models in Pole Loading Software (SPIDA®calc)

2.4.2.1. Additional Asset Extraction via Image Recognition

If, in addition to pole loading and clearance analyses, unique assets need to be identified (telecom panels, splice boxes, power boxes, fuses, switches, etc.), a deep learning object detection model workflow (Figure 14 – Image Recognition Workflow) can be applied to co-acquired spherical images which automatically identifies assets of interest. The image recognition results (Figure 15 – Image Recognition Results View) are quality-controlled and then used to update the geographic information system (GIS). Assets are associated with lidar-rectified poles and spans using lidar depth mapping (when available) or stereo triangulation.

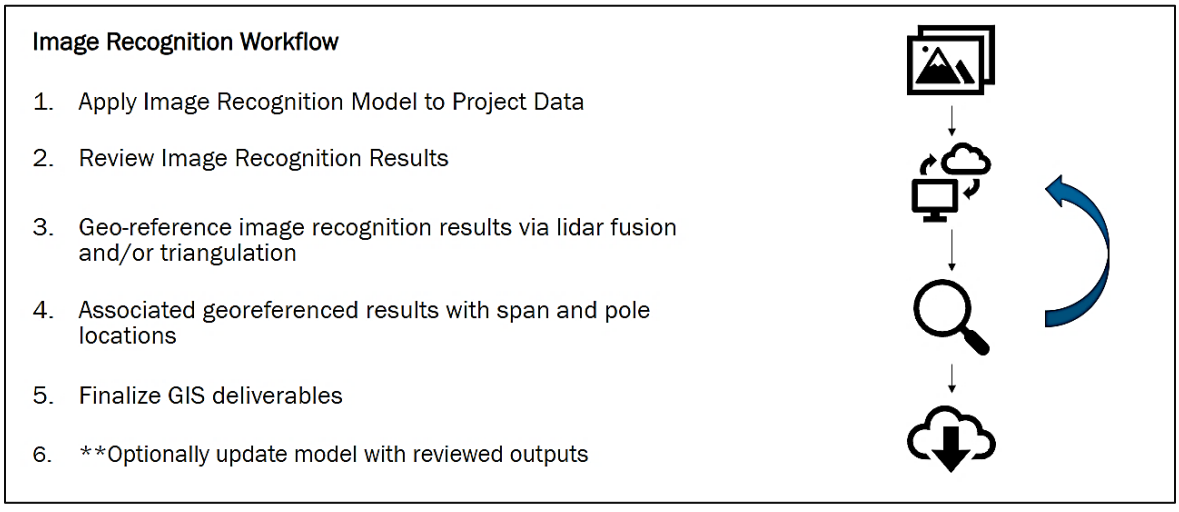


Figure 14 – Image Recognition Workflow

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 15 – Image Recognition Results View

The quality-controlled results (Figure 16 – Image Recognition QC Results View) are also leveraged to further fine-tune the object detection model as the project progresses. Continuous improvement of the model on project data allows for locally unique characteristics to be better captured. This continuous improvement results in higher model precision and accelerates throughput over the project.



Figure 16 – Image Recognition QC View

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Results

An analysis was completed comparing conventional in-situ and remote sensing derived measurements of pole attachment, wire measurements, and clearance distances against in-situ field measurements. The analysis was controlled for inconsistent field measurements (variance >3" between field visits #1 & #2), and other known inaccuracies in the field capture (e.g. inability to access pole due to obstructions such as fences or vegetation). A threshold of 3" was considered an acceptable amount of variance by the Utility for measurements to agree.

To assess the consistency of field measurements, an analysis was done comparing first-visit (field crews) and second-visit (utility engineer present) measures. Twenty-two percent (22%) of field from-ground measurements and twenty-five percent (25%) of field clearance measurements agreed to within 3 inches between visits. This indicates that field measurements are consistent, on average, approximately seventy-seven percent (~77%) of the time.

Using the second-visit field measurements, which were deemed to be truth as they were directed by a utility engineer), an analysis comparing field to lidar-derived measurements was also completed. Here, ninety-seven percent (97%) and ninety percent (90%) of lidar derived from ground and clearance measurements matched field measurements respectively.

Table 7 – Variance of lidar-derived measurements vs. in-situ ground truth measurements

	Variance							
Measurements (from ground)	**Count**	**0"**	**<1"**	**<2"**	**<3"**	**<4"**	**<5"**	**>5"**
Highest Comm Wire	**21**	5	12	4	0	0	0	0
Neutral Wire	**4**	1	1	2	0	0	0	0
Secondary Wire	**12**	4	3	4	1	0	0	0
Secondary Drip Loop	**13**	4	4	3	1	0	0	0
Streetlight	**10**	1	4	5	1	0	0	1
Streetlight Drip Loop	**8**	2	2	3	0	0	0	0
Electrical 4" Riser	**0**	0	0	0	0	0	0	0
Electrical 6" Riser	**0**	0	0	0	0	0	0	0
Total Variance Count	**68**	**17**	**26**	**21**	**2**	**0**	**0**	**1**
Percent Total	**97%**	**25%**	**38%**	**31%**	**3%**	**0%**	**0%**	**1%**
Clearances (between)	**Count**	**0"**	**<1"**	**<2"**	**<3"**	**<4"**	**<5"**	**>5"**
Comm/Electric	**20**	4	7	5	3	0	0	1
Comm/Streetlight	**14**	1	10	1	1	0	0	1
Comm/Streetlight Drip Loop	**11**	5	3	0	1	0	0	0
Total Variance Count	**45**	**10**	**20**	**6**	**5**	**0**	**0**	**2**
Percent Total	**90%**	**22%**	**44%**	**13%**	**11%**	**0%**	**0%**	**4%**

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. Conclusions

To meet the ever-increasing data demands from their clients, RDOF and 5G deployments, cable and telecommunications companies need to expand their networks quickly. The most efficient way, both time and cost-wise, of building this new infrastructure is to utilize existing utility poles. The process of attaching additional fiber and antennas requires adherence to industry regulations aimed at ensuring utility poles are not overloaded and that both utility and telecom networks are reliable and safe. Some of these regulations include an assessment of the distance between different assets that are currently attached to the pole and those wanting to be added. These measurements are conventionally taken in situ by field crews. This project aimed to assess if remote-sensing technologies could achieve comparable measurements, which could potentially expedite the timeline required to obtain approval for additional cable and asset attachments.

Results indicate that comparable measurements can be achieved using lidar data and supporting spherical imagery, with lidar-derived measurements averaging a ninety-three-point five percent (93.5%) agreement with in-situ field measurements. Additionally, inconsistencies inherent in field-derived measurements were also brought to light, with an average of twenty-three-point five percent (23.5%) disagreement between same feature measurements and served to highlight another benefit of utilizing high-accuracy lidar data, that of consistency. As a result, utilities and telecommunications companies will have a high degree of confidence when using remote sensing data that pole loading requirements are being met and a safe and reliable network is being constructed.

While these preliminary results are very encouraging, some of the challenges associated with the use of mobile lidar data, such as back-yard poles that cannot easily be captured from street-based collected, laser penetration impediment by thick vegetation and weather conditions such as precipitation suggest that a combination remote-sensing and focused field-based approach may be the ideal solution that ensures complete network coverage while reducing costs and the time required to obtain the necessary measurements.

To further assess the applicability and efficacy of using mobile lidar and spherical imagery for this application, future analyses can be conducted with telecommunications partners to determine if this workflow results in a reduction in the asset attachment inspection timeline. With data collection currently achieving 40 miles per day and the extent of automation currently utilized in data processing and analysis workflows, the outcome looks quite promising.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AOI	area of interest
DEM	digital elevation model
DXF	drawing interchange format
EO	exterior orientations
GIS	geographic information system
GNSS/GLONASS	global navigation satellite system/ Globalnaya Navigazionnaya Sputnikovaya Sistema
GPS	global positioning system
IMU	inertial measurement unit
JPEG	Joint Photographic Experts Group (digital image file format)
JSON	JavaScript object notation
lidar	light detection and ranging
MMS	mobile mapping suite
NESC	National Electric Safety Code
NV5G	NV5 Geospatial
PPP	precise point positioning
PPSM	points per square meter
PRF	pulse repetition frequency
PRR	pulse repetition rate
QA	quality assurance
QC	quality control
RGB	red, green, and blue
RTK	real-time kinematic
RDOF	Rural Digital Opportunity Fund
SBET	smoothed best estimate of trajectory
SPIDA®calc	Pole loading program by software company SPIDA

Bibliography & References

[1]*5G Network Coverage Planning and Analysis of the Deployment Challenges*; Md Maruf Ahamed, Saleh Faruque, Gianmarco Romano; National Library of Medicine; 2021 October 3; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8512478/.

[2]Federal Communications Commission; FCC 18-133; *Declaratory Ruling and Third Report and Order*; 2018 September 27; https://docs.fcc.gov/public/attachments/FCC-18-133A1.pdf.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Leveraging Machine Learning for Network Traffic Forecasting

A Technical Paper prepared for SCTE by

Diane Prisca Onguetou, PhD
Independent Consultant
onguetou@umich.edu

Achintha Maddumabandara
Co-Author
Principal Data Analyst / Rogers Communications
2400 32 Ave NE, T2E 6T4, Calgary AB, Canada
achintha.maddumabandara@sjrb.ca

Jeffrey Lee
Co-Author
Manager, Data Acquisition and Reporting / Rogers Communications
2400 32 Ave NE, T2E 6T4, Calgary AB, Canada
jeffrey.lee@sjrb.ca

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Case Studies 4
   - 2.1. The Network Graph 4
   - 2.2. Data Collection Over Time 6
   - 2.3. Traffic Overview and Modelling 8
   - 2.4. The Study 9
     - 2.4.1. Main Objective 9
     - 2.4.2. Accuracy 10
     - 2.4.3. ML Pipeline 10
3. CAGR Approach to Network Traffic Forecasting 10
   - 3.1. Global CAGR 10
   - 3.2. Limitations of Global CAGR Approach 11
   - 3.3. A Distributed CAGR Strategy 12
4. Exploring Machine Learning Time Series Models 14
   - 4.1. Seasonal Decomposition of Time Series (STL) 14
   - 4.2. Forecasting Models for Network Traffic Time Series 17
     - 4.2.1. Holt Winters Exponential Smoothing 17
     - 4.2.2. Seasonal Auto Regressive Integrated Moving Average (SARIMA) 18
   - 4.3. Which Model is Best for Network Traffic Forecasting? 18
5. Key Insights 20

Abbreviations 21

Bibliography & References 21

List of Figures

Title **Page Number**

Figure 1 – The Network Topology 5
Figure 2 – Distribution of Connectivity at a Hubsite 6
Figure 3 – Availability of Historical Data 7
Figure 4 – Number of Hubsites and Intraconnects (based on data collection) 8
Figure 5 – Evolution of Network Connectivity (based on data collection) 8
Figure 6 – Historical Traffic Evaluation for Several Links 9
Figure 7 – Global CAGR Estimate 11
Figure 8 – Global vs. Local CAGR 12
Figure 9 – Global vs Local CAGR Predictions for internal longhaul link 13
Figure 10 – Global vs Local CAGR Predictions for international link 13
Figure 11 – Global vs Local CAGR Predictions for internal metro link 13
Figure 12 – A Naïve Look at Total Traffic Periodicity 14
Figure 13 – Seasonal Decomposition for Total Traffic (additive) 15
Figure 14 – Seasonal Decomposition for the internal long haul link (additive) 16
Figure 15 – Seasonal Decomposition for the international link (multiplicative) 16
Figure 16 – Seasonal Decomposition for the internal metro link (multiplicative) 17
Figure 17 – Comparison of ML Model Predictions for internal long haul link 19

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 18 – Comparison of ML Model Predictions for international link 19

List of Tables

Title **Page Number**

Table 1 – CAGR Accuracy Checking [Total Traffic]..... 12

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Network traffic modelling and forecasting play a crucial role in maintaining network performance, optimizing resource allocation, and providing seamless user experience. Perhaps the most widely used method for network traffic forecasting is Compound Annual Growth Rate (CAGR). While CAGR can provide a view of network traffic growth, it does not provide the required context for informed decision making that can differentiate services from the competition. Adhoc analysis of network traffic models or conventional approaches may not capture the complex network behavior that is observed today. Machine learning (ML) network traffic models can be a powerful aid to align network behavior with organizational goals.

CAGR has limitations such as period choice sensitivity, network traffic volatility and seasonality. Incorrect traffic predictions increase the chance of suboptimal investments in capacity, reliability, and/or security. Network burst patterns based on high-profile events such as smartphone updates, major video game releases, or live global sporting events (e.g., the 2022 FIFA World Cup) may be overlooked. CAGR accuracy can be improved by calculating each network segment. In addition, ML network models can provide the additional context required to make informed decisions.

This paper proposes using ML to enable automated iterative calculations and model attributes such as trends and seasonality, failure events, subsequent interactions between the primary and failover links, and network burst patterns. This provides the additional context that is missing from CAGR alone to make the most informed business decisions.

This study considers a portion of a real Internet backbone. It analyzes traffic patterns within four consecutive years, using insights and findings to predict monthly network traffic in the fifth year. Section 2 describes the network under consideration and the data collection process. It discusses challenges incurred in this exercise and highlights influences that deteriorate data quality and the performance of any prediction. The section ends with a traffic modeling exercise and the presentation of accuracy metrics for future model evaluation. Section 3 revisits the traditional CAGR-based approach to network traffic forecasting. It exposes the limitations of the global CAGR strategy and discusses possible improvements for that method. Section 4 explores ML alternatives to network traffic forecasting. It shows how ML models can help to uncover seasonal traffic patterns and provide better forecasting. It emphasizes the importance of choosing the forecasting model appropriate to the data's nature, the presence or absence of seasonality, the prediction horizon, and the complexity of patterns in the traffic. This section also shows models in action and compares the performance of a few ML time series forecasting models when predicting traffic for the reference Internet long-haul. Section 5 summarizes critical intakes for effective and reliable forecasting and indicates future investigations. The paper ends with a list of abbreviations and references.

2. Case Studies

2.1. The Network Graph

The new insights and approaches developed in this paper are tested on a large service provider network with millions of users and a mix of almost all types of internet traffic. The network has hundreds of links with varied capacities, from multiple 10Gs for smaller hubsites to tens of 100Gs for backbone links. Link costs are also varied as some are on the service provider's own fiber infrastructure while others are leases from other providers. Figure 1 presents a logical view of the reference network, with three highlights for future illustrations throughout the paper. Highlighted links all originate in the same city. One represents

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

an internal metro link between two hubsites of the same city. Another link pictures the connection of two cities through the national long haul. The third highlight is typical of international links.

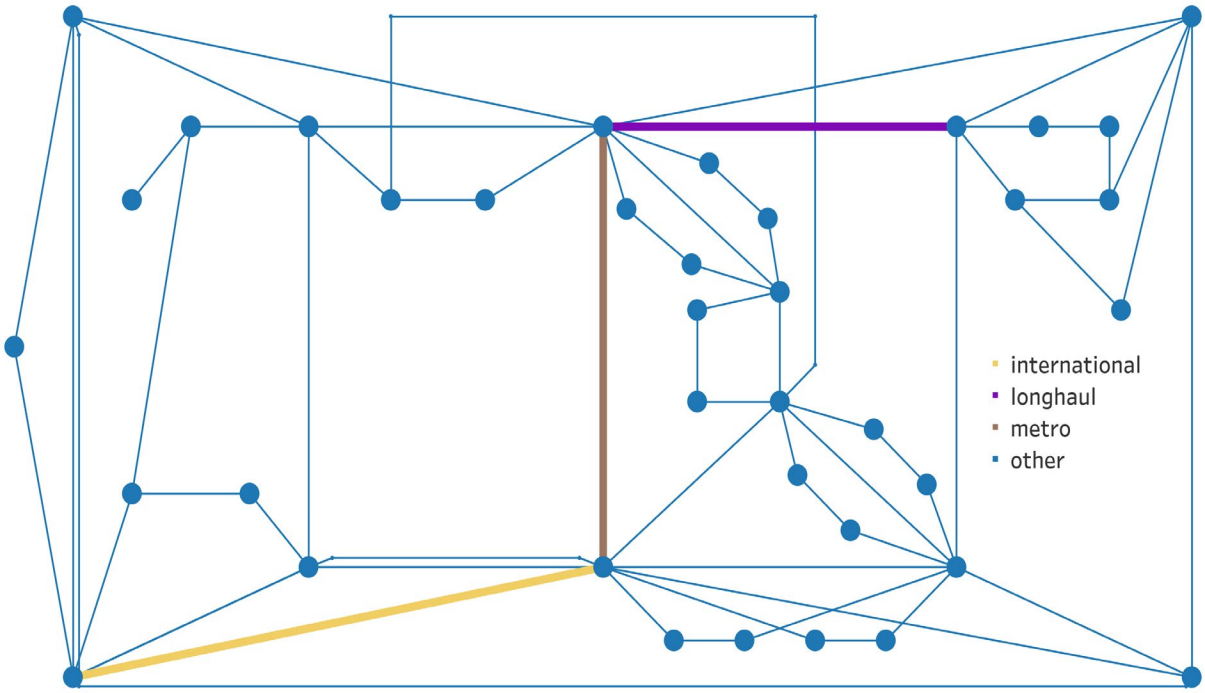


Figure 1 – The Network Topology

Regarding network connectivity, Figure 2 shows the connectivity distribution at a hubsite for the reference network. The connectivity or the degree at a hubsite is the number of links originating or terminating at the given hubsite. For example, the international hubsite degree is 7 while the national hubsite has connectivity 9 for the international link referenced in Figure 1. The histogram in Figure 2 indicates connectivity of 4, 5, or 8 for most hubsites. The graph under study is highly connected, with an average hubsite connectivity of 5.38.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

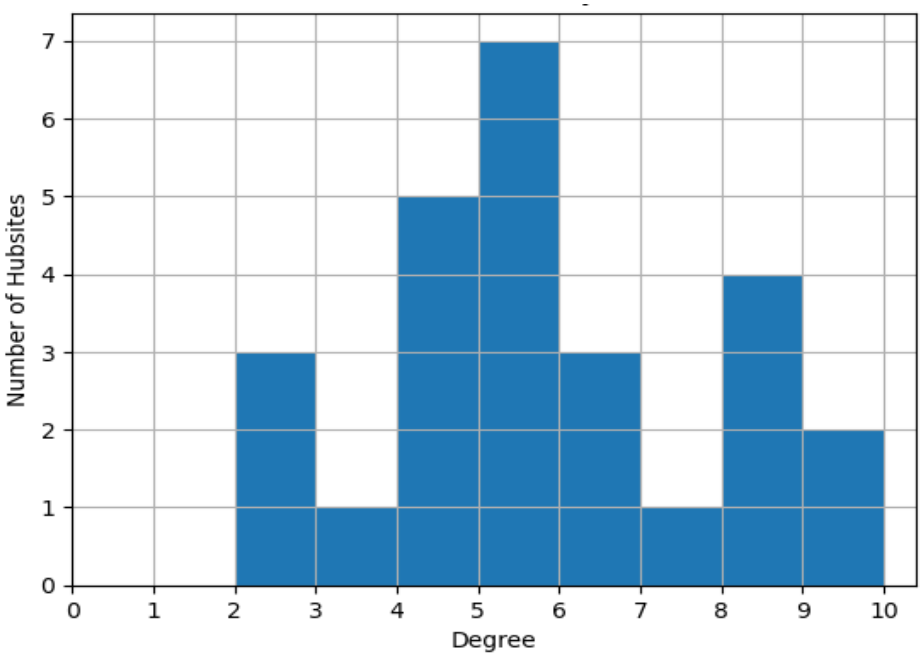


Figure 2 – Distribution of Connectivity at a Hubsite

2.2. Data Collection Over Time

Links are, in fact, bundle-intraconnects between hubsites. Data are measurements of upstream and downstream traffic in bits per second. Using the set of intraconnects from the above-described network, the heatmap in Figure 3 shows data availability as of June 30th, 2023. A dark shade indicates data records on the given date, while a lighter shade is for unavailable data.

Many links in Figure 3 show no data record for multiple reasons. Some hubsite devices are probably missing from the database due to human errors, and high volumes of historical data may have been purged to free up disk space. One data source from this study is an old legacy Network Management System (NMS) with limited storage. The NMS in question also requires manual device entry before being managed. The industry transition from NMS to powerful SDN controllers and orchestrators is recent. Another reason for missing data relates to network normal operations and changes. Network activities include the commissioning, decommissioning, and migration of routers. Changes also involve the turn up/down of new links and capacity augmentation on existing links with different identifiers, etc. Besides network device changes, technology evolves; some bundle intraconnects may have existed in another form (e.g., 1GE to 10GE to 100GE to bundle-ethernet of 100GE).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

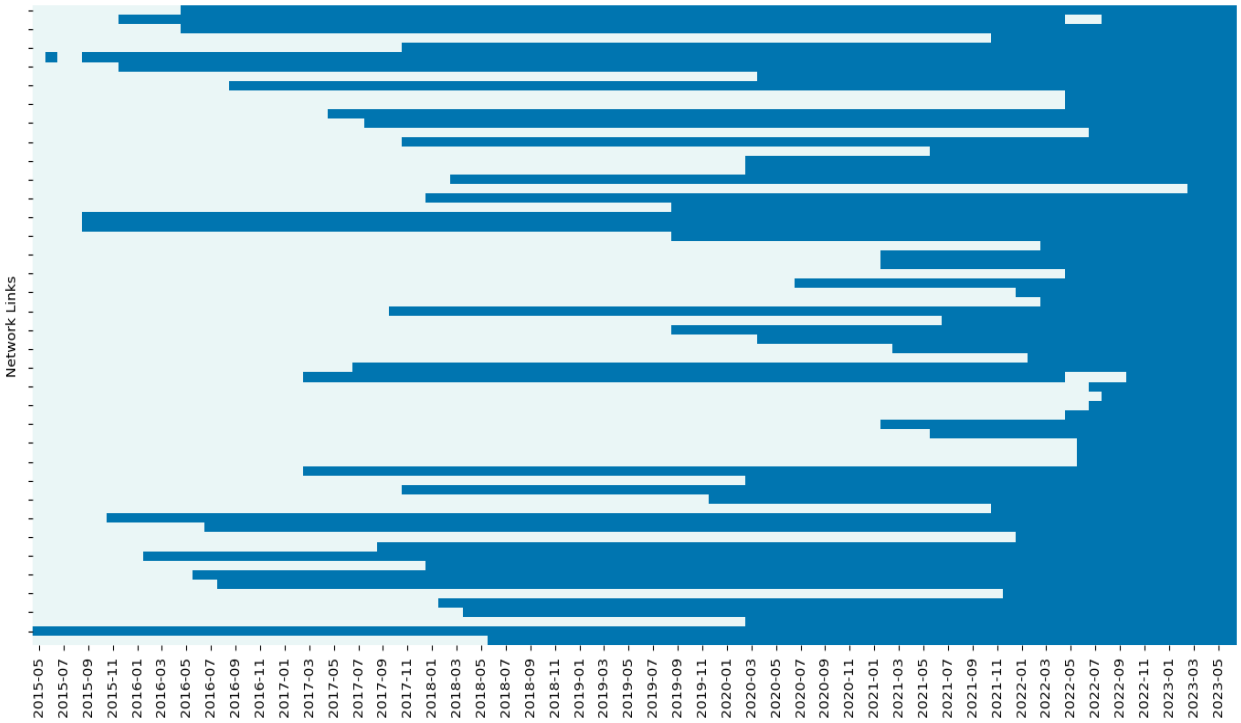


Figure 3 – Availability of Historical Data

Figure 4 and Figure 5 well illustrate legacy network tracking system inconsistency. It is highly improbable that the Internet backbone under examination had less than 30 nodes and interfaces before mid-2017. It is more realistic that either network elements were missing from the tracking system or experienced name changes over time. Considering this assumption, a reconciliation exercise might significantly improve the dataset quality. Reconciling data is arduous and time-consuming. It also requires some logic and automation to maintain the resulting dataset and future data collection clean for analysis.

Data acquisition is beyond this paper's scope. The study focuses on a subset of links with traffic traces for the previous five years as of June 30th, 2023. Those are about half of the links shown in Figure 3, including international, longhaul, and metro reference links in Figure 1, with seven years of historical data.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

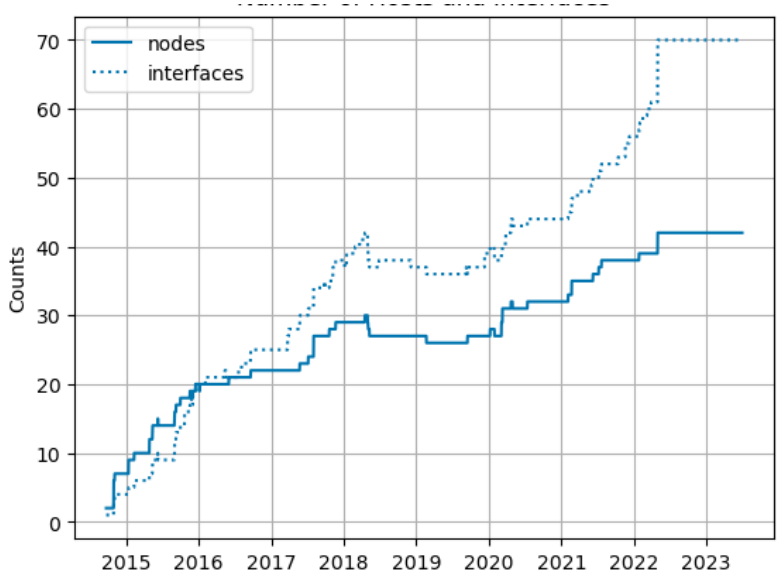


Figure 4 – Number of Hubsites and Intraconnects (based on data collection)

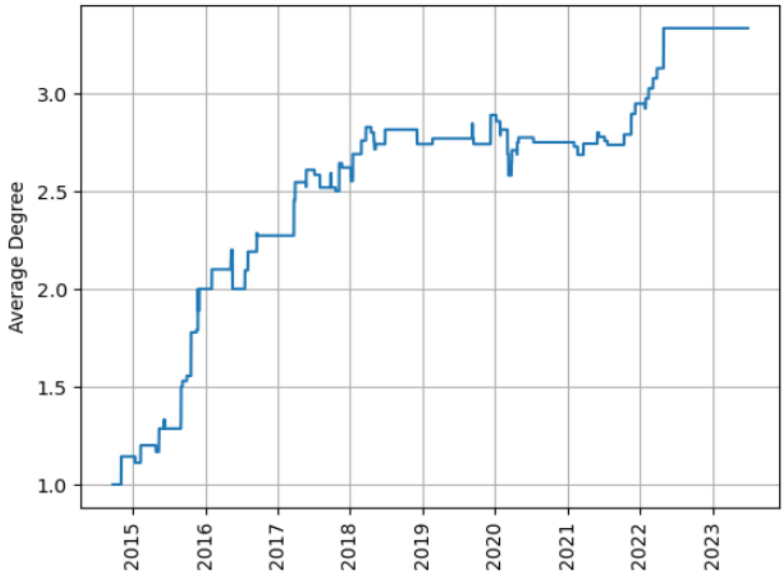


Figure 5 – Evolution of Network Connectivity (based on data collection)

2.3. Traffic Overview and Modelling

As previously stated, our study considers a subset of intraconnects with data from July 1st, 2018, to June 30th, 2023. Upstream and downstream traffic measurements are collected from each hubsite of the link. Links are bidirectional; the dataset is further simplified by approximating the link to the one intraconnect showing the highest data rate between the four traffic measurements for upstream and downstream. That

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

maximum data rate pictures the worst-case scenario and thus is considered traffic usage over the simplified link.

On the other hand, the lowest granularity for traffic, common to all our data sources for this study, is daily aggregations. An ideal granularity is 5-minute samples, which is unrealistic to expect in 10+ years of storage-limited archives before adopting cloud computing. Regarding aggregation functions, one generation collector provided daily maximum, average, and several percentiles, while the other tool only provided maximum and average. Picking up daily maximums allowed consistency along the concatenated data set.

The daily maximums are further aggregated in monthly measurements since it is common to forecast network traffic for the medium to long term—a quarter, a year, and up to five years. In making monthly aggregates, it is ideal to consider the maximum daily 95th percentile (P95) or even take the compound P95 over all 5-minutes samples within given month. Depending on the application, this allows to control high profile and failure events causing traffic increases on (failover) links. Given the nature of the prior-described dataset for this study, reversing and taking the 95th percentile represents a decent monthly aggregate. Figure 6 shows the traffic trends for longhaul, international, and metro reference links. Note the traffic jump in mid-2020, maintained for an entire year on the three links, due to work-from-home during the COVID-19 pandemic.

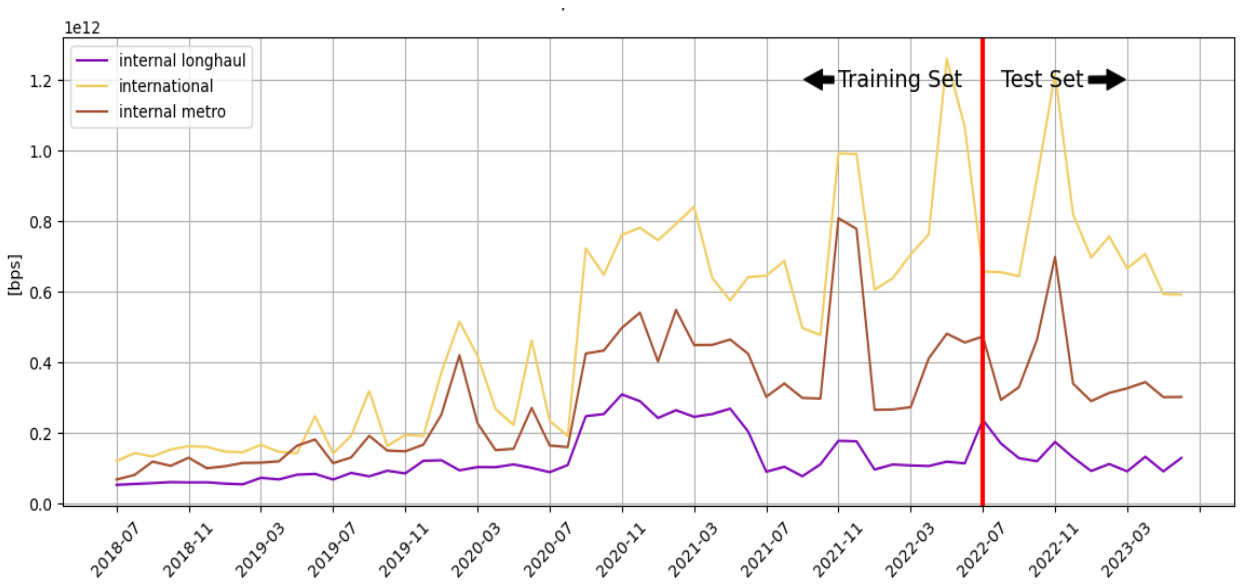


Figure 6 – Historical Traffic Evaluation for Several Links

2.4. The Study

2.4.1. Main Objective

Five years or 60 months of historical traffic measurements are available for the 30 selected links. The first four years or 48 months are used to train the models studying past traffic behavior and patterns. Insights gained from the observations support traffic predictions for the remaining 12 months. In other words, traffic measurements from the first four years make the training set, and the 5th year of data is the test set or comparison set. Goals include revisiting the traditional CAGR approach and discussing why ML is necessary, then building new ML-based forecasting models and conducting a comparative discussion.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.4.2. Accuracy

Accuracy is critical in comparing traffic predictions with measurements in the 5th year. One metric is the root mean squared error (RMSE) which measures the average difference between a statistical model's predicted and actual values, as shown in the following equation. A good predictive model shows a RMSE much smaller than the mean and variance of the training set.

$$RMSE = \sqrt{\frac{\sum(actual\ value - predicted\ value)^2}{number\ of\ predictions}}$$

Although widely used to determine the accuracy of a given predictive model, it may be challenging to compare the performance of different models using RMSE.

2.4.3. ML Pipeline

All ML pipelines were implemented within Amazon Web Services (AWS) cloud computing platform. The authors used S3 for data storage, Athena for data access, and SageMaker to train and test ML models. scikit-learn, statsmodels, pmdarima, and TensorFlow Python ML modules were used for the implementation.

3. CAGR Approach to Network Traffic Forecasting

CAGR is one of the most popular forecasting methods to predict traffic in the cable industry. Total traffic across the entire network at a specific time makes a time series, a regression function is fitted to that time series, and a growth rate is derived to determine CAGR. The resulting CAGR coefficient is used anywhere in the network and whenever forecasting is needed.

3.1. Global CAGR

Simple CAGR implementations typically rely on either linear or exponential regressions. Figure 7 provides time series for the sum of traffic, i.e., blue curve, across the network—or more precisely, for the subset of 30 links under consideration. It is noticeable that fitting an exponential regression function to total traffic, in bit counts aggregates over months, is more appropriate than a linear regression line. The solid black curve corresponds to the exponential regression function that best fits the training set of data. Exponential regression equation is $y = b * a^x$, where regression coefficient a is the growth factor for the monthly periodicity under consideration. $CAGR = ((a + 1)^{12} - 1) * 100$ derives annual growth from the monthly growth rate.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 7 – Global CAGR Estimate

3.2. Limitations of Global CAGR Approach

CAGR is straightforward to implement and is one of the best-known approaches to estimating compound traffic growth rates. However, there is much room for improvement when using the resulting growth factor for traffic forecasting. One drawback is the tendency to naively apply the estimated CAGR to the most recent measurement. For illustration, the solid gray line in Figure 7 shows predictions from applying the global CAGR to traffic measurement in June-2022, the most recent trace on hands. On the contrary, see how dotted gray lines adjust using the same growth factor but with the minimum and the peak measurements over the past twelve months (i.e., from July-2021 to June-2022). Traffic levels are too low or too high in seasons, and spending aligns between the regression and dotted prediction curves.

Table 1 also reveals poor accuracy for the CAGR approach. In all cases, RMSE is higher than the standard deviation although much smaller than the mean. Before applying the global CAGR to individual links, we can already see that its prediction accuracy is questionable. A good practice is tracking predictions over actual values to re-evaluate any forecasting model.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 – CAGR Accuracy Checking [Total Traffic]

Strategy for Predictions	[Mean, std]	RMSE
Exp. reg. curve	[5.06 Tbps, 2.05 Tbps]	3.95 Tbps
Most recent date		2.53 Tbps
Minimum over past 12 months		2.36 Tbps
Peak over past 12 months		5.01 Tbps

3.3. A Distributed CAGR Strategy

The other concern is considering the total-traffic CAGR as the growth factor of every link across the network. Figure 8 illustrates the problem: the x-axis corresponds to traffic peaks for individual links over the 48 months of training, and the y-axis gives CAGR in percentage. The red line is the global CAGR, as discussed in section 3.1. Scatter plots are CAGRs resulting from applying the exponential regression function to each link; 2/3rd are more than 10% above or below the global CAGR. The observation remains true in areas with the most concentration of links traffic or maximum traffic measurements, expected to influence total traffic trends greatly.

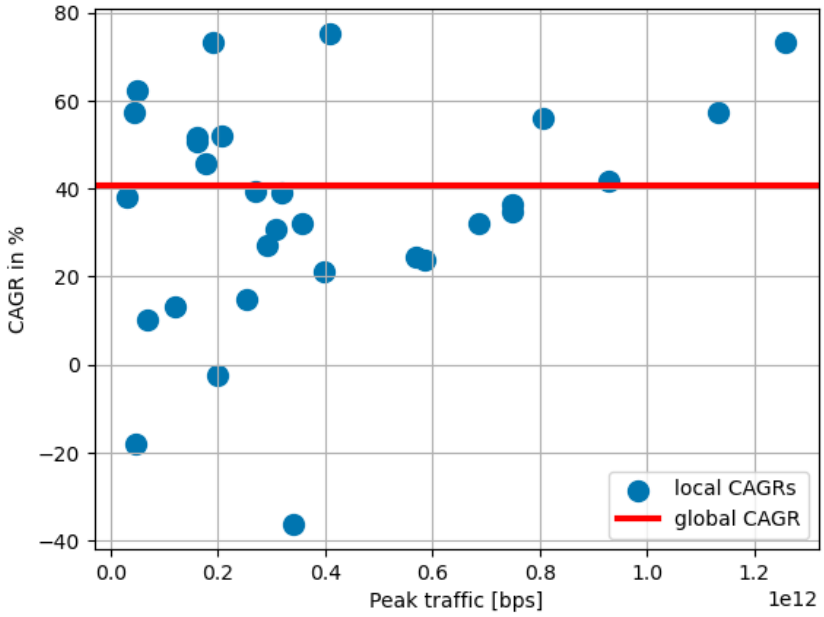


Figure 8 – Global vs. Local CAGR

Figure 9 is for the longhaul illustration, is for the international reference link, and Figure 11 is for the metro link use case. The three charts show predictions when applying the global CAGR in gray vs. the local CAGR in green to the prior-mentioned links. The black curve is for the exponential regression applied directly to the link instead of the total traffic. As before, solid curves are for cases when CAGR is applied to the most recent measurement, while dotted lines are for maximum and minimum measurements within the previous year. The actual traffic measurements match the colors for internal longhaul, international link, and internal metro in Figure 1 network topology. Comparing the green and gray lines

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

illustrates the gap between local and global CAGR predictions. One might be closer or farther from the actual measurements, depending on how similar local and total traffic trends look alike.

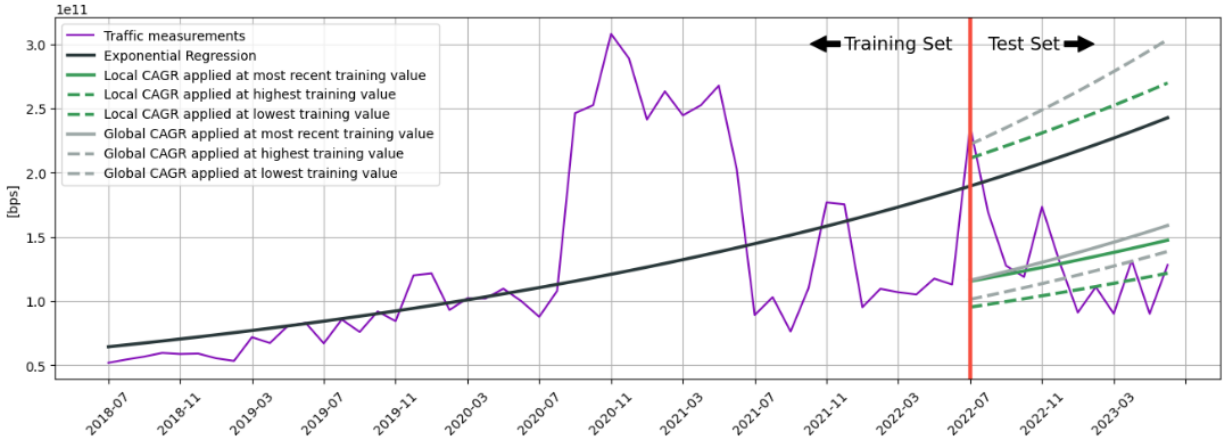


Figure 9 – Global vs Local CAGR Predictions for internal longhaul link

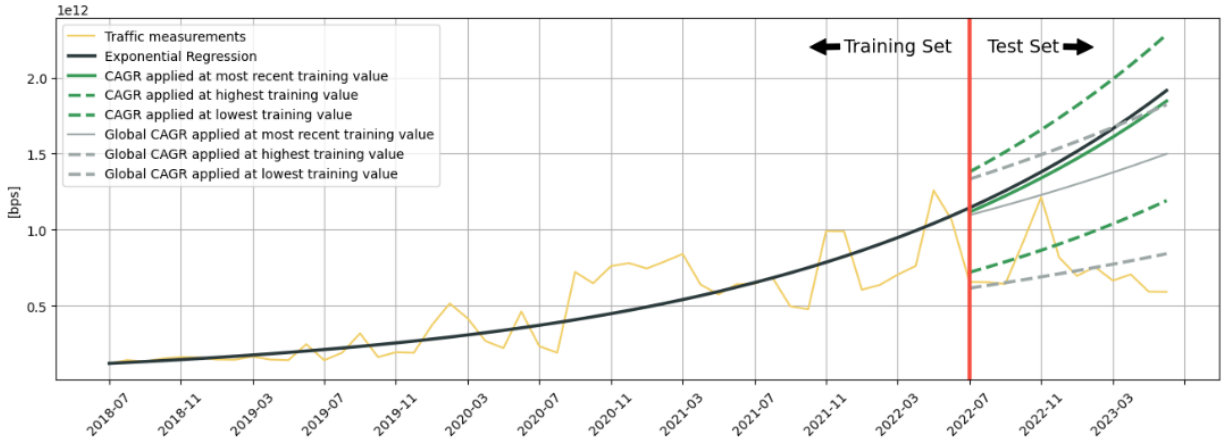


Figure 10 – Global vs Local CAGR Predictions for international link

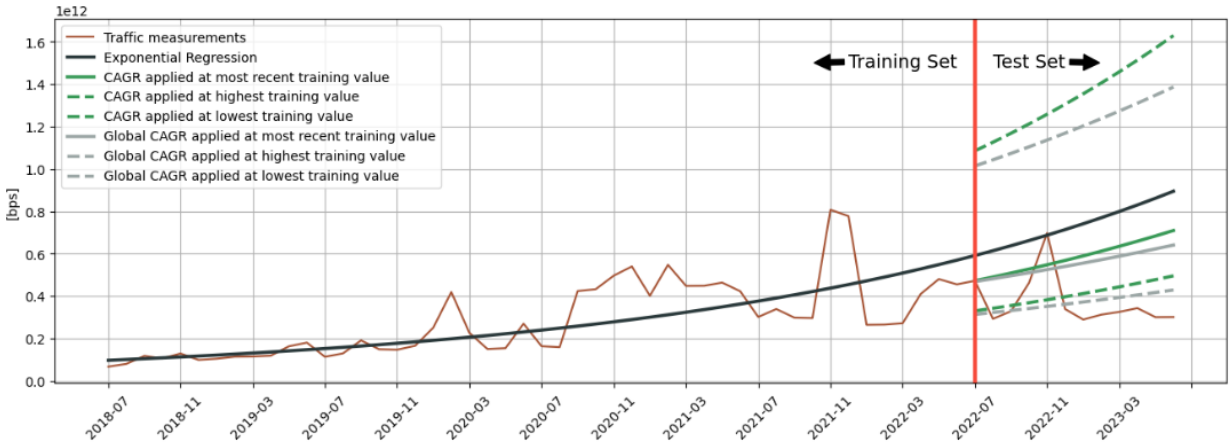


Figure 11 – Global vs Local CAGR Predictions for internal metro link

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In summary, the simplicity of CAGR makes it appealing to cable operators. Trials in Section 3 show that a local CAGR, specific to its given link, performs much better than the total-traffic CAGR. The other intake is to define a logic behind the choice of reference traffic measurements to which the growth factor is applied; otherwise, extend regression curves to future times.

4. Exploring Machine Learning Time Series Models

The level of accuracy for CAGR-based predictions is still very high. The reason is that linear and exponential regressions supporting CAGR calculation focus on the trend only, while intuitions and other empiric observations suggest a certain periodicity in Internet traffic. Machine Learning offers a broad range of models, including time series models, specialized in analyzing and forecasting data collected over time, such as daily, weekly, monthly, or yearly observations.

A typical quick verification to check for periodicity in network traffic is to plot total traffic trends over months or weeks of a year for various years, over days of the months for multiple months, or even days of the week for different weeks. Figure 12 illustrates this principle for the network use case. Comparing yearly trends over months confirms slight seasonality in the data; November shows higher activities than other months, but nothing is clearly defined.

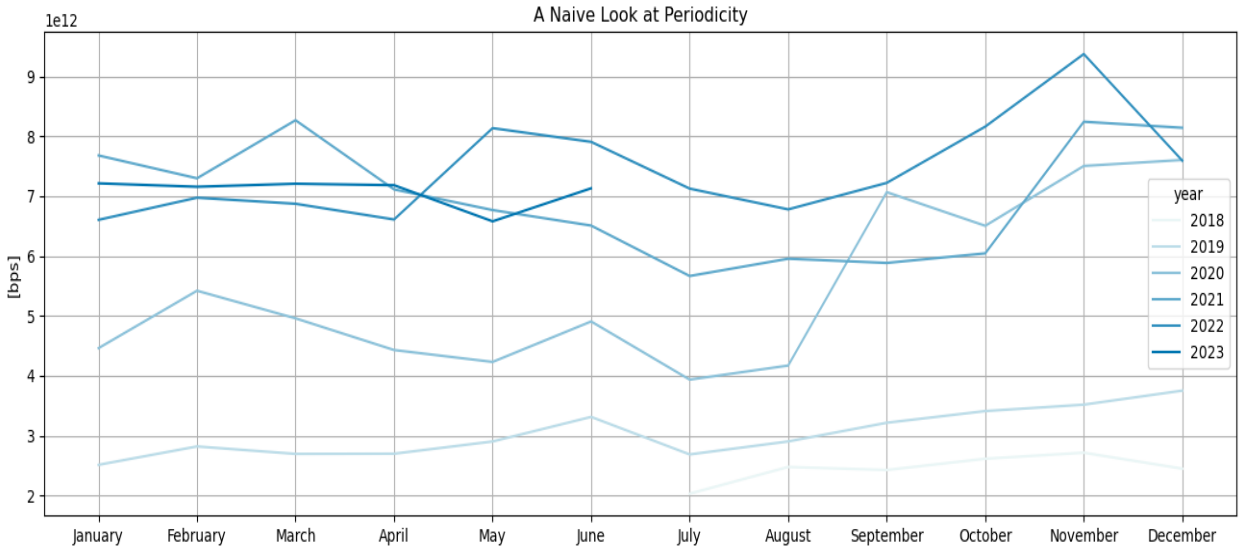


Figure 12 – A Naïve Look at Total Traffic Periodicity

4.1. Seasonal Decomposition of Time Series (STL)

Seasonal Decomposition Time Series (STL) is an ML technique that uses filters to handle seasonality and trend variations in a time series. STL separates trend, seasonal, and residual components from time series data. The trend component gives the overall movement over time, the long-term direction or pattern in the data. The seasonal component is the repeating and predictable patterns at regular intervals, such as daily, monthly, or yearly cycles. The residual component is random fluctuations or noise after removing the trend and seasonal components.

Figure 13, Figure 14, Figure 15 and Figure 16 show decomposed time series applying STL to total traffic and traffic traces from long haul, international, and internal metro sample links. There are two STL options: additive vs. multiplicative; Figure 13 and Figure 14 result from the additive model, while the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

other two come from the multiplicative version. In the additive case, adding trend, seasonal and residual components produce the actual observation; in the multiplicative model, the multiplication of the components produces the actual observation.

STL allows deep insights into underlying patterns and variations within the data. There is a clear yearly cycle in all illustrations; the low season is from July to August, and traffic starts ramping up in September, when students return to school, to reach its peak in November, maybe due to extensive online shopping around Back Friday. There are also local maximums in May and June that might be due to graduation or final games for Stanley Cup, NBA, and NFL sports. The international and metro links also show peaks in February and March; is this happening during the Super Bowl final and the release of new versions of smartphone OS and video game updates? Cable operators may want to confirm the events influencing traffic peaks by decomposing traffic time series at the day or the week granularity. Regarding trends, COVID-19 lockdowns significantly influence traffic trends, even a year after everything returned to normal. However, that lockdown period does not affect the periodicity component.

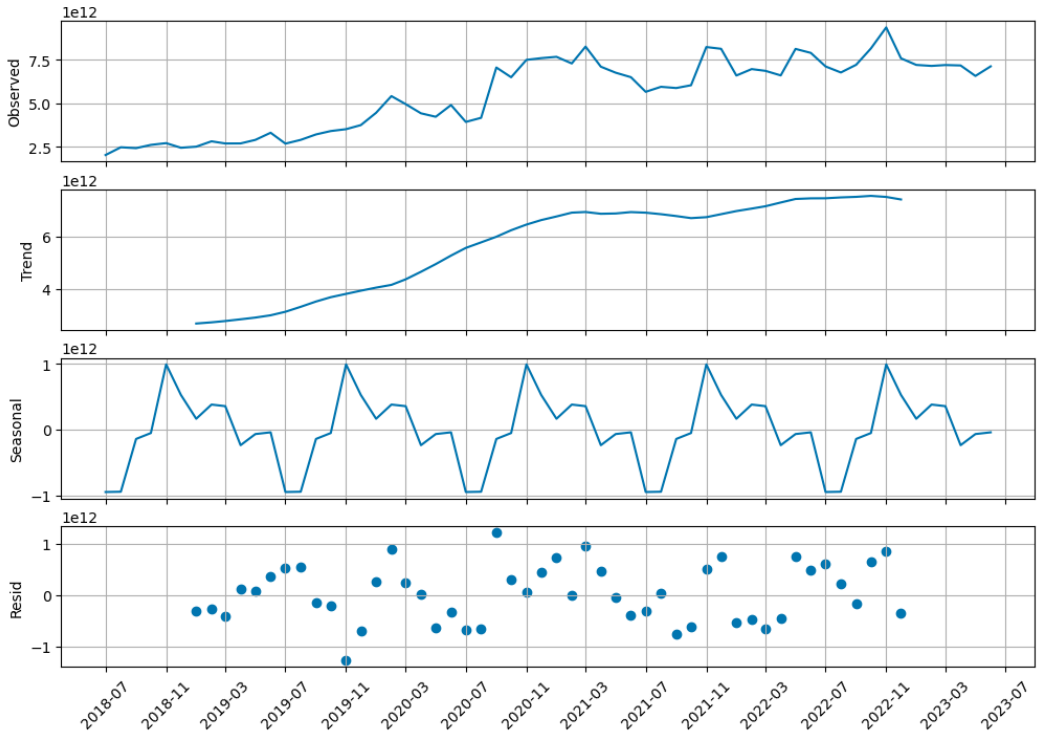


Figure 13 – Seasonal Decomposition for Total Traffic (additive)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

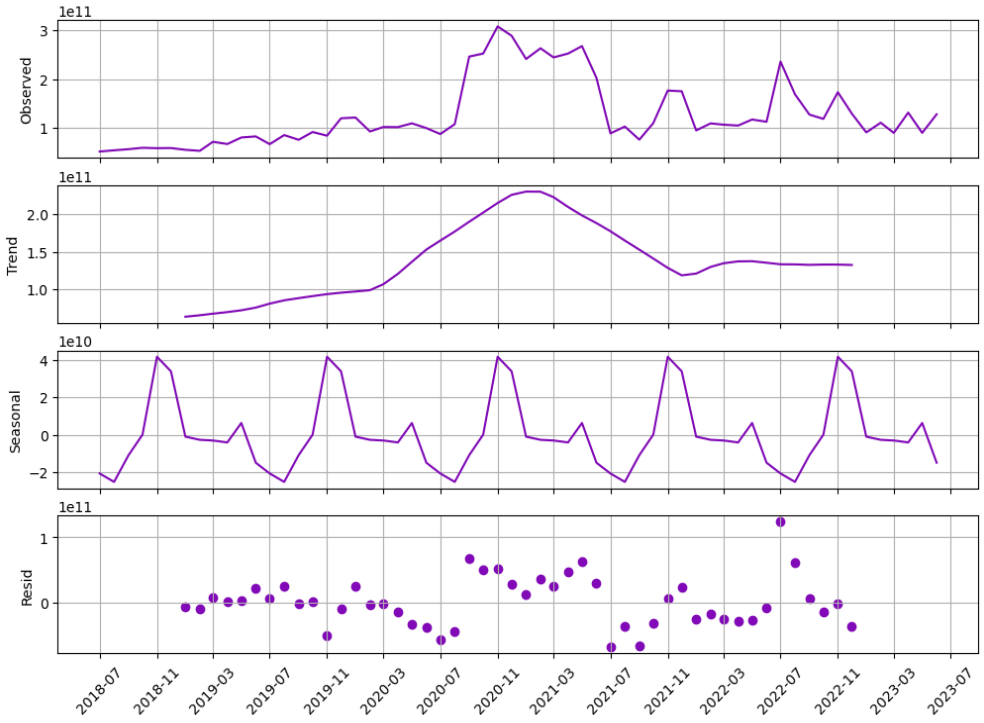


Figure 14 – Seasonal Decomposition for the internal long haul link (additive)

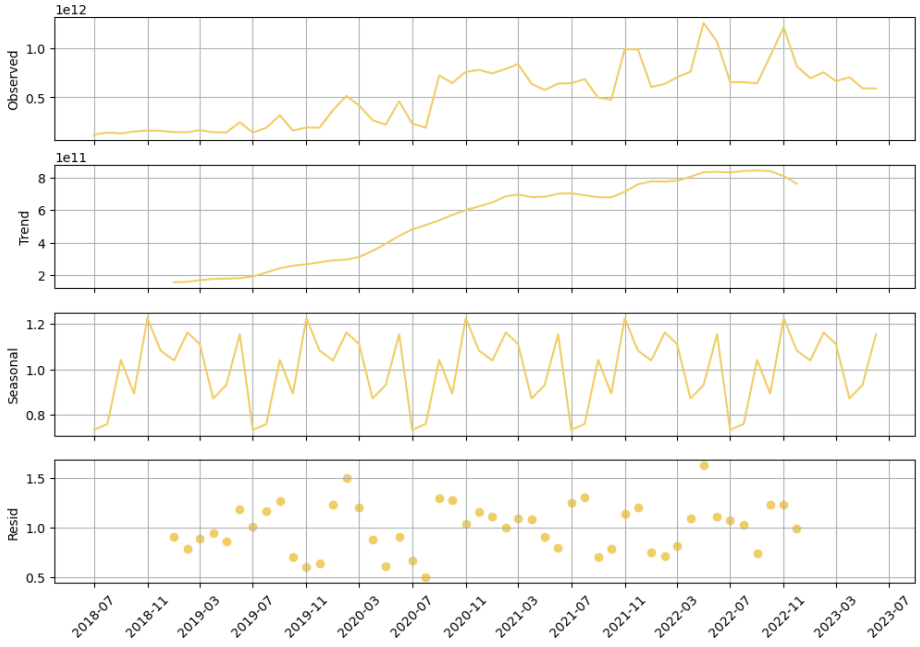


Figure 15 – Seasonal Decomposition for the international link (multiplicative)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

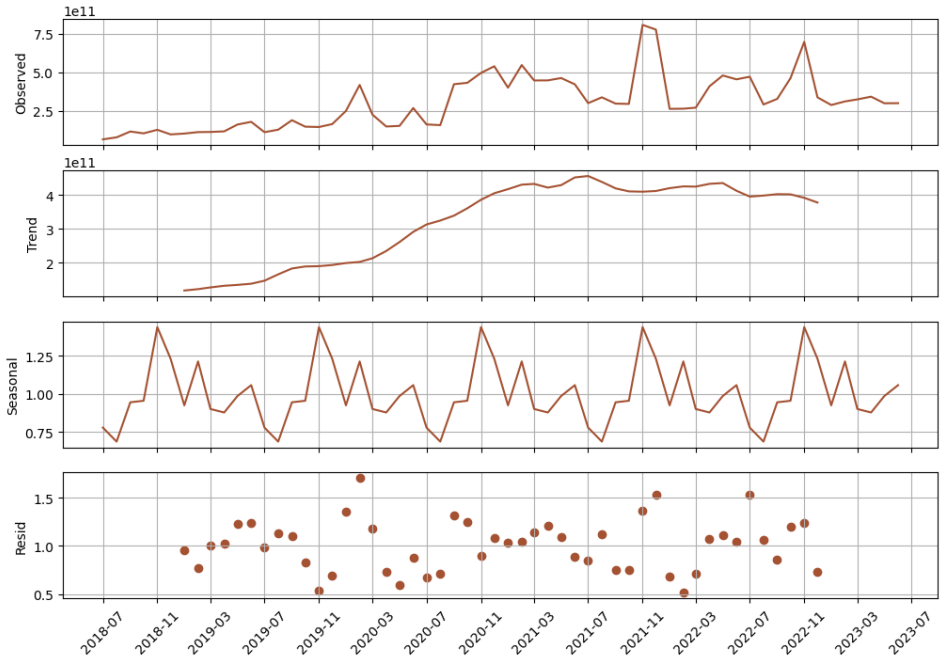


Figure 16 – Seasonal Decomposition for the internal metro link (multiplicative)

4.2. Forecasting Models for Network Traffic Time Series

As earlier stated, time series models consider the sequential nature of the data and aim to capture patterns, trends, and seasonality to make predictions about future values. STL does not directly predict the future but prepares for more accurate forecasts and predictions by identifying both trend and seasonality. There are various time series forecasting models, each with strengths and suitable applications. A few examples appropriate for network traffic forecasting are seasonal autoregressive integrated moving average (SARIMA), holt-winters exponential smoothing (ETS), Prophet from Facebook, long short-term memory (LSTM), and vector autoregression (VAR) extending autoregression to the analysis of interdependencies among multiple time series. The choice of the most appropriate model depends on the nature of the data, the presence of seasonality, the desired forecasting horizon, and the complexity of patterns in the time series.

4.2.1. Holt Winters Exponential Smoothing

Exponential smoothing (ETS) is a family of models using weighted averages of past observations to forecast future values, with different methods for handling trends and seasonality. Holt-Winters is a popular extension of ETS that consists of three components: level, trend, and seasonality. The level or smoothing component represents the time series' overall baseline or average value. Based on the current observation and a smoothing parameter alpha, the smoothing component updates at each time step. The trend component captures the time series' direction and rate of change. It is updated using a smoothing parameter (beta) that adjusts the trend over time. The seasonality component accounts for repeating patterns in the data; it updates with a smoothing parameter (gamma) that adapts to seasonal changes.

The data's presence and type of seasonality instill variations in forecasting with Holt-Winters. No seasonality Holt-Winters ETS is suitable for time series data without any seasonal patterns. Another variant is the additive seasonal model, appropriate when the seasonal fluctuations are consistent in amplitude and added to the trend and level components. The multiplicative seasonal variant is suitable for

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

this traffic study, where the seasonal fluctuations increase and decrease with the level of the traffic time series. Python library ***statsmodels*** provides a framework for the Holt-Winters ETS multiplicative model, ready to be fitted in the data to generate predictions. Section 4.3 discusses Holt-Winters model performance.

4.2.2. Seasonal Auto Regressive Integrated Moving Average (SARIMA)

SARIMA extends the Autoregressive Integrated Moving Average (ARIMA) model and handles time series data with both non-seasonal and seasonal patterns, making it more suitable for datasets that exhibit recurring seasonal variations. ARIMA and SARIMA require making time series stationary before they can be applied. Trend and seasonality are removed from stationary time series to keep the mean, variance, and autocorrelation constant.

SARIMA model has two sets of hyper-parameters, (p, d, q) and (P, D, Q, s); where p is the order of the autoregressive (AR) component capturing the linear relationship between the current observation and its lagged values, d is the degree of non-seasonal differencing required to make the time series stationary, and q is the order of the moving average (MA) component capturing the linear relationship between current observation and the residual errors from past observations. P, D and Q are respectively SAR or seasonal AR, seasonal differencing, and SMA or seasonal MA; they are similar to p, d and q but for the seasonal part. The last parameter s gives the length of the seasonal cycle ($s = 12$, for the use case since data are monthly aggregations with yearly seasonality.)

Python library ***statsmodels*** provides a framework for SARIMA. If this simplifies the model fitting and predicting, it still requires setting and tuning hyper-parameters for prediction accuracy. Literature shares a simple strategy encapsulating SARIMA into a for loop, feeding it with a range of possibilities for each hyperparameter, and keeping the combination of hyperparameters with the least RMSE. Section 4.3 compares SARIMA performance with other time series models for network traffic forecasting.

4.3. Which Model is Best for Network Traffic Forecasting?

Figure 17, Figure 18 and **Error! Reference source not found.** give a closer look at the predictions for longhaul, international, and metro reference links. In all charts, exponential regression shows the worst performance, and ML time series techniques significantly improve predictions. SARIMA performance is inconsistent; the model reproduces previous traffic values after the MA window size. AR and MA hyperparameters require more tuning to perform. Holt-Winters ETS offers a good balance between performance and a simple implementation without further tuning.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

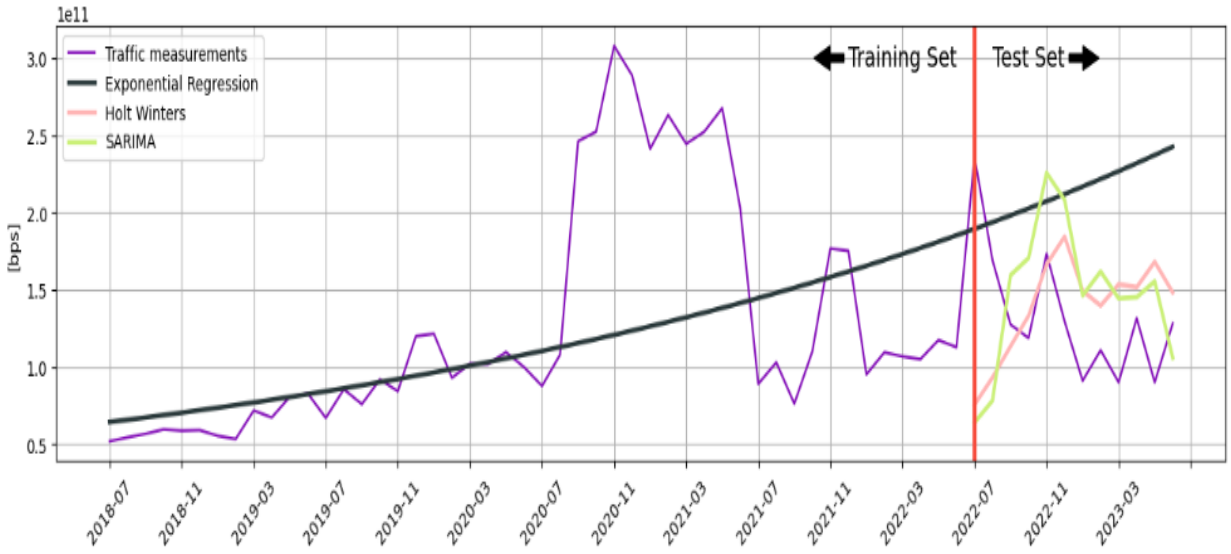


Figure 17 – Comparison of ML Model Predictions for internal long haul link

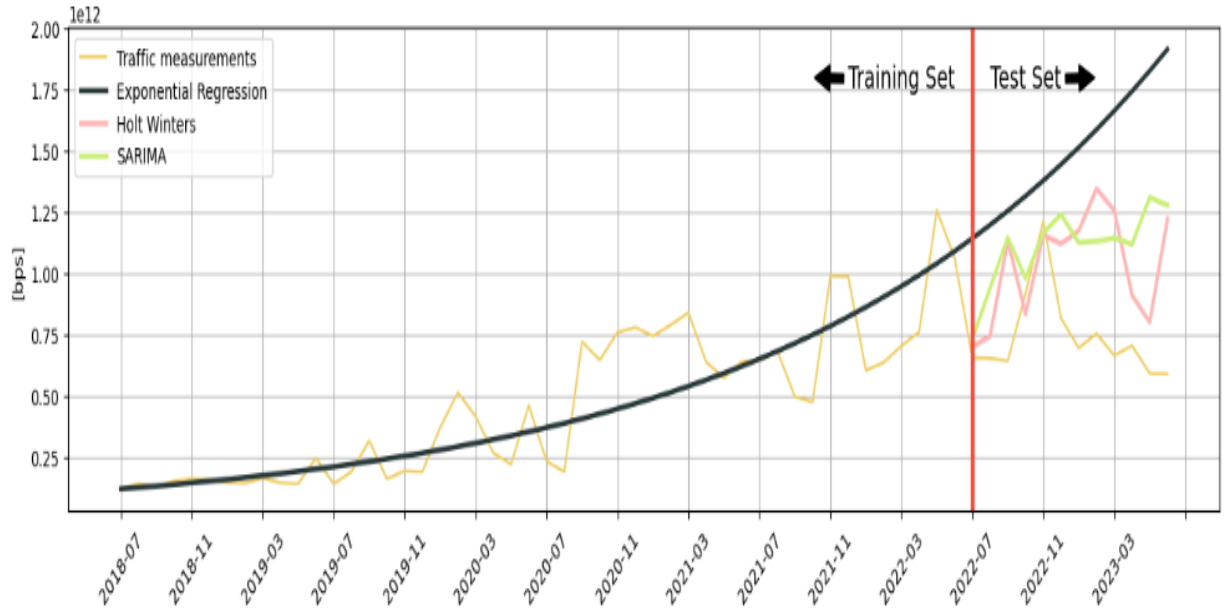
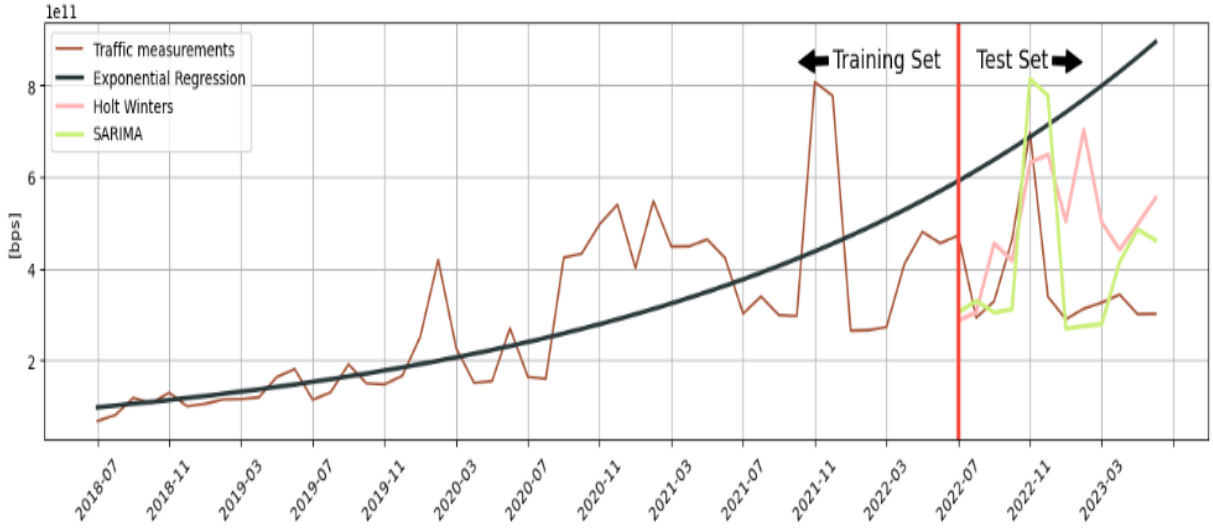


Figure 18 – Comparison of ML Model Predictions for international link

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



5. Key Insights

The following recommendations are based on our study forecasting traffic on a subset of 30 bundle-intraconnects from an internet backbone.

Most people will agree that the insights are only as good as the data used, so data cleaning is crucial to the success of any analysis. Reconciling missing and hidden information from the past can be time-consuming and frustrating, but the outcome is well worth the hassle. Part of clean data collection must also consider network changes, technology evolution, and network management and collection tool migrations. One may want to archive the lowest granular datasets and create suitable aggregations.

Network growth estimated for the entire network is not appropriate for the capacity planning of individual network links. This paper shows that different links have different growth rates, so applying a blanket growth rate for each link may result in inaccuracy in traffic forecasting. The authors recommend estimating the growth rate for individual links separately.

CAGR may be appropriate when determining the growth of the entire network, but it is less accurate for capacity planning. However, it is still widely used by cable operators due to its simplicity. This paper demonstrates how forecasting accuracy is impacted when applying CAGR on the most recent data sample. The authors encourage applying it on a window summary statistic instead.

CAGR's low performance in traffic forecasting starts with blindness to seasonality when calculating growth rates. ML can extract seasonality and trend, making it more appropriate for network traffic forecasting. No need to be intimidated by the term "Machine Learning." In the end, it is doing statistical analysis more efficiently. ML models are readily available in popular programming languages like Python and R or other proprietary programs.

When choosing the appropriate model, attention should be given to the nature of the data set. In this study, Holt-Winters ETS combines simplicity in understanding and implementation with efficiency in forecasting. SARIMA performs as well but requires careful tuning of hyperparameters.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AR	autoregressive component
ARIMA	autoregressive moving average
AWS	Amazon web services
CAGR	compound annual growth rate
ETS	exponential smoothing
MA	moving average
MAPE	mean absolute percentage error
ML	machine learning
NMS	network management system
RMSE	root mean squared error
RNN	recurrent neural network
SAR	seasonal autoregressive component
SARIMA	seasonal autoregressive moving average
SAM	seasonal moving average
STL	seasonal decomposition of time series
SDN	software-defined networking
VAR	vector autoregression
WFH	working from home

Bibliography & References

Norwell, MA: Kluwer Academic Publishers, 2001: *Principles of Forecasting: A Handbook for Researchers and Practitioners*; J. Scott Armstrong

Texts (2nd ed): *Forecasting Principles and Practice*; Rob J Hyndman and George Athanasopoulos

O'Reilly September 2006: *Introduction to Machine Learning*; Andreas C. Muller; Sarah Guido

O'Reilly March 2018: *Machine Learning with Python Cookbook*; Chris Albon

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Machine Learning Model for Customer Claim Prediction in HFC Subscribers

A Technical Paper prepared for SCTE by

Dr. Claudio Righetti
Chief Scientist
Telecom Argentina S.A.
crighetti@teco.com.ar

Matilde Cuenca
Sr Data Scientist Analytics
Telecom Argentina S.A.
crighetti@teco.com.ar

Dr. Diego Martinez Heimann
Tech Scientist Analytics
Telecom Argentina S.A.
crighetti@teco.com.ar

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

Abstract 3

Content 3
1. Introduction 3
2. Problem Description and Data Science 4
   2.1. Data sources 5
   2.2. Model 6
3. Prediction Process 8
   3.1. Next Step 9
4. MVP 10
   4.1. Proactive Case Handling 10
5. Model Evaluation – Results and Discussion 11
   5.1. Three-Week Precision 11
   5.2. Perception of Service Degradation 13
   5.3. NPS 14
6. Conclusions and Future Work 15
7. Acknowledgments 15

Abbreviations 16

Bibliography & References 16

List of Figures

Title **Page Number**

Figure 1 - AIOps 5
Figure 2 – ROC curve obtained for a single day prediction in the training phase. 8
Figure 3 – Current data handling tools utilized within the GCP environment. 8
Figure 4 – Proposed configuration for full scalability and automation. 10
Figure 5 – Proactive attention flow designed for the trial MVP. 11
Figure 6 – Three-week model precision calculated for a single week of list emissions. 12
Figure 7 – Model precision measured over the three weeks following the prediction. 13
Figure 8 – Results for the perception of service degradation (P+S) versus not perceived (NP) obtained with the survey. 14

List of Tables

Title **Page Number**

Table 1 – Training dataset for two CM on a two-day lag case 6
Table 2 – Example of daily results for accumulated feature importance results. 9

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abstract

In this technical paper, we propose a machine learning-based approach for predicting customer claims in hybrid fiber coaxial (HFC) subscribers. We are currently finalizing a proof of concept of the model in some areas of our network.

The proposed approach involves collecting data from multiple sources such as network logs, customer service records, hourly collected information from over 3.5 million cable modem (CM) data over cable service interface specification (DOCSIS) and others. We have been able to overcome many technical challenges, two of the most difficult tasks have been dealing with an extremely unbalanced dataset and label noise.

We use several machine learning algorithms, and finally the one selected was extreme gradient boosting (XGBoost) to build the ML model. As a result, we obtain a daily list of customers with high probability to start a claim (95%). From this list, with Customer Experience (CX) and Field Service teams, we can make proactive calls to solve customer problems remotely or to send a technician to their home if necessary.

The proposed approach can help HFC network service providers to proactively identify potential issues in their subscribers' connections and take preventive measures to avoid customer claims that end up in the generation of a technical service ticket. This can lead to improved customer satisfaction, reduced churn rates, and lower operational costs. We measured through surveys how was the experience of our customers and we found that this proactive action has a positive impact on their satisfaction.

Furthermore, the approach can be extended to other types of networks and where predictive maintenance and customer experience management are critical.

Content

1. Introduction

In our networks and services, the artificial intelligence (AI) has the potential to change, the way we operate, and to become the foundation of the transformation that leads to the fourth industrial revolution. But this requires hard work, a long-term commitment, and a deep cultural change.

In the landscape of broadband telecommunications, the DOCSIS protocol has been pivotal for providing high-speed data transfer over existing coaxial cable systems. The widespread adoption of DOCSIS has revolutionized the way internet service providers (ISPs) deliver their services to residential and business customers. However, the complexity of DOCSIS networks, coupled with increasing demand for seamless, high-quality internet service, has posed several challenges for ISPs. One significant area of concern is the predictability of service issues that lead to customer complaints—a critical metric that directly impacts customer satisfaction and churn rates.

To preemptively address the root causes of service issues, proactive network maintenance (PNM) has emerged as an invaluable tool. PNM has been incorporated into DOCSIS since 2005.

Telecom Argentina has been participating in the CableLabs PNM working group since 2013, adopting the best practices proposed. It’s worth noting that for the working group, improving user experience based on PNM data is a constant challenge.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

PNM uses advanced diagnostics and analytics to monitor the health and performance of DOCSIS networks. By analyzing metrics such as signal-to-noise ratio (SNR), downstream and upstream power levels, and partial service availability and others, PNM can identify problematic network conditions before they escalate into major outages or other service-affecting issues. However, PNM tools often generate a wealth of data that can be difficult to interpret manually.

In this context, the application of AI techniques offer a promising avenue for leveraging PNM data more effectively. AI can automate the analysis of intricate network behavior, thus providing actionable insights into the future state of network health, including the likelihood of customer complaints.

To create a more holistic view of network health and customer experience, we introduce additional data dimensions such as customer experience index (CEI) HFC, Wi-Fi experience index (WEI) [1] , location, user profile, and modem characteristics (CPU, memory etc.) are increasingly being integrated into the analytical framework. These metrics offer nuanced insights into customer satisfaction and can serve as important predictors for future complaints.

The overarching objective is to shift the paradigm from a reactive customer service model to a proactive one. By leveraging advanced AI algorithms to sift through hourly updates from multiple data sources, our predictive model aims to preemptively identify potential service issues. This proactive approach offers threefold benefits: enhancing customer satisfaction, reducing customer churn, and minimizing operational costs associated with service claims. The paper further details the model's prediction process, its validation techniques, and the successful implementation of a minimum viable product (MVP) to empirically demonstrate its efficacy in proactive customer service.

In the sections that follow, we will delve into the methodologies, technologies, and evaluation metrics that have been instrumental in shaping this ambitious project, highlighting its significance and potential impact on the telecommunications industry.

2. Problem Description and Data Science

In this section, we describe the datasets we use in this MVP. A related work is CableMon [1]. At Telecom Argentina, we employ artificial intelligence for IT operations (AIOps) as our operational framework (Figure 1):

[1] Both the CEI and the WEI are metrics that we have developed at Telecom Argentina. The CEI is an indicator of our customers' experience on the HFC network, while the WEI measures the experience in Wi-Fi usage. It is calculated by taking into account the most important devices in the home, which are defined by their frequency and intensity of use. For these devices, a series of metrics are calculated that show interference, signal strength, the age of the devices, usage of both frequency bands (2.4 GHz and 5 GHz), and packet loss during transmission. These metrics are weighted using a polynomial equation, resulting in a value between 0 and 100, with 100 representing perfect Wi-Fi or HFC network performance.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

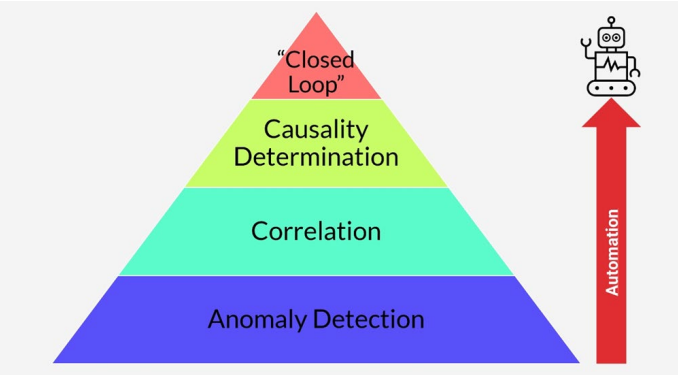


Figure 1 - AIOps

2.1. Data sources

For the training of the predictive model, a collection of anonymized modem-level data is included. The primary source of information considered consists of the proactive network maintenance (PNM) data obtained from the entire spectrum of HFC devices aforementioned. This data is sourced via a universally established standard inherent to the DOCSIS framework. The DOCSIS standard encompasses provisions wherein a cable modem termination system (CMTS) can initiate queries to a DOCSIS-compliant cable modem to retrieve specific performance metrics. These are systematically stored within a local management information base (MIB) in alignment with DOCSIS conventions. Consequently, external processes can harness the simple network management protocol (SNMP) to execute queries on MIBs associated with each CM or CMTS, thereby acquiring comprehensive performance insights [2, 3].

We use this PNM data:

- Timestamp: The time when a PNM data is collected, provided in an hourly basis.
- MAC-ADDRESS: The hashed MAC address of the queried CM.
- SUM_BYTES: Total bytes accumulated in the upstream (US) or downstream (DS) direction.
- AVG_TX_US: Average CM's signal transmission power in the upstream direction.
- AVG_RX: Average received signal power in the CMTS.
- MAC_CER: Maximum package loss percentage.
- MAX_CCER_US: Maximum package correction percentage.
- MIN_SNR: Minimum signal to noise ratio.
- TIMEOUTS: Number of T3 and T4 timeouts the CM has experienced since its last reboot.
- SYSUPTIME: Time since the CM is on.

In addition to the PNM data, we use some other sources of information referred to each individual CM:

- HFC Customer Experience Index (CEI) [2]
- WiFi Experience Index (WEI)
- Location.
- User profile.
- Modem characteristics.

[2] In the CEI, we include additional PNM data.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Hired service specifications.

Finally, we use as the target of the model, the records of customer tickets generated over the last days which ended in a technical service (TS) being dispatched to the client's location. The relevant fields in each record include the ticket creation time, current status (and closing time if the ticket was resolved), and the distinctive identifier of the customer's account.

2.2. Model

As previously stated, our goal is to predict, on a daily basis, which devices in our HFC network—complying with DOCSIS 3.0 or 3.1 specifications— are likely to generate a customer claim that will escalate to a technical service (TS) assignment within the next few days. Currently, our HFC network comprises approximately 3.5 million devices, which collectively result in a large volume of daily calls to our call center for various reasons. On average, 10% of these interactions escalate to a TS assignment each day. Given the sheer number of devices on our network, the number of such escalations represents less than 0.1% of total daily interactions. This presents us with a highly unbalanced classification problem.

To reduce the volume of information processed, we aggregate the available hourly data into daily values for each individual MAC address. These aggregated daily values include the mean, maximum, minimum, and standard deviation calculated for each of the proactive network maintenance (PNM) data variables.

Given the nature of the available data and the challenges we face, we opted for a technique known as 'Time Series to Supervised Learning.' This approach transforms time-dependent data into a format suitable for conventional supervised learning algorithms. Specifically, it reformulates the time series dataset so that each row represents a unique sample, complete with input features and a target output. This is achieved by generating 'lag features,' which are created based on a predetermined number of previous time steps (or 'lags') to use as predictors for current values.

For each time step, new columns are added to the dataset, each containing the values from prior time steps. This results in a restructured dataset with shifted values, allowing us to apply traditional machine learning algorithms for predictive analysis. This method capitalizes on the inherent temporal dependencies found in time series data, enabling us to employ a wide range of machine learning techniques for forecasting tasks.

An example on how the training dataset looks on a daily basis is shown on Table 1. In this case, a sample representing two CM with a two-day lag is displayed, as if the only predictive feature was SNR.

Table 1 – Training dataset for two CM on a two-day lag case.

ID	date (t-2)	SNR (t-2)	tgt (t-2)	date (t-1)	SNR (t-1)	tgt (t-1)	date (t)	SNR (t)	target
A	20221015	1	0	20221016	2	0	20221017	3	1
A	20221016	2	0	20221017	3	1	20221018	4	1
A	20221017	3	1	20221018	4	1	20221019	5	1
A	20221018	4	1	20221019	5	1	20221020	6	0
B	20221015	7	0	20221016	8	0	20221017	9	0
B	20221016	8	0	20221017	9	0	20221018	10	0
B	20221017	9	0	20221018	10	0	20221019	11	1
B	20221018	10	0	20221019	11	1	20221020	12	1

Features named as previous targets (t-2, t-1) indicate wether the CM had a TS in course on the corresponding date. Then, in the case of the client labeled A, the data shows that a claim was started on

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

October 17th and remained open until the 19th, when it was closed. For client B, a claim was initiated on October 19th, and it was still open on the last day analized (October 20).

For the prediction dataset we only use one day (file) for each CM, and the target column becomes the model target. In typical conditions, during the active stages of the project (discussed further ahead in the work) we used around 1 million CM for model training, with a time-window of 40 days. In the case of the prediction dataset, we worked with a 300 thousand CM set.The Time Series to Supervised Learning approach has many benefits, some of which include: compatibility with traditional machine learning algorithms, simplification of the feature engineering processes, ease of variable interpretability and feature importance analysis, among others. In our case we selected a time window of five days for the lag duration, and periodically retrained our model coincident with that time.

In terms of model selection, after several tests the choice was made for an XGBoost model, which is known to be particularly well suited for imbalanced classification problems due to several intrinsic characteristics, among which can be mentioned that it is a strong ensemble model that makes use of regularization techniques, especially important in unbalanced datasets, consistent handling of missing values, its robustness to outliers due to the nature of decision trees, etc.

For hyperparameter tuning, we employed a time series cross validation (TSCV) approach featuring a 5-fold split and a 5-day lag gap to offset the temporal dependence introduced by our time series to supervised transformation. This setup was designed to closely emulate the intended predictive application. The optimal hyperparameters we identified are as follows: scale_pos_weight set to 740, **reg_alpha** at 0.004, **n_estimators** at 900**, max_depth** at 4, and **learning_rate** at 0.031.

Traditionally, accuracy has been the default metric for evaluating classification models. However, in cases of class imbalance, relying solely on accuracy can be misleading. The model may achieve high accuracy simply by predicting the majority class, which is generally not the desired outcome in such scenarios. To mitigate this, we selected recall as our refitting parameter. Recall is particularly useful for focusing on the model's ability to correctly identify positive instances, which is often the more critical objective when one class is significantly underrepresented.

Our model's performance was assessed under a typical training scenario, yielding a receiver operating characteristic - area under the curve (ROC-AUC) score of 0.97, as indicated in Figure 2. Additionally, we achieved a precision of 0.11 and an F1-score of 0.13 when using a 90% probability threshold and a one-day prediction window. It's worth noting that the duration of the prediction window emerges as a significant factor in evaluating precision outcomes, a point we will explore further in subsequent discussions.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

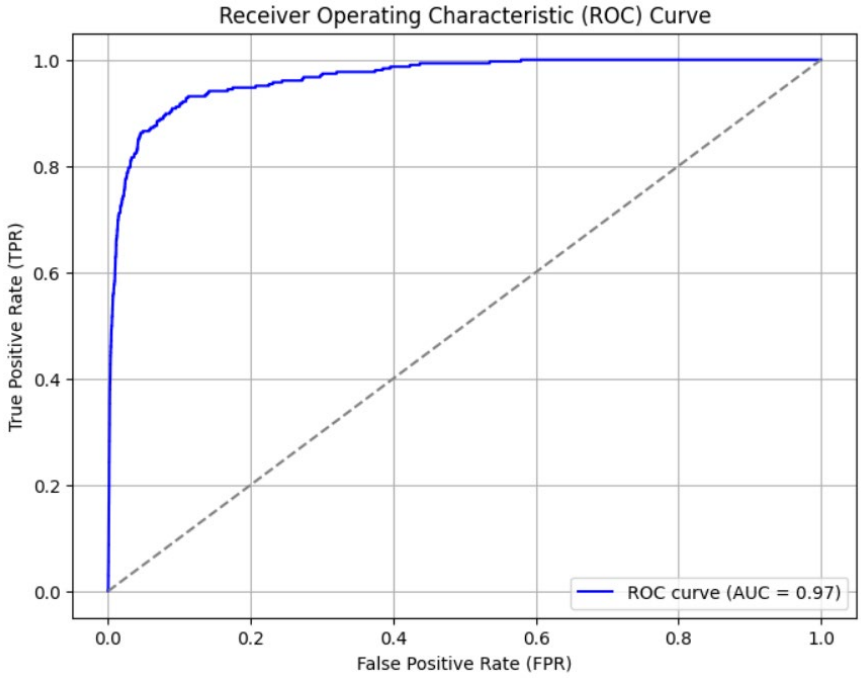


Figure 2 – ROC curve obtained for a single day prediction in the training phase.

3. Prediction Process

This MVP is entirely hosted on the Google Cloud Platform (GCP). As previously mentioned, various data sources—including technical metrics from cable modems, customer information, claims history, and both WiFi and Customer Experience Indexes—are imported into Google Cloud Storage from our on-premise data lake. The data is then processed using Google Cloud Functions and stored in BigQuery for streamlined access. For the analytics component, we leverage Vertex Notebooks to perform daily predictive analyses as well as weekly model training.

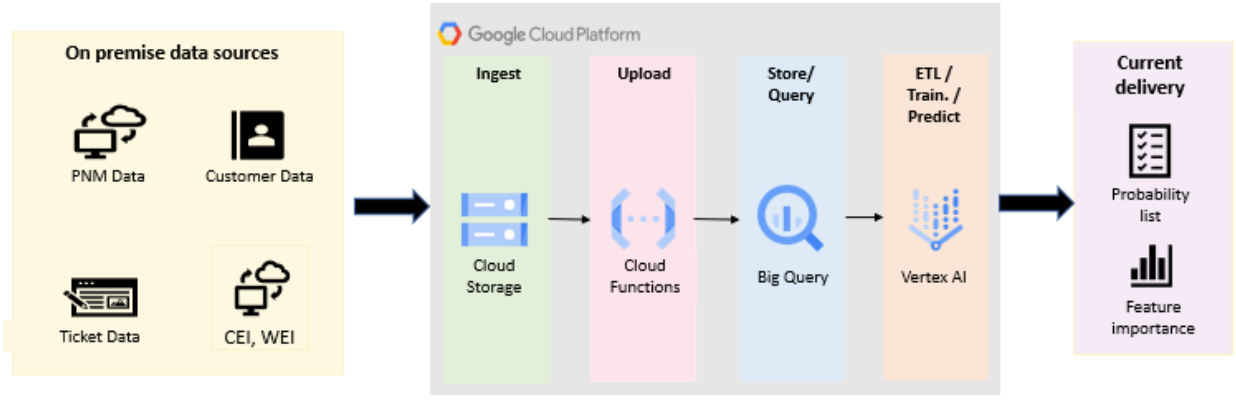


Figure 3 – Current data handling tools utilized within the GCP environment.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Through this configuration, we were able to perform a weekly training over a portion of the entire available data comprising the information of 1M CM for time period of approximately 40 days (of the 3.5M HFC devices available). On the other hand, given the specific characteristics of the experimental procedure (discussed in the following section), the daily predictive process was performed over a set of 300k CM with a lag time window of five days.

Table 2 – Example of daily results for accumulated feature importance results.

Feature	Description	Weight
STDpnm_data.SYSUPTIME_t	SYSUPTIME standard deviation for the predicted day (t)	187
week_day_number_t	Day of the week in a [0-6] range	137
WEI	Wi-fi experience index	127
CEI	Customer experience index	125
MINpnm_data.SYSUPTIME_t	Minimum SYSUPTIME value recorded during the predicted day (t)	125
AVG_CEI_by_node	Mean CEI value for the CM node over past week	104
MINpnm_data.SUM_BYTES_DS_t	Minimum downstream SUM_BYTES for the predicted day (t)	101
STDpnm_data.AVG_RX_US_t	Standard deviation for the upstream AVG_RX value for the predicted day (t)	96
STDpnm_data.SYSUPTIME_t2	SYSUPTIME standard deviation for two days before the predicted day (t2)	90
MINpnm_data.SNR_US_t4	Minimum upstream SNR value for four days before the predicted day (t4)	85

The daily output of the model consisted of the list of customers with the highest calculated probabilities of starting a claim that would eventually end in a TS. The length of the list, or equivalently, the probability threshold was determined by the amount of calls that could be made on a daily basis by the field service teams dedicated to the project. Along with this list, we made use of the Shapley additive explanations (SHAP) library [4] to calculate and add the information of the most relevant features for each prediction individually; this information was further analyzed in order to try to understand the origin of each perceived service inconvenient and develop automated strategies for proactive attention, as will be discussed in the following sections. An example of the results obtained for this feature in an accumulated case for a single prediction day are shown in Table 2. The final index in some feature names (t, t2, t4) indicates whether each variable corresponds to 0 to 4 days prior to the prediction date, according to the time series to supervised learning approach.

3.1. Next Step

To achieve scalability across the entire HFC network, the project is planning a transition from the current extract, transform, load (ETL) stage to a Google Composer-based process. This new stage will leverage the data build tool (DBT) library [5] to handle raw data using standard SQL queries in BigQuery, a shift that is expected to significantly enhance both performance and resource efficiency. Additionally, the existing train/predict operations are being converted into automated workflows through Vertex Pipelines. This automation will streamline the end-to-end process, from model training and prediction to final output. The architecture of this proposed system is illustrated in the figure 4.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

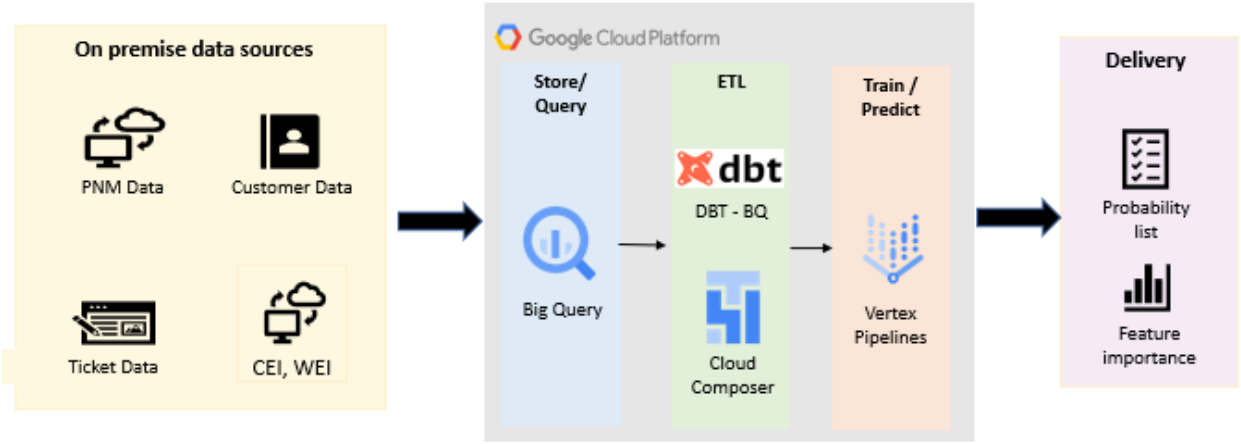


Figure 4 – Proposed configuration for full scalability and automation.

4. MVP

To deepen our understanding and optimization of the MVP process, we developed a proactive attention protocol in collaboration with multiple internal teams, including Quality Management, Customer – Technical Support, Field Service – Selected Technical Bases, Customer Voice, Service, and Field Services. Through iterative refinement of the process, we investigated and identified the most effective procedures, culminating in the development of a final filtering and attention protocol. The key conditions established for the targeted customer segments were:

- Not have an opened or closed claim in the 5 days previous to the listing.
- Not have an ongoing ticket.
- Have a calculated probability higher than 80%
- Not have a massive service interruption reported.
- Have schedule availability from the FS for the following 24/48 hours after the call.

Regarding this last restriction, it was found during the early project stages that if a proactive call is made to a customer which cannot be responded to in a short time period, this tends to be more harmful to the customer perceived experience than not taking action at all.

4.1. Proactive Case Handling

Having established both the daily outcome and the attention process, a trial was performed in a group of selected geographic regions where the service availability was enough to assure dispatch in one or two days after the communication.

For this trial minimum value product (MVP) a subset of approximately 300 thousand CM was established for which the model was weekly trained and daily predictions were made. Following the lessons learned in early stages, a process was implemented for the proactive attention where each customer was contacted according to a specifically designed flow, shown in Figure 5.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

First, the prediction list was swept via phone calls made by four Customer Service (CS) representatives. In this step, the average sweep capacity was around of 30% and the effective contact was 30% of that amount. Next the customer was asked if he had perceived a degradation (PD) of the service in the 72 hours prior to the call. For the affirmative answers (excluding the negative ones and the ones that had had a problem but this was solved independently of our intervention), the customer was inquired on whether he was willing to accept the proactive management of the problem.

The customer was then asked if he was available to perform a series of checks and procedures to further assess the service condition, and lastly, a remote diagnosis was performed via automated tools in order to establish if the interaction was due to finalize with the dispatch of a FS or, in certain conditions, the problem was able to be fixed via remote solutions conducted by our CS teams.

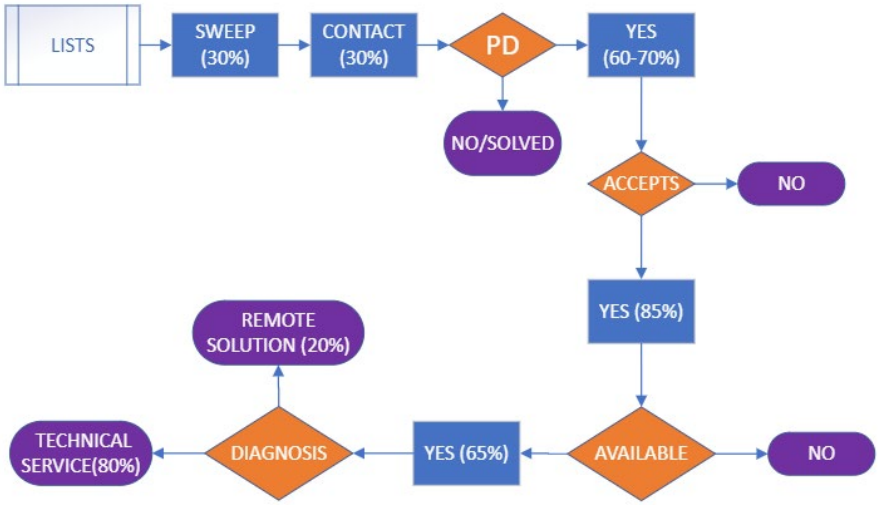


Figure 5 – Proactive attention flow designed for the trial MVP.

With the objective of making the most profit from the customer interaction as possible, a set of ad hoc procedures were developed for the proactive case handling. For example, in cases where the customer was aware of a service difficulty, but he/she was unavailable, or didn't want a home visit, and under certain configurations of the predictive features that were understood to be highly linked to a service dispatch, the technical visit was automatically arranged or the remote solutions available were applied. Some of the key results obtained during this trial are discussed in the following section.

5. Model Evaluation – Results and Discussion

This section focuses on the evaluation of the model's performance and ensuing discussions. To evaluate the MVP, we defined an *ad-hoc precision*.

5.1. Three-Week Precision

During the evaluation phase of our model, we observed that clients frequently did not initiate contact immediately the day after a prediction was made; rather, they often reached out several days or even weeks later. Another noticeable trend was that repeated predictions for a client within a single week increased the likelihood of that client eventually filing a claim, thereby improving the model's predictive accuracy for that particular group. Given these observations and the importance of minimizing unnecessary technician dispatches—which represent a waste of both time and resources—'custom' precision emerges as a

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

particularly relevant metric for assessing the model's effectiveness. Consequently, precision has become one of our key performance indicators for model evaluation.

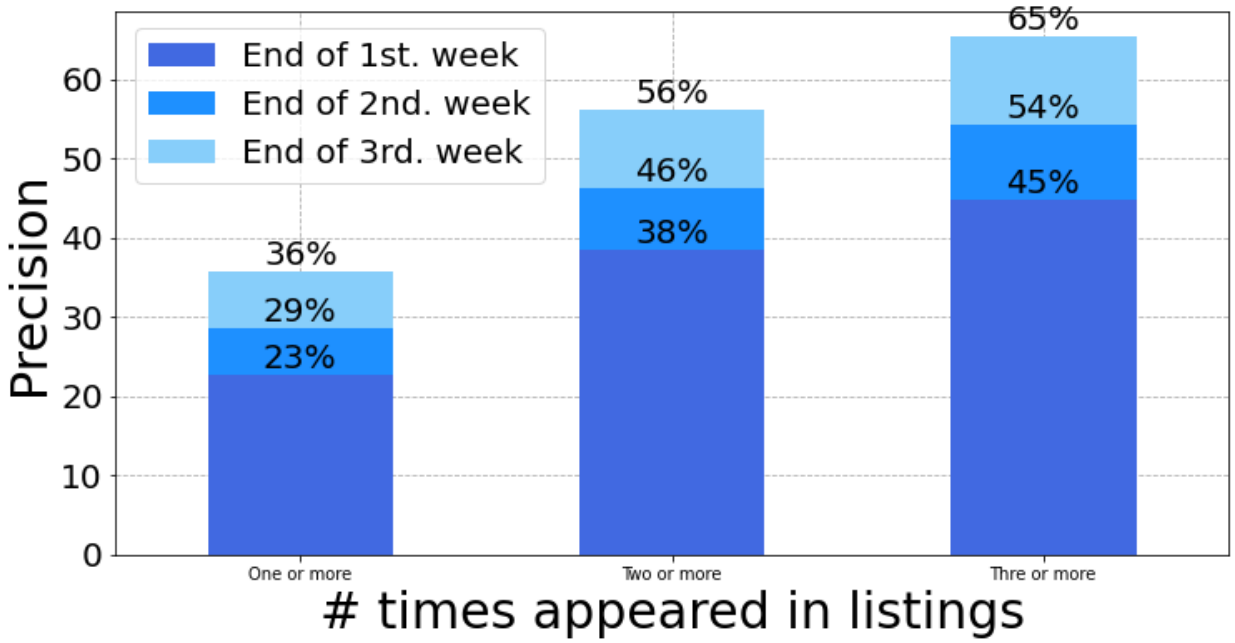


Figure 6 – Three-week model precision calculated for a single week of list emissions.

To account for the inherent delays in customers' claims reporting, our evaluation includes communications that take place up to three weeks after the initial prediction. As a result, a custom precision metric has been developed. Moreover, to mitigate any potential biases arising from proactive outreach initiatives, precision is assessed solely within a control group.

True positive cases are classified as those where clients file a claim subsequent to a prediction. These cases are further categorized based on the time elapsed between the prediction and the actual filing of the claim. We also differentiate among cases based on the frequency of predictions for the same client within a given week—whether they were predicted once, twice, or multiple times within the same week, as shown in Figure 6.

Our data show that while the majority of customer communications occur within a week, a significant proportion take place after one or two weeks. This observed increase in precision among clients who were predicted multiple times aligns with our earlier observations.

To simplify the evaluation metric into a single, unified value, we chose to define our 'custom' precision as the ratio of customers who, having appeared at least twice on our weekly prediction lists, initiated a call leading to a technical Service (TS) assignment within the three weeks following their inclusion on the list. Results for this tailored metric are illustrated in Figure 7. Here, we display the temporal evolution of precision, broken down by the minimum probability thresholds used to filter customers onto the lists. For instance, the curve corresponding to a 0.85 threshold includes only cases where the calculated probability exceeds 85%. It's readily apparent that higher thresholds correspond to increased levels of precision.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

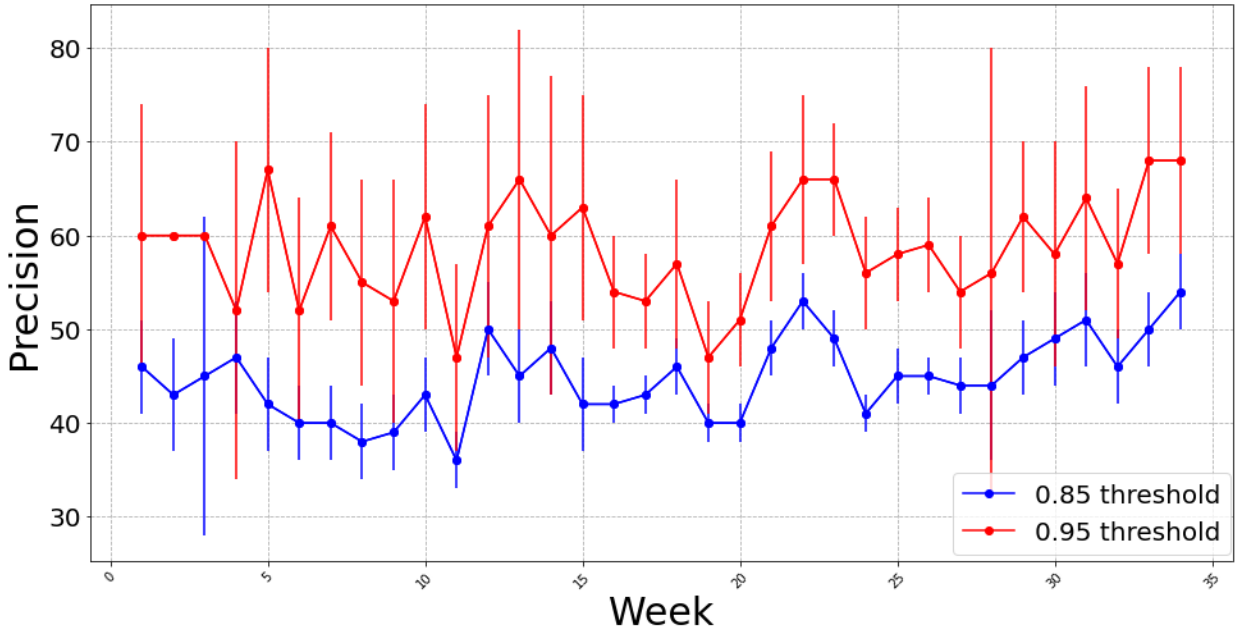


Figure 7 – Model precision measured over the three weeks following the prediction.

The data reveals that, on average, for a threshold of 0.95, our model achieves a precision of approximately 60%. This means that 6 out of 10 predicted customers do indeed file a claim in the weeks following their appearance on the prediction lists. Considering that, as will be further discussed, not every customer initiates a claim, we believe this result underscores the reliability of our model. Additionally, we conducted experiments simulating random customer lists and analyzed the percentage of received calls. We found that for a random sample, the precision falls within the range of 5-10%, which reinforces our initial impression regarding the model's effectiveness.

5.2. Perception of Service Degradation

Given that many clients experience difficulty in establishing contact due to data center congestion, the study augments the precision evaluation of contacted customers by assessing their perception of service malfunctions, indicated as the percentage of perceived malfunctions among those contacted (perceived degradation of service, P), data shown if Figure 8. The value also includes within the perceived curve, those cases where the customer reported to have noticed a malfunction in the service but this was automatically solved in the days passed between the event and the communication (solved cases, S). Finally, the orange curve in the figure represents those cases where the contacted client did not feel that the service had presented any inconvenience in recent times (Not Perceived, NP). It general, it can be seen that the predictive model, after some adjustments made midway in terms of parameter tuning and other refinements, is able to correctly identify between 6 and 7 out of 10clients that are feeling a problem in the service provided.

As previously indicated, the MVP incorporates solicitation of client feedback through surveys, with the aim of validating not only the existence of service-related concerns from their perspective but also their inclination to initiate claims. A comparative analysis between precision and perception underscores that the latter consistently exhibits a marginally higher value than the former. One of the relevant questions within the survey is the inquiry into customers' predisposition to initiate claims following the perception of service degradation. This aspect affords insight into the distinction between precision and the subjective recognition of degradation. It emerges that a considerable number of clients, around 65% of the queried population,

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

express disinclination towards pursuing claims, citing their awareness of the formidable challenges entailed in establishing effective communication with the data center. This can give us an upper threshold for the precision because even if a customer is perceiving a malfunction, he may not be willing to initiate a claim.

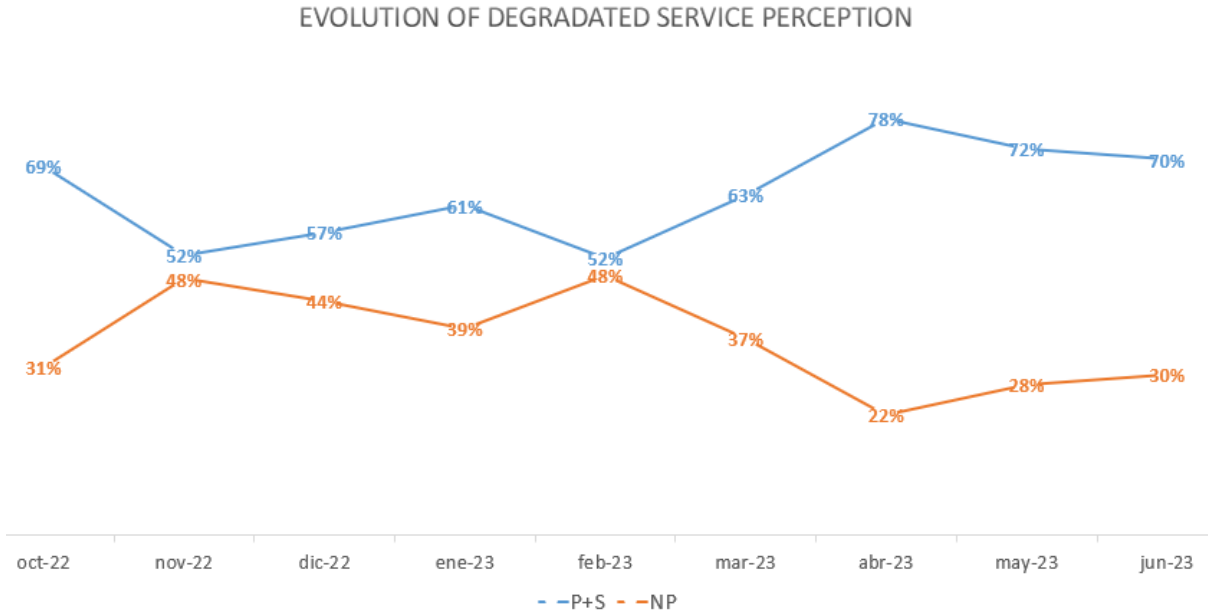


Figure 8 – Results for the perception of service degradation (P+S) versus not perceived (NP) obtained with the survey.

5.3. NPS

Another focal point of our MVP investigation was the impact of proactive customer engagement on customer experience, as quantified through the net promoter score (NPS). We designed a specialized survey to compare the NPS changes between proactive service and traditional customer service channels. This involved crafting precise questionnaires and establishing control groups for a rigorous evaluation.

Our key findings indicate that proactive customer engagement leads to a modest improvement in problem resolution rates when the case is resolved through technical service (TS) — 44% versus 41% for traditional channels. However, the resolution rates jump significantly to 80% when the issue is addressed through one of the various remote solutions evaluated during the MVP phase.

Complementing this is our second observation: the NPS scores generally rise when using proactive engagement channels. Specifically, the NPS value shifts from -9 in reactive channels to 1 in proactive ones. This improvement is even more pronounced in cases resolved remotely, which have an NPS score of 20, compared to an NPS score of -2 for cases requiring a physical visit.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Conclusions and Future Work

Our predictive model successfully generates a daily list of customers at high risk of filing claims, each accompanied by pertinent variables. This information empowers our Customer Service and Field Service teams to proactively mitigate potential issues. Through remote problem resolution and targeted technician deployment, we have made strides in boosting customer satisfaction. This is corroborated by our survey results and NPS metrics, which show a marked improvement in customer satisfaction, particularly when issues are resolved remotely.

The MVP stage has provided compelling evidence of the model's effectiveness in preemptively identifying service disruptions and elevating customer satisfaction levels. Moreover, the model's scalability is promising, especially with the automation of the ETL process. The Future Work are:

Algorithm Refinement: There is room for fine-tuning the predictive algorithm to enhance its precision. This includes the integration of additional data sources to augment the model's accuracy.

Root Cause Analysis: Preliminary observations indicate that the model's feature importance calculations could serve as a valuable tool for root cause analysis in early diagnostic procedures, especially when combined with Field Service (FS) reports.

Remote Procedures: We've initiated research into other remote interventions, such as conditional CM resets. These shows promise for improving network maintenance, but more comprehensive studies are needed for conclusive results.

Automated Customer Outreach: A promising avenue for future exploration is the automation of customer interactions through digital platforms. This could exponentially increase the volume of data collected and thereby enhance the statistical significance of our findings.

7. Acknowledgments

We would like to express our sincere gratitude to all those who have contributed to the success of this MVP: Myriam Ramirez, Marcos Avalo, Silvana Regules Silva, Sebastian Seijas, Claudio Miguel Ambrogio, Matias Miguel Parisi, Federico Gomez, Alicia Frassia, Paola A Fusari, Mariela Fiorenzo, Martin Juiz, Leonel Gonzalez Biot and Bruno Pallotta for their tireless efforts, collaboration and commitment to innovation.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

CMTS	cable modem termination system
HFC	hybrid fiber coaxial
FN	fiber node
RF	radio frequency
ISP	internet service provider
DOCSIS	data over cable service interface specification
PNM	proactive network maintenance
CM	cable modem
SNMP	simple network management protocol
MIB	management information base
CEI	customer experience index
WEI	Wifi experience index
TS	technical service
FS	field service
ETL	extract, transform and load
MVP	minimum viable product
CS	customer service
NPS	net promoter score
SHAP	Shapley additive explanations
AIOps	artificial intelligence for IT operations
XGBoost	extreme gradient boosting

Bibliography & References

[1] *CableMon: Improving the reliability of cable broadband networks via proactive network maintenance*, J. Hu et al.; Proceedings of the 17th USENIX Symposium on Networked Systems Design and Implementation (NSDI ´20).

[2] *DOCSIS CableLabs;* Best Practices and Guidelines, PNM Best Practices: HFC Networks (DOCSIS 3.0); Technical report, CM-GL-PNMP-V03-160725, 2016.

[3] *Management Information Base for Data Over Cable Service Interface Specification (DOCSIS)*; W Sawyer; Cable Modem Termination Systems for Subscriber Management. RFC 4036, 2005.

[4] Available at https://shap.readthedocs.io/en/latest/#

[5] Available at https://www.getdbt.com/

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Minimum Bit Rate Impact on Outdoor Wi-Fi Network

A Technical Paper prepared for SCTE by

Sunil Narayanan, Aniyambeth
Architect 5, Systems Architecture - WIFI Architecture
Comcast Cable Communication LLC
1800 Arch St, 15.133 Philadelphia, PA 19103
281-799-7025
SunilNarayanan_Aniyambeth@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Minimum Bit Rate Lab results ... 3
3. Minimum Bit Rate Field Trials ... 5
   - 3.1. MBR Urban Area field trial ... 6
     - 3.1.1. RSSI VS Throughput Scatter Plot ... 6
     - 3.1.2. Ekahau Plot – Signal Strength ... 7
     - 3.1.3. Ekahau Plot – SNR ... 7
   - 3.2. MBR Suburban area field trial ... 8
     - 3.2.1. RSSI VS Throughput Scatter Plot ... 8
     - 3.2.2. Ekahau Plot – Signal Strength ... 9
     - 3.2.3. Ekahau Plot – SNR ... 10
4. Conclusion ... 10

Abbreviations ... 11

Bibliography & References ... 11

List of Figures

Title **Page Number**

Figure 1 - Rate Vs Range Test Results ... 4
Figure 2 - MBR Field Test locations ... 5
Figure 3 – Urban RSSI vs Throughput Scatter Plot ... 6
Figure 4 – Urban Ekahau Plot – Signal Strength ... 7
Figure 5 – Urban Ekahau Plot - SNR ... 7
Figure 6 – Suburban RSSI vs Throughput Scatter Plot ... 9
Figure 7 – Suburban Ekahau Plot – Signal Strength ... 9
Figure 8 – Surburban Ekahau Plot – SNR ... 10

List of Tables

Title **Page Number**

Table 1 – Legacy Rates and the Basic Rates ... 4
Table 2 – MCS Rate Shift for Different MBR Values ... 5
Table 3 – MBR Urban Area Field Trial ... 6
Table 4 – MBR Suburban Area Field Trial ... 8

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The Wi-Fi operates on 2.4GHz and 5GHz which are unlicensed as per FCC (Federal Communications Commission) part 15. Th recent addition of 6GHz gave the Wi-Fi industry many more opportunities to explore. The IEEE formulated the 802.11 standards that became the guidelines for both STA (stations) and AP (Access Point) vendors to be able to design devices that operate on these allocated bands with full compatibility as per other focused bodies like WFA (Wi-Fi Alliance) and WBA (Wireless Broadband Alliance). The clients operating on these bands have basic rates defined. Here are Basic rates for 2.4GHz and 5GHz band:

- The default basic rates (2.4Ghz): 1Mbps, 2Mbps, 5.5Mbps, 11Mbps
- The default basic rates (5GHz): 6Mbps, 12Mbps, 24Mbps

Basic Rates are the most backward compatible rate by which clients and APs can transmit and receive the RF (radio frequency) signals over the air. Typically, the management frames are transmitted at the basic rate defined. The default basic rate for 5GHz is 6Mbps, and for 2.4GHz 1Mbps is the basic rate for DSSS and HR-DSSS and when operating on ERP-OFDM PHY the basic rate is 6Mbps. The management frames will always be sent out at the lowest selected rate. So, selecting the minimum bit rate will help your BSS (basic service set) perform optimally in given RF conditions. As you increase your minimum bit rate, you can see your management frames are now transmitted at higher speed, reducing the management overhead, and minimizing the airtime utilization. This results in higher throughput and reduced cell edge connection, which improves the customer experience.

The biggest challenge on an outdoor Wi-Fi network is to provide a good customer experience for a client throughout its journey. That is, associating to an access point, experiencing high quality content without degraded Wi-Fi experience, and seamlessly transitioning or disassociating from the current access point to another access point nearby or connecting to a cellular network. There are multiple Wi-Fi optimization techniques and adjusting the minimum bit rate or basic rate is one of the techniques.

2. Minimum Bit Rate Lab results

The hypotheses with which the lab test was conducted are as follows:

The change in minimum bit rate (MBR) would impact the management frames, which will help the client to start the open system authentication process at a higher SNR (signal to noise ration) value, which will result in reliable and efficient RF transmissions between the STA and AP. In the case of Wi-Fi to Wi-Fi roaming and Wi-Fi to LTE/cellular roaming, the transition is smoother with MBR if designed for a good overlap.

1. The management frame overhead is reduced which in turn will reduce the channel utilization.
2. The MAC layer boundary of the BSS is now shrunk due to increase in MBR value as MAC (media access control) layer is impacted. The PLCP (physical layer convergence procedure) PHY frame is not impacted as this is always transmitted at 1Mbps for 2.4GHz and 6Mbps for 5GHz.
3. The AP should now report higher SNR and RSSI (hertz) uplink STA statistics and similarly STA will show higher SNR and RSSI values in downlink.
4. The higher SNR should show a shift in the data rates to higher MCS (modulation coding scheme) values as throughput would increase.
5. Reduces sticky client scenarios at the cell edge.

Here is a table that shows the legacy rates and the basic rates.

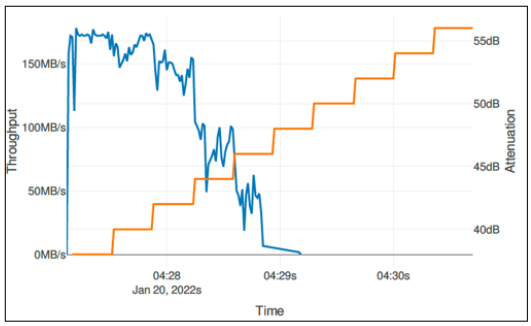
© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 – Legacy Rates and the Basic Rates

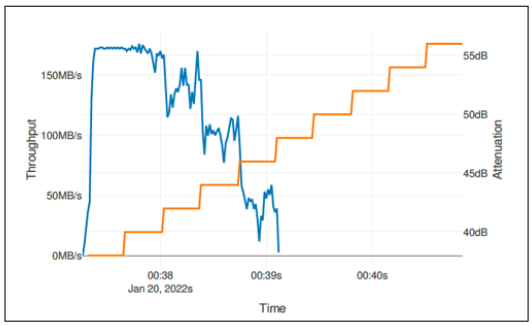
Operating Mode	Legacy Rates (Mbps)	Basic Rate (Mbps)
802.11b	1.2,5.5.11	1,2,5.5,11
802.11b, 802.11n/g	1,2,5.5,11,6,9,12,18,24,36,48,54	1,2,5.5.,11,6,12,24
802.11a, 802.11n/a, 802.11ac	6,9,12,18,24,36,48,54	6,12,24

The goal was to study the 5GHz radio behavior with varied MBR settings, hence Rate Vs Range test was conducted with a STA connecting to an AP with higher MBR value and simulated STA moving away from the AP.

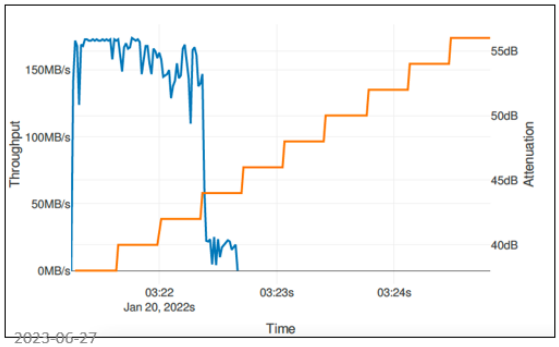
Note: 2.4GHz was turned off and the test was conducted only on 5GHz radios.



6Mbps (MBR)



12Mbps (MBR)



24Mbps (MBR)

Orange line – Attenuation in dBm

Blue line – Throughput in Mbps

Figure 1 - Rate Vs Range Test Results

The observation from the above test was that the sticky client behavior was noted highly in 6Mbps and the stickiness of STA to the AP reduces as the MBR value increases. This helps remove the lower rates at the cell edge, improving data rates, reducing retransmission over the air, and minimizing the airtime utilization.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Here is the table that shows the MCS rate shift for different MBR values:

Table 2 – MCS Rate Shift for Different MBR Values

RSSI (dBm)	SNR	6Mbps (MBR)	12Mbps (MBR)	24Mbps (MBR)
-30	57	9	9	9
-69	18	4	4	4
-71	16	4	4	4
-73	14	3	3	2
-75	12	2	2	
-77	10	1	1	
-78	9	1		

3. Minimum Bit Rate Field Trials

The test methodology included stationary and walk tests out in the field where Comcast outdoor APs are installed.

Two mobile devices along with Ekahau and Aircheck G2 and a few other applications were utilized during this test. We selected the following test locations based on client density and classified them into urban and suburban areas.

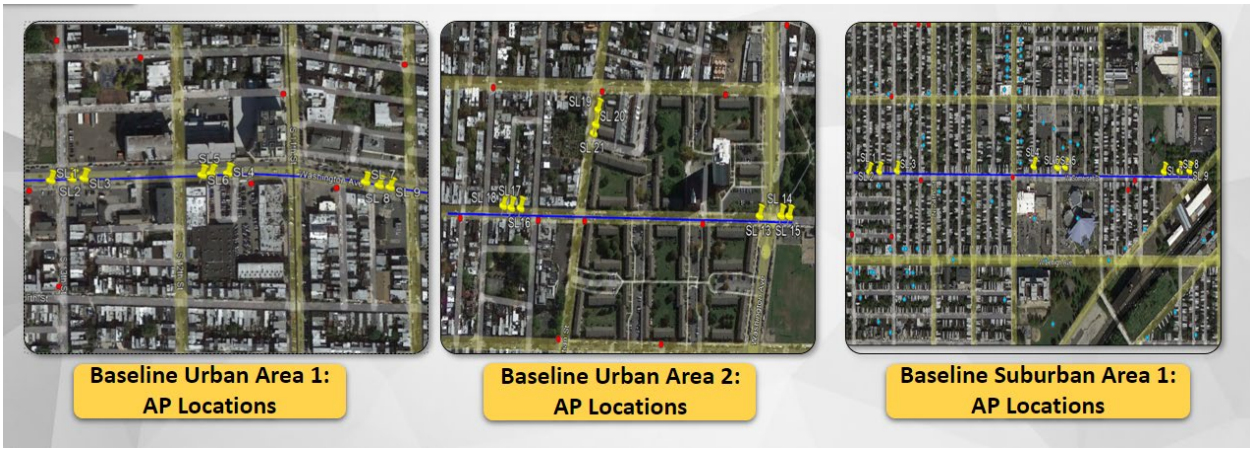


Figure 2 - MBR Field Test locations

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.1. MBR Urban Area field trial

Here is the summary of tests conducted in the urban areas. The metrics that were measured are listed below, and with a change in the MBR value you could see a gradual shift in the metrics.

Table 3 – MBR Urban Area Field Trial

Phase	MBR06		MBR12		MBR24	
	Overall	Outdoor APs	Overall	Outdoor APs	Overall	Outdoor APs
Average RSSI	-68.53dBm	-67.05dBm	-68.59dBm	-68.00dBm	-64.18dBm	-62.66dBm
Average Wi-Fi DL	42.72Mbps	56.53Mbps	23.03Mbps	26.75Mbps	41.29Mbps	51.57Mbps
Average SNR	21.31dB	21.60dB	20.34dB	25.00dB	22.06dB	25.6dB
nPerf DL	18.3Mbps	26.7Mbps	11.5Mbps	11.1Mbps	26Mbps	33.6Mbps
nPerf UL	5.5Mbps	7.9Mbps	3.6Mbps	4.0Mbps	9Mbps	12.2Mbps
YouTube Streaming%	95.4%	95.2%	92.37%	94.3%	96.73%	98.74%
Sample	11691	7392(62.7%)	11067	6755(61%)	9994	6502(65.1%)
>MCS 2 RSSI Threshold (>=-76dBm)	75.12%	78.8%	80.69%	73.21%	94.6%	95.3%
Wi-Fi Disconnected% | LTE %	3.8% | No LTE		1.7% | No LTE		2.13%| No LTE	

3.1.1. RSSI VS Throughput Scatter Plot

For MBR12 and MBR24 the scatter plot shows lower samples of connected clients below -80dBm RSSI values and have higher downlink speeds reported.

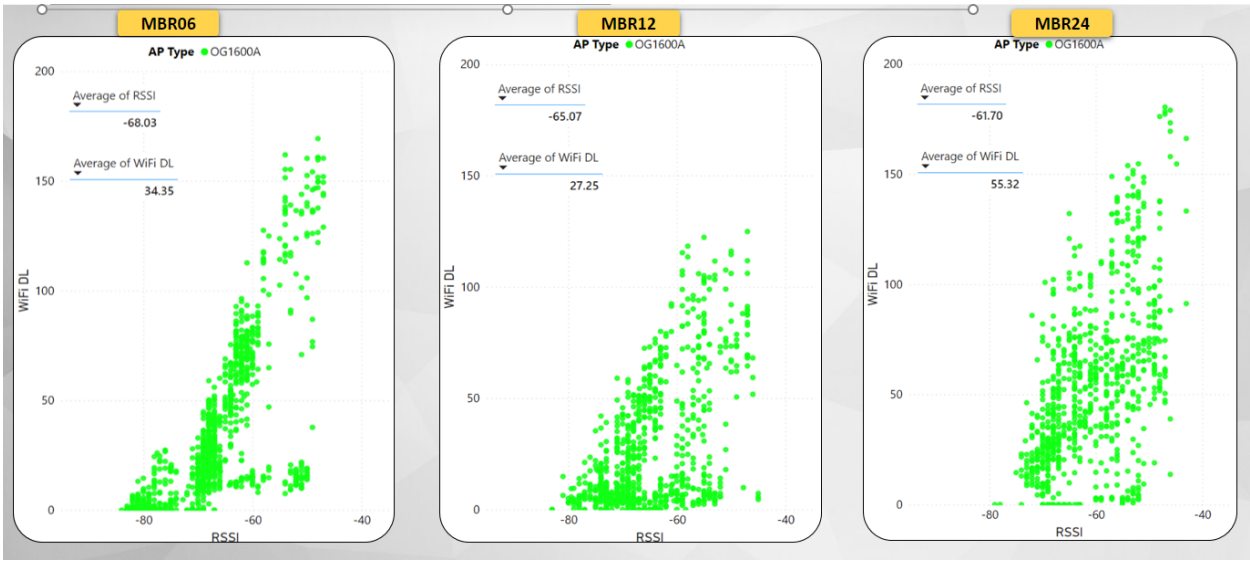


Figure 3 – Urban RSSI vs Throughput Scatter Plot

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.1.2. Ekahau Plot – Signal Strength



Figure 4 – Urban Ekahau Plot – Signal Strength

3.1.3. Ekahau Plot – SNR



Figure 5 – Urban Ekahau Plot - SNR

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Key takeaways of implementing MBR in an Urban area are as follows:

1. The number of samples with low RSSI connected to outdoor was reduced during MBR 24 testing.
2. During MBR24, a lower number of samples observed at RSSI<-75dBm.
3. MBR24 showed 77% samples above 10Mbps, compared to 57% in MBR12 testing.
4. Coverage and SNR improved during MBR24.
5. The YouTube media quality on HD 1440 improved on MBR 24 testing.

3.2. MBR Suburban area field trial

Here is the summary of tests conducted in suburban areas. When compared to urban areas, we see few variations; this is due to the RF condition in the location, the client density, and the distribution of indoor and outdoor APs.

Table 4 – MBR Suburban Area Field Trial

Phase	MBR06		MBR12		MBR24	
	Overall	OG1600 APs	Overall	OG1600 APs	Overall	OG1600 APs
Average RSSI	-67.3dBm	-67.83dBm	-68.76dBm	-70.24dBm	-68.95dBm	-71.27dBm
Average Wi-Fi DL	34.3Mbps	44.65Mbps	25.5Mbps	37.12Mbps	19.98Mbps	24.58Mbps
Average SNR	18.8dB	20dB	18dB	21dB	18.08dB	16.4dB
nPerf DL	12.4Mbps	10Mbps	16.5Mbps	17.2Mbps	13.6Mbps	11.6Mbps
nPerf UL	5.1Mbps	7.4Mbps	4.7Mbps	7.2Mbps	3.8Mbps	4.3Mbps
YouTube Streaming%	99.6%	99.08%	96.6%	95.14%	94.97%	97.22%
Sample	10054	5037(50%)	9515	3867(40.6%)	9530	3733(39.2%)
>MCS 2 RSSI Threshold (>=-76dBm)	88.4%	82%	85.27%	75%	84.17%	77.5%
Wi-Fi Disconnected% | LTE %	115(1.1%)		160(1.6%)		198(2.0%)	

3.2.1. RSSI VS Throughput Scatter Plot

From the scatter plots it appears that the MBR6 shows better throughput and samples below -80dBm, unlike the urban area.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

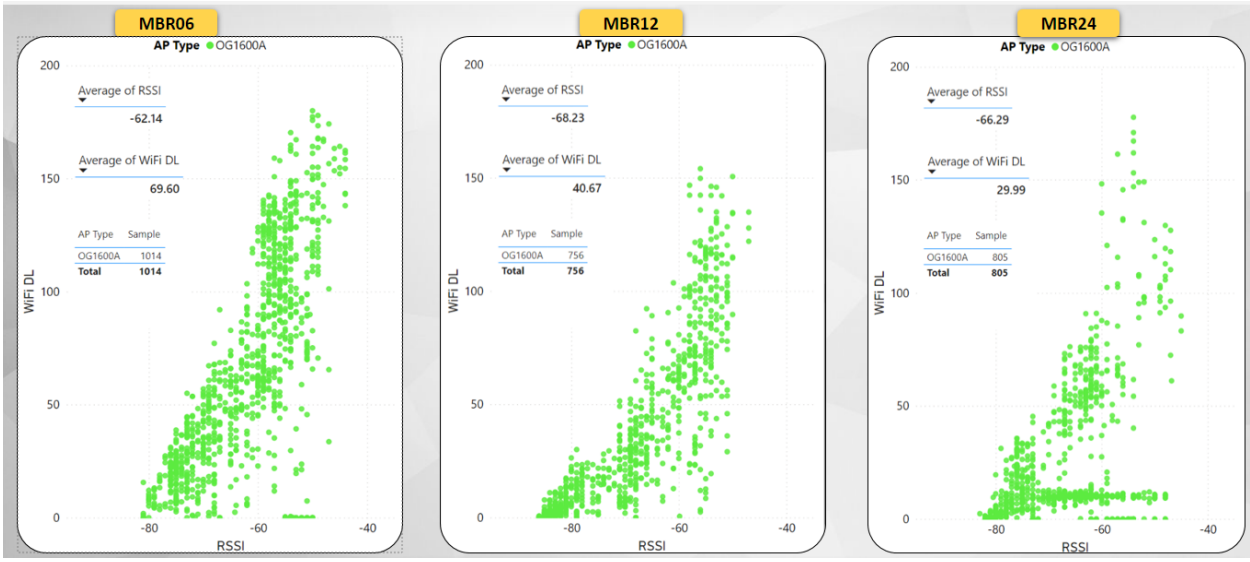


Figure 6 – Suburban RSSI vs Throughput Scatter Plot

3.2.2. Ekahau Plot – Signal Strength

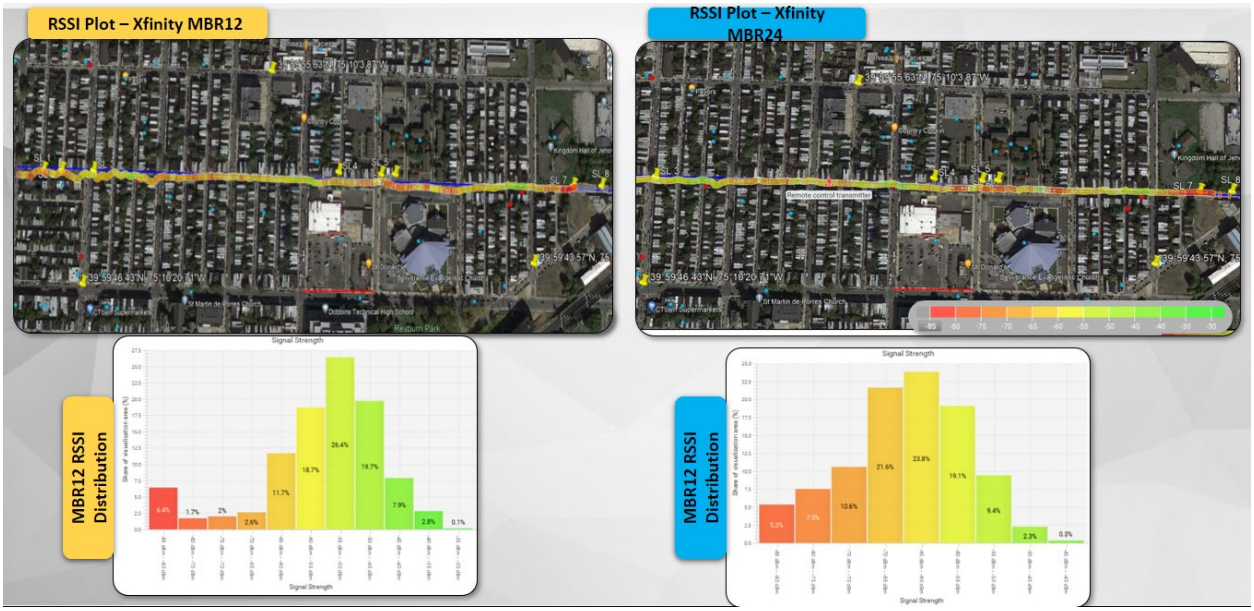


Figure 7 – Suburban Ekahau Plot – Signal Strength

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.2.3. Ekahau Plot – SNR



Figure 8 – Surburban Ekahau Plot – SNR

Key take aways from the Suburban field test are as follows:

1. The MBR6 or MBR12 values provided better results when compared to MRB24.
2. The samples on MRB24 have lower count compared to MBR12 but the difference is less than one percent.
3. Down load average was higher in 6Mbps compared with other MBR values.
4. Youtube streaming was high and at 6Mbps MBR value.
5. The DL throughput when MBR6 or MBR12 was set was higher than MBR24 configuration.
6. The SNR and RSSI distribution were better when MBR12 was set when compared to MBR24

4. Conclusion

The Outdoor Wi-Fi network has many variables that would constantly change the RF environment. Some of them can be controlled and are predictable, but most of the variants need constant observation and testing. Hence, to optimize such a network, we need multiple telemetry from the field and from Wi-Fi metrics. MBR would be one of the techniques that would help the client connection; based on your requirements and the RF environment you can have varied MBR settings to achieve your requirements. In our Urban testing the results favored 24Mbps as the MBR value, mostly because of the client density and the high AP density. In the Suburban testing, the results favored 6Mbps and 12Mbps as the MBR values, mostly because the APs are distributed and no Wi-Fi overlaps.

Additional study is needed to understand the 2.4GHz and 6GHz band and its impact with the MBR settings. The 5GHz study was conducted with bandwidth set to 40Mhz, hence for 20MHz and 80MHz design additional testing is required as the noise floor would vary with channel bounding which will impact other Wi-Fi metrics.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AP	Access Point
BSS	Basic Service Set
DL	Down Link
FCC	Federal Communications Commission
HD	High Definition
IEEE	Institute of Electrical and Electronics Engineers
MAC	Media Access control
MBR	Minimum Bit Rate or Minimum Basic rate
MCS	Modulation Coding Scheme
PHY	physical layer
PLCP	physical layer convergence procedure
RF	radio frequency
RSSI	Received signal strength indicator
SCTE	Society of Cable Telecommunications Engineers
SNR	signal to noise ratio
UL	up link
WBA	Wireless Broadband Alliance
WFA	Wi-Fi Alliance

Bibliography & References

https://www.fcc.gov/document/5-ghz-unlicensed-spectrum-uniiThe ARRL Antenna Book, 20th Ed.; American Radio Relay League

https://www.cwnp.com/cwap-study-guide404/

https://www.cwnp.com/certifications/cwdp

https://www.ekahau.com/blog/channel-planning-best-practices-for-better-wi-fi/

https://sourceforge.net/arris/og1600/home/Home/

https://documentation.meraki.com/MR/Radio_Settings/Minimum_Bitrate_Control

https://fccid.io/UIDOG1600CT/User-Manual/manual-3227914

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Motivational Metrics for Security

Driving Progress Without Burning Bridges

Matt Carothers
Senior Principal Security Architect
Cox Communications
6305 Peachtree-Dunwoody Rd, Atlanta, GA 30338
+1 (404) 269-7220
matt.carothers@cox.com

Brad Boucher
Vice President and Deputy Chief Information Security Officer
Cox Communications
6305 Peachtree-Dunwoody Rd, Atlanta, GA 30338
+1 (404) 269-8880
brad.boucher@cox.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 3
2. Metrics Literacy 3
   2.1. Nominal 3
   2.2. Ordinal 3
   2.3. Interval 4
   2.4. Ratio 4
3. Setting Goals 4
   3.1. Socialize the Goal and Explain the Why 4
   3.2. Set Goals Collaboratively 5
   3.3. Ensure Partners Understand the Goal Posts Will Move 5
   3.4. Establish Principles and Guardrails Early to Enable Scale 5
   3.5. Look for Incremental Progress 5
4. Measuring Progress Clearly and Fairly 6
   4.1. Use Data Rather than Words 6
   4.2. Give Credit Where Credit is Due 6
5. Communicating Metrics 7
   5.1. Show What You Are Not Doing 7
   5.2. Show Both Magnitude and Velocity 7
   5.3. Know Your Audience 7
      5.3.1. Company Wide Application Cyber Metrics 7
      5.3.2. Filterable Application Cyber Metrics 8
      5.3.3. Application Specific Metrics at a Glance 8
      5.3.4. Detailed Vulnerabilities for an Application 9
6. Conclusion 9

List of Figures

Title **Page Number**

Figure 1 -- Range Compression Hides Progress 6
Figure 2 -- Company Wide Overview 8
Figure 3 – High Level View 8
Figure 4 – Mid Level View 9
Figure 5 – Detail View 9

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

In the modern workplace, cybersecurity teams operate in matrixed organizations. Security teams drive and track progress without direct authority over developers, system administrators, and other employees responsible for implementing security policies. Many people focus narrowly on the measurements themselves without considering the impact they have on those being measured. This leads to ineffective systems that make employees feel scrutinized and punished.

This paper discusses techniques for designing metrics that not only measure security in a meaningful way but also address the human element. It takes you on a journey through our security program, tells you what we learned from the mistakes we made along the way, and leaves you positioned for success.

Why is it so Hard to Measure and Report on Cybersecurity?

Cybersecurity measurement, metrics, and reporting are still in their infancy. As a comparison, consider the accounting industry and associated financial reporting norms and regulations. We take for granted that we can compare the financial health of companies through standardized balance sheets, income statements and cash flow statements. Building on those financial statements, finance professionals use ratios such as earnings per share and return on invested capital as key financial metrics to quickly compare companies and assess their relative performance or overall financial health.

Entities like the Financial Accounting Standards Board (FASB) and the Securities and Exchange Commission (SEC) developed those practices over the course of nearly 100 years following the stock market crash of 1929 and the subsequent Great Depression. In comparison, control frameworks like the National Institute of Standards and Technology (NIST) Cyber Security Framework (CSF) are still relatively early in their lifecycles, so we must give ourselves (and the teams we are trying to measure) a little grace and remember that we are still in the early stages of this journey.

2. Metrics Literacy

Even once we determine what needs to be measured, we often do so in a way that makes no sense and yields bad results. Not all mathematical operations apply to all measurements. Using operations like multiplication and division on scales that do not support them leads to bad decision making. To understand why, we need to take a moment to discuss the four types of measurements.

2.1. Nominal

Nominal measurements express set membership. They can be taken by asking simple yes-or-no questions. For example, by asking “Is EDR installed on this server?” for every server in our environment, we create a count of servers belonging to the set with EDR.

2.2. Ordinal

An ordinal measurement tells us one thing is greater than another, but not by how much. For example, let’s consider movie ratings. We can say a five-star movie is better than a one-star movie, but we cannot say by how much. Is a five-star movie five times better than a one-star movie? Will we get the same enjoyment from watching five one-star movies as we would from one five-star? Likely not.

In the cybersecurity world, the Common Vulnerability Scoring System (CVSS) serves as a great example of an ordinal scale. A CVSS score of 10 is more severe than a score of 5, but we cannot say it is exactly

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

twice as bad. Nor could we ever say with a straight face that 10 vulnerabilities with a CVSS score of 1 are just as bad as a single vulnerability with a score of 10. Beware of scoring systems that add, subtract, multiply, or divide CVSS scores!

2.3. Interval

Interval measurements allow us to not only compare two values, but to add and subtract them. The Celsius temperature scale provides a good example. The difference between 10 degrees and 20 degrees is the same as the difference between 90 degrees and 100 degrees. Take care, however, because multiplication does not work on interval scales. 20 degrees is not twice as hot as 10 degrees on the Celsius scale because on that scale the zero point does not mean "no heat." Rather, it is arbitrarily set at the freezing point of water.

2.4. Ratio

Ratio measurements allow us to use not only addition and subtraction, but also multiplication and division. Take mass, for example. 20 kilograms is exactly twice as much as 10 kilograms. This is because zero kilograms means "no mass." The same is true for the Kelvin temperature scale, where zero means "no heat."

3. Setting Goals

Now that we have a good grasp on the math, let us talk about the people. It is a rare information security organization with the authority to simply dictate what must be done. Instead, cybersecurity leaders must build bridges to other departments within the enterprise and create performance metrics that motivate success.

3.1. Socialize the Goal and Explain the Why

The best way to meet expectations is to set expectations. Cybersecurity is no different. It is imperative to socialize the key objectives of a cybersecurity program to its stakeholders and partners. Cybersecurity frameworks contain hundreds of cybersecurity controls, but they rarely explain why they are needed. Even if they contain risk statements, the risks are typically not expressed in business-friendly terms and offer little context.

When establishing the cybersecurity controls needed in a technology environment, take the extra steps to explain why they are important and how they improve the security posture and health of your organization. In these conversations, create a real dialogue with technology teams. Allow them to ask provocative questions, and provide thoughtful answers, even if it means taking on the occasional homework assignment to clarify the position. Working out challenges to the merits of a control at the beginning of the journey is far easier than at the end. "I never agreed we needed that control anyway" must be avoided, particularly when regular measurements and executive reporting begin.

In the same way that explaining the value of security controls is important, cyber security teams must also take the time to understand the "why" from stakeholders. Other teams may have completely valid reasons for pushing back on security controls, and a one-sided conversation will not convince them. Take the time to understand and appreciate the perspectives of others and winning them over will be much easier.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.2. Set Goals Collaboratively

Collaborative target setting helps boundary partners appreciate that cybersecurity teams understand that security is just one of many business imperatives. A team that helps set its own targets, especially when they feel that competing priorities were heard and thoughtfully considered, will be much more amenable to meeting those targets. For example, the rollout of a new security tool might interrupt the launch of a critical new service. Identifying this conflict early is always better than in the middle of a struggling implementation.

This can certainly be a complex conversation, but the dialogue created in these discussions typically helps both sides learn and build trust. At the end of the day, security organizations want to protect and enable this business, not disable it. Finding the right pace of change and remediation in your organization and culture can help prevent heated battles in the trenches. If the pace of cybersecurity improvement is not moving fast enough, escalate up not down.

3.3. Ensure Partners Understand the Goal Posts Will Move

Security is a continuous process and not a one-time activity. Your partners must be aware that the targets can and will change. If they are not aware from the outset, they will feel increasingly frustrated as it seems their achievements are being snatched away.

For example, say that your security team identifies 1,000 unknown devices on the network and negotiates with the server administration team to document them in an asset management database. The team completes 900 of them, and your measurement shows 90% completion. The next day, your team gains access to another internal tool and discovers an additional 1,000 unknown assets. Suddenly, the denominator is 2,000 instead of 1,000, and the team is now only at 45%. These types of issues will occur in any sufficiently complex organization, and partners must be prepared to face them.

In these cases, you may even add new metrics or provide additional context on readouts to clearly explain that even though the goal posts moved, significant progress was achieved. Taking the above example further, adding a metric that shows number of assets mapped as well as the number of newly discovered assets during a time period acknowledges both great progress being made and a shift in the end target. Going the extra mile to ensure that positive progress is celebrated while the goal posts are moving is important to keep teams motivated.

3.4. Establish Principles and Guardrails Early to Enable Scale

In alignment conversations, work to establish principles and guardrails that can be referenced later when conflicts arise. An example of a principle or guardrail is, “Our internet facing applications must be free of critical and high-risk coding vulnerabilities.” Note that the statement does not address specific vulnerabilities. Instead, it addresses wide classes of vulnerabilities. It is a simpler concept that more boundary partners can understand and adopt, and it prevents security teams from being bogged down arguing about each specific issue individually down the road.

3.5. Look for Incremental Progress

When establishing these guardrails and principles, one must often take incremental wins instead of immediately swinging for the fence. For example, an organization may not even be willing to entertain estimating the cost of what it would take to patch all vulnerabilities throughout the environment every 30 days. But most organizations would find it immediately more palatable to mandate patching only of critical and high-risk vulnerabilities for internet facing systems every 30 days. In this case, patching

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

internet facing systems would be a defensible first step in a risk-based approach to improving the organizations cybersecurity posture, while also allowing operations teams to learn what it takes to deliver.

4. Measuring Progress Clearly and Fairly

4.1. Use Data Rather than Words

Qualitative (words) measurements can be subjective and imprecise. False communication can occur when two people believe they agree on something but actually hold different ideas of what words mean. Qualitative measurements also suffer from the issue of range compression. If we break our measurements down into buckets like "low", "medium", and "high" or visualize performance with buckets like "green", "yellow", and "red" we lose the ability to show any progress within a bucket.

For example, let us imagine we have a population of 10,000 servers in a data center, and our goal is to motivate the owners to enter information into an asset database. We create a dashboard with a traffic light. The light is red below 25% compliance, yellow between 25% and 75%, and green above 75%. Due to range compression, the teams performing the work will see no change in the status from the time they complete server 2,501 and server 7,500. Furthermore, as soon as they enter server 7,501, the light turns green, and they have no further motivation to continue.

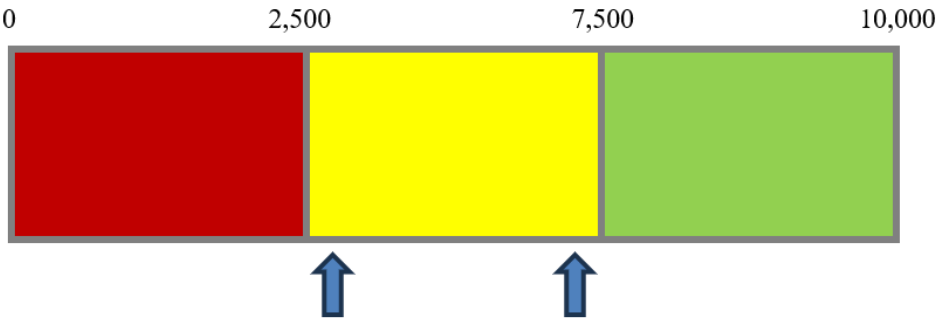


Figure 1 -- Range Compression Hides Progress

4.2. Give Credit Where Credit is Due

Ensure you acknowledge work as it is done. One metric the authors designed internally involved patching compliance. They measured at the application level rather than the device level, so for any given application, multiple devices might need to be addressed. They only counted an application as complete when all associated devices were complete. This created a situation where an administration team could spend weeks updating dozens of servers, but if even one server could not be patched for any reason, the entire application registered as a failure. Understandably, some teams decided that if they could not patch every server for a given application, they may as well not patch any of them. Our Information Technology (IT) partners dubbed it the Brad Punitive Scoring System (BPSS). The authors took the hint and retired the BPSS in favor of measuring devices rather than applications.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Communicating Metrics

Simply measuring progress does not suffice if one cannot communicate the results of the measurements in a way that drives change. Here are some things the authors have learned about effective communication in their security journey.

5.1. Show What You Are Not Doing

Another way to ensure our partners understand security is never done is to show them measurements for work that has not started yet or for which we have yet to take any measurements. It sets the stage for later engagements without blind siding them.

5.2. Show Both Magnitude and Velocity

Leaders do not only want to see today's score. They want to know how it compares to yesterday's. They want to understand if the number is going up or down and by how much. As such, dashboards should include not just numbers but the amount of change since the last presentation.

5.3. Know Your Audience

When we make metrics, we typically have at least three views. There is of course a detailed view that provides insight into an exact metric and what is driving the positive or negative score of the metric. We then also create aggregate views for mid-level managers and senior leadership. By creating useful aggregate views upfront, we provide the right level of detail to the right level of leader at the right time. If we are escalating poorly performing metrics, we have polished dashboards or summaries ready to go for leaders to enable them to ask the right questions of their teams and help drive decisions and efforts to get us closer to our desired goals.

The following examples show a hypothetical vulnerability management dashboard with multiple views intended for different audiences. Note that is conveys both measurements and areas where measurements are lacking, and it expresses not only the current values but the rate of change.

5.3.1. Company Wide Application Cyber Metrics

This view is intended for senior technology leadership.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

All Application Control Summary

			Visibility				Identity and Access Management						Vuln Mgmt	
Application Categories	# of apps	# of LBs, Servers, Containers	Web Application Firewall		EDR/Logs		Multi-Factor Auth		Priv Access Mgmt		Identity Provider Federation		Patching	
Public Facing	85	3,257	95%	-5	75%	-4	94%	-	86%	-2	59%	+2	74%	+16
Core Services	32	1,267	n/a		80%	+6	88%	+6	93%	+17	n/m		78%	+3
Business Critical	75	6,431	n/a		88%	+5	n/a		88%	+15	n/m		55%	-12
Business Tier 2	250	8,560	n/a		81%	+2	n/a		82%	+6	rba		rba	
Business Tier 3	425	12,654	n/a		83%	+4	n/a		82%	+6	rba		rba	

n/a Control not applicable
n/m Control not yet measured
rba Deprioritized, risk based approach

Figure 2 -- Company Wide Overview

5.3.2. *Filterable Application Cyber Metrics*

This view in intended for various levels of mid-level management. It is the same view as the summary, but filterable by business unit and/or leader.

All Application Control Summary - Filterable by Org and/or Leader

Filter by Business Unit/Organization: Tech Organization 1
Filter by Leader: Leader 1

			Visibility				Identity and Access Management						Vuln Mgmt	
Application Categories	# of apps	# of LBs, Servers, Containers	Web Application Firewall		EDR/Logs		Multi-Factor Auth		Priv Access Mgmt		Identity Provider Federation		Patching	
Public Facing	14	536	100%	-	98%	+6	100%	-	86%	-2	59%	+10	79%	-15
Business Critical	12	1,029	n/a		95%	+5	n/a		89%	+4	n/m		65%	+10
Business Tier 2	74	2,534	n/a		94%	+2	n/a		87%	+3	rba		rba	
Business Tier 3	119	3,543	n/a		83%	+4	n/a		82%	+1	rba		rba	

n/a Control not applicable
n/m Control not yet measured
rba Deprioritized, risk based approach

Figure 3 – High Level View

5.3.3. *Application Specific Metrics at a Glance*

This view moves from categorizing applications into buckets to providing application specific scores and is intended for application owners or managers drilling down to identify problem areas.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Public Facing by Application Summary Drill Down

Filter by Business Unit/Organization: Tech Organization 1 | Filter by Leader: Leader 1 | Filter by Application Category: Public Facing

Application Name	# of LBs, Servers, Containers	Web Application Firewall	EDR/Logs	Multi-Factor Auth	Priv Access Mgmt	Identity Provider Federation	Patching
Application 1	75	100%	98%	100%	89%	100%	98%
Application 2	164	100%	95%	100%	100%	0%	98%
Application 3	45	100%	100%	100%	97%	0%	65%
Application 4	32	100%	100%	100%	98%	100%	45%
Application 5	15	100%	89%	100%	79%	100%	38%

Figure 4 – Mid Level View

5.3.4. Detailed Vulnerabilities for an Application

This view provides the specific vulnerabilities that need to be addressed for an application. This view is meant for operations personnel or developers to act upon.

Aged Vulnerabilities by Host

Compliance	Age	Time Left/Past Due	DNS / Container Image	Environment	Cluster Name	CVE ID	Sev	Title
Non-compliant	181+	1 year past due	app.example.com:4443/exp/store	Production	eks-exp-dmz-p-1	CVE-2022-29458	4	ncurses 6.3 before patch 20220416 has an
Non-compliant	181+	1 year past due	app.example.com:4443/exp/store	Production	eks-exp-dmz-p-1	CVE-2022-29458	4	ncurses 6.3 before patch 20220416 has an
Non-compliant	181+	7 months past due	app.example.com:4443/exp/coam	Production	eks-exp-dmz-p-1	CVE-2022-42916	4	In curl before 7.86.0, the HSTS check could
Non-compliant	181+	7 months past due	app.example.com:4443/exp/local	Production	eks-exp-dmz-p-1	CVE-2022-42916	4	In curl before 7.86.0, the HSTS check could
Non-compliant	181+	7 months past due	app.example.com:4443/exp/store	Production	eks-exp-dmz-p-1	CVE-2022-42916	4	In curl before 7.86.0, the HSTS check could
Non-compliant	181+	7 months past due	app.example.com:4443/exp/store	Production	eks-exp-dmz-p-1	CVE-2022-42916	4	In curl before 7.86.0, the HSTS check could
Non-compliant	181+	7 months past due	app.example.com:4443/exp/wirel	Production	eks-exp-dmz-p-1	CVE-2022-42916	4	In curl before 7.86.0, the HSTS check could
Approaching Non-compliant	31-59	2 days left	app.example.com:4443/exp/store	Production	eks-exp-dmz-p-1	CVE-2023-1999	4	There exists a use after free/double free in
Approaching Non-compliant	31-59	2 days left	app.example.com:4443/exp/store	Production	eks-exp-dmz-p-1	CVE-2023-1999	4	There exists a use after free/double free in
Approaching Non-compliant	31-59	10 days left	app.example.com:4443/exp/store	Production	eks-exp-dmz-p-1	CVE-2023-3138	4	A vulnerability was found in libX11. The sec
Approaching Non-compliant	31-59	10 days left	app.example.com:4443/exp/store	Production	eks-exp-dmz-p-1	CVE-2023-3138	4	A vulnerability was found in libX11. The sec
Approaching Non-compliant	31-59	10 days left	app.example.com:4443/exp/wirel	Production	eks-exp-dmz-p-1	CVE-2023-3138	4	A vulnerability was found in libX11. The sec
Compliant	31-59	25 days left	app.example.com:4443/exp/store	Production	eks-exp-dmz-p-1	CVE-2023-35945	4	Envoy is a cloud-native high-performance e
Compliant	31-59	25 days left	app.example.com:4443/exp/store	Production	eks-exp-dmz-p-1	CVE-2023-35945	4	Envoy is a cloud-native high-performance e
Compliant	31-59	25 days left	app.example.com:4443/exp/wirel	Production	eks-exp-dmz-p-1	CVE-2023-35945	4	Envoy is a cloud-native high-performance e

Figure 5 – Detail View

6. Conclusion

Building a cybersecurity metrics program is a journey. The highest performing security teams establish deep relationships with both business and technology partners based on trust, empathy, and a common commitment to improving cybersecurity in the organization.

When designing a metrics program, it is important to create metrics collaboratively using sound math and ensure those metrics are evaluated fairly. This builds confidence in the metrics and helps stakeholders focus on what makes the metric high or low instead of challenging the metric itself. Metrics should also be tailored for your audience and provide context on both magnitude and velocity. Having the right level of detail for the right audience and showing how metrics are progressing (or not progressing) over time is key to driving productive conversations.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Finally, the most important part of your cybersecurity metrics program is alignment. Alignment on the goals of the program. Alignment that those goals will change over time because technology and attackers are always changing. Alignment that cybersecurity isn't just achieved by cybersecurity and instead achieved by cybersecurity, technology and the business working together. And lastly, alignment that the journey is going take time. It is extremely important to recognize and celebrate the progress that is made incrementally. Each day, each week, each month, you will be improving your security posture and the resilience of your organization.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Multi Layered Unsupervised Machine Learning for Detection of Real Time Network Service Impairments

An Operational Practice prepared for SCTE by

Eric Frishman
Senior Principal Engineer
Comcast
1800 Bishops Gate, Mount Laurel, NJ 08054
215-510-6347
Eric_frishman@cable.comcast.com

Devangna Kaushish
E4 Product Development Engineering
Comcast
1800 Bishops Gate, Mount Laurel, NJ 08054
901-605-7629
Devangna_kaushish@cable.comcast.com

Russell Harlin
Senior Engineering Manager
Comcast
1050 Enterprise Way, #500, Sunnyvale, CA 94089
650-766-7256
Russell_harlin@comcast.com

Aditya Vallabhajosyula
E4 Software Development and Engineering
Comcast
1050 Enterprise Way, #500, Sunnyvale, CA 94089
aditya_vallabhajosyula@comcast.com

Eswaramoorthy Subramaniam
Principal Engineer Product Development
Comcast
11951 Freedom Dr, Ste 900, Reston, VA
Eswaramoorthy_Subramaniam@cable.comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Background ... 3
3. Exploratory Data Analysis ... 4
4. ML Modeling: Primary and Secondary Variables ... 7
5. 1st ML Layer: Outlier Distribution Analysis ... 8
6. 2nd ML Layer: Grouped Outlier Distribution Analysis ... 9
7. 3rd ML Layer: Unique Element Distribution Analysis ... 9
8. Large Scale Implementation ... 11
9. Conclusion ... 11

Abbreviations ... 11

List of Figures

Title **Page Number**

Figure 1 – ML Workflow ... 3
Figure 2 – High Level ML Structure ... 4
Figure 3 – Traffic Samples ... 5
Figure 4 – Traffic and Density Views ... 6
Figure 5 – Baseline Modeling ... 7
Figure 6 – ML Layer 1 Outlier View ... 8
Figure 7 – ML Layer 2 Grouped Outliers ... 9
Figure 8 – Outlier Detection Examples ... 10
Figure 9 – Large Scale Design ... 11

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Finding and categorizing network impairments is a significant challenge. Using Machine Learning (ML) to study network traffic patterns, we can quickly discover places where behaviors depart from statistical norms and highlight them. As the divergences may stem from common underlying characteristics themselves, ML modeling can further separate and classify situations to speed in diagnoses and resolution, leading to improved customer experience and satisfaction as well as more efficient use of staff resources.

In this paper and presentation, we will walk through project and process flow, and ML considerations.

2. Background

Unsupervised ML techniques offer great methodologies of discovery to distinguish underlying behavior through data organization, unlabeled classification, density analysis, and outlier detection. Though these methods do not specifically identify or label outcomes, these techniques utilize sophisticated hunt and search capabilities to first highlight areas of interest and thereafter parse and filter them into more organized and simplified data structures. From these updated structures, user feedback through semi-supervised ML learning additionally refines the outcomes into appropriate anomaly labeled classifications. Thus, we begin with unsupervised techniques and through our workflow we finish with supervised and labeled solutions as outlined in Figure 1.

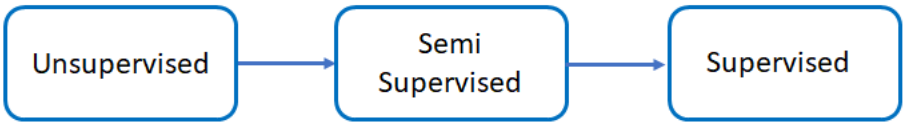


Figure 1 – ML Workflow

With the above broadly structured approach to both organize and thereafter classify the information for outlier detection, the generalized solution is highly flexible and adaptable to a wide variety of situations and data sets. In addition, an unsupervised ML leading stage can simplify extremely large data sets of hundreds of millions of data points into more reasonably organized and explorable subsets.

In the following diagram, Figure 2, the flow from left to right highlights the fundamental functional components and ML structures we used to detect outlier traffic behavior within a Cable Model Termination System (CMTS). The benefit of this approach is layering suitable ML techniques to refine, adapt, and classify broad and small-scale outliers.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

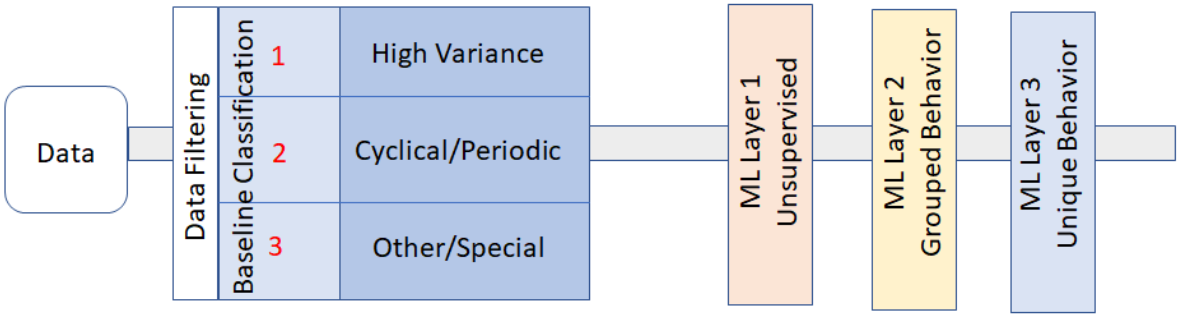


Figure 2 – High Level ML Structure

3. Exploratory Data Analysis

Data exploration is a key critical first step in developing robust solutions. The understanding and consideration of signal data changes, variations, and patterns with respect to time and frequency, as well as how outliers manifest themselves, permits a stronger selection of appropriate ML tools, techniques, and solutions. By choosing to create baselines against the lowest granular element in the network, the cable modem, higher level aggregates and analysis become possible, corresponding to ML Layer 2 within Figure 2. We used anonymized cable modem traffic data to better understand usage patterns and impacting outliers. Data usage, with strong outlier detection methods, can be turned into high value data streams when other informational sources and data are integrated Quality of Service (QOS), latency, bandwidth utilization, software releases, hardware versioning, etc.).

The data set consists of millions of elements which are evaluated for every time interval. This results inhundreds of millions to a billion analyses and status determinations in a day. With large data output volumes, a critical consideration is reducing both the False Positive and False Negative rates in alerting/alarming. This is a primary concern since appropriately focusing precious operational resources reduces mean time to resolve (MTTR). False positives, even in small numbers, may become prohibitive unless efforts are well aligned. Therefore, any aspect which improves signal-to-noise ratios through data cleaning and robust model development is critical. For simple signals, ML modeling does not pose a significant challenge. For very complex signals, variable design and simplifications though classification assist in running models effectively at scale.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 3 shows a few examples of cable modem traffic variations:

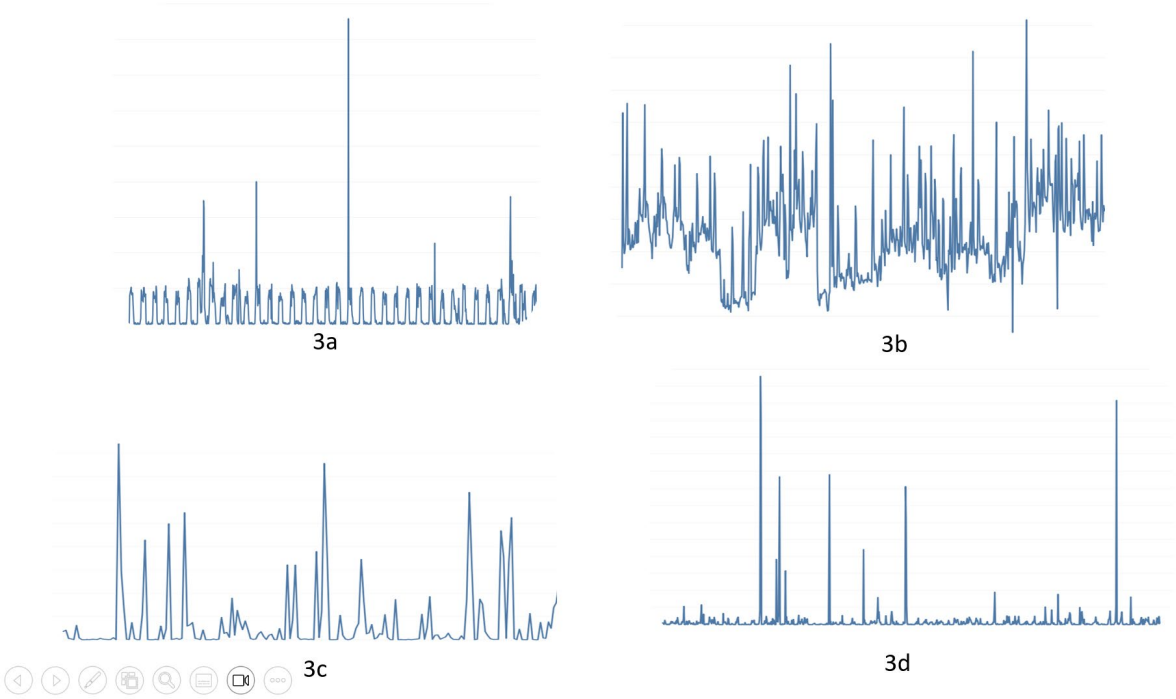


Figure 3 – Traffic Samples

The above examples have a variety of individual features. Note that all of the behavior within these graphs is expected without any outliers or points of interest. Within Figures 3a and 3b, there are components of cyclical traffic behavior as well as a variety of signal changes and variations. Figure 3a has high periodicity with some higher usage spikes. Figure 3b shows persistent higher usage with large fluctuations. Day of the week, as well as hour of day usage patterns, may be evident in some examples, however, usage itself is rarely a constant and is subject to instantaneous change.

Figures 3c and 3d emphasize very low traffic usage preceding much heavier volumes with significant amplitude differences. Traffic usage is often quite bursty with large volumes in short periods of time. Traffic peaks may be periodic, though in many cases sporadic with large magnitude differences and long and short time pauses in between peak usage.

In many cases historical signal usage is only loosely correlated with potential current expectation forecasts. Therefore, when modeling individual elements, to account for these potentially significant changes and shifts, individual element modeling must incorporate appropriate biases when possible. Adaptable models which characterize signal components with respect to intensity, symmetry, time, variance, and signal randomness may create robust signal expectations. Please note that long-time interludes of low usage may not be representative of significant change, and simply reflective of normal usage.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

As noted, there are a variety of attributable signal differences at the cable modem level. Due to high variance elements having much lower predictability, we decided to model high variance elements as their own class separately. Categorically defining baseline characteristics enables specific signal modeling refinements for complex behavior patterns.

Figure 4 depicts snapshot views of signal characteristics based on signal intensity and time distributions.



Figure 4 – Traffic and Density Views

Through variable design, relative descriptive features are developed to generalize, differentiate, and classify the raw signal data into categories. In the four examples in Figure 4, there are distribution density bar analyses included on the horizontal time axis and the vertical traffic volume axis. Darker colors indicate higher data density presence as well as positioning within the respective axis. We quantify and qualify important signal aspects using the density qualities and signal position references to create assignable model classifications. In comparing Figures 4a and 4b against Figures 4c and 4d, the time signal density distributions along the horizontal time axis appear very different with the highly variant signals showing great concentration in very narrow bands. Traffic densities on the vertical traffic axis may often be largely unpredictable but are recorded characteristics. While sometimes highly variant signals show periodicity, this was not shown to be reliably predictive of outlier detection in many cases. In contrast, signals of lower variability had greater predictability in periodicity and usage but were also subject to data shifts simply because of usage changes. Through separation, each class may be distinctly modeled.

The vertical traffic density mappings show signal concentrations and divergences, with high variability. Using both traffic intensity and time density mappings, we create model classifications with respect to groups as well as individual element models. The time density characteristics are a key mapping for differentiating highly variant signals and the combination of both the horizontal and vertical densities contribute to an overall understanding of cyclical behavior and persistence.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The first stage within the Figure 2 solution is focused on signal classification and reliable baseline model development using historical patterns and density signal mappings. Signals which are highly variant may be so in both intensity and/or time. While highly variant signals might be periodic, a noteworthy element is that the length of time between usage may often be sporadically distributed with long periods of time between large amplitude peaks. Because of this, low periods of utilization may have little predictive value and modeling these areas may hinder model convergence. Instead, low traffic periods may be compressed to quantify the peak amplitudes better.

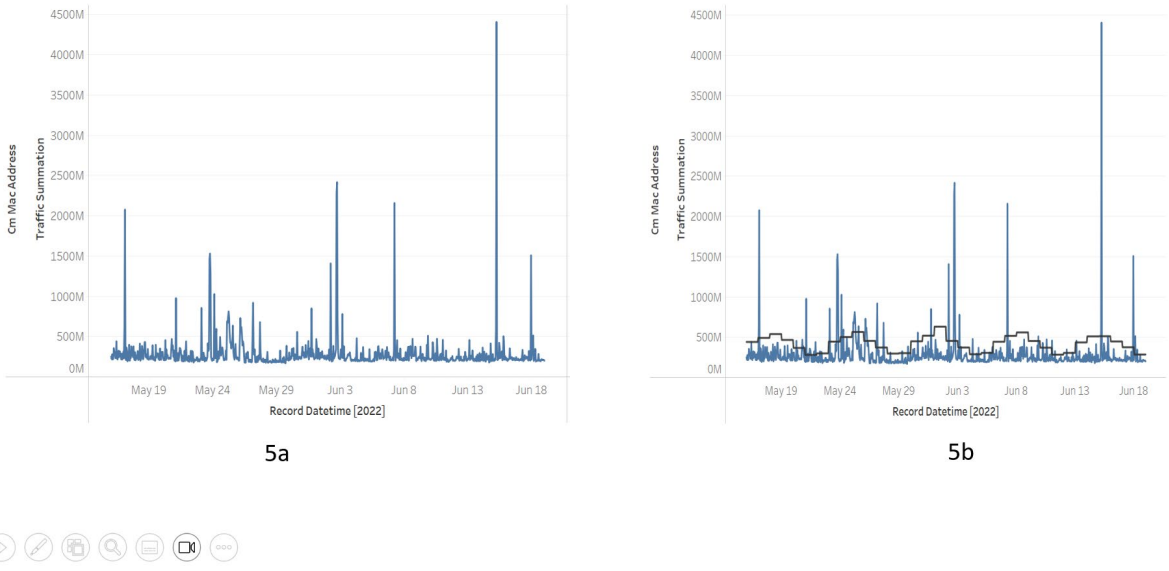


Figure 5 – Baseline Modeling

The graphics in Figure 5 depict the raw usage signal in Figure 5a as well as the interpolated baseline in black, Figure 5b. The baseline includes aspects of time-based periodicity including day of the week and hourly components. Note that there are instances of higher intensity beyond the baseline model. These aspects and their interpretations are managed within Layer 1 of the ML model for appropriate outlier detection.

4. ML Modeling: Primary and Secondary Variables

After the creation of individual baselines there will be instances where signals may exceed modeled expectations as shown in Figure 5. These areas are not necessarily indicative of outliers but high variability and lower predictability. Our objective is to characterize both signal characteristics and signal changes. We do this by creating primary and secondary evaluation variables analyzed through sequential ML layers. This progression moves from unsupervised/unknown descriptive features to classified subsets. The primary variables relate to signal changes with respect to the developed modeled baselines including changes of intensity, time, density, and position. These variables relate to characteristics of unsupervised modeling in ML Layer 1 in Figure 2. The secondary variables consist of grouped common classification sets evaluated by ML Layer 2 in Figure 2. These secondary characteristics include software type, hardware type, physical network layers [CMTS, service group, virtual group, etc.], Ip Version, or any other grouped classification of interest.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Variation metrics are derived to reflect how much outlier traffic signals deviate from the expectations using ML Layer 1, for each baseline. There are both parametric and non-parametric methods for these derivations, and the metrics create and define the appropriate boundaries in Figure 2, ML Layer 1, as described in Section 5.

5. 1st ML Layer: Outlier Distribution Analysis

Using the modeled baselines for each traffic subgroup, we then created intensity boundaries to define appropriate outliers. A baseline sets the key characteristics of expectation for outcome modeling. However, outlier sensitivity may be specific to project or end user perspectives. Because element traffic behavior may exceed developed baseline models simply because of instantaneous changes in pattern usage, ML's non-linear solution space helps to clarify outcomes. With the integration of density-based ML methods to define intensity levels, and through semi-supervised feedback through sampling, zones or areas of true outlier interest may be quickly derived, explored, and refined, at a cable modem level.

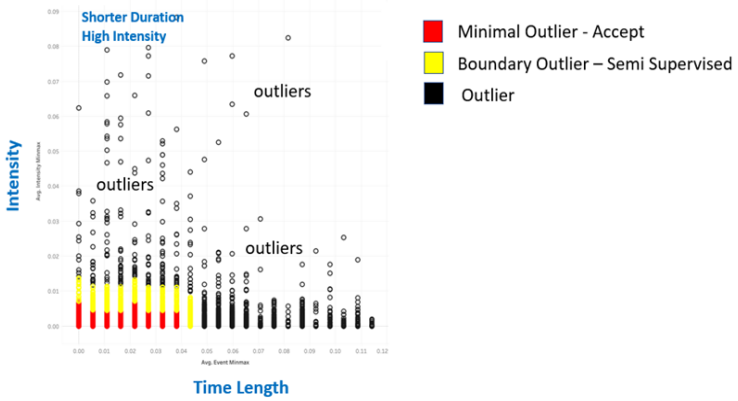


Figure 6 – ML Layer 1 Outlier View

Figure 6 has simplified the primary variables, which may include relative signal position, velocity and acceleration components, or other characteristics, into a two-dimension perspective showing the intensity of variation (y-axis) and time dimensions (x-axis). Please note that the axes have been normalized. Since each sample point falls outside of baseline expectations, all are technically outliers. However, because of the variability of signals, and the complexity of modeling, the simple presence and position of a sample outside of a modeled boundary may not be sufficient in intensity to classify it as an outlier. The great benefit of unsupervised modeling is that characteristics of these key variables and density changes throughout the distribution organize and identify specific areas of interest. With a statistical sampling of the distribution at various intensity points based on time, intensity, or both variables, semi-supervised client feedback may be propagated into the distribution so that millions of data points are classified into an appropriate data stratum, using only a few representative points.

In Figure 6, the red points reside closely to the baseline boundaries with respect to intensity and/or time. For outliers that occur in shorter time durations, oftentimes greater magnitudes of intensity might be necessary for alerting, as opposed to lower intensity outliers which may only be of interest if they exist for longer durations. Therefore, areas that are highlighted in red may not be of interest. Yellow demarked boundaries might be of sufficient intensity and are dependent on semi-supervised client

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

feedback. The black marked areas are reflective of outliers which should be alarmed. The boundary definitions between red, yellow, and black areas are quickly discernible through sampled feedback.

6. 2nd ML Layer: Grouped Outlier Distribution Analysis

The first ML layer defined potential issues at a cable modem level against baselined attributes and key performance indicators (KPI), in this case, intensity and time. ML layer 2 in Figure 2 organizes and evaluates additional shared descriptive features among device groups. For these secondary descriptive variables, grouped baselines and correlative analysis may provide insight into common behavior shifts. These characteristics may include groupings of different network layers [CMTS, Service Group, Virtual Group, etc.], software releases, hardware associations or other similar collections. The first ML layer defines individual performance, while the second layer organizes outlier groupings for correlated behavior shifts.

For network layer associations, the highest level of incident impact will most often characterize the affected network level. An impacted lower-level service group usually will not result in a higher-level CMTS aggregate data alarm. However, CMTS level issues may propagate across many lower network levels; this differs from descriptive grouped collections of software or hardware that would show effects within their specific classifications.



Figure 7 – ML Layer 2 Grouped Outliers

Figure 7 shows historical outlier counts from ML Layer 1 organized and modeled into a network CMTS ML Layer 2. The dark red highlights show baseline outlier events quickly rising and being detected. Large scale element counts are easily monitored and tracked using ML strategies.

7. 3rd ML Layer: Unique Element Distribution Analysis

The first layer of ML provides single-element interpretation of outlier or normal expectation status for each data point. Because the model baselines provide a firm foundation, the characterization of group behavior is then easily interpreted by ML layer 2 for shifts in anomaly counts and type qualifications. With strong individual model baselines, outlier evaluations at the cable model level, the lowest common denominator, become possible despite this resolution's high variance and variability.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The statistics and evaluations from the proposed multi layered model indicated a 0.05% outlier issue rate at ML Layer 2 during a normal operating cycle. At this issue rate, outlier presence is sufficiently low such that individual cable modem investigations become operationally possible through an additional layer of ML modeling and filtering. Figure 2 ML Layer 3 creates individual element density models and expectations for alerting confirmation at a cable modem granularity.

Figures 8 shows areas of interest based on intensity and/or time where baselines differed from expectations.

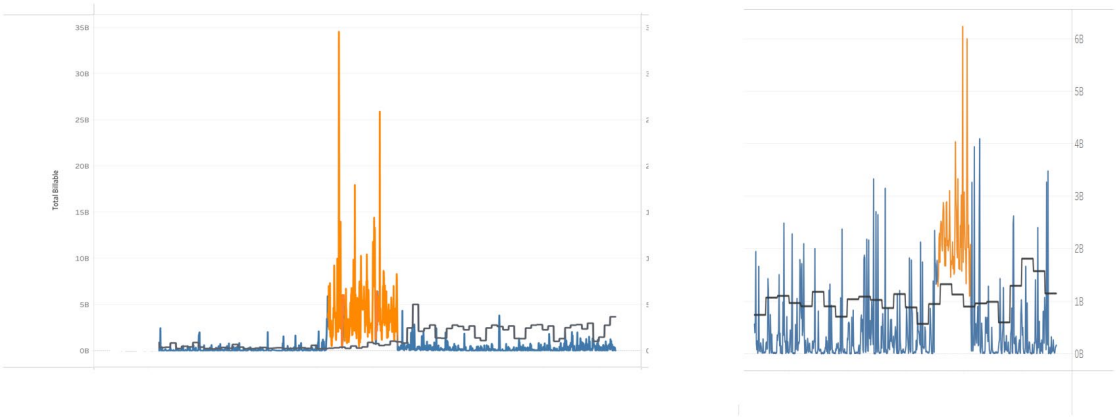


Figure 8 – Outlier Detection Examples

Figure 8 shows a divergence from baseline expectations in both magnitude and time whereas, Figure 8B reflects a time-based length event. While visual outliers may be easy to discern, these examples exist within hundreds of millions of examples, and the machine learning process is quickly able to highlight them.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

8. Large Scale Implementation

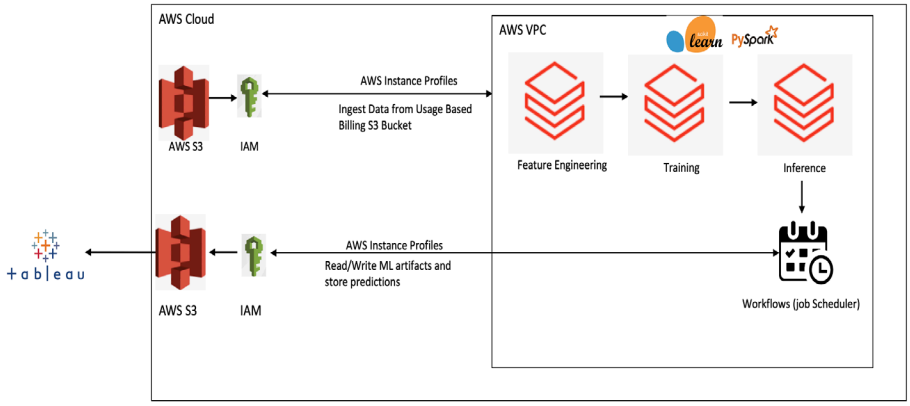


Figure 9 – Large Scale Design

Figure 9 shows the implementation of the model through Databricks on AWS as it easily processes large datasets and multiple models quickly and efficiently. The design is comprised of feature engineering, model training, and model inference. The predictions are stored in AWS S3, and Tableau is used for visualizations.

The model is in a preproduction phase at present. The addition of centralized logging and alerting will complete the deployment.

9. Conclusion

Through this paper, we have explored how it is possible to model highly variable and dynamic signals by creating a construct that initially develops modeled baselines using interpreted signal characteristics, creates data organization through unsupervised ML models, and then parses the information through outlier detection ML Layers 2 and 3 for additional refinement and characterization. With these methods, it is possible to provide reasonable coverage for very large systems and optimize outputs such that operational investigation of outliers is both reasonable and possible.

We have shown that the described methodology has captured potential network anomalies of a software issue because of the ability to model at a very granular resolution, and capture grouped behavior and outliers well. Finally, with the ability to model at granular levels, data sets and models become value added data streams when joined with latency, dropped packet segments, and other Quality of Experience (QOE) metrics.

Abbreviations

CMTS	cable model termination system
MTTR	mean time to resolve
QOE	quality of experience
QOS	quality of service

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Navigating the XGSPON Journey From an MSO Perspective

A Technical Paper prepared for SCTE by

Derek Lee
Manager, Wireline Access Technology & Engineering
Rogers Communications Inc.
8200 Dixie Ave., Brampton ON L6T0C1 Canada
derek.lee@rci.rogers.com

Giancarlo Urbani
Director, Cable Access & Transport Engineering
Rogers Communications Inc.
8200 Dixie Ave., Brampton ON L6T0C1 Canada
giancarlo.urbani@rci.rogers.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
1.1. Why PON? Business Drivers behind expediting PON introduction 4
1.2. Primer: PON vs HFC Networks 5
2. Architecture Comparison GPON vs. EPON Standards 6
2.1. Important Facts About PON Standards for MSOs 6
2.1.1. DOCSIS Provisioning Over EPON (DPoE) 8
2.1.2. RF Over Glass (RFoG) 9
2.1.3. Delivering Services Over PON 10
2.1.4. Service Syndication 10
2.2. The Case for EPON 11
2.2.1. Re-use DOCSIS Provisioning Platforms 11
2.2.2. Re-use DOCSIS OSS Platforms for Network Performance Monitoring 11
2.2.3. ONU Interoperability 11
2.2.4. Familiar Vendor Ecosystem 12
2.3. The Case for GPON 12
2.3.1. Global Market Dominance 13
2.3.2. Business Services Synergy 14
2.3.3. Wireless Network & Services Synergy 14
3. Expediting Time To Market 14
3.1. Microwave Backhaul 14
3.2. RFoG to PON Migration 15
3.3. 3.3 Distributed Split Design & Design Automation 17
4. Lessons Learnt In Introduction of XGSPON 20
4.1. Implementing XGSPON 20
4.1.1. Enclosure and POP site 20
4.1.2. Clamshell OLT 20
4.1.3. Network Latency 21
4.2. Operationalizing XGSPON 21
4.2.1. Network Performance Monitoring 22
4.2.2. Power Supplies 22
4.2.3. Operations Teams 22
4.2.4. SFU and MDU Customer Premise Installations 23
5. Conclusion 24

Abbreviations 24

List of Figures

Title **Page Number**

Figure 1 – Traditional CMTS, DAA, and PON 6
Figure 2 – PON Physical Network Architecture 7
Figure 3 – Logical Comparison of DOCSIS and DPoE Network 8
Figure 4 – Logical Network Diagram of DPoE EPON Network 9
Figure 5 – Logical Network Diagram of XGSPON Network 12
Figure 6 – GPON vs EPON Industry Landscape Comparison[1] 13
Figure 7 – Digital Microwave Backhaul for PON 15

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Figure 8 – RFoG to XGSPON Migration 16
Figure 9 – Using Combiner for RFoG to XGSPON Migration 16
Figure 10 – Underground Microduct Infrastructure for brownfield and greenfield 17
Figure 11 –Example of Design Automation Tool 19
Figure 12 – Customer Premise Installation 23

List of Tables

Title **Page Number**

Table 1 – Air Jetting Distance Chart 18

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

1. Introduction

More than ever Multiple System Operators (MSOs) are faced with unprecedented network capacity demand. Internet traffic growth, online gaming, video streaming services with higher resolution and higher bit rates, and video conference calls from work-from-home customers are among the contributors to consumer demand. Beyond the need to support the actual required capacity, MSOs are often pressured by their telco competitors on speed offerings from their fiber-to-the-home networks.

While MSOs continue to extend the runway of their Hybrid Fiber Coax (HFC) infrastructure with new DOCSIS technologies such as DOCSIS 4.0 and Distributed Access Architecture (DAA), many cable operators have also started exploring or deploying Fiber-to-the-Home (FTTH) Passive Optical Network (PON) as a way to address customer capacity demand, as well as to compete with telco's high speed service offerings.

This paper will share some of the technical analysis process and decisions Rogers made in navigating our journey to deploy FTTH 10Gbps symmetrical passive optical network (XGSPON), and the technology, deployment and operational lessons learnt along our journey.

Note that this paper is not intended to be a technical deep dive of the gigabit-capable passive optical network (GPON) or ethernet passive optical network (EPON) standards or implementation best practice, as there are plenty of resources readily accessible from CableLabs, ITU, IEEE and other industry bodies. Instead, this paper will focus on sharing Rogers's PON journey from defining the PON architecture, the pros and cons of all the technical assessments and decisions we had to make, and the lessons learnt from a cable operator's perspective in implementing and operationalizing PON.

1.1. Why PON? Business Drivers behind expediting PON introduction

For decades cable operators have successfully leveraged the high bandwidth of Hybrid Fiber-Coax (HFC) network and DOCSIS technology to provide triple-play services: high-speed internet, video and voice, to residential customers successfully, often beating its telco competitors with significant edge in service penetration over their DSL offerings. However, with heavy investments from the telcos in passive optical network (PON) deployments starting in the early 2000s, telcos who have made the investment in PON deployment now have a very competitive service offering in their PON footprint.

Competitive Pressure from Telcos

The major business driver for cable companies to invest in PON is competitive pressure on internet speed. Telcos coming from a DSL access network in the past could not compete with cable providers' DOCSIS technology in the 1990's and early 2000's. However for telcos who have started deploying PON to replace their DSL network, they have now come to a point where they can provide very competitive service offering in their PON footprint. Current generation 10G PON, as we will discuss further in this paper, has a line speed of symmetrical 10Gbps to each subscriber's home. Some PON service providers, including Rogers, are already offering 8Gbps symmetrical services because of the low utilization and small serving group (usually 32 homes). These high-speed service offerings present a challenge to cable operators in trying to match their speed while cable operators are actively modernizing their networks to prepare for introduction of next generation DOCSIS 4.0 technology.

Since PON networks are expensive to implement, for the telcos that have paid the sunk cost to build the optical fiber to the home network already, their primary motive is to recuperate as much of the sunk cost as fast as possible via increased service penetration with aggressive marketing. Many telcos have begun to offer their ultra-high speed PON services (such as 8Gbps) at very competitive rate comparable to cable

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

operators' "mass-market" offering (such as 1Gbps/50 Mbps service tier). This marketing strategy is putting pressure on cable operators trying to compete with DOCSIS service offerings.

Changing Consumer Usage Pattern

While HFC technologies continue to evolve, PON offers operators and customers many benefits, thanks to its ability to support multi-gig speeds and low latency experiences. While multi-gig is not critical today, emerging technologies may drive its importance in the future, such as:

- Many employees now work from their home offices full-time. While the typical conference call does not require more than 3-5 Mbps per call, some employees' jobs require them to send and receive large multi-gigabyte files to perform their job functions. Data scientists, for example, often need to download large data files daily. Video and graphic designers often need to download and upload large video files. The difference between a 30 Mbps upstream pipe and a 2 Gbps upstream pipe means the upload time is 67x faster. With the line between 'home' and 'office' becomes intertwined, the need to offer residential high speed symmetrical multi-gig internet service is becoming increasingly important.

- Many Rogers residential customers are also part-time or full-time video content generators on YouTube, Vimeo or other platforms. The need to frequently upload high resolution videos drives the need for a multi-gig symmetrical service.

- In 2023, the gaming industry is valued at over $160B and is among one of the highest revenue-generating industries globally. A good online gaming experience requires high bandwidth (especially with more and more gaming platforms shifting from a download model to cloud gaming model), but most importantly a low latency network. DOCSIS network averages ~30-45 ms in round trip latency today. With the emergence of low latency DOCSIS (LLD) cable operators aim to significantly reduce the latency to ~10 ms in an actual network. Today, PON offers low latency performance, and continues to be a popular solution with the gaming market. These business drivers show cable operators should at the very least consider deploying PON in strategical markets to remain competitive with telcos' offerings.

1.2. Primer: PON vs HFC Networks

While it is not the objective of this paper to compare DOCSIS technology to PON technology, this section should help provide the background for PON network and its fundamental technical differences from DOCSIS HFC networks.

In a traditional hybrid fiber coax (HFC) network, a CMTS (Cable Modem Terminal System) transmits an amplitude modulated (or often referred to "analog fiber") optical signal over fiber from the headend to an optical node in the field, which then converts the optical signal to radio frequency (RF) signal and transmits over coaxial cable to the customer homes, usually over cascade of active amplifiers and passive components such as multitaps and splitters.

The distributed access architecture (DAA) is a slight variation, where the CMTS .cCore will transmit 10GE signal over 'digital fiber' to the remote PHY nodes in the field, thus extending the 'digital fiber' further into the access network and increasing the bandwidth. However, the remote PHY node still transmits analog RF signal over coaxial cable as the last mile into the customer home.

Passive Optical Networks (PON) is an access network that is fully digital over an end-to-end fiber medium. The BNG (Broadband Network Gateway) in the headend transmits 10GE (or higher speed

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Ethernet signals such as 40GE or 100GE) digital optical signals to an optical line termination (OLT), which may be located inside a site facility or in the field. The OLT transmits the PON optical signal to customer homes over fiber, via a local convergence point (LCP) which is an optical splitter / coupler, where the signal is split / combined to feed a combining group of typically 32 or 64 homes.

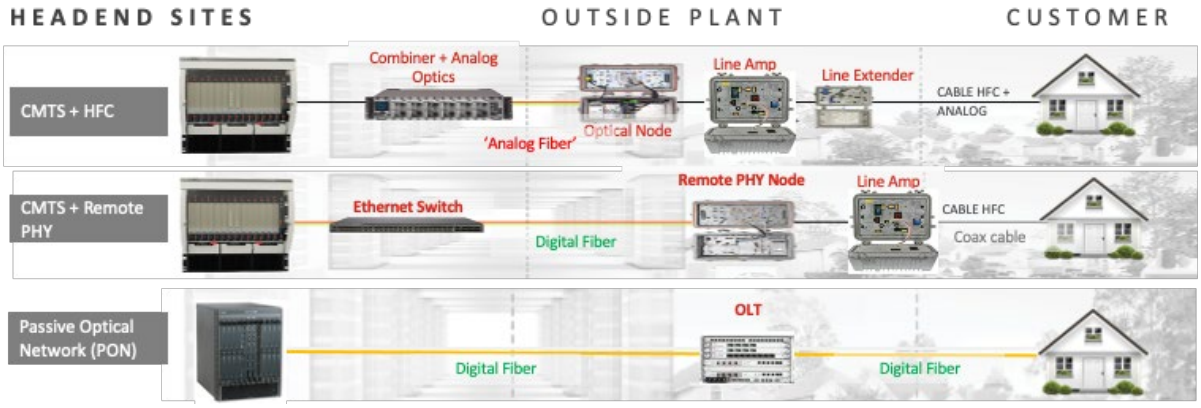


Figure 1 – Traditional CMTS, DAA, and PON

2. Architecture Comparison GPON vs. EPON Standards

2.1. Important Facts About PON Standards for MSOs

While there were various PON standards (such as ATM PON, BPON etc.) in the long history of PON technology development, there are only 2 relevant standards in 2023: GPON and EPON standards. The physical architecture design of either PON standard is largely the same : a broadband network gateway (BNG) in the primary hub (PHUB) to connect between the core IP network and the access network, an optical line terminal (OLT) to aggregate and terminate access connections, a local convergence point (LCP) to split and couple to 32 or 64 homes passed, and direct fiber from LCP to each individual customer home to terminate the optical signal on a customer premise equipment (CPE) called an optical network terminal (ONT) or optical network unit (ONU). This is common to both GPON and EPON, and their respective 10G or 25/50G standards.

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

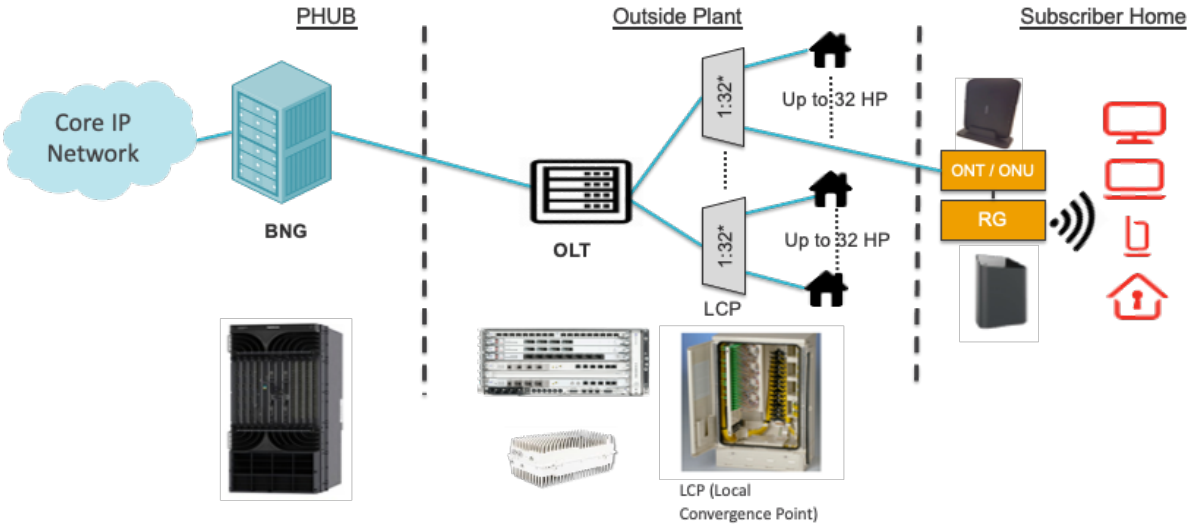


Figure 2 – PON Physical Network Architecture

While the physical network design is largely similar between GPON and EPON standards, the supporting provisioning, policy and other OSS and BSS platforms are very different. Any cable operator who plans on deploying PON must first understand the pros and cons of each standard and decide which of these 2 standards to adopt. Deciding on the PON standard is often the biggest decision cable operators have to make in their fiber-to-the-home PON strategy.

i) **Gigabit Passive Optical Network (GPON)** is an ITU standard G.984 that was originally standardized in 2003. The original GPON standard supports a line rate of 2.5 Gbps downstream, and 1.25 Gbps upstream.

The ITU GPON standard has evolved over the years to support 10 Gbps (Initially asymmetrical 10/2.5 Gbps XGPON or XGPON1, and the more popular 10/10 Gbps Symmetrical XGSPON ITU G.987). The next generation of 25Gbps and 50Gbps (ITU G.9804) products are expected to be generally available in 2023.

The GPON family supports co-existence of different speeds on the same PON, meaning when operator upgrades from 2.5G GPON to 10G XGSPON, both 2.5G and 10G ONTs (customers) can co-exist on the same optical distribution network (ODN), and allows upgrading of PON customers from 2.5G to 10G (and subsequently to higher speed such as 25G) on a per ONT basis.

ii) **Ethernet Passive Optical Network (EPON)** is an IEEE standard 802.3 that was originally standardized in 2004. The original EPON standard supports a line rate of 1 Gbps symmetrical.

Over the years the EPON standard has also evolved to support 10 Gbps which includes the asymmetrical 10/1 Gbps EPON and the more popular symmetrical 10G EPON. Like the ITU GPON standard, the IEEE standard EPON is developing 25G/50G EPON products and they are expected to be generally available in 2023 as well. And similar to GPON, different speeds of EPON can co-exist on the same optical distribution network (ODN) without having to upgrade the entire PON with higher speed ONU.

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

While it is sometimes misinterpreted that telcos adopt GPON and cable operators adopt EPON, that is not the case at all. Most cable operators in Europe, and South and Central America have deployed GPON instead of EPON. The largest current EPON deployments globally are in Asian markets including China, Japan and Korea, where EPON is equally adopted by telcos and cable companies. In fact, even among the cable operators in North America, Comcast and Charter are the only operators for large scale EPON deployment.

2.1.1. DOCSIS Provisioning Over EPON (DPoE)

In the early 2010s CableLabs working group developed a DOCSIS specification over EPON (DPoE) standard that is an extension to the EPON standard. The objective of DPoE is to help cable operators easily manage the provisioning and performance management aspects of EPON network by abstracting and translating EPON configuration information into virtual CMTS and virtual cable modems, so that cable operators can continue to use cable provisioning and network management platforms they are familiar with to manage the EPON network, without having to introduce network platforms like policy and charging rules function (PCRF), subscriber profile repository (SPR) etc., which are foreign to a lot of cable operators.

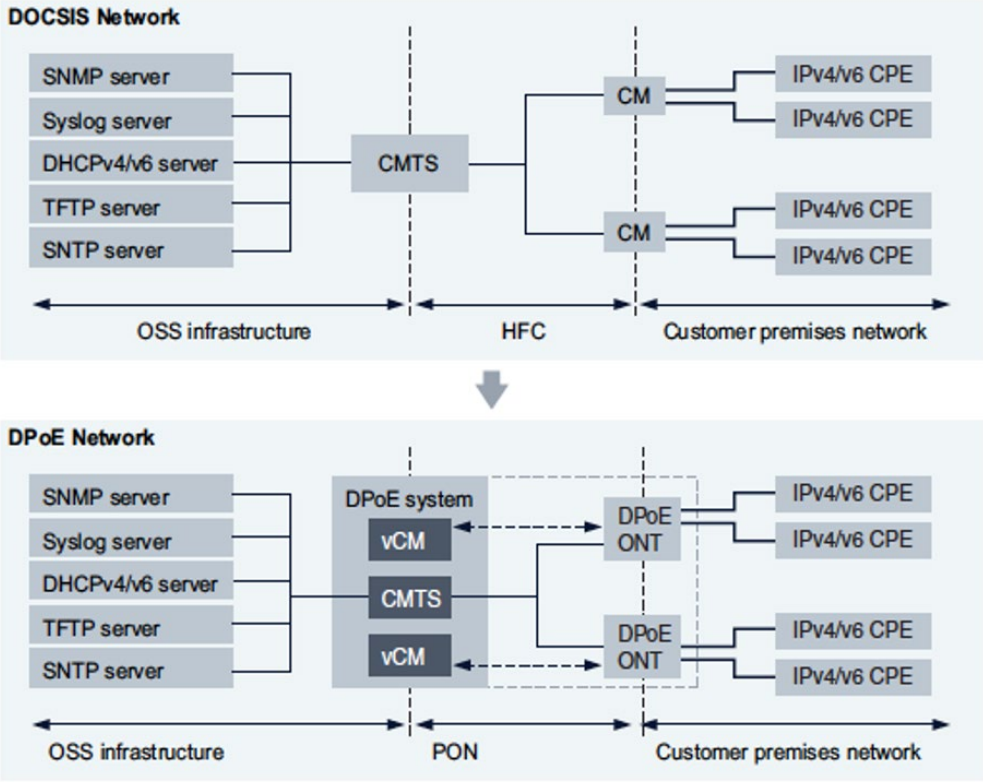


Figure 3 – Logical Comparison of DOCSIS and DPoE Network

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

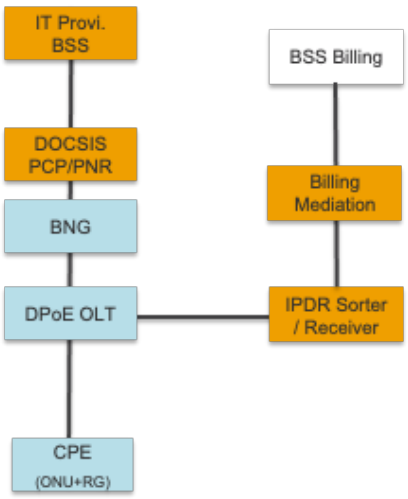


Figure 4 – Logical Network Diagram of DPoE EPON Network

The other significant advantage of DPoE is that it successfully standardizes the interoperability between different vendors' DPoE OLTs and ONUs. In the GPON standard, even though the OLT–ONT communication and management interface ONT management control interface (OMCI) is fully standardized, almost all major OLT vendors have inserted their own proprietary measures and information on top of the OMCI to enhance their ONT management, As a result, 3rd party ONTs are rarely 'plug and play' when paired with a different vendor's OLT. In fact, OLT–ONT interoperability is always challenging and unpredictable among GPON vendors, and many operators have chosen to avoid interoperating OLT and ONTs from different vendors. DPoE EPON completely eliminates this interoperability issue.

For cable operators considering the EPON standard, DPoE should be a natural decision due to all the above advantages to a cable operator. However, cable operators should carefully consider the advantages and disadvantages of EPON (or, DPoE EPON) vs. GPON before making architectural decision on either standard. This will be addressed in great detail in the second half of this section.

Subsequent to DPoE, CableLabs has also released a DOCSIS provisioning over GPON (DPoG) standard with the objective to allow cable operators to maintain their current provisioning and network management platforms while interfacing with a GPON access platform. However, DPoG was never widely adopted, and since in 2023 the eco-system is almost non-existent, the standard is no longer relevant.

2.1.2. RF Over Glass (RFoG)

In the early days of fiber to the home (FTTH) deployment, some MSOs chose to deploy RF over Glass (RFoG) as an alternative or precursor to PON deployment. While RFoG allows cable operators to deliver fiber to the home without major changes to their DOCSIS – HFC ecosystem, RFoG essentially uses an analog modulated (or 'analog') fiber to replace the last mile coax cable from the fiber node out into the customer home. The RF optical signal is terminated at a customer ONU, converted to RF over coax, and feeds respective cable modem or set top box equipment. This does not provide any real advantage over an HFC network, with the exception of the potential of slightly more RF spectrum.

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

RFoG lacks the bandwidth and speed of true PON technology, since it is still using quadrature amplitude modulation (QAM) modulation over fiber, and further, it creates optical beat interference (OBI) issues in the upstream that prohibits the deployment of orthogonal frequency-division multiple access (OFDMA) in the same serving group. Cable operators that plan to turn on OFDMA in the upstream for their DOCSIS customers will need to introduce some kind of OBI suppression mechanism on the CMTS, or migrate their RFoG customers to PON.

The main advantage of RFoG is it allows operators to offer RF services such as traditional video services and DOCSIS voice service. For operators that must continue to offer RF services in a FTTH setting, RFoG is the only option.

RFoG can also be overlayed with PON deployment over different wavelengths on the same access fiber for operators that need to provide high speed internet services over PON and traditional RF services over RFoG to the same household. The downside to that is it essentially doubles the network cost by having parallel CMTS and PON network equipment serving the same customer, all the way from headend access platform to customer CPE. It also doubles the number of network equipment that needs to be operated and maintained, thus inflating the operating cost and complexity.

In 2023 RFoG is no longer a viable technology, and many operators are removing RFoG from their network and replacing it with PON deployment.

2.1.3. Delivering Services Over PON

One challenge cable operators faced in the early days of PON was how to provide triple-play (internet, voice and video) services, since PON is a purely digital access platform, there is no backward compatibility for legacy RF services such as traditional QAM based video or DOCSIS voice.

PON can provide triple-play using internet protocol television (IPTV) and digital session initiation protocol based (SIP) voice services. However, if the operator does not support IPTV and SIP voice services yet, then the only remaining solution is to deploy parallel PON and RFoG access networks to serve the same customer home, and offer high speed internet over PON, and legacy video and voice services over RFoG. As aforementioned, this is not recommended as it doubles the network cost and operating cost.

As for business services, OLT is often superior to traditional DOCSIS network since many OLTs are capable layer 2 and layer 3 devices and offer operators an abundance of layer 2 and layer 3 business services to customers. PON also offers defined quality of service (QoS), and lower latency.

PON is also a capable access technology to complement or support wireless services. The high bandwidth and low latency of PON make it a viable option for LTE / 5G wireless service backhaul, and potentially mid-haul or front-haul using 25G/50G PON. PON also supports timing and frequency synchronization protocols such as Sync-E and IEEE 1588, allowing wireless operators to provide Precision Time Protocol (PTP) timing over the PON network to their wireless platform in absence of other medium such as GPS antenna. Note that the GPON family has much better wireless services support compared to the EPON family, this will be discussed in more detail in later sections.

2.1.4. Service Syndication

Note that although Comcast has adopted EPON standard for their PON deployment, their syndicated IPTV and voice services are access network agnostic and work equally well over GPON or EPON networks. The Comcast PON solution uses a 2 box solution, where the ONU (for EPON, or ONT for

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

GPON) terminates the PON signal, and the Comcast XB CPE will be enabled with EWAN (Ethernet-WAN) mode, which connects its WAN connection to the ONT/ONU, and functions as a home wifi gateway.

2.2. The Case for EPON

For many cable operators EPON may be the natural choice to deploy PON, because throughout the years CableLabs has been very supporting in defining DPoE EPON standard, and many North American cable companies have indeed deployed EPON as well.

In analyzing EPON vs. GPON standards, the main advantages of EPON from a .cCable operator's perspective are as follows:

2.2.1. Re-use DOCSIS Provisioning Platforms

Through DPoE's simulation of ONUs as virtual cable modems (vCM) and OLTs as virtual CMTS (vCMTS), cable operators can reuse the existing DOCSIS provisioning system (PVP/PNR, as an example) and its workflow to complete automated provisioning. Implementation of DPoE does effectively allow the current DOCSIS provisioning workflow to work with very little customization or changes over EPON.

For cable operators to adopt GPON standard for their PON deployment, they will need to introduce new provisioning platforms such as policy and charging rules function), (PCRF subscriber profile repository (SPR), and Policy and Charging Enforcement Function (PCEF), which is usually supported by the broadband network gateway (BNG). This requires new technology platform introduction, training and operationalization.

2.2.2. Re-use DOCSIS OSS Platforms for Network Performance Monitoring

Similarly, through DPoE standard cable operators can re-use current DOCSIS OSS systems to monitor network performance of OLTs and ONUs as vCMTS and vCMs, through the same OSS interfaces such as internet protocol data record (IPDR) and simple network management protocol (SNMP).

The GPON ecosystem does not support IPDR, hence cable operators will need to introduce new element management system (EMS) and integrate that with existing OSS platforms to monitor network alarms and performance KPIs. Subscriber usage metering is also not provided over IPDR, thus there needs to be additional integration between usage metering OSS platform and the Policy and Charging Enforcement Function (PCEF)/BNG.

2.2.3. ONU Interoperability

DPoE standard has successfully enforced interoperability among different ONU vendors so cable operators that adopt DPoE EPON will have the advantage to easy integration if they wish to introduce different ONU vendors.

OLT-ONT interoperability is very weak in the GPON vendor eco-system, regardless of 2.5G, 10G or 25/50G PON platforms. Extensive testing and integration are expected among most OLT/ONT vendors to achieve interoperability. Even when full interoperability is finally validated, there needs to be continuous interop testing with new firmware versions on the OLT and ONT side. Interoperability is even more difficult for GPON vendors that sell in both the ONT and OLT spaces due to the competitive environment. GPON operators that desire to keep dual GPON vendors often resort to geographical

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

division instead of interoperation between different vendors' OLT and ONT equipment for ease of operation simplicity.

2.2.4. Familiar Vendor Ecosystem

Due to popularity of DPoE EPON among North American cable operators, many of the largest EPON vendors are also large DOCSIS CMTS vendors. In fact, many DOCSIS CMTS vendors also manufacture DPoE EPON module for their distributed access architecture (DAA) remote PHY nodes to support an upgrade roadmap from DAA to PON. This familiar vendor eco-system gives cable operators an advantage of sales volume and vendor relationship compared to introducing new GPON vendors who may not be familiar to cable companies.

2.3. The Case for GPON

While GPON standards (2.5G GPON, XGSPON, or upcoming 25G/50G PON) are traditionally unpopular among cable operators, there are good reasons for cable operators to seriously consider GPON before finalizing the standard and architecture.

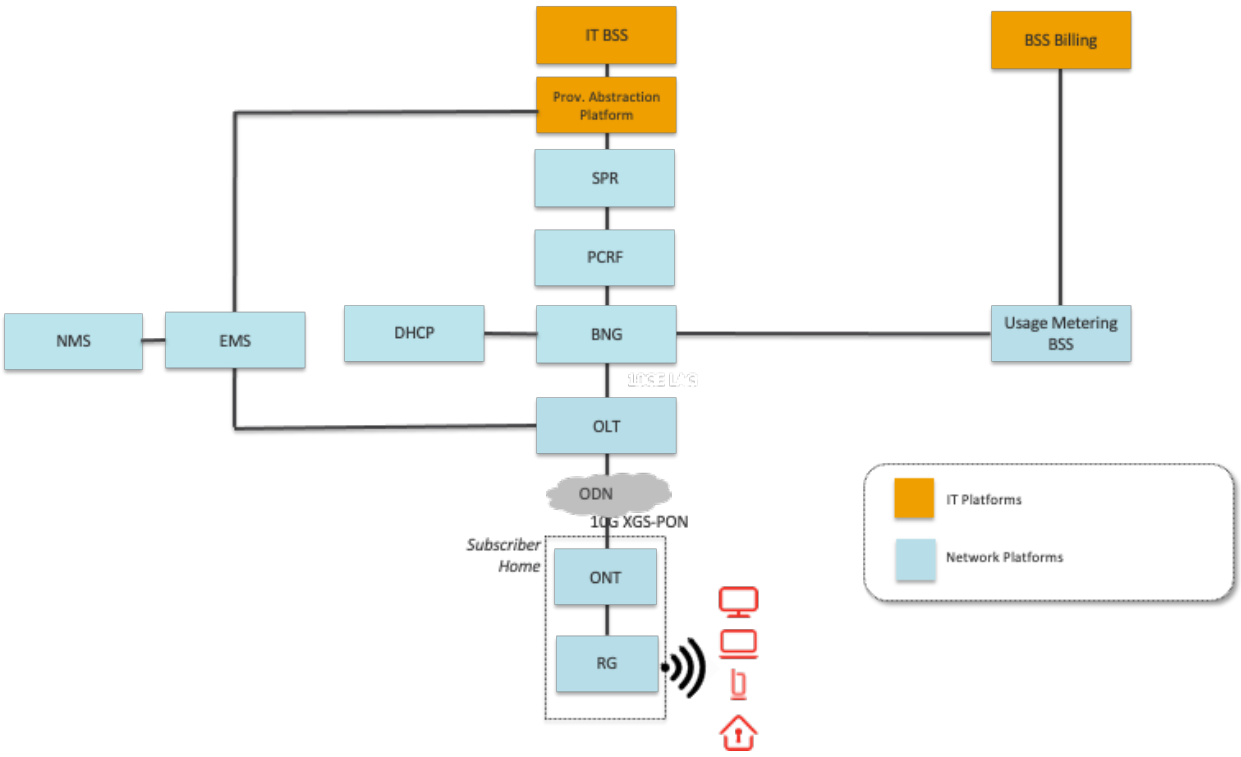


Figure 5 – Logical Network Diagram of XGSPON Network

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

2.3.1. Global Market Dominance

In 2023, GPON standards (2.5GPON, XGSPON combined) dominate over 92% of global market share compared to EPON standards (1GEPON, 10GEPON combined) with less than 8% market share.[1] This is expected to increase to 99% dominance for GPON standards in 2026 as the Asian market moves further away from EPON to GPON.

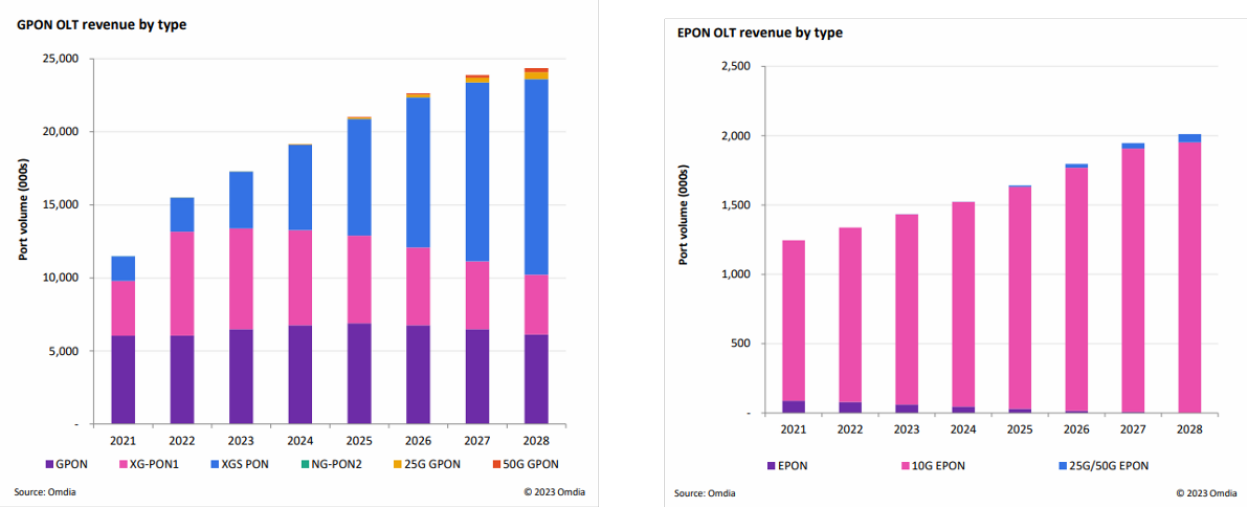


Figure 6 – GPON vs EPON Industry Landscape Comparison[1]

Contrary to popular beliefs where telcos deploy GPON and cable operators deploy EPON, actually a majority of EPON deployments were in the Asian-pacific market where China, Korea and Japan started deploying EPON in the late 2000s. Outside of the Asian-Pacific market GPON standards have over 95% market share. However, over the past 7 years with a lot of Asian-Pacific operators switching from EPON to XGSPON for their new deployments, this further widens the gap between GPON and EPON standards in global markets.

With a lopsided market split, operators need to be concerned about long term development and vendor support of the EPON platforms. GPON platforms have a much healthier vendor ecosystem, and operators can expect to have better equipment pricing due to volume, and better supply chain structure. GPON standards can also expect more research and development resources invested in it, thus supporting more new features to fill customers' and service providers' needs. This is evident from the difference in vendor ecosystems actively engaged in providing upcoming 25G/50G PON equipment from both the GPON and EPON standards.

Likewise for network platforms to supplement the GPON architecture, such as PCRF and SPR, are also commonly deployed by many GPON providers across the globe. They have a large vendor ecosystem with continuous development and support, competitive pricing and supply chain structure. Compared with DOCSIS's provisioning (PCP/PNR) provisioning platform, PCRF-SPR offers a vastly better vendor ecosystem for future support and roadmap.

[1] Source : Report by independent industry researcher Omida in *"PON OLT Vendor Landscape Report – 2023"*

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

2.3.2. Business Services Synergy

While EPON and GPON may be on par for residential service offering, the GPON platform provides improved business service support and synergy; not as much from a standard definition perspective, but from a market demand and equipment offering perspective. As an example, it is very common for GPON vendors' OLT chassis to support PON access cards for residential or small / medium business customers, and also point to point ethernet cards for large or enterprise business customers, thus installing 1 OLT node service, a provider will be able to provide a residential or small./medium sized business with PON 'best effort' offerings, and large business and enterprise businesses with guaranteed service level agreement ethernet services over point-to-point ethernet, plus wireless (more on that in next section) services. An EPON vendor OLT chassis typically focus on residential services, but the same OLT often lack business-oriented support and features such as point-to-point ethernet card support, layer 3 and MPLS services etc. When Rogers considered deploying DPoE EPON, it became very apparent that in order to offer most of our current enterprise and business services, we would have to overlay the EPON PON network with a separate equipment network to satisfy our business services needs. This doubles the access equipment cost, fiber wavelength usage, and maintenance and operation cost and complexity. Being able to provide all these services from a single PON network is an advantage for the GPON standard.

This may be a lesser issue in suburban areas where residential areas and business areas rarely overlap, but in a city it becomes expensive from a network build cost and maintenance cost perspective to overlay 2 separate fiber networks to serve the residential and enterprise customers in the same geographic area.

2.3.3. Wireless Network & Services Synergy

While this may not be the requirement for every cable provider, the GPON standard is a better strategic choice for wireless x-haul and providing timing and sync for 5G Wireless networks. The ITU standards include the stringent frequency, phase and latency requirements to support wireless x-haul and wireless synchronization. GPON network can be used to provide timing and synchronization via ONTs to wireless nodes that have no access to GPS antenna. The GPON (XGSPON and 25/50G PON) network can also be used for backhaul and mid-haul of the wireless network. This is lacking on the EPON network, especially on vendor support. EPON ONUs that support timing and sync standards (IEEE1588v2 or Sync. E standard) are non-existent because of lack of demand from EPON operators.

Wireless services may not be on top of many cable operators' priority list, but note that this also eliminates the PON network owner from a potential revenue stream to provide x-haul and network synchronization to other wireless providers.

3. Expediting Time To Market

Implementation of FTTH PON network is not only costly but is also time consuming especially in brownfield deployments. This section explores options where cable operators may be able to expedite time to market with PON offerings with a few 'out of the box' options.

3.1. Microwave Backhaul

Although connecting an OLT to the IP distribution network over fiber is preferred, having alternative solutions, such as microwave backhaul, that provide sufficient capacity and accelerate deployment timelines are often very helpful. Wireless transport often provides lower cost solutions when connecting difficult to reach areas such as islands, mountainous communities or even urban communities impacted

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

by construction moratoriums. Microwave backhaul for PON services is particularly useful for quadruple play service operators such as Rogers with extensive wireless infrastructure and in-house wireless backhaul expertise; OLT equipment serving the nearby community is installed in the existing wireless shelter and connected to a high-capacity microwave link mounted on the collocated cellular tower or rooftop sites. The microwave link is part of the mobile backhaul network that provides IP connectivity to the BNG at the core location.

Modern multi-band microwave equipment combines the best characteristics of different frequency bands to extend high capacities over longer distances. Combination of high-capacity channels in millimeter E-band with high availability channels in the traditional11-32 GHz frequency bands allows for microwave link capacities up to 20 Gbps over 10 km reach.

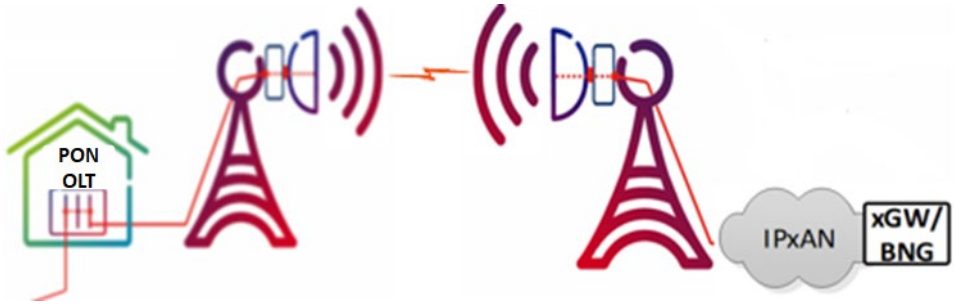


Figure 7 – Digital Microwave Backhaul for PON

3.2. RFoG to PON Migration

RFoG homes that cable operators have built over the past 2 decades may be the low hanging fruit to convert to PON, because the optical distribution network (ODN) to the customer homes already exists. Furthermore, RFoG does not offer any technical advantage over HFC DOCSIS offerings but creates its own limitations such as it inhibits turning on OFDMA for upstream.

Transitioning from RFoG to PON can be challenging for MSOs. The goal is to build a PON network parallel to the RFoG network while reusing as much of the existing RFoG network as possible to minimize costs. In general, the steps are as follows:

1. Build the parallel PON network.
2. Stop selling RFoG to new customers.
3. Migrate existing RFoG customers to PON.
4. Decommission RFoG plant.

While building the PON network, the link to the remote OLT should utilize existing RFoG links that can be freed up whenever possible. Options for reutilizing links include leveraging existing protection links and converting dual fiber RFoG feeds to single fiber feeds and repurposing the freed fiber for PON.

Migrating existing customers from RFoG to PON will be the largest challenge for MSOs. Ideally, as soon as the PON network is built, the customers will all be converted and the RFoG network can be

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

decommissioned. This is unlikely to happen as customers may be slow to migrate, and steps will need to be taken to allow for coexistence of RFoG and PON customers until the last RFoG customer switches.

If a second fiber is available from the splitter location to the customer premise, using this fiber to feed PON to the customer is the most simplistic approach. It does, however, require additional splitters to connect the OLT to the customer premise.

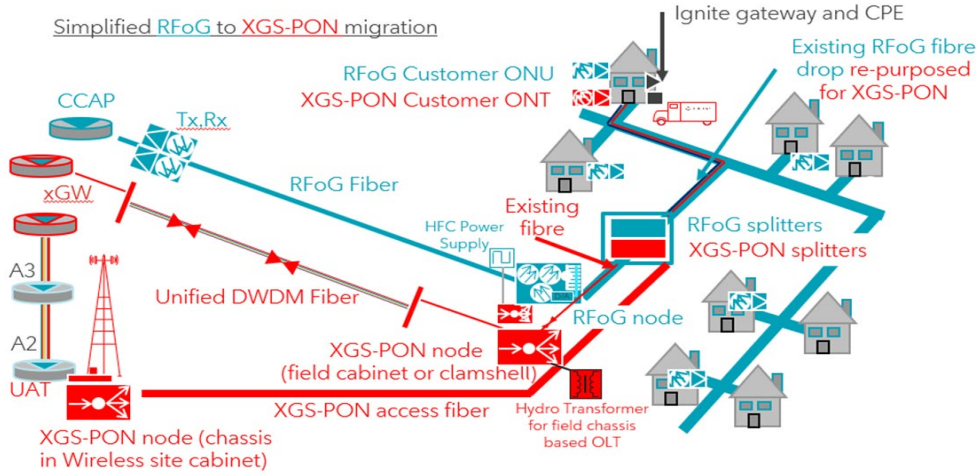


Figure 8 – RFoG to XGSPON Migration

If there is only a single fiber between the splitter and each home, combiners can be used to carry the PON wavelengths on the same fiber as the RFoG wavelengths. One example is shown below.

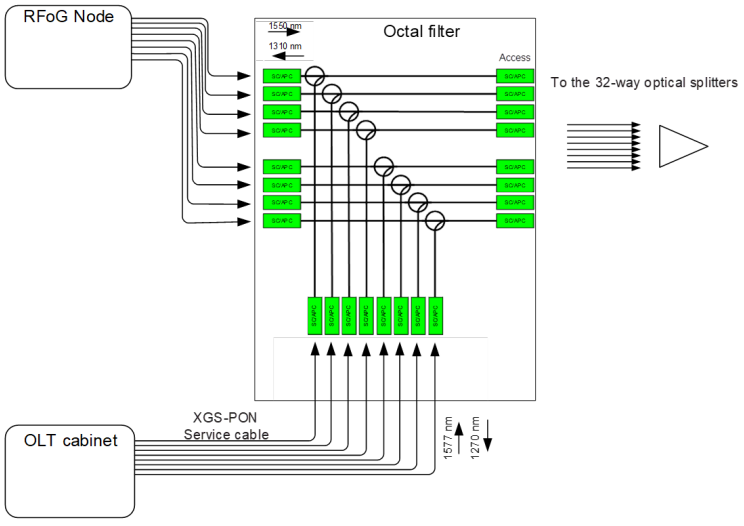


Figure 9 – Using Combiner for RFoG to XGSPON Migration

A disadvantage to this method is the additional insertion loss incurred on the RFoG path, making it necessary to re-align the upstream path. A second potential issue is that the SCTE specifies that RFoG ONUs must receive 1540 nm to 1565 nm but does not limit the upper band of the ONU downstream

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

receiver. Some manufacturers allow a wider band that allows the XGSPON downstream wavelength (1577 nm) to pass through to the ONU, impairing the RFoG signal. To mitigate, an optical filter may be added in-line to RFoG customers to filter out the XGSPON wavelength.

Benefits of migrating RFoG plant to XGS-PON:

- Provides 10 Gbps symmetrical capacity that will readily scale in the future.
- Provides unified fiber DWDM wavelength capacity for XGS-PON, DWDM based R4B EON, and any future adjacent R-PHY nodes.
- Harvests valuable DOCSIS ports for conventional node segmentation work.
- Avoids future costs and potential customer impact because RFoG technology is near end-of-life and end-of-support and it will soon become difficult and expensive to procure parts.
- Harvests RFoG equipment from early RFoG to XGS-PON migrations that may be used to segment RFoG plant as required before those areas are also migrated.
- Avoids future product impact because RFoG cannot be upgraded to DOCSIS 4.0, is limited to a spectrum of 1.0 GHz DS and 42 MHz US and cannot accommodate OFDMA.

3.3. 3.3 Distributed Split Design & Design Automation

Rogers has standardized on a distributed split architecture using air blown fiber (ABF) and a 2nd generation micro duct infrastructure for both greenfield and brownfield underground deployments for its speed of implementation, lower cost per unit and availability of multiple vendors to de-risk supply. The splitters are housed in fiber distribution hubs (FDH) and are available in multiple split ratios (1:2, 1:4, 1:8, 1:16, 1:32).

Micro duct is available in different configurations and sizes, and MSOs will need to select sizes and configurations that make sense for them based on their topology. Rogers has standardized on 7/3.5 mm (OD/ID) hard wall micro duct for distribution and a 14/10 mm micro duct for larger count transport cables. To access the micro duct from the sheath, a window cut is performed, and the duct is cut and spliced along with the toning wire. The unused end of the tube must be capped. After the micro duct lateral has been run, the air blown fibre (ABF) can be jetted to the side of the house.

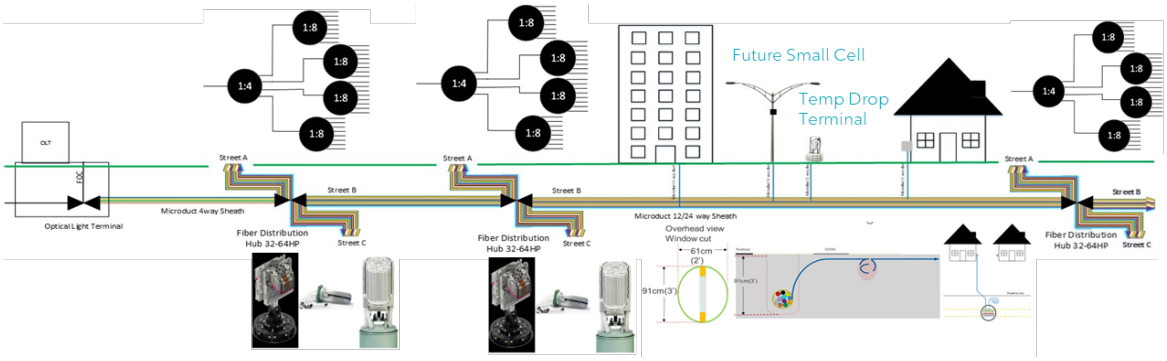


Figure 10 – Underground Microduct Infrastructure for brownfield and greenfield

Benefits include:

- capital cost reduction per home pass due to less street furniture
- less splicing and avoids additional ducts due to tether aggregation
- standardizes design and product components across greenfield and brownfield deployments

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

- improved equipment installation efficiency compared with BAU methods
- Fusion splitter; improved reliability, no tube distribution centres
- uses duct all the way to the home
- leverages existing pedestals for temporary drops. Maximizes splitter feed efficiencies

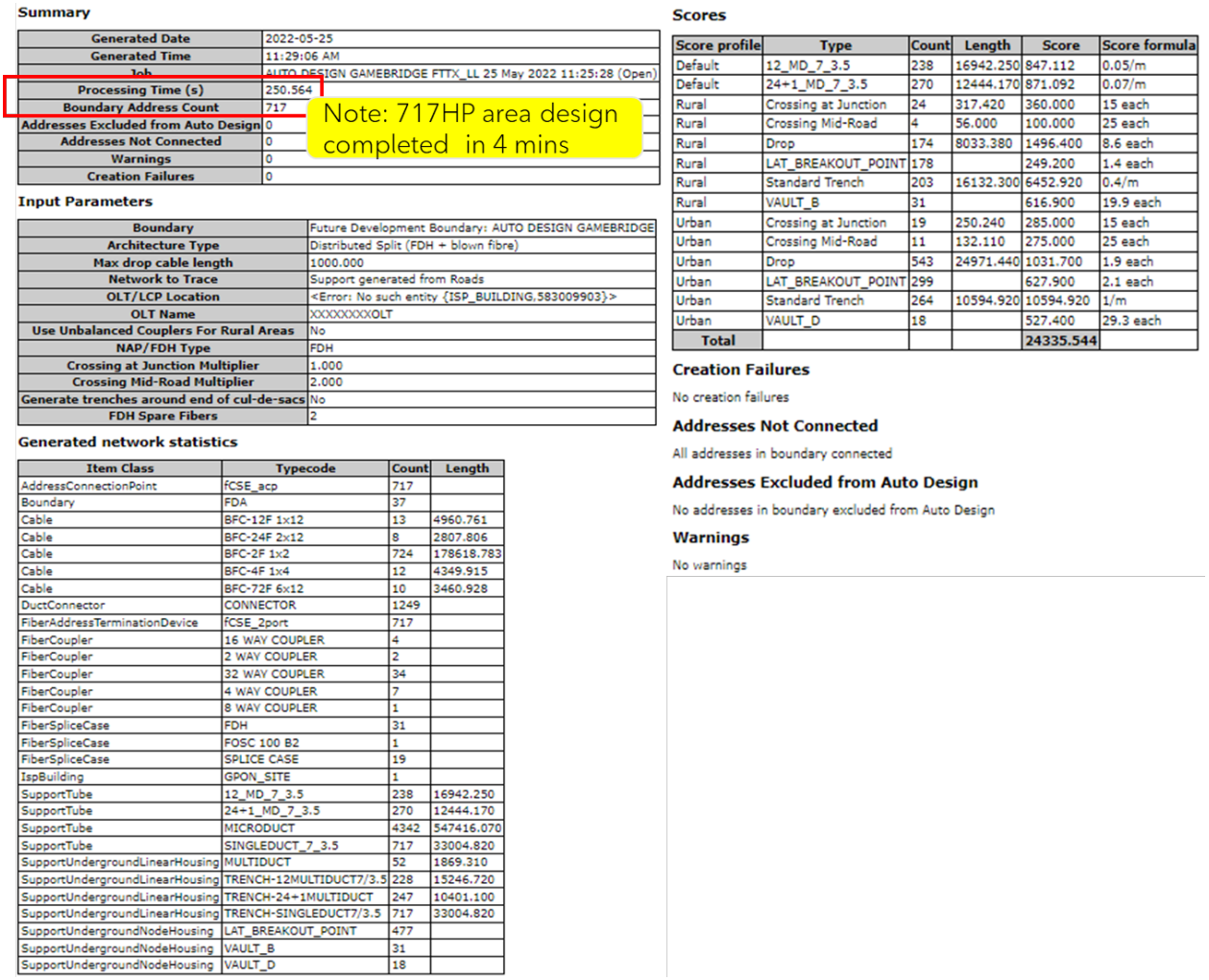
Blowing distance is heavily dependent on the size of the ABF and the size of the duct. The table below shows standard diameters and blowing distance for various fibre counts and duct size.

Table 1 – Air Jetting Distance Chart

AIR JETTING DISTANCE CHART			
CABLE TYPE and TYPICAL OD	CABLE DIAMETER (millimeters)	TUBE OD/ID (millimeters)	AIR JETTING DISTANCE (meters)
2 Fibre Unit	~1.1	5/3.5 and 7/3.5	1000
4 Fibre Unit	~1.1	5/3.5 and 7/3.6	1000
12 Fibre Unit	~1.6	5/3.5 and 7/3.7	600+
24 Fibre Unit	~2.05	5/3.5 and 7/3.8	600
48f-144f micro cable	≤ 8	14/10	1500
288f + micro cable	≤ 9.6	16/12	1400
Assumptions: up to 15 changes of direction 15-27% penalty, factored to account for bore and plough depth variation Recommendation is no more than 15 bends per km Recommend use of MD lubricant for micro cables			

Standardizing on splitters and micro duct sizes allows for the utilization of automated design tools. Automation can be accomplished by inputting a series of rules, criteria, and constraints that normally are factored into the design of a PON area. The automated design tool can then optimize the layout of the network, selecting the most cost-effective design and placement of equipment including splitter locations and ducts. Using AI/ML, automation tools can simulate many designs in a fraction of the time it would take to manually design a single layout. The tools also have the ability to transfer the design directly into MSOs' design/inventory platforms and can generate design reports and bill of materials (BOMs.)

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.



Summary

Generated Date	2022-05-25
Generated Time	11:29:06 AM
Job	AUTO DESIGN GAMEBRIDGE FTTX_LL 25 May 2022 11:25:28 (Open)
Processing Time (s)	250.564
Boundary Address Count	717
Addresses Excluded from Auto Design	0
Addresses Not Connected	0
Warnings	0
Creation Failures	0

Note: 717HP area design completed in 4 mins

Input Parameters

Boundary	Future Development Boundary: AUTO DESIGN GAMEBRIDGE
Architecture Type	Distributed Split (FDH + blown fibre)
Max drop cable length	1000.000
Network to Trace	Support generated from Roads
OLT/LCP Location	<Error: No such entity {ISP_BUILDING,583009903}>
OLT Name	XXXXXXXXOLT
Use Unbalanced Couplers For Rural Areas	No
NAP/FDH Type	FDH
Crossing at Junction Multiplier	1.000
Crossing Mid-Road Multiplier	2.000
Generate trenches around end of cul-de-sacs	No
FDH Spare Fibers	2

Generated network statistics

Item Class	Typecode	Count	Length
AddressConnectionPoint	fCSE_acp	717	
Boundary	FDA	37	
Cable	BFC-12F 1x12	13	4960.761
Cable	BFC-24F 2x12	8	2807.806
Cable	BFC-2F 1x2	724	178618.783
Cable	BFC-4F 1x4	12	4349.915
Cable	BFC-72F 6x12	10	3460.928
DuctConnector	CONNECTOR	1249	
FiberAddressTerminationDevice	fCSE_2port	717	
FiberCoupler	16 WAY COUPLER	4	
FiberCoupler	2 WAY COUPLER	2	
FiberCoupler	32 WAY COUPLER	34	
FiberCoupler	4 WAY COUPLER	7	
FiberCoupler	8 WAY COUPLER	1	
FiberSpliceCase	FDH	31	
FiberSpliceCase	FOSC 100 B2	1	
FiberSpliceCase	SPLICE CASE	19	
IspBuilding	GPON_SITE	1	
SupportTube	12_MD_7_3.5	238	16942.250
SupportTube	24+1_MD_7_3.5	270	12444.170
SupportTube	MICRODUCT	4342	547416.070
SupportTube	SINGLEDUCT_7_3.5	717	33004.820
SupportUndergroundLinearHousing	MULTIDUCT	52	1869.310
SupportUndergroundLinearHousing	TRENCH-12MULTIDUCT7/3.5	228	15246.720
SupportUndergroundLinearHousing	TRENCH-24+1MULTIDUCT	247	10401.100
SupportUndergroundLinearHousing	TRENCH-SINGLEDUCT7/3.5	717	33004.820
SupportUndergroundNodeHousing	LAT_BREAKOUT_POINT	477	
SupportUndergroundNodeHousing	VAULT_B	31	
SupportUndergroundNodeHousing	VAULT_D	18	

Scores

Score profile	Type	Count	Length	Score	Score formula
Default	12_MD_7_3.5	238	16942.250	847.112	0.05/m
Default	24+1_MD_7_3.5	270	12444.170	871.092	0.07/m
Rural	Crossing at Junction	24	317.420	360.000	15 each
Rural	Crossing Mid-Road	4	56.000	100.000	25 each
Rural	Drop	174	8033.380	1496.400	8.6 each
Rural	LAT_BREAKOUT_POINT	178		249.200	1.4 each
Rural	Standard Trench	203	16132.300	6452.920	0.4/m
Rural	VAULT_B	31		616.900	19.9 each
Urban	Crossing at Junction	19	250.240	285.000	15 each
Urban	Crossing Mid-Road	11	132.110	275.000	25 each
Urban	Drop	543	24971.440	1031.700	1.9 each
Urban	LAT_BREAKOUT_POINT	299		627.900	2.1 each
Urban	Standard Trench	264	10594.920	10594.920	1/m
Urban	VAULT_D	18		527.400	29.3 each
Total				**24335.544**	

Creation Failures

No creation failures

Addresses Not Connected

All addresses in boundary connected

Addresses Excluded from Auto Design

No addresses in boundary excluded from Auto Design

Warnings

No warnings

Figure 11 –Example of Design Automation Tool

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

4. Lessons Learnt In Introduction of XGSPON

4.1. Implementing XGSPON

This section shares some lessons learnt in Rogers' journey of implementing XGSPON. They are not definitive guidelines in implementing XGSPON network; however, they may be helpful to other cable operators in similar circumstances and conditions.

4.1.1. Enclosure and POP site

PON networks have a fixed optical link budget between the OLT and the customer homes, depending on the split ratio. In the case of 1:32 split ratio, most OLTs have to be within ~ 18km of customer homes. While this may be adequate for telcos to place their OLTs in their central office (CO) locations because they have been forced to move their COs close to their customer homes due to DSL technology limitations, most cable operators cannot serve the majority of their customers within 18km of hub sites.

Thus, the majority of cable operators will have to install their OLTs in environmentally controlled enclosures or point of presence (POP) sites in the field if they are available. This is due to the fact most chassis-based OLTs (or even micro-plug based Virtual OLTs) are not temperature hardened, and for most operators this means the OLT must be installed in a temperature-controlled POP site, or in an active environmentally controlled enclosure. Environmentally controlled enclosures complicate network planning and increases the deployment cost, because of the cost and size of the enclosure. Cable operators must also provide alarms and telemetry from the power supplies and batteries that may be inside the enclosure.

Cable operators are therefore recommended to locate the OLTs in close-by shelters, POP sites, multiple dwelling unit (MDU) telecom closets if the circumstances allow. However, that is not always the case, and installation of active enclosures may be inevitable as part of PON deployment.

Alternatively, many PON vendors offer clamshell OLTs which can be pole or strand mounted or can easily mount in passive enclosures or outside of buildings, because the clamshell already protects the OLT against environmental elements and human intrusions. However, pole mounted OLTs are harder to access especially in winter months in areas where there is snow. They will require a bucket truck or ladder for access. Clamshell OLTs have their advantages and disadvantages, which will be discussed in more details in the next section, but clamshell OLT is another viable option for field deployment.

4.1.2. Clamshell OLT

Many PON vendors offer clamshell OLTs which can be very appealing to cable operators whose access network is accustomed to deploying amplifiers, HFC nodes or remote PHY nodes using clamshells. Most clamshell OLTs can also leverage direct current (DC) square-wave cable power plant as well, which is an advantage to cable companies.

Clamshell OLTs eliminate the need to build enclosures in the field and provide an "all in one" solution that provides protection and security against outside elements and human intrusion. They can also be pole mounted (and in some cases, even strand mounted) offering more implementation options.

On the other hand, when evaluating clamshell OLTs, Rogers finds the per-home cost of the clamshell OLTs only make sense for relatively small subdivisions (<250 homes passed). As the size of the OLT grows, the per home cost of the clamshell OLT could begin to rise quickly, to a point it may cost 2x the

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

OLT equipment per-port-cost of a chassis based OLT for larger nodes (>500 homes passed), even including the associated active enclosure and power supply costs.

Cost aside, note that clamshell OLTs most often also lack the full features offered in the chassis based alternatives, such as business service support, wireless x-haul support etc. Operators should also expect that clamshell OLTs will be slow in supporting new features such as 25/50G PON compared to chassis-based solution due to lower market demand.

Operationally, clamshell OLTs are typically non-modular, meaning if there is any port or card failure the entire OLT module needs to be replaced, impacting more customers as compared to a modular chassis-based OLT.

Cable operators are recommended to assess the cost-benefits of both environmental enclosure solutions vs. clamshell OLTs based on their own circumstances and requirements. The advantages of clamshell based OLTs may not outweigh the disadvantages, and they may be best used as a corner case solution.

The clamshell OLT and chassis-based OLT cabinets have different costs and implementation considerations. Depending on the situation, the clamshell OLT or chassis-based OLT cabinet will be the most cost optimal. Choose a clamshell OLT or a chassis-based OLT cabinet for the most cost optimal long-term solution.

For example: In MDUs, space limitations and existing power supply electrical connections suggest clamshell OLTs will be most cost optimal; however, chassis-based OLT cabinets on the boulevard feeding multiple MDU buildings would be cost optimal.

4.1.3. Network Latency

Due to a reduced number of customers in a service group, the fiber medium and a reduction in number of network elements between the customer CPE and the broadband network gateway (BNG), PON offers excellent latency performance. Operators can expect 2-4 ms of round-trip latency with PON between the CPE and the core IP network, and ~10 ms in the worst case scenario.

Very low latency is a strength for PON; however, an operator should pay attention to the distance between the OLT and the BNG in the network design. relative to latency performance. Since the OLT to BNG link is simply a standardized Gigabit Ethernet (10GE typically) the PON connection will work over long distance over 100km as long as the signal level is within the link budget of the optics, but the distance will have an impact on the service latency. In one of Rogers's early PON trials customers were experiencing latency of over 40ms, which is very high for PON. It was later discovered the long haul transport network. routed the OLT–BNG link over a route of 100+km through another province. Fixing that issue improved the latency from 40 ms to under 5 ms.

4.2. Operationalizing XGSPON

This section discusses the lessons Rogers learnt in operationalizing XGSPON. Since PON as an access technology is quite different than what MSO operations teams are used to, the intent for this section is to share some lessons learnt in our journey of operationalizing PON that it may be helpful to other cable operators.

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

4.2.1. Network Performance Monitoring

Whether cable operator chooses to deploy GPON or EPON, building the OSS platform to capture network alarms and health telemetries must be part of the new technology introduction "day 1" requirements.

In terms of network health key performance indicators (KPI), because a PON signal is 100% digital from PHUB to customer CPE, the network KPIs are simpler compared to the network KPIs for an RF signal in a typical HFC plant. Nevertheless, PON network telemetries are not the same as DOCSIS telemetries, and an operator must work with the PON vendor to clearly understand the KPIs that are essential to maintaining the network.

In terms of service KPIs, since services do not change whether they are offered over HFC or PON network, the service KPIs are mostly consistent and agnostic of the access network. Operators can expect to use the same OSS system and same service KPI dashboards to monitor the services.

With the introduction of PON, an MSO must: 1. Define and test new KPIs for network performance, capacity, and triage; and 2. Develop new dashboards for Critical/Major/Minor alarm reporting

4.2.2. Power Supplies

Traditional HFC networks have transponders used for power plant monitoring. Typically, in a DOCSIS HFC network an external cable modem (CM) is used to backhaul the power supply monitoring signal out-of-band on the DOCSIS payload channel back to the PHUB, where it can then connect to the operator's management network and reach the power supply monitoring polling platform. The PON network does not have an identical solution for power plant monitoring, but a similar architecture that replaces the CM with ONT can be applied.

With the different power plants in the field, a single solution must be developed to resolve this issue. Working with multiple vendors, Rogers is implementing a solution with a single application that will be capable of monitoring multiple vendors' power plant that will improve network performance and efficiency within the network.

An operator should pay attention to IT platform details such as white-listing the power supply monitoring ONT, and allowing traffic to traverse between public data networks and private management networks. If practicable, leverage existing HFC power supplies to power the new clamshell OLTs.

During the customer migration period while there are both RFoG nodes and clamshell OLTs being powered, ensure the power supply maximum current draw is not exceeded. But if necessary, battery standby time does not have to meet the 4-hour standby specification, and is waived for the migration period.

Another consideration with OLT power supplies is whether the MSO can avoid local power metering and continue to categorize them as unmetered scattered loads by their local utility provider.

4.2.3. Operations Teams

Traditionally, MSO field technicians are mostly familiar with coax cable and RF signal, being that the majority of their experiences are with the HFC network. Transitioning coax cable technicians to make the switch from coax to fiber can be very challenging.

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Working with fiber cable in the field requires a new set of tools that cable operators must equip their frontline workers. Technicians may not be familiar with the new tools and require proper training. In addition, it requires technicians to be meticulous with a different frame of mind when dealing with fiber compared to coax cables. To help technicians overcome these challenges, proper training must be available, starting from basics such as emphasis on proper fiber cleaning.

In Rogers's experience, coax technicians were not accustomed to cleaning fiber and connectors. Dirty connectors were a major reason for service calls during early deployment of PON networks. Once our maintenance technicians were properly trained on fiber network maintenance we observed a significant reduction in service calls on PON network.

4.2.4. SFU and MDU Customer Premise Installations

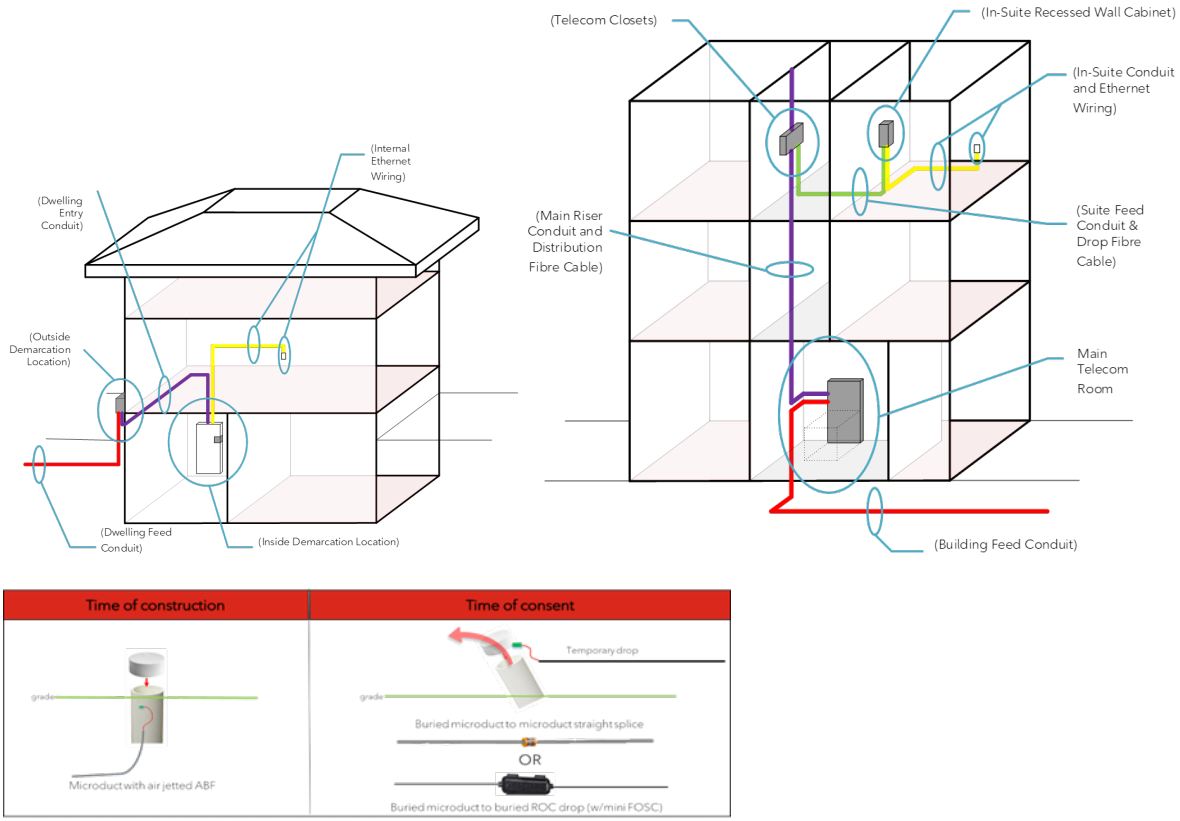


Figure 12 – Customer Premise Installation

For non-drop consents, Rogers established a low-cost container at the property line for single family units to use a temporary connection point when service is requested at a later date. Proper task procedure is required for lateral non-consent returns.

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

5. Conclusion

As cable operators and telcos continue to evolve their access networks to provide faster, lower latency, and more resilient and available services, PON offers a viable and high performance option.

While PON, and its associated network architecture and supporting network platforms may not be intrinsically native to traditional cable operators, it also is not an overly complex technology to master and operate.

To summarize, these are the takeaways of this paper:

- Cable operators are recommended to evaluate the pros and cons of GPON vs. DPoE EPON per their own requirements and constraints subjectively and select the standard that best fits their circumstances.
    - GPON standards have a significant advantage in global market deployment and will continue to dominate future PON deployments worldwide
    - DPoE provides cable operators with a fast track towards provisioning, BSS and OSS integration, and offer much higher interoperability among OLTs and ONTs from different vendors
- Cable operators can strategically expedite time to market with PON deployment with 'low-hanging fruit' use cases such as RFoG to PON conversion and using microwave as the OLT backhaul.
- Some lessons learnt in Rogers own journey of PON introduction were also shared in this paper, including the decision on appliance vs. clamshell OLTs, decisions on environmentally controlled enclosure vs. POP site installation, and some SFU/MDU installation guidelines.

This paper has shared some technical evaluation and decisions Rogers has made in introducing PON technology, and also lessons learnt in implementing and operationalizing PON from the perspective of a MSO. The hope is this paper and the experiences shared can be helpful as other cable operators navigate their path towards introducing PON.

Abbreviations

ABF	Air blown fiber
BNG	Broadband Network Gateway
BOM	Bill of material
BSS	Business Support System
CM	Cable Modem
CMTS	Cable Modem Terminal System
CO	central office
DAA	Distributed Access Architecture
DC	direct current
DOCSIS	Data Over Cable Service Interface Specification

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

DPoE	DOCSIS Provisioning Over EPON
EMS	element management system
EPON	Ethernet Passive Optical Network
FDH	fiber distribution hubs
FTTH	Fiber to the home
GPON	Gigabit Passive Optical Network
HFC	Hybrid Fiber Coax
IEEE	Institute of Electrical and Electronics Engineers
ITU	International Telecommunication Union
KPI	key performance indicator
LCP	Local Convergence Point
MDU	Multiple Dwelling Unit
MSO	Management Services Organization
OMCI	ONT management control interface
ODN	optical distribution network
OLT	Optical Line Termination
ONT	Optical Network Terminal
ONU	Optical Network Unit
OSS	Operations Support System
PCRF	Policy and Charging Rules Function
PHUB	primary hub
PON	Passive Optical Network
POP	Point of Presence
PHY	physical layer
QoS	Quality of service
RF	radio frequency
RFoG	radio frequency over glass
SCTE	Society of Cable Television Engineers
SFU	Single family unit
SIP	session initiation protocol
SPR	Subscriber Policy Repository
XGSPON	10Gbps symmetrical passive optical network

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Network Transformation and the Road to 10G

A Technical Paper prepared for SCTE by

Ted Boone
AVP, Engineering
Cox Communications, Inc.
ted.boone@cox.com

Dr. Bill Wall
Principal Architect
Cox Communications, Inc.
bill.wall@cox.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Overview of the Cox HFC Network ... 3
3. Services and HFC Architecture on the Road to 10G ... 3
4. HFC Access Network Spectrum Roadmap ... 4
5. HFC DOCSIS Evolution Stages ... 5
   - 5.1. HFC DOCSIS 3.1 Sub-split Network Overview ... 5
   - 5.2. HFC DOCSIS 3.1 Mid-split Network Overview ... 5
   - 5.3. HFC DOCSIS 3.1 High-split 1 GHz Network Overview ... 6
   - 5.4. HFC DOCSIS 4.0 Ultra-high-split 1.8GHz Network Overview ... 7
6. All-IP Transition ... 7
7. Target High-split and D4.0 Premises Architecture ... 9
   - 7.1. Preparing the home for DOCSIS 4.0 ... 9
8. The Road to 10G and CPE Lifetimes ... 11
9. Network Footprint Evolution ... 13
10. Conclusion ... 13

Abbreviations ... 14

List of Figures

Title **Page Number**

Figure 1 – Cox HFC Access Network Overview ... 3
Figure 2 – Services and HFC Architecture Mapping ... 3
Figure 3 – HFC Access Spectrum Roadmap ... 4
Figure 4 – D3.1 Sub-split Network ... 5
Figure 5 – D3.1 Mid-split Network ... 5
Figure 6 – D3.1 High-split 1 GHz Network ... 6
Figure 7 – D4.0 Ultra-high-split 1.8GHz Network ... 7
Figure 8 – Target High-split and D4.0 Premises ... 9
Figure 9 – Analysis of Single Point-of-Entry for D4.0 Modems ... 10
Figure 10 – Video Device Transitions ... 11
Figure 11 – Data Device Transitions ... 12
Figure 12 – Network Footprint Transition ... 13

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

This paper provides an overview of the Cox hybrid fiber-coax (HFC) network and its transformation towards delivering the 10G vision. The network roadmap maps different customer service tiers to HFC configurations, such as the sub-split, mid-split, high-split, and ultra-high split 1.8GHz configurations. The goal of achieving 10G is enabled by DOCSIS® 4.0 (D4.0), 1.8 GHz plant, ultra-high split, All-IP video migration, and other network component upgrades. The All-IP migration program and customer premise equipment align with this network transformation, with a recommendation to implement the premise architecture as All-IP and single-outlet. This long-range plan aims to achieve the All-IP 10G Network via DOCSIS 4.0, providing multi-gig services with 10G capacity across much of the HFC network.

2. Overview of the Cox HFC Network

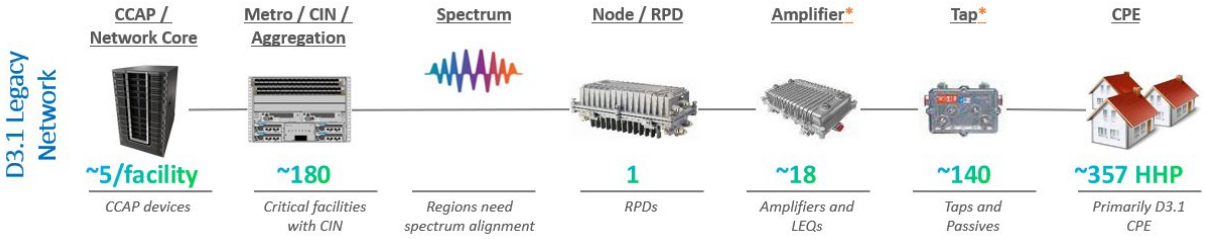


Figure 1 – Cox HFC Access Network Overview

An average node in the Cox HFC access network passes around 360 households or businesses and uses 140 taps and passives along with 18 amplifiers or line-extenders. Roughly 5 Cable Modem Termination Systems (CMTS) or Convergence Cable Access Platform (CCAP) devices in each of 180 critical facilities across 21 markets connect these nodes across the metro and backbone networks to the primary data centers and the broader internet.

3. Services and HFC Architecture on the Road to 10G

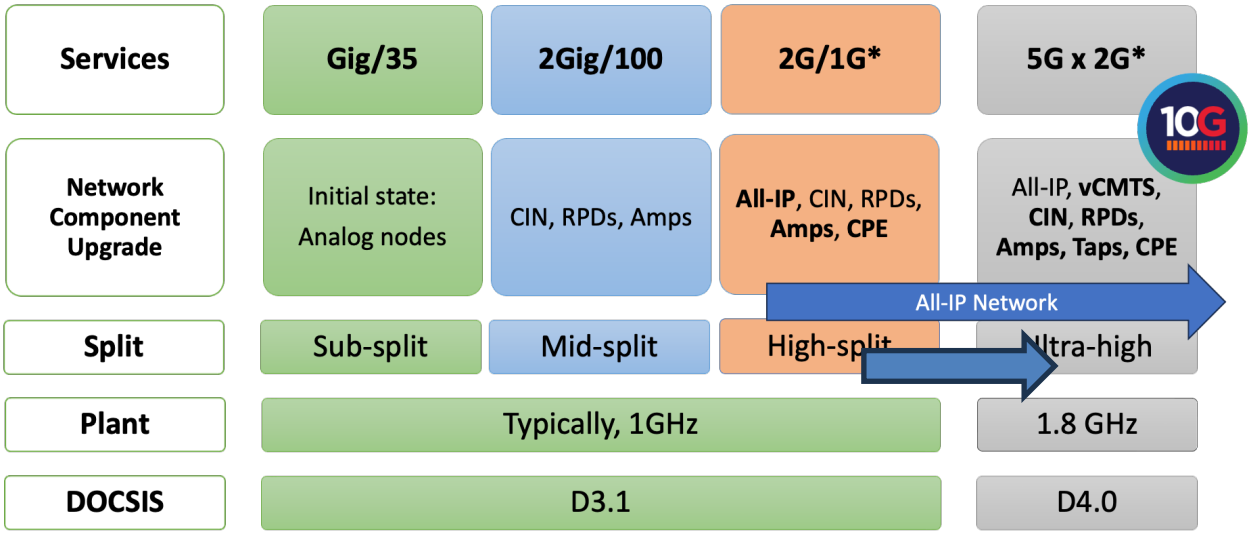


Figure 2 – Services and HFC Architecture Mapping

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

This high-level service roadmap above maps these various access network components and technologies to the service they ultimately enable for our customers. The goal of 10G is enabled by D4.0, 1.8 GHz plant, ultra-high split, All-IP and a host of network component upgrades.

While it is necessary for us to develop the capability to deploy these interim technologies along the way, most of our network will only see a couple of these configurations. And all will ultimately realize the 10G configuration.

4. HFC Access Network Spectrum Roadmap

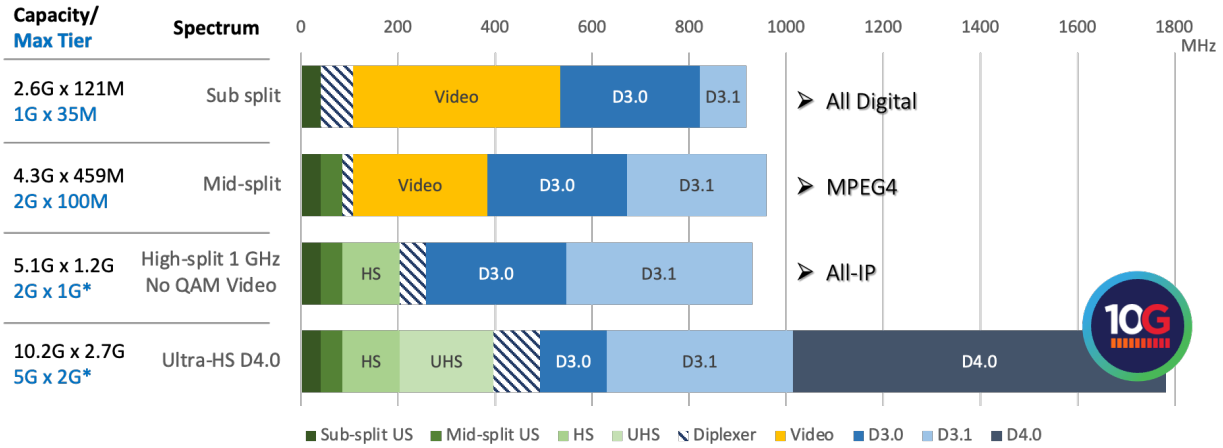


Figure 3 – HFC Access Spectrum Roadmap

Not quite half of Cox's HFC network is sub-split and 1 GHz or 860 MHz plant, with 120 Mbps upstream data capacity, and not quite 3 Gbps downstream data capacity and with a significant amount of the spectrum supporting legacy video. This configuration enables services tiers up to 1 Gbps/35 Mbps.

The next major network configuration was enabled by several initiatives: a conversion to all MPEG-4 compressed legacy video spectrum down to around 250 MHz, which enabled downstream data capacity expansion with additional OFDM blocks. And, of course, the mid-split upgrade program expands upstream data capacity more than 3-fold. The resulting total downstream+upstream node capacity approaches 5 Gbps, allowing high penetration of service tiers up to 2 Gbps by 100 Mbps.

The next major capability: high-split. With a commitment to reclaim and repurpose legacy video spectrum for upstream and downstream data, our network can be configured as high-split with actives similar to the mid-split upgrade and implemented with existing DOCSIS 3.1 (D3.1) silicon. This enables service tiers of up to 2 Gbps/1 Gbps. This step significantly de-risks the D4.0 timeline.

Lastly, as multi-gigabit downstream and upstream speed tiers are needed via HFC, we upgrade taps to 2 GHz, actives to 1.8 GHz ultra-high-split, and nodes to D4.0 to enable a 10 Gbps by 2.7 Gbps node capacity offering speeds up to 5 Gbps/2 Gbps. This is the realization of the "All-IP 10G Network".

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. HFC DOCSIS Evolution Stages

5.1. HFC DOCSIS 3.1 Sub-split Network Overview

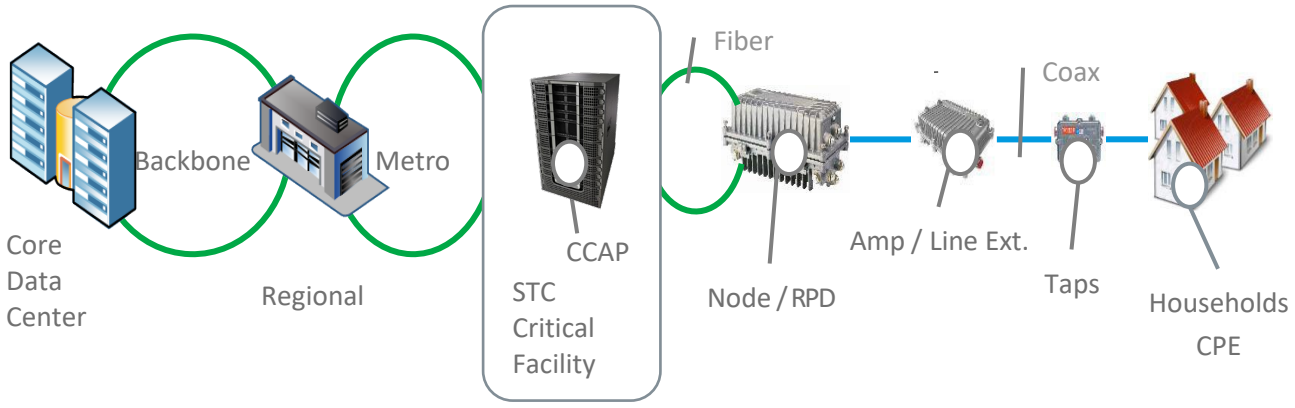


Figure 4 – D3.1 Sub-split Network

The figure above outlines the typical DOCSIS 3.1 sub-split network. This standard architecture connects equipment at the customer premises via the coax drop to the tap, usually though amplifiers, to an analog node, which is connected via the access fiber ring to the CCAP housed in the critical facility. Those facilities are connected via the metro fiber ring to the regional data center(s) which are connected via the backbone fiber ring to core data centers.

5.2. HFC DOCSIS 3.1 Mid-split Network Overview

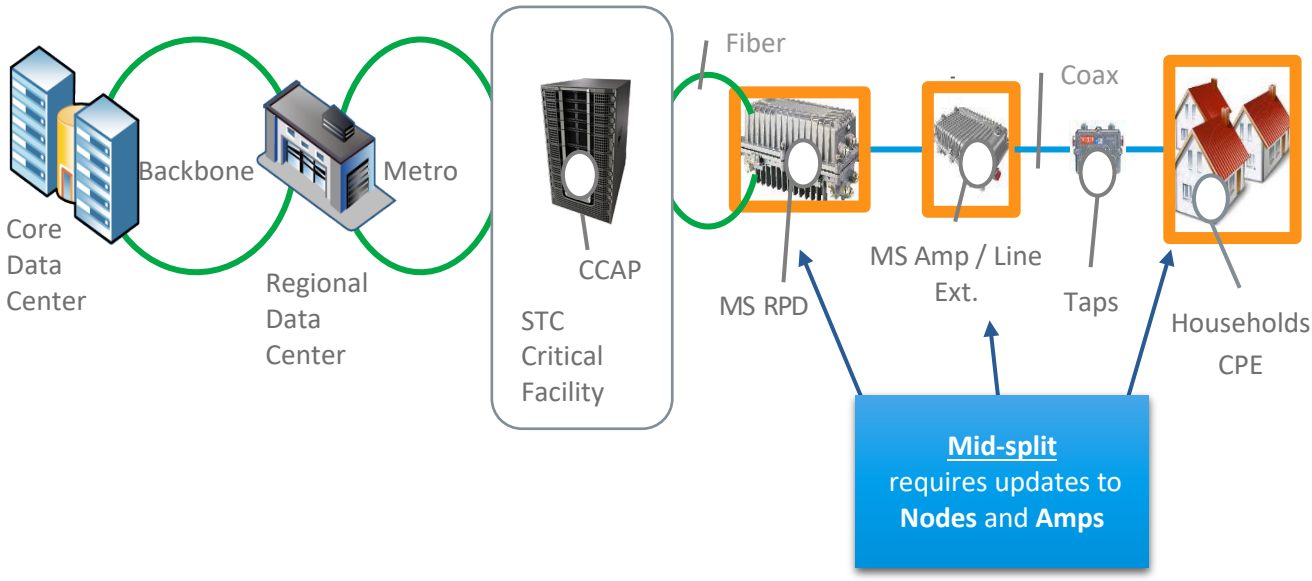


Figure 5 – D3.1 Mid-split Network

The figure above highlights the network component changes necessary to upgrade to the mid-split distributed access architecture (DAA) network. The node is upgraded to a mid-split Remote PHY device (RPD) with a redundant 10 GbE fiber-optic connection to the CCAP. Amplifiers are upgraded to mid-split. And, of course, D3.1 mid-split customer premises equipment (CPE) are required.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.3. HFC DOCSIS 3.1 High-split 1 GHz Network Overview

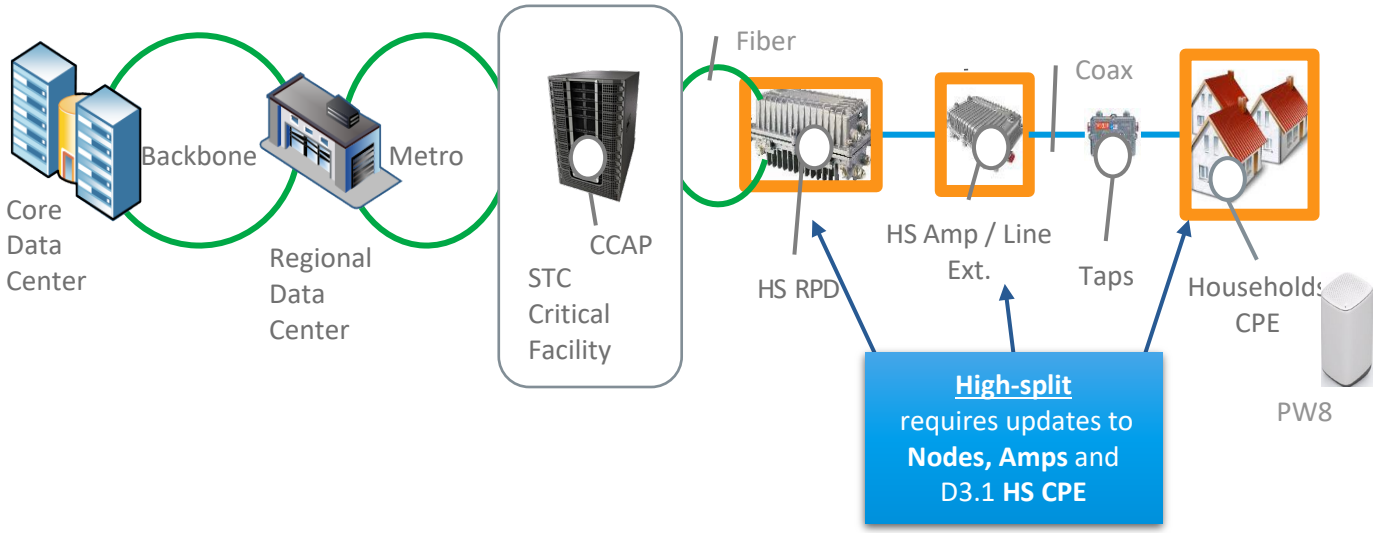


Figure 6 – D3.1 High-split 1 GHz Network

This figure above highlights the network component changes necessary to upgrade to the high-split DAA network. This transition is very similar to mid-split. The sub-split node is upgraded to a high-split RPD with a redundant 10 GbE fiber-optic connection to the CCAP. Amplifiers are upgraded to high-split 1.8GHz. And, of course, D3.1 high-split CPE are required. The high-split capable premises also begins the migration to single-outlet in preparation for a future transition to D4.0 1.8 GHz.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.4. HFC DOCSIS 4.0 Ultra-high-split 1.8GHz Network Overview

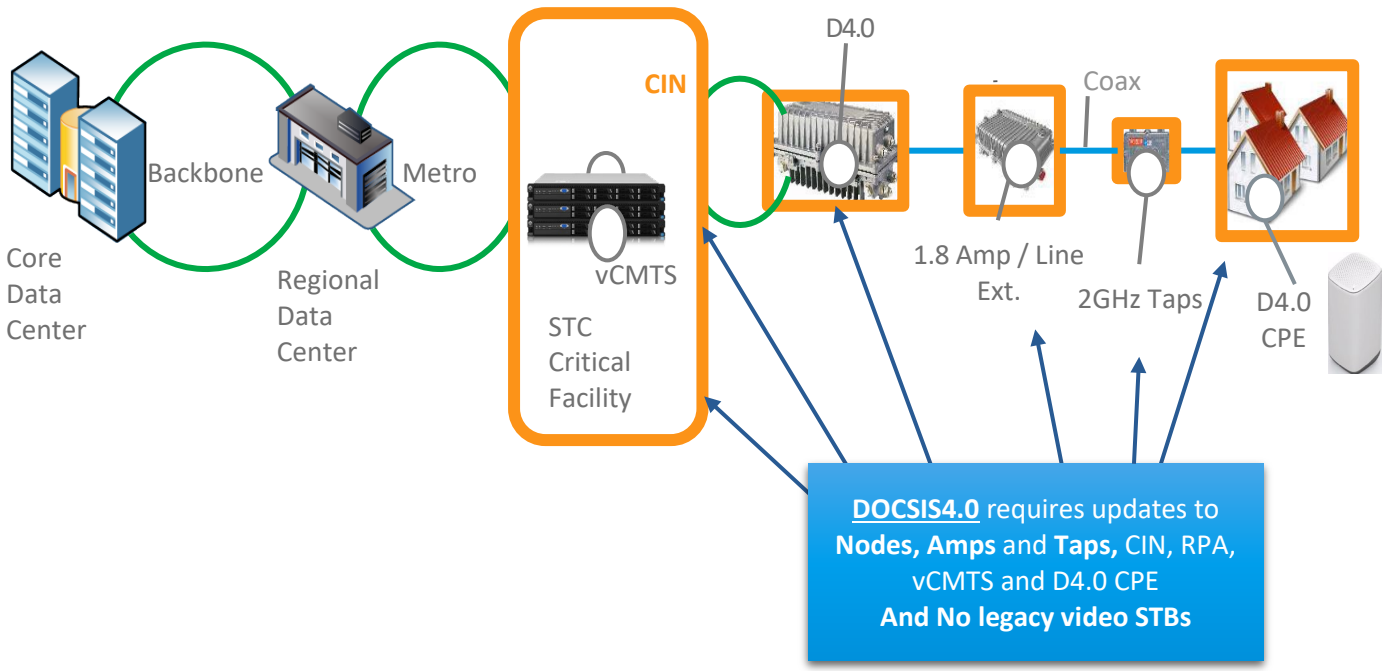


Figure 7 – D4.0 Ultra-high-split 1.8GHz Network

This final figure highlights the network component changes necessary to upgrade to the D4.0 ultra-high-split 1.8 GHz network. The D4.0 architecture will be supported by a virtualized CMTS (vCMTS) in the critical facility with 25 GbE converged interconnect network (CIN). The node is upgraded to a D4.0 ultra-high-split 1.8 GHz RPD. Amplifiers are upgraded to ultra-high-split 1.8GHz. Taps and passives are upgraded to 2.0 GHz. And, of course, D4.0 ultra-high-split 1.8 GHz CPE are required. The D4.0 ultra-high-split premises generally requires single-outlet.

6. All-IP Transition

Cox began deploying IP video approximately three years ago. Currently IP video subscribers represent about 14% of our video subscriber base. DOCSIS 4.0 is designed as a point-of-entry technology, and as such can't support splitters within the home that would connect to QAM based video set tops, thus requiring IP video within the home distributed via Wi fi or Ethernet. Cox is also planning to move homes that subscribe to tiers requiring high-split service to single point-of-entry and IP video. This decision is based on two primary factors: 1) to remove the possibility of high split modem upstream transmissions interfering into sub-split or mid-split video CPE within the home, and 2) prepare the home for eventual conversion to DOCSIS 4.0.

While the above can be done on a home-by-home basis, there is a significant advantage of converting all video subscribers in a market (or even on a node-by-node basis) to All-IP. When Cox converted all video to MPEG4 from MPEG2, we recovered more than 1/3 of the spectrum devoted to QAM video delivery. Moving to All-IP video allows us to recover approximately 250 MHz of spectrum to devote to DOCSIS. That video traffic moves to DOCSIS, but the amount of bandwidth used is far less. We observe that the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

average IP video subscriber uses approximately 6 Mbps of DOCSIS capacity at busy hour. On average we have less than 100 video subscribers per service group, or less than 600 Mbps of DOCSIS video traffic. This represents approximately 1/3 of the capacity of a 192 MHz OFDM channel.

While there is significant expense in replacing QAM-based video CPE with IP CPE, analysis has shown this is a cost-effective way to increase the DOCSIS bandwidth of the network. The majority of the Cox network is 1 GHz, with a small percentage at 860 MHz. Our mid-split 1 GHz plant has sufficient downstream capacity to offer a 2 Gbps top tier; however, moving the plant to high-split reduces the downstream available spectrum to the point that a 2 Gbps top tier could only be offered if the plant is converted to All-IP video. 860 MHz plant only has the capacity to offer 2 GHz downstream tier in a mid-split configuration if the plant is converted to All-IP. With an ultra-high split DOCSIS 4.0 deployment, legacy 1 GHz DOCSIS CPE would have very little downstream bandwidth available to them unless legacy QAM video was removed.

There are other advantages of moving to All-IP video. It simplifies the network by having all services delivered over IP, reduces complexity in the headend, eliminates the need for CableCard and out-of-band signaling, and provides a common method of delivering video to both our HFC and passive optical networking (PON) networks.

Cox is embarking on a multi-year transition plan to remove video CPE that is exclusive to QAM based video delivery and replace it with Wi-Fi™ enabled IP video devices. Some currently QAM based CPE, for example the Contour 2, contain DOCSIS 3.0 modems and can transition to IP delivery.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

7. Target High-split and D4.0 Premises Architecture



Figure 8 – Target High-split and D4.0 Premises

One rather important implication that results from this network roadmap is the future premises architecture. Since our long-term plan realizes the All-IP network, which means reclaiming legacy video spectrum and replacing legacy STBs, the straight-forward recommendation is this:
The high-split and D4.0 1.8 GHz premises architecture also be All-IP and single-outlet. Single-outlet means essentially all splitters and house-amps can be eliminated. All-IP means no legacy video QAM devices, no MoCA (Multimedia over Coax Alliance), and existing POE (Point-of-Entry) filters should be removed. Wi-Fi becomes the primary connection within the home, and Ethernet the secondary.

7.1. Preparing the home for DOCSIS 4.0

The DOCSIS 4.0 specifications were developed with the concept that the CPE is a point-of-entry device, that is, it is the only device in the home connected to the network and there are no splitters, amplifiers, or filters between it and the network. With a potential upstream frequency of up to 684 MHz and a maximum transmit power of 65 dBmV, a pre-equalized maximum power spectral density (PSD) was specified for the ultra-high split bands. Based on the analysis in the figure below, it shows that operation as a single point-of-entry device is required. Another compelling reason for both D4.0 modems and high-split modems to be single point-of-entry is that their upstream transmit frequencies fall within the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

downstream passband of sub-split and mid-split legacy video CPE. The transmit levels of these modems can overload the wideband front end of the legacy CPE within the same home, producing errors across the entire band.

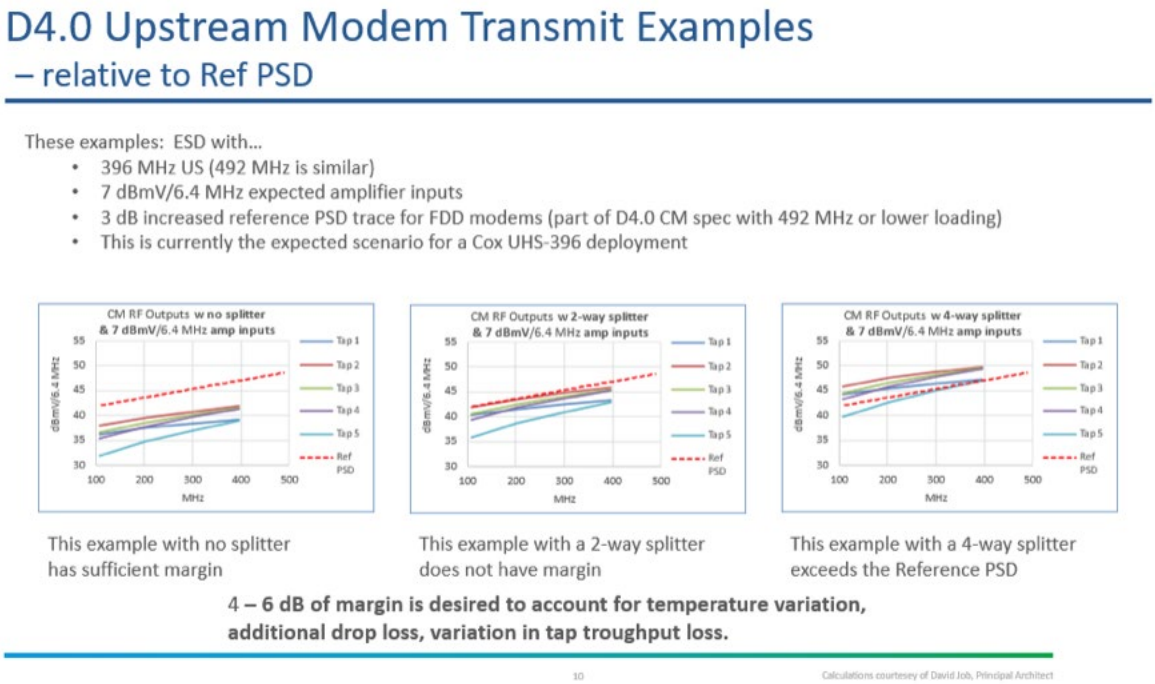


Figure 9 – Analysis of Single Point-of-Entry for D4.0 Modems

Single point-of-entry also significantly reduces the number of potential ingress points within the network, improving network performance and reliability. Early estimates are that 75% of ingress originates in the home, and single point-of-entry may reduce that ingress by 75%. Trials are underway to validate these assumptions.

Removal of splitters, MoCA point-of-entry filters, and house amps would require a technician to enter the home; however, Cox would like to preserve the ability for customers to self-install. Approximately 70% of current installations are self-installs. We are exploring different methods to accomplish a significant number of self-installs. One approach would have a technician, any time they enter a customer's home for any reason, convert that home to single point-of-entry whenever possible. That way the home would be ready for a self-install whenever the customer needs to upgrade their service.

High-split modems and D4.0 modems operating in high-split mode are not limited by the PSD requirements of ultra-high split and can in most cases work behind splitters. Legacy video equipment must still be removed. It would be possible to create self-install kits including IP video set-tops that could work in homes not yet wired for single outlet. In the case of D4.0 modems, they would be limited to service tiers supported by high-split and would require a technician visit to remove splitters if a higher service tier is required. A more immediate service call would be required if a house amplifier or MoCA POE filter was in place.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

8. The Road to 10G and CPE Lifetimes

How do services and spectrum transformation initiatives align to programs and CPE?

This series of figures attempts to illustrate that relationship between video and data CPE volumes over their lifetimes and the technology programs to which they relate.

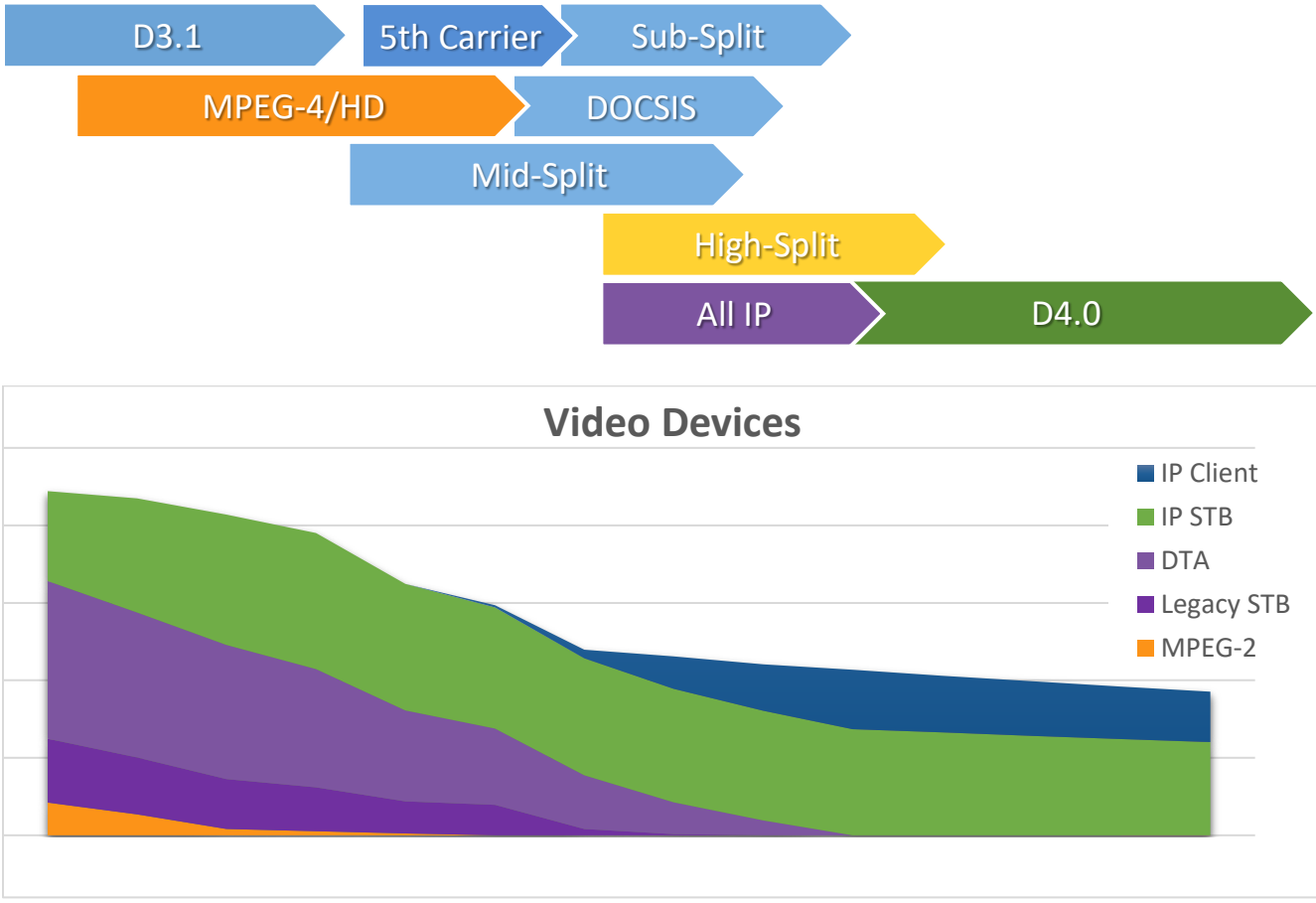


Figure 10 – Video Device Transitions

For example, the MPEG-4 program (orange) spans the end of life of MPEG-2 boxes. Likewise, All-IP, including legacy video retirement (purple) spans the end of life of legacy set-top boxes (STBs) and digital terminal adapters (DTAs), with only IP STBs and IP clients remaining long-term.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

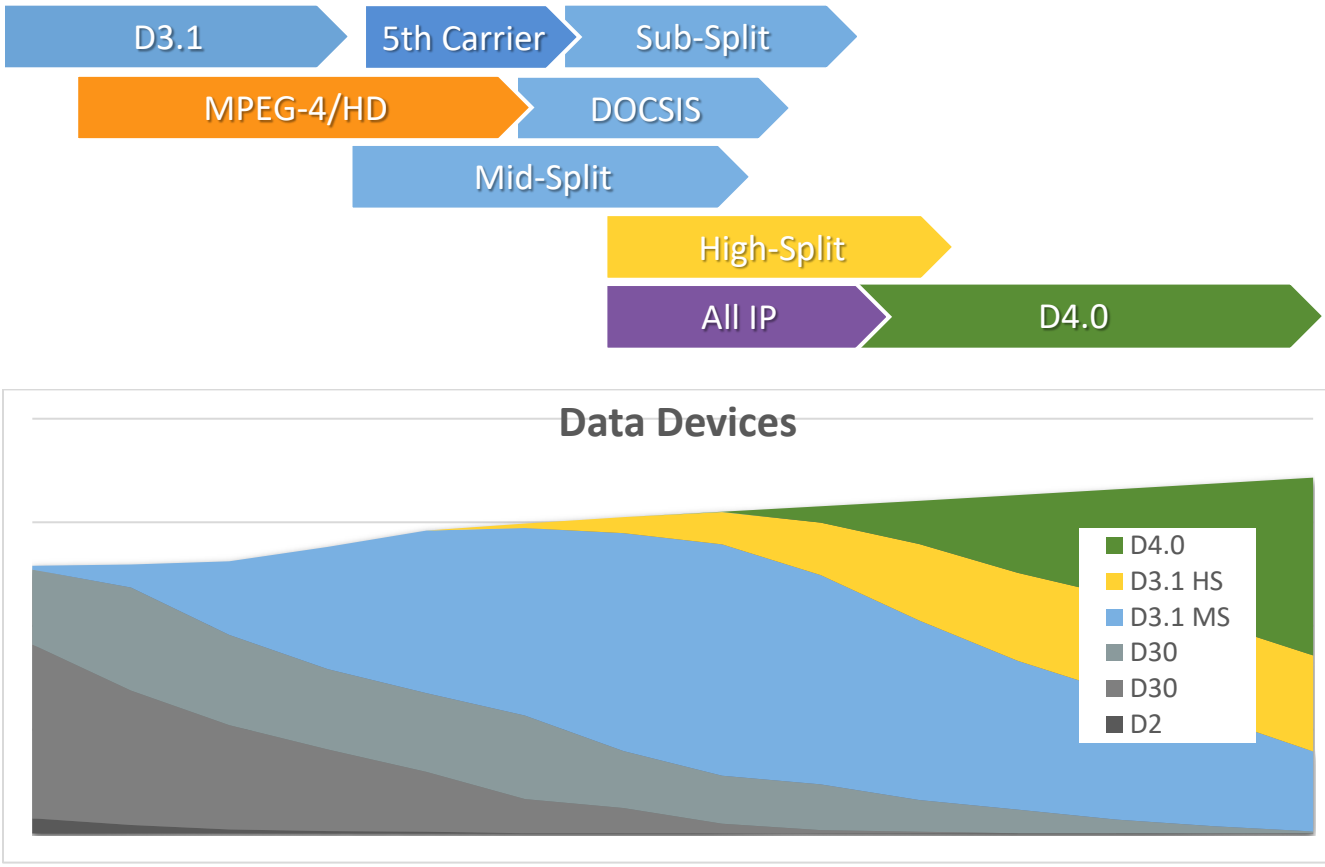


Figure 11 – Data Device Transitions

The introduction, and expansion of the D3.1 gateway (blue) with the D3.1 program supports sub-split OFDMA, OFDM Expansion, and Mid-split. High-split (yellow), which paces legacy video retirement, leverages the D3.1 high-split 1.2 GHz gateway.

And of course, the D4.0 Ultra-high-split program (dark green), requires legacy video retirement completion, and leverages the D4.0 ultra-high-split 1.8GHz cable modem (CM).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

9. Network Footprint Evolution

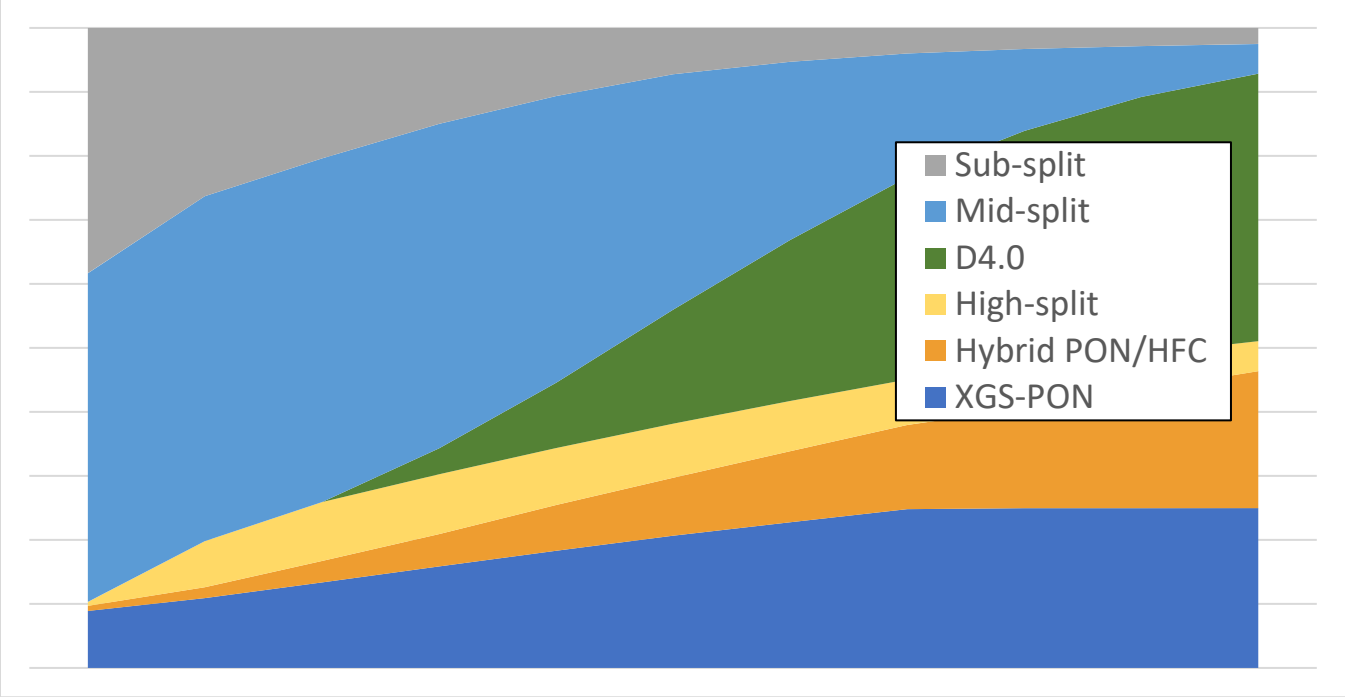


Figure 12 – Network Footprint Transition

The figure above demonstrates the enterprise-wide technology footprint that would be available given the long-range network plan. A healthy amount of mid-split 1 GHz footprint is already network ready and supporting multi-gig service tiers. High-split footprint grows as the All-IP transition commences. And, of course, the D4.0 and fiber-to-the-premises (FTTP) footprints as well, which are capable of 5 Gbps or better.

Once legacy video reclamation is complete, a preponderance of the network will be capable of 2 Gbps down or better. And moving rapidly forward with 5 Gbps or better.

10. Conclusion

This paper has outlined a long-term plan for Cox to get to a 10G network through stages of development that ends with a DOCSIS 4.0 1.8 GHz All-IP network that complements our XGS-PON deployments. The plan includes node actions that both solve network congestion and provide increasing speed tiers as needed to address competition. The result is a more reliable network that is designed to the state of the art.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

CCAP	Converged Cable Access Platform
CIN	Converged Interconnect Network
CM	Cable Modem
CMTS	Cable Modem Termination System
CPE	Customer Premises Equipment
D3.1	DOCSIS 3.1
D4.0	DOCSIS 4.0
DAA	Distributed Access Architecture
dBmV	Decibel millivolts
DOCSIS	Data over Cable Service Interface Specification
FTTP	Fiber to the Premises
Gbps	Gigabits per second
GHz	Gigahertz
HFC	Hybrid Fiber-Coax
IP	Internet Protocol
Mbps	Megabits per second
MHz	Megahertz
MoCA	Multimedia over Coax Alliance
MPEG	Motion Picture Experts Group
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
POE	Point-of-Entry
PON	Passive Optical Network
PSD	Power Spectral Density
QAM	Quadrature Amplitude Modulation
RPD	Remote PHY Device
STB	Set-Top Box
vCMTS	Virtualized Cable Modem Termination System
XGS-PON	10-Gigabit-capable Passive Optical Network

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Next Gen 10G Networks and 1.8 GHz HFC Design Challenges

The Design Path to 10G

A Technical Paper prepared for SCTE by

G. Keith Auzenne
Principal Engineer lll
Charter Communications
14810 Grasslands Drive Englewood, CO 80112
360-389-8958
Gerald.Auzenne@charter.com

Esteban Sandino
Distinguished Engineer
Charter Communications
14810 Grasslands Drive Englewood, CO 80112
720-518-2318
Esteban.Sandino@charter.com

Diana Linton
Principal Engineer I
Charter Communications
14810 Grasslands Drive Englewood, CO 80112
720-536-1027
Diana.Linton@charter.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Leveraging the Existing OSP – What are the Objectives? ... 3
3. Design Challenges – Legacy OSP vs 1.8 GHz Upgrades ... 3
   3.1. Gain, Cable Loss, and Distribution Spacing ... 5
4. Service Drop Engineering Profiles ... 7
5. Performance ... 9
6. Step-Down: TCP & Channel Power ... 10
7. Booster Type Actives for 1.8 GHz Design ... 13
8. Design Modeling Legacy OSP With 1.8 GHz Proposed Upgrades ... 14
9. Performing Network Due Diligence Ahead of Budgeting ... 15
10. Conclusion ... 16

Abbreviations ... 17

Bibliography & References ... 17

List of Figures

Title **Page Number**

Figure 1 - Gain, Cable Loss and Trunk / Express Spacing ... 5
Figure 2 - Distribution Example with and without Passive Loss ... 6
Figure 3 - Distribution Example with Booster ... 6
Figure 4 – Performance Indicators ... 10
Figure 5 – Step-Down Frequency(s): Constant, Single, Multiple ... 11
Figure 6 – Step-Down, Virtual- and Actual-Tilt, & Levels ... 11

List of Tables

Title **Page Number**

Table 1 - Gain, Cable Loss and Trunk / Express Spacing ... 4
Table 2 - Example of Service Drop Engineering ... 8
Table 3 - Example of Modem / Gateway RX & TX Levels ... 8
Table 4 - Example of Tap Output Levels and Modem / Gateway RX Levels ... 9
Table 5 - Channel Power 61 dBmV and TCP Calculations ... 12
Table 6 - Channel Power 60 dBmV and TCP Calculations ... 13

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Operators worldwide are working through the 10G initiative and are in the process of developing and deploying scalable, cost-effective technology platforms to power future 10G wired networks. These efforts are coupled with physical network upgrades to leverage today's nearly ubiquitous Hybrid Fiber Networks (HFC) deployed throughout the world. The ultimate goal is to support delivery of 10G services and enable new immersive digital experiences over low-latency, reliable and secure connections at speeds of 10 Gbps. The introduction of DOCSIS 3.1 and DOCSIS 4.0 have given operators the ability to move towards the path to 10G.

This paper outlines Outside Plant (OSP) architectural, technical, and physical challenges that must be addressed when designing the next generation of HFC networks to achieve higher split orders, greater speeds and increased bandwidth capacities. The importance of leveraging the *existing HFC infrastructure and maintaining legacy RF levels below 1.8 GHz*, and why this is crucial during the design process, will be explained. In addition, we will focus on the importance of managing total composite power (TCP) and other performance indicators when operating higher powered 1.8 GHz hybrids and amplifiers. And the importance of full spectrum RF levels, tilts and step-downs, to ensure best performance while minimizing distortions, will also be discussed.

Through much research, extensive design specification creation and experimentation coupled with network modeling over more than 30 years, one thing stands out; the design process is one of the most important phases of proposed outside plant upgrades. It is true that HFC networks are thought of as forgiving. Meaning, even though the spacing of actives from past bandwidth expansions may have been pushed to or past the limits of their capability, somehow the networks continue to deliver. With 1.8 GHz upgrades, this tolerance to compromised networks is greatly reduced. Performance and distortion parameters are more critical. Insertion and attenuation losses at such high frequencies accumulate much more rapidly. Well tested design specifications and the adherence to them at 1.8 GHz becomes more crucial to ensure healthy, bi-directional 10G capable networks.

2. Leveraging the Existing OSP – What are the Objectives?

- Increase operational bandwidth for greater network capacity while maintaining a budget.
- Increase bi-directional speeds for faster data transfer.
- Achieve greater reliability through improved network stability.

How does an operator do this?

- Understand our current OSP infrastructure. This identifies the physical and technical challenges to overcome in the design process which directly affects OSP construction and splicing costs.
- Perform electronics product testing, evaluations, and network design modeling to see what is needed from an RF signal level viewpoint.
- Work through the R&D process with electronics manufacturers to be sure they are building what is needed.
- Test and evaluate product samples to be sure they meet or exceed our requirements.

3. Design Challenges – Legacy OSP vs 1.8 GHz Upgrades

Design challenges must be approached systematically to understand what needs to be overcome, met, or exceeded in creating effective 1.8 GHz design specifications. Once the specifications are created, 1.8 GHz design modeling of existing lower bandwidth networks will allow us to see what the physical and technical plant demands will be as part of the proposed upgrade.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 illustrates how the maximum cable spacing between RF actives decreases as operating bandwidth increases. Gain has increased with every new generation of actives. However, as we go higher in bandwidth, coaxial cable attenuations, in addition to passive device insertion losses, will increase. As a result, it can be more difficult to maintain physical legacy amplifier locations, which is especially evident when re-spacing a 750, 870, 1002 or even 1218 MHz outside plant to accommodate 1.8 GHz.

Table 1 - Gain, Cable Loss and Trunk / Express Spacing

Generational System Amps – Gain Max – dB in Cable at Frequency			
MHz	**Gain dB**	**.750 Cable Loss dB/100'**	**Cable Feet**
750	37	1.48	2500
870	40	1.61	2484
1002	43	1.74	2471
1218	48	1.94	2474
1794	49-50-51	2.45	2000+

Relocating actives can be expensive. If design specifications are not created to optimize greater distances and higher losses between current active locations, the network upgrade costs will increase exponentially due to added construction and splicing costs.

In Table 1, the differential between 750 MHz and 1794 MHz spacing with .750 cable is 500'. To turn the 500' of .750 cable into dB we would do the following: 5-100's (cable feet) X 2.45 (dB loss per 100' @ 1794 MHz) = 12.25 dB. In theory this would mean we would be 12.25 dB short on our actives input to maintain the current location of the previously spaced 750 MHz active with a new 1.8 GHz active.

In an actual design, 1.8 GHz electronic equalizer (EEQ) losses at forward frequencies need to be considered when looking at Table 1. In addition, RF level input requirements of a manufacturer's 1.8 GHz active, as well as legacy frequencies below 1794 MHz, must remain within specification.

In the example below in Figure 1, actives A and B are 1.2 GHz actives spaced at 1218 MHz, while the actives gain is 48 dB. The spacing realized due to equalizer losses at all frequencies and amplifier input requirements is 45.0 dB of .750 cable. The result is 2325' of cable spacing.

Actives C and D are 1.8 GHz actives spaced at 1794 MHz, while the actives gain is 49 dB. The spacing realized due to EEQ losses at all frequencies and amplifier input requirements is 47.5 dB of .750 cable. The result is 1950' of cable spacing. What is shown here is a 1.8 GHz active will fall 500' or 12.25 dB short of making the spacing of the previous 1218 MHz active's location. We now have the below choices in design to make.

- Relocate the 1.8 GHz active to a location (pole) where it is no greater than 1950' of .750 cable spacing to satisfy the 1794 MHz input requirement.
- Upsize the .750 cable as necessary to a larger coaxial cable to reduce the amount of attenuation to satisfy the 1794 MHz input requirement maintaining 2325' of cable footage.
- Add a booster type amplifier at a location along the 2325' of .750 cable that increases the amplitude enough within the 2325' of .750 cable spacing to satisfy the input requirement of the 1.8 GHz active.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

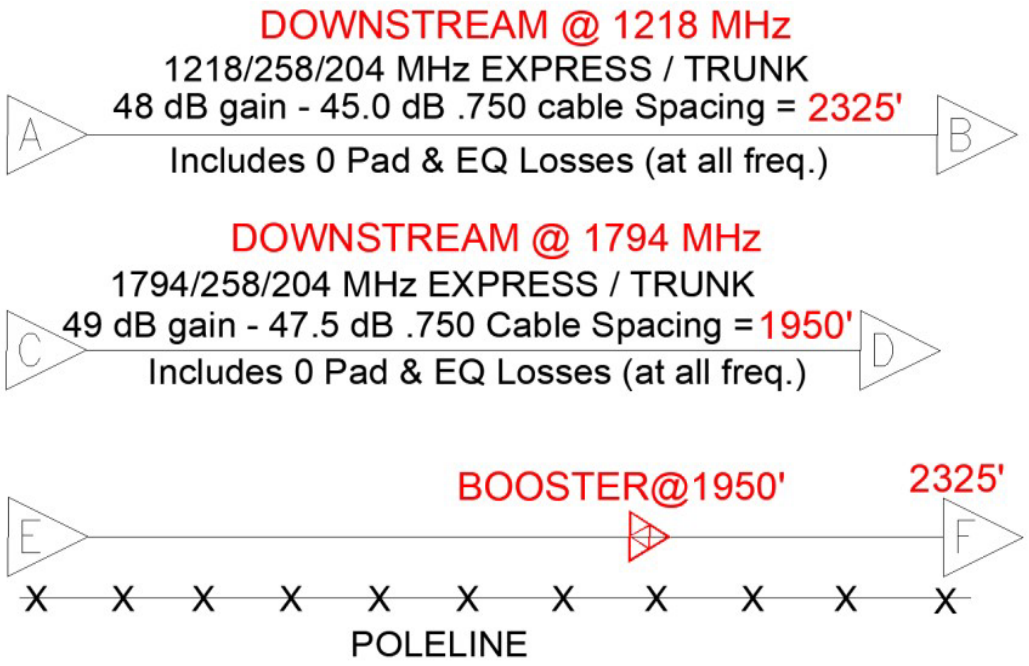


Figure 1 - Gain, Cable Loss and Trunk / Express Spacing

3.1. Gain, Cable Loss, and Distribution Spacing

Upgrading the distribution portion of the network can be especially challenging due to higher attenuation through smaller coaxial cabling and 1.8 GHz insertion losses through the frequent placement of taps and splitting devices. This is the section in the network where we must extract RF to serve our subscriber base. To that end, some points in the network will require certain levels of RF to be extracted or RF signals to be directed depending on the distribution requirements.

- Hot taps extraction of higher RF levels
- Directional couplers (DCs) or splitters delivering required input levels to additional downstream actives
- DCs or splitters feeding all passive legs (eliminating actives) requires higher RF levels
- Feeds and considerations for hardline-fed Wi-Fi transmitters
- Other dependencies upon an RF feed from the coaxial distribution (e.g., MDUs, hotels, etc.)

Figure 2 illustrates two things: 1218 MHz and 1794 MHz gain and spacing on P3.500 cable without any passive loss. Moreover, what happens when we introduce passive loss which is unavoidable in actual distribution. The higher losses at 1.8 GHz through the frequent placement of passives contributes to our depth of reach. In fact, adding passive devices can cut our total cable distance in half. The below examples would be significantly compounded if splitting were added to the passives shown.

It is imperative to know the gain of the chosen manufacturer's 1.8 GHz actives. Not all manufacturers' actives necessarily have the same gain. And not all manufacturers' actives have the same performance characteristics. Each of these considerations contribute heavily to defining network actives input and output RF design levels.

The below examples compare 1218 MHz and 1794 MHz. Even greater differences will occur between 1794 MHz and frequencies below 1218 MHz.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

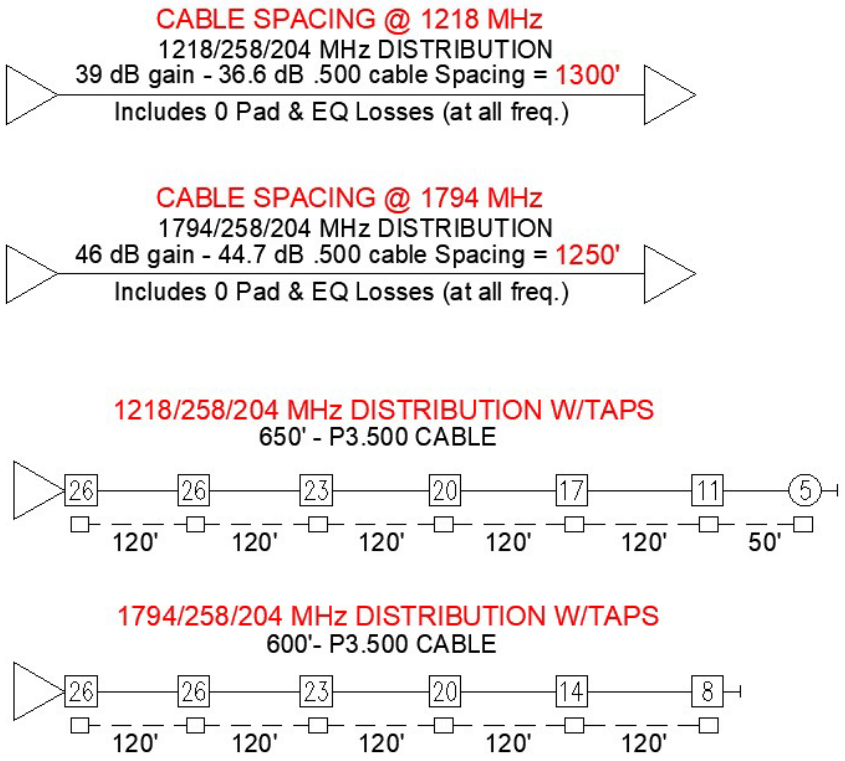


Figure 2 - Distribution Example with and without Passive Loss

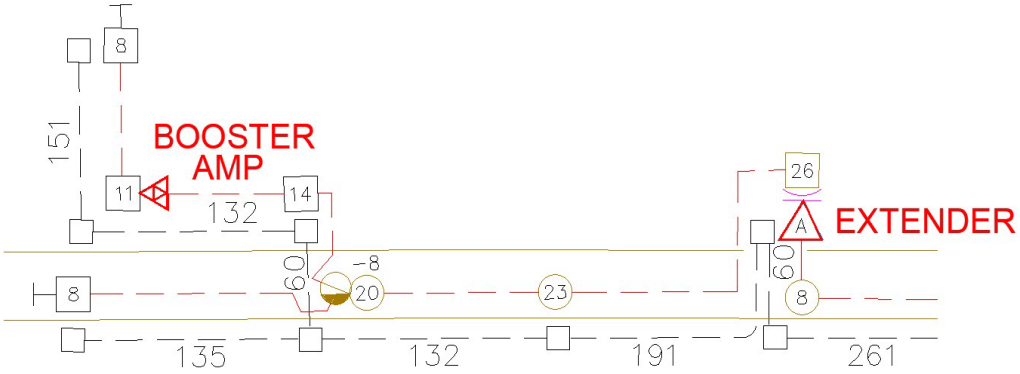


Figure 3 - Distribution Example with Booster

In Figure 3, there is not enough available RF for the design software to reach the last two taps. In 1.8 GHz design we may have a few choices to remedy the problem, but the most obvious choice would be to use a low gain booster type active. A line extender could also work, but that is much more gain and greater expense than necessary. We try not to knowingly waste gain.

Depending on the manufacturer, the gain in line extenders can sometimes be less than the gain in trunk or express type amplifiers. This means there is an additional challenge if standardization is utilized throughout the network with the intent for all actives to run at identical levels. The only way to achieve standardized levels would be to increase the input level requirement in the design software specifications for actives with lower gain in order to achieve the same output levels of actives with higher gain.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Example: A particular system type active might have 49 dB of operational gain at 1.8 GHz. A line extender might have 45 dB of gain at 1.8 GHz. The gain differential is 4.0 dB. To achieve the same output levels in both actives, the input level for the line extender would have to be 4.0 dB higher due to the lower gain. Having a higher input level requirement for any amplifier adds to the design challenges. These are the types of particulars that must be considered in design specifications.

4. Service Drop Engineering Profiles

We must understand distribution challenges in order to create design specifications that will meet or exceed the proper modem RX and TX RF levels at the premises or business.

Most networks have combinations of shorter and longer lengths of service drop line. Whatever these average distances are, as well as the service drop cable size, will help to define the RF levels needed from tap ports to adequately serve subscribers. Creating drop engineering profiles is integral in the development of design specifications to be sure we are maintaining proper modem RX and TX specifications.

Another component that merits further discussion in an operator's upgrade design specifications plan is the migration of DOCSIS modems, such as transitioning from legacy DOCSIS 3.0 or DOCSIS 3.1 modems to DOCSIS 4.0. It will be highly unlikely that an operator can swap all customers in a planned upgrade node or service area from earlier DOCSIS modems to a DOCSIS 4.0 modem immediately after the physical 1.8 GHz OSP upgrade. This may take weeks or months to accomplish.

In older existing plant below, 1.8 GHz splitters are common at the point-of-entry (POE) or even beyond that point located in the internal premises wiring. Sometimes splitters are in attics, walls or co-located with equipment. These splitters can be two-, three- or four-way splitters. Each addition of splitter ports constitutes additional RF signal loss. A two-way splitter will have about 3.5 dB of loss per port and a four-way splitter will have about 7.0 dB of loss per port.

The below bullets outline what is needed to create drop engineering profiles within your networks for legacy frequencies, as well as upcoming 1.8 GHz:

- Define average drop distance from tap port to premises POE or demark.
- Define average interior drop wiring distance to set-top box or modem located within the premises.
- Define average house type splitter number of ports at POE.
- Define average drop cable size RG-59, RG-6 or both.

Calculation example for premises (legacy modem) at 1.0 GHz:

RG6 drop cable loss / 100' = 6.55 dB

Two-way splitter loss @ POE = 3.5 dB

If the distance to the premises is 100' + 50' of interior wiring to the set-top or modem, that is a total of 150' of drop loss that must be accounted for in our drop engineering. 6.55 dB loss X 1.5 (150') = 9.82 dB. If we add a two-way house splitter (3.5 dB of loss), the total loss is (9.82 + 3.5) = 13.32 dB.

Calculation example for premises (D4 modem) at 1.8 GHz:

RG6 drop cable loss / 100' = 8.49 dB

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

If the distance to the premises is 100' + 50' of interior wiring to the set-top or modem, that is a total of 150' of drop loss that must be accounted for in our drop engineering. 8.49 dB loss X 1.5 (150') = 12.73 dB.

With 1.8 GHz upgrades, we focus on DOCSIS 4.0 modems, also known as D4 gateway. In theory, these gateways feature point-to-point wireless transmission of RF signals via Wi-Fi to internal premises devices. Internal premises wiring may still be used in our calculations only because we don't know exactly where the gateway will be placed within a premises. They are often called point-of-entry (POE) devices. However, they may not always be placed right at the POE. They may be placed elsewhere in the premises. Splitters can be eliminated because they are no longer needed. If they continue to be used at the premises, the additional loss must also be accounted for.

An operator must know what the minimum and maximum modem receive (RX) levels are for certain DOCSIS generations of modems that reside in their networks or, at the very least, what modem RX design levels specifications have been used in the past as a guideline that need to be maintained in a 1.8 GHz upgrade. The idea is to prevent truck rolls and customer complaints during and after the 1.8 GHz upgrade.

Using the two drop calculation examples given above, we can show them below in Table 2:

Table 2 - Example of Service Drop Engineering

Service Drop Engineering		
MHz	1794	1002
RG6 Loss / 150' (dB)	12.73	9.82
Loss 2/W Splitter (dB)	N/A	3.5
Total Drop Loss (dB)	12.73	13.32

The next items to define before we can build our tap output levels specifications are the legacy modem(s) RF levels input (RX) requirements that still need to be met in an upgrade and the proposed RX levels needed for the DOCSIS 4.0 gateway. At the time of writing this document, many operators are working diligently to define what these levels are. The sample levels shown below are not specific to any operator and are not final numbers. They are just placeholders to illustrate an example of the type of information that needs to be decided upon to build design specifications for tap output RF levels based on service drop engineering.

Table 3 - Example of Modem / Gateway RX & TX Levels

CPE / MODEM SPECIFICATIONS MIN/MAX FOR CALCULATED RX & TX									
MHz	1794	1794	396	396	1218/1002	1218/1002	204	204	RX
CPE / MODEM	Min RX dBmV	Max RX dBmV	Min RX dBmV	Max RX dBmV	Min RX dBmV	Max RX dBmV	Min RX dBmV	Max RX dBmV	Min RX dBmV
DOCSIS 3.1	N/A	N/A	N/A	N/A	-8.0	12.0	30.0	48.0	-8.0
DOCSIS 4.0	-6.0	15.0	30.0	49.0	N/A	N/A	N/A	N/A	-10.0
	21 dB Window		19 dB Window		20 dB Window		18 dB Window		

From these exercises we can now experiment with tap output levels needed to satisfy legacy DOCSIS modems as well as DOCSIS 4.0 gateways. It is simple: whatever min/max modem RX levels are decided

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

upon for legacy modems, as well as DOCSIS 4.0, will determine the target tap output level range to maintain in our design to assure we are delivering within those tolerances.

Table 4 illustrates the math. As before, these are not levels or recommendations specific to any operator. They are just exemplary placeholders to show how the math works.

Table 4 - Example of Tap Output Levels and Modem / Gateway RX Levels

Defining Tap Output Levels		
MHz	1794	1002
RG6 Loss / 150' (dB)	12.73	9.82
Loss 2/W Splitter (dB)	N/A	3.5
Total Drop Loss (dB)	12.73	13.32
Tap Output dBmV	13.00	12.00
Minus Drop Loss = (Modem RX)	*0.27*	*-1.32*
Tap Output dBmV	10.00	9.00
Minus Drop Loss = (Modem RX)	*-2.73*	*-4.32*

The above modem RX examples would be typical numbers that should keep legacy, as well as upcoming DOCSIS 4.0 modems, within their desirable RX range.

For design purposes, *maximum* RX levels are generally in the +12.0 to +15.0 dBmV range. Taps immediately spliced to an active or hot tap with potentially higher than normal output levels should be monitored to be sure they are not causing excessive RX levels at the modem. The drop engineering we created is necessary for calculating tap output levels to be sure we will not exceed the maximum or drop below the minimum RX for particular modems.

Remedies for controlling excessive high tap output levels and modem RX vary across operators. One method of controlling excessively high tap spigot outputs is to increase the tap's value which will lower the drop port output dBmV. Using this method will also require the designer to be aware of upstream tap spigot input levels to be sure the higher tap face-plate value is not causing the modem to transmit out of its intended range. The higher the upstream loss in a network, the higher the modem will transmit to overcome the loss. Another method is to install internal plugin tap simulators which attenuate the tap spigot RF output levels on the downstream. These do not affect the downstream or upstream tap hardline (distribution cable) ports. They do however affect the upstream frequencies RF input from the modem through the tap drop spigots. As stated, care must be taken to ensure modems are not attempting to transmit out of their intended range.

5. Performance

With the increase in total bandwidth to 1.8 GHz, we now have a much wider spectrum to test and validate acceptable input and output RF signal levels in which to operate our amplifiers. As mentioned prior, in an OSP upgrade from previous legacy bandwidths out to 1.8 GHz, we must identify our network challenges. Those challenges can be summarized in a few words (e.g., **passive loss and coaxial cable attenuation at higher frequencies**). Amplifiers in a 1.8 GHz network upgrade will run at higher RF output levels than their previous counterparts. But we must be aware of the performance and TCP rules around establishing the new higher levels. The below checkmarks exemplify certain critical performance indicators to be mindful of.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

	Reference
✓ TCP – Total Composite Power	dBmV
✓ NPR – Noise Power Ratio	dB
✓ CCN – Carrier to Composite Noise	dB
✓ CIN – Carrier to Intermodulation Noise	dB
✓ CTN – Carrier to Thermal Noise	dB
✓ CNR – Carrier to Noise Ratio	dB
✓ MER – Modulated Error Ratio	dB
✓ BER – Bit Error Ratio	dB

Figure 4 – Performance Indicators

6. Step-Down: TCP & Channel Power

The term "**step-down**" is in reference to an absolute decibel amount below the expected operational output of an active device at a specific frequency(s). Step-down is needed in high output 1.8 GHz modules to limit the TCP (Total Composite Power) of these modules and prevent the overdrive of RF amplifiers and unacceptable distortion levels. TCP is the combined amount of power of all signals within a specific frequency range (bandwidth).

It is vital in 1.8 GHz UHS (Ultra-high Split) design to maintain legacy cable modems as well as the new DOCSIS 4.0 gateway RX and TX specifications levels. The main thing to keep in mind is, yes, we are designing for greater total bandwidth and a higher split order in 1.8 GHz design, but we cannot ignore or cause out-of-spec situations for our legacy customers that still may be on DOCSIS 3.0 or 3.1 modems. When a node area is physically upgraded in the OSP it will take some time, possibly months or more, to transfer hundreds of customers in a given node area to DOCSIS 4.0 gateways. For this reason, the legacy customer premise equipment must continue to function. Receive (RX) and Transmit (TX) levels MUST be maintained and work for multiple generations of CPE.

1.8 GHz UHS actives have a significant increase in power over any previous generations of active devices. This increase brings with it challenges. While we need to harness the power, we really need to ***manage*** the power. This is done through step-down.

The goal is to utilize this higher power while maintaining distortion and performance acceptance levels. To exceed performance limits would mean introducing unwanted distortion into the network. Through a cascade of amplification, distortion only increases. It will never improve until the source or cause of the problem is changed. Without step-down, extending RF levels linearly out to 1.8 GHz could result in extremely high TCP levels exceeding 74 dBmV out of the RF hybrids. This is well beyond today's RF hybrid and amplifier technology.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

DOCSIS 4.0 introduced the concept of RF level adjustments to manage TCP through signal level reductions to help improve amplifier performance. Operator "A" might use a single 6 dB step-down at 1.0 GHz OR 1.2 GHz. Operator "B" might use a 3 dB step-down at 1.0 GHz and a second 3 dB step-down at 1410 MHz, and so on.

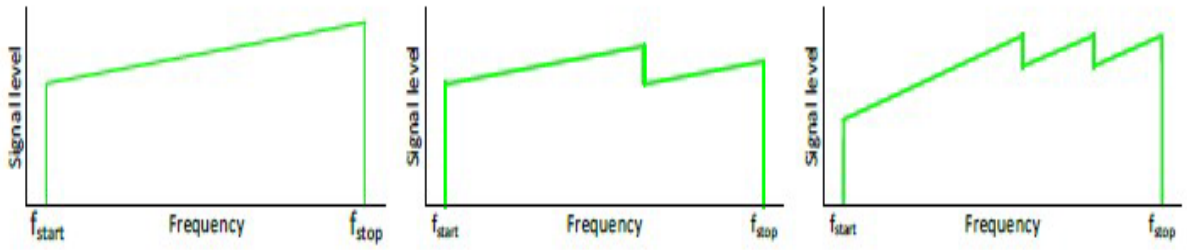


Figure 5 – Step-Down Frequency(s): Constant, Single, Multiple

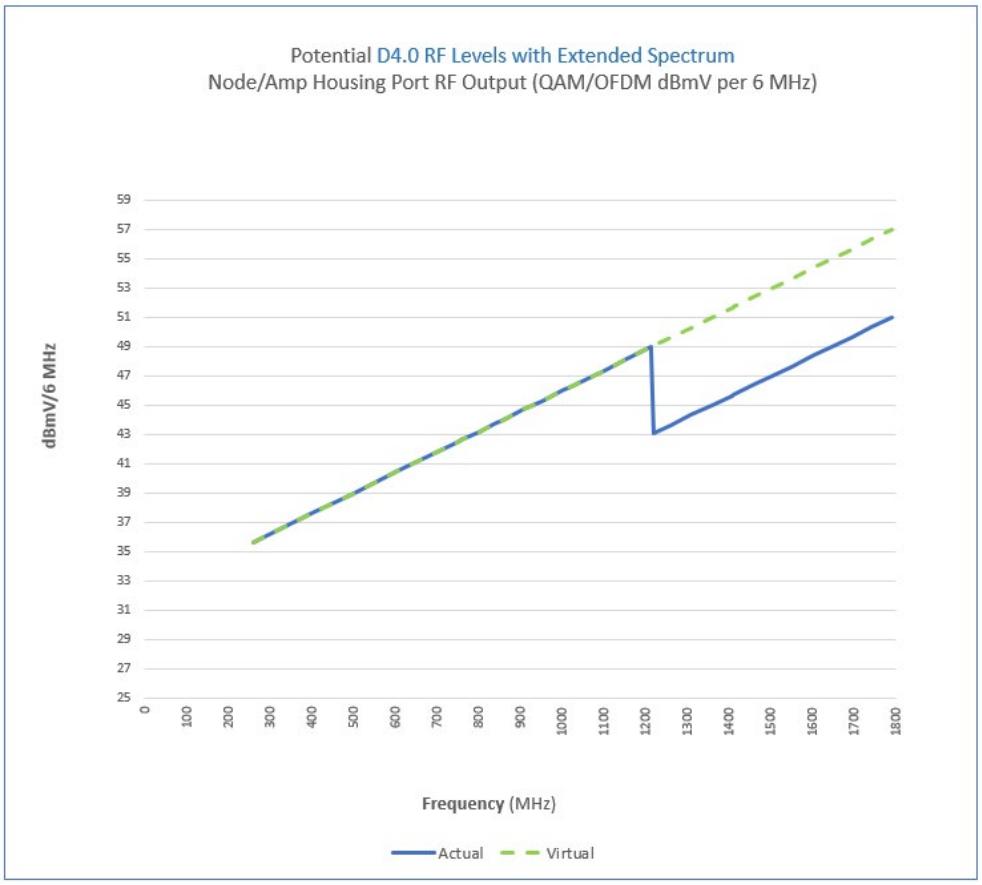


Figure 6 – Step-Down, Virtual- and Actual-Tilt, & Levels

Step-down occurs at the RF source right at the RPD (Node). It is already reduced as it travels downstream to the actives. In Figure 6, what we call "Virtual Tilt and Levels" is the dashed green line. "Actual Tilt and Levels" is the solid blue line. In this example there is a 6 dB step-down beginning at 1221 MHz and extending to 1800 MHz.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Virtual tilt (green line) is the RF power we would use in design software levels specification files. This is the amount of RF power the software will use to perform its calculations as we space and locate actives downstream in the network on trunk or express and extract tap values in our distribution.

In design, we don't subtract the step-down amount until we leave the tap. The RF power to 1.8 GHz is a straight, dashed, virtual line shown above in Figure 5. When we deduct the step-down amount, we begin at the lowest frequency where step-down is applied (actual). This deduction is subtracted from the tap output level from any frequency that is affected by the step-down.

Example: If we have a single 6 dB step-down starting at 1221 MHz, that means we have to deduct 6 dB starting at 1221 MHz and from all frequencies leading up to 1800 MHz. The same applies if we have multiple frequencies where step-down occurs. We would subtract the amount of step-down at the first lowest frequency where it is applied, then add that to the next highest frequency. This is now an accumulative amount of step-down all the way to 1800 MHz. So, two 3 dB step-downs would accumulate to 6 dB of total step-down at 1800 MHz.

At the operator level, calculating the exact frequency(s) and how much step-down to apply should be performed by those experienced in performance calculations. The step-down frequency(s) and step-down amount (dB) directly affect TCP dBmV, as shown in Table 5 and Table 6 below.

In addition, there is the TCP (dBmV) of the hybrid chips themselves, as well as the TCP at the active's housing port to be concerned with. Know that there is internal split loss from the internal power amp to the housing port. This usually ranges between 3 and 4 dB.

In Table 5, we are showing a 4 dB differential between the hybrid and the port. Note: Depending on variables, operators may have differing approaches and results when determining their TCP limits, forward tilt and operational levels.

Table 5 - Channel Power 61 dBmV and TCP Calculations

Channel Power @ 1791 MHz	Channel Power @ 1791 MHz	Channel Power @ 1791 MHz	Channel Power @ 1791 MHz
61 dBmV	61 dBmV	61 dBmV	61 dBmV
Tilt 57-1791 MHz	Tilt 57-1791 MHz	Tilt 57-1791 MHz	Tilt 57-1791 MHz
24.2 dB	24.2 dB	24.2 dB	24.2 dB
6 dB Step @ 1002 MHz	6 dB Step @ 1221 MHz	9 dB Step @ 1221 MHz	6 dB Step @ 1410 MHz
TCP @ Hybrid = 72.77 dBmV	TCP @ Hybrid = 73.46 dBmV	TCP @ Hybrid = 71.95 dBmV	TCP @ Hybrid = 74.53 dBmV
TCP @ Port = 68.77 dBmV	TCP @ Port = 69.46 dBmV	TCP @ Port = 67.95 dBmV	TCP @ Port = 70.53 dBmV

Looking at Table 5, the channel power at 1791 MHz is 61 dBmV. This is a virtual level out of the hybrid. This level will directly affect the virtual RF output levels we will be setting our 1.8 GHz actives output levels to operate in our design specifications files (actives' RF output levels are not the same as channel power). At the present time, manufacturers are indicating the amount of TCP that is tolerable at the port is between 68 and 70 dBmV.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Looking at Table 6, we have lowered the channel power by (1) dB to 60 dBmV. This also lowers our TCP. Our tilt of 24.2 dB remains the same. When we adjust channel power +/- we are raising or lowering the overall power levels, but not affecting the total tilt between the high and low forward frequencies. The ratio between them remains the same.

Table 6 - Channel Power 60 dBmV and TCP Calculations

Channel Power @ 1791 MHz	Channel Power @ 1791 MHz	Channel Power @ 1791 MHz	Channel Power @ 1791 MHz
60 dBmV	60 dBmV	60 dBmV	60 dBmV
Tilt 57-1791 MHz	Tilt 57-1791 MHz	Tilt 57-1791 MHz	Tilt 57-1791 MHz
24.2 dB	24.2 dB	24.2 dB	24.2 dB
6 dB Step @ 1002 MHz	6 dB Step @ 1221 MHz	9 dB Step @ 1221 MHz	6 dB Step @ 1410 MHz
TCP @ Hybrid = 71.77 dBmV	TCP @ Hybrid = 72.46 dBmV	TCP @ Hybrid = 70.95 dBmV	TCP @ Hybrid = 73.53 dBmV
TCP @ Port = 67.77 dBmV	TCP @ Port = 68.46 dBmV	TCP @ Port = 66.95 dBmV	TCP @ Port = 69.53 dBmV

7. Booster Type Actives for 1.8 GHz Design

The concept of booster type amplification is not new. Booster actives have been around since the 1970s. The use case for them was the same as it is today. They simply make up for a small amount of gain and cable footage allowing the next downstream active to realize its required RF input level and avoid relocation. In earlier years, cascades were longer and not as much of a concern as they are in today's modern networks where performance and distortion can add up rather quickly due to higher powered amplification and much greater channel loading.

Manufacturers of 1.8 GHz actives are still working through a plethora of ideas and concepts around what they perceive to be the best, most cost-effective booster product they can build given the obstacles it must overcome in 1.8 GHz networks. The devices must be:

- Cost effective (much less than a line extender)
- Smaller in size (not much larger than a typical line splitter or DC)
- Simple to install and set up (nowhere near the complexity of a conventional active)
- Small amount of gain, 11 to 15 dB (just enough to do the job)
- Good on performance (will contribute some, but minimal performance degradation)
- Power efficient (will not add significantly to current loads)
- Minimal bells and whistles (avoid complexity, keeping the price point as low as possible)
- Flexible in terms of where it can be placed in the network to deliver, but adds minimal noise contributions

Creating boosters in design software is much different than that of conventional actives. Conventional actives are completely predictable in terms of performance contributions along with their required RF input levels and fixed RF output levels. Boosters are none of these things. They are a complete moving target dependent upon where they are placed in the network and what the RF level inputs happen to be at that location. For this reason, the RF outputs will vary. Boosters are simply: RF input + Gain = RF output. This works the same on the downstream and upstream, subject to the gain in each direction. Boosters typically will have 2x to 4x the gain on the downstream as they do on the upstream. The gain is generally

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

fixed. In some units there can be attenuation options. Some also have fixed or variable equalization. But in general, they are much simpler than a line extender.

Boosters will typically amplify the entire downstream and upstream spectrum, similar to conventional actives, but with much less gain. As stated, booster gain is typically between 11 and 15 dB at 1.8 GHz and dropping significantly as we go lower in frequency.

Design managers will have to train their designers on *how, when,* and *where* to deploy boosters. They cannot just be placed anywhere in a network with regard to maintaining performance. They are not intended to be used for every RF signal short-fall or design problem. A network can quickly get into trouble if too many boosters are used incorrectly, adding to cascade depth, performance degradation and power loads.

8. Design Modeling Legacy OSP With 1.8 GHz Proposed Upgrades

There must be careful coordination and synchronization between calculating TCP and performance measurements along with proposed RF network design levels and tilts. Depending upon current legacy plant conditions and previous design practices, proposed 1.8 GHz upgrades can go smoothly and stay within a projected budget, or be problematic and present significant cost-over-runs.

Once preliminary design specifications are created based on acceptable TCP and other performance indicators, network design modeling should be performed on actual legacy nodes or networks. This is a great way to test the design specifications for their effectiveness as well as present the opportunity to adjust them where needed.

The effectiveness of the design outcome in terms of making spacing of previous actives and providing service taps with minimal additional amplification are key to keeping construction, materials and equipment costs as low as possible.

Module drop in upgrades are the most common type of network expansion that has been the standard protocol for the past 25 or more years. These past drop in upgrades were somewhat simpler and easier to create design specifications for because the bandwidth expansions were generally 200 MHz or less.

Example Past Bandwidth Increases:

- 450 MHz to 550 MHz
- 550 MHz to 750 MHz
- 750 MHz to 870 MHz
- 870 MHz to 1002 MHz
- 1002 MHz to 1218 MHz

This coupled with the fact that electronics manufacturers increased the gain in the higher bandwidth amplifiers helped to ensure enough gain for an upgrade expansion, as shown previously in Table 1, to maintain existing actives spacing. As long as the legacy housings could accommodate the newer modules drop in, very little needed to change in the physical plant to facilitate this type of upgrade budget model. Construction and splicing costs could be kept to a minimum.

The proposed upgrade from a 1.0 GHz low- or mid-split system to a 1.8 GHz high-split or ultra-high-split system introduces certain facts that must be addressed as discussed herein. In this event, the downstream is an immediate 800 MHz increase in total bandwidth (1.0 GHz to 1.8 GHz). The upstream has its

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

challenges as well, increasing from 42 MHz or 85 MHz to 204 MHz or 396 MHz. Again, higher bandwidth, greater RF losses.

Rebuilds generally happen when much of the existing coax needs to be replaced for various reasons. The outside plant may need significant plant extensions or additional coax added due to a newer architecture, such as going from an older Trunk and Bridger network to all Express architecture. Or from lengthy cascaded older networks to reduced Node + 0, Node + 2, Node + 4; or even going to N+6 would be a cascade reduction in some older networks.

Working through the gain and cable spacing examples shown previously will greatly help define successful and cost effective 1.8 GHz design and deployment. An operator must make decisions and budget accordingly regarding a complete rebuild or network upgrade. The differentials between these cost models are substantial.

9. Performing Network Due Diligence Ahead of Budgeting

No matter how much OSP due diligence is performed ahead of an upgrade or rebuild, something is bound to remain an unknown until it is known. Below are some of the more common items for consideration that can affect bandwidth expansions and costs:

- An operator must decide if their current design and plant mapping records are recent or accurate enough to perform a design upgrade to 1.8 GHz. Be mindful that if outside plant conditions have changed enough over time and have not been updated or reflected in the mapping records, it may be more difficult and costly to make certain 1.8 GHz design adjustments once construction is underway. A new field walkout or survey may be needed prior to beginning the design process.
- Limitations in 1.8 GHz active cascades. Can they meet the existing amplifier cascades? Is reducing lengthy cascades in older networks to Node + 0, Node + 2, Node + 4 or even N+6 part of the upgrade requirement? It is not recommended to exceed five to six 1.8 GHz actives in a cascade due to performance issues. This should be discussed with your chosen vendor(s).
- Prevent or limit node splitting as much as possible. In addition to the cost of the nodes themselves, node splitting generally means adding or extending new fiber optic cabling to feed them.
- Are high homes passed or service group counts in need of reductions or balancing to prevent future network bottlenecks or choke points?
- Avoid the relocation of current active locations. Moving active locations can be costly due to de-splicing and re-splicing. In addition, potential re-design at the new location(s), as well as the old, may now be needed and cause re-splicing and or equipment revisions of downstream legs.
- In many cases the chosen 1.8 GHz actives for the upgrade are duplicates of the current actives in terms of their type, configuration, and porting. Avoid re-splicing of current coaxial cabling configurations at actives locations to the greatest extent possible. Try to design by emulating the current cabling configurations while adhering to the new network design specifications.
- Adjust tap levels requirements using similar logic as illustrated in the drop engineering exercises shown previously. If it is discovered later that the modems are having RX or TX issues, costly design and construction revisions may become necessary.
- Leverage the existing outside plant coaxial cables to the greatest extent possible to avoid the placement of new or additional cabling.
- Be sure to use the manufacturer's powering specifications from their 1.8 GHz actives equipment specifications. Do not assume they are the same as previous generations of the same actives type. The newer 1.8 GHz actives in many cases use less power than previous generations of actives and may save the need to place additional power supplies.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Create a set of design rules with respect to an upgrade. This will help prevent designers from making assumptions due to lack of guidance. The final designs will be much more consistent and cost effective if the rules are followed and assure best network performance.
- 1.8 GHz bandpass lab testing of legacy connectors, pin seizure mechanisms and splices has proven to be a factor. The manufacturers have addressed these issues within new amplifier and passive device housings, including backwards compatibility replacements in some older equipment.

If, however, you have certain types of splices, such as splice blocks, pin connectors or any equipment that is connected and intended to continue to use as is without revisiting them, to assure 1.8 GHz bandpass, you may find that they do not pass beyond 1.0 GHz. Or there can be a significant roll-off due to the connector. This can be a noteworthy cost increase if not considered or allocated for in budget planning.

10. Conclusion

This paper presents some of the more crucial and sometimes overlooked concerns regarding the creation of RF design specifications with respect to upcoming bandwidth expansions. It focuses on the approach from a pragmatic point of view using simple, real-world examples and calculations. Challenges and hurdles exist in any planned upgrade and there are a multitude of factors to consider. The greater the proposed bandwidth increase, the greater the challenges.

There are also numerous physical OSP challenges discussed herein that must be identified ahead of a planned upgrade and corresponding budgeting. The physical revisions needed depend on the design. Physical plant revisions or additions can contribute to the highest single budgetary costs.

In closing, the main message is this: we must do our due diligence. Leveraging the existing infrastructure and overall configuration of the OSP to be upgraded is to our biggest advantage and will help to ensure the least costly budget.

The creation of effective but accurate RF design specifications that honor performance is key to healthy bi-directional networks that can deliver the products and services intended.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

bps	bits per second
CPE	customer premises equipment
dB	decibel
dBmV	decibels relative to one millivolt
DC	directional coupler
DOCSIS	data over cable service interface specification
DS	downstream
EEQ	electronic equalizer
gbps	giga bits per second
GHz	gigahertz
HFC	hybrid fiber coax
MER	modulated error ratio
MHz	megahertz
NPR	noise power ratio
OFDM	orthogonal frequency division multiplexing
OSP	outside plant
POE	point of entry
QAM	quadrature amplitude modulation
R&D	research and development
RF	radio frequency
RX	receive
TCP	total composite power
TX	transmit
UHS	ultra-high split
US	upstream
W	watts

Bibliography & References

How Can NPR Improve Future 1.8 GHz Extended Spectrum Amplifier Characterization, Diana Linton; Broadband Library: https://broadbandlibrary.com/how-can-npr-improve-future-1-8-ghz-extended-spectrum-amplifier-characterization/

Signal-to-noise ratio (S/N or SNR), Robert Sheldon and John Burke: https://www.techtarget.com/searchnetworking/definition/signal-to-noise-ratio

Understanding Band Splits in Two Way Networks, Ron Hranac; Broadband Library: https://broadbandlibrary.com/understanding-band-splits-in-two-way-networks/

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Next Generation Neighbor Interference Prediction Tools

From iHAT to nHAT

A Technical Paper prepared for SCTE by

John Chrostowski
Executive Director Connect Engineering
Comcast
John_Chrostowski@comcast.com

Richard Primerano
Principal Engineer
Comcast
Richard_Primerano@Comast.com

Kang Lin
Manager, Quality Assurance
Comcast
Kang_Lin@Comcast.com

Jay Zhu
Principal Software Engineer
Comcast
Jay_Zhu@Comcast.com

Javed Nazim
Network Engineering
Comcast
Javed_Nazim@Comcast.com

Jon-En Wang
Comcast
Senior Director, Quality Assurance
Jon-En_Wang@cable.Comcast.com

Dustin Tracy
Comcast
Senior Broadband Engineer
Dustin_Tracy@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Mid-Split Adjacent Channel Interference Background 4
3. iHAT Background 6
   3.1. Summary of Comcast Mid-Split Deployments and iHAT Results 8
4. iHAT Improvements and Optimization 9
   4.1. Field Test Results 9
   4.2. STB Front End Concerns 9
   4.3. 2nd Harmonic Distortion 10
   4.4. OUDP Pulse 11
5. Summing up the iHAT Improvements 13
6. Moving to FDX 13
7. Determining Interference Thresholds 15
   7.1. STB 15
      7.1.1. Test Setup and Parameters 15
      7.1.2. Test Results 17
   7.2. Cable Modems 19
      7.2.1. Test Setup and Automation 19
      7.2.2. Initial Test Results 21
8. FDX Neighbor Interference Implementation and nHAT 29
9. Conclusion 34
10. Abbreviations 35

Bibliography & References 36

List of Figures

Title **Page Number**

Figure 1 - ACI In-Home Interference 5
Figure 2 - ACI Example Power Calculation 5
Figure 3 - Typical Splitter Isolation Lab Measurement 6
Figure 4 - Upstream Capture Showing OUDP Burst 7
Figure 5 - Set-top Box Downstream FBC with OUDP and Downstream Video 8
Figure 6 - iHAT Summary Dashboard 8
Figure 7 - MER Degradation with 4 MHz OUDP Pulse 10
Figure 8 - Codeword Error with 4 MHz OUDP Pulse 10
Figure 9 - Set-top Box Field Full Band Capture Measurement 12
Figure 10 - 20 MHz OUDP Pulse Measurement Example 13
Figure 11 - ACI Power Addition vs. OFDMA and FDX Channels 14
Figure 12 - FDX Neighbor Interference Paths 14
Figure 13 - Test Setup for Adjacent Channel Interference on STBs 15
Figure 14 - Power vs Time Capture of Waveform with 10 ms Period and 50% Duty Cycle 16
Figure 15 - FDX Upstream and Downstream Channel Lineup scenarios for testing neighbor interference scenarios 17
Figure 16 - FDX Neighbor Interference CM Test Setup 20
Figure 17 - 288 MHz FDX: CM-2 SC-QAM Average MER changes (4 ms cycle period) 24

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 18 - 288 MHz FDX: CM-2 CER Over Time (4 ms cycle period, 10% duty cycle, 5 dB PSD Difference) 25
Figure 19 - nHAT Flow Diagram 31
Figure 20 - Example ROCI Topology with FDX Customers 32
Figure 21 - ROCI Analysis of Neighbor Interference Levels 32
Figure 22 - nHAT ROCI Analysis ACI Delta Same Tap 33
Figure 23 – nHAT ROCI Analysis ACI Delta Adjacent Tap 34

List of Tables

Title **Page Number**

Table 1 – Setop Field Video Tiling vs ACI Frequency 9
Table 2 – Setop Lab Video Tiling vs ACI Frequency 11
Table 3 - Period and Duty Cycle for Upstream Burst Signals 16
Table 4 - TCP Delta Threshold for Map 2 and DS Video at 495 MHz 18
Table 5 - TCP Delta for Map 2 for Various Video Channels 18
Table 6 - TCP Threshold Comparison for Map 2 and Map 3 (DS Video 495 MHz) 18
Table 7 - CM Neighbor Interference Test Parameters 20
Table 8 - 288 MHz FDX: Downstream Spectrum Configuration 22
Table 9 - 288 MHz FDX: 4 ms Cycle Period Test Results 22
Table 10 - 288 MHz FDX: 10 ms Cycle Period Test Results 22
Table 11 - 288 MHz FDX: 70 ms Cycle Period Test Results 23
Table 12 - 288 MHz FDX: 10% Duty Cycle Created the Most Impact (CM 2 Example, 4 ms) 24
Table 13 - 384 MHz FDX: Downstream Spectrum Configuration 27
Table 14 - 384 MHz FDX: 4 ms Cycle Period Test Results 28
Table 15 - 384 MHz FDX: 10 ms Cycle Period Test Results 28
Table 16 - 384 MHz FDX: 70 ms Cycle Period Test Results 29

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

During the planning for activation of the mid-split spectrum on the cable plant, adjacent channel interference (ACI) between the mid-split upstream spectrum and legacy devices on the network that use the same spectrum on the downstream was anticipated. Extensive testing was conducted on DOCSIS® and set-top devices to understand the impact on the performance of these devices and the customer-impacting thresholds. Based on this testing, tools have been developed to identify this interference and prevent negative customer experiences when the mid-split spectrum is activated.

With millions of mid-split activations in the field, we have been able to see how well the tools that look for interference perform. We also have gained better insight into how ACI impacts the RF front end and performance of the CPE devices and how in-home wiring can have a large impact on ACI.

As the industry moves to DOCSIS 4.0 technology, ACI involves not just co-located in-home devices but also neighboring devices. In this paper, we will review our current prediction tools for ACI, field and lab test results of CPE devices in the presence of ACI, and improvements which have been made to the prediction tools. Most important will be how these tools and lessons learned can be adapted to understand and predict neighbor interference to ensure an optimal customer experience as the upstream spectrum is expanded further in DOCSIS 4.0 technology.

2. Mid-Split Adjacent Channel Interference Background

As the industry moves to increase bandwidth with the full capabilities of DOCSIS 3.1 and 4.0 devices, these devices will be deployed simultaneously with older, less capable devices: cable modems, gateways, and set-top boxes. Different revisions of these devices have different RF filtering and front ends, and it is important to ensure interoperability with differing upstream channel loads and RF bandwidths.

In mid-split systems, set-top boxes are more susceptible to adjacent channel interference (ACI). The front end of the set-top boxes typically has a lower downstream band edge of 54 MHz, which overlaps with the mid-split upstream frequency band of 5-85 MHz. With the activation of the mid-split spectrum, energy from the Mid-Split cable modem/gateway at 54-85 MHz can make its way across the in-home splitting network and impact the operation of the set-top boxes. This is shown in Figure 1. Another contribution to the interference problem is that the upstream transmission from the cable modem is at a much higher power level compared to the downstream receive level at the CPE devices.

Typically, a high transmit level is the result of high upstream path loss; these conditions can also lead to low downstream receive levels. A change in path loss results in a 2-to-1 relationship between the interfering signal level at the set-top box vs. the downstream receive level at the set-top box. As an example, if a cable modem is transmitting at its maximum power of 65 dBmV, with 25 dB splitter isolation, the total composite power of the interfering signal level at the set-top box is 40 dBmV. With a nominal input level of -10 dBmV/6 MHz into the set-top box, which equates to a total power of 11.7 dBmV, the ACI is 29.3 dB higher than the downstream total power. This is shown in Figure 2. This level difference, plus the bursty time domain nature of this interfering signal, can impact the AGC of the set-top box along with overdriving the front end, causing distortion.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

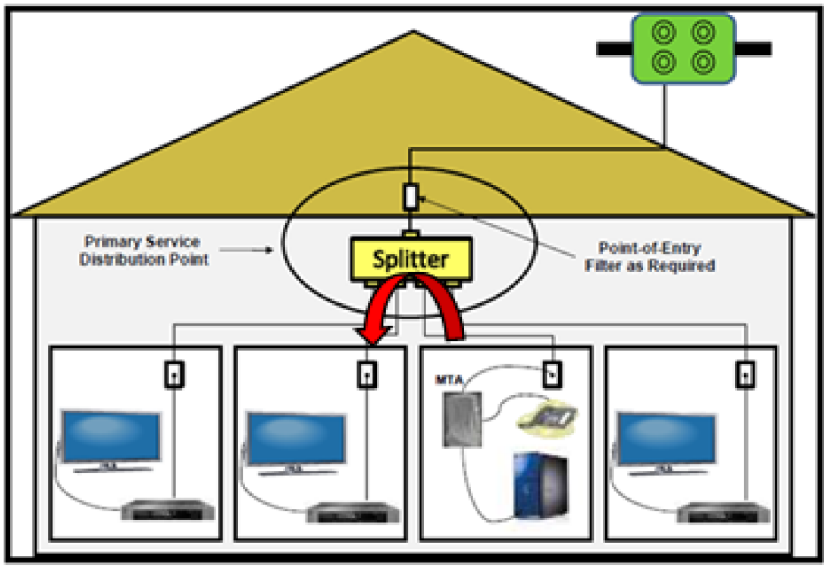


Figure 1 - ACI In-Home Interference

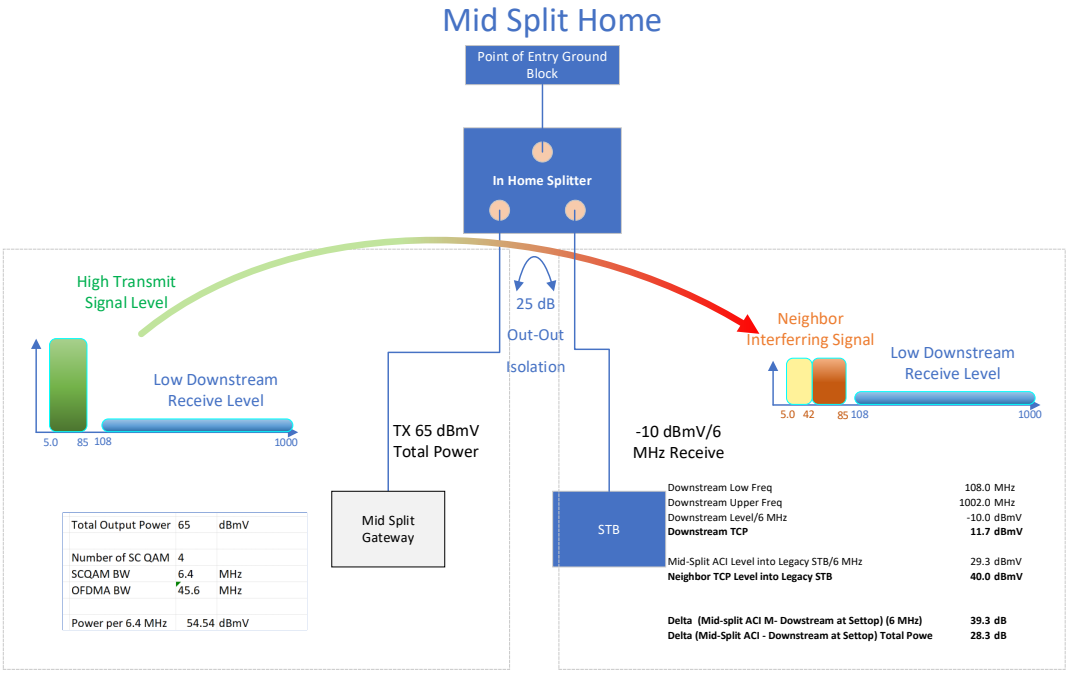


Figure 2 - ACI Example Power Calculation

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. iHAT Background

To be proactive and minimize any negative customer experience during mid-split upgrades and activation, Comcast developed an In-Home Assessment Test tool called iHAT. This tool has two main functions. The first is to check and validate that a mid-split device can use the mid-split spectrum and OFDMA channel on the upstream. Second is an ACI video interference test for any set-top boxes in the same home as the mid-split cable modem or gateway.

The test to validate the OFDMA channel is straightforward, with a check if the OFDMA channel is bonded. The test to check for video interference is more involved. The original concept for iHAT video interference was a brute force test, which consisted of running a speed test and simultaneously monitoring the SNR of the adjacent set-top box. This had several issues. First, the speed test needed to use the entire upstream spectrum. Second, the speed test had to be long enough to affect the set-top box SNR. Third, in order to run a speed test and use the entire upstream spectrum, a specific boot file had to be loaded onto the cable modem and then reverted to the customer boot file after the test. Fourth, the STB had to be tuned to a specific channel to monitor the SNR. This was customer-impacting in many ways. Loading the boot files required a cable modem reboot. Using the entire upstream spectrum to check for interference had the possibility of impacting high-speed data customers, and changing the STB tuner to a specific channel can impact a customer's viewing experience.

A better way to measure video interference had to be developed. DOCSIS 3.1 technology has an OFDMA Upstream Data Profile (OUDP) functionality which allows the CMTS to schedule an upstream OFDMA burst which can be used for measurements. Comcast had developed and used OUDP signaling to facilitate upstream leakage measurements for high-split applications. OUDP signaling coupled with downstream spectrum capture on the set-top boxes allows adjacent interference measurements without impacting the customer's experience. The OUDP burst originally was 1.6 MHz wide at 80 MHz. The narrow bandwidth was chosen to minimize adjacent interference and minimize unwanted power at the set-top box. 80 MHz was chosen to correspond to the lowest isolation point in the spectrum. RF splitters typically have lower isolation as frequency is increased, and 80 MHz is near the worst isolation point of the splitter in the mid-split band. Figure 3 and Figure 4 illustrate the upstream spectrum with the OUDP burst and typical in-home splitter isolation.

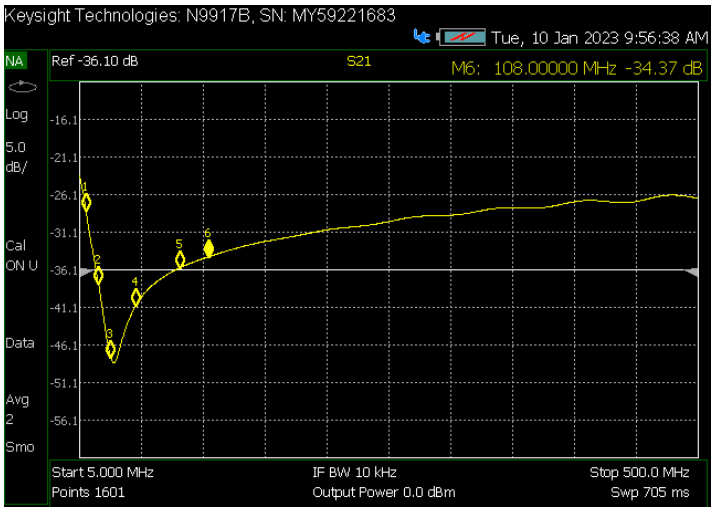


Figure 3 - Typical Splitter Isolation Lab Measurement

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 4 - Upstream Capture Showing OUDP Burst

Figure 4 - Upstream Capture Showing OUDP Burst shows the upstream band with 4-SC QAM channels, OFDMA and OUDP burst. Through extensive testing measuring interference thresholds on set-top boxes in the lab, we have established a threshold set for pass/fail for interference, which the customer sees as video tiling. From the set-top box full band capture, the delta between the OUDP burst, and downstream spectrum is measured and compared against the pre-determined threshold. If the delta exceeds the threshold, the device is steered to the 4 SC-QAM sub-split bonding group via a DBC command, and the OFDMA channel is not used. If the delta is lower than the threshold, the device remains with the OFDMA channel active. In the example in Figure 5 - Set-top Box Downstream FBC with OUDP and Downstream Video, the ACI delta is ~ 25 dB, which is very high and would trigger the device to be steered to sub-split.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

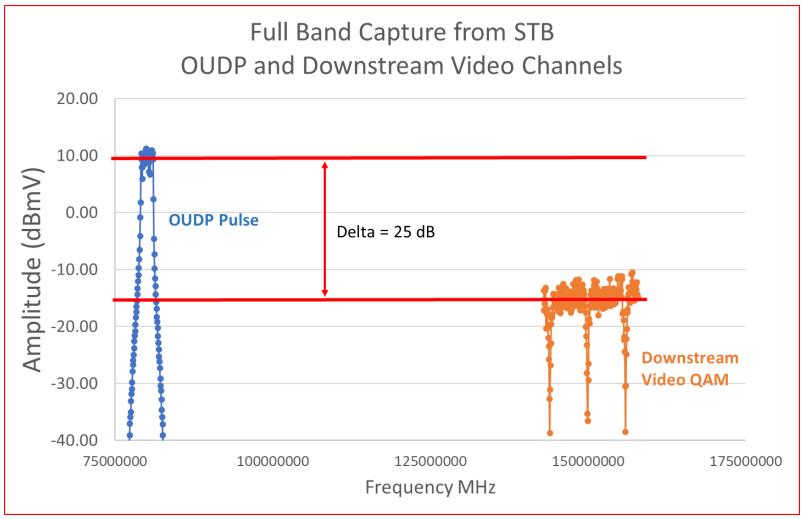


Figure 5 - Set-top Box Downstream FBC with OUDP and Downstream Video

3.1. Summary of Comcast Mid-Split Deployments and iHAT Results

Comcast has a well-defined automated process for activating mid-split. Over the past year, Comcast has tested over 2.4 million mid-split devices, and iHAT is run for every mid-split RPD activation with results available in several dashboards.

The iHAT summary dashboard for all these tests is shown in Figure 6 - iHAT Summary Dashboard. We see:

- 77.4% Pass
- 15.5% OFDMA blocked, (typically in-home drop amps blocking the OFDMA)
- 1.7% Video failures, (based on the OUDP test measurements)
- 3.9% Could not be tested

We also look at CM high transmit power to ensure the devices that do not have enough power to support the extra mid-split spectrum are not forced to use OFDMA.

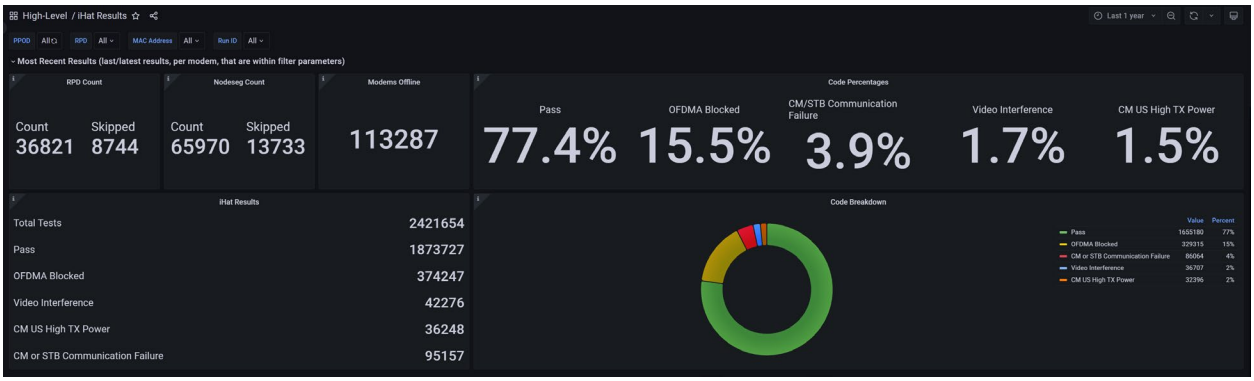


Figure 6 - iHAT Summary Dashboard

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. iHAT Improvements and Optimization

4.1. Field Test Results

Over the course of these 2.4 million deployments, observations, updates, and optimizations have been made to the functionality of iHAT. After reviewing the data and feedback from customers and field technicians, we investigated instances where iHAT was not as accurate as we needed on the video interference test. After evaluating many customer accounts, we identified two main areas for improvement to accuracy.

4.2. STB Front End Concerns

Our original test data for video interference was based on understanding that the burstiness and duty cycle of the interfering signal affected the AGC of the set-top box, which would affect the entire downstream band, and all video channels would be impacted. The data used to determine the thresholds for video interference was taken in the middle of the downstream video band.

Table 1 shows test data from 4 customer accounts that experienced video interference. Three of these accounts used RDK set-top boxes, and one account had a 3rd party video device, which also experienced video interference.

Table 1 – Setop Field Video Tiling vs ACI Frequency

Customer	CM	STB	DS RX dBmV	US TX dBMV	111 MHz	129 MHz	291 MHz	333 MHz
1	RDK D31 modem A	STB model A	-4	45	Y	Y	N	N
2	Retail modem A	Cable Card based video device	0-6	44	Y	Y	N	N
3	RDK D31 modem A	STB model B	-6.5	47.8	Y	Y	Y	Y
4	RDK D31 modem AB	STB model B	-9	50	Y	Y	Y	Y

As can be seen from this data, two accounts showed video interference at lower video channels, and two accounts showed video interference across the video spectrum; this result implied there was more to the interference than just AGC impact on the set-top box. Additional testing was completed in the lab, which confirmed that the interference from the adjacent mid-split spectrum caused a 2nd harmonic distortion, affecting video channels 2X the frequency from the narrow OUDP interfering signal as shown in Figure 4 and Figure 5.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.3. 2nd Harmonic Distortion

Figure 7 - MER Degradation with 4 MHz OUDP Pulse shows the degradation of MER and codeword error rate on the video channels in a STB vs. the input frequency of the interference. In a lab home network configuration, a 4 MHz wide OUDP pulse was generated at different frequencies in a cable modem, and corresponding frequencies of an adjacent set-top box were measured. A 4 MHz pulse was used to validate that only frequencies 2X the OUDP pulse in the video spectrum were impacted.

With an OUDP pulse with a 58.5 MHz center frequency, the entire spectrum is affected somewhat, but 2X the OUDP pulse frequency, 117 MHz, is most impacted. This second order distortion component of the interference continues as the OUDP pulse is cycled from 67.5 through 73.5 MHz. In each instance, the video channel 2X the video frequency is the most impacted. A similar correlation to the affected frequency is measured when looking at the codeword errors. Video frequencies 2X the OUDP pulse frequency were most impacted.

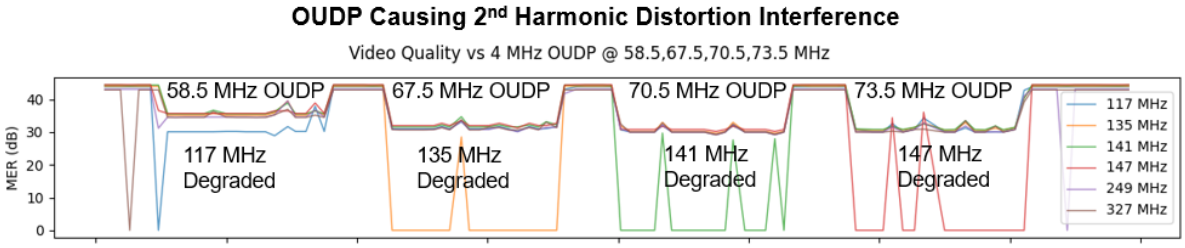


Figure 7 - MER Degradation with 4 MHz OUDP Pulse

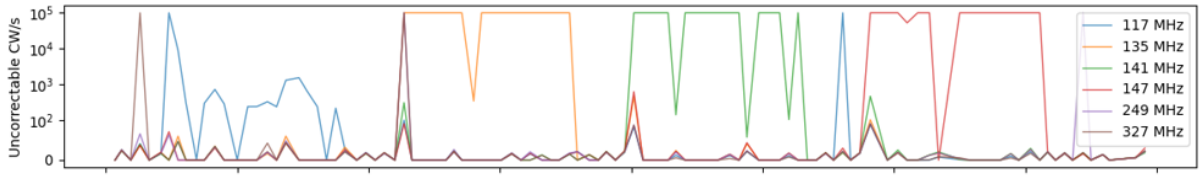


Figure 8 - Codeword Error with 4 MHz OUDP Pulse

In addition to measuring MER and codeword errors, testing was also performed for correlation to actual video tiling. A group of set-top boxes were set up, and the video was monitored, with adjacent channel interference applied at various frequencies. Video tiling was present on channels that were 2X the adjacent channel interference, correlating to the MER and uncorrectable codeword errors seen. This data is detailed in Table 2.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 2 – Setop Lab Video Tiling vs ACI Frequency

Channel Number	Channel Frequency (MHz)	US OFDMA Freq (MHz)	OFDMA Power (dBmV)	STB/TV1 Video Status	STB/TV2 Video Status
800	141	60	59	No Tiling	No Tiling
800	141	70	52	Video Tiling	Video Tiling
800	141	80	59	No Tiling	No Tiling
306	159	60	59	No Tiling	No Tiling
306	159	70	59	No Tiling	No Tiling
306	159	80	50	No Tiling	Video Tiling
306	159	80	54	Video Tiling	Video Tiling

4.4. OUDP Pulse

One of the early updates to iHAT and the video interference test was to ensure that running the test itself caused minimal to no impact on the customer experience. The OUDP pulse used to measure the amount of interference at the adjacent set-top box was narrow in width, with a duration of the minimal amount of time needed to complete an accurate full band capture and measurement at the set-top box. As mentioned previously, the frequency chosen for the interference measurement was chosen for worst-case in-home splitter isolation, which is typically at higher frequencies. In investigating field-related video tiling, emphasis was placed on field measurements of the OUDP pulse and set-top box full band capture. During the field investigation, speed tests were performed with set-top full band captures recorded. Where the isolation of a splitter when measured with a network analyzer is typically at higher frequencies, looking at the full band capture of a speed test in the field, which utilizes the full upstream band, the isolation, and the measured spectrum can be significantly different from a splitter 2-port measurement in the lab. Figure 3 - Typical Splitter Isolation Lab Measurements shows the splitter isolation measurement in the lab. Figure 9 - Set-top Box Field Full Band Capture Measurement shows a full band capture measurement from the field, which in essence, shows the isolation of the in-home network. The red marker is at 80 MHz, and with an OUDP pulse generated at 80 MHz, the level measured on the full band capture is 7 dB lower than the maximum peak of the interfering signal at the setup box.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

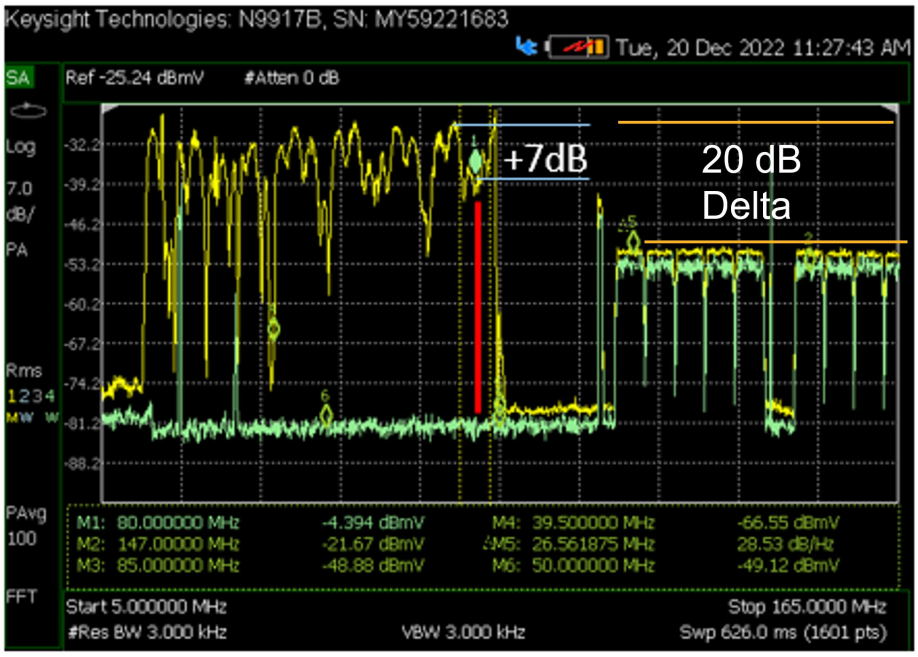


Figure 9 - Set-top Box Field Full Band Capture Measurement

Measurements were made in the lab with various in-home network topologies, with ports both terminated and unterminated, which validated the field results.

To improve the measurement accuracy of the isolation and interfering signal at the set-top box, a wider OUDP pulse is necessary to encompass the variation in amplitude seen with frequency. With an OFDMA bandwidth of 39.4-85 MHz, using this entire spectrum for the OUDP pulse and corresponding set-top box full-band capture would be the most accurate and provide the exact total interference level. This could be capacity and customer-impacting and also require a longer time to do an accurate full-band capture. Out of several measurements, a 20 MHz pulse was chosen and encompassed most of the variation in amplitude seen in Figure 9 - Set-top Box Field Full Band Capture Measurement while still minimizing the energy at the front end of the setup box during the measurement.

With a 20 MHz wide OUDP measurement, a different methodology to measure the interfering signal and energy at the set-top and corresponding ACI delta is needed. One method is to calculate the total power of the interfering signal. Another method is to measure the peak across the 20 MHz band. Due to the possibility of an interfering signal at a specific frequency causing 2nd Harmonic distortion, measuring the peak across the 20 MHz OUDP interference measurement was chosen. The full band capture of the 20 MHz OUDP band is broken down into ten 2 MHz segments, and then the segment with the highest power is used for the ACI delta measurement. An example of this is shown in Figure 10, the original 20 MHz wide pulse centered at 80 MHz is shown in gray, and the 20 MHz pulse is split into 2 MHz increments. In this example, the peak of the 20 MHz pulse and worst interference is below 80 MHz. When comparing the two methods, the original pulse at 80 MHz vs peak across the 20 MHz pulse from 60-80 MHz, the maximum power across the 20 MHz is ~ 2 dB higher than the original OUDP pulse, improving the accuracy of the ACI measurement.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

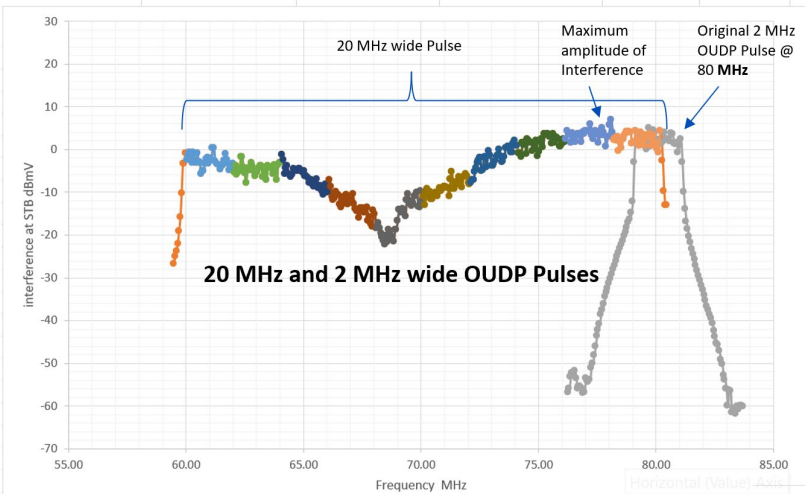


Figure 10 - 20 MHz OUDP Pulse Measurement Example

5. Summing up the iHAT Improvements

Using the above field and lab test data, the iHAT methodology was updated, and the field measurements improved.

1) The OUDP pulse was updated to be 20 MHz wide instead of the original 1.6 MHz pulse, and the peak across a 2 MHz band is used for the ACI measurement.
2) When using the lab set-top box data to calculate the interference thresholds, frequencies at 2X the upstream frequency is now used for threshold measurements.
3) When measuring the full band capture and interference, similarly, frequencies within 2X the OUDP pulse of 60-80 MHz are used to calculate the interference delta.

6. Moving to FDX

Full Duplex DOCSIS (FDX) technology poses additional challenges with adjacent channel interference. With Mid-Split systems, interference is confined to each individual mid-split customer's in-home network. The total power addition of the mid-split spectrum is high enough to cause interference on the mid-split customer's set-top box, but the total power of the mid-split spectrum is not enough to cause interference on neighbor home set-top boxes.

With a mid-split system consisting of four 6.4 MHz wide SC-QAMs and a single OFDMA channel 45.5 MHz wide, the power added by the OFDMA channels is 4.4 dB. With six 96 MHz wide FDX channels added, the additional power is 9.6 dB, or a total of 14 dB additional power with both the OFDMA and FDX channels. Figure 11 shows power addition for the OFDMA and FDX channels in the upstream.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

	BW(MHz)	Power level per 6.4 MHz bandwidth		
SCQAM1	6.4	45.00	dBmV	
SCQAM2	6.4	45.00	dBmV	
SCQAM3	6.4	45.00	dBmV	
SCQAM4	6.4	45.00	dBmV	
		SC QAM TOTAL POWER (dbmV)	51.02	dBmV
MS OFDMA	45.6	Total Power SC + MS OFDMA =	55.46	dBmV
FDX1	96	Total Power SC +MS OFDMA + FDX1 =	59.17	dBmV
FDX2	96	Total Power SC +MS OFDMA + FDX1+FDX2 =	61.14	dBmV
FDX3	96	Total Power SC +MS OFDMA + FDX1+FDX2+FDX3 =	62.49	dBmV
FDX4	96	Total Power SC +MS OFDMA + FDX1+FDX2+FDX3+FDX4 =	63.52	dBmV
FDX5	96	Total Power SC +MS OFDMA + FDX1+FDX2+FDX3+FDX4+FDX5 =	64.35	dBmV
FDX6	96	Total Power SC +MS OFDMA + FDX1+FDX2+FDX3+FDX4+FDX5+FDX6 =	65.05	dBmV
				dBmV
		Total Power addition of FDX Channels	9.59	dBmV
		Total power additoin of FDX and OFDMA	14.03	dBmV

Figure 11 - ACI Power Addition vs. OFDMA and FDX Channels

In an FDX system, the FDX customer will be an all-IP customer with no additional gateways or set-top boxes. The FDX customer will have no DOCSIS devices that will be affected by the upstream FDX spectrum generated by their FDX cable modem. Instead, neighbor homes may be affected. Figure 12 - FDX Neighbor Interference Paths shows this potential interference path to neighboring homes.

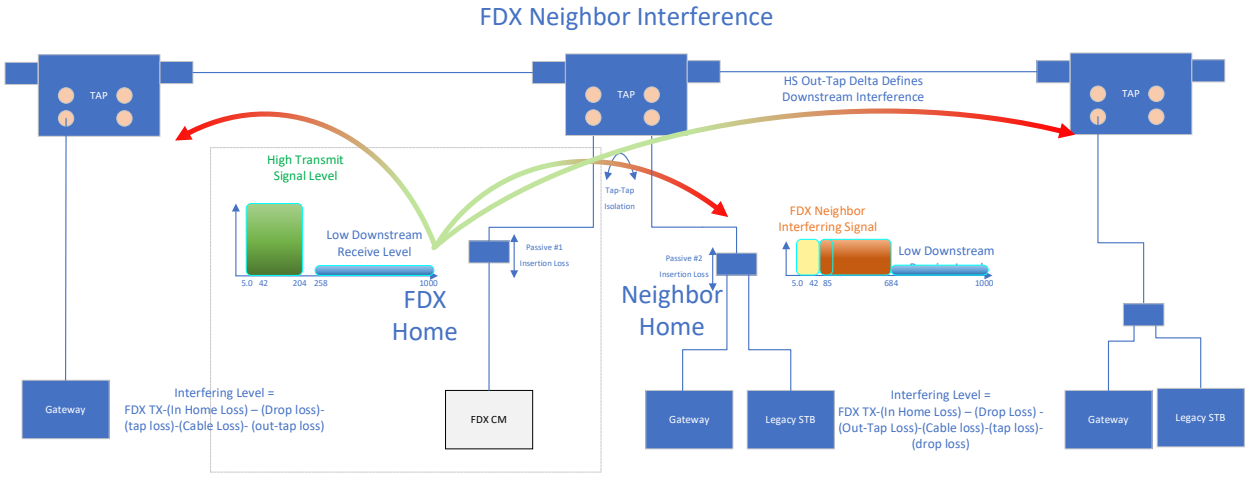


Figure 12 - FDX Neighbor Interference Paths

As can be seen, the neighboring homes contain set-tops with a sub-split front end, and DOCSIS 3.0 and 3.1 technology gateways can be susceptible to ACI from FDX signals from a nearby FDX device. To fully characterize the impact on these neighboring homes, Comcast labs in Downingtown and Dry Creek tested both gateways and set-top boxes to determine the thresholds for ACI interference. These results are presented in section 7 of this report.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

7. Determining Interference Thresholds

7.1. STB

7.1.1. Test Setup and Parameters

Five different STB models from different vendors were selected based on ranking in terms of the total quantity deployed in the Comcast network. As discussed previously, both impact from the AGC and the distortion of the STB front end are part of this study. Since composite distortions will have frequency dependency due to the channel line-up and also due to the order (second, third, etc.), video is monitored at multiple channels corresponding to different analog frequencies.

Downstream video signals are generated from an RPD node. Upstream interfering carriers are generated using a DOCSIS cable load generator (CLGD), while a second CLGD is used to generate downstream FDX signals (Figure 13).

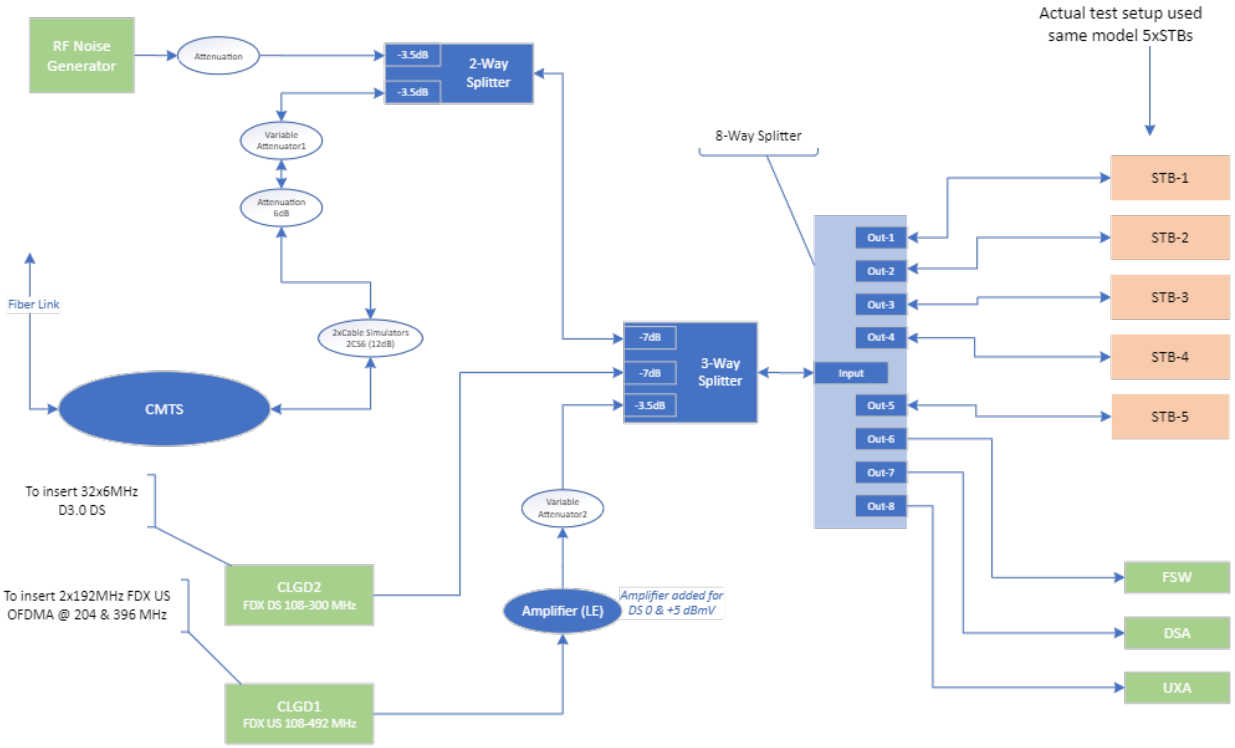


Figure 13 - Test Setup for Adjacent Channel Interference on STBs

The downstream signal levels are kept constant while the upstream carrier levels are increased until video tiling is observed. Upstream carriers utilize a burst profile as specified in the CableLabs PHY specifications, as shown below in Table 3 for waveforms with periods of 10, 70, and 200 milliseconds. Furthermore, a period with four milliseconds was added to the test parameters as real-time measurements of prototype FDX cable modems show that transmissions with periods as low as four milliseconds are possible.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 3 - Period and Duty Cycle for Upstream Burst Signals

US FDX Period (ms)	US FDX On Time (ms)	US FDX Duty Cycle (%)
10	1	10
10	5	50
10	9	90
70	7	10
70	35	50
70	63	90
200	20	10
200	100	50
200	180	90
4	0.4	10
4	2	50
4	3.6	90

Waveforms generated by the CLGD for each period and duty cycle are verified on a real-time analyzer. Initially, due to RAM limitations, the CLGD was not able to generate upstream FDX signals with the correct timing, and a firmware upgrade was necessary to implement a new algorithm for generating the waveforms. Figure 14 is an example of the time domain measurement of a waveform with a duty cycle of 50% and a period of 10 milliseconds.

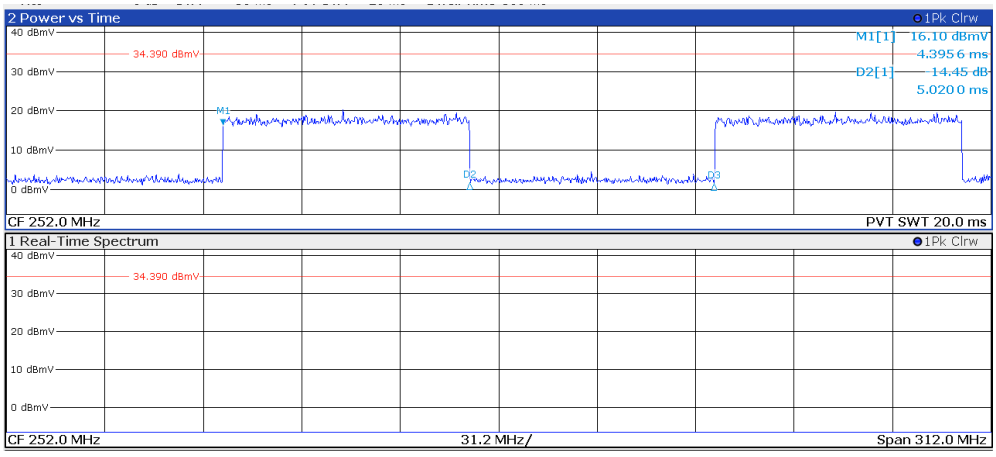


Figure 14 - Power vs Time Capture of Waveform with 10 ms Period and 50% Duty Cycle

Five different channel line-up scenarios were used in the test, shown in Figure 15 below. While these may not be the actual channel line-ups that will be deployed, they represent various amounts of upstream spectrum utilization to ensure that all cases are considered in the design.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

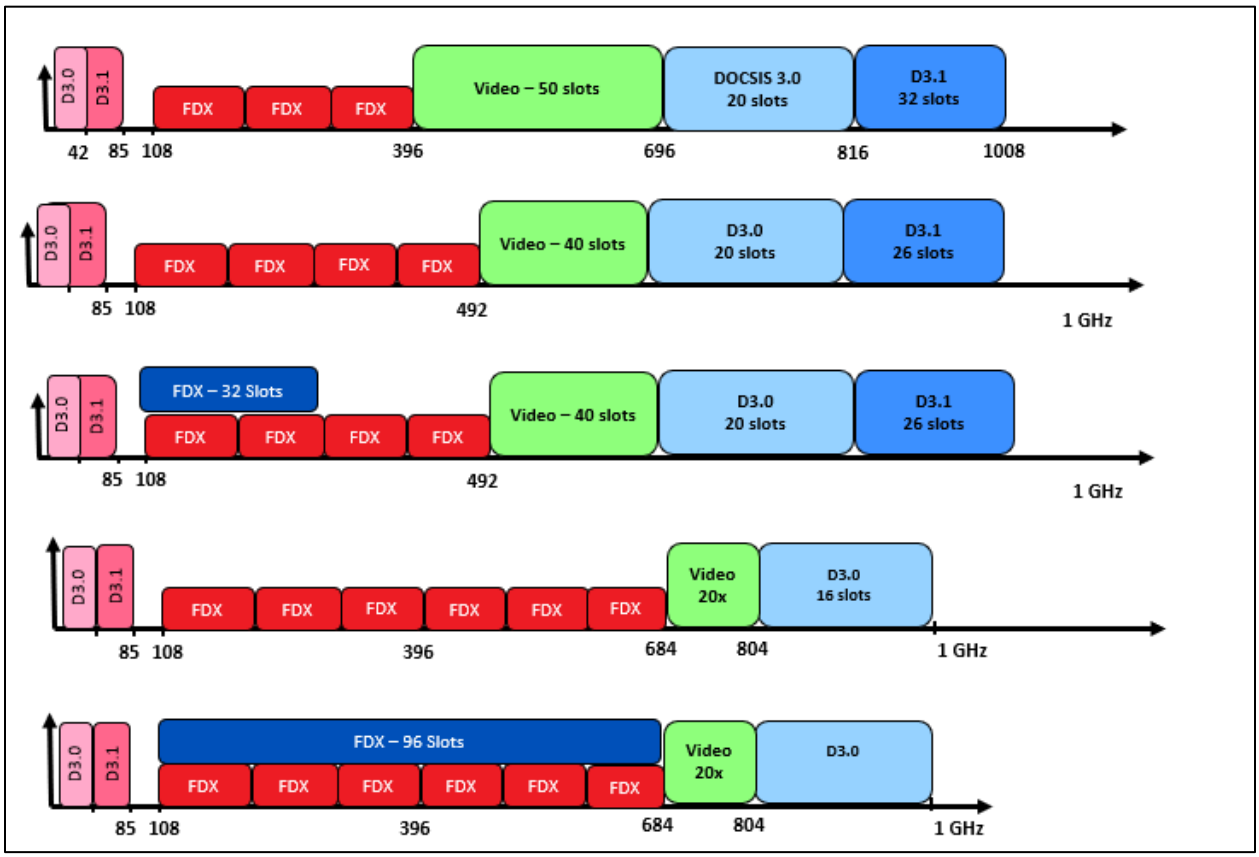


Figure 15 - FDX Upstream and Downstream Channel Lineup scenarios for testing neighbor interference scenarios

7.1.2. Test Results

All measurements are in terms of total composite power (TCP) to provide for a consistent way to compare signal levels where interference to video quality is visible . TCP is measured for the upstream signal present at the STB front end and also measured for the downstream spectrum at the STB front end. The delta for these two TCP values is defined as the threshold value at which ACI impacts the STB and is the value that the neighbor Health Assessment Test (nHAT) uses as reference in its detection algorithm.

Thousands of measurements were taken for different channel maps, RF levels, downstream video frequencies, waveform periods, and duty cycles. This was set up using a Design of Experiment (DOE) approach, and the data was tabulated so that it could be analyzed using tools such as pivot tables and queries. It is not practical to include all the data in one report, so subsets of the data were selected and presented below to highlight factors that impact the TCP threshold for ACI interference.

The susceptibility of the STB to video tiling varies for each STB model. This is likely due to different chipsets and RF front-end designs. These performance differences are significant and can be in excess of 10 dB for different STB models, as shown in Table 4.

Furthermore, Table 4 also shows that both the period and duty cycle impact the TCP threshold. In the case of the STB with the lowest TCP threshold, the lower duty cycle of 10% and shorter period results in the lowest threshold.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 4 - TCP Delta Threshold for Map 2 and DS Video at 495 MHz

Period	10 ms			70 ms			200 ms		
Model / Duty Cycle	10%	50%	90%	10%	50%	90%	10%	50%	90%
1_Model A	14.9	14.1	13.2	9.5	10.1	8.4	10.3	9.7	11
2_Modle B	14.9	14.1	13.2	6.6	8.2	8.4	8.3	10.7	11
3_Model C	15.7	14.1	14.1	8.3	14.1	14.3	8.3	8.6	12
4_Model D	13.5	13	13.2	4.5	5.2	8.4	5.4	5.9	6.1
5_Model E	4.4	7.1	12	4.5	5.2	12.5	5.4	5.9	8.1

Multiple video channels at analog frequencies ranging from closest to the FDX upstream signals to farthest away were measured. Table 5 shows the relationship of frequency to TCP delta. For the particular channel map (Map 2), the worst-case performance is not always at the lowest frequency (closest to the ACI signals) and is an indication that, in addition to the AGC, distortion from the STB front end is also a contributor to ACI.

Table 5 - TCP Delta for Map 2 for Various Video Channels

Video Ch Freq	495 MHz	549 MHz	651 MHz	729 MHz
1_Model A	9.5	7.4	9.8	9.6
2_Modle B	6.6	7.4	8.7	8.4
3_Model C	8.3	8.3	10.8	10.9
4_Model D	4.5	5	5.7	6.4
5_Model E	4.5	4.3	5.9	5.5

As shown in Figure 15, Map 2 and Map 3 have the same upstream FDX spectrum. Map 3, however shows that a portion of the FDX upstream spectrum is also utilized for downstream FDX signals. Table 6 shows the TCP threshold measurements for the downstream video channel at 495 MHz. For the STB with the lowest TCP threshold, Map 3 resulted in an improvement to the TCP threshold of the STB.

Table 6 - TCP Threshold Comparison for Map 2 and Map 3 (DS Video 495 MHz)

Map	2			3		
Duty Cycle	**10**	**50**	**90**	**10**	**50**	**90**
1_Model A	14.9	14.1	13.2	13.4	13	12.8
2_Modle B	14.9	14.1	13.2	13.4	13	12
3_Model C	15.7	14.1	14.1	14.2	13.6	13.8
4_Model D	13.5	13	13.2	12	13	12
5_Model E	4.4	7.1	12	10.3	11.9	12

In conclusion, TCP threshold data was captured for all five maps, tested with different periods and duty cycles, at different RF input levels to the STB, and with video monitored at different frequencies. This

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

data is then implemented as thresholds for the new nHat tool used to support FDX activations and to allow for proactive remediation of potential ACI issues.

7.2. Cable Modems

When the FDX CMs utilize OFDMA channels, the total composite power of the spectrum can fluctuate frequently to affect the neighboring pre-DOCSIS 4.0 CMs' downstream performance. This is because frequent total composite power fluctuations on legacy devices can cause analog-to-digital converter (ADC) saturations and result in uncorrectable codeword errors. Several factors, such as the tuners and AGC algorithms used by the pre-DOCSIS 4.0 cable modems, can result in variations in devices' tolerance levels to the neighbor interference bursts. To understand different CM models' performance under neighbor interference, we performed extensive tests to collect data for determining their thresholds.

7.2.1. Test Setup and Automation

We designed test procedures for searching CM tolerance thresholds under different test setups and conditions. As the test procedures are highly repetitive and time-consuming if performed manually, we successfully developed test automation software to reduce the workload and testing time.

The lab setup created for neighbor interference testing for pre-DOCSIS 4.0 cable modems consists of a mid-split RPD, a set of cable modems under test, a traffic generator, and an upstream burst signal generator to simulate FDX upstream OFDMA bursts. The test automation software communicates with the test devices and equipment through various interfaces to control the traffic generation life cycles, power level, attenuation, cycle period, and duty cycle of the burst signal, and arbitrary waveforms for each interference band. For configuration changes and subscriber management, the test automation software utilizes our virtual cable modem termination system (vCMTS) application programming interfaces (APIs) to perform downstream OFDM modulation profile changes and dynamic bonding changes (DBCs). It also communicates with a network power switch's APIs to perform automated power cycling to reset CM states completely when necessary. The high-level illustration of the test setup is shown in Figure 16 - FDX Neighbor Interference CM Test Setup.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

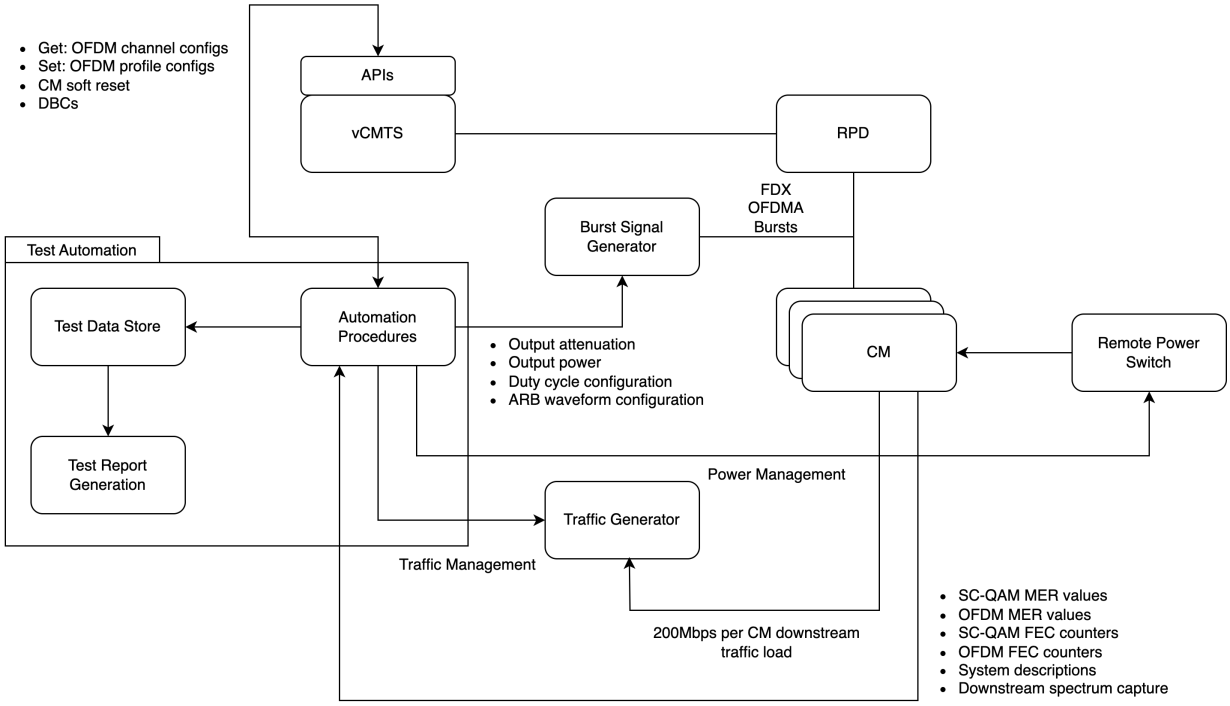


Figure 16 - FDX Neighbor Interference CM Test Setup

In order to collect increasing forward error-correction (FEC) codeword counters and calculate statistically meaningful codeword error ratio (CER) values, the traffic generator is configured by the test automation software to send 200 Mbps downstream user datagram protocol (UDP) traffic on each test device for 10 minutes during each round of the tests. This contributes to the majority of testing time.

The test automation software integrates with all required equipment controls and error handling that allow it to continuously run tests 24 hours a day, seven days a week, without human intervention. Once the tests are complete and the data is collected, interactive test reports are generated by the test automation software to include summarized test results and detailed CM performance metrics.

The test parameters in Table 7 are used to cover spectrum and interference characteristics variations while searching for devices' neighbor interference tolerance thresholds. Currently, the test automation is configured to search linearly through incremental interference output power levels; this provides a complete view of the CMs' performance while the interference characteristics change and provides valuable insights for determining the thresholds. Alternatively, when there is a need to identify the CMs' failing points quickly, binary search can be used while changing the interference's power level to reduce the test time.

Table 7 - CM Neighbor Interference Test Parameters

Parameter	Value
Interference power spectral density difference range	20 dB (1 dB steps)
Interference spectrum	192 MHz, 288 MHz, 384 MHz, 576 MHz
Burst cycle periods	4 ms, 10 ms, 70 ms
Duty cycle	10%, 50%, 90%

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

For each spectrum configuration, 162 rounds of tests are run to cover all test parameter combinations. Given that each round of tests performs a 10-minute downstream traffic session, it can be estimated that the test automation covers all test parameter combinations within 30 hours, considering configuration and wait times on the test equipment.

In addition to the procedure designed to test with periodic neighbor interference bursts, we designed another test procedure to explore how the legacy devices tolerate sudden total composite power changes after they are adapted to quiet FDX bands. This test procedure is performed by turning the signal generator on after being muted for 4 minutes. It examines the maximum CM front-end tolerance under extreme changes and provides reference data points of bottom-line CM thresholds. Both described test procedures are run for DOCSIS 3.1 CMs and DOCSIS 3.0 CMs.

7.2.2. Initial Test Results

In this section, we discuss selected results from testing 288 MHz and 384 MHz of FDX upstream neighbor interference with DOCSIS 3.1 CMs, respectively. These two spectrum configurations are the candidates to be implemented in the initial FDX deployment phase. We also focused on researching the CMs' tolerance levels with simple interference characteristics as a starting point. This was done by injecting the neighbor interference signals in the downstream direction to affect the CMs' AGC immediately. Alternatively, as planned for future tests, the interference source can be combined to form port-to-port isolation to simulate a more realistic scenario in terms of echoes and harmonics or replaced by real FDX devices that are configured to utilize upstream FDX OFDMA bands at different levels.

The performance metrics collected from the CMs are downstream SC-QAM and OFDM FEC counters and MER values. These metrics are collected by the test automation software using simple network management protocol (SNMP) version 3 and are included in the test reports for passing/failing a test and providing detailed CM performance data. The FEC codeword counters of the downstream channels are used to calculate the CER values, which indicate the user impact from the traffic-loss probability perspective. The downstream MER values are expected to decrease as the power of the neighbor interference increases; this is because the AGC on the CMs adds attenuation to adapt to the increased total composite power. Degraded MER values can potentially affect the codeword errors on the SC-QAM channels as they approach the MER threshold for 256-QAM modulation order and can reduce the spectral efficiency of the OFDM channels as their modulation orders are downgraded by our profile management application (PMA) to ensure optimal robustness while maximizing the total capacity.

For each round of tests run for a certain combination of test parameters, the test failing criteria are defined as:

- Showing greater than or equal to $9e^{-7}$ CER on either SC-QAM or OFDM channels
- Significantly impaired SNMP (Simple Network Management Protocol) reporting capabilities

The passing/failing conditions are aggregated in the top-level test summaries, shown in Table 9, Table 10, and Table 11, by PSD differences between the interference and the rest of the legacy downstream spectrum; this means that failing one of the duty cycle tests also fails other tests conducted under the same PSD difference.

7.2.2.1. 288 MHz FDX Neighbor Interference Test Results

In this test, the signal generator was configured to simulate 288 MHz (3 OFDMA bands) worth of FDX upstream bursts. The arbitrary waveforms were generated for combinations of 4 ms, 10 ms, and 70 ms burst cycle periods and 10%, 50%, and 90% duty cycles. A static, 1024-QAM flat modulation profile was

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

assigned to the OFDM channel during the test. The CM downstream receive-power level was configured to -2 dBmV to accommodate the signal generator's maximum output power while staying in the receive-power range required by the DOCSIS PHY layer specifications. The spectrum configuration is shown in Table 8. The test results are listed in Table 9, Table 10, and Table 11.

Table 8 - 288 MHz FDX: Downstream Spectrum Configuration

Type	Start Frequency (Lower Bound)	End Frequency (Upper Bound)	Width
Reserved for FDX	108 MHz	396 MHz	288 MHz
Video QAM channels	396 MHz	696 MHz	300 MHz
Downstream SC-QAM channels	696 MHz	816 MHz	120 MHz
Downstream OFDM channel	816 MHz	1002 MHz	186 MHz

Table 9 - 288 MHz FDX: 4 ms Cycle Period Test Results

Reference PSD Diff. (per 6 MHz)	Calculated FDX TCP (dBmV)	Calculated Legacy TCP (dBmV)	TCP Diff. (dB)	CM 1	CM 2	CM3	CM 4	CM 5	CM 6	CM 7	CM 8
+7 dB	21.81	18.04	3.77	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL
+6 dB	20.81	18.04	2.77	FAIL	FAIL	FAIL	FAIL	FAIL	PASS	PASS	FAIL
+5 dB	19.81	18.04	1.77	PASS	FAIL	PASS	PASS	FAIL	PASS	PASS	PASS
+4 dB	18.81	18.04	0.77	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
+3 dB	17.81	18.04	-0.23	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
+2 dB	16.81	18.04	-1.23	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
+1 dB	15.81	18.04	-2.23	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS

Table 10 - 288 MHz FDX: 10 ms Cycle Period Test Results

Reference PSD Diff. (per 6 MHz)	Calculated FDX TCP (dBmV)	Calculated Legacy TCP (dBmV)	TCP Diff. (dB)	CM 1	CM 2	CM 3	CM 4	CM 5	CM 6	CM 7	CM 8
+9 dB	23.81	18.04	5.77	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Reference PSD Diff. (per 6 MHz)	Calculated FDX TCP (dBmV)	Calculated Legacy TCP (dBmV)	TCP Diff. (dB)	CM 1	CM 2	CM 3	CM 4	CM 5	CM 6	CM 7	CM 8
+8 dB	22.81	18.04	4.77	FAIL	FAIL	FAIL	FAIL	FAIL	PASS	FAIL	FAIL
+7 dB	21.81	18.04	3.77	FAIL	FAIL	FAIL	FAIL	FAIL	PASS	FAIL	FAIL
+6 dB	20.81	18.04	2.77	FAIL	FAIL	FAIL	FAIL	FAIL	PASS	FAIL	FAIL
+5 dB	19.81	18.04	1.77	FAIL	FAIL	PASS	FAIL	FAIL	PASS	FAIL	PASS
+4 dB	18.81	18.04	0.77	FAIL	FAIL	PASS	PASS	FAIL	PASS	FAIL	PASS
+3 dB	17.81	18.04	-0.23	FAIL	FAIL	PASS	PASS	FAIL	PASS	FAIL	PASS
+2 dB	16.81	18.04	-1.23	PASS	PASS	PASS	PASS	FAIL	PASS	PASS	PASS
+1	15.81	18.04	-2.23	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
0 dB	14.81	18.04	-3.23	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS

Table 11 - 288 MHz FDX: 70 ms Cycle Period Test Results

Reference PSD Diff. (per 6 MHz)	Calculated FDX TCP (dBmV)	Calculated Legacy TCP (dBmV)	TCP Diff. (dB)	CM 1	CM 2	CM 3	CM 4	CM 5	CM 6	CM 7	CM 8
+6 dB	20.81	18.04	2.77	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL
+5 dB	19.81	18.04	1.77	FAIL	FAIL	FAIL	FAIL	FAIL	PASS	FAIL	FAIL
+4 dB	18.81	18.04	0.77	FAIL	FAIL	PASS	PASS	PASS	PASS	FAIL	PASS
+3 dB	17.81	18.04	-0.23	FAIL	PASS	PASS	PASS	PASS	PASS	FAIL	PASS
+2 dB	16.81	18.04	-1.23	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
+1	15.81	18.04	-2.23	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
0 dB	14.81	18.04	-3.23	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS

It was observed that in the test results, different CMs can have significant tolerance thresholds to the neighbor interference, which can be as large as 7 dB worth of total composite power difference (between CM 5 and CM 6).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Diving into the detailed test data, as a reference, the average MER value of the SC-QAM channels on CM 2 dropped approximately 3 dB as the neighbor interference power increased by 8 dB, such changes are shown in Figure 17 - 288 MHz FDX: CM-2 SC-QAM Average MER changes (4 ms cycle period). We also observed that the 10% duty cycle created more impact than the 50% and 90% duty cycles. An example from CM 2's test results is shown in Table 12.

Table 12 - 288 MHz FDX: 10% Duty Cycle Created the Most Impact (CM 2 Example, 4 ms)

Reference PSD Diff. (per 6 MHz)	Duty Cycle	OFDM Avg. MER	SC-QAM Avg. MER	SC-QAM Corrected	SC-QAM Uncorrectable	OFDM Corrected	OFDM Uncorrectable
+5 dB	90%	42.36 dB	40.58 dB	0.000e+00	0.000e+00	5.149e-01	0.000e+00
+5 dB	50%	42.12 dB	40.55 dB	0.000e+00	0.000e+00	5.387e-01	0.000e+00
+5 dB	10%	42.25 dB	40.54 dB	1.901e-04	**1.293e-03**	5.555e-01	**5.193e-06**

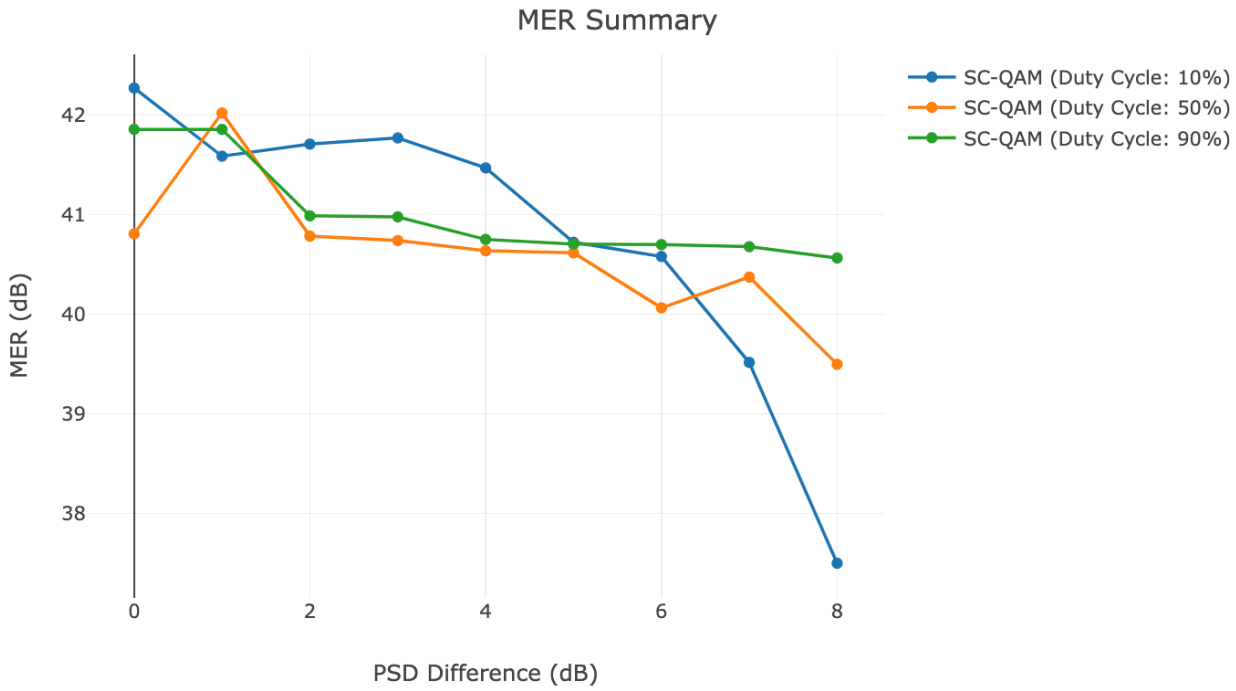


Figure 17 - 288 MHz FDX: CM-2 SC-QAM Average MER changes (4 ms cycle period)

In the per-CM test report generated for 4 ms cycle period, 10% duty cycle, and +5 dB PSD difference for CM 2, it can be observed that all downstream channels produced uncorrectable codeword errors in the selected time window, as shown in Figure 18 - 288 MHz FDX: CM-2 CER Over Time (4 ms cycle period, 10% duty cycle, 5 dB PSD Difference):

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

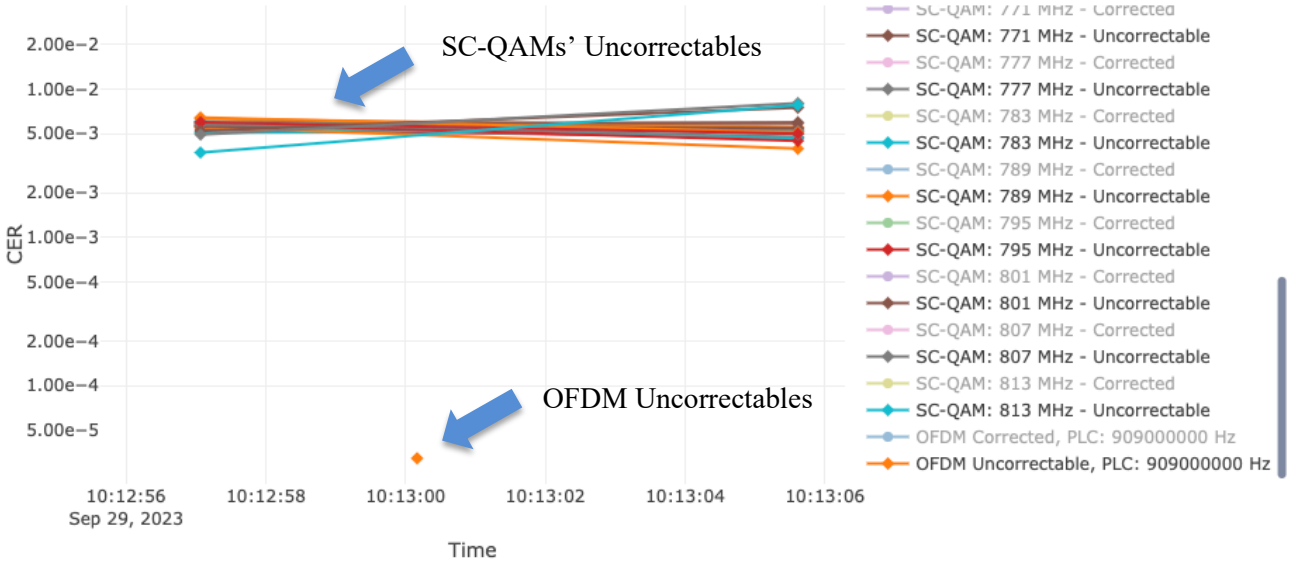


Figure 18 - 288 MHz FDX: CM-2 CER Over Time (4 ms cycle period, 10% duty cycle, 5 dB PSD Difference)

7.2.2.2. 384 MHz FDX Neighbor Interference Test Results

In this test, the signal generator was configured to simulate 384 MHz (4 OFDMA bands) worth of FDX upstream bursts. The same arbitrary waveforms, cycle periods, and duty cycles were used. The spectrum configuration is shown in

Table 13. The test results are listed in

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 14, Table 15, and

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 16.

Table 13 - 384 MHz FDX: Downstream Spectrum Configuration

Type	Start Frequency (Lower Bound)	End Frequency (Upper Bound)	Width
Reserved for FDX	108 MHz	492 MHz	384 MHz
Video QAM channels	492 MHz	732 MHz	240 MHz
Downstream SC-QAM channels	732 MHz	852 MHz	120 MHz
Downstream OFDM channel	852 MHz	1002 MHz	150 MHz

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 14 - 384 MHz FDX: 4 ms Cycle Period Test Results

Reference PSD Diff. (per 6 MHz)	Calculated FDX TCP (dBmV)	Calculated Legacy TCP (dBmV)	TCP Diff. (dB)	CM 1	CM 2	CM 3	CM 4	CM 5	CM 6	CM 7	CM 8
+6 dB	22.06	17.29	4.77	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL
+5 dB	21.06	17.29	3.77	FAIL	FAIL	PASS	FAIL	FAIL	PASS	PASS	PASS
+4 dB	20.06	17.29	2.77	FAIL	FAIL	PASS	PASS	PASS	PASS	PASS	PASS
+3 dB	19.06	17.29	1.77	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
+2 dB	18.06	17.29	0.77	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS
+1 dB	17.06	17.29	-0.23	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS

Table 15 - 384 MHz FDX: 10 ms Cycle Period Test Results

Reference PSD Diff. (per 6 MHz)	Calculated FDX TCP (dBmV)	Calculated Legacy TCP (dBmV)	TCP Diff. (dB)	CM 1	CM 2	CM 3	CM 4	CM 5	CM 6	CM 7	CM 8
+6 dB	22.06	17.29	4.77	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL
+5 dB	21.06	17.29	3.77	FAIL	FAIL	FAIL	FAIL	FAIL	PASS	FAIL	FAIL
+4 dB	20.06	17.29	2.77	FAIL	FAIL	PASS	PASS	FAIL	PASS	FAIL	PASS
+3 dB	19.06	17.29	1.77	FAIL	FAIL	PASS	PASS	PASS	PASS	FAIL	PASS
+2 dB	18.06	17.29	0.77	FAIL	FAIL	PASS	PASS	PASS	PASS	PASS	PASS
+1 dB	17.06	17.29	-0.23	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 16 - 384 MHz FDX: 70 ms Cycle Period Test Results

Reference PSD Diff. (per 6 MHz)	Calculated FDX TCP (dBmV)	Calculated Legacy TCP (dBmV)	TCP Diff. (dB)	CM 1	CM 2	CM 3	CM 4	CM 5	CM 6	CM 7	CM 8
+5 dB	21.06	17.29	3.77	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL	FAIL
+4 dB	20.06	17.29	2.77	FAIL	FAIL	PASS	FAIL	FAIL	PASS	FAIL	PASS
+3 dB	19.06	17.29	1.77	FAIL	FAIL	PASS	PASS	PASS	PASS	FAIL	PASS
+2 dB	18.06	17.29	0.77	PASS	FAIL	PASS	PASS	PASS	PASS	FAIL	PASS
+1 dB	17.06	17.29	-0.23	PASS	FAIL	PASS	PASS	PASS	PASS	PASS	PASS
0 dB	16.06	17.29	-1.23	PASS	PASS	PASS	PASS	PASS	PASS	PASS	PASS

In the 384 MHz FDX testing results, the total composite power difference increased by 2 dB at the same PSD difference levels compared to the spectrum configuration used for 288 MHz FDX interference tests, as the interference spectrum increased from 288 MHz to 384 MHz and the legacy downstream spectrum decreased from 606 MHz to 510 MHz. This caused a subset of CMs to show a 2 dB lowered threshold in the normalized PSD differences under certain cycle period settings. However, this effect is not consistently reflected by all test CMs across all test conditions at the same degree. Because multiple factors can result in variations in CMs' performance, their different responses to different neighbor interference characteristics can cause non-linear effects in the results.

Leveraging the automation software that we developed for this testing initiative minimizes the effort of building a rich dataset of interference tolerance thresholds for different CM models under various test conditions. And the CM thresholds for nHAT can potentially be determined automatically based on this dataset and associated policy configurations.

8. FDX Neighbor Interference Implementation and nHAT

Comcast's terminology for validating the neighbor accounts and devices for interference with implementing FDX is nHAT (neighbor Health Assessment Test). This testing will be similar to that completed for iHAT with a few differences.

Different thresholds will be needed for the set-top boxes, and thresholds will also need to be determined for pre-DOCSIS 4.0 Devices.

OUDP signaling will be used to measure neighbor interference. Where in the updated iHAT test, a 20 MHz wide OUDP measurement is used, for nHAT, three 2 MHz OUDP pulses across the FDX spectrum from 108-684 MHz will be utilized to determine the level of interference at the neighboring devices; this will allow for variation across frequency and also minimize the amount of time needed for the full band capture.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

With FDX neighbor interference having the potential to impact many neighboring devices, minimizing and optimizing the number of neighbors to test is needed. Comcast uses a mapping graph database called ROCI (Routing of Cable Infrastructure), which will be used to determine the neighbor devices which are susceptible to ACI and which neighboring devices to be tested. If ROCI is not available for a specific node or segment, the entire node segment will be evaluated.

Figure 19 - nHAT Flow Diagram shows the nHAT flow, which includes determining which devices to test using ROCI, running the test, pass/fail actions, remediation, and activation.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

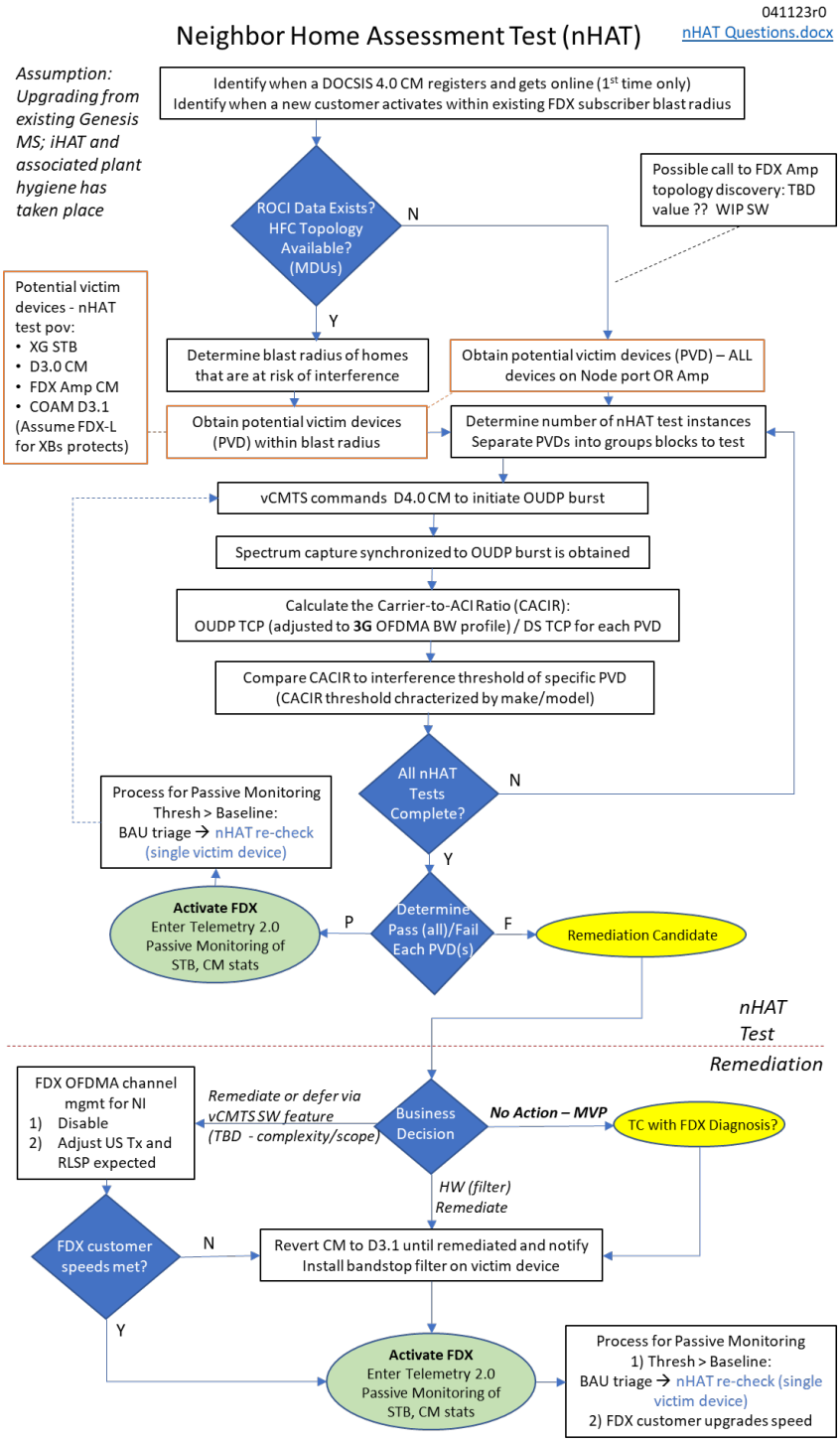


Figure 19 - nHAT Flow Diagram

Pre-analyzing the network is beneficial to help determine the potential for neighbor interference. Using the plant topology and devices which are in ROCI, an estimate of interference can be determined. Figure 20 - Example ROCI Topology with FDX Customers shows an example plant topology with two FDX customers. The red dot is the node, purple are actives, green are taps and light blue are customers. ROCI

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

contains plant information including cable types and lengths and device model types. In this example, customer one has three neighbors off the same tap and customer two is the only customer on their tap.

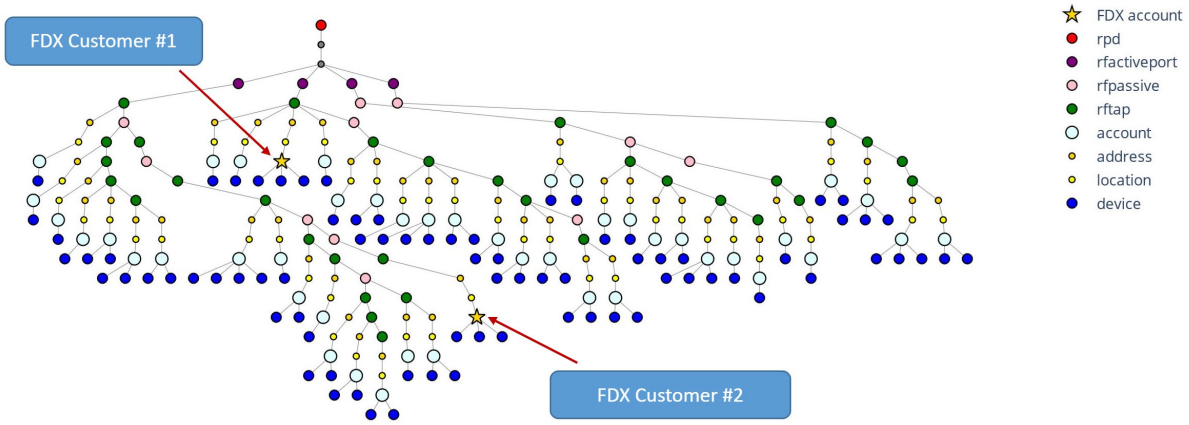


Figure 20 - Example ROCI Topology with FDX Customers

Drilling down further, the plant details can be seen. See Figure 21 - ROCI Analysis of Neighbor Interference Levels. Cable spans and tap values are used to determine the amount of isolation between the FDX customer and neighbors. In this example, the drop loss is unknown and a nominal value of 5 dB is used.

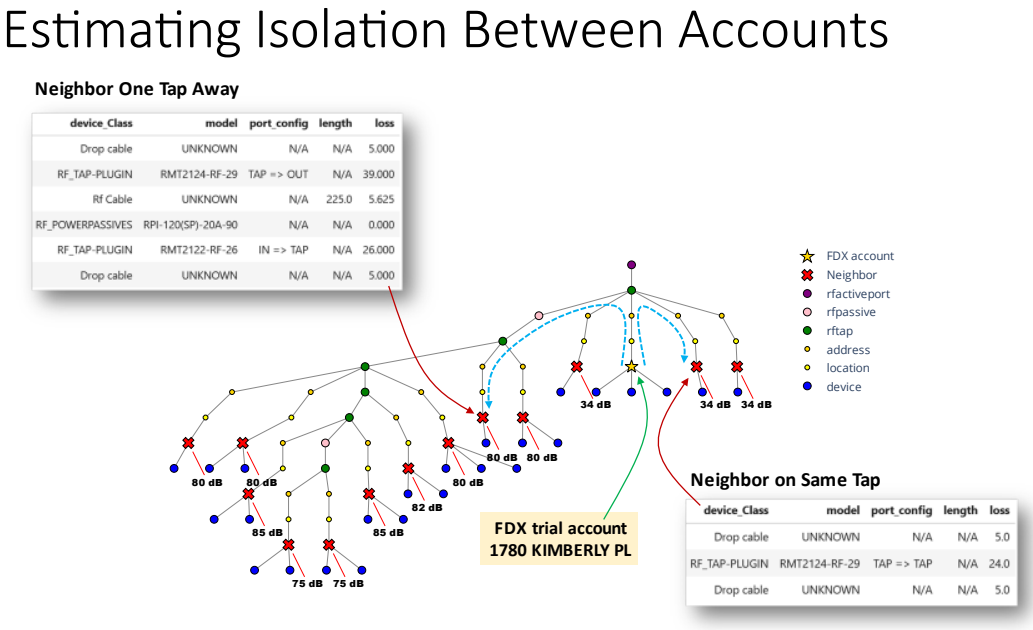


Figure 21 - ROCI Analysis of Neighbor Interference Levels

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Using the measured transmit levels for the FDX customer and receive levels for the neighbor accounts, the level of interference can be estimated.

For neighbors on the same tap see Figure 22 - nHAT ROCI Analysis ACI Delta Same Tap:

- FDX transmit PSD = -28 dBmV/Hz
- Neighbor-to-neighbor isolation = 34 dB
- Received interference level = -62 dBmV/Hz
- Downstream mean at legacy device = -69 dBmV/Hz
- FDX/DS PSD delta = 7 dBmV/Hz
- FDX/DS Power delta = 3 dBmV

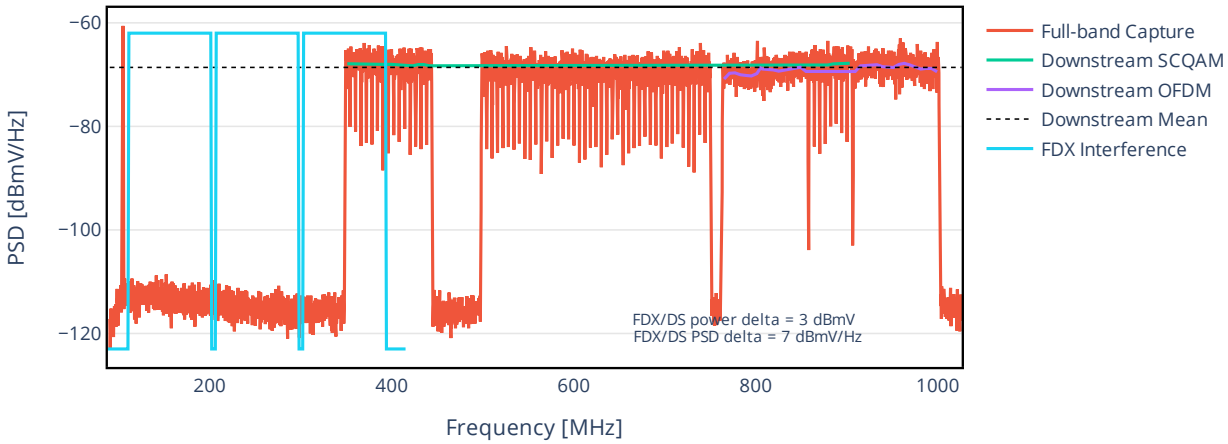


Figure 22 - nHAT ROCI Analysis ACI Delta Same Tap

Thresholds are still being determined for FDX neighbor interefence. Initial data shows a power spectral density delta of 7 dBmV and total power delta of 3 dBmV between the interfering signal and the neigbor, which is below the thresholds being measured in the lab for interference for both the set-top box and cable modems.

For neighbors on the downstream adjacent tap see Figure 23 – nHAT ROCI Analysis ACI Delta Adjacent Tap:

- FDX transmit PSD = -28 dBmV/Hz
- Neighbor-to-neighbor isolation = 80 dB
- Received interference level = -109 dBmV/Hz
- Downstream mean at legacy device = -69 dBmV/Hz
- FDX/DS PSD delta = -46 dBmV/Hz
- FDX/DS Power delta = -40 dBmV

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

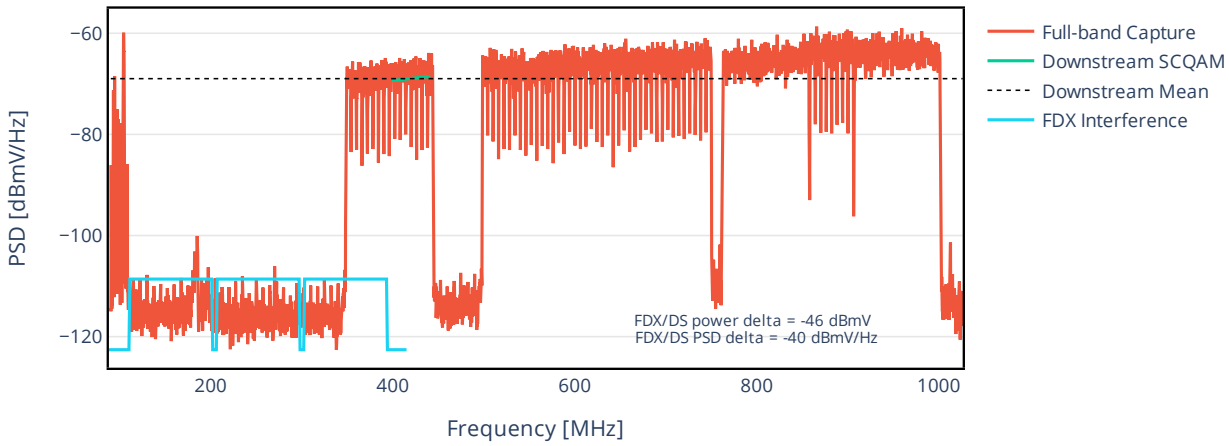


Figure 23 – nHAT ROCI Analysis ACI Delta Adjacent Tap

Both the power spectral density and total power delta are well below the threshold of interference and testing is not required on this device.

9. Conclusion

Understanding, calculating, detecting, and remediating adjacent channel interference are all needed as the industry advances and implements newer revisions of DOCSIS which re-use existing spectrum while the cable plant has a large number of older DOCSIS and video devices co-existing on the systems. With over 2.4 million mid-split activations, Comcast has developed a very good understanding of how to detect and monitor ACI in mid-split applications. This knowledge and experience are being used to be proactive in the modeling, calculations, and detection of interference in FDX systems. The interference in FDX systems may be significantly more impactful as it has the potential to affect both neighboring set-top devices and cable modems and not just devices in a single home as in the mid-split systems.

Comcast pioneered the use of OUDP signaling first for leakage applications in high-split systems and then again in the iHAT tool to evaluate the health of mid-split homes and accounts. The continued use of OUDP signaling in nHAT for FDX systems is key to ensuring ACI is detected before there is any negative customer impact.

The discussion in this paper touched on many of the improvements which are being made to iHAT and the additional complexities of measuring interference in FDX systems. The development of tools and processes in bringing mid-split to scale was an excellent learning experience for interoperability with different DOCSIS capabilities and also to bring this technology to scale across the network. For FDX, work continues on determining the interference thresholds for both the cable modems and set-top boxes. Work also continues on developing the processes and tools needed to bring FDX to scale. FDX has been deployed on a small scale to date and full-scale deployments are planned for Q4 2023 and 2024.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

10. Abbreviations

ACI	adjacent channel interference
AGC	analog-to-digital converter
AGC	automatic gain control
ARB	Arbitrary
BW	bandwidth
CACIR	carrier to ACI ratio
CER	codeword error rate
CM	cable modem
CMTS	cable modem termination system
COAM	customer owned and maintained
CPE	customer premises equipment
CW	continuous wave
CW	codeword
dB	decibel
dBc	decibel from Carrier
DBC	dynamic bonding change
dBmV	decibel Millivolt
DOCSIS	data over cable service interface specification
DS	downstream
HFC	hybrid fiber-coax
HS	High-Split
FBC	full band capture
FEC	forward error correction
FDX	full duplex DOCSIS
Hz	hertz
iHAT	In Home Health Assessment Test
MER	modulation error ratio
MHz	megahertz
ms	millisecond
MS	mid split
MVP	minimum viable product
nHAT	neighbor home health assessment test
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
OUDP	OFDMA upstream data profile
PMA	profile management application
PSD	power spectral eensity
PVD	potential victim device
QAM	quadrature amplitude modulation
RDK	reference design kit
RF	radio frequency
RLSP	return level setpoint
ROCI	routing of cable infrastructure
RPD	remote physical device
R-PHY	remote physical layer
RX	receive

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

s	second
SC	single carrier
SC-QAM	single carrier quadrature amplitude modulation
SCTE	Society of Cable Telecommunications Engineers
SNMP	simple network management protocol
STB	set-top box
SW	software
TBD	to be determined
TC	Trouble call
TCP	total composite power
TX	transmit
UDP	user datagram protocol
uV/m	microvolt per meter
US	upstream
vCMTS	Virtual cable modem termination system
WIP	work in progress

Bibliography & References

Bringing the Mid-Split Factory Online to Rapidly Produce Terabytes, SCTE Cable-Tec Expo 2022

A Proactive Network Management Scheme for Mid-split deployment, SCTE Cable-Tec Expo 2020

Rapid and Automated Production Scale Activation of Expanded Upstream Bandwidth, SCTE Fall Technical Forum 2021

Operating Legacy Cable Modems in an FDX Environment, SCTE Fall Technical Forum 2019

DOCSIS® 3.0 *Physical Layer Specification*, CM-SP-PHYv3.0-C01-171207, December 7, 2017

DOCSIS® 3.1 *Physical Layer Specification*, CM-SP-PHYv3.1-I20-230419, April 19, 2023

DOCSIS® 4.0 *Physical Layer Specification*, CM-SP-PHYv4.0-I05-220328, March 28, 2022

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Open Radio Access Network (RAN)

How Fast Can We Get There?

A Technical Paper prepared for SCTE by

Mark Poletti
Director, Mobile Networks and Principal Wireless Architect
CableLabs
858 Coal Creek Circle, Louisville, CO 80027
+1 303-550-6413
m.poletti@cablelabs.com

Omkar Dharmadhikari, CableLabs, Principal Wireless Architect

Srinivas Sriram, Delartech, Wireless Architect

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Background 4
2.1. Distributed and Centralized RAN (D-RAN and C-RAN) 5
2.2 5G RAN Split Architecture 6
2.3 Virtualized RAN (vRAN) 9
3. Open RAN 10
3.1 O-RAN ALLIANCE 10
3.2 Open RAN Architecture 11
3.3 Key Benefits 13
3.4 Deployment Scenarios 14
4. Open RAN Ecosystem Activity 15
4.1 Global Mobile Network Operator Activity 15
4.2 Industry Test Beds 17
4.3 Government Policy and Investment 18
5. Multi-Vendor Open RAN integration at CableLabs 20
5.1. Lab Overview 20
5.2. Industry Test Results 22
5.3. Observations 24
5.4 Lessons Learned 26
6. Open RAN Transition 26
6.1. Transition from vRAN to Open RAN 26
6.2 Open RAN Transition Approach 27
7. Conclusion 31

Abbreviations 32

Bibliography & References 33

List of Figures

Title **Page Number**

Figure 1. Integrated eNB vs Distributed RAN 5
Figure 2. Centralized RAN 6
Figure 3. Disaggregated RAN 6
Figure 4. 3GPP 5G RAN functional split options 8
Figure 5. Deployment configurations of 5G RAN functional split options 8
Figure 6. Lower and higher layer splits with separate user and control plane in the CU 9
Figure 7. Various vRAN Deployment Architectures 10
Figure 8. 3GPP and O-RAN Alliance coexistence 11
Figure 9. Logical Open RAN Architecture 12
Figure 10. Traditional to Virtualized to Containerized Platforms 13
Figure 11. CableLabs/Kyrio Test Lab Facilities 21
Figure 12. Open RAN subsystem/system test configurations 21
Figure 13. CableLabs/Kyrio OTIC test and certification capabilities 22

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 14. CableLabs/Kyrio lab setup for standalone wraparound tests 23
Figure 15. CableLabs/Kyrio lab setup for e2e integration 24
Figure 16. Open RAN transition approach 28
Figure 17. Hardware First transition path to Open RAN 29
Figure 18. Software First transition path to Open RAN 30
Figure 19. Direct transition path to Open RAN 30

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

As mobile operators transition from 4G to 5G, one of the biggest parts of the mobile network experiencing transformation is the Radio Access Network (RAN). This transformation includes virtualization, disaggregation, open interface, intelligence, and cloudification, which are foundational components to an Open RAN architecture and complements the service-based architecture envisioned for 5G.

Today, mobile networks typically consist of proprietary software integrated onto customized hardware designed by a single mobile network supplier. Network upgrades or fixes require systematic deployment and validation to each of the network subsystem functions and, in some cases, lead to an entire network overhaul. Mobile operators are locked into the development, release pace, and schedule of the network supplier, which can significantly impact network operation, offered services, and deployment plans. This industry dynamic increases costs, slows innovation, and reduces competition.

The industry is attempting to change this single vendor paradigm ('vendor lock') by developing industry standards to disaggregate proprietary software from the hardware and open proprietary interfaces between network subsystems and functions to allow for network supplier interoperability.

With a recent increase in mobile network operators (MNOs) trials, industry Plugfests, and standards development, the industry has quickened the pace towards achieving Open RAN, which has the vision of RAN components as a virtual software application capable of working on any bare metal hardware operated with orchestration and automation. The mobile industry is at a tipping point where non-traditional network suppliers are gaining momentum in developing Open RAN solutions that are maturing and becoming commercially viable, reliable, and scalable to the operator community.

MNOs are assessing many factors and dependencies that are key to the transition of mobile networks to Open RAN, including the commercial and economic viability of Open RAN. As part of the considerations, MNOs must leverage current infrastructure investments to expand networks while developing a road map for future growth and purchase cycles.

For several years, CableLabs evaluated the progress of Open RAN architecture and conducted several lab investigations in collaboration with Telecom Infra Project (TIP), O-RAN Alliance, and the National Telecommunications and Information Administration (NTIA). To assist in this assessment, CableLabs undertook an investigation of Open RAN transition approaches and key indicators for assessment with learnings from lab activities of industry generated and representative common operator deployment scenarios. As a result, this paper will address the following topics[1]:

- Analyze the current state of the Open RAN ecosystem.
- Discuss learnings of external and internal CableLabs Open RAN test activities.
- Develop an Open RAN transition approach with assessment criteria.

2. Background

The evolution of wireless communication networks has led to the introduction of fifth generation (5G) technology, which promises to deliver unprecedented speeds (enhanced Mobile Broadband (eMBB)), ultra-low latency (ultra-reliable Low Latency Communications (URLLC)), and massive connectivity (massive Machine Type Communication (mMTC)). To support the use cases promised by 5G networks,

[1] Note that Open RAN economics are out of scope of this paper and have been addressed in detail in previous CableLabs Technical Briefs

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the RAN plays a crucial role. Traditionally, the RAN has been dominated by proprietary and vertically integrated solutions offered by a handful of established vendors. However, with the emergence of Open RAN, the landscape is undergoing a significant transformation by reducing the barriers to entry and broadening the vendor ecosystem.

To better understand the emergence of Open RAN, an overview of the transitional changes to the RAN architecture in 4G and 5G is described below, which introduces foundational elements of centralization, disaggregation, and virtualization into the RAN architecture.

2.1. Distributed and Centralized RAN (D-RAN and C-RAN)

During the rollout of 4G networks, the RAN transitioned from a 'monolithic' integrated eNB (i.e. LTE base station), where radio elements are integrated and housed in a single enclosure with coax to the antenna, to a Distributed RAN (D-RAN) architecture where the radio elements of the RAN is split into the Radio Unit (RU) and Base Band Unit (BBU), as shown in Figure 1. The RU consists of the lower physical layer of the radio protocol stack, converts RF signal into data signal and vice versa, and includes components such as RF filters and high-power amplifiers connected to the antenna.

The BBU manages the entire base station, including signal processing, operation, and maintenance, and consists of all the radio protocol stack layers except the lower physical layer. Each cell site with its radio functions is distributed and connected with a single backhaul connection to the core network. The BBU-RU interface replaced the traditional coax cable with optical fiber managed by the Common Public Radio Interface (CPRI) protocol. The BBU and RU can be housed in the same enclosure or separated by a short distance for instance, residing on a tower close to the antenna.

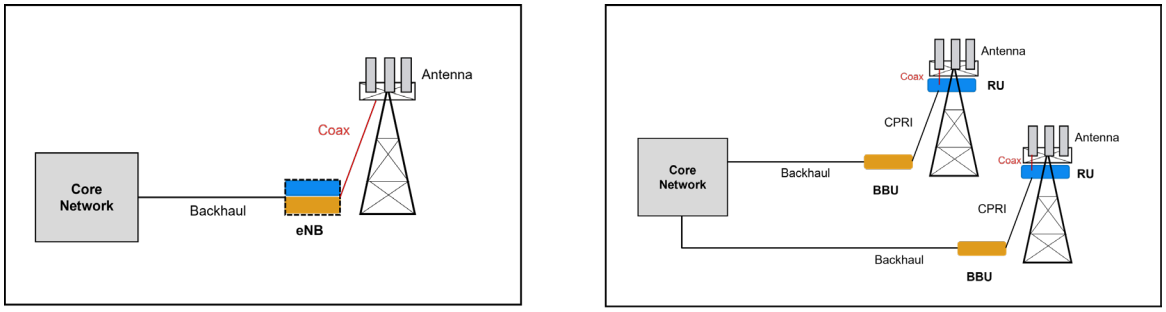


Figure 1. Integrated eNB vs Distributed RAN

The D-RAN later evolved to a Centralized RAN (C-RAN), where a 'pool' of BBUs is centralized at a local hub or data center close to the cell site, as shown in Figure 2. This enables the Radio Units (RUs) to be up to several miles away from the BBUs. The RUs became Remote Radio Units (RRUs) since they can be separated by a relatively long distance from the BBUs. This allows for the cell site to only include the RRU and the antenna. This also results in a new interface between the RRU and BBU pool called the 'fronthaul'.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

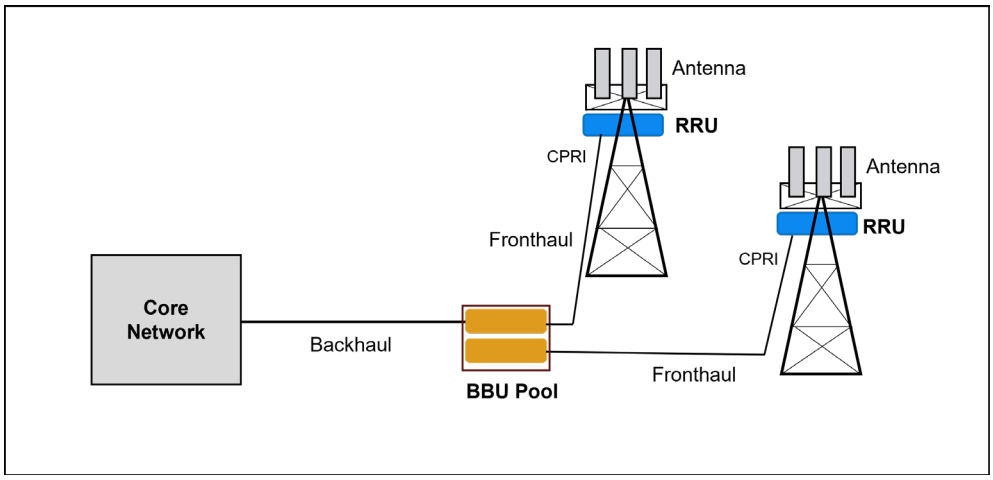


Figure 2. Centralized RAN

The C-RAN architecture can pool multiple RRUs to a single BBU pool and extend beyond a 1:1 mapping to a 1:N mapping of BBU to RRU. This provides capital and operational expenditure (CAPEX and OPEX) benefits of reduction in deployment and maintenance cost by requiring fewer routers, reduced electricity usage, and reduced real estate needs. The end-to-end latency can be further reduced with edge computing where an application server can be placed in or close to a hub or data center. Advanced features such as Coordinated Multi-Point (CoMP) can be supported due to low latency communications made feasible by the proximity of the BBUs in a data center. In addition, it improves spectral efficiency and reduces inter-channel interference since centralized BBUs can share resources dynamically with multiple RRUs using joint scheduling and processing. The 5G RAN (i.e., gNB) also utilizes the integrated RAN, D-RAN, and C-RAN architectures similar to 4G LTE.

2.2 5G RAN Split Architecture

Disaggregation

The 3rd Generation Partnership Project (3GPP) 5G Release 15 specifications laid the initial foundation for the 5G RAN which includes further disaggregation of the RAN into multiple components, as shown in Figure 3.

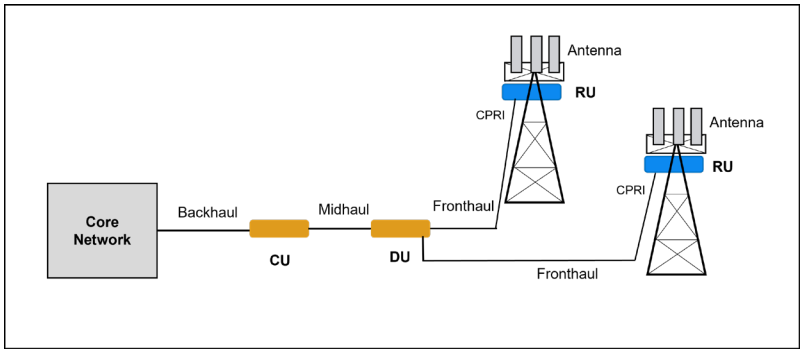


Figure 3. Disaggregated RAN

The 3GPP architecture divides the gNB into two logical components, the gNB-Central Unit (gNB-CU) and the gNB-Distributed Unit (gNB-DU). The gNB-CU implements the upper layers of the radio protocol

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

stack for the Control Plane and User Plane and includes the Radio Resource Control (RRC), Service Data Adaptation Protocol (SDAP), and Packet Data Convergence Protocol (PDCP) layers. The gNB-DU implements the lower layers of the radio protocol stack for the Control Plane (CP) and User Plane (UP) and includes the Radio Link Control (RLC), Media Access Control (MAC), and Physical (PHY) Layer. A new interface is introduced called the 'Midhaul' that connects the CU to the DU. From the 3GPP perspective, the gNB-DU includes the entire PHY layer, however, this can be further divided into Upper PHY and Lower PHY based on implementation-specific requirements. The Upper PHY manages baseband processing (previously in the BBU) and the Lower PHY manages RF processing. The gNB-CU can interface directly with the core network or with other gNB-CUs and gNB-DUs. In addition, the gNB-DU can directly interface with multiple RUs and other gNB-DUs and gNB-CUs.

From a 3GPP perspective, the gNB-CU and gNB-DU utilize an F1 interface between them and further separates the CP and the UP. The CP protocol uses F1 Application Protocol (F1AP) and the UP protocol uses General Packet Radio Service (GPRS) Tunneling Protocol. Thus, this decomposition goes into a gNB-CU-UP and gNB-CU-CP interfaces on the F1 with an E1 interface between them that is managed by the E1 Application Protocol (E1AP).

The benefits of disaggregation of the gNB include increased flexibility, scalability, and interoperability to realize 5G use cases and enable the RAN to support multiple logical paths over the same physical path that allows for innovative service delivery to multiple user types over the same network. 3GPP Release 16/17/18 will include a fully featured 5G network using Service Based Architecture (SBA) that will include virtualization and cloudification of the network that lays the foundation for the core to be extended into the RAN.

Functional Splits

With the advent of 5G, the industry realized the need for further standardization of the radio protocol stack layers within the disaggregated RAN and created functional splits. 3GPP TR 38.801 provides definitions of these multiple 'split' variants to the 5G RAN architecture shown in Figure 4. The functional RAN split options provide a standardized arrangement of radio stack layers within the disaggregated RAN to allow for flexibility in design and deployment of the 5G RAN to meet capacity, coverage, and latency requirements.

The architecture of a 5G RAN depends on the architectural split selected by the MNO. The 3GPP standard provides multiple options or 'splits' of functionality based on building blocks shown in Figure 4. The various functions can then be performed by the CU, DU, or RU depending on the option chosen. As shown in the lower portion of Figure 4, common splits include: (1) 5G high layer split (F1) (option 2) with a CU and DU/RU; (2) 5G low layer split (Fx) (option 7) with a CU/DU and RU; and (3) 5G cascaded split (option 2 and option 7) with a standalone CU, DU, and RU. The higher layer splits are well suited for low bandwidth requirements and are latency tolerant - ideal for low latency, edge computing, and non-mobile use cases such as Fixed Wireless Access (FWA). While the lower layer splits are well suited for high bandwidth requirements and are latency constrained – ideal for mobile use cases such as Coordinated Multipoint Access (CoMP).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

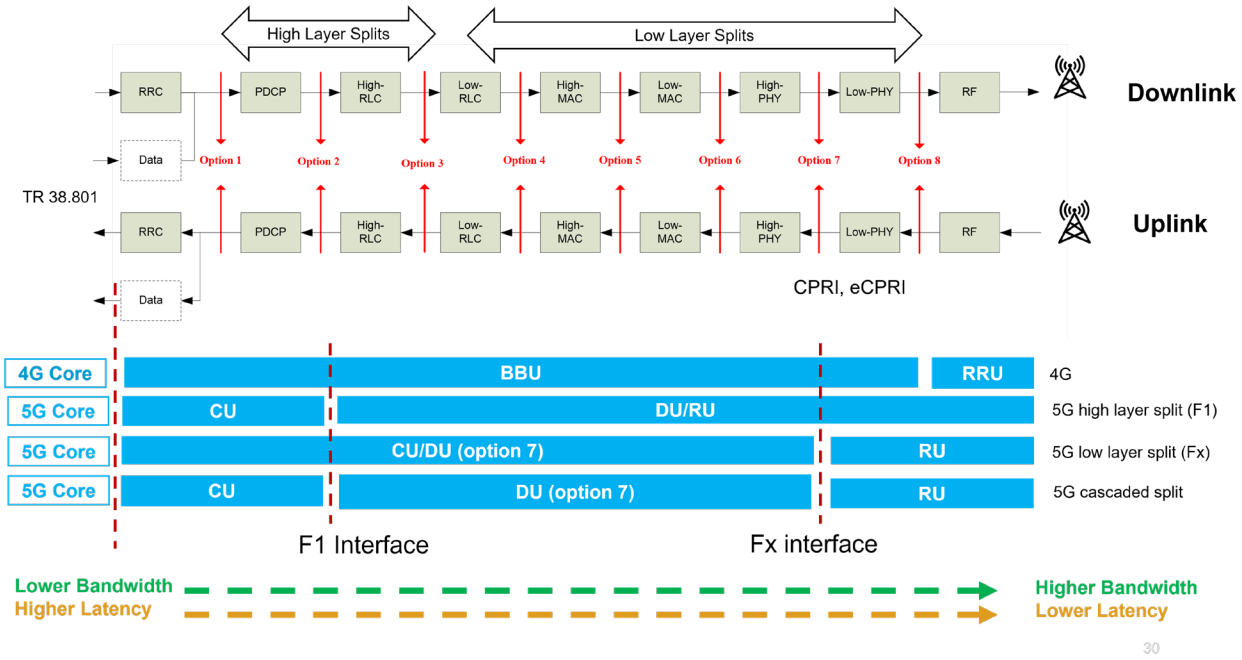


Figure 4. 3GPP 5G RAN functional split options

Various configurations and usages of 5G RAN functional splits are compared with the traditional monolithic 4G BBU-RU architecture in Figure 5. The choice of which option to use depends on several factors including deployment scenarios, network constraints, and intended supported services.

- Support for specific Quality of Service (QoS) per offered services (e.g., low latency, high throughput)
- Support for specific user density and load per geographical area (RAN level coordination need)
- Transport network availability – ideal to non-ideal
- Application type – Real-time or Non- Real Time Radio Access
- Network feature requirement – such as Carrier Aggregation, Coordinated Multipoint Access (CoMP), Enhanced Inter-Cell Interference Coordination (eICIC)

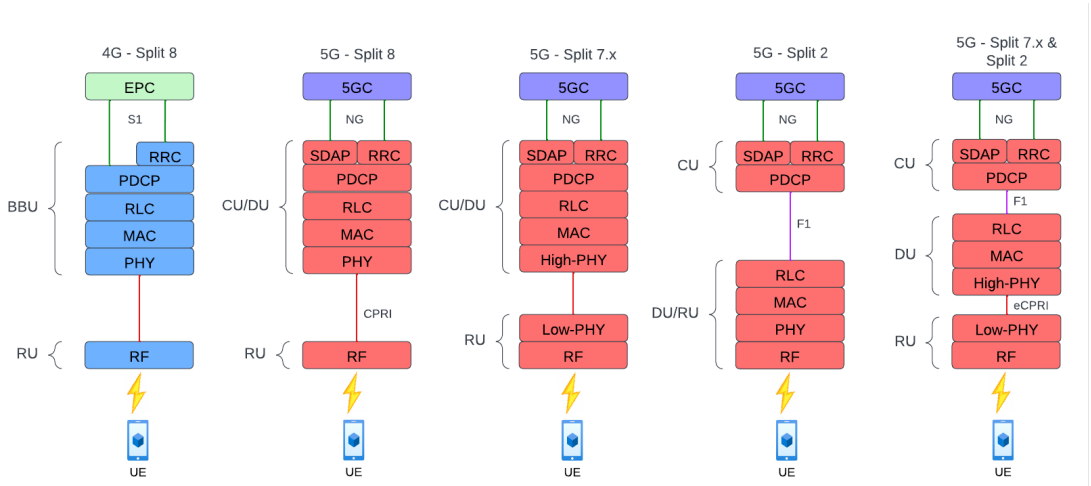


Figure 5. Deployment configurations of 5G RAN functional split options

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3GPP only supports split option 2, which is commonly used by many operators today with an integrated DU and RU and standalone CU to support high bandwidth, low latency deployment scenarios such as densification of small cells. As previously mentioned, the 3GPP also supports gNB-CU-UP and gNB-CU-CP on the F1 interface as further decomposition of the CP and UP to the option 2 split as shown in Figure 6.

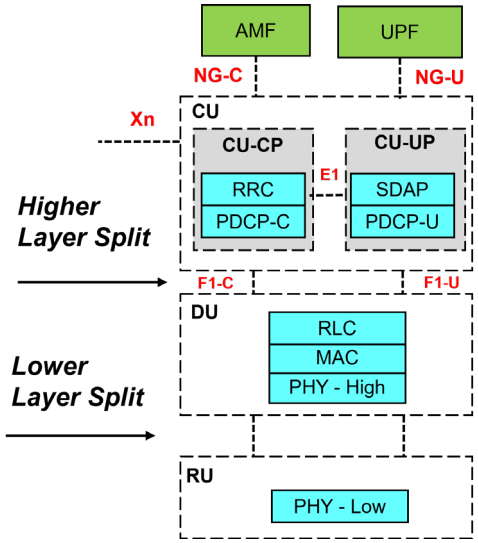


Figure 6. Lower and higher layer splits with separate user and control plane in the CU

Furthermore, as part of this transition, the CPRI specification was improved to enhanced CPRI (eCPRI), a new protocol to connect the DU and RU that uses ethernet as its physical layer and runs over fiber. eCPRI provides a faster link rated at 10-25 Gbps and a more efficient use of fronthaul bandwidth (point to multipoint from point to point). 3GPP does not address the fronthaul interface.

2.3 Virtualized RAN (vRAN)

Another transformation of the RAN is the virtualization of the RAN functions by decoupling the software from the hardware using virtualized network functions, referred to as virtual RAN (vRAN). This includes using network function virtualization (NFV) or containers to deploy CU and DU over bare metal servers such as an x86 server (e.g., commercial-off-the-shelf (COTS)) and operating network functions as software on a server platform. Figure 7 illustrates various vRAN deployment architectures using COTS hardware, including option 8, option 7x, and option 2 and 7.x splits.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

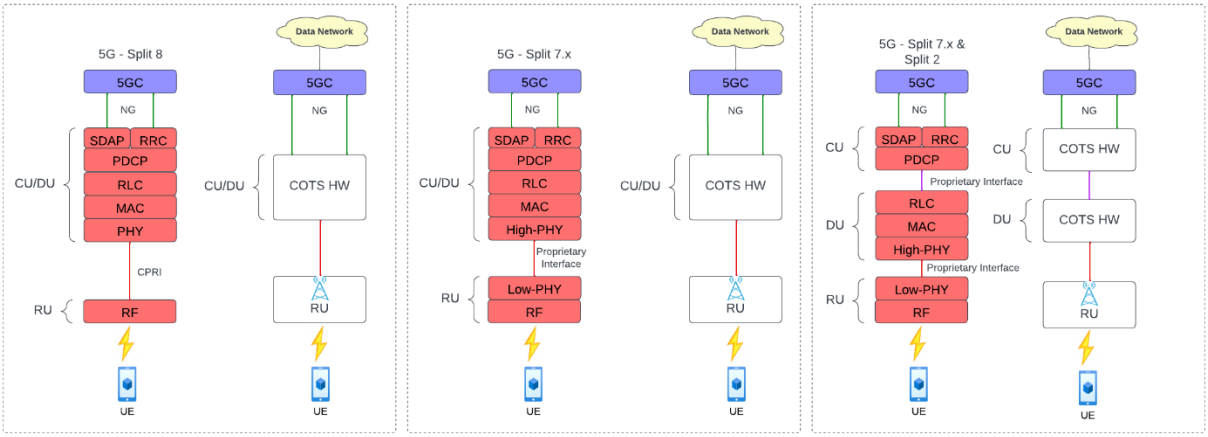


Figure 7. Various vRAN Deployment Architectures

3. Open RAN

Open RAN is the next shift in RAN architecture that builds upon RAN disaggregation and virtualization by adding the key elements of open interfaces, intelligence (i.e., automation, artificial intelligence, machine learning), and cloudification. When all these elements are integrated into an Open RAN architecture, it enables service providers the opportunity to create solutions for open, multi-vendor networks with enhanced flexibility and control.

Open RAN conforms to O-RAN Alliance specifications that build on the baseline functionality defined by 3GPP required for a complete and high-quality specification in order to guarantee multi-vendor inter-operability. This includes many of the 5G RAN functional split options, not including option 2, and gives the Open RAN ecosystem the flexibility to design any of these splits to meet deployment architectures defined by MNOs.

3.1 O-RAN ALLIANCE

The O-RAN Alliance has emerged as a prominent force in the Open RAN ecosystem, driving industry-wide collaboration and standardization efforts. By promoting openness, interoperability, and intelligence in RAN deployments the Alliance is reshaping the landscape of wireless networks, empowering operators with greater choice, flexibility, and innovation opportunities.

The O-RAN Alliance introduces the fronthaul interface, automation, and the RAN intelligent controller (RIC) to operate, optimize, and conduct configuration/performance management. There are several Working Groups (WGs) that address the interfaces, test cases, and specifications. Figure 8 illustrates the 3GPP defined interfaces and the O-RAN Alliance interfaces. O-RAN Alliance specifications are primarily based on 3GPP specifications and requirements.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

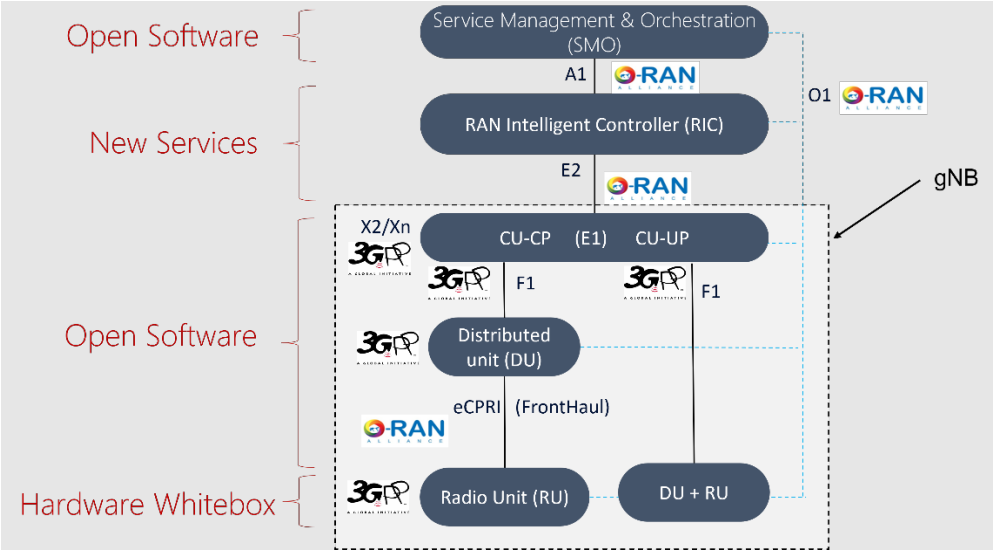


Figure 8. 3GPP and O-RAN Alliance coexistence

3.2 Open RAN Architecture

Open RAN logical architecture refers to the high-level design and organization of components within an Open RAN ecosystem. It defines the logical relationships, interfaces, and functionalities between different elements to enable interoperability, flexibility, and scalability. The Open RAN logical architecture is designed to decouple hardware and software, allowing for the integration of best-of-breed components from multiple vendors. The overall vision of Open RAN architecture is to utilize RAN components as a virtual software application capable of working on any bare metal hardware operated with orchestration and automation. This is accomplished with the key elements of virtualization, disaggregation, open interfaces, intelligence, and cloudification, which are foundational components to an Open RAN architecture.

At the core of the Open RAN logical architecture, as shown in Figure 9, is the concept of functional split, described previously, which separates the RAN functions into different components. These components typically include the RU, DU, CU, and Management and Orchestration (MANO) layer. The RU is responsible for radio signal transmission and reception, while the DU handles baseband processing and digital signal processing tasks. The CU centralizes functions such as scheduling, beamforming, and network optimization. The MANO layer oversees the management and orchestration of the entire Open RAN system, including resource allocation, configuration management, and service provisioning.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

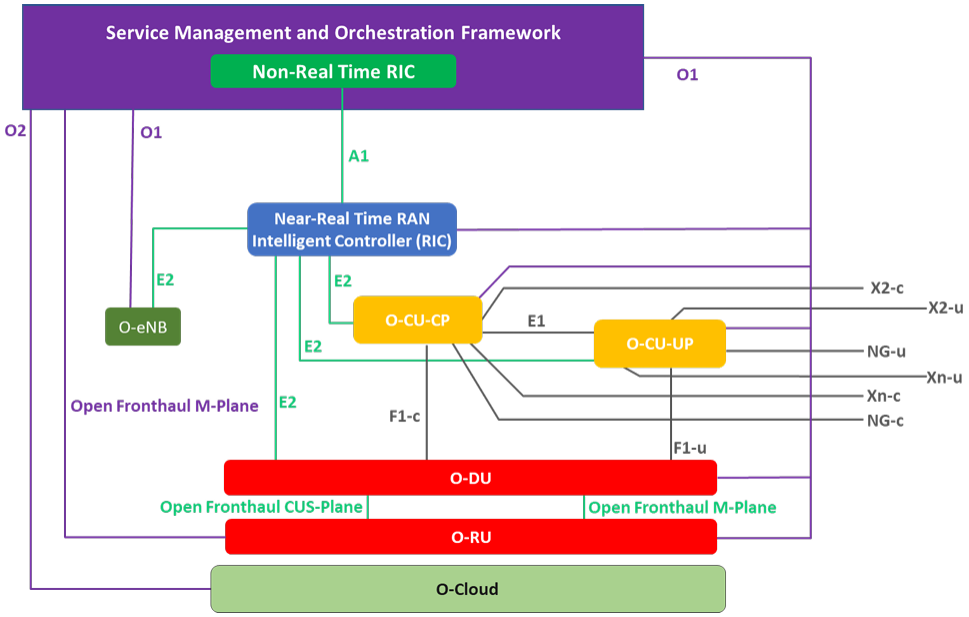


Figure 9. Logical Open RAN Architecture

Open interfaces and standardized protocols play a crucial role in the Open RAN logical architecture. These interfaces enable seamless communication and interoperability between different components from various vendors. By adopting open interfaces, Open RAN promotes vendor diversity, encourages competition, and reduces vendor lock-in. The standardized protocols ensure that the components can effectively communicate and exchange information, facilitating the dynamic allocation of resources and the efficient operation of the network.

Overall, the Open RAN logical architecture provides a blueprint for building open, flexible, and interoperable RAN solutions. It enables network operators to deploy and manage RAN components from different vendors, fostering innovation, reducing costs, and promoting a more competitive telecommunications ecosystem.

Cloudification

One key element to Open RAN is the adoption of Containerized Network Functions (CNFs) and the shift towards virtualization and Cloudification. The transition from Physical Network Functions (PNF) to Virtual Network Functions (VNF) to CNFs represents a significant evolution in network architecture and deployment models. Initially, traditional telecommunication networks relied on PNF, which involved dedicated hardware for each network function. However, the industry recognized the need for more flexibility, scalability, and cost-efficiency, leading to the adoption of VNFs. VNFs virtualized network functions, enabling them to run on standard servers and be dynamically instantiated or scaled as needed. This shift allowed operators to reduce hardware dependencies and achieve greater agility.

The subsequent transition from VNFs to CNFs further advances network virtualization and cloudification. CNFs leverage containerization technology, such as Docker or Kubernetes, to encapsulate network functions into lightweight, isolated containers. CNFs offer numerous benefits, including faster deployment times, efficient resource utilization, and easier management and orchestration. By embracing CNFs, operators can leverage the scalability and portability of containerization, enabling them to build cloud-

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

native architectures and take full advantage of cloud-based infrastructure and services. This is illustrated in Figure 10.

Cloud RAN is yet another evolution in 5G RAN in that the hardware will be a containerized platform that will host RAN radio stacks as a CNF application. This further separates the software from the hardware.

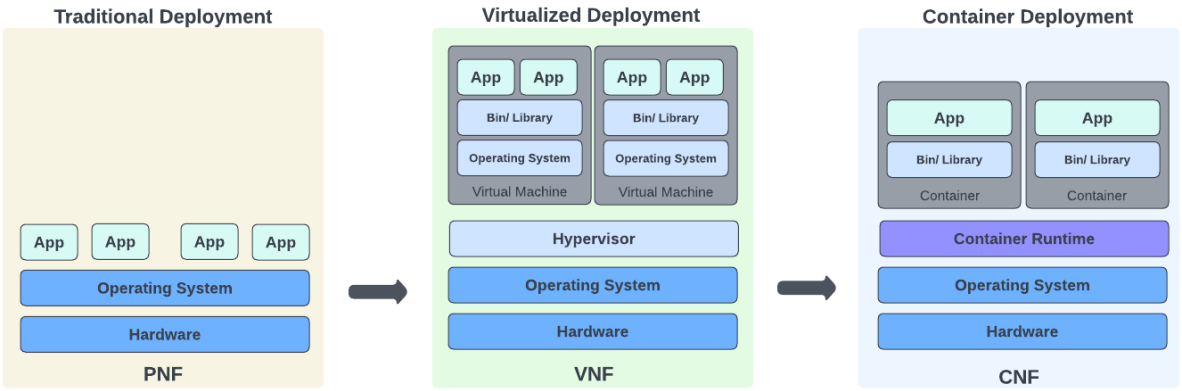


Figure 10. Traditional to Virtualized to Containerized Platforms

Open RAN has embraced the concept of CNF as part of its journey towards virtualization and cloudification. Open RAN aims to disaggregate and open up traditionally closed and proprietary network components, fostering interoperability and innovation. CNFs align well with the principles of Open RAN by enabling the deployment of flexible, scalable, and easily manageable network functions. By embracing containerization and cloud-native architectures, Open RAN networks can leverage the benefits of virtualization and drive the next generation of open and innovative telecommunications infrastructure.

3.3 Key Benefits

When fully integrated, Open RAN offers many key benefits that are critical to service providers' ability to deploy RAN architecture in a flexible, customized, open manner. Many of the key benefits and attributes are described below.

1. *Enhanced flexibility*: Open RAN enables multi-vendor interoperability, allowing network operators to choose hardware and software components from different vendors. This flexibility promotes innovation, competition, and prevents vendor lock-in.
2. *Cost reduction:* By decoupling hardware and software components, Open RAN reduces the dependency on proprietary equipment and enables the use of off-the-shelf, commodity hardware. This leads to cost savings in network deployment and maintenance.
3. *Improved scalability*: Open RAN architecture enables the deployment of networks in a more modular and scalable manner. Operators can easily add or remove network elements based on demand, optimizing network resources, and reducing capital expenditures.
4. *Accelerated innovation*: Open RAN encourages the development of new solutions by allowing different vendors to contribute and integrate their innovations into the network ecosystem. This fosters rapid development, improved feature sets, and faster time-to-market for new technologies.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. *Increased vendor diversity:* Open RAN promotes a diverse vendor ecosystem by enabling the participation of smaller or niche vendors who may specialize in specific network components. This diversification helps prevent monopolies and encourages healthy competition.
6. *Interoperability and standardization*: Open RAN adheres to open standards and specifications, ensuring interoperability between different vendors' equipment. This facilitates seamless integration and compatibility between components, leading to more efficient and reliable network operations.
7. *Easier network optimization*: With Open RAN, operators have greater visibility and control over network performance. They can optimize specific network functions, such as radio resource management, to improve efficiency, capacity, and overall user experience.
8. *Rapid deployment and upgrades*: Open RAN simplifies network deployment and upgrade processes by enabling the use of software-defined approaches. Operators can remotely deploy software updates and new features, minimizing service disruptions and reducing operational costs.
9. *Enhanced network customization:* Open RAN allows operators to customize network features and functionalities based on their specific requirements. This customization capability enables tailored solutions for different use cases, such as private networks or IoT applications.
10. *New Technology Interworkin*g: Open RAN facilitates integration with emerging technologies such as cloud computing, artificial intelligence, and edge computing. This integration enables new methods for network optimization, intelligent traffic management, and enhanced user experiences.

3.4 Deployment Scenarios

Open RAN architecture coupled with 5G RAN split options provide flexibility to the operator to customize an architecture to meet various deployment scenarios with minor enhancements to the RAN. The list below describes some common Open RAN deployment scenarios:

1. *Macro Cells*: Open RAN can be deployed in macro cell environments, which typically cover larger geographic areas and provide wide-area coverage. By adopting Open RAN, operators can introduce flexibility and vendor interoperability in their macro cell networks, enabling the integration of different hardware and software components from various vendors.

2. *Small Cells*: Open RAN can also be used in small cell deployments. Small cells are low-power, short-range base stations that enhance network capacity and coverage in densely populated areas or indoor environments. Open RAN allows for the disaggregation of small cell hardware and software, promoting interoperability and enabling operators to choose the best components for their specific needs.

3. *Private 5G Networks*: Open RAN can facilitate the deployment of private 5G networks. These networks are typically used by enterprises, industries, or government entities for dedicated connectivity and specific use cases like industrial automation, smart cities, or critical infrastructure. Open RAN provides the flexibility to customize and optimize network solutions for private deployments, enabling organizations to tailor their networks to their unique requirements.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. *Rural Connectivity*: Open RAN can play a crucial role in expanding connectivity in rural or underserved areas. Traditional network equipment can be expensive and challenging to deploy in such regions. Open RAN offers a more cost-effective and flexible solution, allowing operators to leverage a mix of hardware and software components from different vendors to build and operate networks in rural areas.

5. *Neutral Host Networks*: Open RAN can support neutral host network deployments. Neutral hosts provide shared infrastructure and services that enable multiple operators or service providers to utilize the same network infrastructure. Open RAN facilitates the integration of multiple operators' equipment and allows for interoperability, enabling efficient sharing of network resources and reducing deployment costs.

6. *Network Densification*: Open RAN can assist in network densification efforts by deploying additional base stations in high-traffic areas to increase capacity and improve user experience. Open RAN's flexibility and interoperability enable operators to deploy cost-effective and efficient solutions for network densification, accommodating growing data demands and enhancing network performance.

The flexibility, interoperability, and cost efficiencies offered by Open RAN make it a compelling choice for various network environments, helping operators address specific challenges and deliver improved connectivity and services.

4. Open RAN Ecosystem Activity

The Open RAN ecosystem is fluid with a broad range of industry activity that includes commercial MNO deployments, industry testbeds and trials, and government policy and investment. These activities demonstrate proof-of-concept and end-to-end validation of Open RAN subsystems and systems and show the progress towards a full set of Open RAN features. This section provides highlights of some examples of industry activities.

4.1 Global Mobile Network Operator Activity

Several global MNOs have deployed greenfield and brownfield commercial Open RAN networks while others have made public announcements of field trials, partnerships, and commitments to Open RAN deployments. This section highlights representative examples of these industry activities and addresses the progress and level of fully integrated Open RAN elements in each.

Greenfield Deployments

Two MNOs lead by deploying the industry's first greenfield Open RAN commercial networks. The first, Rakuten, a Japanese MNO[2], integrated Open RAN elements into commercially launched Open RAN 4G and 5G networks using disaggregated, virtualized RAN on a cloud native architecture using accelerated automation and intelligence. The network offers virtualized Open RAN software that can be implemented on any COTS component. While this network has successfully proven several Open RAN elements, it has yet to prove fully integrated Open RAN across multi-vendors on a cloud native platform with a fully developed RIC.

[2] https://symphony.rakuten.com/open-ran

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The second, DISH Wireless, a US MNO, integrated Open RAN elements into a commercially launched Open RAN 5G Standalone (SA) network that includes disaggregated, virtualized features on a cloud native architecture using automation and intelligence[3]. While this network succeeded in utilizing these Open RAN elements, it is still validating the remainder of the Open RAN elements such as open interfaces, using commercial-off-the-shelf products, and multi-vendor interoperability with a fully developed RIC, at scale and with reliability.

Brownfield Deployments

A leading Japanese MNO[4] [5], DoCoMo, integrated Open RAN elements into their existing commercial 4G and 5G network deployments using disaggregation, virtualized, and intelligent (i.e., MANO) features with multiple vendors, however it uses purpose-built hardware and partially compliant Open RAN interfaces with partnered vendors within its network. DoCoMo announced that their continued brownfield deployment efforts will include development of software onto COTS hardware, fully compliant open interfaces, cloud native architecture, and continued automation and application development of the SMO/RIC.

A leading US MNO, Verizon, deployed a disaggregated and virtualized 5G RAN on COTS hardware on its own cloud platform using automation and is working towards development of complaint open interface between Open RAN elements.[6] [7] Their goal is to use Open RAN architecture in parts of the network serving dense urban areas.

Continued Commitment by MNOs

Many of the world's largest operators have made public announcements of field trials, partnerships, and commitments to Open RAN deployments.

A leading European MNO, Vodafone, deployed a brownfield Open RAN network trial[8] that utilizes a disaggregated RAN with virtualized software and Open RAN compliant RUs across its trial 4G and 5G network[9]. The RUs support low and mid band spectrum and include massive MIMO capability. Vodafone's initiative is to use Open RAN in at least 2,500 sites by the end of 2027[10].

In 2021, a group of five leading European MNOs (Vodafone, Deutsche Telekom (DT), Orange, Telefónica, and TIM) signed a Memorandum of Understanding (MoU) and formed the Open RAN MoU Group within TIP. The group's focus is to develop and define Open RAN solutions to further implementation of Open RAN technologies and networks. The formation of the group demonstrates the MNOs' commitment to Open RAN deployment and addresses collaborative Open RAN technical priorities in multiple Releases[11]. Release 1 (2021) focused on the main scenarios and technical requirements for each of the building blocks of a multi-vendor RAN. Release 2 (2022) mainly focused on intelligence, orchestration, transport, and cloud infrastructure, addressing energy efficiency goals and targets to support sustainable Open RAN. Release 3 (2023) primarily focused on developing further

[3] https://about.dish.com/2023-02-23-DISH-Wireless-Expands-Cloud-Native-Open-RAN-Network-With-Mavenir-Open-vRAN-Software-Solutions
[4] https://www.docomo.ne.jp/english/info/media_center/pr/2019/0918_00.html
[5] Open_RAN Docomo.pdf
[6] https://www.mobileworldlive.com/featured-content/home-banner/verizon-gears-up-for-open-ran-field-trials/
[7] https://news.samsung.com/us/samsung-verizon-charge-ahead-with-vran/
[8] https://www.vodafone.com/about-vodafone/what-we-do/technology/open-ran
[9] https://news.samsung.com/global/vodafone-and-samsung-begin-mass-open-ran-rollout-across-the-united-kingdom
[10] https://www.sdxcentral.com/articles/news/samsung-strikes-5g-vran-open-ran-deal-with-vodafone/2023/09/
[11] https://telecominfraproject.com/openran-mou-group/

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

requirements on Service Management and Orchestration (SMO) and RIC and enhancements to security, cloud infrastructure, O-CU/O-DU, and O-RU.

A leading European MNO, Deutsche Telekom, completed an 'Open RAN Town' field trial[12] with several vendors to investigate the development, capability, and operations of a fully integrated multi-vendor Open RAN solution based on O-RAN Alliance specifications. The trial resulted in a successful integrated Open RAN solution with key learnings of various aspects of Open RAN that include maturity, integration, scalability, security, and energy efficiency.

4.2 Industry Test Beds

There are several leading industry test beds that are instrumental in advancing and testing proof of concept and compliance to Open RAN architecture, which include OTICs, TIP OpenRAN, and SONIC labs.

Open Testing and Integration Centres (OTICs)

The O-RAN Alliance established OTICs[13] which are vendor-independent, open, and neutral lab environments that support Open RAN ecosystem development to advance the adoption of Open RAN. There are 15 globally authorized OTICs, one of which includes CableLabs/Kyrio as the first established OTIC in North America.

OTICs provide various levels of conformance, interoperability, and end-to-end testing (e2e) of Open RAN components and subsystems against O-RAN Alliance interface specifications and award O-RAN Alliance certificates and badging, which include:

- Conformance testing: OTICs can test the conformity of RAN equipment to O-RAN Alliance interface specifications.
- Interoperability testing: OTICs can test the interoperability of RAN equipment from different vendors (or the same vendor) using O-RAN Alliance interface specifications.
- Performance testing: OTICs can test the performance of RAN equipment under different load conditions.
- Security testing: OTICs can test the security of RAN equipment against known vulnerabilities.

As of this writing, twelve badges have been awarded across six different vendors for open fronthaul conformance, interoperability, and e2e validation, eight of which have been awarded in 2023. This demonstrates that open fronthaul is commercially available and the maturity of Open RAN is gaining momentum.

OTICs also host periodic Global Plugfests[14], co-hosted by the O-RAN Alliance, to showcase proof of concept, interoperability, or end-to-end validation of Open RAN solutions. Plugfests demonstrate many levels of validation of Open RAN solutions with multiple vendors.

TIP Community Labs

TIP Community Labs were established by various players in the ecosystem (i.e., operators, vendors, academia) to support the development, testing, and deployment of open, disaggregated Open RAN

[12] https://www.telekom.com/en/company/details/bundled-in-a-white-book-learnings-from-o-ran-town-1026846
[13] https://www.o-ran.org/testing-integration#learn-otic
[14] https://plugfestvirtualshowcase.o-ran.org/2023/SPRING

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

solutions[15]. There are 14 TIP Community Labs worldwide, each having their own set of test equipment and specialized lab focus with industry partners. TIP developed their own certification process called 'Open RAN System Certification process (SCOPE)' that is complementary to the O-RAN Alliance certification process. It builds on O-RAN Alliance certification and test plans and offered at a TIP-accredited Open RAN system certification lab that is focused on pre-launch validation and post-launch regression and release upgrade validation.

SONIC Labs

The United Kingdom (UK) developed the SmartRAN Open Network Interoperability Centre (SONIC) Labs as an independent lab to host the development of Open RAN solutions through Digital Catapult[16], a company funded by the UK Department for Science, Innovation and Technology (DSIT), Department of Culture Media Sport (DCMS), and Ofcom.

SONIC Labs completed two phases of interoperability testing with multiple Open RAN vendors. The first phase, SONIC-1, was an initiative to conduct end-to-end testing of Open RAN solutions with multiple vendors across three separate field testbeds. The goals of SONIC-2 are to continue integration, multi-vendor interoperability of Open RAN solutions in indoor and outdoor environments, encourage innovation towards quicker deployment, and economic study.

4.3 Government Policy and Investment

Countries across the globe, including the United Kingdom, United States, Germany, Canada, Australia, India, South Korea, and Singapore are establishing policies and making available funding to the Open RAN ecosystem to advance the technology, development, and technology of Open RAN. Several governments generally view the RAN infrastructure vendor supply chain as an oligopoly and consider Open RAN as an economic, security, and supply chain issue, not a technical issue. The intent of many of the government programs is to de-risk the vendor oligopoly and supply chain and enable more resilient and secure networks.

United Kingdom

The UK DCMS announced its '5G Supply Chain Diversification Strategy'[17] in 2020 that consists of three focus areas: (1) supporting incumbent suppliers, (2) attracting new suppliers into the UK market, and (3) accelerating the development and deployment of open-interface solutions. It is backed by a £250 million investment in the project, which is highly focused on the development of Open RAN solutions.

United States

The United States government views mobile network infrastructure suppliers consisting of closed source software and hardware with proprietary solutions with vendor-lock and few choices to operators and subject to security and supply chain concerns. To overcome this vendor-lock, the National Telecommunications and Information Administration's Institute for Telecommunication Sciences (NTIA/ITS) in collaboration with the Department of Defense Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)) created a multi-vendor interoperability event called the '5G

[15] https://telecominfraproject.com/clabs/
[16] https://www.digicatapult.org.uk/about/press-releases/post/digital-catapult-and-ofcom-unveil-first-of-its-kind-open-ran-testing-facilities-in-the-uk/
[17] https://www.gov.uk/government/publications/5g-supply-chain-diversification-strategy/5g-supply-chain-diversification-strategy

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Challenge' to influence the adoption of Open RAN solutions over a two-year timeframe (2022 and 2023).[18] The 5G Challenge included a total of $10 million in prize awards to contestants that demonstrate compliance to various levels of conformance, interoperability, performance, and reliability.

To further develop wireless innovation, promote competition, and strengthen supply chain resilience, Congress created the Public Wireless Supply Chain Innovation Fund[19], a $1.5 billion grant program. The NTIA is leading the implementation of the Innovation Fund and issued the first grants of the ten-year program in August 2023.[20]

Singapore

Singapore's Infocomm Media Development Authority (IMDA) in partnership with the National Research Foundation, Singapore (NRF) created a S$70 million Future Communications R&D Programme (FCP) that built an OTIC lab at the Singapore University of Technology and Design. The government investment is to promote the technological capabilities of Singapore and support the global Open RAN ecosystem.[21]

Germany

The German Federal Ministry for Digital and Transport (BMDV) has invested €17 million funding to the i14y Lab in Berlin[22]. The i14y Lab is an O-RAN Alliance OTIC that hosts Plugfests to validate the technical readiness of individual vendor solutions and explore multi-vendor interoperability in new or refined test scenarios. The i14y Lab is a joint effort with Deutsche Telekom, Vodafone, and Telefonica along with some of the vendor community to focus on test and integration of disaggregated network solutions.

Coordinated Efforts by Governments

UK, US, Canada and Australia

The UK, US, Canada, and Australia released a 'Joint statement on telecommunications supplier diversity'[23] that states a joint commitment to "ensuring the security and resilience of our telecommunications networks, including by fostering a diverse supply chain and influencing the development of future telecommunications technologies such as 6G. Collectively, we recognize that open and interoperable architectures are one way of creating a more open, diverse and innovative market."

US and India

The US and India share a commitment to "a vision of secure and trusted telecommunications, resilient supply chains, and global digital inclusion." The countries committed to Joint Task Forces focused on Open RAN, including an Open RAN field trial in India.[24]

---

[18] https://5gchallenge.ntia.gov/
[19] https://www.ntia.gov/page/innovation-fund
[20] https://www.ntia.gov/page/innovation-fund/grant-programs/notice-of-funding-opportunity
[21] https://www.sutd.edu.sg/Research/Research-News/2023/2/south-east-Asia-first-O-RAN-Open-testing-centre
[22] https://www.i14y-lab.com/
[23] https://www.gov.uk/government/publications/joint-statement-between-the-united-kingdom-australia-canada-and-the-united-states-of-america-on-telecommunications-supplier-diversity/joint-statement-on-telecommunications-supplier-diversity
[24] https://www.whitehouse.gov/briefing-room/statements-releases/2023/09/08/joint-statement-from-india-and-the-united-states/

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

US, South Korea, and Japan

The leaders of the US, South Korea and Japan committed to bolstering trilateral collaborative research and development in science, technology engineering, and mathematics, including Open RAN.[25]

These recent commitments and investments by many governments demonstrate the growing importance of Open RAN across the globe.

5. Multi-Vendor Open RAN integration at CableLabs

CableLabs and its Kyrio subsidiary have a strong history of holding vendor agnostic, neutral host interoperability events for the industry. Starting with interoperability events for DOCSIS technologies, CableLabs now hosts events for optical technologies, Wi-Fi technologies, Citizens Broadband Radio Service (CBRS) and Spectrum Access System (SAS) interoperability for 3GPP technologies, and most recently Open RAN.

The key to accelerating Open RAN adoption is to demonstrate comparable performance of an integrated 5G RAN with a multi-vendor setup utilizing a disaggregated, virtualized RAN. CableLabs and Kyrio were the host lab for the 5G Challenge in 2022 and 2023 and provided a state-of-the-art test lab, expert staff in wireless networks and Open RAN, and performed all of the testing of contestant's systems. In addition, CableLabs and Kyrio provided expertise to NTIA-ITS and DoD when designing the challenge, the rules, and the grading process. CableLabs and Kyrio also became the first established OTIC Lab in North America and hosted several O-RAN Alliance Plugfests. From these experiences, CableLabs is recognized as an industry leader in the Open RAN ecosystem as system integrators, independent and neutral host of a fully featured Open RAN test lab, Open RAN subsystem and system test experts, and contributors to Standards Development Organizations (SDOs) (i.e., O-RAN Alliance, TIP, 3GPP). Our labs and expertise help industry develop the building blocks required to accelerate the adoption of Open RAN. This section provides an overview of the lab, completed tests and results, observations, and lessons learned from these Open RAN activities.

5.1. Lab Overview

The Open RAN Interoperability lab is a virtual lab that is part of CableLabs 10G lab and consists of multiple racks housing servers and test equipment that includes a 5G core emulator, User Equipment (UE) emulators, CU, DU, and RU emulators, fronthaul analyzer, Radio Frequency (RF) spectrum analyzers, RF vector signal analyzers and generators, a UE diagnostics tool, a commercial 5G SA core, and two 5G vRANs, as shown in Figure 11.

[25] https://www.whitehouse.gov/briefing-room/statements-releases/2023/08/18/the-spirit-of-camp-david-joint-statement-of-japan-the-republic-of-korea-and-the-united-states/

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 11. CableLabs/Kyrio Test Lab Facilities

The lab test equipment offers a full set of Open RAN features from two different Open RAN test emulator vendors, Viavi and Keysight, that demonstrate compliance to 3GPP and O-RAN Alliance specifications as well as development and proof of concept activities. This configuration provides flexibility and capacity in testing by offering two 'swim lanes' of simultaneous testing to support multiple vendors.

As shown in Figure 12, the test equipment is configured to support (1) standalone subsystems for conformance and interface testing, (2) paired subsystems for interoperability (IOT) testing, and (3) end-to-end integration, interoperability, functional, and performance testing.

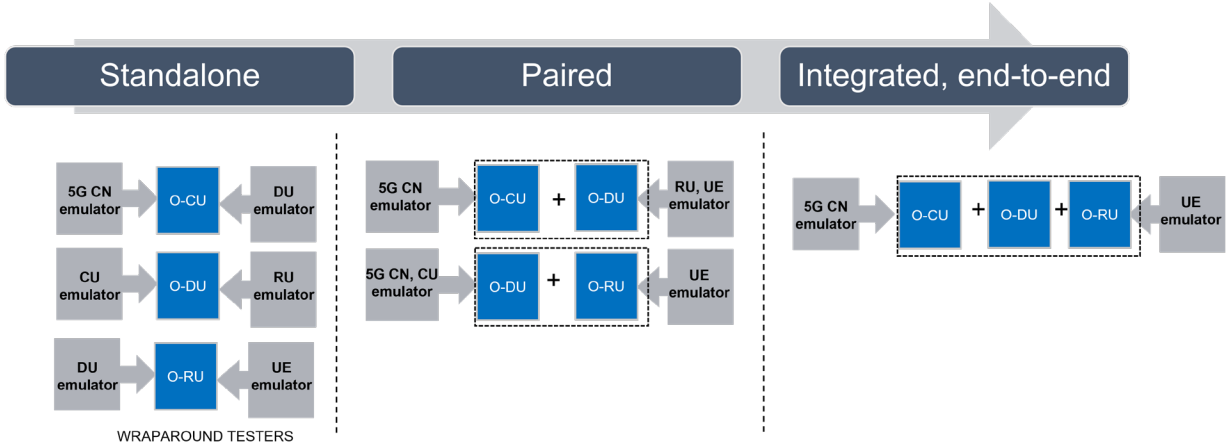


Figure 12. Open RAN subsystem/system test configurations

As shown in Figure 13, the lab supports the entire range of OTIC certification and badging tests and testing of all the O-RAN Alliance interfaces, conformance, IOT, and e2e interoperability across single and paired vendors in addition to multiple Open RAN vendors. The lab capabilities are used as a framework for Open RAN development and interoperability testing beyond the 5G Challenge and OTIC Plugfests, including internal R&D projects and requests from Open RAN vendors and MNOs.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

	Subsystem	IOT	E2E System	Functional	Performance	Certification	Badging
O-RU	■	■	■	■	■	■	■
FHGW			□	□	□		
O-DU	■	■	■	■	■		■
O-CU-CP	■	■	■	■	■		■
O-CU-UP	■	■	■	■	■		■
O-CU-CP & O-CU-UP	■	■	■	■	■		■
O-RU & O-DU	■	■	■	■	■		■
O-DU & O-CU-CP & O-CU-UP	■	■	■	■	■		■

	RF	OFH-C	OFH-U	OFH-S	OFH-M	F1-C	F1-U	Xn-C	Xn-U	E1	NG-C	NG-U
O-RU	■	■	■	■	■							
FHGW		■	■	■	■							
O-DU		■	■	■	■	■	■					
O-CU-CP						■		■		■	■	
O-CU-UP							■		■	■		■
O-CU-CP & O-CU-UP						■	■	■	■		■	■
O-RU & O-DU	■	■	■	■	■	■	■					
O-DU & O-CU-CP & O-CU-UP		■	■	■	■			■	■		■	■
Transport xHAUL		□	□	□	□	■	■	■	■	□	□	□

Figure 13. CableLabs/Kyrio OTIC test and certification capabilities

5.2. Industry Test Results

The 2022 5G Challenge successfully demonstrated multi-vendor end-to-end interoperability across an Open RAN system comprised of three different 'cold' vendors (never been integrated with each other in the past) and conducted data sessions using various protocols (UDP, TCP, RTP, streaming video). This was accomplished in a 5-week period (compared to several months in other industry labs) which demonstrated the maturity and variety of available Open RAN vendors in the ecosystem. Details of the findings are described in the NTIA Technical Memorandum 23-568: 5G Challenge Preliminary Event: Evaluating Modular, Interoperable, Multi-Vendor, Open RAN Solutions[26].

The 2023 5G Challenge demonstrated multi-vendor end-to-end interoperability with four pairs of Open RAN systems (i.e., CU/DU and RU) with seven different vendors conducting data sessions using various protocols (UDP, TCP, RTP, streaming video), traffic loading, and stability in varying RF conditions. Two pairs of multi-vendor Open RAN systems continued on to demonstrate mobile handover (Xn and NG) between the two different Open RAN gNBs, which is known to be the industry's first. These functional, performance, loading, stress, and reliability tests demonstrate the maturity and variety of available Open RAN vendors in the ecosystem.

CableLabs/Kyrio OTIC Plugfests[27] followed a similar track by conducting conformance, interface, and e2e interoperability tests with nine different vendors. Tests included multi-vendor e2e interoperability (mobility, performance, load, QoS), standalone conformance, RIC (xApp, energy savings), and paired integration. For several years, over 25 different Open RAN vendors testing in 15 different OTIC labs have been conducting proof-of-concept, interface validation, paired integration, e2e interoperability, and RIC

[26] NTIA Technical Memorandum 23-568: 5G Challenge Preliminary Event: Evaluating Modular, Interoperable, Multi-Vendor, Open RAN Solutions
[27] https://plugfestvirtualshowcase.o-ran.org/2023/O-RAN_Global_PlugFest_hosted_by_University_of_New_Hampshire

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

testing. This level of activity, including a recent uptick, demonstrates the progress and maturity of the Open RAN ecosystem[28].

5.2.1. Standalone Tests

As shown in Figure 14, wraparound testers are used to conduct conformance and interface testing with a single CU, DU, or RU as the Device Under Test (DUT). Wraparound testers are used to validate compliance to 3GPP and O-RAN Alliance specifications at an interface level by emulating all the other components encompassing the DUT. The goal is to validate the basic functionality against the specifications with regards to the messages and Information Elements (IEs) that are carried across the interfaces supported by the subsystem. O-RAN Alliance Work Group 4 (WG4) and Work Group 5 (WG5) specifications and test cases are used as the baseline for tests and reference with the lab setup shown in Figure 14.

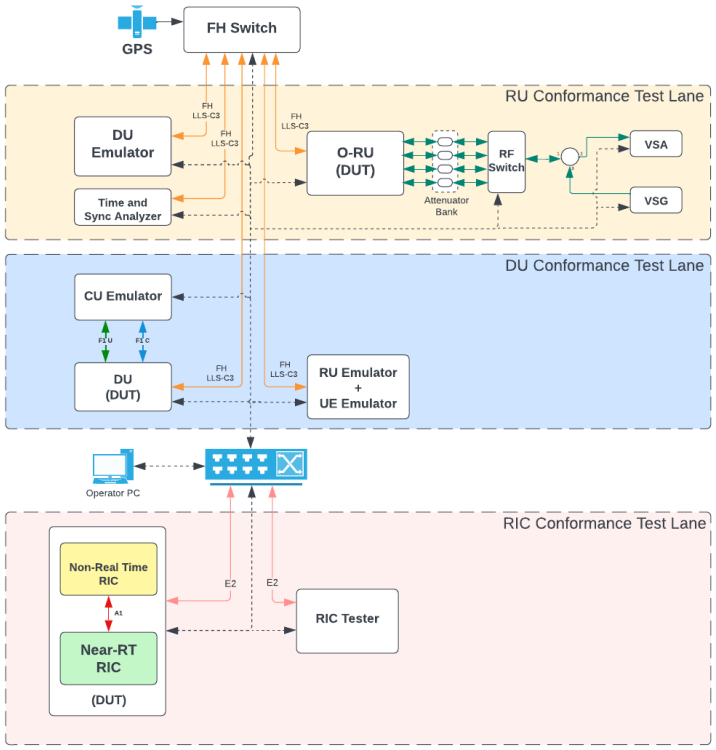


Figure 14. CableLabs/Kyrio lab setup for standalone wraparound tests

[28] https://www.o-ran.org/what-we-do

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.2.2. End-to-end Integration

As shown in Figure 15, Core and UE emulators are used to conduct end-to-end integration testing with paired Open RAN subsystems and systems. O-RAN Alliance WG5 and Test and Integration Focus Work Group (TIFG) specifications and test cases are used as the baseline for tests and reference with the lab setup shown in Figure 15. Objectives of the multi-vendor sub-system tests are summarized below:

- Interoperability and baseline functionality to successfully set up an e2e network that could pass different data traffic types such as TCP, UDP, FTP, ICMP, and RTP.
- e2e network performance with a multi-vendor setup that includes single UE stationary and mobile uni-and bi-directional throughputs in different radio conditions, multiple UE aggregated cell throughput, e2e network loading, stability, and reliability using different traffic models to simulate real world scenarios.
- Mobility scenarios for connected mode mobility and idle mode reselection for both intra and inter frequency over N2 and Xn interfaces.

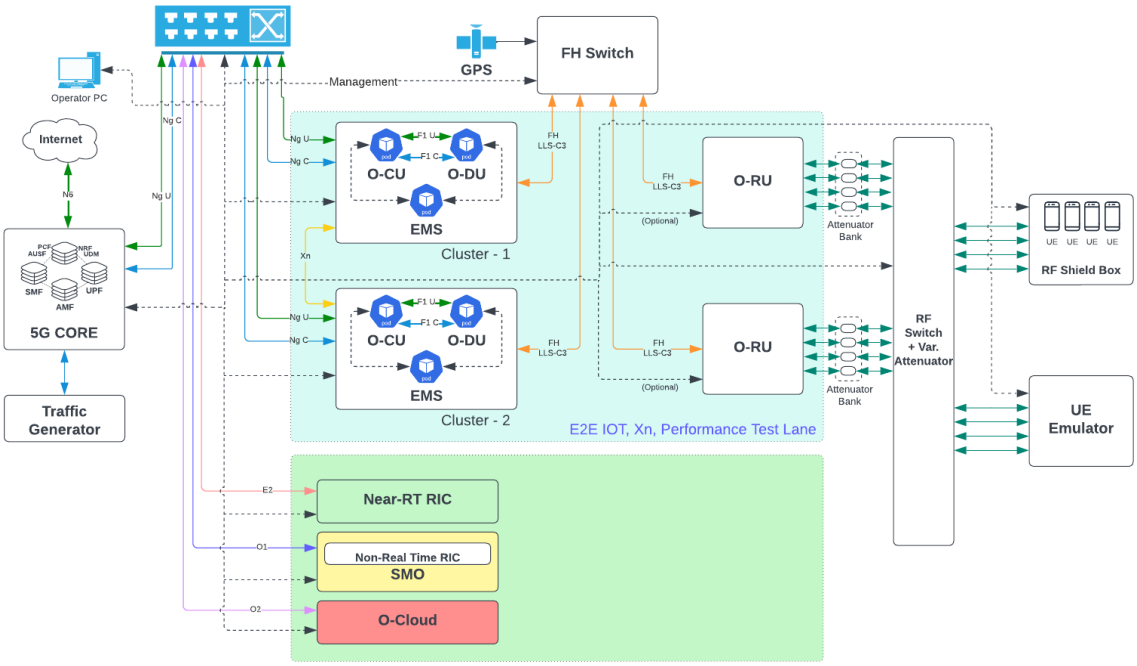


Figure 15. CableLabs/Kyrio lab setup for e2e integration

5.3. Observations

This section provides a summary of observations from Open RAN test activities at CableLabs to date.

Baseline Functionality - 100% of vendors successfully validated the 3GPP defined messaging for 5G system procedures across a multi-vendor e2e setup.
All the vendors successfully demonstrated the base functionality needed to establish an e2e setup that could allow the UEs to register/deregister, establish PDU sessions, perform service request and paging procedures, re-establish network connection post RLF, and perform state transitions. The system procedures were thoroughly validated for messaging and IEs exchanged between the multi-vendor

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

disaggregated RAN subsystems. Though certain challenges arose when the subsystems coming from different vendors were compliant to different specification releases or versions (within a release) and when vendors had different implementations considering the flexibility allowed within the specifications, the obstacles were overcome with collaboration and the adaptive mindset of the vendors.

Performance - 80% of vendors achieved peak throughputs (close to expected theoretical) and acceptable performance at cell edge across the multi-vendor e2e setup.
This was one of the key evaluation metrics to showcase if Open RAN multi-vendor disaggregated RAN could have a comparable performance when measured against the monolithic single vendor RAN. Leveraging the ORAN TIFG specification, the performance was evaluated with regards to throughput in both uplink and downlink for different traffic types (e.g., TCP, UPD, FTP) with UEs at different radio conditions. The performance was evaluated with stationary/mobility scenarios and with single/multiple UE(s). Key Performance Indicators such as RTT (for ICMP traffic), Jitter (for RTP traffic), and MOS score (for video traffic) were evaluated to complete a more thorough assessment of network performance.

Handover - 100% of vendors demonstrated mobility over Xn and N2 across the multi-vendor e2e setup.
The one aspect that has not been tested or vetted by the industry to date is the handovers across multi-vendor setups. The Xn is traditionally a closed interface with significant implementation differences across vendors and customized behavior tailored towards specific operator deployments. A successful demonstration of Xn and N2 connected mode mobility (handovers) and idle mode mobility (cell reselection) for both intra and inter-frequency scenarios across a multi-vendor setup was completed.

Stability - 70% of vendors sustained multiple UE connections across the multi-vendor e2e setup without dropping a call for long duration test runs.
Though the stability testing is done in a controlled lab environment, taking interference from neighboring networks out of the equation, simulating multiple UEs at different radio conditions, loading the network, and ensuring the stability over long durations, demonstrates the ability of the Open RAN multi-network setup to sustain in real deployments.

Reliability - 50% of vendors with the ability for multiple UEs in different RF conditions to run traffic across the multi-vendor e2e setup.
Though the reliability testing was not done on a cluster of nodes with thousands of UEs simultaneously connecting to the network and encountering interference from the neighboring deployments and environmental factors, as would be the case in real-world deployments, the reliability testing allowed us to run different traffic models and test the ability of the multi-vendor setup to sustain the network connectivity and performance offering some insights into how reliable the e2e setup is.

A limited number of vendors bring a software-only solution utilizing the bare metal servers
More than 90% of the vendors prefer to bring in their own hardware along with their software solutions. The limited integration time allowed in plugfests or the 5G Challenge is one of the prime reasons for vendors opting to ship their own hardware. This is also an indication of the software and hardware inter-dependency (software solutions may need a specific type of server, CPU version, etc.)

A majority of vendors required hardware accelerators
More than 50% of the vendors require hardware accelerators to boost the performance of the vendor's system to the expected level.

Average integration time for a vendor to integrate with test platform - 1 week (5 business days)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Assuming vendors ship hardware in a timely manner or have the software ready to install on the bare metal servers at the test lab, all the vendors were able to integrate with the test platforms within five business days. This also considers the readiness of the test lab with regards to the test setup, networking (IP address and VLAN assignments), power needs, and required cabling and SFPs in place.

Average integration time for multi-vendor integration - 1 week (5 business days)
With the same consideration mentioned for integration with a test platform, the multi-vendor integration took the same amount of time, indicating the integration process should not take longer if the best practices are followed with the required system integration expertise available at the test lab. The only additional factor in the case of multi-vendor integration is parameter configuration alignment to be able to successfully communicate over the spec-defined interfaces. While in case of integration with the test platform, the test platform can always be aligned to adapt to the vendor configuration, in case of multi-vendor integration a handshake is required to agree on specific parameters which can take longer in certain cases when the vendor solutions are not flexible to adapt such as having hard coded parameters that require software patch to align with the other sub-systems during integration.

5.4 Lessons Learned

As Open RAN continues to gain traction and maturity in CableLabs and similar labs worldwide, the following were general lessons learned:

- O-RAN Alliance test plans/cases and specifications are a solid foundation to support development efforts and demonstrate subsystem/system validation. Creating a common test plan, requirements, and features used by all industry labs would help operators adopt Open RAN faster.
- Plugfests and interoperability labs provide a valuable environment and means for subsystem/system development and interoperability.
- Achieving alignment of implementation/interpretation of 3GPP/O-RAN Alliance specifications between vendor subsystem and test emulators is a common challenge.
- As Open RAN continues to mature, developing certified trusted solutions for operators that include features and requirements for a common deployment profile will help accelerate Open RAN adoption.

6. Open RAN Transition

As MNOs continue to build out 5G networks there are many key elements of RAN architecture to consider as they evaluate deployment approaches, purchasing cycles, business strategy, customer needs, and deployment scenarios. As Open RAN continues to mature, the path of MNO RAN buildout and growth strategy may overlap with the developments in Open RAN architecture. This section proposes a transition approach from vRAN to a fully integrated Open RAN architecture.

6.1. Transition from vRAN to Open RAN

The transition from vRAN to Open RAN can be characterized by evaluation indicators, several of which are summarized below.

1. *Virtualization of Baseband Processing:* In vRAN, baseband processing functions are virtualized and run on standard servers or cloud infrastructure. However, in Open RAN, the focus shifts to disaggregating the RAN components, allowing for multi-vendor interoperability and greater flexibility.

2. *Interoperability and Standardization*: Open RAN emphasizes interoperability between different vendors' hardware and software components. This involves adopting open standards and

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

interfaces such as O-RAN Alliance specifications. The transition from vRAN to Open RAN involves the adoption and implementation of these standardized interfaces.

3. *Multi-vendor RAN Deployment*: One of the key characteristics of Open RAN is the ability to mix and match equipment from different vendors. This allows operators to select best-of-breed solutions and avoid vendor lock-in. The transition from vRAN to Open RAN involves enabling multi-vendor deployments and ensuring compatibility between various vendors' equipment.

4. *Cloud-Native Architecture:* While vRAN leverages virtualization, Open RAN takes it a step further by adopting cloud-native architecture principles. Open RAN solutions are designed to be highly scalable, flexible, and capable of running on cloud platforms. Transitioning from vRAN to Open RAN involves embracing cloud-native architectures and containerization techniques.

5. *Open Interfaces and APIs*: Open RAN relies on open interfaces and APIs (Application Programming Interfaces) to enable interoperability and facilitate the integration of different RAN components. These interfaces and APIs allow for easier customization, innovation, and the introduction of new functionalities. The transition from vRAN to Open RAN involves implementing and utilizing open interfaces and APIs.

The transition from vRAN to Open RAN is an ongoing process, and the significance of these indicators may vary depending on the specific context and timeline of an MNO's implementation.

6.2 Open RAN Transition Approach

A transitional approach to Open RAN architecture is illustrated in the diagram in Figure 16. It is based on the five key elements of Open RAN architecture, namely, Virtualization (V), Disaggregation (D), Open Interface (O), Cloudification (C) and Intelligence (I). These elements are depicted as a 5-piece 'element dial' that is used as a reference during transition and are arranged in a matrix-like structure, with hardware depicted along the horizontal path and software along the vertical path.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

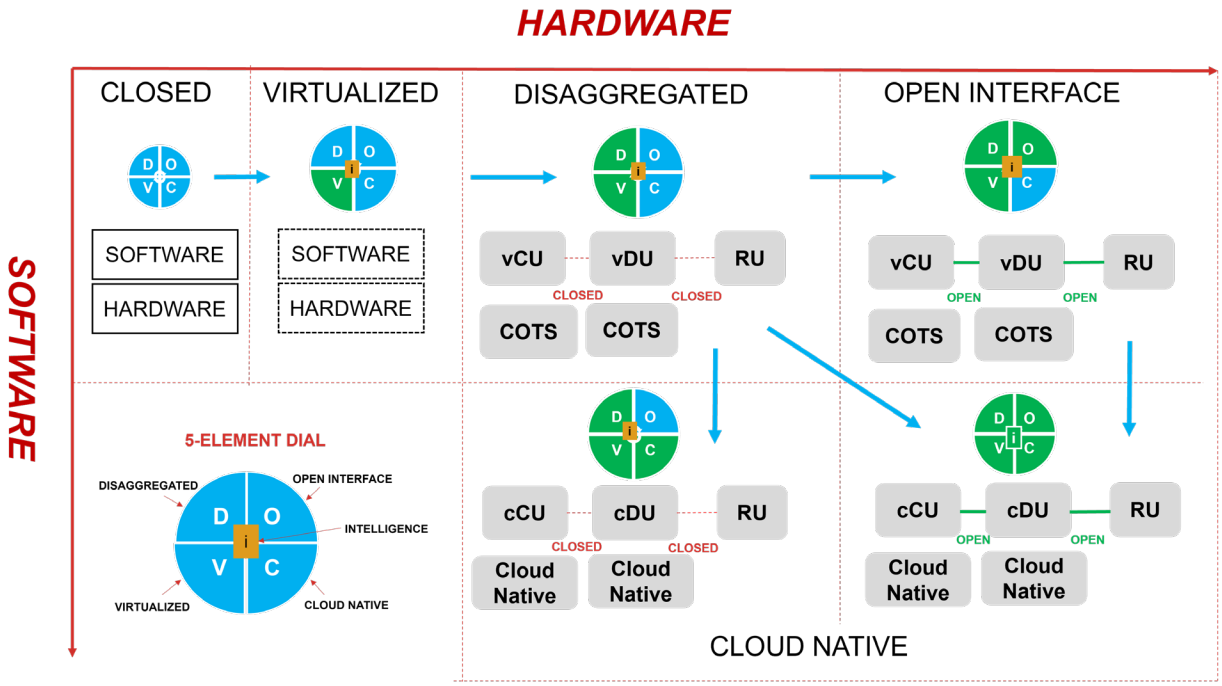


Figure 16. Open RAN transition approach

The transition starts with a closed, proprietary RAN with no Open RAN elements and leads to an outcome with all five elements. There are three main paths leading to the 5-element outcome with each path adding an element to the Open RAN architecture as it is built. As an element is added to the architecture, its color changes to green on the 'element dial'. The paths can follow a hardware or software progression. The intelligence element (i.e., automation, orchestration) is included at the center of the 'element dial' and remains the same color until the 5-element outcome is realized since its implementation is independent of the other elements which are more architecture components of Open RAN. While the intelligence element is a key part of each transitional architecture, its composition can vary as open or closed until the final 5-element outcome where it must be open across multi-vendor interfaces to realize the vision of Open RAN.

As a closed RAN progresses towards a fully integrated Open RAN architecture, the path can add each of the following elements:

Virtualization:

- This transition path involves the shift to a virtualized environment.
- Initially, the focus is on virtualizing software functions, enabling them to run on general-purpose hardware. This step prepares the network for further openness and flexibility.

Disaggregation:

- This path emphasizes the separation of hardware and software components.
- COTS hardware is introduced, allowing software functions to run on more flexible and open platforms. Disaggregation paves the way for future scalability and customization.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Open Interface:

- Open interfaces facilitate interoperability between different vendor components, enabling seamless integration. This step enhances the network's flexibility and promotes innovation.

Cloud Native:

- This path involves transitioning to a cloud-native architecture.
- Disaggregated components are combined with cloud-native principles, making use of containerization and microservices. This step optimizes resource utilization and scalability, setting the stage for more advanced capabilities.

Three distinct transition paths are proposed below:

1. **Hardware First:** The first transition path, as shown in Figure 17, starts with a virtualized RAN and follows a direction that adds disaggregation, then open interfaces and finally cloudification to the RAN architecture. This is called 'Hardware First' since it follows the path of building virtualized, disaggregated RAN components on the same bare metal servers as it transitions to disaggregation and open interfaces before changing hardware to a cloud native platform.

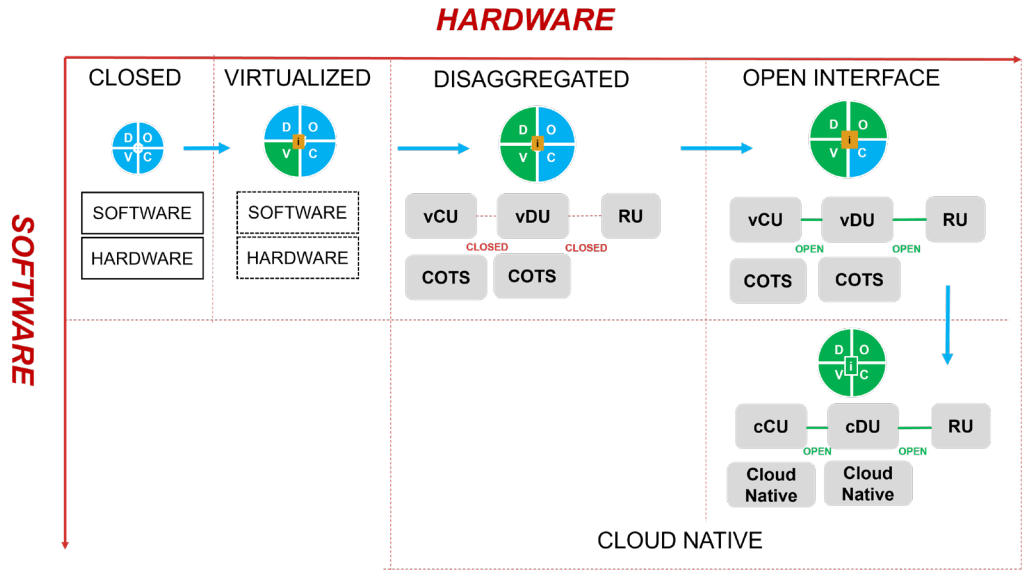


Figure 17. Hardware First transition path to Open RAN

2. **Software First:** The second transition path, shown in Figure 18, starts with a virtualized RAN and follows a direction that adds disaggregation, then cloud native, and then open interfaces to the RAN architecture. This is called 'Software First' since it follows the path of building virtualized, disaggregated components on bare metal servers and transitions to a cloud native platform and then transitions to an open interface.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

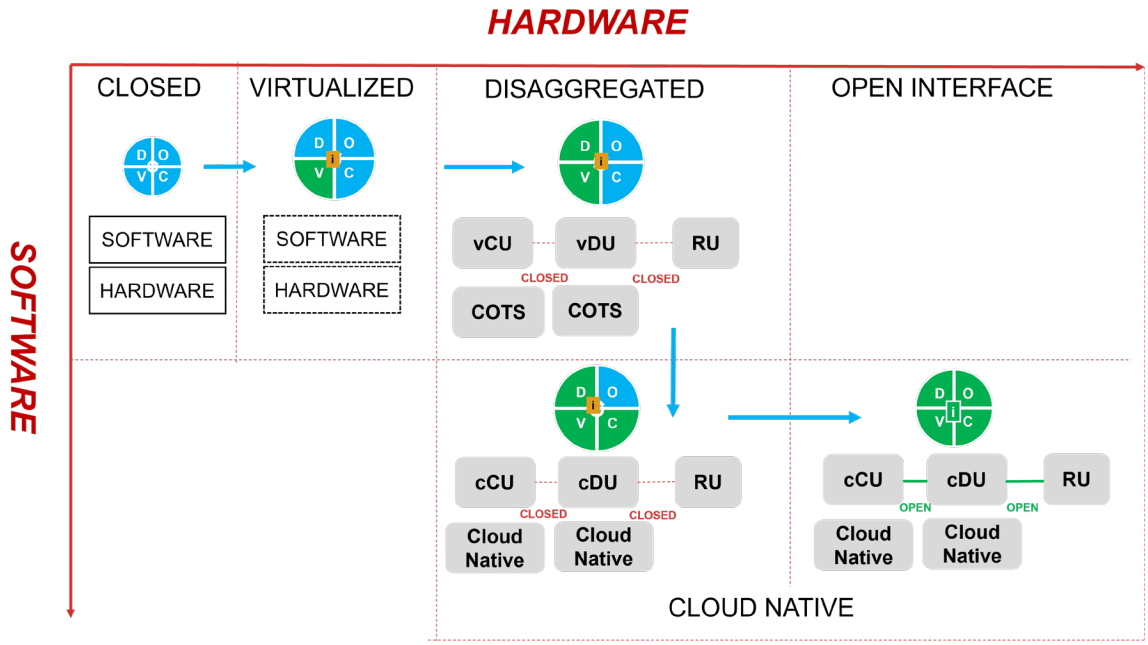


Figure 18. Software First transition path to Open RAN

3. **Direct Path:** The third transition path, shown in Figure 19, starts with a virtualized RAN and follows a direction that adds disaggregation on bare metal servers, then transitions directly to open interfaces and cloud native RAN architecture. This is called 'Direct Path' since it follows the path of building virtualized, disaggregated components on bare metal servers and then transitions directly to an open interface, cloud native platform.

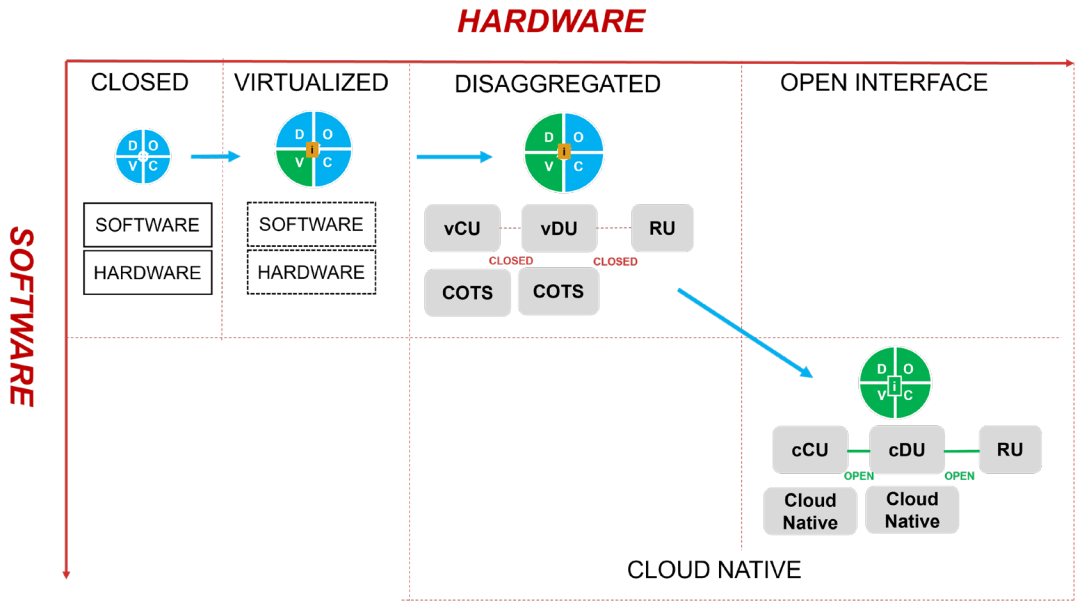


Figure 19. Direct transition path to Open RAN

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Each transition path offers a strategic sequence of steps to guide the migration from a closed vRAN deployment towards a more open and adaptable Open RAN network architecture. The MNO RAN transition journey is at its own pace and selected path based on business goals, network needs, and network deployment strategy. The overall transition approach is intended to provide paths that make sense to an MNO and offer various levels of integration that can be temporary or permanent.

7. Conclusion

The Open RAN ecosystem continues to evolve towards the vision of RAN components as a virtual software application capable of working on any bare metal hardware operated with orchestration and automation. A fully featured Open RAN architecture complements 3GPP 5G RAN specifications and incorporates five key elements into RAN architecture, namely, disaggregation, virtualization, open interface, intelligence, and cloudification. Key benefits include a shift away from monolithic infrastructure and a move towards multi-vendor, plug-and-play interoperable components that offer flexibility, vendor diversity, and innovative freedom.

The Open RAN ecosystem is vibrant with a broad range of industry activity that includes commercial mobile network operator deployments, industry testbeds and trials, and government policies and investment. These activities demonstrate proof-of-concept and end-to-end validation of Open RAN subsystems and systems. Several mobile network operators across the globe have deployed greenfield and brownfield commercial Open RAN networks while others have made public announcements of field trials, partnerships, and commitments to Open RAN deployments.

Governments such as the United Kingdom, United States, Australia, Germany, Canada, Japan, South Korea, and Singapore are investing in and promoting the development, adoption, and deployment of Open RAN.

CableLabs/Kyrio is a recognized industry leader in the Open RAN ecosystem as a system Integrator, independent and neutral host of a fully featured Open RAN test lab, Open RAN subsystem and system test experts, and contributors to SDOs. Test results, observations and lessons learned from these Open RAN test activities demonstrate progress towards Open RAN adoption proving multi-vendor e2e interoperability with multiple vendors.

A proposed Open RAN transition approach offers three transition paths to achieve the outcome of fully integrated Open RAN architecture, namely, Hardware First, Software First, and Direct Path. Each transition path offers a strategic sequence of steps to guide the migration from a closed vRAN deployment towards a more open and adaptable Open RAN network architecture. The choices among these paths depend on the network operator's goals, existing infrastructure, and desired outcomes. The overall transition approach is intended to provide paths that make sense to an MNO and offer various levels of integration that can be temporary or permanent.

Future study includes methods to develop operator deployments of multi-vendor e2e interoperability solutions based on common deployment profiles, common industry test plans, and further detailed total cost of ownership analysis. The future of wireless networks starts with Open RAN. The potential for secure and reliable networks built on multi-vendor components that are easy to upgrade to future generations of network technologies is recognized by many governments, mobile network operators, vendors, and labs. Working together, the ecosystem can bring this potential to fruition in the near future.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

3GPP	3rd Generation Partnership Project
RAN	Radio Access Network
CU	Centralized Unit
DU	Distributed Unit
RU	Radio Unit
eMBB	Enhanced Mobile Broadband
gNB	gNodeB
URLLC	Ultra-reliable Low Latency Communications
mMTC	Massive Machine Type Communications
CPRI	Common Public Radio Interface
SCTE	Society of Cable Telecommunications Engineers
vRAN	Virtualized RAN
SDO	Standards Development Organizations
TIP	Telecom Infra Project
IMDA	Infocomm Media Development Authority
5G	Fifth-generation technology standard for broadband cellular networks
US	United States
UK	United Kingdom
NFV	Network Function Virtualization
MANO	Management and Orchestration
COTS	commercial-off-the-shelf
SMO	Service Management and Orchestration
PNF	Physical Network Functions
MOCN	Multi-Operator Core Networks
MORAN	Multi-Operator RAN
RIC	RAN Intelligent Controller
e2e	end-to-end
DCMS	Department of Culture Media Sport
NTIA/ITS	National Telecommunications and Information Administration's Institute for Telecommunication Sciences
CBRS	Citizens Broadband Radio Service
SAS	Spectrum Access System
UE	User Equipment
RF	Radio Frequency
WG	Work Group
TIFG	Test and Integration Focus Work Group

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

NTIA Technical Memorandum 23-568: 5G Challenge Preliminary Event: Evaluating Modular, Interoperable, Multi-Vendor, Open RAN Solutions.

NTIA 2023 5G Challenge: Host Lab Configuration.

NTIA 2023 5G Challenge: RU Test Plan

NTIA 2023 5G Challenge: CU+DU Test Plan

NTIA 2023 5G Challenge: E2E Test Plan

O-RAN ALLIANCE: Global PlugFest – Fall 2022

O-RAN ALLIANCE: Global PlugFest – Spring 2023

O-RAN ALLIANCE: Overview of Open Testing and Integration Centre (OTIC) and O-RAN Certification and Badging Program

O-RAN Architecture Description 9.0, O-RAN.WG1.OAD-R003-v09.00

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23 CELEBRATING 40 YEARS

2023 Fall Technical Forum SCTE • CABLELABS • NCTA

Operational Sustainability: On Achieving Optimal Leverage of the Power Grid with Wi-Fi CPE

A Technical Paper prepared for SCTE by

James R. Flesch
Director Adv. Tech, office of CTO/Home
Commscope
3871 Lakefield Drive, Suwanee, GA 30024
jr.flesch@commscope.com

Brian Carroll
Sr. Manager, HW Eng'g/R&D
Commscope
101 Tournament Drive, Horsham, PA 19044
215 323-1427
brian.carroll@commscope.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Scaling the Wi-Fi AP Hardware for Fractional Operations 6
   2.1. Installing “Always ON” Test and Measurement 7
   2.2. Proportionate Network Processor Operations 9
   2.3. Partial Band Coverage 9
   2.4. Pruning Spatial Streams 9
   2.5. The Resulting Block Diagram 10
   2.6. Controllable Impacts to Power Budget 10
   2.7. Considerations on the Standby -> Active Sweet Spot 15
3. Managing the Power Consumed: Software Behaviors 16
   3.1. Band-Specific Characteristics 17
   3.2. Wi-Fi Roll of Clients and Services 18
   3.3. Band Capacity Trigger Points 20
   3.4. But What About 6 GHz? 22
   3.5. Proposed Flow for Implementation of AP Power Management 25
4. Conclusion 29

Abbreviations 30

Bibliography & References 31

List of Figures

Title **Page Number**

Figure 1 – WAN Bitrate Cyclic Variation Over One Week 5
Figure 2 – Brute Force AC Management 5
Figure 3 – Distribution of Maximum Power Consumed by TBC AP for Various Functional Blocks (~ 35 W total) 7
Figure 4 – Deep Hibernation of AP “Live” Sections With IoT-only Monitoring 8
Figure 5 - AP Full Block Diagram with Granular Power Controls 10
Figure 6 – Dissipation Assumptions for Major AP Operational Modes 11
Figure 7 – 20W (>40% Reduction off of Max) Active Mode Power Distribution 12
Figure 8 – Standby Power Profile, 6W 13
Figure 9 – Deep Sleep Power Profile (Memory Off), 1.6W 14
Figure 10 – Device “Off”; Power << ¼ Watt 15
Figure 11 – Power Cost Matrix per AP Functional Circuit 16
Figure 12 – Mix of Bands, Services and Clients in Wi-Fi Space 19
Figure 13 – “Realistic” Channel BW (MHz) and Bitrate Capacities (Mbps) for 2SS Clients from a 4SS AP 21
Figure 14 – 24-hr Partitioning of AP Operations (hours) 21
Figure 15 – Boilerplate Bitrate vs Power for AP Operational Permutations over 24 hrs 22
Figure 16 – Relative “Realistic” Band Efficiencies, Bitrate/W 23
Figure 17 – Tipping Point Calculation for Xfer of Service from Band X to 6 GHz 24
Figure 18 – AP Block Diagram Indicating Green Operations 26
Figure 19 – Polling Logic to Turn AP ON/OFF 27

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 20 – Moderation of AP Profile to Fit Services Mounted 28

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Sustainability has established itself as a critical aspect of customer premise equipment (CPE) design goals, worthy of equal consideration to elements such as packaging aesthetics, thermal behavior, economic value proposition and overall electronic performance. While for some time this has drawn attention to items such as single-use plastic exposure, recyclable or otherwise non-hazardous material exploit in hardware implementation and some fair amount of device and accessory packaging cleverness, the interest now has expanded to include operational power signature which conforms to device service usage demands. Adapting power grid draw to match measured (or anticipated) application performance on a "just necessary" basis (and so eliminate wasteful standby consumption which – on a scaled basis – irresponsibly overtaxes the energy grid) is now the stuff of practical necessity. This type of considered and throttled behavior on the part of CPE immediately substitutes perceptions of thoughtless energy banditry with environmentally aware, green energy iconography -- which clever SPs can leverage in the marketing realm to amplify their brand's public association with responsible (and astute) corporate behavior. It also aligns electronic CPE with application-fitting power consumption as exhibited by other in-home electrical appliances – substituting managed operations for the manual ON/OFF cycling associated with their draws on the power grid.

There is a power consumption history associated with this class of CPE (Wi-Fi access points (APs) in particular) which must be acknowledged – an undeniable progression of unmanaged power sink which classically has not regulated itself during periods of constant feature aggregation and employment. As Wi-Fi medium access control layers (MACs) and bands have evolved and the number of spatial streams increased, so too has the power requirement to operate these devices. What used to cost approximately 5W for single band APs a dozen years ago has now risen to over 30 watts (W) for a tri-band concurrent (TBC) device with multiple internet of things (IoT) radios.

For this paper, reformulation of the power footprint for always ON Wi-Fi APs (be they wide area network (WAN)-attached or other) is examined to investigate approaches into making this class of device a better power dissipation partner in the home. We will examine tactics which lever hardware, firmware and software – the latter both device-native and cloud-hosted – with an eye to fitness for purpose (i.e., meeting the bitrate and latency requirements of services mounted at any particular time) and efficacy of power-saving results. The general form of solution – to appropriately hibernate various resources when service demand does not mandate their exploitation – will be a guiding rubric throughout this dissertation. Note that a collateral goal will be seamless transition from various hibernation states to appropriate service support levels which will eventually manifest as "slow to hibernate, quick to wake" behavior; the goal being such does not express itself in human-observable debits in local area network (LAN) performance, nor requires human intervention to achieve.

To enhance the end value proposition, the device will be outfitted with self-monitoring of the power consumed during operations and such will be made available to a cloud-based observer for both reporting purposes and as an aid in formulating pre-emptive behaviors based on prior history and time-of-day (ToD) considerations – the latter a key telltale input used to determine responsivity requirements of transitional device operations (to/from hibernation). Motivation for the implications of this operating litmus can be found in the following figure showing diurnal/nocturnal cable network operating bitrate demands (from March 2020) as can be inferred from ToD service group (SG) waxing and waning over the course of multiple 24-hour periods (and further grasping that such represents ensemble LAN back/front service hauls over that same period):

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

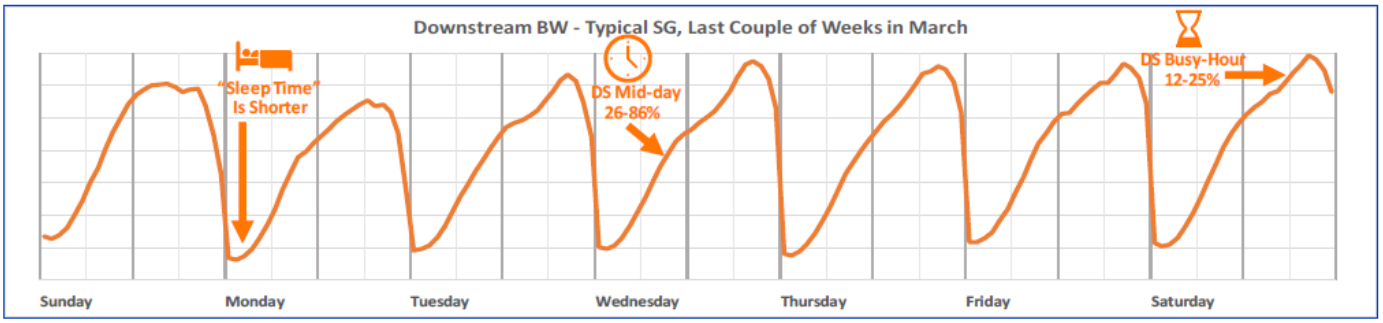
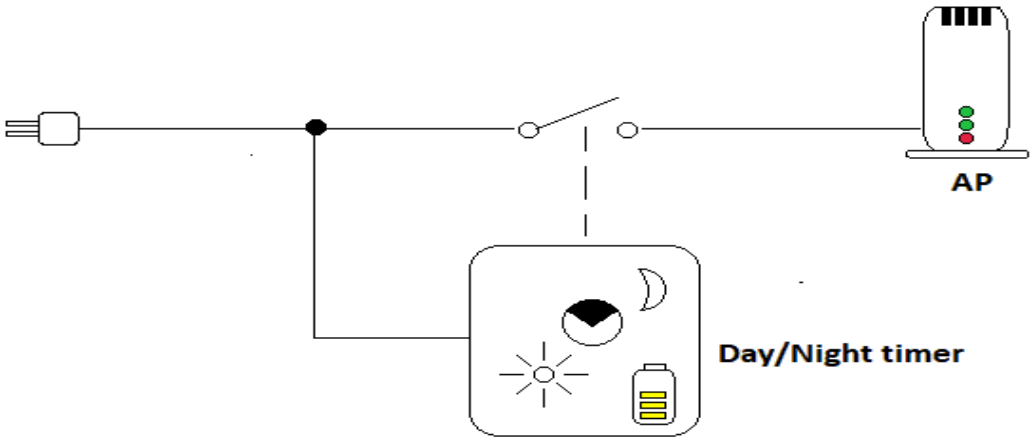


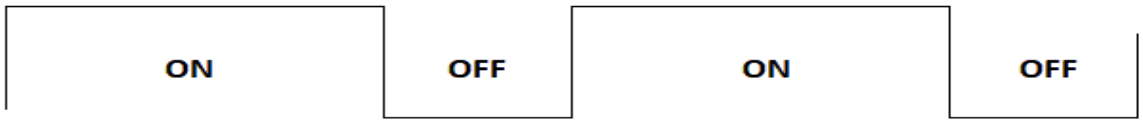
Figure 1 – WAN Bitrate Cyclic Variation Over One Week

Clearly, there is a ToD-based cyclic demand for wireless bitrate services which can form the master basis for modulation of power consumption on the part of a Wi-Fi AP.

These diurnal cycles of peak usage beg a simple brute-force solution – and it is fair game to point out the opportunity. As with all appliances, a simple ON/OFF strategy would serve to minimize power draw of the AP and, in cases where no keep-alive 24-hour services have it for a dependency, end up guiding us to the asymptote of power savings of which we should be mindful in targeting savings. For example, the expedience of 16 hours ON and 8 hours OFF buys us ~ 33% worth of power savings for the device and may be implemented by leverage of an explicit switch or a ToD-gating function:



Diurnal-based AC duty cycling



20-35% bill reduction with brute force AC management

Figure 2 – Brute Force AC Management

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Assuming, however, that AP availability for network attachment is associated with some support for 24-hour applications, we will proceed to break out design approaches which move across an operational spectrum from some minimally effective hibernation state to maximal performance. The European Code of Conduct (eCOC), for example, targets a bit more aggressive results in power use mitigation than that above – from mid-30W type of maximum dissipation to 8W or less in standby – so additional performance mining along this front should be conducted.

2. Scaling the Wi-Fi AP Hardware for Fractional Operations

In Wi-Fi APs, legacy design approaches have always proceeded on the basic calculation of how best to achieve maximum Wi-Fi signaling coverage for a given material cost point. Power supply specification, then, builds on presumptions of high duty cycle transmit leverage of all radio frequency (RF) spatial streams (as concurrent use), in concert with simultaneous support of all LAN and IoT interfaces. Central processing unit (CPU) core functionality is presumed to full extent of the available processing Dhrystone million instructions per second (DMIPs), and all corollary memory support is biased and clocked at maximal rates as well. Typically, this also invokes full use of active cooling (where such is provided) and so fan speed is also ratcheted to its upper limit. Against these artificially coincident maxima, packaging is then reduced to a three-dimensional (3D) envelope which balances best antenna farm orthogonality (as in- and cross-band isolations, typically the result of best physical separation of all coincidentally polarized elements) along with a maximally tolerable external package touchpoint temperature (or margin to silicon core temperatures, whichever is the more restrictive). Service provider (SP) architects routinely request small-as-possible industrial designs (IDs), so a balancing act occurs between a size minimization push and temperature/RF orthogonality pulls. There is a sort of comfortable challenge here, if it amounts to an incomplete accommodation for what we now wish to additionally accomplish: scaling the electrical draw of these maximal functions to dissipate appropriately fractional power loads when operational demands (service mounts) recede from these outer edge-of-performance considerations – thereby minimizing standby power use.

In this countervailing discipline, determination of required functional support from the AP must be pulled into a matrix which ends up defining which circuits/capability must be ON – and if frangible, to what order of inclusion -- for particular service profiles to be enabled. Bulk switchable aspects might include functional modules such as IoT radio support, battery charging circuits, Wi-Fi band(s), LAN ports, etc. Tunable aspects would include items such as number of RF spatial streams to enable per band and amount of CPU power (as clock speed based or other segmentation) and perhaps the amount of memory to apply to reduced operations. (This latter nut is a bit tougher to crack; responsivity in fractions of a second to some few seconds up to full service demand does not permit time for reconfiguration of the AP so unused software (SW) application containers in still-active random access memory (RAM) would remain an unavoidable – power consuming -- constant. (As circumstances have it, such memory hibernation consideration does not yield significant power savings in any event and so the complications associated with its consideration are rendered moot – refer to the next section.)

Establishing a budgetary feel for various circuits' impact to power dissipation is a necessary step. The following figure provides a weighting comparison of different functional elements to total power for a representative tri-band high end AP with 12 spatial streams (SS) (maximum dissipation around 35 Watts):

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

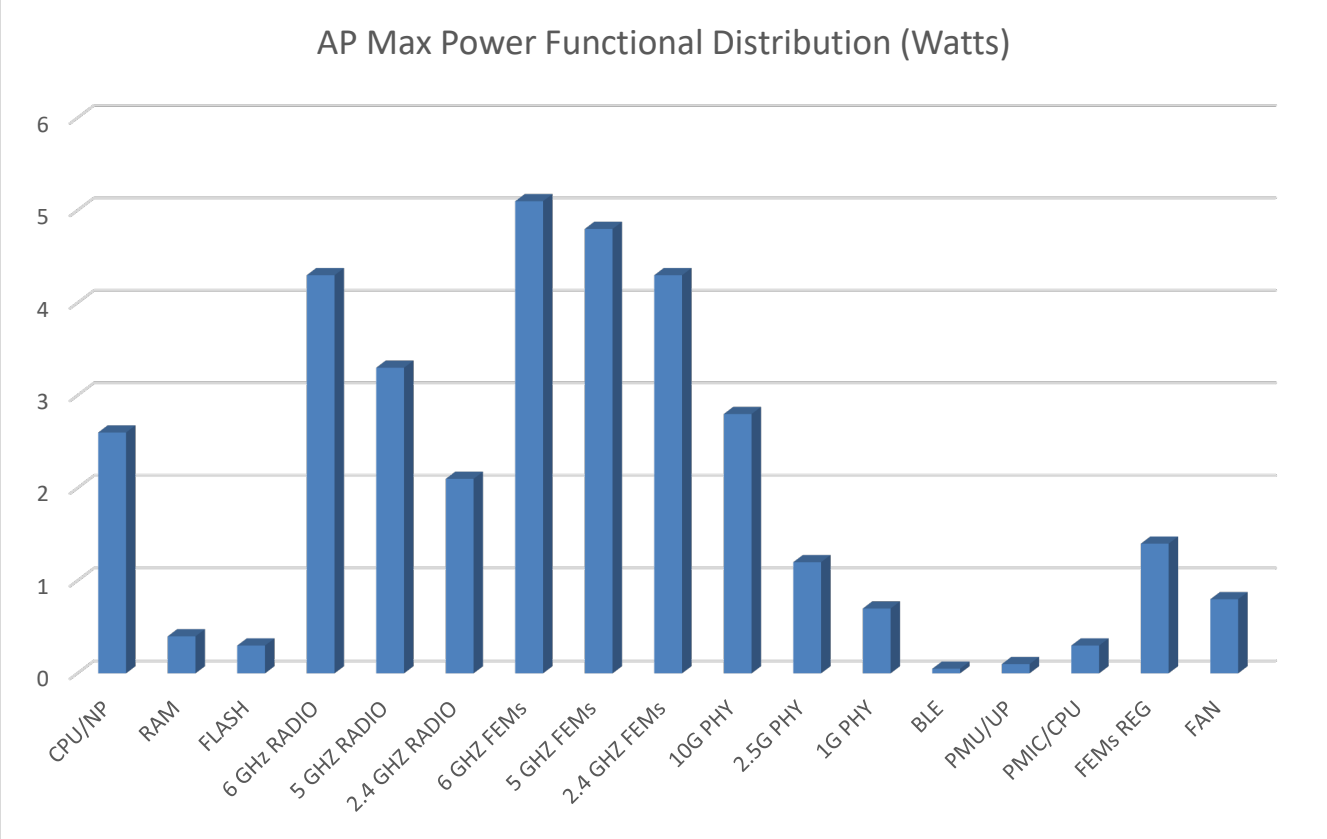


Figure 3 – Distribution of Maximum Power Consumed by TBC AP for Various Functional Blocks (~ 35 W total)

As can be seen, there are some broad power-management opportunities along the radio band, SS deployment, Ethernet ports and network processor fronts, with the majority benefit accruing to Wi-Fi radio management. These opportunities will be examined in more detail following.

2.1. Installing "Always ON" Test and Measurement

Recall the comment during introduction that self-reporting energy consumption will be a key performance aspect of the changes wrought to the AP. This metric is easily distilled, however, by simply monitoring voltage and current at the AP's power entry point (an external alternating current/direct current (AC/DC) brick form factor conversion being a common enough implementation). But the question of where to route the measured analog-to-digital converter (ADC) values and subsequent calculation (along with the hosting environment for this circuitry) needs to be answered as well; for the time being, we can assume a tagged sequence of calculated power values generated over some repetitive sampling interval is stored away for eventual routing to a cloud entity for analysis. How this measurement and storage is effected depends on the nature of the standby operations – whether these include IoT and wireline operations, for example, and if so, which physical layers (PHYs) are required to be supported.

Unfortunately, critical 24-monitoring typically accompanies medical and security services which straddle at least two different PHYs: Z-wave or Zigbee for security and Bluetooth Low Energy (BLE) for medical/wearable devices. (It is possible that in the very near future, Matter's Thread meshes will be able to see to both of these client communities, but available legacy devices currently occupy the different radio camps described above). If the IoT hub support is separately hosted, a very standard arrangement

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

sees a single- or dual-band Wi-Fi backhaul or wireline Ethernet used to provide connectivity to the WAN. (Note that this is not guaranteed to be the case; arguably, WAN connectivity for IoT services might leverage direct-to-outside personal area network (PAN) – long range wide area (LoRa), for example -- or long term evolution/fifth generation (LTE/5G) committed cellular backhauls but for purposes of evaluating AP hibernation energy costs we will tread the path of presuming it proxies the ultimate WAN connection for IoT as well as 802.11-based radio services.) Given at least vestigial monitoring tasks during nocturnal operation (such being an appropriate time to idle – to the greatest extent possible – AP network processing and routing functions), the substitute of a cheap, very low power custodial processor to oversee AP hibernation and measure power use seems appropriate. The following figure details a possible approach to the very deepest level of hibernation, where nothing aside from IoT monitoring is presumed:

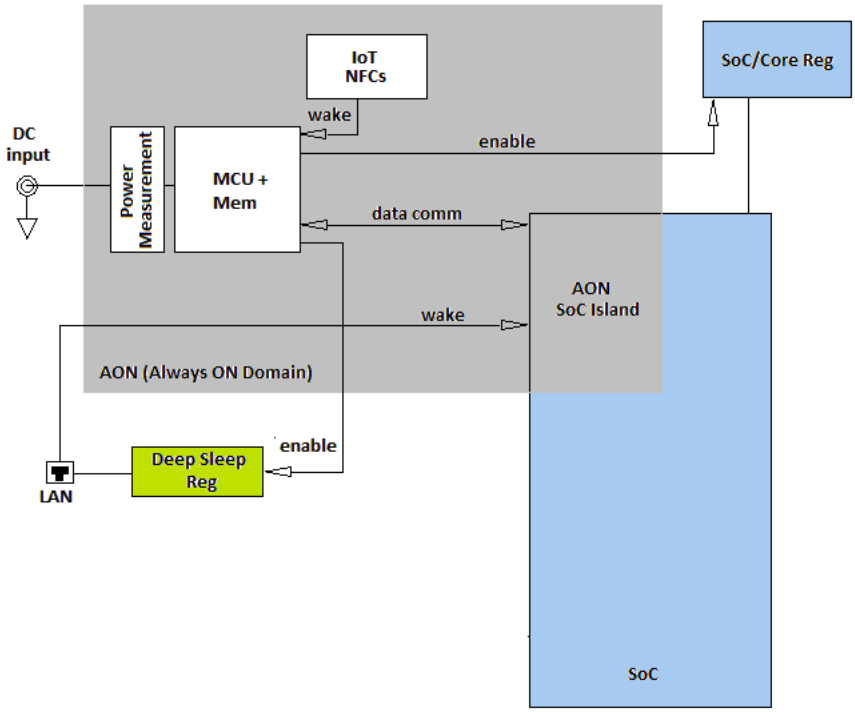


Figure 4 – Deep Hibernation of AP "Live" Sections With IoT-only Monitoring

In practice, it may also be a very good idea to permit some level of Wi-Fi awake triggering to occur, so the figure above would be modified to include minimal detection of at least the 2.4 gigahertz (GHz) Wi-Fi band in addition to the IoT and LAN emergency detection scheme. Note that from a WAN connectivity standpoint, the AP is effectively OFF at this point and so the timestamped power measurements would likely become bookended at OFF cycle start (power down) and OFF cycle end (just prior to power up). The decision on memory powering comes down to power up responsiveness. At an additional cost of ~ 0.7 W (0.4 W if only RAM is kept alive), the APs ability to light up with full containerized application space (versus a period of re-installing the software container environment) may be a worthwhile idle power investment.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.2. Proportionate Network Processor Operations

During full or partial hibernation, the expectation is that executive management of the AP would see reduced activity from one or more of its multiple IP or other service flows and so the Network Processor/CPU could throttle back operations accordingly, maintaining sufficient awareness to react when service demands re-mount. This type of throttling typically involves clock speed reduction at a minimum (power dissipation being proportionate to clock speed) – but may also involve retirement of certain cellular functions as well, depending upon SoC power control granularity. The scalability is rooted in internet protocol (IP) packet traffic tending and so is proportionate to mounted service density; the goal is to invoke just enough DMIPS to see to the packet segmentation and re-assembly (SAR) and routing requirements (among other keep alive operations).

2.3. Partial Band Coverage

The halo product in AP space is considered a tri- or quad-band device (2.4 GHz / 5 GHz Lo / 5 GHz Hi / 6 GHz) with a MAC epoch of 6/6E (Wi-Fi 7 coming early next year). With currently available client devices spanning all these bands (rarely if ever all supported concurrently per client device, however), operation of the AP during peak hours could mean pressing all radio bands into service for a household with multiple devices. But clever "service aware" scheduling can counter this broad swath approach by examining the actual band service and interference profiles in a given household along with the quality of service (QoS) expectations of the various services -- herding capable clients onto as few radio bands as would meet the ensemble QoS proposition. (This herding is sometimes referred to as "band coalescing".) The operational behavior will be outlined in SW sections following, but from a hardware (HW) standpoint, the ability to isolate power dissipation on a per-band basis is key. This involves idling PHY and MAC operations on a per-band basis for those bands without active clients and assumes that device advertisement is best (first) serviced on the 2.4 GHz band – the expectation being that this band is supported on all clients regardless of Wi-Fi MAC epoch. For those rare cases where such client band support does not exist at 2.4 GHz, a minimal monitoring cadence can be maintained on 5 GHz as well. In "perfect hibernation", all client devices would be serviced at only 2.4 GHz and AP transmit opportunities (TXOPs) would be as infrequently subscribed as the minimal service keep-alive requirements permitted – and this, at a single active spatial stream and at a narrow bandwidth (BW) which still meets QoS demands for the near-dormant services.

Deviations from this asymptote would only occur for escalations of connectivity (alarm propagation of one type or another or the onset of service-heavy diurnal periods). Referring to the Figure above breaking out dissipation category, note both the granularity and impact to device power use based upon hibernating select band utilizations; avoiding 6 GHz, for example, in this standard power-capable TBC AP prunes roughly ¼ of the *total* power use from the budget by itself (if the band is fully de-activated). In the forthcoming couple of years, the adoption rate of 6 GHz-capable clients (coupled with the onset of post-Wi-Fi 5 MACs) will still lag significantly relative to the legacy band (and MACs) leverage, so the sacrifice as first resort seems advisable; the developing premises service profiles will indicate when (and how pre-emptively) 6 GHz use should occur.

2.4. Pruning Spatial Streams

Even in those cases where a particular band is drafted for use, circumspect employment of less than a full complement of spatial streams may still provide the necessary bitrate support while hibernating transmitters; however, this becomes a complex problem to undo – bitrate is inversely related to airtime, which is the variable (x SS) which defines the power profile for a given RF band. Remember too that bitrate is directly related to channel BW – increasing this to the maximum possible results in the shortest

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

airtime transmission for a given packet length but consumes the most transmit power. For now, we defer the algorithmic impacts to the SW section.

In cases where internal frontend module (iFEM) (frontend modules (FEMs) internal to the radio SoC) technology is available, such simplifies the selection of spatial width to transmit over; however, it must be noted that these cases do not typically support power levels associated with fixed AP devices and are more intended for mobile, battery-powered devices. One further aspect is noteworthy: recent developments in FEM technology have seen improvements in idle (receive mode) operation, so full disabling of the external FEMs does not yield power savings during receive operations as it would have only a year or so ago (i.e., receive mode operations have become much more power efficient); nonetheless, given the inexpensive nature of pass device disabling of the bias for FEM blocks, it still behooves us to consider them for complete granular control of RF chain power usage (if the radio chip does not natively support individual idling of RF chains itself).

2.5. The Resulting Block Diagram

As a result of the considerations outlined in prior sections, we end up with the following AP ensemble block diagram which includes the granular controls required to selectively enable or disable the various contributing elements to total power dissipation:

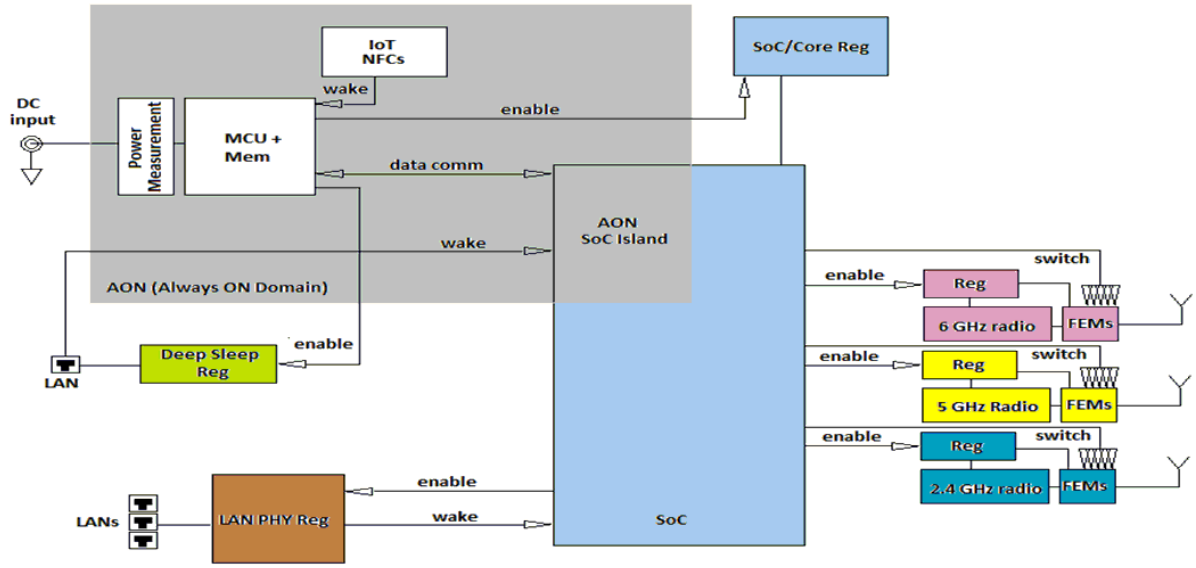


Figure 5 - AP Full Block Diagram with Granular Power Controls

2.6. Controllable Impacts to Power Budget

Selective enablement of the various control aspects detailed above result in a cascade of power operating points for the AP. The impact of major blocks is summarized below, along with the presumptions made for partial leverage of each of the subgroups:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Power Management Configurations						
	Off Mode	Deep Sleep	Network Standby	Active	On Full	Conditions
SoC	N	10%	50%	75%	100%	CPU Load
DDR4	N	N	Y	Y	Y	
eMMC	N	N	Y	Y	Y	
6G Wi-FI IC	N	N	25%	50%	100%	Duty Cycle
5G Wi-FI IC	N	N	25%	50%	100%	Duty Cycle
2.4G WiFI IC	N	N	25%	50%	100%	Duty Cycle
6GHz FEM	N	N	1%	50%	85%	Duty Cycle
5GHz FEM	N	N	1%	50%	85%	Duty Cycle
2.4G FEM	N	N	1%	50%	85%	Duty Cycle
10G WAN PHY	N	100	100	1000	10G	Port activity
2.5G LAN PHY	N	100	100	1000	2.5G	Port activity
1G LAN PHY	0	0	0	1	2	Ports Active
2.4G BLE IoT	Y	Y	Y	Y	Y	
PMU	Y	Y	Y	Y	Y	
PMIC	N	N	Y	Y	Y	
Losses	N	N	Y	Y	Y	
Fan	N	N	25%	50%	100%	
Power (W)	0.15	1.6	6.2	20	34.5	

Figure 6 – Dissipation Assumptions for Major AP Operational Modes

The categories shown include On Full, Active, Network Standby, Deep Sleep and Off Mode. On Full is a typical specification description which presumes worst-case concurrent operations and is used to investigate full circuit cooling and maximal Wi-Fi simultaneous use at high power levels. Ethernet componentry is also full exercised, as are all onboard regulators and the IoT subsystem. It is only a design bogie (and never encountered in normal operations), but nonetheless establishes the outer power envelope in which an unmanaged AP will inevitably waste power under partial loading.

Active Mode is a bit more circumspect in traffic handling – though it still tends towards the "over-exercised" portion of the operational spectrum. Wi-Fi use is still quite heavy and concurrent across all 3 bands, with roughly 2 gigabits per second (Gbps) packet traffic split across the two faster Ethernet ports. In "real" operational considerations, this represents a heavily loaded device (as might be the case during

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

evening peak use cycles after 5 pm and before 11 pm or so). The power fingerprint by section looks as follows:

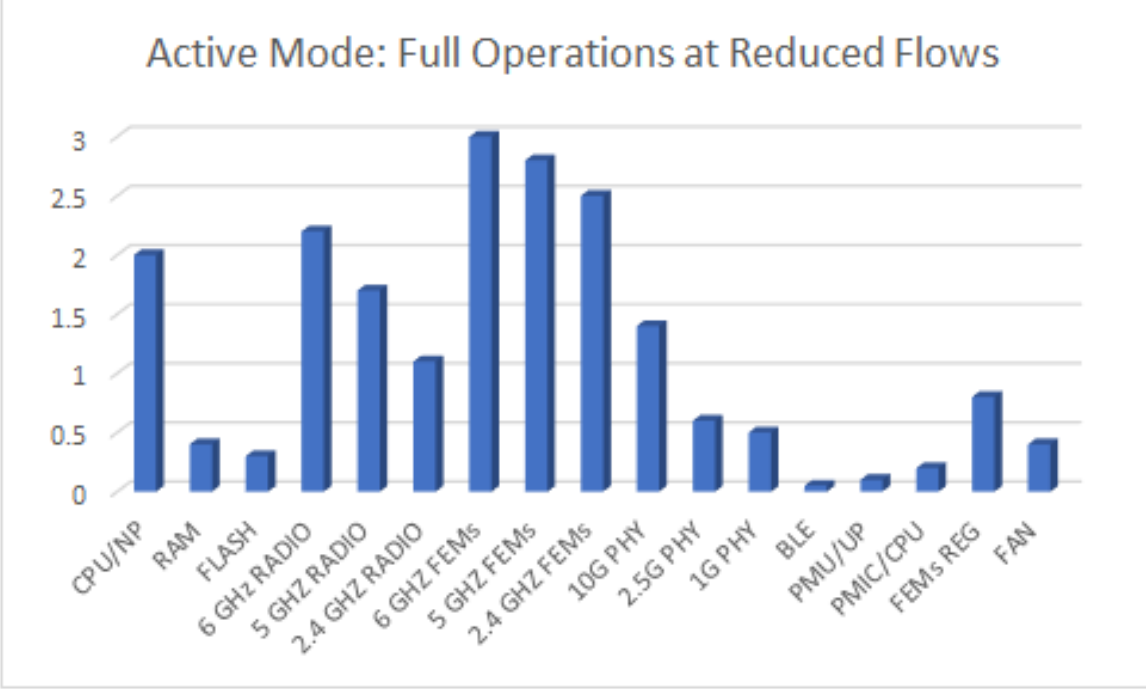


Figure 7 – 20W (>40% Reduction off of Max) Active Mode Power Distribution

Network Standby is, in declining power order, the next state shown in the table and essentially stands as the gateway between an operating AP and one of the two Sleep modes shown. Examining the table-supplied operational states definitions, we see that the device, while still active on all Wi-Fi bands is essentially just advertising presence and awaiting service mounts. At 6 W dissipation, it fundamentally establishes a lower bound on the most common operating range for the high-end AP: 6 – 20 W dissipation; it is this range we will attempt to shift lower by virtue of a slower cadence to awaken and lever the various Ethernet ports and Wi-Fi bands. Here is what comprises the contribution strata to this lower power level:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

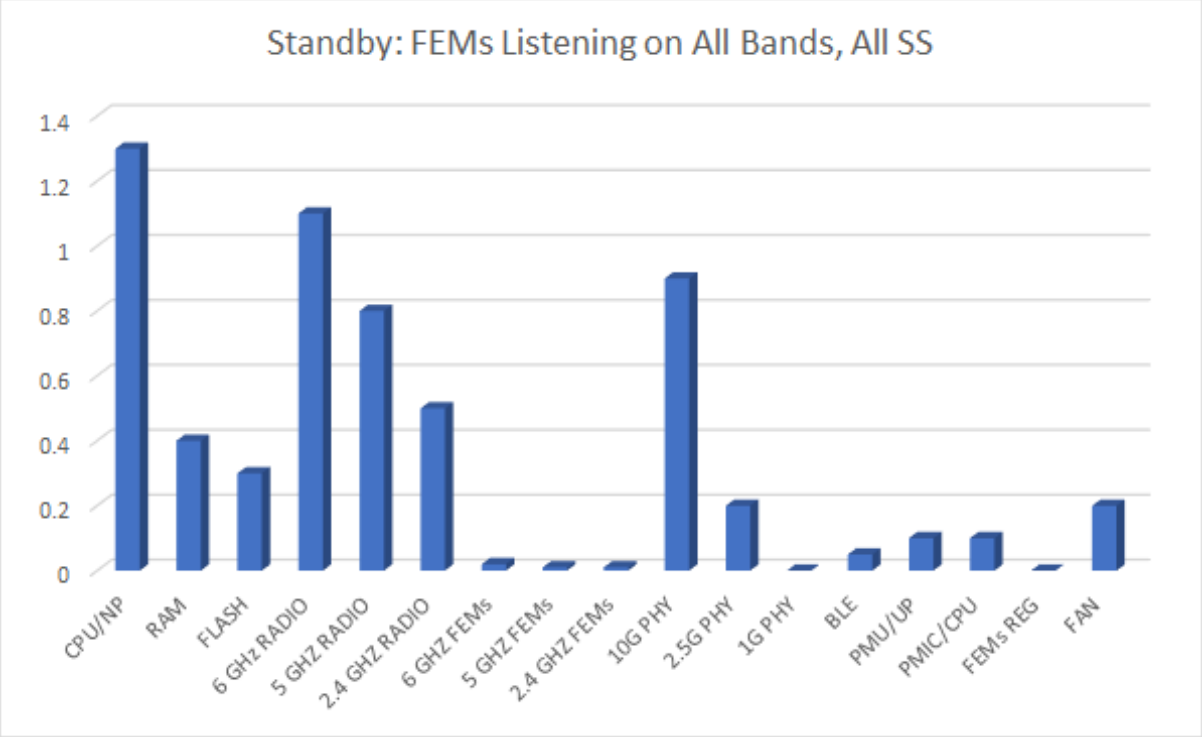


Figure 8 – Standby Power Profile, 6W

This brings us to the most dormant device states: Deep Sleep and "Off" (italicized in acknowledgment of the fact that some awareness is yet available). Regarding the former, the AP still has WAN connectivity; but to exploit this, memory may have to be energized in order to forward IoT-based alarms. Note that leaving the memory up (allowing immediate dispatch of same) would only involve a power penalty of some 0.7 W (moving the power cost to 2.3 W from 1.6 W) so this exigency might well be worth the additional power consumption if it involves medical or security alarm considerations. The alternative would be to lever the small MCU memory footprint to store the necessary routing information to facilitate this action. The profile absent energizing the main memory looks as follows:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

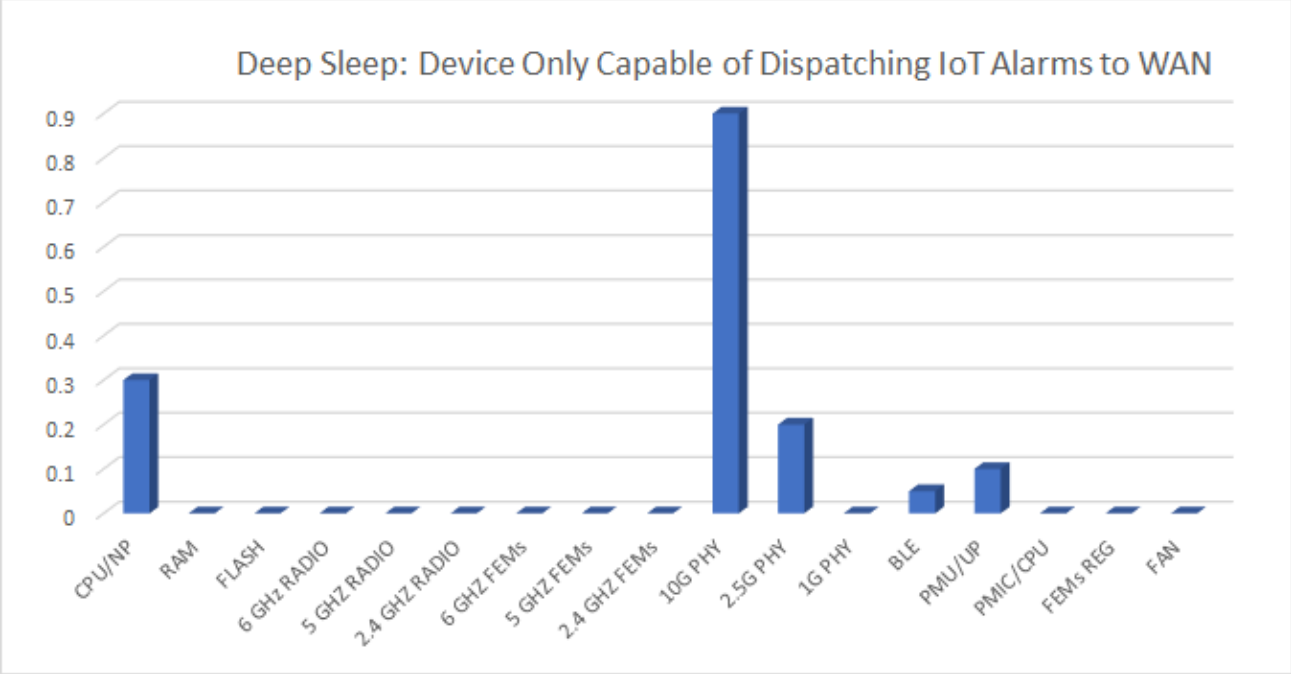


Figure 9 – Deep Sleep Power Profile (Memory Off), 1.6W

Finally, we have the near-fully-lobotomized variant of AP operations associated with the "Off" mode. This fundamentally recreates the ToD-driven brute-force ON/OFF control described in the introductory notes and involves a reconfiguration of the device for it to migrate to, for example, the standby mode. This reconfiguration would involve container image reprofiling so that all ancillary IP services (those sitting astride flows, for example) would be re-instantiated. The exit from this state to something less custodial (or dormant) in fact represents the ToD point where prior history indicates full AP leverage (or "normal" operation) is imminent. The near-empty power landscape for "Off" looks as follows:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

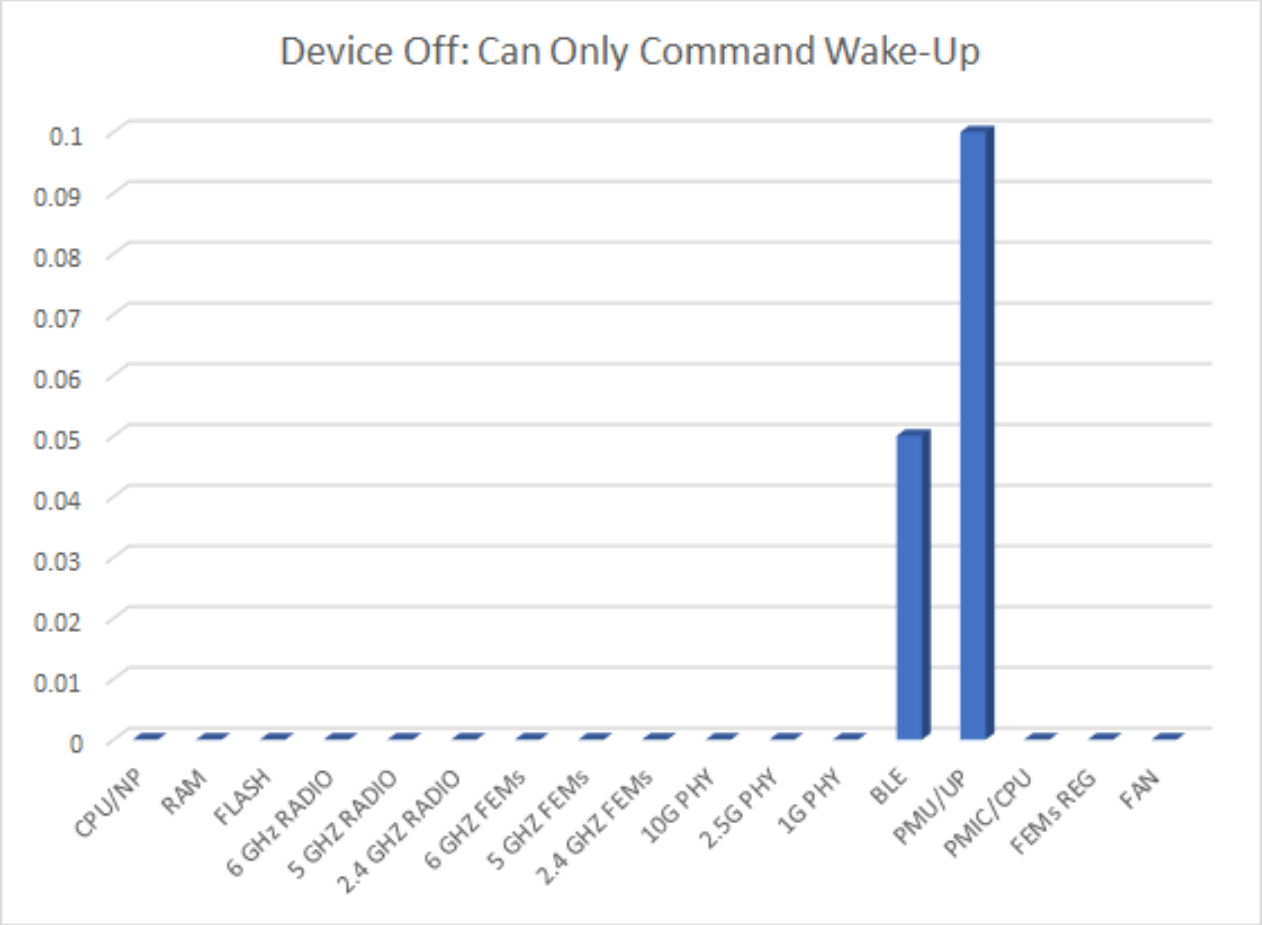


Figure 10 – Device "Off"; Power << ¼ Watt

Note that the IoT is energized such that an application to command power up based upon, say, a cell phone would be possible (to skip past the TOD-driven auto-awakening and allow the user to migrate the device to Standby or Active on a commanded basis).

2.7. Considerations on the Standby -> Active Sweet Spot

Between the 6W of Standby power consumption and the 20W listed for the Active state lie many options for parking the AP's behavior. Below is the power cost matrix which breaks out dissipation of the various circuit functions on a per profile basis:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Circuit Block	Full	Active	Stndby	Dsleep	Off
CPU/NP	2.6	2	1.3	0.3	0
RAM	0.4	0.4	0.4	0	0
FLASH	0.3	0.3	0.3	0	0
6 GHz RADIO	4.3	2.2	1.1	0	0
5 GHZ RADIO	3.3	1.7	0.8	0	0
2.4 GHZ RADIO	2.1	1.1	0.5	0	0
6 GHZ FEMs	5.1	3	0.02	0	0
5 GHZ FEMs	4.8	2.8	0.01	0	0
2.4 GHZ FEMs	4.3	2.5	0.01	0	0
10G PHY	2.8	1.4	0.9	0.9	0
2.5G PHY	1.2	0.6	0.2	0.2	0
1G PHY	0.7	0.5	0	0	0
BLE	0.05	0.05	0.05	0.05	0.05
PMU/UP	0.1	0.1	0.1	0.1	0.1
PMIC/CPU	0.3	0.2	0.1	0	0
FEMs REG	1.4	0.8	0	0	0
FAN	0.8	0.4	0.2	0	0
Power	34.55	20.05	5.99	1.55	0.15

Figure 11 – Power Cost Matrix per AP Functional Circuit

Note that the FEMs represent a frangible line entry, i.e., as the AP has four each spatial streams per band, the power indicated is equally assigned over four devices – each of which may be selectively disabled. This will provide us with additional control granularity for enabling traffic in a particular band, especially as most of the client connectivity represents two spatial streams. (An extender in a mesh network would be a likely four SS link; reduction in spatial density requiring some added consideration since it would impact network performance for several potential clients.) Note that ~7W are tied up in FEM leverage in the Active state, along with 5W in having all bands actively tended to (relative to Standby mode); it is permutation in these areas in particular which will motivate the establishment of certain software behavioral triggers in the following sections.

3. Managing the Power Consumed: Software Behaviors

With the AP's hardware being suitably modified for granular power consumption control, it is now time to discuss how device management can be structured for politic application of assets in various situations to effect responsible standby behavior yet still seamlessly aggregate service demands as these mount. Recall the mention of ToD implications: based on this qualifier, the AP's triggers (and poll delays or time constants) for various asset employment or retirement are bound to diverge as should the expectations for coverage (on average) over the course of 24 hours. Responsivity during peak demand should be necessarily more aggressive at applying a heft of operational assets than might be manifest in device

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

reactionary behavior in periods which experience low average use. Deep packet inspection (DPI) against known service fingerprints also provides the necessary clues (along with historical context) as to the level of QoS (defined along the two axes of bitrate and latency of delivery) expectation the handled traffic has of the LAN.

Note that software administration for the conservation efforts is typically hosted in two domains: AP/local and cloud. The partitioning for these efforts comes by way of offloading situational planning as much as possible into the cloud, with the expectation that this appropriately leverages large-scale machine learning assets in the cloud and increases the level of dormancy in the local device during periods of hibernating service demand. For example, packet differentiated service code point (DSCP) marking for selective QoS handling would already be largely hosted within the cloud, and it then becomes more straightforward to cross this and related packet tagging with ToD and history-related service mount observations and supply the AP with trigger profiles appropriate to when latency-sensitive services are expected to manifest within a given household so that it may be resource-prepped to serve these needs when they arise. (This might involve pre-emptive energizing of the otherwise dormant 6 GHz band when an opportunity arises to mount a highly latency-sensitive service on a client which supports such connection). ToD would also be a bellwether for how deep a hibernating sleep the AP exhibits over the course of the day; maximal retirement of capability would be expected when humans are least likely to seek services of the AP, obviously.

Based on our prior observation of generic ToD usage, a blanket template which sees to lowest power usage from 11 pm to 7 am seems appropriate. The implication here is that, even if one cannot achieve absolute zero power use during this time, such is the goal and one should be at one's most reluctant to apply dissipative resources to any activation during this period; furthermore, gaps in demand should be met with rapid withdrawal of capability since the probability is that any use is likely to be a short, single-link exception to otherwise deep sleep of the AP. (Note that if sporadic history of overnight use is manifest in the device operational history, such may preclude resort of the deepest sleep mode, since this involves the greatest effort at re-initialization of the device upon wake-up – and would perhaps cause more aggravation than would be appreciated by what might be a hard-pressed nocturnal user).

3.1. Band-Specific Characteristics

The three available major bands for Wi-Fi – 2.4, 5 and 6 GHz – manifest different operating environments and serve as residences for differing populations of client devices (and hence, services). 2.4 GHz is the oldest and narrowest band (at 80 MHz total) with the narrowest permissible channels, addressable by all extant MACs and capable of the longest service throws (due to the combination of effective isotropic radiating power (EIRP) and longer relative wavelength) but with the lowest service bitrates – yet also tends towards oversubscription and behaves as a host for overlapping basic service sets (OBSS) and non-WiFi energies (IoT PHYs associated with Bluetooth, Zigbee and Thread). Typically, all Wi-Fi clients have access to this band but because of legacy support, this also implies that a mix of all Wi-Fi MACs can be expected to be encountered (negating many of the significant partial channel and scheduling exploit benefits of the newer MACs – 6 and 7).

5 GHz Wi-Fi also supports all of the MACs 4-7 and features a much wider bandwidth (775 megahertz (MHz), if the seldom exploited U-NII-4 band is included). The shorter wavelength struggles a bit more with signal throw than the 2.4 but there are many 80 MHz wide channels to serve (with appropriate observation and keep-out administered by the dynamic frequency selection (DFS) considerations associated with co-use of some band sections by the military). With a dense enough short/mid-path-length spectral density (64-quadrature amplitude modulation (QAM) and above) and the widest channel BWs, this potentially yields Gbps+ link rates if the stars align on the interference front (meaning no

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

spectrum puncturing and low contention). Lower latency services like videoconferencing and gaming can park here, along with trunk links associated with in-home Wi-Fi mesh extension. It is more point-to-point than 2.4 GHz but has vastly more capacity in its nearly 10x additional BW and wider channels.

Finally, there is the ultrawide 6 GHz band with its prodigious ~ 1.2 GHz of BW (in the US) and home to multiple 160 and even 320 MHz channels – along with MUCH better channel access determinism due to the fact that access to it is only available to Wi-Fi 6E or 7 link endpoints (and so subject only to their schedule-able MAC uses and not to previous Wi-Fi 4/5 MACs). Even with simple single user (SU), priority-unaware round-robin packet dispatch discipline, links in this band can host low single digit millisecond-grade latencies. The availability of standard power trunk links means multiple-Gbps extension(s) are no problem to instantiate, implying that bulletproof QoS coverage of up to 8,000 square foot (sf) (or even larger) premises for qualifying client populations (those of 6 GHz capability) is easily achieved. In particular, even in contentious environments, the ability to schedule exchanges (absent accommodation for at-will accesses by legacy MACs) mitigates random interferences and means such a time division multiplexing (TDM) mechanism can be paired with the frequency division multiplexing (FDM) flexibility of channel puncturing and fine-grain multi-resource unit (MRU) sub-assignment to greatly improve the usability of very wide channel BWs (160 and 320 MHz) in an extender-based coverage scenario (especially useful in single frequency concatenated meshes). As mentioned before, in early adoption days (until perhaps 2026 or even later) the band will be greatly underutilized and very specifically new client targeted, so less than full exploit – especially under constrained power auspices – ought to be expected. In fact, unless single-band 6 GHz clients specifically emerge (almost all will be dual- if not triband), a transitory scheduling effort at using only the first two of the available 3 Wi-Fi bands (especially in low service demand scenarios) makes good sense from a dissipation mitigation standpoint. (This assumes Wi-Fi 6/7 MAC benefits can be applied to the lower bands where applicable, of course.) This is a bookmark only; The applicability of 6 GHz exploit in a coming algorithmic section will be examined – where service priority will also demand an accounting.

3.2. Wi-Fi Roll of Clients and Services

Recall the previous mention of a budgetary matrix of network capacity/capability versus the roll call of mounted services in the premises and the role this would play in modulating power demand in the AP. Nearly every data-consuming service mounted is hosted by a client, said pairing producing parametric telltales which will define the nature of the link delivery for data to and from the client. The raw poll of parameter assignment precedes an ordering of demand details which will set the band and spatial stream leverages to be applied to the Wi-Fi distribution by the AP – and it is this profiling which will be monitored and instantiated with a bias not just blindly toward maximum performance, but with an eye to conservation of applied power to achieve minimal successful disposition of the ensemble services matrix.

Wi-Fi services have long been categorized under Wi-Fi multimedia (WMM) policies for four priorities of airtime access: audio, video, best effort and background (in descending priority order). We can align more or less with this thinking, bearing in mind that priority will be manifest along the dual axes of bitrate and latency (best QoS associated with highest bitrate and minimal latency). A catalog of a representative mix of service delivery follows:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

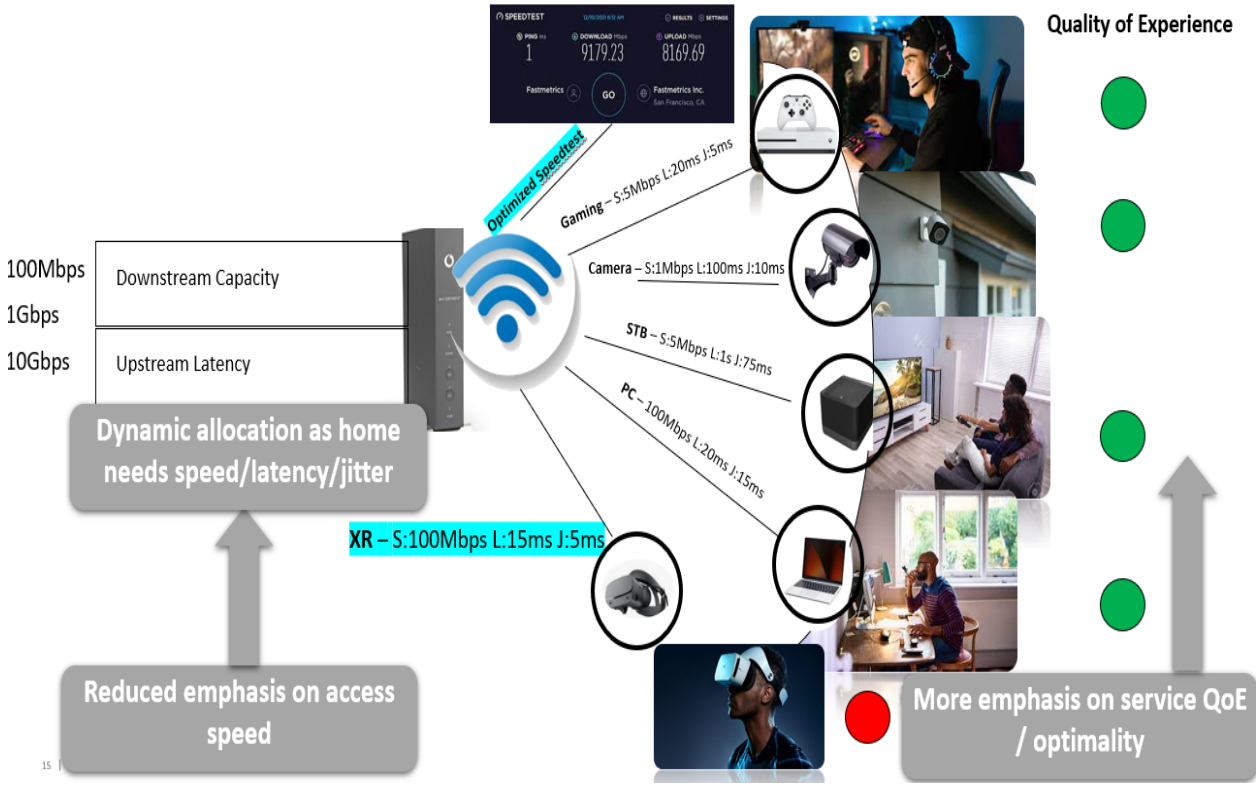


Figure 12 – Mix of Bands, Services and Clients in Wi-Fi Space

In considering delivery requirements, note that both determinism (scheduling conformance to latency bounds) and raw bitrate improve as one advances up in band from 2.4, through 5 and onto 6 GHz. In regard to latency, this can be traced to the reduced occurrence of coincident airtime accesses and an increasing spread of available channels across overlapping basic service set identifications (OBSSIDs) as one moves up-band (allowing channel choice to dodge the worst interference). With respect to bitrate, this evolves in direct reaction to increasing channel bandwidths as one goes up-band. The likelihood of airtime access contention from OBSS or rogue sources also declines over that upwards ramp as well, so ultimately the raw number of services to draw upon the AP will progress along a waterfall where exhaustion of the resource packet delivery "pool" (capacity) available in each band is progressively depleted from a bitrate operating point perspective, thus dictating when the next band up-spectrum is activated to accept some of the load. The pre-emptions to this progression end up being high QoS-based services and client "map-ability" to the various bands, the latter involving exceptions for clients which either cannot connect or connect poorly in some lower band (as in lack of SS resources or, rarely, outright band skip) in deference to operation in a higher band. (In non "green" operations, remember that QoS considerations would dominate the rationale for mapping of band-to-client, with best achievable margins-to-performance the guiding principle and no operational band dormancy tolerated). These two antipathetic strategies actually motivate the potential resort of what amounts to a selectable Econ/Norm switch in AP operational strategy, to permit user input on AP behavior – similar perhaps to the "Econ" mode in some vehicles which maps fuel conservation ahead of outright engine performance. The selection of "Econ" in the AP sense would tend to pack lower bands and subvert full exploit of SS resources where possible – in order to conserve power; "Norm" would move to spread services and pre-

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

emptively push more latency-sensitive traffic to 6 GHz (when client-supported), exploit multi-link operations (MLO) as the situation permits and *always* lever full SS capability.

In any event, the task at hand involves the establishment of service requirement to band capacity ratings. Typical rating of band performance involves use of maximum PHY rates to describe supported bitrates, but these are only achieved at high EIRP and a client service population in immediate proximity to the AP, with all transfers at maximum allowable channel bandwidth and SS leverage; as such, they are unrepresentative and seldom achievable in "normal" use cases. It is therefore better to examine actual airtime costs for connection of a service to a particular client (even though this client may be mobile and subject to some link operating point variability), which at its core involves link EIRP, path losses, modulation and coding scheme (MCS), channel BW and available SS count. Given that some clients can mount more than one wireless service at a time, the airtime budget cost tagging would tabulate as client/service/band/channel/EIRP/MCS/BW/SS. With this history-based 8-tuple for each IP flow, the onset of any client/service pairing (excepting first-time connections) can be pre-emptively assessed for airtime budget impact to existing (or predicted) service mounts based on prior connection and then appropriate band rank ordering can occur for the extant service roster (who gets assigned where, in other words). There is also value in telemetry which monitors delivery performance vis-à-vis packet losses and roundtrip time (RTT) – which provide reliable metrics on contending wireless energies on the channel in question.

Note that, to this point, considerations have been put in place to characterize only SU packet dispatch from the AP (no MRU subgrouping and no MLO binding – the former available in Wi-Fi 6 and 7 and the latter exclusive to Wi-Fi 7). MRU assignment becomes a key adjunct of multi-user (MU) leverage of wideband single channel operations and MLO coupled with MRU can be thought of as a very rare joker card (at least in the coming couple of years) – most invoked for creation of high-capacity multiband trunk links to Wi-Fi 7 capable extenders (operating at peer standard power, perhaps, and capable of MLO/puncturing exploit) and perhaps otherwise reserved for emergency sidecar carriage of data for those few initial clients capable of exercising the MLO option. In a green operational sense, the most likely utilization of these features would fall (by design) to peak use timeframes (the 5-11 pm timeframe) and would otherwise tend to be avoided under conservation auspices unless and until all bands are energized by the AP. The two aspects to tuck away are: 1) for the next couple of years, the number of Wi-Fi clients at MAC epochs 6E and 7 will be relatively low; and 2) scheduling aside, the more advanced aspects of Wi-Fi 7 (very wide BW with MU operations) are not energy-friendly.

3.3. Band Capacity Trigger Points

Each Wi-Fi premises will manifest a history of wireless BW adoption as the active periods progress from early morning to late evening – and such will bear a statistical fingerprint of which a proper AP power scaling algorithm must advantage itself. But as a sort of generic profile (and to provide context where perhaps statistics are not well founded or exhibit wide variance), it may do to project expectations based upon the suggestion provided by Figure 1 in the introduction. Recall that a relatively smooth and monotonic progression (during the diurnal period) of increasing bitrate consumption is forecast by that chart; marrying this to our "increasing bitrate pool" observation above suggests a progression of band leverage that goes as 2.4 GHz only, 5 GHz only, 2.4 + 5 GHz, 5 GHz + 6 GHz (if such step is client-supported) and finally all bands. Note that this progression acknowledges no client band mismatching exceptions and so must be taken with a grain of salt. Accepting this guidance, one can nonetheless calculate band capacity under certain assumptions and generate expansion/collapse trigger points for AP band use. Further stipulation includes that the channel BWs to be used for capacity calculation will be 20 MHz at 2.4 GHz, 80 MHz at 5 GHz and 160 MHz at 6 GHz, with spectral density of no more than 5 bps/Hz (presuming average path losses and median MAC epochs across all subscribed clients) and an

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

airtime transmit density maximum of 25% for 2.4 GHz, 50% for 5 GHz and 90% for 6 GHz (to maintain some margin against transmit backoff accumulation and related deferrals/collisions/OBSS energy – and account for the improvement in this contention behavior as one moves up-band). SS setting at 2 seems advisable as well (though this may be a slight overreach in cases where 3 or less SS are employed by the AP in linking with 2SS-equipped clients). Measurements in the Commscope Wi-Fi house, for example, suggest when spatial streams match, the effective goodput rate essentially reaches 70% of maximum; this translates to 95% or better if a 4x4 is linking with a 2x2 device. The "realistic" band capacities in the generic case then end up being the following:

Band	BW	Capacity
2.4 GHz	20	50
5 GHz	80	400
6 GHz	160	1440

Figure 13 – "Realistic" Channel BW (MHz) and Bitrate Capacities (Mbps) for 2SS Clients from a 4SS AP

Note that, as SWAGs go, these are extremely tenuous dart throws and historical data establishes by far a better – and more premises-appropriate -- view of the wireless landscape and the path losses associated with actual client location distribution. It is not uncommon – in some multiple dwelling unit (MDU) instances especially – to discover that the raw level of OBSS contention at 2.4 GHz renders the band nearly useless for all but background and best effort traffic. It is also true that old "unicorn" single stream clients located far from the AP in any premise can wreck the basic service set (BSS) airtime aperture for other clients in the 2.4 GHz band. As regards to the airtime density factor, this is reasonably inflated (so: improved airtime accesses) as one goes up-band due to MAC maturation and improved channel location options in those instances. Nonetheless, these estimates give us rough per-band bitrate pool capacities we can consider for band energizing triggers as AP utilization ramps up over the course of a day.

Given this stepped view of bitrate demand as posed by the daily cycling, we can perhaps divide the daily cycle into 4 "bins":

Period	Start	End	Duration
Start Up	700	1200	5
Mid-day	1200	1700	5
Evening Pk	1700	2300	6
Night/Idle	2300	700	8

Figure 14 – 24-hr Partitioning of AP Operations (hours)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Referring to our band progression mentioned above, this would seem to drive a band-to-ToD mapping (at least in regard to the Active portion of the AP's operations, between 0700 and 2300 hours) if one, at first cut, regards all clients as at least dual-band capable (and then deals with exclusions as they arise). In addition to the progressive band energizing, there will also be circumspection in assigning full SS resources to a given band until nearing the pool limit (or being driven there by packet loss telemetry which suggests a move to marginally higher capacity is warranted). A boilerplate migration (with SS options) most appropriate to early 6 GHz days (so minimal posed unique-to-band traffic) would look as follows:

	Time	Time	Hrs		Bitrate	Power	Bitrate	Power
Period	Start	End	Duration	Band	w/2SS	w/2SS	w/4SS	w/4SS
Start Up	700	1200	5	2.4	35	10	50	13.2
Mid-day	1200	1700	5	5	280	10.9	400	14.9
Evening Pk	1700	2300	6	2.4+5	315	14.2	450	20.7
Night/Idle	2300	700	8	none	0	0.2	0	0.2
					Mbps	W	Mbps	W

Figure 15 – Boilerplate Bitrate vs Power for AP Operational Permutations over 24 hrs

Note that, for purposes of initial analysis, leverage of the 6 GHz band is not done at all here. It is also presumed that ToD (or historical data) will provide an opportunity to more deeply hibernate the AP, and then awaken it prior to the start-up period (energizing the main CPU/system-on-chip (SoC) and memory and taking the time to establish the correct container profiles). The suggested transitions above provide some permutations as regards SS leverage and so those options (with bitrate benefits and power cost) are shown. In the light-duty case, where all traffic is resolvable to 2SS service only on progressive employ of the 2.4 and 5 GHz bands, one ends up achieving an average daily power use of 191 Whr. For perspective, an unmanaged AP over the course of 24 hours would consume just over double that amount, at around 390 Whr (assuming it can collapse to no better than 6W standby during OFF time). An interesting telltale could be provided by AP's self-reporting "Norm" and "Econ" mode power data over the course of 24 hours, providing users with some value guidance (if they wish to perform the experiment for their particular premises).

3.4. But What About 6 GHz?

The previous section detailed all the reasons for eschewing leverage of the 6 GHz band until some critical mass point of 6 GHz client service is (in future) involved; it turns out that this trigger point happens quickly and for very good reason. If one takes our prior figure on "realistic" bitrate capacity on the various bands (to see when one might begin energizing more bands based on expected capacity exhaustion over the course of a day) and add in the power cost for powering up the band exclusively with 4 SS worth of spatial coverage, a calculation that amounts to a transducer gain (as Mbps/W) worth of band efficiency and the result provides a startling modification of our boilerplate plan can be made:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Band	BW	Capacity	Power	Bitrate/W
2.4 GHz	20	50	10.35	4.83
5 GHz	80	400	11.25	35.56
6 GHz	160	1440	11.95	120.50
			W, 4SS @ 50%	Mbps/W

Figure 16 – Relative "Realistic" Band Efficiencies, Bitrate/W

The blunt takeaway is that catering to 2.4 GHz clients on behalf of their legacy installation costs dearly on the power efficiency front and instead operation of that band ought to be considered the least desirable option (a huge opportunity power cost) for any service. This observation underscores the true cost of continuing to use 2.4 GHz-only clients in a modern Wi-Fi mesh; add to this the difficulty of providing decent low latency service in the band due to its multi-PHY oversubscription and narrow spectrum/channels and perhaps the best one can do for users in pursuit of AP energy efficiency is to have them substitute newer 5- or 6-GHz clients for older 2.4-GHz devices.

So, what does this do to the boilerplate approach? Essentially, it puts three bookmarks in place right away: if there are no exclusively 2.4 GHz clients in the home LAN, one should move initially to energizing only the 5 GHz band for Wi-Fi use instead of the 2.4 and move the dual- and tri-band clients to 5 GHz service. Second, one should consider 2.4 GHz only for exclusive-to-band clients and those whose service radius otherwise exhausts the AP's throw. (Remember that, on the services front, backhaul of 2.4 GHz-based security camera feeds likely hits both of these parameters). Third, to the extent that both lower bands *must* be energized anyway, it becomes worthwhile to query how much of the day's capacity demands can be met with only 2SS on each of the first two bands (recall the "realistic" estimate that this would serve up to 315 Mbps or so worth of bitrate – though, of course, admittedly perhaps not to acceptable latency limits).

All of this gets to the question of 6 GHz, succinctly put: "when should it be used in a power conserving paradigm?" The answer becomes a calculation on opportunity cost; given that new FEMs almost exclusively only dissipate in transmit mode (very much reduced standby operation power), the 6 GHz band should be energized when the service load power cost undercuts operational power cost in an alternate band by an amount greater than the cost associated with 6 GHz radio chip dissipation – approximately 2 W. Based on our previous calculations for band-specific bitrate-to-power efficiencies, this occurs at approximately 101 Mbps for a 6 GHz-based replacement of a 5 GHz-based service and only 10 Mbps for a 2.4 GHz-based service (though it would be rare to find a client with only 2.4 and 6 GHz connectivity – and those multiband capable should have already been migrated to the more efficient 5 GHz band). The algebra looks as below:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

$$\frac{Bitrate_{SERVICE}}{Transducer_X} >= \frac{Bitrate_{SERVICE}}{Transducer_6} + 2$$

$$\vdots$$

$$Bitrate_{SERVICE} >= \frac{2 * Transducer_X * Transducer_6}{Transducer_6 - Transducer_X}$$

$Bitrate_{SERVICE}$ → Necessary amount of service bitrate to motivate the 6 GHz band use

$Transducer_X$ → W/Mbps for band x

$Transducer_6$ → W/Mbps for 6 GHz band

Figure 17 – Tipping Point Calculation for Xfer of Service from Band X to 6 GHz

(Quick sidebar: the same tipping point calculation can be done for 2.4 GHz to 5 GHz loading using a 1.5 W offset factor instead of the 2 W above and the result ends up being 8.4 Mbps; this underscores the comments above that the 2.4 GHz band is best used for 2.4 GHz-only client legacy work – and after that, really suited only for best effort and background LAN tasks for DBC/TBC clients as a sort of emergency offload. As soon as service bitrates approaching this low Mbps are reached, it makes more sense from a power-efficient packet delivery standpoint to immediately move the service to a higher band if the client supports it.)

So now there are some directions for the considered exploit of the 6 GHz band in a power-conserving paradigm. First, if one has 6-GHz band only clients, there is obviously no choice but to enable the AP's use of the band. Second, once one has exhausted the available 4SS capacity pool of 2.4/5 GHz for multiband clients with 6 GHz capability, there is need to transfer some of this load (overflow) to 6 GHz to those clients capable of leveraging it. Finally, even if the pool capacity has not been exhausted for 2.4/5 GHz services but such have accumulated either 101 Mbps worth of 6 GHz client service offload of the 5 GHz band – or (unusually) 10 Mbps of service offload from the 2.4 Ghz band (for, say, a triband-capable client), then the 6 GHz band should be enabled, and these (and subsequent band-matching) service prospects moved to it. Note that this offload applies equally for trunk link replacement – so while 100 Mbps seems like a fair amount, the onset for a 5 GHz trunk link replacement (assuming the extender in question is 6 GHz-capable) can happen quickly.

One final note on 6 GHz exploit and this has to do with considerations on an alternate axis from raw bitrate support: latency. Unlike the determination of opportunity cost on the bitrate capacity front, the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

consideration for latency typically comes down to two less calculable considerations. The first is an absolute LAN latency limit requirement of less than 5 msec or so; if this is the case (as would be true for perhaps gaming or related simulation work) and the service has a 6 GHz mount option for its leveraging client, then such should be chosen. The second consideration is based on telemetry alarms regarding QoS SLA in a dual-band concurrent (DBC)/TBC scenario, where the SLA is being routinely violated despite even 4SS connectivity to the client in any other band option; in this case, power considerations must be sidelined and if the client is 6 GHz-capable, the band necessarily must be enabled, and the affected client/service pairing moved to 6 GHz to (hopefully) resolve the QoS violation.

3.5. Proposed Flow for Implementation of AP Power Management

It is now time to describe behavior for power managing a TBC AP; the hardware is suitably granular in control and candidate leverage points exist along with an understanding of the merits and parametric weights associated with activation of these controls – enough at least to propose a general flow diagram which addresses shifting cyclic service loads over the course of their (daily) frequency of occurrence. As previously noted, the guidance here will not be to maximize performance unilaterally but to attempt to scale the AP's power use to acceptable levels of accommodation for the services which progressively draw higher bitrates from early morning until late evening of every day. The principal drivers for AP change will be ToD, client population band dependencies, service roster, QoS SLA(s), premise service history, efficiency of the various bands and link telemetry of ongoing transfers. In general, ToD would be used to start and stop the AP in alignment with historical patterns for start/stop (for the latter, the presumption would be that the traffic would have already stopped; otherwise, the ToD STOP becomes an arming signal for the Sleep process which awaits a timeout period of inactivity before commanding shutoff). The proposed control block diagram goes as follows:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

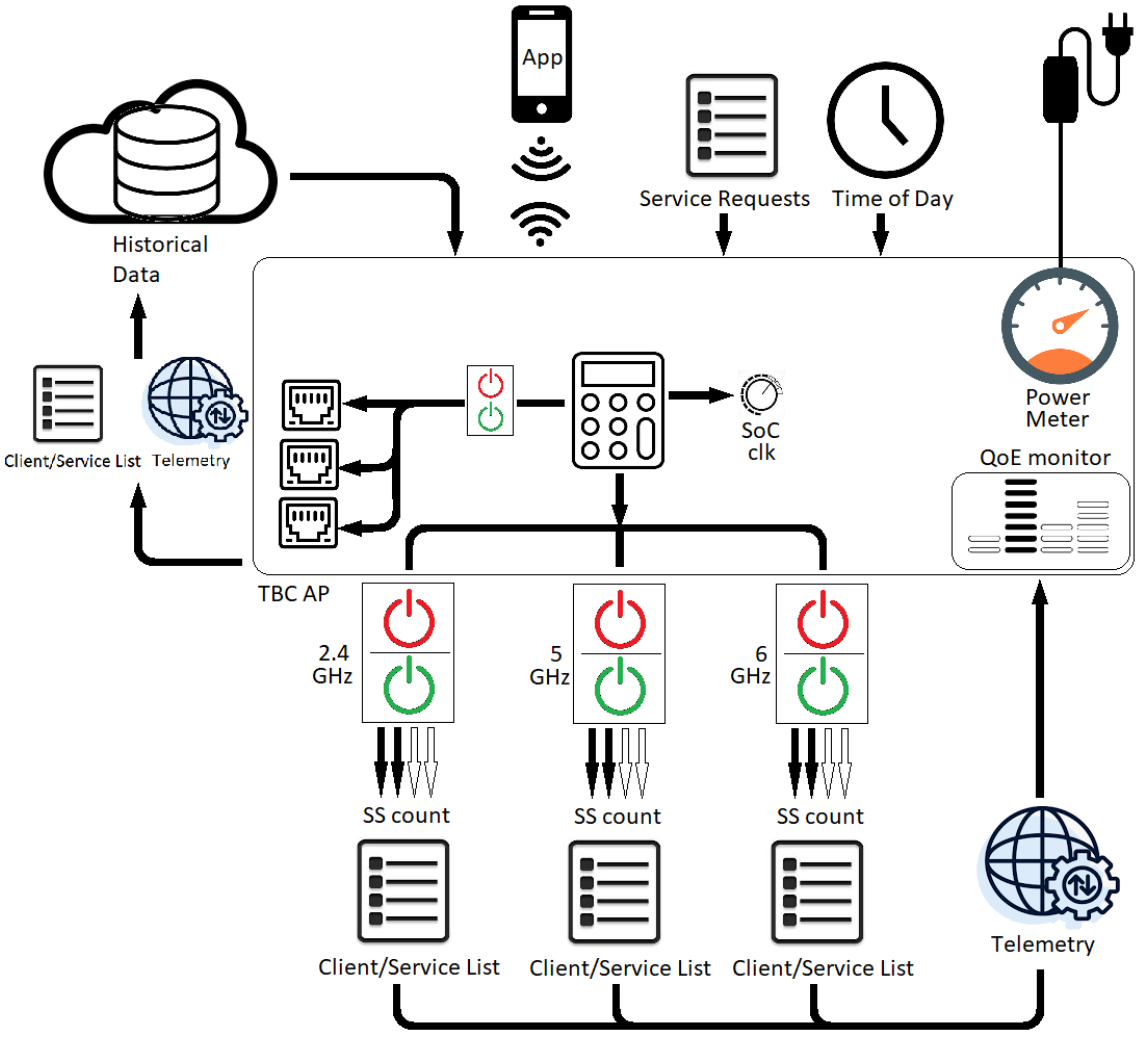


Figure 18 – AP Block Diagram Indicating Green Operations

The general flow of decision-making for transit of the above block would consist of time-driven ON and OFF cycles and then operations within the ON domain. The general ON/OFF block flow (which adjusts AP state based on ToD and history) goes as follows:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

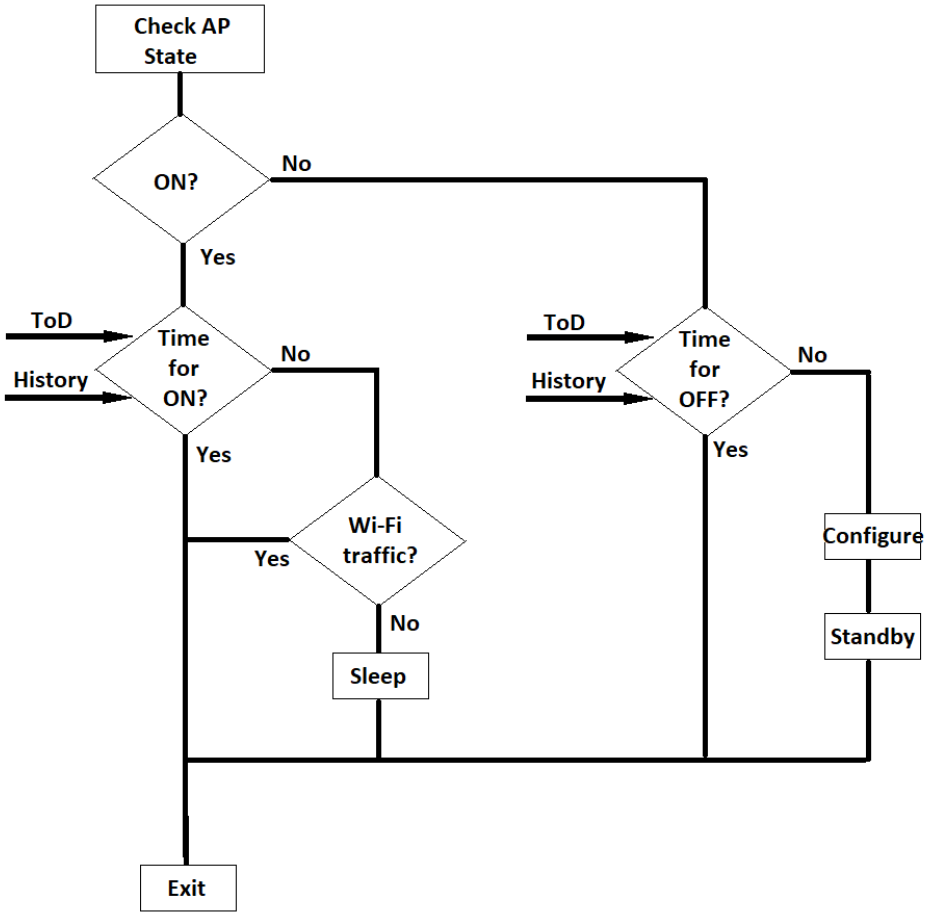


Figure 19 – Polling Logic to Turn AP ON/OFF

Note that, as referred to previously, a phone application over BLE could be used to interdict the automated pacing of AP availability and selected profile or mode, either to pre-emptively awaken the device ahead of standard schedule, change the Econ/Norm mode or to extend operating hours (as either a day extension or an abrupt, atomic nocturnal event). The value of storing history over a week-long period (or longer) allows the AP to adjust to weekend schedules, where daily work usage patterns (and the related service profiles) might find themselves significantly altered from midweek norms.

The next illustration addresses the regular diurnal progression from Standby through end-of-day operations; the summarized operation strategy looks as follows:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

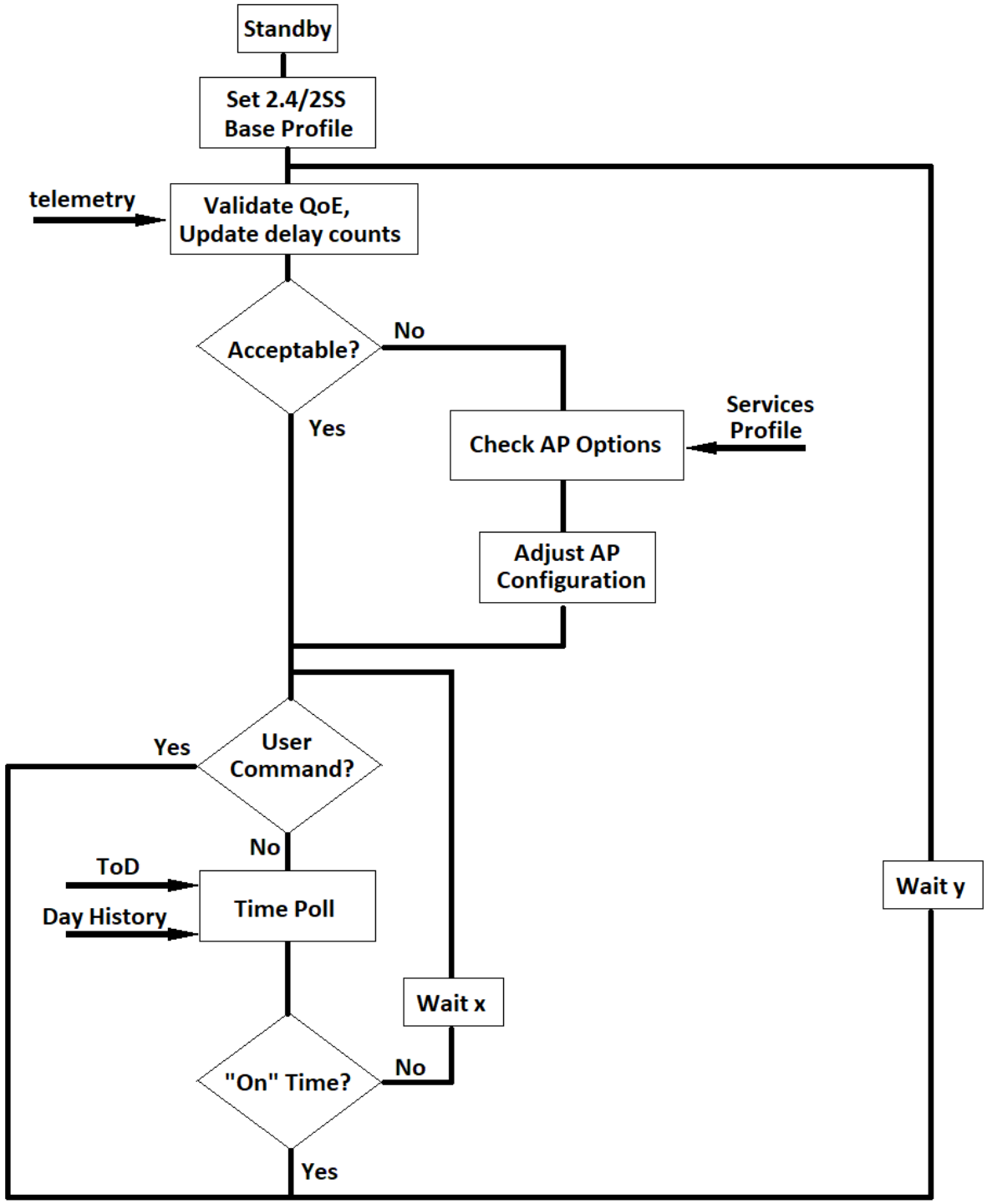


Figure 20 – Moderation of AP Profile to Fit Services Mounted

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The hold times for the various AP states can work against "speed to upgrade" versus "speed to downgrade" considerations associated with ON or OFF state and the ToD. Once committed to 4SS service on a given band during diurnal operations, for example, the AP could be set to only slowly decommission some of the spatial coverage, pending expected resumption or expansion of day services. In inverse consideration, if 2SS is already committed to a given band and hints of QoE miss are noted, upgrade to 4SS quickly may be advisable (as would energizing, and moving to, an available up-band to at least 2SS, if such matches the client band coverages).

The progression of 2.4, 5, 2.4+ 5, 5+6 and all bands – subject to client band matching and tipping point calculations (band efficiency fitness for services large enough to be offloaded from lower bands) – provides the basic guidance in these determinations.

4. Conclusion

The European Code of Conduct regarding expectations for CPE provides guidance on what expectations for responsible powering of APs should target; and while US cost of electricity undercuts the European markets by up to 3x (tending to mute the energy responsibility outlook domestically), the opportunity for homogeneous design elements (including SoC functional partitioning) in product sold on a common basis around the world, coupled with shared appreciation for similar specification, seem likely to eventually land the majority portion of these eCOC KPI into US Energy Star profiles. Anticipating such, pre-emptive SP endorsement and patronage of industry efforts to reduce the energy profile of deployed CPE network endpoints would almost certainly experience market reward for responsible citizenship. In supporting task-appropriate energy consumption, the AP signals that it belongs in a 21st-century roster of responsible home electrical appliances – perhaps a bit late to the electrical use sustainability party, but ultimately a conscientious and clever device in that regard.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

3D	three dimensional
5G	fifth generation
AC	alternating current
ADC	analog-to-digital converter
AP	access point
BLE	Bluetooth low energy
bps	bits per second
BSS	basic service set
BW	bandwidth
CPE	consumer premises equipment
CPU	central processing unit
DBC	dual-band concurrent
DC	direct current
DPI	deep packet inspection
DSCP	differentiated service code point
eCOC	European code of conduct
DFS	dynamic frequency selection
DMIPs	Dhrystone millions of instructions per second
EIRP	effective isotropic radiated power
FDM	frequency-division multiplexing
FEMs	frontend modules
Gbps	gigabits per second
GHz	gigahertz
HW	hardware
Hz	hertz
ID	industrial design
iFEM	internal frontend module
IoT	internet of things
IP	internet protocol
KPI	key performance indicator
LAN	local area network
LoRa	long range, wide area
LTE	long term evolution
MAC	medium access control
Mbps	megabits per second
MCU	microcontroller unit
MCS	modulation coding scheme
MDU	multi-dwelling unit
MHz	megahertz
MLO	multi-link operation
MRU	multi-resource unit
MU	multi-user

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

NFC	near field communications
OBSS	overlapping basic service sets
OBSSID	overlapping basic service set identification
PAN	personal area network
PHY	physical layer
QAM	quadrature amplitude modulation
QoE	quality of experience
QoS	quality of service
RAM	random access memory
RF	radio frequency
RTT	round-trip time
SAR	segmentation and re-assembly
sf	square feet
SG	service group
SLA	service level agreement
SoC	system on a chip
SP	service provider
SS	spatial streams
SU	single user
SW	software
SWAG	scientific wild-a$$ guess
TBC	tri-band concurrent
TDM	time-division multiplexing
ToD	time of day
TXOPs	transmit opportunities
W	watt
WAN	wide area network
Whr	watt-hour
WMM	Wi-Fi multimedia

Bibliography & References

Broadband Capacity Growth Models, John Ulm, Zoran Maricevic, Ram Ranganathan, SCTE Technical Paper 2022

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23 CELEBRATING 40 YEARS

2023 Fall Technical Forum SCTE • CABLELABS • NCTA

Pairing IoT and AI to Reduce Network Maintenance Costs

A Technical Paper prepared for SCTE by

Goutam Agarwal
Principal Enterprise Architect
Rogers Communications
2400 32 Ave NE, Calgary, AB T2E 6T4
goutam.agarwal@rci.rogers.com

J. Clarke Stevens
Independent Consultant
cstevens63@yahoo.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Network Maintenance Challenges 5
2.1. Network Maintenance Program Evoution 5
2.2. Network Maintenance Program Challenges 6
3. Key Use Cases for Network Maintenance Improvements 8
4. How can IoT and AI help Cable Operators? 9
4.1. Accurate Input (IoT) 9
4.2. Automated Diagnosis (AI) 9
4.3. Automated Solutions (AI) 10
4.4. Improved Remediation (IoT) 10
4.4.1. Fix Problems via Actuators 10
4.4.2. Precisely Locate Affected Equipment and Cabling 10
4.4.3. Reroute Traffic Around the Problem 10
4.4.4. Instantly Update Configuration and Inform Affected Customers When Network Restored 10
4.4.5. Anticipation 11
4.4.6. Avoidance 11
5. What Do Telco Technology Leaders Need To Do To Start Their AI/IoT Transformation Journey? 12
5.1. Lessons Learned 12
5.1.1. Embed AI at the Core of Business Processes 12
5.1.2. Strategic Ownership and Alignment 13
5.1.3. Use Agile Process and Principles 13
5.1.4. Treat Talent as a Scarce Resource 13
5.1.5. Future-Proof Data and Systems 13
5.2. Technology Best Practices 13
5.2.1. Software Driven Network Operations 14
5.2.2. Cloud Native Principles 14
5.2.3. Use of Open Data Model and Standards 14
5.3. Challenges and limitations 14
5.3.1. Data Privacy 15
5.3.2. Cybersecurity Risks 15
5.3.3. Ethical Considerations 15
5.3.4. Data Management and Scalability 16
5.4. Collabrative Governance Approach 16
6. Conclusion 17

Abbreviations 18

Bibliography & References 18

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Figures

Title **Page Number**

Figure 1: Network Maintenance Program Evolution 5
Figure 2: Network Maintenance Program Challenges 6
Figure 3: Cable Operator's Integration Approach for AI and IoT 9
Figure 4: Key Lessons Learned 12
Figure 5: Key Technology Best Practices 14
Figure 6: Key Challenges & Limitations 15
Figure 7: Next Gen Network Operations Management 16

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The most fundamental component of a cable service provider's products is the network. The networks we provide are ubiquitous in the lives of our customers. Any network disruption causes inconvenience, but as more people work from home, run businesses over the network, or manage critical services online, an outage can increasingly have a serious financial effect or even put people's lives at risk.

It is imperative that our networks provide the features, quality, and reliability our customers have come to expect. There are many obstacles to maintaining quality and reliability: Data is frequently erroneous or arrives too late to be of significant value; finding qualified employees for a more complex portfolio of products and getting those people to the right place at the right time to fix problems is increasingly difficult and expensive; diagnosing problems is often a matter of experience that is isolated to a few experts. It is expensive to send people to fix plant problems, and the prescribed fix may not always be effective.

One possible solution is to provide the necessary technical capability to the technicians to cut through complexity and reduce cost, while allowing the network to self-diagnose and in many cases repair problems without the need to roll a truck. Such tools are built on the concept of the Internet of Things (IoT). This includes sensors and actuators in the network and Artificial Intelligence (AI) algorithms to interpret and correctly diagnose or even automatically repair network problems.

While instrumenting the network is an up-front capital cost, it can pay for itself by providing an accurate real-time view of the status of the network. AI can interpret the collected data and use that data to provide an accurate diagnosis of the problem, isolate a precise location for the fix, then potentially reroute traffic while a repair is made and even automatically make the repair. This results in a more reliable network that can be maintained at a lower cost.

In this paper, we will explore existing and potential IoT solutions that can be installed in the network, and how IoT and AI can be combined to build network infrastructure that is simpler and less expensive to maintain. We will frame this as a progressively improving solution from a reactive response, to planned upgrades, then moving to proactive prevention and finally to a predictive solution that anticipates and resolves problems before they happen. This approach migrates the network from a resource that exhibits frequent downtime into a resource that is rarely down while being less expensive to maintain.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2. Network Maintenance Challenges

2.1. Network Maintenance Program Evoution

Traditionally, Cable Operators have tried to evolve their network maintenance program from a reactive to more proactive maintenance, with predicative being the elusive goal. Figure 1 illustrates a common view into the evolution of network maintenance programs.

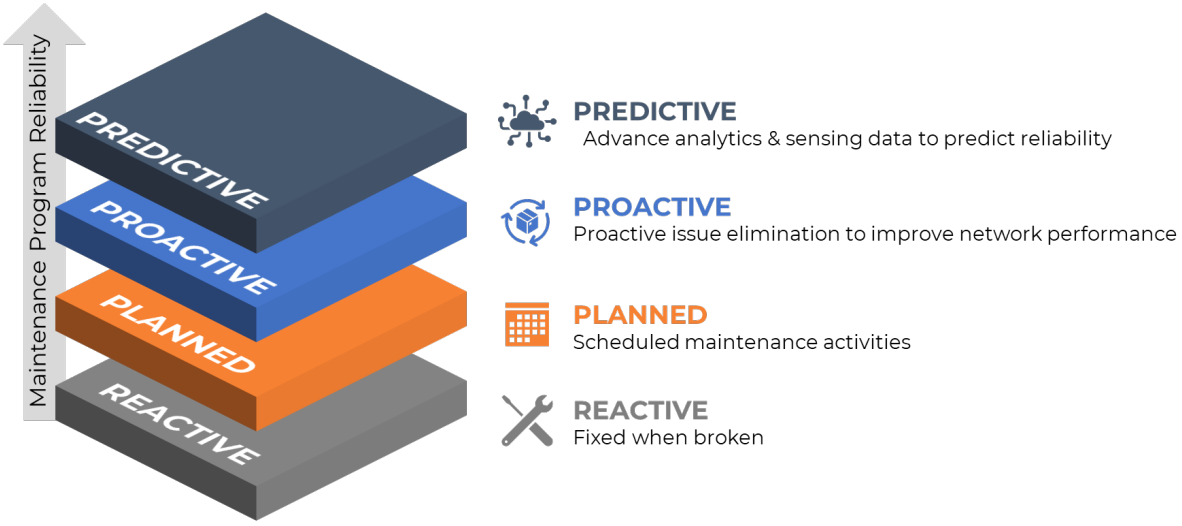


Figure 1: Network Maintenance Program Evolution

- **Reactive Maintenance** – Allowing parts to run to failure.
  As the least technologically advanced and most common level of network maintenance, reactive maintenance involves repairing or remediating parts or equipment only after they have broken down or have been run to the point of failure. Reactive maintenance relies on people and using existing tools and components to their limits. However, this can lead to catastrophic machine damage as humans make mistakes and network components get damaged due to vibration, overheating, potentially causing unknown damage to network assets beyond the failed parts.

- **Planned Maintenance** – Attempt to prevent problems before they occur.
  A planned, time-based preventative maintenance approach can help avoid broken network assets and decrease downtime by replacing parts at regular, pre-planned intervals. While planned maintenance may be more cost-effective than reactive strategies, it also can be more difficult to justify as the parts are replaced before they have reached their end-of-life and requires an inventory of spares to be managed. These can also result in greater planned downtime.

- **Proactive Maintenance** – Treating the root case, not the symptom.
  Proactive maintenance strives to identify and address the problems that can lead to those breakdowns in the first place, such as configuration misalignment and, environmental conditions. By identifying and addressing the root causes of many component failures, proactive maintenance typically helps to prevent the wear and tear that leads to equipment failure, ultimately decreasing failures and downtime. This can also help in improving operational efficiency by reducing unnecessary repairs and improving the lifespan of network assets. Proactive maintenance requires organizational support to be successful and often requires more extensive training and robust change management practices.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- **Predictive Maintenance** – Fix issues before failure (analyze trends, predict failures)
  Predictive Maintenance can help break the inefficient cycles of older strategies by enabling cable operators to maximize the useful life of their network assets while avoiding *both planned and unplanned* downtime, reducing cost and increasing operational efficiency. On a very high level, Predictive Maintenance analyzes operators' network data collected from the connected equipment to predict when a network asset needs maintenance or replacement. But simply gathering the information from sensors and systems is not enough to yield the benefits of Predictive Maintenance; the ability to aggregate and then analyze data can be crucial to predicting malfunctions. This often requires new capabilities for creating, handling, and making use of data.

2.2. Network Maintenance Program Challenges

According to a McKinsey Report [2], Network maintenance and service operations account for 60 to 70 percent of most telcos' operating budgets. The industry has already faced a decade-plus of increasing cost pressure, and the returns on necessary infrastructure investments are barely outpacing the cost of capital. To stay ahead, operators will need to make critical investments in network maintenance programs to offer efficient and effective processes to drive costs down while improving customer and employee experience. With the evolution of network maintenance programs, there has been an addition of complexity in the forms of modern technology, new services and delivering the services in a reliable manner. This scaling of the network and complex environment comes with considerable challenges that inhibits cable operators' ability to exit the reactive maintenance world (scale of reactive items challenging their ability to resolve). What ends up happening is the teams have very little remaining capacity to deal with proactive/predictive items. In Figure 2, below we highlight some of the key challenges that cable operators face on daily basis which can be resolved with the help of AI and IoT as the key technology enablers -

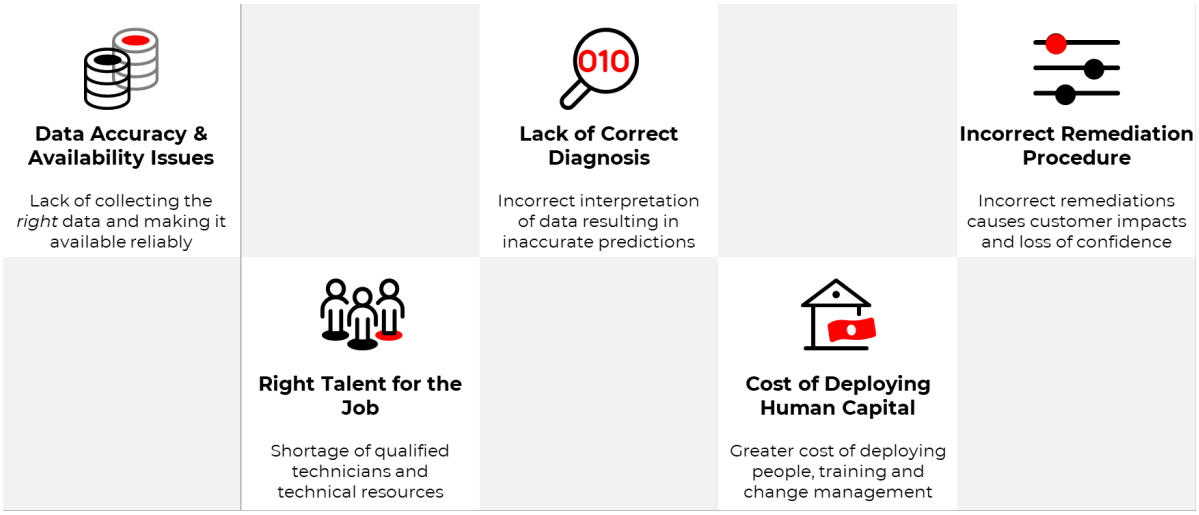


Figure 2: Network Maintenance Program Challenges

- **Data Accuracy & Availability Issues** –
  In the past, network maintenance often required time-consuming manual data crunching and resource intensive processing to generate insights from data being collected. The strategies relied heavily on having in-depth knowledge of the data and expectations that the data is accurate and

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

available to the users when needed. The decreasing cost of sensors, computing power, and bandwidth, coupled with increasing technological advancements, has enabled us to capture more data, more often. Data accuracy, integrity, real-time data flow and data availability have been challenges with which cable operators struggle as networks increase in complexity necessitating more sophistication in network maintenance.

- **Lack of Correct Diagnosis** –
  Simply gathering the data / telemetry from sensors and connected devices is not enough to yield the benefits of proactive / predictive network maintenance. One should be able to aggregate, analyze data and predict diagnosis with a high degree of confidence. Building trust in the process and models doing the predications takes time and effort – developing validation scenario and test cases – and may require novel approaches until the space matures with open-source and vendor-supported models.

- **Incorrect Remediation Procedures** –
  After gathering the data and being able to correctly predict correct diagnosis – not guesses (false positives) – about network assets and health, the next step is to provide the right remediation steps (manual or automated) to drive optimal efficiency for network maintenance. Remediation processes require a much deeper knowledge of the network and customer dynamics, so that remediation not only happens on the right network asset but also at the right time considering customer experience impacts. Remediation also involves proactive communication of impacts to customers and paths to resolution. Coalescing with the ubiquity of networks and the unpredictability of customer usage patterns, Cable Operators have been challenged in balancing network remediations with customer experience.

- **Right Talent for the Job** –
  As the network becomes more complex, Cable Operators are challenged with shortage of right talent for the job (e.g., right network technicians for remediation, or skilled technical resource required for the data synthesis and analysis, cloud native technologists), coming into competition with the *hyperscalers* (e.g., AWS, Azure) and other core technology companies. Some of the operators are viewing network maintenance as a technology problem but talent, process and organizational change are equally if not more important. Reducing the requirement for specialized talent and replacing it with accurate and effective automation can help.

- **Cost of Deploying Human Capital** -
  Operationally lean and efficient network operations and maintenance is what Cable Operators have been striving for. The cost of deploying human capital to manage the network complexity has been on a constantly increasing curve and is, in a capital-strapped ecosystem, one of the key challenges being faced by the network operators.

In the subsequent sections we will discuss how the strategic use of IoT and AI can help Cable Operators to overcome some of these challenges and to be able to run lean network operations supported by proactive and predictive network maintenance.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Key Use Cases for Network Maintenance Improvements

In the telecommunications industry, network operations and maintenance are often said to be one of the most complex aspects of the business. Historically, most successful telcos tend to outperform at this task with close coordination across business units. It might sound simple but is a herculean task and requires expensive manpower and resources. We present below a sample of the key use cases that cable operators face in detecting flaws in the network, network security, network optimization, while offering real-time improvements in managing day-to-day network operations. Central to these use cases are the more efficient resolution of issues and providing enhanced customer service and satisfaction.

- **Fault Detection, Prediction and Resolution**
  Service impairments and faults are inevitable in a Cable Operator's network. Operations primary consideration is maintaining a functioning network and that faults do not result in large costs, whether through maintenance costs themselves or penalties for breaching SLAs. Given these concerns, three areas emerge for measuring fault detection, prediction, and resolution:
    1. **Customer experience**: how can the cable operator limit the customer impact radius and be able to provide a highly confident estimate of restoration of the services.
    2. **Human capital cost**: The cost of network engineers fixing the problem, as well as those customer-facing roles dealing with complaints.
    3. **Speed**: How quickly the Cable Operator can identify the problem and therefore solve it; measured as Mean Time to Repair (MTTR)

- **Network Optimization**
  Network optimization is about how to route traffic and balance workloads across the available infrastructure and assets to try and deliver the highest quality of most cost-effective service. It is possible to optimize the network manually, but with tens of thousands of nodes, headends and hub sites this would mean the whole team of network engineers doing nothing but re-optimizing the network. With the high cost of staffing network operations, manual network optimization is not practical. There is a need for self-optimizing network which can provide real-time, event-based network traffic balancing. Measurements for the success of self-optimizing network fall under the following areas:
    1. **Cost Optimization**: Maximize the use of existing network assets; predicting and optimizing network traffic that would otherwise incur higher costs.
    2. **Quality of Service**: Service prioritization during demand peaks or disruptions and reverting to default when the event is over; prioritizing firmware updates to provide least services disruption to customers.

- **Network Planning and Upgrades**
  Predicting / forecasting customer demands for products and services to assist cable operators in rolling out network upgrades that are designed to enable new services by using the right deployment model between mid/high-splits, DOCSIS 4.0, FTTP overbuilds. Proactive network optimization and upgrade planning can help operators switch from site-centric rollout planning - based on engineering guidelines - to a customer value-centric rollout that identifies priority sites using technical needs, customer experience metrics, and financial data, factoring in regulatory-approval timelines and local vendor delays.

These are only a subset of numerous compelling use cases that can be leveraged by Cable Operators to stay competitive in an increasingly challenging market. In the subsequent sections we will discuss how, by investing in AI and IoT and adopting best practices, Cable Operators can future proof their operations and remain competitive in the years to come.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. How can IoT and AI help Cable Operators?

AI and IoT has the potential to revolutionize the Cable industry. The integration of Artificial Intelligence (AI) in Internet of Things (IoT) introduces new dimensions of efficiency, automation, and intelligence to our daily lives. IoT provides the interface between the computing power of AI and the real world. It allows reality to be accurately measured in real-time, then uses actuators to allow changes to be affected in the real world. Artificial intelligence has revolutionized the way machines learn, reason, and make decisions. When combined, Artificial Intelligence and Internet of Things opens a realm of possibilities, enabling intelligent, autonomous systems that can analyze vast amounts of data and take actions based on their insights.

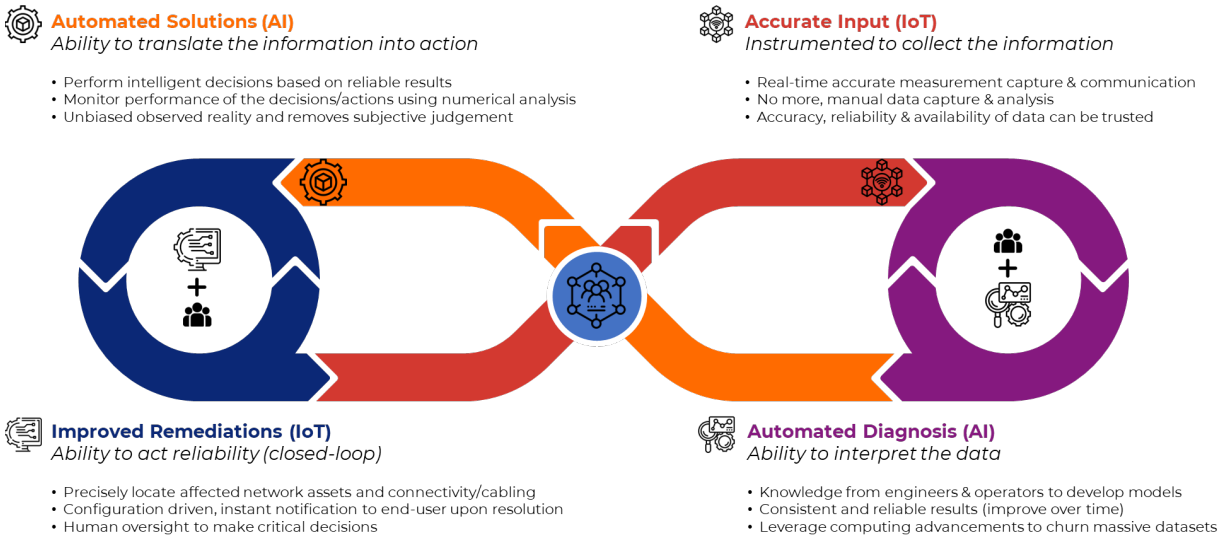


Figure 3: Cable Operator's Integration Approach for AI and IoT

4.1. Accurate Input (IoT)

The Internet of Things (IoT) is the universe of networked sensors and actuators. One of the most important aspects of IoT is its ability to take accurate direct measurements and communicate those measurements immediately. Before the advent of the IoT, measurements were often made by bulky analog instruments, data was transcribed by hand or automated in batches and analyzed later. There were many ways to introduce errors or lose important real-time context. IoT sensors, on the other hand, can be tiny, digital, and numerous. Their accuracy, reliability, and instantaneity mean the data can be trusted and evaluated in real-time and used to make instant decisions.

4.2. Automated Diagnosis (AI)

Instant decisions can only be trusted if the data is accurate, and the decisions are reliable. AI gives us the techniques we need to make decisions based on enormous amounts of data and verified results. Machine Learning (ML) is one form of AI that is based on statistical analysis. It identifies patterns in input data and observes results. When the data consistently produces reliable results, making decisions on the data seems "intelligent." Often, the sheer amount of data is too much for humans to process, but this is where computing power shines. The ability to train, verify and optimize models on sheer volume of data substitutes for experience and the expected results can be relied upon with a measurable level of confidence.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.3. Automated Solutions (AI)

There are many AI techniques beyond ML. These techniques can be individually used on the same reliable data to draw different conclusions. The reliability of each of these techniques can also be assigned a numerical probability; correlations between techniques can be automatically verified. This analysis can be used to further increase confidence in proposed solutions. This algorithmic approach leverages unbiased observed reality and removes subjective judgement. While mistakes can be made, the consequences of possible mistakes can also be assessed – humans in the loop can make critical decisions based on objective numerical analysis.

4.4. Improved Remediation (IoT)

Better information (more accurate, timelier, and in greater volume) leads to improved remediation opportunities. Better AI allows us to clearly see the various remediation solutions and objectively evaluate them.

4.4.1. Fix Problems via Actuators

With innovative actuators in the network (*which can be an IoT device or code that can make configuration changes*), solutions can sometimes be automatically and immediately applied. One simple example is signal amplification. It is relatively simple to design amplifiers that can be remotely adjusted. Similarly, wave frequency can be adjusted remotely to affect the efficiency of the network or compensate for cable defects. If there is redundant equipment in the network, new equipment can be switched on to resolve some problems and faulty equipment simply switched off until it can be replaced.

4.4.2. Precisely Locate Affected Equipment and Cabling

Frequently, understanding the location of the problem is as important as understanding the problem itself. Signal reflection analysis and other techniques can be used to precisely find the location of a cable cut or kink. Network mapping and analysis can precisely locate a faulty piece of networking equipment.

4.4.3. Reroute Traffic Around the Problem

Sometimes, a truck roll is required to fix the problem. However, the "fix" is often not immediately necessary. It may be acceptable to temporarily avoid the problem until a complete fix can be made. Customers are more likely to be satisfied with partially degraded service for a short time than with a completely disabled service, even if the outage is short.

4.4.4. Instantly Update Configuration and Inform Affected Customers When Network Restored

When problems occur, one of the biggest frustrations for customers is the lack of information. Understanding the likely duration of the problem, which areas are affected and timely notifications when service is restored are all critical in customer satisfaction. Customers are often more forgiving when you keep them informed. IoT and AI can be valuable tools here too. Automated calls or messages to customers when service is restored can reduce the time to inform customers and avoid the time and trouble it would take for people to make those contacts.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.4.5. Anticipation

While it is good to solve observed network problems quickly and correctly, it would be even better to anticipate problems. IoT can provide valuable information for planning the network. Signal levels can be observed in real time. Equipment loads can be identified. AI can be used to suggest where problems are likely to occur. Redundant equipment can be installed or located strategically. Technicians can be on call when weather is predicted to be a problem. Good managers already do this, but if real-time sensors are used, managers can have better information about potential threats and be more strategic about preparations.

4.4.6. Avoidance

The corollary is that while responding quickly and accurately to network problems is great, avoiding problems is generally a better customer solution and cheaper for the operator. IoT and AI are critical components required for achieving this goal. A constant stream of data from IoT sensors provides a real-time view of the network. AI can monitor this data and identify early warning signs of potential problems. Then these worrisome signs can be addressed as regularly scheduled network maintenance during regular hours. The problems can potentially be avoided. Furthermore, this maintenance can be publicized to customers and used as a tool to retain existing customers or attract new ones.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. What Do Telco Technology Leaders Need To Do To Start Their AI/IoT Transformation Journey?

As we have discussed so far, the fusion of Artificial Intelligence and the Internet of Things can help Cable Operators optimize their networks by automatically adjusting network settings and configurations to improve performance and reduce costs. AI algorithms can further be used to analyze vast amounts of data generated by telecommunication networks, providing valuable insights into network performance, and helping to identify and resolve issues in real-time. This can significantly improve network reliability and reduce downtime, ultimately leading to enhanced customer satisfaction. All of this is easier said than done. In this section we will discuss as a Cable Operator AI leader some of the lessons learned, best practices that must be followed and the challenges / limitations that will be faced. We will also discuss the collaborative governance process that will help us define the new ways of working with using AI/IoT as part of our existing process.

5.1. Lessons Learned

In recent years, several Cable Operators and telcos have invested in AI and IoT. Most have launched limited proof-of-concept experiments to address pockets of problems and broadly these experiments or proofs-of-concept are done in silos. The toughest challenge most operators face is scaling the initial trials to create operational visibility, improve network reliability and generate financial impact. Here are some of the key lessons learned for technology leaders in the telecommunications industry.

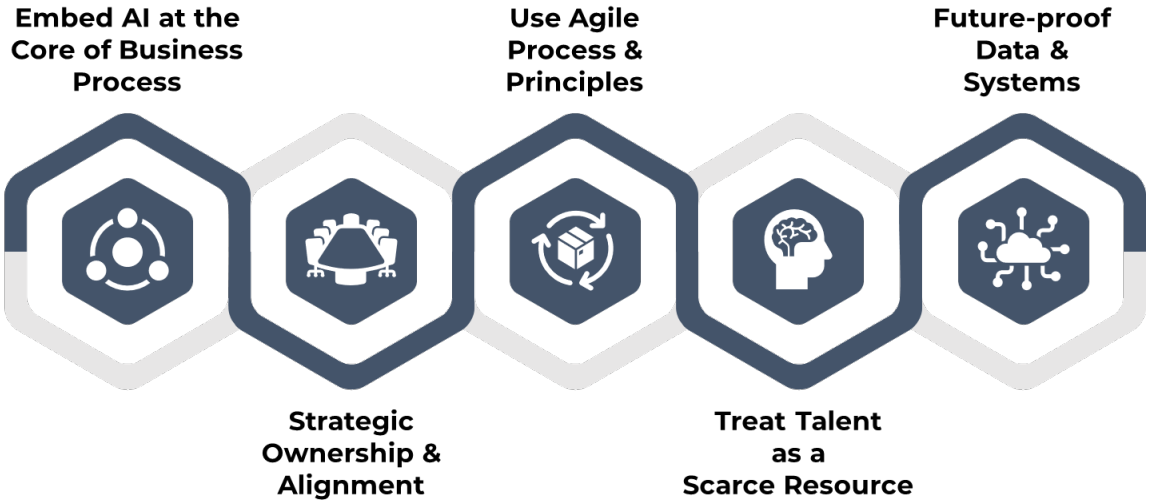


Figure 4: Key Lessons Learned

5.1.1. Embed AI at the Core of Business Processes

The early winners in the telecom industry have reinvented themselves by embedding AI at the very heart of not just their products, but their key processes. Cable operators must first identify priorities in consultation with business visions and goals to determine where AI will create the highest value. They should roll out use cases end-to-end in those priority areas so that they can optimize their processes and provide more value to customers. One huge benefit of embedding AI as part of the core business process is the training & upkeeping of the AI model i.e., the model can be quickly adapted to real life situations and

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

complexity changes. Operators who have tried approaching AI as a niche process and introduced separate implementation programs have seldom been successful.

5.1.2. Strategic Ownership and Alignment

Using AI at scale requires a new employee mindset and culture; it is important to prepare people for change. To do that, getting the C-Suite to own the AI deployment process is critical. One way of ensuring AI ownership at the top is to make the company's executive committee members accountable for delivering outcomes and value throughout the AI journey.

5.1.3. Use Agile Process and Principles

Agile methods of working are a catalyst, if not a prerequisite, for Cable Operators to unlock AI's power. Agile allows the adoption of a flexible and value-driven approach and enables working across functions – capabilities that are necessary to scale AI. Cable Operators must build cross-functional, agile teams focused on tangible AI-driven products or processes. They can scale up/down the teams as they do AI usage to ensure that they have sufficient resources for both along the way thus catalyzing the rapid adoption of AI in the organization.

5.1.4. Treat Talent as a Scarce Resource

Cable Operators must attempt to attract and retain digital talent in competition with the tech sector globally. To become more attractive workplaces, some Operators have rewritten their employee value propositions, revised their HR and promotion policies, and made work more flexible. Additionally, Operators must invest heavily in communications, skill building, and on-the-job training for employees and managers. Operators can also look at as AI for leveling up its employees, performing the function of an assistant by offering advice and guidance which will improve the effectiveness and performance.

5.1.5. Future-Proof Data and Systems

Operators must ensure that their data and technology assets are ready for AI. That requires two elements: Establishing a data governance structure that will ensure the data is usable, consistent, and valuable; and, revamping the IT architecture to make it more flexible and robust. That will enable AI-driven Operators to break down the siloed complexity of legacy IT systems and scale their AI applications.

5.2. Technology Best Practices

Legacy operational practices are not fit for purpose in modern cloud-native networks, whose increased complexity means that network engineers will be unable to have full knowledge of all the network stacks. For network operators to overcome these challenges, we present below a list of key AI technology-related best practices.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

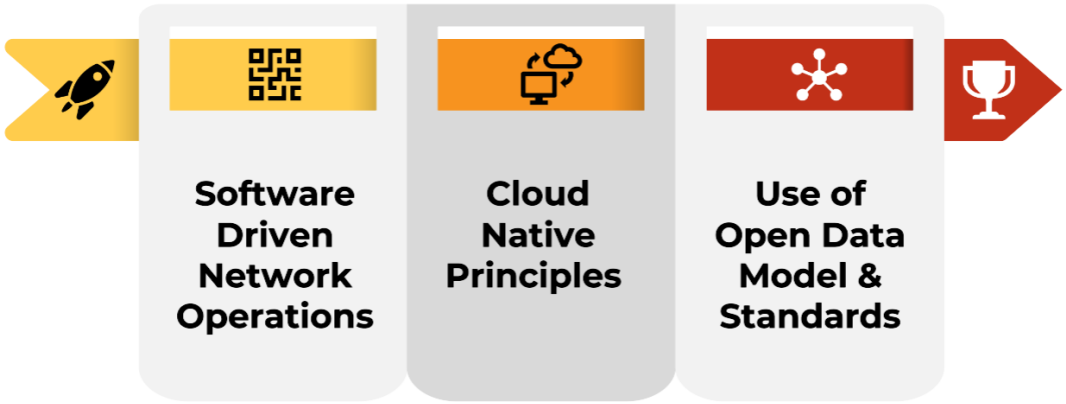


Figure 5: Key Technology Best Practices

5.2.1. Software Driven Network Operations

Software-driven network (SDN) operations must be employed, where intent-based automation replaces the current manual approaches in managing networks. The application of Artificial Intelligence to operations (AIOps) is the key to making installation, deployment, and operation part of an automated lifecycle management. The use of AI and IoT will be essential for analyzing network data, identifying network issues, and optimizing network performance. It will help take preventive action and fix network issues before they cause service disruption or downtime.

5.2.2. Cloud Native Principles

Cloud-native principles and technology have proven to be an effective accelerator in building and continuously operating the largest networks in the world. Using 12-Factor design principles [1], many telcos have realized operational effectiveness and business value with portability and scalability, simplified hand-offs between network engineering and operations and seamless continuous deployments.

5.2.3. Use of Open Data Model and Standards

Another prerequisite for the successful transformation of telco operations is the adoption of a common, open data model for network data based on industry standards, building open integration points for more uniform automation across the network stack.

5.3. Challenges and limitations

While the integration of Artificial Intelligence in Internet of Things offers numerous advantages, it also presents certain challenges and limitations. It is important to understand and address these issues to ensure the successful deployment and utilization of AI in IoT systems.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

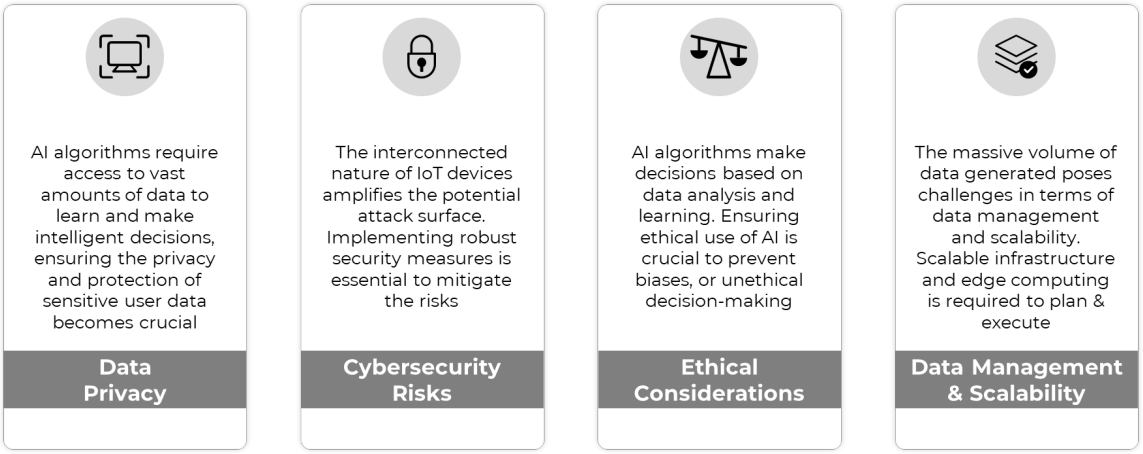


Figure 6: Key Challenges & Limitations

5.3.1. Data Privacy

AI algorithms require access to vast amounts of data to learn and make intelligent decisions. However, ensuring the privacy and protection of sensitive user data becomes crucial. Organizations must implement robust data encryption, secure data transmission protocols, and stringent access control mechanisms to safeguard user information and prevent unauthorized access.

5.3.2. Cybersecurity Risks

The interconnected nature of IoT devices increase the potential attack surface for cybercriminals and amplifies the risks of breaches. AI-enabled IoT systems can become targets for malicious activities, such as data breaches, unauthorized access, or manipulation of critical operations. Implementing robust security measures, including intrusion detection systems, encryption, and regular security updates, is essential to mitigate these risks, minimize the area of attack, minimize data exposure and exfiltration.

5.3.3. Ethical Considerations

AI algorithms in IoT devices make decisions based on data analysis and learning. However, ensuring ethical use of AI becomes crucial to prevent biases, discrimination, or unethical decision-making. Organizations must adhere to ethical guidelines, fairness principles, and transparent AI practices to avoid unintended consequences and maintain trust among users. Additionally, aspects of human-machine interactions, poses some key challenges –

- **Transparency and Explainability** - AI algorithms can be complex and difficult to interpret, resulting in challenges in understanding how an individual result was arrived at. A 1979 presentation slide attributed to IBM said, "computers can never be held accountable... therefore a computer must never make a management decision." Ensuring transparency and explainability of AI-driven decisions in IoT systems is crucial for user trust and accountability. Organizations must strive to develop AI models that provide clear explanations for their decisions.
- **Human-machine collaboration** - As AI becomes more integrated into IoT systems, striking the right balance between human control and AI autonomy becomes essential. Organizations must design interfaces and interactions that facilitate effective collaboration

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

between humans and AI-powered IoT devices. This involves understanding user needs, preferences, and the ability to override or intervene when necessary.

5.3.4. Data Management and Scalability

The massive volume of data generated by IoT devices poses challenges in terms of data management and scalability. Consider the following challenges:

- **Data storage and processing** - AI algorithms require substantial computational power and storage capacity to process and analyze IoT-generated data. As the number of connected devices increases, managing the sheer volume of data becomes a daunting task. Organizations must invest in scalable infrastructure and efficient data storage solutions to handle the ever-growing data streams.
- **Bandwidth and network limitations** - Transferring large volumes of IoT data to the cloud for AI processing can strain network bandwidth and lead to latency issues; in some situations, this can aggravate strained network conditions into cascading failures. This becomes particularly challenging in scenarios where real-time decision-making is required. Edge computing, where AI computations are performed closer to the data source, can help alleviate bandwidth constraints and reduce latency.
- **Integration with legacy systems** - Integrating AI capabilities into existing IoT systems or legacy infrastructure can be complex. Legacy systems may lack the necessary compatibility or processing power to handle AI algorithms effectively. Organizations must carefully plan and execute integration strategies, ensuring seamless interoperability between AI-driven IoT systems and legacy infrastructure.

5.4. Collabrative Governance Approach

Networks are the new utility – ubiquitous, pervasive, and critical. Like any other utility, customers expect it to be available whenever needed and to be operationally reliable. A utility that is performing as expected is essentially invisible. The goal should be zero down time and operational excellence. This high standard is possible with Next-Gen Operations Management. To operationalize this structure and successfully integrate AI and IoT into the business, it is important to consider current operations. A well-considered transition that minimizes current operational disruption would produce better, more sustainable results.

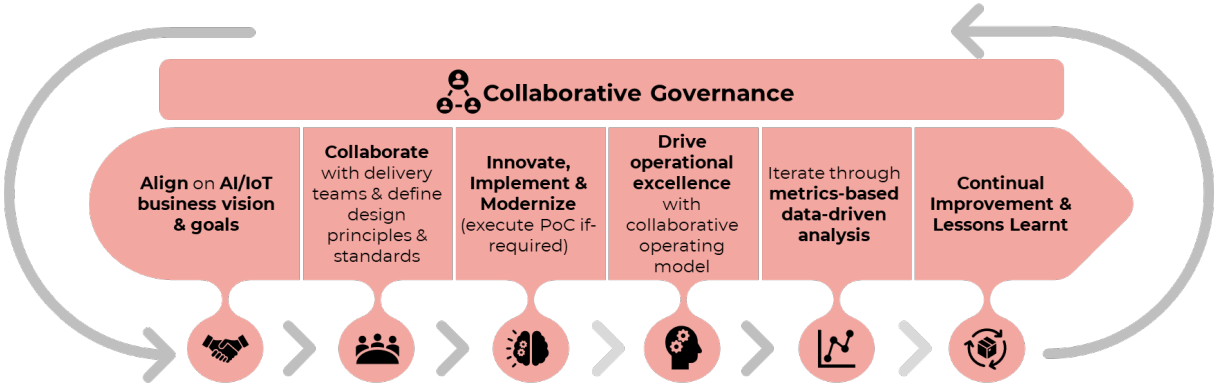


Figure 7: Next Gen Network Operations Management

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Introducing IoT and AI into operations management is a critical change that will achieve the goals identified above, but it is important that we do not abandon process and governance. IoT and AI are not replacements for these critical operational elements. Rather they should be leveraged to ensure that the efficiencies anticipated can be realized. Good process is important when deciding where IoT and AI can best be used. Good governance helps keep critical corporate goals in sight and ensures that progress toward those goals is measured. Governance also helps ensure that data is protected and used appropriately.

Communication and collaboration are essential. Adoption of a new technology cannot be accomplished without the buy-in and assistance of Network Operations (end-users, those who will be using the system). Operations must understand the motivation and must be consulted every step of the way. A proof-of-concept (PoC) trial may be advisable to verify the details of how the technology will be installed, integrate with existing process, how will the end-user experience look like and how will the success of AI/IoT be measured.

6. Conclusion

In this paper, we have attempted to make a case for migrating from a manually maintained network to a Next-Gen Operations Management approach based on IoT and AI. We believe such an approach improves reliability, reduces downtime, and lowers network maintenance costs. One of the unique opportunities that cable operators can use to incorporate IoT and AI as the people and process enabler in their network is DOCSIS 4.0 (D4.0). Most of the cable operators will need to rebuild their network to a certain degree to support D4.0, this provides a great opportunity to embed as much IoT and AI capability as possible as part of the network design. This decision needs to happen now as most of the cable operators are working with vendors to define the specifications of new generation network equipment. While this strategy requires additional capital investment to instrument the network and install the AI/ML platform, the anticipated benefits are enormous and something that cannot be underestimated.

The network strategy can migrate from a reactive response to customer-reported problems to an automated predictive system that can anticipate and prevent problems before they are noticed by customers. Between these two extremes, steps can be taken to simplify the transition. Initially, reactive responses can be migrated to planned improvements and proactive maintenance.

Taking the correct steps can be guided by an IoT platform that allows accurate and instantaneous measurement of key network parameters. Current knowledge of the network is key to an accurate and effective response. AI / ML techniques take this accurate data and provide automated, effective, and validated actions that can resolve network problems. Once the network is relatively stable, these same tools can be used to prioritize proactive network maintenance, then proactively identify network conditions that may predict future problems.

This automation allows a limited human work force to work more effectively and efficiently while being able to correctly resolve issues (a.k.a. first time right) that they may not have personally encountered in the past. All of this can be accomplished while lowering network maintenance costs over time.

IoT and AI can be combined to provide the most reliable network for our customers while reducing costs, time, and people necessary to solve problems. Additionally, the core IoT and AI platforms used to improve operational maintenance can be used to improve other internal processes or as new revenue generating services for our customers.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

3GPP	3rd Generation Partnership Project
AI	Artificial Intelligence
AIOps	Artificial Intelligence Operations
DOCSIS	Data Over Cable Service Interface Specification
FTTP	Fiber to the Premises
HR	Human Resources
IT	Information Technology
IoT	Internet of Things
MTTR	Mean Time to Repair
OpenRAN	Open Radio Access Network
SDN	Software Driven Network
SLA	Service Level Agreement

Bibliography & References

[1] https://12factor.net/

[2] *How AI is helping revolutionize telco service operations*, Jorge Amar, Tomás Lajous, Shreya Majumder, and Zachary Surak (https://www.mckinsey.com/)

Making maintenance smarter Predictive maintenance and the digital supply network, Chris Coleman, Satish Damodaran, Mahesh Chandramouli, Ed Deuel, A Deloitte series on digital manufacturing enterprises

AI and IoT: How do the internet of things and AI work together?, Tom Nolle, https://www.techtarget.com/

Exploring the dynamic fusion of AI and the IoT, Kerem Gülen, https://dataconomy.com/

Transforming Telcos with Artificial Intelligence, Sumit Banerjee, François Candelon, Rocío Lorenzo, Angelos Cacouros, and Soufiane Harguil, Boston Consulting Group

3 key AI use cases in telco networks, Matt Bamforth (https://stlpartners.com/)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

PNM Upstream Data Analysis

A Practical Solution for Automatically Locating Upstream Noise and Ingress

A Technical Paper prepared for SCTE by

Larry Wolcott
Comcast Fellow Engineer
Comcast Corporation
183 Inverness Dr W, Englewood, CO 80112
Larry_Wolcott@cable.comcast.com

Rob Gonsalves
Director 2 of Product Engineering
Comcast Corporation
183 Inverness Dr W, Englewood, CO 80112
Robert_Gonsalves@cable.comcast.com

Jonathan Leech
Principal Engineer 2, Software Engineer
Comcast Corporation
183 Inverness Dr W, Englewood, CO 80112
Jonathan_Leech@cable.comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Background of Upstream Noise on Coaxial Networks 4
2.1. The Noise Funnel 4
2.1.1. Additive Noise, multiple ingress points 6
2.2. Ingress vs. Egress – Reciprocity 6
2.3. Troubleshooting Process 6
2.3.1. Service Impact 7
2.3.2. Upstream Noise Mitigation 7
3. Spectrum Capture and Analysis 8
3.1. Downstream Frequency Spectrum and Diplex Filters 8
3.2. Upstream Frequency Spectrum 9
3.3. Attenuation 9
3.4. Port-to-port Isolation 10
4. UDA Theory of Operation 10
4.1. Why is this Important? 10
4.2. Detection Closest to the Point of Ingress 10
5. UDA Sensitivity and Performance 11
5.1. Setup 11
5.1. Measurements 12
5.2. Alternate Setup 13
5.3. Results 15
6. Analysis 16
6.1. UDA Detection 16
6.2. SC-QAM Detection 17
6.3. Noise Power Computation 17
6.4. UDA Issue Detection 17
6.5. Results 17
6.5.1. Issues Detected 17
6.5.2. Device Compatibility 17
6.5.3. Performance Parameters 18
6.5.4. Issues & Solutions 18
7. Field Examples 18
7.1. Common Mode Disturbance 19
7.2. Switching Regulator Noise 20
7.3. Citizens Band Radio 21
7.4. Spurious Noise 22
7.5. High-energy Narrow Band Interference 23
7.6. Unknown Interference 24
7.7. Shortwave Radio 24
8. Future Work 26
8.1. Additional analysis 26
8.2. FM detection and power level analysis 26
9. Conclusion 26

Abbreviations 27

Bibliography & References 27

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Figures

Title **Page Number**

Figure 1 – Broadcast Nature of the Downstream 4
Figure 2 – Upstream Operation 5
Figure 3 – Reverse Funneling 6
Figure 4 – Spectrum analysis test setup 11
Figure 5 – Spectrum analyzer max hold from 0 MHz to 120 MHz with energy being generated in 6 MHz to 91 MHz 12
Figure 6 – UDA Enabled capture (top) compared to non-UDA enabled capture (bottom) 12
Figure 7 – Direct comparison of UDA and non-UDA noise measurements between UDA (blue) and non-UDA (orange) spectrum capture analysis 13
Figure 8 – Minimum power detection levels for UDA (blue) and non-UDA (orange) spectrum capture analysis 13
Figure 9 – Alternate setup, using reversed 2-way splitter, injecting five tones with a signal generator (XM) 14
Figure 10 – Injected tones measured at the upstream burst receiver 15
Figure 11 – Injected tones measured using UDA at the cable modem (GW) attached to the 2-way splitter 15
Figure 12 – Common mode disturbance at burst receiver, around 24 MHz 19
Figure 13 – Same common mode disturbance at cable modem, around 24 MHz 19
Figure 14 – Switching regulator noise at the burst receiver, centered around 10 MHz 20
Figure 15 – Switching regulator noise, centered around 10 MHz 20
Figure 16 – CB Radio Ingress at the burst receiver, around 26.9 MHz 21
Figure 17 – CB Radio Ingress, around 26.9 MHz 21
Figure 18 – Powerline gap noise at the burst receiver, below 25 MHz 22
Figure 19 – Powerline gap noise, below 25 MHz 22
Figure 20 – High-energy narrow band interference at burst receiver, around 15 MHz 23
Figure 21 – High-energy narrow band interference, around 15 MHz 23
Figure 22 – Unknown interference between 35 MHz and 42 MHz at burst receiver 24
Figure 23 – Unknown interference between 35 MHz and 42 MHz 24
Figure 24 – Shortwave radio ingress, 9.33 MHz at the burst receiver 25
Figure 25 – Shortwave radio ingress, 9.33 MHz 25
Figure 26 – FM Ingress matching example 26

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Upstream noise and ingress can pose significant challenges for cable operators in maintaining coaxial cable networks. While Proactive Network Maintenance (PNM) offers numerous advantages for optimizing cable plant performance, it has historically lacked a solution for the reactive tracking of noise. Traditional upstream noise and ingress troubleshooting processes are primarily manual, labor-intensive, and potentially affect customers' service. Introducing the Upstream Data Analysis (UDA), a capability using PNM data that has been in development for a decade. The UDA technique is designed to automatically locate upstream ingress and noise sources. By utilizing the Data Over Cable Service Interface Specification (DOCSIS®) spectrum analyzers embedded in cable modems, the UDA samples the upstream frequency spectrum near the points of ingress, rather than at the receiver. This approach enables quick identification of individual premises and drop cables contributing to upstream noise, potentially resulting in operational efficiency improvements in network maintenance. The authors will provide an analysis of the PNM UDA's capabilities, limitations, and opportunities.

2. Background of Upstream Noise on Coaxial Networks

2.1. The Noise Funnel

The following three figures and text have been adapted from "Understanding and Troubleshooting Cable Upstream RF Spectrum" [3]. Figure 1 illustrates the broadcast nature of the downstream. That is, downstream signals (purple arrows) originating in the fiber optic node (node) are transmitted throughout the coaxial cable feeder portion of the network. In this diagram, the downstream radio frequency (RF) (circled purple) is being measured at the cable modem locations. Note that the measurements exclude the upstream portion (circled green) of the RF spectrum, being blocked by the modem's diplex filter.

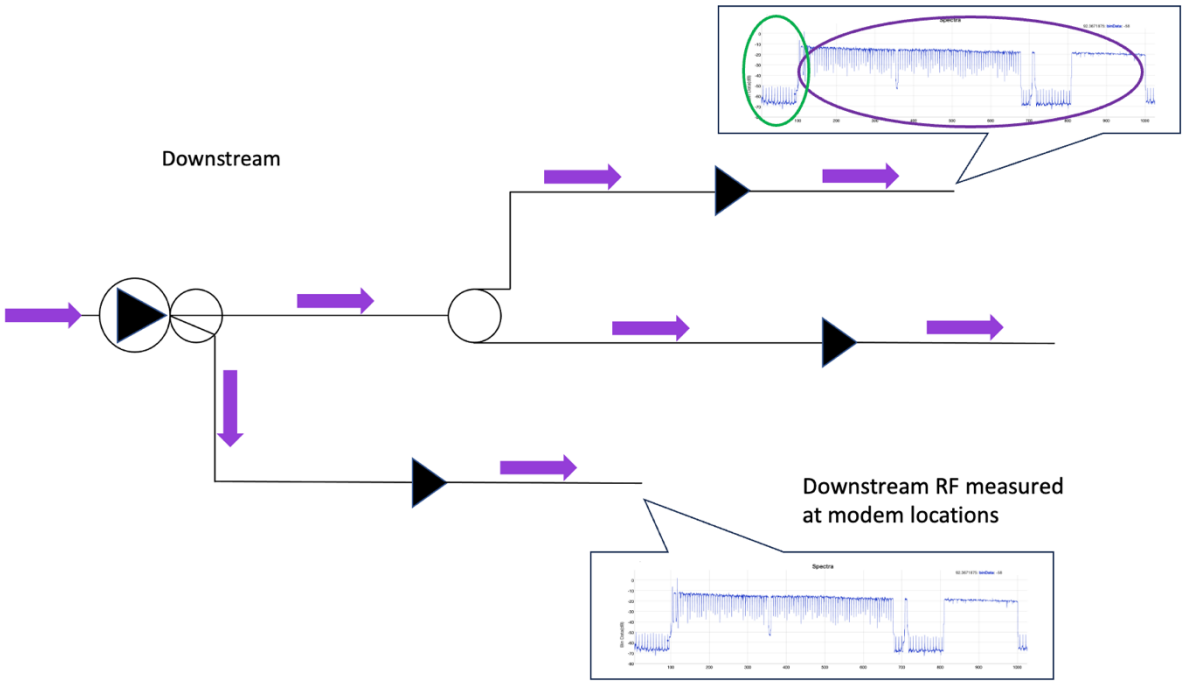


Figure 1 – Broadcast Nature of the Downstream

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 2 graphic illustrates upstream operation. Signals coming from subscribers in various parts of the node's service area (blue, green, and red arrows) travel upstream toward the node. Those signals can originate in cable modems, set-tops, status monitoring transponders, and other devices. (Note: In practice, some sort of time division multiplexing is typically used to prevent upstream signals on the same frequencies from interfering with one another. The intended signals operate with grants, using frequency division multiplexing and/or time division multiplexing, so that at any one time and frequency only one intended signal is present.) This figure illustrates the upstream measurements and analysis being done in the receiver. In this typical example, the upstream RF spectrum capture, SNR/MER and FEC, are all measured at the receiver.

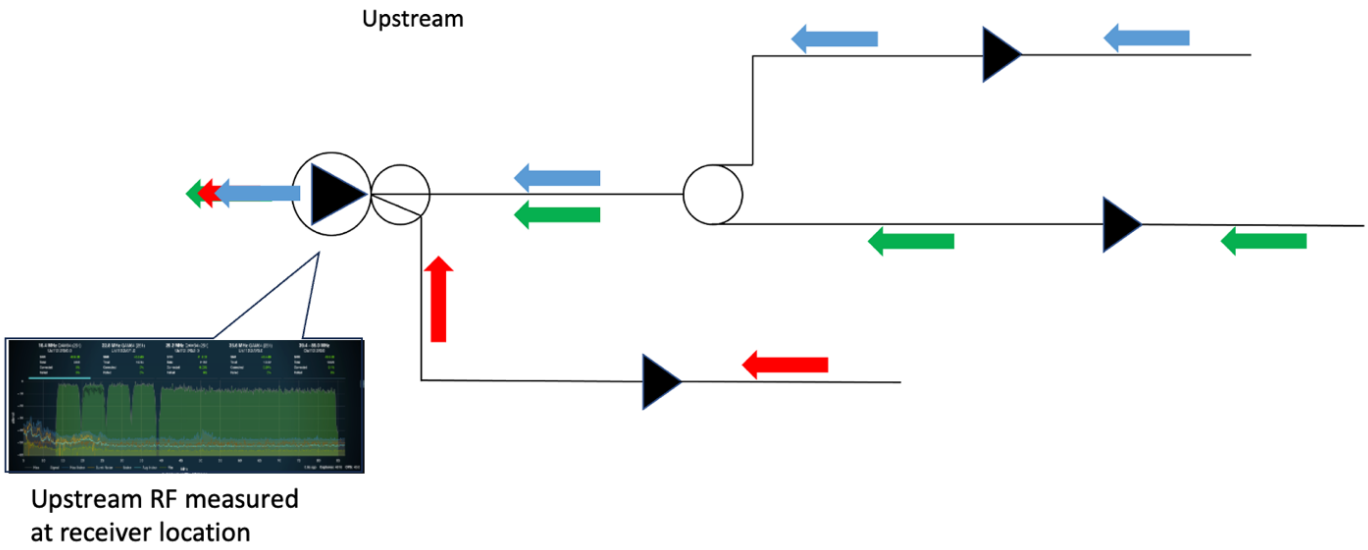


Figure 2 – Upstream Operation

In both the upstream and downstream directions there are intended signals, and undesired signals. Examples of undesired signals would be amplifier noise, ingress, or spurious emissions. Figure 3 illustrates interference of some kind (see the lightning bolt images in Figure 3) that is originating in the lower part of the node's service area. Note the interference is not present in other parts of the coax plant, just the feeder amplifier at the bottom of the graphic. In the upstream direction, all the undesired signals in the upstream frequency spectrum travel to the node – that is, they "funnel" toward the node. Once the interference (lightning bolt in the graphic) reaches the node it affects all signals coming from all parts of the node's service area.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

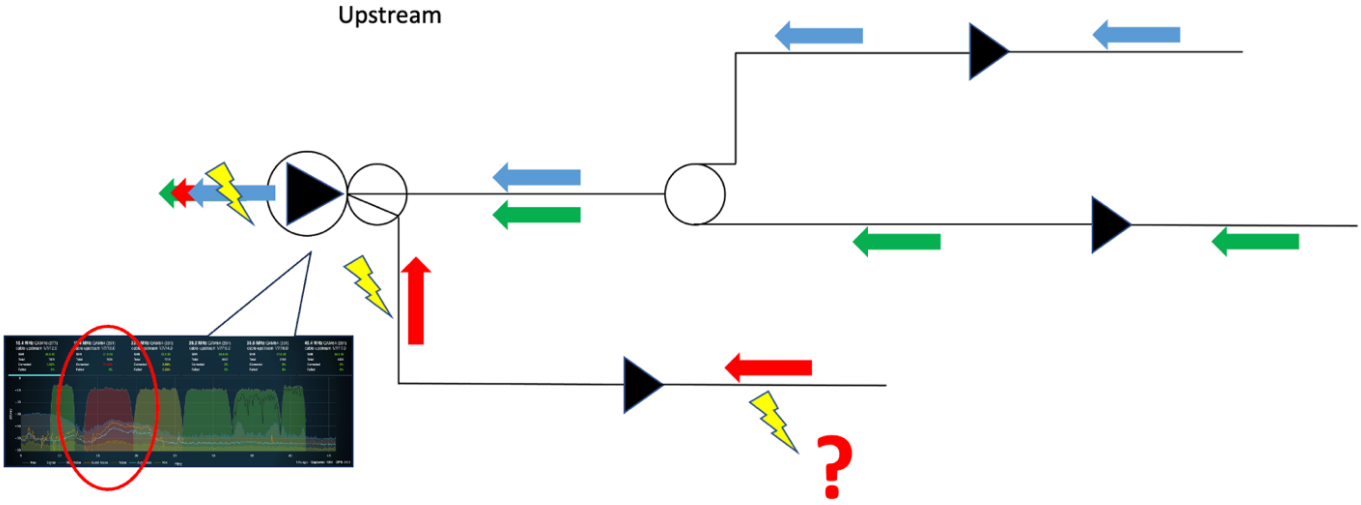


Figure 3 – Reverse Funneling

2.1.1. Additive Noise, multiple ingress points

The examples in Figure 1, Figure 2, and Figure 3 show a simple Node + 1 architecture. In other words, there is a maximum of one amplifier in cascade after the node. Calculating downstream carrier-to-noise ratio is simple, considering just one amplifier after the node. In the upstream direction, the CNR calculation must include the contribution of all return actives in the node's service area. In the graphics, there are three amplifiers after the node, so all three must be accounted for when calculating upstream CNR. This is, of course, related to the reverse funneling just discussed. (For more about calculating downstream and upstream CNR, see SCTE 270 2021r1 Mathematics of Cable.) In this figure, note that the upstream ingress is being detected and measured at the receiver. At this point of measurement, there is no way to discriminate the source of the ingress.

2.2. Ingress vs. Egress – Reciprocity

Antenna reciprocity is a fundamental principle of electromagnetics that asserts the transmission and reception characteristics of an antenna are identical. In the context of a coaxial cable network, this principle becomes crucial for understanding ingress and egress. Ingress refers to unwanted signals entering the network, which can cause significant noise and degrade network performance. On the other hand, egress, or signal leakage, refers to the unintended radiation of signals from the network. Antenna reciprocity helps to link these two concepts, as it implies that a point on the network that is vulnerable to ingress is equally likely to have signal leakage or egress. Thus, identifying ingress sources can also help determine potential areas of signal leakage and allow for appropriate remedial actions to ensure optimal network performance.

2.3. Troubleshooting Process

The process of tracking upstream noise in a coaxial cable network typically starts at the node or headend, where the technician begins by identifying and assessing the noise level in the upstream spectrum. This

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

can be done using specialized network diagnostic tools and monitoring software. Once an abnormal noise level is detected, the technician moves down the network, systematically isolating each segment.

Isolation typically involves disconnecting or disrupting the service for each segment one by one. After each disconnection, the technician will check the noise level at the node again. If the noise level decreases significantly or disappears, the disconnected segment is likely to be the source of the noise. If the noise level remains the same, the technician continues the process with the next segment.

This manual troubleshooting method can indeed be time-consuming and disruptive to the network services, especially if the network is extensive. It also requires significant technical expertise to accurately interpret the readings and correctly identify the noise source. Hence, newer methods that leverage automated diagnostic tools and advanced signal processing techniques can be adopted to increase the efficiency and accuracy of the process. These automated systems allow for continuous monitoring and can even predict potential noise issues based on historical and real-time data, possibly reducing the need for manual intervention.

2.3.1. Service Impact

When a coaxial network segment is isolated during troubleshooting, all DOCSIS connections on that segment are disrupted. This disruption has several impacts on the operation of cable modems.

One frequent problem is upstream T3 timeouts. In normal operation, a cable modem sends periodic "ranging" requests to the cable modem termination system (CMTS) to ensure optimal communication. However, when the network is interrupted, these requests cannot reach the CMTS, which results in a T3 timeout error. This signifies that the modem did not receive a response to its ranging request within a specified time.

Similarly, ranging response failures can occur. These happen when the cable modem sends a ranging request but either does not receive a response, or the response is incorrect or malformed. This is another indication that the modem cannot communicate properly with the CMTS.

Eventually, these problems lead to the cable modem re-initializing its connection. This process involves modem resetting and attempting to re-establish communication with the CMTS. This includes another round of ranging requests and responses, setting of upstream and downstream frequencies, and synchronization of time. It is a process that can be time-consuming and disruptive for end-users, particularly if it happens frequently.

In summary, interruptions to a coaxial network can have significant impacts on DOCSIS connections. This can lead to a range of problems, including upstream T3 timeouts and ranging response failures, to cable modems needing to re-initialize their connections. Such disruptions can negatively affect network performance and customer satisfaction, emphasizing the importance of efficient and accurate network troubleshooting.

2.3.2. Upstream Noise Mitigation

Mitigating noise ingress sources in a coaxial cable network is a crucial part of maintaining optimal network performance. Once the source of noise ingress has been localized, various techniques can be implemented, each having different implications to consider.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.3.2.1. Noise Filters or Traps

These high-pass filters are designed to block the lower frequency spectrum where noise typically resides, allowing only the higher-frequency upstream communications from cable modems to pass. However, the use of noise filters can limit the available spectrum for each subscriber, potentially leading to channel congestion on the reduced set of channels that are permitted to pass.

2.3.2.2. Attenuation or Pads

By reducing the power level of signals entering the network, these pads can lessen the amount of noise infiltrating the system. But they also reduce the power levels received at the modem, necessitating additional transmit power to ensure the signal reaches the intended receiver at the desired level. This approach can lead to insufficient signal power and reduced performance.

2.3.2.3. Disconnecting Noisy Connections

"Hot drops" are connections left by previous customers that, though no longer active, can still inject noise into the system. One straightforward mitigation technique is to completely disconnect these noisy connections. While this solves the immediate noise problem, it complicates the process when a customer wants to re-establish service, requiring a truck roll for reconnection.

2.3.2.4. Automated Profile Management Application (PMA)

This technique allows customers to operate with minimal upstream errors by dynamically adjusting the modulation profile of the upstream channels. While this increases the robustness of the connection, it reduces its capacity due to the lower modulation scheme. PMA is a balancing act between network stability and network capacity. If in time, that the maximum network capacity becomes reached, the noise will ultimately require mitigation to regain that capacity.

3. Spectrum Capture and Analysis

Full Band Capture (FBC) is an important feature used in DOCSIS PNM tools. It enables the cable modem to capture and digitize the entire spectrum available to its receiver, not just the DOCSIS data channels. This allows for a more comprehensive and in-depth analysis of network conditions, including noise and interference issues.

It works by taking advantage of built-in capabilities within many modern cable modems. These modems have an integrated tuner that can rapidly scan the entire downstream frequency range, digitize the signal, and then perform a fast Fourier transform (FFT) on the captured data. This process effectively converts the time-domain signal into the frequency domain, making it possible to analyze the signal's frequency spectrum and identify many impairments.

Once the digitized data has been captured and transformed, it is sent back to the network operator's PNM system. The system can then analyze the data in real-time to identify and diagnose a wide range of network issues, including ingress noise, micro-reflections, and frequency response issues.

3.1. Downstream Frequency Spectrum and Diplex Filters

FBC primarily operates on the downstream spectrum, due to the architecture of a DOCSIS system which can rely on diplex filters for frequency separation.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In a DOCSIS system, the downstream (from the cable operator to the user) and upstream (from the user to the cable operator) data transmissions occur over separate frequency bands. This separation allows simultaneous transmission and reception of data. The device responsible for managing this separation is the diplex filter.

A diplex filter is designed to direct the higher frequency downstream signals to the downstream path (receiver) and lower frequency upstream signals to the upstream path (transmitter). Consequently, the diplex filter effectively isolates the downstream and upstream paths from each other, ensuring that signals intended for one path do not interfere with the other.

Given this arrangement, FBC primarily works on the downstream spectrum because the downstream receiver is shielded from the upstream signals by the diplex filter. The receiver has a broader frequency range than it can capture, and this range contains all the downstream signals that the modem is designed to receive. Once captured, the modem then digitizes the received signals and analyzes them to monitor and assess the quality of the network's downstream spectrum. As a result, FBC provides a comprehensive view of the network's downstream health, which can be invaluable in troubleshooting and proactively managing the network.

3.2. Upstream Frequency Spectrum

While the upstream spectrum is traditionally blocked by the diplex filter from reaching the downstream receiver, advances by DOCSIS receiver chip and modem manufacturers have allowed for limited upstream spectrum analysis. These techniques enable an assessment of the upstream spectrum, providing valuable information from then end-of-line locations in the network and thus contributing to a more comprehensive analysis of the overall network health and localization of ingress sources.

There are diverse ways of implementing spectrum capture from frequency regions that are obscured by filters. These may include additional receivers and RF paths, software, timing, and other methods outside this document's scope.

3.3. Attenuation

Attenuation in coaxial cables refers to the reduction in signal strength as the signal propagates through the cable. This attenuation is influenced by multiple factors, including the frequency of the signal, the length of the cable, the type of cable used, and even the temperature. In general, higher frequencies suffer more attenuation than lower frequencies over the same distance.

As a result of this frequency-dependent attenuation, spectrum measurements taken at different points within the network can look quite different, even if the source of the signal is the same. For example, a high-frequency noise ingress signal that is strong at the point of entry into the network might be significantly weaker when measured further along the network due to the higher attenuation of high frequencies. Conversely, a low-frequency signal might appear consistent in strength across multiple measurement points.

Moreover, the effect of attenuation can also vary based on the specific location of the measurement relative to the point of ingress. If the measurement is taken closer to the point of ingress, the impact of the noise source might be higher due to less distance for attenuation to occur.

Therefore, understanding the impacts of attenuation is crucial for accurate network troubleshooting. Knowing how signal strength changes with frequency and distance can help network technicians to

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

pinpoint the source of noise ingress, determine the severity of the issue, and select the most appropriate mitigation strategy.

3.4. Port-to-port Isolation

Port-to-port isolation is an essential aspect of managing signal integrity in a coaxial RF network. Taps, which are used to distribute signals to individual customers, are specifically designed to have port-to-port isolation. This means they can reject noise or unwanted signals from traveling across input ports, limiting their potential to interfere with other parts of the network.

In the context of a coaxial RF network, when a signal, including any potential noise or ingress, arrives at a tap, it is split into different paths. Each path corresponds to a different output port that is connected to a customer's drop cable. The port-to-port isolation feature of the tap ensures that signals (and potential noise or ingress) on one output port do not interfere with the signals on the other output ports.

In addition to splitting the downstream signals, taps also combine upstream signals from multiple customers. These upstream signals are then sent back towards the node or headend. Here again, port-to-port isolation plays a key role. It allows each customer's upstream signals to be combined without being contaminated by ingress or noise from the other output ports.

Port-to-port isolation can affect measurements taken at various locations within a network. For example, if you measure the signal at a particular tap output port, the measurement will not be significantly affected by noise or ingress entering the network through the other output ports due to the isolation. This means that localized noise issues can be more accurately diagnosed and addressed without being masked or confounded by noise issues on other parts of the network.

So, port-to-port isolation is a crucial mechanism that enables UDA to localize noise sources by helping segment the measurements.

4. UDA Theory of Operation

4.1. Why is this Important?

As previously discussed in Section 2.1, the upstream noise funneling nature of our frequency-divided RF network can be challenging to troubleshoot and repair. Section 2.3 explains the operational practices required to maintain this return spectrum. Collectively, the upstream noise localization and mitigation efforts represent a sizable portion of network maintenance activities for cable operations. By improving the ability to locate these noise sources, cable operators can offer more reliable service, higher capacity, and faster speeds, while improving the operational efficiency for our network maintenance technicians.

4.2. Detection Closest to the Point of Ingress

The DOCSIS PNM specification for FBC does not limit the analysis of upstream RF spectrum from the cable modem’s measurements. However, the presence of the diplex filter in some modem designs precludes the upstream spectrum from being measured using the downstream receiver. This of course, is intentional to protect the sensitive downstream receivers from being exposed to the upstream transmitter’s burst energy. Given the low power of the cable downstream RF energy levels, this most certainly would create a problem in some modem designs. However, as described in Section 3.1, there are some cable modem designs which have been implemented in ways that facilitate the measurement and cable modem’s reporting of the upstream spectrum.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Having the benefit of port-to-port isolation and attenuation properties of the coaxial cable system, these factors can work together as a solution to localize upstream ingress in certain conditions.

5. UDA Sensitivity and Performance

To better understand the potential benefits and limitations of using UDA for localizing upstream noise sources, we must first characterize the sensitivity and performance of the UDA implementations.

5.1. Setup

Figure 4 shows the RF signal path including source (outlet), splitters (including a 2-way and a 4-way splitter), two spectrum analyzers and full band capture compliant cable modems.

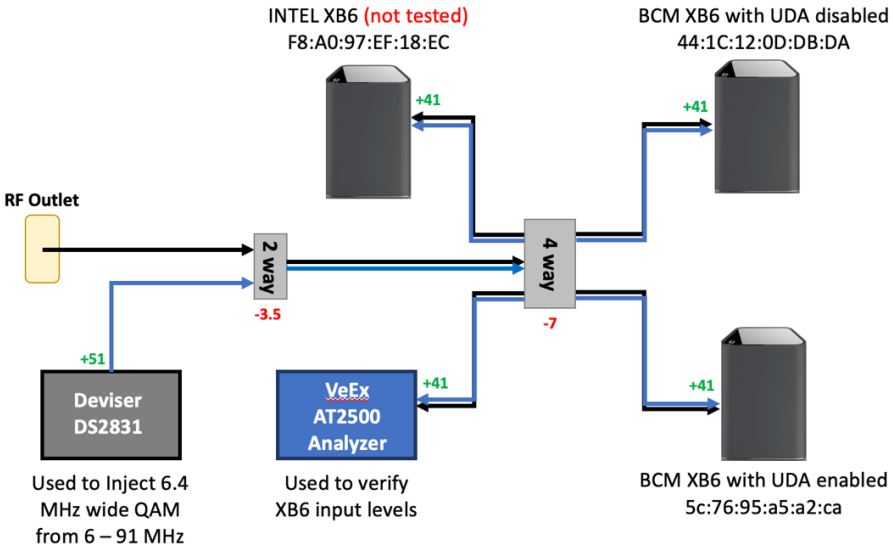


Figure 4 – Spectrum analysis test setup

A flat, 6.4 MHz SC-QAM (single carrier quadrature amplitude modulation) input signal with an RF level of 41 dBmV was achieved at the input of each spectrum analyzer and each XB6 cable modem. The RF spectrum was tested over a frequency range of 6MHz – 91MHz. Figure 5 shows the max hold signal level over the course of the entire test to ensure the input to each device was flat across the whole frequency range tested.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

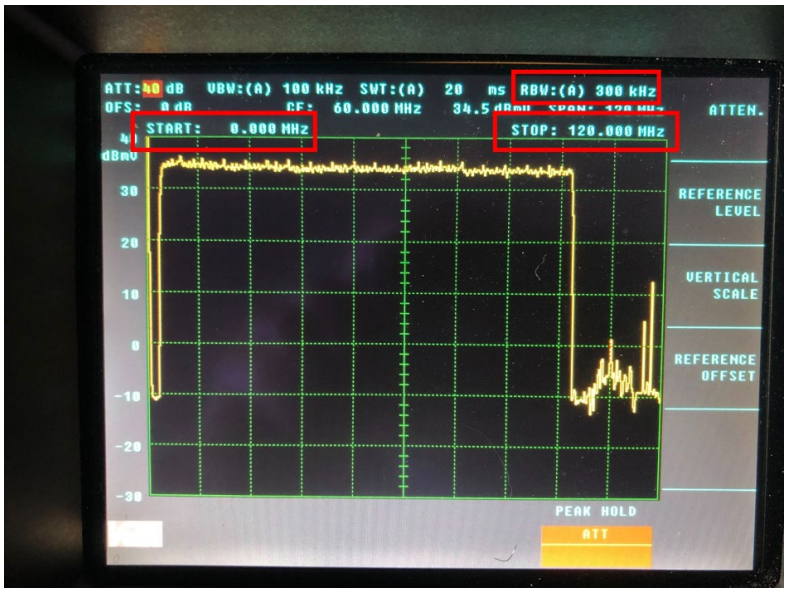


Figure 5 – Spectrum analyzer max hold from 0 MHz to 120 MHz with energy being generated in 6 MHz to 91 MHz

5.1. Measurements

In Figure 6 the UDA enabled devices show 15 – 38 dB more sensitivity, depending on frequency, compared to the same devices with UDA disabled. After the 42 MHz diplex filter cutoff both devices show the same level of sensitivity.

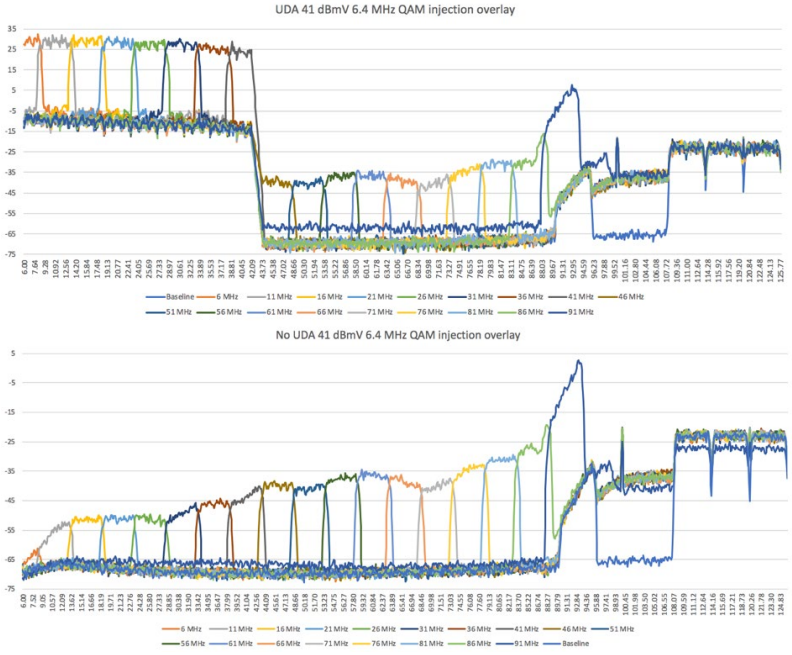


Figure 6 – UDA Enabled capture (top) compared to non-UDA enabled capture (bottom)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 7 and Figure 8 show the minimum signal power levels required to be visible above the noise floor on a UDA and non-UDA device.

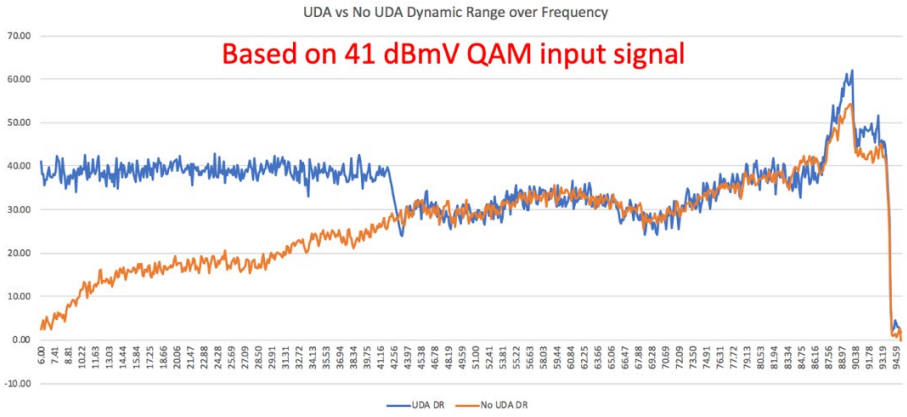


Figure 7 – Direct comparison of UDA and non-UDA noise measurements between UDA (blue) and non-UDA (orange) spectrum capture analysis

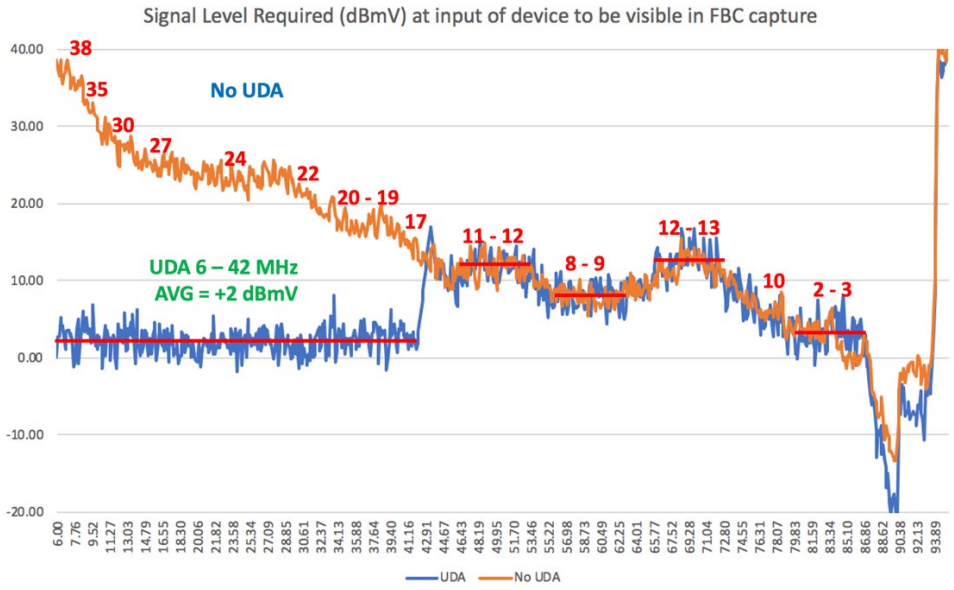


Figure 8 – Minimum power detection levels for UDA (blue) and non-UDA (orange) spectrum capture analysis

5.2. Alternate Setup

To further validate sensitivity on alternate types of equipment and later versions of software, a simplified test was performed using a field-deployed cable modem. In this test, a 2-way splitter was reversed so the spectrum capture sensitivity could be evaluated while circumventing the isolation properties of the two output ports (Figure 9). Five tones were injected at the output port of the 2-way splitter and measured at

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the upstream burst receiver (Figure 10). Figure 11 shows the bins captured at the UDA-enabled gateway (GW) attached to the input port of the 2-way splitter. A signal level meter (XM) as used to create and inject five tones in the rolloff and guard bands of four upstream SC-QAM channels, The five tones are clearly detected with exceedingly high sensitivity, improved over the initial setup using older cable modem hardware and software.

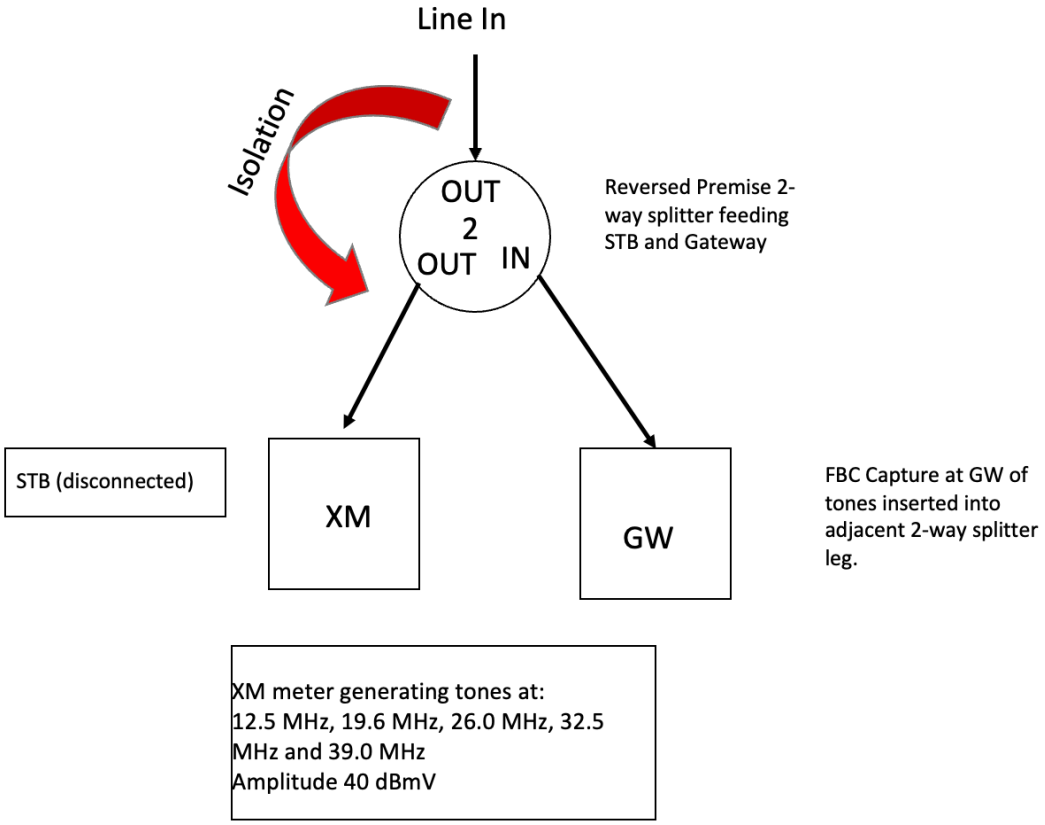


Figure 9 – Alternate setup, using reversed 2-way splitter, injecting five tones with a signal generator (XM)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 10 – Injected tones measured at the upstream burst receiver

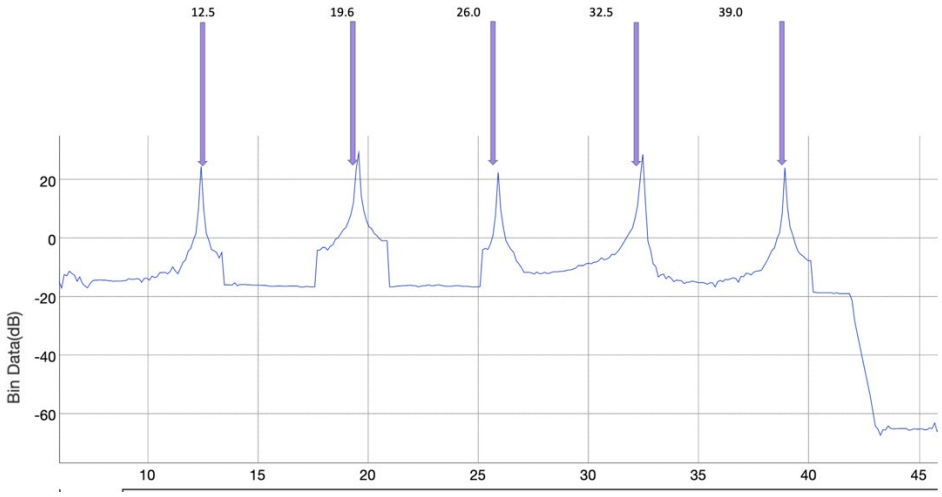


Figure 11 – Injected tones measured using UDA at the cable modem (GW) attached to the 2-way splitter

5.3. Results

Based on the outcome of the test setup and measurements, the following points can be made.

- UDA enabled cable modems offer 15 dB to 38 dB better noise sensitivity in the 5 MHz to 42 MHz band compared to non-UDA devices.
    - Same models tested (Device 1 with UDA enabled and Device 2 with UDA disabled, no other models tested)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- UDA and non-UDA devices have similar sensitivity after diplex filter cutoff frequency.
- If upstream noise originates from the drop side feed or the same outlet as the UDA device and is sufficiently strong to affect the return performance at the node, it should be observable in the FBC spectrum specific to that device with UDA.
- Noise beyond the diplex cutoff frequency will need to be 8 dB to 13 dB higher to be visible due to added filter rejection in that frequency range.
- If noise is on another RF splitter leg and subject to port-to-port isolation, visibility will be drastically reduced (additional testing required to characterize further).
- Attenuation between the point of ingress and point of measurement will also be a determining factor in the usefulness of UDA to detect an impairment.

6. Analysis

Having a better understanding of the UDA sensitivity and performance provides important context for the following analysis.

A statistically significant number of cable modems were analyzed. More than 25% of the sample population of FBC capable modems also supported UDA.

One problem the authors encountered while analyzing the upstream spectrum was the presence of energy produced within the OFDMA (orthogonal frequency division multiple access) channel area. For this reason, that portion of the spectrum as well as some adjacent spectrum, was ignored. Further work needs to be done to better understand this issue.

Another issue encountered was that SC-QAM and OFDMA bursts from nearby devices were often visible, including what appeared to be spurious bursts (unexplained energy in otherwise clean spectrum) was sometimes present. The amplitude of a burst from a nearby device can vary due to transmit power, distance, timing, and port-to-port isolation from the nearby device,

Finally, the spectrum captures are stitched together from multiple, smaller samples. Since the smaller samples are captured at distinct times from each other, this meant that the width of each stitch, approximately 4 MHz, was often the maximum width of the noise or burst that could be observed. Note that this is narrower than the typical width of an SC-QAM channel (6.4 MHz).

For the above reasons, the authors implemented QAM detection logic that was purposefully simplistic and aggressive, trading a high false negative rate for a low false positive rate. In other words, the detector would rather mis-classify noise bursts as QAM versus classifying QAM bursts as noise. Furthermore, the analysis only focused on noise bursts within noisy upstream channels with an average MER of less than 30 dB. Again, this is all to minimize the false positive rate.

6.1. UDA Detection

The first step of the analysis is UDA detection. Due to the much higher noise floor in UDA-capable modems, there is significantly higher power in UDA-capable modems, The UDA detection logic computes the average power in the range of interest, which is 6 MHz to 38 MHz, and compares it to a threshold of 6.363 dBmV. Total UDA power is computed as 10 * log10(sum linear values in UDA spectrum range). The total UDA power threshold (6.363) was computed as an average dBmV of >= -18 within the UDA range (6 - 38 MHz). The spectrum analyzer resolution bandwidth is 117.1875 kHz and 32 MHz of total observed spectrum, comprised of spectrum 273 bins.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Next, the UDA noise floor was computed using the 15th percentile of the values within the UDA spectrum range. The resulting value in dB varies between modems with the majority samples in the range of –14 dB to –18 dBmV. The computed UDA noise floor is used later in the analysis. Further refinement might include removing tilt from the UDA spectrum, or otherwise accounting for tilt in the analysis.

6.2. SC-QAM Detection

The next step of the analysis is SC-QAM detection. The SC-QAM detection logic looks for four or more consecutive samples above a threshold of more than 10 dB above the noise floor. Additionally, single values less than the threshold are permitted, so long as the next sample exceeds the threshold. The regions of spectrum containing SC-QAM bursts are replaced with the UDA noise floor. Further refinement of the SC-QAM detection logic might include detecting the guard bands and the expected lower energy therein.

6.3. Noise Power Computation

Only samples within the spectrum of noisy channels (defined above) are considered. The linear values are summed after the UDA noise floor is removed.

6.4. UDA Issue Detection

The UDA noise power was compared to three thresholds, 8 dB for MINOR, 10 dB for MAJOR, and 12 dB for SEVERE to establish severity.

6.5. Results

The goal was to evaluate the UDA detection method (Section 6.1) of upstream noise in a DOCSIS coaxial cable network. The analysis was conducted for a duration of one week, running UDA analysis hourly.

6.5.1. Issues Detected

a) The UDA analyzer detected between 30 and 200 issues per hour, many of which were repeat occurrences.
b) Over 50% of the UDA issues detected correlated with existing FM radio ingress detection, confirming a known indicator of potential ingress issues.
c) Notable interference was identified at 9.33 MHz, attributed to World's Last Chance Flat-Earthers.
d) Despite the vast range of narrowband ingress detected, the ingress canceler effectively filtered noisy channels, improving the demodulator's performance.

6.5.2. Device Compatibility

Over 25% of the modems with full-band capture support also featured UDA compatibility. Although many device models supported UDA, 90% of these were from the top four models. The two latest UDA firmware models represented approximately 33% of the UDA-compatible devices.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6.5.3. Performance Parameters

a) Spectrum Range: 6 MHz to 38 MHz
b) Upstream MER Threshold: 30 dB.
c) SC-QAM Power dB Threshold: 10 dBmV.
d) Minimum consecutive SC-QAM samples required: 4.
e) UDA total peak power threshold: 6.363 dBmV.
f) Noise criteria were segmented as: MINOR (>8 dB), MAJOR (>10 dB), and SEVERE (>12 dB).
g) For SC-QAM bursts adjacent modems, amplitude varied, whereas OFDMA bursts from the same modem equated to TX power.

6.5.4. Issues & Solutions

The SC-QAM detector, while effective for certain types of interference, was particularly adept at pinpointing narrowband ingress from sources such as citizen band (CB) radio and shortwave radio.

Spurious bursts, which are unexplained data within a clean spectrum, can arise from multiple sources including stitching errors, gain setting errors during data collection, and other potential bugs in modem UDA functionality.

Regarding stitching, issues arose when the sample thickness was less than the SC-QAM width, leading to partial SC-QAM bursts. The extent of these bursts was dependent on channel alignment with stitch points.

To minimize false positives, especially where SC-QAM bursts are misidentified as noise, the SC-QAM detector was kept deliberately simple. This might result in high false negatives where noise bursts could be misidentified as SC-QAM.

7. Field Examples

The following examples were selected from the population described in Section 6. Each of these examples include a spectrum capture from the upstream burst receiver and corresponding capture from the cable modem's location. These examples are intended to illustrate the usefulness for improving the localization of certain types of upstream noise.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

7.1. Common Mode Disturbance

Figure 12 and Figure 13 provide an example of a common mode disturbance (CMD) interference which appears as a noise hump around 24 MHz These are typically caused by ground faults in the presence of loose connectors on certain types of modems, set-top boxes, and other types of customer equipment. In this example, all modems are equally impacted by the impairment, however, the primary source of noise can be localized to a single customer location.



Figure 12 – Common mode disturbance at burst receiver, around 24 MHz

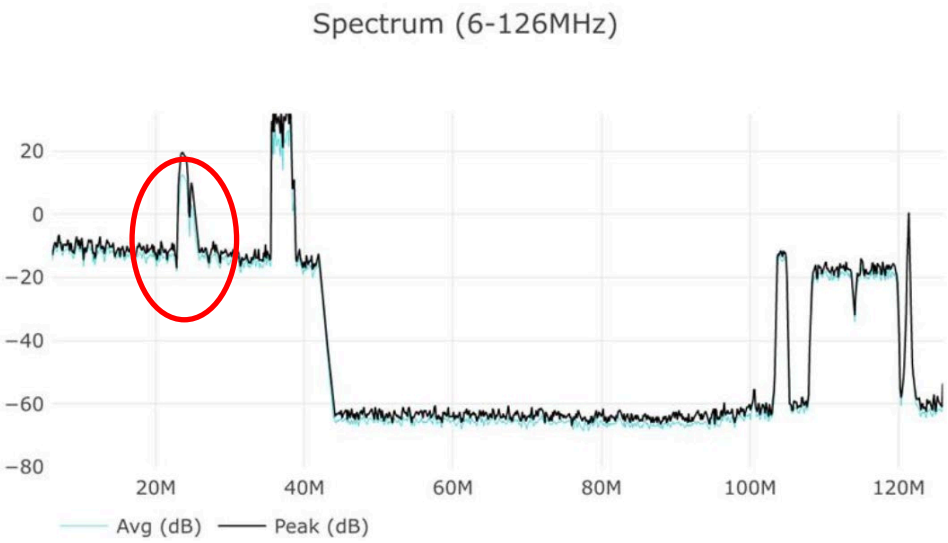


Figure 13 – Same common mode disturbance at cable modem, around 24 MHz

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

7.2. Switching Regulator Noise

Switching regulator noise is among the most observed types of noise in cable networks. Some switching (or switch-mode) power supplies in customer premises equipment, wall warts and other devices can generate interference that can affect cable network operation. The interference often appears as harmonics or spurious signals that are spaced at intervals of the power supply's switching rate – for instance, every 50 kHz. When switching regulator noise enters the network as ingress, it often appears in the upstream spectrum.

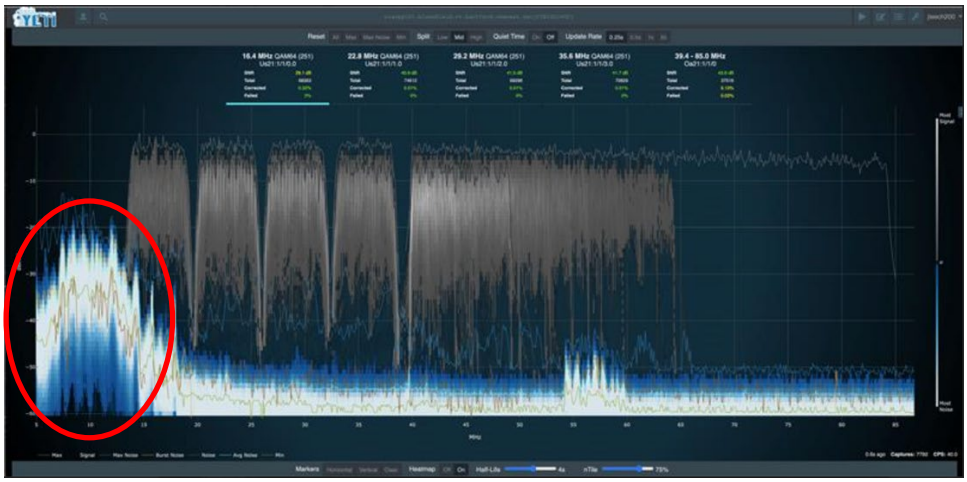


Figure 14 – Switching regulator noise at the burst receiver, centered around 10 MHz

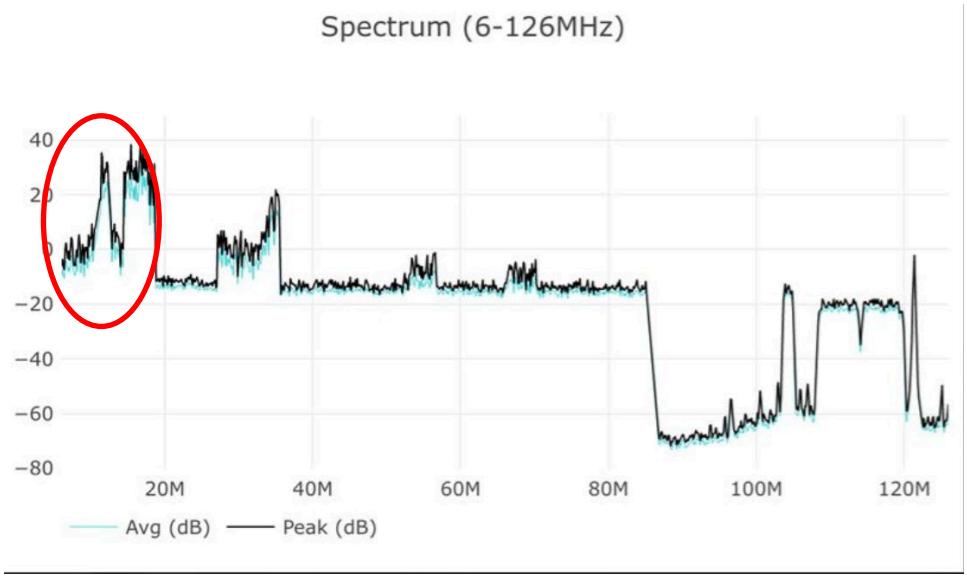


Figure 15 – Switching regulator noise, centered around 10 MHz

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

7.3. Citizens Band Radio

CB radio operates in the frequency range between 26.965 MHz and 27.405 MHz



Figure 16 – CB Radio Ingress at the burst receiver, around 26.9 MHz

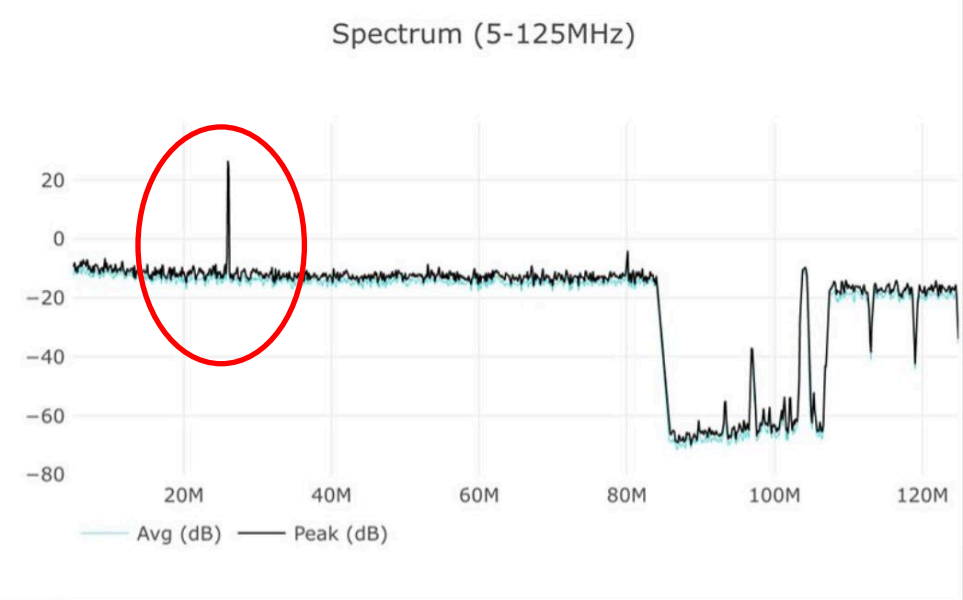


Figure 17 – CB Radio Ingress, around 26.9 MHz

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

7.4. Spurious Noise

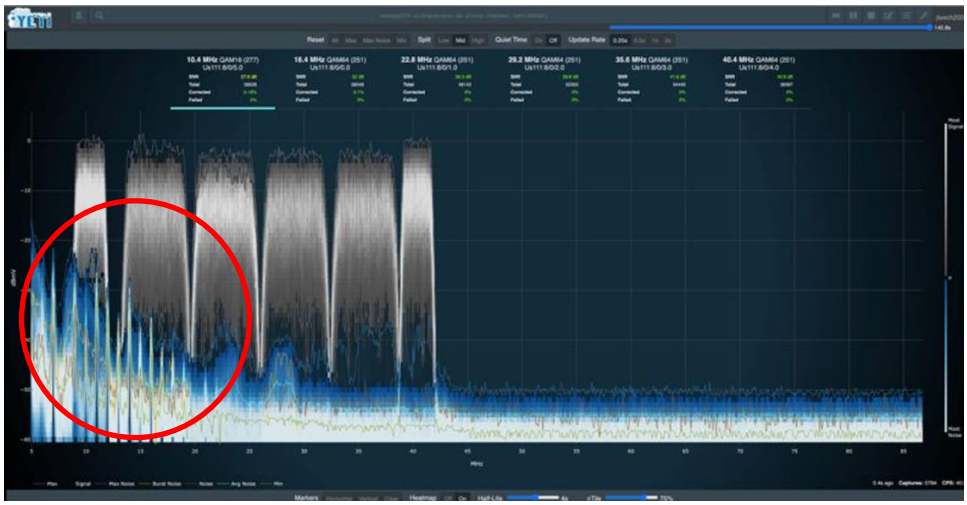


Figure 18 – Powerline gap noise at the burst receiver, below 25 MHz

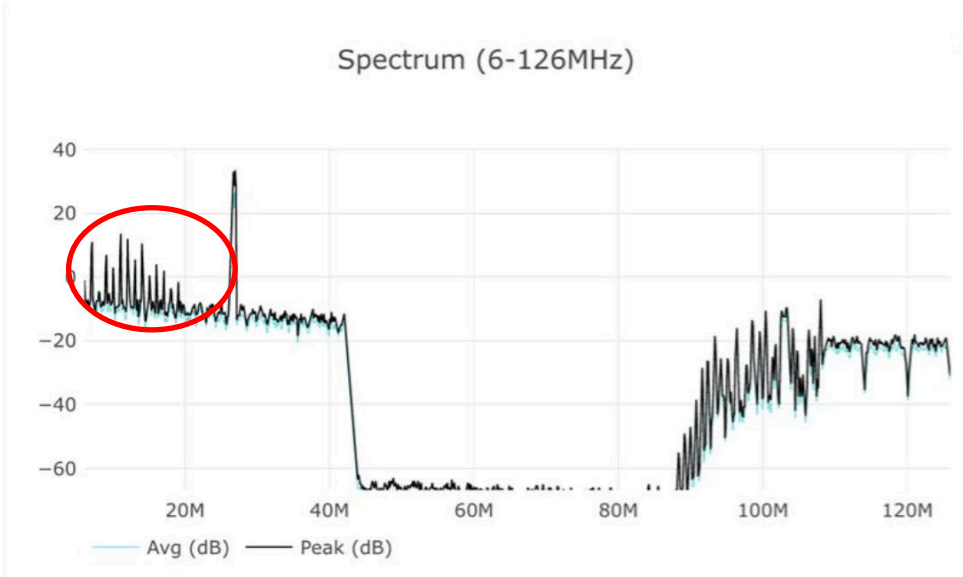


Figure 19 – Powerline gap noise, below 25 MHz

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

7.5. High-energy Narrow Band Interference



Figure 20 – High-energy narrow band interference at burst receiver, around 15 MHz

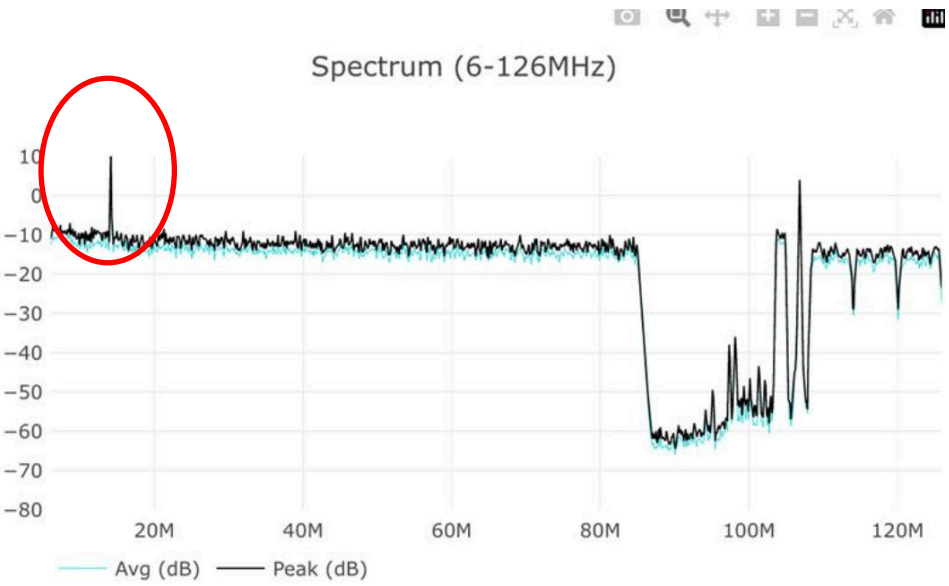


Figure 21 – High-energy narrow band interference, around 15 MHz

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

7.6. Unknown Interference

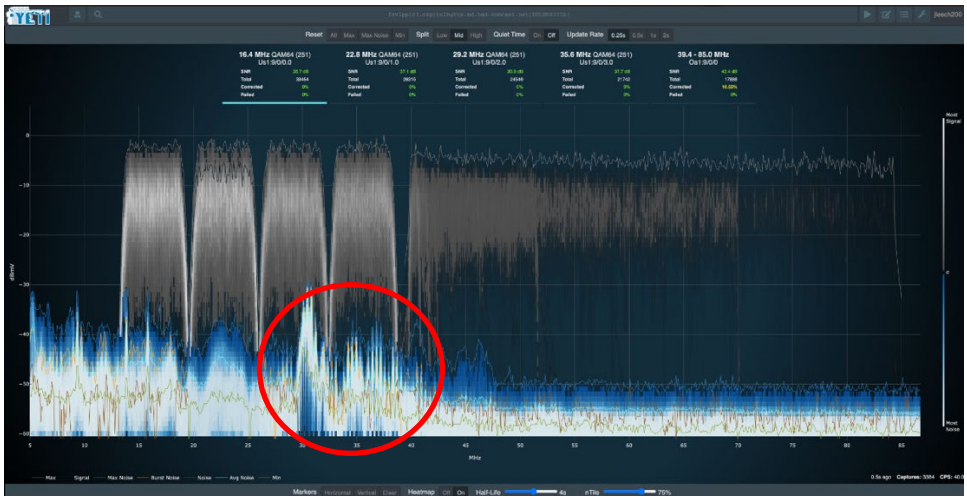


Figure 22 – Unknown interference between 35 MHz and 42 MHz at burst receiver

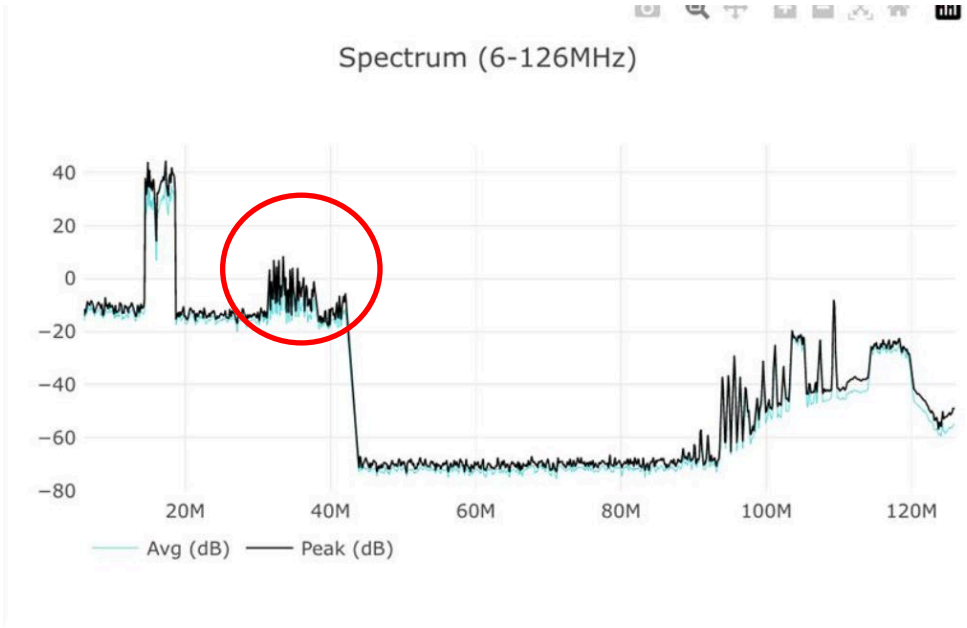


Figure 23 – Unknown interference between 35 MHz and 42 MHz

7.7. Shortwave Radio

Shortwave radio was detected frequently with varying severity. One interesting example is from a shortwave broadcast from a half-million-watt transmitter in Maine, which causes ingress issues throughout the Northeastern United States. The following figures shows the 9.330 MHz transmission.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

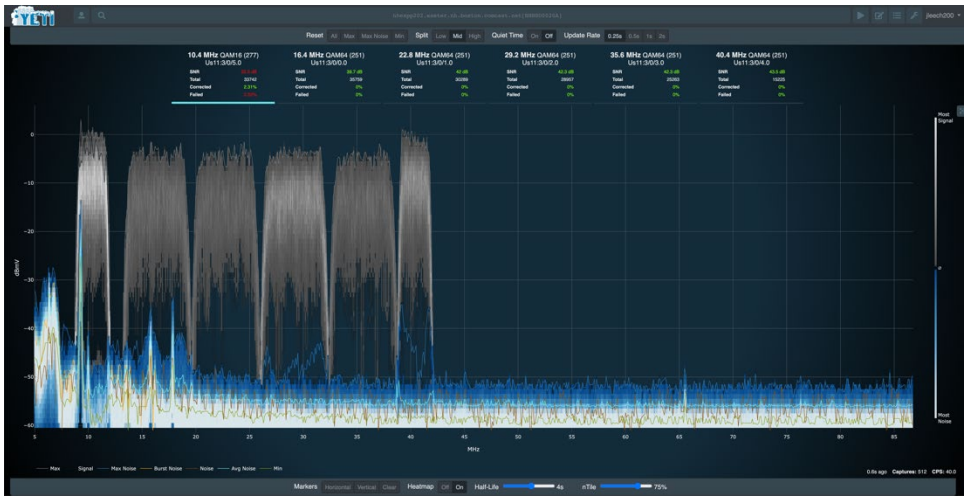


Figure 24 – Shortwave radio ingress, 9.33 MHz at the burst receiver

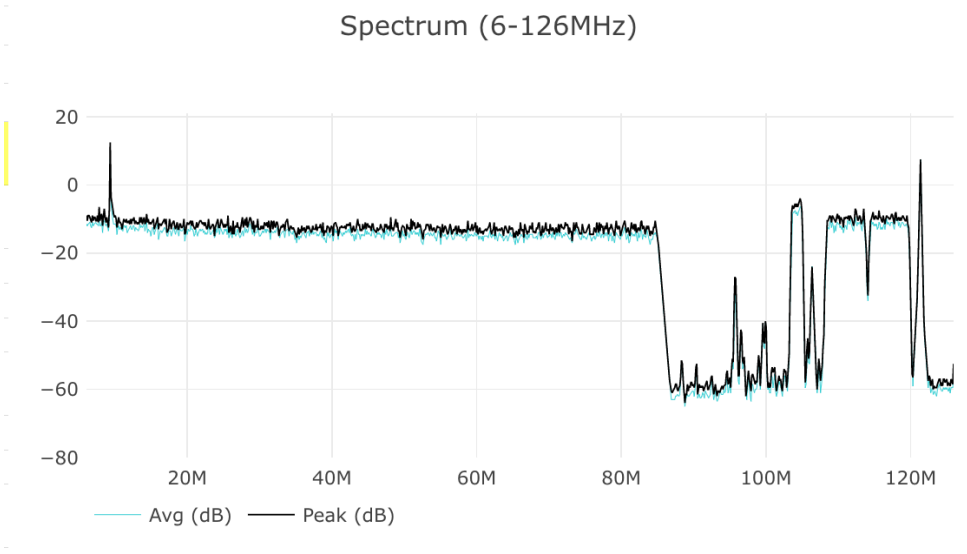


Figure 25 – Shortwave radio ingress, 9.33 MHz

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

8. Future Work

8.1. Additional analysis

With encouraging initial results, additional areas of interest have been identified.

- Further refine SC-QAM detection and find more noise with acceptable false positive rate.
- Expand the UDA spectrum range to include OFDMA frequency spectrum.
- Examine history, histogram and exponential decay of values.
- Evaluate raw poll values (e.g., not averaged, to remove vs exclude SC-QAM bursts).
- Correlate to upstream burst receiver noise captures.

8.2. FM detection and power level analysis

Given the high correlation (~50%) of FM ingress to other types of interference, FM radio ingress can provide valuable insight into the detection and location of other types of upstream noise. At the time of writing, the authors, along with SCTE NOS (Network Operations Subcommittee) WG 7 (PNM) are performing FM power level analysis useful in pursuing this goal.

Figure 26 shows an example of two different FM ingress problems. The example on the left illustrates two neighbors that have individual ingress problems, while the example on the right is typical of a common network problem.

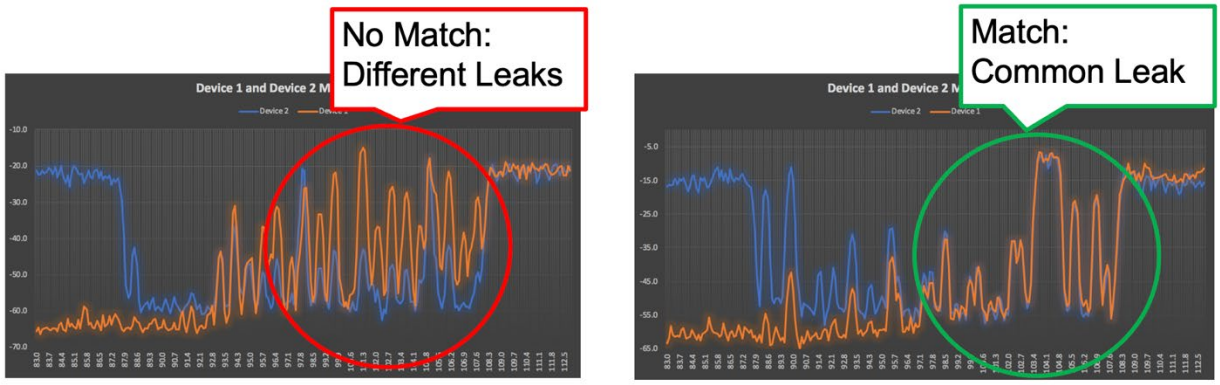


Figure 26 – FM Ingress matching example

9. Conclusion

This exercise has proven that localizing upstream noise with upstream spectrum analysis near the sources of ingress does work! However, there are several limiting factors which reduce the overall effectiveness, including:

- Port-to-port isolation can hide ingress even with the same house across splitters.
- Ingress from drop locations that do not have active equipment will be un-detectable.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Limited support within deployed cable modem population (25% at time of writing).

Most cable operators would immediately recognize the importance and value of improving the process of detecting, locating, and mitigating upstream noise. This should serve as encouragement that we, as an industry, can finally prevail in the relentless pursuit of this long-standing foe of reliable cable service.

Abbreviations

CB	citizens band radio
CM	cable modem
CMD	common mode disturbance
CMTS	cable modem termination system
CNR	carrier-to-noise ratio
CPE	customer premises equipment
DOCSIS	Data Over Cable Service Interface Specification
dBmV	decibels relative to one millivolt
FBC	full band capture
FEC	forward error correction
FFT	fast Fourier transform
FM	frequency modulation
GW	gateway
MER	modulation error ratio
MHz	megahertz
NOS	network operations subcommittee
OFDMA	orthogonal frequency-division multiple access
PMA	profile management application
PNM	proactive network maintenance
SC-QAM	single channel quadrature amplitude modulation
RF	radio frequency
SNR	signal-to-noise ratio
STB	set-top box
SCTE	Society of Cable Telecommunications Engineers
UDA	upstream data analysis
WG	working group
XM	Xfinity signal meter

Bibliography & References

[1] PNM Best Practices: HFC Networks (DOCSIS® 3.0) CM-GL-PNMP-V03-160725, July 25, 2016, Cable Television Laboratories, Inc.

[2] SCTE 280 2022 Understanding and Troubleshooting Cable RF Spectrum, 2022

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

[3] SCTE NOS WG 7, Understanding and Troubleshooting Cable Upstream RF Spectrum, Unreleased

[4] A Comprehensive Case Study of Proactive Network Maintenance, Heslip, Thomas, Gonsalves, Wolcott, SCTE CableTec Expo 2016

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Remote Diagnosis of Adjacent Channel Interference in a High-Split System

A Technical Paper Prepared for SCTE by

Lei Zhou
Principal Architect
Charter Communications
14810 Grasslands Drive, Englewood, CO USA
lei.zhou@charter.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
1.1. Mid-Split and High-Split DOCSIS Networks 4
1.2. Adjacent Channel Interference in Mid-Split and High-Split Systems 5
1.3. Diagnosis of ACI 5
2. Characteristics of High-Split ACI 5
2.1. Scenarios of High-Split ACI 5
2.1.1. In-Home ACI 6
2.1.2. Neighbor ACI 6
2.2. Impact of ACI 6
2.2.1. Automatic Gain Control Maladjustment 6
2.3. Numerical Evaluations of High-Split ACI Impact to STB 8
2.3.1. STB Model-A 9
2.3.2. STB Model-B 13
2.3.3. STB Model-C 15
3. Detection of ACI Using Telemetry 16
3.1. Useful MIBs 17
3.1.1. HS-CM Transmit Power MIBs 17
3.1.2. SS-CPE Receive Power and Tuner Status MIBs 17
3.2. Availability of Upstream Traffic Generator 18
4. Conclusion 19

Abbreviations 19

Bibliography & References 20

Acknowledgement 20

List of Figures

Title **Page Number**

Figure 1: High-Split, Mid-Split and Sub-Split Spectra and ACI 4
Figure 2: In-Home and Neighbor ACI Scenarios 6
Figure 3: AGC in Full-Band Receiver 7
Figure 4: QAM Signal Level and AGC Changes Under Different ACI Patterns 7
Figure 5: Test Setup for In-Home ACI Evaluation 8
Figure 6: Model-A Tuner SNR Sample Series Under Fixed Rx Power and Variable HS-CM Tx Power (ACI level) 10
Figure 7: Model-A Tuner AGC Sample Series Under Fixed Rx Power and Variable HS-CM Tx (ACI level) 10
Figure 8: Model-A Tuner SNR Sample Series Under Variable Rx Power and Fixed HS-CM Tx Power (ACI level) 11
Figure 9: Model-A Tuner AGC Sample Series Under Variable Rx Power and Fixed HS-CM Tx Power (ACI level) 11
Figure 10: Model-A Tuner SNR Sample Series Under Different Rx and HS-CM Tx Power but Fixed Power Ratio 12
Figure 11: Model-A Tuner AGC Sample Series Under Different Rx and HS-CM Tx Power but Fixed Power Ratio 12

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 12: Model-B Tuner SNR Sample Series Under Different Rx and HS-CM Tx Power Levels 13
Figure 13: Model-B Tuner AGC Sample Series Under Different Rx and HS-CM Tx Power Levels 14
Figure 14: Model-C Tuner SNR Sample Series Under Different Rx and HS-CM Tx Power Levels 15

List of Tables

Title **Page Number**

Table 1: Upstream Configurations of the Test Setup 8
Table 2: Model-C STB Tuner SNR drop due to ACI under different HS-CM transmit power and STB receive power; Data corresponding to video tiling are highlighted with colors. 16

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

As the demand for network bandwidth continues to grow, the stress on the upstream direction of transmission becomes more pressing. This trend calls for more spectral resources for the upstream channels in a traditionally asymmetric coax-based access network. The DOCSIS® specification has defined mid-split and high-split spectrum allocation schemes to expand the upstream capacity. The work of this technical paper is to facilitate the deployment of mid-split and high-split DOCSIS networks with a proposed solution of managing potential radio frequency interferences among customer premises devices operating on the new spectral chart.

1.1. Mid-Split and High-Split DOCSIS Networks

Most Hybrid Fiber Coax (HFC) networks on which DOCSIS technology is operating allocates frequencies between 5 and 42 MHz for the upstream channels and above 54 MHz for the downstream. In the 42 – 5 = 37-MHz wide frequency band, typically four bonded 6.4-MHz Single-Carrier Quadrature Amplitude Modulation (SC-QAM) channels are configured to offer up to 140 Mbps of upstream bandwidth.

A *mid-split* spectrum allocation scheme was introduced in the DOCSIS 3.0 specification [1]. It moves the upper edge of the upstream band to 85 MHz, practically doubling the upstream spectral bandwidth. The DOCSIS 3.1 specification [4] pushed the upper edge further to 204 MHz and is referred to as the *high-split* scheme. Retrospectively, the legacy scheme with the 42 MHz upstream edge is called *sub-split*. The spectrum allocations between the DOCSIS upstream and downstream in sub-split, mid-split and high-split schemes are illustrated in Figure 1.

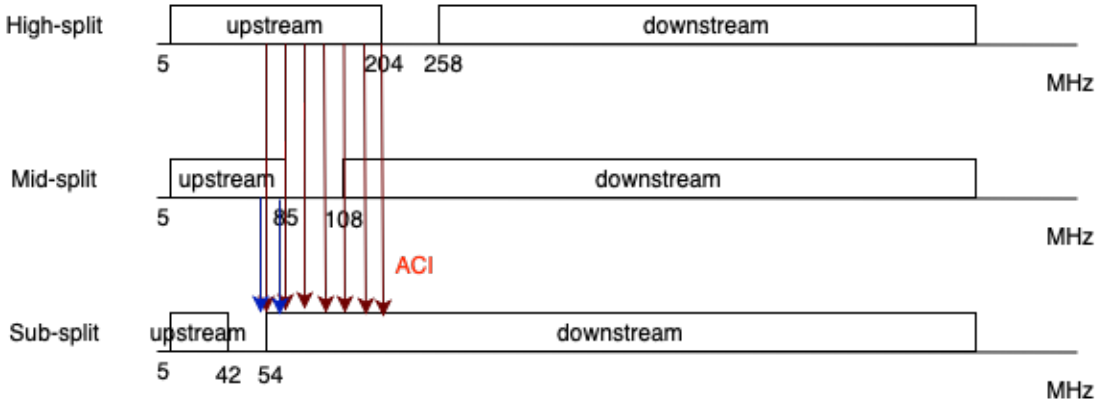


Figure 1: High-Split, Mid-Split and Sub-Split Spectra and ACI

The mid-split and high-split schemes ease the deployment of more efficient modulation and multiple access control technology, OFDMA, as they supply extra spectral bandwidth to accommodate the new signaling while keeping the legacy SC-QAM channels occupying the band below 42 MHz intact to support the large number of non-OFDMA-capable devices. When OFDMA is used in a high-split system, up to 2.5 Gbps upstream capacity can be achieved in DOCSIS 3.1 networks.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1.2. Adjacent Channel Interference in Mid-Split and High-Split Systems

A challenge to the deployment of a mid-split or high-split scheme is spectral compatibility. In a mid-split or high-split HFC network, the upstream radio frequency (RF) signal can go up to 85 or 204 MHz, respectively. However, the receiver of a sub-split customer premises equipment (SS-CPE) such as a video set-top box (STB) accepts RF signals starting from 54 MHz. The overlap of the upstream and downstream transmission bands, as shown in Figure 1, means that the SS-CPE may "see" RF power transmitted from mid-split or high-split cable modems (MS-CMs or HS-CMs) in the 54 – 85 or 54 – 204 MHz range.

Although downstream services will have been excluded from the frequencies below 108 or 258 MHz, the downstream lower edge of the mid-split or high-split system, a SS-CPE may still be susceptible to such interfering RF energy if it comes into and impairs the RF font-end of the SS-CPE receiver. This effect is referred to as *Adjacent Channel Interference* (ACI). Strong enough ACI can cause service impairment. For example, a sub-split set-top box may show video tiling when a neighboring MS-CM or HS-CM is transmitting in the upper part of the upstream band of a mid-split or high-split system. While an uncommon scenario, ACI can also exist from a HS-CM to a MS-CPE, because the upstream band of the high-split system partially overlaps the downstream band of the mid-split system.

Preventing ACI from causing service impairments is critical to the deployment of mid-split and high-split networks on which SS-CPE will be co-existing with MS-CMs and HS-CMs. Calibrating the downstream power for the worst case, ACI may not be practical as the output RF power at a node may have already been maximized for other reasons, such as to achieve maximum household-per-node efficiencies. ACI can be rejected or attenuated by installing a band-stop filter at the input of the SS-CPE to eliminate signals between 54 – 85 or 54 – 204 MHz. Then identification of the ACI-prone devices is highly desired because installing a band-stop filter on all SS-CPE devices in a network is prohibitively expensive.

1.3. Diagnosis of ACI

Many studies of mid-split and high-split ACI have been conducted to establish baselines and thresholds of ACI power in relation to margins when service impairments are observed. The metrics used by those studies may not be easily obtainable in a production network due to the lack of readily available telemetry data from the SS-CPE, MS-CMs, HS-CMs and other RF components. Consequently, the methods in those studies may not be suitable for the task of identifying SS-CPE devices that may be impacted by ACI, and subsequently should be targeted for proactive ACI mitigation.

This paper investigates an alternative method for remote diagnosis of ACI on a per SS-CPE basis using existing telemetry data from online devices. The method helps identify the devices suffering from ACI. Then ACI mitigation can target those problematic sites or devices so that high-split or mid-split can be enabled for them. The proposed method is generally applicable to both mid-split and high-split scenarios. The paper will focus on the high-split case.

2. Characteristics of High-Split ACI

2.1. Scenarios of High-Split ACI

High-split ACI is the 54 – 204-MHz RF signal emitted from a HS-CM and sinking into a SS-CPE. There are primarily two scenarios where ACI may manifest itself and both are illustrated in Figure 2.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

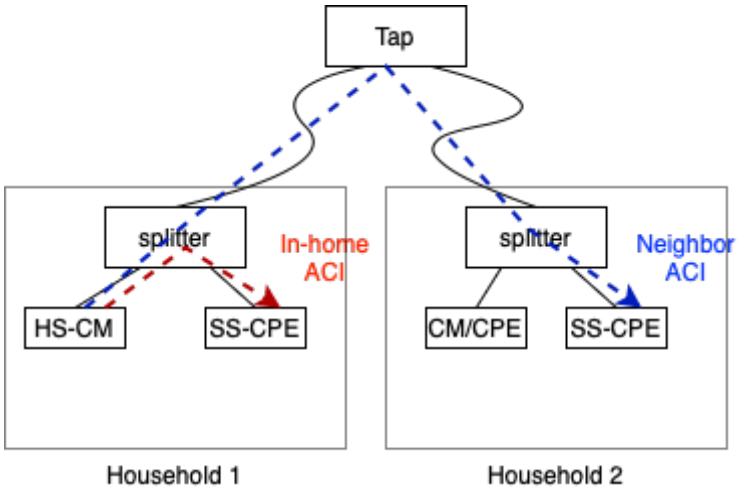


Figure 2: In-Home and Neighbor ACI Scenarios

2.1.1. In-Home ACI

The splitter in a customer home, which both the HS-CM and the SS-CPE are connected to, provides a signal path for the HS-CM's upstream to "leak" to the SS-CPE's receiver, when the port-to-port isolation of the splitter does not provide enough attenuation of the RF signals. The port-to-port isolation value of a typical consumer-grade splitter is about 25 dB. The cabling loss between the HS-CM and SS-CPE via the splitter also provides some attenuation of the ACI. Depending on the length of the coaxial cables, that loss is typically of 1-2 dB.

2.1.2. Neighbor ACI

Neighbor ACI happens between the HS-CM of one customer home and the SS-CPE of another through the tap shared by the two homes. The mechanism is like that of the in-home ACI, but the RF leakage happens between the ports of the tap.

The typical port-to-port isolation value of a tap is similarly 25 dB. There is much more cabling loss attributed to the two drop cables from the tap to the two homes, and the two in-home cables from the in-home splitters to the two devices. The insertion loss of the ground blocks and of the splitters in the two homes should also be accounted for when calculating the RF attenuation. The total attenuation is usually high enough to suppress the effect of ACI. ACI due to leakages of multiple hops of taps has less of an impact. Consequently, neighbor ACI is usually less of a concern than in-home in high-split systems.

2.2. Impact of ACI

The level and extent of the impact by ACI depend on the type of SS-CPE receiver and the pattern of ACI. There are several RF front-end architectures used in SS-CPE receivers, including the single-channel receiver, wide-band or block receiver and full-band receiver. The full-band receiver is prevalent in SS-CPE devices deployed in cable networks.

2.2.1. Automatic Gain Control Maladjustment

Figure 3 shows an example full-band RF receiver identifying the critical components of the RF front-end. The components include a Variable Gain Amplifier (VGA) to boost a weak input signal, an equalizer (EQ) to remove inter-symbol interference and an automatic gain control (AGC) logic to maintain the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

signal level to the QAM tuners. The AGC logic is a negative feedback loop: it calculates the difference between the output signal and a reference, integrates the error and uses the integral to tune the VGA and EQ. When in equilibrium, the AGC logic by design should get the optimal Signal to Noise ratio (SNR) and Modulation Error Ratio (MER) for QAM signals.

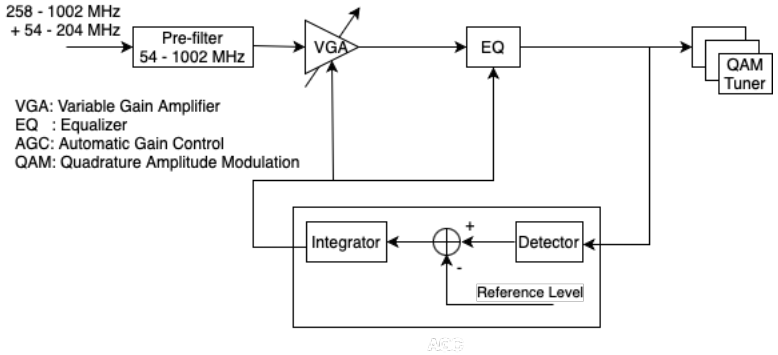


Figure 3: AGC in Full-Band Receiver

When ACI in 54 – 204 MHz frequencies enter the RF front-end of the SS-CPE receiver, the added energy will drive the AGC logic off the equilibrium. The AGC will then adjust to lower the gain of the VGA in response to the increased in-band power.

Several cases of the AGC adjustment due to ACI are illustrated in

Figure 4.

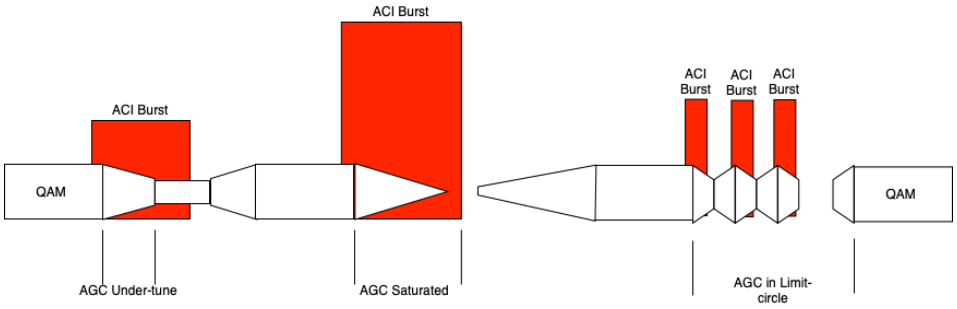


Figure 4: QAM Signal Level and AGC Changes Under Different ACI Patterns

Generally, the adjustment will result in reduced QAM signal levels to the input of the tuner. That also means reduced SNR and MER. In the case that the QAM signal level falls below a certain threshold, bit errors, packet loss and finally, service impairment will occur. When the ACI power is so high that it saturates the VGA completely, leaving no room for the AGC to maneuver, the QAM signal will be clipped. Signal clipping will result in excessive bit errors, packet loss and service interruption.

Another important case is when ACI comes in a sequence of short bursts. In this case, the AGC may step into a limit-circle and not converge to a steady state. The fluctuations in the QAM signal levels become added noise and significantly reduce the SNR at the QAM tuner.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.3. Numerical Evaluations of High-Split ACI Impact to STB

The susceptibility of three models of sub-split STBs to high-split ACI from one HS-CM is evaluated in lab, with a focus on identifications of indices and value threshold that can be used for detections of service-impairing ACI situations. The test setup is depicted in Figure 5.

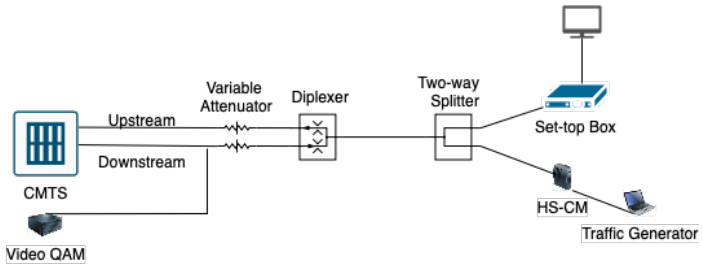


Figure 5: Test Setup for In-Home ACI Evaluation

Two variable attenuators are used to control the transmit and receive power of the two devices. A laptop is connected to the HS-CM and serves as a traffic generator to simulate upstream transmissions. The STB video output is monitored on a TV for subjective quality verification. The high-split upstream is configured with four SC-QAM channels and two OFDMA channels; the details of the channel parameters are shown in Table 1.

Table 1: Upstream Configurations of the Test Setup

SC-QAM 1	Center Frequency: 16.1 MHz Bandwidth: 6.4 MHz Type: ATDMA
SC-QAM 2	Center Frequency: 22.5 MHz Bandwidth: 6.4 MHz Type: ATDMA
SC-QAM 3	Center Frequency: 28.9 MHz Bandwidth: 6.4 MHz Type: ATDMA
SC-QAM 4	Center Frequency: 35.3 MHz Bandwidth: 6.4 MHz Type: ATDMA
OFDMA 1	Frequency: 39.6 – 88.0 MHz Subcarrier Spacing: 50 KHz Rolloff Period: 256 symbols Symbol Frame: 16 symbols Cyclic Prefix: 96 symbols
OFDMA 2	Frequency: 108.475 – 203.475 MHz Subcarrier Spacing: 50 KHz Rolloff Period: 96 symbols Symbol Frame: 16 symbols Cyclic Prefix: 256 symbols

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The test procedure includes polling the HS-CM and the STB for metric data periodically for a specified duration. In the middle of polling, the traffic generator is triggered to send traffic upstream for a short period. The metric data polled includes STB receive SNR, AGC of the active QAM tuner, which is tunned at 321 MHz, and the QAM channel frequency closest to the upper edge of the upstream OFDMA channel of the system under test. The procedure is run under different HS-CM transmit power and STB receive power settings, which are set by the attenuators.

2.3.1. STB Model-A

Model-A STB supplies tuner SNR and AGC data through a diagnostic shell over SSH. The test procedure polls the tuner status by SSH logging in the device and running diagnostic command at 5-second intervals for 120 seconds. At the lapse of 20 seconds, the HS-CM upstream transmission is triggered and lasts for 70 seconds.

Figure 6 and Figure 7 show the sample series of the STB tuner SNR and AGC values under a fixed STB receive power level -5.0 dBmV and three HS-CM transmit power levels 41, 43 and 45 dBmV. The subjective video verification results under each parameter set are annotated in the legend labels. The data reveals the effect of increasing ACI levels to the tuner SNR and AGC at a fixed receive power. The STB tuner SNR drops and AGC increases nearly linearly as the ACI increases. Video tiling is the result when the SNR drops below a threshold (around 28 dBmV). The AGC value is a unitless quantity, but its decrement is clearly correlated to the ACI levels and consistent with the SNR decrement.

Figure 8 and Figure 9 show the sample series of the STB tuner SNR and AGC values under four STB receive power levels -2, -5, -6 and -7 dBmV and a fixed HS-CM transmit power level 45 dBmV. The data reveals that the STB tuner SNR is impacted by the same ACI more at a lower receive power. The tuner SNR values are rather close under different receive power levels when ACI is not present, but the decrements of the SNR are bigger for lower receive power levels.

This fact is further corroborated by Figure 10 and Figure 11, where the STB tuner SNR and AGC values under two sets of different STB receive power and HS-CM transmit power levels with the same power ratio (difference by the dB values). At lower receive power, the STB tuner experiences more SNR decrements and AGC increments even under lower ACI levels (translated from lower HS-CM transmit power). This observation may indicate that the Carrier to Interference Ratio (CIR) alone may not be enough to characterize the impact of ACI. The baseline receive power and SNR are needed for the diagnosis of an STB's susceptibility to ACI.

The STB tuner AGC increase almost linearly with the decrements of the receiver power, and it can be completely saturated by the ACI when the receive power is too low. An AGC saturation case is captured in Figure 9 when the receive power is -7 dBmV.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

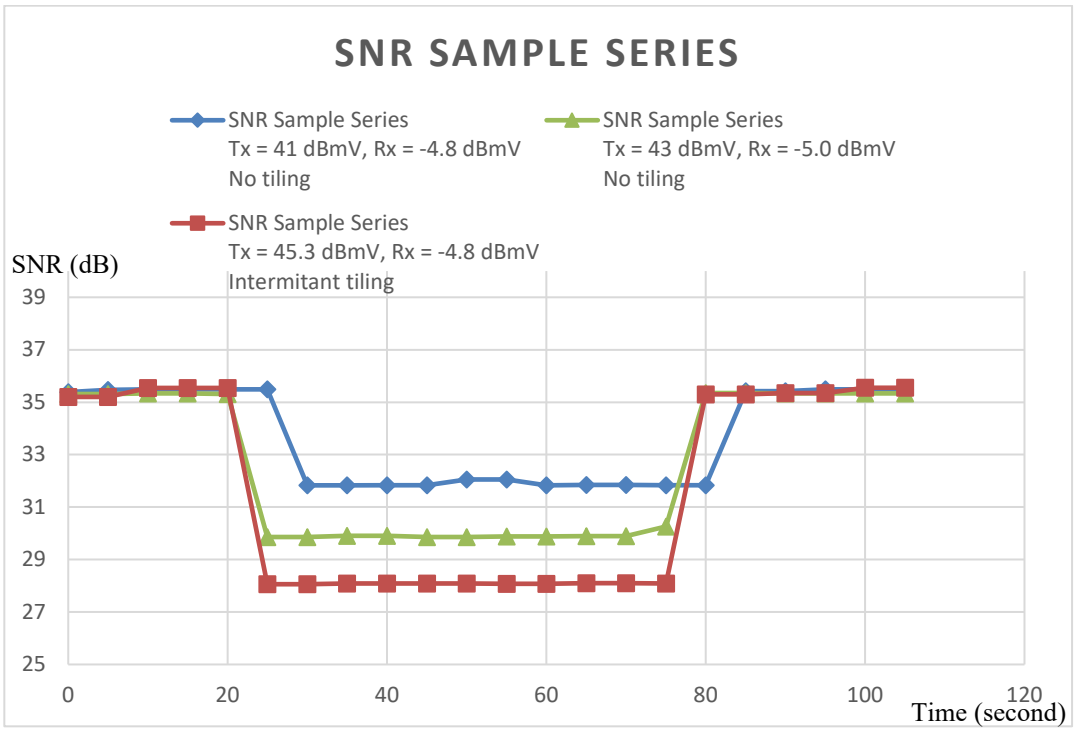


Figure 6: Model-A Tuner SNR Sample Series Under Fixed Rx Power and Variable HS-CM Tx Power (ACI level)

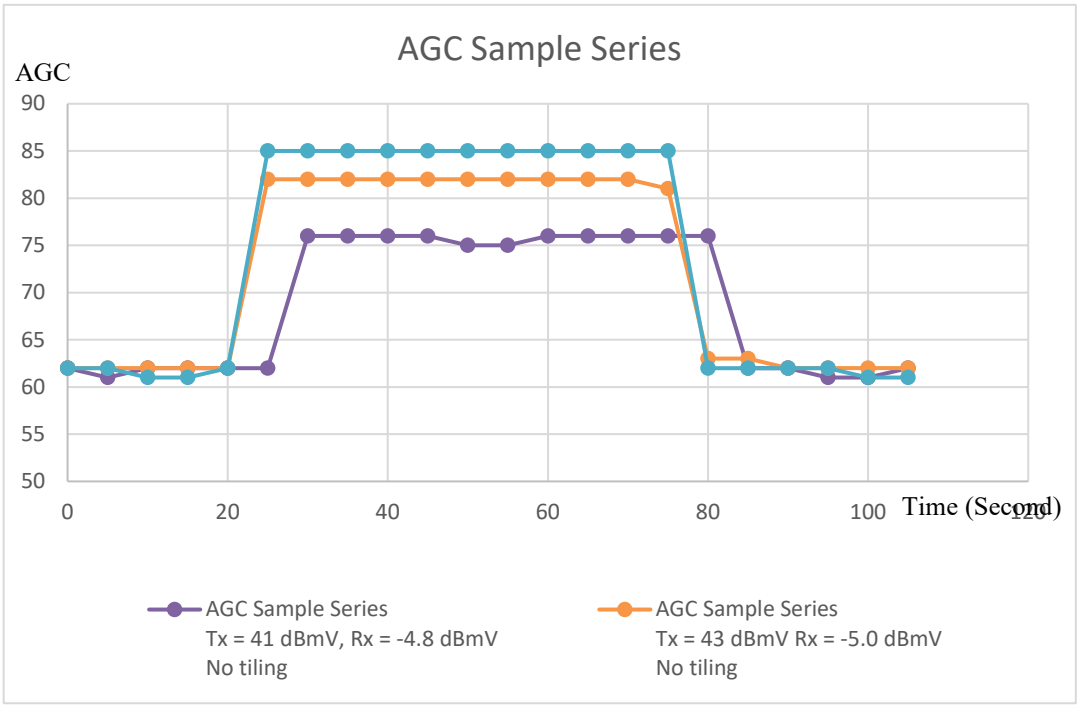


Figure 7: Model-A Tuner AGC Sample Series Under Fixed Rx Power and Variable HS-CM Tx (ACI level)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

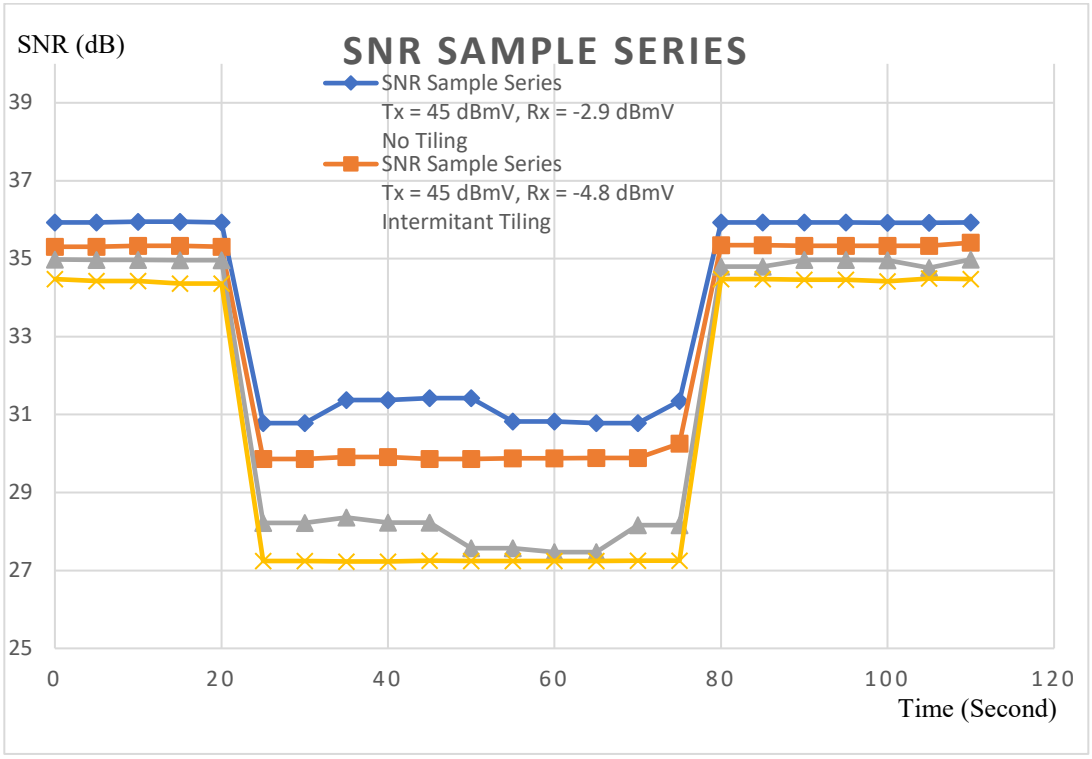


Figure 8: Model-A Tuner SNR Sample Series Under Variable Rx Power and Fixed HS-CM Tx Power (ACI level)

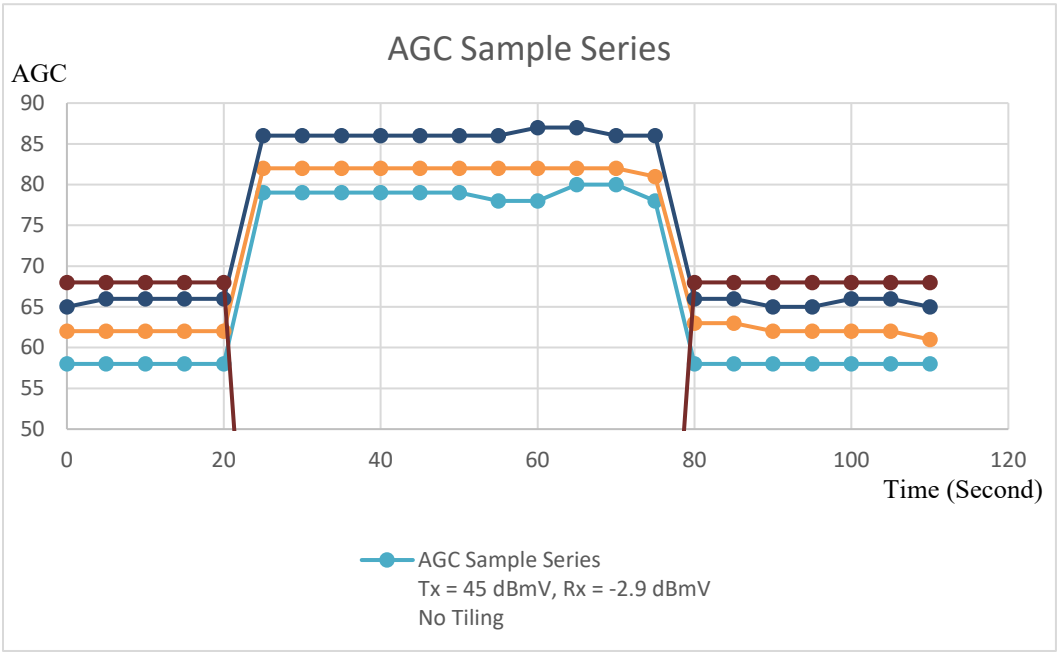


Figure 9: Model-A Tuner AGC Sample Series Under Variable Rx Power and Fixed HS-CM Tx Power (ACI level)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

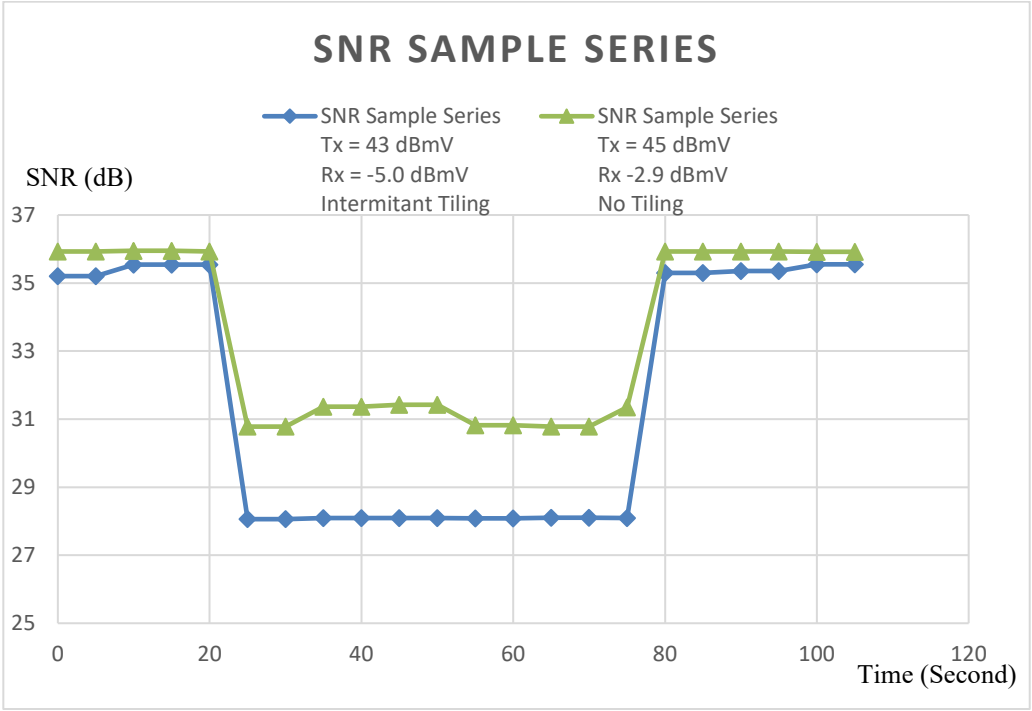


Figure 10: Model-A Tuner SNR Sample Series Under Different Rx and HS-CM Tx Power but Fixed Power Ratio

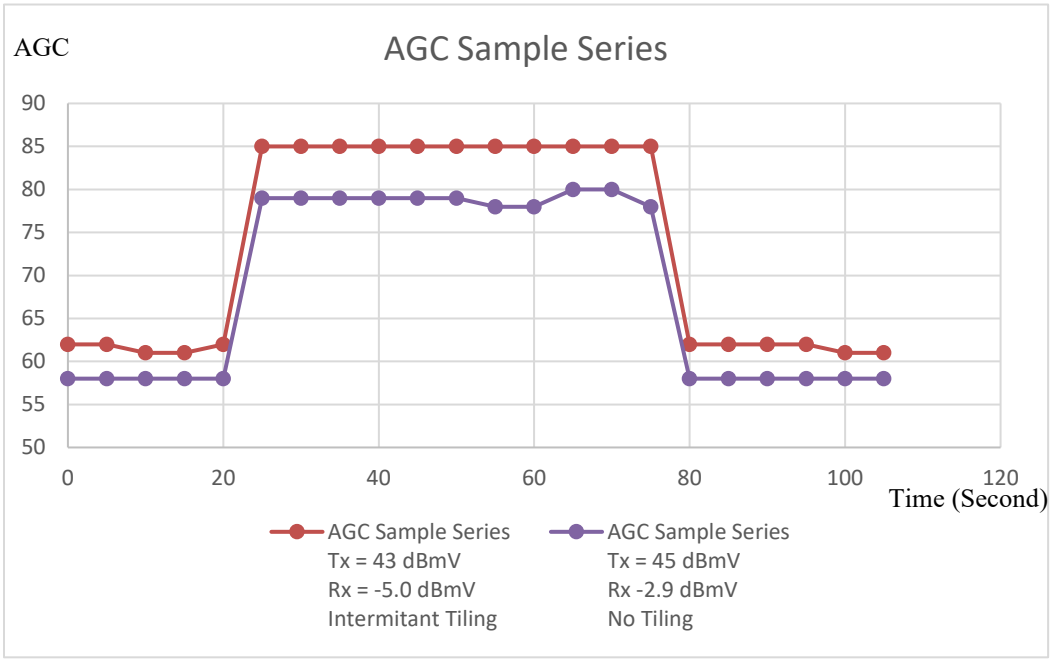


Figure 11: Model-A Tuner AGC Sample Series Under Different Rx and HS-CM Tx Power but Fixed Power Ratio

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.3.2. STB Model-B

Model-B STB has similar hardware and software architecture as Model-A. The tuner SNR and AGC data are also obtainable through a diagnostic shell over SSH. The STB is tested with the same sets of receive and SS-CM transmit power levels and same data polling procedure. The results are summarized in Figure 12 and Figure 13.

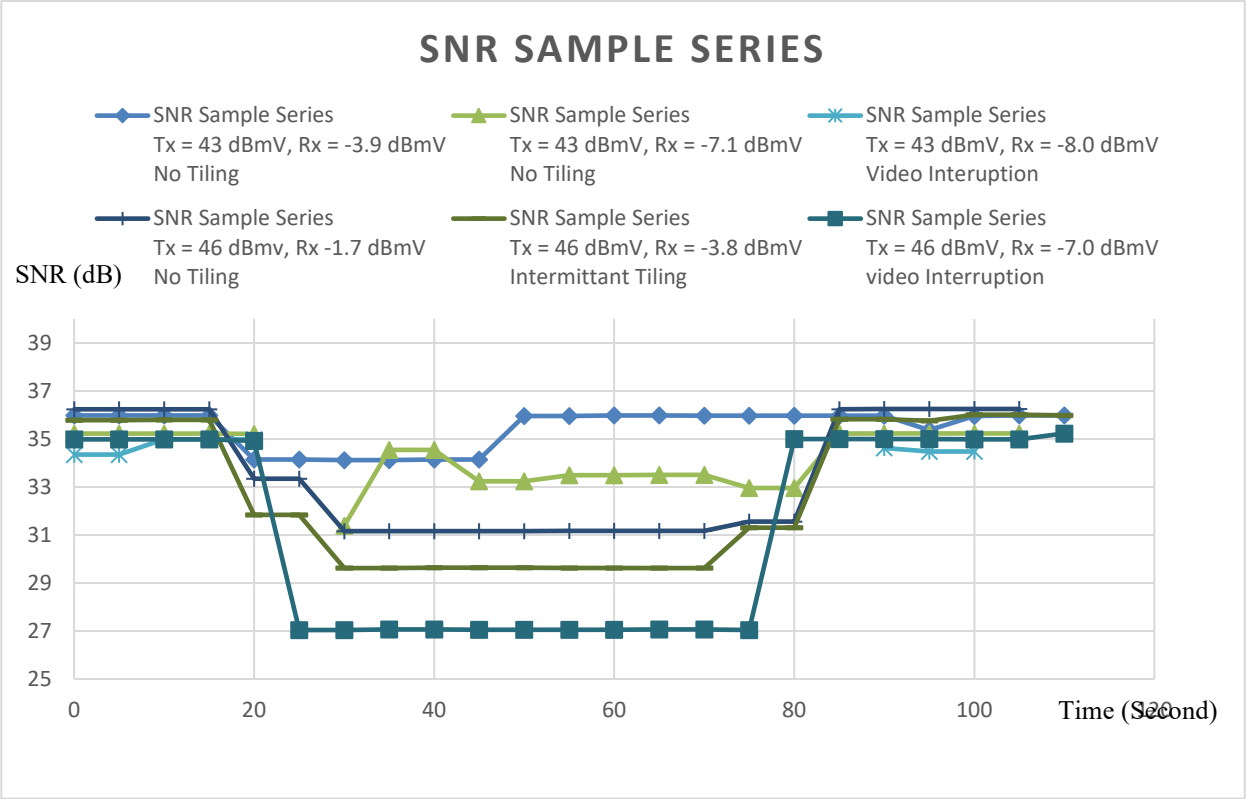


Figure 12: Model-B Tuner SNR Sample Series Under Different Rx and HS-CM Tx Power Levels

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

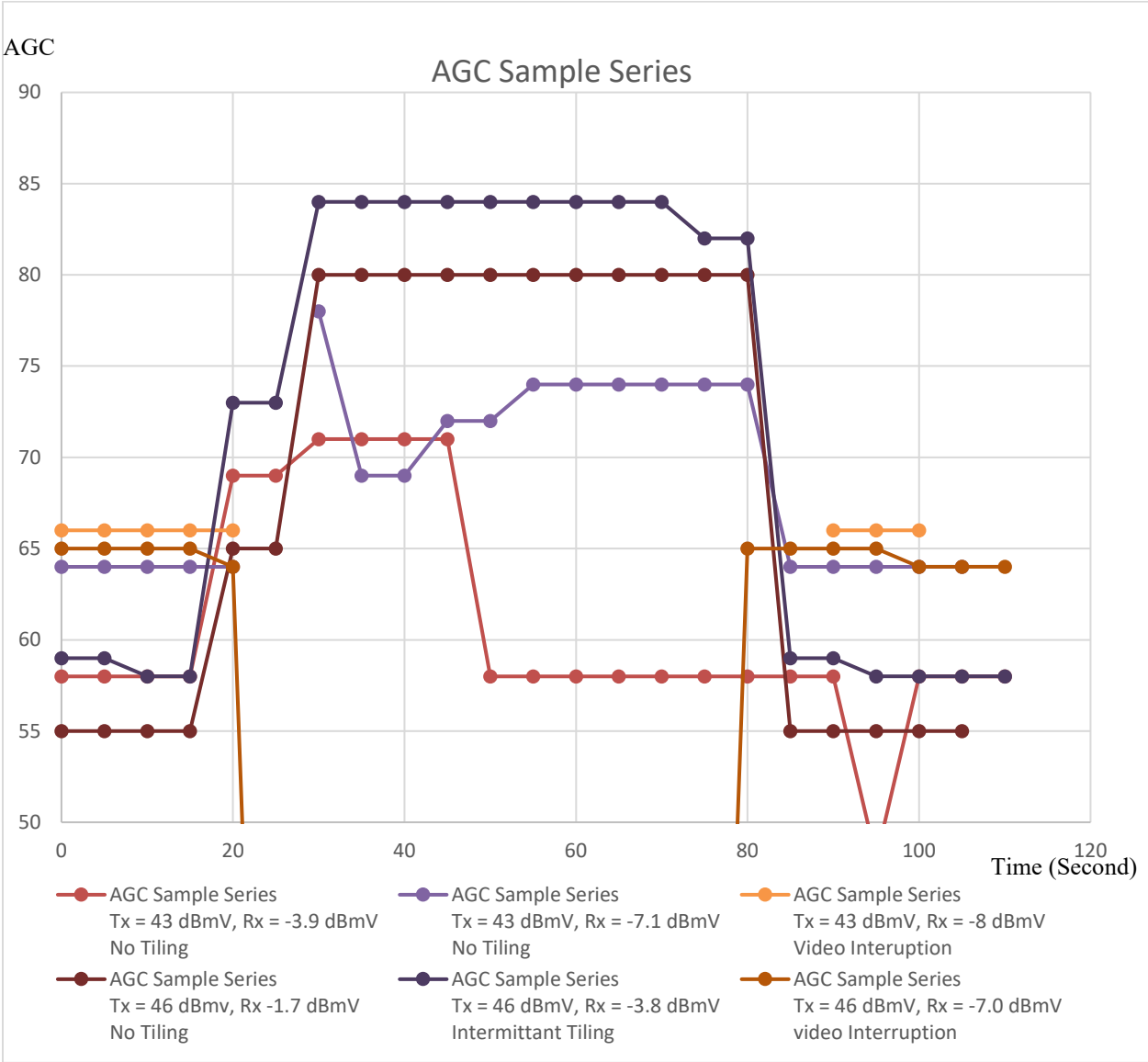


Figure 13: Model-B Tuner AGC Sample Series Under Different Rx and HS-CM Tx Power Levels

Model-B STB shows more dynamics in SNR and AGC values under ACI. Though most of the conclusions for Model-A apply to it, Model-B is slightly more tolerant to SNR drops in the low receiver power region.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.3.3. STB Model-C

The obtained SNR series are plotted in Figure 14, with the subjective video verification results annotated in the labels.

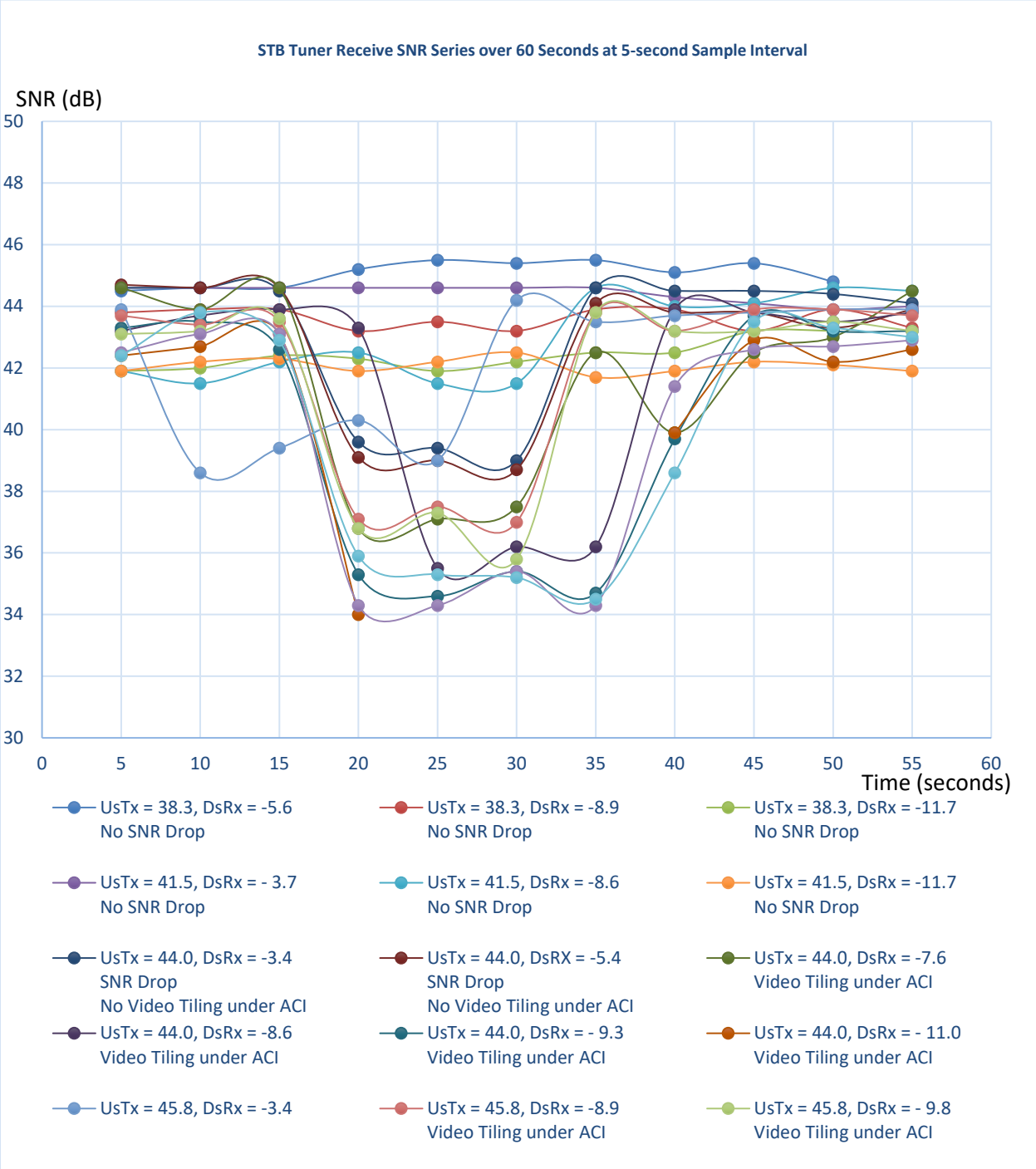


Figure 14: Model-C Tuner SNR Sample Series Under Different Rx and HS-CM Tx Power Levels

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Model-C STB is of different hardware and software architecture from the previous two under tests. It supports Simple Network Management Protocol (SNMP) and provides Management Information Base (MIB) objects for tuner SNR. The tuner AGC MIB is defined but only gives a fixed value that does not reflect the changed input to the RF front-end. The STB is tested with the same sets of receive and SS-CM transmit power levels as those for Model-A and Model-B. The data collection procedure is somehow different in that MIB polling is performed at 5-seconds for only 60 seconds and HS-CM upstream transmission is triggered at 10-seconds lapse for only 20 seconds.

The data reveals the following observations:

1) SNR drops happen when the HS-CM is transmitting, thus generating ACI. The SNR drops only occur when the HS-CM transmit power is above a certain threshold (44 dBmV in the test results) and the STB receive power is below a certain threshold (-3.4 dBmV in the test results). It is noteworthy that high transmit power or low receive power alone does not necessarily translate to SNR deterioration. This fact may be due to the nonlinearity of the AGC.
2) SNR drops do not necessarily translate to uncorrectable FEC errors. Only when the STB receive power is at the low-value region (below -7.6 dBmV in the test results) do uncorrectable FEC errors occur with SNR drops. It has been confirmed with subjective verification that video tiling happens when uncorrectable FEC errors occur.

The test data is also processed for characterizations of the magnitude of the SNR drop due to ACI, as summarized in Table 2.

Table 2: Model-C STB Tuner SNR drop due to ACI under different HS-CM transmit power and STB receive power; Data corresponding to video tiling are highlighted with colors.

STB Rx Power (dBmV) → **SNR (dB) Drop at ACI** ↓CM Tx Power(dBmV)	-4	-5	-8	-9	-10	-11	-12
38	**0**	**x**	**x**	**0**	**x**	**x**	**0**
42	**0**	**x**	**x**	**0**	**x**	**x**	**0**
44	**5.2**	**5.7**	**7.2**	**5.2**	**8**	**8.7**	**x**
46	**0**	**x**	**x**	**6.3**	**6.6**	**8.2**	**7.5**

The SNR drops are calculated as the difference between the average SNR when no traffic is generated and the average SNR when traffic is sent. All averages are taken for the lowest three values.

3. Detection of ACI Using Telemetry

Generally, the ACI of coexisting HS-CMs and SS-CPE is not easily resolvable. An efficient solution should be individualistic, which means that the high-split mode should be turned on or off for customer premises equipment individually, based on each site's unique RF and network conditions. A procedure should be developed that can remotely evaluate each customer home for ACI vulnerability so that interference mitigation measure, such as using a band-drop filter, can be targeted only to the vulnerable homes.

In Section 2.3, the key metric indices used to evaluate the impact of ACI, which include STB receive power, HS-CM transmit power, STB Tuner SNR and AGC, can be used in a production environment for targeted interference mitigations and proactive network managements. To implement the diagnostic

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

procedure of that in Section 2.3, a telemetry to monitor the metric indices and a method to exert upstream RF signals are needed.

As described in Section 2.3, two STB models, Model-A and Model-B, support remote access to the status of their tuners through SSH. The format of the data obtained through the diagnostic shell is not standardized, so a data parsing script shall be developed to make the data available for analytics purposes. STB Model-C supports SNMP and provides MIBs for the STB tuner status. The standard communication protocol and data model are already employed by many operation support systems so they offer easier data collection and analytics.

The following subsection will summarize some standard and proprietary MIB objects for STB receive power, HS-CM transmit power, STB Tuner SNR and AGC.

3.1. Useful MIBs

3.1.1. HS-CM Transmit Power MIBs

The transmit power of the HS-CM can be obtained through two MIB objects for OFDMA and SC-QAM channels.

.iso.org.dod.internet.private.enterprises.cableLabs.clabProject.clabProjDocsis.docsIf31Mib.docsIf31MibObjects.docsIf31CmUsOfdmaChanTable.docsIf31CmUsOfdmaChanEntry.docsIf31CmUsOfdmaChanTxPower

.iso.org.dod.internet.mgmt.mib-2.transmission.docsIfMib.docsIfMibObjects.docsIfCmObjects.docsIfCmStatusTable.docsIfCmStatusEntry.docsIfCmStatusTxPower

There is one (maybe of less interest) which gives the 1.6MHz power spectral density (PSD) associated with SC-QAM channels.

.iso.org.dod.internet.private.enterprises.cableLabs.clabProject.clabProjDocsis.docsIf31Mib.docsIf31MibObjects.docsIf31CmUsOfdmaChanTable.docsIf31CmUsOfdmaChanEntry.docsIf31CmUsOfdmaChanTxPower

3.1.2. SS-CPE Receive Power and Tuner Status MIBs

Depending on the type of the SS-CPE, the tuner information is reported in different MIB trees. For cable modems, DOCSIS has two MIBs for receive power on OFDM and SC-QAM channels.

.iso.org.dod.internet.private.enterprises.cableLabs.clabProject.clabProjDocsis.docsIf31Mib.docsIf31MibObjects.docsIf31CmDsOfdmChannelPowerTable.docsIf31CmDsOfdmChannelPowerEntry.docsIf31CmDsOfdmChannelPowerRxPower

.iso.org.dod.internet.mgmt.mib-2.transmission.docsIfMib.docsIfMibObjects.docsIfBaseObjects.docsIfDownstreamChannelTable.docsIfDownstreamChannelEntry.docsIfDownChannelPower

For the STB, the receive power of the QAM tuner is usually vendor proprietary. Though Open Cable Application Platform (OCAP) has defined the following MIB table:

.iso.org.dod.internet.private.enterprises.cableLabs.clabProject.clabProjOpenCable.ocStbHostMibModule.ocStbHostMibObjects.ocStbHostSystem.ocStbHostInterfaces.ocStbHostServiceProgramInfo.ocStbHostInBandTunerTable

it is not widely supported by STB vendors.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The STB Model-C provides the following MIBs on status of the QAM tuner. Noted that the MIBs, eSTBInBandTunerAGCValue and eSTBInBandTunerAGCState, do not reflect the true AGC status of the STB tuner.

1.3.6.1.4.1.1166.1.300.12.9.5 eSTBInBandTunerTable

1.3.6.1.4.1.1166.1.300.12.9.5.1 eSTBInBandTunerEntry

1.3.6.1.4.1.1166.1.300.12.9.5.1.1 eSTBTunerIndex

1.3.6.1.4.1.1166.1.300.12.9.5.1.2 eSTBInBandTunerModulationMode

1.3.6.1.4.1.1166.1.300.12.9.5.1.3 eSTBInBandTunerCarrierLock

1.3.6.1.4.1.1166.1.300.12.9.5.1.4 eSTBInBandTunerPCRLock

1.3.6.1.4.1.1166.1.300.12.9.5.1.5 eSTBInBandTunerDataLock

1.3.6.1.4.1.1166.1.300.12.9.5.1.6 eSTBInBandTunerEMMDataPresent

1.3.6.1.4.1.1166.1.300.12.9.5.1.7 eSTBInBandTunerFrequency

1.3.6.1.4.1.1166.1.300.12.9.5.1.8 eSTBInBandTunerAGCValue

1.3.6.1.4.1.1166.1.300.12.9.5.1.9 eSTBInBandTunerAGCState

1.3.6.1.4.1.1166.1.300.12.9.5.1.10 eSTBInBandTunerSNRValue

1.3.6.1.4.1.1166.1.300.12.9.5.1.11 eSTBInBandTunerSNRState

1.3.6.1.4.1.1166.1.300.12.9.5.1.12 eSTBInBandTunerCorrectedErrors

1.3.6.1.4.1.1166.1.300.12.9.5.1.13 eSTBInBandTunerUncorrectedErrors

1.3.6.1.4.1.1166.1.300.12.9.5.1.14 eSTBInBandTunerLongTermErrors

3.2. Availability of Upstream Traffic Generator

To diagnose ACI, controlled transmissions in the upstream direction from the HS-CM must be performed. In the test described in Section 2.3, a customer client device is used as a traffic generator. This method may not be practical in production, as accessing a customer client or installing a test client in the customer home is not generally allowed. On the other hand, the HS-CM usually implement some speed test applications for device and network monitoring purposes. These applications offer good tools to generate controlled upstream traffic.

Besides the Layer-3 or -4 packet generator internal or external to the cable modem, a DOCSIS technology, OFDMA Upstream Data Profile (OUDP), can readily serve the purpose. The OFDMA protocol needs to be aware of the frequency and time-fading conditions of the channel to perform many tasks such as channel equalization and data profile calculation. That means that periodic upstream transmissions of “training signals” are required. There are two forms of training signals: the upstream probe and the OUDP test. The upstream probe is a carrier-only burst, which is primarily for ranging-related functions. The OUDP test is in the form of user data and is transmitted using the assigned upstream data profiles and interval usage codes (IUCs) so it can be used to mimic user traffic to sound the upstream channel at both the physical and the medium access control layers. The OUDP test can be

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

controlled by the Cable Modem Termination System (CMTS) via the MAP messages with mini-slot assignments, just like any upstream user data transmissions.

The advantages of the OUDP test signal include: 1) As part of the DOCSIS built-in messaging it does not require external devices or additional software functions for packet generation, 2) the bandwidth and duty cycle can be controlled from the CMTS and 3) it behaves the same as customer traffic.

4. Conclusion

This paper investigated an alternative scheme for remote diagnosis of in-home ACI in high-split DOCSIS networks. It is found that the tuple of STB receive power, HS-CM transmit power and STB tuner SNR and AGC forms metric indices to evaluate the impact of ACI to a QAM tuner. Collecting this metric data using telemetry, coupled with data analytics, support remote diagnoses of ACI in production for targeted interference mitigation and proactive network management.

Abbreviations

ACI	Adjacent Channel Interference
AGC	automatic gain control
CM	Cable Modem
CLI	command-line-interface
CMTS	Cable Modem Termination System
DOCSIS	Data Over Cable Service Interface Specification
EQ	equalizer
FEC	Forward-Error-Correction
HS-CM	high-split cable modem
HFC	Hybrid Fiber Coax
IUC	interval usage code
MER	Modulation Error Ratio
MS-CM	mid-split cable modem
MIB	Management Information Base
OFDMA	orthogonal frequency division multiple access
OUDP	OFDMA Upstream Data Profile
PSD	power spectral density
QAM	Quadrature Amplitude Modulation
SNR	Signal to Noise ratio
SS-CPE	sub-split customer premises equipment
SNMP	Simple Network Management Protocol
SC-QAM	Single-Carrier Quadrature Amplitude Modulation
STB	set-top box
VGA	Variable Gain Amplifier

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

[1] Data-Over-Cable Specifications DOCSIS 3.0 Physical Layer Specifications, CableLabs, 2006.

[2] Data-Over-Cable Specifications DOCSIS 3.0 MAC and Upper Layer Protocols Interface Specifications, CableLabs, 2006.

[3] Data-Over-Cable Specifications DOCSIS 3.0 Operations Support System Interface Specification, CableLabs, 2006.

[4] Data-Over-Cable Specifications DOCSIS 3.1 Physical Layer Specifications, CableLabs, 2013.

[5] Data-Over-Cable Specifications DOCSIS 3.1 MAC and Upper Layer Protocols Interface Specifications, CableLabs, 2013.

[6] Data-Over-Cable Specifications DOCSIS 3.1 Operations Support System Interface Specification, CableLabs, 2013.

[7] OpenCable Home Networking MIB Specification, CableLabs, 2013.

[8] TR-069 CPE WAN Management Protocol, Broadband Forum, 2018.

Acknowledgement

The author would like to express appreciations to Justin Stiles, Principal Engineering; Chad Wells, Director, Engineering; Carlos Moro, Dir, Engineering and Michael Littleford, Senior Manager, Network Engineering for their support on lab setup, equipment provisioning and tuning, and data collections. And to John Williams, VP Engineering; Nick Segura, Principal Engineer III and Justin Stiles, Principal Engineer for their proof review of the draft.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Remote PHY Device Automation

From Theory to Deployment

A Technical Paper prepared for SCTE by

Rick Morrison
Manager, Network Automation Systems
Rogers Communication
2400 32nd Avenue, Calgary, Alberta, T2E 6T4
403 750 4500
rick.morrison@rci.rogers.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
1.1. Motivation 4
1.2. Objectives of DAA 4
2. Approach 5
2.1. Methodology 5
2.2. Deployment Milestones 5
3. Delivery Considerations 7
3.1. Release Pipeline Automation 7
3.1.1. Automating Builds and Testing 8
3.1.2. Automating Environment Turn-Up 9
3.1. Low-Code Paradigm 10
3.2. Cohesion and Coupling 10
4. Process and Automation Architecture 12
4.1. R-PHY Fiber node Processes 12
4.1.1. Core Build Process 13
4.1.2. Network Migration Strategy 14
4.1.3. New R-PHY Fiber Node Launch Process 14
4.2. Automation Architecture 16
4.2.1. Auto Configuration Orchestration 18
4.2.2. Provisioning Orchestration 19
5. Results 21
5.1. Scaling R-PHY Fiber Node Additions Over Time 21
5.2. Accelerated Node Activation Time 22
5.3. Errors in Auto-Configuration & Provisioning 22
5.4. Complexity and Level of Effort 23
6. Reflections 24
6.1. Positive Collaboration and Alignment Feedback 24
6.2. Continuous Improvement and Feedback Culture 24
6.3. Adherence to Industry Standards and Models 24
6.4. Influence of Conway's Law 24
6.5. Reliability of Automation Systems 25
6.6. Foundations and Essentials 25
7. Concluding Insights 25

Abbreviations 26

Bibliography & References 27

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Figures

Title **Page Number**

Figure 1: Multi Domain Service Orchestration (MDSO) Workload Instance 9
Figure 2: Cohesion and coupling 11
Figure 3: Decoupling teams' impact on schedule 12
Figure 4:R-PHY fiber node processes 13
Figure 5: DAA Core Build Process 14
Figure 6:Network Migration Strategy 14
Figure 7: Analog & R-PHY network topology 15
Figure 8: New R-PHY node process 15
Figure 9: Orchestration System Interactions 16
Figure 10: Auto configuration UI 17
Figure 11: Autoconfiguration Sequence 18
Figure 12: Provisioning sequence 19
Figure 13:Output of Provision 20
Figure 14: R-PYH Device Additions over time 21
Figure 15:Time to Activate a Node 22
Figure 16:Error Rate 23

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

As functionality moves away from the core, hubs are shifting from housing specialized RF networks to commodity fiber-based IP networks.

Trading RF for IP based networks simplifies the physical work and time required by a technician to setup and maintain the physical aspect. As a tradeoff, additional complexity is added in configuring and maintaining software configuration.

In this paper, we will discuss how we reduced complexity and eliminates repetitive, manual tasks by embracing an approach that automates the delivery and management of infrastructure. We will also present some essential steps and best practices that were used in managing the rollout of automation.

1.1. Motivation

We had two major drivers for automation The first was the ongoing need for network capacity growth which was not sustainable with the status-quo. The second was the need to improve node activity reliability.

1.2. Objectives of DAA

To understand why we picked the automation options we did; it is important to review the overall objectives we had of distributed access architecture (DAA) and how those influenced the approach.

1. Network Automation. DAA significantly simplifies our hub architecture allowing us the leverage to increasingly automate provisioning, turn up and streamline our overall process.

2. Multi-gig symmetric broadband services. A DAA architecture is required to deploy full duplex DOCSIS (FDX) or Extended Spectrum DOCSIS (ESD) and deliver multi-gigabit symmetric speeds over our HFC Network.

3. Facilities Improvement. DAA does not eliminate the need for us to scale and upgrade our facilities, however it reduces the scope and cost of these upgrades significantly.

4. Leveraging DAA for our future. DAA extends an IP network deep into the plant. This network can support multiple services for residential and commercial use-cases. In addition, DAA is a steppingstone to virtual cable modem termination system (vCMTS) and the edge cloud.

In the context of automation and considering the above objectives it was important to choose what to automate first. We chose to automate the process that introduced new R-PHY fiber nodes to begin with.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2. Approach

2.1. Methodology

Creating a synergistic and effective team is a shared aspiration, yet achieving this level of collaboration is often challenging. Our objective was to establish a working framework that would foster cooperation among network, software, and operational teams.

Within this framework, we set forth the following objectives:

1. Adopt continuous improvement methodologies by releasing regular improvements to deployment and operational tool sets.

2. Implement feedback systems from our previous production deployments, and from industry knowledge sharing, to continuously optimize/enhance future deployments.

3. Development of an architecture that supports a transition to next generation technologies such as vCMTS and FDX/ESD.

4. Automation-first philosophy in the development of our deployment processes. Leverage existing operational support system (OSS) investments to provide high-value automation opportunities. Engage with network teams to support identification of high value automation opportunities leveraging lean methodologies.[22]

Creating a single project team comprised of both software and network teams was crucial to fostering collaboration. This approach allowed teams to collaboratively analyze all facets of solutions, provide mutual support, acquire new proficiencies, and align priorities effectively.

2.2. Deployment Milestones

We adopted a strategy of "continuous improvement" and "iterative process" as we worked through our objectives. To make things manageable, we followed a common approach of "crawl, walk, run" where work was divided into smaller, "bite-sized" steps to achieve our goals.

Our main goals were split into distinct stages:

1. **Network lab trial**: We started by evaluating the various components of the network. This process ensured the interop between network components from a variety of vendors. This was a very manual, but necessary to understand what it is we were building including:

    - Understand how we were going to support frequent firmware upgrades that address frequent bug fixes and new features associated with a new technology.

    - Understand how we were going to troubleshoot and monitor for issues by using command line interfaces (CLI), element managers (EMS) and consolidated logging.

    - Identify required components and gaps for future automation.

    - Create sequence diagrams and models for inventory and orchestration based on the lab trial workflows.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2. **Minimal viable product (MVP):** This is the basic version of our solution that we can use for testing and ensure functionality. We validated that it could manage video, voice, and data services to proceed into customer trials.

    - First automation use case was to add an R-PHY fiber node to the back office and network.

3. **High touch trial (10 nodes):** This stage included a highly knowledgeable cross-functional team to support the trial. This team was also tasked with identifying process and preparing for the next stage of ramp up.

    - Learn to iterate functionality of monitoring & activation tools and processes.
    - Heavy focus on integrating technology with operational processes.
    - Introduction of automation.

4. **Operational ramp up**: Here we aimed to share knowledge across a broader group, including additional regions. The focus was on quality vs quantity and velocity of rolling out new nodes.

    - New fiber node activations without the direct support of development teams
    - Introduce an easy-to-use service order (SOM/SO) user interface for an improved user experience.

5. **Business as usual - Scale Deployment**: This phase involved scaling up to thousands of R-PHY fiber nodes. The goal was to target critical markets first, and then start to prepare our transition to next generation technologies under the DAA umbrella.

    - Introduce R-PHY fiber nodes that supported N x N configurations for future considerations including the use of 2x10GE links back to the CIN for additional capacity.
    - Update R-PHY models to align to standards.[2][3][4][5][6]
    - Tune automation and inventory systems to align with forward looking processes and architecture.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Delivery Considerations

In the realm of confidently delivering automation products iteratively, our attention was directed towards three key areas:

1. **Automating the release pipeline**: Our efforts were concentrated on streamlining the release pipeline through automation. This not only expedited the delivery process but also improved its reliability.

2. **Utilizing low-code techniques:** We explored the use of low-code techniques as part of our strategy. This approach aimed to strike a balance between efficiency and customization, enhancing our overall development.

3. **Emphasizing Cohesion and Coupling:** Recognizing the significance of a cohesive and well-coupled system architecture, we integrated this principle into our strategy to ensure robustness and maintainability.

An invaluable lesson we grasped early in the process was the dual importance of automated testing and an automated release pipeline. This concept of "automating the automation" significantly contributed to the confidence of a successful release, and efficiency of our release lifecycle.

3.1. Release Pipeline Automation

A release pipeline serves as a framework through which software transitions from development through to production. This process incorporates continuous integration and delivery (CI/CD) along with automated testing supporting frequent and confident software releases.

In the absence of this automation, the process of releasing software took us around three months. However, with the implementation of an automated release pipeline, this timeline was drastically reduced to 15 minutes. The integration of an automated pipeline also facilitated the support of concurrent development efforts and automated testing activities.

This was accomplished by:

1. **Automating builds and testing:** We automated the process of building software and conducting unit & mock testing. This streamlined approach expedited the development cycle while maintaining quality.

2. **Automating multiple environment**: We introduced automation through deploying software solutions in containers using Kubernetes, enabling the rapid creation of multiple testing and production environments. This approach enhanced efficiency and consistency in deployments.

3. **Decoupling solutions and domains**: We made a conscious effort to separate different solutions and domains, considering both distributed and monolithic architectures. This separation allowed for greater flexibility, scalability, and adaptability in our software development. Most importantly, this separation helped keep cognitive loading manageable and contained with the knowledgeable domain.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. **Adopting cloud native and 12 factor App principles:** By embracing cloud-native practices and adhering to the twelve-factor app principles, we ensured that our software was designed to fully leverage the capabilities of cloud native and promote scalability, resilience, and maintainability.[7][40]

These efforts collectively exemplify the immense impact an automated release pipeline has on the software development cycle, not only in terms of speed and quality, but also in terms of enabling modern development practices that align with industry.

3.1.1. Automating Builds and Testing

The automation of the build process occupies a leading role within the CI/CD framework. These automated procedures form the initial sequences of steps intentionally designed to swiftly identify any discrepancies within code arising from recent modifications.

The following procedure describes the steps encompassed:

1. **Detailed ticket documentation:** A meticulous description of the task's objectives and rationale is documented. This documentation allows future traceability, and any team member to comprehend and fulfill the task's requirements effectively.

2. **Environment setup for building and testing:** An environment is prepared to accommodate both the building and testing processes.

3. **Code is written in version control and merged:** Code, along with unit and other tests, are written and related to task tickets. Select code is merged to a branch for automated testing**.**

4. **Automated code reviews and static analysis:** These are implemented through rules based static analysis tools and custom build unit & logic tests. Manual peer reviews are still a critical component of development, but can be decoupled from the automated process, or reduced.

5. **Compilation and building**: The code is compiled and built, transforming it into a fully functional, containerized software product.

6. **Automated documentation generation and publication**: Documentation is automatically generated and published wherever feasible.

7. **Artifact tagging and registry storage**: All other binary artifacts are tagged and stored within a registry for systematic organization.

8. **Deployment and testing of artifacts and configuration**: The containerized artifacts and configuration are deployed and subjected to comprehensive, automated testing.

9. **Deactivation of build and test environments**: The build and test environment are deactivated and automatically torn down.

10. **Team notification of outcomes**: The outcomes of the entire process are communicated to the team through notification services.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

We also implemented API contract testing of all systems integrated into the automation solution by constructing scenarios involving successful and erroneous use cases, utilizing tools like RESTAssured [32] and Mockito[25]. In consideration of the "practical test pyramid" [39], we ensure that we adhere to both upstream and downstream contracts where possible. This approach reduces the necessity for resource-intensive, end-to-end testing, thereby optimizing testing efforts, while improving release confidence.

3.1.2. Automating Environment Turn-Up

This shift negates the need for developers to endure infrastructure delays or depend on the development progress of other systems.

By embracing versioning practices with environments, adhering to cloud native principles, and adopting declarative definitions, the outcome is infrastructure that is predictable and reliable. Notably, this infrastructure can be swiftly deployed within seconds, fostering and environment conducive to experimentation and innovation.

Over the span of a year, we transitioned from a substantial monolithic structure to a hybrid distributed structure that profoundly elevated our velocity and confidence.

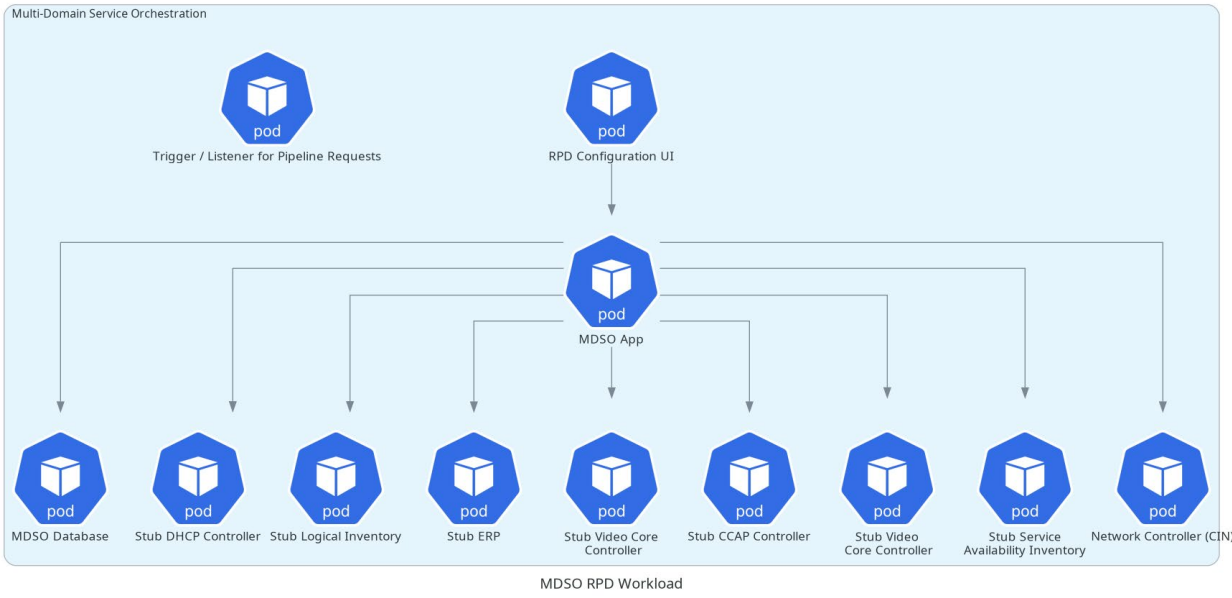


Figure 1: Multi Domain Service Orchestration (MDSO) Workload Instance

One or many instances (Figure 1) of a workload is meticulously brought into existence, tested, and systematically torn down each time a code merge or commit occurs. This parallel operational framework fosters the rapid simultaneous development of multiple features, ensuring a high rate of efficient development.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.1. Low-Code Paradigm

In consideration of the lifecycle of an R-PHY fiber node, several low code strategies were adopted to accelerate development and simplify changes. [24]

Some of the low-code techniques we leveraged:

1. **Utilizing templates and models:** such as YAML (yet another markup language), YANG (yet another next generation) or HOT (heat orchestration templates) we were able to reduce the amount of code. [11][13][14][17]

2. **Auto-generated adapter code and tests:** derived from OpenAPI standards (e.g., REST swagger). [36]

3. **Industry standards:** benefited from constructs described from TMF638 including resource and customer facing services (RFS/CFS) and TOSCA.

4. **Declarative syntax:** The declarative approach defines the desired state of the system, including resources and properties infrastructure should have. In contrast an imperative approach defines a sequence of specific commands needed to achieve the desired configuration.

By adopting these techniques unnecessary and duplicate coding through layers is avoided.

3.2. Cohesion and Coupling

Cohesion and decoupling are two key concepts in software architecture, which serve to streamline intricate software components into more manageable work segments. This arrangement has several advantages:

- Reduction of cognitive load[9][37][35]
- Facilitation of parallel development through domain contracts and mock procedures
- Determination of optimal team sizes to maximize skill-based contributions.

To address the structing of tasks, work packages were subdivided into skill defined domains and explicit API (application interface) contracts were established between domains. This process defined two layers of architecture:

1. **The multi-domain orchestrator:** This orchestrator is responsible for the composition of all the services, coordinating resources declaratively defined in domains.

2. **The domain orchestrator(s):** Responsible for domain-specific orchestration, this layer operates through imperative sequences. It abstracts imperative actions into a declarative API which the multi-domain controller utilizes.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

This arrangement (see Figure 2) ensures that domain-specific knowledge is concentrated within a domain orchestrator, alleviating cognitive burden within a specific domain. Qualifying cognitive load is intricate, underscoring the importance of garnering feedback from teams to calibrate the balance of complexity.

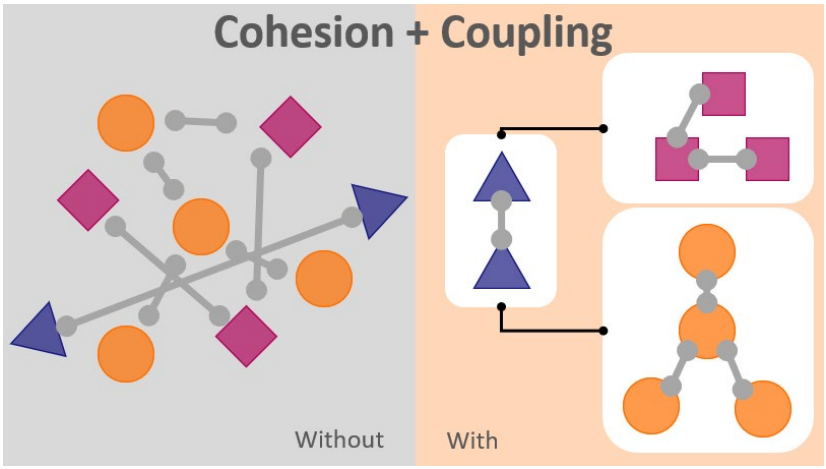


Figure 2: Cohesion and coupling

Decoupling is crucial because it simplifies the intricacies of the solution, rendering them easier to grasp and manage. On the other hand, coupling (or cohesion) serves to streamline the solution by aligning closely related elements together.

Key strategies for (de)coupling:

1. **Decoupling systems and domains for change management:** By disentangling systems and domains, the ability to enact changes without affecting external systems.

2. **Feature decoupling via API versioning:** API versioning allows for the separation of features, enabling modifications to specific components without impacting others until ready. This allows teams to develop at their own pace.

3. **Decoupling development through contracts and mock services:** Defining contracts and implementing "mock services" that simulate external systems or APIs enables teams to develop independently and at their own pace.

4. **System decoupling for resilience:** Decoupling systems ensures that if one system experiences an outage, the dependent system either generates an error or queues events.

5. **Selective coupling of closely related elements:** Closely related elements should be grouped together (ex: in a specific domain of orchestration), are strategically coupled to bolster confidence in real-time network provisioning.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Cohesion and coupling stand as fundamental attributes of our architecture that support velocity, confidence in delivery and reliability.

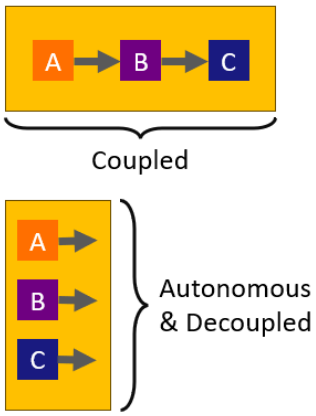


Figure 3: Decoupling teams' impact on schedule

In the realm of coupled or serial development, a sequential progression prevails (see Figure 3). This entails a situation where Team A's progress hinges on Team B's advancement, while simultaneously Team B is reliant on Team C's completion of tasks. This sequential reliance can result in notable delays in overall completion.

In contrast, the approach of autonomous development adopts a parallel trajectory. Here, Team A, Team B and Team C operate independently within their domain. This independence empowers teams to develop concurrently and reduces the aggregate time required to accomplish a project.

Autonomous development practices were instrumental in optimizing efficiency and expediting our success.

4. Process and Automation Architecture

The establishment of a process automation architecture is important to articulate a clear understanding of the business requirements and applications of the involved in automation efforts.

This section will describe these processes with focus on activations of R-PHY fiber nodes.

4.1. R-PHY Fiber node Processes

The journey of process improvement takes it first step by identifying existing business processes. Thought a detailed analysis of the operational landscape, inefficiencies and opportunities can be pinpointed. Armed with this insight, optimizing, and enhancing a business process unfolds.

Our goal was two-fold: Introduce a streamlined process alongside automation efforts.

At a high level we identified processes associated with introducing a new R-PHY node as follows:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

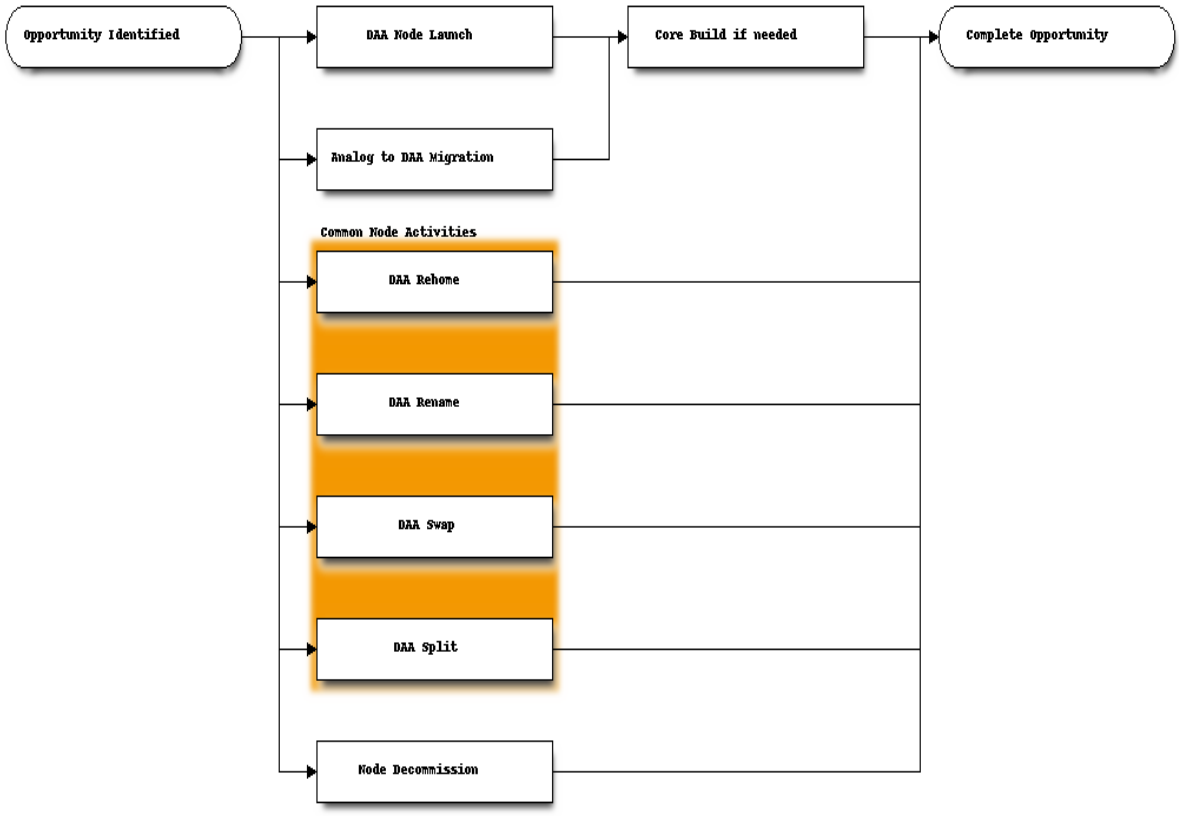


Figure 4:R-PHY fiber node processes

The crux of our automation endeavors was centered around the processes of adding & removing new R-PHY fiber nodes and the analog-to-DAA migration process.

This particular emphasis was attributed to our aspiration to scale up the deployment of R-PHY fiber noes within an established customer base, while upholding a sustainable pace.

The decommissioning aspect holds particular significance, as it supports our ability for regression testing and frequent updates, making it a pivotal use case to address.

4.1.1. Core Build Process

The focal point of the core build process centers on the deployment of the essential infrastructure (i.e., distributed converged cable access (D-CCAP) chassis and a video core) to accommodate DAA and R-PHY fiber nodes. [23]

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

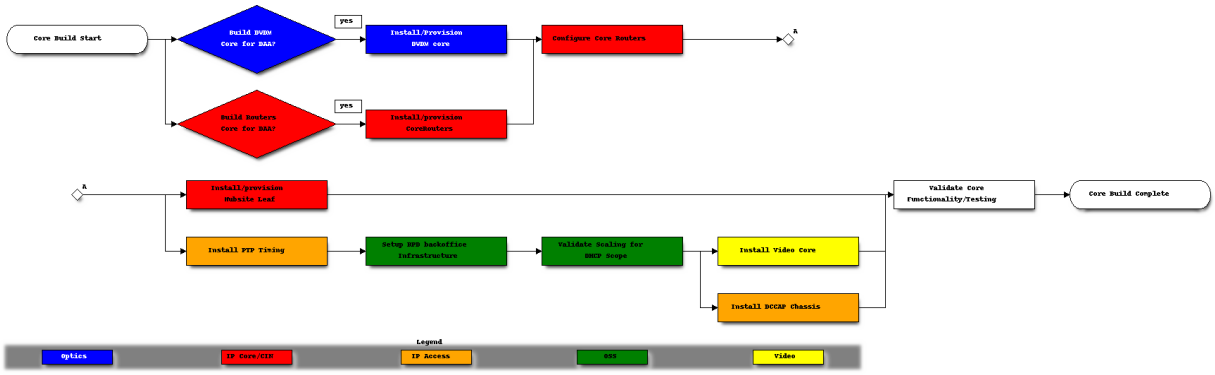


Figure 5: DAA Core Build Process

In our observation, the core build process (see Figure 5) was like the conventional build-out processes. The core build process is mainly physical cabling and physical documentation. As such this process was not a priority for automation.

4.1.2. Network Migration Strategy

R-PHY deployments are one step in a comprehensive plan (Figure 6) to further the evolution of the network. It is important to highlight during the sub split and mid split phases we are "preparing" the physical aspects of the plant to support R-PHY (see Figure 6).

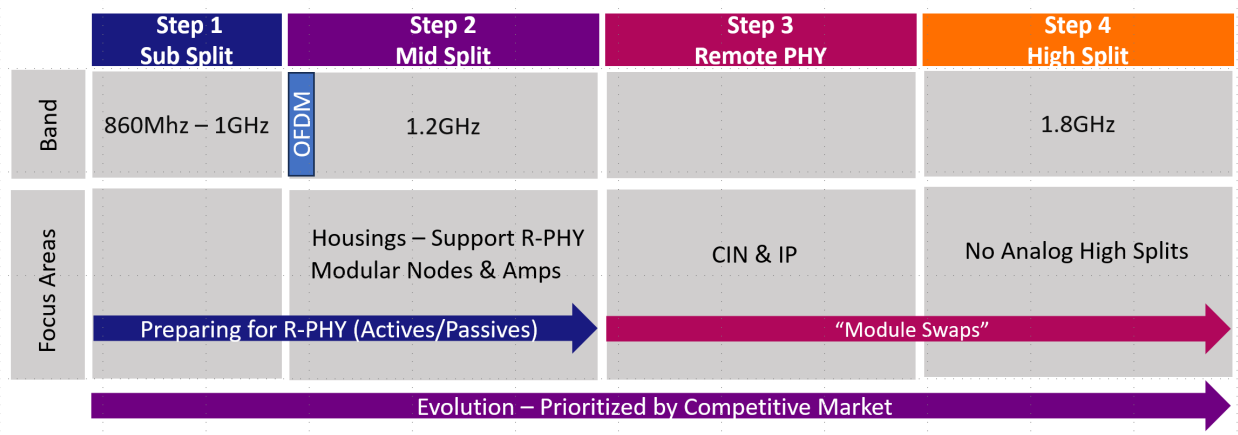


Figure 6:Network Migration Strategy

Our strategy has been to prepare housings, RF actives and passives to support bandwidth changes and modularize components where possible. This strategy has simplified deploying R-PHY by reducing the complexity when deploying a R-PHY node.

4.1.3. New R-PHY Fiber Node Launch Process

The "R-PHY Node Launch" and "Analog Conversion" processes (see Figure 7) emerged as the central contenders for automation priority. This choice was largely driven by the intricacies inherent to these processes and the desire to rollout DAA.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

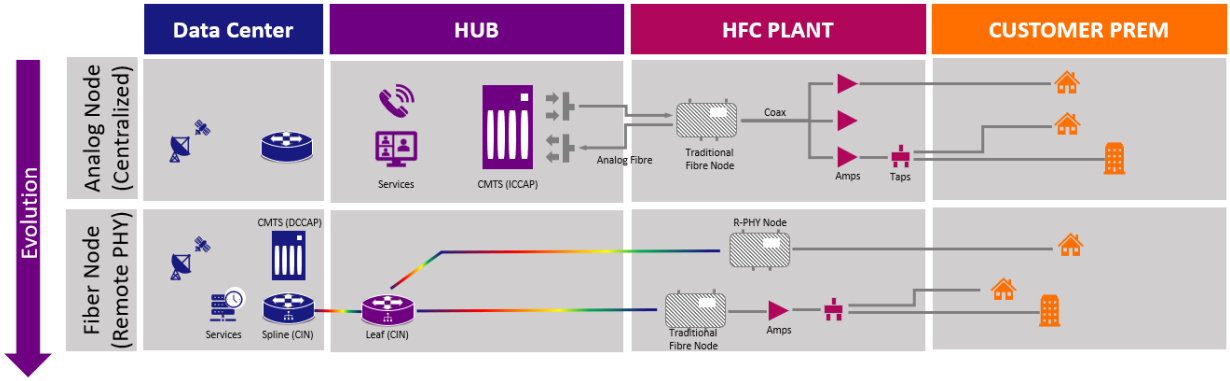


Figure 7: Analog & R-PHY network topology

A notable distinction between analog fiber nodes and R-PHY fiber nodes configuration is the substation of RF with IP technologies between the fiber node and CMTS. This entails replacing tangible physical construction with software-based configuration procedures.[2]

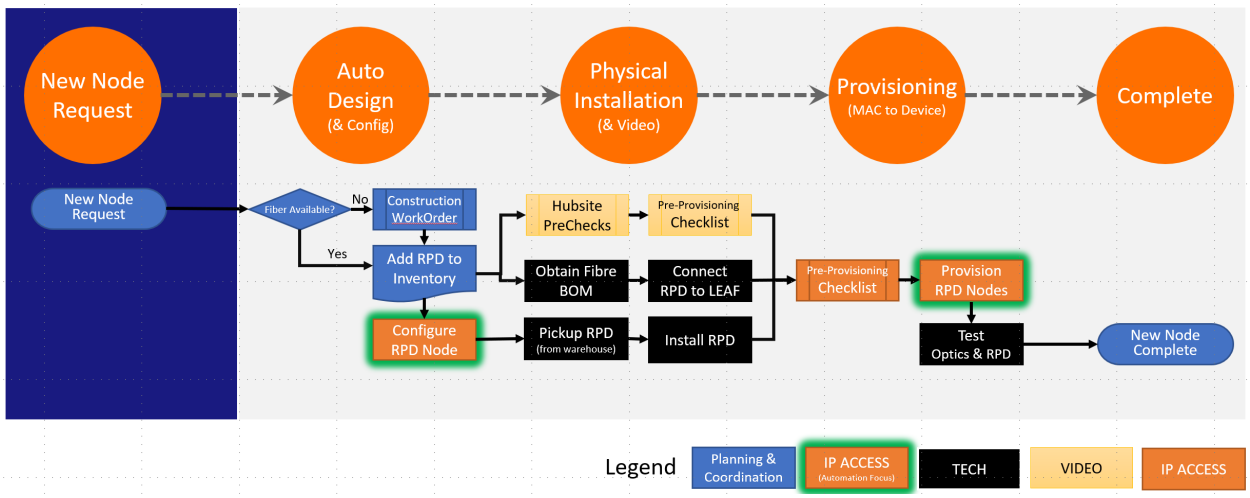


Figure 8: New R-PHY node process

At a broader perspective, a new R-PHY fiber node request (see Figure 8) entails an activity which triggers documentation of the fiber node within spatial inventory. This documentation, combined with current state inventory serves as the foundation for guiding network technicians in the installation of cables, and automating the design and configuration of the fiber node.

Upon the completion of the physical aspects of the installation, the activation process commences by associating the MAC (media access control) address of the R-PHY node with the video core and CMTS (cable modem termination system).

The entire process from beginning to end can span anywhere from a couple of days to a month depending on the complexity of the physical work.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The two automated tasks (configure and provision) replace approximately eight hundred lines of configuration with a handful of attributes per R-PHY device (device name, hub site & mac address or serial number of the R-PHY device).

Future automation will include pre-provisioning and the supplementary validation. Our judicious approach dictates that continue improvements be sequenced in accordance with diminishing returns of value, while simultaneously upholding the overarching architectural objectives.

4.2. Automation Architecture

Upon examination of the "new node process," the most substantial value derived from automation would be and was realized through the automation of the auto-configuration and auto-provisioning tasks. These two tasks would also be accelerated by leveraging prior OSS modernization investments.

As part of our strategy, we opted to continue utilizing our existing tools for monitoring and node segmentation tasks, as these required minimal or no alteration at this time.

Domain orchestrators and the multi-domain service orchestrator (MDSO) are as stateless as possible, with inventory systems being entrusted with resource state information.

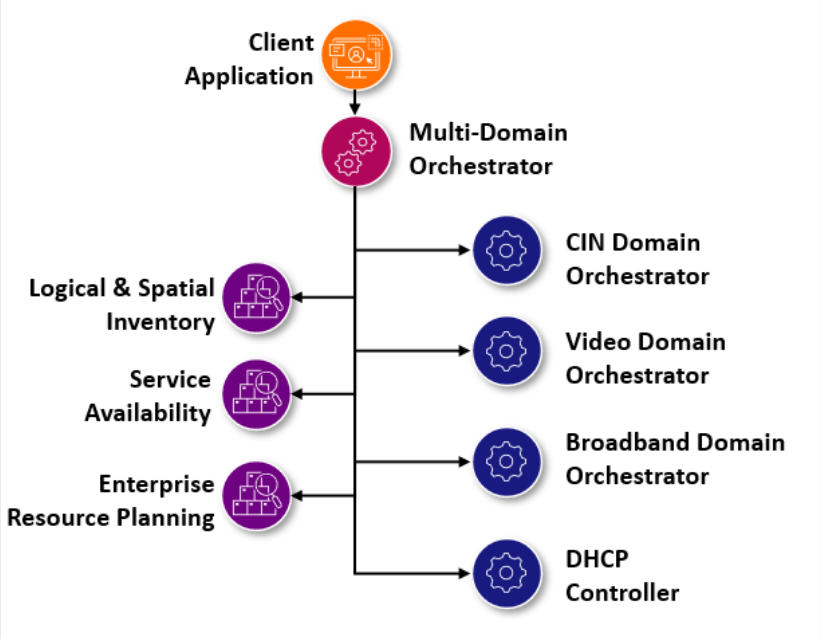


Figure 9: Orchestration System Interactions

This architecture (see Figure 9) is function based, and therefore tool agnostic. Some of the platforms used are:

1. Client Application - Vaadin, Quarkus running on Openshift.[41] [31]

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2. Multi-Domain Orchestrator – HPE Service Director running on Openshift and deployed with ArgoCD and Tekton. [12][31][1][38]

3. CIN Domain Orchestrator – Cisco Network Service Orchestrator[28]

4. DHCP Controller – customer interface into Cisco CNR[8]

5. Broadband Domain Orchestrator – custom interface running on Kubernetes[21]

6. Logical & Spatial Inventory – Netcracker RI and SpatialNet[26][34]

7. End to End Testing – SOAPUI[33]

A minimalistic client application (see Figure 10) serves as a conduit for initiating either configuration or provisioning activities.

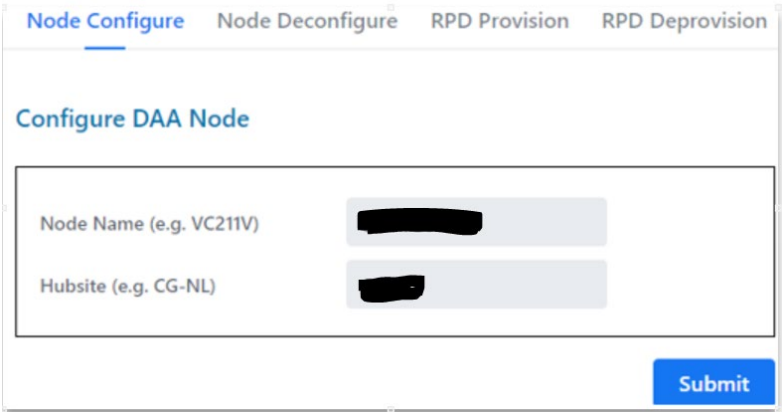


Figure 10: Auto configuration UI

The "configure node" function solicits for the new node name, and hub site location. These details facilitate the retrieval of spatial, RF spectrum and logical inventory data and automation of design and configuration tasks via the MDSO.

The "provision node" function seeks input of the new node name and either serial number or MAC address. These details activate the RPD by associating the MAC address to relevant equipment.

The multi-domain orchestrator functions as a vigilant overseer, tracking errors and the status of each domain function through the orchestration process.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2.1. Auto Configuration Orchestration

The R-PHY fiber node configuration will configure the node and make it available for the R-PHY fiber node device to provision. This involves creating the circuit between node and the CIN device followed by assigning the node to the CIN port and assigning the CCAP device to the R-PHY fiber node. After the end of this process, the CIN port assignment status will be set to 'Reserved' and the node status will be set to 'Planned' in inventory.

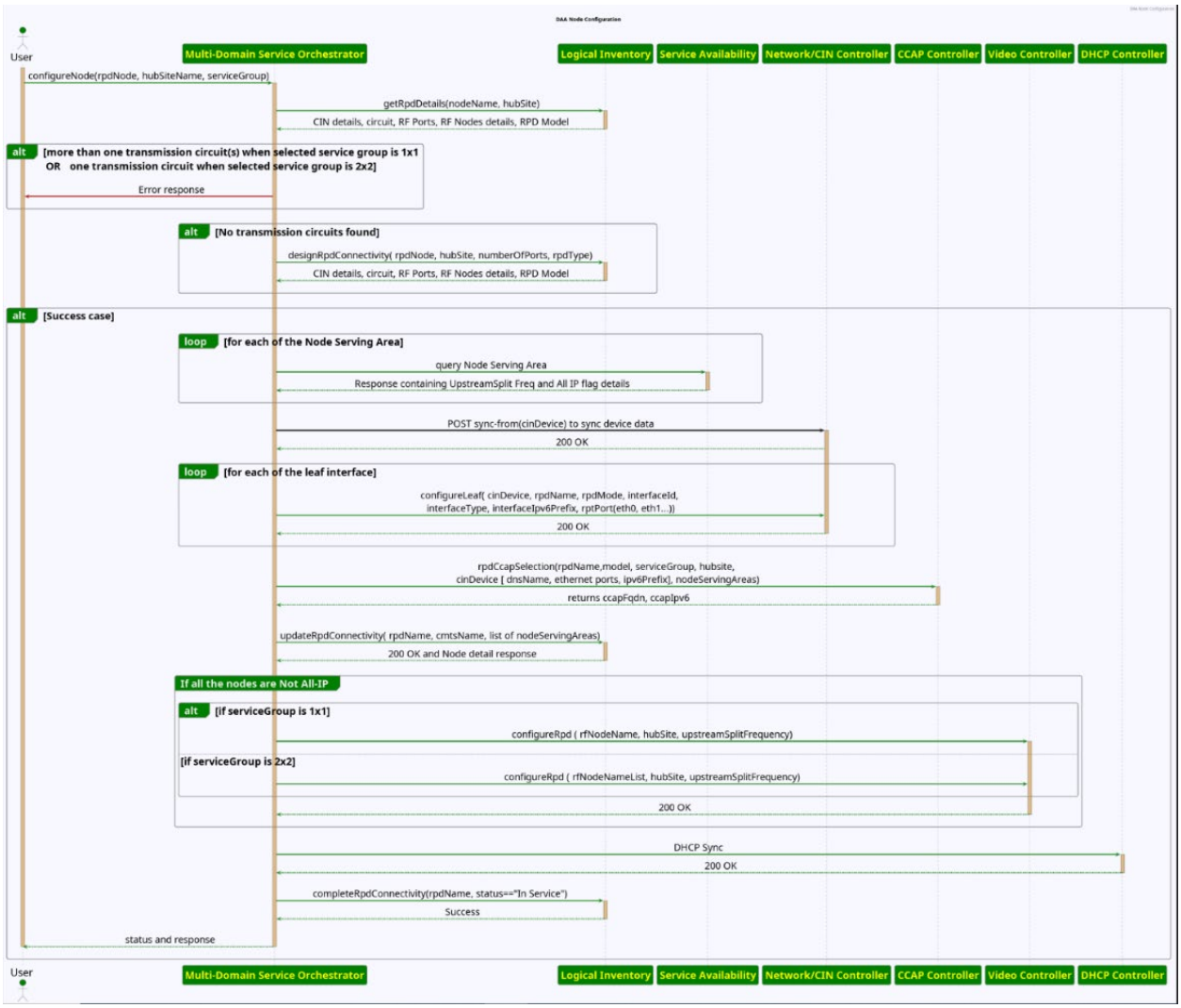


Figure 11: Autoconfiguration Sequence

This task replaces the manual effort (see Figure 11)required to:

1. Create the node in our logical inventory system.
2. Add the node to video on demand.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Configures the leaf port with ipv6, midv2, multicast and DHCP.

4. Reserves a cable-mac on the CMTS for the R-PHY fiber node.

5. Updates the DHCP server.

This task enables field technicians to start and complete their work to connect fiber cables to the CIN and install the physical R-PHY fiber node device.

4.2.2. Provisioning Orchestration

Provisioning is the last task before the R-PHY fiber node can be activated. This automation associates the MAC address of the R-PHY fiber node with the network.

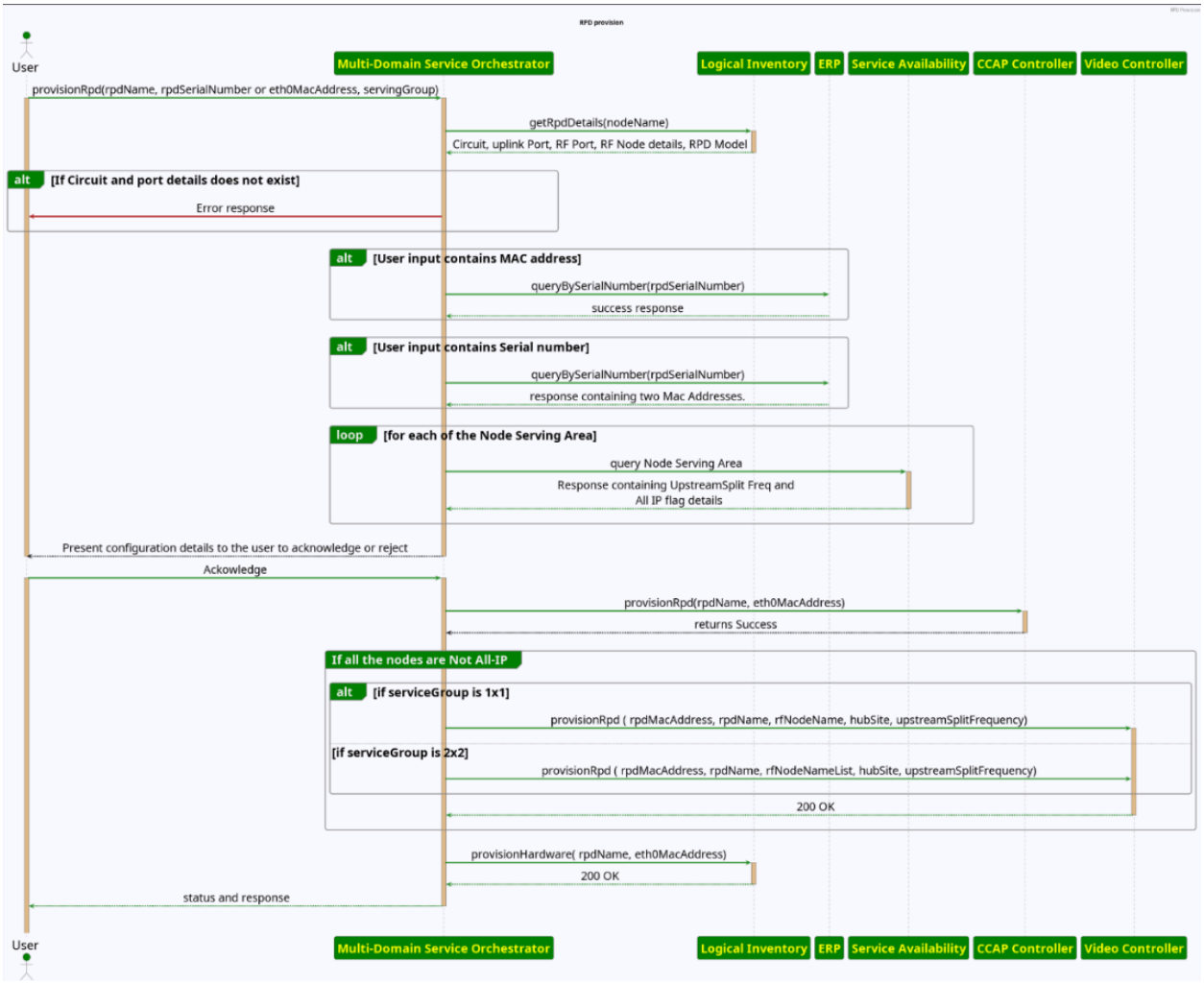


Figure 12: Provisioning sequence

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

This task (see Figure 12) replaces the manual workflow for the following:

1. Get MAC Address of R-PHY fiber node from serial number (SN) located in enterprise resource planning (ERP) (used for video core).
2. Get Circuit and R-PHY fiber node details from logical inventory.
3. Get video service availability details (all IP or legacy).
4. Provisions the R-PHY fiber node on the CCAP Chassis.
5. Provisions the video core.
6. Updates Logical Inventory with MAC to R-PHY fiber node relationship & label it as provisioned.

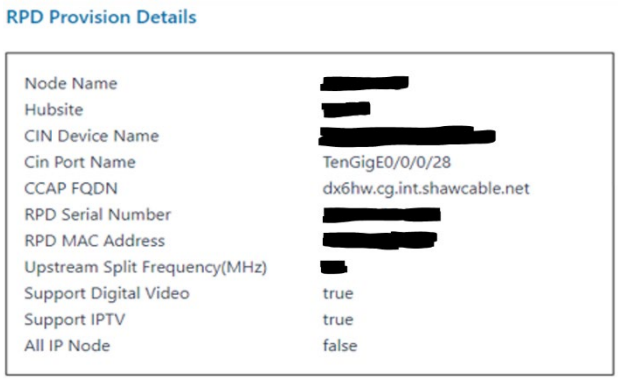


Figure 13:Output of Provision

The above output (see Figure 13) describes key variables and inventory information used to automate fiber node provisioning.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Results

5.1. Scaling R-PHY Fiber Node Additions Over Time

Across the past three years, a noteworthy milestone has been that the inclusion of 1000 new R-PHY devices was achieved. This progressive achievement has realized at a monthly rate spanning from 20 to 50 R-PHY nodes per month.

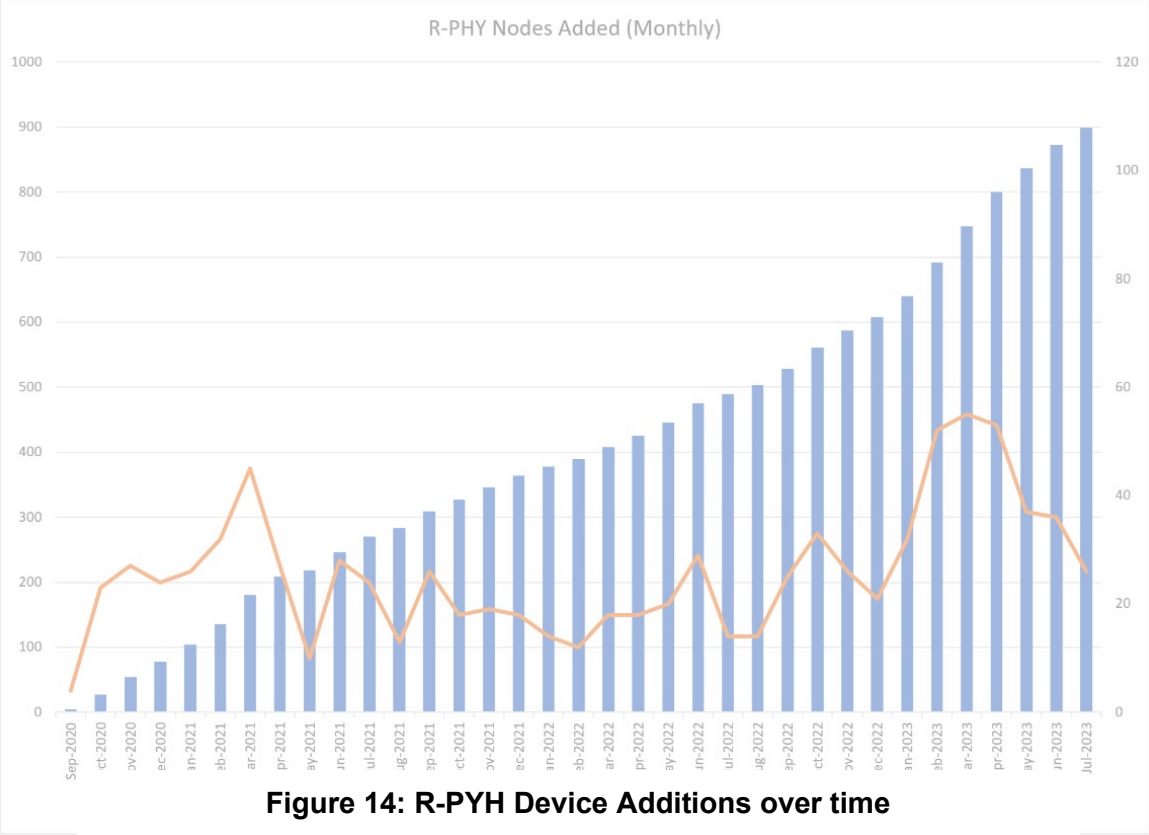


Figure 14: R-PYH Device Additions over time

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.2. Accelerated Node Activation Time

A remarkable transformation has unfolded in terms of the time taken for R-PHY node activation. The following chart describes the comparison of deploying a R-PHY node with automation, estimated R-PHY node without automation, and a traditional analog fiber node.

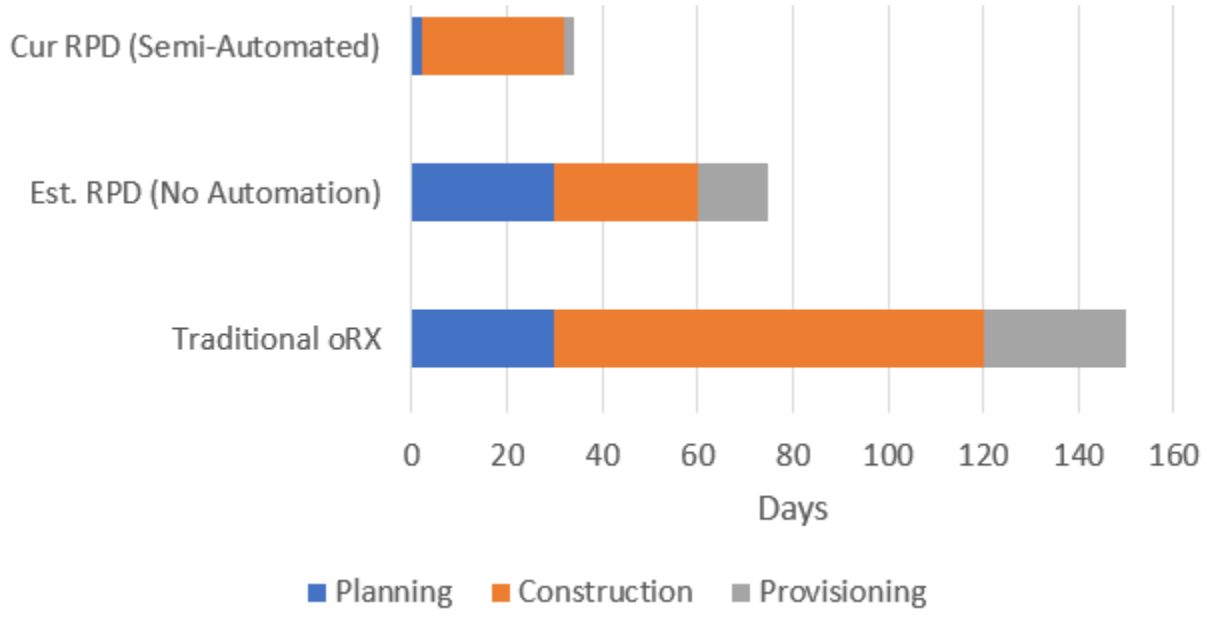


Figure 15:Time to Activate a Node

Presently, the automated "new R-PHY fiber node" can conclude within a matter of days, contingent on the intricacies of construction.

5.3. Errors in Auto-Configuration & Provisioning

Initial stages witnessed a notable frequency of transactional errors, surpassing the tally of successful instances. This trend primarily stemmed from delayed, absent, or inaccurate logical and spatial inventory. In response several critical lessons emerged:

1. Shifting from traditional analog fiber nodes, automation necessitates documentation ahead of installation. This transition ensures inventory data is preemptively available for automation.

2. Delays stemming from inventory systems, taking up to 24 hours for updates, prompted a realization that provisioning must transition to a near-real-time paradigm, warranting the updating of back-office systems.

3. A minor number of errors were attributed to unique use-cases. Continued focus on resolving these through continual improvement ensures stability.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

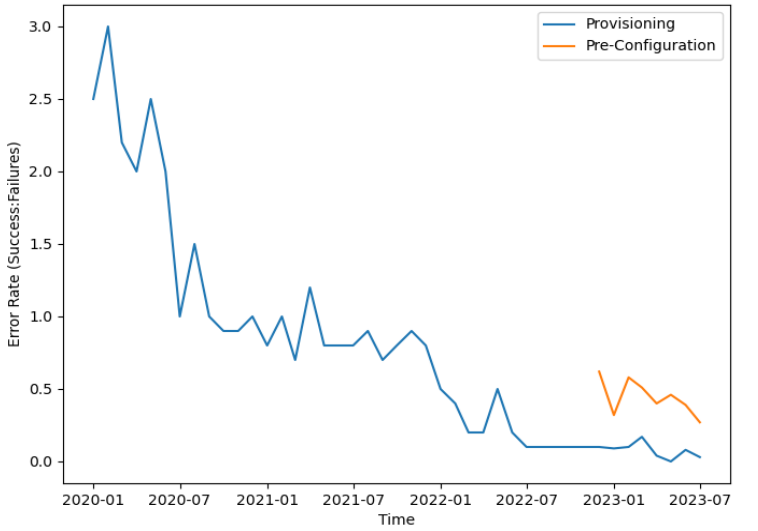


Figure 16:Error Rate

Subsequent efforts have led to a significant reduction in errors, a trajectory we endeavor to sustain while simultaneously introducing new features within the solution.

5.4. Complexity and Level of Effort

It is difficult to imagine the magnitude of effort required to maintain, support, and expand the DAA solution. To appreciate this, we undertake a comparative analysis between the magnitude of configuration lines on managed devices and the manual documentation and task-team requirements for deploying new R-PHY fiber nodes.

Through orchestration, dedication to inventory management, process mapping, and domain automation we have a highly sustainable process. In lieu of managing approximately 800+ lines of configuration within various systems, we have distilled the process down to a mere handful of parameters.

This optimization, achieved through streamlined processes and automation has resulted in a notable reduction in the effort required for operational continuity. This has been observed by reducing the amount of team required to launch a new node from approximately 25 to 7 and the number of manual tasks from approximately 70 to 20.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Reflections

During each iteration feedback was obtained from the program and documented in retrospectives & lessons learned. Retrospectives can be categorized under the following themes:

6.1. Positive Collaboration and Alignment Feedback

The orchestration of R-PHY devices engaged a multitude of teams: from planning, project and fiber coordination to IP teams, OSS, network technicians, video and more. This intricate ensemble necessitated a robust focus on collaboration.

Key elements encompassed fostering trust, defining roles and responsibilities, and cultivating an environment where the coalescence of efforts led to successful deployments, swift problem resolution, and a sense of meaningful contribution.

Vision and goal setting played a pivotal role by establishing the trajectory and minimizing tangential pursuits. Steering and working committees further facilitated efficient prioritization, enhancing alignment, and maintaining project trajectory.

6.2. Continuous Improvement and Feedback Culture

The culture of continuous improvement, underpinned by a keen responsiveness to feedback, resulted in the delivery of incremental successes, heightened agility, and augmented confidence in outcomes.

The iterative cycles and milestones offered scrutiny of each phase's insights and celebrations of collective accomplishments.

6.3. Adherence to Industry Standards and Models

Leveraging industry standards and models proved instrumental in expediting delivery and integration endeavors. However, lessons emerged from the choice of "custom models" due to the maturity of some models at the time resulted in additional effort down the road. In retrospect, embracing and maturing emerging standards may facilitated a smoother progression.

Additionally, the pursuit of transactional automation underscored the need to overcome the limitations of CLI or SNMP, promoting the evolution of the OSSI model [30] to enhance orchestration in the DOCSIS technology sphere.

6.4. Influence of Conway's Law

The resonance of Conway's Law, "the design of a system reflects the communication structure of the organization," often subtly unfolds in project dynamics.

Initial milestones emphasized technology and process, deferring user experience considerations to subsequent iterations which resulted in avoiding negative consequences of Conway's Law.[10]

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6.5. Reliability of Automation Systems

As software systems encompass more systems and integrations, the probability of low availability increases. This concern held sway during the initial iterations, fueling proactive measures.

Optimizing start-up and recovery times, coupling & cohesion strategies and scrutinizing individual transactions errors emerged as successful strategies to mitigating this concern.

It is crucial to control the number of dependent systems in orchestration chains to maintain reliability expectations.

6.6. Foundations and Essentials

Initial stages underscored the significance of comprehending the "DAA Solution" before automating it. Having a working network solution in place before embarking on any automation proved invaluable. This approach furnished rapid support for DAA initiatives, magnifying the importance of an established groundwork.

Early architectural insights played a pivotal role by identifying gaps, API contracts, and sequence diagrams, culminating in expedited implementation.

7. Concluding Insights

In culmination, the collaborative efforts of diverse teams have positioned us well to support next generation technologies and services. Automation has emerged as our ally, propelling the swiftness and consistency of R-PHY fiber node additions to new heights while ensuring unwavering predictability and reliability.

However, as automation becomes more widespread, it is also becoming apparent that the technology that enables it is in many ways the easiest part.

An effective automation initiative is overwhelmingly based on people—including culture, process, capabilities, and skill sets.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

API	application programming interface
CCAP	converged cable access platform
CD	continuous delivery (not deployment)
CI	continuous integration
CIN	converged interconnect network
CLI	command line interface
CMTS	cable modem termination system
COTS	commercial off the shelf
DAA	distributed access architecture
D-CCAP	distributed converged cable access platform
DOCSIS	data over cable service interface specification
EMS	element management system
ESD	extended spectrum DOCSIS
FDX	full duplex
Heat	cloud formation declarative based template from OpenStack
HFC	hybrid fiber / coax network
HOT	heat orchestration templates
I-CCAP	integrated converged cable access platform
IP	internet protocol
MAC	media access control
MIB	management information base
MOP	method of procedure
NED	Network element driver
NETCONF	network configuration protocol
NOC	network operations center
OMS	order management system
ORX	optical receiver
OSS	operational support system
PHY	physical layer
PTP	precision time protocol
REST	representational state transfer
RESTCONF	restful network configuration protocol
RF	radio frequency
R-PHY fiber node	remote physical layer device
SNMP	simple network management protocol
SOAP	simple object access protocol
SOM	service order management
SSH	secure shell
SSL	secure sockets layer
TLS	transport layer security
TOSCA	topology and orchestration specification for cloud applications
VOD	video on demand
YAML	yet another markup language
YANG	yet another next generation

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

[1] Argo CD (Argo CD | Argo (argoproj.github.io))

[2] CableLabs DOCSIS DCA-MHAv2, Remote PHY Specification, CM-SP-R-PHY-I12-190307,

[3] CableLabs Flexible MAC Architecture System Specification, CM-SP-FMA-SYS-I03-220126.

[4] CableLabs FMA MAC Manager Interface Specification, CM-SP-FMA-MMI-I03-220126.

[5] CableLabs FMA OSS Interface Specification, CM-SP-FMA-OSSI-I02-220602

[6] CableLabs R-PHY Node Model (http://mibs.cablelabs.com/YANG/DOCSIS/R-PHY/)

[7] Cloud Native (www.cncf.io)

[8] Cisco CNR DHCP (Cisco Network Registrar - Cisco)

[9] Cognitive Load Theory (The Evolution of Cognitive Load Theory and the Measurement of Its Intrinsic, Extraneous and Germane Loads: A Review (tudublin.ie))

[10] Conway's Law on Organizational Structure (Conway's law - Wikipedia)

[11] Heat Orchestration Template (HOT) specification (Heat Orchestration Template (HOT))

[12] HPE Service Director (Telecom Operations, Network Automation, HPE)

[13] IETF RFC 6020, YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF), October 2010.

[14] IETF RFC 7950, The YANG 1.1 Data Modeling Language, August 2016.

[15] IETF RFC 7951, JSON Encoding of Data Modeled with YANG, August 2016.

[16] IETF RFC 7952, Defining and Using Metadata with YANG, August 2016.

[17] IETF RFC 8040, RESTCONF Protocol, January 2017.

[18] IETF RFC 9195, A File Format for YANG Instance Data, February 2022.

[19] ITU-T, "Precision time protocol telecom profile for frequency synchronization", ITU-T Recommendation. 8265.1, July 2014.

[20] Jenkins Pipeline Automation (Jenkins)

[21] Kubernetes (Kubernetes)

[22] Lean Methodologies (Lean - Wikipedia)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

[23] Legacy & Video Core Automation (Paper - Leveraging Legacy Video in Digital Access Architecture Networks - NCTA Technical Papers)

[24] Low Code Methods (What is Low-Code? | IBM)

[25] Mockito Framework (Mockito framework site)

[26] Netcracker Resource Inventory (Netcracker - Home)

[27] Node Provisioning and Management in DAA (Paper - Node Provisioning and Management in DAA - NCTA Technical Papers)

[28] NSO Network Service Orchestrator (Cisco Network Services Orchestrator - Cisco)

[29] OpenAPI Initiative (Home - OpenAPI Initiative (openapis.org))

[30] OSSI CableLabs (Specifications Search – CableLabs))

[31] RedHat Openshift (Red Hat OpenShift Container Platform)

[32] RESTAssured Test Services (REST Assured (rest-assured.io))

[33] SOAPUI Test Suite (ReadyAPI | SmartBear Software)

[34] SpatialNet (Synchronoss - Operator Branded Cloud and Messaging Experiences)

[35] Software Engineering, cognitive load (Dave Farley's Weblog | Thoughts on Continuous Delivery and Agile development.)

[36] Swagger Specification (API Documentation & Design Tools for Teams | Swagger)

[37] Team Topologies Regarding cognitive loading (Team Topologies and Team Cognitive Load Assessment — Team Topologies)

[38] Tekton CI/CD (Tekton)

[39] Test Pyramid: The Practical Test Pyramid (martinfowler.com))

[40] Twelve Factor App (The Twelve-Factor App (12factor.net))

[41] Vaadin UI (Vaadin | Discover the Web App Framework Built for Java)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Remote PHY Going the Distance

A Technical Paper prepared for SCTE by

Marek Hajduczenia, PhD
Network Architect, Distinguished Engineer
Charter Communications
14810 Grasslands Dr, Englewood, CO 80112
marek.hajduczenia@charter.com

Glen Hardin
VP, Technical Fellow
Charter Communications
14810 Grasslands Dr, Englewood, CO 80112
Glen.Hardin@charter.com

John Chapman
CTO Broadband & Fellow
Cisco Systems
300 East Tasman Drive, San Jose, CA 95134
Glen.Hardin@charter.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. RPHY Background 4
2.1. A Brief History 4
2.2. How RPHY Manages DOCSIS REQ-GNT Delays 5
2.3. DOCSIS Latency Measurement (DLM) 6
2.4. DOCSIS Predictive Scheduling (DPS) 7
3. RPHY Service Architecture 7
3.1. Service Delivery Model 8
3.1.1. RPD Placement: Option 1 10
3.1.2. RPD Placement: Option 2 11
3.1.3. RPD Placement: Option 3 13
3.1.4. vCMTS Placement: Centralized 15
3.1.5. vCMTS Placement: Distributed 16
3.2. PTP Architecture 17
3.3. RPD Spectrum Plans 19
4. RPHY Performance Measurement Scenarios 20
4.1. Impact of Network Latency on HS Services 23
3.2 Impact of Network Latency on LS Services 24
3.3 Impact of PTP Path Pinning 24
5. Conclusions 28
6. Appendix I – Historical Calculations on REQ-GNT Delays 29

Abbreviations 31

Definitions 31

Bibliography and References 32

List of Figures

Title **Page Number**

Figure 1 – RPHY and DOCSIS Delays 6
Figure 2: RPHY service architecture 7
Figure 3: High-level service architecture used in the Charter network 10
Figure 4: Option 1: Migration from iCMTS to vCMTS with hub-based RPDs 11
Figure 5: Option 2: Migration from iCMTS to vCMTS with node-based RPDs 12
Figure 6: Option 3: In-building RPD-fed coaxial distribution plant 14
Figure 7: In-building RPD deployed with NID 15
Figure 8: Regional and national DC locations for centralized vCMTS placement 16
Figure 9: Distributed vCMTS placement 17
Figure 10: PTP distribution architecture 18
Figure 11: Low-split RF plan with FC+ analog video services 19
Figure 12: High-split RF plan with no analog video services 19
Figure 13: MER for test location in production network 20
Figure 14: Impairment generator, latency / PDV added to data path to all IP traffic 21
Figure 15: Impairment generator, latency / PDV added to data path to PTP traffic only 21

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 16: Test topology with in-line impairment generator 22
Figure 17: Asymmetric routing problem 23
Figure 18: Upstream and downstream HS throughput with 0, 100, and 700 miles network asymmetry 25
Figure 19: Upstream and downstream LS throughput with 0, 100, and 700 miles network asymmetry 26
Figure 20: Upstream data throughput with PTP path pinning, HS service plan 27

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Why RPHY go the distance?

The answer to this question has a dramatic impact on the plant topology, network implementation, compute scaling and, most importantly, the operational performance of the overall system. One of the cost reduction opportunities is in the consolidation of HFC hub operations. HFC hubs can be consolidated, simplified, or even completely decommissioned by transitioning traditional integrated CMTS (iCMTS) platforms to virtualization solutions. This drives the opportunity for the complete hub-collapse and aggregating more customers into larger metro sites.

Charter has been evaluating different technologies with Remote PHY (also referred to as R-PHY, R PHY, or simply RPHY) being one of those. Charter also examined the distance question of RPHY design in detail from a theoretical and practical perspective. While there are general statements to that effect, the relationship between specific distances and performance has not been identified to date.

Spectrum Enterprise has been deploying vCMTS with subtended video cores and RPHY devices for years with many of its hospitality clients. The commercial application of RPHY has given Charter real-world data of actual deployments. This data is very relevant to the residential application for data services. Additionally, significant studies have been conducted using impairment generators in trial deployments, validating the impact of symmetric and asymmetric latency on the data throughput, video services, etc.

This paper explains the challenge at hand, identifying key technical risks and limitations, and looks at the lessons learned from the existing video centric RPHY deployments within the Charter network. Details on the internal studies with impairment generator are also presented, exploring the limits of RPHY architecture performance, providing a much more positive picture of the system-level performance.

2. RPHY Background

RPHY is the new and more future-proof HFC architecture. RPHY replaces the previous generation of analog fiber with digital IP based fiber and moves the function of generating RF spectrum to the fiber node.

2.1. A Brief History

The RPHY architecture was invented by John Chapman of Cisco in 2001 [1] [7] [8] [9] and was originally part of the Modular Headend Architecture (MHA). The main protocols required for the operation of RPHY include:

- **R-DEPI**, the remote downstream external PHY interface, a pseudowire that connects the CMTS Core to the RPD [2],
- **R-UEPI**, the remote upstream external PHY interface, a pseudowire that connects the RPD upstream to the CMTS Core [3],
- **GCP**, the generic control plane, that is the management protocol for the RPD [4], and
- **R-DTI**, the remote DOCSIS Timing Interface, that uses IEEE 1588 to manage the DOCSIS timestamp between the CMTS Core and the RPD [5].

DEPI, UEPI, and DTI were written in 2004. GCP was written in 2013 at a Pete's coffee shop. DEPI and DTI were standardized as part of MHAv1 at CableLabs in 2005 [10]. UEPI and GCP, along with a

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

"Remote" prefix, were standardized at CableLabs as part of CDOCSIS in 2014 [11] and MHAv2 in 2015 [12].

2.2. How RPHY Manages DOCSIS REQ-GNT Delays

All modern access systems, such as WiFi6, mobile 5G, PON, and DOCSIS, use a scheduled upstream. They all work in a similar manner. The DOCSIS system coupled with a RPHY system is shown in Figure 1.

The delays of this system were analyzed in extensive detail during the MHAv1 process. The delays, based on 2004 CMTS technology, are described in the MHAv1 DEPI specification, Appendix I, "DEPI and DOCSIS System Performance" [13] and in Appendix I of this white paper.

The first thing to be mentioned is that all the delays shown, apart from the three RPD queueing delays, are common to RPHY, RMACPHY, and I-CMTS systems. That is because they are DOCSIS delays. The REQ and MAP get special processing in the RPD that allows them to skip the data queue which provides very little delay. The RPD upstream data queuing is bringing a packet from a 100 Mbps DOCSIS interface to a 10 Gbps Ethernet interface, so it also has negligible delay.

Let's spend a moment looking at DOCSIS delays. Let's also assume that MAP messages are sent every 2 milliseconds on average. When a packet arrives at the CM, it is processed and then it has to wait up to a full MAP time for a REQ slot. When the REQ arrives at the CMTS, it is processed and put into a scheduler queue where it waits for up to a full MAP time for the scheduling process to initiate. A MAP is then built and sent to the CM.

The CM needs to receive the MAP in advance of when it needs to use it. The CMTS does this by creating MAP far enough in the future that the MAP will never arrive late. This is known as MAP advance. It includes engineering margin and can be as much as a MAP time. The MAP then has to play out and it can be up to a MAP time before the data packet is sent from the CM if it is at the end of the MAP. The CMTS then receives the data packet, processes it, and sends it out.

All in all, there are at least four MAP times chained together which creates a delay of 4 to 8 ms, in addition to PHY and queueing delays. Very fast CMTSs can get this time sequence down to 5 ms best case. On average, this delay is 12 ms, and if there is contention on the REQ channel, DOCSIS delays can quick add up to 50 ms. [15]

The additional delay from the RPD could be less than 0.25 ms, which is less than a 5% impact.

The summary of the delay analysis is that most of the request-grant delay in absorbed in the DOCSIS system itself. These measurements were shown in [15]. For systems less than the DOCSIS specification limit of 100 miles, the RPHY system adds less than 10% to the overall delay. In practice, since upstream delay is rounded off to MAP intervals, this additional delay may not even be seen in a short loop system.

The delay time in DOCSIS is much higher than RPHY, which makes RPHY nearly equivalent to DOCSIS.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

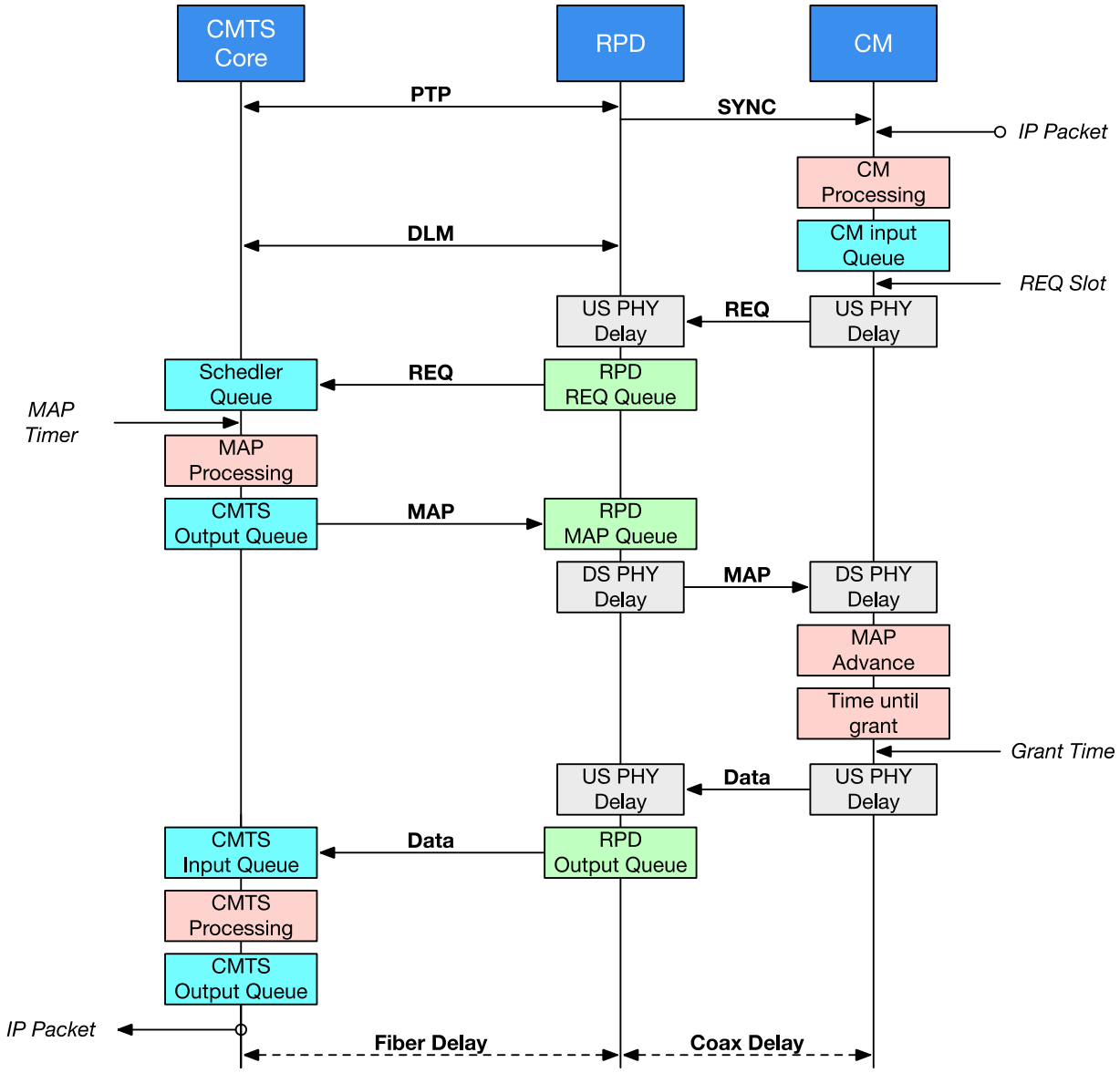


Figure 1 – RPHY and DOCSIS Delays

2.3. DOCSIS Latency Measurement (DLM)

The CMTS Core and the RPD are synchronized together to the reference PTP clock [16] [17]. This means that both the CMTS Core and the RPD are time-synchronized. The DEPI specification contains a special header operation called DLM that allows the timestamps to be recorded and compared at each point. The difference between the two is an accurate measurement of the delay of the fiber optic cable.

When the measured fiber delay is greater than the equivalent of 100 miles, the CMTS Core uses value from DLM to calculate a larger MAP advance time.

*It is DLM that allows the CMTS to make RPHY work with **almost any length of cable**.*

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.4. DOCSIS Predictive Scheduling (DPS)

DOCSIS predictive scheduling is a further enhancement that is not in the DOCSIS standard but is used by several CMTS vendors. DPS issues extra grants to active CMs, allowing these CMs to accomplish two objectives:

1. Send data packets from before receiving a schedule grant, or even before requesting upstream bandwidth.
2. Send request messages as a piggyback message and to not have to worry about contention in the system request slot. This will reduce both short-term latency and the long-tail latency.

Just one grant slot sent per MAP interval, or several intervals, to an active CM will low the upstream latency significantly.

DPS is a further performance enhancement that lowers the upstream latency to just a single delay of the cable length instead of tripling the cable length delay.

3. RPHY Service Architecture

RPHY is a type of distributed access architecture (DAA), whereby the physical DOCSIS layer is moved from the hub (typical location where CMTS is deployed) to the edge of access network, i.e., much closer to the end customers. In the RPHY architecture, the integrated CMTS (I-CMTS) is therefore divided into the physical layer, aggregation layer and routing/MAC layer, with each of these aforementioned layers mapped into new functional elements:

- the physical layer is implemented in the RPHY device (RPD),
- the aggregation layer is implemented using the IP/MPLS transport layer, and
- the routing/MAC layer is implemented using a virtual CMTS (vCMTS) software element running on a generic compute platform deeper in the carrier network.

This service architecture allows for the processing functions (schedulers, software management, bandwidth scheduling, etc.) to remain centralized, while the low-level DOCSIS physical layer processing is to be distributed in the most cost-effective manner, as shown in Figure 2. The use of the digital fiber-based transport IP/MPLS network between the centralized and distributed elements of this architecture substantially increases the reach of the DOCSIS vCMTS, eliminating the need for analog fiber deployment, analog fiber amplifiers, etc.

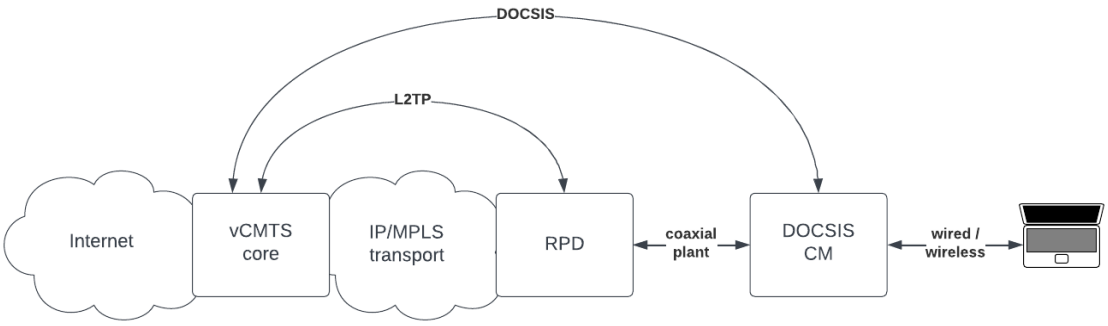


Figure 2: RPHY service architecture

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The majority of low-level physical layer DOCSIS processing is now concentrated within the RPD at the network edge, which plays two important roles:

- In the downstream direction, the RPD converts IP-encapsulated data received from the vCMTS core into analog signaling and transmits the resulting RF towards individual end-user devices (DOCSIS CMs, video receivers, etc.)
- In the upstream direction, the RPD performs the reverse process, converting any analog signals received from individual end-user devices (primarily DOCSIS CM) into IP-encapsulated digital data that is then transmitted towards the vCMTS core.

The RPHY service architecture is highly flexible in which individual RPDs can be configured independently from other RPDs in the same or other serving groups. This facilitates delivering traffic to and aggregating traffic from customers sharing the same service lineup. Through an efficient use of unicast and multicast L2TP data tunnels between individual RPDs and a vCMTS core, specific content can be distributed to multiple locations at the same time, taking full advantage of underlying IP multicast capabilities of modern transport networks.

This architecture does, however, have a strong timing dependency, whereby the vCMTS and individual RPD must be tightly synchronized to permit centralized schedulers to properly stack transmissions from individual end-user devices attached to the RF portion of the network. At this time, given the existing level of technical maturity of this technology, the timing synchronization aspect remains the only technical challenge for any operator planning to deploy the RPHY service architecture. Careful design of the timing distribution is required, as discussed in more detail in the following sections.

3.1. Service Delivery Model

As indicated before, the service delivery model relies on a distributed CMTS platform where the service processing and routing functions are centralized within the vCMTS and all the DOCSIS-specific media conversion functions (PHY) are pushed to the edge of the network (RPD). All communication between individual RPDs and the vCMTS cores takes place on the IP/MPLS network core (when vCMTS cores are deployed in the regional/metro DCs) or the hub-level converged interconnect network (CIN, when vCMTS cores are deployed at the hub level).

In this case, the RPHY service architecture shares the packet-switched network infrastructure with other data services, thereby taking advantage of the economies of scale in carrier networks. The high-level service architecture is shown in Figure 3 where the vCMTS cores are located in a market-centralized facility (marked as data center (DC) vCMTS) or at individual hub sites (marked as hub vCMTS), and individual RPDs are deployed at one of the possible locations:

- Option 1 (hub location): At the hub, representing a direct 1:1 replacement of iCMTS platforms
- Option 2 (field location): In the field (using a strand mount or street cabinet format), replacing one of amplifier nodes and shortening the amplifier cascade
- Option 3 (on-prem location): At the customer premises, connecting to intra-building coaxial distribution network

Individual options outlined above present the evolution of residential DOCSIS deployment models (Option 1 and Option 2) as well as different application scenarios (typical residential deployments – see Option 1 and Option 2, and community deployments – see Option 3). Individual options are discussed in more detail in the following subsections.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The placement of the vCMTS cores is operator-specific and depends on several factors, including:

- Aggregation scale required to achieve proper economies of scale
- Availability of data center (DC)-class facilities in the network and their location
- IP/MPLS transport network design and its capacity
- Acceptable splash radius in case of any DC outages and general risk aversion levels

Note that the decision to use a centralized (vCMTS deployed in regional DC locations) versus distributed (vCMTS deployed in the hub) delivery model does not necessarily mean that only a single vCMTS deployment schema is present in the given network. For any operator, it is likely that both the distributed and centralized models will be used for different applications.

The existing residential high-density deployments translate more naturally into a distributed, CIN-based vCMTS deployment with a high number of RPDs replacing existing iCMTS DOCSIS ports at the given hub or replacing individual amplifier nodes (RPD Option 1 or Option 2).

It is also possible that both the centralized and distributed vCMTS architectures eventually converge to a certain degree, sharing the compute infrastructure but following their own service models to best fit the target customer population. Charter's enterprise services more naturally follow the centralized model, with a much lower number of strategically deployed RPDs (RPD Option 3) aggregated into a single vCMTS location per market.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

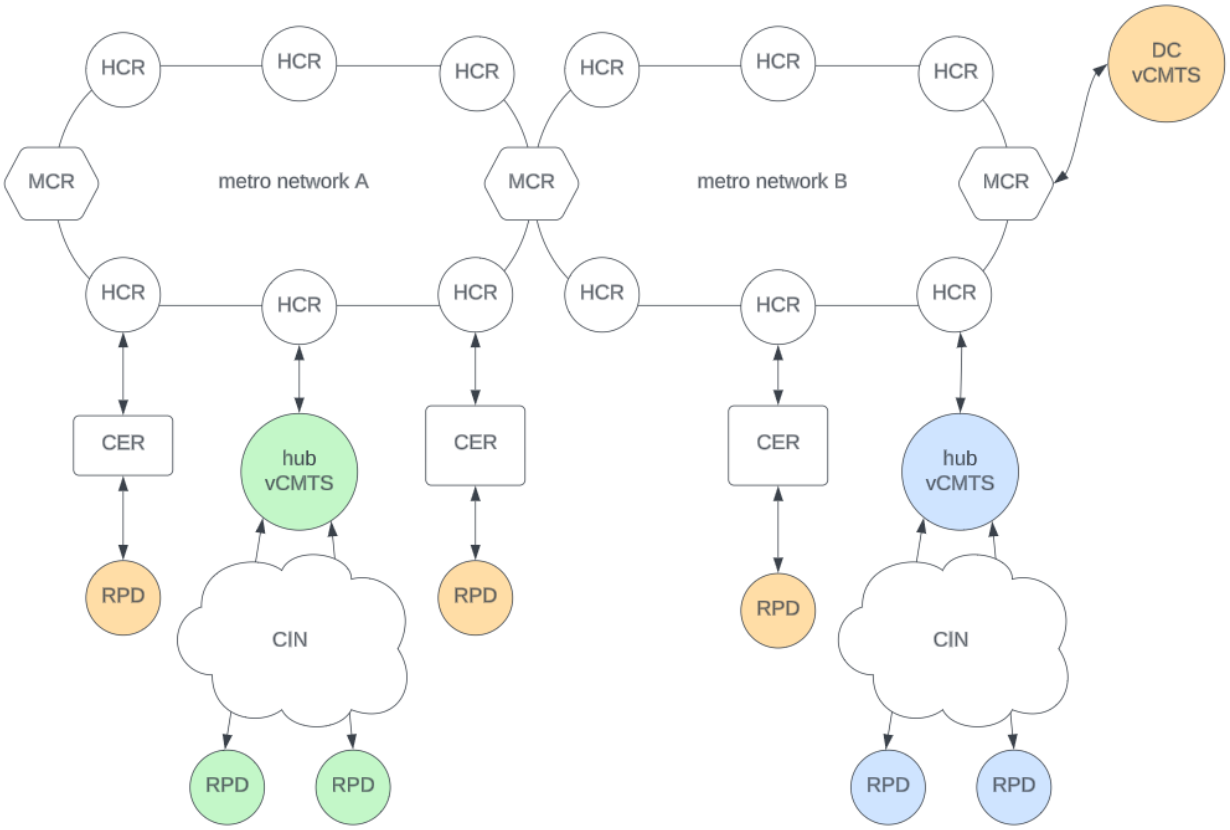


Figure 3: High-level service architecture used in the Charter network

3.1.1. RPD Placement: Option 1

This particular option represents a direct replacement for an iCMTS platform deployed today at the hub level, whereby the routing and MAC layer functions are extracted from the iCMTS chassis and re-deployed as the vCMTS either at the same hub or pushed deeper into the network, while individual RF iCMTS ports are replaced with RPDs connecting to field coaxial cabling, as shown in Figure 4.

While this option may seem a bit questionable in terms of operator advantages, it does provide a steppingstone towards the deployment of RPDs closer to the customer premises. This option presents a learning opportunity for field operations and network operation center (NOC) teams, extending the life span of the existing outdoor coaxial plant while transitioning towards a more scalable and distributed service architecture.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

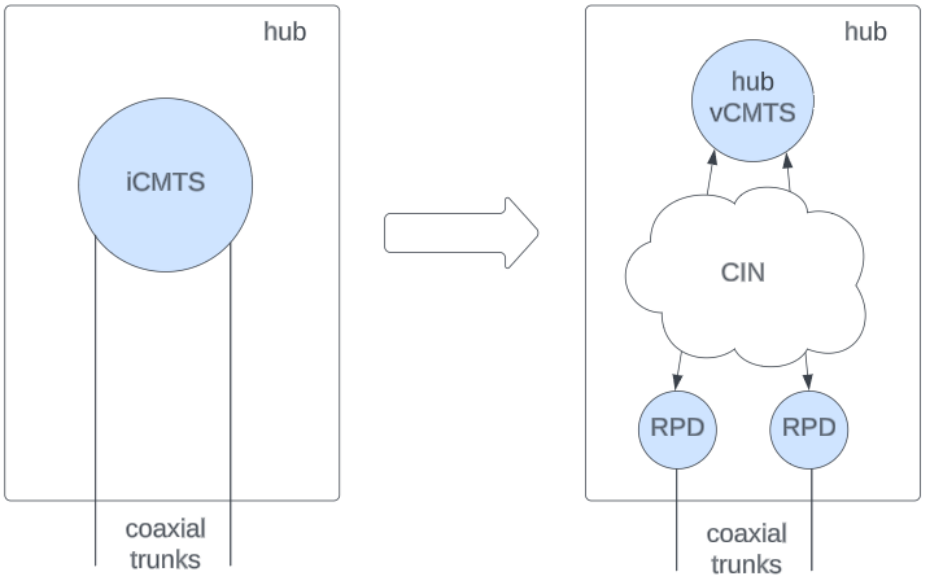


Figure 4: Option 1: Migration from iCMTS to vCMTS with hub-based RPDs

In the hub-based option, RPDs require only short distance digital optical connectivity to the hub network (CIN), resulting in none of the optical reach challenges characteristic of Option 2, as discussed in the next section. Also reducing the dependency on IP/MPLS network for connectivity.

3.1.2. RPD Placement: Option 2

Once the RPHY service architecture elements have been fully integrated into a regular network operational model, individual RPDs can be pushed further out into the field. The selection of which of the RF amplifiers in the node+N architecture gets replaced depends heavily on operator goals. A node+X (where X < N, i.e., the amplifier cascade becomes shorter) architecture eliminates some of the challenges associated with maintaining a large amplifier cascade (noise amplification and accumulation, pulse shaping, powering, management and everyday operations), but faces all the limitations of existing coaxial plant builds. The advantages and disadvantages of node+X architectures have been discussed in detail in literature already. Figure 5 shows an example of a migration from iCMTS with node+3 coaxial architecture into vCMTS with node RPDs and node+1 coaxial distribution network, providing all the advantages of shortened amplifier cascade and the use of digital feeding fibers.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

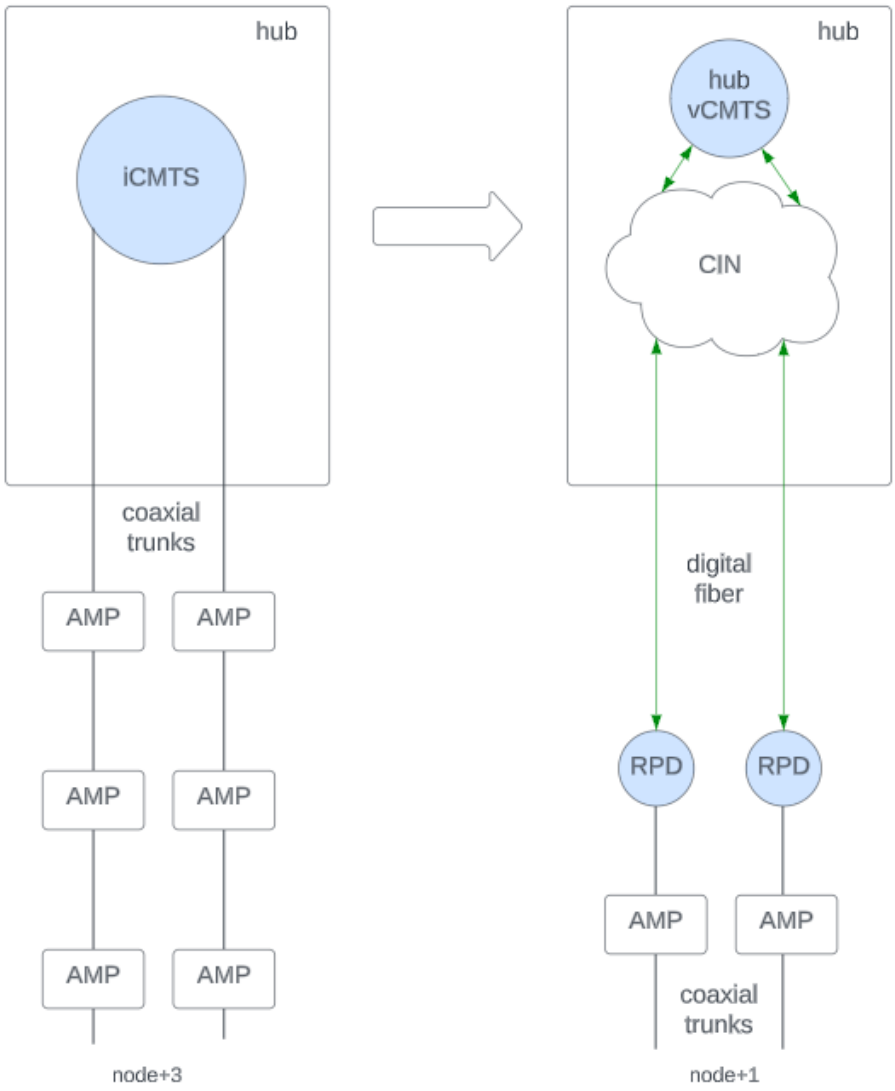


Figure 5: Option 2: Migration from iCMTS to vCMTS with node-based RPDs

It is important to note, though, that irrespective of the placement of the RPD (i.e., how many amplifiers remain in the coaxial distribution network), the RPD must be fed with a digital fiber using either a more common single mode fiber (SMF) pair with respective digital optics or a less common single SMF with the so-called bidirectional (commonly referred to as BiDi) optics. Where analog fiber is available, it can be easily converted into digital fiber by replacing active electronics at the ends of the fiber pair(s).

A single fiber path may be further reused for many RPDs by using standard DWDM optics, putting up to 20 bidirectional wavelength pairs on the same fiber path and using the methodology commonly used today in the enterprise-grade access networks. Depending on operator preferences, fixed DWDM or various types of tunable DWDM optics may be used.

Additionally, the bandwidth requirements per RPD also play an important role in the RPD deployment decisions. As DOCSIS 4 supports 10 Gbps data rates, the digital RPD backend must also support

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

compatible data rates. With future planned DOCSIS 4.0 deployments targeting 10/1 Gbps data rates, a single RF port RPD must support at least 10G Ethernet. RPDs supporting two or more RF ports must obviously support higher aggregate data rates on their digital backend.

Intensity modulated optics are distance-limited, with typical 10G Ethernet SMF optics operating up to 80 km and with specially designed long-reach variant achieving 120 km. Along with the increase in the data rate, the viable distance decreases though, with 25G Ethernet optics typically limited to 40 km, 50G Ethernet optics limited to 20 km and 100G Ethernet optics rates at 10 km. This problem can be solved through the use of coherent optics, supporting the reach of hundreds of kilometers, though at the cost of increased optical module size and power consumption.

Considering the cost-sensitive aspect of the RPHY service architecture and the evolutionary character of the RPHY deployment, it is expected that 10G Ethernet backhaul is the most viable and cost-effective option for RPDs today. The use of longer reach and higher power 120 km optics is possible, though that might present additional thermal challenges, especially in the case of RPDs deployed in street cabinets.

3.1.3. RPD Placement: Option 3

Option 3 represents an RPD deployment model specifically tailored to community services, whereby an RPD is deployed at the customer premises (a basement attachment point, telecommunication cabinet, etc.) and feeds in the intra-building coaxial network. A large number of student dorms, hotels, multi-tenant buildings, etc., are still wired with coaxial cabling, representing the lowest cost access medium to individual residential units, eliminating the need for costly and time-consuming indoor rewiring, as shown in Figure 6.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

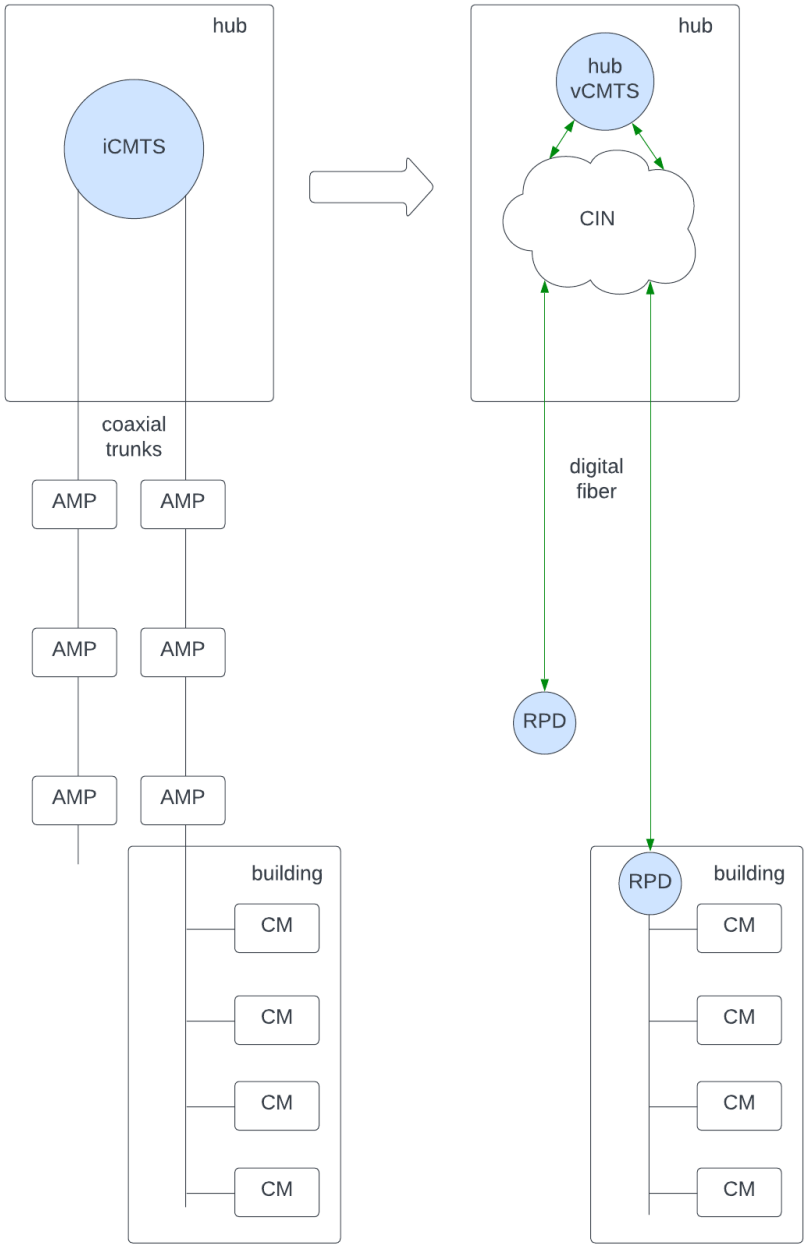


Figure 6: Option 3: In-building RPD-fed coaxial distribution plant

The challenges with the optical transport to individual RPDs as outlined in the description of Option 2 are still applicable in this case but may be circumvented to a certain degree through the deployment of a network interface device (NID), such as an aggregation router, feeding the RPD, as shown in Figure 7. In this case, the NID may implement more advanced coherent optics for extended reach and/or capacity in

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the digital section of the fiber (marked green in Figure 7), allowing for the use of lower cost gray optics in the RPD (marked red in Figure 7) and limiting the distance between the NID and the RPD to typical in-building scope of a few hundred feet at most. In the majority of deployments, the NID and RPD are physically collocated, allowing for the use of much cheaper and reliable short reach multi-mode optics.

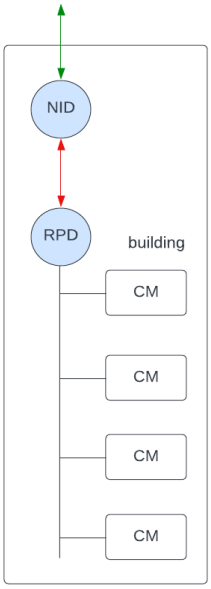


Figure 7: In-building RPD deployed with NID

3.1.4. vCMTS Placement: Centralized

In the centralized vCMTS deployment model, individual vCMTS cores are deployed in the regional or national data centers, as shown in Figure 8. In such a scenario, individual RPDs are deployed following one of the three aforementioned options, and the resulting L2TP sessions are then transported over the IP/MPLS transport network, covering metro, regional and national network segments, as applicable.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

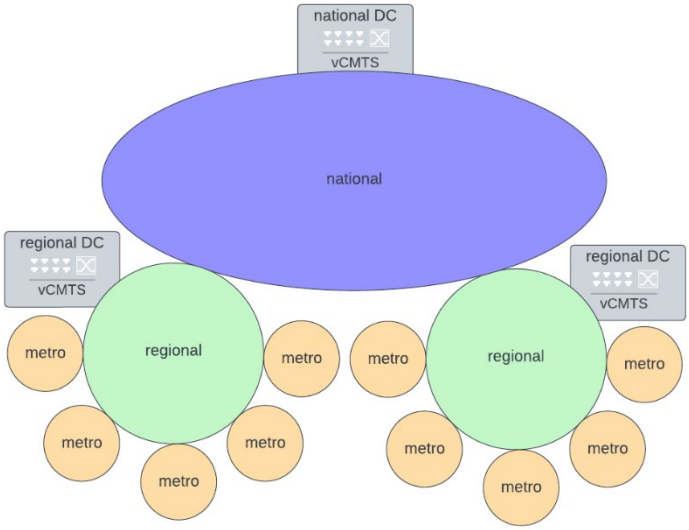


Figure 8: Regional and national DC locations for centralized vCMTS placement

The centralized vCMTS deployment model is best suited for low and medium density deployments of the RHY architecture, minimizing the vCMTS core cost and providing a deployment seed for this particular technology. Down the road, when the service density increases and a much larger number of vCMTS cores are required to serve the customer population, individual customers can be moved seamlessly to one of the distributed vCMTS units located closer to the network edge.

One of the drawbacks of this deployment model is the need to carry all the L2TP sessions from individual RPDs back to the centralized vCMTS cores, and thus aggregating a large volume of traffic at DC locations. Along with the increase in the number of supported RPDs, the aggregate volume of traffic ingressing and egressing the DC may become excessive, requiring transition to the distributed vCMTS deployment model covered in the following section.

Note that this deployment model does not affect the traffic flow in the network with any significance. Most networks are built around the hub-and-spoke architecture, carrying all Internet-bound traffic towards POP as effectively and quickly as possible. With the centralized vCMTS code deployment, such Internet-bound traffic is carried in L2TP sessions towards the target vCMTS core within the DC, where it is then decapsulated, and egresses the DC towards the typically co-located POP.

3.1.5. vCMTS Placement: Distributed

The distributed vCMTS placement model caters specifically to high-density customer populations, where a large number of vCMTS cores deployed at the hub limit the logical distance individual L2TP sessions have to travel over the IP/MPLS transport network to just the local CIN. L2TP is a non-zero overhead transport protocol and a very high volume of customer traffic transported over L2TP does have an impact on the network capacity. In the upstream direction where individual customer datagrams are typically small, the L2TP encapsulation results in a higher overhead when compared to the downstream direction.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

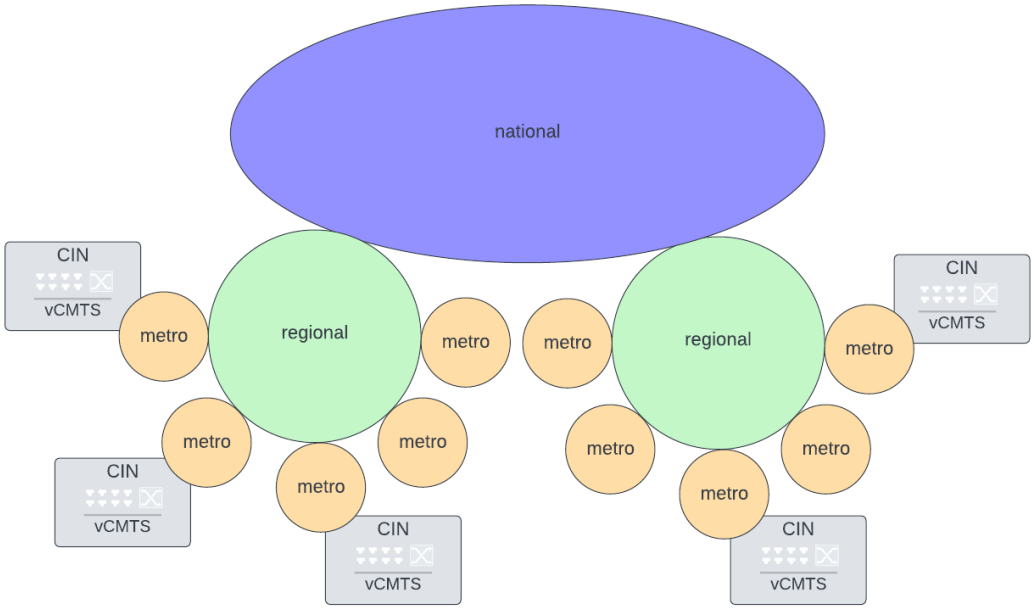


Figure 9: Distributed vCMTS placement

3.2. PTP Architecture

The Precision Time Protocol (PTP), standardized in IEEE Std 1588, defines an interoperable protocol for the distribution of phase, frequency and/or time of day information across various types of communication networks, including a range of specific network profiles. At Charter, PTP is distributed using the unicast IP encapsulation option (compliant with ITU-T G.8275.2 profile), traversing the IP/MPLS core network between the GPS-fed PTP Grandmaster clocks (GMC) attached to the metro core routers (MCRs), vCMTS cores collocated within the DC or local hubs (depending on the market and the delivered services), and RPDs deployed using one of the three options discussed before.

PTP traffic further traverses hub core routers (HCRs) and boundary clocks (BC) deployed where individual PTP domains need to be segmented to minimize the impact of packet delay variation (PDV) on the resulting PTP service quality. The high-level view of the aforementioned PTP architecture is shown in Figure 10.

Each metro network is equipped with two redundant GMC units, providing protected PTP sources for all PTP clients deployed within the given metro network. The PTP architecture comprises two layers, core PTP network (shown in red in Figure 10) and distribution PTP network (shown in green in Figure 10), providing a very high number of PTP clients without taxing GMC units with direct client connections.

Through the proper IP prefix allocation and reachability limitation via access lists and/or prefix visibility scoping, only PTP BC units are allowed to connect to GMC units within the given metro network. Individual PTP clients are permitted to connect to PTP BC units only, protecting GMC units from supporting thousands of PTP clients in large-scale deployments.

Additionally, PTP traffic is prohibited from traversing backbone networks (interconnecting individual metro networks, not shown in Figure 10), effectively constraining PTP traffic exclusively to the given

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

metro. This self-imposed limitation increases slightly the number of GMC units that need to be deployed but constrains PDV to much smaller values typical to metro networks, resulting in improved PTP performance. Moreover, backbone routers are typically more challenging to optimize for PTP transfer, even in the presence of properly designed CoS/QoS policies.

Furthermore, to minimize PDV for PTP traffic and optimize its performance across all metro networks, updates to network QoS/CoS policies are needed, prioritizing PTP traffic over any customer traffic by configuring individual routing platforms to put PTP traffic into real-time queues.

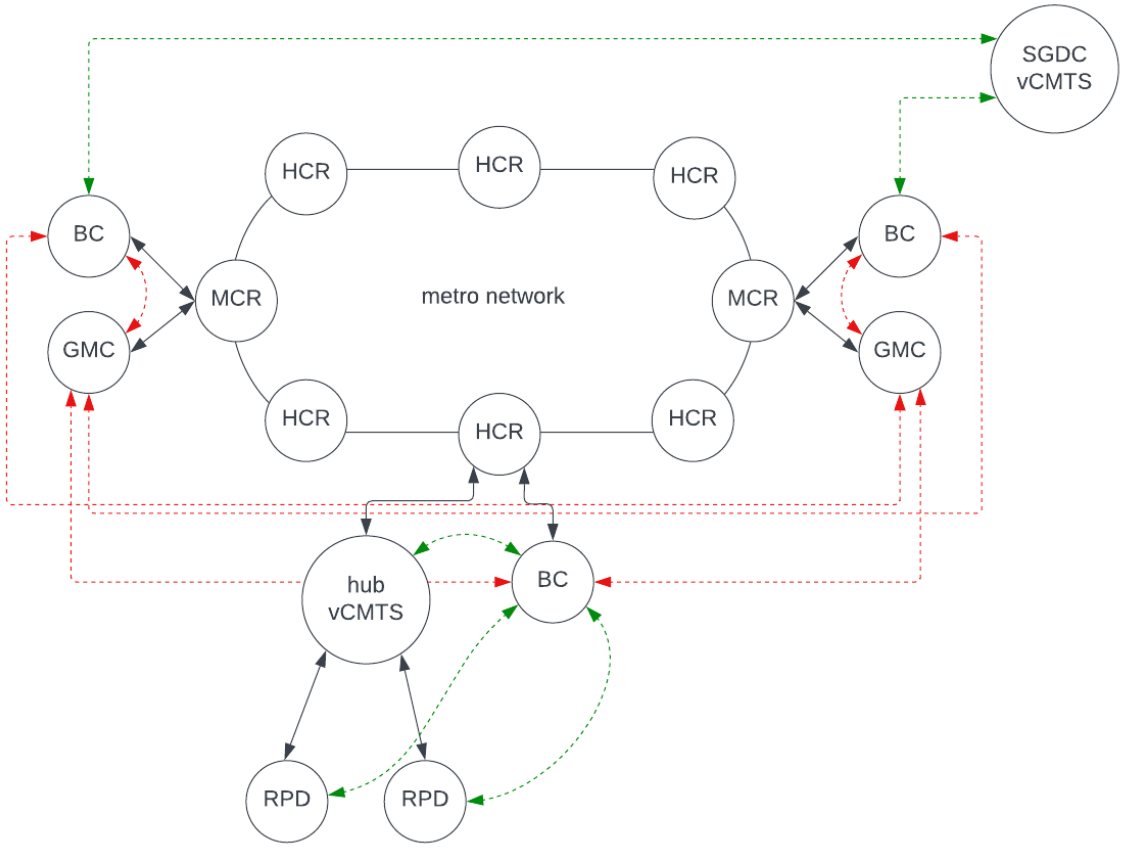


Figure 10: PTP distribution architecture

PTP traffic is critical for the proper operation of data (digital) services delivered using the RPHY architecture, where individual RPDs must be tightly synchronized (in terms of time of day as well as frequency, with phase synchronization not required) with the vCMTS core they are connected to. The synchronization is required, since the DOCSIS scheduler is hosted within the vCMTS cores deployed within DC locations, while the RPDs are deployed outside of DC locations, with the distance between vCMTS cores and individual RPDs ranging from a few miles to a few hundreds of miles in practical deployments. The said distance is primarily constrained by PDV that PTP traffic is subject to, as well as the network design and segmentation.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

PTP is not required for video services, given that typical RPDs are capable of operating in the so-called asynchronous mode, where there is only downstream traffic transmitted towards RPDs, requiring no synchronization between the RPD and the vCMTS core.

3.3. RPD Spectrum Plans

Each RPD can be configured independently with a specific RF spectrum plan, including any combination of digital (DOCSIS) and analog (video QAM) services. To keep things simple, a finite number of RPD RF spectrum plans are typically defined.

For Spectrum Enterprise applications, there are two primary RF spectrum plans in use: the low-split (LS) spectrum plan with Fiber Connect Plus (FC+) analog video services (see Figure 11) and the high-split (HS) spectrum plan with digital services only (no analog video, see Figure 12).

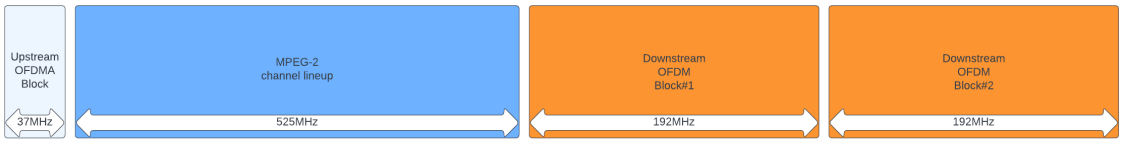


Figure 11: Low-split RF plan with FC+ analog video services

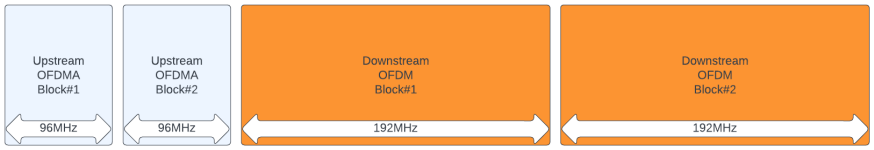


Figure 12: High-split RF plan with no analog video services

In comparison of a low split spectrum plan which delivers a complete analog video channel lineup (about 525 MHz worth) with MPEG2-encoded content, providing sufficient RF capacity for two downstream 192 MHz-wide OFDM blocks and a single reduced size (so-called *skinny*) upstream 37 MHz-wide OFDMA block, with no DOCSIS primary QAM channel. There is also no ATDMA in the upstream since all connected DOCSIS CMs are D3.1 compliant and there is no requirement for any backward compatibility at all. Apart from providing analog video services, this spectrum plan can theoretically provide around 2.5 Gbps downstream and around 180 Mbps upstream data rates, assuming 4k QAM constellations under sufficient modulation error ratio (MER).

The HS spectrum plan eliminates the analog video channel lineup, making much lower-frequency RF spectrum available for the upstream OFMDA blocks. Effectively, this spectrum plan provides sufficient RF capacity for two downstream 192 MHz-wide OFDM blocks and two full size upstream 96 MHz-wide OFDMA blocks. Therefore, this full digital spectrum plan provides around 2.5 Gbps downstream and around 1.2 Gbps upstream data rates, assuming 4k QAM constellations under sufficient MER (41 dB or more).

During testing, with properly engineered intra-building coaxial plant, much higher MER was achievable, reaching the average of 49 dB, as shown in Figure 13. Given the relatively well-controlled character of the intra-building coaxial plant, as well as much less challenging environmental conditions, maintaining the minimum MER required for 4K QAM constellation is not expected to be very complicated.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

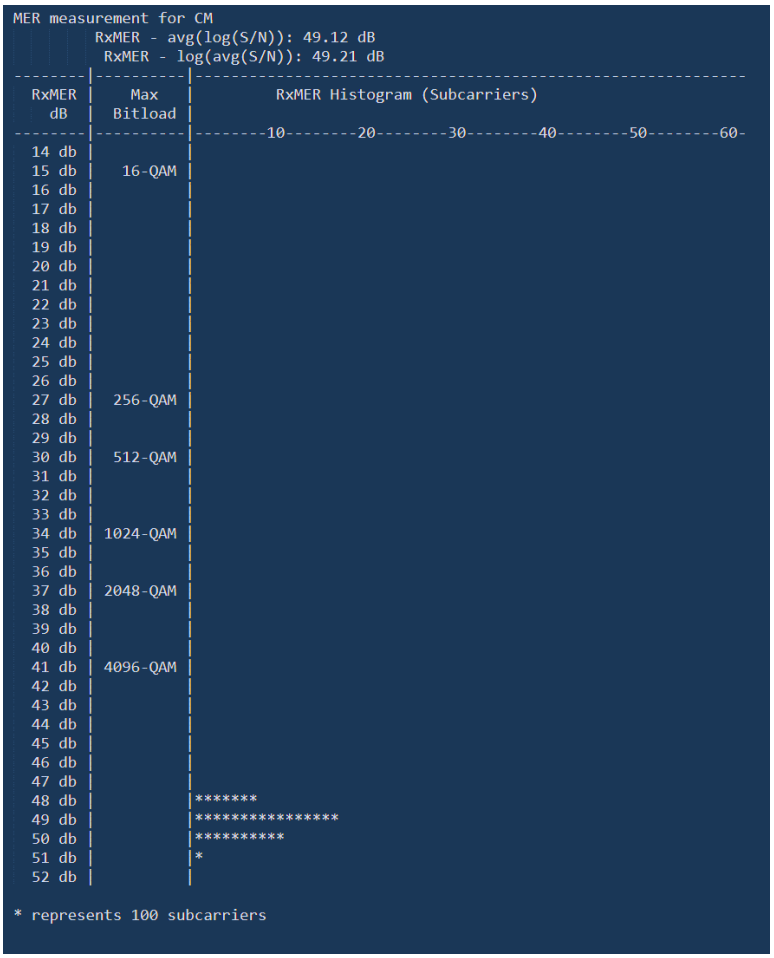


Figure 13: MER for test location in production network

The application of the specific RF spectrum plan in the given customer location (site) depends on the requirement to support the analog video service, whereby customers using analog TV receivers are provided with the LS spectrum plan and the other customers are defaulted to the HS spectrum plan. The main difference between these two spectrum plans lies, obviously, in the lack of support for analog video channel lineup in the HS spectrum plan as well as higher upstream capacity for digital services. The downstream capacity for both spectrum plans was designed to be exactly the same, utilizing the full capability of existing off-the-shelf DOCSIS 3.1 cable modems.

4. RPHY Performance Measurement Scenarios

All of the results presented in this paper were collected in a live production network where a test customer site with two RPDs was deployed with distinct RF plans: RPD-A with the low-split (LS) spectrum plan with FC+ analog video services, and RPD-B with the high-split (HS) spectrum plan with digital services only (no analog video). A vCMTS core used for these tests was deployed in the regional DC (centralized deployment) model and was shared by a number of FC+ video-only RPDs.

In order to eliminate any testing variability associated with the use of public infrastructure (public speed test servers, for example) and the intermediate public Internet, all testing was conducted to an iPerf 3 server connected to the very same routing platform the vCMTS core is connected to. A 10 Gbps+ capable

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

server, connected using four 10GE interfaces, was used to eliminate any potential bandwidth bottlenecks and provide an optimum path from the multi-chassis routing platform stack.

At the test customer premise, a 10 Gbps capable NID terminated the DWDM optical link and then fed the RPD via the inline impairment generator. The inline impairment generator is capable of adding IP-level latency (see Figure 14) and/or PDV to all or selected packets. For example, it is possible to add PDV to PTP traffic while not affecting other traffic (see Figure 15). This particular device is used in the test bed to emulate different options for network distance between the RPD and the vCMTS and observe the impact of the said network distance on the service performance (data throughput as measured using the iPerf 3 server).

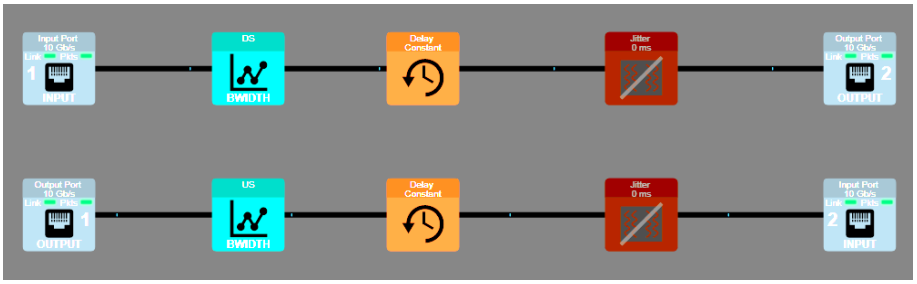


Figure 14: Impairment generator, latency / PDV added to data path to all IP traffic

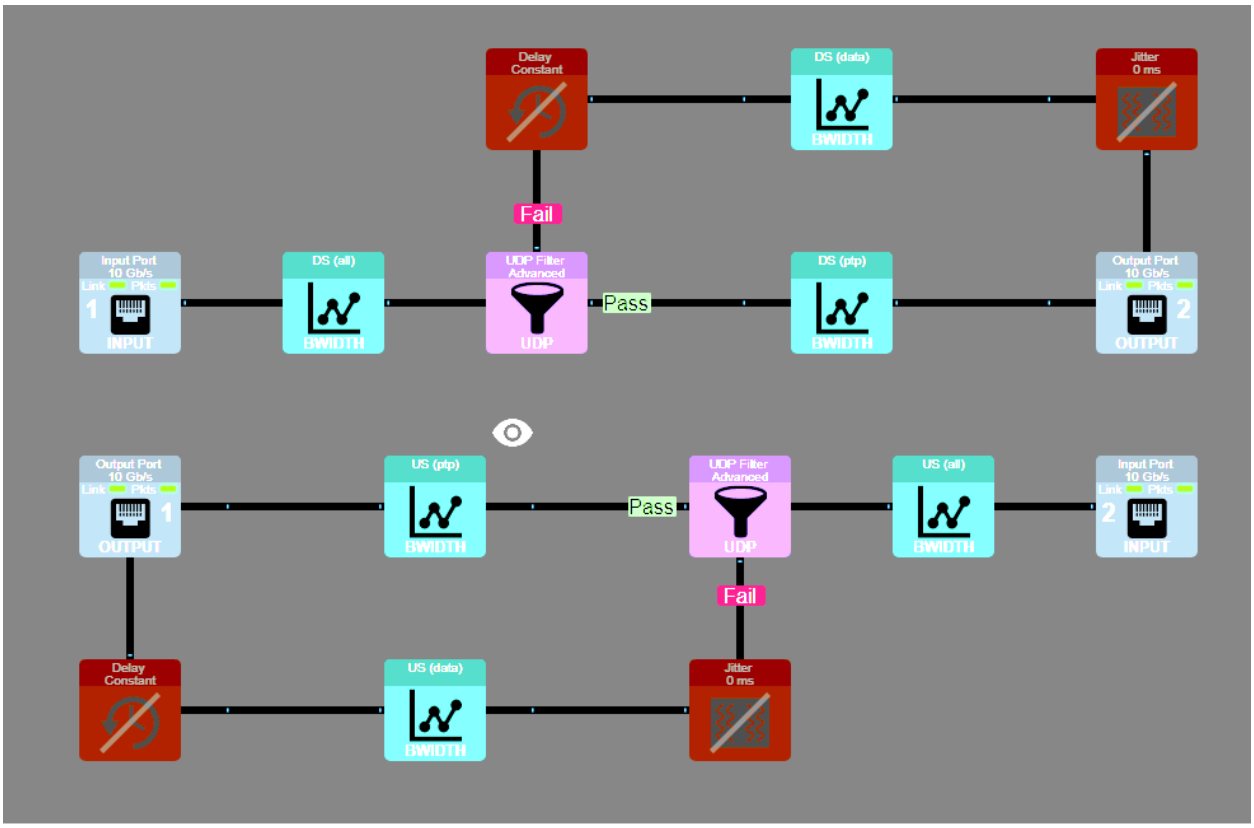


Figure 15: Impairment generator, latency / PDV added to data path to PTP traffic only

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

RPD A/B then fed individual coaxial distribution plants with connected DOCSIS 3.1 CMs. One of the DOCSIS CMs then fed a test laptop via 2.5GE connection to maximize the available throughput.

The resulting test topology is shown in Figure 16.

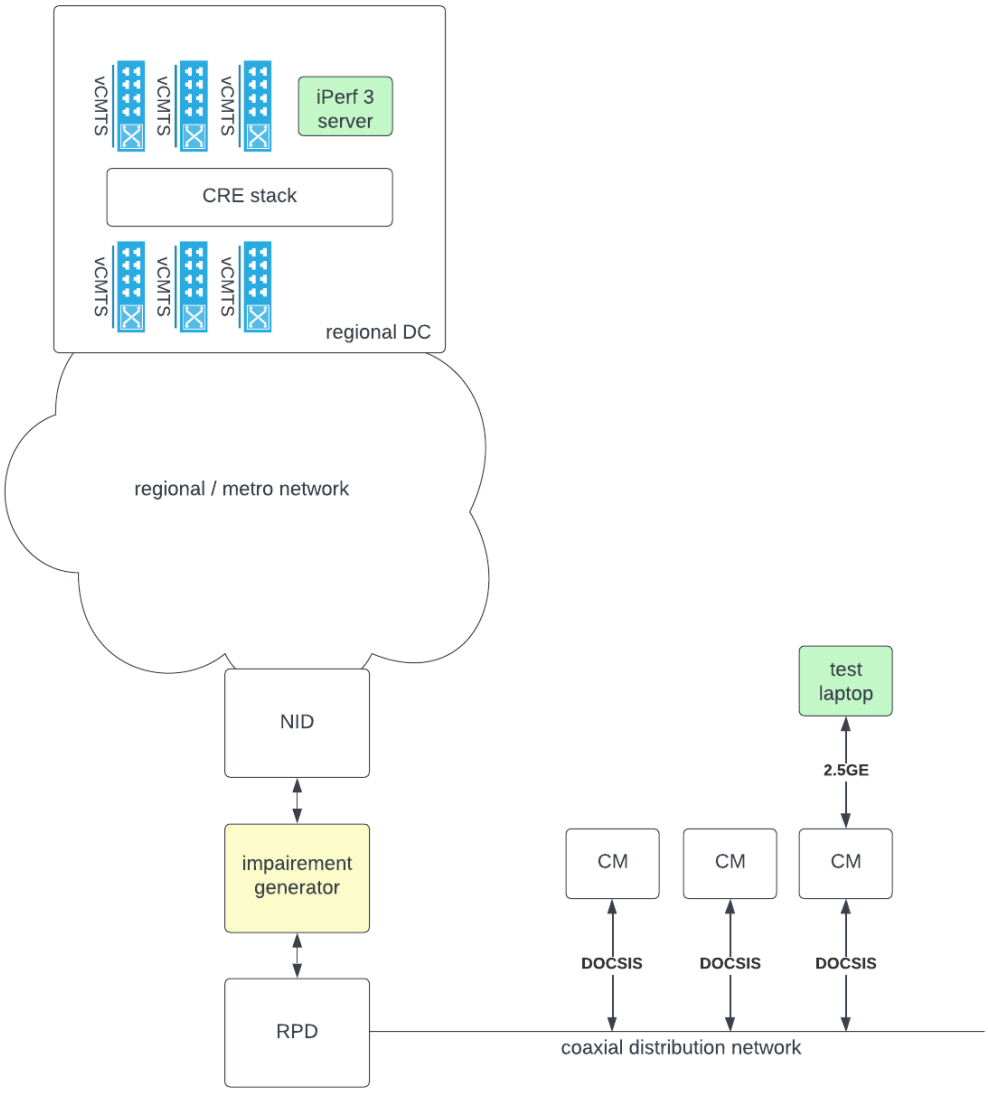


Figure 16: Test topology with in-line impairment generator

In each scenario described below, the service performance (throughput) was measured for a number of different extra latency points, corresponding to the distance between the vCMTS and the RPD equal to: 0, 2.5, 126, 158, 190, 221, 253, 285, 316, 348, 379, 506, 632 and 1264 miles. These individual distance points were derived from added one-way latency assuming the fiber propagation delay of 1.5ns/foot.. Measurements past the 1,300 mile mark were limited by the ability of the vCMTS to maintain communication with the RPD. Each measurement was taken at least five times and averaged to eliminate

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the potential impact of any single measurement outlier. Measurement at zero mile-added latency corresponds to inherent network latency (regional and metro) associated with the test data circuit.

Additionally, the impact of network latency asymmetry was examined. When enabled, the operation of ECMP may result in the downstream and upstream data streams taking different paths across the provider network, resulting in different latency. This problem is shown in Figure 17 , where the downstream direction (green path) is different than the upstream direction (red path), typically resulting in a different network latency. Asymmetric network latency does have an impact on the PTP, and specifically, on the synchronization precision since PTP is not able to compensate for the network latency asymmetry.

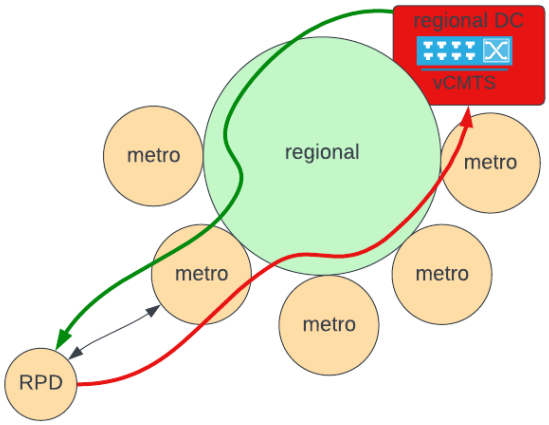


Figure 17: Asymmetric routing problem

Three scenarios are examined in each case: a fully symmetric (ideal) network, a network with 100 miles of asymmetry (around 0.8ms, a value that can be easily accumulated in the production network) and the worst-case scenario of 700 miles asymmetry (around 5.6ms) observed to date in the production network.

All the presented data rate figures represent L3 throughput and are inherently lower than the maximum L2 Ethernet data rate for the link between the test laptop and the DOCSIS CM (2.5 Gbps).

4.1. Impact of Network Latency on HS Services

The observed downstream and upstream service performance in the HS service scenario is shown in Figure 18 for scenarios with symmetric and asymmetric (100 miles and 700 miles worth of latency asymmetry) conditions.

The downstream direction performance is mostly independent from the network latency symmetry, showing the highest performance for the symmetric latency, and decreasing slightly along with the increase in the latency asymmetry. This is an expected behavior since the downstream direction does not require any scheduling and the observable throughput is only limited by the performance of the TCP itself. There is also very limited dependence on the network latency value (network diameter size), likely due to the TCP windowing as well as the use of parallel test streams filling the test circuit very efficiently.

The upstream direction exhibits much stronger correlation with the increase in the network latency, where with sub 400 miles of latency, the throughput decreases from ~1250 to ~1050 Mbps and then falls off much quicker, reaching around 550 Mbps for 1264 miles marker. The dependency on the network latency

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

asymmetry is less expressed than in the downstream direction, likely due to the extra buffering available in the upstream direction along the data path, and especially in the DOCSIS CM.

3.2 Impact of Network Latency on LS Services

The observed downstream and upstream service performance in the LS service scenario is shown in Figure 19 for scenarios with symmetric and asymmetric (100 miles and 700 miles worth of latency asymmetry) conditions.

The observations for the HS covered in the previous section equally apply to the LS scenario, with the only major observable difference in the upstream data rates supported by the system. The decrease in the upstream throughput comes from the use of very limited upstream RF spectrum (single skinny OFMDA block) when compared to the HS scenario with two full OFDMA blocks.

3.3 Impact of PTP Path Pinning

In this scenario, the PTP traffic was isolated from the data traffic and was subject to symmetric network latency, while the data traffic was subject to the previously examined three network latency scenarios: symmetric latency, 100 miles of network latency asymmetry and 700 miles of network latency asymmetry. This scenario corresponds to the PTP traffic engineering design, in which PTP traffic is transported between RPD and the source clock in a bidirectional data tunnel, eliminating the side effect of ECMP operation in the network. While this approach does not eliminate the impact of the network latency itself, it helps manage the impact of network latency asymmetry on PTP traffic, eliminating any potential synchronization inaccuracies between the RPD and the vCMTS core.

Since the downstream transmission direction is not impacted by the RPD/vCMTS synchronization accuracy, only the upstream direction examined for the HS scenario is shown in Figure 20.

Comparing the upstream performance for the HS scenario without PTP path pinning (see Figure 18), LS scenario without PTP path pinning (see Figure 19) and both scenarios with PTP path pinning (see Figure 20), it is visible that the elimination of the network latency asymmetry for PTP traffic improves the throughput stability in the upstream direction. This makes, in turn, the upstream throughput less sensitive to the network latency on the data plane.

PTP path pinning does require additional network configuration, where pre-computed and symmetric data paths are built for PTP traffic between the PTP source (BC) and PTP client (RPD, vCMTS) over the IP/MPLS transport network. Various approaches may be taken as far as the implementation aspect is concerned, including selective MPLS, RSVP-TE, etc., and each method has its own challenges and scaling requirements.

In the simplest implementation, each RPD would have its own PTP path built across the transport network. While it is a simple approach, it is not scalable, since it would require thousands of PTP paths to be configured and maintained, presenting an additional operational challenge.

A more scalable approach builds a PTP path between the given hub and the clock source, aggregating all PTP traffic from all PTP clients served from the given hub. In this approach, a single PTP path can serve hundreds to thousands of PTP clients, depending on the scale of the given RPHY deployment.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

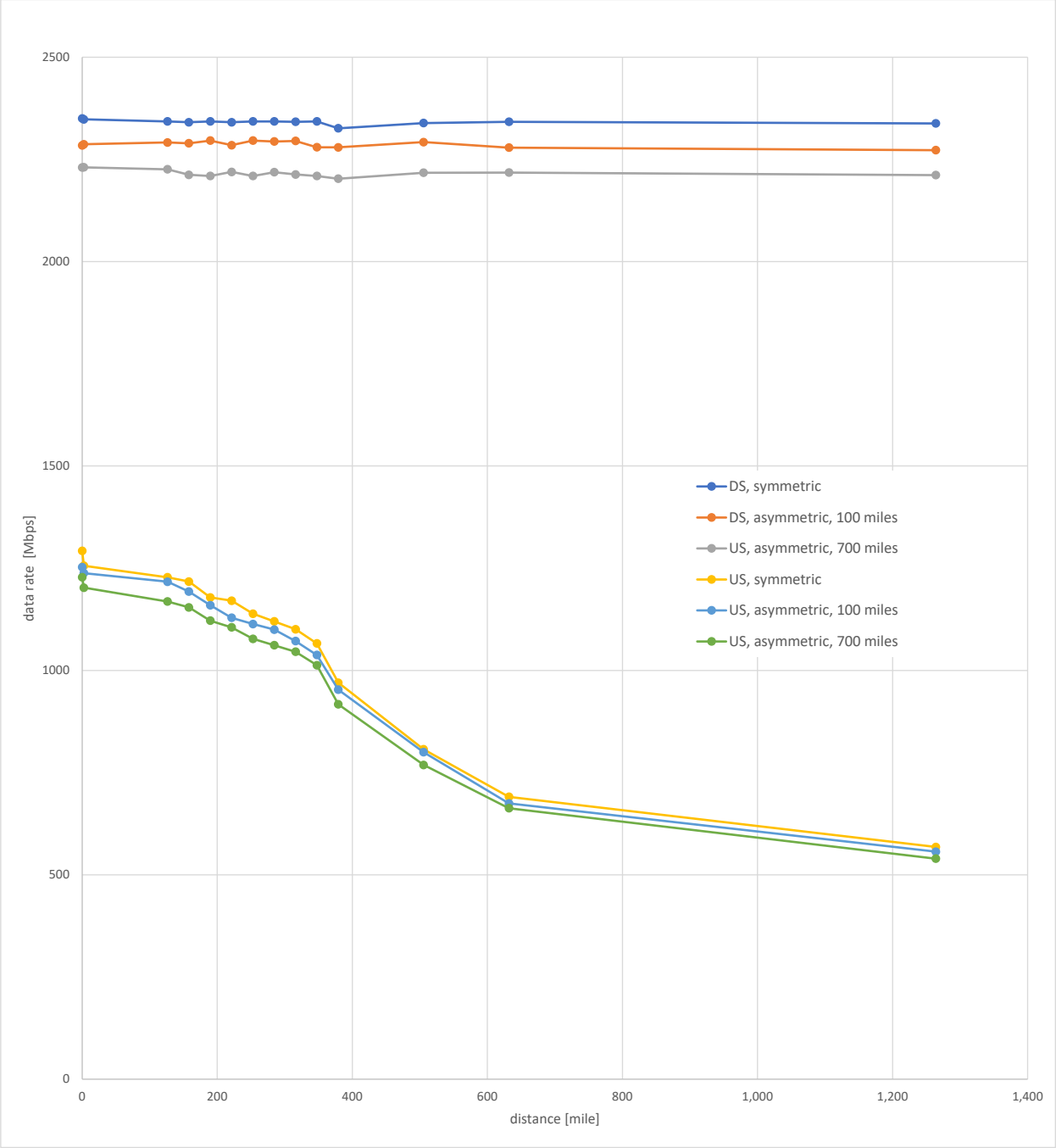


Figure 18: Upstream and downstream HS throughput with 0, 100, and 700 miles network asymmetry

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

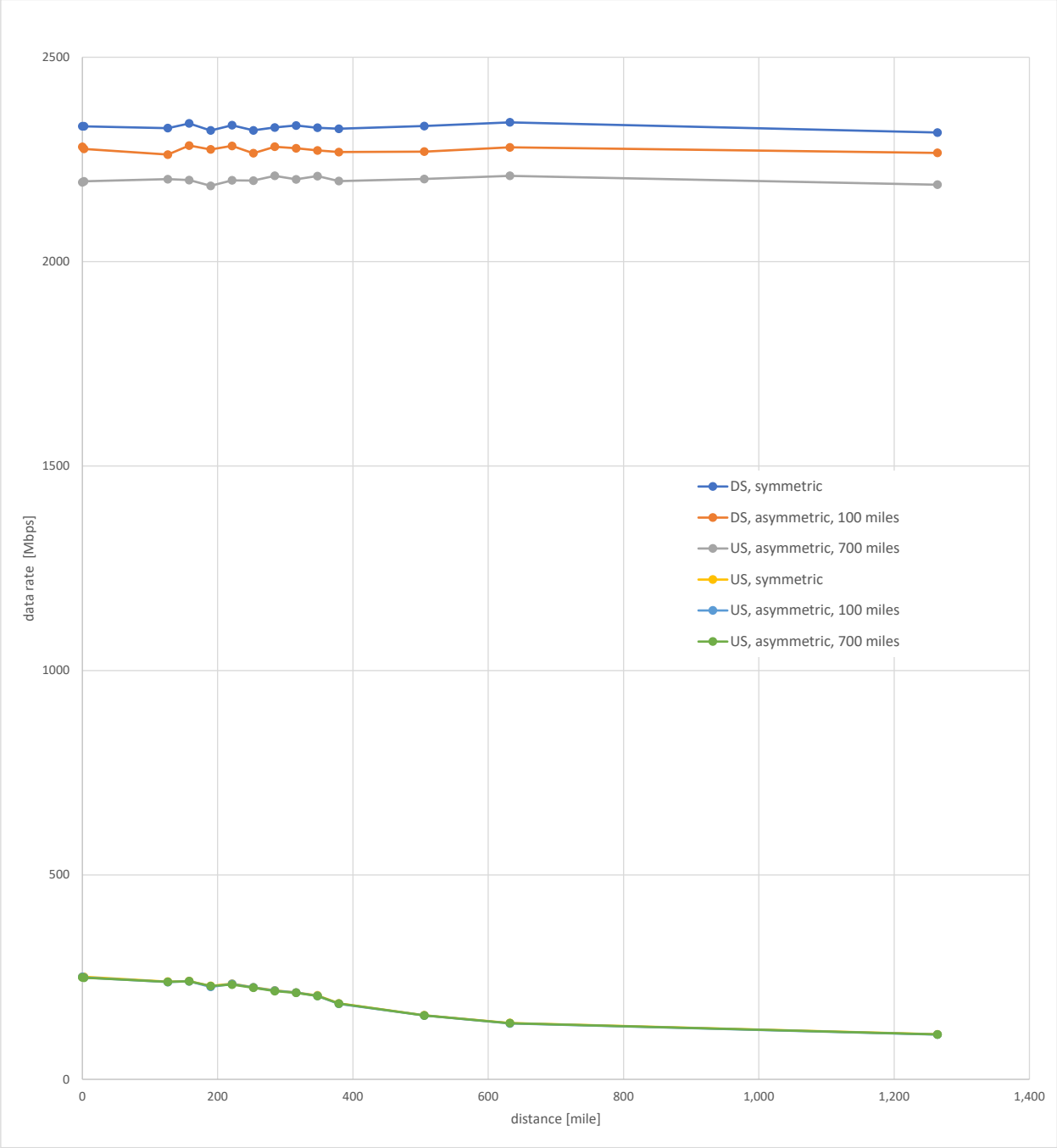


Figure 19: Upstream and downstream LS throughput with 0, 100, and 700 miles network asymmetry

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

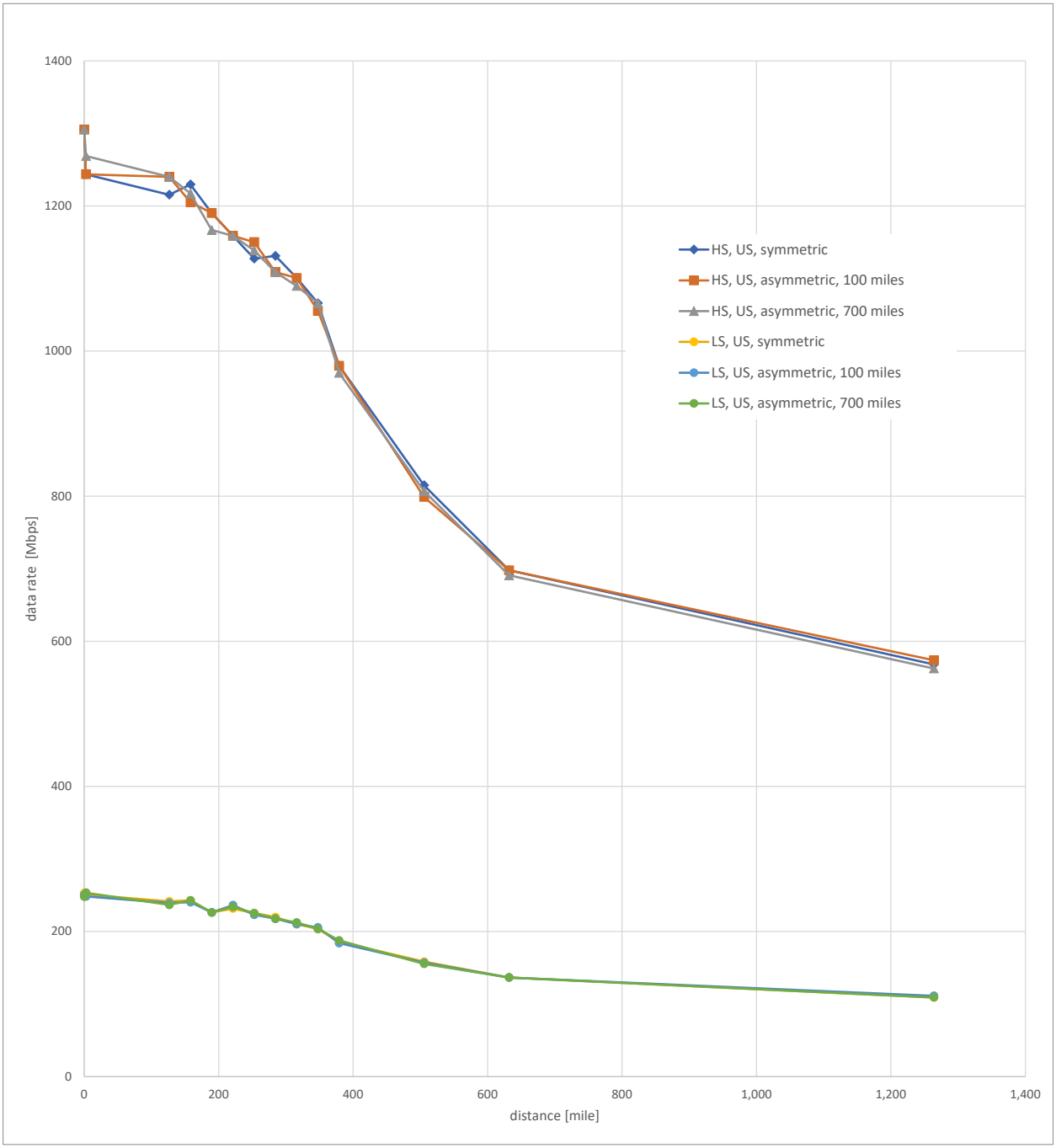


Figure 20: Upstream data throughput with PTP path pinning, HS service plan

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Conclusions

The RPHY service architecture can operate at distances between the vCMTS and RPD, far exceeding typical DOCSIS distances supported by the latest generation of iCMTS. The downstream direction is not adversely affected by the increased distance between the vCMTS and RPD, while the upstream direction exhibits a drop in performance past ~400 miles of distance, likely related to the ability of the vCMTS to schedule the upstream direction effectively.

An effective service area (distance between the vCMTS and the RPD) of around 200 miles provides an opportunity to retire a portion of existing iCMTS units and collapse these DOCSIS plants into a much smaller number of vCMTS nodes. The exact savings will be largely operator-dependent and vary from market to market, driven primarily by the number of iCMTS ports being removed, customer density, availability of IP/MPLS fiber transport between individual hubs and the vCMTS location, as well as any additional work done on the outside coaxial distribution network.

When replacing existing iCMTS ports with RPD ports at the hub (Option 1), the perceived gains from the RPHY architecture are somewhat limited, representing only a disaggregation of the iCMTS platform itself. Along with the move of individual RPDs into the field and decreasing the size of the amplifier cascade, or the deployment of RPDs at the customer premises (Option 3), the aggregate data rates within the architecture will increase and the number of RPDs will grow as well, especially as individual serving groups are further split to increase their capacity.

There is also an observable improvement in the upstream direction performance when the PTP path pinning is used to eliminate the impact of network latency asymmetry on the vCMTS upstream scheduler.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Appendix I – Historical Calculations on REQ-GNT Delays

For reference, here are some delay calculations for RPHY done in 2004. These numbers are out of date and for lower speed interfaces. However, this is included to capture historical documents and to show how the inner structure of the CMTS works. These numbers can be combined with Appendix I of [13].

Remote PHY Delay Budget	EDCS-376514
Author Name, e-mail	John T. Chapman, jchapman@cisco.com
Date	July 20, 2004 (original: May 4, 2004)
Filename	remote-phy-delay-040720a.xls

Inputs:		
DS bit rate	40	Mbps
DS MPEGoIP packet size	1366	bytes
DS MAP Min Length	2000	usec
DS MAP Size	200	bytes
DS Early MAP Release	yes	yes/no
US bit rate	30	Mbps
US REQ packet Size	250	bytes
US Concatenation Size	8000	bytes
US Early REQ Release	yes	yes/no
Ethernet Switch Intrinsic Delay	25	usec
Ethernet Switch Hops	1	0 to n

Outputs:	**Today**	**Remote PHY**	**Delta**	**units**
REQ-GNT Delay Sub-total	**5107**	**5931**	824	usec
% increase of delay budget			16	%
Total With No Concatenation				
MAP Interval Time	2000	2000	0	usec
Rounded up to MAP intervals	**6000**	**6000**	0	usec
% increase of delay budget			0	%
Performance				
net PPS rate	167	167	0	PPS
US bit rate with 1500 byte packets	**2.0**	**2.0**	0	Mbps
Total With Concatenation				
MAP Interval Time	2133	2133	0	usec
Rounded up to MAP intervals	**6133**	**6133**	0	usec
% increase of delay budget			0	%
Performance				
net PPS rate	163	163	0	PPS
US bit rate at max concatenation	**10.4**	**10.4**	0	Mbps

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

REQ-GNT Delay Element (usec)	Today		Remote PHY	
	Sub-total	Total	Sub-total	Total
HFC Plant Round Trip Delay	1600	1600	1600	1600
CM Delay		267		267
CM intrinsic delay	200		200	
CM packet tx time (onto QAM)	67		67	
CMTS Delay		3240		
CMTS US PHY delay	200			
CMTS US REQ Queuing & Processing	1000			
CMTS MAP Advance Margin	1000			
CMTS Downstream PHY delay	1000			
CMTS MAP tx time (onto QAM)	40			
MAC Card Delay				2036
CMTS US REQ Queuing & Processing			1000	
CMTS MAP Advance Margin			1000	
CMTS Packet Encapsulation Delay			0	
CMTS switching intrinsic delay			25	
CMTS MAP tx time (onto GE)			11	
DS Remote PHY Delay				1738
Edge QAM intrinsic delay			400	
Edge QAM Queuing Delay			263	
Edge QAM Downstream PHY delay			1000	
Edge QAM MPEG-TS packet tx time			75	
US Remote PHY Delay				227
Remote US PHYdelay			200	
Remote US intrinsic delay			25	
Remote US packet tx time			2	
GE Switch Delay				63
GE Switch US instrinsic delay			25	
GE Switch US packet tx time			2	
GE Switch DS instrinsic delay			25	
GE Switch DS packet tx time			11	
REQ-GNT Delay Sub-total		**5107**		**5931**

Notes:

DS FEC delay of ~1 msec is included in the MAP advance time

MAP Size is not exactly 2ms. Instead, it is equal in length to an integral number of packets and IE slots.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

BC	boundary clock
BiDi	bidirectional
CAPEX	capital investment
CER	commercial edge router
CIN	content integrated network
CM	cable modem
CMTS	cable modem terminal system
CoS	class of service
DC	data center
DOCSIS	data over cable service interface specifications
DWDM	dense wavelength division multiplexing
FC+	fiber connect plus
GMC	grandmaster clock
HCR	hub core router
HD	high definition
HFC	hybrid fiber-coax
HS	high split
L2TP	layer 2 transport protocol
LS	low split
MCR	metro core router
MER	modulation error ratio
MMF	multi mode fiber
MPLS	multi protocol label switching
NID	network interface device
NOC	network operation center
OPEX	operational expense
OPEX	operational expense
PDV	packet delay variation
PHY	physical layer
PTP	precision time protocol
QoS	quality of service
RF	radio frequency
RPD	remote PHY device
RPHY	remote PHY
SMF	single mode fiber
vCMTS	virtual CMTS
iCMTS	integrated CMTS

Definitions

Downstream	Information flowing from the hub to the user
Upstream	Information flowing from the user to the hub

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography and References

[1] John T Chapman, "Packet Fiber Node", US Patent 7,209,442, filed Jun 27, 2001, issued April 24, 2007 [link]

[2] John T Chapman, "Downstream Remote Physical Interface for Modular CMTS (DEPI)", US Patent 7,688,828, filed May 17, 2005, issued March 30, 2010 [link]

[3] John T Chapman, "Upstream physical interface for modular cable modem termination system (UEPI), US Patent 7,639,617, filed May 17, 2005, issued Dec 29, 2009 [link]

[4] John T. Chapman, "Generic Control Protocol (GCP)", US Patent 9246700, filed Sept 12, 2023, issued Jan 26, 2016 [link]

[5] John T. Chapman et. al, "Managing time offset and frequency drift in asynchronous docsis remote phy network environments (DTI)", US Patent 9,722,739, filed April 15, 2015, issued Aug 1, 2017 [link]

[6] John T Chapman, Alon Berstein, et. al., "Virtual Cable Modem Termination System", US Patent 9,369,339, filed Sept 30, 2013, issued Jun 14, 2016 [link]

[7] John T Chapman, "Modular CMTS Architecture", NCTA Spring Technical Forum, 2005 [link]

[8] John T. Chapman, "DOCSIS Remote PHY", *SCTE Cable-Tec Expo Fall Technical Forum*, Atlanta, GA, Oct 2013.

[9] John T. Chapman, "Remote PHY for Converged DOCSIS, Video and OOB", *NCTA/SCTE Technical Forum*, Los Angeles, Jun, 2014. [link]

[10] CableLabs, Modular Headend Architecture Specifications, 2005 [link]

[11] CableLabs, CDOCSIS Specifications [link]

[12] CableLabs, Remote PHY Specifications (MHAv2) [link]

[13] CableLabs, "Downstream External PHY Interface Specification", CM-SP-DEPI-I08-100611 [link]

[14] CableLabs, "DOCSIS 3.1 Physical Layer Specification", Section 1.2.1 Broadband Access Network, CM-SP-PHYv3.1-I20-230419 [link]

[15] John T. Chapman, Gerry White, Hang Jin, "Impact of CCAP to CM Distance in a Remote PHY Architecture", *SCTE Cable-Tec Expo Fall Technical Forum*, New Orleans, Oct, 2015. [link]

[16] IEEE Std 1588: IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems; Institute of Electrical and Electronics Engineers, published 2019

[17] ITU-T G.8275.2: Precision time protocol telecom profile for phase/time synchronization with partial timing support from the network, International Telecommunication Union, Telecommunication Standardization Sector, published 2022

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23 CELEBRATING 40 YEARS

2023 Fall Technical Forum SCTE • CABLELABS • NCTA

Revolutionizing Access Networks: How TFDM Coherent PON Combines the Best of TDM and WDM PONs

A Technical Paper prepared for SCTE by

Zhensheng (Steve) Jia, Ph.D.
Fellow and Director of Advanced Optical Technologies
CableLabs
858 Coal Creek Circle, Louisville CO, 80027
303.661.3364
s.jia@cablelabs.com

Haipeng Zhang, Ph.D.
Lead Architect
CableLabs
858 Coal Creek Circle, Louisville CO, 80027
303.661.3796
h.zhang@cablelabs.com

L. Alberto Campos, Ph.D.
Fellow
CableLabs
858 Coal Creek Circle, Louisville CO, 80027
303.661.3377
a.campos@cablelabs.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction and Motivation 4
   1.1. PON Evolution 4
   1.2. Coherent Optics Technology Evolution 5
   1.3. Coherent PON 6
   1.4. Coherent TDM-PON and Coherent WDM-PON 7
   1.5. Coherent TWDM-PON and Coherent TFDM-PON 8
2. TFDM CPON Technology and Architecture Overview 10
3. TFDM CPON Key Technology Development 11
   3.1. Burst Signal Processing in TFDM Subcarriers 12
   3.2. Subcarrier Recognition, Pre-equalization, and Power Rebalancing 15
   3.3. ONU Cost Reduction Through Remote Optical Carrier Delivery 19
4. Advantages, Challenges, and Discussions 23
5. Conclusions 25
6. Acknowledgements 25

Abbreviations 26

Bibliography & References 27

List of Figures

Title **Page Number**

Figure 1 – Standardized PON Evolution 4
Figure 2 – The Next Application Fields for Coherent Optics 6
Figure 3 – CPON Supporting Various Applications 7
Figure 4 – Examples of PON Architecture: (a) Time-Division Multiplexing (TDM); (b) Wavelength-Division Multiplexing (WDM) 8
Figure 5 – Examples of PON Architecture: (a) Time and Wavelength Division Multiplexing (TWDM); (b) Time and Frequency Division Multiplexing (TFDM) 9
Figure 6 – Principle of TFDM CPON based on digital subcarrier multiplexing scheme: (a) architecture; (b) example of flexible data/bandwidth allocation 10
Figure 7 – TFDM DSP flow: (a) Tx subcarrier generation DSP; (b) Rx subcarrier demodulation DSP 11
Figure 8 – (a) TFDM burst Tx DSP procedures; (b) TFDM burst Rx DSP procedures; (c) Preamble design in TFDM burst frame 13
Figure 9 – TFDM burst optimization and performance evaluation: (a) Double-Correlation pattern in TFDM burst operation under different synchronization symbol length; (b) TDM burst BER performance under residual CD; (c) TDM burst BER performance under CFO 14
Figure 10 – Proposed DSP procedures featuring subcarrier recoginition, pre-equalization, and power rebalancing in: (a) TFDM coherent burst receiver; (b) TFDM coherent transmitter 16
Figure 11 – Process of CR-FC, Pre-EQ, and PR of TFDM signal 17
Figure 12 – CR-FC algorithm performance: (a) channel boundary detection with large CFO; (b) BER versus frequency detuning using different methods 18
Figure 13 – (a) BER performance with and without Pre-EQ + PR; (b) BER performance 18
Figure 14 – (a) TFDM CPON scheme with remote optical tone delivery; (b) subchannel design for DS and US transmission 19

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 15 – (a) Experimental setup with optical spectrums of TFDM DS and US channels; (b) BER versus ROP in downstream; (c) BER versus ROP in upstream 21
Figure 16 – (a) Residual CFO for proposed OIL scheme vs. regular ECL-based system; (b) BER vs. ROP per channel for proposed OIL scheme without CFO correction compared with regular ECL-based system with CFO correction 22
Figure 17 – Technical summary of DSC based TFDM CPON 23

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction and Motivation

The modern world heavily depends on the swift and dependable sharing of information. The growing need for faster data speeds and greater network capacity consistently surpasses the capabilities of existing systems [1, 2]. Passive optical networks (PONs) have experienced significant advancements in the last twenty years and are now considered highly appealing as access network solutions for providing high-speed data and video services [3, 4]. A PON is an economical and efficient optical communication network technology that uses fiber optics to deliver various broadband services to users. PON architecture adopts a point-to-multipoint (P2MP) structure and employs passive optical components to establish connections between service providers and customers. Passive fiber optic splitters play a crucial role in PON architecture as they divide the signal in a way that each port receives the same data signal, although depending on the splitter design, different ports may have varying power levels. To meet the increasing bandwidth demand driven by data intensive applications such as video streaming, 5G mobile Internet, and cloud networking, several generations of PON systems have been standardized through the efforts of two major organizations: the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) and the IEEE 802.3 Ethernet Working Group, as shown in Figure 1.

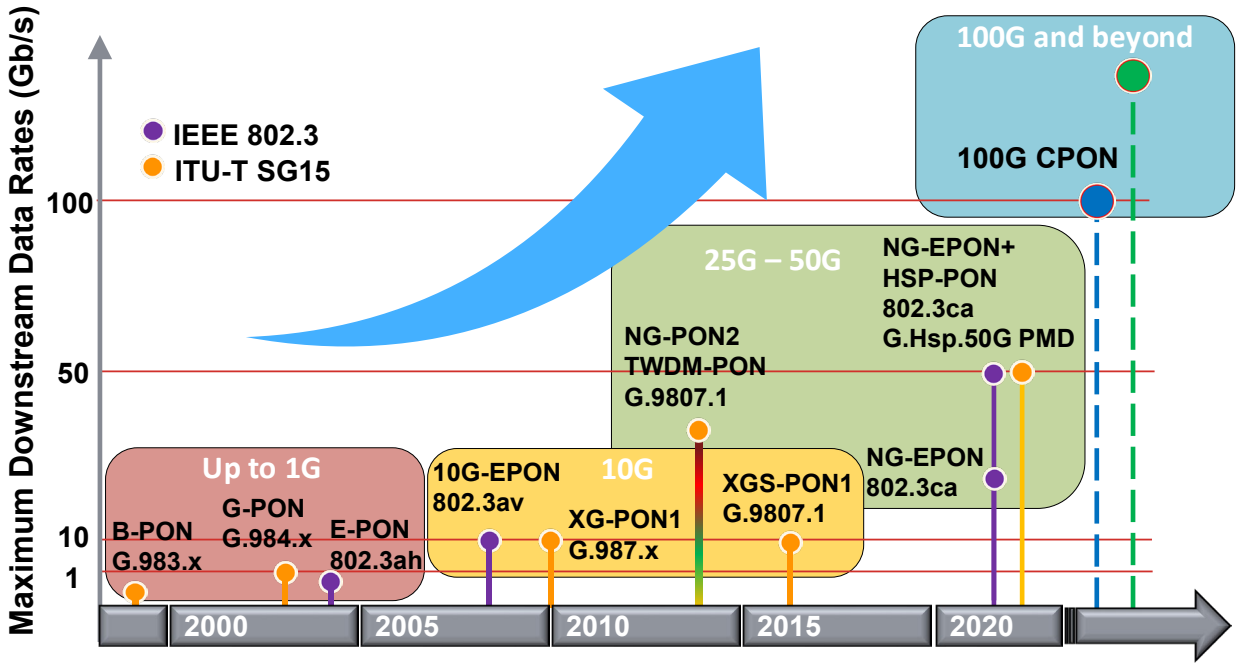


Figure 1 – Standardized PON Evolution

1.1. PON Evolution

The initial standardized PONs, namely asynchronous transfer mode PON (APON) and broadband PON (BPON) defined in G.983, were introduced by the ITU-T in the late 1990s. These systems offered a symmetric line rate ranging from 155 Mb/s to 622.08 Mb/s. Following this, the widely deployed gigabit PON (GPON) defined in G.984 was standardized in 2003, offering downstream speeds of up to 2.5 Gb/s. It had a maximum reach of 20 km and a maximum split ratio of 1:64. The IEEE standardized the Ethernet PON (EPON) as 802.3ah in 2004. EPON utilized ethernet frames and provided a symmetric line rate of 1.25 Gb/s, with a maximum reach of 20 km and a split ratio of 1:32. In 2009, the 10G EPON standard (802.3av) was released, offering a symmetrical line rate of 10.3125 Gb/s. The ITU-T introduced the next-

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

generation PON (XG PON) in 2010 (G.987), providing downstream speeds of 10 Gb/s and upstream speeds of 2.5 Gb/s. The next-generation PON stage 2 (NG-PON2) was standardized in 2015 (G.989), utilizing time- and wavelength- division multiplexing (TWDM) architecture to support 4-8 wavelengths with a broadband speed of 10 Gb/s per wavelength. The next-generation symmetric PON (XGS PON), a symmetric version of XG PON, offered a symmetrical line rate of 10 Gb/s for both downstream and upstream. Recently, the IEEE 802.3ca Nx25G EPON Task Force defined a system based on a 25 Gb/s line rate, aiming to standardize a 100G-EPON initially but scaled back to a 2x25 Gb/s system [5]. On the ITU-T side, in 2021, a suite of G.9804 Recommendations (G.hsp) was developed, focusing on a 50 Gb/s line rate PON system to lay the groundwork for future ITU-T PON systems [6]. The Common Transmission Convergence (ComTC) layer in G.9804.2 is designed to be agnostic of transmission rates, wavelength channels, and signal modulation, making it applicable to future time-division-multiplexing (TDM) and TWDM PON systems.

The current PONs utilize intensity modulation–direct detection (IM-DD) technology to achieve a cost-effective and efficient balance. By prediction, PON technology of 100 Gb/s per wavelength and beyond is expected to be required soon to meet the ever-increasing demands for higher capacities. Lately, Nokia has demonstrated 100 Gb/s PON based on IM/DD pulse amplitude modulation 4-level (PAM-4) and flexible forward error correction (FEC) code rate [7]. However, as PON development progresses, certain factors come into play. These factors include the need for increased launched power, improved forward error correction (FEC), and higher sensitivity receivers to meet around 29-dB loss budget and optical path penalty required for higher data rates. This power budget is essential to ensure compatibility with existing PON infrastructure, as deploying fiber networks entails significant investment. Nonetheless, this trend changes when data rates exceed 25 Gb/s, as IM-DD technology faces physical challenges that result in higher costs and power consumption for the overall system architecture and optical design. To tackle these challenges, several requirements and considerations emerge. These include the need for high sampling rate digital-to-analog converters (DAC) and analog-to-digital converters (ADC) to enable signal processing in the digital domain. Additionally, digital signal processing algorithms play a vital role in compensating for device bandwidth limitations and mitigating transmission impairments. Due to limited link budgets, optical amplification becomes necessary in both centralized optics and customer premise equipment. The congestion in the O-band for both downstream and upstream adds complexity to wavelength planning and resource management. Furthermore, the utilization of higher bandwidth opto-electronic components becomes imperative to minimize implementation penalties. Hence, when aiming for data rates reaching and exceeding 100 Gb/s on a single wavelength, the utilization of IM-DD technology becomes considerably difficult.

1.2. Coherent Optics Technology Evolution

On the other hand, the introduction of coherent optical technology has had a profound impact on optical transmission systems, leading to extensive upgrades and the implementation of dense wavelength division multiplexing (DWDM) networks operating at speeds of 100 Gbps, 200 Gbps, and 400 Gbps per wavelength. Initially developed for long-haul applications, coherent optics has evolved and is now being widely employed in metro networks. This transition has been facilitated by advancements in complementary metal-oxide semiconductor (CMOS) manufacturing, simplified design approaches, cost reduction in opto-electronic components, and the demand for more efficient optical transport technologies. As a result, coherent solutions are now being integrated into new market segments, particularly for short-haul applications in edge and access networks. This progression from long-distance to short-distance models has been observed previously in the optical industry, as DWDM technology initially emerged in long-haul scenarios before being adopted in metro and edge access networks.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

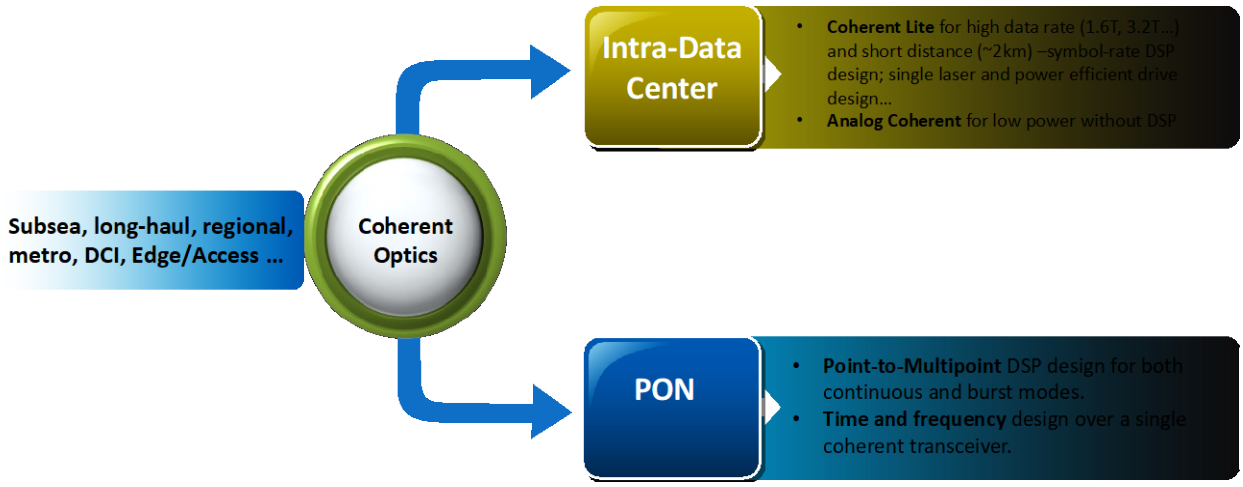


Figure 2 – The Next Application Fields for Coherent Optics

Building upon the foundation laid by long-haul technology development, coherent optics is set to follow a similar natural progression into access networks as shown in Figure 2. The forthcoming horizons of coherent optical networking are anticipated to encompass intra-data center communication and point-to-multipoint PONs. Progress has been made in the advancement of power-efficient designs targeting intra-data center application through the utilization of coherent lite, analog coherent, and mixed-domain coherent technologies. These innovations hold the promise of achieving data transmission rates of up to 3.2Tb/s over distances of less than 2 km within intra-data center environments, all while maintaining comparable advantages with conventional IM-DD systems in terms of cost, power efficiency, and latency. On the other hand, the inception of CableLabs' Coherent PON (CPON) project, initiated two years ago, is emblematic of a collective endeavor aimed at realizing the next-generation 100G point-to-multipoint access architecture.

1.3. Coherent PON

CPON is the application of coherent optical transmission principles to PON technology, offering numerous advantages for access networks. CPON utilizes multiplexing in multiple dimensions (optical amplitude, phase, and polarization) to encode information, enabling higher data rates with low-bandwidth analog and digital signal processing techniques. By employing a local oscillator, CPON achieves significant coherent gain, resulting in superior receiver sensitivity and a high-power budget without the need for excessive launched power. The recovered signal facilitates digital compensation of linear transmission impairments, such as chromatic dispersion (CD) and polarization-mode dispersion (PMD), with minimal optical path penalty, enabling the implementation of ultra-long reach PON.

CPON technology sets itself apart from IM-DD technology by providing several benefits for network operators. These include expanded use cases and coverage areas through higher link budgets, the potential for cost reduction by scaling back or eliminating expensive facilities for distant communities, the ability to increase subscriber counts without a linear increase in fiber usage through wavelength stacking in the C-band, and the flexibility to support convergence needs for optical signal transport functions at the network edge, extending PON's applications beyond traditional residential deployments.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

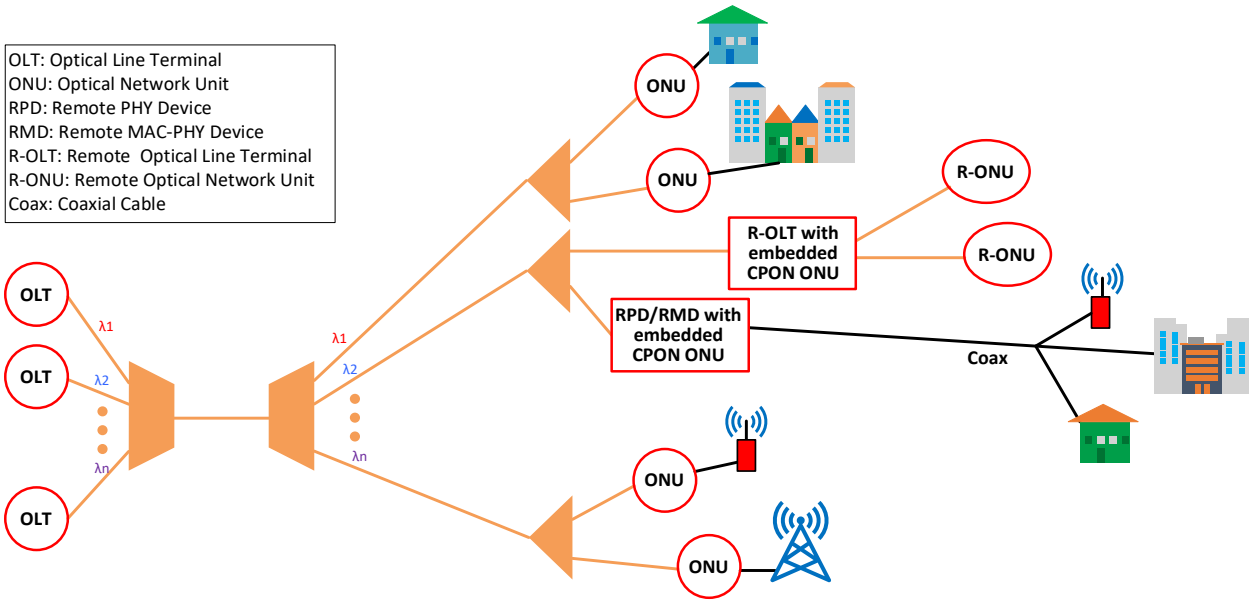


Figure 3 – CPON Supporting Various Applications

In addition to fiber to the home (FTTH) covering multiple dwelling units (MDUs) and single-family units (SFUs), CPON has the capability to support a wide range of applications, as depicted in Figure 3. For instance, CPON can facilitate backhaul connectivity for remote devices within an operator's network, enabling connectivity and backhaul of remote OLT traffic. It can also handle the traffic collected from Remote PHY devices (RPDs) and Remote MAC-PHY devices (RMDs) in a distributed Converged Cable Access Platform (CCAP) architecture. Moreover, CPON can provide mobile backhauling services by offering base station connectivity and carrying mobile backhaul traffic. It is designed to coexist with existing fiber-based technologies like legacy PON systems or wavelength-division multiplexing (WDM) point-to-point links, while also supporting the stacking of multiple CPON wavelengths on a single fiber using WDM. These features greatly enhance deployment flexibility across various scenarios, from high-density interconnects in densely populated urban areas to low-density deployments over rural areas.

1.4. Coherent TDM-PON and Coherent WDM-PON

There have been reports on two types of coherent PON: coherent TDM-PON [8, 9] and coherent WDM-PON [10, 11]. A coherent TDM-PON architecture, as illustrated in Figure 4 (a), is based on a commonly used power splitter-based optical distribution network (ODN). In this architecture, the OLT transmits the downstream signal continuously to each ONU on a single wavelength channel. For upstream transmission, each ONU is allocated a specific timeslot to send a burst of upstream data on another wavelength channel. Alternatively, Figure 4 (b) depicts a coherent WDM-PON architecture where a dedicated wavelength is assigned to each ONU, enabling a virtual point-to-point optical connection. This architecture utilizes a wavelength-routed ODN with inherent wavelength routing capability through wavelength splitters. Coherent WDM-PON and TDM-PON both exhibit significant performance improvements compared to traditional IM-DD PON, thanks to the utilization of coherent detection. On the other hand, although coherent TDM-PON offers a practical approach by efficiently sharing bandwidth in the time domain, it may require a new scheduling algorithm to reduce latency in larger coverage groups. Coherent WDM-PON allows for more flexible allocation of bandwidth but necessitates the use of multiple wavelengths and colored optics, which can result in additional costs and operational complexities.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

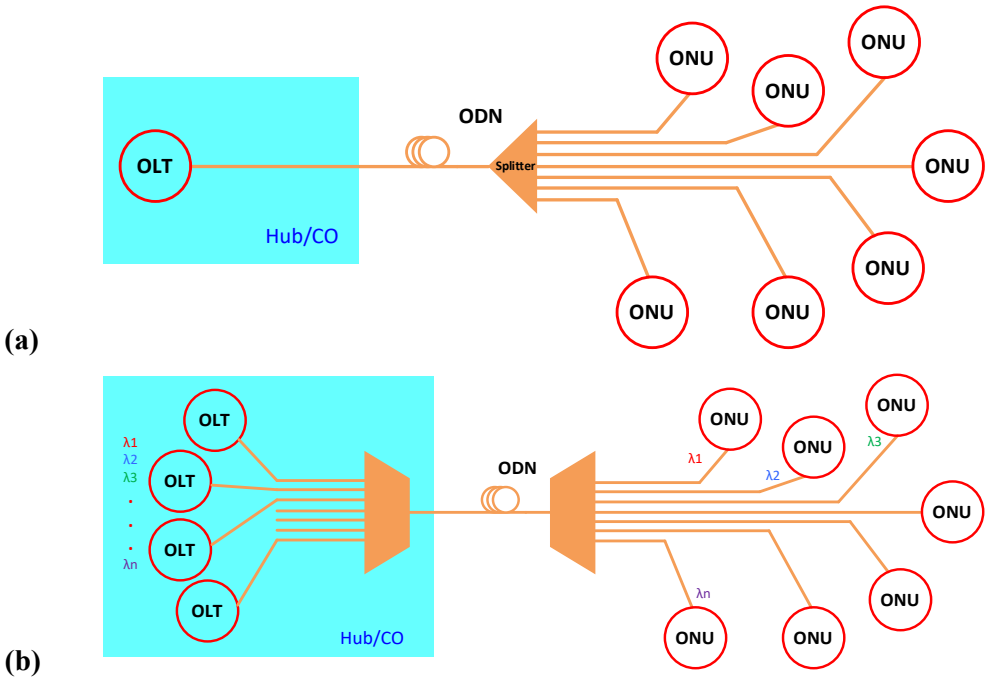


Figure 4 – Examples of PON Architecture: (a) Time-Division Multiplexing (TDM); (b) Wavelength-Division Multiplexing (WDM)

1.5. Coherent TWDM-PON and Coherent TFDM-PON

The next generation access network can benefit from combining the strengths and addressing the limitations of TDM-PON and WDM-PON to achieve more versatile bandwidth allocation. Time and Wavelength Division Multiplexing (TWDM) stands as an alternate form of multiplexed PON architecture harnessed by NG-PON2, enabling multiple signals to be transmitted simultaneously over a single ODN. Coherent technology can be implemented in TWDM-PON, which combines both TDM and WDM technologies mentioned earlier. In TWDM-PON, multiple optical signals are transmitted over the same fiber by assigning different wavelengths to each signal. Each wavelength channel can carry its own independent data stream, and time slots within each channel are used to transmit data from different users or services. Another promising solution for coherent PON is Time and Frequency Division Multiplexing (TFDM), which utilizes digital subcarrier (DSC) multiplexing technology [12-15]., TFDM is a technique that combines TDM and frequency division multiplexing (FDM). In TFDM, multiple optical signals are transmitted over the same fiber by allocating different frequency bands to each signal. Each frequency band can carry its own independent data stream, and time slots within each band are used to transmit data from different users or services. both TWDM and TFDM are techniques that leverage combinations of time, wavelength, and frequency division multiplexing to achieve higher data transmission capacities in PONs as shown in Figure 5. The key difference lies in whether they primarily use wavelength channels (TWDM) or frequency bands (TFDM) to separate data streams. In TWDM architecture, for each wavelength channel, a dedicated transmitter and receiver are necessary for both directions. This involves the use of optical mux/demux and tunable optical filters for effective channelization. The spacing between each wavelength can be as wide as 100GHz or more, allowing for the use of cost-effective optical filtering devices. The commercial reality of NG-PON2 reveals a setup with four channels operating at a

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

100 GHz spacing. It adopts a 10G/10G TWDM approach, facilitating 10 Gbps upstream and downstream data flow, exclusively tailored for 10G ONUs.

However, in the case of TFDM, frequency division multiplexing occurs in the digital domain, where a single transmitter and receiver can handle any given DSC. The inherent coherent selection filtering capability of TFDM enables its direct application within the power splitter based ODN. This TFDM coherent systems based on the DSC partition data into multiple narrower intermediate digital carriers. These carriers are then modulated onto the optical carrier in a parallel fashion. This technique occupies a comparable bandwidth to single-carrier modulation, utilizing identical modulation formats and total transport capacity. As shown in Figure 5 with four DSCs, while the OLT handles the transmission and reception of all DSCs, individual ONU nodes have the flexibility for modulating and demodulating only a specific subset of these sub-carriers. TFDM CPON offers considerable flexibility by leveraging the sharing of bandwidth in both the time and frequency domains, eliminating the need for colored optics. By employing TFDM DSCs in the frequency domain, different frequency bands can be assigned to network services or groups of ONUs with distinct latency or capacity requirements. This approach can significantly reduce scheduling latency and traffic blocking rates. Furthermore, TFDM technology enables efficient bandwidth granularity and utilization, flexible link budget designs, and scalable architecture. The enhanced flexibility and capability of coherent TFDM technology positions it as a compelling candidate for future access networks.

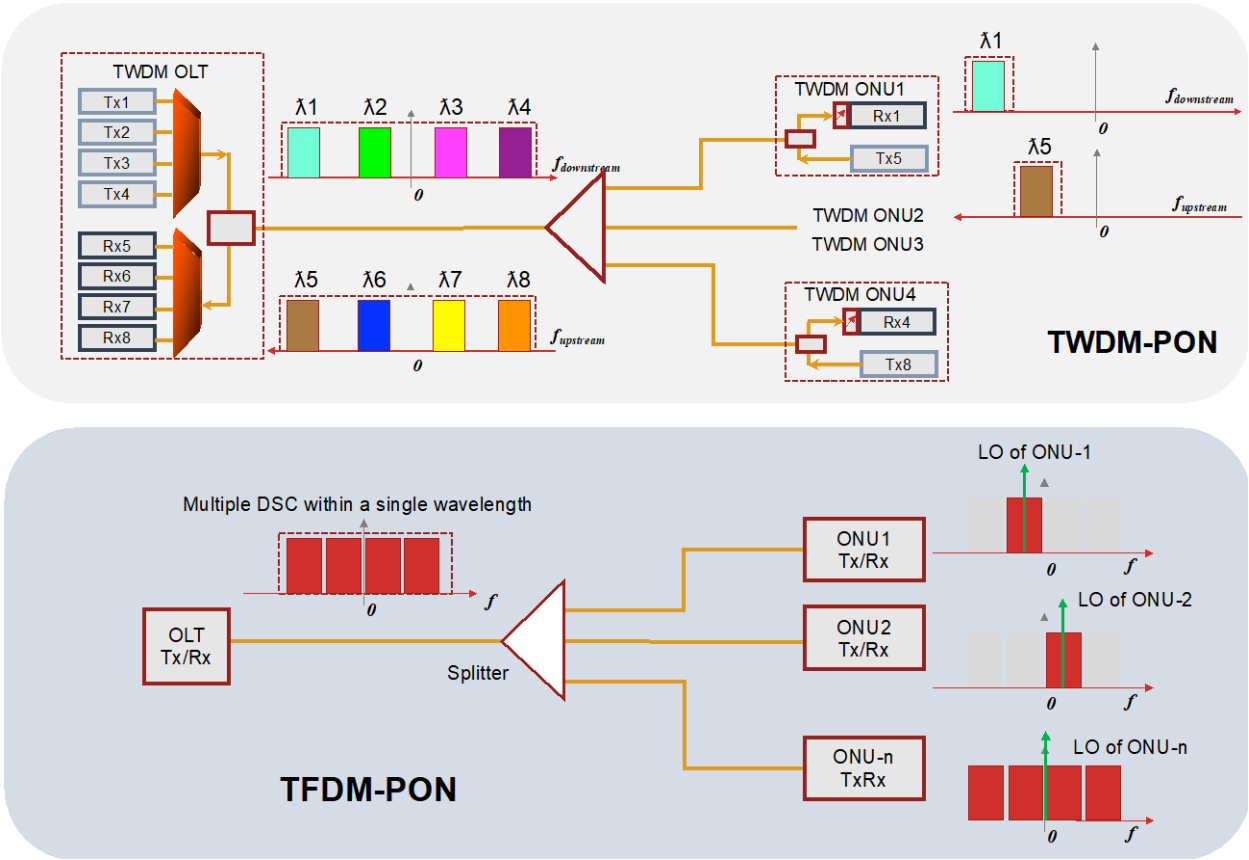


Figure 5 – Examples of PON Architecture: (a) Time and Wavelength Division Multiplexing (TWDM); (b) Time and Frequency Division Multiplexing (TFDM)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2. TFDM CPON Technology and Architecture Overview

Leveraging DSCs in frequency domain and time domain burst processing, TFDM CPON offers unprecedently high degree of flexibility in terms of capacity and resource allocation. Figure 6 depicts the structure of a TFDM CPON that utilizes DSC multiplexing technology. In Figure 6 (a), the architecture of a 100G CPON with four subcarriers is presented, as an example. In the downstream direction, the four subcarriers are generated and modulated on a single wavelength in the optical line terminal (OLT) using a coherent I/Q modulator. The coherent receivers in optical network units (ONUs) detect the multi-subcarrier signal downstream and each sub-band can be demodulated individually. In the upstream direction, each ONU can generate and modulate one to four subcarriers, with data rates ranging from 25 Gb/s to 100 Gb/s. Consequently, the aggregated upstream signal comprises four subcarriers, which may originate from the same ONU or different ONUs. Unlike IM/DD technology, which requires multiple receivers and different wavelengths for different subcarriers from different ONUs, TFDM CPON only needs a single broadband coherent receiver at the OLT-side to detect the four subcarriers, even if they come from different ONUs. This represents a significant distinction between the two methods in terms of multiple sub-carrier detection. Figure 6 (a) also illustrates the transmitter (Tx) and receiver (Rx) configuration of the OLT, which is identical to the setup in the ONU. Consequently, the total downstream and upstream capacities are both 100Gb/s and are shared among multiple ONUs using a hybrid TFDM scheme, as shown in Figure 6 (b). Each ONU can achieve a peak data rate of 100Gb/s in both the downstream and upstream, similar to what is achievable in TDM PON. Additionally, the proposed TFDM CPON offers more flexible bandwidth allocation. It enables two-dimensional bandwidth allocation, allowing for the establishment of dedicated channels for end users with, i.e., low-latency or high-capacity requirements.

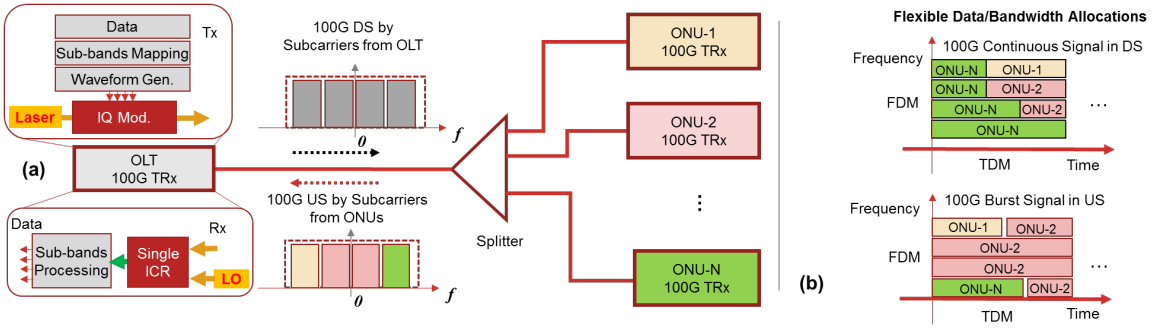


Figure 6 – Principle of TFDM CPON based on digital subcarrier multiplexing scheme: (a) architecture; (b) example of flexible data/bandwidth allocation

Figure 7 (a) and (b) illustrate the digital signal processing (DSP) involved in subcarrier generation on the Tx side and demodulation on the Rx side. In the signal generation process, the data is initially mapped and modulated onto each subcarrier. To reduce the bandwidth on each subcarrier, Nyquist pulse shaping is applied with a 0.1 roll-off factor. Following pre-equalization, the four subcarriers with a baud rate of 6.25 GBd are upconverted to four intermediate frequencies: -15 GHz, -5 GHz, 5 GHz, and 15 GHz. On the receiver side, the inverse processing flow is performed. The subcarriers are first filtered out and separated, and then down converted to baseband. They are processed independently using several key signal recovery functions.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

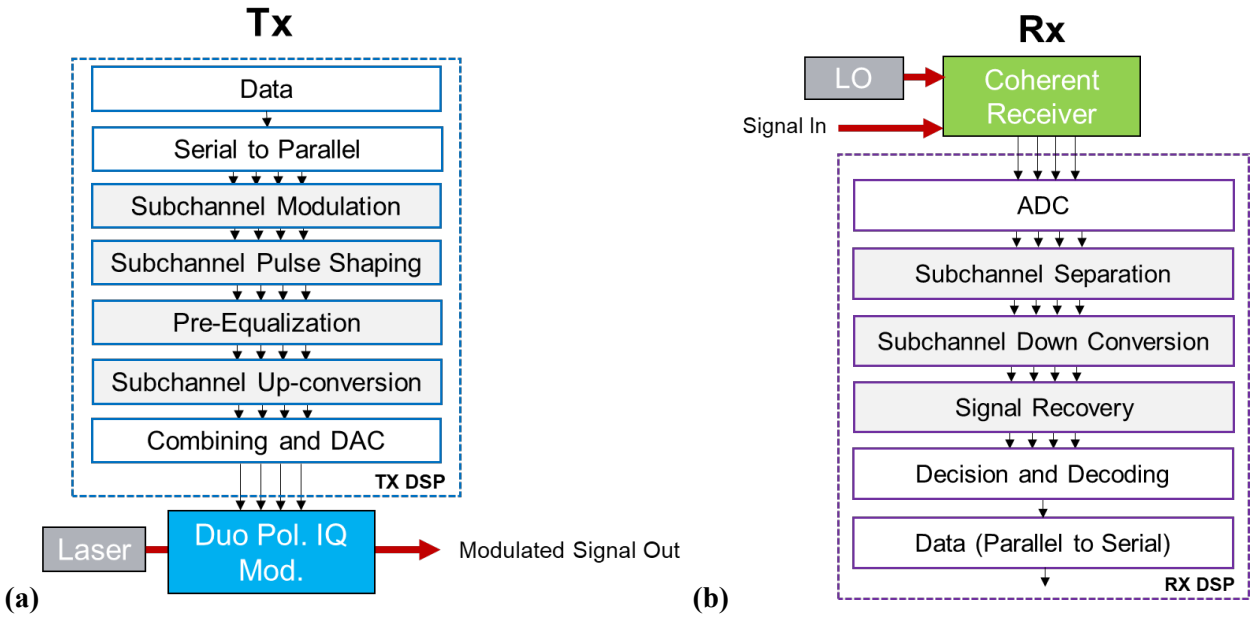


Figure 7 – TFDM DSP flow: (a) Tx subcarrier generation DSP; (b) Rx subcarrier demodulation DSP

The proposed coherent TFDM CPON introduces support for asymmetric ONU/OLT hardware setups and a pay-as-you-go cost model based on bandwidth subscription. Unlike traditional WDM-PON or TWDM-PON, all the ONUs and OLT in the proposed system operate on the same wavelength grid with slight frequency tuning. The coherent detection's inherent wavelength selective feature eliminates the need for optical filtering or wavelength selective components for frequency selection. This coherent TFDM PON system enables more flexible bandwidth sharing utilizing both time and frequency. It is fully compatible with widely deployed TDM PON and can operate on power splitter-based ODNs without requiring additional components. With only one transceiver on the OLT-side and one transceiver on the ONU-side, the system can provide lower overall scheduling latency due to the utilization of frequency division multiplexing.

3. TFDM CPON Key Technology Development

Coherent TFDM technology has undergone significant advancements in past few years, enabling enhanced performance and flexibility in CPONs. In this paper, we highlight three key developments in TFDM technology. Firstly, TFDM Burst Transmitter/Receiver, which is a crucial advancement in TFDM CPONs. This technology enables time domain multiplexing in DSCs, allowing for efficient transmission and reception of burst signals. Secondly, subcarrier frequency window detection and calibration, pre-equalization, and power rebalancing. To maximize the performance of TFDM CPONs, accurate subcarrier frequency window detection and calibration techniques are essential. These methods ensure precise alignment of the frequency windows across the subcarriers, minimizing frequency offset and power imbalance, and improving overall system performance. Lastly, low-cost ONU light sources through remote master tone delivery and optical injection locking. One of the challenges in deploying coherent optics for access networks is the high cost associated with ONU light sources. TFDM technology has introduced a novel solution by leveraging remote master tone delivery and optical injection locking to replace high-cost laser sources and reduce the complexity and cost of coherent ONUs. Through injection locking, the child laser (Fabry-Perot (FP) laser) closely adopts the optical frequency and linewidth characteristics of the parent laser (ECL laser). At the OLT end, seed (or parent/master) laser

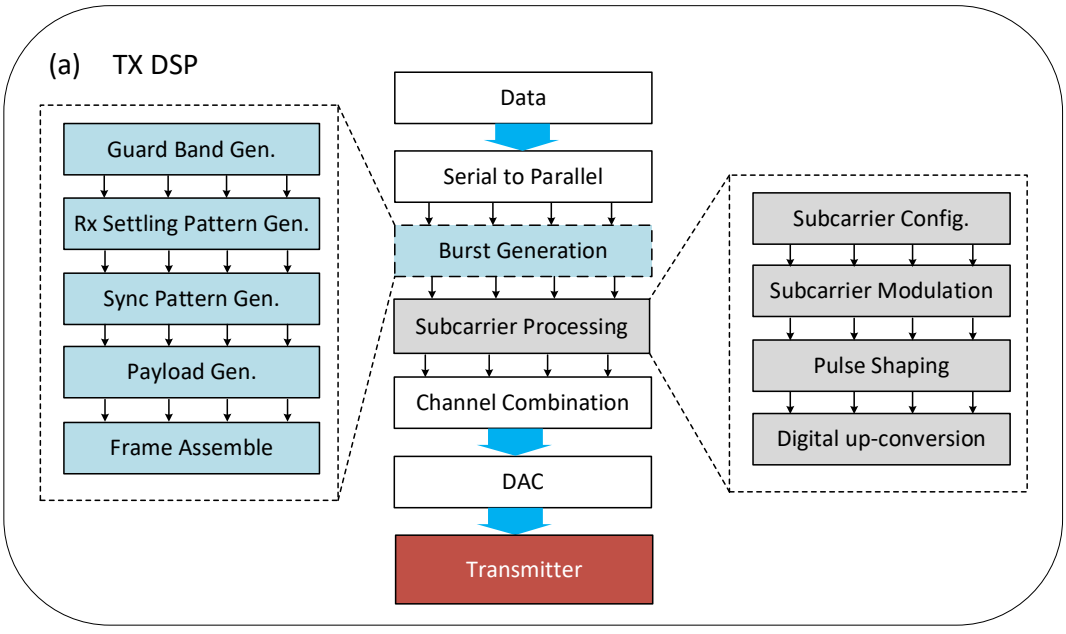
© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

sources are generated for both downstream LO and upstream injection carrier seed. Through remote delivery, at the ONU end, only low-cost child lasers are required. Through design, the OIL process can amplify both tones (LO and upstream carrier), eliminating the need for extra optical amplifiers.

These key TFDM technology developments have significantly contributed to the advancement of CPONs. These advancements are expected to drive the adoption of TFDM technology in the emerging era of high-capacity and flexible optical access networks.

3.1. Burst Signal Processing in TFDM Subcarriers

One of the key challenges for TFDM CPON is the realization of upstream coherent burst-mode transmission and detection in DSCs. Figure 8 depicts the DSP processes of coherent TFDM transmitter and receiver that features burst signal processing and burst preamble design. In Figure 8 (a), the TFDM transmitter involves burst frame generation and subcarrier processing, with the former exclusively for upstream transmission. Burst frame generation includes creating patterns for the guard band, receiver settling, synchronization, and payload data. The frames are assigned to DSCs and undergo Nyquist pulse shaping, digital up-conversion, and digital-to-analog signal conversion for transmission. In the TFDM burst receiver as shown in Figure 8 (b), analog signals are down-converted to baseband and subjected to Fast Fourier transform (FFT), digital filtering, and inverse FFT (IFFT). Burst signal detection, primarily for upstream burst reception, starts with power-based frame detection, followed by chromatic dispersion (CD) compensation, clock recovery, and burst frame synchronization. Once the burst signal is received, payload signals are processed using regular coherent receiver DSP methods. The structure of the TFDM signal's burst frame is depicted in Figure 8 (c). A guard time is incorporated between adjacent burst frames. For TFDM signals operating at 6.25 GBd (0.16 ns per symbol), a guard time of 640 symbols (equivalent to 102.4 ns) is allocated. This guard time allows the transmitter (Tx) to be turned on/off as required. When the burst frame initiates, the first portion is dedicated to Rx settling, typically starting with automatic gain control performed by the burst-mode transimpedance amplifier (BM-TIA). Once the BM-TIA reaches a steady-state, the OLT receiver proceeds with burst-mode signal processing for coherent detection of upstream burst signals. The synchronization section encompasses three major components: frame synchronization, state of polarization (SOP) estimation, and frequency-offset estimation (FOE).



© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

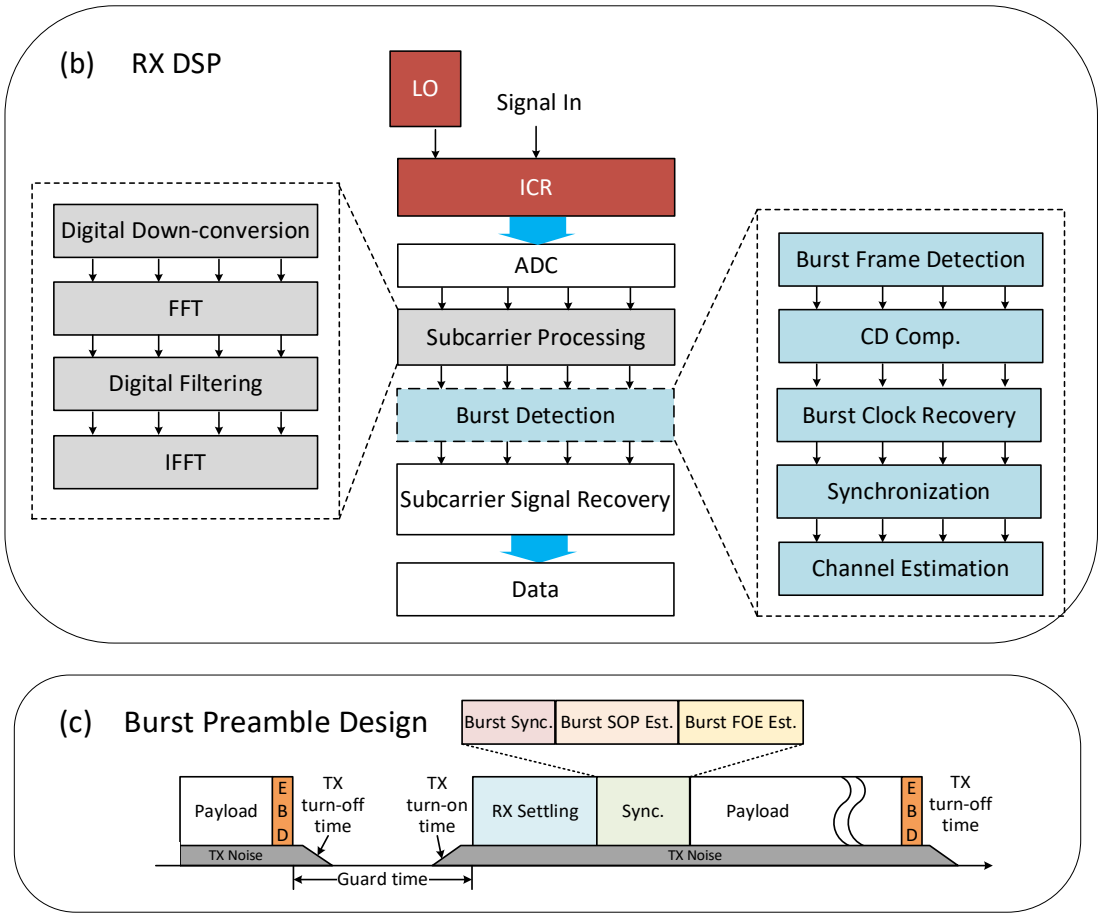


Figure 8 – (a) TFDM burst Tx DSP procedures; (b) TFDM burst Rx DSP procedures; (c) Preamble design in TFDM burst frame

Previous research utilized an auto-correlation based algorithm for TDM burst synchronization, demonstrating good performance [16]. In this study, a double-correlation based algorithm is employed for TFDM burst signal synchronization. This approach aims to achieve more stable and robust burst detection, especially when dealing with significant carrier frequency offset (CFO) and CD resulting from fiber transmission [17]. The synchronization peak $L(\mu)$ using the double-correlation method can be mathematically expressed as Equation (1):

$$L(\mu) = \sum_{i=1}^{L-1} \left\{ \sum_{k=i}^{L-1} \left| r_{\mu+k}^{*} s_k r_{\mu+k-i} s_{k-i}^{*} \right| - \sum_{k=\mu+i}^{\mu+L-1} |r_k||r_{k-i}| \right\} \quad (1)$$

In this equation, $\{s_k\}$ represents a fixed frame synchronization pattern with a length of L symbols, and $\{r_k\}$ denotes the received signal [17]. However, when the sync frame length L is large, the computational complexity becomes impractical for real-world applications due to the high number of iterations required by L-1. To address this, the L-1 term is replaced with a factor M, representing the number of iterations in the double-correlation algorithm, where $1 \leq M < L\text{-}1$. This modification allows for a more feasible calculation as demonstrated in Equation (2):

$$L(\mu) = \sum_{i=1}^{M} \left\{ \sum_{k=i}^{M} \left| r_{\mu+k}^{*} s_k r_{\mu+k-i} s_{k-i}^{*} \right| - \sum_{k=\mu+i}^{\mu+M} |r_k||r_{k-i}| \right\} \quad (2)$$

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In Equation (2), the M-factor represents the number of iterations in the double-correlation synchronization algorithm. Increasing the M value enhances the performance of the algorithm, but it also leads to a substantial increase in computational complexity. To strike a balance between synchronization robustness and computational complexity, we have selected M=4 for the subsequent experiments.

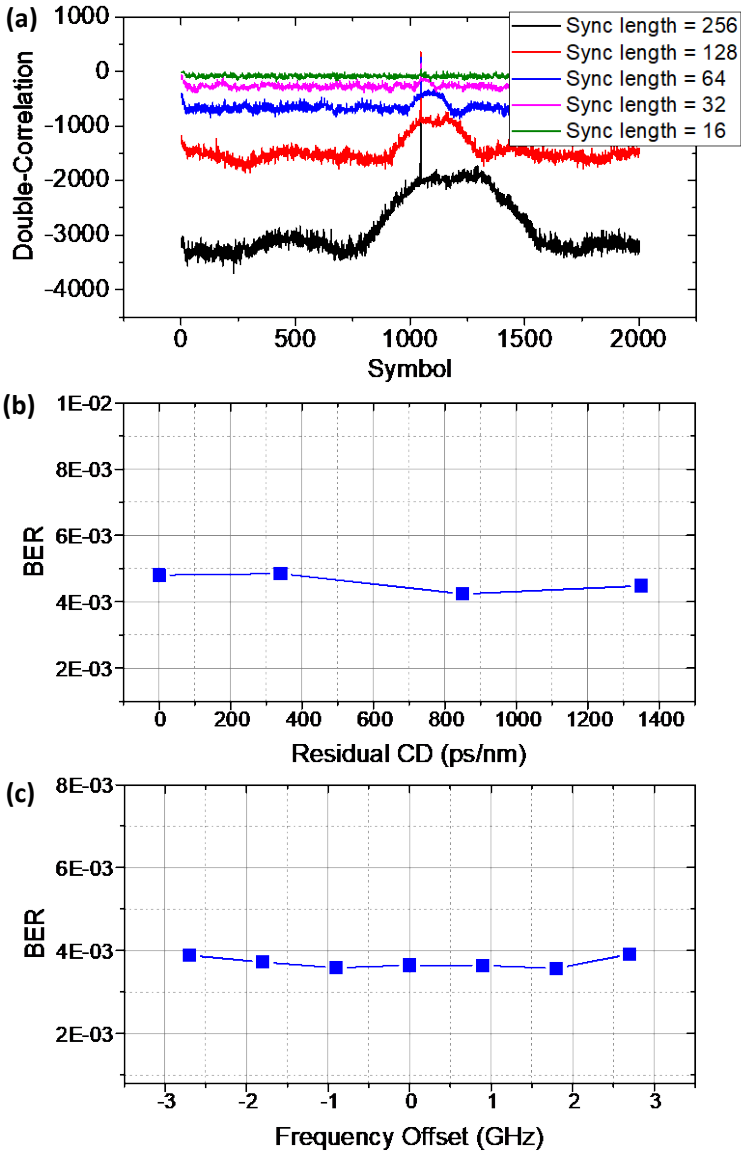


Figure 9 – TFDM burst optimization and performance evaluation: (a) Double-Correlation pattern in TFDM burst operation under different synchronization symbol length; (b) TDM burst BER performance under residual CD; (c) TDM burst BER performance under CFO

To examine the effect of the burst preamble design on the performance of the TFDM system, we conducted experiments using various lengths of preambles. For the Rx settling section, we utilized Rx settling length in number of symbols of 1024, 512, and 256, which correspond to time durations of 163.84 ns, 81.92 ns, and 40.96 ns, respectively (assuming a symbol rate of 6.25 GBd, where 1 symbol equals 0.16 ns). To achieve reliable synchronization, we tested sync length of 256, 128, 64, 32, and 16 symbols, corresponding to time durations of 40.96 ns, 20.48 ns, 10.24 ns, 5.12 ns, and 2.56 ns, respectively. In

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 9 (a), a collection of double-correlation patterns illustrates the TFDM burst operation at different synchronization lengths: 256, 128, 64, 32, and 16 symbols. It is worth noting that the length of the Rx settling does not significantly impact the synchronization results. Consequently, the results presented in Figure 9 (a) employ an Rx settling length of 512. Based on the observations in Figure 9 (a), for a reliable and consistent synchronization of the TFDM coherent burst, the synchronization length in symbols should be greater than 32, corresponding to a duration of 5.12 ns. We also evaluated the performance of the proposed synchronization algorithm for the TFDM burst receiver in the presence of transmission impairments, such as CD or CFO. Figure 9 (b) illustrates the system bit error rate (BER) under different residual CD values caused by varying fiber transmission distances. The results indicate that the double-correlation synchronization algorithm ensures stable operation even with uncompensated CD, as no significant impact on performance was observed. Figure 9 (c) presents the BER results at different CFO values. Similar to the CD test results, the presence of CFO does not affect synchronization or system BER performance. It is important to note that the TFDM burst receiver incorporates channel recognition and frequency-window calibration processes that will be discussed in the following section. These processes enable accurate calibration of TFDM subcarrier frequency windows and enhance the receiver's tolerance to carrier frequency detuning.

3.2. Subcarrier Recognition, Pre-equalization, and Power Rebalancing

In TFDM, laser drifting leads to subcarriers not being at their desired locations, especially in the ONU upstream direction in the case of different ONUs have independent laser sources. Drifting and collisions between adjacent subcarriers can result in crosstalk and must be avoided to prevent system outages. On the other hand, At the OLT receiver's end, power imbalances between DSCs coming from different end nodes result in additional penalties. To overcome the challenges posed by power imbalances between subcarriers in a system, as well as random frequency drifting caused by laser diodes in different ONUs, a novel TFDM burst transceiver algorithm has been proposed and experimentally demonstrated. This transceiver incorporates several key features, including subcarrier pre-equalization, power rebalancing, and frequency window detection and calibration. Figure 10 illustrates the processes involved in the proposed CPON TFDM receiver and transmitter, which consist of three main functions: 1) channel recognition and frequency window calibration (CR-FC), 2) channel pre-equalization (Pre-EQ), and 3) channel power rebalancing (PR).

Figure 10 (a) provides a detailed description of the CPON TFDM burst receiver in the OLT. In the receiver, CR-FC occurs between the analog-digital conversion (ADC) and subcarrier down conversion stages. The CR-FC process begins by extracting a set of 4096 points from the received data stream, which are then transformed into the frequency domain using Fast Fourier transform (FFT). The next step involves smoothing the channel distribution in the frequency domain and generating a binary spectrum map. In this binary map, 0s represent signal components below a preset threshold, while 1s represent signal components above the threshold. The binary spectrum map, along with a set of predefined rules, helps eliminate noise peaks, optical carrier residue, and harmonics in the frequency domain. The channel boundaries are determined by identifying abrupt changes on the binary map. By locating rising and falling edges on each frequency channel, the carrier frequencies are calculated, and the frequency windows are calibrated accordingly.

With CR-FC completed, the optical power in each subcarrier is estimated by integrating the square of the absolute signal values within each calibrated frequency window. This power estimation is then transmitted downstream and utilized for subcarrier power rebalancing in subsequent processes. The next steps in the receiver include digital down conversion, baseband filtering, burst detection by locating and synchronizing the burst signals in each subcarrier, and coherent DSP for upstream signal recovery. Within the subcarrier DSP, the channel response is extracted during the execution of the constant modulus

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

algorithm (CMA). The OLT sends control signal feedback to the ONU, which includes the estimated power information obtained from the CR-FC step and the channel response acquired from CMA. Figure 10 (b) illustrates the DSP process in the ONU transmitter, where the control signal feedback received from the OLT is utilized for performing Pre-EQ and PR during subcarrier data processing. Using the channel response information provided by the control signal, Pre-EQ is implemented by applying a reverse channel response to the signals. This step adjusts the signal amplitudes in the ONU transmitter based on the received channel response. Subsequently, the ONU transmitter utilizes the estimated power information obtained from the CR-FC step to further adjust the signal amplitudes. This adjustment ensures a completely rebalanced optical power distribution among the subcarriers.

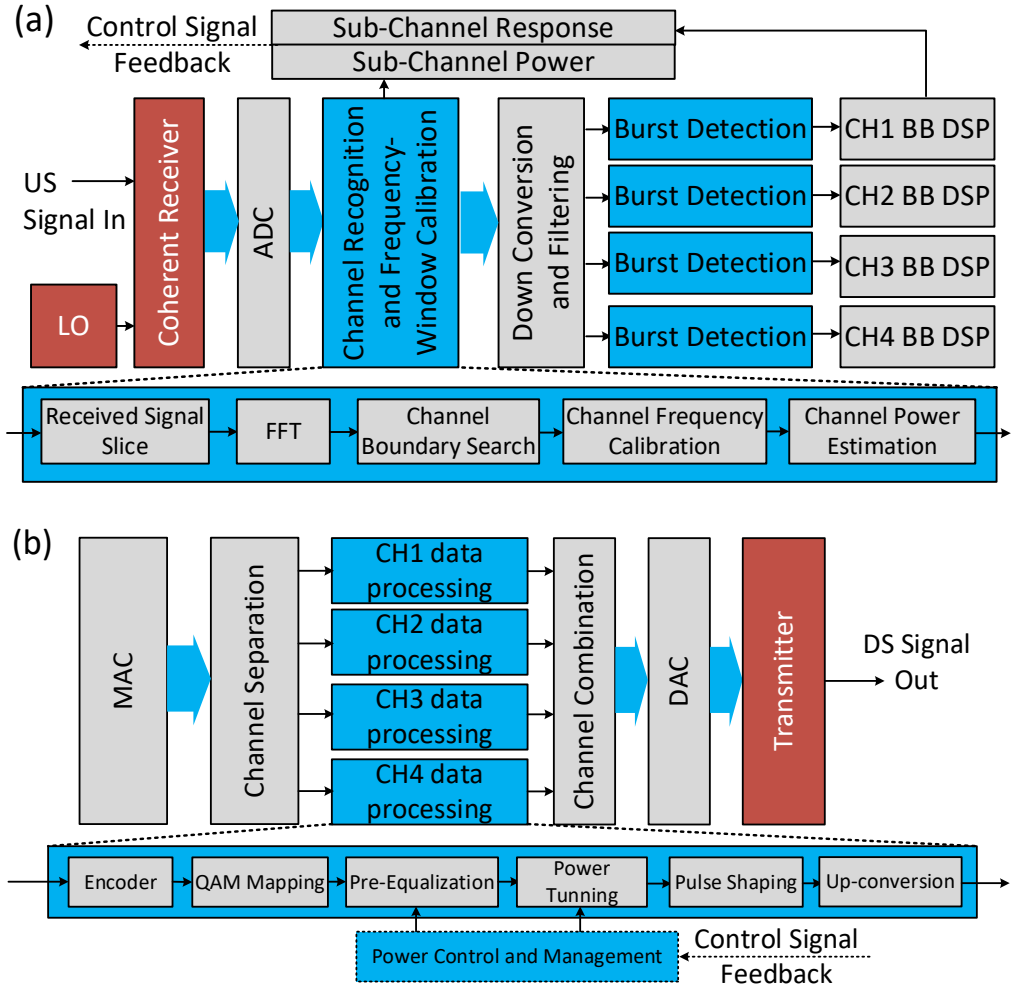


Figure 10 – Proposed DSP procedures featuring subcarrier recoginition, pre-equalization, and power rebalancing in: (a) TFDM coherent burst receiver; (b) TFDM coherent transmitter

Figure 11 displays the electrical spectra of TFDM signals and illustrates the process of subcarrier recognition, pre-equalization, and power rebalancing. In the unprocessed TFDM signal depicted in Figure 11 (i), the spectrum and power distribution in the frequency domain are uneven. The first step of the process involves performing CR-FC to identify the boundaries of the subcarriers, as shown in Figure 11 (ii). Following CR-FC, subcarrier pre-equalization (Pre-EQ) is carried out by applying the reverse response of the subcarrier information obtained from the CMA step. This Pre-EQ step results in a

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

flattened signal within each subcarrier, as illustrated in Figure 11(iii). Once the subcarrier boundaries are identified and accurate power control is achieved through Pre-EQ, the final step is to rebalance the power in each subcarrier by adjusting the transmitter signal amplitude at the ONUs. As a result of this power rebalancing, a fully rebalanced TFDM signal spectrum is obtained, as depicted in Figure 11 (iv).

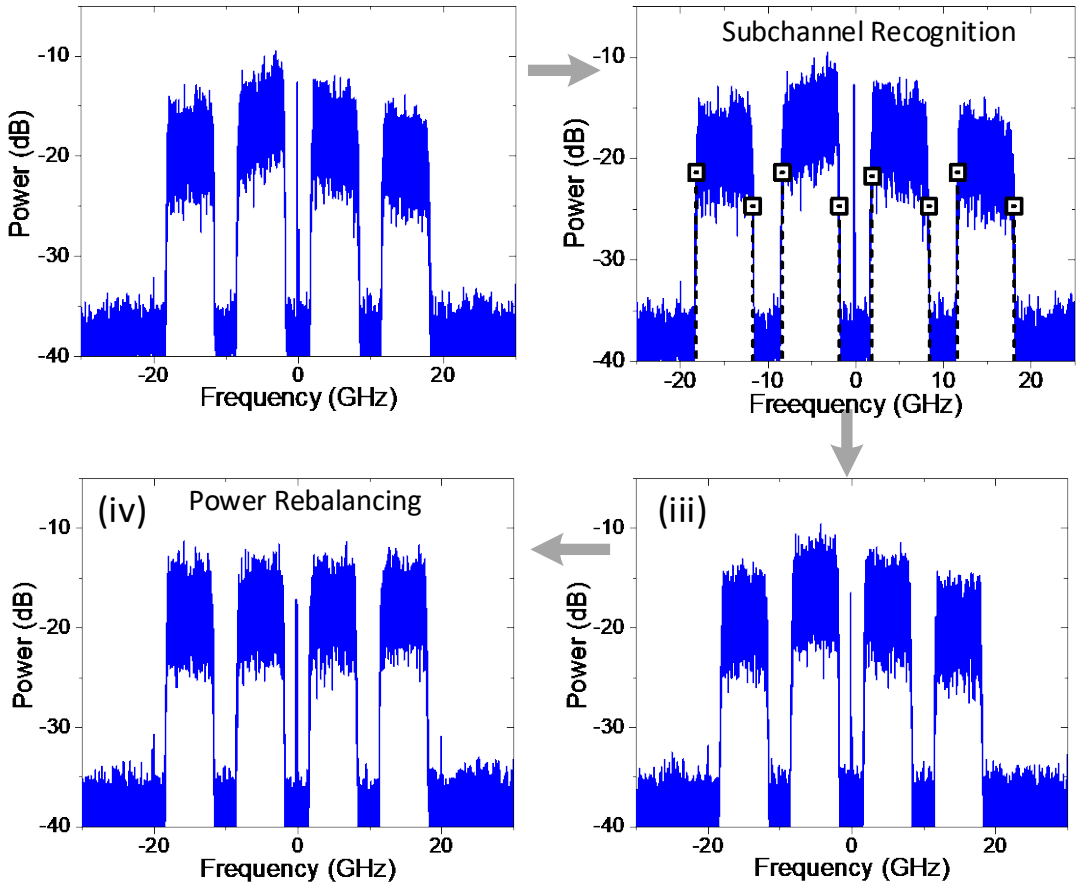


Figure 11 – Process of CR-FC, Pre-EQ, and PR of TFDM signal

Figure 12 demonstrates the performance of the proposed CR-FC algorithm. In Figure 12 (a), we intentionally shifted the center frequency of one subcarrier (CH3) by 2.2 GHz from its original frequency. With the implementation of CR-FC, the boundaries of CH3, as well as other channels, were successfully identified. Figure 12 (b) presents BER performance of upstream transmission plotted against CH3 frequency detuning, ranging from -2.4 to +2.4 GHz. To assess the performance of CR-FC, we compare three methods for CFO estimation and correction: 1. Conventional fourth-power carrier estimation (FPCE) at baseband (BB) DSP after down conversion; 2. Frequency-pilot-tone-based CFO estimation, where pilot tones are inserted at -20, -10, 10, and 20 GHz, followed by FPCE in the BB DSP; and 3. The proposed CR-FC followed by FPCE in the BB DSP. Based on the BER performance depicted in Figure 12 (b), using only FPCE exhibits the lowest CFO tolerance, primarily due to the reduction in frequency detection range caused by the fourth-power operation. Pilot-tone + FPCE improves CFO estimation and correction performance. However, when encountering a large CFO, the pilot tones end up overlapping with adjacent sub-channels, limiting the CFO tolerance. In contrast, the proposed CR-FC + FPCE demonstrates significantly improved CFO tolerance and outperforms the other two methods. The CR-FC algorithm, showcased in Figure 12, provides subcarrier detection and frequency window calibration

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

across a wide frequency range. It plays a crucial role in facilitating dynamic burst detection and enhances the TFDM system's ability to tolerate CFO variations.

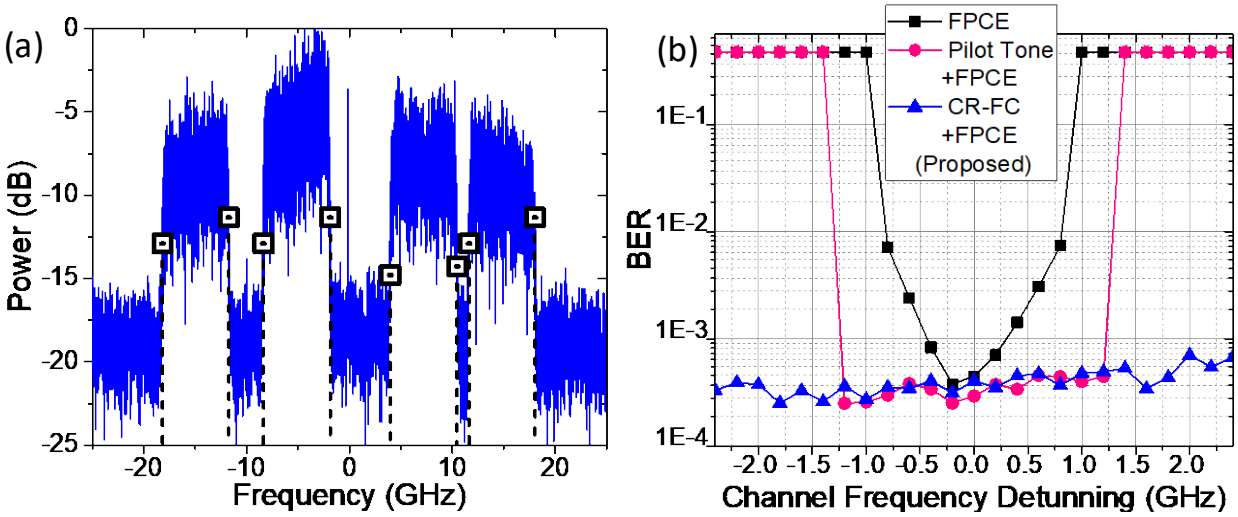


Figure 12 – CR-FC algorithm performance: (a) channel boundary detection with large CFO; (b) BER versus frequency detuning using different methods

Figure 13 illustrates the test results of BER performance versus received optical power (ROP) for a TFDM CPON optical link utilizing 80 km of single-mode fiber (SMF). Figure 13 (a) displays the BER versus ROP curves for all four subcarriers transmitted from two ONUs, both with and without the implementation of Pre-EQ and PR algorithms. Without Pre-EQ and PR, a noticeable difference in power budget of up to 4 dB can be observed among the subcarriers. However, with the application of Pre-EQ and PR, the power budget difference between the subcarriers is minimized, resulting in an overall improvement in system performance. Figure 13 (b) showcases the BER versus ROP performance for the system with Pre-EQ and PR in two scenarios: back-to-back (B2B) transmission and 80 km fiber transmission. In both cases, for all ONUs and subcarriers, the transmission over the 80 km fiber link does not introduce any significant penalty in terms of BER performance. The plot also includes reference points such as the staircase hard-decision (HD) forward error correction (FEC) threshold (BER = 4.5E-3) [18] and the concatenated soft-decision (SD) FEC threshold (BER = 1E-2) [19] for comparison. Overall, the results depicted in Figure 13 highlight the positive impact of Pre-EQ and PR algorithms on minimizing power budget differences between subcarriers and improving the BER performance of the TFDM CPON optical link, even over long-distance fiber transmissions.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

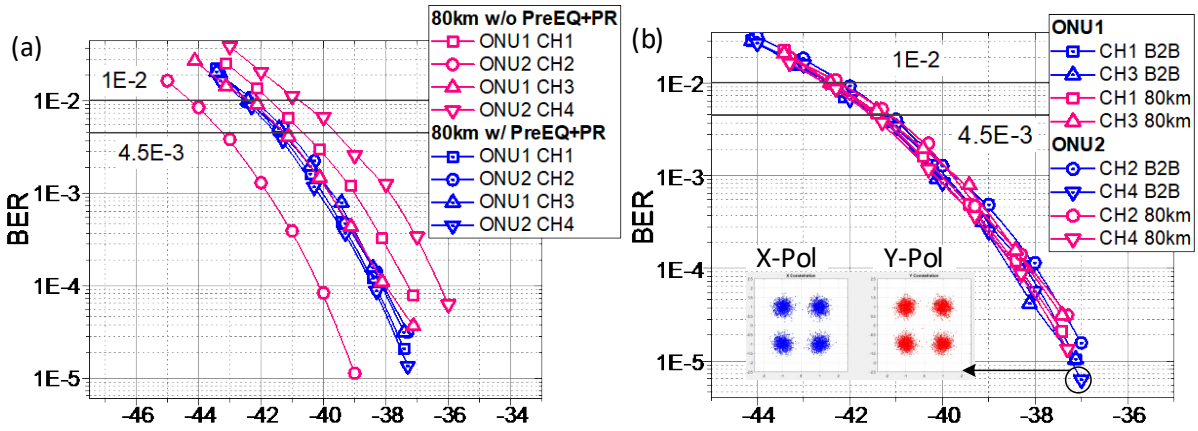
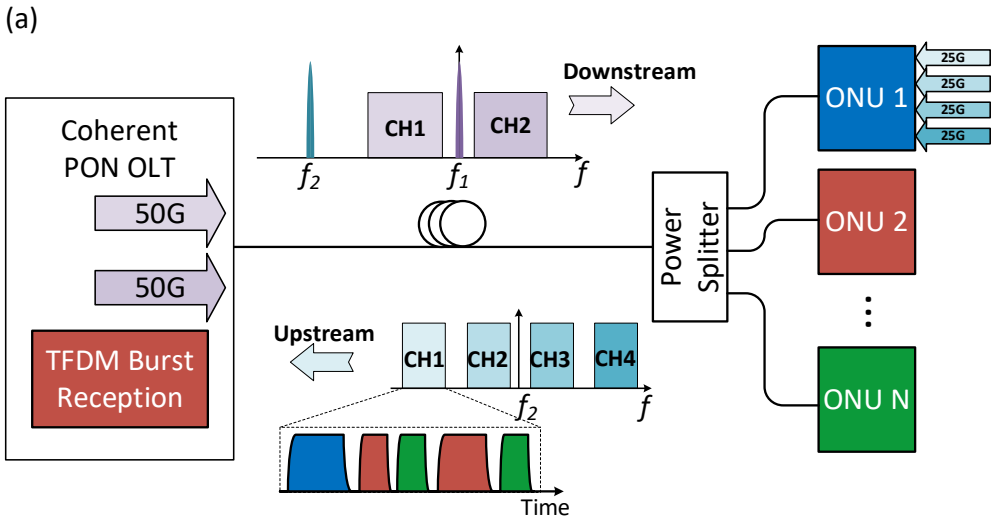


Figure 13 – (a) BER performance with and without Pre-EQ + PR; (b) BER performance

3.3. ONU Cost Reduction Through Remote Optical Carrier Delivery

Despite efforts to simplify coherent optics for short-haul applications, the high cost of current products remains the primary barrier to widespread adoption in access networks. One major cost contributor is the use of high-quality optical sources, such as external cavity lasers (ECLs) which are essential components of coherent modules. However, optical injection locking (OIL), a technique that locks a laser to an external optical tone through injection of external light, presents an opportunity for low-cost coherent optics in PON applications, especially for customer premise devices. Leverage OIL and TFDM subcarriers, we propose and experimentally demonstrate a new TFDM CPON architecture that involves delivering two optical tones to the ONUs from a remote location. By amplifying these optical tones using OIL, we can provide the ONU with an optical carrier and a local oscillator (LO), effectively replacing expensive lasers like ECLs with affordable Fabry-Perot laser diodes (FP-LDs). Experimental investigation shows that the proposed architecture performs comparably to a regular ECL-based system, without any significant degradation in performance. Furthermore, this architecture offers the advantage of frequency synchronization between the ONU light sources and the OLT lasers, which helps mitigate random frequency drifts caused by independently operated lasers in ONUs.



© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

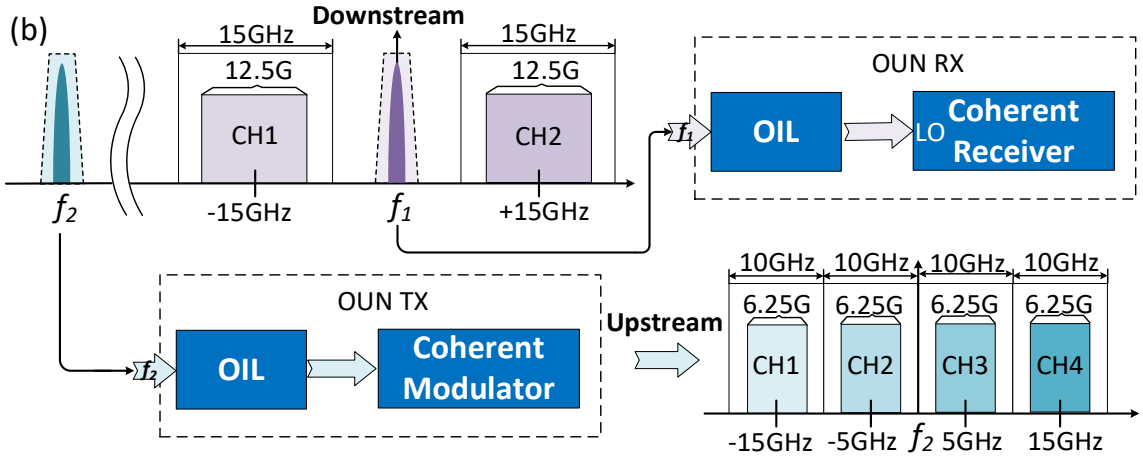


Figure 14 – (a) TFDM CPON scheme with remote optical tone delivery; (b) subchannel design for DS and US transmission

Figure 14 (a) illustrates the basic concept of the proposed TFDM CPON architecture with remote optical tone delivery. By utilizing the flexible bandwidth allocation capability of TFDM CPON, various subcarrier configurations can be employed to accommodate different capacities and link budgets. In the downstream direction, two TFDM subcarriers are generated on a single wavelength in the OLT. Two unmodulated optical tones centered at frequencies *f1* and *f2* for remote delivery are combined with the TFDM subcarriers for downstream transmission. One of the optical tones at frequency *f1*, which matches the optical carrier frequency of the downstream TFDM signal, is used to generate the LO for downstream signal detection at the ONU through OIL. The other optical tone at frequency *f2* is employed to generate an optical carrier through OIL at the ONU for upstream signal transmission. Importantly, both tones are amplified during the OIL process, eliminating the need for an additional optical amplifier. In the upstream direction, four TFDM subcarriers are generated at the ONU, utilizing a carrier frequency of *f2*. Figure 14 (b) depicts the arrangement of the TFDM subcarriers. The two subcarriers designated for downstream transmission occupy two 15 GHz frequency windows centered at -15 and +15 GHz with respect to *f1*. Each downstream subcarrier is modulated with a 50 Gb/s dual-polarization (DP) quadrature phase shift keying (QPSK) signal at a baud rate of 12.5 GBd, resulting in an aggregate downstream data rate of 100 Gb/s. In the upstream direction, the four subcarriers are each distributed within a 10 GHz frequency window, with center frequencies at -15, -5, +5, and +15 GHz relative to *f2*. Each upstream subcarrier is modulated with a 25 Gb/s DP-QPSK signal at a baud rate of 6.25 GBd, yielding an aggregated upstream data rate of 100 Gb/s. The frequency spacing between *f1* and *f2* is set to 100 GHz to align with the ITU DWDM (dense wavelength division multiplexing) frequency grid.

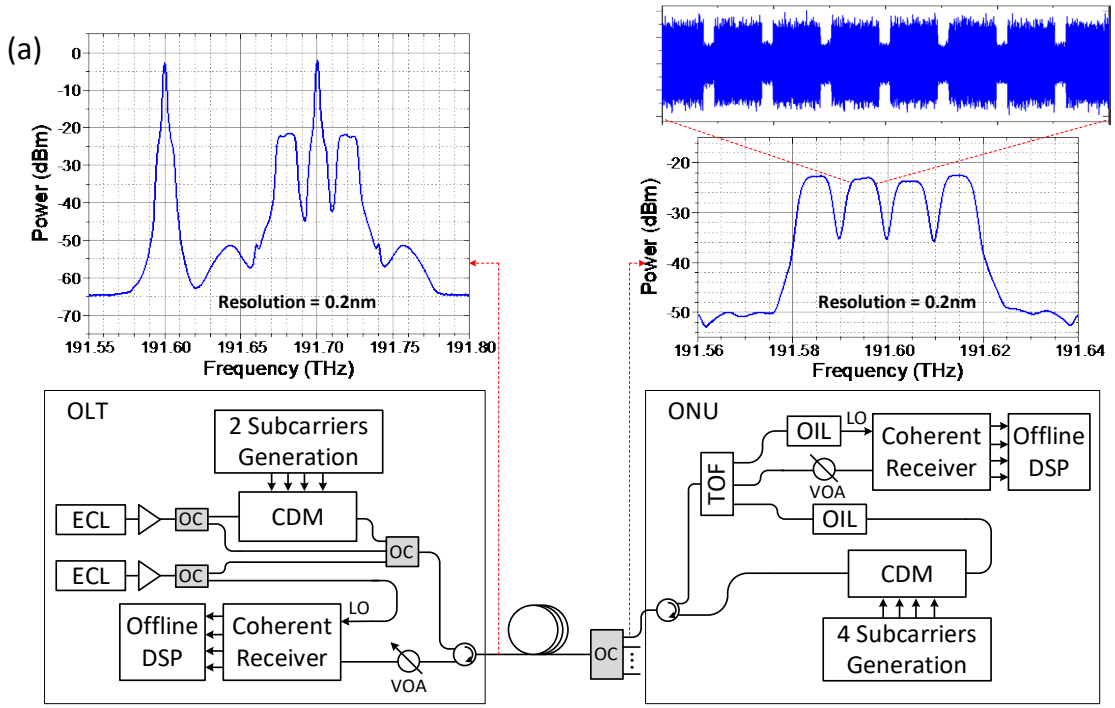
Figure 15 (a) shows the experimental configuration of the proposed TFDM CPON. Two ECLs serve as the light sources at the OLT. One ECL generates downstream TFDM signals using a coherent driver modulator (CDM), while the other ECL provides the LO for upstream signal detection at the OLT receiver. The ECL outputs also serve as optical master tones for injection locking at the ONUs. To establish the experimental setup, a 50 km fiber link and a 1x32 passive optical splitter are employed for the optical distribution network (ODN). At the ONU side, a multiport tunable optical filter (TOF) is utilized to separate the downstream TFDM signals and the two optical tones. The downstream TFDM signals are detected by a coherent homodyne receiver using the LO generated through OIL. The upstream TFDM signals, on the other hand, are transmitted using another OIL setup coupled to a CDM. Both the downstream and upstream signals undergo offline DSP using appropriate codes. The optical spectra of the downstream (broadcast) and upstream (burst) TFDM signals are depicted in the insets of Figure 15 (a). Although the experiment employs off-the-shelf discrete components for demonstration purposes, it is

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

worth noting that advanced photonic integration platforms have the potential to combine these components, leading to the development of low-cost commercial products.

To demonstrate the effectiveness of using an OIL laser as the ONU LO for downstream signal detection, the BER performance was measured against ROP per channel for downstream TFDM subcarriers. The experiments were conducted over a 50 km fiber transmission with a 32-way optical splitter, and the results of one subcarrier (CH1) are shown in Figure 15 (b). Note that the performance of the other TFDM subcarrier (CH2) is almost identical and is therefore not shown here for simplicity. Additionally, back-to-back (B2B) BER versus ROP measurement was carried out using the OIL LO. The performance of the system using a regular ECL as the LO was also included in the charts, along with reference thresholds for staircase HD FEC (BER = 4.5E-3) and concatenated SD FEC (BER = 1.2E-2). The goal was to compare the performance of the proposed architecture with remotely delivered optical tone and OIL for ONU LO against the performance of a system using a regular ECL LO. The experimental results demonstrated the functionality of the proposed architecture in downstream transmission. It was observed that there was no significant degradation in performance when using the remotely delivered optical tone in conjunction with OIL for ONU LO, as compared to using a regular ECL LO.

Figure 15 (c) presents the results of BER versus ROP per channel for upstream burst TFDM signals (CH1, other subcarriers perform nearly identical and will not be shown here) using an OIL-based ONU transmitter enabled by the remotely delivered optical tone. The results include both fiber transmission (50 km/32 split) and B2B scenarios. Additionally, the BER performance of the subcarrier using a regular ECL as the transmitter laser is included for comparison. Similar to the downstream broadcasting results discussed earlier, the upstream burst transmission employing the proposed OIL-based transmitter demonstrates minimal performance degradation. Overall, the experimental findings indicate that the proposed OIL-based transmitter/LO, utilizing the remotely delivered optical tone, achieves comparable performance to a traditional ECL-based transmitter/LO in terms of BER, both at the HD-FEC threshold and the SD-FEC threshold.



© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

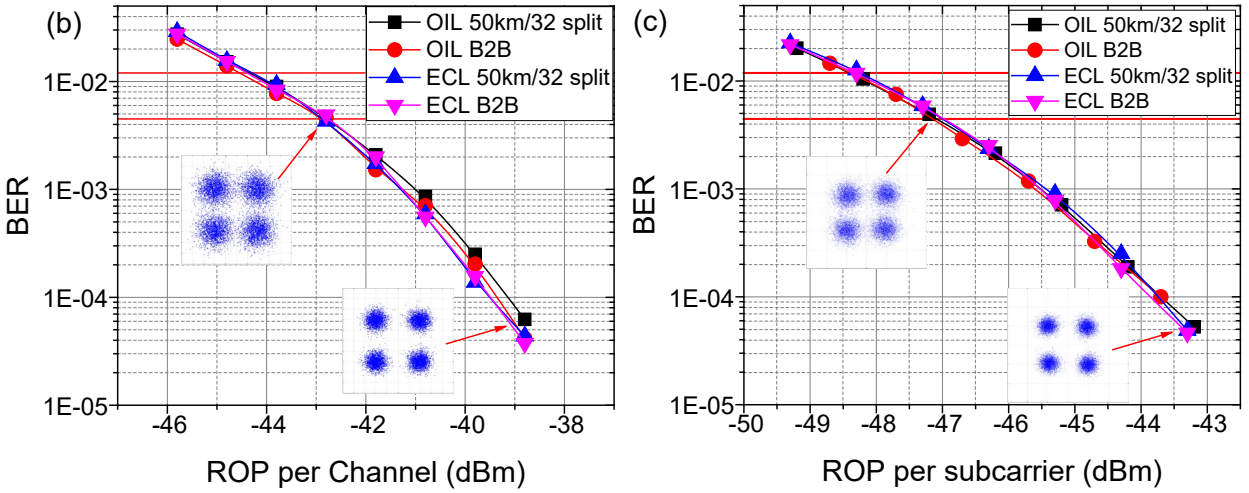


Figure 15 – (a) Experimental setup with optical spectrums of TFDM DS and US channels; (b) BER versus ROP in downstream; (c) BER versus ROP in upstream

The proposed OIL-based system offers an additional advantage by achieving frequency locking between the ONU transmitter/LO and the OLT light sources. This leads to minimal optical frequency offset between them. In contrast, a regular ECL-based system exhibits a much larger carrier frequency offset (CFO) of approximately 0.34 GHz, as depicted in Figure 16 (a). This substantial CFO makes signal recovery impossible without CFO compensation. On the other hand, the proposed OIL-based system demonstrates a residual CFO of only 0.12 MHz, enabling simplification of the coherent receiver DSP by eliminating the need for CFO compensation. This simplification is achieved without significant performance degradation. Figure 16 (b) illustrates the BER performance of the OIL-based system without CFO compensation for both downstream and upstream burst transmissions. In comparison to the ECL-based system with CFO compensation, the OIL-based system exhibits similar performance levels. However, it offers the significant advantages of simplified receiver DSP complexity and cost savings in ONU hardware.

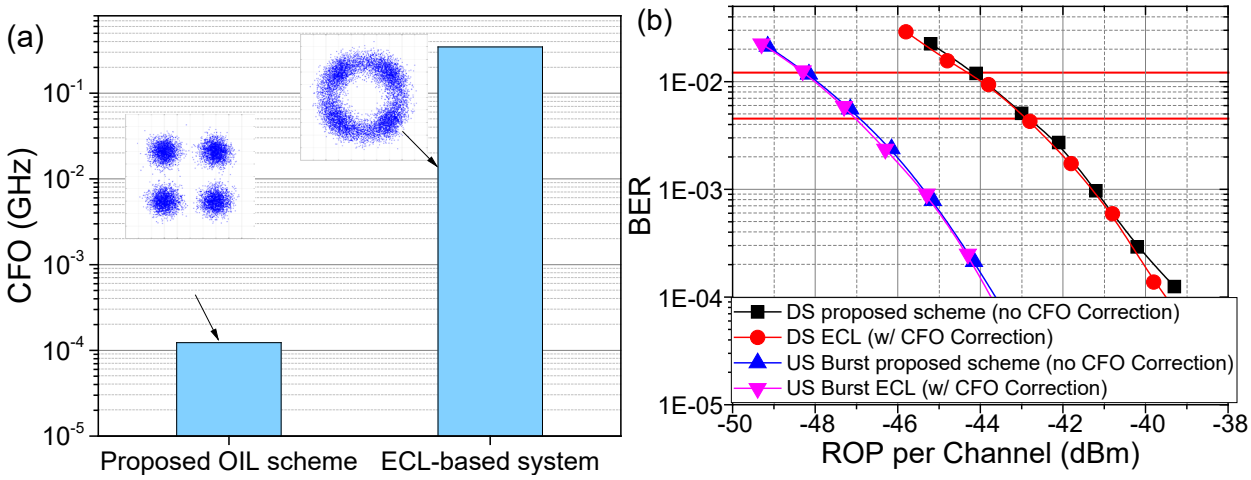


Figure 16 – (a) Residual CFO for proposed OIL scheme vs. regular ECL-based system; (b) BER vs. ROP per channel for proposed OIL scheme without CFO correction compared with regular ECL-based system with CFO correction

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. Advantages, Challenges, and Discussions

DSC based TFDM is a promising advancement in optical transmission technology and architecture that leverages subcarrier implementation in digital domain to enhance data transmission flexibility and efficiency. In this part, the benefits of subcarrier implementation in TFDM are summarized, including efficient bandwidth granularity and utilization, flexible link budget designs, low-latency dedicated bandwidth services, reduced power consumption, and scalable architecture. However, the implementation of TFDM also introduces challenges such as subcarrier frequency and gain control, impairment calibration to mitigate frequency mirroring, increased complexity of DSP and scheduling control, and transmitter power reduction due to modulation clipping, as shown in Figure 17. Balancing these benefits and challenges is crucial for the industry to harness the long-term potential of TFDM while addressing current market demands and infrastructure readiness.

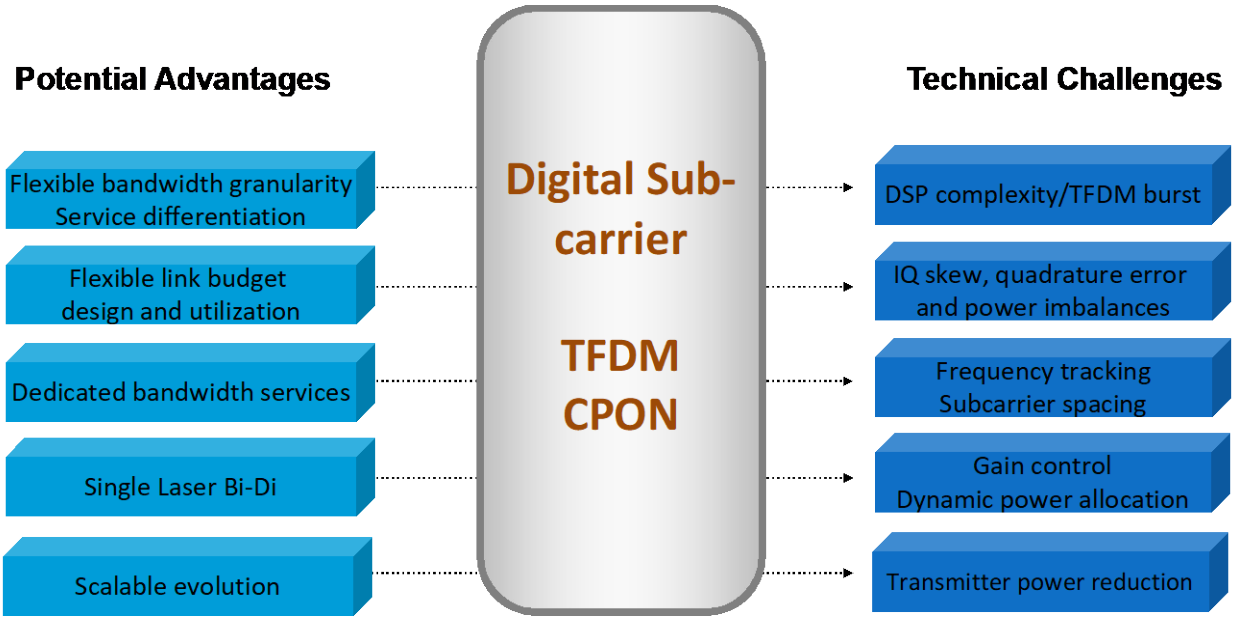


Figure 17 – Technical summary of DSC based TFDM CPON

DSC implementation can enhance bandwidth utilization compared to traditional single-carrier options. The finer granularity of subcarriers, leads to finer data capacity quantization, thereby reducing data wastage and allowing for increased data throughput within a given bandwidth allocation. Subcarrier implementation opens the door to flexible link budget in the PON designs, fostering improved network performance. The allocation of subcarriers can be optimized to achieve specific outcomes. For example, by allocating 50% of the subcarriers to carry 50% of the overall aggregated capacity, a 3dB additional link budget can be achieved by transmitting the similar amount of optical power as 100% subcarriers. Additionally, the emphasis of subcarrier power in the downlink can be effectively managed to balance power allocation across the network.

A subcarrier model provides a way to achieve low latency and/or high security dedicated bandwidth services through segmentation in subcarriers and their configuration for specific service needs while utilizing the same optical PON infrastructure. This model can be reconfigured on a per ODN basis, enabling tailored services to meet varying demands. DSC implementation also contributes to reduced power consumption, particularly during periods of lower peak bandwidth demands. The design of a DSP that scales power utilization based on active DSC usage enhances energy efficiency. The incorporation of

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

DSCs establishes a scalable architecture that allows for independent expansion of OLT and ONU bandwidth options through increasing the supported number of subcarriers.

The flexibility and efficiency of this TFDM approach come with technical implementation challenges. Despite the substantial advantages subcarriers provide, certain hurdles emerge, particularly in the domain of subcarrier frequency control. The possibility of collisions between neighboring subcarriers introduces the risk of crosstalk, which could potentially lead to disruptions in the system's operation. To mitigate such issues, the allocation of adequate guard bands and meticulous management of laser jitter become essential prerequisites. Ensuring uniform signal strength across DSCs is also essential for optimal system performance. A per subcarrier gain control mechanism on both the transmitter and receiver sides is necessary to equalize the power of received DSC channels before subsequent DSP processing [20].

Additional impairments stemming from manufacturing and initialization processes, such as frequency mirror image due to IQ skew, quadrature error, and IQ power imbalances, need to be accurately calibrated out to maintain signal integrity and quality. Further, modulation clipping, resulting from the enhanced peak-to-average power ratio (PAPR) in the multiple subcarrier scheme, can lead to a reduced fiber link budget due to the reduction in the transmitter output power. Careful system-level consideration and optimization are necessary to mitigate this impact. It is also noted that the introduction of subcarriers introduces complexity to DSP and scheduling control systems. The need for 2-dimensional resource scheduling to manage diverse subcarriers adds intricacy to system design and management.

DSC TFDM has the potential to transform the landscape of coherent optical technology application in the access networks. The benefits of subcarrier implementation, including enhanced bandwidth utilization, flexible link budget designs, and low-latency services, are compelling. However, the challenges related to subcarrier frequency and gain control and others must be addressed. Achieving equilibrium between the enduring merits of TFDM and the pragmatic factors of market demand, infrastructure preparedness, and economic viability is pivotal to shaping the future of coherent optical transmission systems and networks.

5. Conclusions

In this paper we present the TFDM technology in the context of CPONs. The CableLabs CPON project, initiated in May 2021, aims to future-proof cable's access architecture by developing specifications for 100G PONs and multi-vendor interoperable devices. Currently two data multiplexing methods are being considered for emerging CPONs: TDM and DSC-based FDM combined with TDM. While the traditional TDM approach has been widely deployed in existing PONs, it faces challenges in future CPONs due to potentially high scheduling latency and limited flexibility. In contrast, the TFDM solution, which combines the benefits of TDM and FDM by utilizing subcarriers in the frequency domain, enables the allocation of different frequency bands to network services or ONU groups with varying latency and capacity requirements while maintaining the simplicity of traditional grant request and bandwidth allocations techniques. This approach eliminates the need of contention resolution, significantly improves scheduling latency and traffic blocking rates, and facilitates the coexistence of mobile and video streaming services on a converged CPON platform with enhanced flexibility. The advantages associated with incorporating TFDM subcarriers, such as improved bandwidth utilization, adaptable link budget configurations, low power consumption, and dedicated channel for low-latency services, are compelling. Nevertheless, the hurdles linked to managing subcarrier frequencies and gains, among other technical challenges, necessitate resolution. Attaining a balance between the strengths of TFDM and the practical considerations of market requirements, infrastructure readiness, and economic feasibility stands as a critical factor in shaping the development of coherent optical transmission systems and networks in the future.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

This paper reviews the technology concept of TFDM CPONs, which support asymmetric ONU/OLT hardware configurations and pay-as-you-go ONU cost based on bandwidth subscription. Unlike traditional WDM-PON or TWDM-PON, the coherent TFDM-PON system operates at the same wavelength grid with slight frequency tuning, ensuring compatibility with widely deployed TDM-PONs and power splitter-based ODNs. Additionally, key technology developments in TFDM CPONs are discussed: Firstly, a burst transmitter/receiver for TFDM CPON enables the time domain multiplexing and processing inside DSCs. Secondly, frequency window detection and calibration schemes, along with pre-equalization and phase recovery, enhance the tolerance to carrier frequency detuning and effectively minimize the power budget differences between subcarriers. Lastly, a novel TFDM CPON architecture addresses a major challenge in deploying coherent optics for access networks by providing low-cost ONU light sources through remote master tone delivery and optical injection locking. The major advantage is to have the opportunity of an asymmetric design where complexity, performance and capabilities are moved to the OLT while the ONU complexity is kept low.

The advancements achieved in TFDM technology establish it as a strong contender for forthcoming next-generation CPONs, offering capacities of 100G and beyond. This promises adaptable bandwidth distribution across temporal and spectral dimensions over a single coherent transceiver. Furthermore, this study delves into the advantages, challenges, and potential remedies linked to DSC integration within TFDM systems. By tackling these challenges in conjunction with sound business strategies, it becomes very promising that TFDM-based CPONs possess the capability to fortify optical access architecture for the future.

6. Acknowledgements

Special acknowledgments for fruitful discussions and assistance with all Aspects of this study: Dr. Curtis Knittle, Chris Stengrim, Matt Schmitt, Karthik Sundaresan, Steve Goeringer as well as the dedicated members of the Working Group (WG) and Operator Advisory Group (OAG) from CableLabs CPON Project.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

ADC	analog-to-digital converter
B2B	back-to-back
BER	bit error rate
CD	chromatic dispersion
CDM	coherent driver modulator
CFO	carrier frequency offset
CMA	constant modulus algorithm
CPON	coherent passive optical network
CPR	carrier phase recovery
CR-FC	channel recognition and frequency window calibration
DAC	digital-to-analog converter
DSC	digital subcarrier
DSP	digital signal processing
DWDM	dense wavelength division multiplexing
ECL	external cavity laser
FEC	forward error correction
FFT	Fast Fourier transform
FOE	frequency-offset estimation
FP-LD	Fabry-Perot laser diode
FTTH	fiber to the home
Gb/s	gigabits per second
IM-DD	intensity modulation and direct detection
ODN	optical distribution network
OIL	optical injection locking
OLT	optical line terminal
ONU	optical network unit
P2MP	point-to-multipoint
PAPR	peak-to-average power ratio
PMD	polarization-mode dispersion
PON	passive optical network
PR	power rebalancing
Pre-EQ	pre-equalization
QAM	quadrature amplitude modulation
QPSK	quadrature phase shift keying
ROP	received optical power
Rx	receiver
SOP	state of polarization
TDM	time-division multiplexing
TFDM	time-and-frequency-division multiplexing
TWDM	time-and-wavelength-division multiplexing
Tx	transmitter
TOF	tunable optical filter
WDM	wavelength-division multiplexing

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

[1] Z. Jia and L. A. Campos, “Coherent Optics for Access Networks,” Routledge & CRC Press, Nov. 01, 2019.

[2] Z. Jia and L. A. Campos, “Coherent Optics Ready for Prime Time in Short-Haul Networks,” IEEE Network, vol. 35, no. 2, pp. 8–14, 2021.

[3] J. Zhang and Z. Jia, "Coherent Passive Optical Networks for 100G/λ-and-Beyond Fiber Access: Recent Progress and Outlook," in IEEE Network, vol. 36, no. 2, pp. 116-123, March/April 2022.

[4] N. Suzuki, H. Miura, K. Matsuda, R. Matsumoto and K. Motoshima, "100 Gb/s to 1 Tb/s Based Coherent Passive Optical Network Technology," in Journal of Lightwave Technology, vol. 36, no. 8, pp. 1485-1491, 15 April15, 2018.

[5] IEEE 802.3ca, Physical Layer Specifications and Management Parameters for 25 Gb/s and 50 Gb/s Passive Optical Networks, 2020.

[6] ITU-T G.9804.2, Higher Speed Passive Optical Networks: Common Transmission Convergence Layer Specification, Sept. 2021.

[7] R. Borkowski et al., "Operator Trial of 100 Gbit/s FLCS-PON Prototype with Probabilistic Shaping and Soft-Input FEC," European Conference on Optical Communication (ECOC), 2021, pp. 1-4.

[8] R. Matsumoto, K. Matsuda, and N. Suzuki, “Fast, Low-Complexity Widely-Linear Compensation for IQ Imbalance in Burst-Mode 100-Gb/s/λ Coherent TDM-PON,” Optical Fiber Communication Conference (OFC) 2018, paper M3B.2.

[9] R. Koma, M. Fujiwara, J.-I. Kani, K.-I. Suzuki, and A. Otaka, “Burst-Mode Digital Signal Processing That Pre-Calculates FIR Filter Coefficients for Digital Coherent PON Upstream,” Journal of Optical Communications and Networking, vol. 10, no. 5, pp. 461–470, May 2018.

[10] M. Luo, D. Wu, W. Li, T. Zeng, L. Zhou, L. Meng, Z. He, C. Li, and X. Li, "Demonstration of Bidirectional Real-Time 100 Gb/s (4×25 Gb/s) Coherent UDWDM-PON with Power Budget of 44 dB," Optical Fiber Communication Conference (OFC) 2019, paper Th3F.2.

[11] K. Matsuda, R. Matsumoto, and N. Suzuki, “Hardware-Efficient Adaptive Equalization and Carrier Phase Recovery for 100-Gb/s/λ-Based Coherent WDM-PON Systems,” Journal of Lightwave Technology, vol. 36, no. 8, pp. 1492–1497, Apr. 2018.

[12] J. Zhang, Z. Jia, H. Zhang, M. Xu, J. Zhu and L. A. Campos, "Rate-Flexible Single-Wavelength TFDM 100G Coherent PON Based on Digital Subcarrier Multiplexing Technology," 2020 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 2020, paper W1E.5.

[13] D. Welch et al., "Point-to-Multipoint Optical Networks Using Coherent Digital Subcarriers," in Journal of Lightwave Technology, vol. 39, no. 16, pp. 5232-5247, 15 Aug.15, 2021.

[14] M. Xu, Z. Jia, H. Zhang, L. A. Campos and C. Knittle, "Intelligent Burst Receiving Control in 100G Coherent PON with 4×25G TFDM Upstream Transmission," 2022 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 2022, paper Th3E.2.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

[15] H. Zhang, Z. Jia, L. A. Campos and C. Knittle, " Low-Cost 100G Coherent PON Enabled by TFDM Digital Subchannels and Optical Injection Locking," 2023 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 2023, paper W1I.4.

[16] J. Zhang, Z. Jia, M. Xu, H. Zhang, L. A. Campos, and C. Knittle, “High-Performance Preamble Design and Upstream Burst-Mode Detection in 100-Gb/s/λ TDM Coherent-PON,” Optical Fiber Communication Conference (OFC) 2020, paper W1E.1.

[17] Z. Y. Choi and Y. H. Lee, "Frame synchronization in the presence of frequency offset," in IEEE Transactions on Communications, vol. 50, no. 7, pp. 1062-1065, July 2002.

[18] International Telecommunication Union (ITU-T) G.709.2 recommendation.

[19] Optical Interworking Forum 400G ZR standard.

[20] T. Duthel, et al, "DSP Design for Point-to-Multipoint Transmission," 2023 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 2023, paper W1E.1.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Scaling DAA: Smart, Continuous Network Health Monitoring for vCMTS with Machine Learning

A Technical Paper prepared for SCTE by

Ilana Weinstein
Data Scientist
Comcast
1800 Arch St, Philadelphia, PA
ilana_weinstein@comcast.com

Ramya Narayanaswamy
Director, Data Science
Comcast
1800 Arch St, Philadelphia, PA
ramya_narayanaswamy@comcast.com

Aaron Tomkins
Data Scientist
Comcast
1800 Arch St, Philadelphia, PA
aaron_tomkins@comcast.com

Nivedhitha Sridhar
Data Scientist
Comcast
1800 Arch St, Philadelphia, PA
nivedhitha_sridhar@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Methodology ... 3
   - 2.1. Anomaly Detection and Data Science ... 3
   - 2.2. Data Overview ... 4
   - 2.3. Solution Architecture ... 5
   - 2.4. Anomaly Detection Algorithms ... 6
     - 2.4.1. Cable Modem Online Signal ... 7
     - 2.4.2. Customer Calls (Interactive Voice Response) ... 7
     - 2.4.3. Quality of Experience ... 8
   - 2.5. Understanding Anomalies ... 10
3. Performance and Impact Findings ... 11
4. Future Work ... 12
   - 4.1. Labeling Tools ... 12
   - 4.2. Expanding Metrics Monitored ... 13
   - 4.3. Alarm Correlation and Causation ... 15
5. Conclusion ... 16

Abbreviations ... 17

Bibliography & References ... 18

List of Figures

Title **Page Number**

Figure 1 – Software Deployment and Continuous Monitoring Integration ... 5
Figure 2 – Anomaly Detection Architecture ... 6
Figure 3 – CM Signal Oscillating Post-Deployment Anomaly Example ... 7
Figure 4 – Spike in IVR Post-Deployment Anomaly Example ... 8
Figure 5 – QoE Hourly and Daily Patern Detected by Bayesian Framework ... 9
Figure 6 – QoE Display Slow Degredation Post-Deployment Example ... 10
Figure 7 – Traffic/Switch Anomaly Detection Development – Top Two Models on Leaf-RUR Traffic Metric ... 15

List of Tables

Title **Page Number**

Table 1 – Example Supporting Data for Anomaly (Corresponds to Figure 4) ... 11
Table 2 – Evaluation Score by Model for Traffic Metric (Leaf-MAGG) – Leading Methods Based on F1-score Highlighted ... 14

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

On Comcast's road to 10G, the virtual cable modem termination system (vCMTS) platform plays an integral part in ensuring the latest Data Over Cable Service Interface Specification (DOCSIS) technology is available to offer higher speeds and the best service to customers. The vCMTS platform is expanding rapidly, which enables enhanced flexibility, scalability, and cost-effectiveness. Therefore, ensuring the stability of this platform is critical, given its significance. However, the introduction of software upgrades often poses challenges for network operators, as it requires meticulous monitoring to detect and address any potential anomalies that may arise in the post-upgrade phase. To overcome these challenges, several systems have been put in place, such as the Automated Network Health Checks. These systems monitor the network's health immediately after deployments or software upgrades, and tools are available for alerting based on a set of key performance indicators (KPIs). While these instantaneous checks are valuable, certain KPIs and telemetry metrics may take hours to days to reveal underlying service degradation. At present, there is limited visibility into detecting these anomalous trends on Distributed Access Architecture (DAA), which could indicate a slow degradation of the platform's health. This paper proposes a comprehensive solution that leverages data science, machine learning, and big data techniques to continuously monitor and detect anomalous trends that could indicate a slow deterioration of platform health and then alert the operations team of any unusual activity.

2. Methodology

2.1. Anomaly Detection and Data Science

Identifying anomalies in data is an important task in many fields, and a variety of approaches have been developed over time to try to solve this problem. Understanding these various approaches to anomaly detection is crucial in determining the most suitable solution for the given problem. This section will briefly cover these approaches, their advantages in and out of the long-term network monitoring domain and background that contributes to the design of the anomaly detection system. Although different survey papers have categorized various machine learning approaches for anomaly detection in different ways, some of the more common approaches have been categorized as: classification-based, clustering-based, nearest-neighbor based, knowledge-based, statistical, and deep learning-based (Bhuyan et al. 2014 and Chandola et al. 2009).

The advantages and viability of a particular machine learning approach are often dependent on the problem context and format of the available data. Some problem contexts contain very high-dimensional data, while others are univariate; in some contexts, anomalies are clearly labeled in the training set, while in others they must be inferred. Classification-based approaches often work well when true anomaly labels and many features exist in the training data, while many statistical approaches can work well despite the absence of these. Furthermore, adjustments or modifications to an approach are often necessary to better fit the intricacies of a particular problem context.

In the context of detecting anomalous trends in some very different network metrics that don't have direct relationships with each other, there can be many benefits to utilizing a different approach that naturally goes with each metric. One benefit is it gives the freedom to customize each approach to that metric instead of limiting the options to modifications that only work in generality. However, as different as some metrics may be, they often can still benefit from and inform each other; thus, there is also an advantage in not keeping models siloed from each other. To that end, additional value can be achieved by building a higher-level aggregating model that is able to combine the insights from the metric-specific models and help produce an actionable summary of what is going on with the network. In this way, one

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

can both get interpretable information at a granular, metric-specific level as well as leverage the relationships between metrics to get a higher-level summary of the network overall.

2.2. Data Overview

The DAA architecture offers rich telemetry, from before the headend to the device. The DAA telemetry is multidimensional; each metric monitored in the anomaly detection system is a timeseries with different properties, both in terms of frequency (ranging from 15-second telemetry to hourly polls to event-based data), the infrastructure level, and the sources of the data. See Eppes et al. 2022 for more details around the DAA infrastructure. The infrastructure levels in scope for anomaly detection encompass the cable modem (CM), remote physical device (RPD), physical pod (PPOD), etc.

The anomaly detection pipeline leverages an existing telemetry collection process (Stehman et al. 2021) that is built on top of cloud computing tools like Apache Spark, which facilitates the parallel processing needed to operate on data at this scale in an efficient manner on a distributed framework. The process collects the data from each source by polling the different data sources on an automated schedule; it then aggregates the telemetry data on the PPOD-level and the RPD-level into a standardized format for further analysis. The four major metrics that are in the current scope of this project are: 1. CM Registration States 2. Interactive Voice Response (IVR) 3. Quality of Experience (QoE) 4. Traffic/Switch Metrics.

The CM registration status data consists of 15 second polls, aggregated up to a 5-minute level indicating the registration state of the CM at each polling interval. It captures the state of the CM's attempt to pair with the CMTS, polls the CM at periodic intervals, and assigns a status based on the response and the DOCSIS specifications of the modem. The polls are aggregated such that the highest impact state in a 5-minute window is considered. This data tends to have a pattern to it – specifically, the nightly reboots that some device types undergo during the maintenance window. These would need to be accounted for as non-anomalous but would also need to factor into an algorithm such that, during that time, if more devices than expected go into an offline state or begin to waver, that would still be flagged as an anomaly.

Next metric is the IVR metric, which comprises of support calls that are made by customers. The IVR is a technology that allows a computer to interact with people who call using voice and keypad inputs. These systems are used in servicing high call volumes, reducing cost, improving the customer experience, and providing a self-service experience. The system uses a routing mechanism to determine where the call is directed based on the customer input. It also collects information on whether there is a current outage in the area and provides that information to callers. High call volumes or specific call categories can indicate regions that are experiencing increased dissatisfaction when comparing call volume before and after a deployment. By identifying these pain points, the operations teams can take measures to further investigate where the issues are or corroborate with other metrics to determine the need for a rollback.

The QoE measure is an aggregate score which considers a wide range of metrics to determine the device's performance. Largely Wi-Fi based, it considers network performance factors such as the data rates, signal strength, packet error rate, any noise or interference on the channel, channel utilization broadband throughput, and the topology of the network. It also factors in the device type, capabilities of the device, DOCSIS version, current and historical data usage, and other specific environmental factors.

The last major group of metrics is the traffic metrics. These metrics serve as counters, keeping track of the packets flowing between various components within the switch leaf architecture. Specifically, the focus of these metrics lies in monitoring unicast traffic, encompassing packets transmitted between the residential U-Ring router (RUR), leaf switches, vCMTS host, and other elements of the architecture.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

When anomalies arise from this network connection or the RUR device, they indicate potential impairments or disruptions in the link between the RUR router and the leaf switch. Additionally, another pair of traffic metrics considered is the octet in/out counters, which quantify the total number of octets received or transmitted via the interface, including both header and frame characters. Discontinuities in the value of this counter can occur during management system re-initialization or other connection drops.

It is worth noting that traffic metrics exhibit distinct patterns over time. Generally, they demonstrate a diurnal cycle with peak values during the day and lower values during the night. Furthermore, traffic tends to increase during holidays and weekends. Hence, any effective time series anomaly detection model must possess the capability to comprehend these patterns, decompose the seasonal elements, establish a baseline, and subsequently identify data points that deviate from the expected norms. The use of these metrics will be further discussed in Section 4.2.

2.3. Solution Architecture

Analyzing near real-time telemetry on several KPIs for anomaly detection is a daunting task. To address this challenge, a sophisticated yet flexible workflow solution has been developed to apply an anomaly detection model to each KPI and trigger alerts to relevant stakeholders.

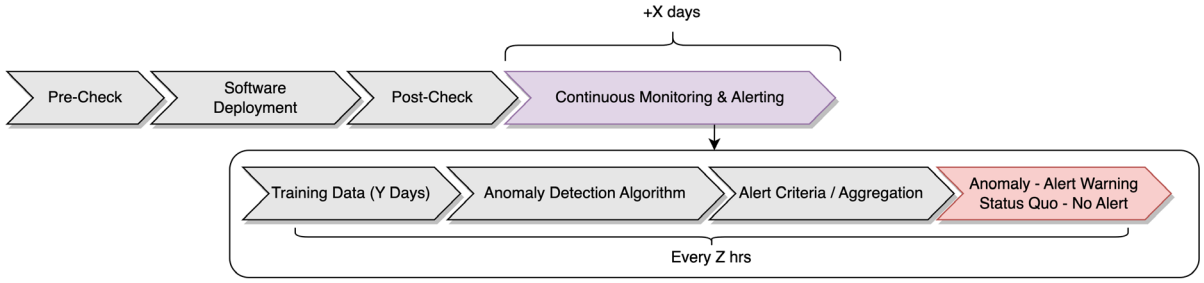


Figure 1 – Software Deployment and Continuous Monitoring Integration

The architecture uses a powerful analytics platform that has big data processing capabilities, machine learning tools. Aside from the platform's core data abilities, the anomaly detection solution is developed and deployed in this platform due to a few reasons. Firstly, it is on the same platform as the telemetry collection solution as mentioned in Section 2.2, providing continuity and centralization of tools. Second, it is designed for collaboration, making it easy to share and contribute to across the data sciences team. Third, the platform makes it easy to interact with the system, deep-dive and perform analyses on the models, as well troubleshoot updates. Lastly the platform has workflow and scheduling capabilities, which is essential in automating and scaling the anomaly detection system.

The workflow runs at set intervals and has the capability to extract, transform and load (ETL) data and run models at different times. The first step in the workflow is determining which KPIs are required for anomaly detection and which PPODs are of focus in that run. A PPOD is included in the workflow if it had a software deployment in the previous week. Each PPOD is monitored for an entirety of a week; see Figure 1 for a PPOD's network monitoring timeline. If the KPI is scheduled for anomaly detection, the telemetry data is ingested and prepared for anomaly detection. Next the respective anomaly detection model is applied; each KPI has an individual anomaly detection model, Section 2.4 will discuss these models in more detail. In order the keep the ML models from going stale, there is also automatic re-training based on criteria developed with the help of subject matter experts (SMEs); this includes the amount of time that has passed since the last training, if there is a new deployment to the PPOD, or if there is an extreme change in the historical data such as an RPD move, etc. If there are anomalies

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

detected, each alert is assigned a universally unique identifier and prepared for alerting and reporting. The alert information as well as helpful metadata will go into a data storage system. An informative visualization of the anomaly, unique to each KPI, is also generated at the time of anomaly detection and saved to a storage container.

A single run can contain a handful of anomalies from multiple KPIs, so it is important that the alerts are presented in an organized manner and provide information that aids prioritization of alerts for investigation. The compiler step in the workflow handles the compiling of alerts and communicating them to stakeholders for validation; utilizing the unique alerts and information saved internally at the time of alert generation. There are two parts to the report, the first organizes the anomalies' visualizations into a document by PPOD and orders by KPI type. By presenting the anomalies in a clear and concise format, it allows easy sharing of information and enables a common understanding of the detected anomalies. The second part of the report is a spreadsheet that organizes the sheets by KPI/alert type and sorts by the percent of CMs that can be tied to the anomaly. The spreadsheet also contains additional supporting data relevant to each anomaly type as well as correlation to existing network monitoring events, providing valuable context to the anomalies. Details around the supporting data will be discussed further in Section 2.5. Presenting the alerts in this way promotes effective communication, enables informed decision-making, and ultimately contributes to maintaining a resilient and high-performing network.

In the final phase of compiling the anomaly detection alerts report, the report is automatically shared with stakeholders through a cloud-based collaboration platform for communication. In this phase of the system, the report is shared for not only investigation but also validation which will be discussed in Section 3. See Figure 2 for the full cycle of the anomaly detection workflow.

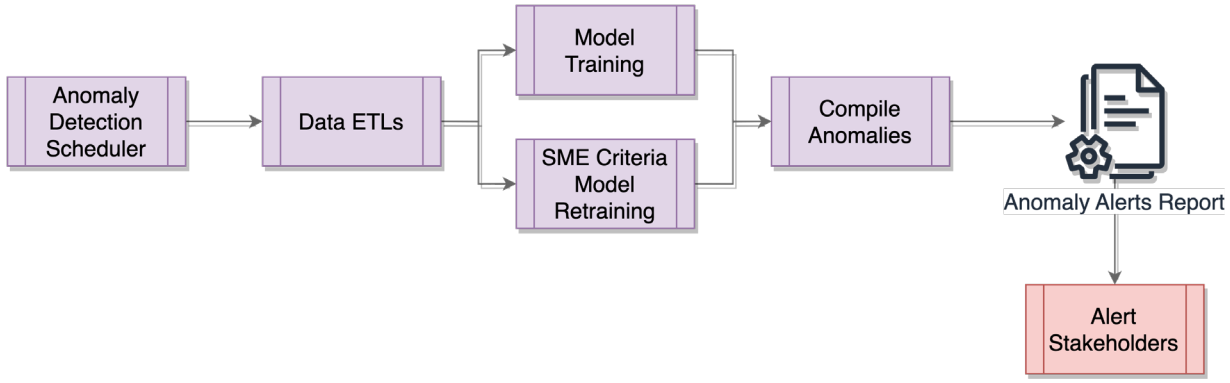


Figure 2 – Anomaly Detection Architecture

2.4. Anomaly Detection Algorithms

This section will discuss the details of each anomaly detection model to the KPIs introduced in Section 2.2. The models range from statistics-based to unsupervised machine learning models and all come with nuances that are important in building an accurate and reliable anomaly detection system.

For post-deployment long term monitoring, all models follow similar criteria for validating and alerting an anomaly. Since deployments can result in different behavior immediately after, the first is omitting anomalies up to 12 hours after a deployment. Any behavior in those hours is caught in the instantaneous network health checks or is expected, so they do not warrant an alert. The second is incorporating threshold criteria for alerting, requiring the anomaly to reach a severity threshold in time and/or maximum

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

customer impact (these data points also play into ranking later). Section 4 will mention the future work of incorporating causality, further improving the accuracy in distributing anomalies in front of the right stakeholder.

2.4.1. Cable Modem Online Signal

The initial anomaly detection algorithm was developed for the CM Online signal. Instead of relying solely on thresholds, this algorithm considers historical patterns and the severity of fluctuations in the CM signal. As mentioned in Section 2.2, when an RPD is re-booted, all CMs will go offline momentarily, but since it is instant and expected it should not be considered an anomaly. Similarly, during maintenance windows, it is possible to see the same pattern for a given RPD each day which is also not anomalous. The CM algorithm overcomes this by utilizing statistical methods for anomaly detection, providing a lightweight yet powerful model with results that can be easily interpreted.

As mentioned in Section 2.1, standard deviation in time-series anomaly detection provides a lot of power and is the core of the CM anomaly detection model. To reach that predictive power, pre-processing is the first step which includes removing extreme outliers, normalizing, and differencing the data. After pre-processing, the following method is applied to each timestamp. Initially, a fixed number of hours is looked back on, and a rolling window of approximately one hour is applied. Next, the standard deviation of each window is calculated. If a large percentage of the windows exceeds the standard deviation threshold, the timestamp is considered an anomaly. By adjusting the window size and applying appropriate thresholds, the model becomes less sensitive to sudden or small-scale anomalies, thereby increasing accuracy in identifying anomalies that warrant alerts in this domain.

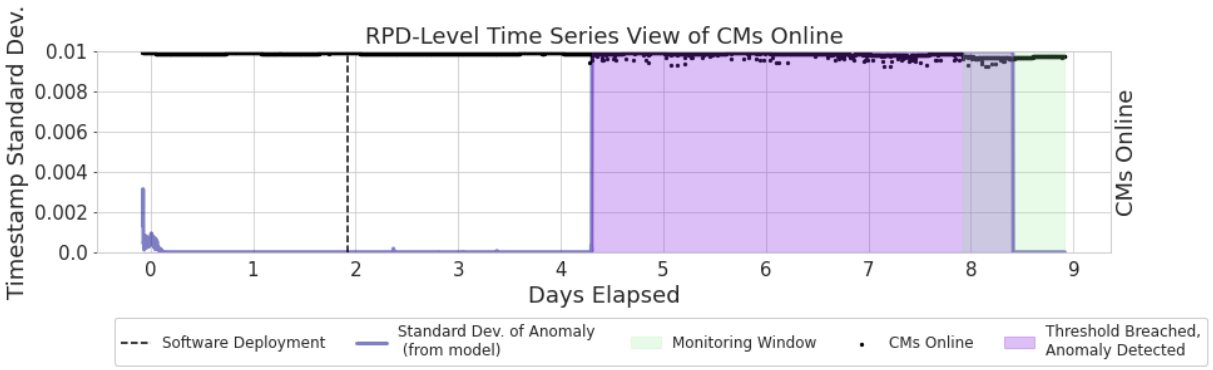


Figure 3 – CM Signal Oscillating Post-Deployment Anomaly Example

Figure 3 exemplifies a historical CM anomaly that occurred several days after the software deployment. The anomaly persisted for several days and is caught when applying the model, alerting during the monitoring window. The model's ability to catch this known anomaly highlights its potential to alert stakeholder from the initial day of occurrence. It is worth noting that examples of this scale are now rare as they are actively flagged and addressed immediately with the CM anomaly detection model in production.

2.4.2. Customer Calls (Interactive Voice Response)

As mentioned in Section 2.2, IVR works as a comprehensive metric when monitoring the network post upgrade. Customers can call in at any time of the day for a multitude of reasons, either network platform related or not (billing questions, in-home support, etc.). Detecting anomalies on the PPOD in and out of peak hours is crucial as a spike in calls can be very telling of the state of the network. Since customers for

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

each PPOD exemplify different historical patterns, that is not related to the software deployment, a simple threshold-based model does not apply. The anomaly detection system utilizes a generalized extreme studentized deviate (ESD) detector for IVR at a PPOD level. The ESD is a statistical method used for detecting outliers or anomalies in a univariate dataset. The IVR model takes a statistical approach for similar reasons as CM Online Signal; the model is easy to understand, implement, and flexible due to significance level and maximum number of outliers thresholds within the model. To incorporate severity of the customer calls, the algorithm is applied to the rolling percentage of IVR to total number of CMs in the PPOD. The model accurately detects impactful customer call anomalies that then can be correlated back to a deployment.

Figure 3 illustrates an instance of an anomaly (red dot) occurring after a software deployment (dashed vertical line). The green shaded area represents the monitoring window within which the anomaly detection system analyzes the metrics, if there is an anomaly within the window then it is alerted on. It is evident that the anomaly detected is valid in comparison to historical trends prior to the deployment, but the question still lingers if this anomaly is directly tied to the deployment. This example will be further discussed in Section 2.5.

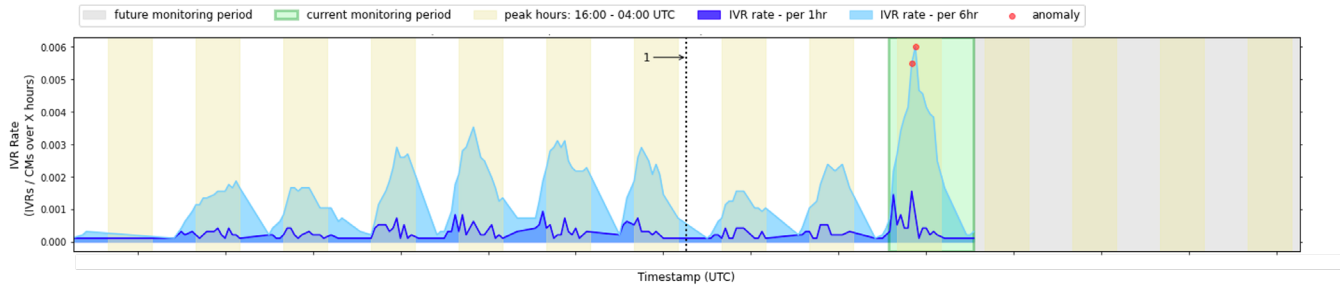


Figure 4 – Spike in IVR Post-Deployment Anomaly Example

2.4.3. Quality of Experience

Quality of experience is a metric derived from multiple network performance statistics (latency being one) that represents the experience of residential high-speed data (HSD) customers. When certain thresholds are crossed, a QoE issue occurs at the cable modem level. When traffic gets congested or some event adversely affects network health, the number of CMs with QoE issues will grow. However, the QoE metric tends to follow regular patterns depending on the hour of the day, day of the week, holiday schedule, etc. The challenge is being able to capture an adverse effect on the network at any time and not just during peak hours when some upper bound is pushed too far. To that end, one goal is to be able to accurately model what is typical at every hour and every day of the week. Furthermore, since the network is constantly evolving, another goal is to capture this evolution while still catching any slow deterioration in network health. While capturing cyclic patterns in the data is typically handled by the model itself, capturing the evolution of the network can be jointly tackled by both the model and various heuristics in the application layer. Some logic in the application layer we use to trigger model retraining includes data shifts caused by RPD movements related to the PPODs, irregular historical patterns that cause exploding forecast uncertainty, and redeployments of PPODs.

Several univariate modeling approaches that don't require covariates were considered. One approach that was considered is a mixture model (e.g., Gaussian mixtures). However, these models have natural limitations when it comes to capturing dependencies between timestamps (sequential dependency); much of their advantage also disappears when we limit them to a single metric. One way to convert a mixture model to capture sequential dependency is to form a hidden Markov or attempt to add more flexibility

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

with a long short-term memory (LSTM) or other deep learning architecture. Although these approaches have the benefit of using a single model on a variety of different time series, one downside is the lack of interpretability when trying to understand model predictions and align them with domain expertise. Deep learning approaches can also run into issues with their lack of consistency in their uncertainty estimates. Frequentist autoregressive approaches (e.g., ARIMA) are another option, and they provide a better uncertainty framework. However, such models can be tricky to adapt when multiple cyclic patterns are present; they also can run into similar interpretability issues since their data generative process is very statistical and often obscured by various transformations. After further understanding the need for a QoE anomaly detection model and the advantages and disadvantages of some approaches, it led to the development of the model currently in production.

A dynamic linear (state space) model was developed to capture the multiple cyclic patterns in the data and get reasonable model interpretability. To acquire uncertainty estimates in these patterns, the model was made Bayesian. In the Bayesian framework, such a model can suggest interpretable latent variables that evolve according to a pre-specified pattern (e.g., hourly repetitions) that generalizes well and is controlled by simple parameters (such as regression coefficients). See Figure 5 for these detected patterns in the QoE data with the model. The Bayesian variation then considers the observed data to be composed additively by the latent variables and computes the posterior distribution of all parameters simultaneously; this computation of the posterior takes into account both the likelihood of the observed data (based on our model specification) and priors (having mainly a regularization role here). Some key components that can help with anomaly detection are 'dynamic' (latent components can evolve), 'linear' (model specification that can generalize) and 'Bayesian' (priors that can regularize and model that can quantify uncertainty). Another benefit of this approach is it can be tuned to specify the sequential evolution of the latent states to explicitly get a stable level suitable for anomaly detection.

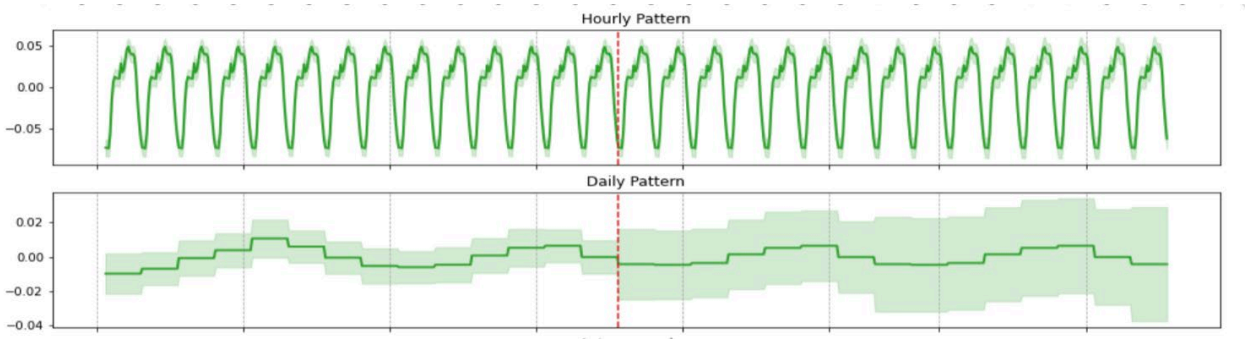


Figure 5 – QoE Hourly and Daily Patern Detected by Bayesian Framework

Since this is a statistical model without true anomaly labels, one of the challenges is identifying and accounting for anomalies that occur in the training data when learning the normal behavior of the system. Two common model-native solutions include setting the prior variance of latent or observation noise dynamically based on certain heuristics and using more flexible noise distributions to prevent outliers from having too much influence on fit. Another general solution, which takes advantage of workflows that retrain models at regular intervals, is to impute anomalies in the new training data using the prediction from the model that existed prior to this. Of course, this solution can have a cold start effect when there is no pre-existing model, and in that situation the model-native solutions mentioned previously can help with this.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

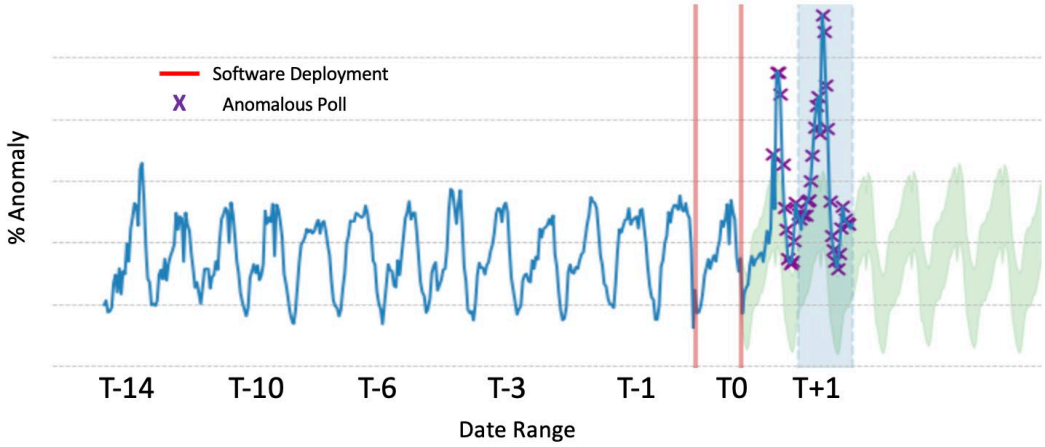


Figure 6 – QoE Display Slow Degredation Post-Deployment Example

Figure 6 is a QoE anomaly example, the blue represents the ratio of QoE events to CMs in the PPOD, the red vertical line indicates a software deployment, and the purple marks are anomalies detected. Timestamps are labeled as anomalies if the actual QoE ratio falls outside of the confidence bands (green, generated by forecast). The training data relies on empirical data prior to the deployment and retrains based on a handful of SME criteria (re-training step seen in Figure 2). These model characteristics result in a confident and telling QoE anomaly alert.

2.5. Understanding Anomalies

When constructing an anomaly detection system, it is crucial to instill confidence in individual alerts when presenting them to stakeholders. As the number of alerts scale with deployments, it is crucial to provide additional supporting data to help prioritize and understand the scenario. Therefore, the alert system provides useful metadata within the final report.

To leverage the existing alerting systems, each anomaly is accompanied by a list of events that occurred within a few hours in its vicinity. It is important to have a wide correlation window since it is possible for events to be correlated but not occur at the same time. An example use case would be if there is a customer power outage (CPO) due to a storm event, it is possible for the power supply (PS) and node reside on a different part of the power grid than some customers that they service. If the storm event affects the grid with the PS (reaches end of discharge) but not the customers, then those customers would call in causing an IVR anomaly. Knowing this information, the anomaly alerted on is most likely not due to the network deployment being monitored. By using systems that track these events, the end-user would be able to make their own assumptions based on the events provided.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 – Example Supporting Data for Anomaly (Corresponds to Figure 4)

alertId	PPOD	Alert Type	Deployment Group Priority	Existing Events	CM Info
example_uuid	example_PPOD	IVR	LOW	Storm Outage Event #A, RPD Unreachable #B	X% offline DOCSIS 1.0, Y% offline DOCSIS 3.0, Z% offline DOCSIS 3.1

Another indicator of whether or not the anomaly is a side effect of the deployment, is the breakdown of CM behavior by DOCSIS version. The supporting data includes the count of offline CMs around the time of the anomaly, which can be especially valuable when an entire group experiences connectivity issues. This information guides subject matter experts toward a better understanding of the underlying cause behind the anomaly.

Lastly, it is important to note that the deployments being monitored are scheduled in groups of PPODs. The supporting data also includes the anomaly priority for each PPOD in the report. If multiple PPODs within a group exhibit an anomaly compared to a single PPOD, it can be argued that the latter represents an isolated event, while the former group is more likely to display anomalies related to the deployment. Therefore, a group priority alert is assigned to each PPOD, with a high priority indicating multiple PPODs experiencing alerts, medium priority signifying a threshold crossing, and low priority assigned otherwise. By incorporating these measures, the end-users can maintain awareness of anomalies without prematurely drawing conclusions. This approach enables them to proactively respond to alerts in a manner aligned with the available information and take appropriate actions accordingly.

Looking at Figure 3 again from Section 2.4.2, it is important to note that while these anomalies are valid, they can be influenced by various causal events unrelated to the deployment itself. In this example, the supporting data presented in Table 1 guides the operations team in identifying potential causes. The data includes information on existing events, cable modem (CM) details, and the priority of the deployment group. In this example, the existing events prove to be the most valuable in assisting the team to pinpoint a potential root cause (storm outage) and go from there.

3. Performance and Impact Findings

Detecting anomalies and evaluating the effectiveness of anomaly detection algorithms are pivotal steps in the development and deployment of models. However, this process presents challenges due to the absence of labeled anomalies, which can manifest as diverse patterns and trends such as contextual, point-based, seasonal, cyclical, or collective anomalies. Consequently, conventional evaluation metrics like precision, recall, F1-score, or AUC-ROC (area under receiver operating characteristic curve) may not be directly applicable in this context.

To overcome this challenge of producing highly confident models without labeled data, the anomaly detection approach relies on empirical data and trend-based training to produce high-confidence results leading to increased performance. As mentioned in Section 2.4, this is done by either predicting future values and creating upper and lower confidence bounds (QoE model) or incorporating thresholds based on the empirical data (CM and IVR model).

As described in 2.4.3, to further validate the identification of an anomaly, there is criteria to determine whether the empirical data for a specific period preceding the anomaly occurrence meets the requirements for exclusion, training or re-training. The network is ever changing and producing anomaly alerts from high-certainty training data is crucial.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

By considering these exclusion and re-training criteria, we ensure that the anomalies identified by the algorithm are reliable and meaningful. This approach combines empirical data analysis, trend-based predictions, and criteria-based validation to enhance the accuracy and effectiveness of anomaly detection in real-world applications. The QoE example mentioned previously encompasses all of these high-performance efforts.

In addition to the above steps, measuring the effectiveness of anomaly detection algorithms involves correlating them with other events (Section 2.5) captured through existing systems and tools. Correlating with these events helps build confidence in the detected anomalies and prevents disregarding an anomaly that lacks corresponding events.

When assessing the impact of modeling and detecting anomalies, the focus returns to the original objective outlined in Section 1. The goal is to identify service degradation over a period of time that may not be detected in instantaneous checks immediately after a deployment. Some degradations are gradual and may go unnoticed by existing tools. These degradations have a negative impact on customer experience and, if undetected, can be masked by other changes being rolled out. Figure 6 is a great example of a gradual degradation; as noted, the empirical data did not display a similar behavior as post deployment behavior, and the percent of QoE events slowly increases resulting in an impactful anomaly detection alert.

To assess a subset of anomalies, network operations experts validated ~250 alerts over a span of two months. To date, these identified anomalies have demonstrated a 98% accuracy and low recall, implying that the algorithm is sensitive in capturing subtle changes based on empirical data. By incorporated expert feedback, various enhancements have been implemented. These include the grouping of anomalies based on deployment date and type, assigning severity based on the number of periods of degraded service, and considering the frequency of anomaly occurrences.

Another important aspect is making the detected anomalies actionable by identifying their root causes and establishing correlations with other anomalies. This effort will be further discussed in a subsequent section.

4. Future Work

In the preceding sections, we provided an overview of the existing long-term monitoring system. Now, we turn our attention towards future endeavors that aim to propel the system closer to its ultimate objective of effectively suggesting rollbacks for DAA software deployments and offering correlation and causality for detected anomalies. One of these future endeavors is the incorporation of additional metrics and attempting supervised learning.

4.1. Labeling Tools

The goal for future models and metrics is to incorporate supervised learning for anomaly detection, which can significantly increase the confidence and accuracy of generated alerts. The process of supervised learning necessitates labeled data. As mentioned in Section 2.1, labeled data plays a crucial role in training an effective anomaly detection model. It provides ground truth information about which events are normal and which events are anomalous. It also enables the training of supervised learning models, especially for classification algorithms. By using these labeled examples, the model can establish a baseline, and learn the patterns, features and examples of which deviations beyond the baseline patterns are expected and which ones are unexpected. The availability of labeled data also helps in performance

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

evaluation – performance metrics like accuracy, precision, recall, F1-score, etc. can be used to understand and compare different models.

However, the collection of labeled data comes with its set of challenges, mainly revolving around the requirement of human involvement, i.e., the need for subject matter experts to flag the anomalies. This challenge gets larger when factoring in scale, because as the size of the dataset increases, labeling becomes more challenging and time consuming. It is also a challenge to represent the different types of patterns, trends, and anomalies that exist in real-world data.

To address some of these issues, two user interfaces (UIs) were internally designed to make event labeling easier. The first is a larger tool that operates on the cloud and facilitates collaboration and a shared directory to store the labels in. This is the tool that is intended to be the source of truth for future labeling efforts across multiple projects in data sciences. In the process of development for anomaly detection efforts, there is a second internal tool based purely in Python, HTML and JavaScript that incorporates various features to increase the simplicity of the labeling process. The process creates a standalone UI that displays time series views of a selected set of metrics of interest and provides features such as zooming, time range highlighting, event labeling, providing additional annotation, and the ability to export the changes into a CSV file.

In order to standardize training set creation across future metrics, criteria are put in place to select time series samples with at least 28 continuous days of data and represent a wide range of samples – samples with no anomalies, samples that only had anomalies in the latter two weeks, and time series samples with anomalies throughout the time range.

The next section will discuss the on-going efforts to expand the metrics monitored and discusses the first metric to utilize the labeling efforts for model development. By utilizing the developed labeling tools, there is a substantial labelled data set covering multiple scenarios across the platform footprint. Since the occurrence of anomalies is rare, the labeled data results in an imbalanced target variable. Thus, future approaches need to take this into account by either leveraging existing labels to take a semi-supervised approach or creating synthetic samples through up-sampling/down-sampling techniques, potentially improving future models' performance.

4.2. Expanding Metrics Monitored

Having a set of labeled anomalies to experiment with, this section will introduce an analysis on traffic metrics (previously mentioned in 2.2) and demonstrate how to use these labels to confidently select a model for production. This analysis compares a handful of models, both a set of naïve baseline models and forecasting models. Since this work is on-going and experimental, understanding each model evaluated is out of scope for this paper. The baseline models are light since they rely on simple heuristics (similar to CM Signal and IVR); they are interquartile range (IQR), delta-based, and Auto-Regression. The baseline captures regular patterns, traffic volumes, and expected behaviors. The set of forecasting models (similar to QoE) include ARIMA, exponential smoothing (Holt 2004), Facebook Prophet (Taylor and Letham 2017), Theta (Assimakopoulos and Nikolopoulos 2000), and Season-trend Decomposition (Hyndman and Athanasopoulos 2021). These methods involve forecasting the time series, computing residuals or error from the true values, and then setting thresholds on the residuals to identify where the data points are much further from where they would be expected to be.

Evaluating event-based predictions presents another set of challenges, such as differences in temporal alignment or event durations. Imbalanced data is another challenge, since the occurrence of events is rare. Additionally, since the values being compared are pairs of true and predicted events, there is no concept

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

of a true negative, which necessitates some specific metrics to provide a more meaningful assessment of model performance.

To this end, the event matching criterion utilized a margin of 20 minutes, where predicted events were considered a match to true events if the predicted event occurred within 20 minutes of the true event. These matched events were then evaluated using precision, recall, F1-Score, and Jaccard index. All evaluation scores fall within a range of zero to one.

Table 2 – Evaluation Score by Model for Traffic Metric (Leaf-MAGG) – Leading Methods Based on F1-score Highlighted

metricGroup	eventMethod	Precision	Recall	Jaccard Index	F1 Score
Leaf-MAGG	IQR	0.0594	0.0507	0.0281	0.0547
	ARIMA	0.1192	0.1305	0.0665	0.1246
	Facebook Prophet	0.9438	0.1065	0.1058	0.1913
	Exponential Smoothing	**0.573**	**0.3384**	**0.2702**	**0.4255**
	Season-Trend Decomposition (Moving Average)	**0.556**	**0.327**	**0.2593**	**0.4118**
	Auto-Regression	0.482	0.237	0.1889	0.3178
	Delta from Previous Poll (%)	0.7976	0.2548	0.2393	0.3862

See Table 2 for the evaluations of the models for each facet of traffic metrics. Given that the data is imbalanced, the models are primarily evaluated on the F1-score as it is considered best practice in evaluating imbalanced data sets. It is evident that exponential smoothing is the best model across the board, although the score is low that can be attributed to the size and nature of the sample data set. See Figure 7 for the experiment's top two models' results on a sample traffic metric. The top graph is the labeled event, the traffic metric experiences anomalous behavior after a software deployment. The middle and bottom graph are the results of the exponential smoothing and season-trend decomposition models respectively. Most anomalous timestamps are detected (green), some missed (red) and there aren't any false positives which is beneficial in avoiding false alarms. The model for this metric is currently being developed, but by having labeled data the experimental models can be evaluated for selection and ultimately used for supervised ML.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

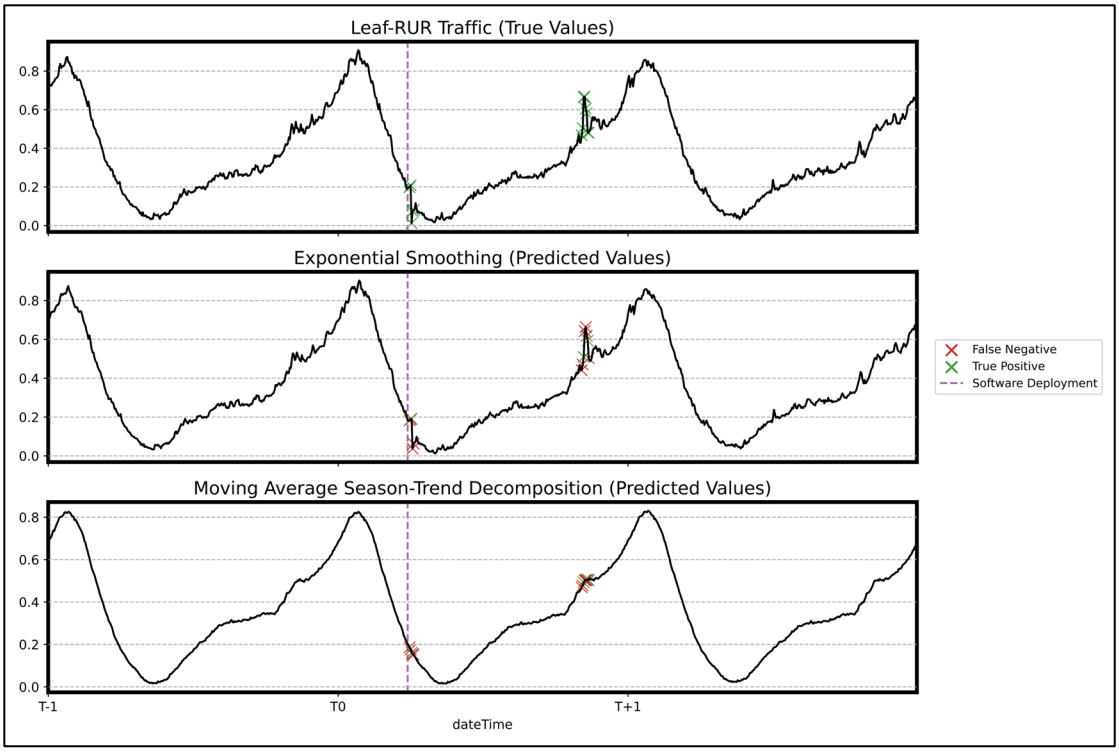


Figure 7 – Traffic/Switch Anomaly Detection Development – Top Two Models on Leaf-RUR Traffic Metric

4.3. Alarm Correlation and Causation

There are a handful of existing efforts and ideas on enhancing anomaly detection capabilities within the network infrastructure at Comcast, one is expanding existing KPIs discussed in sections 2.4 and 4.2 for measuring platform health. To accomplish the goal of autonomous anomaly detection holistically around different systems - i.e., platform health at home or radio frequency plant, the detection system will monitor additional metrics such as network latency or host utilization which will allow a full analysis of network performance capabilities in a more comprehensive manner.

With this mindset towards proactive problem-solving extending beyond just end-user consumption but also analyzing correlation between different metrics and root-causes, the causality analysis should lead not only towards improvements regarding service reliability but also reduced customer downtime whenever hardware issues seemingly appear randomly.

Essentially through extensive correlative investigation between detected anomalies along with pattern identification/detection throughout event occurrences, the anomaly detection system can actively prevent future instances of the same pattern tends to be on the horizon. Gaining insight into the causes behind different types of anomalies within the complex vCMTS platform architecture requires an approach where relationships are analyzed between various network metrics like latency spikes and host utilization rates. Also of note is establishing any existing correlations which would assist in identifying corresponding contributory factors towards such anomalous trends seen. Once common trends exist among different types of anomalies within our system environment via this data-driven approach, proactive measures could be employed through enhanced capacity planning/resource allocation strategies which optimize future performance.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In better distinguishing between isolated incidents and systemic issues awaiting triage, causal analysis can be conducted on the occurrences of these anomalous events. This helps in pinpointing specific changes to the network like software/hardware updates or hardware failures responsible as anomaly root causes. Once identified, it becomes easier to deliver tailored-fit solutions such as targeted software updates or hardware maintenance procedures geared towards ensuring comprehensive performance optimization. Producing the root cause can also minimize investigation time and speed up recovery.

When these analytical methods are deployed across the network environment for all components, this will help identify recurring themes or common factors across multiple anomalies that occur when specific software updates are implemented. If discovered early enough, improvements like enhanced testing protocols could ensure future anomalies are prevented from occurring again.

Lastly, implementing daily full-footprint anomaly detection offers precise insights into our entire platform through its simultaneous evaluation of all relevant metrics and KPIs across subsystems. Doing this across the entire platform will lead to a much clearer view of system behavior while providing improved precision towards preventing future problems from arising. Full-footprint analysis has potential to provide deep insights pertaining to complex interdependencies among various components resulting in unexpected behaviors showcasing potential issues throughout platforms as opposed to isolated incidents.

Incorporating daily full-footprint (across all PPODs) analyses ensures superior accuracy and efficiency in detecting these anomalous occurrences since they enable establishing base norms and benchmarking them against new telemetry. Adopting this methodology empowers the organization with holistic understanding aligning all integral components, reinforcing fault-tolerance capabilities by extensive cross-functional validation rather than just viewing each KPI and anomaly as an individual siloed outpost.

This approach also enables the identification of anomalies associated with cascading or correlated effects across diversified workplace components which could result in unexpected behavior. By expanding the scope of the analysis, the anomaly detection system will not only identify sporadic or intermittent occurrences but frequently mitigate associated impediments to maintain the relevant networks' overall operational stability. Hence it can be summarized that full-footprint analysis proves imperative in detecting irregularities facilitated through intricate mechanics within the system, providing better user experiences while ensuring platform stability by capturing otherwise-infrequent aberrations of behavior affecting overall performance.

5. Conclusion

This paper presented a comprehensive analysis of anomaly detection post software deployment to detect slow deterioration in system performance. The sections covered big data architecture, ML techniques used to detect anomalies, challenges associated with detecting anomalies, particularly in the absence of specific patterns and the existence of various anomaly types. To address these challenges, the system relies on empirical data, trend-based training, and the establishment of validation criteria to ensure the accuracy and effectiveness of the anomaly detection algorithm.

However, the work does not end here. As part of future endeavors, the application of the anomaly detection algorithm is planned to be expanded to encompass the entire virtualized platform daily. This expansion will provide a more comprehensive understanding of system behavior and enable the identification of anomalies that may not be captured in smaller subsets of data. Furthermore, efforts will be made to broaden the range of monitored KPIs incorporating additional metrics to enhance the overall analysis of system performance and correlate between different anomalies and identify root cause for the anomalies detected.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

By addressing these future research directions, this work aims to contribute to the ongoing improvement of network anomaly detection techniques and the overall performance of virtualized platforms.

Abbreviations

ARIMA	auto-regressive integrated moving average
AUC-ROC	area under receiver operating characteristic curve
CM	cable modem
CPE	customer premise equipment
CPO	commercial power outage
DAAS	distributed access aggregate switching
DAA	Distributed Access Architecture
DOCSIS	Data-over-cable Service Interface Specification
ESD	extreme studentized deviate
ETL	extract, transform and load
HAGG	hub aggregator
HSD	high speed data
IQR	inter-quartile range
IVR	interactive voice response
KPI	key performance indicator
LSTM	long short-term memory
MAGG	mother/master aggregator
ML	machine learning
PPOD	physical pod
PS	power supply
QoE	quality of experience
RPD	remote physical device
RUR	residential U-Ring router
SME	subject matter expert
UI	user interface
vCMTS	Virtual Cable Modem Termination System

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. International Journal of Forecasting, 16(4), 521-530.

Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2014). Network Anomaly Detection: Methods, Systems and Tools. IEEE Communications Surveys & Tutorials, 16(1), 303-336. doi: 10.1109/SURV.2013.052213.00046.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection: A Survey. ACM Computing Surveys, 41. doi: 10.1145/1541880.1541882.

Eppes, M., et al. (2022). Scaling DAA: Automated Network Health Check for vCMTS Platform. Paper presented at SCTE Expo 2022.

Holt, C. (2004). Forecasting Seasonals and Trends by Exponential Weighted Moving Averages. International Journal of Forecasting, 20, 5-10.

Hyndman, R.J., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3.

Stehman, M., et al. (2021). Solving The Mysteries of the Distributed Access Architecture. Paper presented at SCTE Expo 2021.

Taylor, S. J., & Letham, B. (2017). Prophet: Forecasting at scale. American Statistician, 72(1), 37-45.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Scaling of Current PON Technologies to Meet the CPON Goals of 100Gbps and Beyond

A Technical Paper prepared for SCTE by

Karthik Sundaresan
Distinguished Technologist
CableLabs
858 Coal Creek Dr, Louisville, CO, USA
3036619100
k.sundaresan@cablelabs.com

Zhensheng (Steve) Jia, Ph.D.
Fellow
CableLabs
858 Coal Creek Dr, Louisville, CO, USA
3036619100
s.jia@cablelabs.com

Kevin Noll
Principal Access Architect
Vecima
771 Vanalman Avenue Victoria, B.C. Canada
250 881 1982
Kevin.Noll@vecima.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Overview of EPON 4
   2.1. EPON PHY Layer 5
   2.2. EPON MAC Layer 5
   2.3. Provisioning (OAM, DPoE OAM, IEEE 1904.1, 1904.4) 7
3. Overview of ITU-T PON 7
   3.1. ITU-T PMD Layers 8
   3.2. TC-Layer 8
   3.3. ONU Management 9
   3.4. 25GS-PON MSA 10
4. Brief Introduction to CPON 10
   4.1. PHY Technology 10
   4.2. Target Applications 11
5. Comparison and commonalities 12
   5.1. Comparison and commonalities of EPON and GPON 12
   5.2. One PON 13
   5.3. MAC Layer Approach for CPON 13
      5.3.1. Why two MACs for CPON? 14
      5.3.2. Delta Specifications 14
6. Scaling Considerations for CPON 15
   6.1. EPON based MAC Layer 15
      6.1.1. Understanding EPON Overheads 16
      6.1.2. Scaling of EPON technology for CPON data rates 19
   6.2. GPON based MAC Layer 20
      6.2.1. Understanding GPON Overheads 20
      6.2.2. Scaling of GPON technology for CPON data rates 24
7. Challenges and Future Considerations 25
   7.1. FEC Evolution & Comparison 25
   7.2. Channel Bonding 26
   7.3. Single Carrier vs Digital Sub-carrier 27
   7.4. Other challenges 28
8. Conclusion 28

Abbreviations 29

Bibliography & References 30

List of Figures

Title **Page Number**

Figure 1 – Nx25G EPON PHY and MAC Layers 6
Figure 2 – Nx25G EPON SIEPON vs OAM 7
Figure 3 – GPON ComTC, PHY and MAC layers 9
Figure 4 –CPON application scenario illustrations 11
Figure 5 –PHY, MAC& OSS technology stacks for GPON& EPON 12
Figure 6 – Two Different PON Camps 13
Figure 7 – A Drop-in CPON PMD for both EPON & GPON tecnology stacks 15

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 8 – Analysis of the line,frame , data rates on EPON tecnology stack 16
Figure 9 – Envelope Quantum Format in EPON 16
Figure 10 – Transmission Sequence consisting of three frames 16
Figure 11 – EPON Physical coding sublayer 17
Figure 12 – ONU Burst Structure 18
Figure 13 – Fragmentation at SA Layer,XGEM Encapsulation in FS Payload 21
Figure 14 – Downstream FS Frame Format 21
Figure 15 – Structure of the BWMap 22
Figure 16 – Upstream FS Frame Format 22
Figure 17 – ComTC Downstream PHY Framing 22
Figure 18 – Upstream PHY Framing, PHY Bursts and Preamble Format 23
Figure 19 – Upstream FS Burst with FEC 23
Figure 20 - FEC Evolution in PONs 26
Figure 21 – Spectral Illustration of DSC and SC 27

List of Tables

Title **Page Number**

Table 1 – Downstream and Upstream Wavelength allocation EPON 5
Table 2 – Downstream and Upstream Wavelength allocation NG-PON2 8
Table 3 – ITU-T-G.9804 50G-PON Wavelengths 8
Table 4 – 25GS-PON wavelengths 10
Table 5 – MCRS Envelope Quantum Capacity 17
Table 6 – 25GMII Clock rate 17
Table 7 – PCS Layer Data rate calculations 18
Table 8 – Upstream PHY Layer Burst Overhead 18
Table 9 – Downstream Data rate calculations 19
Table 10 – Upstream Data rate calculations 19
Table 11 – Potential Multipliers for Data rates 20
Table 12 – Relationship of Various PON Line Rates 24
Table 13 – FEC in Optical Communication Links 25

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Passive Optical Network (PON) is a fiber-optic telecommunications network that uses point-to-multipoint architecture to provide high-speed data, voice, and video services to subscribers. PON is the dominant solution to provide broadband services using fiber to the premises (FTTP). Operators look to expand access to more people and geographically remote areas in a cost-effective method. There is a need for new PON technology with higher capacities and longer reach and CableLabs has launched the Coherent PON (CPON) project to future-proof operator networks and deliver 100Gbps over PON (with coherent optical technology). Adapting existing standards to this new CPON technology is an advantageous approach. This needs an investigation of existing standards, to understand how to scale them to the speeds and capacities of CPON, which is the main goal of this paper.

EPON (Ethernet Passive Optical Network) and GPON (Gigabit Passive Optical Network) are two main PON technologies and cable operators are deploying both. While these PON technologies share similar architecture and topology they differ in many ways. This paper discusses the similarities and differences for the latest version of these technologies (ITU-T [GPON-G.HSP] and IEEE [Nx25G-EPON]). The focus will be on understanding the different layers of each technology and discussing the functional equivalence between the two and areas where they differ from each other.

Reusing technology layers from existing standards saves development time. Also, reusing many of the components of ASIC/silicon design implementations will reduce the cost of this new CPON solution. This paper highlights different factors of the existing GPON and EPON technologies that will need to be modified to reach these goals. This analysis includes a study of aspects such as the framing sublayer, fundamental line rates, effective data rates, FEC encoding/decoding, upstream scheduling/granting mechanisms, channel bonding, and provisioning etc. This paper aims to give an understanding of the current technologies and how those could be scaled to meet 100Gbps and beyond with CPON.

2. Overview of EPON

EPON is a point-to-multipoint technology that delivers 1/10/25/50Gbps upstream and downstream in FTTH/FTTP networks. 10G-EPON (defined in IEEE Std 802.3av-2009) supported backward compatibility with 1G-EPON (defined in IEEE Std 802.3ah-2004), allowing operators with existing 1G-EPON deployments to add 10G-EPON using either wavelength division multiplexing (WDM) for fully independent and parallel operation, or using time division multiplexing (TDM) for time-shared operation in the upstream direction.

The latest evolution in EPON technology is Nx25G EPON (defined in IEEE Std 802.3ca-2020) that enables 50Gbps in symmetric and asymmetric configurations, while maintaining backward compatibility with deployed 10Gbps EPON systems. Using a combination of 25Gbps (downstream or upstream) channels and/or 10Gbps (upstream only) channel, the resulting PON system can supports symmetric and/or asymmetric MAC data rates of:

- 25Gbps, downstream and 10/25Gbps, upstream (25G-EPON)
- 50Gbps, downstream and 10/25/50Gbps, upstream (50G-EPON)

Collectively, these systems are referred to as Nx25G-EPON, with parallel 25G-channels, operating on separate wavelength channels, and supporting channel bonding for an increased aggregate MAC data rate.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.1. EPON PHY Layer

In [Nx25G-EPON] technology, wavelength allocation allowing concurrent operation for downstream and upstream PHYs are shown in Table 1 below.

Table 1 – Downstream and Upstream Wavelength allocation EPON

Direction	Wavelength name	Center wavelength (nm)	Wavelength range (nm)
Downstream	DW0	1358	± 2
	DW1	1342	± 2
Upstream	UW0	1270	± 10
	UW1	1300	± 10
	UW2	1320	± 2

Most commonly, ODNs today are designed for compatibility with 10G-EPON PR30 power budget, supporting ODNs with 29dB optical loss. Nx25G-EPON accommodates channel insertion losses equivalent to 10G-EPON-defined PR20 and PR30 power budgets.

The physical coding sublayer (PCS) in Nx25G-EPON is responsible mainly for converting the data stream received from xMII into codewords, which can then be passed through PMA, PMD, and finally transmitted on the medium. In the receive direction, PCS performs the reverse function, i.e., decodes the received data and recreates the original data stream, hands off to the MCRS and then towards the Ethernet MAC and any associated MAC clients.

[Nx25G-EPON] uses 256b/257b line coding, the interleaver and de-interleaver are realized by using omega networks and it uses a stronger quasi-cyclic QC-LDPC FEC (with a much longer codeword but also providing a much higher degree of protection against bit errors).

2.2. EPON MAC Layer

Nx25G-EPON operates in a point-to-multipoint mode, delivering logically tagged Ethernet frames to subtended ONUs using a broadcast across the passive fiber ODN. In the upstream (from the ONUs towards the OLT), Nx25G-EPON operates in the multipoint-to-point mode. ONUs are allowed to transmit data only when they are explicitly granted by the OLT. The Nx25G-EPON architecture is based on multiple channels bonded together, 25Gbps in the downstream and 10Gbps/25Gbps in the upstream. Figure-2 shows the PHY and MAC layer components on the EPON network devices.

The multi-channel reconciliation sublayer (MCRS) provides logical channel bonding between any MAC and any number of underlying physical layers (channels), data is striped across multiple physical channels in both directions. MCRS is an interface between the MAC sublayer and one or more xMIIs. xMII is a generic term for the Media Independent Interfaces, for 25 Gb/s implementations, it is called 25GMII MCRS adapts the bit-serial data stream received from the MAC layer to the parallel format of the PCS service interface, effectively stripping Ethernet frames across multiple physical channels. Each and every PCS instance represents an MCRS channel, carrying data units between OLT and ONU.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

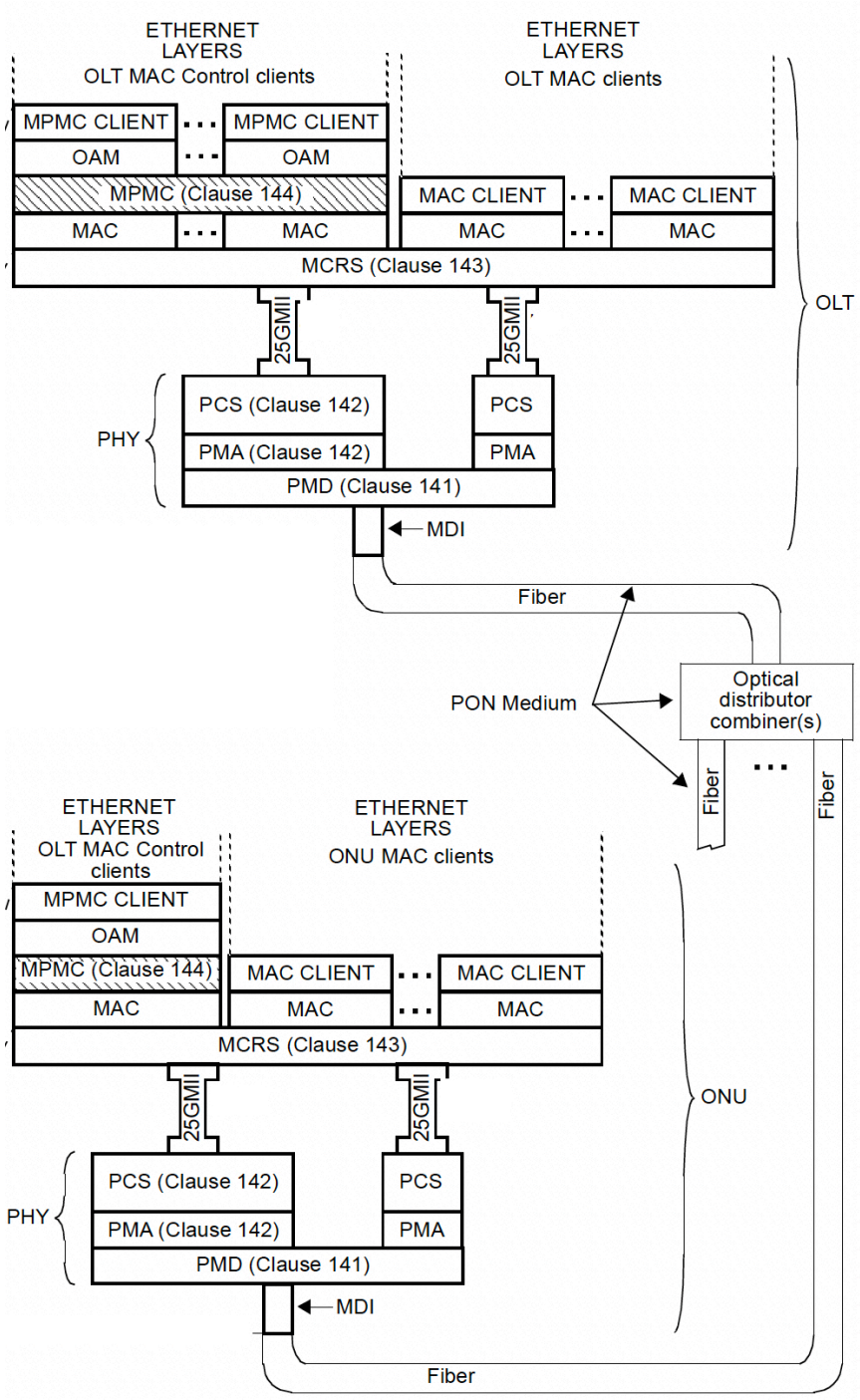


Figure 1 – Nx25G EPON PHY and MAC Layers

Source:*IEEE-Std-802.3ca-2020:Figure 144–2*

All multi-channel variants of the system rely on channel bonding, allowing for a single higher speed MAC (e.g., 50Gbps transmitting over a pair of bonded 25Gbps channels in a manner that is completely transparent to MAC). The resulting system architecture is scalable, allowing future revisions of the system to add individual channels, as well as replace channel definitions, data rates, FEC encoding, etc.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.3. Provisioning (OAM, DPoE OAM, IEEE 1904.1, 1904.4)

IEEE Standard for Service Interoperability in EPON (SIEPON) defined in IEEE Std 1904.1 & IEEE1904.4, describes the system-level requirements needed to provide service-level, multi-vendor interoperability of EPON equipment. It complements the existing EPON standards, which enable the interoperability at the Physical Layer and Data Link Layer.

SIEPON enables system-level interoperability, covering equipment functionality, traffic engineering, service-level quality of service/class of service (QoS/CoS) mechanisms, service and equipment management etc. SIEPON provides requirements for the operation of ONU and OLT devices. To foster interoperability across the PON among equipment from different vendors, it specifies the flow of information exchanged between the ONU and OLT as well as their associated behaviors.

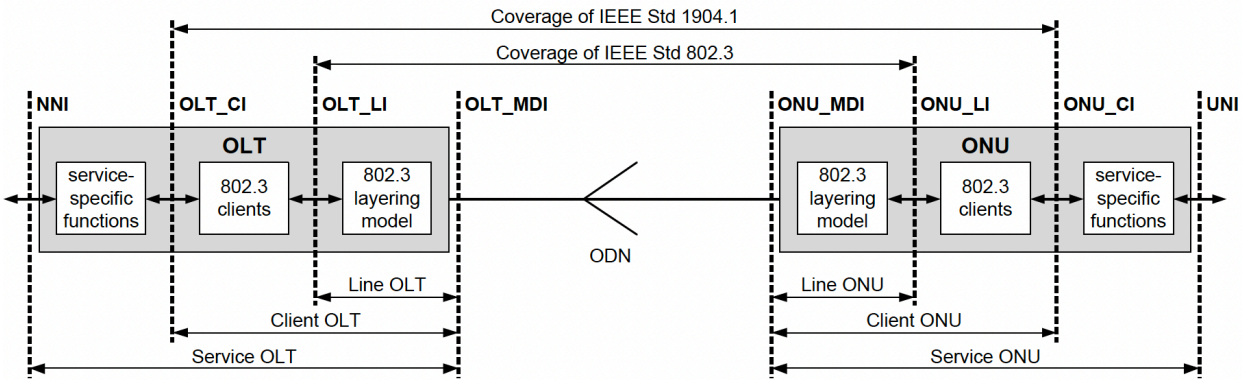


Figure 2 – Nx25G EPON SIEPON vs OAM

Source:*IEEE-Std-1904.1-2017:Figure 5-1*

IEEE Operations, Administration, and Maintenance (OAM) sublayer [IEEE 802.3], provides mechanisms useful for monitoring link operation such as remote fault indication and remote loopback control. OAM provides network operators with the ability to monitor the health of the network and quickly determine the location of failing links or fault conditions. The OAM provides data link layer mechanisms that complement applications that reside in higher layers.

3. Overview of ITU-T PON

The history of PON within the ITU-T dates back as far as 1987 when passive optical networking was proposed by British Telecom. The first PON specification, ITU-T G.982, (ATM PON or APON), was released in 1996. APON operated at 1.5Mbps and was followed in 1998 by BPON (Broadband PON) which operated at 622Mbps in the downstream direction. Several operators like Verizon deployed BPON prior to the release of the GPON, G.984 in 2003. GPON, operating at 2.5Gbps downstream and 1.25Gbps upstream, enjoys massive deployments around the globe to this day.

However, XGS-PON, (2016) has come to the forefront in the past several years and is positioned to be the most widely used PON technology ever. XGS-PON, specified in ITU-T G.9807, operates at 9.95328Gbps (typically referenced as a 10Gbps PON) in the upstream and downstream direction.

NG-PON2 was released as ITU-T G.989 in 2015. NG-PON2 was intended to add certain resiliency features and to boost aggregate network capacity up to 40Gbps of total capacity. While one large operator is committed to deploying NG-PON2, it will never see widescale deployment.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.1. ITU-T PMD Layers

The PMD layer of GPON, XGS-PON, and NG-PON2 are all based on intensity modulated laser transmission over single mode fiber with direct detection at the receiver (IM-DD). These support a range of optical power budgets to enable split ratios up to 1:128 and transmission distances up to 20km (40km with PON extender devices).

GPON uses a wavelength of 1490±10 nm for downstream transmission. G.984.2 specified 1310±50 nm for upstream transmission, but this has been narrowed in recent years to 1310±20 nm which opens some spectrum for new PON standards. XGS-PON uses a wavelength of 1575-1580nm for downstream transmission and 1280-1290nm for upstream transmission.

NG-PON2 achieves 40Gbps of aggregate capacity by allocating 4 lanes, each operating at 10Gbps. These lanes are allocated in upstream plus downstream pairs, sometimes referred to as channel pairs, Table 2.

Table 2 – Downstream and Upstream Wavelength allocation NG-PON2

Channel	Downstream Wavelength	Upstream Wavelength
1	1596.34nm	1532.68nm
2	1597.19nm	1533.47nm
3	1598.04nm	1534.25nm
4	1598.89nm	1535.04nm

50G-PON, specified in ITU-T G.9804, continues the use of IM-DD but increases the downstream nominal line rate to 49.7664Gbps. The currently released specification supports only 24.8832Gbps and 12.4416Gbps in the upstream but the specification for 49.7664Gbps in the upstream is expected to be released soon.

The optical spectrum for PON is very crowded in the O-band, but 50G PON is wedged in with a couple of options as shown in Table 3.

Table 3 – ITU-T-G.9804 50G-PON Wavelengths

Channel	Wavelength	Line Rate	Coexistence Impact
Downstream	1340-1344nm	49.7664Gbps	
UW2	1260-1280nm	12.4416Gbps	Displaces XGS-PON
UW1	1290-1310nm		Displaces GPON
UW2	1260-1280nm	24.8832Gbps	Displaces XGS-PON
UW1	1290-1310nm		Displaces GPON
UW1 Narrowband	1298-1302nm		Coexist with XGS-PON and GPON
Upstream	Work in Progress	49.7664Gbps	

3.2. TC-Layer

Each ITU-T PON specification includes a Transmission-Convergence Layer (TC Layer). The function and design of each is fundamentally the same but incorporates changes specific to the technology and needs at the time of the specification's release. The TC layer adapts the upper layer client entities to the underlying PMD and also includes the functions necessary to operate and manage the PON. [GPON-

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

G.HSP], defines a Common TC Layer (ComTC) that is the most advanced, intended to be applied to 50G PON. ComTC defines three sublayers as shown in Figure 3. The Service Adaptation (SA) sublayer is responsible for encapsulating user data into XGEM frames and multiplexing XGEM frames into the framing sublayer.

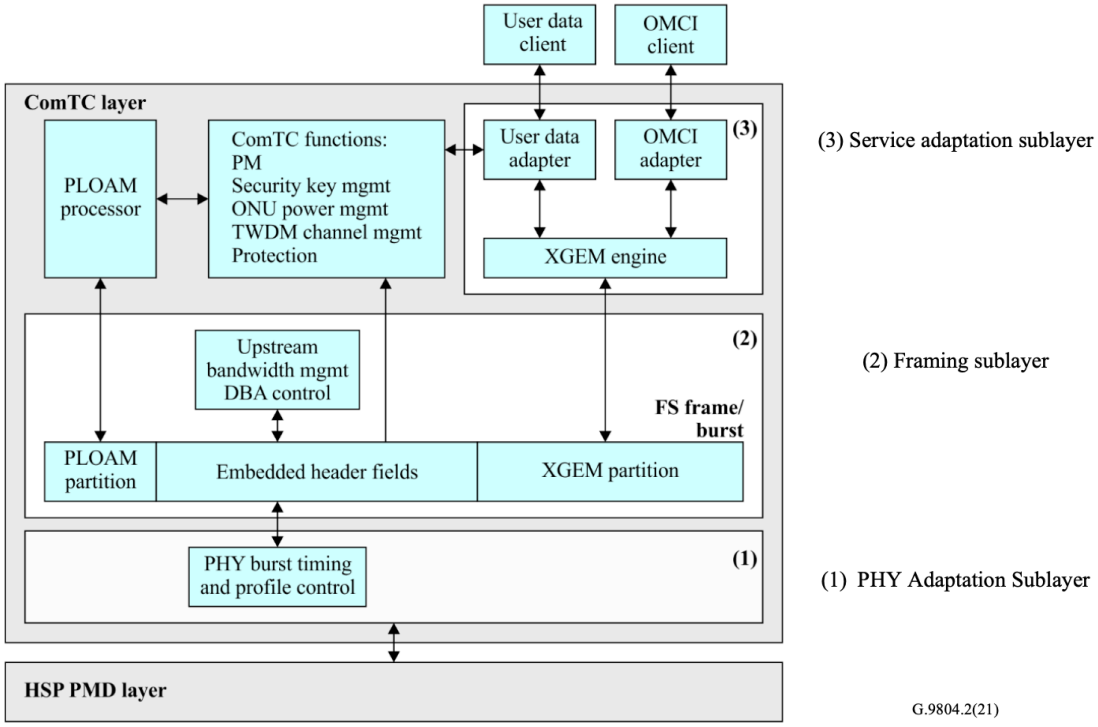


Figure 3 – GPON ComTC, PHY and MAC layers

Source:*ITU-G.9804.2-Figure6-3*

In the Framing Sublayer (FS), the XGEM data is encapsulated with overheads that are necessary to manage operation of the PON, primarily the Physical Layer OAM channel, which carries the information to manage the FS and PMD layer, and Embedded OAM data, which carries bandwidth allocation and dynamic bandwidth assignment fields. The PHY Adaptation sublayer enriches the bit stream to enable synchronization and time alignment of the bit stream and provides coding, scrambling and Forward Error Correction.

3.3. ONU Management

The Optical Network Unit (ONU) is the network terminal equipment located at the customer-side of the PON. The ONU is physically separate from the OLT and must be configured and managed, this is is supported by the ONU Management and Control Interface (OMCI) defined in ITU-T G.988.

OMCI has three main components: One is the Management Information Base (MIB) which specifies the database that enables configuration management, fault management, performance management and security management of the ONU. The MIB is described in terms of Managed Entities (ME) and their relationship to one another. The second component of the OMCI is the protocol used to convey the MEs between the ONU and the OMCI management function which is traditionally located in the OLT. The third component is the OMCI Management and Control Channel (OMCC) over which OMCI is

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

conveyed. Since OMCI is simply a client to the ComTC layer, the OMCC is a special case of an XGEM port on the PON.

3.4. 25GS-PON MSA

The 25GS-PON MSA (October-2020) developed outside the ITU-T as a "delta spec". It is written to adopt the relevant GPON specifications and to adapt the EPON PHY to them by calling out specific differences that are detailed in the new specification. The 25GS-PON MSA is a delta specification that provides the "glue" between: [Nx25G-EPON] for PMD layer and FEC (LDPC), ITU-T G.9807.1 (requirements) for TC layer (G.987.3) a.k.a. XG(S)-PON 10G and ITU-T G.988 for OMCI and the BBF TR-385 for Yang models. Key differences in the 25GS-PON PMD layer are the wavelengths and the line rate as shown in Table below.

Table 4 – 25GS-PON wavelengths

Channel	Wavelength	Line Rate Gbps	Coexistence Impact
Downstream (DW0)	1358nm +/- 2nm	24.8832	
Upstream UW1	1300nm +/- 10nm	24.8832 or 9.95328	XGS-PON
Upstream UW2	1270nm +/- 10nm	24.8832 or 9.95328	GPON
Upstream UW3	1286nm +/- 2nm	24.8832 or 9.95328	GPON+XGS-PON

25GS-PON holds closely to the TC Layer for XGS-PON found in ITU-T G.9807.1. The most significant change in the TC layer is the adoption of LDPC FEC from [Nx25G-EPON]. The majority of the remaining changes are necessary only to adjust frame byte counts, preserve alignment while preserving the 125µs frame size.

4. Brief Introduction to CPON

The advent of coherent optics technology in fiber transport has brought about a paradigm shift in optical transmission systems, facilitating a comprehensive overhaul in the widespread DWDM networks operating at data rates such as 100Gbps, 200Gbps, and 400Gbps per wavelength. Over the past decade, the scope of coherent optics has transcended long-distance transports, extending its purview to metropolitan networks, and now into optical edge networks. The next domain that coherent is poised to enter is the realm of FTTP, primarily characterized by point-to-multipoint PON, a revolutionary step for 100G PON and beyond. The utilization of coherent optics within a PON, known as Coherent PON (CPON), yields numerous advantages by implementing multi-dimensional modulation and coherent detection and digital signal processing techniques to achieve unprecedently high degree of flexibility in terms of transmission distance, split ratio, and the optical access architecture.

4.1. PHY Technology

Compared to alternative direct-detect solutions that just use amplitude (one dimension only) to represent the signal, coherent optical solutions use a local laser source as a reference to achieve linear conversion of the optical field instead of optical power used in direct detection. This enables modulation and detection using four independent degrees of freedom, including amplitude and phase in two polarizations. With a local oscillator, significant coherent gain is provided along with wavelength selection without the need of an optical filter. Additionally, power fading induced by chromatic dispersion in direct-detection systems is no longer an issue because of optical field recovery in coherent detection. These features coherent optics bring to optical transport systems, enable greater modulation efficiency and receiver sensitivity.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Coherent detection for short-haul networks enables a superior receiver sensitivity that allows for extended power budget. The multi-dimensional signal recovered by coherent detection also compensates for linear transmission impairments such as chromatic dispersion and polarization mode dispersion. The efficient use of fiber spectral resources results in more optical spectrum available for future use and enables future network upgrades through the use of multi-level advanced modulation formats.

CPON exhibits substantial potential for optimization and advancement across the optical, electrical, and digital domains. This includes extended reach, eliminating the necessity for repeaters, reducing real estate costs associated with cabinets and flexibility to accommodate diverse deployment scenarios through adaptable topology. It introduces operational simplicity by minimizing the number of ports and links that require management and prolonged technological lifespan, negating the frequent need for upgrades.

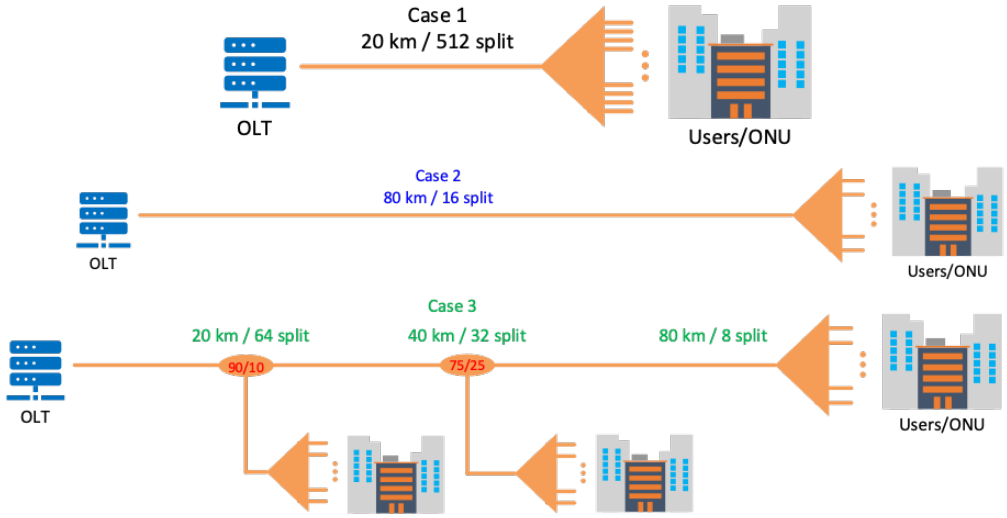


Figure 4 –CPON application scenario illustrations

Figure 4.1 illustrates 100G CPON target topology options. Case 1 involves opting for an elevated split ratio over a limited distance. For instance, a 512 split configuration is achievable at a transmission distance of 20 km, well-suited for situations requiring a high-density network in close proximity. Conversely, Case 2 demonstrates an alternative strategy emphasizing a reduced split ratio over an extended transmission distance. A 16 split configuration is demonstrated at 80 km of transmission distance, catering to rural applications where longer reach is imperative. Case 3 introduces a hybrid mode wherein various split ratios are judiciously combined across different transmission distances. Specifically, this involves a CPON port initially tapped at 20 km using a 90/10 passive splitter, thereby accommodating 64 splits. This is followed by a 40 km tap employing a 75/25 passive splitter, facilitating 32 splits. Finally, at a distance of 80 km, an additional 8 splits are supported. Case 3 is indicative of a distributed CPON architecture, strategically designed to optimize the conveyance of optical power over passive ODNs. In this dynamic context, each coupler diverts a predetermined portion of the CPON signal to a localized splitter equipped with a suitable number of ports, meeting demands of local ONUs. The integration of adjustable and remotely controlled couplers introduces a layer of automation, streamlining operational procedures.

4.2. Target Applications

CPON supports fiber-to-the-home (FTTH) connectivity, direct fiber connection to either multiple dwelling units or single-family units. In brownfield FTTH deployments with existing IM-DD PON, a

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

CPON network could be overlayed using Coexistence modules, allowing the operator to gracefully migrate certain customers to increase revenue opportunities or lower congestion for other customers on the ODN. CPON can support back-haul connectivity to the aggregation node. The aggregation node can contain Remote PHY devices (RPDs) and Remote MAC-PHY devices (RMDs), to support distributed Converged Cable Access Platform (CCAP) architectures. CPON can be utilized to provide base station connectivity, carry back-haul traffic in 5G RAN and also carry traffic from the mid-haul/front-haul segment of 5G RAN. Beyond these, CPON can find applications in Wi-Fi access, edge computing, IoT-based Smart City connections, etc. enhancing connectivity across diverse scenarios. The 100Gbps target is only a starting point, and the plan is to scale further in the future.

5. Comparison and commonalities

Both the GPON and EPON systems enable flexibility to support higher performance physical media dependent (PMD) interfaces without impacting the definition of the associated MAC and higher level PHY layers. This section looks at some of the commonalities, past approaches at unification and a potential path forward for a new technological development like CPON.

5.1. Comparison and commonalities of EPON and GPON

GPON and EPON provide a system that can be adapted to different line rates, number of wavelength channels, and signal modulation etc. Both GPON and EPON systems are layered in a similar fashion, supporting physical layer technology which physical transmission of data over the optical medium and a data link layer technology to enable point to multi point communication between OLTs and ONUs.

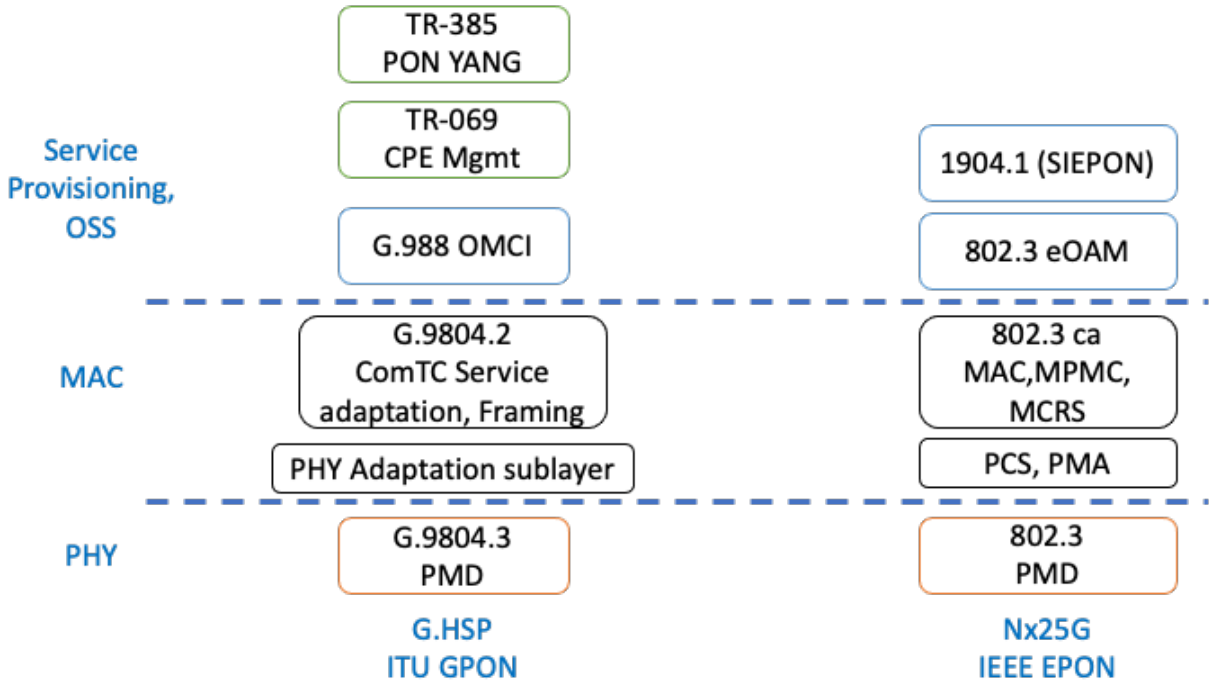


Figure 5 –PHY, MAC& OSS technology stacks for GPON& EPON

Figure 5 summarizes the various technology layers in GPON and EPON and tries to align the functionality, the horizontal lines to indicate equivalence. At the physical layer, both technologies define

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

a PMD (physical media dependent) layer. Above the PMD, is the layer which connects the MAC to the PHY, PHY Adaptation sublayer (in GPON) and PCS, PMA layers (in EPON), these define the FEC parameters, burst profiles, interleaving etc. The MAC layer above that defines how the data is framed and sent across one or multiple PHY channels. EPON allows for combining multiple PHY channels together (MCRS) , and at this time GPON does not have a solution for aggregating multiple channels.

5.2. One PON

There are two main camps in the development of PON standards: ITU-T PON and IEEE EPON, Figure 6. These camps share common requirements, particularly in terms of ODN and PMD aspects, the higher-cost components in deployment. The similarity in network architecture contributes to similar ODN designs, which demands similar PMD technical requirements. This allows for leveraging the same shared pool of optical components.

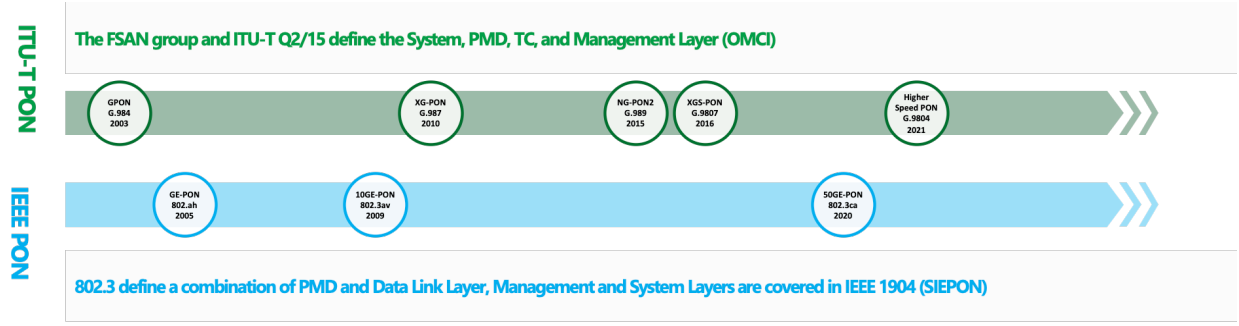


Figure 6 – Two Different PON Camps

However, the two camps differ significantly in their approaches for various aspects of the layered structure of the upper layer and network management. This encompasses distinct framing, timing, adaptation, and PLOAM message handling methods. The desire to unite these two camps has been present for a long time, because having a single PON standard, referred to as One PON, offers operators the advantage of achieving broader PON deployment while maintaining a clear evolutionary path. This, in turn, would lead to increased volume and subsequently drive down costs. However, this endeavor has not proven successful thus far.

Based on market deployment statistics from both OLTs and ONUs, there is approximately an 8:1 ratio in favor of ITU-T PON compared to IEEE EPON. Given ongoing trends, it seems challenging for this unification to succeed in the immediate future as well. The reasons include the diverse composition of members within the two camps, differing design philosophies, and the distinct FTTx/PON roadmaps pursued by various operators in both established and new deployment areas. Due to the complexities and challenges involved in designing and developing the next-gen PON, particularly the need for a common Physical Layer (optoelectronic components and ASICs) to share costs, industry experts are still working towards convergence. The revolutionary 100G or 200G CPON might hold the potential to realize this convergence, given its capability to innovate and address these challenges.

5.3. MAC Layer Approach for CPON

Both EPON and GPON have their own advantages and are used in various deployments around the world. The choice between the two depends on factors such as the required data rates, compatibility with existing infrastructure, and specific service offerings. Additionally, there has been an evolution in both families of PON technologies to provide higher data rates and capabilities. At this point, we are recommending a dual

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

path approach, to build a solution which will fit in with both the GPON and EPON MAC and higher layer technology stacks.

5.3.1. Why two MACs for CPON?

EPON and GPON remain among the most widely adopted PON technologies. The opportunities for PON equipment vendors are continuing in developed and developing countries. Some operators are selectively adding FTTH while other operators see an opportunity for PON infrastructure to support both residential and non-residential subscribers and applications. The adoption of 10G PON (XGS PON and 10G EPON) in developed markets motivated vendors, providing the opportunity to support higher speeds at lower cost compared to 1G EPON and 2.5G GPON.

While the current GPON and EPON market opportunity is sufficiently large, we reason that the market is unlikely to accept a third PON standard, and economically support a third type of PON technology. Hence our thought process is to align the new CPON technology with the current EPON and GPON technology, by adding the new CPON PHY layer technology to fit and work with both the existing EPON and GPON higher layer MAC technologies.

Reusing as much as possible from existing standards lowers the time to produce the technology versus completely creating a new PON MAC and higher layer from scratch to go with the new CPON PHY. Reusing portions of the existing silicon design components also reduces relative cost for a CPON solution.

5.3.2. Delta Specifications

The CPON working group is taking on a delta approach to ‘glue’ the CPON PMD under the two existing PON standards: IEEE [Nx25G-EPON] and ITU [GPON G.HSP]. Reusing technology and staying focused on getting the new CPON PHY to work with existing MAC layers minimizes the time to market.

EPON and GPON technologies have been heavily vetted by multiple PON vendor companies. As a precedent, the 25GS-PON MSA effort reused much of those technologies, and this approach was proven by product demonstrations in the market in about two years after the 25GS-PON MSA was formed.

For CPON, a separate delta specification will be required for each standard, i.e., a single CPON PMD specification will be interfaced with both [Nx25G-EPON] and [GPON-G.HSP]. The idea is to create a CPON PMD specification that is a “drop in” for an [Nx25G-EPON] PMD and [GPON-G.HSP] PMD, Figure 7. The proposal is to re-use the [Nx25G-EPON] LDPC mother code, with codeword length and puncturing adjustments as needed. The goal is to specify “delta” details for any higher layer adjustments, such as: Data rate, Framing, Synchronization, Channel skew, Grants: Gates, BWmap, etc. For channelization [Nx25G-EPON] defines two channels, each consisting of a downstream and upstream wavelength, originally intended to support up to four channels (pairs). G.9804.3 PMD Table 9-5 on specifies one downstream wavelength and two upstream wavelengths TWDM extends to 8 channel pairs.

As regard management and OSS protocols for both of these systems, work will likely be needed to develop management layer plane configuration for the new CPON physical layer/ PMD. Further detailed study is needed for ITU-T PON (with PLOAM and OMCI) and IEEE EPON (with OAM and SIEPON).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

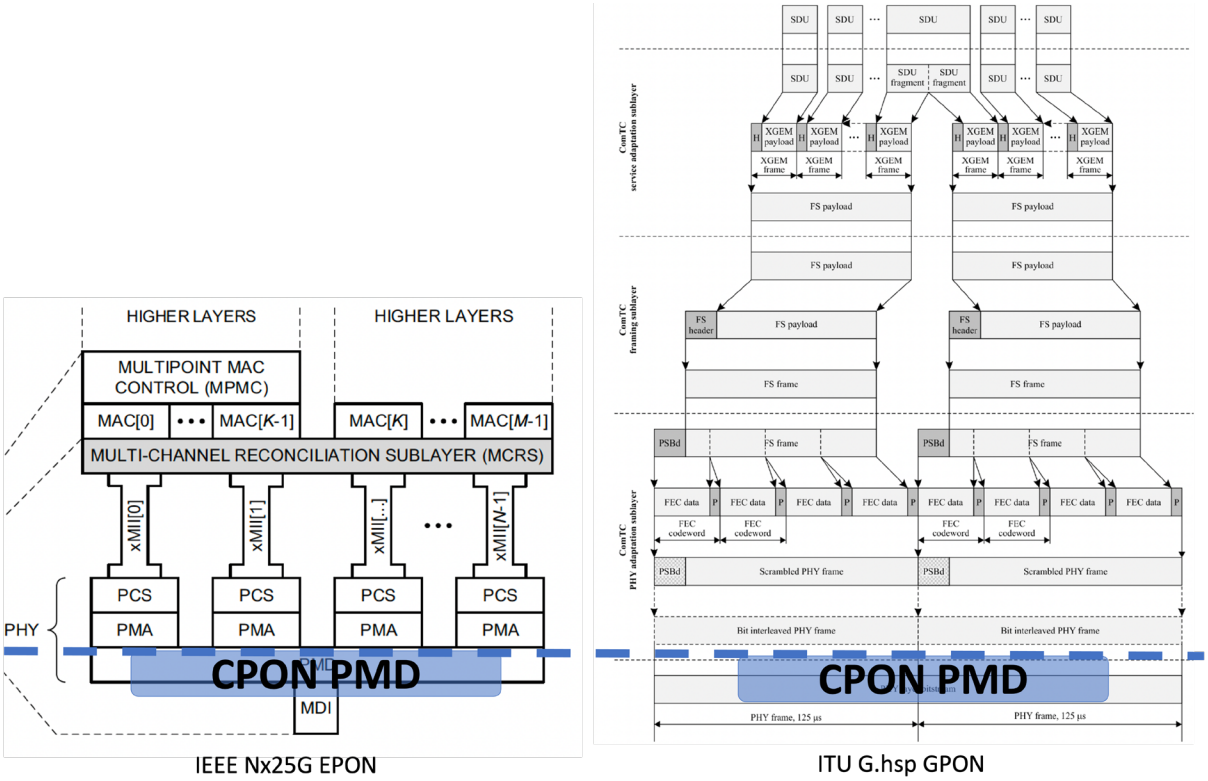


Figure 7 – A Drop-in CPON PMD for both EPON & GPON tecnology stacks

Source:*IEEE-Std-802.3-2022,and ITU-G.9804.2*

6. Scaling Considerations for CPON

Scaling from existing (EPON and GPON) MAC layer technologies to the latest CPON is possible with careful consideration. In this section we think through how the current technologies work and how scale them to the speeds and capacities of CPON. With a goal of at least scaling to 100G and beyond in future generations it needs tweaking of the current EPON and GPON implementations for the future.

6.1. EPON based MAC Layer

This section starts to dig deeper into the [Nx25G EPON] technology PHY-MAC interfaces and look at the framing and data rates. We investigate the MCRS, 25GMII, and the PCS layers to understand the current capacity and overheads. We then discuss methodologies on how to scale the existing Nx25G EPON technology to scale to 100Gbps and beyond to meet the goals of CPON.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

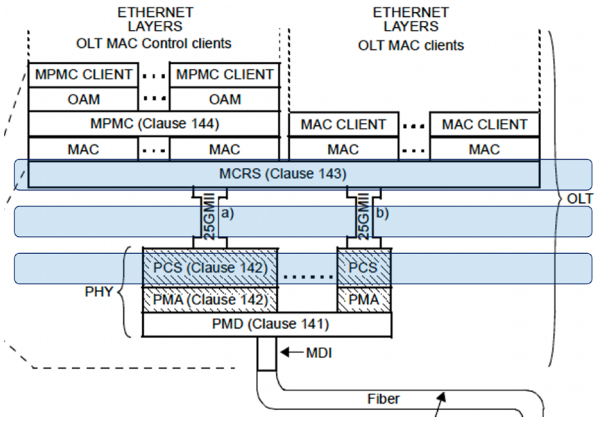


Figure 8 – Analysis of the line,frame , data rates on EPON tecnology stack

Source:*IEEE-Std-802.3ca-2020:Figure 144–2.*

6.1.1. Understanding EPON Overheads

6.1.1.1. Multi-Channel Reconciliation Sublayer

The MCRS adapts the bit-serial data stream received from an Ethernet MAC to the parallel format of the PCS service interface, across multiple physical channels. Each PCS instance represents an MCRS channel, that operates on envelope quantum (EQ) data units (consisting of eight control bits followed by 64 bits of data). Effectively, a single EQ comprises data from two consecutive medium independent interface (MII) transfers, with re-arranged data and control symbols, Figure 9.

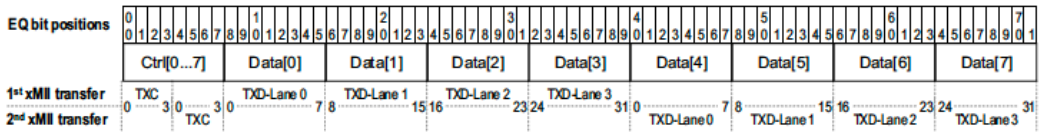


Figure 9 – Envelope Quantum Format in EPON

Source:*IEEE-Std-802.3ca-2020:Figure 143–2.*

The user data is split up into multiple of these EQ data units, with an envelope start header (ESH) and multiple envelope continuation headers (ECH). Only the ESH needs to be counted in overhead calculations, ECH doesn't count as it is used in place of a data preamble.

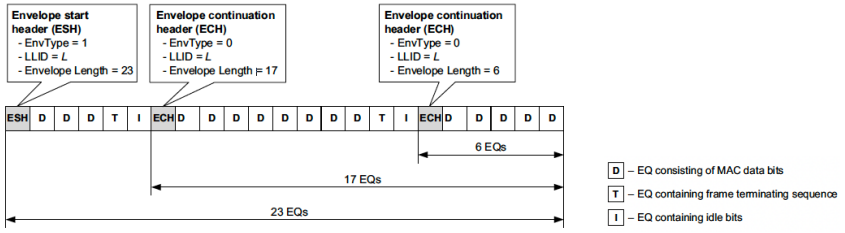


Figure 10 – Transmission Sequence consisting of three frames

Source:*IEEE-Std-802.3ca-2020:Figure 143–3.*

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 5 – MCRS Envelope Quantum Capacity

Layer	Size(bits)
User data	14272
EQ Data	64
EQ Control	8
EQ total bits	72
Number of Data EQs (= user data/ EQ Data)	223 (EQs)
EH (ESH or ECH = 1 EQ)	64
Number of EH	1 (EQ)
Control	8
EH + Number of data EQs	14336

For user data we are starting with 14272 bits, just as an example as this happens to fit within a codeword after some of the overheads.) and is transformed into 14,336 bits of data coming out of the MCRS.

6.1.1.2. Specific Media Independent Interface (xMII)

The Media Independent Interface (xMII) provides an interconnection between the MAC layered and the PHY layers. The Nx25G EPON 25GMII is capable of supporting 25Gbps data rates. The structure of each of the 25GMII interfaces in an MCRS system has a clock rate of 390.625 MHz, Table 6.

Table 6 – 25GMII Clock rate

Layer	Rate
MAC Data Rate	25Gbps
Time to move 1 EQ of data. 64 bits @ 25Gbps (2x32 + 2x4 ctrl)	2.56ns
25GMII Clock rate	390.625MHz

6.1.1.1. Physical Coding Sublayer PCS

The Physical Coding Sublayer (PCS) defines the 256/257B transcoder, scrambler, the FEC encodings and burst transmission sync patterns etc. used within EPON.

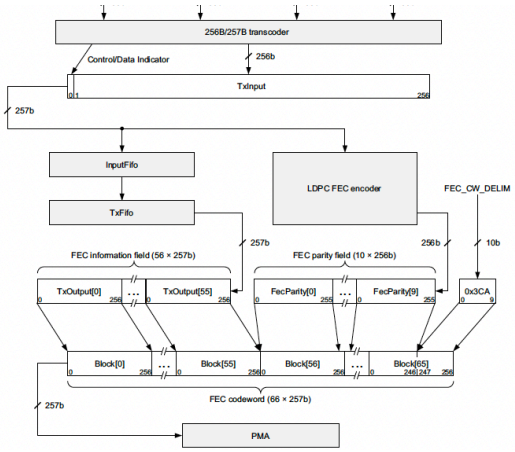


Figure 11 – EPON Physical coding sublayer

Source: *IEEE-Std-802.3-2022:Figure 142–4*

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 7 walks through the transformation the user data goes through at the PCs layer. The PCs layer operates on 256 bits at a time and ultimately outputs LDPC FEC code words off the size 16,962 bits.

Table 7 – PCS Layer Data rate calculations

Layer	Size(bits)
PCS User Data	256
Data after 256b/257B encoding	257
Number of FEC data blocks in a codeword	56 blocks
FEC data per codeword (256 * 56)	14336
FEC payload (257 * 56)	14392
FEC Party block size	257
Number of Parity blocks in a code word	10 blocks
FEC Parity	2570
FEC CW Size (FEC Payload + FEC Parity)	16962

Now for the upstream direction in the PCS layer, there is an additional upstream burst overhead. The ONU upstream burst begins with a synchronization pattern and ends with a termination sequence, which are not FEC protected. The synchronization pattern comprises of SP1 zone, optimized for laser on (Ton) and automatic gain control (AGC, Trx_settling); SP2 zone, optimized for clock and data recovery (T_{CDR}); and SP3 zone, optimized for the start-of-burst delimiter (SBD) pattern. Each SP zone is a multiple of 257 bits, aligning with the PCS line code of 256B/257B. The upstream burst ends with an end-of-burst delimiter (EBD). When received at the OLT, the EBD pattern allows for the rapid reset of the OLT FEC synchronizer, preparing the OLT for the next incoming upstream burst.

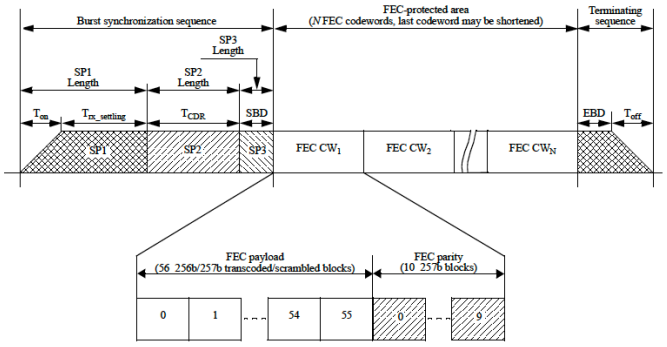


Figure 142–3—ONU burst structure, 3 zones

Figure 12 – ONU Burst Structure

Source:*IEEE-Std-802.3ca-2020:Figure 142-3*

The range of the upstream PHY overhead in time for 25G EPON is around ~500ns. (528.330ns as calculated in the table below). For CPON the expected range is in the range of 300 ~ 500ns.

Table 8 – Upstream PHY Layer Burst Overhead

Upstream	bits	Time(ns)
SYNC Pattern size	257	9.968
SP1 (20*SP)	5140	199.370
SP2 (20*SP)	5140	199.370
SP3 (1*SP)	257	9.968

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Upstream	bits	Time(ns)
Burst Sync Sequence (SP1+SP2+SP3)	10537	408.708
EBD	257	9.968
Laser off time, Toff	771	29.905
Terminating Sequence (EBD+Toff)	1028	39.874
Guard time (between bursts)	2056	79.748
US PHY Overhead per burst (BurstSyncSeq+TermSeq+Guard)	13621	528.330

6.1.1.2. Data rate Calculations

Given overheads in the downstream and upstream and the clock rates, one can calculate the effective data rate obtained in EPON. The 'Line Rate' is defined as the number bits/second transmitted on the fiber. The 'Data Rate' is the effective user data throughput.

For downstream data rate, Table 9, shows the line rate after the xGMII, and uses that rate to find the time it takes to transmit one full code word. Based on that time it calculates the data rate or the effective data throughput obtained in the downstream.

Table 9 – Downstream Data rate calculations

Layer	Rate
Line rate after 25GMII (25Gbps * 64/66)	25.78125Gbps
Time to transmit 1 CW. (FEC CW Size / Line rate)	657.92ns
Data rate for data bits (FEC Data per codeword / Time to transmit 1 CW)	21.790Gbps

For upstream data rate, Table 10, calculates the data throughput in a similar fashion.

Table 10 – Upstream Data rate calculations

Upstream	Rate/Size
Upstream line rate	25.78125Gbps
FEC CW Size	16962bits
US PHY Overhead per burst	13621bits
US FEC CWs (assume number of codewords in burst)	100cw
US FEC CWs + PHY burst Overhead	1709821bits
FEC Data bits only	1433600 bits
Transmit Time for US FEC CWs (FEC+PHYoverhead / upstream line rate)	66320.3297ns
US Data rate (data bits only) (FEC data / Transmit time for USFEC CWs)	21.616Gbps

6.1.2. Scaling of EPON technology for CPON data rates

As proposed if we build a CPON PMD to the EPON MAC, then we have a few options in terms of changing existing rates within EPON to allow us to scale to the CPON 100Gbps+ target service rates. Here are some of the options to change scale, for CPON starting with [Nx25G-EPON] based upper layers.

- If CPON decides to go down the route of multiple subcarriers then it can use 4 subcarriers, each at 25Gbps, and perhaps 8 subcarriers in the future.
- If CPON decides on a single channel PON i.e., just one carrier, then the MAC would need to change to handle higher speeds, this can be done in a couple of ways.
    - CPON could increase from the current 25 Gbps MAC data rate to a 100 Gbps data rate. This would mean that we would need to increase the xGMII Clock rate by four times,

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

390.625x4 → 1562.5MHz. It needs to be investigated if it is possible to cost effectively build designs at this high clock rate.

- Another option is to design a new xGMII, with a larger bit width: i.e., from supporting 64-bit transfers now to supporting 256-bit transfers. This is a bigger effort given that the XGMII interface is vetted by the EPON community and needs to go through that process.

It also makes sense to align with the 100 Gigabit Ethernet Applications and make sure that the line rates can support this. Table 11 shows the line rates and data rates achievable by simply multiplying the number of 25Gbps channels, it shows some of the choices in getting to the 100Gbps mark. As we can see in the table below, a multiplier of 4 gets us to a line rate above 100Gbps second and multiplier of five gets us to a data rate above a 100Gbps.

Table 11 – Potential Multipliers for Data rates

	Line Rates Gbps		Data Rates Gbps	
Multiplier	**DS**	**US**	**DS**	**US**
1	25.781	25.781	21.790	21.616
2	51.563	51.563	43.580	43.233
3	77.344	77.344	65.370	64.849
4	103.125	103.125	87.160	86.465
5	128.906	128.906	108.949	108.081
6	154.688	154.688	130.739	129.698
7	180.469	180.469	152.529	151.314
8	206.250	206.250	174.319	172.930
9	232.031	232.031	196.109	194.547
10	257.813	257.813	217.899	216.163

6.2. GPON based MAC Layer

This section starts to dig deeper into the ITU [GPON-G.HSP] technology PHY-MAC interfaces, framing and data rates. In attempting to evaluate the potential performance of a 100Gbps PON that adopts the ITU-T model of PON, it is necessary to understand the Common Transmission Convergence (ComTC) layer and its sublayers: Service Adaptation, Framing, and PHY adaptation sublayers. We investigate these below to compute the current capacity and overheads and discuss methodologies on how to scale these to 100 Gbps and beyond.

6.2.1. Understanding GPON Overheads

6.2.1.1. XGEM Framing

The Service Adaptation layer receives user data, Service Data Units (SDU), from client applications. Once received, the SA sublayer encapsulates the SDU in an XGEM frame, Figure 13. SDUs are allowed to be fragmented, but most modern networks are based on Ethernet frames, and ComTC does not permit Ethernet frames to be fragmented. There is nothing in the XGEM frame that would need to change to enable ComTC to scale up to 100Gbps.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

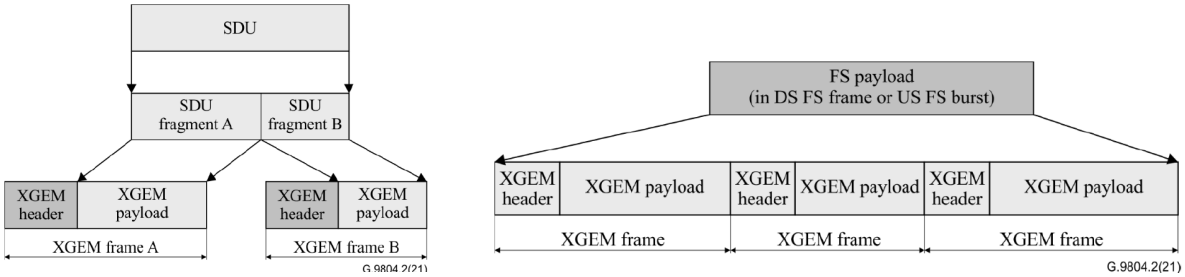


Figure 13 – Fragmentation at SA Layer,XGEM Encapsulation in FS Payload

Source:*ITU-G.9804.2-Figure 9-4, Figure 9-1*

After XGEM encapsulation occurs, the XGEM frames are assembled into the Framing Sublayer (FS) payload. The FS payload, Figure 13 is simply a concatenation of XGEM frames. In the downstream, the number of XGEM frames in the FS Payload is dependent on the line rate and the PHY frame size. The PHY frame size is always 125µsec and in the downstream PHY frames are transmitted in a continuous stream. Therefore, as the line rate increases, the amount of data that can be transmitted in a PHY frame will increase accordingly. In the upstream, data is transmitted in bursts that can be smaller than the PHY frame, but the same scaling principle can be applied, i.e., as the line rate increases, the amount of data that can be transmitted in a given time period will increase accordingly.

6.2.1.2. Framing Sublayer

The framing sublayer is slightly different in the upstream versus the downstream.

6.2.1.2.1. Downstream FS Frame

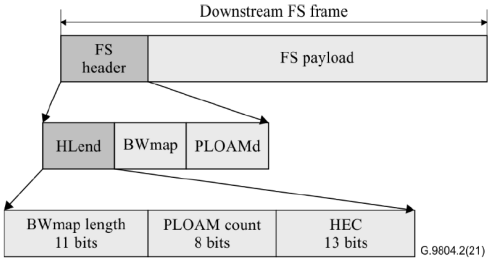


Figure 14 – Downstream FS Frame Format

Source:*ITU-G.9804.2-Figure 8-1*

The downstream FS frame contains the FS payload discussed above and an FS Header, Figure 14. The FS header carries several fields, but the BWMap is of particular interest relative to scaling ComTC to support 100Gbps operation, (BWMap contains structures that convey upstream grant opportunities to ONUs).

6.2.1.2.2. BWMap

The BWMap, Figure 15, contains a number of allocation structures, each containing a grant start time and grant size. Interpretation of these two values is directly related to the line rate of the PON and the PHY frame size of 125µsec. The Start Time field(16bits) potentially divides a PHY frame into 65536 “slots”, the ComTC divides the PHY frame into 12150 “slots”, each equal to one-time quanta (Q_0).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

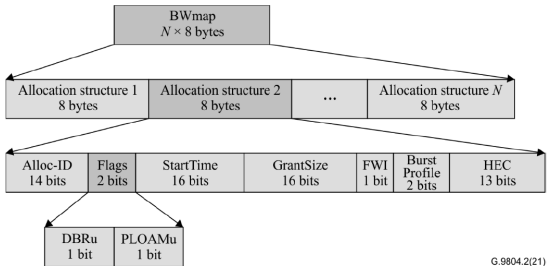


Figure 15 – Structure of the BWMap

Source:*ITU-G.9804.2 -Figure8-2*

Similarly, Grant Size is expressed as a value from 0 to 12149. The interpretation of GrantSize is dependent upon the line rate, so it follows that the interpretation of GrantSize will need to be adjusted for 100Gbps operation.

6.2.1.2.3. Upstream FS Frame

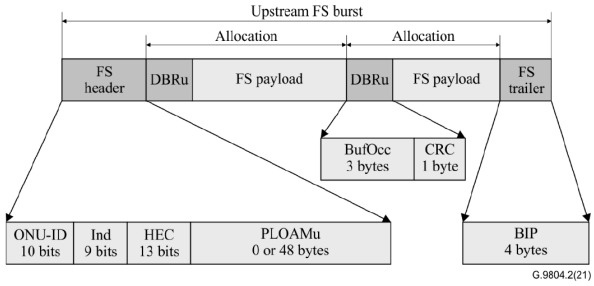


Figure 16 – Upstream FS Frame Format

Source:*ITU-G.9804.2-Figure8-5*

The upstream FS header contains the ONU-ID of the transmitting ONU, a series of Ind (ication) bits and the upstream PLOAM (PLOAMu) message. Because the upstream is a time division multiplexed system, the upstream FS frame, unlike the downstream FS frame, varies in size based on the Grant Size.

6.2.1.3. PHY Adaptation Sublayer

The PHY Adaptation sublayer connects the framing sublayer to the PHY layer by coding the FS frame and then encapsulating it with additional bit patterns necessary to ensure correct reception by the receiver.

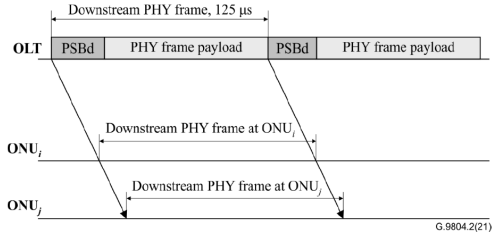


Figure 17 – ComTC Downstream PHY Framing

Source:*ITU-G.9804.2-Figure10-1*

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The downstream is a continuous stream of (125µs) PHY frames, Figure 17, containing a PSBd and a payload which is a bit stream after the FEC encoding, scrambling, and interleaving of the FS frame. In enabling 100Gbps operation, little would need to change in the PHY frame payload computation (aside from potential changes in the FEC), and these are not fundamentally related to the line rate. The PSBd contains data related to the PHY layer operation, timing, and alignment, and are not related to the line rate and not a concern in scaling to 100Gbps.

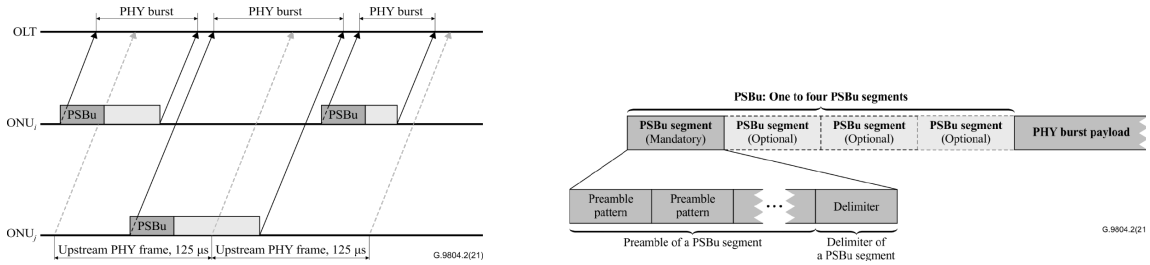


Figure 18 – Upstream PHY Framing, PHY Bursts and Preamble Format

Source:*ITU-G.9804.2-Figure10-5, 10-6*

The upstream PHY frame (Figure 18) is very similar to the downstream PHY frame. The key differences are replacement of the PSBd with a PSBu, the burst-nature of the upstream channel, and the upstream PHY burst varying in length based on the grant size issued by the OLT. These factors make the upstream PHY burst a key item for study when scaling to 100Gbps.

The PSBu provides a necessary function to enable burst reception at the OLT. It is the bit pattern used by the upstream receiver to perform clock data recovery, equalization, and automatic gain control.

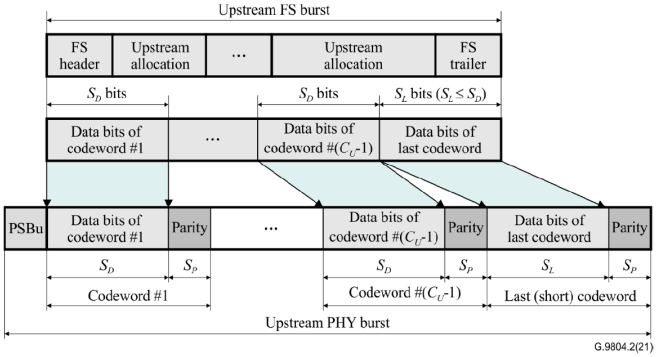


Figure 19 – Upstream FS Burst with FEC

Source:*ITU-G.9804.2-Figure10-11*

The PSBu can be customized on a per ONU basis to allow the operator to adjust for network and equipment characteristics. The PSBu length varies based on operator configuration but has a maximum length that is determined by the line rate (preamble in a PSBu segment can be up to $1000Q_0$).

Another important factor in assessing the performance of PON is the Forward Error Correction (FEC). [GPON G.HSP] specifies the use of LDPC. The general function of FEC is achieved by encoding the data to be transmitted and adding parity bits to the outgoing data stream, Figure 19. An LDPC(17280, 14592) code is used for 50G PON, meaning that a codeword is 17280 bits long and includes 14592 bits of data and 2688 bits of parity.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6.2.2. Scaling of GPON technology for CPON data rates

6.2.2.1. ComTC Changes to Support 100Gbps

Performance of PON can be evaluated on a few different dimensions. The major concerns are the impact of PON-specific overheads like FEC, framing, and physical layer overhead (preamble + guard time). Relative to scaling from 50G to 100G it will be important to identify the areas of ComTC that will need to be adjusted. It is important to state the assumption that the 125µsec frame size is immutable across GPON specifications, and that assumption will hold for CPON. Another aspect of the ITU-T series of PON specifications is the relationship of the line rate among the specifications, all derived from the SDH hierarchy, demonstrated in Table 12.

Table 12 – Relationship of Various PON Line Rates

PON Specification	Upstream Line Rate/ SDH Level	Downstream Line Rate/ SDH Level
G.984 GPON	1.24416Gbps = OC-24	2.48832Gbps = OC-48
G.9807 XGS-PON (10G PON)	9.95328Gbps = OC-192	
25GS-PON	24.8832Gbps = 10xOC-48	24.8832Gbps = 10xOC-48
G.9804 V.HSP (50G PON)	ρ_0 = 12.4416Gbps = 10xOC-24	
	2xρ_0 = 24.8832Gbps	
	4xρ_0 = 49.7664Gbps	

However, none of these line rates live up to the name given to each, especially when overheads are considered, e.g., XGS-PON is a “10G” PON, but the maximum user data capacity is approximately 8.8Gbps when FEC is enabled. This need not be the case for 100Gbps CPON or any future PON.

Specific parameters that need to be addressed to adapt ComTC for 100Gbps support are:

- Line Rate to Support 100Gbps User Data capacity
- BWMap Calculations
- Upstream Physical Layer Overhead Parameters

6.2.2.2. Line Rate

The main contributor to reduced data capacity is FEC coding (added parity bits) and physical layer overhead. To simplify the current study, framing overheads will be ignored, but an analysis attempting to achieve 100Gbps user capacity should take framing overheads into account.

The literature is rich with proposals and specifications for FEC intended for use in 100Gbps communication systems. E.g., the [OpenZRMSA], specifies OFEC at a coding rate of (4096,3552). While not proposing coherent transmission, Nokia has demonstrated 100Gbps PON, [FLCS PON], using PAM4 and FEC based on the EPON LDPC (17664, 14592). Since CPON is yet to make a FEC choice, we adopt the [Nx25G-EPON] FEC for analysis. This uses a codeword length of 17,664 bits with 14,592 information bits and 3,072 parity bits.

A simple calculation shows that the FEC overhead is: $FEC\ Overhead\ \% = \frac{3072}{14592} \times 100 = 21\%$

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

To achieve 100Gbps of user data capacity, the nominal line rate would need to be at least 21% higher than 100Gbps, or at least 121Gbps. For the purpose of further study, we propose a nominal line rate of 10 times the ComTC fundamental line rate of 12.4416Gbps.

In other words, $\phi = 10$, $\rho_0 = 12.4416Gbps$

such that the nominal line rate for 100Gbps CPON might be $\rho_0\phi = 124.416Gbps$.

This is an option for further study as CPON is expected to be able to scale to higher rates such as 200Gbps, 400Gbps, etc.

7. Challenges and Future Considerations

This section discusses other related considerations to scaling the current EPON and GPON technology to the CPON technology.

7.1. FEC Evolution & Comparison

Forward error correction has been a powerful tool in the cable industry for many years. In fact, perhaps the single biggest performance improvement in the DOCSIS 3.1 specifications was achieved by changing the FEC (previously Reed-Soloman) to a new coding scheme with improved performance: low-density parity check (LDPC). Similarly, FEC has also become an indispensable element for high-speed optical transmission systems, especially in the current coherent optical transmission age. Since this is the first "marriage" of coherent and PON technologies, let's review current FEC usage in these two fields.

Table 13 – FEC in Optical Communication Links

HD: hard decision SD: soft decision	Coding overhead rate	Net coding gain (NCG) dB @10E-15	pre-FEC BER threshold
GFEC (HD)	6.69%	6.2	8.0E-5
EFEC (HD)	6.69%	8.67	2.17E-3
100G: Staircase FEC (HD)	6.69%	9.38	4.5E-3
200G: oFEC (SD)	15.3%	11.1 for QPSK 11.6 for QAM16	2.0E-2
400G: cFEC (SD)	14.8%	10.4 for QPSK 10.8 for QAM16	1.22E-2

On the coherent side, CableLabs adopted Hard-Decision (HD) Staircase FEC at 100Gbps, achieving an NCG of 9.38dB with pre-FEC BER of 4.5E-3, verified in the P2P Coherent Optics Interoperability Event. At 200Gbps, openFEC (oFEC) with NCG of 11.1dB for QPSK and 11.6dB for 16QAM is used in CableLabs' P2P Coherent Optics PHYv2.0 Specification, suitable for various scenarios. This oFEC was also standardized for metro applications by Open ROADM. For 400G, OIF adopted cFEC with inner Hamming and outer Staircase codes. The 400G IA offers NCG of 10.8dB and pre-FEC BER of 1.22E-2 for dual-polarized 16QAM, tailored for DCI. Table 13 summarizes standardized FEC performance in coherent optical fiber transmission systems.

On the PON side, the evolution of FEC has seen several advancements over time. The first generation of FEC for PONs were based on Reed-Solomon (RS) codes (255, 239), which offer good performance at low bit error rates (BER), Figure 20. As the demand for higher data rates and the corresponding link budgets in PONs has increased, the need for more powerful FEC has also increased.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

At 10G era, RS code moves to RS (255,223) and the truncated form (248, 216) with the NCG to 7.1dB. Second-generation FEC for PONs are based on LDPC codes, which offer significantly better performance than RS codes at high BERs. The [Nx25G-EPON] standard has adopted high-performance, hard-decision (HD), LDPC FEC with a BER threshold of about 1E-2 and a rate of 0.849 and 9dB of coding gain [DSP 50G].

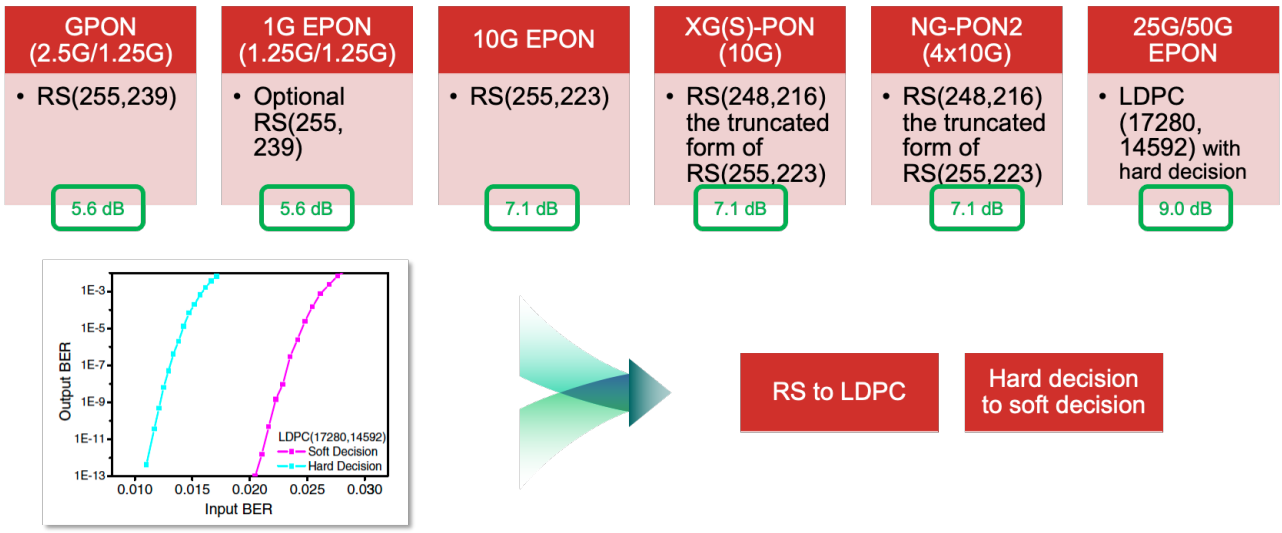


Figure 20 - FEC Evolution in PONs

In 50G and beyond PON, the link budget is more challenging to achieve. Therefore, soft-decision (SD) FEC is preferred to fully exploit the DSP function because SD FEC uses the full information of the received signal to make decisions about whether or not there is an error. This allows SD FEC to correct more errors than HD FEC, which can help to improve the link budget. Flexible FEC code rate has also been proposed to allow the OLT to choose the FEC code rate between 0.733 to 0.889 that is best suited for the current channel conditions [DSP 50G]. This can help to improve the BER and achieve higher throughput.

So, what kind of FEC should CPON choose? The requirements include being able to provide enough NCG to meet the different link budget requirements in CPON applications. The industry is currently considering how to leverage FECs that are already in use, especially the LDPC FEC for PONs. The LDPC FEC from [Nx25G-EPON], 25G-PON MSA, and [GPON-G.HSP] has been shown to be effective in other PON standards, only small changes are needed for the existing chip design and the compatible operation with PMD and MAC layers.

7.2. Channel Bonding

In [Nx25G-EPON], the latest evolution of EPON allows for 25G, 50G and potentially 100G capacities by introducing the concept of channel bonding. This strategy has already found utility in the cable industry through DOCSIS, which now has proven the channel bonding in multiple versions of DOCSIS. Channel bonding in EPON facilitates the aggregation of multiple 25G wavelengths to increase throughput. An ONU could be enabled to both transmit and receive data at a rate of 100G by aggregating four 25G wavelengths. Channel bonding serves as a mechanism for augmenting the speed of the next generation PON in a gradual manner, thereby extending the network's operational longevity. [Nx25G-EPON] specifies a simple and efficient channel bonding method for dynamic bonding of multiple channels. The

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

MCRS, see 6.1.1.1, enables multiple MACs to interface with multiple xMIIs. This bonding method is controlled by the OLT scheduling and does not require any explicit provisioning or configuration. Under a channel-bonded approach, the dynamic bandwidth allocation (DBA) functions will need to orchestrate upstream transmissions across one or more wavelengths simultaneously.

The question for CPON technology is if we start with a single carrier or multiple subcarriers with channel bonding enabled. [Nx25G-EPON] can accommodate new PMDs with different characteristics. E.g., A single carrier TDMA PON architecture with 100Gb/s line rate or an architecture where each ONU can send and receive on multiple sub carriers. Currently [GPON-G.HSP] does not have support for Channel bonding. For CPON, we continue to analyze, investigate and debate the path forward. Enabling support for channel bonding from day one even if we choose to have a single carrier looks to be prudent so that next generation of CPON technology could easily build off of that base capability. Enabling channel bonding is an easy way to future proof the MAC layer of CPON allowing for easy extensibility.

7.3. Single Carrier vs Digital Sub-carrier

There are two potential options for CPON– conventional single carrier (SC) or digital sub-carrier (DSC) PONs, Figure 21. SC is the most common type of PON and uses a single wavelength to transmit data to and from multiple ONUs. The data is divided into time slots, and each ONU is assigned a specific time slot to transmit and receive data. DSC is a newer type of PON that uses two or more frequency bands in digital domain over a single coherent transceiver to transmit data to and from multiple ONUs. The data is divided into time slots and frequency bands, and each ONU is assigned a specific time slot and frequency band to transmit and receive data.

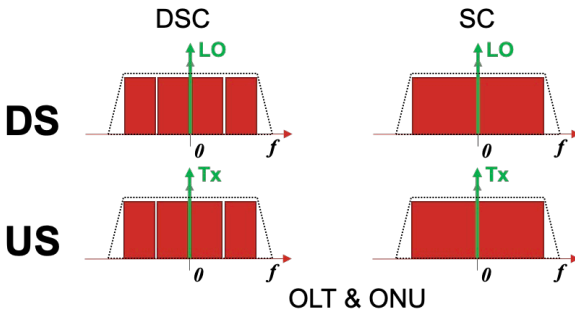


Figure 21 – Spectral Illustration of DSC and SC

The comparison between DSC and SC implementations in CPON presents distinct advantages and trade-offs. DSC offers several benefits, such as the reduction of burst overhead in upstream transmission for time-based sources of overhead. By allowing bandwidth trading, DSC facilitates additional link budget while maintaining overall capacity. It introduces the potential to employ sub-carriers for individual customers/services, enhancing network flexibility. However, there are notable concerns associated with DSC. It is anticipated that DSC implementations might lead to a reduction in nominal transmission (Tx) power from increased peak-to-average-power (PAPR), thus affecting the link budget by a magnitude of 3dB or more. DSC entails increased complexity, including subcarrier frequency and power control, as well as the need for tracking laser phase noise.

SC technology offers distinct advantages, it simplifies the achievement of the target link budget, as it avoids the PAPR loss inherent to multi-carrier approaches. Additionally, excess bandwidth can be exploited for power enhancement through shaping, contributing to improved performance. Seamless

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

integration of SC modulation with existing MAC layers eliminates compatibility concerns. However, SC modulation faces its own set of challenges. Its lack of flexibility restricts the physical division of customers or services and the extension of reach. SC modulation introduces increased burst overhead stemming from time-based sources of overhead.

For system level design, the integration of Semiconductor Optical Amplifier (SOA) amplification, (external or integrated at the OLT), and potential low-cost implementation of Silicon Photonics, introduce additional factors that warrant consideration in the selection of the most suitable modulation technique for specific architectures and operational demands. Balancing these benefits and challenges is crucial to harness the long-term potential of CPON while addressing current market demands and infrastructure readiness.

7.4. Other challenges

The transition from legacy PON networks needs to be managed and operational systems need to be adapted to the new CPON technology. For EPON, the SIEPON specifications enable interoperability at the PHY and MAC layer between the OLT and ONU. This system-level and network-level standard, allows interoperability of the transport, service, and control planes across multiple vendors. Operators have built back-office software tools to provision services on the PON network on top of these 802.3 ah/av/ca, SIEPON and DPOE technologies. This is a fair amount of investment/expertise that operators may like to preserve. They also would like the ability to provision services across different types of PON networks and perhaps even other access networks using the same/similar provisioning systems.

On the GPON side, OMCI is common to the various iterations of GPON technology and other industry bodies, (such as the BBF), support these technologies with test plans and other work related to making the equipment interoperable. For any new technology, as the PHY matures many of the interoperability issues come at the PLOAM layer or at the OMCI layer. Work needs to be done for the CPON technology as well to allow for interoperability. This includes defining common OMCI messaging, third party testing, etc.

Deploying the fiber infrastructure is by far one of the biggest investments for an operator. New PON systems thus must be compatible with the existing ODN, which may include the trunk fiber connections between hub site to remote fiber nodes. This minimizes infrastructure costs and enables a non-disruptive migration toward CPON deployments. Three forms of coexistence are required for CPON systems: coexistence with IM-DD (legacy) PON systems, coexistence with P2P DWDM systems, and coexistence with other CPON systems (aka CPON stacking). Coexistence of CPON systems with other PON and DWDM systems is essential for operators to be able to migrate to CPON without having to replace their entire fiber infrastructure and deploy CPON systems in a variety of network scenarios.

8. Conclusion

As the PON evolves toward a data rate of 100Gbps or higher, PON technology based on coherent optical technology is a very promising solution due to its performance and potential. The cable industry is working towards developing new CPON technology for use in its networks. CPON is a fundamental change to physical layer, with the transmission technology moving from the IM-DD technology to coherent optical technology. The CPON technology will solve for not just for access network applications but also for transport applications like backhaul.

The new CPON PHY layer will need a supporting MAC layer to build a complete solution. We propose both a GPON based MAC layer and an EPON based MAC layer. This approach enables quicker time to market by implementors by reusing large digital logic components from previous implementations. It

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

also allows operators who are already invested in either GPON or EPON to reuse a lot of the provisioning and back-office systems. This paper analyzed each of the EPON and GPON technologies and got an understanding of the EPON stack (MCRS, PCS, PMA) as well as the GPON stack (ComTC Service adaptation, framing and PHY adaptation layer).

EPON technology can be scaled to 100Gbps speeds by using a couple of different options. With multiple subcarriers approach, CPON technology can use 4 subcarriers, each at 25Gbps and use the EPON MAC layer as is. If we go down the single carrier approach, the MAC implementation will need to change to handle higher speeds, by increasing the MAC data rate to 100 (from 25Gbps today) which means increasing the xGMII clock by 4 times to 1.5625GHz. Another option is to design a new xGMII, with a larger bit width to support 256-bit transfers.

On the GPON side, the technology can be scaled to 100Gbps speeds by choosing a nominal line rate of 10x the ComTC fundamental line rate of 12.4416Gbps. Many of the other framing formats remain unchanged from the latest GPON technology. From this analysis we can prove the feasibility of both technologies, that they can work /scale for CPON data rates.

The choice of FEC is a big decision for the technology development, with many choices from Coherent and PON technology. The industry is currently considering leveraging the FEC that has already been developed, the LDPC FEC from 802.3ca, 25G-PON MSA, and G.HSP (all using the same mother code) has been shown to be effective in other PON standards, and only small changes are needed for the existing chip designs and compatible operation with PMD and MAC layers.

We also recommend the technology support a data rate of 100Gbps traffic (vs. a 100Gbps line rate target which only obtains a lower user data rate after FEC overheads). The 100Gbps goal is only a starting point, and the CPON technology can scale this further (to 200/300/400Gbps and beyond) to accommodate service targets for the future. As these specifications are developed the next generation requirements needs to be thought through, to allow for smooth evolution of CPON generations.

Abbreviations

ASIC	Application Specific Integrated Circuit
bps	bits per second
CPON	Coherent Passive Optical Network
DPoE	DOCSIS Provisioning of EPON
DSP	digital signal processing
EPON	Ethernet Passive Optical Network
FTTH	Fiber to the Home
FTTP	Fiber to the Premise
FEC	forward error correction
HD	Hard Decision
GPON	Gigabit-capable Passive Optical Network
LDPC	Low Density Parity Check
MAC	Medium Access Control
MCRS	Multi-Channel Reconciliation Sublayer
NCG	Net coding gain
ODN	Optical Distribution Network (ODN)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

OLT	Optical Line Terminal
ONU	Optical Network Unit
OAM	Operations, Administration, and Maintenance
PCS	Physical Coding Sublayer
PHY	Physical Layer
PSB(d/u)	Physical Synchronization block. (downstream or upstream)
PMA	Physical Media Attachment
PMD	Physical Medium Dependent sublayer
P2MP	point-to-multipoint
WDM	Wavelength division multiplexing
xGMII	(1/10/25/50Gbps) Media Independent Interface

Bibliography & References

[Nx25G-EPON] IEEE Std 802.3ca-2020: *IEEE Standard for Ethernet Amendment 9: Physical Layer Specifications and Management Parameters for 25 Gb/s and 50 Gb/s Passive Optical Networks - or -* IEEE Std 802.3-2022: *IEEE Standard for Ethernet.*

[GPON-G.HSP] ITU-T Recommendation G.9804.2: (09/2021) *Higher speed passive optical networks – Common transmission convergence layer specification ,* ITU-T Recommendation G.9804.3: (09/2021) *50-Gigabit-capable passive optical networks (50G-PON): Physical media dependent (PMD) layer specification*

[DSP50G] B. Li et al., "DSP enabled next generation 50G TDM-PON," *in Journal of Optical Communications and Networking, vol. 12, no. 9, pp. D1-D8, September2020, doi: 10.1364/JOCN.391904.*

[FLCS-PON] R. Borkowski et al., "*World's First Field Trial of 100 Gbit/s Flexible PON (FLCS-PON)," 2020 European Conference on Optical Communications (ECOC),* Belgium, 2020, pp. 1-4, doi: 10.1109/ECOC48923.2020.9333413.

[OpenZRMSA] Open ZR+MSA Technical Specification version 2.0, July2022, https://openzrplus.org/site/assets/files/1091/openzrplus_2p0.pdf

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Shifting the Conversation from Speeds, Feeds, and WiFi 7 to a Multidimensional Model for Flawless Application Performance

A Technical Paper prepared for SCTE by

Adam Hotchkiss
VP, Product Solutions
Plume Design, Inc.
325 Lytton Ave, Palo Alto, CA 94301, USA
adam@plume.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Abstract 3
2. Introduction 3
3. Evolution of 'speeds and feeds' and their relevance 3
4. The Role of WiFi in modern QoE 3
5. Unpacking the Seven Dimensions of Full Stack Optimization QoE 4
6. Full Stack Optimization: A Revolutionary Approach 6
7. Conclusion 13

Abbreviations 14

Bibliography & References 14

List of Figures

Title **Page Number**

Figure 1 - Full Stack Optimization dimensions 6
Figure 2 – Dynamic Deep Packet Inspection CPE architecture 7
Figure 3 - Households running real-time applications 8
Figure 4 - Streaming application market share 8
Figure 5 - Video conferencing application market share 9
Figure 6 - Network topology with adaptive WiFi functionality 10
Figure 7 - Internet latency measurements for real-time applications 11
Figure 8 - Prioritized packet delivery for low latency applications 11
Figure 9 - Latency improvement for different application types in congested WiFi network 12
Figure 10 - Measurements for Application Performance 12

List of Tables

Title **Page Number**

Table 1 – Application session statistics 7

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Abstract

Central to every 21st-century home are ‘smart’ applications that require robust, secure, and dependable internet to operate smoothly. These applications encompasses everything from 4K video streaming to advanced heating controls and video conferencing. With ever-increasing network complexity, households are progressively dependent on Communications Service Providers (CSPs) for expertly-managed connectivity solutions to ensure superior Quality of Experience (QoE) across floorplans, devices, and apps.

Our industry spends an astounding amount of time discussing WiFi 7’s expansive speed and performance as a means to solve for evolving application use in the home, but that’s the wrong conversation. WiFi 7 is only an enabling technology to easily connect and send data between devices. We must fundamentally shift the conversation around delivering flawless application experiences for every device in the consumer’s home, regardless of WiFi standards.

This paper proposes a comprehensive QoE delivery approach down to the application level, harnessing the potential of Plume’s Full Stack Optimization. It further offers technical guidelines for CSPs on the deployment of cloud-based learning algorithms. These algorithms, paired with precise Key Performance Indicator (KPI) metrics, leverage capabilities of any WiFi networking protocols, and open-source platforms to address the intricate challenges of ensuring optimal connectivity.

2. Introduction

In today's digital age, households are no longer just physical spaces but digital ecosystems. Real-time applications such as Netflix, FaceTime, Zoom, and various online gaming platforms have emerged as mainstays. With this transformation, there's an elevated demand for a new way to optimize and measure true consumer Quality of Experience (QoE). Historically, the answer lay in amplifying “speeds and feeds”, but such strategies now exhibit diminishing returns in ensuring optimal user experience.

3. Evolution of 'speeds and feeds' and their relevance

For decades, when users encountered issues with their online experience, the primary solutions were straightforward: boost the internet bandwidth or transition to the next-gen WiFi technologies. This method predominantly produced satisfactory results; however, now it is faltering in plain sight. Recent observations indicate diminishing efficacy due to two critical shifts:

1. Internet bandwidth–the speed at which data enters a home–now often surpasses the bandwidth needs within the home's internal network
2. Modern applications have evolved. They no longer demand raw bandwidth, but require consistent, low-latency data streams for optimal QoE.

Consequently, we find ourselves witnessing a paradigm shift, where the emphasis migrates from solely amplifying aggregate capacity to ensuring a consistent, application-specific delivery on each device, regardless of its spatial position and environmental conditions in the home.

4. The Role of WiFi in modern QoE

As technology continually advances, newer iterations of WiFi naturally bring improvements. However, is there a direct and significant correlation between WiFi evolutions and QoE? While WiFi 7 has its advantages, it's critical to recognize that WiFi, in essence, is merely a medium for transmission of data

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

between two points. WiFi specifications are centered around bandwidth and connectivity specifications, caring little about the individual applications that run on top.

For example, below is the typical WiFi technology benefit comparison chart:

WIFI TECHNOLOGY BENEFIT COMPARISON

	WiFi 4	WiFi 5	WiFi 6	WiFi 6E	WiFi 7
Year	2007	2013	2019	2020	2024
Protocol	802.11n	802.11ac	802.11ax	802.11ax	802.11be
Bands	2.4/5GHz	5 GHz only	2.4/5GHz	2.4/5/6*GHz	2.4/5/6 GHz
Peak Speed	600 Mb/s	6.9 Gb/s	6.9 Gb/s	6.9 Gb/s	46 Gb/s
Channel Widths	20/40 MHz	20/40/80/160 MHz	20/40/80/160 MHz	20/40/80/160 MHz	20/40/80/160 MHz/ 320 MHz
Security	WPA2	WPA2	WPA3	WPA3	WPA3
Key Features	4X4 MIMO LDPC Error Correction 64-QAM	8X8 MIMO 4X DL MU-MIMO Beam Forming[8] 256-QAM	8X8 MIMO 8X DL MU-MIMO Beam Forming OFDMA TWT 1024-QAM	8X8 MIMO 8X DL MU-MIMO Beam Forming OFDMA TWT *6 GHz added, USA ONly.	16X16 MIMO 16X DL MU-MIMO Multi-AP Multi-RU Puncturing Multi-Link Beam Forming OFDMA TWT 4096-QAM

The table[1] predominantly emphasizes parameters such as modulation rates, maximum bit rates, potential number of MIMO streams, and the spectrum utilized. While these elements are crucial for enhancing the bandwidth capacities between devices, they do not directly address the core KPIs of application QoE, which is central to the consumer experience. Furthermore, these specifications are largely theoretical; for instance, the prospect of households benefiting from a 16x16 MIMO and 46 Gbps throughput remains distant. Therefore, relying solely on WiFi technologies to address the growing application QoE challenge does not meet the comprehensive needs essential for delivering a Full Stack QoE experience.

5. Unpacking the Seven Dimensions of Full Stack Optimization QoE

To genuinely grasp and elevate Application Performance (i.e. QoE), it's imperative to understand and address its seven dimensions:

Space: Network performance shouldn't be hostage to a home's design or size. It should consistently deliver across various architectural layouts.

Time: It’s not just about speed, but also about unwavering performance irrespective of external temporal disruptions.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Load: As households contain an array of devices and applications, the network must dynamically adapt, recognizing and responding to myriad load scenarios.

Position: Device mobility within a home shouldn't compromise the QoE. Consistent, seamless application delivery is paramount.

Latency: In an era of real-time applications, latency becomes a critical KPI. Swift response times are non-negotiable.

Interoperability: A high QoE system must integrate seamlessly with multiple vendor hardware without compromising stability or performance.

Efficiency: Optimal utilization and allocation of resources underscore the efficacy of a QoE system.

Many current network systems predominantly addresses space and load requirements by integrating mesh extenders to enhance WiFi RF coverage. While effective in augmenting capacity in regions of the home distant from the primary gateway, this strategy falls short in delivering a fully optimized user experience. The performance of such systems remains largely determined and constrained by the type of WiFi technology deployed, and fails to capitalize on the capabilities of an advanced Full Stack Optimization framework.

To achieve a holistic system optimization, algorithmic adjustments are required to accommodate temporal variations induced by environmental factors, enable uninterrupted mobility across any position, and facilitate real-time application performance by reducing latency. Furthermore, Full Stack Optimization should be universally compatible across diverse Customer Premise Equipment (CPE) and System on Chip (SoC) vendors, incorporating a range of WiFi technologies and radio configurations.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

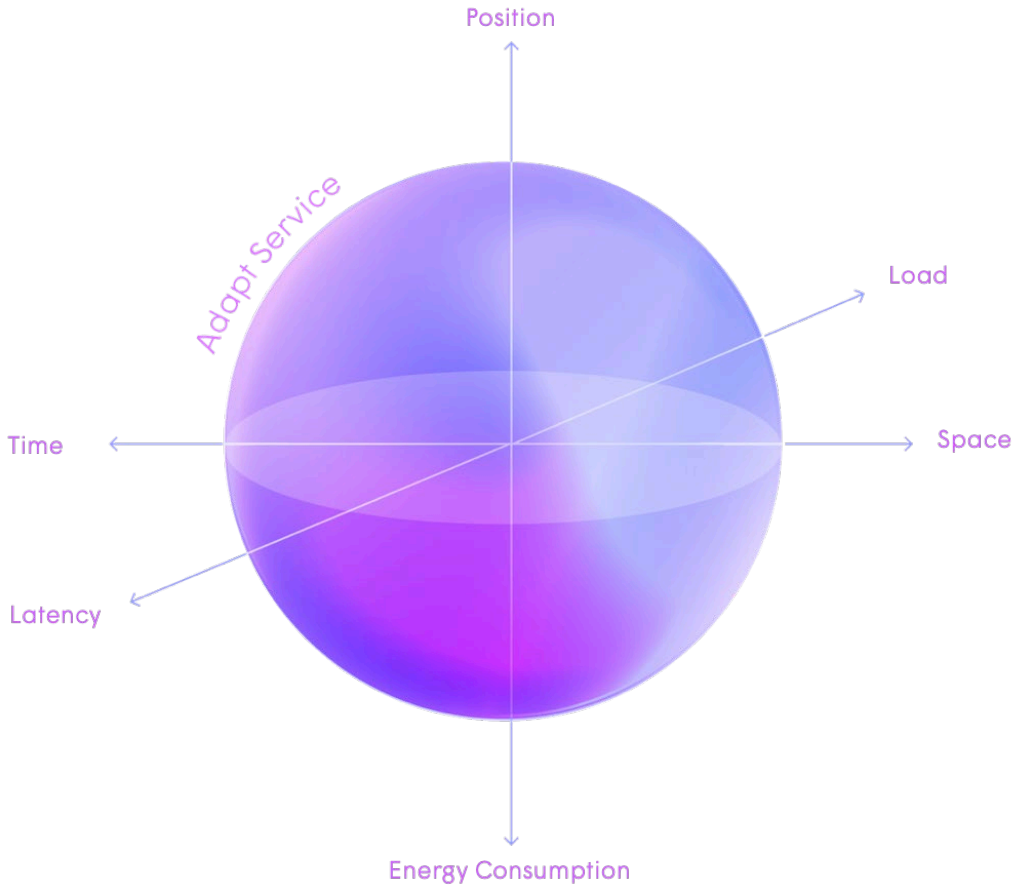


Figure 1 - Full Stack Optimization dimensions

6. Full Stack Optimization: A Revolutionary Approach

Full Stack Optimization embodies a comprehensive strategy for enhancing QoE, requiring:

1. Visibility across the stack: Merely relying on WiFi statistics is myopic. By employing Dynamic Deep Packet Inspection (DDPI), networks gain detailed insight into the devices, applications, and people who use them [2]. This granular understanding, when fused with WiFi metrics, empowers cloud-based algorithms to proactively adapt and optimize the network landscape.

Plume utilizes OpenSync in the CPE to perform DDPI functions to identify applications in real time on the network [3]. Packets from each IP traffic flow may be requested by any number of CPE user space applications for processing according to the desired functions, which may include cybersecurity protection, parental controls or application optimization functions. Applications are identified by inspecting anywhere from 1 to 12 packets in each application session. After identification, application sessions may be passed through without change, blocked or modified by manipulating the IP headers of the traffic flow according to a policy applied from the cloud. The traffic flow packet processing task is offloaded from the CPU by training the SoC hardware acceleration engine to take over the handling of the packets when all user space packet inspection demands are satisfied. After application identification is complete the application is monitored and scored for QoE delivery in the network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

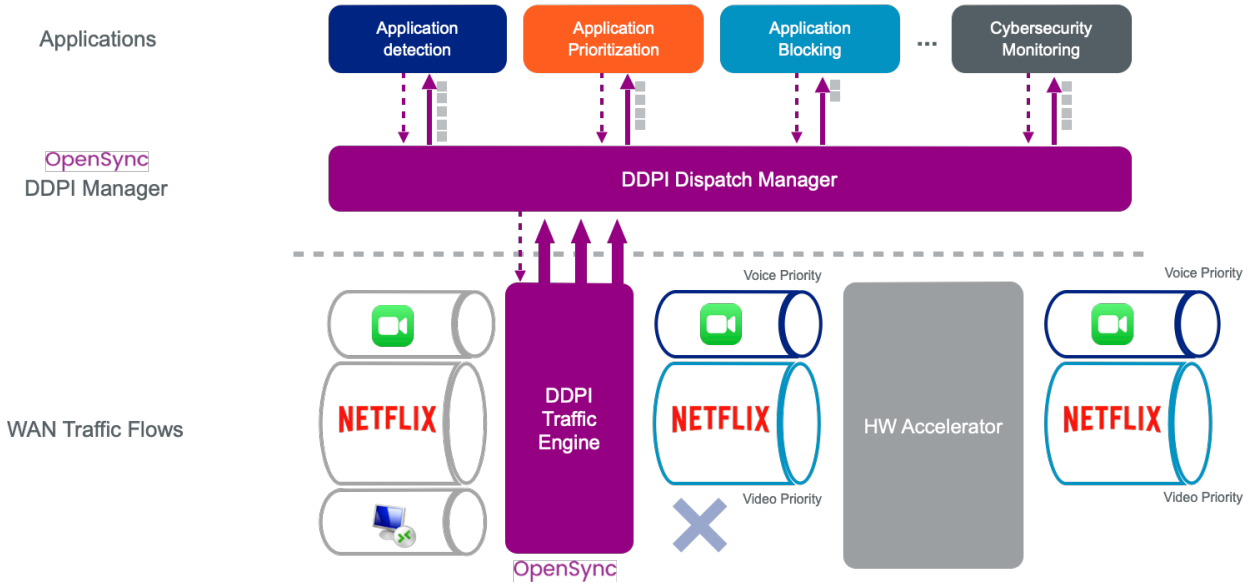


Figure 2 – Dynamic Deep Packet Inspection CPE architecture

To quantify the shift in the home to real-time applications we measured applications running in Plume managed households over a 30 day period in July and August 2023. Application information was detected and measured by using DDPI in OpenSync enabled CPEs for households with one or more APs and subscribed to Plume's HomePass service. The results show:

Table 1 – Application session statistics

Application Statistic	Value
Average number of application sessions	863 per HH/day
Average number of real-time applications sessions	64.7 per HH/day
Percentage of total applications that are real-time	78%

While the number of real-time application sessions ran in a location is only 7.5% of the total sessions, the real-time applications make up 78% of the applications in the home. To provide more clarity, we measured application usage on US based Plume customer homes over a 30-day period in July and August 2023 to better understand the importance of optimizing around real-time applications.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

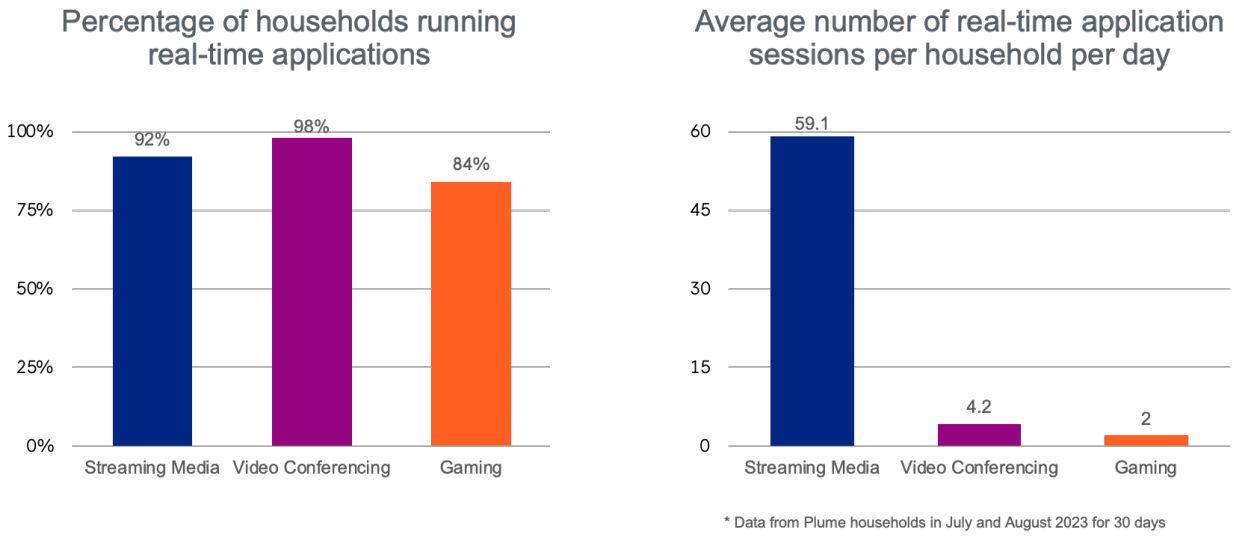


Figure 3 - Households running real-time applications

Additionally, the percentage of households running real-time applications is overwhelmingly high as shown above, with video conferencing being run at least once in 98% of the homes and gaming present in 84% of the homes over a 30-day period of monitored Plume subscribers.

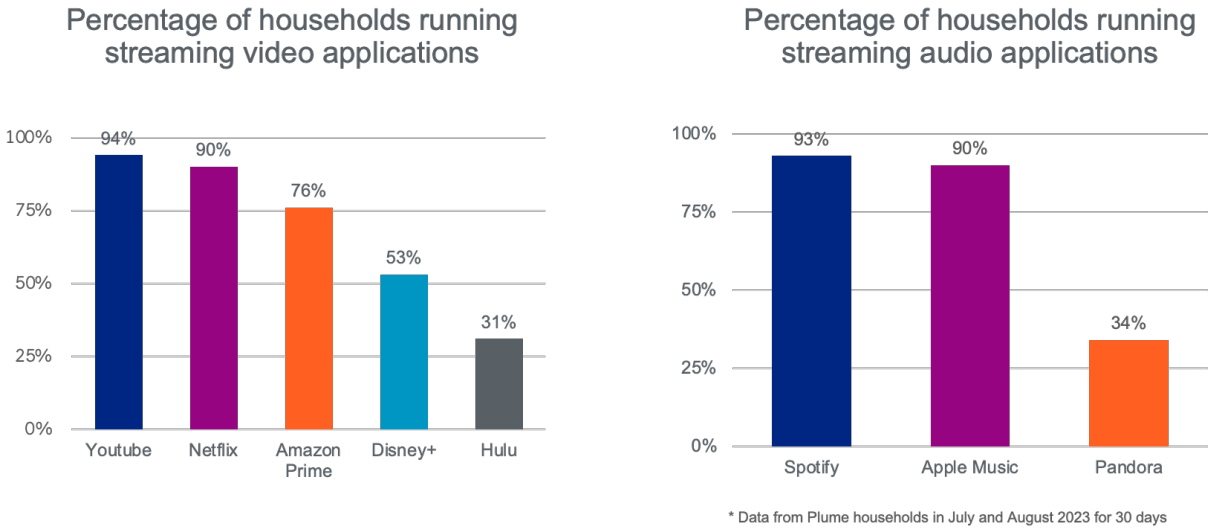


Figure 4 - Streaming application market share

Streaming Media services are dominated by YouTube, Netflix, Spotify, and Apple Music, each with at least one stream played in 90% of households or higher over the 30-day period.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

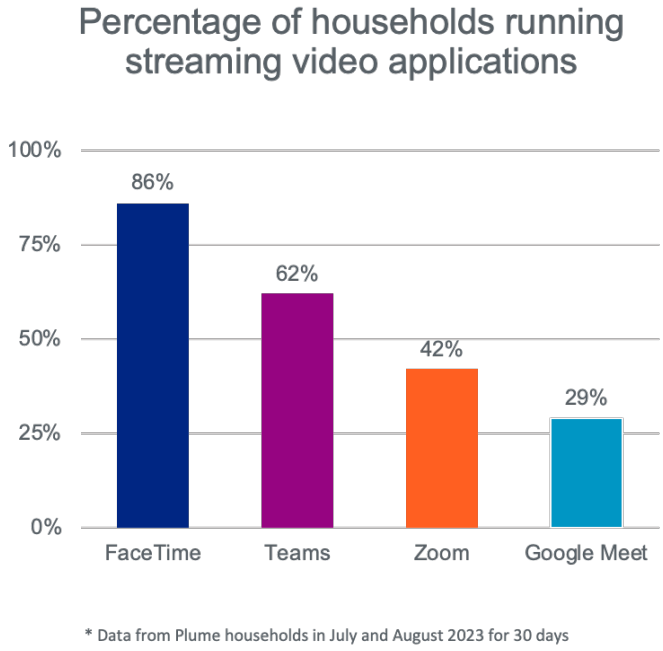


Figure 5 - Video conferencing application market share

Video Conferencing was dominated by personal devices using FaceTime, but followed by more business-related video conferencing services such as Microsoft Teams, Zoom, and Google Meet.

Based on the widespread and continued growth of real-time applications in households, high consumer QoE may benefit from giving these applications special consideration- particularly through the implementation of low-latency delivery. The efficacy of commonly utilized real-time applications, as gauged by their performance and ensuing QoE, is intrinsically tied to the latency and its consistency [4] [5].

2. Resource Manipulation and Allocation: To effectively accommodate a range of application needs, networks must extend their capabilities beyond mere WiFi utilization. In the current landscape, numerous applications demonstrate heightened sensitivity to latency, making it imperative for networks to offer pathways with minimized delay. Essential to achieving optimal application performance is the network's agility in identifying, classifying, and prioritizing applications according to their distinct requirements. Predictive algorithms, empowered by cloud-based AI and ML models, are instrumental in facilitating this dynamic adaptation.

The foundation of network optimization is a high-bandwidth and low-congestion WiFi network [6]. WiFi bandwidth allocation is computed for each access point, device, and expected applications running in the home based on previous network usage by the household. Learning models predict the needed WiFi capacity and compute the best use of spectrum, topology, and channel bandwidths to meet the application demand.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

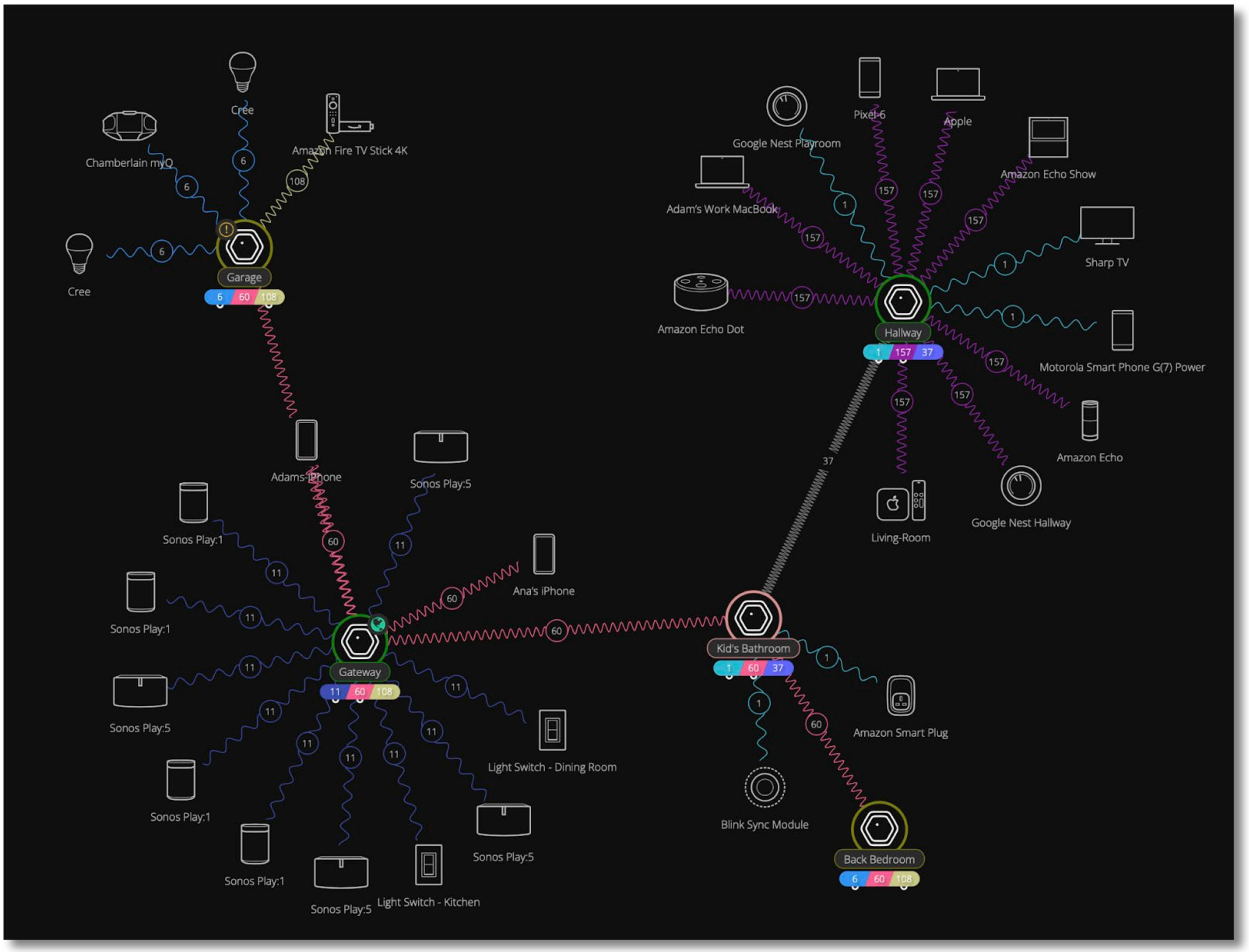


Figure 6 - Network topology with adaptive WiFi functionality

The figure above shows an example of a typical home network topology utilizing 7 different WiFi channels in 2.4GHz, 5GHz and 6GHz with WiFi 5, WiFi 6, and WiFi 6E technologies utilized. Use of different WiFi channels provides the highest amount of capacity with lower latency.

Apart from robust WiFi connectivity, latency optimization plays a pivotal role in delivering superior QoE for real-time applications. For instance, a latency exceeding 100 milliseconds from the end device to the internet server can significantly compromise the QoE in gaming, especially for cloud-based services. Likewise, real-time video conferencing experiences degradation with latency surpassing 200 milliseconds [4] [5].

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

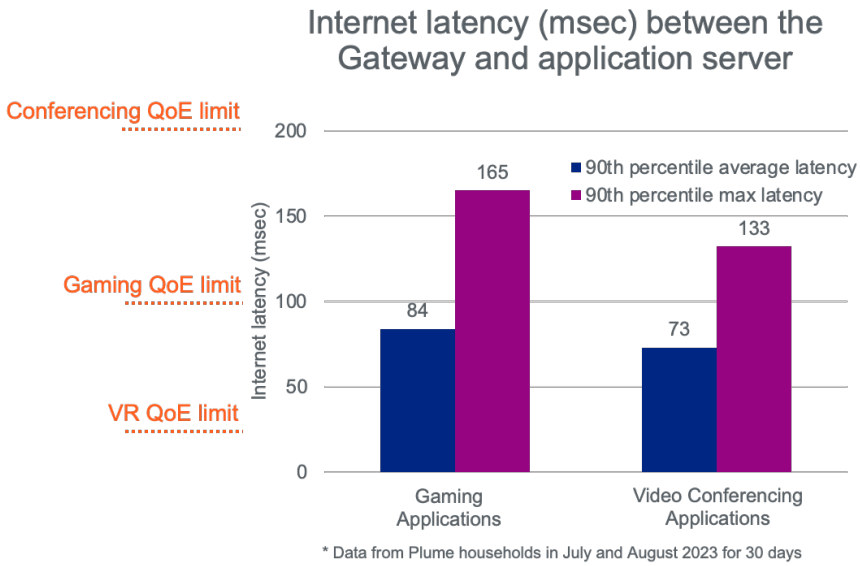


Figure 7 - Internet latency measurements for real-time applications

Measured latency between the household gateway and the internet server already measured maximum latency of 165 msec and 133 msec for gaming and video conferencing applications respectively, leaving little-to-negative latency margin in the home.

Utilizing packet prioritization for designated latency-sensitive applications can significantly mitigate latency issues during periods of temporary WiFi network congestion. Enhanced latency performance is achieved when packets from these identified applications receive precedence over those from non-latency-sensitive sources at the queuing points of each CPE interface. There are several queuing mechanisms available, such as WiFi WMM queues, OFDMA schedulers, and MLO with WiFi 7 [7]. In scenarios where the bandwidth on the WAN interface is restricted, as seen in certain DSL, DOCSIS, and Fixed Wireless Access (FWA) connections, a distinct queue can be set up for latency-sensitive traffic. This ensures that extensive buffered data does not delay the processing of latency-sensitive packets [8].

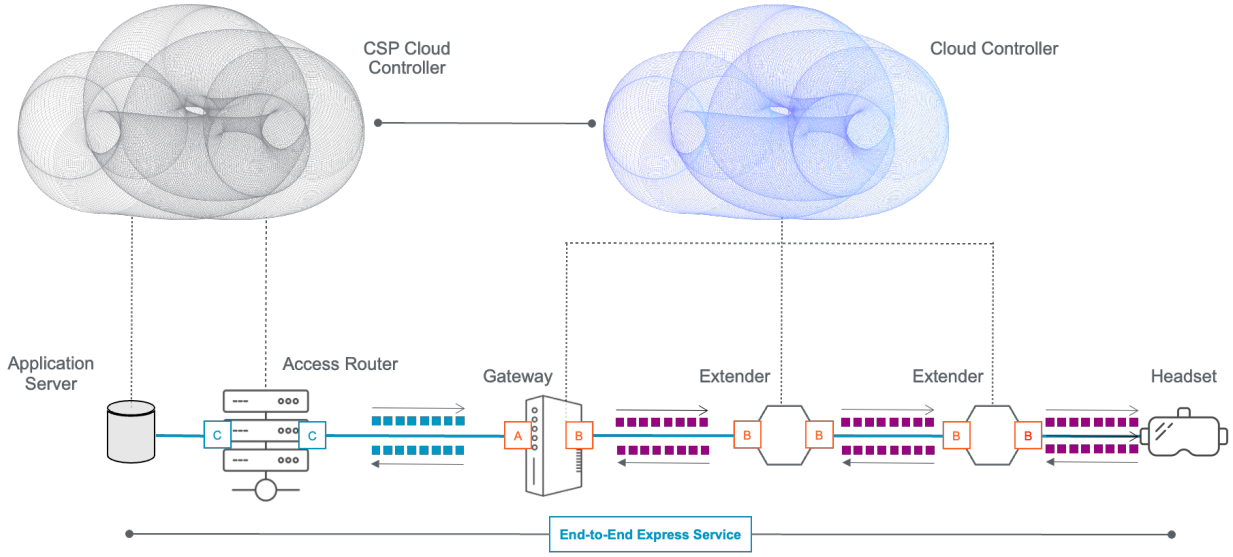


Figure 8 - Prioritized packet delivery for low latency applications

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

As demonstrated and measured in the figure below of packet latency versus different levels of prioritization, it is not adequate to simply throw more bandwidth at the problem to solve consumer expectations. When application session prioritization is implemented, the maximum in-home latency for prioritized application can be reduced by more than 60% in the presence of high congestion compared to best effort transmission. Full Stack Optimization must continuously provide active latency reduction to match the requirements of each individual application session for each consumer household.

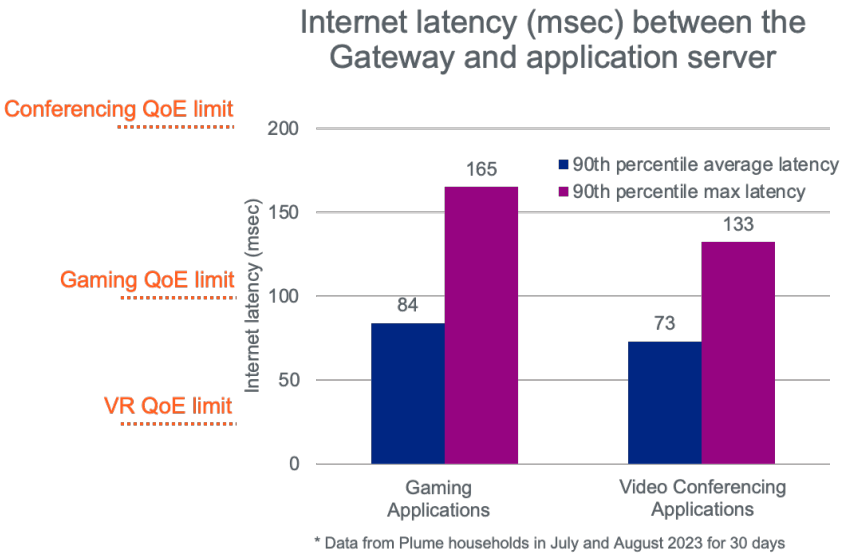


Figure 9 - Latency improvement for different application types in congested WiFi network

3. Monitoring and Assessment: Beyond generic metrics like speed tests, it's essential to delve deeper. Real-time monitoring of every application's performance, encompassing parameters like latency, throughput, and consistency, is crucial. This comprehensive monitoring translates into actionable insights, shaping continual refinement strategies.

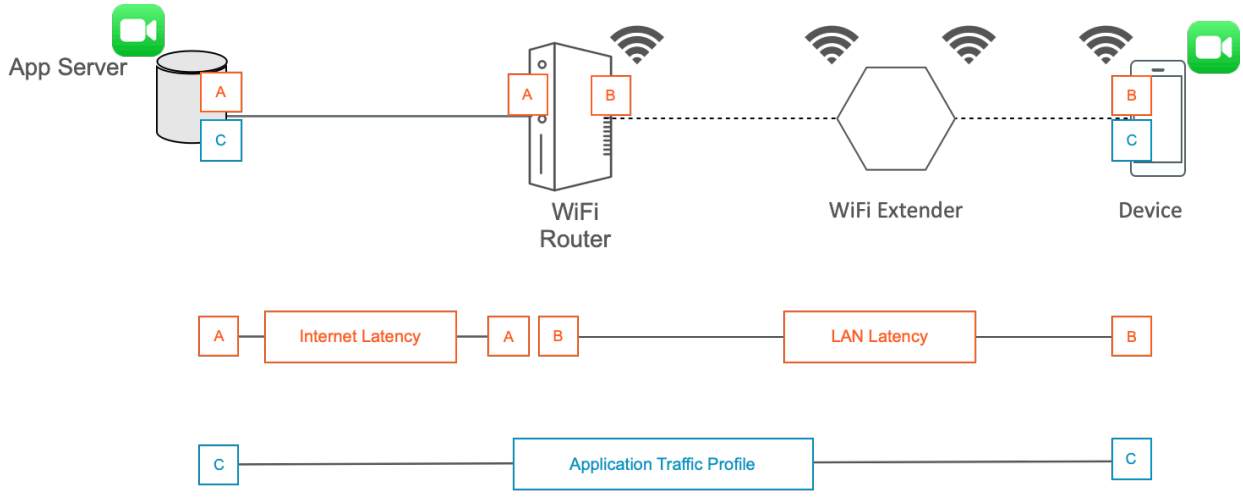


Figure 10 - Measurements for Application Performance

In the assessment of Application Performance, an inclusive evaluation is required, encompassing both latency and throughput measurements. Latency metrics are measured from both the Wide Area Network

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

(WAN) segment of the internet and the Local Area Network (LAN) within premises. This dual-segment assessment is pivotal in pinpointing the origins of excessive latency that may impede optimal Application Performance. Application throughput is determined from the internet application server directly to the consumer device, facilitated by real-time surveillance of the packets traverse the WiFi router.

The culmination of these metrics - throughput, latency, and their consistency - produces an Application Performance score tailored to the distinct necessities of the specific application. For example, Voice Over Internet Protocol (VOIP) calls are predominantly weighted on latency and the uniformity in packet delivery, notwithstanding their lower throughput. In contrast, high-definition video streaming places a substantial emphasis on the unwavering delivery of high throughput and can tolerate a considerable degree of latency. Finally, cloud gaming applications require a combination of both ultra-low latency control packet feeds, and low latency high bandwidth video streams simultaneously in different directions. Application Performance scores may be used by CSP for call center troubleshooting, subscriber self-help, and network level KPIs to measure the delivery of application performance across different subscriber types and equipment configurations.

7. Conclusion

The burgeoning prevalence of real-time applications across households emphasizes the urgency for a measurement standard that's both elevated and consistent. The days of singular strategies, like bandwidth enhancement or WiFi upgrades, are waning. Engineers and network architects must embrace a multi-dimensional approach. The seven-dimensional model, integrated within the Full Stack Optimization feature set, charts a promising path toward a future where each application's performance is seamless, user-centric, and consistent.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

QoE	Quality of Experience
DDPI	Dynamic Deep Packet Inspection
LAN	Local Area Network
CSPs	Communication Service Providers
MHz	Megahertz
GHz	Gigahertz
Mb/s	Megabit per second
MIMO	Multiple-Input Multiple-Output
AP	Access Point
AI	Artificial Intelligence
ML	Machine Learning
FWA	Fixed Wireless Access
WAN	Wireless Area Network
OFDMA	Orthogonal Frequency-Division Multiplexing
MLO	Multi-Link Operation
WMM	Wireless Mobility Management
VOIP	Voice Over Internet Protocol

Bibliography & References

[1] https://www.truecable.com/blogs/cable-academy/wifi-7-vs-wifi-6-whats-the-difference

[2] Tomasz Bujlow, Valentín Carela-Español, Pere Barlet-Ros, Independent Comparison of Popular DPI Tools for Traffic Classification, http://tomasz.bujlow.com/publications/2014_journal_elsevier_comnet_independent_comparison.pdf

[3] OpenSync 3.0 Overview, EUB-020-013-00 OpenSync 30 Overview.pdf, Flow Service Manager – FSM 53-55

[4] Normoyle, A., Guerrero, G., & Jörg, S., Player Perception of Delays and Jitter in Character Responsiveness, proceedings of the ACM Symposium on Applied Perception (SAP 2014), 117-124

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

[5] Eric Baccei, Alexander Clopper, Taylan Sel, An Evaluation of Zoom and Microsoft Teams Video Conferencing Software with Network Packet Loss and Latency, Worcester Polytechnic Institute

[6] Adam Hotchkiss, Bill McFarland, The Plume Advantage: Putting the Adapt in Adaptive WiFi, https://content.plume.com/v/the-plume-advantage-putting-the-adapt-in-adaptive-wifi

[7] Monica Paolini, Senza Fili, Wi-Fi 7: The next generation in the evolution of Wi-Fi, https://6ghz.info/wp-content/uploads/2022/07/SenzaFili-IntelTB-Wi-Fi7-The-next-generation-in-the-evolution-of-Wi-Fi.pdf

[8] Amir Mirzaeinnia, Mehdi Mirzaeinia, and Abdelmounaam Rezgui, Latency and Throughput Optimization in Modern Networks: A Comprehensive Survey, September 2020

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Smart Building Internet of Things Deployment Using Low Power Wide Area Networks

Long Range (LoRa) Wireless RF Technology

An Operational Practice prepared for SCTE by

Victor Lopez
Network Engineer, Emerging Technology
Charter Communications
6399 S Fiddlers Green Cir, Greenwood Village, CO 80111
+1(913) 295-7460
Victor.H.Lopez@charter.com

Mohamed Daoud
Director, Emerging Technology
Charter Communications
6399 S Fiddlers Green Cir, Greenwood Village, CO 80111
+1 (720) 699-5077
Mohamed.Daoud@charter.com

Hossam Hmimy
Sr. Director, Emerging Technology
Charter Communications
6399 S Fiddlers Green Circle, Greenwood Village, CO 80111
+1 (720) 536-9396
Hossam.Hmimy@charter.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. LoRa and LoRaWAN Technical Overview 4
3. Methodology and Equipment Setup 9
   - 3.1. Experimental Results at Location 1 10
   - 3.2. Propagation Model Calibration at Location 1 13
4. Results 14
   - 4.1. Design / Planning 14
     - 4.1.1. RF Design for Location 2 14
     - 4.1.2. Networking Design 16
   - 4.2. Pre-commissioning on LNS and App Server integration 17
   - 4.3. Walk Test at Location 2 18
   - 4.4. Application Server's Dashboards 21
5. Conclusions and Best Practices 22

Abbreviations 24

Bibliography & References 25

List of Figures

Title **Page Number**

Figure 1 – LoRa and LoRaWAN Technology Stack [2] 4
Figure 2 – Data Rate Formula For LoRa Transmission [4] 5
Figure 3 – LoRa Architecture [2] 5
Figure 4 – Charter's LoRaWAN Architecture (High Level) 6
Figure 5 – Charter's LoRaWAN Architecture & (High Level) 6
Figure 6 – Comparison of Wireless Technologies [5] 7
Figure 7 – LoRaWAN Channels for North America [2] 8
Figure 8 – Deployment Process 9
Figure 9 – Location 1 Setup and RF Specifications 10
Figure 10 – Location 1 Walk Test Results Signal Strength RSSI 11
Figure 11 – Location 1 Summary Results Signal Strength RSSI 11
Figure 12 – Location 1 Walk Test Results Signal Quality SNR 12
Figure 13 – Location 1 Summary Results Signal Quality SNR 12
Figure 14 – Before Calibration of RSSI (Theoretical ≠ Experimental) 13
Figure 15 – After Calibration of RSSI (Theoretical ≈ Experimental) 14
Figure 16 – Location 2 Setup and Specifications 14
Figure 17 – Location 2 Design 1st Floor 15
Figure 18 – Location 2 Design 2nd Floor 15
Figure 19 – Location 2 Design 3rd Floor 16
Figure 20 – Location 2 Design 4th Floor 16
Figure 21 – Networking Design 17
Figure 22 – LoRaWAN Network Server 18
Figure 23 – LNS and App Server Integration 18
Figure 24 – Location 2 Walk Test Results 1st Floor 19

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 25 – Location 2 Walk Test Results 2nd Floor 19
Figure 26 – Location 2 Walk Test Results 3rd Floor 20
Figure 27 – Location 2 Walk Test Results 4th Floor 20
Figure 28 – Application Server Main Menu View 21
Figure 29 – Application Server Sensors Placement on 2nd Floor 21
Figure 30 – Application Server All Sensors Data 22
Figure 31 – Application Server Sensors Data Room C226 22
Figure 32 – Gateways Placement Strategy 23

List of Tables

Title **Page Number**

Table 1 – US 902-928 MHz Frequencies Plan [6] 8
Table 2 – LoRaWAN Regulation for North America [6] 8
Table 3 – Calibration of the Model Propagation 13
Table 4 – Action Points 23

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Charter Communications designed and deployed a Low Power Wide Area Network (LPWAN) and smart building solution at one of its offices located in North Carolina (location 2), with the goal of enhancing employee comfort and improving operational efficiencies. The solution includes environmental monitoring and room occupancy sensors across four floors. In this paper, we describe the methodology used to deploy the end-to-end solution, therefore, we outline the goals and various components of a smart building solution. Second, we describe the radio frequency (RF) design and dimensioning for network deployment. Third, we show the parameters of the propagation model which were utilized to calibrate the design tool. Fourth, we explain the method of optimization and monitoring the live solution. The entire process is shown in Figure 1. Last, our experimental results also describe the network performance testing (coverage and quality) and best practices to deploy in a high-rise building based on Charter's experience. As Charter evaluates emerging technologies to expand our wireless network portfolio, LoRa technology has been selected to enable many use cases for Internet of Things (IoT) solutions.

2. LoRa and LoRaWAN Technical Overview

One of the prominent wireless technologies of LPWAN for IoT is Long Range, commonly known as LoRa. Specifically, LoRa is purely the radio signal that carries the data into the physical (PHY) layer which is based on a Chirp Spread Spectrum (CSS) modulation technique that was developed by the vendor Semtech [1]. LoRa chip sets are built into LoRa gateways and sensors. Medium Access Control (MAC) is the following layer, commonly known as LoRaWAN, which is managed by LoRa Alliance. LoRaWAN is the communication protocol that controls how data is communicated across the network for the upper layer [2].

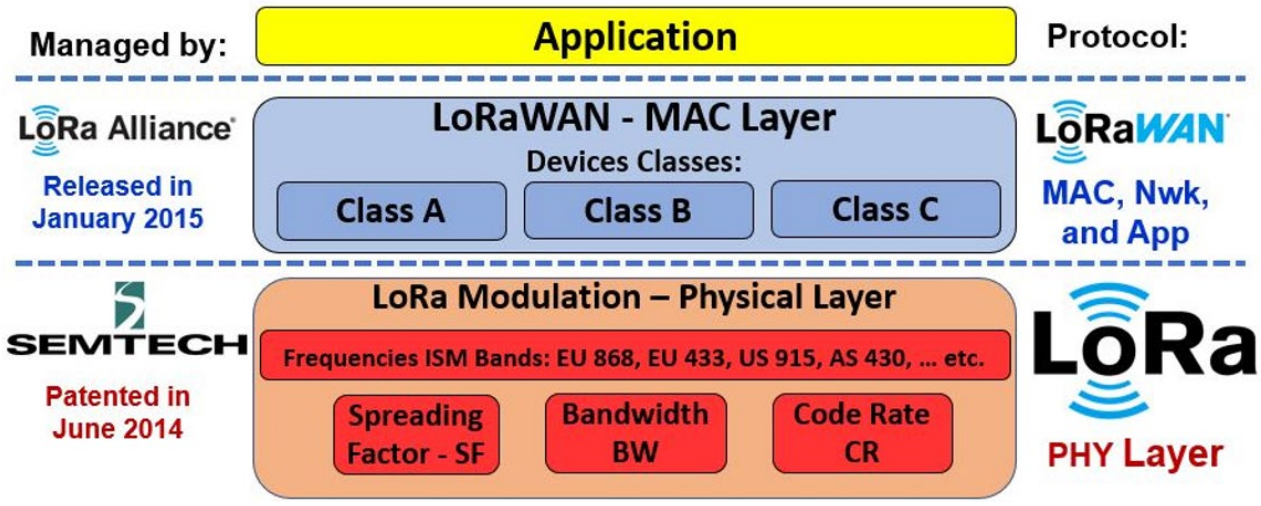


Figure 1 – LoRa and LoRaWAN Technology Stack [2]

In addition, LoRaWAN defines multiple Data Rates (DR) which are a combination of Spread Factor (SF), channel Bandwidth (BW), and Coding Rate (CR) [3]. These parameters together are used to compute the DR also known as LoRa Modulation Bit Rate (Rb) and is given in Figure 2.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Data Rate Formula

$$R_b = SF * \frac{\left[\frac{4}{4+CR}\right]}{\left[\frac{2^{SF}}{BW}\right]} * 1000$$

SF = Spreading Factor (6,7,8,9,10,11,12)
CR = Code Rate (1,2,3,4)
BW = Bandwidth in KHz
(10.4,15.6,20.8,31.25,41.7,62.5,125,250,500)
Rb = Data rate or Bit Rate in bps

Figure 2 – Data Rate Formula For LoRa Transmission [4]

Figure 3 details a typical LoRaWAN architecture, which includes LoRa sensors or end devices, LoRa Gateways (GW), and a network server called LoRaWAN Network Server (LNS). Sensors are connected to one or more gateways via LoRa CSS modulation technique which encodes data on radio waves using chirp (also known as sweep signal which defines the tone in which the frequency changes in a period of time) [2]. It can be of two types, up-chirp (connoting an increase in frequency) and down-chirp (frequency decrease). Gateways are connected to the LNS through standard IP connectivity. The LNS routes data information from the sensor to the associated Application Server (App Server), provides authentication for sensors, and manages security for the entire network using keys such as the DevEUI, the Network Session Key (NwkSKey), and the Application Session Key (AppSKey). LoRaWAN provides end-to-end security services through the use of Advanced Encryption Standard (AES) cryptography.

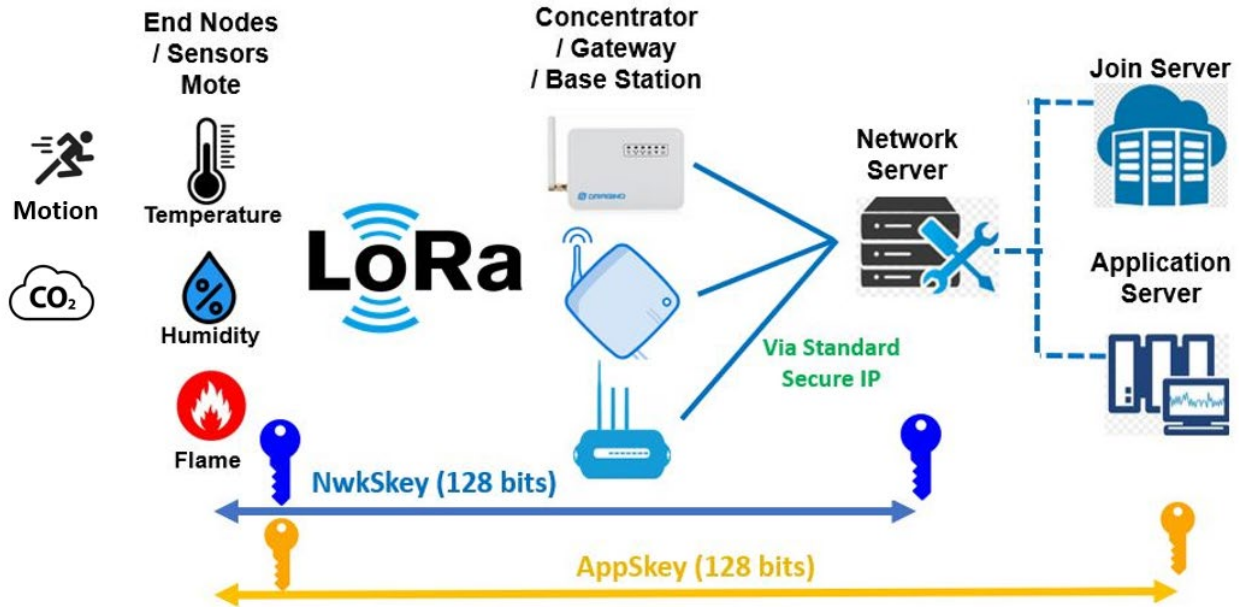


Figure 3 – LoRa Architecture [2]

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In the United States, the LoRa technology works on the unlicensed Industrial, Scientific, and Medical (ISM) bands from 902 to 928 MHz, therefore, there is no Federal Communications Commission (FCC) licensing required. The LoRa gateway's backhaul can be Ethernet, cellular, or both, depending on the GW model. The red outlined sections shown in Figure 4 are part of Charter LoRa Network as a Service (NaaS).

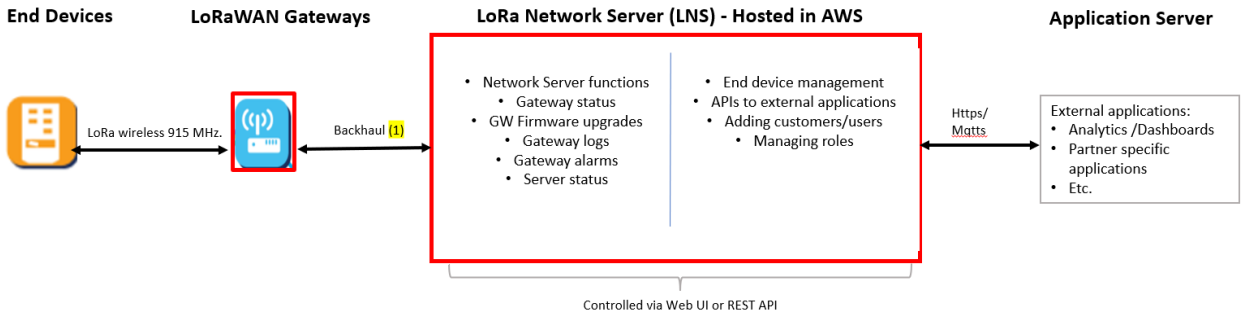


Figure 4 – Charter's LoRaWAN Architecture (High Level)

In this deployment, Browan and Tektelic brand gateways have been used along with multiple LoRa sensors (e.g., temperature, humidity, CO_2, and motion). In addition, a Multitech brand Field Test Device (FTD) has also been used to measure the strength and quality of the received signal – or, in technical terms, the Received Signal Strength Indicator (RSSI) and Signal-to-Noise Ratio (SNR). The sensors' information is pushed from the LNS to the App Server using an HTTP connection. The LNS gathers all the sensors' raw data and transforms it (by decoders sent to the App Server) into actionable insight for optimizing workflow process, improving the efficiency of room occupancy, and reducing operating costs. Therefore, all the data analytics are managed by the App Server, allowing for user-friendly data visualizations to be displayed through dashboards (Figure 5).

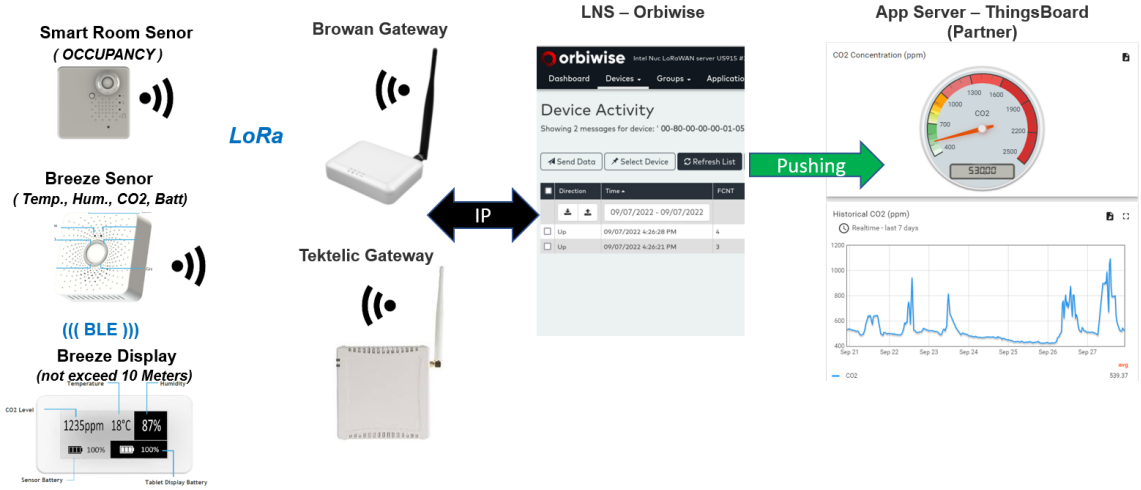


Figure 5 – Charter's LoRaWAN Architecture & (High Level)

The LoRa's transmission based on the spread spectrum technique is carried out by low-power transmitters to send small amounts of data to receivers placed at long distances (Figure 6).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

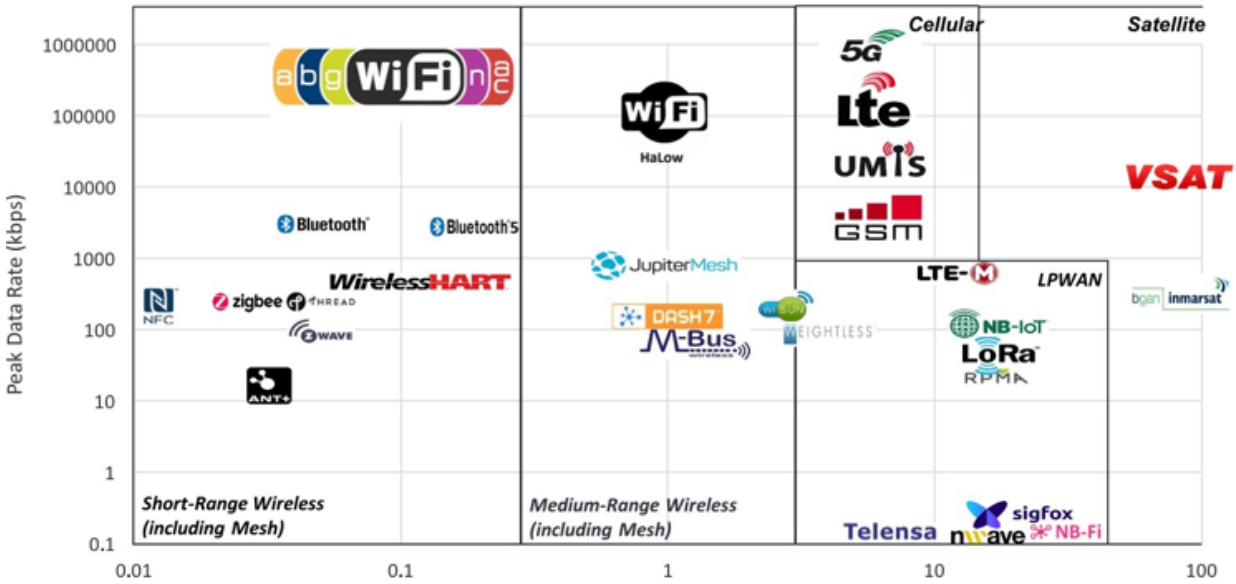


Figure 6 – Comparison of Wireless Technologies [5]

LoRa's key features are detailed below.

- **LoRa is specially designed for:**
  - ✓ Multiple low data rates ranging from 0.3 kbit/s to 50 kbit/s per channel.
  - ✓ Long-Range wireless communication.
    - ➢ Three miles in urban areas.
    - ➢ 10 miles or more in rural areas (depending on the line of sight).

- **Strengths:**
  - ✓ The best link budget of any other standardized wireless communication technology.
  - ✓ Operates under unlicensed frequency ISM bands 915MHz - FCC Part 15.
  - ✓ Deep penetration inside homes/buildings – low freq.
  - ✓ Low-power consumption / long battery life (~ 10 years).
  - ✓ Cost efficiency (infrastructure implementation).
  - ✓ High capacity - multi-tenant interoperability.
  - ✓ Scalability - easy to add sensors and gateways.
  - ✓ Security - embedded end-to-end AES-128 bits encryption.
  - ✓ Allows FUOTA (Firmware Updates Over-The-Air).
  - ✓ Resilience - multiple spreading factors result in interference protection.

- **Weaknesses:**
  - ✓ Not good for high data-rate transmissions.

The LoRaWAN Channels for North America follow the entire FCC Part 15 regulations for the 902 - 928 MHz ISM band, which is better known as the 915 MHz Band. This frequency plan is detailed in Figure 7, Table 1 and Table 2.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

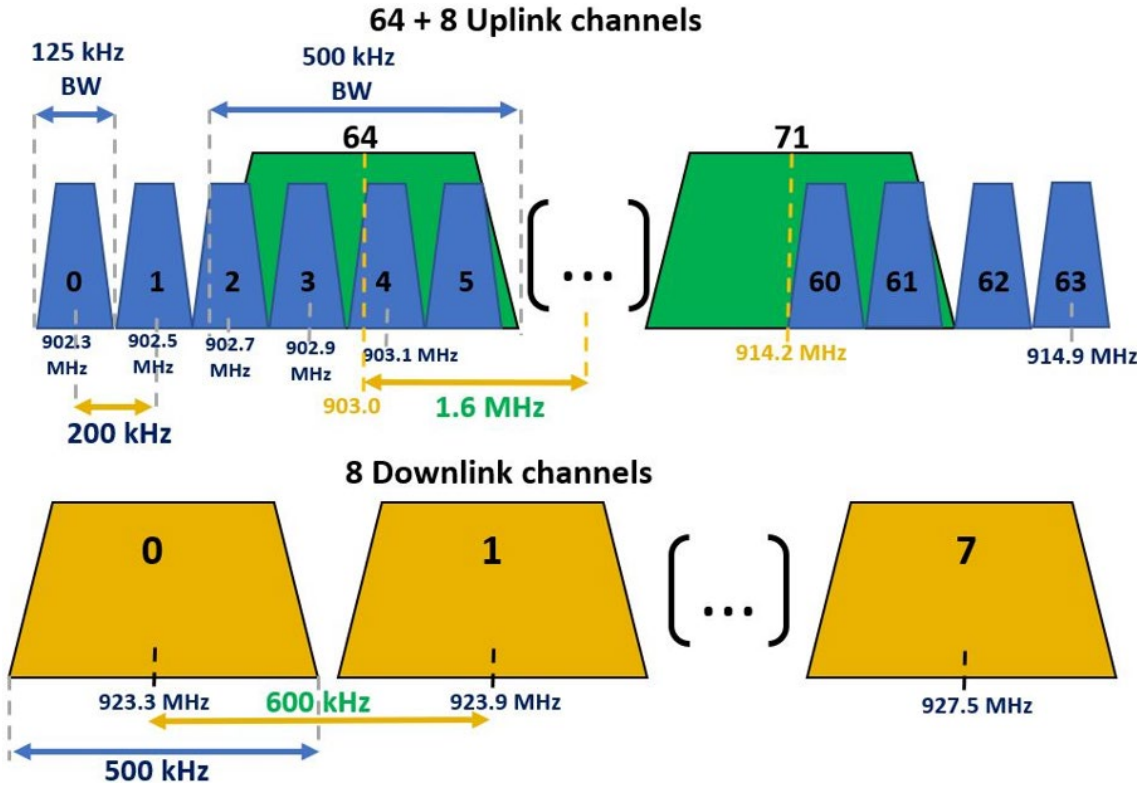


Figure 7 – LoRaWAN Channels for North America [2]

Table 1 – US 902-928 MHz Frequencies Plan [6]

Description	Upstream – 64	Upstream – 8	Downstream – 8
Channels numbered	0 to 63	64 to 71	0 to 7
Number of channels	64	8	8
Frequency starting at	902.3 MHz	903.0 MHz	923.3 MHz
Linearly increment	200 kHz	1.6 MHz	600 kHz
Frequency ending at	914.9 MHz	914.2 MHz	927.5 MHz
Bandwidth	125 kHz	500 kHz	500 kHz
SF varying	SF7 - SF10	SF8	SF7 - SF12
Coding rate	4/5	4/5 - 4/8	4/5 - 4/8

Table 2 – LoRaWAN Regulation for North America [6]

Description	LoRaWAN specification for North America
Frequency Band	902 - 928 MHz
Max. Tx Power Uplink	(30 dBm allowed) 20 dBm is typical
Max. Tx Power Downlink	27dBm
Max. dwell time	400 milliseconds on Up-Links

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Methodology and Equipment Setup

This section details the methodology and equipment setup for the entire LoRa solution deployment process. First, the LoRA network was deployed on the 3rd floor of the Charter Colorado office lab (location 1) in which RSSI and SNR were measured on each floor through the field test device (FTD). Second, using the experimental results of RSSI and SNR, the propagation model was calibrated on the design tool to adjust the theoretical results. Third, once the propagation model was calibrated, the Radio Frequency (RF) and networking design was carried out for the Charter premises in North Carolina (location 2). Fourth, gateways and sensors had been pre-commissioned in our laboratory, as well as the HTTP connection between the LNS and the App Server. In addition, all the dashboards, triggers and alarms need to be tested prior to the actual installation. Fifth, when all the components are working properly, the next step is to mount all gateways and sensors in the planned place. Finally, we validate all the gateways and sensors are up, running, and sending data. Moreover, all the action points were tested. This process is illustrated on Figure 8.

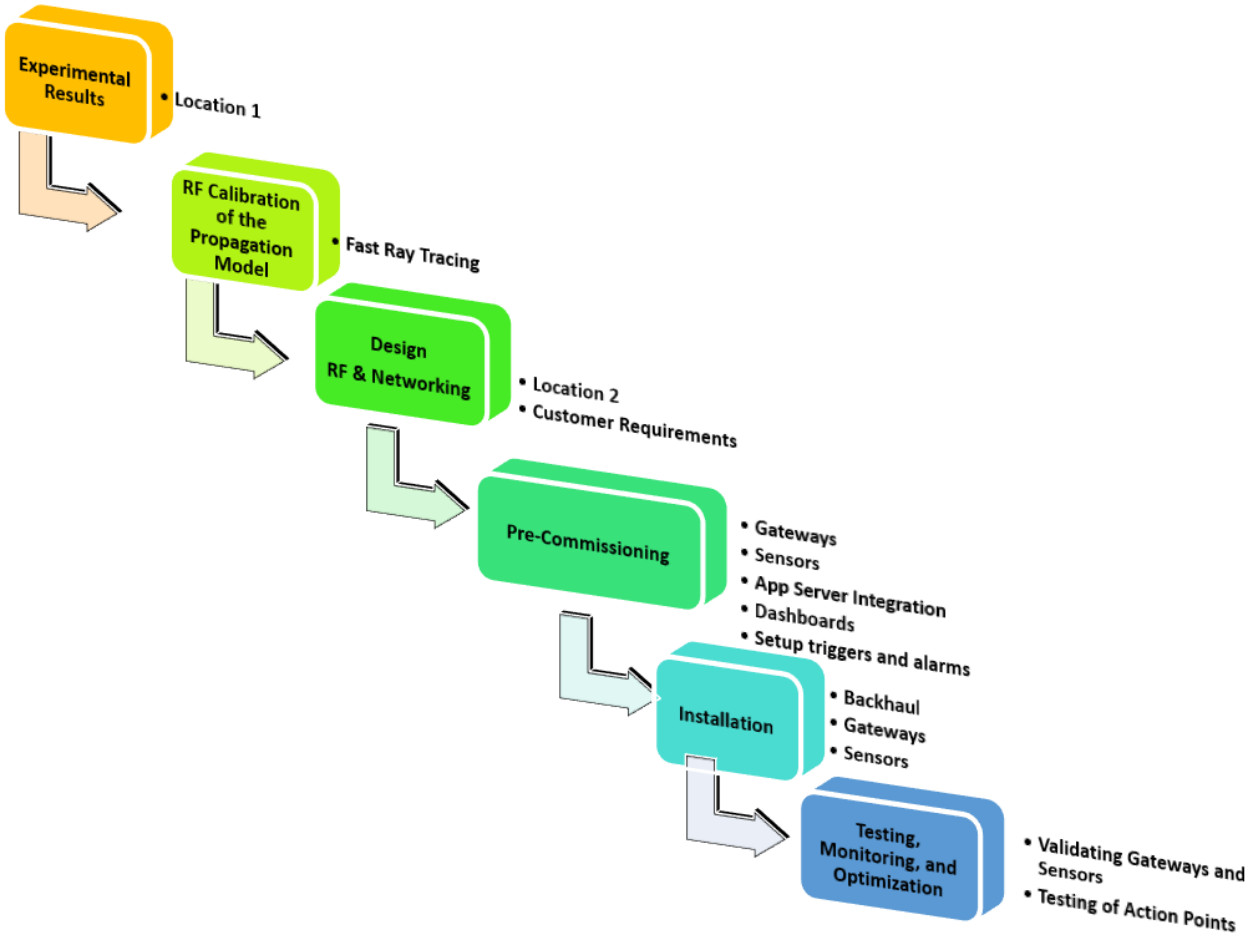


Figure 8 – Deployment Process

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.1. Experimental Results at Location 1

Figure 9 shows the five points of interest where the FTD was used to capture the RSSI and SNR values on each floor at the location 1. In the same way, one gateway was located on the corner of the building on the third floor (green section in the southeast). The legend shows the gateway setup used on the design tool.

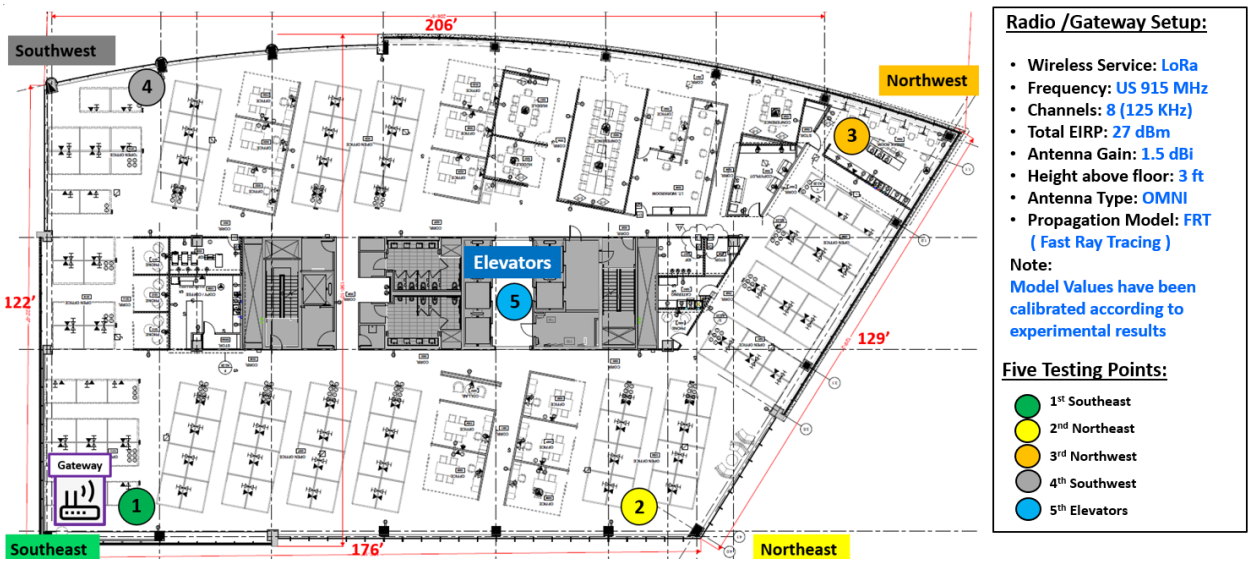


Figure 9 – Location 1 Setup and RF Specifications

Figures 10-13 show the result obtained using the FTD to measure the RSSI and SNR on each floor. The gateway located on the third floor was able to cover the entire floor, as well as one floor below and above, with optimal levels of RSSI and SNR, covering 25,000 sq. ft.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

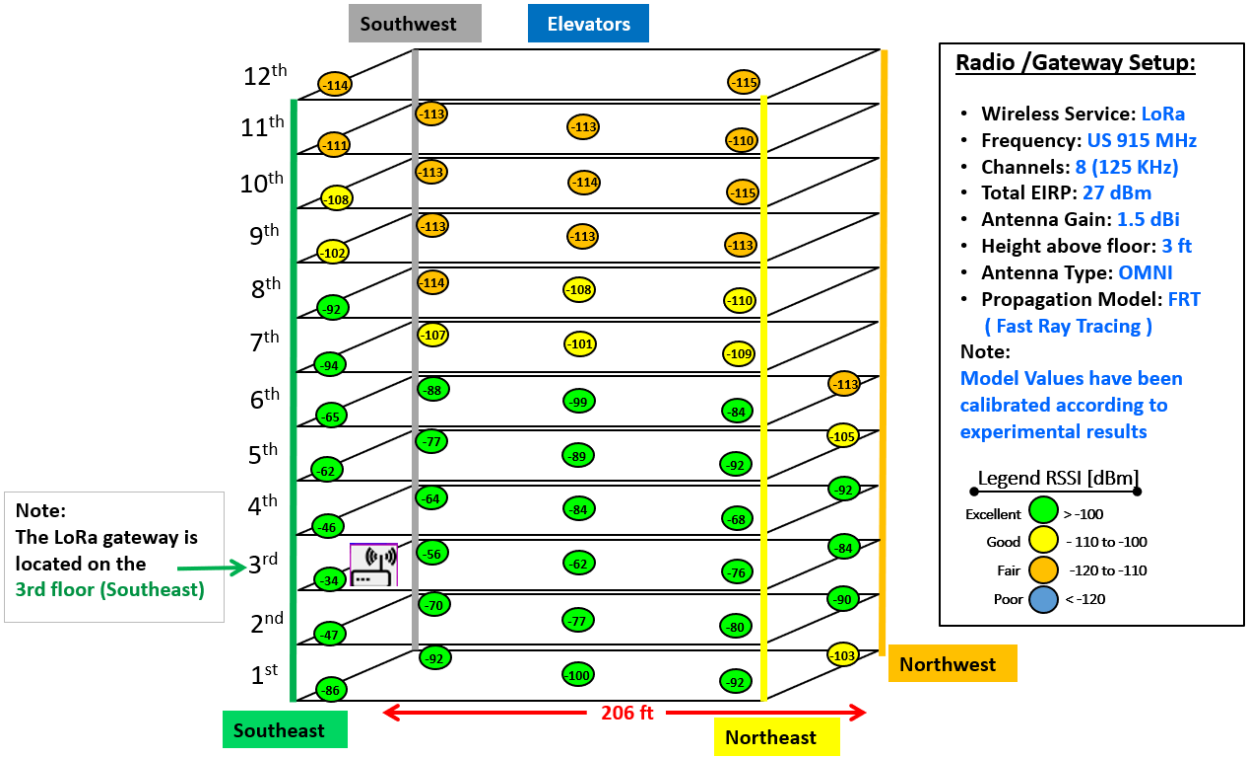


Figure 10 – Location 1 Walk Test Results Signal Strength RSSI

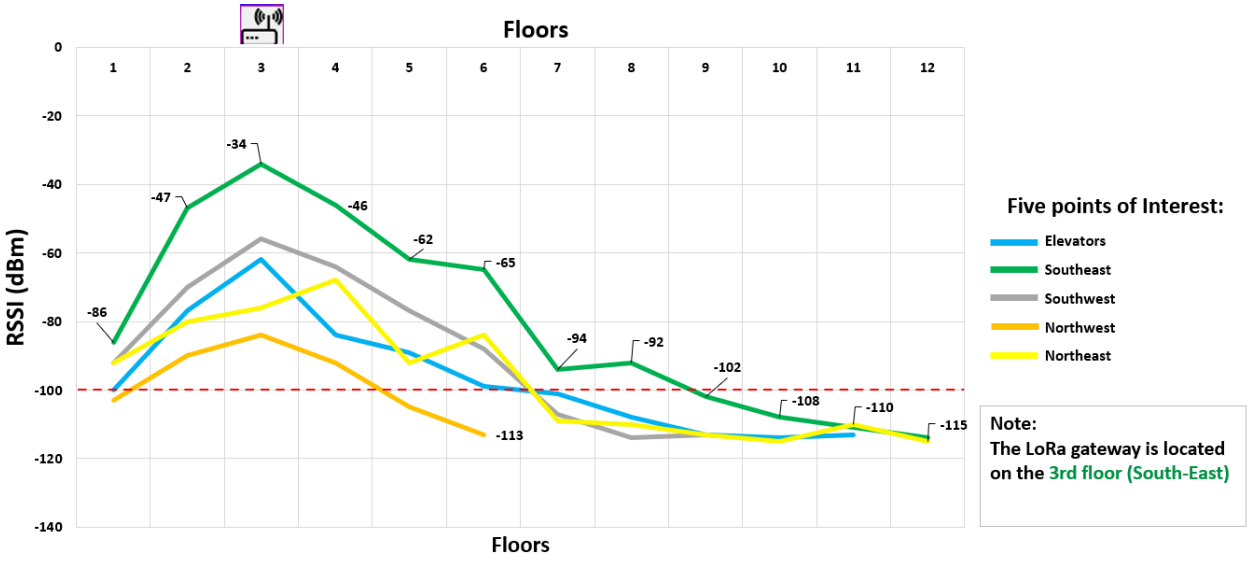


Figure 11 – Location 1 Summary Results Signal Strength RSSI

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

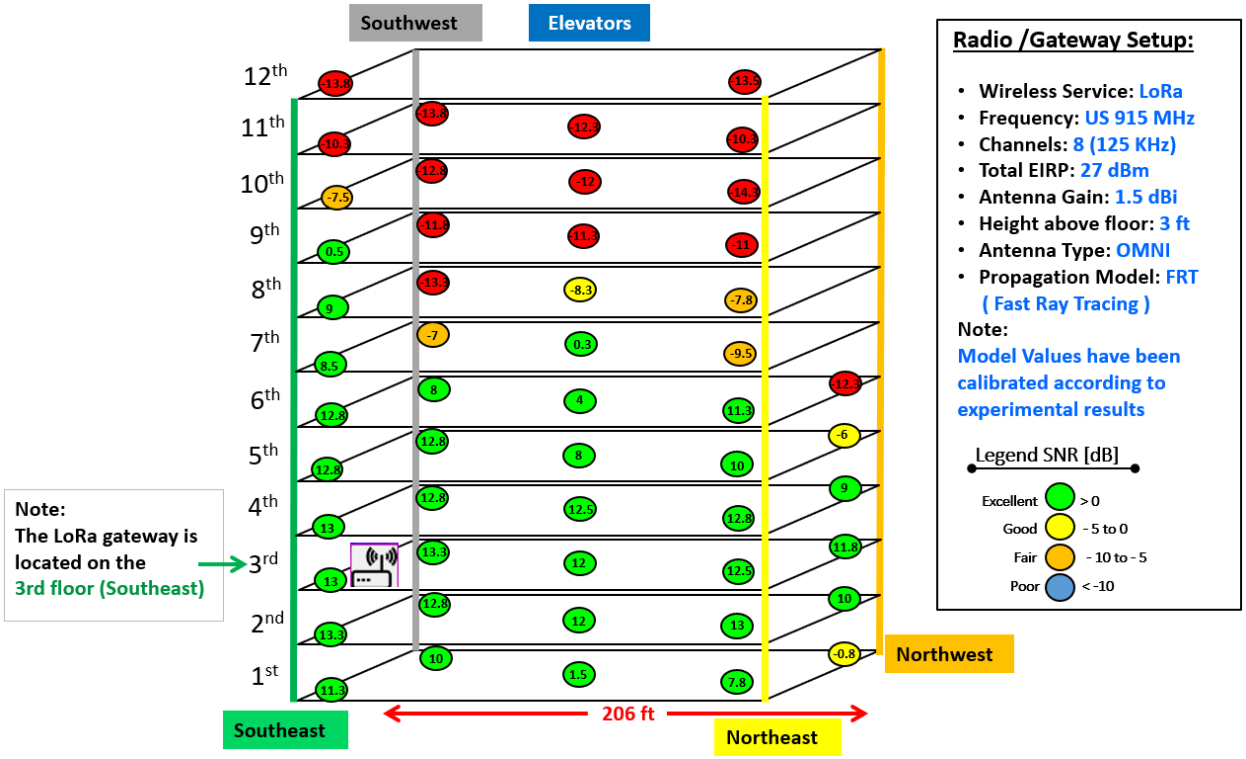


Figure 12 – Location 1 Walk Test Results Signal Quality SNR

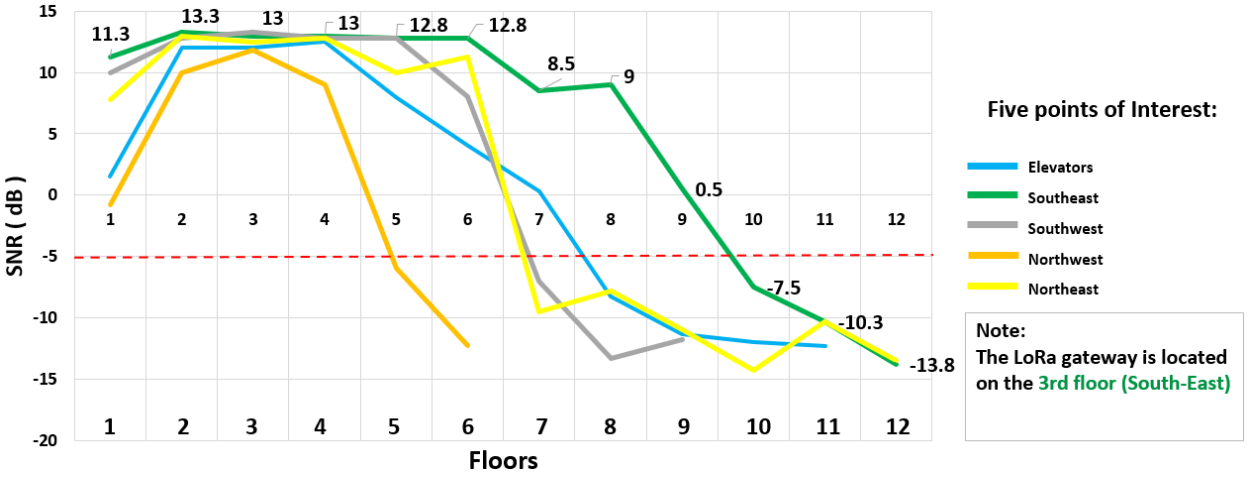


Figure 13 – Location 1 Summary Results Signal Quality SNR

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.2. Propagation Model Calibration at Location 1

All these data surveys were utilized to calibrate the propagation model on the design tool. The path loss parameters and the physical properties of the materials have been calibrated (transmission and reflection losses), as well as model coefficients of the Fast Ray Tracing (FRT). Table 3 details the transmission loss values for different materials and the FRT model coefficients which have been calibrated on the 917 MHz band for LoRa technology. Using the calibrated model, one example of signal strength RSSI is shown on Figure 15.

Table 3 – Calibration of the Model Propagation

Parameters of the Materials (Band 900MHz)		**Model Coefficients of FRT (Band 915MHz LoRa)**
Type of Material	**Transmission Loss [dB]**	Wall loss factor 1 (p1): 4
Drywall	1.75	Wall loss factor 2 (p2): 2
Glass Indoor	2	DLOS (y1): 4.70
Glass Outdoor	7	RP (y2): 3.77
Drywall individual workstation	0.15	DNLOS (y3): 4.27
Concrete Heavy	27	Body loss (Δy1): 4
Elevators (Metal)	75	

- DLOS (y1): Direct line of sight path only.
- RP (y2): Reflected path only.
- DNLOS(y3): Direct non-line of sight path.

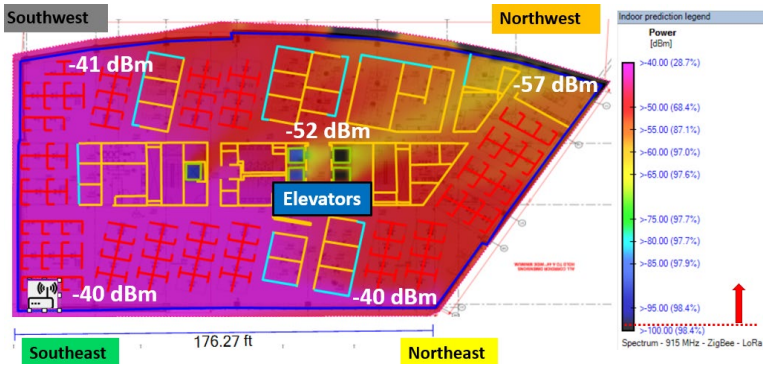


Figure 14 – Before Calibration of RSSI (Theoretical ≠ Experimental)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

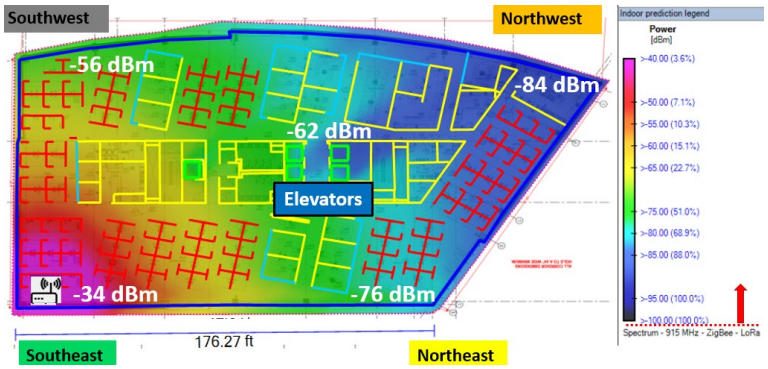


Figure 15 – After Calibration of RSSI (Theoretical ≈ Experimental)

4. Results

Determined by the building size, shape and design, the placement of gateways were strategically located at the location 2 premises to cover the entire building. Four gateways were needed to cover the entire building, and their locations will warrant appropriate signal strength and quality levels.

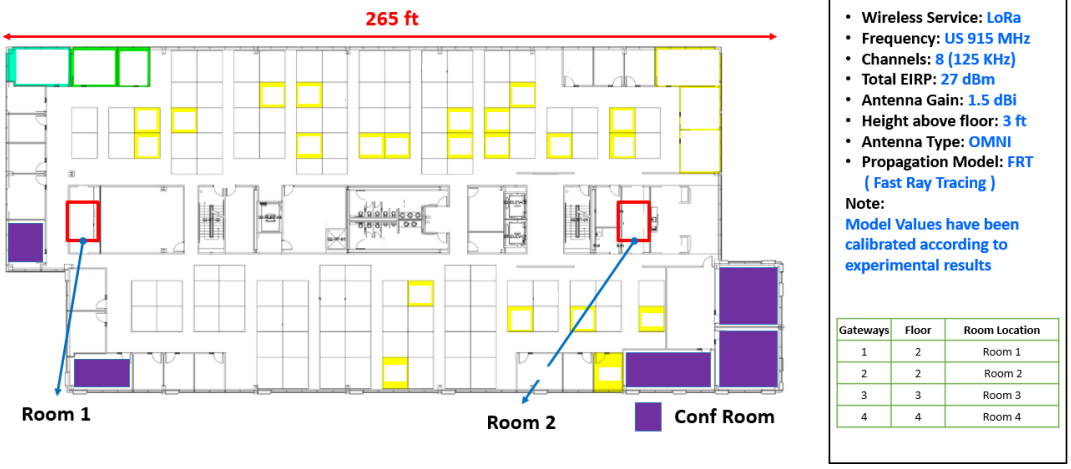


Figure 16 – Location 2 Setup and Specifications

4.1. Design / Planning

4.1.1. RF Design for Location 2

Once the propagation model was finally calibrated, the heatmaps for RSSI, SNR and best server were made for each floor. The results are shown on Figures 17-20. The criterion for the design is focused on RSSI > 97 dBms and SNR > 2 dB, based on the experimental results.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

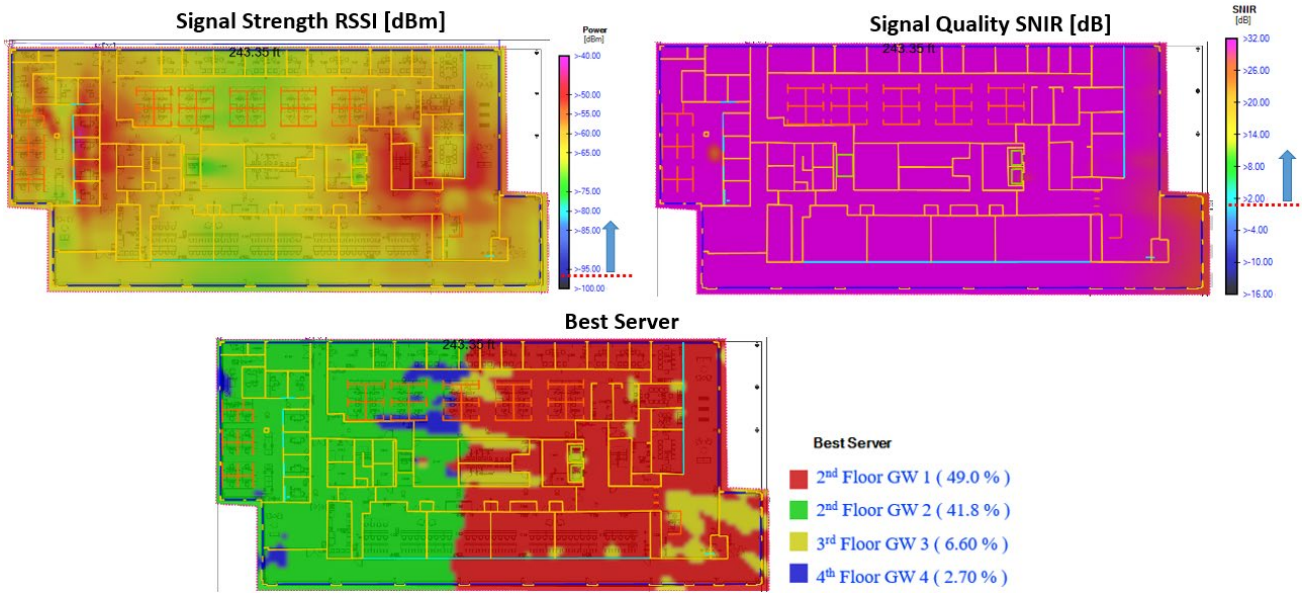


Figure 17 – Location 2 Design 1st Floor

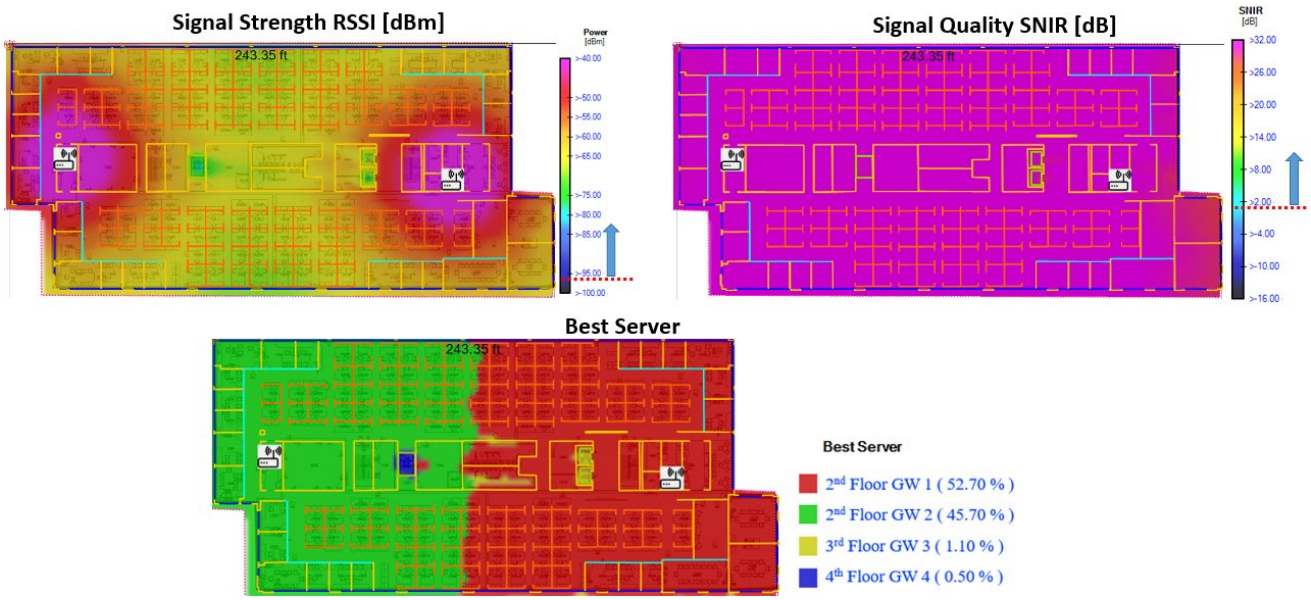


Figure 18 – Location 2 Design 2nd Floor

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

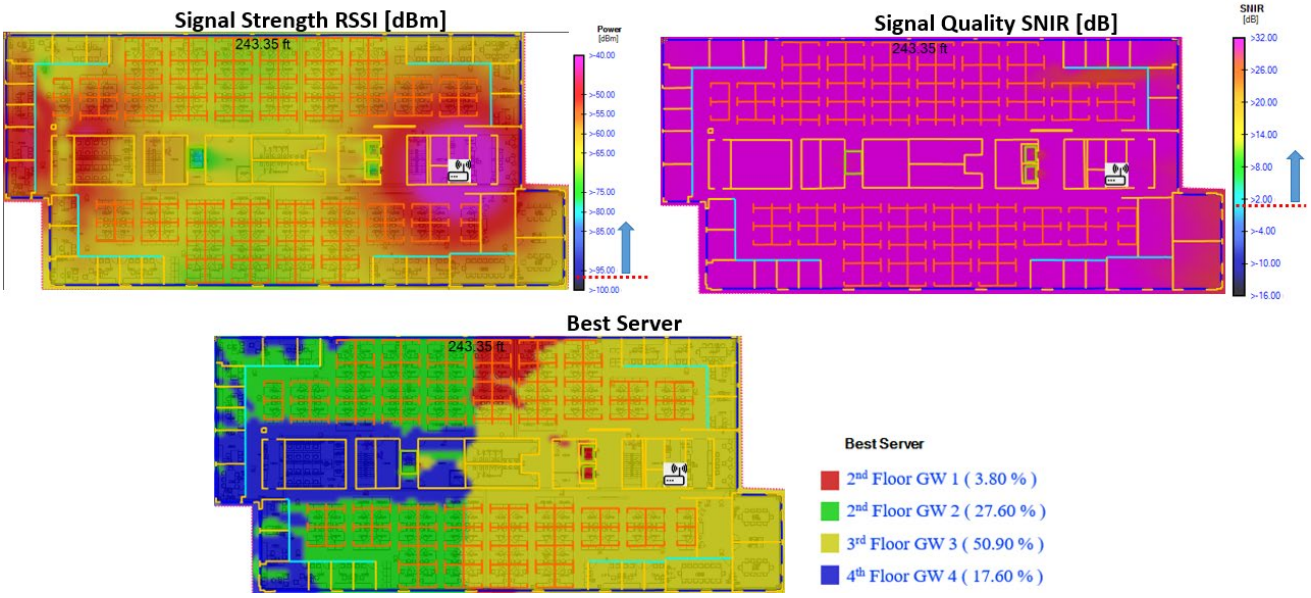


Figure 19 – Location 2 Design 3rd Floor

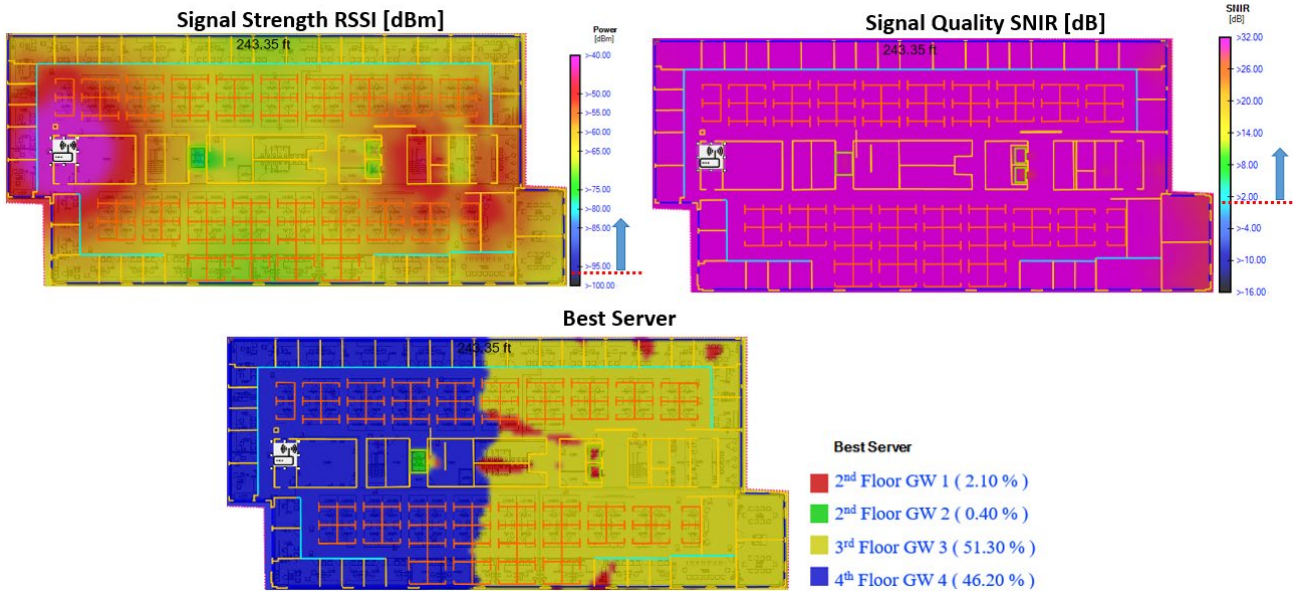


Figure 20 – Location 2 Design 4th Floor

4.1.2. Networking Design

The networking design is based on how many gateways were needed and their respective locations. Therefore, this design required three PoE (Power over Ethernet) switches and one router (Figure 21).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

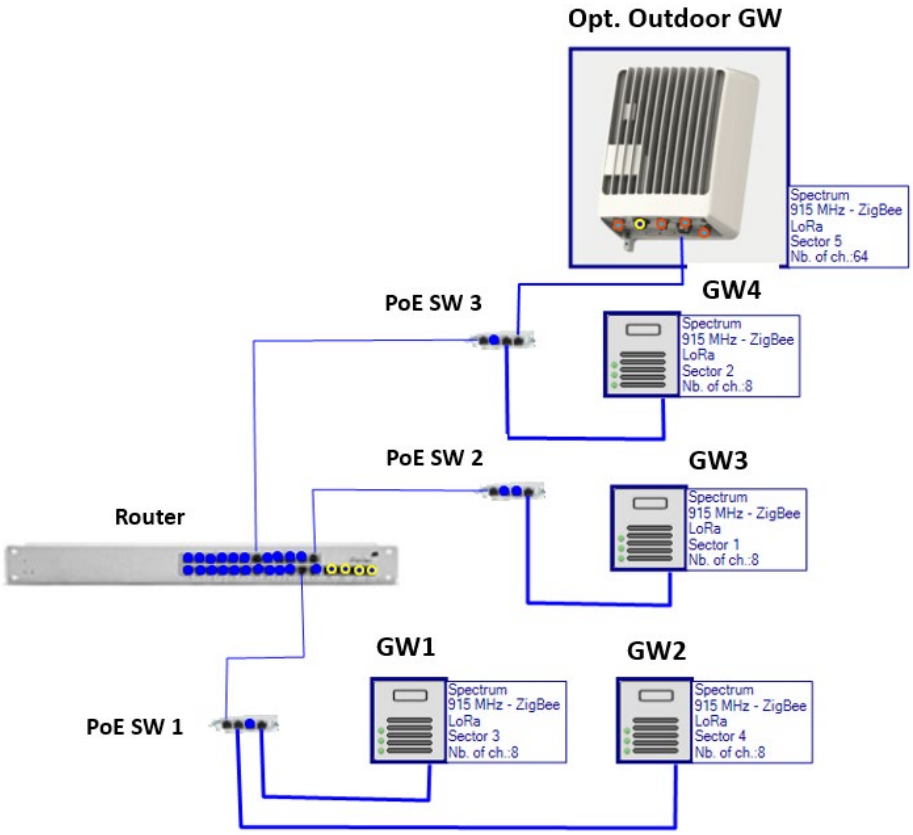


Figure 21 – Networking Design

4.2. Pre-commissioning on LNS and App Server integration

All the gateways and sensors were pre-commissioned, which means the gateways are up and running on the Network Supervision Tool (NST) and the sensors are sending data on the Data Access Sub System (DASS) platform through the GWs (Figure 22). The sensors' information is decoded on the LNS and then sent to the App Server for user-friendly data visualization via dashboards (Figure 23). Finally, all the components (sensors, GW, LNS and App Server) were working together and the next step was to mount all equipment in its planned location at the location 2 premises.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

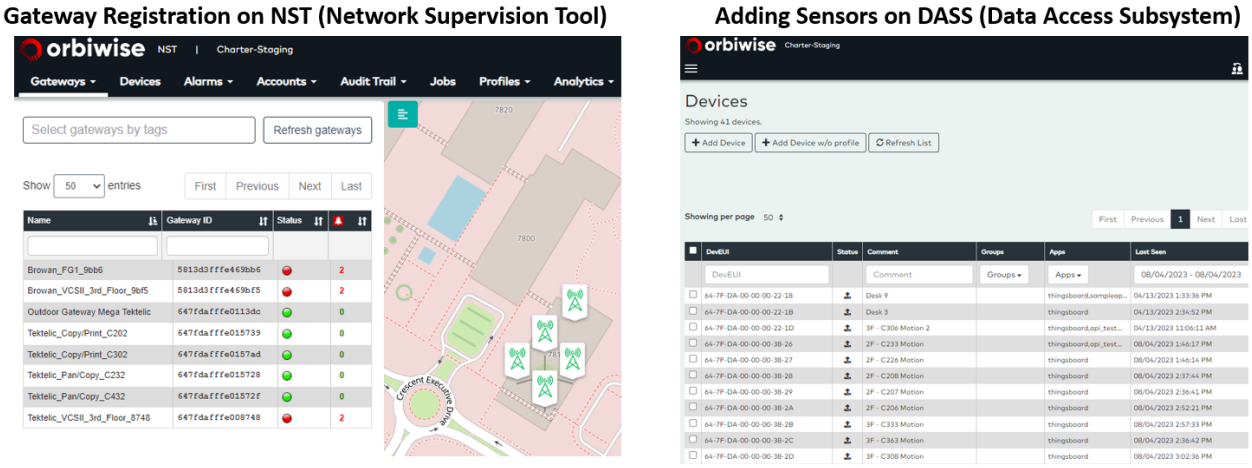


Figure 22 – LoRaWAN Network Server

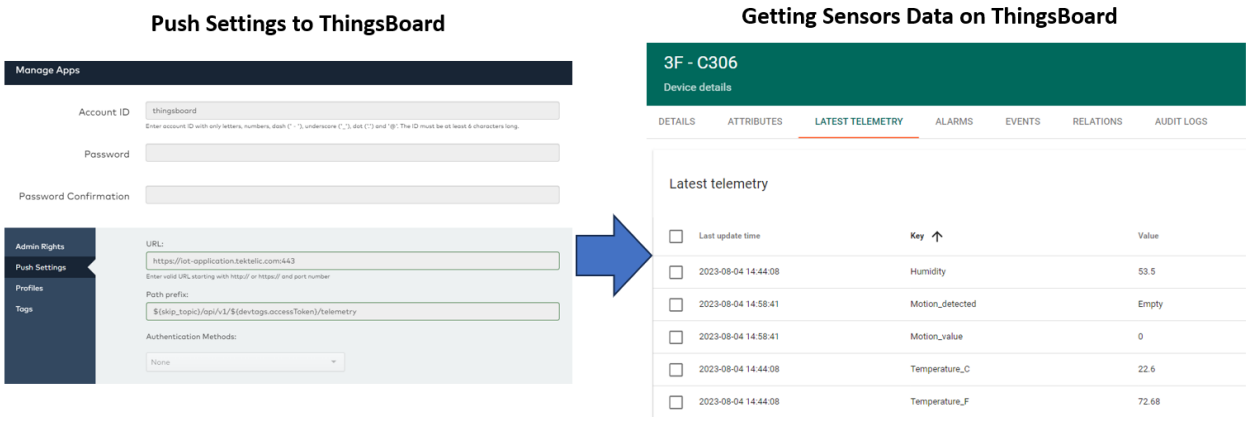


Figure 23 – LNS and App Server Integration

4.3. Walk Test at Location 2

Once all the components are in place and operating properly, the FTD was used to characterize the GWs' signal on each floor at the location 2 premises. Figures 24 - 27 show the actual results of signal strength RSSI (measurement of how strong the sensor can hear a signal from a GW), signal quality SNR and best server. On each floor, several measurements were obtained, and the results show that all the samples are above -100 dBms of RSSI and above 0dB of SNR, which means the entire building is covered with appropriate LoRa signal levels for the suitable sensor functionality.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

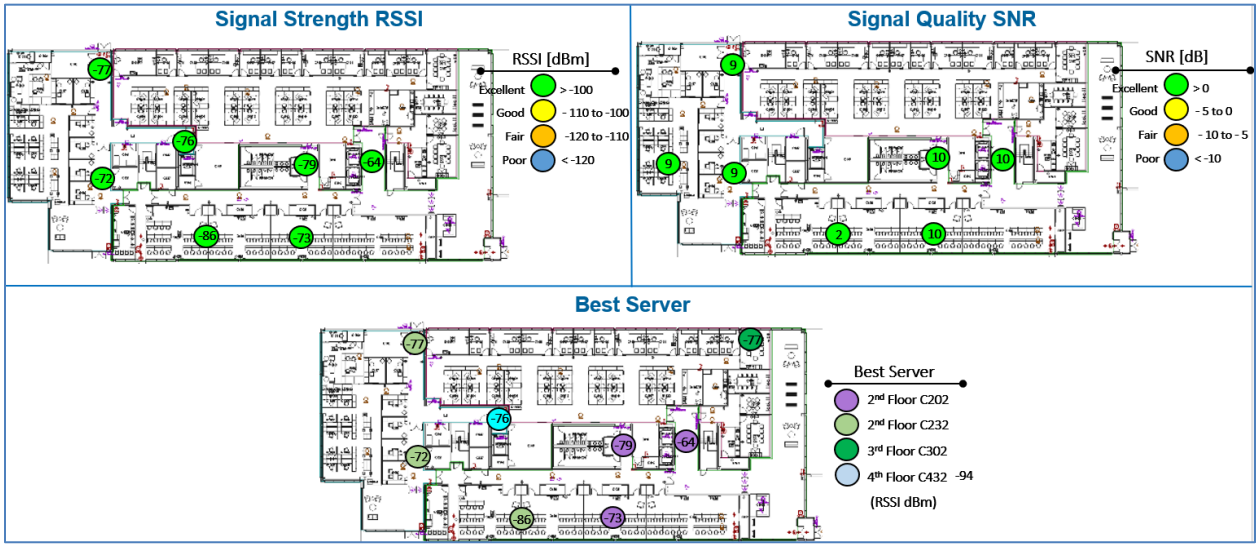


Figure 24 – Location 2 Walk Test Results 1st Floor

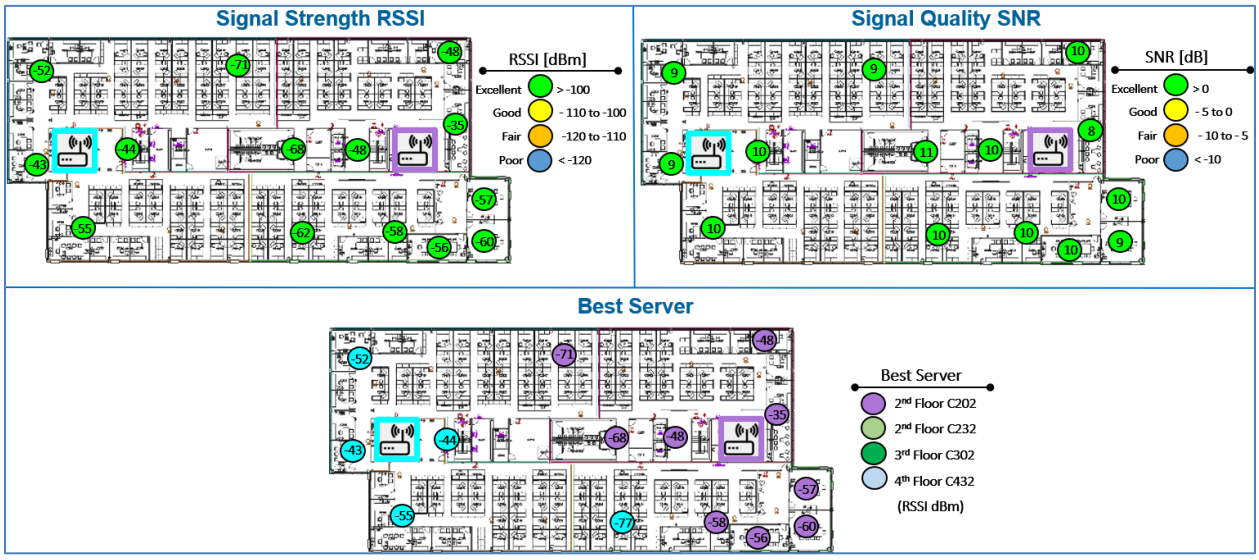


Figure 25 – Location 2 Walk Test Results 2nd Floor

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

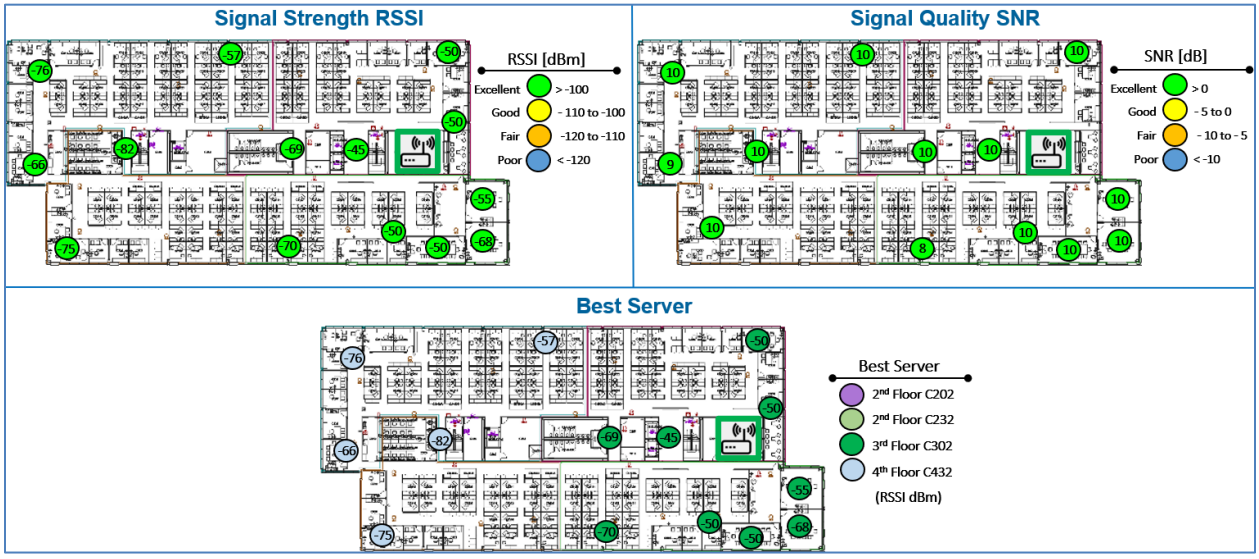


Figure 26 – Location 2 Walk Test Results 3rd Floor

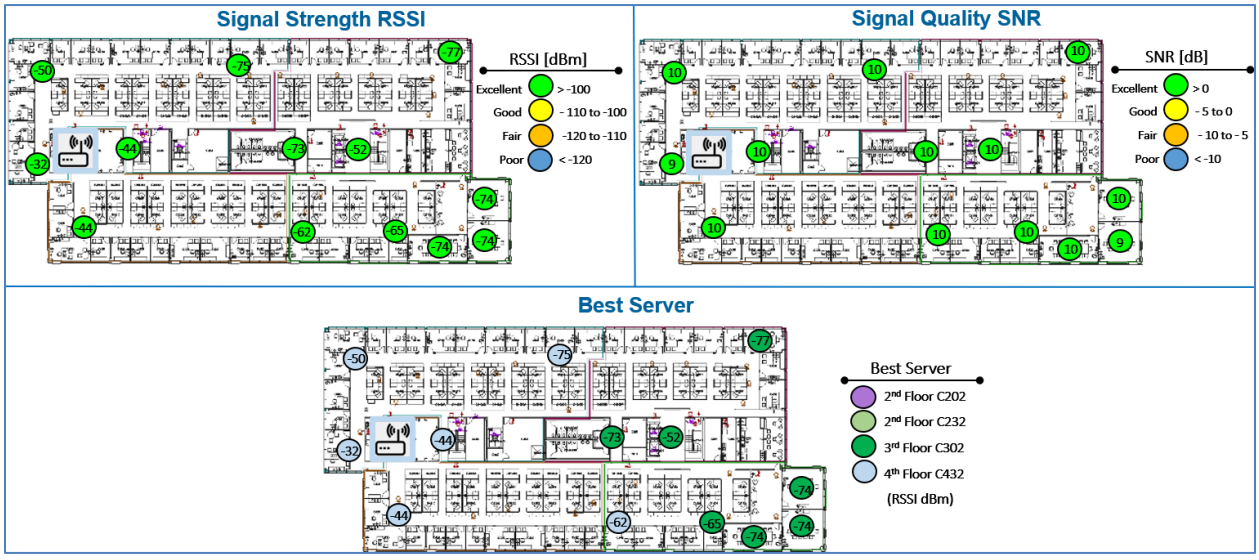


Figure 27 – Location 2 Walk Test Results 4th Floor

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.4. Application Server's Dashboards

The entire campus is exhibited on the main webpage; accordingly, the friendly menu allows a user to dive into specific, building, floor, room, and sensor information (Figure 28). Figure 29 indicates where exactly the sensor is located on the blueprint and allows the user to see real-time and historical data (Figure 30 and Figure 31).

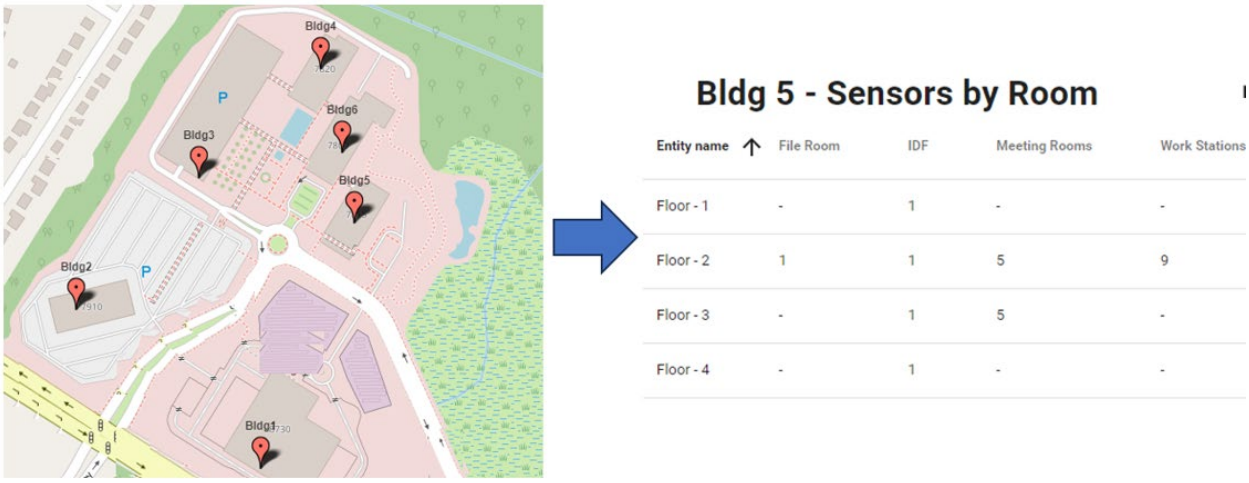


Figure 28 – Application Server Main Menu View

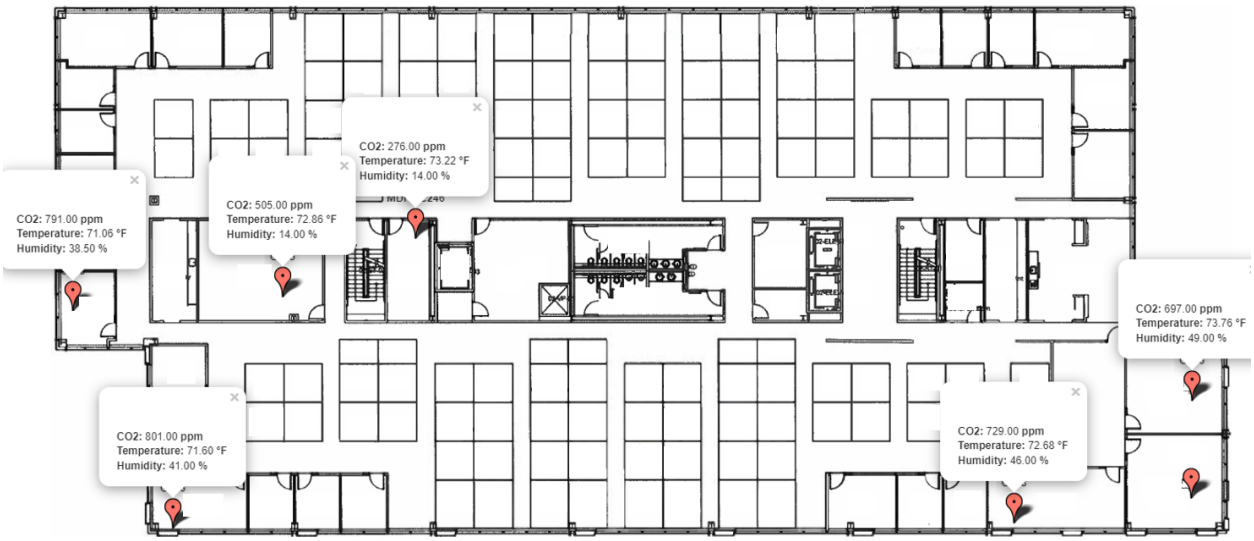


Figure 29 – Application Server Sensors Placement on 2nd Floor

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

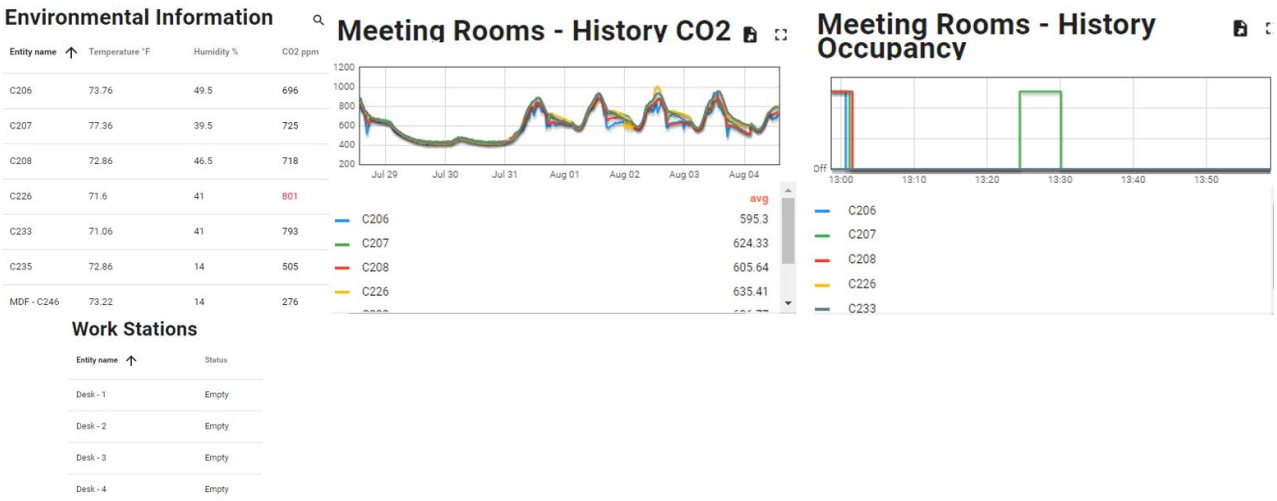


Figure 30 – Application Server All Sensors Data

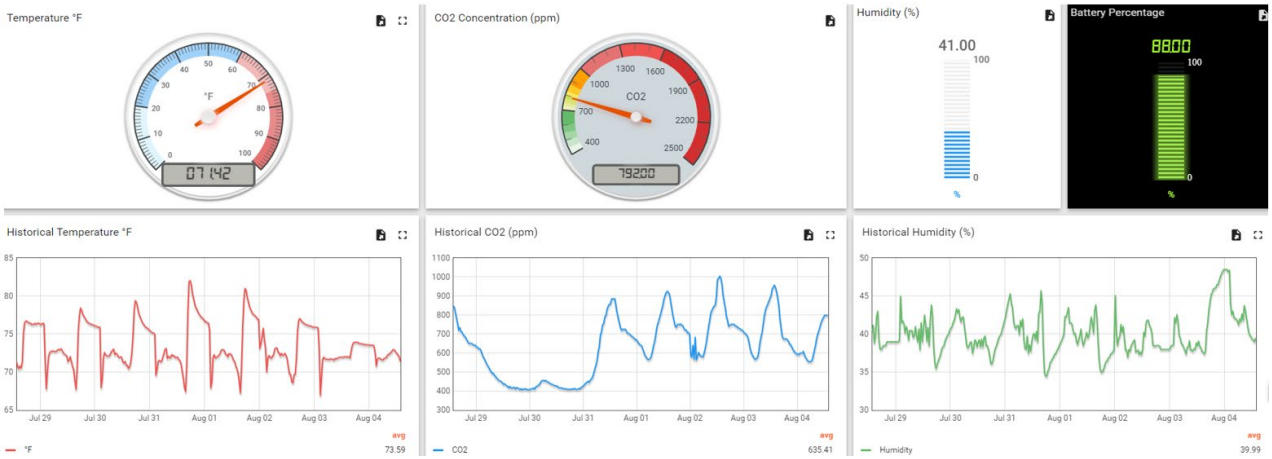


Figure 31 – Application Server Sensors Data Room C226

5. Conclusions and Best Practices

- Using the calibrated propagation model, the measured results show that all the samples are above -100 dBms of RSSI and above 0dB of SNR which means the entire building is covered with appropriate LoRa signal levels for the suitable sensor functionality.

- Based on the results obtained, the planning of the physical gateway placement is significant for developing low-power, long-distance communication via LoRa.

- The gateways are strategically located in order to cover the entire building, and their locations will warrant appropriate signal strength and quality levels. Therefore, one LoRa gateway can cover at least one floor below and one floor above it. Depending on the building size, shape, and design results, it is generally recommended to place gateways on different corners and alternating floors.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

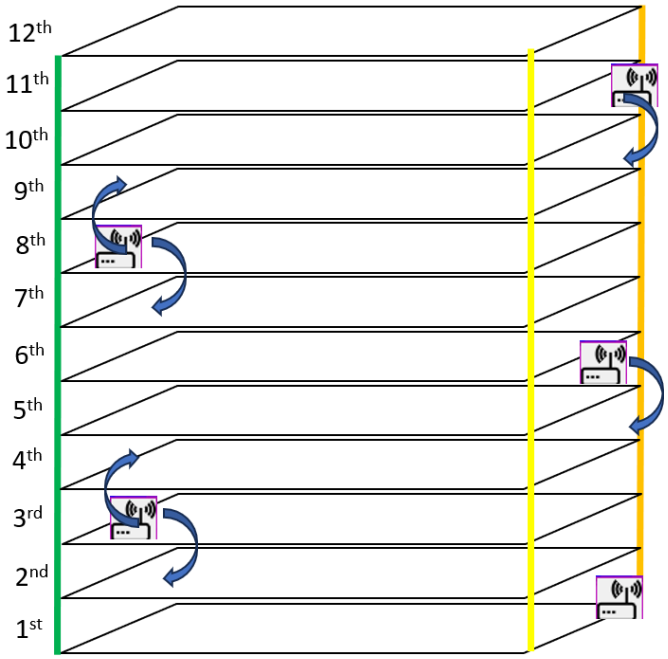


Figure 32 – Gateways Placement Strategy

Action points will be implemented with the goal of enhancing employee comfort, improving operational efficiencies, the efficiency of room occupancy, and reducing operating costs, according to the Table 4 – Action Points.

Table 4 – Action Points

Environmental Variables	Units	Triggers	Actions - Send email notifications +
Temperature	Fahrenheit (F°)	> 80	Turn on the backup AC
		< 62	Turn on the backup Heater
Humidity	Percentage (%)	> 75	Turn on the dehumidifier
		< 25	Turn on the humidifier
CO2	Parts per million (ppm)	> 1000	Turn on the fans and open doors
Remaining battery	Percentage (%)	< 25	Change batteries
Motion	String	Empty	Turn off the lights

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

3GPP	3rd Generation Partnership Project
ABP	Activation by Personalization
ADR	Adaptive Data Rate
AES	Advanced Encryption Standard
AppKey	Application Key
AppSKey	Application Session Key
App Server	Application Server
BW	Bandwidth
CR	Coding Rate
CSS	Chirp Spread Spectrum
DDR	Dynamic Data-Rate
DevAddr	Device Address
DevEUI	Device Extended Unique Identifier
DL	Downlink
DLOS	Direct Line of Sight
DNLOS	Direct Non Line of Sight
DR	Data Rate
FCC	Federal Communications Commission
FRT	Fast Ray Tracing
FTD	Field Test Device
GW	Gateway
GPS	Global Positioning System
IoT	Internet of Things
HTTP	Hypertext Transfer Protocol
ISM	Industrial, Scientific, and Medical
JS	Join Server
KHz	Kilohertz
LNS	LoRaWAN Network Server
LoRa	Long Range
LoRaWAN	Long Range Wide Area Network
LPWAN	Low Power Wide Area Network
MAC	Medium Access Control
MHz	Megahertz
ms	Milliseconds
NaaS	Network as a Service
NB	Narrow Band
NS	Network Server
NwkSKey	Network Session Key

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

OTAA	Over the Air Activation
PHY	Physical
RAN	Radio Access Network
Rb	Bit Rate
RF	Radio Frequency
RP	Reflected Path
RSSI	Received Signal Strength Indicator
RX	Receive
SCTE	Society of Cable Telecommunications Engineers
SF	Spreading Factor
SNR	Signal-to-Noise Ratio
ToA	Time-on-Air
TX	Transmit
UL	Uplink
Wi-Fi	Wireless Fidelity

Bibliography & References

[1] Semtech, “What is Lora?” https://lora-developers.semtech.com/learn/get-started/what-is-lora Accessed: Aug. 17, 2023.

[2] Lopez Chalacan, Victor Hugo. "Performance Evaluation of Long Range (LoRa) Wireless Rf Technology for the Internet of Things (IoT) Using Dragino LoRa at 915 Mhz." (2020)

[3] Lopez, V., et al. "Evaluation of communication delays in LoRaWAN networks for indoor emergency scenarios." Journal of Communications Technology and Electronics 66.Suppl 2 (2021): S149-S158. K. Mekki, E. Bajic, F. Chaxel, and F. Meyer, “A comparative study of lpwan technologies for large-scale iot deployment,” ICT express, vol. 5, no. 1, pp. 1–7, 2019.

[4] RF Wireless World, “Formula/Equations used in LoRa Data Rate Calculator” https://www.rfwireless-world.com/calculators/LoRa-Data-Rate-Calculator.html Accessed: Aug.09, 2023.

[5] K. Mekki, E. Bajic, F. Chaxel, and F. Meyer, “A comparative study of lpwan technologies for large-scale iot deployment,” ICT express, vol. 5, no. 1, pp. 1–7, 2019.

[6] LoRa-Alliance, “RP2-1.0.3 LoRaWAN Regional Parameters”. https://resources.lora-alliance.org/document/rp2-1-0-3-lorawan-regional-parameters Accessed: Aug.09, 2023

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23 CELEBRATING 40 YEARS

2023 Fall Technical Forum SCTE • CABLELABS • NCTA

Sounds of Daily Living: Audio Classification for Deaf/Hard-of-Hearing Customers

A Technical Paper prepared for SCTE by

Adina Halter
Sr. Principal Software Architect
Comcast
1800 Arch Street, Philadelphia, PA 19103
484-832-3518
adina_halter@comcast.com

Abhijeet Mulye
Lead Researcher, Machine Learning
Comcast
1800 Arch Street, Philadelphia, PA. 19103
267-260-2643
abhijeet_mulye@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Audio Classification in the Home ... 3
   - 2.1. Why is Audio Classification Needed ... 3
   - 2.2. Current Sound Awareness Solutions ... 4
   - 2.3. Why Use Home Technology ... 4
   - 2.4. Our Approach ... 5
3. Classification Models ... 6
   - 3.1. Model Training ... 6
   - 3.2. Pre-Processing ... 7
   - 3.3. Neural Network Layer Description ... 8
   - 3.4. Classification Results ... 9
4. Solution Architecture ... 10
5. Conclusion ... 11

Abbreviations ... 12

Bibliography & References ... 12

List of Figures

Title **Page Number**

Figure 1 - Schematic of Training Process ... 7
Figure 2 - Audio Pre-processing ... 8
Figure 3 - Model Summary ... 8
Figure 4 - Training Result: Classification Report and Confusion Matrix for 21 Classes ... 9
Figure 5 - Testing Result: Classification Report and Confusion Matrix for 21 Classes ... 9
Figure 6 - Testing Result: Classification Report and Confusion Matrix for 5 Main Classes ... 10
Figure 7 - Audio Data Flow ... 10

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Approximately 14 million people in the United States are deaf or significantly hard-of-hearing (DHH). This means they are often unaware of important household sounds such as a doorbell, baby crying, microwave beeping, etc. While smart homes are becoming the norm, these solutions are heavily skewed toward the hearing community, largely ignoring the monitoring of commonplace but important household sound cues.

Using existing home cable networking devices, we have been able to successfully detect daily household sounds of interest amidst the noise generated by a busy household. Detection of the sounds of daily living can then be utilized by existing visual alert signalers to notify the deaf or hard-of-hearing customer.

2. Audio Classification in the Home

2.1. Why is Audio Classification Needed

"[Some] people with hearing loss find utility in sound awareness technology that recognizes and notifies them about relevant and important sounds, like safety-related sounds, social/human activity sounds, and non-urgent status sounds."

—Laurene Milan, Senior UX Researcher, Comcast

Sounds provide informative signals about the world around us, affording us the ability to better understand and interact with our environment. Sounds of interest usually fall under one of three categories:

- Safety
- Social presence
- Non-urgent

In situations where auditory cues are inaccessible, it's useful for DHH individuals to be notified about sounds. A 2019 wearable sound study by the University of Washington and Gallaudet University found that almost 75% of DHH were very interested in sound awareness solutions through visual and haptic feedback. They desired contextual awareness of activities around the home. Additionally, they wanted to selectively display sounds based on such things as:

- time of day
- what a DHH individual is currently doing
- how active the house currently is
- the DHH individual's location in the home

While this is extremely advantageous from an accessibility perspective, the advantages of having sound detection in your home benefit everyone.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.2. Current Sound Awareness Solutions

"For people with temporary, situational, or permanent hearing loss, there is no single cost-effective solution that supports continuous sound awareness in the home."

—Laurene Milan, Senior UX Researcher, Comcast

We will explore three commonly used solutions with their abilities and limitations.

Sound recognition settings on smart phones.
Pros: A library of common sounds for alarms, animals, households, and people can be selected from. In limited cases, a custom sound may be added.
Cons: DHH individual must always have the phone nearby. Detection is localized to the individual's current proximity, thus sounds in other areas of the house would not be picked up.

Smart speakers such as Amazon Echo and Google Nest Audio.
Pros: A library of sounds. Routines can be set up and peripherals such as lights can be used for notification.
Cons: Notifications are commonly done audially through the speaker. Optional screens can be purchased for additional cost, sometimes more than the cost of the smart speaker itself.

Sound awareness wrist devices such as Neosensory Buzz.
Pros: Awareness of hundreds of sounds, haptics are used for notification.
Cons: Cost ($999). Haptic only, limited use (doesn't tell time, monitor fitness, or anything else), conspicuous, learning curve to understand what the different haptic signals mean.

Amazon and Google offerings have expanded into the smart home space and now have significant market-share in the US, having sold 18% of all smart-home devices in 2021. For DHH customers who already have a suite of products like this installed, they may already have a viable solution. But for the cable + home security customer, buying a separate home system simply for sound awareness is redundant and may not be cost effective.

Additionally, certain late-deafened customers may not have the technical awareness or the discretionary income to spend on a separate sound awareness system.

2.3. Why Use Home Technology

"I already have light [based tech] at my house. I want a coordinated system that [includes] these devices."

—DHH individual interviewed for University of Washington study

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Based on a 2019 study run by the University of Washington, some in the DHH community stated a preference for a sound awareness system that was integrated into their existing smart home. One person in the study suggested incorporating her security camera sound detection into a sound awareness solution.

A substantial benefit of home technology is that the hardware is multi-purpose, and already owned or leased by households for purposes such as home security. By leveraging already-installed cable-based home security and monitoring systems, we can provide substantially similar sound classification alerts to the consumer.

Home technology is generally hands-free and typically does not require a dedicated wearable device (although smart watch apps can be incorporated if preferred). There is no need to unplug it, recharge it, or turn it off for bedtime.

The research field of artificial intelligence is highly active, with new technical papers being published to the arXiv (pronounced "archive") website every day. As sound classification models improve, we will be able to upgrade our AI models and seamlessly update the system so that detections get more accurate.

Hardware quality also improves every year, with the newer generations of home security devices being more cost effective, energy efficient, and offering higher resolution recording. Any advancement in the quality of microphones, analog-to-digital-conversion microchips, etc. will also likely directly improve the utility of this system.

While our current implementation is an edge-cloud hybrid, future iterations of the feature may live exclusively on-device, as Application Specific Integrated Circuits (ASICs) for AI become less expensive and more commonplace. Edge deployments are better for privacy as personal data does not need to be transmitted and stored on centralized servers. And as edge devices improve, machine learning (ML) and inferencing will be handled on-device.

2.4. Our Approach

Our DHH sound awareness strategy starts with analyzing security camera audio streams for important sounds like a baby crying or a glass breaking. Inferencing and classification is handled by an ML model to recognize and identify specific pre-determined sounds.

Upon classification, the user is alerted of the sound that was detected in their homes through a smart notification such as a mobile app, set-top box/television experience, or home hub display. Users can choose which sounds to be alerted on, and how frequently they would like the alerts. Being able to select the sounds they want to be notified of reduces notification overload.

Audio of interest might include:

- Dog Bark
- Baby Crying
- Water running
- Doorbell
- Washer-dryer/oven timer
- Footsteps
- Conversation

For example, a DHH mother with a newborn child may find it useful to turn their indoor security camera into a baby monitor. An elderly individual residing in a large house and unable to effectively hear high-pitched doorbell sounds may find value in enabling just the doorbell detection feature. And while a DHH dog-owner my find barking detections helpful, a DHH individual without a pet may find these same detections unhelpful, since those are likely from neighboring dogs, or false detections due to television.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Classification Models

Audio classification is a deep learning model trained on hundreds of thousands of samples. Sounds of daily living may be challenging to inference as the solution will focus on appropriate confidence of detection and conveying the level of certainty to the recipient. It is expected that the solution will continuously learn and improve with time.

3.1. Model Training

(We collect, store, and use all data in accordance with our privacy disclosures to users and applicable laws.)

We use Google's YAMNet as a pre-trained base model. YAMNet was selected due to following reasons: pre-trained to predict 521 classes of sounds from YouTube's AudioSet corpus; open source and available under Apache 2.0 license permitting commercial use; efficient to run on CPUs because of Mobilenet v1 architecture as opposed to many deep learning models requiring the use of more expensive GPUs; depth-wise, separable convolutions reduce the number of additions and multiplications compared to ordinary convolutional neural networks (CNN), thus making it efficient to train and infer; and it "only" has ~3.4 million trainable parameters, compared to some of the state-of-the-art models like PANNs, which can have billions of parameters.

In addition to the YouTube/AudioSet corpus which was used in pre-training, we re-trained the model using audio files collected from customer cameras, sound effects owned by NBC Universal studios, and certain proprietary databases. Transfer learning helps us in ensuring a proper "domain transfer" compared to straightforward thresholding and calibration.

Our current model predicts probabilities for the following 17 classes:

- ***Dog**
- Conversation
- Vehicle
- Footsteps
- Bird
- Television
- Wind
- Traffic
- ***Alarm aka "loud sound"**
- ***Kitchen appliance**
- ***Baby crying**
- Rooster
- Cat
- Others
- Running water
- ***Doorbell**
- Silence

Only the classes marked ***bold** are used to notify customers, whereas the remaining 12 classes serve as "negative classes." We empirically determined that rather than using a general negative class, specific negative classes perform better for refining positive results with overlapping frequencies. For instance, adding sounds made by a cat or rooster helped us eliminate certain false detections for the baby crying class which we care about. Similarly for the dog barking class, we added footsteps, pots and pans, etc.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

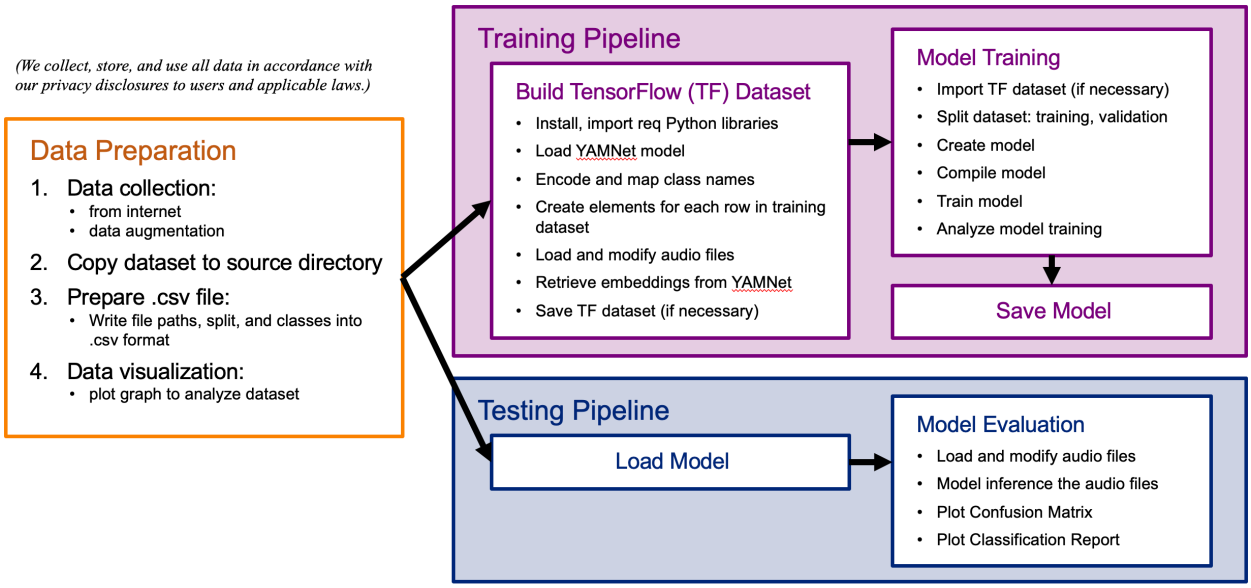


Figure 1 - Schematic of Training Process

Many of the sounds of interest, in particular microwave oven beeps and smoke detector alarms, were rare and difficult to collect organically. We used data augmentation techniques to make the most of training data. Samples were mixed with Gaussian noise, large files were split and mixed into multiple pieces.

While we trained the model on alarm sounds such as smoke detector, we simply labeled them all as “loud sound”.

3.2. Pre-Processing

Pre-processing of wav audio is similar to YAMNet. Audio is converted to a tensor, and Short-Time Fourier transform is applied to convert it to frequency domain. 96ms long frames are extracted with 48ms hop and a periodic Hann window is applied. The log Mel spectrogram thus created is fed to the deep learning model to get a 17x1 tensor prediction.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

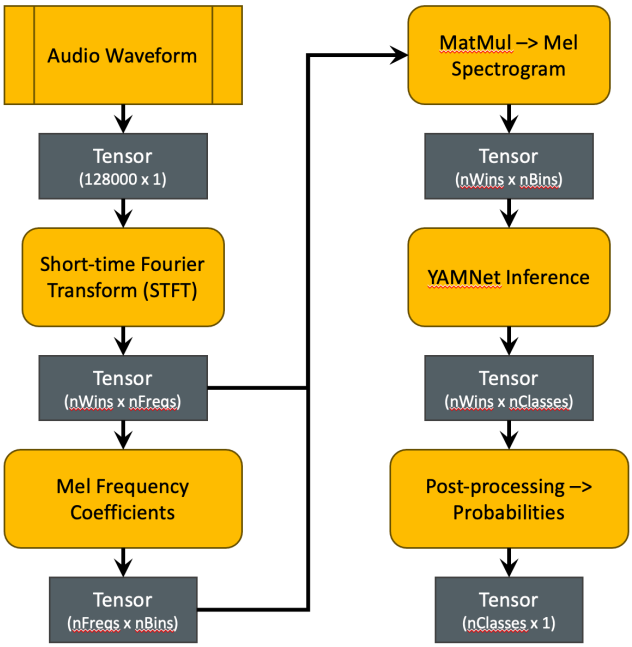


Figure 2 - Audio Pre-processing

3.3. Neural Network Layer Description

The following model summary diagram shows the layers and number of parameters used on top of the embedding generated by YAMNet. We have three dense layers separated by dropouts and Gaussian noise used to prevent overfitting the model. Final layer output 17 scores, one per audio class which can be fed to SoftMax and interpreted as relative probability.

Layer (type)	Output Shape	Param #
layer1 (Dense)	(None, 160)	164000
dropout_6 (Dropout)	(None, 160)	0
gaussian_noise_3 (GaussianNoise)	(None, 160)	0
layer2 (Dense)	(None, 480)	77280
dropout_7 (Dropout)	(None, 480)	0
layer3 (Dense)	(None, 150)	72150
final_layer (Dense)	(None, 17)	2567

Total params: 315,997
Trainable params: 315,997
Non-trainable params: 0

Figure 3 - Model Summary

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.4. Classification Results

Below are sample results of one of our later model versions. These continue to evolve as we perfect our model. We have since taken out some labels or collapsed them into a general negative data classification. We also continue to add more training data for labels, such as kitchen appliance sounds, to improve performance.

Training result				
	Classification Report for 21 classes			
	label	precision	recall	f1-score
	Baby_crying	0.78	0.97	0.86
	Bird	0.85	0.82	0.84
	Cat	0.94	0.84	0.88
	Conversation	0.88	0.74	0.8
	Dog	0.98	0.92	0.95
	Doorbell	0.97	0.91	0.94
	Footsteps	0.95	0.88	0.91
	Kitchen_appliance	0.92	0.88	0.9
	Laughing	0.93	0.94	0.94
	Loud_sound	0.98	0.92	0.95
	Musical	0.78	0.96	0.86
	Rooster	0.98	0.98	0.98
	Running_water	0.87	0.85	0.86
	Shouting	0.72	0.96	0.82
	Silence	0.45	0.76	0.56
	Singing	0.91	0.75	0.82
	Thump_thud	0.69	0.86	0.76
	Traffic	0.84	0.73	0.78
	Wind	0.83	0.59	0.69
	other_furniture_sounds	0.58	0.83	0.68
	other_kitchen_sounds	0.82	0.69	0.75
	accuracy			0.82
	macro avg	0.84	0.85	0.84
	weighted avg	0.84	0.82	0.82

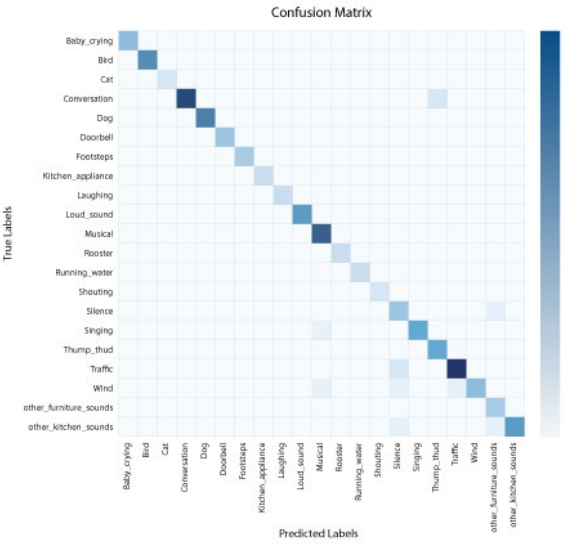


Figure 4 - Training Result: Classification Report and Confusion Matrix for 21 Classes

Testing result				
	Classification Report for 21 classes			
	label	precision	recall	f1-score
1	Baby_crying	0.86	0.75	0.80
2	Bird	0.74	0.82	0.78
3	Cat	0.87	0.79	0.82
4	Conversation	0.52	0.65	0.58
5	Dog	0.98	0.76	0.86
6	Doorbell	0.93	0.93	0.93
7	Footsteps	0.97	0.85	0.90
8	Kitchen_appliance	0.79	0.61	0.69
9	Laughing	0.93	0.94	0.93
10	Loud_sound	0.90	0.82	0.86
11	Musical	0.55	0.98	0.71
12	Rooster	0.96	1.00	0.98
13	Running_water	0.87	1.00	0.93
14	Shouting	0.65	0.93	0.76
15	Silence	0.11	0.33	0.16
16	Singing	0.85	0.72	0.78
17	Thump_thud	0.65	0.88	0.75
18	Traffic	0.85	0.72	0.78
19	Wind	0.77	0.51	0.61
20	other_furniture_sounds	0.52	0.77	0.62
21	other_kitchen_sounds	0.68	0.59	0.63
	accuracy			0.75
	macro avg	0.76	0.78	0.76
	weighted avg	0.80	0.75	0.77

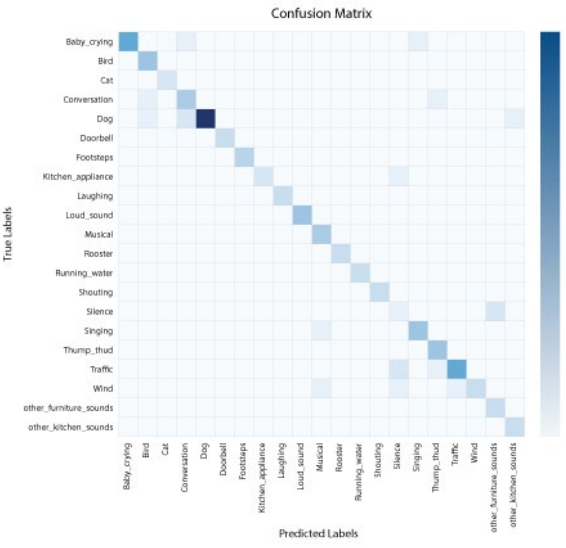


Figure 5 - Testing Result: Classification Report and Confusion Matrix for 21 Classes

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Testing result				
	Classification Report for 5 classes			
	label	precision	recall	f1-score
	Baby_crying	0.86	0.75	0.8
	Others	0.9	0.98	0.94
	Dog	0.98	0.76	0.86
	Doorbell	0.93	0.93	0.93
	Kitchen_appliance	0.79	0.61	0.69
	Loud_sound	0.9	0.82	0.86
	accuracy			0.91
	macro avg	0.89	0.81	0.85
	weighted avg	0.91	0.91	0.9

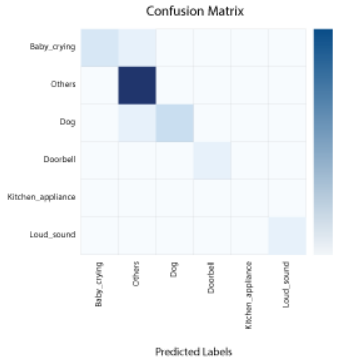


Figure 6 - Testing Result: Classification Report and Confusion Matrix for 5 Main Classes

4. Solution Architecture

We deploy the solution using a hybrid edge-cloud approach that aims to find the right balance of performance, cost, maintainability, and privacy.

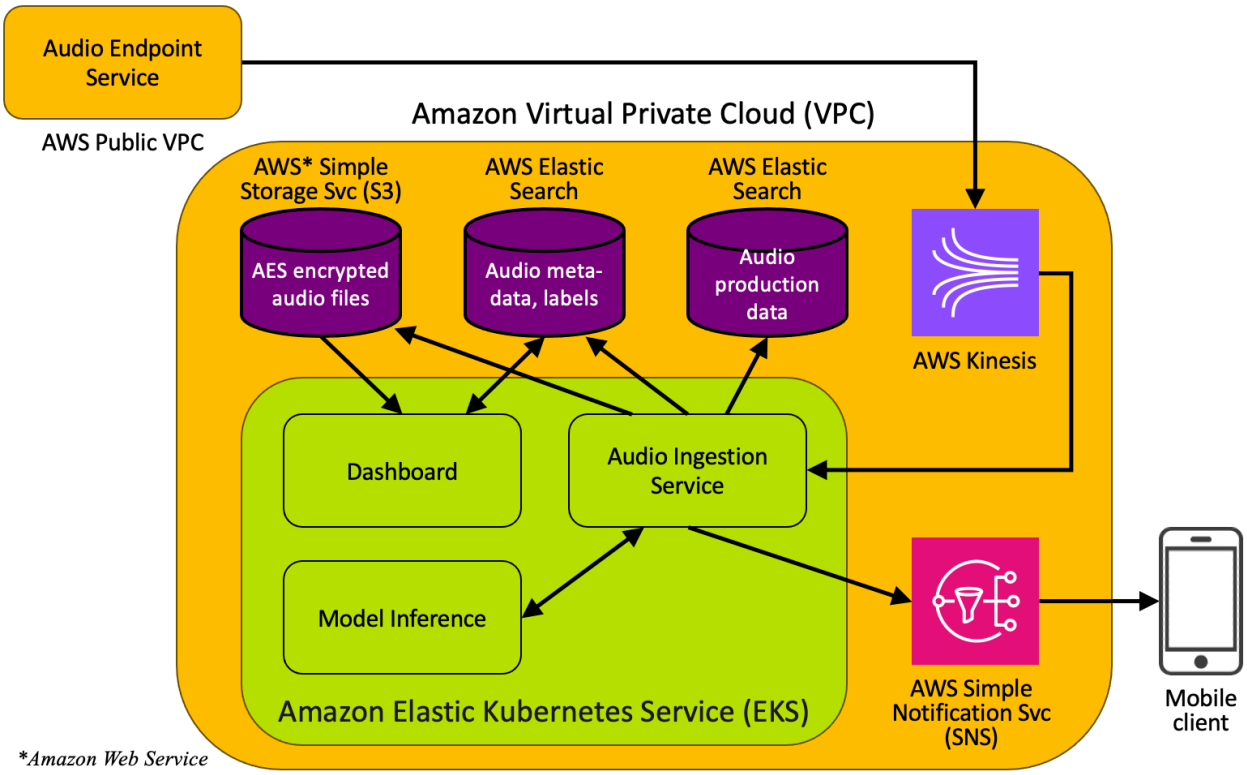


Figure 7 - Audio Data Flow

Although models are trained and validated using TensorFlow/Keras frameworks, we deploy them using the Open Visual Inference and Neural Network Optimization (OpenVINO) framework. OpenVINO is an open-source toolkit developed by Intel. It is designed to optimize and accelerate the deployment of deep learning models for computer vision applications. OpenVINO provides a set of tools, libraries, and runtime components that enable efficient inference on a wide range of hardware platforms, including

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

CPUs, GPUs, FPGAs, and VPUs. Compared to TensorFlow serving, OpenVINO achieves a 40% decrease in latency on the same hardware.

Audio onset is detected on camera using a preset sound amplitude level threshold. When the instant power of the audio in a frame exceeds this threshold, an 8-sec clip is recorded and uploaded to an AWS endpoint. The sampling rate of this audio clip is set at 8khz and the resolution is 16 bits per sample. The recording uses a single (mono) channel. Thus, one clip equals 128kilobytes in size. By utilizing a relatively low sampling frequency compared to standard YAMNet use case (16 khz) we're able to reduce the bandwidth necessary to transmit the audio from edge to cloud by half.

The rest of the pipeline includes an endpoint where enrolled cameras send raw encrypted audio.

See diagram on architecture above (audio data flow).

Amazon Elastic Kubernetes Service (Amazon EKS) is a managed Kubernetes service that makes it easy for you to run Kubernetes on Amazon Web Services (AWS) and on-premises. Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications. Amazon EKS is certified Kubernetes-conformant, so existing applications that run on upstream Kubernetes are compatible with Amazon EKS.

The extended part of the pipeline does the following:

- The audio ingest service calls the audio inference endpoint and writes the metadata and the inference results into Elasticsearch. The ELB framework is used by the ML team to quickly diagnose and troubleshoot false detections.
- Inference results are also stored under the predictions attribute of the audio metadata in the elastic dashboard.
- There are dev and prod clusters that process the data in parallel, and changes are first pushed to and validated on the dev cluster before mirroring them in the prod cluster.
- The audio-inference endpoint runs in both dev/prod clusters. The client for the audio-inference endpoint was written in Python, but the audio-ingest service is written in Go for concurrency and high performance. A little client API (application programming interface) service receives requests from the audio-ingest service, calls the inference endpoint, and returns the results to the audio-ingest service.
- There is also a notification piece responsible for sending SMS messages (via AWS SNS) to the trial customers so, once in production, this piece is not needed.

This hybrid architecture is easy to maintain and improve. A new model can be integrated easily with a continuous integration and continuous delivery (CI/CD) concourse pipeline.

If 50,000 cameras were to be enrolled in the system and upload 1,000 clips daily, the expected throughput would be roughly 600 requests per second. We can handle inferencing on c5.2xlarge node types with under 200ms latency.

5. Conclusion

The classification of sounds of daily living can offer the DHH community the ability to receive and react to sound cues in their environment through their residential cable services. By leveraging existing cable home security services, paired with a ML/AI classification model which analyzes security camera audio

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

detection streams, cable companies can offer a low- or no-cost solution as part of a customer's sound awareness strategy, giving them an opportunity to better understand and interact with their environment.

Abbreviations

AI	artificial intelligence
Amazon EKS	Amazon Elastic Kubernetes Service
API	application programming interface
AWS	Amazon Web Services
CI/CD	continuous integration and continuous delivery
CNN	convolutional neural network
DHH	deaf or hard-of-hearing
ML	machine learning
ML/AI	machine learning/artificial intelligence
OpenVINO	Open Visual Inference and Neural Network Optimization
YAMNet	Tensorflow's pre-trained deep net that predicts over 500 audio event classes, based on Google's AudioSet-YouTube corpus

Bibliography & References

Amazon Dominates the $113 Billion Smart Home Market — Here's How it Uses the Data it Collects. 2022. Katie Tarasov. CNBC

Amazon EC2

Amazon Echo and Alexa Devices

Amazon EKS

Amazon Kinesis

App Helps New and Deaf Parents Know When and Why Their Baby is Crying. 2018. Simi Singer. UCLA Newsroom

Apple iPhone User Guide

arXiv Scholarly Articles. Cornell University

Connected Living: Sounds of Daily Living. 2022. Adina Halter, Sr. Principal Software Architect, Accessibility Innovations, Comcast; Brittany Roots, Product Manager for Accessibility, Comcast

Deaf and Hard-of-hearing Individuals' Preferences for Wearable and Mobile Sound Awareness Technologies. 2019. L. Findlater, B. Chinh, D. Jain, J. Froehlich, R. Kushalnagar, A. Lin

Deaf Parents and Hearing Children. Disability, Pregnancy & Parenthood

Deafness And Hearing Loss Statistics. 2023. J. Wirth (contributor), A. Hall (Editor), L Jorgensen, Au.D., Ph.D. Audiologist (medical reviewer). Forbes Health

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Elasticsearch and ELK Stack

Exploring Sound Awareness in the Home for People Who are Deaf or Hard of Hearing. 2019. D. Jain, A. Lin, R. Guttman, M. Amalachandran, A. Zeng, L. Findlater, J. Froehlich

Google Nest Audio

Google Nest Sound Detection

Home Alerting Devices for People Who are Deaf or Hard of Hearing. 2018. Gallaudet University

How to Enable Sound Detection with your Amazon Echo and Alexa. 2022. Chris Wedel. Android Central

Intel's OpenVINO

NeoSensory Website

PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition. 2019. Q Kong, Y Cao, T Iqbal, Y Wang, W Wang, M. D. Plumbley

Smart Home – Americas. Statista

Sound Awareness Technology Design Considerations for People with Hearing Loss. 2022. Laurene Milan, Senior UX Researcher, TPX User Research, Comcast

YAMNet github repository maintained by Manoj Plakal and Dan Ellis

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23 CELEBRATING 40 YEARS

2023 Fall Technical Forum SCTE • CABLELABS • NCTA

Tackling Upstream Ingress and CPD Issues in High-Split Cable Networks

Challenges and Solutions

A Technical Paper prepared for SCTE by

Claus Lassen
Senior Architect, HFC Networks
Norlys
clalas@norlys.dk

Steve Condra
Senior Director, Product Management
Teleste Intercept, SCTE Member
440 Forsgate Drive, Cranbury, NJ 08512
+16786418099
steven.condra@telesteintercept.com

Arttu Purmonen
Vice President, System Marketing
Teleste Corporation
Telestenkatu 1, Kaarina, 20660
+358405562942
arttu.purmonen@teleste.com

Pekka Hilke, Teleste Corporation

Niko Suo-Heikki, Teleste Corporation

Henri Aittomaa, Teleste Corporation

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Definitions and Categories of Ingress and Common Path Distortion ... 3
   - 2.1. Subcategories of Upstream Ingress ... 4
   - 2.2. Common Path Distortion ... 4
3. Norlys' Experiences ... 5
   - 3.1. The High-Split Roll-Out ... 5
   - 3.2. Ingress and CPD Challenges Following the High-Split Roll-Out ... 7
   - 3.3. Detecting, Finding, and Solving Ingress and CPD Issues after the High-Split Roll-Out ... 8
   - 3.4. Short or Long-Term Interference and "Blind" PNM ... 10
   - 3.5. The Bridge to North American Context ... 10
4. Ingress and CPD in the North American Context ... 11
   - 4.1. Applicability of Norlys' Experiences to the North American Context ... 11
   - 4.2. How we can optimize 'proactive time' within North American networks ... 12
5. Conclusion ... 14
   - 5.1. Limitations ... 14
   - 5.2. Concluding thoughts ... 14

Abbreviations ... 15

Bibliography & References ... 16

List of Figures

Title **Page Number**

Figure 1 - Essential Network Elements ... 6
Figure 2 - Three amplifier categories ... 7
Figure 3 - Probability of Disturbances ... 8
Figure 4 - Logical Flow of Ingress and CPD Detection ... 9
Figure 5 - Illustration of the Foundational Premises ... 11

List of Tables

Title **Page Number**

Table 1 - Classification of Ingress and CPD ... 4
Table 2 – Transponder Alternatives ... 13

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The cable industry plans to leverage Data Over Cable Service Interface Specification (DOCSIS) 4.0 technology over hybrid fiber-coaxial (HFC) networks to meet the demands of future bandwidth-intensive applications. While it remains uncertain which new split frequencies cable operators will deploy in their networks, there is consensus that the upstream frequency band will expand (Ciarla, 2023; Segura, 2022). However, this extended upstream introduces new challenges. Among these challenges, upstream ingress (Ciarla, 2023; Hranac et al., 2022; Segura, 2022; Topazi, 2022), and common path distortion (CPD) stand out as significant issues (Heiler et al., 2022). While these problems have long been recognized, it is crucial for cable operators to continually evolve their strategies for managing them, particularly during the transition to high-split networks and the deployment of distributed access architecture (DAA) (Segura, 2022).

Insights from various articles clearly show that addressing upstream ingress and CPD requires proactive measures, such as the deployment of proactive network maintenance (PNM) technology, which plays a crucial role in detecting and resolving these issues (Ciarla, 2023; Hranac et al., 2022; Volpe, 2019; Walsh, 2020; Wolcott, 2019). In response to the evolving landscape of cable networks, cable operators are striving to optimize workforce efficiency, improve subscriber experiences, and ensure reliable network performance through the adoption of innovative technologies and proactive maintenance practices (Volpe, 2019; Wolcott, 2019).

Norlys, a leading Danish cable multiple-system operator (MSO), has been at the forefront, pioneering large-scale high-split upgrades across its entire cable network infrastructure while concurrently rolling out DAA in Denmark. This case study elucidates the upstream challenges Norlys encountered and the solutions they employed. We then extrapolate these learnings to a North American context. This comprehensive approach is aimed at helping North American cable operators maintain high network uptime for the benefit of their subscribers and preempt issues before they negatively impact the customer experience.

In the context of cable networks, a 'high-split' configuration refers to the utilization of frequencies below 204 MHz for the upstream (return path) and frequencies above 258 MHz for the downstream (forward path). This shift in frequency allocation presents both new challenges and opportunities for cable operators. Moreover, the adoption of DAA, which involves implementing remote PHY devices (RPDs) or remote MAC PHY devices (RMDs) in place of traditional fiber nodes, further transforms the network infrastructure.

This paper is structured as follows: first, we define and categorize ingress and CPD; second, we delve into Norlys' experiences and the problems caused by ingress and CPD; third, drawing on the Norlys' case study, we establish foundational premises to frame the subsequent discussion in a North American context; fourth, we examine the requirements for tackling upstream ingress and CPD issues in North American high-split cable networks. We conclude by outlining the limitations of the paper and discussing future developments that could assist cable operators in managing ingress and CPD in their networks more efficiently.

2. Definitions and Categories of Ingress and Common Path Distortion

In high-split cable networks, upstream ingress is a critical concern, and it refers to the unwanted intrusion of external signals within the 5 to 204 MHz range of the upstream frequency spectrum (Ciarla, 2023; Hranac et al., 2022; Segura, 2022; Topazi, 2022). This interference, originating from a range of sources including both human-made and natural factors, poses significant challenges for cable operators striving to maintain signal integrity and ensure optimal network performance. As highlighted in the articles, upstream ingress is a persistent issue that can lead to interference and disruptions in the cable network. Operators confront

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the challenge of identifying and mitigating the sources of ingress to maintain signal quality and prevent service degradation. Furthermore, CPD complicates these challenges in cable networks, arising from the non-linear behavior of specific network components, particularly oxidized connectors (Heiler et al., 2022).

Table 1 - Classification of Ingress and CPD

Signal	Subcategory	Bandwidth	Origin
Ingress	Short-Term Ingress	Narrowband	Network
			Customer premises
		Wideband	Network
			Customer premises
	Long-Term Ingress	Narrowband	Network
			Customer premises
		Wideband	Network
			Customer premises
CPD	Short-Term CPD	Wideband	Network
	Long-Term CPD	Wideband	Network

2.1. Subcategories of Upstream Ingress

Short-Term Ingress: In our paper, short-term ingress refers to interference that lasts less than the period required to locate its exact source using traditional methods, such as truck rolls. These methods involve a field technician visiting locations that house RF amplifiers or fiber nodes and using measurement devices to determine where ingress enters the network or where CPD is created. As the field technician can only make educated guesses about these locations, they usually have to visit several sites before pinpointing the exact source. The duration of these visits, or 'truck rolls,' varies, so there is no exact time limit; however, short-term ingress typically lasts from one hour to several hours.

Long-Term Ingress: In this paper, long-term ingress refers to interference that persists long enough to be located using traditional methods, typically lasting from hours to days. By defining short-term and long-term ingress in this way, we indicate that traditional methods cannot locate short-term ingress, as by definition, once located, it becomes a long-term issue. Our rationale for this choice will become clear in later chapters.

Both short-term and long-term ingress can be further categorized based on their bandwidth. Narrowband ingress refers to interference with a bandwidth of less than 1 MHz, while wideband ingress implies interference with a bandwidth exceeding 1 MHz. The 1 MHz limit is a pragmatic choice, as broader signals are more likely to disrupt more than just one single-carrier quadrature amplitude modulation (SC-QAM) upstream channel. Typical sources of narrowband ingress include broadcast radio, analog television channels, amateur radio, public safety networks, and baby monitors. Conversely, common sources of wideband ingress can include electrical motors, power switching devices, lightning, and digital equipment, which can generate strong electromagnetic fields. We also classify ingress based on its entry point into the cable network. Here, the term "customer premises" refers to individual residential houses or apartments, while "network" refers to all other locations within the cable television network.

2.2. Common Path Distortion

CPD is caused by non-linearities within the network, frequently due to oxidized connectors (Heiler et al., 2022). This distortion manifests as wideband noise in the return spectrum, adversely affecting network

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

performance. Importantly, CPD mainly stems from parts of the network susceptible to oxidation and corrosion. When high downstream signal levels pass through these damaged network parts, performance is compromised. Contact points prone to CPD are affected by vibrations, changes in humidity, and temperature variations, all of which can influence network integrity. Consequently, both short-term and long-term CPD can be present in a cable network, but it very seldom originates from individual customer premises.

3. Norlys' Experiences

Our paper primarily concentrates on the emergence and resolution of ingress and CPD issues in high-split cable networks. Consequently, our discussion of Norlys' network upgrade will be concise. Our focus will rather be on a detailed examination of the ingress and CPD issues that Norlys encountered following the completion of their high-split rollout.

3.1. The High-Split Roll-Out

In 2016, Norlys operated cable networks using frequencies up to 65 MHz for upstream and up to 860 MHz for downstream. Although Norlys could have increased upstream capacity by further segmenting the network and upgrading to the 85 MHz return path, these methods were deemed to offer only modest improvements. The intense competition among fixed broadband providers, the increasing prevalence of full fiber-to-the-home (FTTH) installations, and a fully deployed downstream spectrum collectively influenced the decision to transition to a high-split (204 MHz/1.2 GHz) cable network. To achieve high spectral efficiency, Norlys decided to combine the high-split with DAA roll-outs. The existing network cables, power supplies, and documentation were primarily in good condition. However, to ensure all upgrade prerequisites were met, Norlys opted to run a few pilot projects before proceeding with the full roll-out. The main roll-out process spanned four years and was completed in 2021.

Following the roll-out, Norlys now operates a modern cable network, composed of 204 MHz / 1.2 GHz intelligent RF amplifiers, 1.2 GHz passives, DOCSIS 3.1 RPDs, and a DOCSIS 3.1 converged cable access platform (CCAP). The network is managed via the CCAP core, PNM, and a network management system (NMS). The NMS and PNM are both accessible, remotely and locally, to field technicians tasked with resolving ingress and CPD issues. Nearly all amplifier cascades have fewer than five amplifiers (N+5), and the majority of cable modems are DOCSIS 3.1 compliant. However, only customers requiring high upstream capacity orthogonal frequency-division multiple access (OFDMA) have had new wall outlet installations. Most customers still have the traditional (65 MHz / 85 MHz) European-style wall outlet with a separate FM radio port, even though FM radio is no longer available in the network, as frequencies below 204 MHz are allocated for upstream use. Figure 1 presents the essential network elements and management systems. FM radio signals are still partly present in the network as a form of ingress that enters the network via the traditional wall outlets.

PNM: The proactive network maintenance functionality, enabled by cable modems and CCAP cores, generates data to facilitate its operation (Volpe, 2019; Wolcott, 2019). PNM plays a crucial role in identifying and alerting network operators to the presence of interference, including both ingress and CPD. While PNM provides valuable insights into the general nature of problems caused by interference, it's important to note that it might not be sufficient for pinpointing the exact source of the interference. For instance, Norlys' experiences indicate that locations where micro-reflections occur may not necessarily align with the points where ingress enters the network or where CPD originates.

NMS: The network management system has the capability to collect and analyze data produced by the connected (intelligent) amplifiers. When combined with PNM data, it provides Norlys with tools to pinpoint

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the sources of ingress and CPD more precisely. The NMS has many additional functions beyond the scope of our article, but for the purpose of locating disturbance sources, it must be capable of adjusting forward and return path gain, attenuation, and signal level of amplifiers. The impact of these changes can be monitored using both the PNM and the NMS. Without connected amplifiers (Segura, 2021), the functionality of the NMS is significantly compromised.

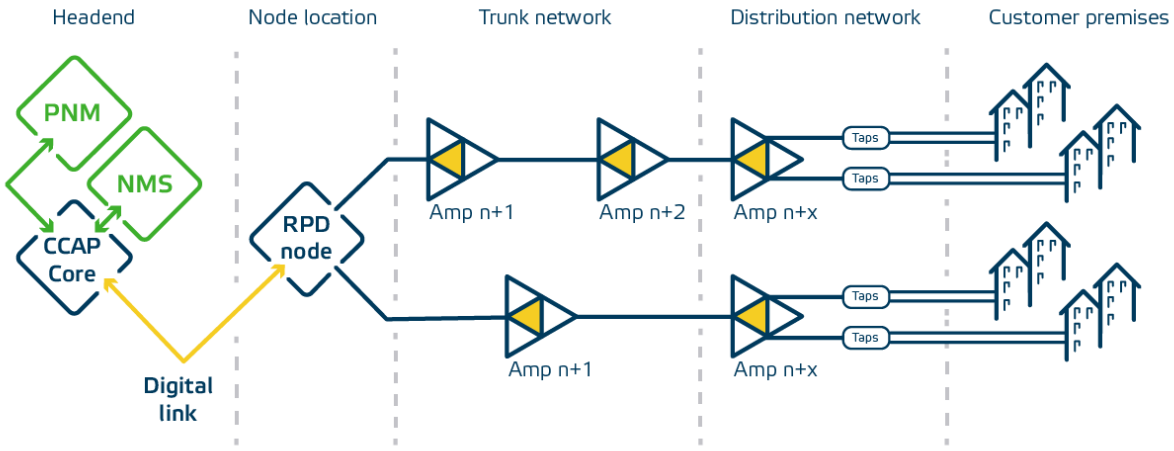


Figure 1 - Essential Network Elements

Intelligent, Connected, and Smart Amplifiers: The terms "intelligent amplifiers" and "smart amplifiers" are often used interchangeably (Segura, 2021). However, in this article, we delineate three distinct categories of amplifiers: "intelligent," "connected," and "smart." We classify them to discuss their unique features and their significance to North American cable operators. "Smart amplifiers" denote those equipped with automatic features, such as automatic level and slope control (ALSC) or return follows forward (RFF), which necessitate a microprocessor. "Connected amplifiers" represent a subset of smart amplifiers that can be remotely interfaced with the NMS, either unidirectionally or bidirectionally. Meanwhile, "intelligent amplifiers" describe a particular type of connected amplifiers fitted with full transponders that support upstream monitoring. As such, they are capable of upstream ingress/CPD analysis. Figure 2 showcases relationships between the three sets of categories. All amplifiers used by Norlys are classified as intelligent amplifiers. They operate in either trunk or distribution modes, with each mode demanding specific configuration templates. Norlys' amplifiers are equipped with DOCSIS 3.0-based transponders and a tuner, essential for enabling upstream measurements. These transponders connect to the NMS via the simple network management protocol (SNMP).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

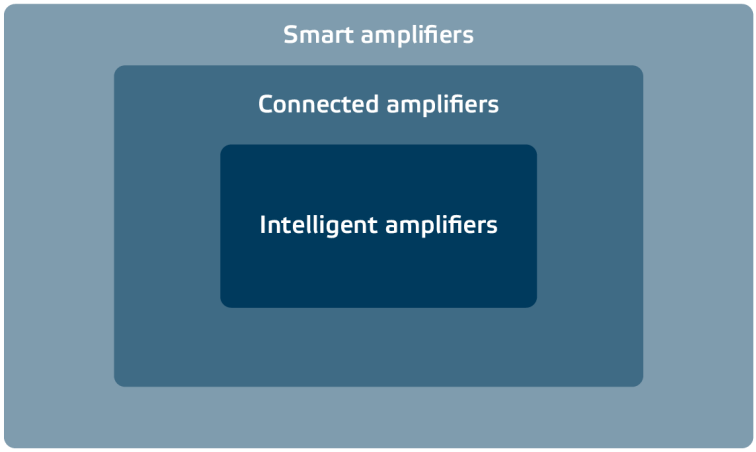


Figure 2 - Three amplifier categories

During the upgrade process, Norlys chose to maintain the levels of downstream signals. However, the upstream signal levels were typically reduced by 5 dB. This adjustment was made to establish a safety margin, ensuring that the cable modem's maximum transmit power wasn't exceeded when using the full upstream spectrum. Additionally, it aimed to mitigate distortion in amplifiers that could arise from increased upstream channel loads. It's important to highlight that this reduction in signal levels led to a decrease in the upstream carrier-to-noise ratio (CNR). To counteract this, Norlys shortened the amplifier cascades and minimized the size of amplifier clusters through segmentation—both measures aimed at reducing noise. Furthermore, the "analog" fiber originally placed between the headend and the optical nodes was substituted with "digital" fiber connecting the CCAP core and the RPDs.

3.2. Ingress and CPD Challenges Following the High-Split Roll-Out

Norlys has identified all previously mentioned categories of ingress and CPD within their network as discussed in Chapter 2. However, there has been a notable decrease in the occurrence of these issues, both during and especially after the completion of the roll-out. Along with a reduction in network issues, the process of resolving them has become more proactive. As a result, subscribers now enjoy higher quality service and experience reduced network downtime. It's critical, however, not to jump to conclusions regarding causality. The improved network quality isn't due to higher upstream frequencies, but to the introduction of new devices, tools, and tightened connectors. Therefore, even if the network had been upgraded while maintaining the old frequency split (65/85 MHz), the network quality would have improved if the same tools (NMS, Intelligent amplifiers, PNM) were utilized. Of course, a massive network upgrade to support the same frequencies doesn't make practical sense. Despite the upgrades, ingress and CPD remain persistent challenges in maintaining network performance. Figure 3 offers an illustrative representation of the likelihood of ingress and CPD occurrences within the network, emphasizing their significance. From the available data, it's clear that ingress is the predominant form of disturbance across three quadrants, and narrowband ingress originating from customer premises dominates. On the other hand, wideband distortion, which could be attributed to either ingress or CPD, primarily originates from the network itself. These findings underline the importance of addressing both ingress and CPD to ensure optimal network performance.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

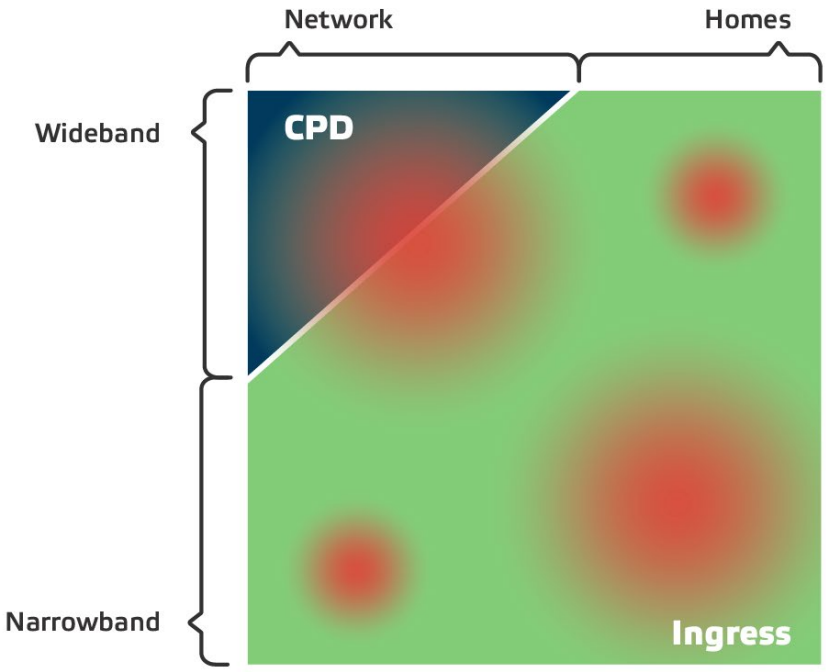


Figure 3 - Probability of Disturbances

3.3. Detecting, Finding, and Solving Ingress and CPD Issues after the High-Split Roll-Out

Instead of subscribers initiating complaints due to poor network quality, most ingress and CPD issues are first detected by the PNM, and occasionally by amplifiers capable of sending ingress alarms to the NMS. The process triggered by the PNM is illustrated in Figure 4. When the PNM sends an upstream interference alarm (1), it indicates the presence of either ingress or CPD. The PNM has access to data generated by the CCAP core, RPDs, and cable modems. This data encompasses, but is not limited to, upstream modulation error ratio (MER) per channel and per subcarrier, codeword error ratio (CER), and the upstream spectrum. Consequently, an initial understanding of whether the interference is related to ingress or CPD can be formed early on and confirmed in subsequent stages of the process.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

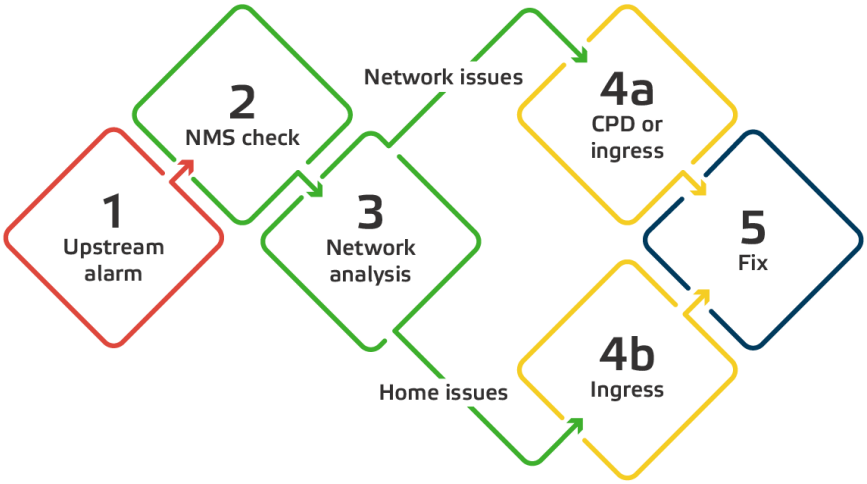


Figure 4 - Logical Flow of Ingress and CPD Detection

Next, the NMS is examined (2) to determine if it shows alarms from RF amplifiers below the specific RPD. In such cases, the NMS, rather than the PNM, can be utilized to analyze the amplifier upstream inputs more accurately, leveraging a remote connection facilitated by a transponder present in each amplifier. (3) The network is dissected amplifier by amplifier, following the amplifier cascade downwards towards the subscribers. This analysis is based on the 6 dB upstream attenuation that can be remotely toggled in every amplifier. While there are other possible levels of attenuation, Norlys has chosen 6 dB. If this attenuation diminishes the interference level, the interference is confirmed to be already present in the amplifier's upstream input, and the subsequent amplifier (N+2) is then examined. This investigation continues until the 6 dB upstream attenuation in the amplifier (N+X) does not affect the level of disturbance. In such an instance, it's known that the issue enters the network between amplifiers N+X and N+X-1. Even though the process involves multiple steps, it doesn't have to be time-consuming, as it can be expedited by the NMS.

When interference permeates the network between amplifiers, be it trunk or distribution, a further CPD analysis (4a) is conducted if the interference has a wideband nature. This analysis is executed using the NMS, which can remotely adjust the downstream level/gain of a specific amplifier, as CPD disrupting the upstream is induced by downstream distortion falling over upstream frequencies. Manipulating the downstream level/gain uncovers if the interference is a result of CPD. This information proves valuable for field technicians, as CPD tends to occur in connections close to the amplifier downstream outputs, where the downstream signal level is high, and the upstream signal is relatively low. Consequently, CPD directs field technicians to inspect connections near the amplifier.

Considering that the NMS operates on servers, it can be accessed by authorized personnel either at the headend or in the field. This allows a single individual to autonomously pinpoint and address network issues. If interference is detected below the last amplifier and it's not related to CPD, the PNM can be employed manually to inspect the parameters of the cable modems. If the last amplifier has multiple ports, the number of cable modems requiring examination can be further narrowed down. In such scenarios, the upstream inputs of the chosen amplifier can be attenuated individually. A 6 dB attenuation can reveal which drop line is the source of ingress. Typically, in these situations, the ingress originates from the subscriber's premises. Any information that reduces unnecessary customer premises visits—often necessitating

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

separately scheduled appointments—is invaluable. Once the problematic customer premises are identified, the necessary repair can be carried out.

3.4. Short or Long-Term Interference and "Blind" PNM

Previously, short-term ingress and CPD were characterized as disturbances lasting shorter than the time needed to identify and rectify the exact source using traditional methods, such as truck rolls. In the context of these traditional methods, "short-term" could span several hours, during which subscribers might face network disruptions or reduced bandwidth. However, Norlys has adeptly reduced the average time to identify the origins of ingress and CPD by synchronizing the functions of the PNM system, NMS, and intelligent RF amplifiers. This proactive and integrated strategy has effectively redefined formerly "short-term-untraceable" issues as "long-term" – not because they last longer, but because they now endure sufficiently long for their sources to be identified and addressed by the faster method.

While rare, there may be instances where the PNM system has limited visibility, making it unable to access particular cable modems or groups thereof. This is more probable with software that hasn't fully matured in the cable modems. Though such issues commonly originate from the cable modems themselves, Norlys acknowledges that intense interference might cause a modem to disconnect. Regardless of the root cause of this "PNM blindness," RF amplifiers equipped with transponders, in conjunction with the NMS, offer a means for remote troubleshooting. In extreme scenarios, remotely deactivating a specific amplifier input can confine noise funneling problems to a particular subsegment below that input. This proactive approach ensures that the broader subscriber base remains shielded from disturbances related to such noise funneling challenges.

3.5. The Bridge to North American Context

Ulaga et al. (2021) underscored the paramount importance of foundational premises in delineating new concepts. These premises are assertions that pinpoint and expound on the core tenets of a phenomenon. In this article, we lay out the foundational premises drawing from key insights from Norlys' experiences. These will steer the conversation in the North American context in the ensuing chapter. Figure 5 depicts these foundational premises:

(1) With traditional methods, the cumulative time needed to address CPD & ingress-related challenges comprises both the time taken to fix the issue and the time to identify its origin. An elongated timeframe heightens the risk of subscribers encountering suboptimal service quality.
(2) By harnessing the PNM, bolstered by the monitoring capabilities of DOCSIS devices (CCAP core, RPD, cable modems), and integrating it with the NMS, which benefits from the monitoring and remote management functions of connected RF amplifiers, cable operators can achieve time efficiency in pinpointing CPD and ingress problems.
(3) This time conservation can be labeled as "proactive time." During this proactive time, the problem is present but has not yet garnered complaints from subscribers.
(4) An extended proactive time results in subscribers benefiting from an increased quality of service.
(5) While PNM provides invaluable insights into the network, it is not wholly adequate if the overarching ambition of network operators is to maximize both proactive time and the quality of service experienced by users.

The experiences of Norlys and the foundational premises prompt us to consider two key questions, which we will tackle in the subsequent chapter: (1) How applicable are Norlys' experiences in the North American context? and (2) How can we optimize 'proactive time' within North American networks?

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

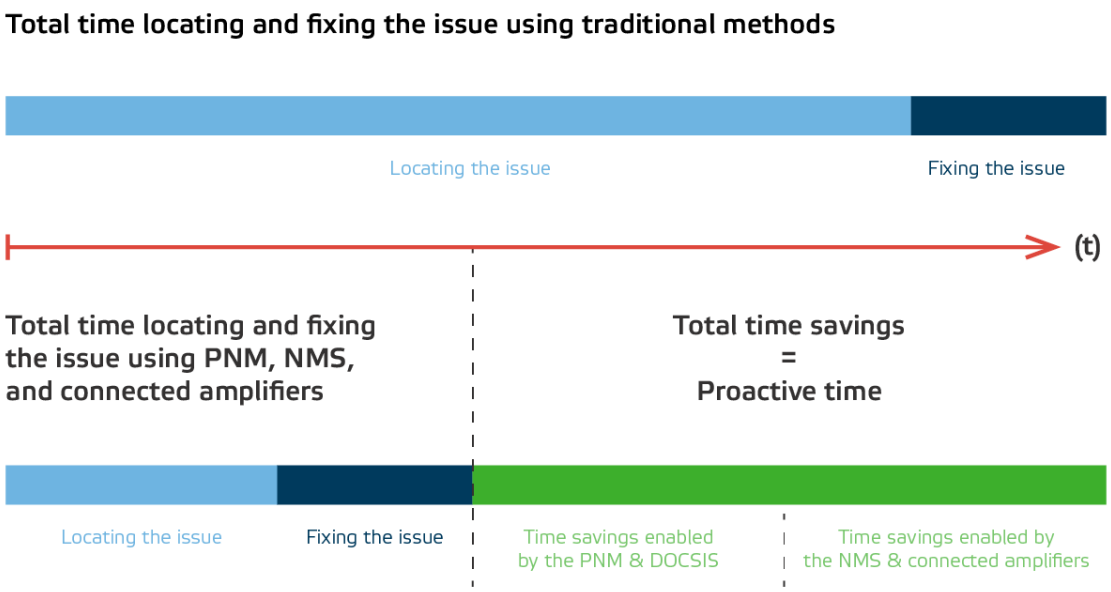


Figure 5 - Illustration of the Foundational Premises

4. Ingress and CPD in the North American Context

While the specific circumstances of Norlys in 2016 may not directly align with the current situation faced by North American operators, many of the issues they encountered are relevant and applicable. The high-split upgrade, combined with the deployment of DAA and the limitation of N+5 cascaded amplifiers behind RPDs, are common aspects in the North American context as well. With upcoming upgrade projects in North America, there is a growing focus on transitioning to a 204 MHz / 1.8 GHz cable network (also known as extended spectrum DOCSIS (ESD)). This presents operators with a unique opportunity to enhance their networks by leveraging the orchestration of PNM, NMS, and intelligent or connected 1.8 GHz amplifiers equipped with smart functionality. It's worth noting that the PNM Working Group has already published operational practices, guidelines, standards, and training content, demonstrating that the benefits of PNM are well-recognized and understood in the North American cable industry (Hranac et al., 2022; Volpe, 2019; Walsh, 2020; Wolcott, 2019). Thus, after discussing the extent to which Norlys' experiences are applicable in North America, we focus on expanding understanding of the cost versus synergistic benefits of NMS and intelligent/connected amplifiers, aiming to maximize proactive time.

4.1. Applicability of Norlys' Experiences to the North American Context

Norlys' coaxial cables are buried underground. This placement makes them less susceptible to environmental factors such as rain, wind, animals, ultraviolet radiation, and outdoor temperature fluctuations, all of which can potentially harm cables and their connectors. Also, amplifiers in North America are more exposed to these elements but temperature fluctuations are a significant issue in Europe too as street cabinets can heat excessively during summer. In some North American systems, moisture infiltrating the connectors might freeze, leading to damage. Hence, there's a higher likelihood of connectors

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

in North American networks oxidizing. Damaged cables and connectors often result in ingress, while oxidized connectors are prone to causing CPD.

For Norlys, FM radio interference originating from customer premises is a predominant source of upstream ingress. In contrast, North American operators might find terrestrial very high frequency (VHF) television more problematic. Here, VHF can seep into the high-split upstream frequency band via homes or impaired outdoor cables and connectors. While various factors suggest North American cable networks could be more at risk from environmental changes, also the strategies to address these issues diverge slightly. For instance, while Europe's underground coaxial cables seldom suffer damage, when they do, mending them is especially tough due to restricted access. In Europe, addressing amplifier-related problems typically involves inspecting street cabinets. In North America, however, mending strand-mounted amplifiers demands specialized tools. Consequently, we foresee that repairs might prove more demanding in North American networks, emphasizing the value of tools that can pinpoint the location of faulty network components.

Customers of Norlys benefit from wall outlets that offer an additional layer of protection against ingress arising from in-home installations. In contrast, in North America, the use of a basic splitter can increase the risk of ingress, especially if the splitter is of subpar quality or has sustained damage. Despite warnings against subscribers tampering with splitters, cables, or attempting DIY installations, these issues continue. In extreme cases, a single individual's DIY attempt can jeopardize service quality for the entire node/RPD area. In summary, while North American operators might face challenges similar to those addressed by Norlys, the frequency and severity of such problems might vary. However, the difference in experiences between Norlys and its North American counterparts might not be more pronounced than the variations seen between operators in the southern and northern parts of North America.

4.2. How we can optimize 'proactive time' within North American networks

Norlys has enhanced its network quality by leveraging PNM, strengthened by the monitoring capabilities of DOCSIS devices. The information they provide is further integrated with the situational awareness of the NMS, which benefits from the monitoring and remote management functionalities of the intelligent RF (204 MHz / 1.2 GHz) amplifiers. While the North American cable industry has led the way in PNM and DOCSIS innovation, a significant challenge remains: How can new 1.8 GHz amplifiers be connected to the NMS? Without this connection, cable operators cannot achieve optimum proactive time, crucial for elevating network and service quality.

When Norlys started its upgrade in 2016, both DOCSIS 3.0 transponders and hybrid management sublayer (HMS) transponders were mainstream solutions. Norlys chose to adopt DOCSIS 3.0-based transponders equipped with integrated upstream analyzers. This choice was influenced by the simplicity and capabilities these transponders offered. For instance, the chosen RPDs did not require support for narrowband digital return (NDR) and narrowband digital forward (NDF). Additionally, there was no need for HMS-capable devices and management systems. The inclusion of an upstream analyzer in the transponders allowed Norlys to conduct more detailed upstream analyses, aiding them in a closer examination of ingress and CPD issues. However, as of this writing, DOCSIS 3.0 transponders are nearing the end of their lifecycle. The current HMS standard faces challenges, including limitations in the ultra-high split frequency range, the security of the supported SNMP version, IPv6 compatibility, and maximum packet/frame size. While delving into solutions for these challenges is beyond the scope of this article, we aim to provide technical insights for the North American cable industry. This industry must deliberate whether future transponders should be HMS or DOCSIS-based, or if both should and can remain available also in the future.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

While Norlys employs intelligent amplifiers (refer to Figure 2 and respective definitions), many of the capabilities they offer would still be available even if future ESD amplifiers were connected. In such a scenario, the adjustments of amplifiers and their impact on upstream ingress and/or CPD would be monitored by RPDs, rather than other amplifiers higher up in the cascade. However, DOCSIS 3.1-based transponders present different advantages compared to their HMS-based counterparts. A deeper understanding of these advantages and disadvantages is rooted in their primary application and is further illustrated in Table 2. DOCSIS has been specifically designed to cater to subscribers seeking ultra high-speed broadband. DOCSIS cable modems are consumer devices powered by electricity within customer premises. While power consumption is a critical factor, it can be offset if the trade-off results in enhanced broadband speeds. In-home installations tend to be more forgiving than challenging outdoor environments. Consequently, the environmental standards that DOCSIS silicon vendors must meet are relatively "lenient" compared to those for outdoor conditions. While the older DOCSIS 3.0-based transponders were closely aligned with HMS-based transponders, the divide between DOCSIS and HMS is widening, when/if DOCSIS 3.1-based transponders come into play. This divergence, along with the gap it represents, will likely expand as newer DOCSIS generations emerge in the future.

The origins of HMS lie in the management of cable television network elements having high density and strict footprint requirements. HMS has been steered towards a communication method that, while low in bandwidth, is exceptionally robust. Although certain DOCSIS 3.1 upstream modulation alternatives also boast significant robustness, using them would not result in "in-band" communication. This means the same spectrum wouldn't be used for both telemetry and subscriber broadband applications. Additionally, the lifecycle of HMS is more focused on network management than on subscribers. As a result, a single HMS generation can span multiple DOCSIS generations. However, the move towards distributed access signals the need for the next version of HMS, which we refer to as HMSv2 in this paper. The specifics of SCTE 25-1 termination, whether it be (1) in the RPD software, (2) in the out-of-band (OOB) gateway device, or (3) realized through direct NDF/NDR-to-HMS conversion in the HMS gateway, still require the collective input of the cable industry to reach a consensus.

Table 2 – Transponder Alternatives

Technology	**DOCSIS 3.1-based**	**HMSv2-based**
Primary Application	Ultra high-speed broadband	Management of HFC devices
Standard/Specification	DOCSIS	Extended SCTE 25-1/2
DS Frequency Range	112…1218 MHz	258…650 MHz
US Frequency Range	5…204 MHz	5…204 MHz
Technical Advantages	• No need to use NDR/NDF • No separate spectrum needed for telemetry • No separate HMS gateway needed	• Low power consumption • Smaller footprint (board space) • Purpose-built robust modulation for telemetry
Technical Disadvantages	• Higher power consumption • Consumer technology that should be field hardened • Less robust RF modulation	• Needs SCTE 25-1 termination • Separate spectrum allocation needed for OOB • Current standard needs updates

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Conclusion

Throughout this article, we have aimed to be transparent and informative, enabling readers to follow our reasoning. However, before concluding, we want to address some limitations of our article.

5.1. Limitations

In Chapter 2, we use a 1 MHz limit to differentiate between narrowband and wideband. We recognize this limit as somewhat arbitrary; it could be narrower (e.g., 0.5 MHz) or wider (e.g., 2 MHz). Nevertheless, we've applied this limit consistently throughout our article, allowing readers to judge if a different threshold would better suit them. This delineation also influences Figure 3. While the figure uses a fourfold table to depict wideband/narrowband and network/home disturbances, we're aware that some disturbances don't fit neatly into these categories. Rather than presenting these issues as binary, we view them as opposite ends of “spectrums”. To communicate this nuance, we've placed the x (Network, Homes) and y axis (Wideband, Narrowband) labels in the corners of Figure 3, instead of centralizing them.

Further clarification is needed for Figure 5, which portrays total time savings as a combination of those enabled by PNM/DOCSIS and by NMS/amplifiers. This illustration doesn't suggest the time savings from each are equivalent. In reality, PNM/DOCSIS may offer more time savings in some instances, while in others, the NMS in conjunction with connected amplifiers might be more beneficial.

One key limitation we'd like to clarify pertains to a potential misreading of section 4.2. This section might suggest that the need for amplifier transponders arises exclusively from ingress and CPD issues, and that the choice between DOCSIS and HMS-technologies is purely technical. However, addressing disturbances in the cable network is just one aspect of the broader value proposition of transponders. These transponders offer multiple functions not detailed in our article, such as (1) remote alarms, (2) full-band capture of the downstream spectrum, (3) remote amplifier hardware diagnostics, and (4) remote adjustments of RF alignments. Beyond these technical aspects, the cost difference between DOCSIS and HMS transponders will undoubtedly be a critical factor when producing the next generation of connected, or even intelligent, amplifiers at scale.

5.2. Concluding thoughts

Through an in-depth case study, coupled with extrapolated findings, our article provides valuable insights for North American cable operators aiming to maintain and even enhance network uptime by promptly addressing ingress and CPD issues. We believe North American operators have a unique opportunity in the upcoming years to elevate the capacity and quality of their networks, especially as high-split and DAA rollouts commence. Additionally, we introduce a categorization of amplifiers to delineate the similarities and differences among "intelligent amplifiers," “connected amplifiers,” and "smart amplifiers." Concurrently, we emphasize the critical nature of determining the connectivity solutions for future amplifiers. The cable industry may not be poised to back both alternatives, namely HMS and DOCSIS transponders.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

ALSC	automatic level and slope control
bps	bits per second
CCAP	converged cable access platform
CER	codeword error ratio
CNR	carrier-to-noise ratio
CPD	common path distortion
DAA	distributed access architecture
DOCSIS	data over cable service interface specification
ESD	extended spectrum DOCSIS
FTTH	fiber-to-the-home
HFC	hybrid fiber-coaxial
HMS	hybrid management sublayer
Hz	hertz
MAC	medium access control
MER	modulation error ratio
MSO	multiple-system operator
NDF	narrowband digital forward
NDR	narrowband digital return
NMS	network management system
OFDMA	orthogonal frequency-division multiple access
PNM	proactive network maintenance
QAM	quadrature amplitude modulation
RFF	return follows forward
RMD	remote MAC PHY device
RPD	remote PHY device
SCTE	Society of Cable Telecommunications Engineers
SC-QAM	single-carrier QAM
SNMP	simple network management protocol

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

Ciarla, R. (2023, Summer). Ingress Again? Making cable shielding effectiveness possible. Broadband Library

Heiler, G., Gadermaier, T., Haider, T., Hanbury, A., & Filzmoser, P. (2022). Identifying the root cause of cable network problems with machine learning. Springer Nature Code of Federal Regulations, Title 47, Part 76

Hranac, R., Rupe, J., Torbet, D., & Volpe, B. (2022, Fall). How Will Proactive Network Maintenance Change Under DOCSIS® 4.0? SCTE Standards Journal.

Segura, N. (2021, Winter). Intelligent Amplifiers for 1.8 GHz HFC Extended Spectrum — A Smart Idea. Broadband Library.

Topazi, C. (2022, Winter). OFDMA and Ingress, Some frequently asked questions (and a few answers). Broadband Library.

Ulaga, W., Kleinaltenkamp, M., Kashyap, V., & Eggert, A. (2021). Advancing marketing theory and practice: guidelines for crafting research propositions. AMS Review, 11(3–4), 395–406.

Volpe, B. (2019, Spring). What Can DOCSIS PNM Do For Me? Broadband Library

Walsh, J. (2020, Winter). Protecting Our Network Heroes, Keeping customers connected and technicians safe. Broadband Library.

Wolcott, L. (2019, Spring). PNM: State-of-The-Art and 2019 Roadmap. Broadband Library

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The Benefits and Challenges of Deploying 5G Small Cells on the HFC Strand Network

A Technical Paper prepared for SCTE by

Cameron Gillis
Sr. Manager, Networks Business
Samsung Electronics America
3245 146th Place SE, Bellevue, WA 98007
408-204-2430
c.gillis@sea.samsung.com

Riley Lehecka
Senior Product Manager
Energy Systems / Enersys
3767 Alpha Way, Bellingham, WA 98226
425-387-8325
riley.lehecka@enersys.com

Charles Chapman
Senior Technical Sales Representative, EnerSys
Chair of the SCTE DSS IoT Smart City working group

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

Table of Contents 2
1. Introduction 4
2. The 5G Strand Use Case 5
   2.1. MVNO Data Offload 5
   2.2. Business Case 7
3. The HMNO Network 8
   3.1. HFC Architecture 8
   3.2. Coax Power Distribution 9
   3.3. 5G CBRS HMNO Architecture 10
   3.4. HMNO DSDS Device Capability 11
   3.5. Wireless Spectrum and DOCSIS 12
      3.5.1. CBRS Spectrum Band 12
      3.5.1. Other Spectrum Bands 14
      3.5.2. DOCSIS 15
4. 5G Strand Small Cell Engineering 16
   4.1. Form Factor - Physical Characteristics 16
      4.1.1. Antenna Considerations 17
      4.1.2. Environmental Considerations 19
   4.2. Power and DOCSIS 20
      4.2.1. Broadband Network Power 21
      4.2.2. RF / Plant Considerations 22
      4.2.3. Directional Couplers and Power Passing Taps 25
      4.2.4. DOCSIS Backhaul 28
      4.2.5. Firmware Considerations 28
   4.3. Antennas 29
      4.3.1. Omni and Sector 29
      4.3.2. Electrical Characteristics 30
5. Network Design and Optimization 35
   5.1. Coverage and Capacity Planning 36
   5.2. Mobility 37
6. Deployment and Operations 38
7. Conclusion 39

Abbreviations 40

Bibliography & References 43

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Figures

Title **Page Number**

Figure 1 - MVNO Offload Distribution – Wi-Fi Offload Only 6
Figure 2 - MVNO Offload Distribution – With MSO CBRS HMNO Network 6
Figure 3 – HFC Network Architecture 8
Figure 4 – Broadband Network Power - Distribution 9
Figure 5 – 5G CBRS Network Architecture 10
Figure 6 – MNO to MSO MVNO Data Offload – DSDS 11
Figure 7 – CBRS Wireless Spectrum 13
Figure 8 – CBRS Channel Bandwidth, Component Carriers, and Carrier Aggregation 14
Figure 9 – Strand Small Cell Dimensions and Cable Clearance 17
Figure 10 – Wireless Antenna – Physical Characteristics 18
Figure 11 – Mechanical Tilt - Omni Strand Antenna 18
Figure 12 – Strand Small Cell Heat and Loading 19
Figure 13 – Broadband Network Power - Overview 21
Figure 14 – DOCSIS - Diplex Splits & Upstream Data Limits 23
Figure 15 – Directional Couplers for Small Cells 25
Figure 16 – Power Passing Taps for Small Cells 26
Figure 17 – Wireless Antenna – Omni and Sector 30
Figure 18 – Electrical Characteristics - Omni Antenna Array 31
Figure 19 – Electrical Characteristics - Sector / Panel Antenna 31
Figure 20 – xTxR Small Cell RF Chain 33
Figure 21 – Effective Isotropic Radiated Power Concept 33
Figure 22 – EIRP Calculation Comparison – FCC, Link Budget, and CBRS 34
Figure 23 – Wireless Coverage and Capacity 36
Figure 24 – MNO to MSO MVNO Data Offload – Mobility 37

List of Tables

Title **Page Number**

Table 1 – Directional Coupler and Power Passing Tap Considerations 27
Table 2 – Omni and Sector Antenna Considerations 29

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Cable / Multiple System Operators (MSOs) offering mobile phone services are some of the fastest growing mobile service providers in the US Market. The two largest MSOs, Comcast and Charter, have a combined total of over 12 million mobile subscriber lines as per their latest quarterly earnings reports. The MSOs are buying spare capacity from existing mobile operators to serve their subscribers, which labels them Mobile Virtual Network Operators (MVNOs). These MSOs have a key advantage over other mobile operators in that they can now utilize new technology to leverage their existing hybrid fiber coax (HFC) network to overlay a 5G mobile network. Comcast, Charter and Cox invested a combined $1.1B in Citizens Broadband Radio Service (CBRS) licenses in 2017 as part of the Federal Communications Commission (FCC) Auction 105[i]. They can now use those frequency licenses and their HFC network to surgically target areas to build out their own mobile network that will coexist with the network capacity that they are buying from existing mobile operators to make them a Hybrid Mobile Network Operator (HMNO).

Strand-mounted 5G small cell radio can be deployed directly on the HFC aerial strand available in many cities in the United States. Attaching this small cell radio to the existing HFC network solves some of the biggest problems a mobile network operator faces when deploying a new greenfield network:

- Site acquisition - hanging on the strand requires no permissions or pole attachment fees.
- Power - the strand-mount small cell radio derives its power from the existing coax infrastructure.
- Backhaul - the strand small cell utilizes an embedded Data Over Cable Service Interface Specification (DOCSIS®) cable modem (CM) for backhaul.

While offering many benefits, deploying a small cell on the communications strand roughly 20 feet above the ground comes with some engineering challenges. This paper will detail the functional components which make up a strand-mounted small cell radio and how those components are packaged together to meet both cable network and wireless network engineering considerations.

Cable network engineering considerations include:

- Optimizing power consumption for maximum performance.
- Supporting flexible cable modem frequency splits across different MSOs.
- Preventing ingress / egress / spurious noise to isolate DOCSIS coax RF from wireless RF.

Wireless network engineering considerations include:

- Antenna design, form-factor, and gain.
- Wireless coverage and capacity planning.
- Utilizing optimum radio network components to maximize data offload performance.

This paper will give the reader an understanding of what needs to be considered as part of designing a 5G wireless network using strand-mounted small cells. It is the goal of this paper to provide a comprehensive guide for the MSOs to refer to as they move forward into expanding their respective convergence portfolios to include MVNO data offload with a strand small cell HMNO network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2. The 5G Strand Use Case

Many MSOs are now experiencing revenue growth and churn reduction associated with bundling mobile services with their broadband Internet offering. Since most MSOs do not own their own mobile network, they have signed MVNO contracts with Mobile Network Operators (MNOs) which gives them permission to roam onto the MNO's wireless network while offering the wireless service under the respective MSO brand.

One of the biggest operational costs for MVNOs is paying for all of the gigabytes (GBs) of data that their customers are consuming on the wireless network. MSOs acting as MVNOs selling mobile services have a huge competitive advantage in that they have existing wired broadband networks. Wherever possible, MSO MVNOs are utilizing technology to off-load data traffic from their MNO roaming partner's wireless network to their HMNO network, thus significantly reducing the monthly GB data bill they have to pay the MNO. Today, Wi-Fi is used to connect their wireless customers to their HFC network today, but their plan is to use small cells to increase the amount of traffic retained on their HFC network.

2.1. MVNO Data Offload

The majority of mobile data offload happens today in the home or office when a smartphone or tablet chooses to use a local Wi-Fi network instead of the mobile network for Internet connectivity. MSO MVNOs facilitate this action by working to ensure that their Wi-Fi networks are always available in locations where their mobile customers frequent. Wi-Fi is well suited for indoor coverage, but has limitations in an outdoor environment and many MSOs are now looking for ways to achieve data offload in outdoor locations.

One way for MSO MVNOs to achieve outdoor data offload is to deploy their own wireless network in high traffic areas where they know their wireless customers frequent.

For outdoor data off-load applications, MSO MVNOs now have the ability to deploy a new generation of small cell radios which are designed to attach to their aerial HFC strand infrastructure, with a similar form factor as their existing nodes or elements. The new strand small cell can be powered directly from the DOCSIS network and utilize an integrated DOCSIS modem to backhaul traffic to the MSO's core network using a single cable (one bill per site attachment: power and backhaul). Similar to Wi-Fi data offload, devices are programmed to connect to the MSO's wireless signal, when available, and offload Internet traffic to that network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The objective of the MSO is to offload outdoor wireless traffic from the MNO network to the MSO's HMNO data offload network to improve the subscriber experience through added coverage and capacity as data demand increases in outdoor wireless coverage Areas of Interest (AOIs). The following figures outline the concept of MVNO data offload in terms of typical distribution between Wi-Fi and 5G for indoor and outdoor before / after adding the MSO's HMNO / MVNO data offload network.

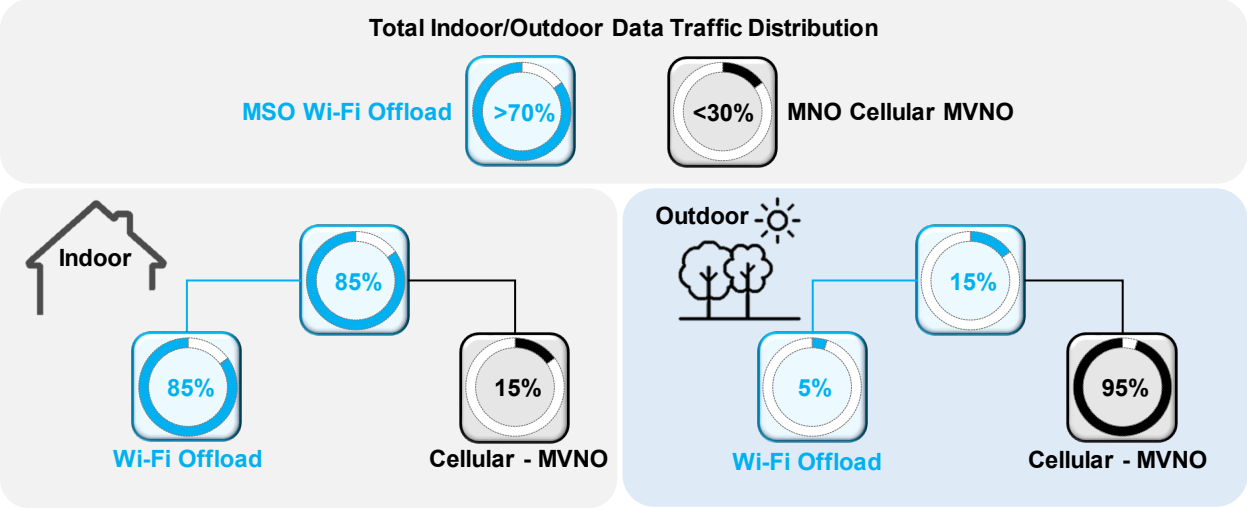


Figure 1 - MVNO Offload Distribution – Wi-Fi Offload Only

As clearly depicted in the above figure, the objective for the MSO is to offload as much outdoor wireless traffic from the MNO network to the MSO's HMNO network as possible. The following figure provides an example of how adding the CBRS strand small cell as an outdoor solution for the MSO's HMNO network can significantly increase the MVNO data offload:

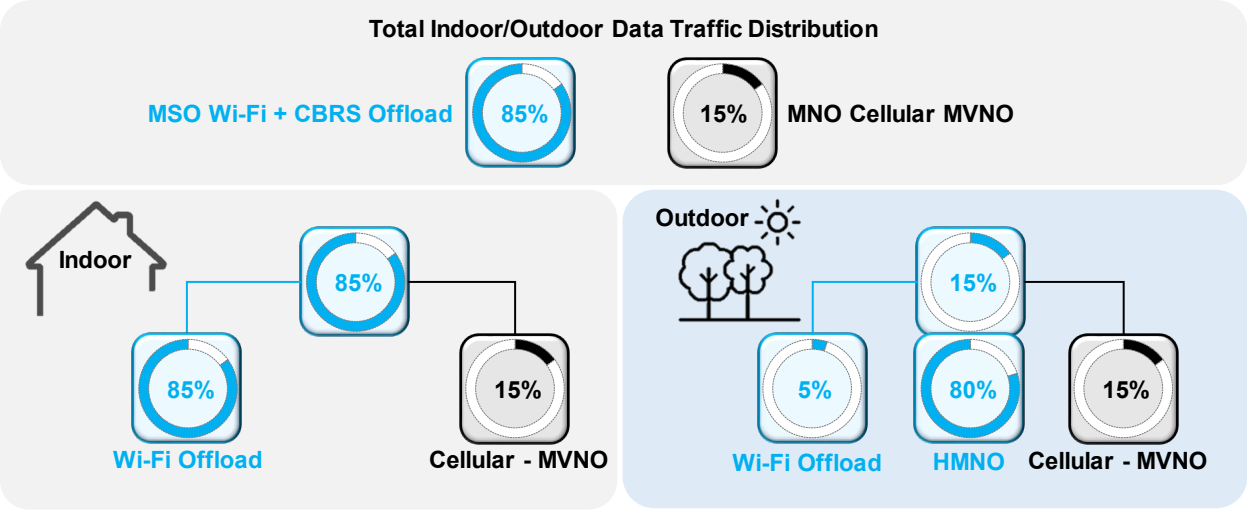


Figure 2 - MVNO Offload Distribution – With MSO CBRS HMNO Network

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.2. Business Case

The business case for deploying CBRS small cells on the HFC network is straightforward. Strand mounted CBRS small cells create operational savings by offloading GBs of data traffic with minimal incremental investment to the MSO's existing infrastructure while providing revenue growth and a more 'sticky' offering with their underlying subscriber base. Since MSO's are deploying within their footprint, this enhances their ability to deliver mobile service to those existing customers. According to a 2020 S&P Global Market Intelligence report, "Using an existing cable network to connect wireless base stations takes one-twentieth as much time and 1% of the cost of laying down new fiber-optic cables"[ii].

There are over 340,000 miles of accessible coax in the United States. HFC passes 96% of all US homes[iii]. Local franchise agreements allow broadband operators to install equipment on the strand without obtaining new permits. This results in huge savings in time and expenditure, as working with the utility and going through a permitting process is typically lengthy and costly. Furthermore, battery backup provides hours of runtime during utility outages and DOCSIS 3.1 provides high speed, low-latency backhaul, that is readily available. These key factors make the strand-mount small cell business case attractive to MSOs.

The integration of the DOCSIS cable modem product with a small cell helps alleviate network congestion and provides higher capacity. It also improves data speeds with the added benefit of extending the cable network's coverage area. The combination of 5G's low latency with Cable's high-speed connectivity enhances the experience of real time services and applications. It paves the way to better supporting the growing number of Internet or Things (IoT) devices and smart technologies with diverse connectivity requirements.

To maximize the MSO's business case for their HMNO network, the strand small cells need to be strategically placed in high traffic areas throughout the target market. This maximizes the data traffic offloaded in the MSO's high traffic AOIs. Once fully operational, the data savings realized by the CBRS HMNO network (reduced monthly data payments to the MNO) needs to generate significant savings to total cost of deploying and operating the HMNO network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. The HMNO Network

3.1. HFC Architecture

From a high-level perspective, the HFC architecture consists of a headend with servers and Cable Modem Termination System (CMTS). Fiber carries the signal from the headend to the optical node. Typically, the power supply sits close to the node, where the transition from fiber to coax takes place. Coax is run to what is called the "last mile", and taps take the signal to the modem inside the customer premises. RF amplifiers are also strategically designed into the network to amplify the signal. Note that the copper coax allows power to go in any direction, which provides the MSOs with the infrastructure required to support the power requirement for all active components.

Since the MSOs have HFC running to the majority of homes and businesses across the country, this allows flexibility in designing and deploying small cells into high density locations where they already have power and bandwidth available. Plant engineers can review key considerations around power budgets and RF to leverage key locations for strand mounted small cells.

The following figure summarizes the HFC network architecture with long reach over fiber and "last mile" with copper RF:

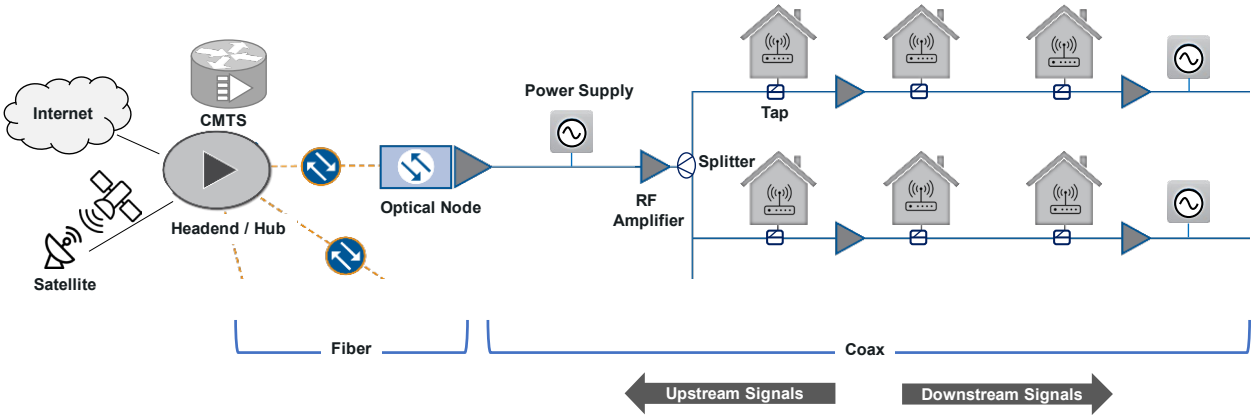


Figure 3 – HFC Network Architecture

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.2. Coax Power Distribution

The coaxial segment of the HFC network is also a power distribution network. The following figure outlines the distribution of broadband network power including the following key points:

- Power supplies are placed at regular intervals along the coaxial network to provide power to nodes, amplifiers, and gateways.
- Power insertion devices are used to multiplex AC power onto the same conductors carrying the RF signals.

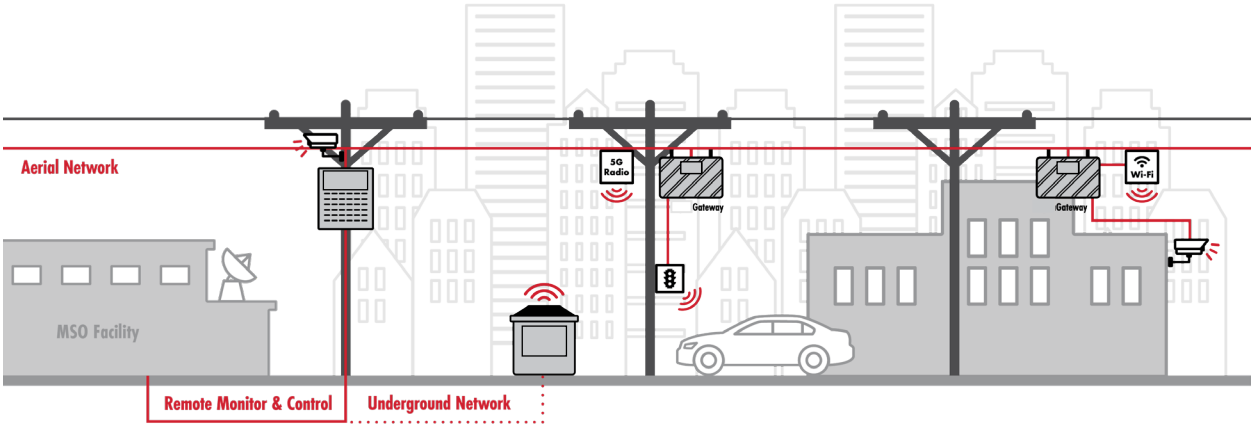


Figure 4 – Broadband Network Power - Distribution

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.3. 5G CBRS HMNO Architecture

The following figure outlines the key network components making up the 5G CBRS HMNO network[iv]:

- 5G Radio Access Network (RAN) – consisting of Next Generation Node B (gNB) radios and User Equipment (UE) wireless mobile devices,
- 5G Core – consisting of the Access and Mobility Management Function (AMF) and User Plane Function (UPF) and the following core network functions:
    - Network Slice Selection Function (NSSF)
    - Network Exposure Function (NEF)
    - NF Repository Function (NRF)
    - Policy Control Function (PCF)
    - Unified Data Management (UDM)
    - Application Function (AF)
    - Authentication Server Function (AUSF)
    - Session Management Function (SMF)
    - Data Network (DN),
- CBRS Spectrum Access System (SAS) - allocates spectrum based on availability and Tier 1, 2, and 3 prioritizations:

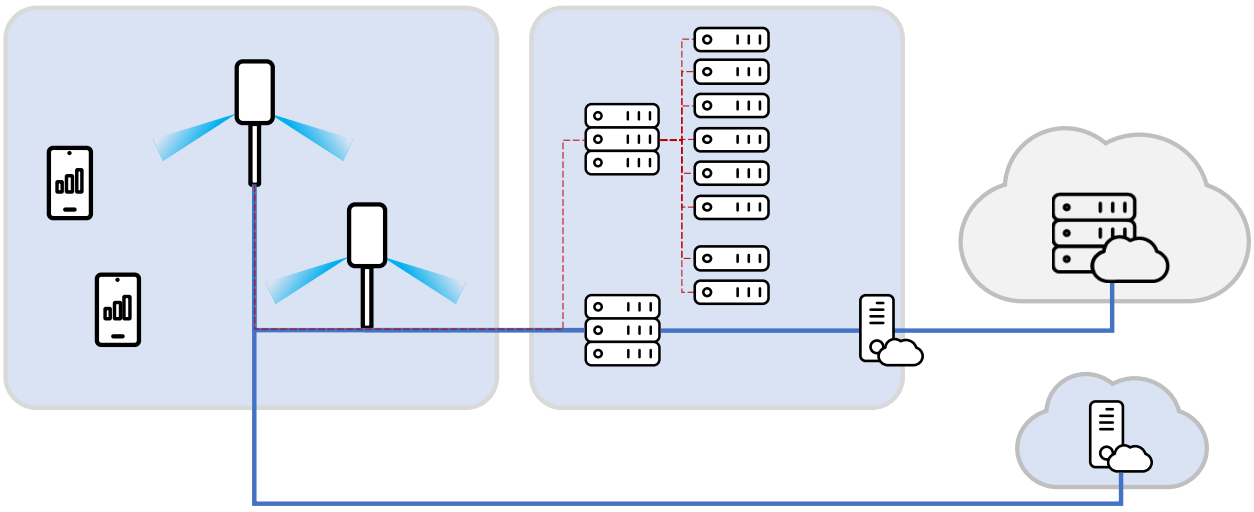


Figure 5 – 5G CBRS Network Architecture

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.4. HMNO DSDS Device Capability

As an integral part of the MVNO data offload ecosystem, end-user devices should support seamless switching between the MSO / HMNO and the MVNO networks. The most commonly used switching technology is Dual SIM Dual Standby (DSDS). In particular, the DSDS functionality manages the subscriber data access to the available wireless networks automatically. This is similar to how Wi-Fi offload works today, where the mobile device will automatically connect to the available Wi-Fi signal and route data traffic over Wi-Fi and save the subscriber from paying for data over the wireless network. This customized DSDS functionality allows the mobile device to manage the data offload preference to favor either the MSO's HMNO network (when available) or "fall back" to the MNO's network in case the MSO's HMNO network is unavailable or an unsupported service (e.g. voice) is required. The following figure summaries the DSDS functionality required for the UE part of the MSO's HMNO network ecosystem to operate effectively:

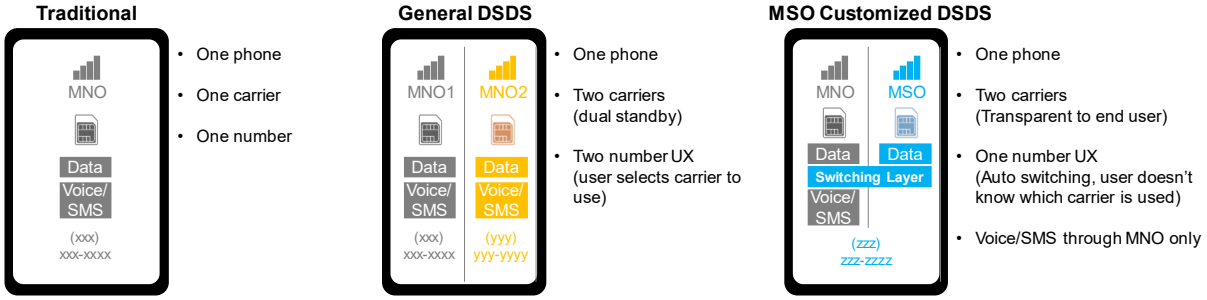


Figure 6 – MNO to MSO MVNO Data Offload – DSDS

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.5. Wireless Spectrum and DOCSIS

When considering the technology and architecture of an HMNO network, the MSO needs to decide on the wireless access spectrum for subscribers to connect their mobile devices to the wireless network. Additionally, the MSO will leverage their DOCSIS infrastructure to provide backhaul for the wireless HMNO network. This will be routed to the Internet and will provide a transparent service experience to the MSO's mobile subscribers. At the time of the writing of this paper, the following sections outline the current and future wireless spectrum and DOCSIS technologies available for the MSO to support an MVNO data offload ecosystem on their respective HFC network.

3.5.1. CBRS Spectrum Band

CBRS (operating from 3550-3700 MHz) is recommended for the MSO's MVNO data offload network due to device availability / network ecosystem maturity, and economies of scale / cost of device and network infrastructure.

3.5.1.1. CBRS Background

The following provides some background on CBRS spectrum:

- The FCC designated the CBRS band to operate in shared spectrum spanning 3550-3700 MHz and, together with the OnGo Alliance[v] and WInnForum[vi], established a product certification program for CBRS equipment and support the coexistence of Citizens Broadband Radio Service Devices (CBSDs). A CBSD is any wireless device that operates on CBRS spectrum. The 5G strand small cell referenced in this paper is classified as a CBSD.
- The 3rd Generation Partnership Project (3GPP) [vii] standards body defined band 48 for CBRS spectrum using 3GPP air interface technologies.
    - NOTE: When referencing 3GPP air interface standards in CBRS spectrum, "Band 48" or B48 refers to 4G/Long Term Evolution (LTE) while n48 refers to 5G/New Radio (NR).
- Both B48 (LTE) and n48 (NR) air interface are based on shared DL and UL Time Division Duplexing (TDD) channels / carriers that adhere to the 3GPP standard implementations of Orthogonal Frequency-Division Multiplexing (OFDM) Modulation Coding Schemes (MCSs).
- CBRS spectrum allocation to wireless operators is managed through a SAS provided by a third-party independent company.
- The SAS will "grant" access to blocks of CBRS spectrum (via channel grants) in increments of 5 and 10 MHz bandwidth; up to 40 MHz of Priority Access License (PAL) channels and up to the full 150 MHz of General Authorized Access (GAA) channels.
- In 2020, the FCC held Auction 105 which made available PALs in the lower 100 MHz (3550-3650 MHz) portion of the CBRS wireless access spectrum:
    - Many MSOs won PAL licenses through the CBRS auction which has an evolved ecosystem that supports the MSO's MVNO data offload ecosystem with DSDS devices.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The following figure summarizes the CBRS spectrum and how it is shared and allocated across the 3550-3700 MHz range:

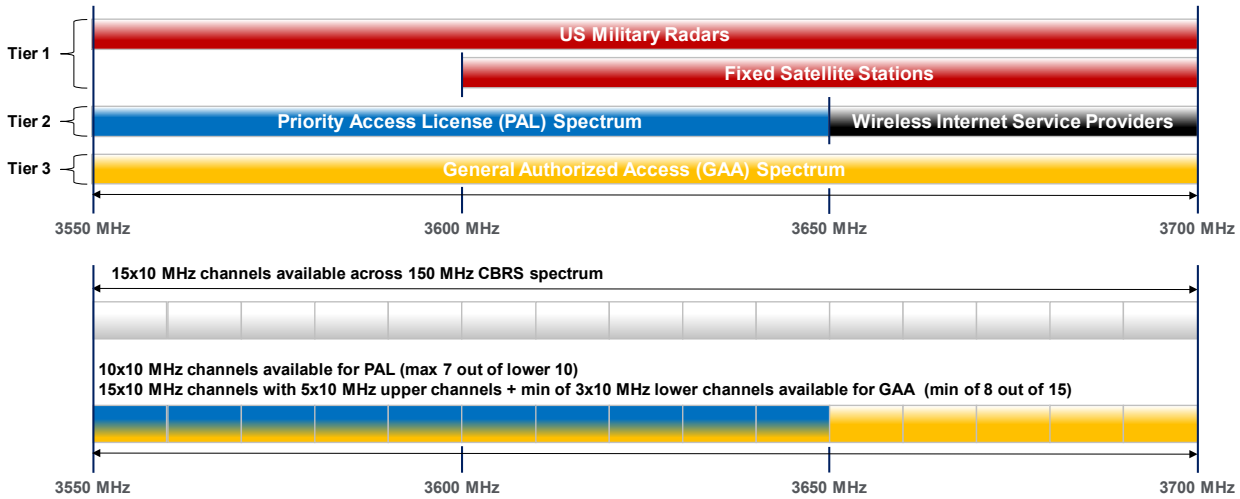


Figure 7 – CBRS Wireless Spectrum

Note that Tier 1, Tier 2 – PAL, and GAA spectrum all overlap. So long as the CBRS spectrum is available with no Tier 1 or Tier 2 spectrum allocations, the CBRS spectrum is available as GAA spectrum allocation.

3.5.1.1. CBRS Air Interface

The following lists the main RF characteristics (Modulation, Numerology, Bandwidth, Carrier Aggregation (CA)) of the CBRS n48 air interface:

- The downlink and uplink use OFDM modulation in a shared Time Division Duplex (TDD) channel.
- The MCS ranges from Quadrature Phase Shift Keying (QPSK) to 256 Quadrature Amplitude Modulation (QAM) constellations. QPSK provides high Forward Error Correction (FEC) for robust signaling and low throughput applications (especially on the uplink) while 256 QAM provides the opportunity to offer high throughput to users on the downlink channel.
- The numerology or Sub Carrier Spacing (SCS) can either be 15 or 30 kHz.
- The channel / carrier bandwidth can be 5, **10**, 15, **20**, **30**, **40**, 50, **60**, 70, **80**, 90, or 100 MHz with typical values highlighted in **bold**.
- Supports stand-alone operations with intra-CBRS band Carrier Aggregation (CA).
- Can also support additional band combinations with CA anchored to other licensed FDD and TDD bands (e.g. when connected to MNO networks).

As mentioned previously, the CBRS band is a shared spectrum band. There is no guarantee that, at any given point in time, large contiguous blocks of spectrum will be available for use. Note that the SAS will dynamically assign actual CBRS PAL and GAA channels based on priority and availability of channels. Therefore, the carrier aggregation capabilities of both the small cell and the end user devices (in particular, support for non-contiguous Carrier Aggregation) should be considered.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The following figure depicts three different examples of channel bandwidth and Component Carrier (CC) and Carrier Aggregation (CA) combinations:

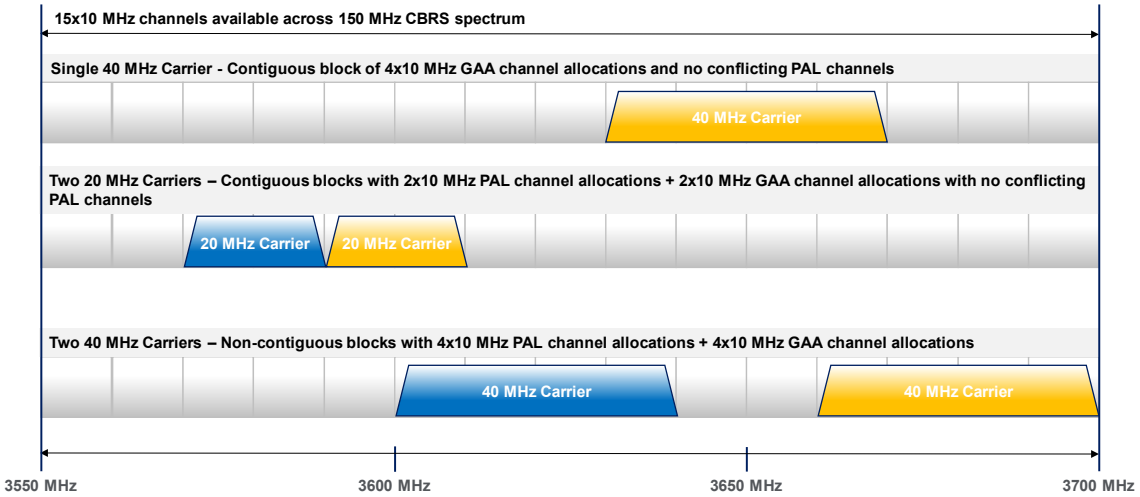


Figure 8 – CBRS Channel Bandwidth, Component Carriers, and Carrier Aggregation

Note that the first example consists of a single CC while the second and third examples consist of two CCs. Three or more CCs may also be assigned based on the capabilities of the gNB and UE. These may either allocate carrier bandwidth to the UEs as separate channels, or allocate paired channels in CA mode for those UEs that support CA. These UEs can take advantage of combining the two carrier bandwidths into one data stream. With reference to the above figure, the following lists three key small cell specifications:

- Operating Frequency Range (OFR) – This is the full range of frequencies which the small cell can operate. In the case of CBRS, the OFR would be 3550-3700 MHz.
- Occupied Bandwidth (OBW) – This is the total carrier bandwidth which the small cell can transmit at any given time. In the above three examples in Figure 8, the OBW would be 40, 40, and 80 MHz, respectively.
- Instantaneous Bandwidth (IBW) – This is the maximum frequency spread / separation (lower most to upper most) that the small cell can transmit simultaneously. In all cases for CBRS small cells, the IBW should be the full 150 MHz so it can support transmitting on the lower most (3550-3560 MHz) to the upper most (3690-3700 MHz) CBRS channels at the same time.

3.5.1. Other Spectrum Bands

The MSOs may continue to evaluate new unlicensed, licensed, or shared / managed spectrum bands:

- Additional spectrum bands may become feasible to add to the MSO's MVNO data offload portfolio (i.e. low-band/below CBRS, sub-6 GHz, mmWave, etc.) which can provide additional overlay / underlay coverage and capacity layers. The availability of these bands will depend on the maturity of the device and network ecosystem (roadmaps), economies of scale (cost of devices), and available network infrastructure to support these additional spectrum bands.
- In terms of 3GPP nomenclature, Frequency Range 1 (FR1) refers to frequency bands operating below 6 GHz while FR2 refers to frequency bands operating above 6 GHz. An example of FR2 would be mmWave spectrum which operates in bands above 28 GHz with the lower unlicensed part of n260 (37.0-37.6 GHz portion of the 37.0-40.0 GHz band) as a future possibility.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.5.2. DOCSIS

The DOCSIS HFC channel line ups are broken into three groups of frequency specific allotments:

- Forward – RF spectrum assigned to deliver downstream traffic from the DOCSIS core to the Customer Premises Equipment (CPE).
- Return – RF spectrum assigned to deliver upstream traffic from the CPE to the DOCSIS core.
- Power – 50 to 60Hz AC power traverses the coaxial cable plant.

The current DOCSIS 3.1 cable plant utilizes 50-60 Hz for the power path. In DOCSIS 3.1, there is a variation between low, mid, and high-split frequencies which are configured based on the upstream data path, summarized as follows:

- Low split is typically 5-42/45 MHz
- Mid split is typically 5-85 MHz
- High split is typically 5-204 MHz

These splits allow upstream throughputs of up to 108 Mbps, 216 Mbps, and 1 Gbps, respectively.

Bonded Single-Carrier Quadrature Amplitude Modulation (SC-QAM) provides 2 or 4 times the upstream capacity, while full Orthogonal Frequency Division Multiple Access (OFDMA) configurations can further increase upstream capacity by 2 times. Corresponding downstream frequency ranges are typically configured from 108-1002 MHz or 258-1218 MHz.

The maximum downstream capacity is 10 Gbps but the current DOCSIS modem hardware utilized in strand small cells limits that speed to 1 Gbps.

MSOs have an advanced DOCSIS network that they've honed over many years to effectively bring high speed internet to consumers. This same network can be utilized for backhaul of small cells. The exciting new advancements in DOCSIS around low-latency active queue management, low latency backhaul, DOCSIS 4.0 extended spectrum Frequency Division Duplexing (FDD), and Full Duplex DOCSIS (FDX) bring further enhancements to support the next generation of wireless.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. 5G Strand Small Cell Engineering

Developing an integrated 5G strand small cell is a complex engineering challenge. There are many design variables that should be considered, (which often compete with each other), in order to develop the optimum and balanced solution.

Everything from physical characteristics and environmental considerations through power, DOCSIS, and antenna design need to be integrated into a compact package that must pass rigorous certification processes (SCTE DOCSIS 3.1 / CableLabs[viii], FCC[ix], and UL[x] to name a few). Once these certifications are completed, the small cell can be confidently deployed into the MSOs HFC network for commercial use.

In particular, the integration of the DOCSIS CM into the small cell requires close collaboration between respective engineering teams (i.e. software and firmware considerations around RF immunity and emissions, heat dissipation, and MSO parametric, outside plant, and DOCSIS requirements). It takes extensive technical expertise and teamwork to make this integration successful.

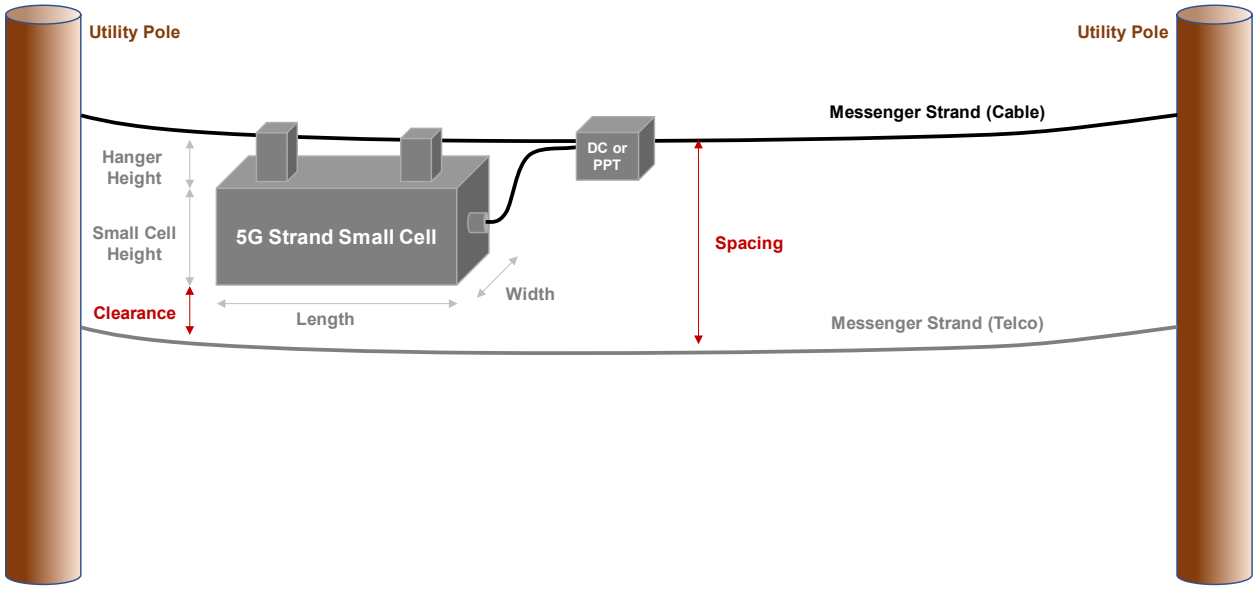
4.1. Form Factor - Physical Characteristics

Since the 5G small cell will be mounted on the aerial HFC cable strand, the following key physical characteristics need to be considered:

- Size – Must meet requirements for the communications space. Must be short and narrow enough to provide clearance to main strand cable bundle above and telco cable level below and have a small enough cross section to minimize the impact of wind loading.
- Weight – Must be light enough for the outside plant (OSP) team to install and maintain. When installed, the small cell must be within the cable tension weight tolerances, including wind loading.
- Mounting brackets – Must be strong enough to handle the load but also be adjustable enough to compensate for small cable sag angles to keep the strand level. Adequate clearance to the main strand cable bundle above and also the telco cable below must be maintained to avoid the possibility of chafing when the strand is swaying in the wind.
- Appearance – Should appear similar to existing HFC strand mounted hardware (such as fiber nodes and amplifiers), to be consistent with HFC outside plant devices.
- Pest resistance – Must avoid “sheltered pockets” and exposed cables where pests can build nests or chew cables. This event could impact the network performance either directly or indirectly as a result of compromised heat dissipation.
- GPS Mounting – Must be mounted clear of main strand bundle and meet Environmental Considerations listed below.
- Additional requirements – Must meet MSO-specific requirements (e.g. connector types, installation guides, local, state, and federal laws and regulations, etc.):
    - Physical distance for service personnel can be maintained so climbing space and RF radiation distances are easily met as the small cells can be mounted away from the pole and other serviceable components.
    - Disconnect switches can be placed in the power path of the small cell to allow easy service shutdown in case of maintenance activities.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The following figure outlines the generic Strand Small Cell Dimensions and Cable clearance and spacing requirements as per N.E.S.C. Rule 235H[xi]:



Separation of Communications
N.E.S.C. Rule 235H

Part of the following material was reprinted with permission from IEEE NESC Safety Code C2-2002, Copyright 2002 *, by IEEE. The IEEE disclaims any responsibility or liability resulting from the placement and use in the described manner.

Current edition of the N.E.S.C. 2002.

235H - Clearance and Spacing Between Communication Conductors, Cables, and Equipment.
1. The spacing between messengers supporting communication cables should be not less than 300 mm (12 in) except by agreement between the parties involved.
2. The clearances between the conductors, cables, and equipment of one communication utility to those of another, anywhere in the span, shall be not less than 100 mm (4 in), except by agreement between the parties involved. **

Figure 9 – Strand Small Cell Dimensions and Cable Clearance

4.1.1. Antenna Considerations

The following key considerations apply to the installation planning and deployment of strand-mounted small cells on the MSO's plant infrastructure:

- Antenna size – The size of a wireless antenna is determined by two basic factors: wavelength and gain. The CBRS frequency band at 3.55 GHz has a calculated estimated wavelength of ~3.33". This small size makes CBRS frequency ideal for supporting higher gain antennas in a small form factor that will fit within the same clearance requirements as any strand-mounted / aerial DOCSIS hardware.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Mechanical – The antennas used for the strand-mounted small cell chassis would be secure, IP67 rated connections and have a radome that protects the antenna elements from dust, water, vibration and shock / impact while being "electrically transparent" so as to not adversely affect the antenna pattern and gain.

The following figure depicts a generic representation of the key physical characteristics (i.e. dimensional considerations) for a simplified single integrated antenna panel / radome. In this example, the antenna panel is mounted on one face of a strand small cell. Note that, depending on the desired electrical characteristics for the strand small cell (i.e. sector/omni, gain, beamwidth, etc.), multiple antenna panels may be mounted at different locations around the strand possibly combined into one or more "antenna arrays" that would provide omni or sector antenna configurations:

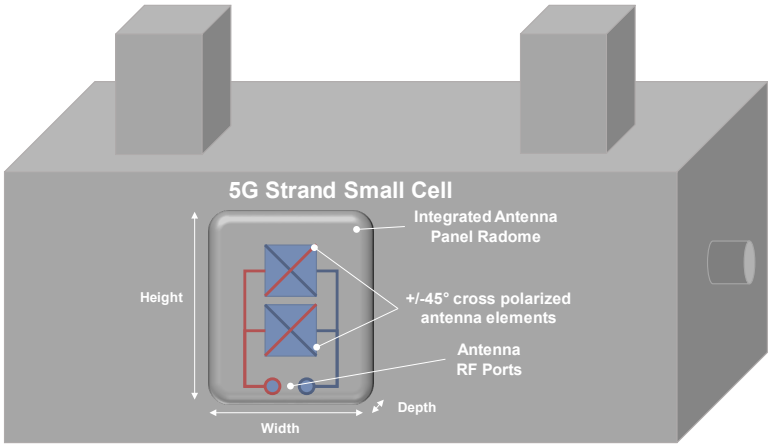


Figure 10 – Wireless Antenna – Physical Characteristics

Another aspect to consider is the mechanical tilting of the strand-mounted small cell. This scenario is depicted in the figure below. Specifically, if the MSO's messenger strand is sagging on a steep enough incline such that the strand small cell "leveling" brackets cannot compensate to make the small cell level (also due to requirement to maintain correct clearance to the telco messenger strand, there will be a forced mechanical tilt such that there will be an "up-tilt" in the direction of the incline, and there will be a steeper "down-tilt" in the downward slope. This will result in the coverage being skewed in both directions, depending on the clutter and obstacles in the horizontal plane. This should be considered during RF network planning in order to select mounting locations that avoid this this scenario.

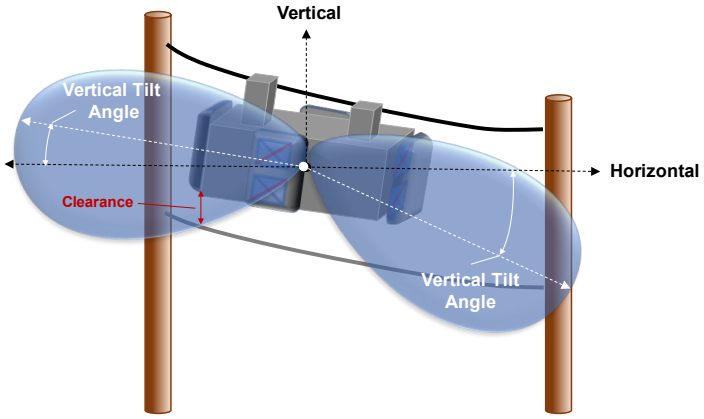


Figure 11 – Mechanical Tilt - Omni Strand Antenna

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.1.2. Environmental Considerations

The following key environmental considerations must be factored into the design of the outdoor strand-mounted small cell exposed to the elements to maintain high reliability / availability (minimize aging):

- Heat dissipation – Must passively dissipate heat at a sufficient rate to regulate the inside temperature of the small cell so the unit does not fail or negatively impact the HFC network, even at high outdoor temperatures (e.g. +55°C).
- Cold start – Must boot up and reliably operate following power outage at below freezing temperatures (typically -40°C).
- IP67 – Must provide 100% ingress protection from dust and water (i.e. salt fog, acid rain tolerance, UV suitability, resistance to infestation).

The following figure depicts the strand small cell internal components that, depending on loading, will impact the internal temperature and required passive heat dissipation to maintain reliable operations across the external temperature range of -40 to +55°C:

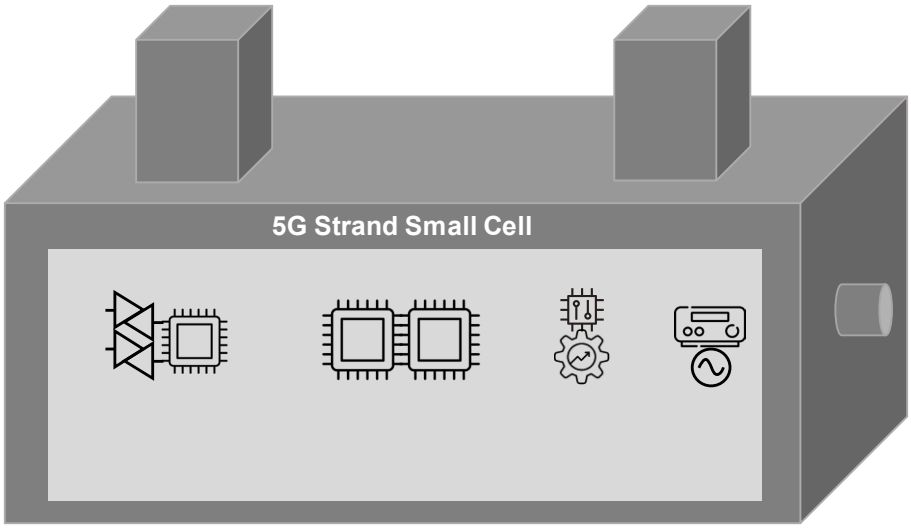


Figure 12 – Strand Small Cell Heat and Loading

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2. Power and DOCSIS

Power is one of the most important design factors for a small cell since it impacts all the key attributes and capabilities of the unit. The small cell with integrated DOCSIS is drawing its power from the existing coax cable plant and should be designed to optimize power consumption while achieving maximum radio performance. One needs to factor in the HFC architecture and various plant considerations to operate within the capabilities and tolerances of the existing infrastructure, to avoid negatively impacting the primary broadband performance, operations, and business model. This is a key consideration to the MSO outside plant team.

Cable broadband power supplies deliver AC power from a utility connection to coaxial cables to power active network elements. This power is typically generated at 90-volts AC in a quasi-square wave form. This network was purposefully designed 40+ years ago so that a licensed electrician was not required, and a low-voltage technician could work on them. A ferro-resonant transformer provides the foundation to the power supply and prevents many disruptions to the network. Often this network is also backed up by batteries that store energy in the event of a utility outage. This makes the HFC network power much more reliable than utility power.

The small cell and all active gear in the HFC plant must be engineered to operate down to 45-volts AC to allow for plant losses. Based upon Ohm's law, the total voltage will decline with resistance as you get further away from the power supply source. Power will also be pulled from the active gear in the upstream coaxial path. The primary point is, based upon how the network is engineered, there is typically a lot of flexibility in where to place the small cell for each individual section of the plant.

The power supply board embedded in the small cell must convert the quasi-square AC power to DC power that can be used by the active components (including the cable modem). Most power supply boards that are used today are based on a switched-mode design. A switched-mode power supply maintains a constant power over a specified input voltage range of the HFC network (45-90 VAC), meaning that as the incoming voltage drops the current pulled will increase.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2.1. Broadband Network Power

There are several advantages to the HFC network power over direct utility power. The first is the simplicity of working with a single vendor; the cable operator. You also don't have to install a meter, wait on additional scheduling, or pay an additional bill. Another advantage is the voltage used in a cable plant complies with the NESC 90 VAC low-voltage standard. This eliminates the requirement for a licensed electrician for the installation.

The design of modern HFC powering systems implement a ferro-resonant transformer with a capacitor tank circuit that provides up to 30 dB of isolation from transient energy events, commonly experienced on direct utility power. Another huge advantage is the ability to connect the power (and communications backhaul) from a single waterproof connection. This may be a hardline or drop cable; the single cable carries all the necessary components to facilitate the connection. Finally, the fact that almost all HFC cable plants in North America leverage a battery backup which provide uninterrupted power over several hours to days without utility connection.

The following figure outlines the main components of broadband network power:

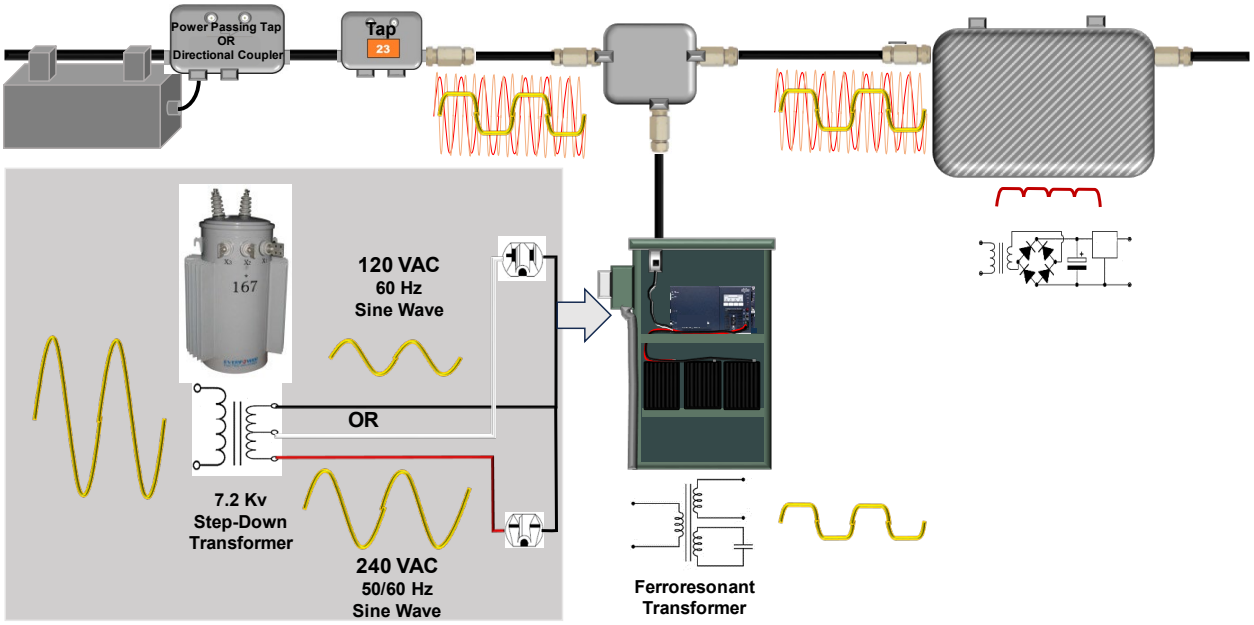


Figure 13 – Broadband Network Power - Overview

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2.2. RF / Plant Considerations

Deployment of RF transmitters and receivers of any kind close to a cable network can be challenging. In the past, any customer who placed their 4G/LTE phone next to a cable box can attest that a high-power transmitter impacts the quality of RF signals.

4.2.2.1. Testing Standards

The SCTE is aware of these situations and has developed many qualification specifications around the cable connected devices to improve the robustness of the DOCSIS network against 4G and 5G wireless networks. As part of this, testing standards have been developed for both, creation of interference to the cable plant, as well as tolerance to interference on the cable plant.

One such standard is the SCTE 40 Digital Cable Network Interface Standard[xii] which defines survivability testing and criteria for cable connected equipment. There is also a testing standard for shielding effectiveness. ANSI/SCTE 48-1 2021[xiii] provides a test method for shielding effectiveness of passives and actives using a Giga Transverse ElectroMagnetic (GTEM) Cell.

There are SCTE LTE field testing recommendations from the SCTE 201 2020 showing how much the potential impact of LTE is considered in cable systems (note that 5G/ NR has essentially the same interference characteristics of 4G/LTE since they both use the same OFDM air interface). Even so, there are additional considerations when placing a >5 W radio next to a hardline or drop cable.

4.2.2.2. RF Isolation

During installation, it is important not to cross the small cell antenna with a drop cable; even quad-shield cable can allow enough energy to propagate to interfere with SC-QAM and OFDM channels. Isolation is an important factor in deploying services. The cable architecture is inherently designed to be isolated from interference, while small cell communications is inherently design to effectively manage wireless interference. However, a potential issue arises when a high-power source is placed close to a point of an ingress to either the cable network or small cell. For example, >5 W of wireless radiated RF signal is significantly higher than the SC-QAM channels running at +40 dBmV in the cable RF signal. Additionally, QAM channels running at +40 dBmV are also much higher than the small cell receiver RF signal levels expected from received handset power levels in the uW range.

4.2.2.3. Diplexing

The cable system has evolved over the years, starting with a simple antenna distribution system to today's networks running multiple split, 2-way plant with distributed CMTS architectures leveraging Converged Cable Access Platform (CCAP) remote Physical (PHY) and remote Media Access Control (MAC) PHY fiber-deep deployments. The DOCSIS 3.1 modem hardware is designed to support the OSP evolution with software adjustable attenuators and electronic multiband diplexers.

The DOCSIS 3.1 CM installed into the strand small cell contain diplexers which isolate and direct the downstream and upstream RF signals. It is extremely important that the CM diplexers support the specific network they will be installed into. As outlined earlier in Section 3.5.2 above, MSOs use different "low", "mid", and "high" frequency splits, summarized as follows:

- Current Mid-split modems typically use upstream frequencies of either 5-42 MHz or 5-85 MHz with downstream frequencies of 108-1002 MHz

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Current High split modems typically use upstream frequencies of either 5-45 MHz or 5-204 MHz with downstream frequencies of 258-1218 MHz

The diplexer frequency ranges are set using the DOCSIS standard leveraging MAC Domain Descriptor (MDD) and type, length, value (TLV) 84 (configuration state of the RF plant) messages sent by the CMTS. Certified DOCSIS modems respond to this command and set their RF front end appropriately.

The following figure summarizes the current (DOCSIS 3.1) and future (DOCSIS 4.0) diplexing splits and upstream data limits:

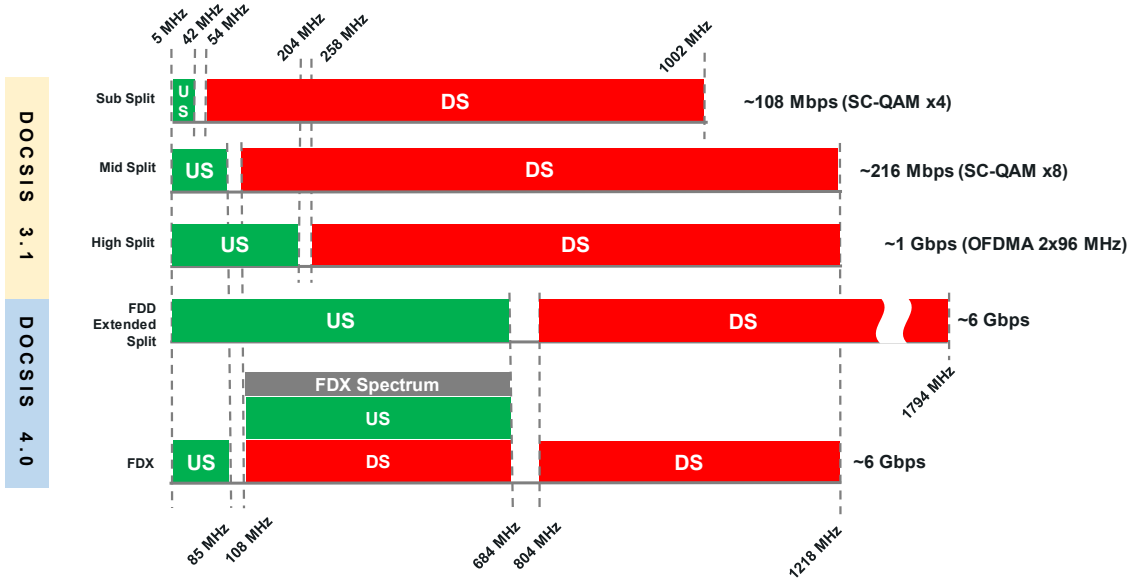


Figure 14 – DOCSIS - Diplex Splits & Upstream Data Limits

4.2.2.4. Power

All but the oldest of these networks leverage an integrated power delivery system used today to provide power to all the HFC equipment. Passive components allow existing HFC deployments to support fiber deep without having to change the existing power infrastructure. Fiber deep architectures, much like small cell, leverage the concept of densification providing additional launch points to improve deployed throughput.

An evolution has occurred over the last decade as improvements in amplifier silicon GASFET and enhancements in the efficiency active power pack supplies used in the HFC equipment has reduced power consumption. However, the deployment of more sophisticated field devices with multiple node points in a single module and the reallocation of the CMTS components to the field have increased power demands. The active field power packs today are mostly based on switch mode designs which vary their current demand as the plant voltage changes to retain constant power. The cable companies understand these variabilities and have dictated that field equipment must operate between 45 VAC and 90 VAC to better control the expected current demands while allowing the expected plant passive losses to allow substantial cable lengths.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The power budget changes over the day and year based on the passive resistance changes due to temperature. As with most wire and electronics, increases in temperature increase passive losses. Leaving headroom in the power budget is critical in assuring reliable plant operation.

Adding equipment like strand-mounted small cells impacts the existing power budgets. Even though HFC power supplies can run close to their rated power indefinitely, both the supplies and cable architecture are more reliable if there is some headroom left for their operation. The important item in adding any of these components is that their impact is more than just their advertised power consumption. The passive loss will impact the voltage of this and all the other actives on the HFC leg, increasing their current demands as well. Therefore, adding a strand small cell device that consumes 95 W at the end of the line (e.g. at 55 VAC), would create a 1.7 AAC draw. Depending on the length of cable, the input voltage could drop the voltage to 47 VAC, increasing its demands to 2 AAC. This would also drop the voltage to all other actives with any cable passive loss, creating an increase in plant current demand.

Most HFC hardline components are designed to operate with up to 15 AAC, with the power supplies themselves operating from 5 AAC to 24 AAC. When adding additional load, through the addition of small cells, upgrading the power system is possible, up to the point you reach the passive current capability. A 15 to 18 A power supply on a single circuit is about as high as possible without having to split the powering to 2 separate circuits. The recommended course of action when reaching close to 15 A load is to upgrade power supplies and/or splitting the demand using additional power sources as required. Additional power delivery improvements can be obtained by the proper use of power feeder cable, with its lower loop resistance.

Today's cable operators are continuously aware of the power demand on their plant through the use of power monitoring systems in each powering location. The powering devices with their embedded cable modems leverage the same DOCSIS networks to immediately notify Network Operation Centers (NOCs) of any change in the utility or HFC powering parameters.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2.3. Directional Couplers and Power Passing Taps

MSO outside plant teams are familiar with using both directional couplers (DCs) and power passing taps (PPTs). They will consider site location, power budget, and existing plant design to determine when to use a DC or PPT to install the small cell.

4.2.3.1. Directional Couplers

Hardline Directional Couplers (DCs) are devices that allow for a percentage of RF power and the AC power line to be coupled to another leg or device on the plant. Their use requires equipment capable of handling 15 A of power, though a pin to F or BAFF[xiv] connector can allow the use of an F connector, if desired. There are many "Mini Macro" small cell deployments that leverage a DC. Those deployments use larger radios which could pull up to 300 W. The hardline connections used with DCs require the same level of complexity as anything else in the hardline architecture. Large tools like coring tools and wrenches are required in their installation. The forming of the hardline to the KS connections is also labor intensive. Finally, removing or replacing a connected device can be time consuming.

Installation of a DC normally requires cutting of the existing hardline, which may mean the shutdown of the power system to avoid shorts. The DC installation time is minutes to hours as all connections should be complete before restoring AC to the HFC plant.

The following figure depicts the use of DCs to power stand-mounted small cells:

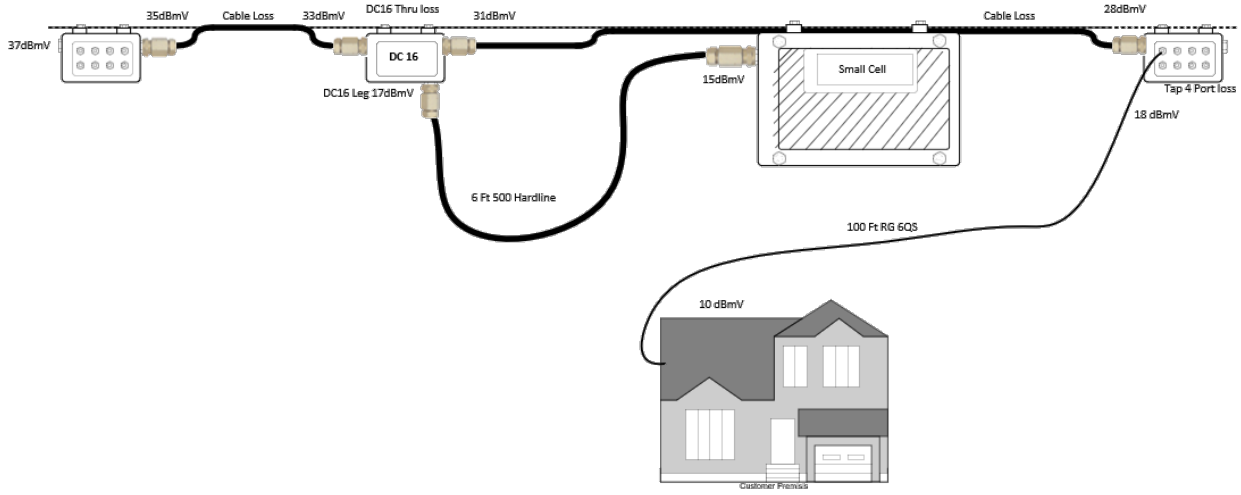


Figure 15 – Directional Couplers for Small Cells

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2.3.2. Power Passing Taps

Power passing taps (PPTs) are another common form of small cell connection to a cable plant. These devices enable the conversion of hardline to drop cable (typically RG 6QS) but are limited to 2 AAC or less. This is well within the 5 A current capability of RG 6 cable and F connectors, but limits the small cell to a 95 W (45 VAC*2 AAC) load and further limits their transmitters to that load.

Note that MSOs need to budget power distribution available to each strand small cell. Therefore, strand small cells connected to power passing taps typically operate below the maximum allowable Category B CBRS EIRP limitations (i.e. EIRPs of < 47 dBm / 10 MHz). However, deployments using power passing taps are quicker and less labor intensive than the hardline DC installations:

- In most cases, an existing tap location is leveraged as the system can run up to 100' of drop cable to the small cell.
- These existing tap faces are replaced with a power passing tap.
- The power passing shunts are installed to allow power to the port where the small cells will be connected, and the drop cable is connected to the small cell.
- The last part of installation is to connect the drop cable to the power passing tap port to assure the center conductor does not short to the body or F connector.
- The shunts may also be placed after the connection is made, but that is not typical.

Interruption in the HFC plant power and RF may only be experienced as the tap face is replaced, normally only a few seconds to minutes.

In both cases, the passive and actives of the cable plant are populated to pass AC to the point the small cell makes connection. This may require opening the field devices and installing shunts.

The following figure depicts the use of PPTs to power stand-mounted small cells:

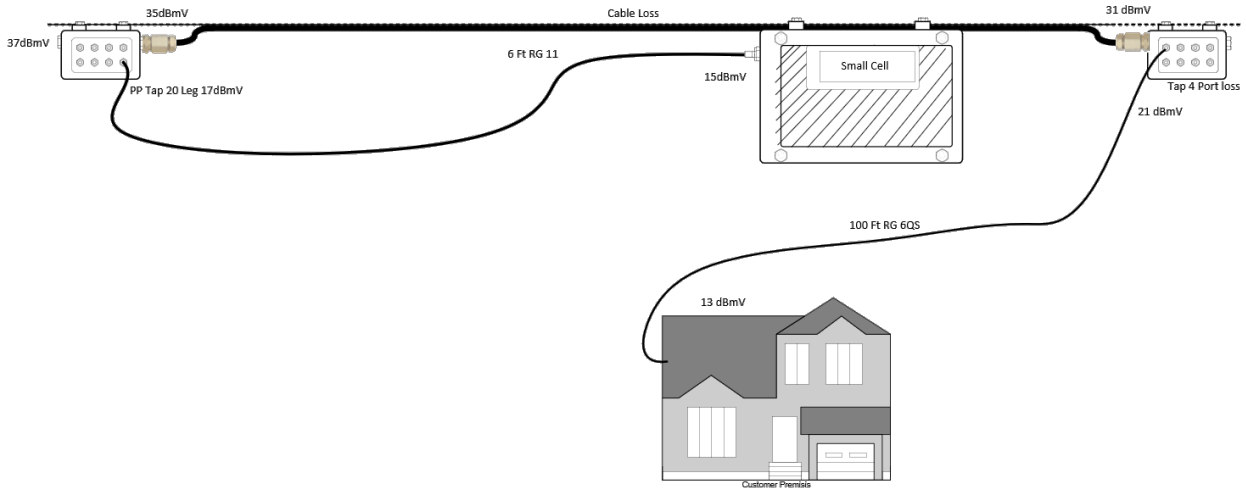


Figure 16 – Power Passing Taps for Small Cells

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2.3.1. Directional Coupler and Power Passing Tap Comparisson

The following table summarizes and compares the key considerations when designing and implementing the DC and PPT power distribution for strand-mounted small cells:

Table 1 – Directional Coupler and Power Passing Tap Considerations

Consideration	Directional Coupler	Power Passing Tap
HFC Components	DC, 1 to 2 Hardline adapters, Heat Shrink	Power Passing Tap Face
HFC Time	30 min to over 1 hour	5 to 15 min
DS RF Plant impacts	2 to 4 dB	0 dB
Potential DS changes	up to 30 additional minutes	0 minutes
HFC power	up to 300W	<100W
HFC Engineering Required	Typical, power, data, RF DS affects may have to be addressed US and DS.	Little, mostly power, data
Drop Cable to Small Cell	420 – 500 hardline, 2 hardline adapters, coring, assembly and heat shrink	RG 6 or RG 11 - Crimp

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2.4. DOCSIS Backhaul

MSO's can leverage their HFC network for supplying both power and backhaul on a single cable to the small cells that comprise the HMNO network. Integrating a DOCSIS CM into the small cell provides the backhaul while integrating a switched-mode power supply will enable the small cell to draw power from the DOCSIS plant via the cable connected to the DOCSIS CM port.

Currently the DOCSIS 3.1 CM is the mainstream standard for backhaul with available throughput of 1 Gbps downstream and 1 Gbps upstream. Next generation DOCSIS 3.1 solutions will add available throughput to achieve up to 2.5 Gbps downstream. As MSOs upgrade their HFC networks to the new DOCSIS 4.0 standard with new splits, the legacy DOCSIS 3.1 CMs can still be supported while new / upgraded strand small cell variants could be added with the improved throughput and features that DOCSIS 4.0 provides.

Since the small cell shares the DOCSIS network with personal and business broadband subscribers, considerations should be made when dimensioning the number of node-splits and devices per cable run. Specifically, the added DL and UL requirements for the strand small cell will require new DOCSIS dimensioning rules to protect existing users while increasing service levels and also ensuring enough bandwidth is allocated. Allocation is based on both the broadband and wireless traffic models (e.g. busy hour (BH)) and utilization for personal, business, and MVNO data offload.

4.2.5. Firmware Considerations

Integrating a DOCSIS CM with a small cell comes with unique challenges. Merging two distinct technologies and interoperability testing (IOT) requires complex engineering knowledge of both technologies. The following summarizes the main considerations:

- Event driven actions that signify when systems go up/down (for example, transferring knowledge from the cable modem to the small cell as to when the modem has either rebooted or lost and re-gained the DOCSIS link).
- The need to maintain backhaul capacity and reliability, including high throughput and low latency. This requires fine tuning of the DOCSIS channels, combined with modifications to modulation profiles and management between the headend CMTS and the CM to ensure that data is sent upstream / downstream in a timely manner required to maintain the wireless network.
- The cable modem software undergoes routine updates as updated requirements from various MSOs arise. This is further necessitated by the continuous forward progress and improvements from the DOCSIS eco-system. Different firmware versions across the same product line can cause variations in the behavior and incremental firmware versions need to be managed to ensure consistent and aligned firmware interoperability between the CMTS, CM, and small cell.
- The DOCSIS architecture leveraged in this solution provides a secure method of updating the DOCSIS cable modems using the "private" cable modem space. All the firmware images are keyed using a manufacturer CVC or CVC chain which should both match the target modem and the firmware file to be applied. The cable operator can further secure the deployment of firmware by requiring a cosigned image so any unapproved firmware would be ignored by the small cell cable modem.
- An integrated CM and small cell enables the possibility of establishing a common protocol between the 2 different platforms. They may share an integrated controller (I2C) channel which can aid with relaying messages, diagnostics, and/or initiating state changes (e.g. reboot or re-initialization of the various Open Systems Interconnection (OSI) layers depending on internal diagnostics).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.3. Antennas

The antennas used for the strand small cell are an essential component of the overall system design as they define the coverage area for each deployed strand small cell and therefore, directly impact the site density, power and backhaul distribution, and resulting business case for the MVNO data offload deployment.

4.3.1. Omni and Sector

There are two fundamental antenna array configurations for strand small cells; omni and sector. Greenfield deployments of cellular networks often start with omni configurations since they provide simplicity with planning and focus on the initial goal of coverage. As the networks mature and require additional capacity, sector configurations can automatically double the data offload in high-density AOIs. As part of the planning, deployment, and operations of an MSO's HMNO network, the MSO should consider both omni and sector configurations to balance the operational simplicity of omni with the capacity benefits of sector antenna configurations. The following table summarizes the considerations between omni and sector antenna configurations:

Table 2 – Omni and Sector Antenna Considerations

Consideration	Omni	Sector
Form Factor	Antennas need to consider the physical and electrical obstruction of the strand small cell chassis (i.e. near field effects that impact resulting pattern)	Antennas can face outwards, away from the strand small cell chassis with minimum impacts of near field effects that cause ripple and nulls in antenna pattern
Coverage	1 sector per small cell, covering all directions	2 or more sectors per small cell with overlapping zones to support handovers between adjacent sectors
Capacity	1 cell capacity for all mobile users in shared geographic coverage area of the small cell	2 or more geographic zones, each with 1 cell capacity so total small cell can support up to double the capacity, depending on the geographic split of mobile users across each zone's coverage area
Planning	Simplified planning and deployment since no antenna azimuth	More complex planning since strand small cell orientation needs to consider the azimuth orientations of the antenna sectors
Optimization	Limited flexibility for changing coverage with main "adjustment knob" being changing RF power levels	More flexibility for changing coverage since have 2 or more RF power levels to adjust

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

For comparison, the following figure depicts representative horizontal antenna patterns for generic Omni and 2-sector antenna arrays:

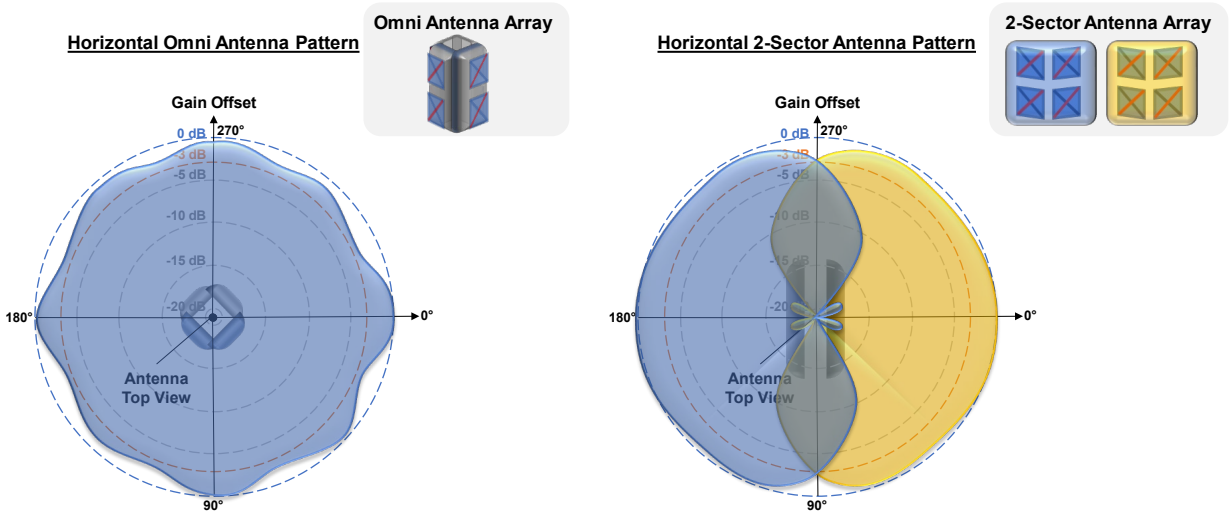


Figure 17 – Wireless Antenna – Omni and Sector

4.3.2. Electrical Characteristics

In order to facilitate reliable coverage and high user throughput (both on the downlink (DL) and the uplink (UL)), the small cell antenna system need to consider the following key factors:

- Gain –Provide enough RF signal boost in both the DL and UL direction such that the resulting Signal to Interference and Noise Ratio (SINR) calculated from the "link budget" provides strong enough signal (at the receiver ends of both the mobile device for DL and the strand small cell for UL) to cover the distance between the strand small cell and the mobile subscriber with a defined minimum "quality of service" in terms of DL + UL throughput and payload capacity for served mobile subscribers that offload to the MSO's MVNO data offload network.
- Antenna Pattern / Beamwidth – Omni antenna arrays should provide consistent gain (low "ripple") in all directions from the strand (360° for omni antenna arrays) while sector antenna arrays should provide consistent gain over a wide horizontal angle from the strand and have sufficient overlap to support smooth handover (HO) between adjacent sectors (i.e. overlapping 2x180° 3dB beamwidth for 2 sectors, 3 x 120° for 3 sectors, 4x90° for 4 sectors, etc.). Due to the nature of antennas (focusing radiated or received RF power), the wider horizontal beamwidth necessitates a much narrower vertical 3 dB beamwidth (typically +/- 10°) so the resulting focused antenna pattern provides sufficient gain and coverage in all directions of the horizontal plane while also providing enough gain in the vertical / elevation angle to provide DL and UL coverage boost to nearby vertical structures and/or terrain (i.e. hills / dips / buildings, etc.).
- Ripple – To be effective, an antenna array pattern should maintain a consistent gain in all desired horizontal directions (typically within 3 dB of the maximum antenna gain). This will facilitate reliable coverage in the desired directions. In the case of an omni configuration, RF planning and strand deployment teams can be assured of meeting the desired coverage, regardless of the small cell orientation. In the case of a desired sector configuration, RF planning and strand deployment teams can be assured of meeting the desired coverage in the targeted directions of each sector.
- Downtilt – Wireless antennas can also be mechanically or electrically down-tilted such that the resulting antenna pattern (focusing RF power) is directed vertically downwards by a small

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

amount (typically 5° down for street level coverage). This is done to help contain / control the RF power being radiated horizontally from the antenna to minimize the overlapping interference zones with adjacent / nearby small cells.

The following figure summarizes the key electrical characteristics of an omni antenna array:

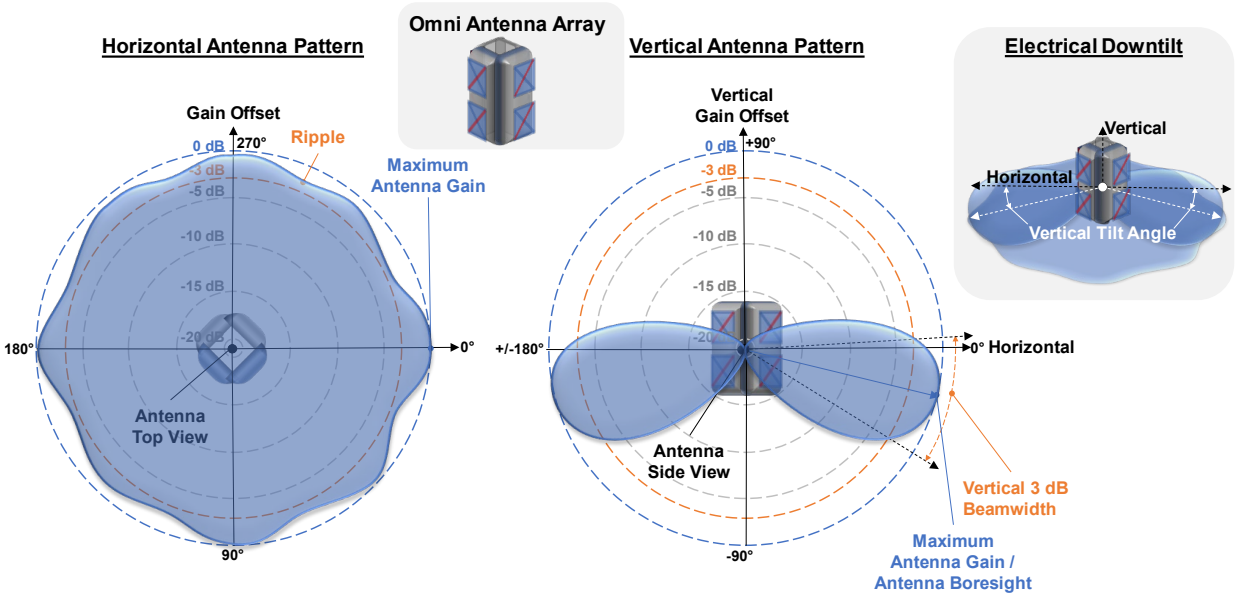


Figure 18 – Electrical Characteristics - Omni Antenna Array

The following figure summarizes the above electrical characteristics of a sector / panel antenna:

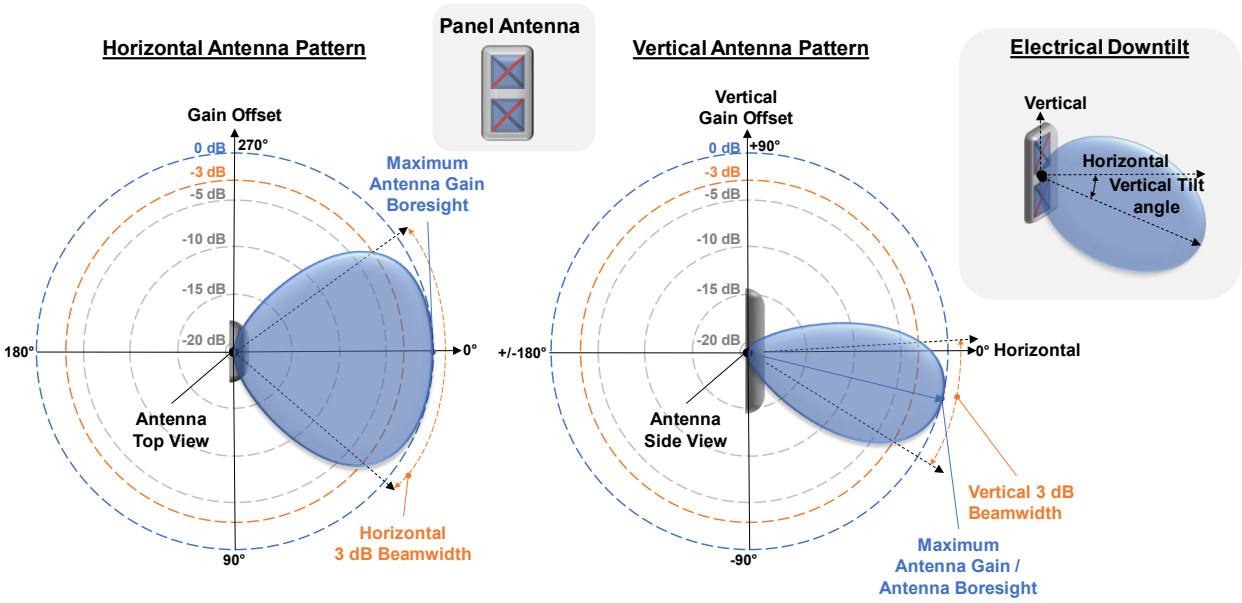


Figure 19 – Electrical Characteristics - Sector / Panel Antenna

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Antenna panels may be mounted in the same electrical plane as the small cell chassis to adhere to the physical size and separation requirements of the strand-mounted small cell. In these cases, multiple antenna panels may be mounted around the strand in an "array" such that the resulting antenna pattern is referred to as "pseudo-omni". The byproduct of the array is the horizontal antenna spacing affects the amount of "ripple" caused by adjacent antenna panels that interfere with each other when they are out of phase.

4.3.2.1. RF Chain (2T2R and 4T4R)

To improve the reliability and consistency of a wireless RF signal, the antenna array in the strand small cell radio will have at least two RF paths between the mobile device and small cell transmitters and receivers. This provides redundancy of the RF signal since, in a wireless environment, there are many reflections and absorptions of an RF signal that result in a phenomenon called "fading" (the RF signal strength will dip and significantly reduce or drop the connection between the mobile device and small cell radio):

- 2T2R provides this redundancy with significant improvement in compensating for fading effects on one path by a technique called "cross-polarization" (orienting the "antenna elements" in the antenna array such that one path is at +45° rotated from the vertical axis of the small cell radio and the second path is at -45° rotated from the vertical axis of the small cell radio). The cross-polarization technique creates an orthogonal (90°) separation of the collected / radiated antenna array RF signal which results in a significant (typically >30 dB) "discrimination" between both paths such that their RF signal strengths are not correlated, thereby allowing the receiver to choose the stronger signal in case one path between the transmitter and receiver has a degraded / weakened signal due to "fading". In addition to diversity of the signal, 2T2R provides the possibility of using MIMO techniques to obtain higher throughput in favorable conditions. In reality, there is some variance on practical gains which is dependent on the "fading environment" the strand small cell radio and mobile device are located. For example, in areas of "clutter" (i.e. buildings, vehicles, trees, houses, street furniture, etc.), there will be a higher chance of fading so the 2T2R transmit and receive diversity will be focused on combating fading rather than using MIMO layers for increased data throughput capacity.
- Employing a 4T4R antenna is more complex design than 2T2R, however, it does offer the added benefit of increasing the robustness of 4-way transmit / receive diversity that will further improve fading immunity while also offering increased possibility of higher throughput due to up to 4 layers (unique signals).
- In summary, the tradeoffs between 2T2R and 4T4R are:
    - 2T2R is simpler and therefore, has higher component reliability while 4T4R is more complex,
    - 4T4R requires 2x the antenna elements, hence more physical real estate,
    - 4T4R may have higher power consumption due to the additional components such as RF power amplifiers, and more power hungry processor(s).
    - 4T4R does offer the possibility of higher throughput and more fading immunity in RF environments with high clutter.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The following figure summarizes the key concepts of the RF chain / "plumbing" for a representative xTxR small cell RF chain:

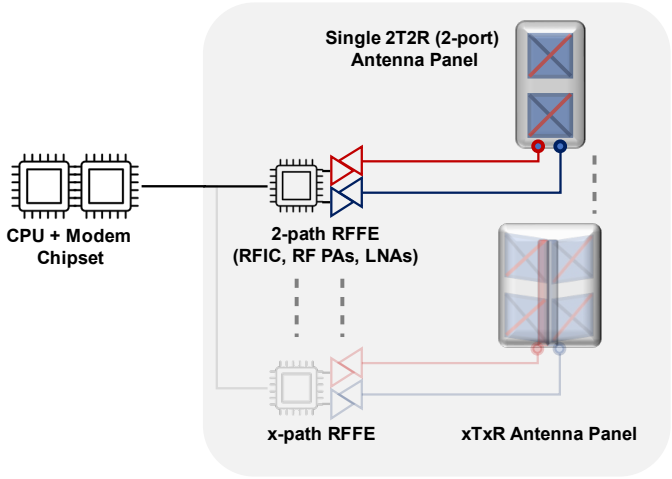


Figure 20 – xTxR Small Cell RF Chain

4.3.2.1. Effective Isotropic Radiated Power (EIRP)

Based on the total conducted RF power, RF chain and antenna configuration and gain, the EIRP is the measured "focused" RF power (typically in dBm units) at the maximum gain / "main boresight" of the antenna. It is as if the conducted RF power directed in that particular direction were radiated in all directions. Since antennas essentially focus radiated RF power from a conducted RF power source, the EIRP reflects the effective power measured after antenna gain (in dBi) above a theoretically isotropic / spherical radiated RF source. The following figure summarizes the EIRP concept and calculation:

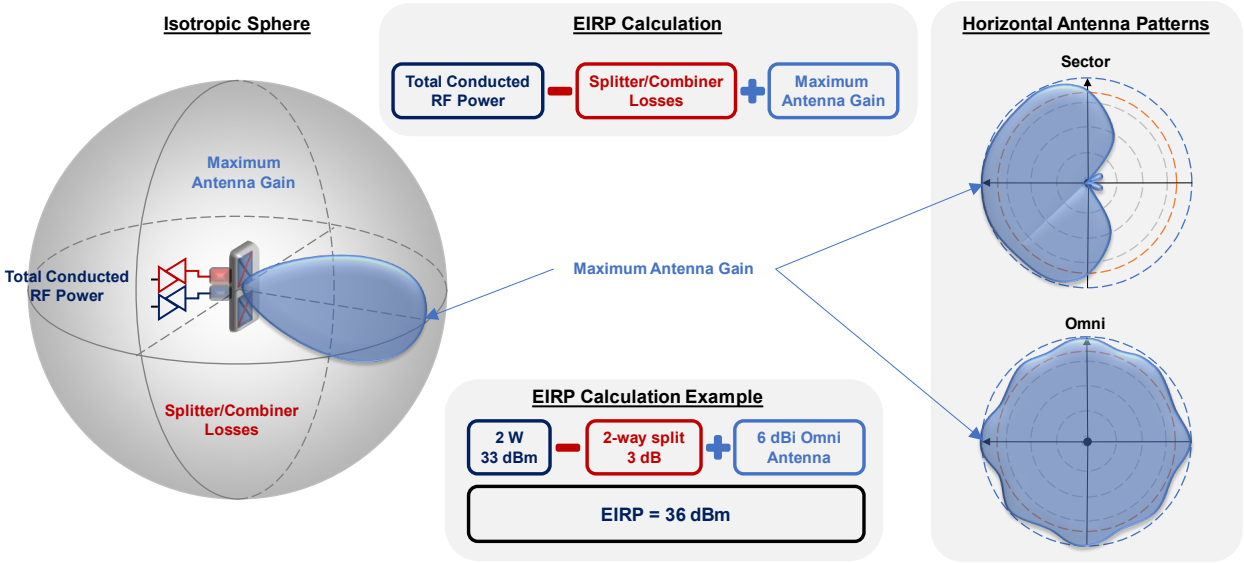


Figure 21 – Effective Isotropic Radiated Power Concept

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In the case of CBRS:

- Two different RF power levels / EIRP categories are available as follows:
  - Category A (indoor CBSDs) – The maximum EIRP must be no more than 30 dBm / 10 MHz
  - Category B (outdoor CBSDs or CPEs) – The maximum EIRP must be no more than 47 dBm / 10 MHz
- All CBRS transmitters above 6 feet height above average terrain (HAAT), and hence effectively all outdoor strand-mounted small cells, must be Category B certified. However, strand-mounted small cells will typically operate below the maximum allowable EIRP of 47 dBm / 10 MHz due to power distribution limitations, as covered in the Sec 4.2.3.

Considering the fundamental EIRP calculation outlined above, the following figure shows the variations on EIRP calculations, depending on the context and how the FCC and CBRS define EIRP and the application of EIRP for use in RF planning through Link Budgets:

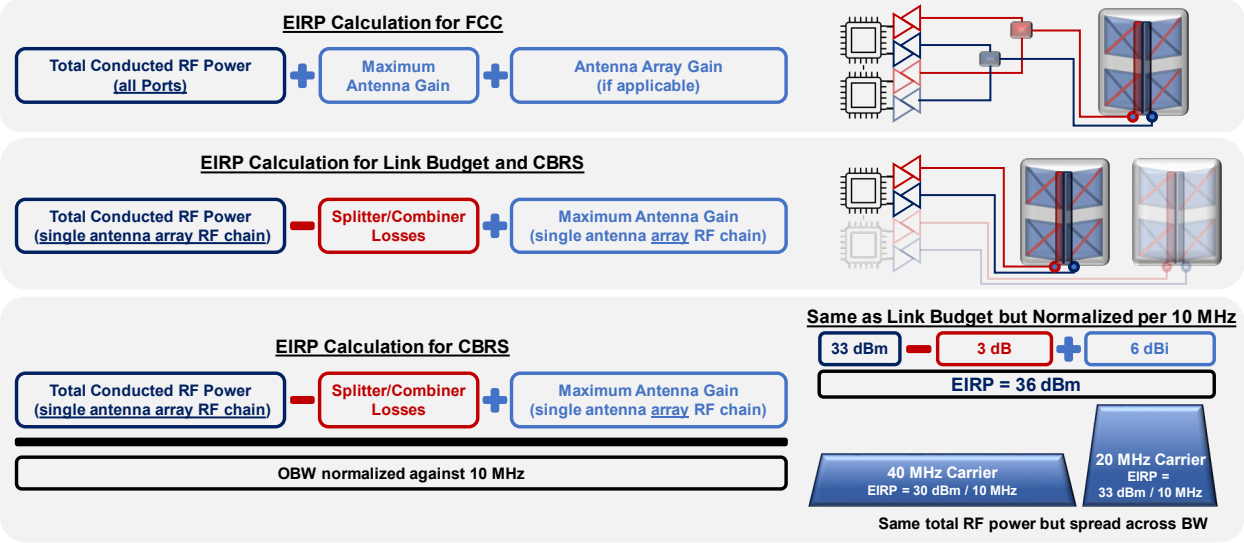


Figure 22 – EIRP Calculation Comparison – FCC, Link Budget, and CBRS

Note the similarities but also the main difference between FCC, Link Budget, and CBRS is how the conducted RF power is distributed in terms of RF chains / splits in antenna arrays and spreading RF power over bandwidth. Essentially, FCC certification means the maximum power spectral density that a CBSD can possibly radiate if all RF power is focused in one direction is within the EIRP limits of the appropriate CBSD Category. The goal of the Link Budget EIRP is to determine the actual coverage based on the RF conducted power focused in a given direction (also used for CBRS SAS planning and channel grants).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Network Design and Optimization

There are several key factors when considering the design, planning, and optimization of wireless coverage and capacity for an MSO MVNO data offload network, including the following:

- Spectrum – This provides the foundation for the HMNO network. Depending on the frequency band and available bandwidth, this will determine both the coverage and capacity capabilities of each small cell, aggregated to the HMNO network level for total network coverage and capacity. In addition to targeted CBRS usage areas, the spectrum assets may include other non-CBRS bands for added capacity.
- AOIs – A typical strategy for the planning and deployment of strand small cells for the MVNO data offload network is to target locations where current MSO MVNO mobile data users are generating the highest traffic. These are typically around public venues (i.e. corporate campuses, shopping malls, sports complexes, universities, etc.). To maximize the return on investment (ROI) and ensure a healthy business case for total cost of ownership (TCO) of the MVNO data offload network, the strand small cells would be deployed in a targeted way.
- Wireless network / RAN architecture – This may consist of a mix of macro, micro, and femto indoor and outdoor radios and Distributed Antenna Systems (DAS) using one or more Radio Access Technologies (RATs). These types of networks are referred to as multi-RAT or Heterogeneous Networks (HetNets) which may include a combination of Wi-Fi and a mix of 3GPP and/or non-3GPP wireless access networks. When correctly designed and optimized, the MSO can leverage the various wireless technology layers and combine them to support robust mobility for their subscribers.
- Mobility – While AOIs may be concentrated areas, they may also be high pedestrian or vehicular traffic zones that would require the strand small cells to hand over (HO) to ensure a smooth mobile user experience. This may include leveraging HetNet / multi-RAT HMNO network capabilities to ensure seamless handovers (HOs) between the networks to provide reliable network experience for the end users.
- Quality of Service (QoS) – This defines the MSO's targeted MVNO data offload capacity that will enable the mobile user experience in terms of throughput and latency (both DL and UL). Several factors impact the QoS and resulting cell and network level capacity for the wireless MVNO data offload network: bandwidth, signal strength, and interference. Depending on the DL/UL link budget (small cell specifications, RF environment, etc.), the resulting signal to interference and noise ratio (SINR) will determine which MCS can be used to maximize the throughput over the air interface. Higher SINR results in higher order MCS that provides more information bits and less forward error correction (FEC) / overhead. For example, a 64 Quadrature Amplitude Modulation (QAM) MCS "constellation" will provide lower throughput but higher FEC compared with 256 QAM since 64 QAM has a larger error vector magnitude (EVM) tolerance.
- DL/UL link budgets – The downlink and uplink coverage / serving area of the strand small cell is determined by the link budget parameters, including:
    - Wireless transmit / receive specifications of the strand small cell and mobile device, with the following two key strand small cell specifications:
        - EIRP for downlink,
        - Receiver sensitivity for uplink,
    - RF environment (i.e. morphology / clutter),
    - Antenna heights of the strand small cell (typically 20 ft.) and mobile user / user equipment (UE) (typically 3 ft.),

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Security – Considering CPEs, other small cells, and other plant equipment are all sharing the same DOCSIS network, IP security (IPsec) tunnels can be established to maintain proper isolation between user and control-plane traffic and virtual private network (VPN) tunnels. Security certificates should be efficiently managed such that the connectivity is reliable.

5.1. Coverage and Capacity Planning

As outlined above, these factors go hand-in-hand in determining the resulting coverage and capacity of the strand small cell network. Depending on the coverage area, QoS, and mobility requirements, the gNB inter-site distance (ISD) can be determined. ISD reflects the average distance between adjacent strand small cells in the MVNO data offload network.

The following figure provides a representative example of how the above-listed factors interact with each other to compare two generic link budgets and resulting coverage/capacity profiles that reflect both uplink limited and downlink limited scenarios:

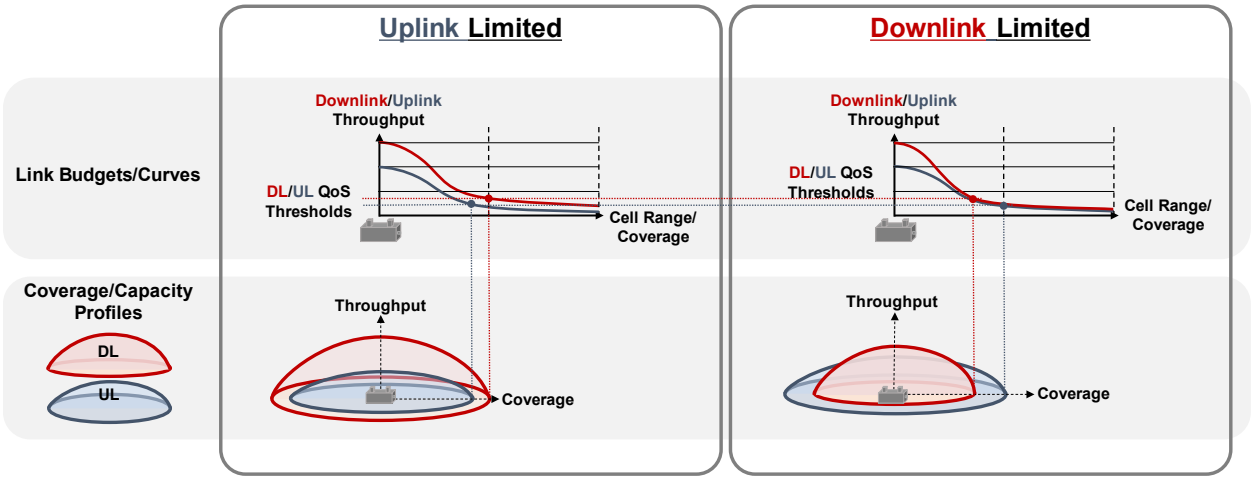


Figure 23 – Wireless Coverage and Capacity

Note that the above coverage exercise is based on the minimum throughput required by a single user at the cell edge. Depending on the combined user traffic demand, the serving area of each small cell may need to be further reduced which would result in increased cell density and smaller ISD.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.2. Mobility

In addition to coverage and capacity, a key attribute and planning requirement for the MSO to carefully consider as part of network planning and deployment for effective MVNO data offload in their HMNO network is the mobility / handover locations and frequencies between the MSO's HMNO network and the MNO's network and is supporting the MVNO. The following figure provides a simplified visualization on the mobility of a user travelling in one direction and handing between the MSO's HMNO and MNO's network:

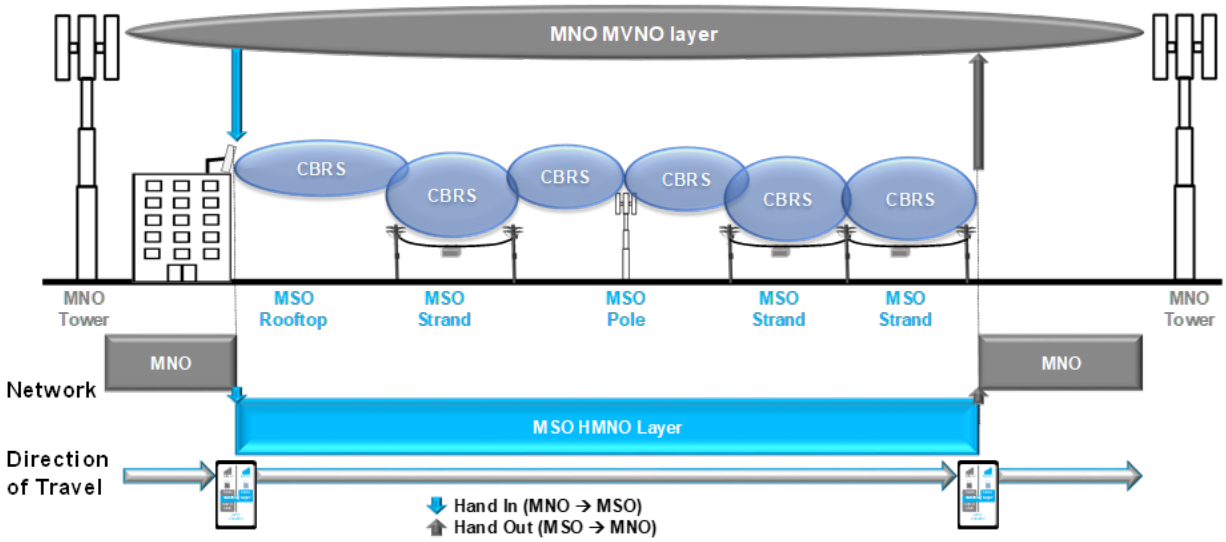


Figure 24 – MNO to MSO MVNO Data Offload – Mobility

The concepts referenced earlier in section 5 (i.e. Spectrum, Network architecture, QoS, ISD,etc.) should be considered as part of Mobility. In addition, the following should also be considered to ensure optimum mobility performance of the HMNO network:

- Neighbors / Adjacent cells / HO Parameters – These should be correctly defined to ensure correct HO and mobility between the different cells.
- Traffic prioritization – Since small cell and broadband traffic will both be sharing the same DOCSIS downstream and upstream resources and also Operations and Maintenance (O&M) control traffic will be sharing the same medium (DOCSIS and wireless), it is important to be able to prioritize traffic layers to ensure reliable network performance for the end user experience.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Deployment and Operations

As outlined previously, the MSO should consider many factors when choosing where to deploy small cells (i.e. user density, power budget, RF engineering, coverage area, etc.). Additionally, operators face the following deployment challenges when deploying small cells:

- The small cell's location will not always coincide with the availability of utility power or fiber for backhaul.
- Zoning, site acquisition, permitting and regulations can delay installations by months.
- Project schedules can be significantly delayed due to delays getting meters installed by the local utility.
- Utility power is prone to outages and disturbances.
- High-speed, low-latency backhaul may be cost-prohibitive to deliver if construction is required.

The major deployment challenges listed above are overcome when the MSOs utilize their existing HFC infrastructure to deploy strand-mount small cells. An important feature that ensures smooth deployment and operations of the MSO's strand small cell network is plug and play (PnP) functionality. As the name implies, both DOCSIS and small cell parameters and features are pre-configured to make the field deployment and operations and maintenance of the strand small cell network more efficient.

Once the strand small cell sites are chosen, the MSO can deploy their Outside Plant (OSP) technicians to install the small cell. The OSP technicians must also be Certified Professional Installer (CPI) certified based on WInnForum accreditation, in order to meet FCC Part 96 rules for Category B CBSD deployment. The CPI certification ensures that the key CBSD installation and configuration parameters (i.e. small cell location, antenna height, EIRP, antenna type (omni / sector), azimuth, etc.) are verified and loaded into the strand small cell and SAS in order to ensure the SAS database accurately reflects the deployed field installation of the strand small cell. This is critically important since the SAS must accurately manage the channel and frequency coordination of channel grants and releases to minimize interference between the MSO's strand small cell CBSDs and other operator's neighboring CBSDs.

Besides CPI certification and additional power and backhaul planning considerations, the MSO OSP teams have the advantage to operate and deploy their new strand small cell network freely, following similar operating procedures as per their mainstream broadband strand mounted HFC network. In particular, the OSP team can efficiently coordinate deployment and operations of both their broadband HFC network and overlaid HMNO network since they do not need additional permissions, permits, leasing, zoning or worry about pole attachments. Once the strand small cells are installed, powered up, and commissioned, they are ready to integrate into the HMNO network to support the MVNO data offload in the AOIs they are deployed.

As commercial MVNO data offload traffic starts to take off in the MSO's new HMNO network, the MSO should also consider the ongoing capacity planning to account for increasing traffic growth on both the DOCSIS and HMNO network and maximize the data offload performance. For the radio access network, this will involve optimizing the channel bandwidth, frequencies, power levels, and handover parameters. For the DOCSIS network, this will involve optimizing plant power and HFC parameters and network architecture that influence downstream / upstream channel capacity. In both cases, the wireless and broadband network architecture evolution and supporting equipment roadmaps will be considered to achieve optimum performance to meet subscriber needs and the MSO business case.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

7. Conclusion

This paper is an effort to provide a comprehensive, yet efficient, compilation of the various concepts that the MSO community should consider as part of adding an HMNO network. This paper addresses both Cable and Wireless network engineering considerations including:

- Optimizing power consumption for maximum performance
- Supporting flexible cable modem frequency splits across different MSOs
- Preventing ingress / egress / spurious noise to isolate DOCSIS coax RF from wireless RF
- Antenna design, form-factor, and gain
- Wireless coverage and capacity planning
- Utilizing optimum radio network components to maximize data offload performance

It is hoped that this paper has achieved the goal; to provide a comprehensive guide for the MSOs to refer to as they move forward into expanding their respective convergence portfolios to include MNVO data offload. Through leveraging the MSO's HFC network (including right-of-way, power, and backhaul), the MSO now has new 5G strand small cell technology available to them as an enabler to further expand their wireless HMNO network portfolio to maximize their MNVO data offload capabilities.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

2T2R	2 transmit 2 receive [2 RF paths for [small cell] radio DL+UL]
3GPP	3rd Generation Partnership Project [standards organization]
4T4R	4 transmit 4 receive [4 RF paths for [small cell] radio DL+UL]
4G	4th Generation 3GPP standard [encompasses RAN and core]
5G	5th Generation 3GPP standard [encompasses RAN and core]
AAC	Amps Alternating Current
AF	Application Function
AMF	Access & Mobility Management Function [3GPP nomenclature for 5G core network element managing control-plane functionality]
ANSI	American National Standards Institute
AOI	area of interest
AUSF	Authentication Server Function
B48	4G-LTE frequency band for CBRS (3550-3700 MHz)
BH	busy hour
bps	bits per second
CA	Carrier Aggregation
CC	Component Carriers
CBRS	Citizens Broadband Radio Service allocated to 3550-3700 MHz band
CBSD	Citizens Broadband Radio Service Device
CCAP	Converged Cable Access Platform
CM	cable modem
CMTS	cable modem termination system
CPE	customer premises equipment
DAS	Distributed Antenna System
dB	decibel [used for relative comparison of RF units on logarithmic scale = 10 * log (ratio of RF unit)]
dBi	dB isotropic [logarithmic unit to define the relative gain of an antenna]
dBm	dB milliwatts [logarithmic unit for RF conducted or radiated power = 10 * log (RF power level in milliwatts)]
dBmV	dB millivolts [logarithmic unit for millivolts]
DC	directional coupler [depending on context]
DC	data center [depending on context]
DL	downlink
DN	Data Network [operator services, internet access, other services]
DOCSIS	Data Over Cable Service Interface Specification
DSDS	dual-SIM dual-standby
EIRP	Effective Isotropic Radiated Power
EMS	element management system
ESD	extended spectrum DOCSIS
EVM	Error Vector Magnitude
FCC	Federal Communications Commission
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FDX	Full Duplex DOCSIS
FEC	forward error correction

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

FR1	Frequency Range 1 [3GPP standard nomenclature for sub-6 GHz]
FR2	Frequency Range 2 [3GPP standard nomenclature for above-6 GHz]
GAA	General Authorized Access
Gb	Gigabit
Gbps	gigabits per second
GB	Gigabyte
gNB	5G Node B [3GPP standards radio]
GTEM	Giga Transverse ElectroMagnetic
HAAT	height above average terrain
HetNet	Heterogeneous Network
HFC	Hybrid Fiber Coaxial
HMNO	Hybrid Mobile Network Operator
HO	Handover
I2C	Inter Integrated Controller
IBW	instantaneous bandwidth
IoT	Internet or Things
IOT	inter-operability testing
IP67	Ingress Protection 6 (dust) 7 (liquids)
IPsec	Internet Protocol Security
ISD	inter-site distance
LTE	Long Term Evolution [4th generation 3GPP air interface standard]
MAPL	Maximum Allowable Path Loss
Mb	megabit
Mbps	megabits per second
MB	megabyte
MCS	Modulation and Coding Scheme
MDD	MAC Domain Descriptor
MHz	megahertz
mmWave	millimeter wave spectrum [3GPP refers to as FR2 - above 6 GHz]
MNO	Mobile Network Operator
MSO	Multiple System Operator
MVNO	Mobile Virtual Network Operator
n48	5G-NR frequency band for CBRS (3550-3700 MHz)
NEF	Network Exposure Function
NGC	Next Generation Core [5th generation 3GPP core]
NOC	Network Operation Center
NR	New Radio [5th generation 3GPP air interface standard]
NRF	NF Repository Function
NSSF	Network Slice Selection Function
O&M	Operations and Maintenance
OBW	occupied bandwidth
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OFR	operating frequency range
OSI	Open Systems Interconnection
OSP	outside plant
PAL	Priority Access License
PCF	Policy Control Function

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

PHY	physical layer
PnP	plug and play
PPT	power passing tap
PSD	power spectral density
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RAN	Radio Access Network [3GPP terminology for wireless network architecture]
RAT	Radio Access Technology
RF	radio frequency [depending on context, may refer to wireless or cable]
RFFE	RF front end [the transmitter / receiver block of a [small cell] radio]
ROI	return on investment
SAS	Spectrum Access System
SCA	Software Communications Architecture
SC-QAM	Single-Carrier Quadrature Amplitude Modulation
SCTE	Society of Cable Telecommunications Engineers
SDR	Software Defined Radio
SDS	Software Defined Systems
SIM	Subscriber Identity Module [3GPP standard nomenclature]
SINR	Signal to Interference and Noise Ratio
SMF	Session Management Function
SRA	Software Radio Architecture
TCO	total cost of ownership
TDD	Time Division Duplexing [shared DL + UL using single frequency]
TLV 84	type, length, value 84 [configuration state of the RF plant]
UDM	Unified Data Management
UE	User Equipment [3GPP terminology for mobile device]
UL	uplink [depending on context]
UL	Underwriter Laboratories [depending on context]
UPF	User Plane Function [3GPP nomenclature for 5G core network element managing user-plane functionality]
uW	microwatts
VAC	Volts Alternating Current
VPN	Virtual Private Network
WInnForum	Wireless Innovation Forum [non-profit wireless standards organization]
WISP	Wireless Internet Service Provider

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

[i] Reference: https://www.fcc.gov/auction/105
[ii] Cable providers using existing network to build wireless cheaper, say panelists; Fogarty; S&P Global; 10/15/2020; <https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/cable-providers-using-existing-network-to-build-wireless-cheaper-say-panelists-60753719>
[iii] Source: https://www.ncta.com/broadband-facts
[iv] Source: https://www.3gpp.org/technologies/5g-system-overview
[v] Reference: https://ongoalliance.org/
[vi] Reference: https://www.wirelessinnovation.org/
[vii] Reference: https://www.3gpp.org/
[viii] Reference: https://www.cablelabs.com/
[ix] Reference: https://www.fcc.gov/
[x] Reference: https://www.ul.com/
[xi] References: http://www.ojua.org/wp-content/uploads/2009/03/rule-235h-history.pdf and https://standards.ieee.org/wp-content/uploads/import/documents/interpretations/ir523.pdf
[xii] Reference: https://www.scte.org/standards/library/catalog/scte-40-digital-cable-network-interface-standard/
[xiii] Reference: https://www.scte.org/standards/library/catalog/scte-48-1-test-method-for-measuring-shielding-effectiveness-of-passive-and-active-devices-using-a-gtem-cell/
[xiv] Definition: BAFF = Industry standard term for hardline to F, female connector

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The New Explosion of Social Engineering: Defensive Techniques to Manage the Risk

A Technical Paper prepared for SCTE by

Abdul Saleem
Development Engineer 4
Comcast India Engineering Center | LLC
Chennai One SEZ, Module 7 & 8, 5th Floor, North Block Phase II,
200 Feet Pallavaram – Thoraipakkam Road, Thoraipakkam, Chennai - 600097
Abdul_saleem@comcast.com

Poornasakthi Sivaraman
Senior Manager, Cybersecurity
Comcast India Engineering Center | LLC
Chennai One SEZ, Module 7 & 8, 5th Floor, North Block Phase II,
200 Feet Pallavaram – Thoraipakkam Road, Thoraipakkam, Chennai - 600097
poornasakthi_sivaraman@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Importance of Addressing Social Engineering in Cybersecurity ... 3
3. Types of Social Engineering Attacks ... 4
4. Psychology Behind Social Engineering ... 5
   - 4.1. Understanding human vulnerabilities and behaviors ... 5
   - 4.2. Identifying deceptive websites or applications ... 6
   - 4.3. Psychological tactics and persuasive techniques employed by social engineers. ... 6
5. Real-Life Examples of Social Engineering Attacks ... 7
   - 5.1. Notable case studies and incidents ... 7
   - 5.2. Impacts and consequences of successful social engineering attacks ... 8
6. Common Red Flags and Indicators of Social Engineering Attacks ... 9
   - 6.1. Recognizing suspicious emails, messages, or phone calls ... 9
   - 6.1. Identifying deceptive websites or applications ... 10
   - 6.2. Behavioral cues and manipulation techniques employed by social engineers ... 11
7. Case Studies of Successful Defense against Social Engineering Attacks ... 12
   - 7.1. Examples of organizations that effectively prevented SE incidents ... 12
   - 7.2. Strategies and measures they employed to protect against social engineering ... 12
   - 7.3. Mandate security awareness and education ... 13
   - 7.4. Implementation of technical measures ... 14
8. Future Trends and Emerging Challenges in Social Engineering ... 15
   - 8.1. Evolving tactics and techniques used by social engineers. ... 15
   - 8.2. Impact of technology advancements (e.g., AI) on social engineering ... 15
9. Conclusion ... 17

Abbreviations ... 19

Bibliography & References ... 19

List of Figures

Title **Page Number**

Figure 1 - Types of Social Engineering ... 5
Figure 2 - Varieties in Social Incidents ... 8
Figure 3 - Threat Actors' Motives and Varieties in Social Engineering Breaches ... 10

List of Tables

Title **Page Number**

Table 1 – Region-wise Cybercrime Incidents on Social Engineering ... 17

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Social engineering [SE] refers to the manipulation and exploitation of human psychology and behavior to deceive individuals or gain unauthorized access to sensitive information, systems, or physical spaces. It involves the use of psychological tactics and persuasive techniques to trick people into disclosing confidential information, performing actions that they wouldn't normally do, or granting access to restricted areas.

SE techniques can be employed through various mediums, including in-person interactions, phone calls, emails, instant messaging, or social media platforms. The primary objective is to exploit human vulnerabilities, such as trust, helpfulness, curiosity, fear, or ignorance, to achieve the attacker's goals. SE is a deceptive and manipulative technique used by malicious individuals or attackers to exploit human psychology and behavior. It involves tricking people into divulging sensitive information, performing actions they wouldn't normally do, or granting unauthorized access to systems or physical spaces. Unlike traditional hacking methods that primarily focus on exploiting technical vulnerabilities, SE targets the human element, taking advantage of our natural tendencies and emotions.

The core principle behind SE is the recognition that humans can be the weakest link in security systems. No matter how robust an organization's cybersecurity measures are, a skilled social engineer can find ways to bypass them by manipulating individuals through psychological tactics. By understanding human behavior, social engineers exploit factors such as trust, curiosity, fear, helpfulness, or ignorance to achieve their objectives.

In this paper we will know about the SE attacks that depend on the attacker's ability to gather information about their targets through methods like reconnaissance, research, or social media profiling. This allows them to customize their approaches and make their attempts more convincing and effective. Awareness and education play a crucial role in defending against SE. Individuals and organizations should stay informed about common SE techniques, regularly update their knowledge of potential threats, and implement security measures such as strong passwords, two-factor authentication (2FA), and employee training programs. By fostering a culture of security awareness and vigilance, individuals and organizations can better protect themselves against SE attacks.

2. Importance of Addressing Social Engineering in Cybersecurity

Addressing social engineering is of paramount importance in the field of cybersecurity. SE refers to the manipulation of individuals to gain unauthorized access to sensitive information or to perform malicious activities. It exploits human psychology, trust, and emotions rather than relying solely on technical vulnerabilities. Here are some key reasons why addressing SE is crucial in cybersecurity:

- **Human Weaknesses:** SE exploits inherent vulnerabilities in human behavior, such as curiosity, trust, and the tendency to be helpful. Attackers leverage these weaknesses to trick individuals into divulging sensitive information, clicking on malicious links, or performing actions that compromise security. By addressing social engineering, organizations can raise awareness among employees and equip them with the knowledge and skills to recognize and resist such manipulations.

- **Insider Threats:** SE attacks can be initiated by insiders who have legitimate access to an organization's systems and data. These individuals, whether intentionally or unintentionally, can be manipulated into performing actions that compromise security. By focusing on social

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

engineering, organizations can mitigate the risk posed by insider threats and reduce the likelihood of internal breaches.

- **Complementary to Technical Measures:** While technical measures like firewalls, antivirus software, and encryption are crucial for cybersecurity, they are not foolproof. Attackers often target the weakest link in the security chain, which is often the human element. SE attacks can bypass technical defenses by exploiting human trust or by impersonating legitimate entities. Combining technical measures with SE awareness and training creates a more robust and holistic security posture.

- **Phishing and Spear Phishing:** Phishing attacks, a form of social engineering, involve sending fraudulent emails or messages that appear to be from reputable sources, enticing recipients to provide sensitive information or perform actions that compromise security. Spear phishing takes this a step further by tailoring the attack to specific individuals or organizations, increasing the likelihood of success. By addressing social engineering, organizations can effectively combat phishing and spear phishing attempts, protecting sensitive data and preventing financial losses.

- **User Awareness and Education:** A critical aspect of addressing SE is educating users about the techniques used by attackers, common red flags to look out for, and best practices to follow. By providing comprehensive cybersecurity training, organizations can empower their employees to make informed decisions, recognize potential SE attacks, and respond appropriately. This creates a security-conscious culture and helps build a human firewall against SE threats.

- **Reputation and Financial Loss:** Falling victim to SE attacks can have severe consequences for organizations. It can lead to data breaches, financial loss, damage to reputation, loss of customer trust, and legal implications. By proactively addressing social engineering, organizations can reduce the risk of such incidents and protect their assets, reputation, and bottom line.

3. Types of Social Engineering Attacks

- **Phishing:** This involves sending fraudulent emails, messages, or websites that appear to be from reputable sources. The attackers aim to trick recipients into providing sensitive information such as passwords, credit card details, or account credentials.

- **Pretexting:** In this technique, attackers create a fictional scenario or pretext to trick individuals into revealing information or performing actions they would not typically do. This can include impersonating a colleague, Information Technology (IT) support personnel, or a trusted authority figure to gain trust and extract information.

- **Baiting:** Attackers offer something enticing, such as a free Universal Series Bus (USB) drive or a promotional item, with malicious software or code embedded. When individuals fall for the bait and use the compromised device, the attacker gains access to their system**.**

- **Tailgating:** Also known as "piggybacking," this technique involves someone following closely behind an authorized individual to gain access to a restricted area without proper authentication.

- **Spear phishing:** A targeted form of phishing attacks where attackers customize their messages to deceive specific individuals or organizations into revealing sensitive information. Unlike traditional phishing, spear phishing involves thorough research on the target to create convincing messages that appear to come from trusted sources. By exploiting human vulnerabilities through

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

personalization, these attacks aim to trick recipients into clicking on malicious links, opening infected attachments, or divulging sensitive data. Protection measures, such as being cautious of unsolicited emails and implementing security protocols, are crucial in mitigating the risks associated with spear phishing.

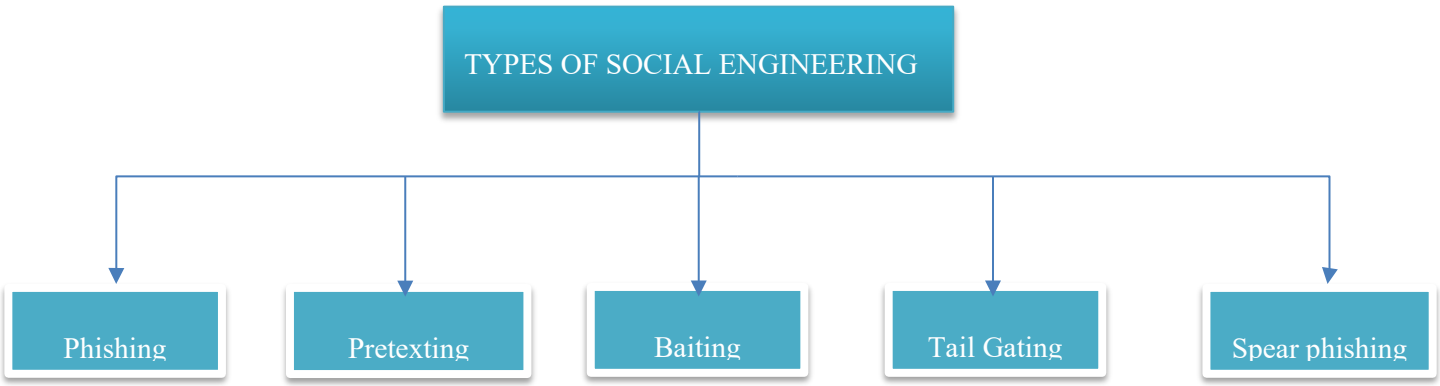


Figure 1 - Types of Social Engineering

4. Psychology Behind Social Engineering

4.1. Understanding human vulnerabilities and behaviors

Understanding human vulnerabilities and behaviors is a fundamental aspect of the psychology behind social engineering. Social engineers exploit psychological principles to manipulate individuals into divulging sensitive information or performing actions that may compromise security.

Humans are inherently social creatures with a natural inclination to trust and help others. Social engineers exploit this tendency by leveraging techniques such as authority, urgency, familiarity, and reciprocity to gain victims' trust and compliance. They may impersonate trusted figures or use persuasive language to create a sense of urgency or importance.

Additionally, social engineers exploit common cognitive biases and heuristics that affect decision-making. These biases include the need for social approval, the desire to avoid conflict or punishment, and the tendency to rely on mental shortcuts rather than engaging in careful analysis.

Another crucial vulnerability is the lack of awareness about cybersecurity risks. Many individuals are unaware of the tactics used in SE attacks and may not recognize the signs of manipulation. This lack of awareness makes them more susceptible to SE techniques.

To combat social engineering, it is essential to educate individuals about these vulnerabilities and train them to recognize and respond appropriately to suspicious requests or manipulative tactics. Building a culture of security awareness and promoting critical thinking can help individuals become more resilient against SE attempts.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.2. Identifying deceptive websites or applications

Social engineers employ a range of strategies to exploit human vulnerabilities and carry out successful SE attacks. One of these vulnerabilities is trust, as humans tend to trust individuals or organizations that appear authoritative or familiar. Social engineers take advantage of this trust by impersonating trusted entities, manipulating victims into divulging sensitive information.

Another vulnerability that social engineers exploit is curiosity. They use intriguing subject lines or messages that pique curiosity, such as urgent or exclusive content. By arousing curiosity, they entice individuals to click on malicious links or open infected attachments, leading to potential security breaches.

In addition to trust and curiosity, social engineers leverage fear to manipulate their targets. They create scenarios that induce panic or urgency, such as threatening legal consequences or claiming compromised accounts. By exploiting fear, they override rational thinking, pushing individuals into making hasty decisions that compromise security.

Furthermore, social engineers take advantage of people's natural helpfulness and cooperation. They often pose as colleagues or IT support personnel, requesting assistance to gain access to confidential information or compromising actions. By appearing friendly and in need of help, they manipulate individuals into unwittingly assisting their malicious intentions.

Lastly, social engineers capitalize on the ignorance of individuals regarding cybersecurity risks and SE techniques. Exploiting this lack of knowledge, they design attacks that deceive victims into revealing sensitive information or clicking on malicious links without realizing the potential consequences. Their success hinges on the victim's unawareness of the tactics being employed against them.

4.3. Psychological tactics and persuasive techniques employed by social engineers.

Social engineers employ a diverse range of psychological tactics and persuasive techniques to manipulate individuals and achieve their desired outcomes. These tactics can be categorized into several key strategies:

Firstly, social engineers often rely on authority to establish credibility and gain compliance. By posing as figures of authority, such as managers, IT personnel, or law enforcement officers, they exploit the inherent trust people have in these positions. This trust allows them to persuade individuals to follow their instructions without questioning or doubting their legitimacy.

Another common tactic is reciprocity, where social engineers offer something of perceived value to their targets. This could be a small favor, a compliment, or even a seemingly innocuous gift. By creating a sense of indebtedness, they leverage the natural human inclination to reciprocate, increasing the likelihood of the target providing the requested information or taking the desired action.

Scarcity is another powerful tactic utilized by social engineers. They create a sense of urgency or scarcity around the situation, such as claiming limited availability or impending deadlines. By instilling a fear of missing out, they manipulate individuals into making impulsive decisions or taking immediate actions to avoid perceived losses or missed opportunities.

Social engineers also exploit the concept of social proof, where people tend to follow the behavior of others, especially in uncertain situations. To establish credibility, they may provide false social proof,

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

such as fabricated testimonials, positive reviews, or endorsements. By showing that others have already complied or endorsed their requests, they encourage the target to follow suit, assuming that it must be the right thing to do.

Familiarity is another tactic employed by social engineers to lower the target's defenses. By gathering information from publicly available sources, such as social media, they create personalized messages that reference shared connections, interests, or even personal details. This familiarity creates a false sense of trust and makes individuals more susceptible to manipulation.

Finally, emotional manipulation is a potent tactic used by social engineers to cloud judgment and override rational thinking. They evoke strong emotions such as fear, excitement, sympathy, or curiosity to push individuals into divulging sensitive information or taking actions they would not otherwise consider. By exploiting these emotions, social engineers manipulate individuals on an emotional level, bypassing their logical reasoning.

Understanding these tactics is crucial in recognizing and mitigating social engineering attacks. By being aware of the psychological strategies employed by social engineers, individuals can better protect themselves and their sensitive information from manipulation and exploitation.

5. Real-Life Examples of Social Engineering Attacks

5.1. Notable case studies and incidents

The LinkedIn Spear Phishing Attack (2016):
Attackers sent spear phishing emails to LinkedIn users, masquerading as legitimate connection requests. The emails contained malicious attachments that, when opened, installed malware and allowed attackers to gather user credentials.

The Twitter Bitcoin Scam (2020):
Attackers compromised high-profile Twitter accounts, including those of Barack Obama, Elon Musk, and Bill Gates. They posted tweets promoting a Bitcoin scam, asking followers to send cryptocurrency to a specified address.

The SolarWinds Supply Chain Attack (2020):
Attackers compromised SolarWinds, a software company, and inserted a backdoor into their software updates. When organizations downloaded and installed the tainted updates, the attackers gained access to their networks, leading to widespread data breaches.

The COVID-19 Vaccine Phishing Campaigns (2020-2021):
Attackers capitalized on the pandemic, sending phishing emails pretending to offer information or access to COVID-19 vaccines. Victims were tricked into providing personal information, financial details, or clicking on malicious links or attachments.

Attack on Twilio (2022):
Attackers gained access to private customer and employee account information by stealing an employee's password. This was done through a broad-based SE attack that involved sending fake IT text messages to Twilio employees.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

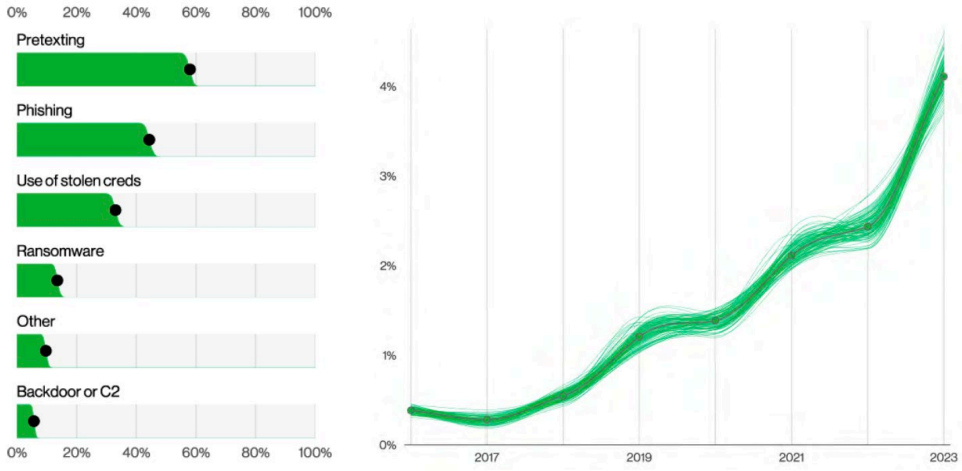


Figure 2 - Varieties in Social Incidents

5.2. Impacts and consequences of successful social engineering attacks

Data breaches: SE attacks often lead to data breaches, where sensitive information such as personal data, financial records, or intellectual property is compromised. This can result in financial losses, identity theft, reputational damage, and legal consequences**.**

Financial loss: SE attacks can result in financial losses for individuals and organizations. Attackers may gain access to banking information, initiate unauthorized transactions, or deceive individuals into transferring funds to fraudulent accounts.

Reputational damage: When an organization falls victim to a SE attack, its reputation may suffer. Breaches and compromises can erode trust and confidence among customers, partners, and stakeholders, leading to a loss of business opportunities and damaged relationships.

Operational disruption: Successful SE attacks can disrupt normal operations, causing downtime and loss of productivity. Attackers may gain unauthorized access to systems, compromise critical infrastructure, or spread malware that affects the organization's networks and services.

Regulatory and legal implications: SE attacks may result in violations of data protection regulations and legal requirements. Organizations could face penalties, fines, and legal actions if they fail to adequately protect sensitive information or if they are found to be non-compliant with applicable laws.

Psychological and emotional impact: Individuals who fall victim to SE attacks may experience psychological and emotional distress. They may feel violated, anxious, or vulnerable due to the invasion of their privacy and the potential consequences of the attack.

Loss of trust and confidence: SE attacks can erode trust and confidence not only within the targeted organization but also among the broader public. People may become skeptical of online communications, hesitate to share information, and become more cautious in their interactions, affecting overall trust in digital platforms and communications.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Common Red Flags and Indicators of Social Engineering Attacks

6.1. Recognizing suspicious emails, messages, or phone calls

Recognizing suspicious emails, messages, or phone calls is crucial in identifying potential SE attacks. Here are common red flags and indicators to watch out for:

Unexpected or unsolicited communication: Be cautious of emails, messages, or phone calls that you were not expecting or did not initiate. SE attackers often reach out to individuals who are not anticipating contact to catch them off guard.

Urgency or pressure: Beware of communications that create a sense of urgency, demanding immediate action or threatening negative consequences if you don't comply. Social engineers often use time-sensitive language to prompt hasty decision-making.

Requests for sensitive information: Be wary of any requests for personal, financial, or sensitive information, such as passwords, social security numbers, or credit card details. Legitimate organizations typically do not ask for such information via email or unsolicited phone calls.

Poor grammar or spelling errors: Many SE attempts originate from non-native English speakers or automated systems, resulting in grammar or spelling mistakes in their messages. Unusual phrasing or frequent errors can be indicators of a suspicious communication.

Unusual sender or caller information: Pay attention to the email address, domain, or phone number from which the communication originates. Social engineers may use slight variations or spoofed addresses that resemble legitimate sources. Be cautious if the sender's or caller's information seems off or unfamiliar.

Requests for money or financial transactions: Exercise caution when receiving requests for money transfers, wire transfers, or unexpected invoices. Verify such requests through alternative means, such as contacting the sender directly using known contact information.

Suspicious links or attachments: Avoid clicking on links or downloading attachments from unfamiliar or suspicious sources. Hover over links to reveal their true destination before clicking and be cautious of file attachments that you were not expecting or that have unusual file extensions.

Unusual or unexpected content: Be wary of emails or messages that contain unusual content or that seem out of character for the sender. Social engineers may attempt to exploit personal or professional relationships by imitating someone you know or by referencing specific events or information.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

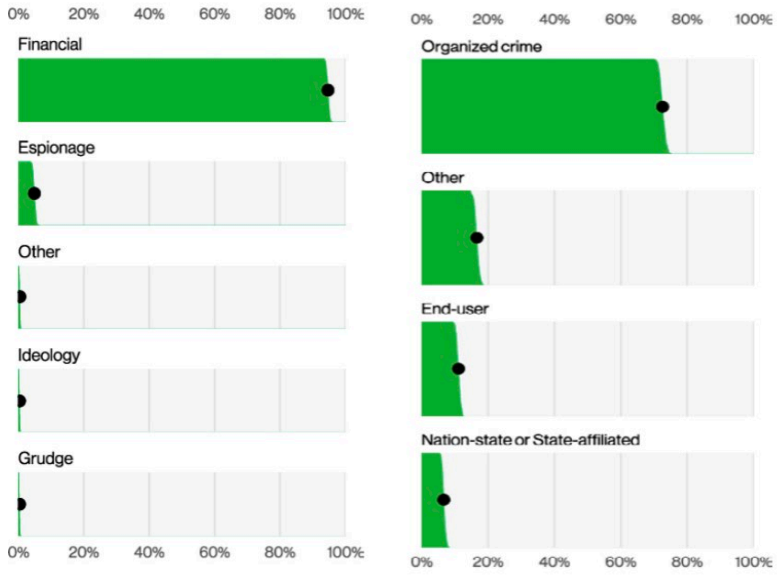


Figure 3 - Threat Actors' Motives and Varieties in Social Engineering Breaches

6.2. Identifying deceptive websites or applications

Identifying deceptive websites or applications is crucial for protecting yourself from SE attacks. Here are some indicators to help you identify potentially deceptive platforms:

Uniform Resource Locator (URL) inconsistencies: Check the website's URL for any misspellings, additional characters, or unusual domain extensions that deviate from the legitimate domain. Deceptive websites often use slight variations to imitate trusted sites.

Poor design and quality: Be cautious of websites or applications with low-quality design, numerous grammar and spelling errors, or an unprofessional appearance. Legitimate platforms typically invest in creating a polished and error-free user experience.

Requests for excessive personal information: Exercise caution if a website or application asks for an unusually large amount of personal or sensitive information that seems unnecessary for the provided service. Only provide information that is relevant and necessary.

Lack of secure connections (Hypertext Transfer Protocol (HTTP) vs. Hypertext Transfer Protocol Secure (HTTPS)): Look for websites or applications that use secure connections indicated by "https://" at the beginning of the URL. Deceptive platforms may use the non-secure "http://" protocol, putting your data at risk during transmission.

Untrusted sources: Download applications or software only from trusted sources such as official app stores or reputable developers. Avoid downloading from unknown or third-party websites, as they may host malicious or counterfeit applications.

Negative reviews or ratings: Check user reviews, ratings, and feedback about the website or application. Negative reviews, complaints about security issues, or reports of suspicious activities are warning signs that should raise concerns.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Unusual behavior or prompts: Be cautious of unexpected pop-ups, excessive requests for permissions, or prompts to install additional software when using a website or application. Such behavior may indicate malicious intent.

Lack of contact information or customer support: Legitimate websites and applications typically provide clear contact information and accessible customer support. If a platform lacks these, it could be a sign of a deceptive website.

6.3. Behavioral cues and manipulation techniques employed by social engineers

Social engineers employ various behavioral cues and manipulation techniques to deceive individuals and manipulate their actions. Here are some common tactics used:

Building rapport and trust: Social engineers often establish rapport and gain trust by engaging in friendly conversations, showing empathy, and finding common interests. This helps create a sense of familiarity and makes individuals more susceptible to manipulation.

Exploiting curiosity and urgency: Social engineers exploit human curiosity by creating scenarios or offering information that piques the target's interest. They may also create a sense of urgency, emphasizing the need for immediate action to prevent potential harm or gain compliance.

Leveraging authority and hierarchy: Social engineers may pose as authority figures or use impersonation to gain compliance. By asserting their position of power or influence, they exploit individuals' natural inclination to follow instructions from higher-ranking individuals.

Exploiting fear and intimidation: Social engineers use fear-based tactics to manipulate individuals. They may create a sense of impending danger or consequences, leading individuals to act impulsively without considering the potential risks.

Creating a sense of reciprocity: Social engineers often initiate small favors or acts of kindness to establish a sense of reciprocity. This creates a psychological obligation in the target's mind, making them more likely to comply with subsequent requests.

Preying on helpfulness and empathy: Social engineers take advantage of individuals' natural inclination to be helpful or empathetic. They may pose as someone in need or exploit sympathy to gain assistance or access to sensitive information.

Exploiting social norms and authority bias: Social engineers manipulate individuals by appealing to social norms and exploiting authority bias. They may present requests as something that is expected or normal within a given context, making it harder for individuals to question or refuse.

Using information gathering and pretexting: Social engineers gather information about their targets through various means, such as online research or pretexting. They then use this information to customize their approach, making their requests seem more legitimate and credible.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

7. Case Studies of Successful Defense against Social Engineering Attacks

7.1. Examples of organizations that effectively prevented SE incidents

Several organizations have effectively prevented SE incidents through proactive measures and vigilant practices. Here are a few examples:

Google: Google implements robust security protocols and continuous employee training to combat SE attacks successfully. They conduct regular phishing simulations, educate employees on identifying suspicious emails and links, and encourage reporting of potential threats. Their proactive approach has helped minimize the impact of SE incidents.

Microsoft: Microsoft has implemented multifactor authentication (MFA) and advanced threat protection mechanisms to defend against SE attacks. They conduct regular security awareness training for employees, focusing on recognizing and reporting phishing attempts. Through these efforts, Microsoft has strengthened their defense against SE incidents.

JPMorgan Chase: JPMorgan Chase has developed a comprehensive security program that includes employee education and awareness programs. They use simulated phishing campaigns to train employees and identify vulnerabilities. Additionally, they employ advanced technologies, such as artificial intelligence (AI) and machine learning (ML), to detect and prevent SE attacks.

Social media platforms (e.g., Facebook, Twitter): Social media platforms have implemented various security measures to protect their users from SE attacks. They use automated systems to detect and block suspicious accounts, employ user education initiatives to raise awareness about common scams, and provide reporting mechanisms for users to flag potentially malicious content.

Government agencies: Government agencies, such as the United States Department of Homeland Security (DHS), have implemented comprehensive security frameworks to combat SE attacks. They conduct training programs for employees, perform risk assessments, and establish incident response protocols to effectively respond to and prevent SE incidents.

7.2. Strategies and measures they employed to protect against social engineering

Here are some strategies and measures employed by organizations to protect against SE attacks:

Robust Security Awareness Training: Organizations conduct regular and comprehensive security awareness training programs for employees. This includes educating employees about SE tactics, red flags, and best practices for identifying and responding to potential threats.

Simulated Phishing Exercises: Organizations conduct simulated phishing campaigns to test employees' susceptibility to phishing attacks. These exercises help identify areas of vulnerability and provide targeted training to improve awareness and response.

Multifactor Authentication (MFA): Implementing MFA adds an extra layer of security to user authentication processes. It requires users to provide an additional form of verification, such as a unique code or biometric data, in addition to their passwords.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Email Security Measures: Organizations implement robust email security measures such as spam filters, anti-phishing technologies, and content analysis tools. These measures help detect and block suspicious emails containing phishing attempts or malicious content.

User Access Controls: Organizations enforce strict user access controls and the principle of least privilege. Employees are granted access only to the resources and information necessary for their roles, reducing the potential impact of successful SE attacks.

Incident Response and Reporting: Organizations establish clear incident response procedures and provide employees with accessible channels to report potential SE incidents promptly. This enables swift action and effective mitigation of threats.

Regular Security Assessments: Organizations conduct regular security assessments, including penetration testing and vulnerability scanning, to identify and address weaknesses in systems and processes. This proactive approach helps identify and remediate vulnerabilities before they can be exploited.

Ongoing Monitoring and Updates: Organizations continuously monitor and update security measures to stay ahead of evolving SE techniques. This includes keeping software and systems up to date with the latest patches and security updates.

By employing these strategies and measures, organizations can enhance their defenses against SE attacks and reduce the risk of successful compromises. It is important to combine technical safeguards with a strong security culture and ongoing vigilance to effectively combat SE threats.

7.3. Mandate security awareness and education

Security awareness and education play a crucial role in mitigating and defending against SE attacks. Here are key strategies in this area:

Employee training programs: Organizations should provide regular and comprehensive training programs to educate employees about SE techniques, red flags, and best practices for identifying and responding to potential threats. Training should cover various attack vectors such as phishing emails, phone scams, and impersonation attempts.

Simulated phishing exercises: Conducting simulated phishing exercises can help employees recognize and respond appropriately to phishing attempts. These exercises involve sending mock phishing emails to employees and analyzing their responses. It provides a practical learning experience and helps identify areas for improvement.

Security policies and guidelines: Establish clear security policies and guidelines that outline acceptable practices for handling sensitive information, sharing credentials, and interacting with unknown sources. Ensure that employees are aware of these policies and regularly update them to address emerging threats.

Ongoing communication and reminders: Consistently reinforce security awareness through regular communication channels such as newsletters, emails, or internal messaging platforms. Remind employees of the importance of remaining vigilant, reporting suspicious activities, and following security protocols.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Incident reporting and response procedures: Implement a clear and accessible incident reporting system that allows employees to report potential SE incidents promptly. Establish well-defined response procedures to ensure swift action in case of an incident, including isolating affected systems, notifying appropriate teams, and initiating incident investigations.

Collaboration with IT and security teams: Foster collaboration between employees, IT teams, and security professionals. Encourage open communication channels where employees can seek guidance, report concerns, and receive timely feedback on security-related issues.

Stay updated on emerging threats: Continuously monitor and stay informed about evolving SE tactics and trends. Regularly share relevant information with employees, providing insights into new attack methods and sharing real-world examples to enhance their understanding and awareness.

7.4. Implementation of technical measures

Implementing technical measures is an essential component of defending against SE attacks. Here are some key technical strategies for mitigation:

Firewalls: Deploying firewalls at network boundaries helps filter incoming and outgoing traffic, blocking known malicious sources and unauthorized access attempts. Firewalls can detect and block suspicious communication patterns and prevent unauthorized access to sensitive systems.

Email filters and spam detection: Implement robust email filtering systems to identify and block phishing emails, spam, and malicious attachments. These filters use various techniques such as sender reputation analysis, content filtering, and attachment scanning to identify and quarantine suspicious emails before they reach users' inboxes.

Intrusion Detection and Prevention Systems (IDPS): IDPS monitors network traffic for signs of suspicious or unauthorized activities. It can detect anomalies, such as unusual login attempts or data exfiltration, and take immediate action to mitigate potential threats.

Endpoint protection: Install and maintain up-to-date antivirus and anti-malware software on endpoints (computers, laptops, mobile devices). These tools help detect and block malicious software, including keyloggers and remote access trojans, which are commonly used in SE attacks.

Web filtering and content categorization: Employ web filtering solutions that restrict access to malicious or inappropriate websites. These solutions can prevent users from inadvertently visiting phishing sites or downloading malicious content.

Two-Factor Authentication (2FA): Implementing 2FA adds an additional layer of security to user authentication processes. By requiring users to provide a second form of verification, such as a unique code sent to their mobile device, it becomes more challenging for attackers to gain unauthorized access even if they possess stolen credentials.

Patch management and system updates: Regularly update operating systems, applications, and firmware with the latest security patches. Keeping systems up-to-date helps protect against known vulnerabilities that attackers may exploit during SE attacks.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

User access controls: Enforce strict user access controls, limiting privileges to only those necessary for job responsibilities. Implement the principle of least privilege, granting users access to only the resources and information they need to perform their tasks, reducing the potential impact of successful SE attacks.

8. Future Trends and Emerging Challenges in Social Engineering

8.1. Evolving tactics and techniques used by social engineers.

As SE continues to evolve, social engineers are adopting new tactics and techniques to deceive individuals and organizations. Here are some emerging trends and challenges in social engineering:

Advanced Phishing Techniques: Social engineers are developing more sophisticated phishing techniques, such as spear phishing and whaling, to target specific individuals or high-profile targets. These attacks involve personalized and convincing messages that are tailored to deceive the recipients.

Pretexting and Impersonation: Social engineers are increasingly using pretexting to create false scenarios or personas to gain trust and extract sensitive information. They may impersonate trusted individuals, such as colleagues, executives, or service providers, to manipulate victims into disclosing valuable data or performing unauthorized actions.

Exploitation of Social Media: Social media platforms provide a wealth of personal information that social engineers can exploit. They leverage this information to craft convincing messages, establish rapport, and increase the chances of successful manipulation.

Voice and Deepfake Technology: Advances in voice and deepfake technology enable social engineers to create highly realistic audio or video impersonations. This allows them to deceive individuals into believing they are interacting with someone they trust, making it more challenging to detect fraudulent communications.

Business Email Compromise (BEC): BEC attacks target organizations, often through compromised or impersonated email accounts of high-ranking executives. Social engineers use these accounts to deceive employees into initiating fraudulent wire transfers or sharing sensitive data.

Expanding Attack Surface: With the increasing adoption of remote work and the Internet of Things (IoT), social engineers have a wider attack surface to exploit. Remote workers may be more vulnerable to manipulation due to the lack of in-person interaction and established security protocols.

Psychological Manipulation and Influence: Social engineers are skilled at exploiting human vulnerabilities, such as fear, curiosity, urgency, and trust. They use persuasive techniques and psychological manipulation to coerce individuals into disclosing confidential information or performing actions against their better judgment.

8.2. Impact of technology advancements (e.g., AI) on social engineering

Technological advancements, including artificial intelligence (AI), have both positive and negative implications for social engineering. Here are some impacts of technological advancements in social engineering.

Enhanced Social Engineering Techniques: AI-powered tools and algorithms can analyze large datasets and generate highly targeted and personalized SE attacks. Social engineers can leverage AI to automate

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the creation of convincing phishing emails, chatbots, or voice simulations, increasing the effectiveness of their deception.

Deepfake Technology: AI-driven deepfake technology can manipulate audio and video content to create highly realistic impersonations. Social engineers can use deepfakes to deceive individuals by imitating trusted individuals or creating false evidence, making it more challenging to distinguish between genuine and manipulated content.

Spear Phishing and AI: AI algorithms can analyze publicly available data from social media platforms, professional networks, and other sources to gather information about individuals. This enables social engineers to create highly targeted spear phishing attacks with customized content and personalized details, increasing the chances of success.

Automated Bot Attacks: Social engineers can deploy AI-powered bots to conduct large-scale SE attacks. These bots can automatically send mass phishing emails, engage in chat conversations, or spread malicious content, amplifying the reach and impact of SE campaigns.

AI-Based Defense Mechanisms: On the positive side, AI can be utilized to develop advanced defense mechanisms against social engineering. AI algorithms can analyze network traffic, detect patterns of suspicious behavior, and identify potential SE attempts in real time, enabling faster response and mitigation.

Behavioral Analysis and User Profiling: AI algorithms can analyze user behavior, including online activities, communication patterns, and preferences, to create user profiles. Social engineers can exploit these profiles to tailor their attacks, making them more convincing and difficult to detect.

Adaptive Social Engineering Attacks: AI-powered SE attacks can adapt and evolve based on the responses and feedback they receive. Social engineers can use machine learning algorithms to continuously refine their tactics and increase their success rates over time.

To mitigate the negative impact of technological advancements on social engineering, organizations and individuals need to adopt proactive security measures. These include implementing AI-based defense mechanisms, conducting regular security assessments, and investing in comprehensive security awareness training programs to educate individuals about the risks associated with AI-driven social engineering.

Additionally, ongoing research and development in AI ethics and countermeasures can help mitigate the risks posed by AI-enabled SE attacks.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 – Region-wise Cybercrime Incidents on Social Engineering

Region	Frequency	Top patterns	Threat actors	Actor motives	Data compromised
APAC 164 with confirmed data disclosure	699 incidents, Internal (9%), Partner (2%), Multiple (2%) (breaches)	Social Engineering, System Intrusion and Basic Web Application Attacks represent 93% of breaches Espionage (39%), Convenience (2%), Grudge (2%), Secondary (1%) (breaches)	External (92%), Secrets (42%), Other (33%), Credentials (29%) (breaches)	Financial (61%),	Internal (56%),
EMEA 637 with confirmed data disclosure	2,557 incidents, Internal (2%), Multiple (1%) (breaches)	System Intrusion, Social Engineering and Basic Web Application Attacks represent 97% of breaches Espionage (8%), Ideology (1%), Fun (1%) (breaches)	External (98%), Internal (37%), System (35%), Other (15%) (breaches)	Financial (91%),	Credentials (53%),
LAC 65 with confirmed data disclosure	535 incidents, Internal (5%), Partner (2%), Multiple (2%) (breaches)	System Intrusion, Social Engineering and Basic Web Application Attacks represent 94% of breaches Espionage (11%), Ideology (2%) (breaches)	External (95%), Internal (32%), Classified (23%), Credentials (23%), Other (19%) (breaches)	Financial (93%),	System (55%),
Norther America 1,924 with confirmed data disclosure	9,036 incidents, Internal (12%), Multiple (9%), Partner (2%) (breaches)	System Intrusion, Basic Web Application Attacks and Social Engineering represent 85% of breaches Espionage (1%), Grudge (1%) (breaches)	External (94%), Internal (50%), Personal (38%), Other (24%) (breaches)	Financial (99%),	Credentials (67%),

9. Conclusion

The new explosion of social engineering represents a grave threat to organizations and individuals alike, necessitating the implementation of robust defensive techniques to mitigate the associated risks. Education and awareness are paramount in combating social engineering attacks. By providing comprehensive training programs, organizations can empower employees to recognize various social engineering techniques, identify red flags, and adopt best practices for securely handling sensitive information. Cultivating a culture of security awareness among individuals equips them to effectively identify and resist social engineering attempts.

In addition to education, implementing strong technical controls is essential in managing the risk of social engineering. Regular software updates and patches, coupled with robust authentication mechanisms and multi-factor authentication, fortify an organization's defense against social engineering attacks. By prioritizing these defensive measures, organizations can decrease the probability of successful social engineering attacks targeting their systems and networks.

Regular security assessments and penetration testing are vital to proactively identify vulnerabilities that may be exploited through social engineering. By actively testing and addressing weak points in their security infrastructure, organizations can strengthen their defenses against social engineering attacks. Moreover, establishing clear policies and procedures for handling sensitive information is crucial. Employees should be trained on proper information-sharing protocols, data protection measures, and incident reporting procedures. Employing access controls and least privilege principles minimizes the risk of unauthorized access to sensitive data through social engineering tactics.

Monitoring and detection mechanisms play a pivotal role in an effective defense against social engineering attacks. Organizations should implement security monitoring systems capable of identifying suspicious activities, detecting unauthorized access attempts, and triggering alerts when social engineering attacks are suspected. Timely detection enables swift response and mitigation, reducing the potential damage inflicted by social engineering attacks. Lastly, having a well-defined incident response plan in place is crucial. This plan should outline predefined steps to be taken in the event of a social

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

engineering attack, communication protocols, and coordination with law enforcement agencies if necessary. Regular testing and updates of the incident response plan ensure its effectiveness in real-world scenarios.

The escalating wave of social engineering necessitates the adoption of comprehensive defensive techniques. Through education, technical controls, policy implementation, monitoring, and incident response planning, organizations can fortify their resilience against social engineering attacks. By proactively addressing these risks, organizations can safeguard their valuable assets and protect against the damaging consequences of social engineering.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AI	Artificial Intelligence
BEC	Business Email Compromise
DHS	Department of Homeland Security
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
IDPS	Intrusion Detection and Prevention Systems
IoT	Internet of Things
IT	Information Technology
MFA	Multifactor Authentication
ML	Machine Learning
SE	Social Engineering
URL	Uniform Resource Locator
USB	Universal Serial Bus

Bibliography & References

H. Aldawood and G. Skinner. *Contemporary cyber security social engineering solutions, measures, policies, tools and applications: A critical appraisal. International Journal of Security (IJS), 10(1):1, 2019.*

Wenke Lee, Bo Rotoloni, *"Emerging cyber threats, trends and technologies", Technical report, Institute for Information Security and Privacy, 2016.*

Nabie Y Conteh, Paul J Schmick, *"Cybersecurity: risks, vulnerabilities and countermeasures to prevent social engineering attacks", International Journal of Advanced Computer Research, Vol.6 pp.23-31, 2016.*

Parker Graeme, Shala Vlerar, *"Social engineering and risk from cyber-attacks", Technical report, PECB, March 2016.*

K. Beckers, L. Krautsevich and A. Yautsiukhin, "*Using attack graphs to analyze social engineering threats", International Journal of Secure Software Engineering (IJSSE), vol. 6, no. 2, pp. 47-69, 2015.*

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

The Smart and Seamless Evolution of the HFC Network – Today and Beyond

A Technical Paper prepared for SCTE by

Gustavo Colly
Technical Solutions Sales Engineer
ATX Networks
Buenos Aires, Argentina
+1 (289) 204-7800 Ext 1300
gcolly@atx.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Traditional Centralized Access Architecture (CAA) ... 3
3. Distributed Access Architecture (DAA) ... 4
4. Envisioning DOCSIS 4.0 (FDD and FDX) ... 5
   - 4.1. Frequency Division Duplex (FDD) DOCSIS ... 5
   - 4.2. Full Duplex (FDX) DOCSIS ... 5
5. Outside Plant Architectures ... 6
   - 5.1. OSP used with FDD ... 6
     - 5.1.1. Bandwidth Migration vs Gain & TCP ... 7
   - 5.2. OSP used with FDX ... 8
6. Residential Drops — Today and Tomorrow ... 8
7. New Requirements on Powering the Network ... 9
8. Industry Transition from CAA to DAA ... 9
9. Conclusion ... 10

Abbreviations ... 11

Bibliography & References ... 12

List of Figures

Title **Page Number**

Figure 1 - Centralized Access Architecture (CAA) ... 3
Figure 2 - Distributed Access Architecture (DAA) ... 4
Figure 3 - Frequency Division Duplex (FDD) Spectrum ... 5
Figure 4 - Full Duplex (FDX) DOCSIS ... 6
Figure 5 - OSP FDD Architecture ... 6
Figure 6 - Bandwidth Migration vs Gain and TCP ... 7
Figure 7 - FDD 1.8GHz RF Output Levels ... 7
Figure 8 - OSP FDX Architecture ... 8
Figure 9 - Residential Drops — Today and Tomorrow ... 8
Figure 10 - New Requirements on Powering the Network ... 9
Figure 11 - Cable Access Revenues ... 9

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

As consumer bandwidth requirements continue to grow, cable operators will need to intelligently and cost-efficiently expand network capacity to accommodate future increases in data traffic — especially upstream traffic. Migrating from CAA (Centralized Access Architecture) to DAA (Distributed Access Architecture) architectures, a strategy that most cable service providers are currently pursuing, provides a near-seamless and cost-effective transition. Reallocating spectrum through band splits and the adoption of DOCSIS® 4.0 will play important roles in that effort by taking advantage of the incredible resiliency of cable's HFC network. Depending on the status of the cable operator's HFC network, dropping in RF amplifier modules to fiber nodes and trunk and distribution amplifiers in conjunction with new taps and passives could also provide cable operators with a seamless evolution of their networks. The ubiquity of coaxial connections makes the continued evolution of DOCSIS standards a logical and economical way to continue to bring world-class broadband services to millions of households. Coax is also capable of supporting current gigabit-class services by preserving/reusing most of the existing network infrastructure. The bottom line is that HFC networks can be competitive with symmetrical Gigabit services through DOCSIS 4.0 standards, which will allow the needed expansion of upstream capacity.

This technical paper will highlight best practices and design criteria for managing the evolution of the HFC network, as well as detailing relevant market trends.

2. Traditional Centralized Access Architecture (CAA)

Traditional Centralized Access Architecture (CAA) has been deployed all over the world and is a proven and secure architecture for cable service providers. But it has reached its limits.

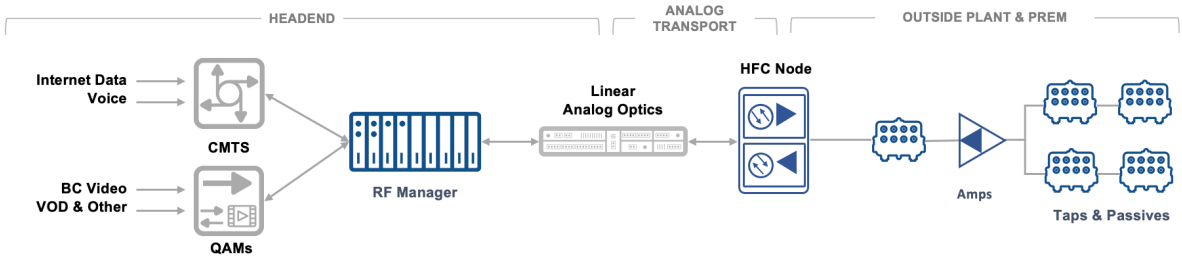


Figure 1 - Centralized Access Architecture (CAA)

In a centralized access architecture, most of the video and data infrastructure remains in the hub or headend location. The main components of it are BC and NC video Edge QAMs, which generate the PHY layer of the video carriers, and the CMTS, which oversees MAC and PHY functions of DOCSIS carriers. CCAP core combines both video and data functions in one device. RF management is used to split/combine the downstream/upstream (DS/US) carriers required by the service group and an optical transmission and reception stage that usually uses linear analog optic technology and oversees the conversion from RF to optical, and vice versa. In the outside plant (OSP), the HFC fiber node connects the fiberoptic cable to the trunk and distribution RF system. It does the same function as the inside plant's (ISP's) optical transmission/reception stage; it converts from an optical signal to an RF signal. OSP taps and passives are being used to distribute RF carriers over the service area with RF amplifiers.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

As mentioned earlier, CAA has limits and they are associated with the limitations of CMTS/CCAP cores, Video Edge QAMs, RF Management products and the inside plant's optical platform. Most of this equipment supports up to 1.2GHz in the downstream direction. In terms of upstream direction, its limit is 204MHz, known as the high-split band. These physical limits, in terms of frequency, are determined by DOCSIS 3.1.

The linear analog optical solutions often used by the ISP for downstream direction consist of RF-modulated signals that will be amplitude modulated (AM) onto an optical carrier. Several studies demonstrate that 4096 QAM OFDM signals are achievable in the link between TX (ISP) and RX (OSP Node). In the upstream direction, a DFB type of laser was widely deployed in reverse transmitters at fiber nodes and can work with 5-85MHz and 5-204MHz bandwidths. Both upstream bandwidths can achieve 1024 QAM Orthogonal Frequency-Division Multiple Access (OFDMA) signals. Digitizing the upstream optical link of HFC using EDR or Remote PHY can help break the power budget limitations of the analog optical link and go a little farther on QAM constellations used.

CMTS/CCAP Cores used in CAA architectures are based on hardware-specific functionalities that impact capacity in DS and US ports. Current D3.1 cable modems support 32 SC-QAM plus two OFDM blocks. Regarding US support, D3.1 CMs support eight SC-QAM signals along with two 96MHz (max) OFDMA blocks.

In numbers, 32 SC-QAMs plus two OFDM full block (192MHz) with 4096 QAM constellation provides approximately 5Gbps DS capacity (1.9Gbps x 2 + 1.2Gbps = 5Gbps downstream capacity). In US, two OFDMA with 1024 QAM constellation plus four SC-QAMs with 6.4MHz width provides approximately 1.575Gbps upstream capacity.

3. Distributed Access Architecture (DAA)

Distributed Access Architecture (DAA) is an evolution of CAA, where the main functions performed in the headend or hub are relocated to the fiber nodes, closer to the subscriber. Moving these operations from the headend/hub enables MSOs to virtualize operations by using generic computer hardware, which relieves space and cooling constraints of the facilities caused by continued exponential growth in node counts and traffic.

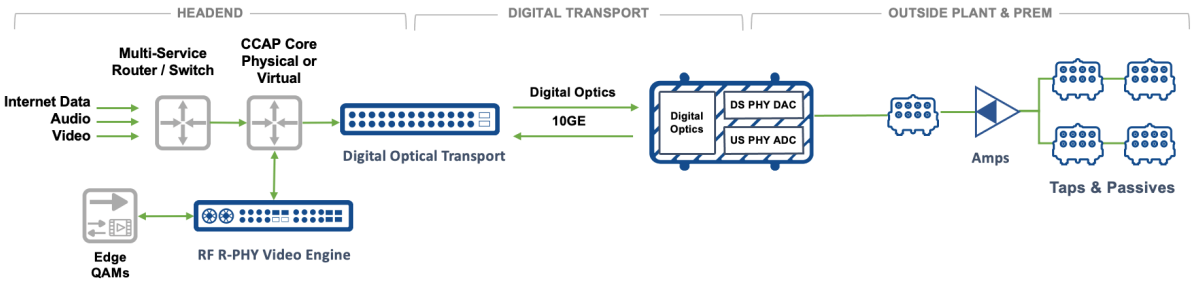


Figure 2 - Distributed Access Architecture (DAA)

The main difference between the two architectures is that the PHY layer of video and data in DAA architectures is now in the fiber node, and a 10Gbps or higher Ethernet fiber link is used to connect all these components, replacing the previous analog optical link. There are two flavors of DAA architectures: R-PHY, which moves the modulation and demodulation to the fiber node and leaves other functionality at the hub or headend, and R-MACPHY, or FMA, that moves the DOCSIS MAC and PHY layers to the fiber node. Both versions are under the Distributed Access Architecture program at CableLabs. All these

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

mentioned differences allow for an improvement in end-of-line (EOL) performance (mainly RxMER), better spectrum efficiency by using higher modulation orders in OFDM/OFDMA blocks and provides flexibility for operators to deploy network functionality. DAA architecture is also the driver for the delivery of 10G symmetrical capacity at the service group level in conjunction with DOCSIS 4.0, which will allow cable operators to achieve 10Gbps speeds downstream and 6Gbps upstream.

4. Envisioning DOCSIS 4.0 (FDD and FDX)

DOCSIS 4.0 technology includes support for Extended Spectrum DOCSIS (ESD), also known as Frequency Division Duplex (FDD), and Full Duplex DOCSIS (FDX) capabilities.

4.1. Frequency Division Duplex (FDD) DOCSIS

FDD mode of operation uses spectrum dedicated to upstream transmission and separate spectrum dedicated to downstream transmission. FDD DOCSIS increases the upper end of upstream spectrum to 684MHz, while the maximum downstream frequency increases to approximately 1.8GHz.

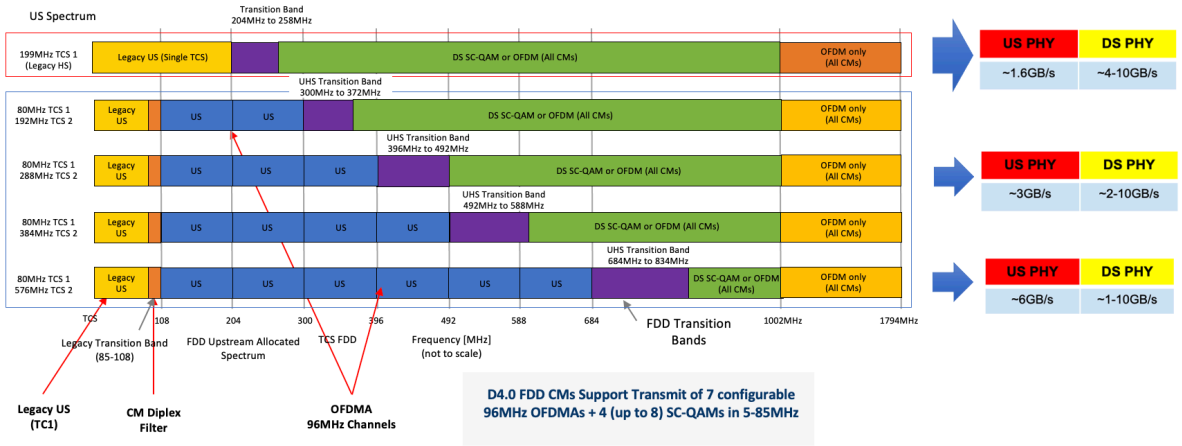


Figure 3 - Frequency Division Duplex (FDD) Spectrum

As shown in Figure 3, depending on the split allocation used, MSOs can get from 1.6Gbps (High-Split) up to 6Gbps (Ultra-High Split 684MHz) upstream provisioned capacity. In the downstream, 1 to 10Gbps can be achieved based on the combination of SC-QAMs with OFDM blocks.

4.2. Full Duplex (FDX) DOCSIS

The FDX specifications were created to provide service providers with the ability to increase the offered upstream speeds without sacrificing the valuable downstream spectrum. This is done by overlapping the US and DS spectrum from 108-684MHz, also known as FDX US/DS overlap band.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

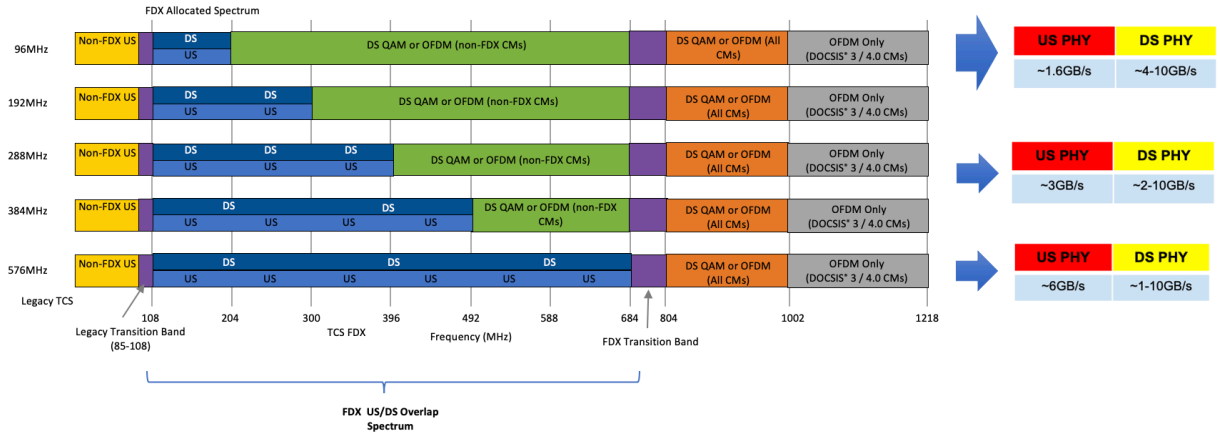


Figure 4 - Full Duplex (FDX) DOCSIS

FDX relies on echo cancellation functionality, which is performed in the fiber node. Echo cancellation is needed because the fiber node is using the same spectrum for simultaneous DS transmissions and US receptions.

5. Outside Plant Architectures

New outside plant architectures will need to support FDD or FDX technologies with the aim of preserving, as much as possible, RF amplifier spacing and taps and passives distribution.

5.1. OSP used with FDD

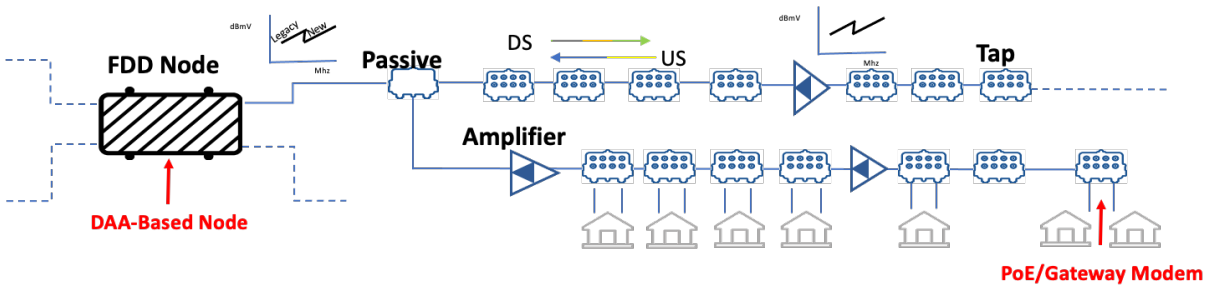


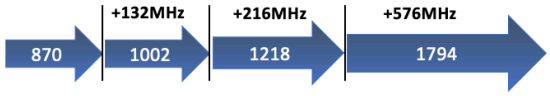
Figure 5 - OSP FDD Architecture

Fundamentally, the OSP Architecture to support DOCSIS 4.0 FDD needs to support N+X architectures and a range of US/DS split bands from 85MHz to 684MHz in upstream spectrums and from 105MHz to 1794MHz in downstream spectrums. It will be driven by the required US capacity, which will determine the US/DS diplexers needed to configure impacted fiber nodes and RF amplifiers. Maintaining "legacy" DS slopes and levels will help to minimize impacts in the OSP. OSP engineers must consider the Total Composite Power (TCP) of their systems to determine if one tilt/level or two levels must be used.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.1.1. Bandwidth Migration vs Gain & TCP

TCP, which is also shorthand to Total Power, is the combined power of all signals in a given frequency range — for instance, the downstream. It is a concern because excessive total power is what overdrives RF hybrids, lasers, set-tops boxes, modems and other devices.



System Amplifier Gain/Level

1.0 to 1.8GHz Migration	1002	1218	1794			
Step down @1.2G (dB)	N/A	N/A	6.0			
Virtual tilt (dB)			20.0			
TCP	60.6	65.6	69.2	Gain w/Internal Cable Profile Tilt		
	Amplifier Output Level (dBmV)			1002	1218	1794
1794MHz			50.0			49
1221MHz			42.6			42.8
1215MHz		49.0	48.5		48	42.8
999MHz	44.0	46.0	45.7	43.0	45.7	40.3
837MHz	41.6	43.1	43.6	41.2	44	38.2
609MHz	38.0	39.5	40.6	38.5	40.9	35.0
495MHz	36.5	37.7	39.1	37.1	39.2	33.2
261MHz	33.0	34.0	36.0	32	33	29.0

Figure 6 - Bandwidth Migration vs Gain and TCP

RF hybrid technology has been supported in the past by increasing gain and RF levels from 870MHz, 1002MHz and 1218MHz, extending the RF levels. That was until the jump to 1.8GHz. A 1.8GHz amplifier equipped with the newest RF hybrid can support around 69dBmV TCP in the output port. This TCP value will dictate the new system design rules of cable operators.

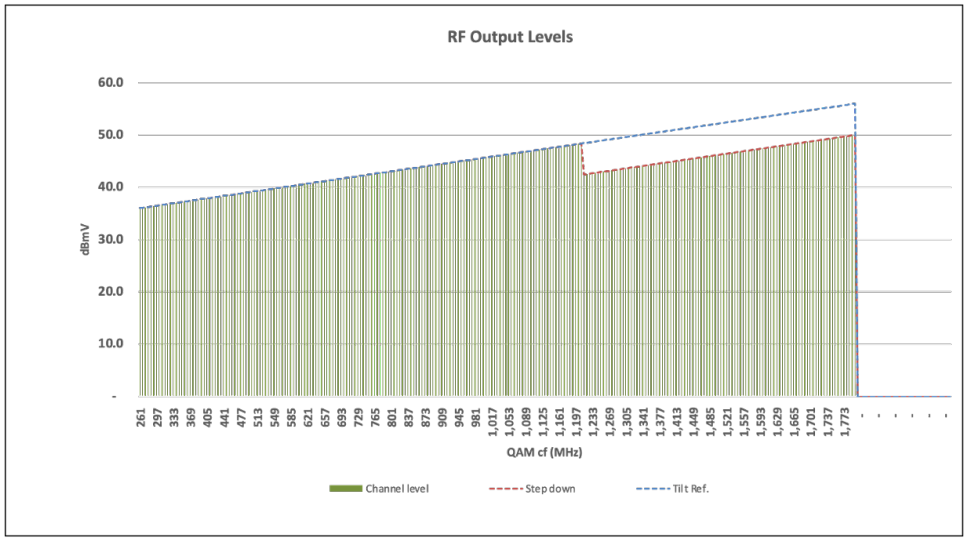


Figure 7 - FDD 1.8GHz RF Output Levels

As an example, Figure 7 shows the RF output levels suggested to use with new 1.8GHz RF amplifier technologies. In this high-split example, legacy levels from 258MHz to 1218MHz were preserved as they are in operation, and levels above 1218MHz to 1794MHz are 6dB lower.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

New RF amplifiers incorporate electronic control of RF levels, which supports the auto alignment of the attenuation and equalization at its various stages. They also support 1.2GHz and 1.8GHz modes to ease the transition to 1.8GHz and to support the actual implementations.

5.2. OSP used with FDX

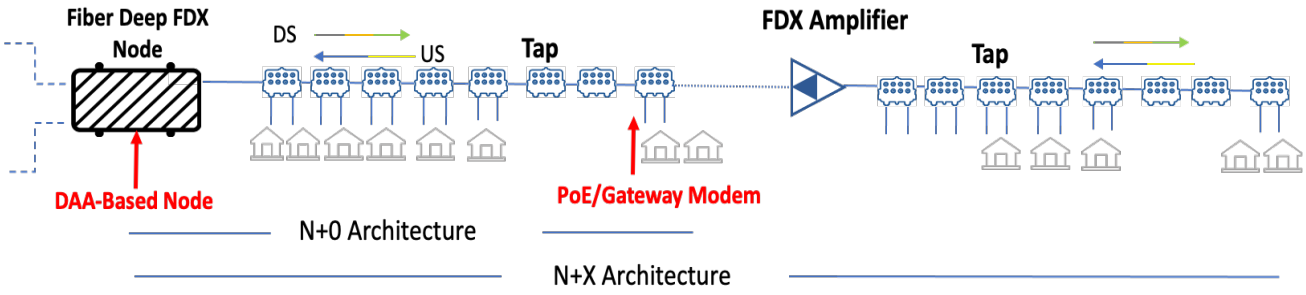


Figure 8 - OSP FDX Architecture

As mentioned previously, FDX is one of two DOCSIS 4.0 flavors available to cable operators. FDX runs DS/US in the same spectrum allocation of 1.2GHz that DOCSIS 3.1 networks use today. It originally was compatible with N+0 architectures, but due to its cost, cable service providers interested in this technology are looking to deploy FDX in N+1 or up to N+4 architectures. FDX will preserve 1.2GHz taps and passives distribution. But it will require a new generation of both fiber nodes and RF amplifiers equipped with echo cancellation capabilities.

6. Residential Drops — Today and Tomorrow

At the customer premise, a significant change, especially in LATAM locals, is the migration from a wired to a wireless home network in which most home devices are connected to a coaxial cable network to a wireless home network.

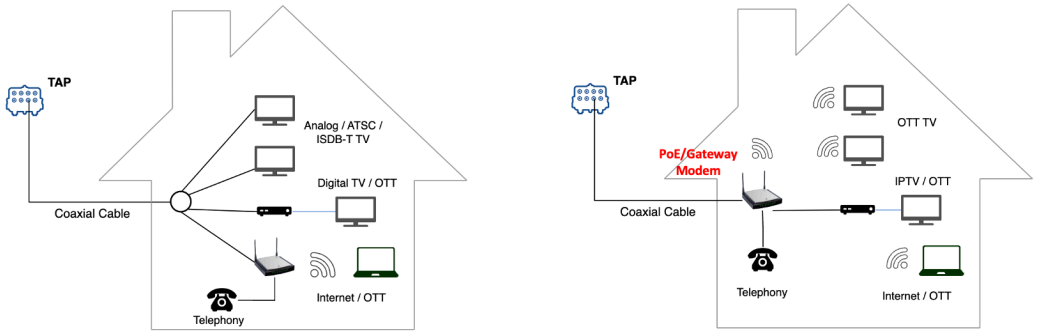


Figure 9 - Residential Drops — Today and Tomorrow

In this environment, a DOCSIS 4.0 modem will be a point of entry (PoE) device, making it the sole hybrid fiber-coaxial (HFC)-terminating device in the household. This PoE gateway incorporates WIFI 6E and WIFI mesh capabilities to support the coverage requested inside the home. Additionally, there will be no need for a splitter network to feed other boxes, such as set-top boxes, and in many instances also the cabling inside the unit.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

7. New Requirements on Powering the Network

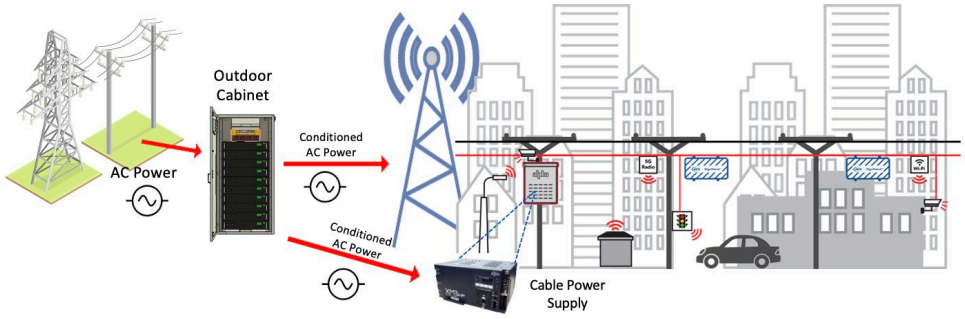


Figure 10 - New Requirements on Powering the Network

Power has become an increasing concern due to the increase in power requirements by modern technologies, as well as the growing need for power to be reliable. Power outages can have disastrous effects, shutting down critical services such as water, energy, communications, transportation and other types of infrastructure. Energy Storage Systems play a big part in establishing a reliable power source.

DAA architectures require the distribution of network components (like RPDs, RMDs, Small Cells, WIFI Access Points, etc.), putting increased power strains on the network. The impact of a power outage in a DAA architecture is more severe and longer lasting than in a CAA architecture. As an example, the booting time of an RPD is a few minutes, which is the time required for all video and data services to be restored. For that reason alone, MSOs have an increasing need for reliable power systems in the OSP.

8. Industry Transition from CAA to DAA

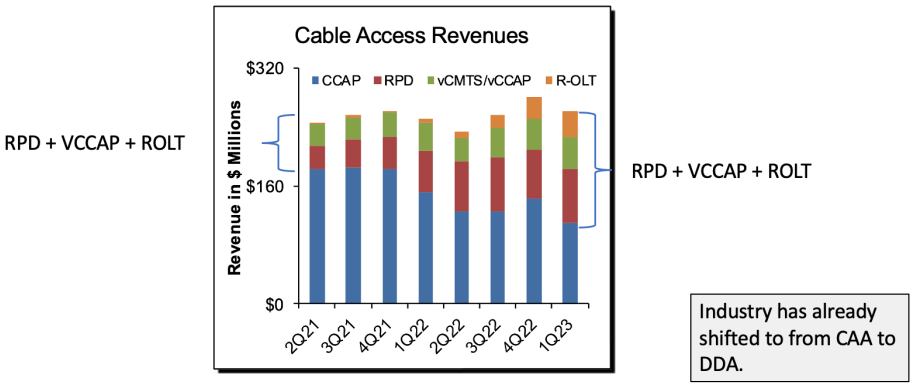


Figure 11 - Cable Access Revenues

An analysis of the cable broadband market over the past few years indicates that the industry has already shifted from CAA to DAA architectures. It was driven mainly by required band splits, which rely on an increase in DAA deployments, particularly the growth of RPDs.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

9. Conclusion

Cable industry innovation, such as the latest release of DOCSIS standards, will enable cable service providers to reach the 10G era of symmetrical multigigabit services, empowering cable operators to meet the increasing bandwidth demands of residential and business subscribers. In terms of architecture, DAA will play a key role in this evolution in conjunction with FDD and FDX technologies. Industry stakeholders are aligned with this evolution and market transition, concentrating R&D, product development and network design and adoption on the continued evolution of the HFC network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AM	amplitude modulation
BC	broadcast
CAA	centralized access architecture
CCAP	converged cable access platform
CM	cable modem
CMTS	cable modem termination system
DAA	distributed access architecture
DFB	distributed feedback laser
DOCSIS	data over cable service interface specification
DS	downstream
EDR	enhanced digital return
EOL	end of line
FMA	flexible MAC architecture
HFC	hybrid fiber coaxial
Hz	hertz
ISP	inside plant
MAC	media access control
MSO	multiple system operator
NC	narrowcast
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
OSP	outside plant
PHY	physical layer
PoE	point of entry
QAM	quadrature amplitude modulation
R-MACPHY	remote MACPHY device
R-PHY	remote PHY
RF	radio frequency
RMD	remote MACPHY device
ROLT	remote OLT
RPD	remote PHY device
RX	receiver
RxMER	receive modulation error ratio
SC-QAM	single carrier QAM
SCTE	Society of Cable Telecommunications Engineers
TCP	total composite power
TX	transmitter
US	upstream
VCCAP	virtualized cable convergence access platform
WIFI	wireless fidelity

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

[Jay Lee/ATX Networks]. (2020, June 24). *A new era of HFC evolution* [Webinar Video]. ATX. https://atx.com/company/news/webinar-the-2050-project-a-new-era-of-hfc-evolution/

Whitley, M. & Whittlesey, P. (2023). [1.8GHz Amplifier System Design Specifications]. ATX Networks.

He, Z., Skrobko, J., Zhang, Q. & Zhang, W. (2016). *The capacity of analog optics in DOCSIS 3.1 HFC networks*. Cisco Systems. https://www.nctatechnicalpapers.com/Paper/2016/2016-the-capacity-of-analog-optics-in-docsis-3-1-hfc-networks

Downey, John J. (2017). *Downstream early lessons learned*. Cisco Systems. https://www.nctatechnicalpapers.com/Paper/2017/2017-docsis-3-1-downstream-early-lessons-learned

Downey, John J. (2020). *The power of distributed access architectures (DAA). Benefits of digital fiber along with remote-PHY*. Cisco Systems. https://www.nctatechnicalpapers.com/Paper/2020/2020-the-power-of-distributed-access-architectures

(2023). [Broadband Access Report]. Dell'Oro Group.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Thinking Of Moving from DPoE to XGS-PON?

A Few Insights from a Cable Operator's Perspective

A Technical Paper prepared for SCTE by

Melody Rock
Manager, Access Technology
Rogers Communication
2400 32 Ave NE, Calgary, AB Canada
403-781-4964
melody.rock@sjrb.ca

Colin Dearborn
Senior Network Architect II, Access Technology
Rogers Communication
2400 32 Ave NE, Calgary, AB Canada
403-538-5297
Colin.dearborn@sjrb.ca

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Optical Distribution Network ... 3
   - 2.1. Transmission Characteristics ... 4
   - 2.2. Wavelength Planning ... 4
   - 2.3. Distributed vs Centralized Split ... 6
3. Deployment Model ... 7
   - 3.1. Greenfield Development ... 7
   - 3.2. Customer Migration from EPON to XGS-PON ... 8
4. Service Catalog ... 9
   - 4.1. Services on RFoG ... 10
   - 4.2. Speed Tiers ... 10
   - 4.3. Specialized Services ... 10
5. Transitioning from DPoE to Native Provisioning ... 11
   - 5.1. Monitoring ... 11
   - 5.2. Provisioning ... 11
   - 5.3. Access CPE Device Management ... 12
   - 5.4. Network Architecture ... 12
6. Program and Organizational Change Management ... 14
   - 6.1. Development to Production Phases ... 14
   - 6.2. Organizational Change Management ... 14
7. Conclusion ... 15

Abbreviations ... 16

Bibliography & References ... 17

List of Figures

Title **Page Number**

Figure 1 – Simplified PON Optical Distribution Network ... 3
Figure 2 – PON Wavelengths and Coexistence ... 5
Figure 3 – Centralized Split Example ... 6
Figure 4 – Distributed Split Example ... 7
Figure 5 – Customer Migration in a Centralized Split Example ... 8
Figure 6 – Customer Migration in a Distributed Split Example ... 9
Figure 7 – DPoE Network Block Diagram ... 13
Figure 8 – GPON Network Block Diagram ... 13

List of Tables

Title **Page Number**

Table 1 – EPON and GPON Characteristics ... 4
Table 2 – Special Services Design Considerations ... 11

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Have you been contemplating implementing XGS-PON in your network but have thus far leveraged Data Over Cable Service Interface Specification (DOCSIS®) provisioning over Ethernet passive optical networking (EPON) (DPoE)?

As cable providers have deep experience in deploying and managing DOCSIS networks, DPoE provides a great method of leveraging DOCSIS knowledge and investments to support fibre to the premises (FTTP) deployments. Despite this advantage, many operators are considering shifting their FTTP technology plans from EPON/DPoE based systems to 10 gigabit symmetrical passive optical network (XGS-PON) for technical or financial reasons.

This paper will explore the key items to consider when making the transition to XGS-PON. Topics that will be explored include technical items such as fiber architecture, provisioning and tools, all the way to cultural and organizational change management (OCM).

As all operator networks and organizations are different, this paper is not intended to provide the answers but instead offer insight and thoughts that operators can take away to help form their own plans.

2. Optical Distribution Network

The optical distribution network (ODN) is the optical transmission medium connecting the optical network terminal (ONT) to the optical line terminal (OLT). It consists of the feeder fiber, drop cables and passive optical splitters fiber. The ODN is a key component of the passive optical network (PON). While the ODN requirements of EPON and XGS-PON are very similar there are several important factors that must be considered before planning a transition from EPON to XGS-PON.

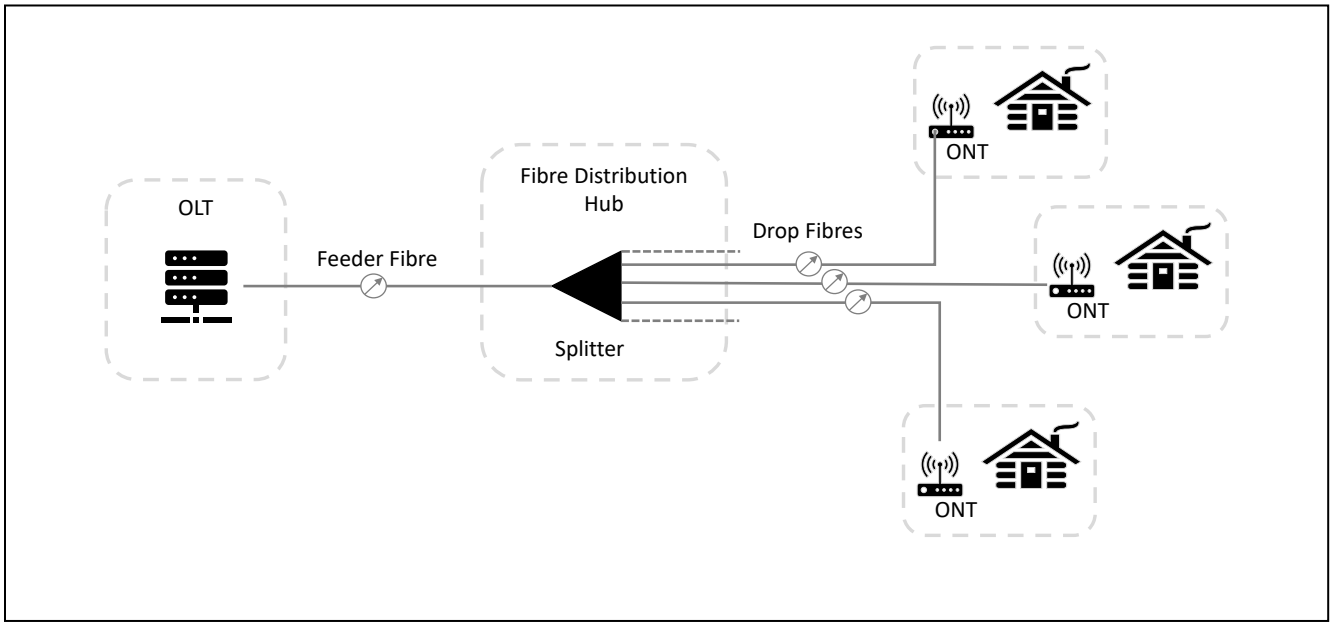


Figure 1 – Simplified PON Optical Distribution Network

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.1. Transmission Characteristics

In the ODN, distances from the OLT to the ONT, and PON capacity are factors in determining the split ratios. The types of customers that will be served on the PON – business and/or residential, and the location of the customers (rural vs urban) may also influence the decision on the split ratios and PON architecture.

Table 1 – EPON and GPON Characteristics

	EPON	10G-EPON	GPON	XGS-PON
Standards	IEEE 802.3ah (2004)	IEEE 802.3av (2009)	ITU-T G.984 (2003)	ITU-T G.987.1 (2016)
Downstream Line Rate	1.25 Gbps	10 Gbps	2.488 Gbps	9.953 Gbps
Upstream Line Rate	1.25 Gbps	10 Gbps	1.244 Gbps	9.953 Gbps
Split Ratio	up to 1:64	up to 1:128	up to 1:64	up to 1:128

EPON/10G EPON and GPON/XGS-PON share transmission characteristics, similar split ratios and optical budgets. As a result, existing EPON/10G EPON ODNs can support XGS-PON services without any modification.

2.2. Wavelength Planning

EPON and GPON technologies use the same wavelengths. 10G EPON and XGS-PON also share the same wavelengths. Therefore, the two technologies cannot share the same physical fiber. Coexistence on the same PON fiber is supported with EPON and 10G EPON, or GPON and XGS-PON. But when migrating from EPON to GPON, or 10G EPON to XGS-PON, then separate PONs and splitters will need to be utilized. When migrating from EPON to XGS-PON, the two technologies can coexist from a wavelength perspective.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

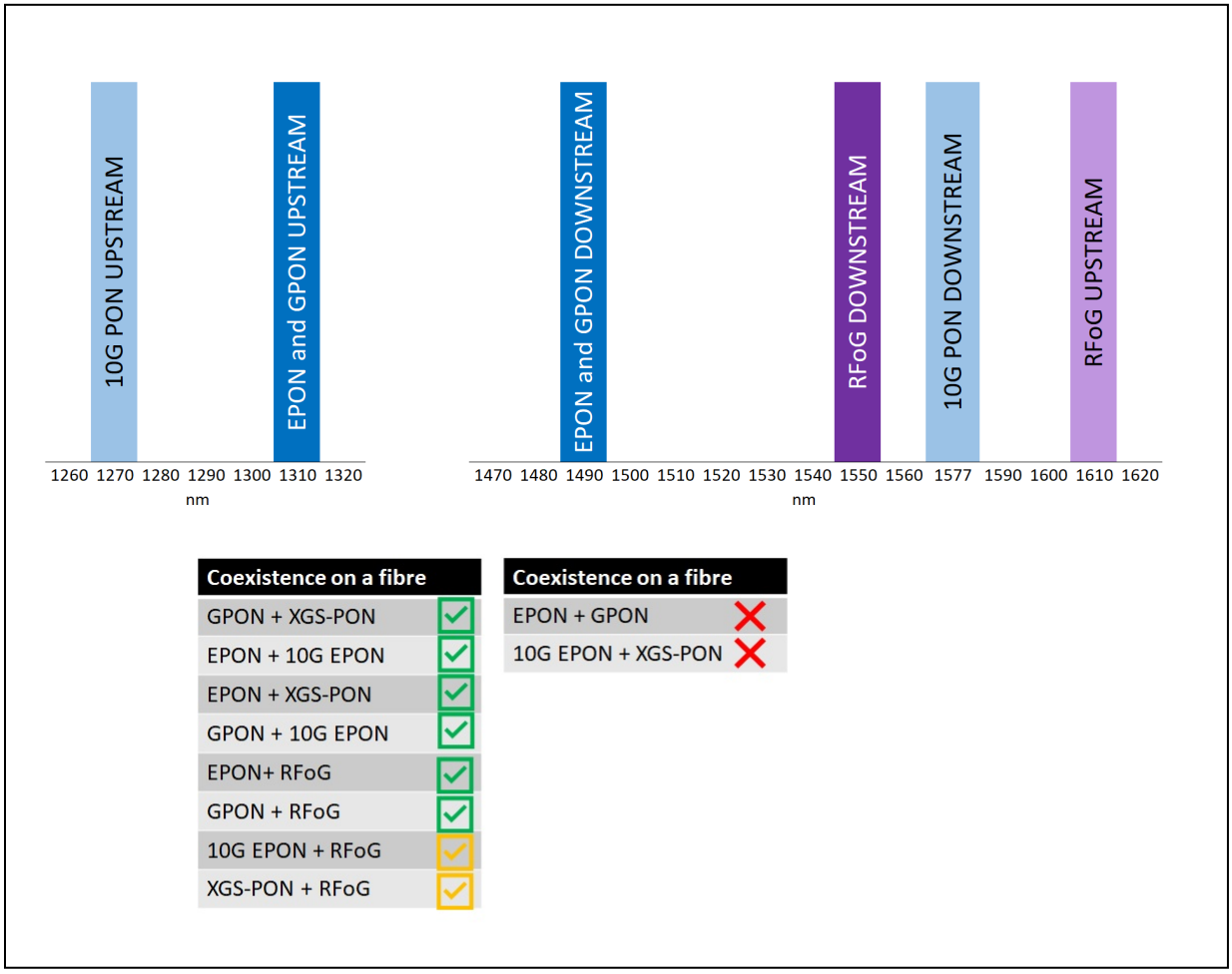


Figure 2 – PON Wavelengths and Coexistence

Radio frequency over glass (RFoG) (ANSI/SCTE 174 2010) may have been used to extend quadrature amplitude modulation (QAM) based services (voice, video and internet) over fiber. For these services, RFoG may be used on its own, with EPON or with GPON. RFoG will also coexist with 10G EPON or XGS-PON but will require wavelength filtering changes due to the proximity of wavelengths used for RFoG and PON. Additionally, operators will need to be cognizant of the RFoG power levels and that the split ratio will need to be kept to 1:32. This could be a limiting factor for PON capacity expansions though, considering that 10G EPON or XGS-PON supports 1:128 split ratios. It should also be noted that if operators use distributed OLT hardware (remote-OLTs) instead of chassis-based OLTs in the hub-site, the analog optics will need to be incorporated into the node housing in the field as well. However due, to RFOG's own wavelengths and optical power characteristics, considerations will not allow coexistence with 10G EPON or XGS-PON. Operators will need to develop a plan to migrate off RFoG especially with PON capacity expansions. Therefore, as ITU and IEEE standards for PON share the same wavelengths this can result in co-existence issues. Operators need to ensure they have a clear wavelength plan defined.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.3. Distributed vs Centralized Split

Operators will need to consider their PON strategy especially around their current split architecture: centralized or distributed splitting.

In a centralized architecture or single stage splitting, one splitter is fed by one OLT PON port and supports all the drops to the customer premises. In Figure 3, one PON port is split to 32 customer premises at one splitter as an example. There would be a higher concentration of fibers from the splitter as opposed to a distributed split. In a centralized architecture, splitters could be dedicated for each wavelength and PON. Good documentation is required to manage the splitters, homes passed of the PON port and the ONT locations.

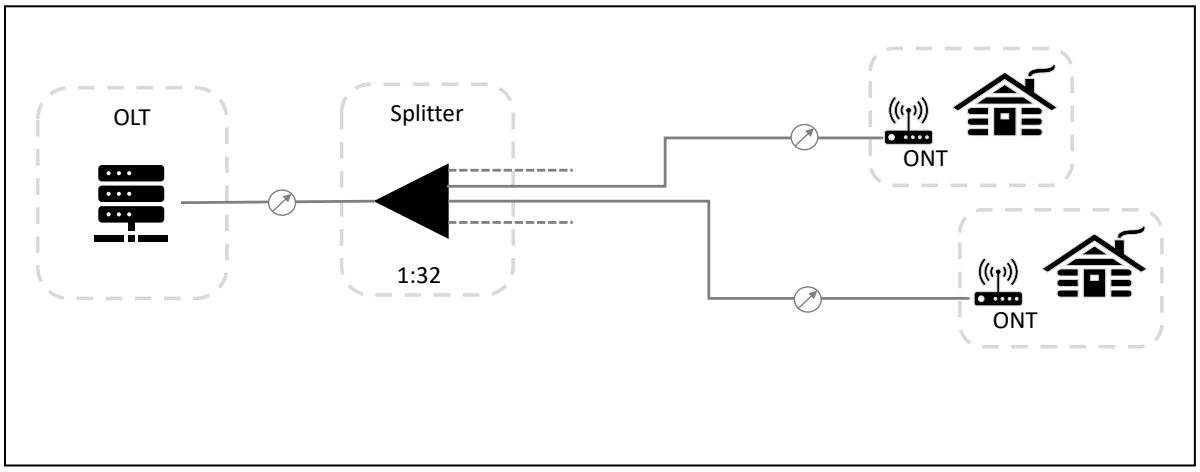


Figure 3 – Centralized Split Example

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In distributed and cascaded splits, there are stages of splitters. The splits ratios at the splitters can be any combination as required of the PON. An example distributed split is shown in Figure 4. In this example, one fiber from the OLT PON port would be split 1:4 at the first splitter, and four 1:8 taps would feed a total of 32 customer premise locations. The splitters can be located anywhere in the field. This architecture reduces the fiber counts required near the first stage splitter. Operators will need to take care in planning any PON technology transitions especially if the PONs cannot coexist.

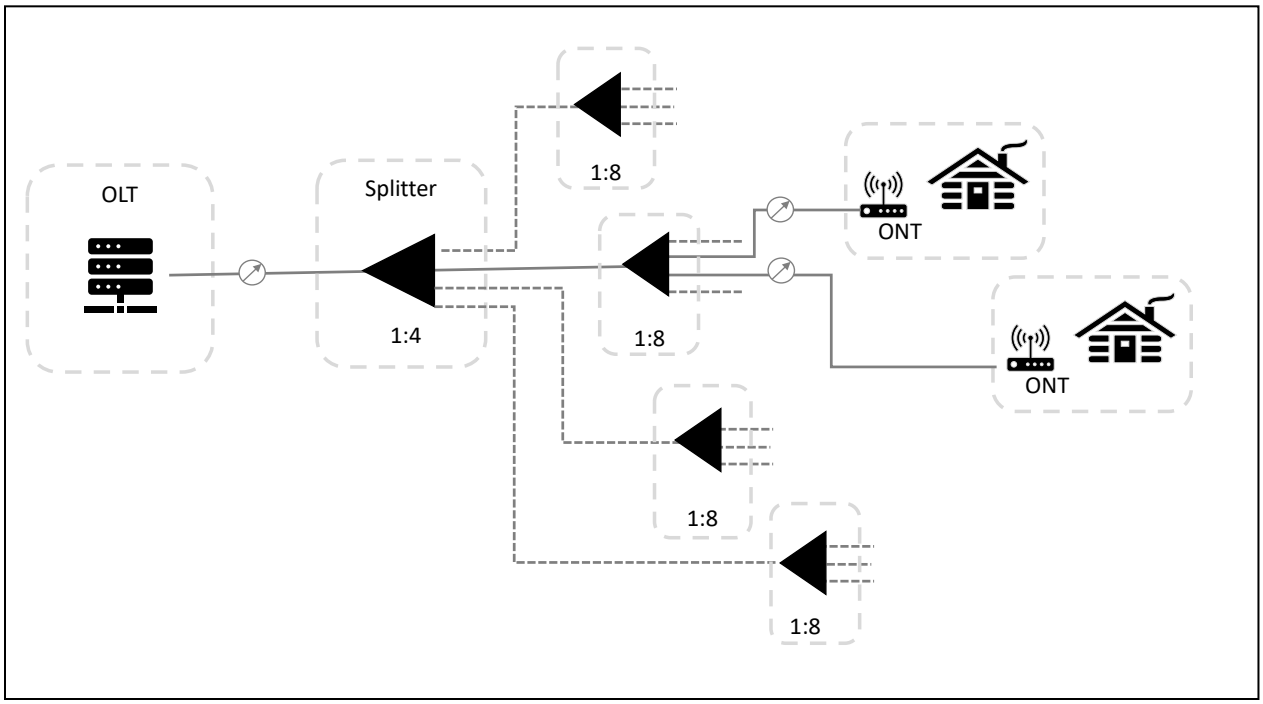


Figure 4 – Distributed Split Example

3. Deployment Model

For PON technology transitions, it will be important to consider the deployment strategies of both new PON customers and of the current PON customer base in an operator's network.

3.1. Greenfield Development

In greenfield deployments, new fiber and PON builds are made to reach new customers. While this is a blank slate, operators will need to determine a wavelength plan of which technologies will be deployed over which fibers and splitters. Proper ODN documentation will be important not only for dependent backend systems, like provisioning, performance and capacity monitoring, but also for operating processes. Good documentation systems will need to be in place to manage the PON capacity and planning. Operators will also need to ensure operating processes and knowledge is in place for the technician dispatch and order fulfillment to understand which customer premise equipment (CPE) to deploy during each customer activation, whether it be the XGS-PON ONT or the EPON ONT.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.2. Customer Migration from EPON to XGS-PON

Operators should consider their PON strategy and how to manage their current PON customer base and technology: will it be capping the growth of the existing PON technology and fully migrating to the new technology? Or will the existing PON technology also be maintained? When considering migration of existing customers, operators will need a plan. There are additional considerations if it is EPON to XGS-PON transitions, or any other migration where PON coexistence is not possible. The existing PON port and fiber cannot be merely reused without field work. EPON/10G EPON and GPON/XGS-PON will have to be on separate splitters as they cannot coexist on the same fiber since they use the same wavelengths, as shown in Figure 2. Operators may want to think about a customer-by-customer transition plan for a mix of customers on EPON, RFOG and EPON, and XGS-PON. In a centralized architecture, splitters can be dedicated to a PON technology. Migrating existing PON customers would involve moving the drop cable from one splitter to the other likely in the same fiber distribution hub. For the first XGS-PON customer activation, a field technician would activate the new splitter for XGS-PON, and then customer-by-customer be able to transition from EPON to XGS-PON. The drop cable would need to be moved for each customer at the splitter. CPE swaps would be required where an XGS-PON ONT would be installed, and the EPON ONT and RFOG ONU, if applicable, would be removed.

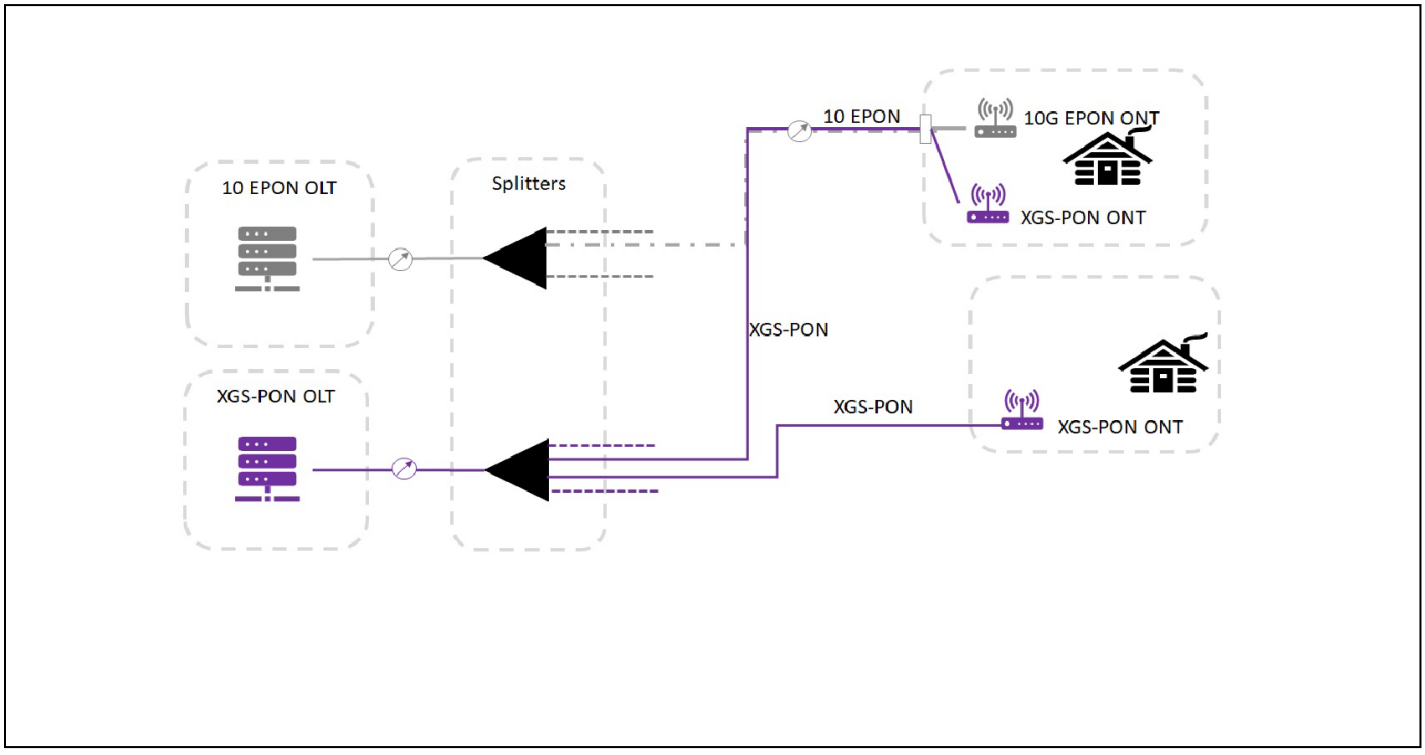


Figure 5 – Customer Migration in a Centralized Split Example

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In a distributed architecture, splitters and taps are located in various locations in the field to reach customers. Migrating existing PON customers would involve careful planning. It will not be operationally easy to do these migrations in large batches without risk of impacting customers for longer periods of time, involving many field resources in coordinated efforts, and/or additional fiber builds. Potentially all customers would be switched over when the fiber to the OLT port is moved. If a new splitter and additional fiber builds from the OLT to the splitter and taps is possible, groups of customers migrating at the same time would be fewer. Perhaps providing the customer with the new ONT equipment and asking them to install is an option but operators will need to consider fiber handling and safety procedures. General anticipation of an increase in customer technical support calls may also need to be planned for, as with any new technology launch.

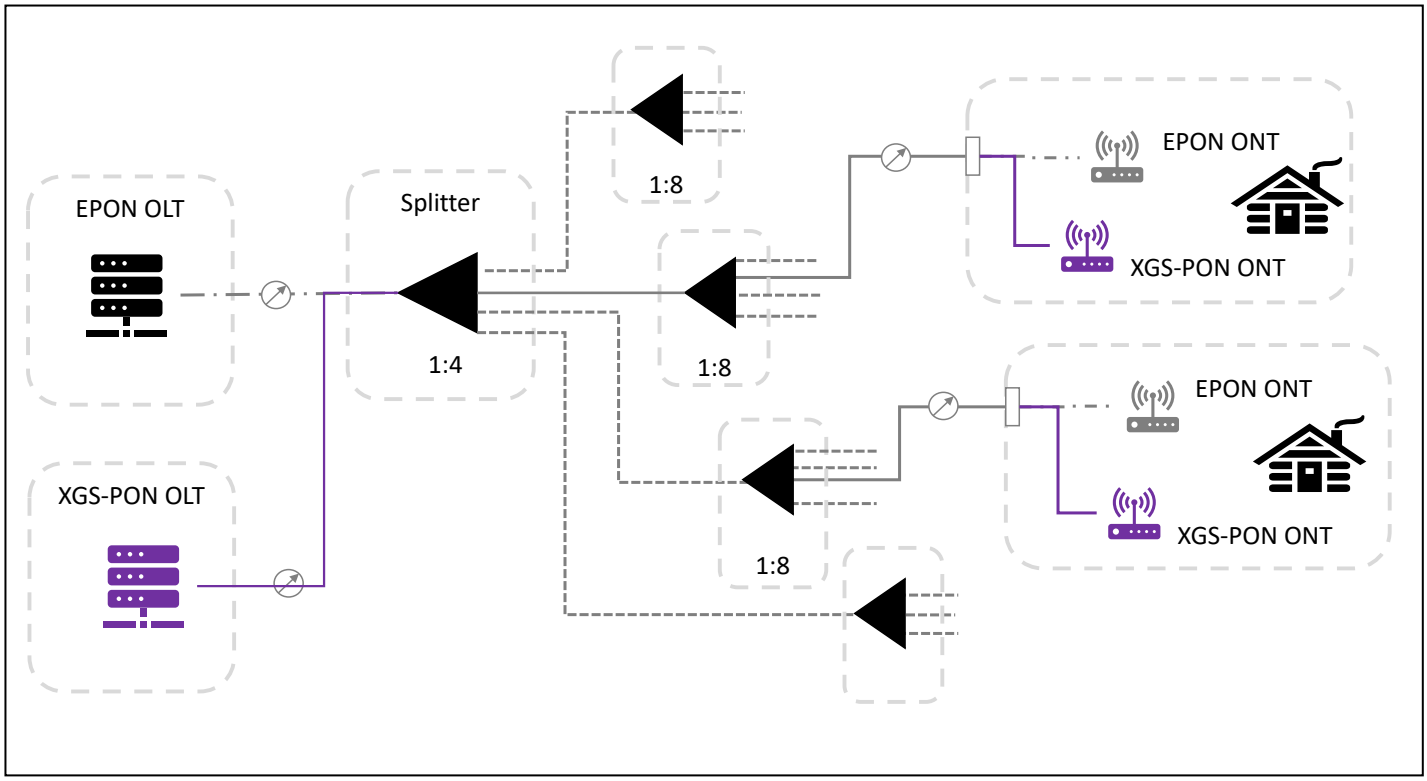


Figure 6 – Customer Migration in a Distributed Split Example

In either split architecture, operators will need to review a plan for capping and growing their PON technologies, and the customer migration strategy for their network scenarios to ensure minimal support requirements.

4. Service Catalog

Operators should review their services to ensure compatibility with the new PON technology. It will be useful to review the entire service catalog in the PON technology transition and determine which are being offered in the new architecture. Re-design and solutioning of each service on XGS-PON may be required. Examples may include speed tiers definitions, IP-only services, and niche services like – hospitality video, elevator service lines, static IPs, lobby cameras, just to name a few. Every service an operator plans to offer will need to be rearchitected and tested prior to launch.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Operators need to have clear plans for all service offerings not just the big three (Voice, Video, Data). Engineering and design teams are not always aware of the work arounds done in the field to support these services (like lobby cameras).

4.1. Services on RFoG

RFoG may have been used with or without an EPON overlay. RFoG allows for quadrature amplitude modulation (QAM) based services to be extended over fiber.

There are several complexities involved with running RFoG. One is with optical beat interference (OBI), when cable modems in an RFoG network transmit simultaneously in the upstream cause interference between the RFoG ONUs, resulting in upstream packet loss. OBI mitigations could include: only using one DOCSIS SC-QAM upstream channel, or specific scheduling at the CMTS capable of detecting OBI issues. A second complexity is with the optical power levels of RFoG and the nearby PON wavelengths. The RFoG optics can overpower those of PON and will result in challenges with RFoG and the 10G PONs. Lastly, RFoG equipment has been difficult to source over time.

Transitioning away from RFoG requires a plan to cap the services being offered on RFoG. The RFoG services will need to be converted to IP services to migrate to PON. So, while the RFoG architecture has served its purpose, as operators transition to all IP-services the need for RFoG is reduced. Migrating from EPON to XGS-PON will be an opportunity to review the transition away from RFoG.

4.2. Speed Tiers

Also required is the decision on which speed tiers to offer on XGS-PON. Each speed tier profile needs to be tested and validated and added in the end-to-end system, from the provisioning system to the billing systems. The operator will need to decide if they want to support all, only some of the current speed tiers offerings, or new tiers on the new PON technology. Servicing tiers over PON allows for potential opportunity to offer higher upstream speeds.

4.3. Specialized Services

There may also be specialized services that operators currently offer that may require re-architecting or review of the solution for XGS-PON. While these services may not be large revenue generating services, their importance should not be discounted and potential redesign should not be overlooked. Example services and their design considerations are listed in Table 2.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 2 – Special Services Design Considerations

Special Services	Design Considerations
Layer 3	Which control system to utilize: BNG, PCRF or other policy control system
Static IPs	Determine how to configure the customer
Hospitality Video	Multicast streaming to the hospitality site will be a priority
Elevator Service Lines, Single Play Voice	Plan for a voice-only service option
Lobby Cameras	Ensuring these can be delivered via IPTV in some manner will be essential.
Third Party Internet	Ensuring the TPIA architecture can efficiently deliver the traffic to the hand off location(s) is important.

In some cases, the CPE swap from EPON to XGS-PON in business customers may be more involved. Customer devices such as routers or firewalls may also need a reconfiguration. Operations will need to plan ahead to avoid delays during the migration. For example, some business customers have third party providers who manage their internal networks. Coordination with customers is important to make sure their IT resources are available during the transitions.

5. Transitioning from DPoE to Native Provisioning

Operators need to think about a plan for the transition from DPoE to GPON provisioning. Here are some things to consider.

5.1. Monitoring

DPoE uses a virtual cable modem (vCM) that emulates a response to all DOCSIS MIBs. For DPoE monitoring the vCM is accessed for metrics. In GPON, the metrics are located at the OLT. An operator would need to plan how and what metrics to monitor on a GPON system. With DPoE, operators may not have fully utilized all the PON related monitoring metrics, especially since DPoE was likely utilized to take advantage of DOCSIS features and making the EPON ONT look and feel like a DOCSIS cable modem. There may not have been an urgent need to dig further into PON specific metrics. New and different metrics like optical levels, signal to noise ratios are not required for DOCSIS. As well, whether to capture and use the native GPON metrics from the OLTs and their element management systems, or to utilize the element management systems themselves will needed to be decided.

In transitioning from DPoE to XGS-PON, new monitoring metrics and modifications to the methods used to collect and report on these metrics will need to be considered.

5.2. Provisioning

Provisioning of PON services varies with technologies. DOCSIS provisioning uses cable modem files and configuration files, while DHCP is used to tell the cable modem where to get the configuration files from. GPON provisioning is done on the OLT and the northbound broadband network gateway (BNG). GPON

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

uses queues for traffic separation or quality of service (QoS), compared to Service Flows used in DOCSIS.

Operators will need to determine how to transition from DOCSIS configuration files which are in the form of type-length-value tuples (TLVs) to GPON configuration. In addition, operators will need to determine a way to differentiate which customers are using which technology and therefore which provisioning system is used – DOCSIS or GPON. A decision will need to be made regarding where each provisioning system will be used, for example at the billing system, or in orchestration or somewhere else. Another decision operators will need to make is whether the operator is building the differentiation decision into their billing system, orchestration, or building a new service to determine this.

5.3. Access CPE Device Management

DPoE systems inherently manage EPON ONT devices via the vCM. The vCM has a management internet protocol (IP) that the DPoE system emulates. In GPON, management of the GPON ONT is done via the OLT and often this is via an element management system (EMS) of the OLT. Any other management systems and any other tools will need to understand how to talk to the EMS when information is required about the ONT itself.

Operators will need to think about how to do policy management, charging, and authentication. policy and charging rules function (PCRF) is responsible making policy decisions and charging rules, which control how subscribers use network resources. The different PON technologies will manage this function differently and operators will require a detailed understanding for integration purposes. An operator will need to decide where this policy is kept, for example: in an element management system, in the BNG via local user database (LUDB), or PCRF.

5.4. Network Architecture

In the PON network architecture, there are a couple of decisions operators may contemplate. With the OLT, there are centralized, chassis-based OLTs typically deployed at a hub site; and distributed OLTs where the OLT is installed in the field like a node. The decision to use one or both may depend on items like the serving area and its distance from the hub site, hub site and current transport architecture, hub site space and power availability etc.

Another consideration is with the broadband network gateway or BNG, which is the traffic aggregator. It is responsible for authenticating subscribers using a LUDB and is the gateway for customer traffic. The BNG could be a layer 3 gateway device that is able to communicate with a PCRF, or operators may decide on a more feature-rich BNG device.

Secondly, operators can decide on whether to use a centralized BNG, or a distributed BNG architecture. A centralized BNG architecture would utilize larger-scale BNGs deployed centrally at hub sites and core locations. A transport network is required to connect to the OLTs. A distributed architecture would use smaller-scale BNG devices, distributed into the regions. These BNGs would be smaller points of failure than a centralized solution. OLTs would likely connect directly to these distributed BNGs. While minimal transport network would be required between the OLT and BNG, additional aggregation points in the core may be required. The design will need to consider expansion and customer scale.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

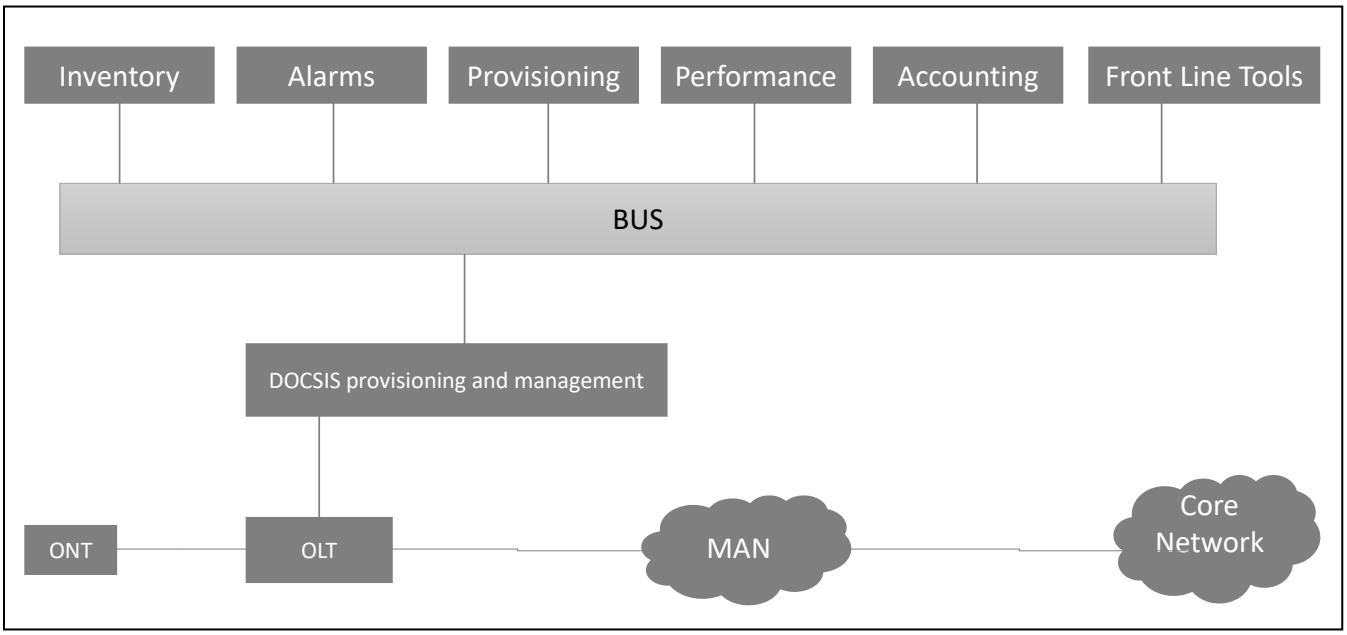


Figure 7 – DPoE Network Block Diagram

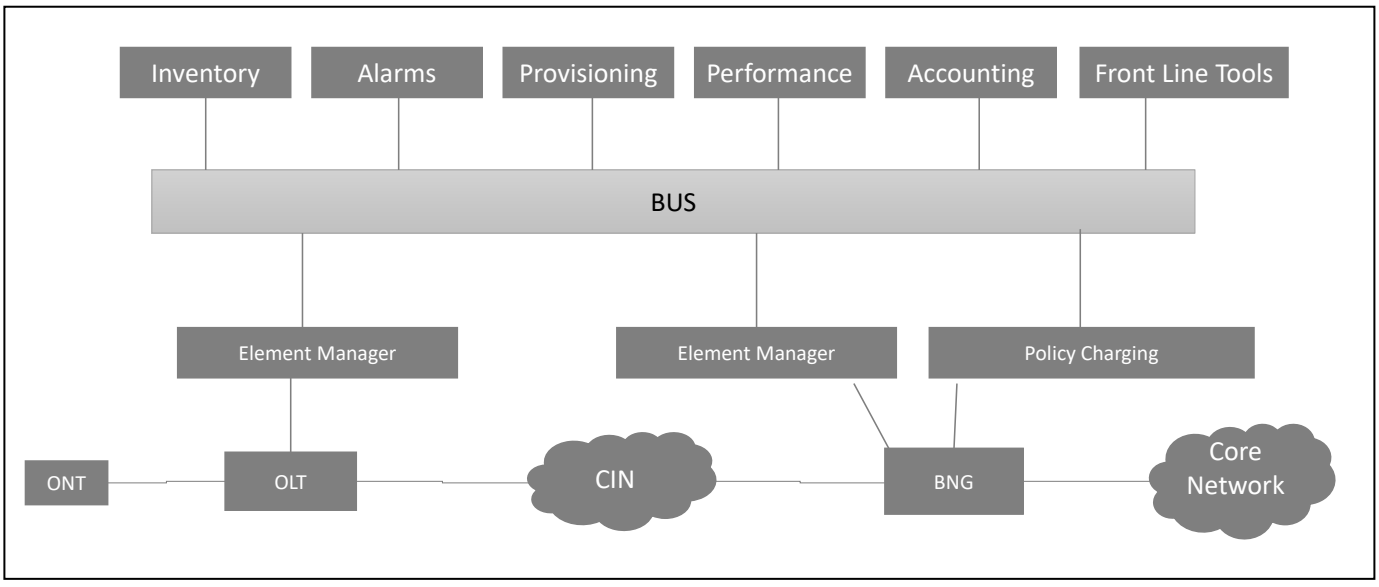


Figure 8 – GPON Network Block Diagram

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Program and Organizational Change Management

Introducing a new technology and architecture will impact many departments. A clear program objective is very important. Stakeholders aligning at the beginning of the program of the minimum deliverables of the program will reduce confusion and debates and make overall delivery more efficient.

Because there can be so much to learn about a new technology, architecture, and processes, clear roles and accountabilities in the program are beneficial, especially Lead Architects in each domain and knowledgeable end-to-end system primes.

6.1. Development to Production Phases

From the development lab through to the first production launch, there is a lot of testing within and between teams. Establishing and coordinating clear testing schedules will be challenging but is important. An agile approach – using short test cycles and reiteration, works well especially when there are dependencies with other team's test results, configurations and time.

Establishing a production trial period in friendly locations, to test and practice production deployment processes and production networks allow for teams to smoke test the system and identify faults that only show up in true production deployments. The potential for unplanned issues should not be underestimated in making changes of this magnitude. This is also an ideal set up for continued regional training of the technology and operating procedures.

For the initial production deployment phase to ensure first customer deployments runs smoothly, a specialized support team could be formed. The purpose of this group could be to ensure that there is a direct line to all the domains and teams for escalations and quick resolution. This could be especially useful when a technician is at a customer's home and requires assistance and quick resolution. It would not be sustainable to maintain this team for a long period of time, so during this phase it is also a good opportunity to review what escalations and issues are experienced and determine what is required to move out of this phase.

6.2. Organizational Change Management

There are new skills for teams to incorporate into their accountabilities in a transition of PON technology. The support tools and processes are different and will require teams to learn and adapt. Operators should plan and think ahead to prepare all the various teams, like network provisioning teams, field teams, and escalation teams.

With the BNG and the access CPE device management functions in either the BNG and PCRF, LUDB or other policy control systems, operators will need to think about roles and accountabilities including subscriber policy and contracts management, network configurations management and associated functions in the element management system. A review of these functional roles may be required by operators.

In many areas, teams will need to learn new troubleshooting and provisioning skills. While integration into existing tools and systems will occur, teams may need to learn CLI or EMS steps on a fundamental basis to integrate systems and also in exceptional troubleshooting situations. Field teams will need a laptop when it was not previously required. Overall escalation processes and teams may be different due to new accountabilities of the technology.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Training is an important aspect and determining the best timing for the training is equally important. Providing the information too early runs the risk that the knowledge may not be retained. And too long after will not be helpful for initial launches. Establishing a regular cadence of refreshers across departments may also be useful.

There is plenty of change between EPON and GPON technologies affecting many roles across the organization that engaging OCM teams early in the program would be beneficial to assist teams through the transition. Developing a plan on how to manage and get in front of the changes impacting teams will alleviate any frustrations during a period of lots of learnings.

7. Conclusion

It is possible to take a project from concept to production in minimal time, even as the whole company is learning along the way. DPoE is an introductory step towards FTTP and with DOCSIS as a foundation it was mostly seamless to integrate. While deploying XGS-PON is unique compared to DOCSIS systems, operators can rest assured that it is doable to integrate the new technology. The more challenging aspects have to deal with the program and change elements rather than the technology itself. Top three items: picking the right battles, establishing a good execution plan, and make training and communication a top priority.

Operators should concentrate on their program's key deliverables. Establishing these at the onset of the program will help keep everyone moving towards the same objectives. As things inevitably will come up, these principles will help with decision making and determining which issues and initiatives to concentrate on.

Along with the overall delivery objectives, a good execution plan including designating key lead primes in each domain will make things smoother. Having lead architects that are knowledgeable of the end-to-end systems will be crucial. One of the hurdles may be with testing between teams: lab testing, pre-production testing, and field validations before adding customers. Of course, these phases are not a new concept, but keep in mind the scale of testing an entirely new end-to-end architecture, and the domains and teams dependent upon one another. There is not a lot common between DOCSIS and XGS-PON. The back-office systems and operations support systems need to be integrated with the new technology. Allocation of adequate time for each team is not easy. They will be dependent upon each other.

Finally, there is lots to learn with switching technologies, aside from the obvious learning about the new technology, but also potentially new processes and accountabilities that are required. Whether for new teams or seasoned architects, there will be a learning curve. Don't underestimate how much there is to learn and get accustomed to with the shift from EPON to XGS-PON. Everything from field technicians needing to learn new troubleshooting skills, to architects needing to learn the functions of domains like the BNG and PCRF. Training will be required early on to even understand the technology enough to design the network architecture, network design and network components. Training is required to get teams familiar with how to implement the new technology and as deployments span throughout the regions.

Tie this together with great communication and some celebrations along the way. While the technologies are different, it is attainable to make the transition.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

BNG	broadband network gateway
CIN	converged interconnection network
CPE	customer premise equipment
DHCP	dynamic host configuration protocol
DOCSIS	data over cable service interface specification
DPoE	DOCSIS provisioning over EPON
EPON	ethernet passive optical network
FTTP	fiber to the premise
GPON	gigabit passive optical network
IEEE	Institute of Electrical and Electronics Engineers
IP	internet protocol
ITU	International Telecommunication Union
LUDB	local user database
MAN	metro area network
MIB	management information base
OCM	operational change management
ODN	optical distribution network
OLT	optical line terminal
ONT	optical network terminal
ONU	optical network unit
PCRF	policy and charging rules function
PON	passive optical network
QAM	quadrature amplitude modulation
QoS	quality of service
RFoG	radio frequency over glass
SCTE	Society of Cable Telecommunications Engineers
SC-QAM	single channel quadrature amplitude modulation
TLV	type-length-value tuples
TPIA	third party internet access
vCM	virtual cable modem
XGS-PON	10G GPON

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

10 Things You Wanted to Know about 10G GPON&10G EPON | FiberMall, https://www.fibermall.com/blog/10-things-about-10g-gpon-and-10g-epon.htm#1_What_is_PON_10G_EPON_and_10G_GPON

Centralized Splitting vs Disbriturbed Splitting in PON Based FTTH networks, https://community.fs.com/blog/centralized-splitting-vs-distributed-splitting-in-pon-based-ftth-networks.html

What is RFoG | RF over Glass basics | RFoG Architecture (rfwireless-world.com), https://www.rfwireless-world.com/Terminology/RFoG-RF-over-Glass-Architecture.html

DOCSIS Provisioning of EPON (DPoE): Architecture and Services, M, Hajduczenia, M. Toy, E. Mallette, C. Knittle; IEEE Communications Magazine, Sept 2012

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23 CELEBRATING 40 YEARS

2023 Fall Technical Forum SCTE • CABLELABS • NCTA

To Kill-A-Watt-Hour, We Recommend to Hot Standby

A Technical Paper prepared for SCTE by

Richard Grivalsky
Principal Engineer
Comcast Cable
Vermont, USA
+1 (802) 316-6553
Richard_Grivalsky@cable.comcast.com

Brian Allen
Engineer III
Comcast Cable
Tennessee, USA
+1 (865) 437-0259
Brian_Allen@comcast.com

Budd Batchelder, Comcast Cable

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Measurement and Validation ... 4
3. Implementing Hot Standby ... 6
   - 3.1. Stakeholder Alignment ... 6
   - 3.2. Pilot Deployment ... 6
   - 3.3. Change Management Program ... 6
   - 3.4. Scaled Deployment of Hot Standby ... 7
   - 3.5. Deployment Impacts ... 7
4. Conclusion ... 8

Abbreviations ... 9

Bibliography & References ... 9

List of Figures

Title **Page Number**

Figure 1 - Dual-Redundant and Hot Standby Platform Operation ... 3
Figure 2 - SCTE Power Pyramid ... 3
Figure 3 - Dual Panel, Mixed Odd/Even, Distributed A/B Power Schema ... 4
Figure 4 - DAA Test Rack Example ... 5
Figure 5 - Dual-Redundant WebView ... 8
Figure 6 - Hot Standby WebView ... 8

List of Tables

Title **Page Number**

Table 1 - 26-Server Trial Results ... 4
Table 2 - Sample of Dual-Redundant Data from 26-Server Trial ... 5
Table 3 - Sample of Hot Standby Data from 26-Server Trial ... 6

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Operational practices have long been to rely on both power supplies, in network equipment, to prevent service interruption in the event of a single power supply unit (PSU) or input failure. Leveraging platform-level controls, such as hot standby, allows for fully redundant powering schemas while reducing energy consumption.

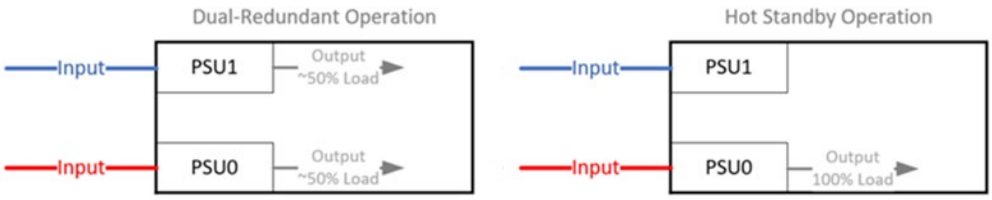


Figure 1 - Dual-Redundant and Hot Standby Platform Operation

With an eye to a greener future, Comcast has set goals to double network energy efficiency by 2030 and to be carbon neutral (Scope 1 and Scope 2 emissions) by 2035. Hot standby presented an opportunity to apply platform-level configuration changes to save power while maintaining reliability. While cable operators can benefit from hot standby at a platform level, when it comes to full-scale solutions some elements in the hardware stack need to be addressed as they do not support hot standby today.

The use of hot standby has the potential to impact 34-50% of the energy pyramid (critical facilities & data centers) by reducing platform consumption by ~6%, which could potentially reduce a cable operator's total consumed electricity by 1.4-3% as more manufacturers integrate the option in additional hardware platforms.

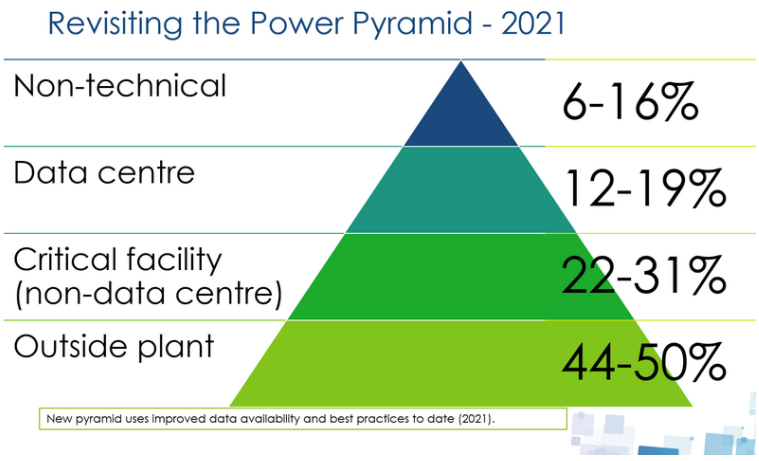


Figure 2 - SCTE Power Pyramid

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Comcast began evaluating energy savings, with hot standby enabled, on several platforms in 2021. Identifying platforms in our next-generation hardware stack and then quantifying the power savings allowed the next phase of conversations to be had, specifically around piloting the energy-saving setting in production equipment.

Our current generation of distributed access architecture (DAA) is put into production with hot standby enabled, reducing power consumption for the hardware stack by ~2%. Today, not all platforms in the stack have the required BIOS or configuration settings that allow for hot standby implementation.

2. Measurement and Validation

Comcast performed rigorous trials of our DAA solution that consisted of a mix of hot standby capable devices as well as network gear that currently does not support hot standby mode in the platform(s).

During these trials, individual circuit-level measurements, as well as combined platform measurements, were captured for 26 network devices in the trial hardware stack. Measurements produced an average energy savings of ~1.7% under maximum central processing unit (CPU) utilization and up to ~6% savings under idle CPU load. Circuit level monitoring was performed for seven days prior to and after initializing hot standby configuration, to compare baseline data in and out of dual-redundant modes. Hot standby was tested under varying loads of simulated traffic.

Table 1 - 26-Server Trial Results

PSU Mode	Aggregated 26-Server Draw (Amperes)	Dual-Redundant to Hot Standby Energy Savings (%)
Dual-Redundant Idle CPU	88.1	--
Hot Standby Idle CPU	82.7	6.5
Dual-Redundant Max CPU	135.3	--
Hot Standby Max CPU	132.9	1.8

Hot standby power supplies were alternated between primary and hot standby, across the power distribution panels, utilizing the odd/even methodology as covered in SCTE 282 2023. This approach offered Comcast additional resiliency while implementing hot standby in production equipment.

DC Distribution Panel 1										
A Side	Platform 1	Platform 3	Platform 5	Platform 7	Platform 9	Platform 11	Platform 13	Platform 15	Platform 17	Platform 19
B Side	Platform 2	Platform 4	Platform 6	Platform 8	Platform 10	Platform 12	Platform 14	Platform 16	Platform 18	Platform 20
Circuit #	1	2	3	4	5	6	7	8	9	10

DC Distribution Panel 2										
A Side	Platform 2	Platform 4	Platform 6	Platform 8	Platform 10	Platform 12	Platform 14	Platform 16	Platform 18	Platform 20
B Side	Platform 1	Platform 3	Platform 5	Platform 7	Platform 9	Platform 11	Platform 13	Platform 15	Platform 17	Platform 19
Circuit #	1	2	3	4	5	6	7	8	9	10

Figure 3 - Dual Panel, Mixed Odd/Even, Distributed A/B Power Schema

Circuit level monitoring was performed via simple network management protocol (SNMP) through a WebView monitoring system, polling the in-rack power distribution panels. This ensured total draw on the DC plant was captured, rather than just that of the platforms themselves. Measurement of the hot standby load savings was recorded at all stages.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

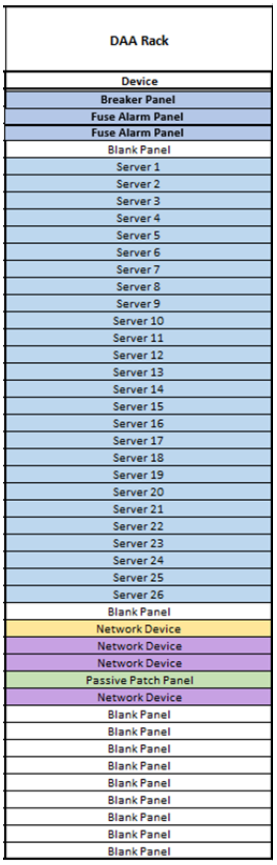


Figure 4 - DAA Test Rack Example

Table 2 - Sample of Dual-Redundant Data from 26-Server Trial

Server	PSU0 (Amperes)	PSU1 (Amperes)	Total (Amperes)
Server1	1.75	1.75	3.50
Server7	1.57	1.68	3.25
Server14	1.86	1.74	3.60
Server25	1.89	1.74	3.63

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 3 - Sample of Hot Standby Data from 26-Server Trial

Server	PSU0 (Amperes)	PSU1 (Amperes)	Total (Amperes)
Server1	0.00	3.32	3.32
Server7	3.04	0.00	3.04
Server14	0.12	2.78	2.90
Server25	0.00	3.03	3.03

Note: panel monitoring sensitivity thresholds resulted in occasional zero-ampere readings for the PSU in hot standby.

3. Implementing Hot Standby

Leveraging SCTE 282 2023, a piloted, data-driven approach was required to implement a successful change management strategy.

3.1. Stakeholder Alignment

Stakeholder identification is critical during the hot standby pilot phase. Impacted teams during the pilot phase include on-site engineering, network operations center (NOC), and engineering leadership. If the pilot is successful and wide-scale adoption is recommended, additional stakeholders including platform teams, procurement, finance, and sustainability reporting should be included.

3.2. Pilot Deployment

The goal of a hot standby pilot deployment should be to collect data to demonstrate energy avoidance while maintaining, or potentially improving, customer reliability. A pilot project plan with scope, milestones, schedule, and measures of success should be documented prior to implementing a hot standby pilot. The plan should be agreed upon by all stakeholders in the pilot phase of the project. Key considerations for the scope of the project include the number of devices to be activated in hot standby, pre- and post-measurements, alarming configuration to eliminate false alarms, and equipment monitoring to verify no-impact implementation.

3.3. Change Management Program

A comprehensive change management program should be rolled out prior to a scaled deployment of hot standby. The stakeholders identified above should be notified of the intent to deploy a hot standby program and given the opportunity to ask questions and/or contribute to the deployment. The Prosci ADKAR® model prescribes five key elements of change management:

- Awareness- of the need for change,
- Desire- to participate and support the change,
- Knowledge- on how to change,
- Ability- to implement desired skills and behaviors, and
- Reinforcement- to support the change.

For a hot standby scaled deployment, project champions should develop a business case justification to garner executive-level approval to expand the effort from the pilot program. The data collected from the pilot program should be presented and extrapolated to quantify the potential benefit of scaled deployment.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Assumptions should be documented. The business case should generate both awareness of the need for hot standby deployment as well as a desire to scale the effort.

A full-on change management program should be implemented once executive approval is secured. User groups such as engineering and NOC teams can be brought through the change management process, including the reason for the change and a description of why the change is beneficial for operations. End users should also be provided with function-specific change guidance, including training to implement hot standby, expectations for schedule, and performance measurements.

3.4. Scaled Deployment of Hot Standby

Scaled deployment of hot standby should be overseen by a dedicated project team with delegated responsibility to end users. Deployment plans should account for other initiatives being implemented at a site(s), as well as stakeholder notification and post-deployment monitoring.

3.5. Deployment Impacts

As covered in SCTE 282 2023, several user communities are impacted when deploying hot standby instead of dual-redundant powering schemas. Hot Standby changes how PSU utilization is viewed, monitored, and alarmed.

Traditionally, end users expect to see circuit-level monitoring, using dual-redundant mode, wherein the circuit-level current draw is approximately equal for each PSU.

With hot standby, end users will have to be retrained to expect the primary PSU circuit-level current draw to match the full load of the platform and a very small draw, for status monitoring purposes, measurable at the hot standby PSU.

Caring for these changes, polling and reporting platforms will require modification to properly display hot standby configured platform power consumption. Alarming modifications will also be needed to appropriately dispatch resources, if a failure is seen, when hot standby is enabled.



© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 5 - Dual-Redundant WebView

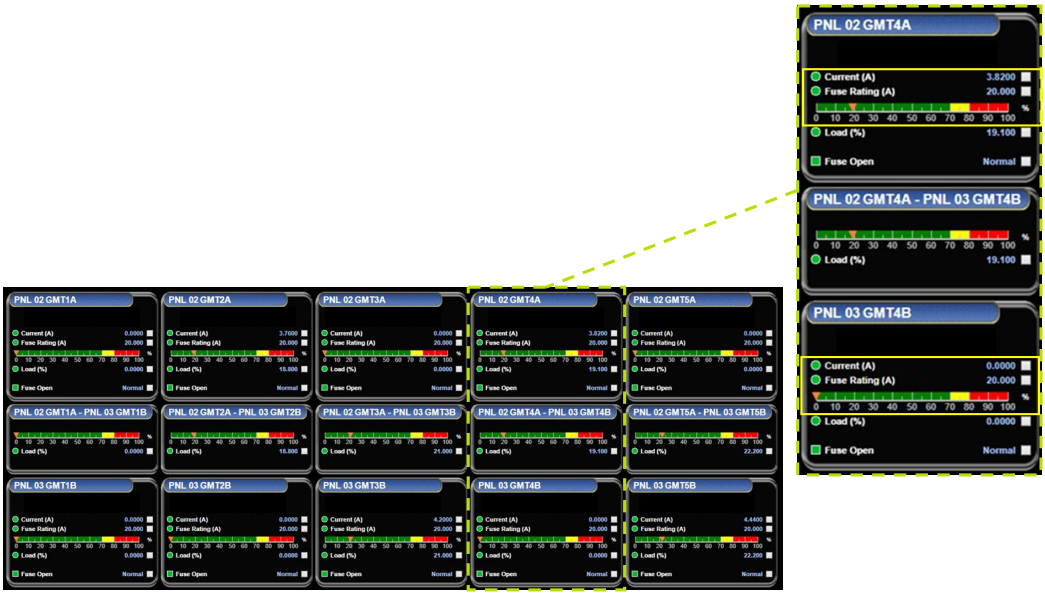


Figure 6 - Hot Standby WebView

4. Conclusion

Our analysis and quantification of energy savings have proven advantageous and impactful toward Comcast's sustainability goals. Operators would benefit from hot standby implementation, whether it drives toward a sustainability goal, or simply to reduce consumed electricity. Our next steps are to continually refresh requests to original equipment manufacturers (OEM) to provide additional platform options for DAA and other network solutions to be deployed with hot standby schemas enabled, striving for the upper limit of the ~3% in consumed energy savings.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

CPU	central processing unit
DAA	distributed access architecture
DC	direct current
NOC	network operations center
OEM	original equipment manufacturer
PSU	power supply unit
SCTE	Society of Cable Telecommunications Engineers
SNMP	simple network management protocol

Bibliography & References

SCTE 282 2023 Implementing Inside Plant Network Platform Hot Standby Powering

Figure 1, SCTE Power Pyramid, https://www.scte.org/standards/development/energy-2020-powering-cables-success/

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Too Many Cooks in the Kitchen

Fostering Organizational Cohesion by Digitizing the RPD

A Technical Paper prepared for SCTE by

De Fu Li
Distinguished Engineer
Comcast
1701 JFK Boulevard Philadelphia, PA 19103
defu_li@comcast.com

Gregory Medders, Comcast

Eric Stonfer, Comcast

Bhanu Krishnamurthy, Comcast

Sinan Onder, Comcast

Mehul Patel, Comcast

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

Abstract ... 3
1. Introduction ... 3
   1.1. iCMTS ... 3
   1.2. vCMTS ... 4
   1.3. Challenges ... 5
2. Scaling Out DAA Deployment ... 5
   2.1. vCMTS Automation ... 5
   2.2. Microbursts ... 6
   2.3. Profile Management Application ... 9
   2.4. Midsplit ... 12
3. Digital Transformation of the Access Network Infrastructure ... 13
   3.1. Concepts ... 14
      3.1.1. Digital Twin ... 14
      3.1.2. Global Object Store for Digital Entities ... 15
      3.1.3. Data Catalog ... 15
   3.2. RPD Digitization ... 15
      3.2.1. Use Cases ... 15
      3.2.2. Model ... 16
4. Conclusion ... 17

Abbreviations ... 18

Bibliography & References ... 19

List of Figures

Title **Page Number**

Figure 1 – MHAv2 Architecture (Source: [1]) ... 4
Figure 2 – Number of production PPODs and RPDs deployed vs time. ... 6
Figure 3 – CIN Based on a Leaf-Spine Network Architecture ... 7
Figure 4 – RxMER Heatmap of a Cable Modem's OFDM Channel ... 10
Figure 5 – vCMTS Application Architecture ... 10
Figure 6 – OFDM profile update in iCMTS vs vCMTS (Source: [1]) ... 11
Figure 7 – Total Count of Modems in OFDM Partial vs Time for a Given RPD ... 12
Figure 8 – (a) Unequalized Probe Data; (b)-(d) Amplitude Response Samples from Three Different Modems ... 13
Figure 9 – Simplified Process for Provisioning a New RPD Based on Capacity Demands ... 16
Figure 10 – Digitized RPD and its Subdomains in a Data Mesh ... 17

List of Tables

Title **Page Number**

Table 1 – Service Group Parameters ... 8
Table 2 – Downstream Data Traffic Model ... 8
Table 3 – DOCSIS MMMs vs Data Traffic Ratios ... 9

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abstract

The distributed access architecture is enabling Comcast's ambitious program to offer symmetric, multi-gig speeds to its customers over a full-duplex (FDX) hybrid fiber-coax network. Over the past years, we have learned many valuable lessons required to activate 1000s of remote physical layer devices (RPDs) a week that we will apply to the D4.0 rollout. However, a production scale deployment of this technology involves many interdependent processes and components, including planning and capacity management, plant design, logistics, construction, provisioning, and operations.

Furthermore, the FDX deployments often take place incrementally over infrastructure that already serves live customers, requiring precise coordination across teams and organizations to ensure a good customer experience. Left unchecked, teams may unintentionally interfere with one another and spoil the work of other groups, much as too many cooks in kitchen may spoil a meal.

Sharing our lessons learned in our mid-split transitions, we describe how a digital transformation of the underlying access network provides a mechanism to avoid the worst manifestations of Conway's law. Because the RPD lifecycle is complex, different domain experts are required at each phase. This can result in ambiguity over who is responsible for the RPD, since it changes over time.

Through a digital transformation of the RPD, we show how a federated data service can be used to democratize data and present guardrails that must be in place to ensure its integrity. Crucially, while the same teams still own the same aspects of the RPD lifecycle, by communicating through the common digital RPD, the scaling implied by Conway's law is mitigated, enabling a smoother transition to FDX. In the future, the digitization of the cable access networks will fundamentally change how Comcast operates, enhancing its ability to provide a reliable, cutting-edge experience to customers.

1. Introduction

Since the days of dial-up, internet service providers have balanced the need to provide reliable services in the current market while anticipating the future growth and data needs of the diverse applications enabled by the internet. Whether due to a surge of downloads prompted by a new gaming platform or the unprecedented shift to working from home during the COVID-19 pandemic, anticipating and exceeding customer internet connectivity requirements is a constant challenge. With the introduction of symmetric multi-gigabit speeds afforded by the most recent standards, cable operators have the opportunity to deliver the fastest speeds at the lowest latency, powering the residential and commercial computing needs of tomorrow.

1.1. iCMTS

The data over cable service interface specification (DOCSIS) standard defines how broadband internet can be distributed over a hybrid fiber-coax (HFC) network. In traditional DOCSIS, the key network elements are the cable modem (CM) and the cable modem termination system (CMTS). As a form of customer-premises equipment (CPE), the CM facilitates connectivity to the internet via the CMTS, a device typically deployed in cable operators' head-end facilities that can serve the control and data needs of thousands of CMs communicating over the shared HFC network.

Earlier generations of CMTSs were built as custom hardware platforms. These are often termed "integrated" CMTS (iCMTS) because they utilize an integrated DOCSIS MAC and PHY function.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

More recently, the distributed access architecture (DAA) specification, also known as the modular head-end architecture version 2 (MHAv2), was introduced [1]. DAA splits the CMTS into two distinct components: the physical (PHY) function and the core function. The remote PHY device (RPD) provides the PHY function, while the core functions consist of CMTS operating on the medium access control (MAC) and internet protocol (IP) layers.

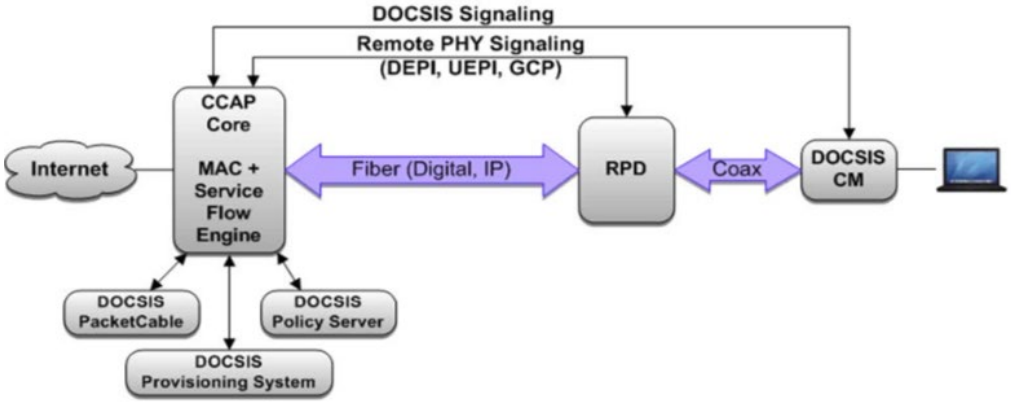


Figure 1 – MHAv2 Architecture (Source: [1])

For the purposes of this paper, iCMTS is used as a representation of custom hardware CMTS platforms.

1.2. vCMTS

The PHY/core split allows significant flexibility in how the CMTS functionality can be implemented. A key feature enabled by the split is the ability to run the CMTS on commercial off the shelf (COTS) hardware, rather than relying on custom hardware. This in turn provides an opportunity to modernize the CMTS, transitioning it from an integrated hardware solution to a collection of software applications or microservices, many of which can even run on a public cloud computing platform. Taken together, this architecture is referred to as the virtualized CMTS (vCMTS).

To smooth the transition from iCMTS, some vCMTS implementations were built to look and feel exactly like an iCMTS through familiar features such as the command line interface (CLI) and the SNMP (Simple Network Management Protocol) interface. In principle, these common interfaces abstract away the iCMTS/vCMTS distinction, allowing the multiple system operators (MSOs) to deploy and operate a vCMTS in a black box way as with iCMTS. However, treating a vCMTS like "just another CMTS" would neglect the significant opportunities provided by modern software and cloud-native architecture.

Adopting a modern, containerized architecture allows us to leverage open-source infrastructure tooling and software. The core vCMTS applications are deployed using Kubernetes, which enables features such as high-availability and in-service software upgrade. From an end-user perspective, one of the biggest advantages of the move to cloud-native architecture is real-time observability. For example, we monitor logs using the Elasticsearch, Fluentbit, and Kibana (EFK) stack. Rather than relying on SNMP polling, real-time telemetry about the customer experience is obtained using Prometheus, a time series database (TSDB). Additionally, end-to-end automation of a vCMTS is very broad in scope, including infrastructure as code, site stand up, deployment, incident and change management, alerting, and automated remediations.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

While some of the microservices in the vCMTS ecosystem can (and do) run in the public cloud, there remain some aspects of a vCMTS that are difficult to operate in a public cloud environment. Very precise timing must be maintained between the RPD and the core (e.g., for DOCSIS upstream bandwidth allocation maps (MAPs)). For this reason, routing communication from the RPDs to the core over the internet backbone (or even over dedicated connections to a public cloud provider) introduces variation in latency and congestion that could impact the subscriber experience. For this reason, Comcast operates a hybrid model utilizing both the public and private cloud. The data plane always resides in the private cloud, however, the control and management plane functions are more flexible, allowing them to utilize the public cloud where appropriate.

1.3. Challenges

Achieving symmetric gigabit speeds is not without its challenges. In addition to the complexities of qualifying new CMs and CMTS for the DOCSIS 4.0 spec, FDX-capable RPDs and amps must be installed in the field. Network equipment between the RPD and the vCMTS, including switches, optics, and network interface cards (NICs), must have the capacity to support many customers with symmetric, multi-gigabit services. The utilization of the radio frequency (RF) spectrum must be changed, since legacy video services are often served in the same frequencies of the new FDX band.

While each of these changes alone is challenging, perhaps the most challenging aspect is the coordination required to deploy these deeply impactful changes while maintaining an outstanding customer experience. In this paper, we will discuss the strategies Comcast is using to deliver the most reliable services via next generation access technologies.

2. Scaling Out DAA Deployment

2.1. vCMTS Automation

The ability to leverage COTS hardware in a vCMTS provides many benefits, such as reduced cost, reduced power consumption, protection against vendor lock-in, and flexibility to change hardware during supply chain crunches, such as those that have occurred since the COVID-19 pandemic. However, these benefits come at a cost since operationalizing a vCMTS is more complicated than a black box iCMTS. At a high-level, the vCMTS edge cloud can be broken down into:

- The primary pod (PPOD): A collection of servers connected to a leaf-pair, collectively forming the Kubernetes cluster in which the vCMTS workloads run [2].
- A converged interconnect network (CIN): the network elements connecting the PPOD to the access network devices (e.g., the RPD).
- The RPDs themselves, which connect to the coax network and ultimately the end users.

In our previous paper “Humanoids Optional: Deploying vCMTS at Scale with Automation” [3], we discussed how learnings from the DevOps movement were used to automate the PPOD and CIN operations [4]. By focusing on flow (automated testing and deployment) and feedback (observability and risk-aware deployment), automation has enabled us to exponentially grow the number of vCMTS clusters deployed at Comcast.

From their inception, the PPODs are assembled in a standard way in the warehouse and shipped to the head-end. Cluster standup pipelines begin by validating all physical layer connections; from that point forward, no humans are directly involved in the initial cluster configuration or any subsequent changes. Rather than assigning operations engineers to manage clusters or network elements individually, the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

PPOD and CIN join a logical fleet that makes up the vCMTS edge cloud. Software and network engineers are then empowered to initiate changes to software versions or network configuration templates against the entire edge cloud. The changes are incrementally rolled out across the fleet, with automated health checks ensuring that the change is successful and does not degrade services.

This extensive investment in automation is motivated by our initial experience operating a vCMTS. During our first vCMTS trials, we approached vCMTS operations much like we would a traditional iCMTS, with heavy reliance on operations engineers to configure and maintain PPODs. While the goal was to maintain identical clusters (aside from intrinsic differences, such as which IP addresses were assigned to the clusters), small differences in the hand-configured systems accumulated over time into more significant variation between the clusters.

Over time, this configuration drift made what should have been trivial changes very difficult to implement. Learning from these initial experiences, we invested heavily in end-to-end automation of a vCMTS. Now that each cluster is deployed identically and all changes are automated, we have essentially eliminated outages caused by human error. Additionally, the automated vCMTS standup enables us to bring new vCMTS clusters online within a few hours of it being connected in the head-end, enabling the rapid expansion of the vCMTS edge cloud (Figure 2).

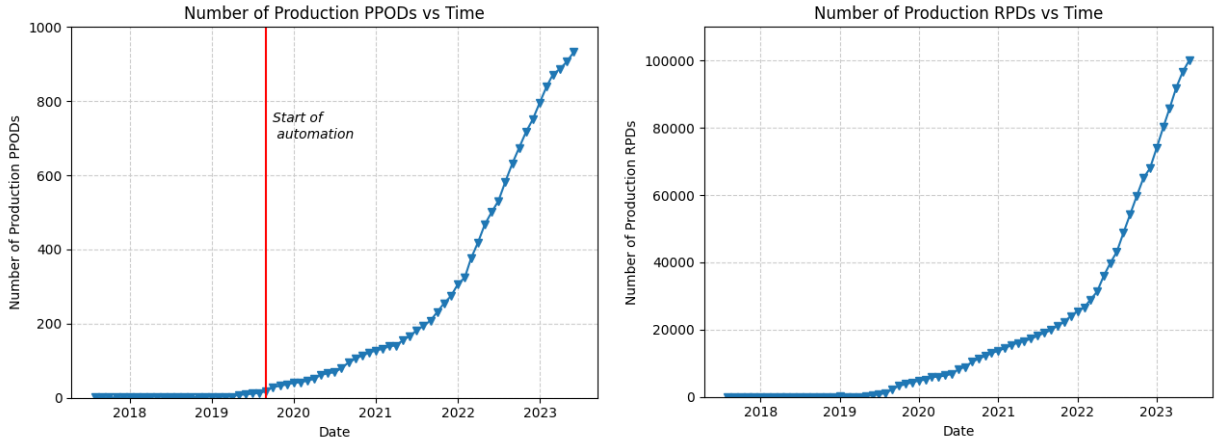


Figure 2 – Number of production PPODs and RPDs deployed vs time.

2.2. Microbursts

End-to-end automation simplifies the debugging of unexpected issues when they arise. Because every system is identical, issues are often either observed in all systems (in which case, the automated health checks that are integrated into every change catch the issue early in the deployment) or in one system.

When issues are present in only one system, the cause is typically attributed to differences in the underlying usage (e.g., number of RPDs, differences in CM models or behaviors). We encountered one such unexpected issue while scaling out our DAA deployment; in a single site, we began observing an increase in CM flapping, where modems are going offline and reregistering or going into partial mode.

To understand the cause of the CM flapping it is helpful to review the CIN architecture. As shown in the simplified figure below, the CIN is based on a typical data center leaf-spine architecture. The hosts are connected to a pair of leaf switches. The leaf switches are then connected to a pair of spine switches. Each

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

RPD is connected via a 10 Gigabit link to an aggregation switch; other than between the aggregation switch and the RPD in the field, each network hop has a redundant path in case of failure.

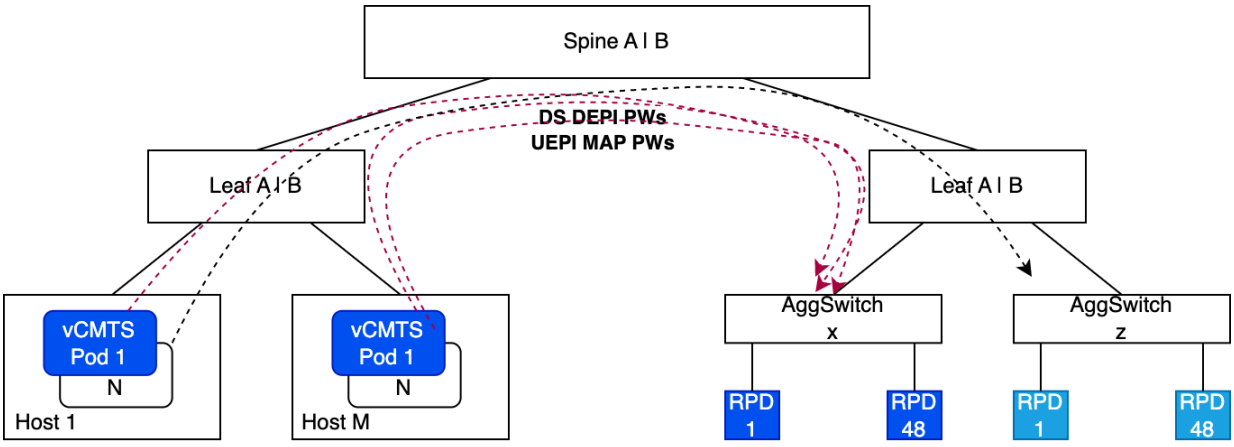


Figure 3 – CIN Based on a Leaf-Spine Network Architecture

Network traffic patterns tend to be site specific (i.e., tied to a specific CIN domain) and are influenced by subscriber and RPD population density. Traffic between the PPOD and RPD is also asymmetric, with a higher bitrate in the downstream direction from the provider toward the subscriber. When investigating the CM flapping, we noticed that this particular site was experiencing microbursts (short spikes in the network traffic) that were saturating the egress queue of the spine. These microbursts can lead to congestion and dropped packets; when the congestion across the CIN is severe enough, this manifests as momentary modem flapping and out-of-sequence packet events.

When investigating the microburst phenomenon, we found that in some situations the DOCSIS map traffic can be a key contributor to microburst activity. Below, we perform a simple calculation to illustrate the network traffic associated with maps. Table 1 shows an example 1x2 service group (SG) configuration with 48 downstream (DS) single carrier quadrature amplitude modulation (SC-QAM) and 6 upstream (US) SC-QAM per port. As was common at the time, every 4th DS SC-QAM was primary capable.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 – Service Group Parameters

DS SC-QAM	48	32, 36, 40, 44, 48	
Primary Capable SC (every 4th or 8th)	4	12	
US SC-QAM per Port	6	4, 5, or 6 upstreams	
US Ports	2	1x1 or 1x2 SGs	
Voice Calls	0		
Voice Frame	20	msec, G.711 or G.722, default to G.711 as the worst case.	
modems	500	D2.0 is less than 5%; assuming all are 8 channel capable 3.0+	
RNG SM-Internval	14	sec	
T4 Multiplier	2		
Map Size	4	msec	
MDD Interval	2	sec	Secondary MDD is small.
UCD interval	2	sec	
OCD on PLC	200	msec	
DPD on PLC	200	msec	
OCD on ProfiA	500	msec	
DPD on ProfA	500	msec	

To examine the impact of subscriber utilization of the node, we model the downstream data in Table 2. We note that packet size in the downstream direction tends to be bimodal, with packet sizes between 1024 to 1518 bytes accounting for the majority of downstream data. Combined with the layer 2 tunneling protocol (L2TP) header overhead, we arrive at the total packets per second and bytes per second in the downstream direction.

Table 2 – Downstream Data Traffic Model

Subscriber traffic parameters			
SG Capacity (bits/s)	SG Avg Utilization	L2TP Header Overhead - no pkt coalecing, IPv6 L2TP Header	
2,500,000,000	20.0%	67	
DS Packet Size (bytes)	Percent of DS traffic in that size bin	Packets per second (PPS)	BPS /w L2TP Header
1024 - 1518	87.65%	43,101	461,352,085
512 - 1023	1.26%	1,026	6,849,967
256 - 511	1.31%	2,135	7,694,329
128 - 255	2.90%	9,465	19,573,107
65 - 127	5.80%	37,760	49,239,583
< 64	1.08%	10,547	11,053,125
	Totals	104,034	555,762,196

Combining the results from previous tables, in Table 3 we break down the DS DOCSIS traffic for our representative service group into two categories: DOCSIS control mac management message (MMM) data and subscriber data. Notably, while the amount of data utilized by control messages is small relative to the actual subscriber data (7%), the overhead of handling the DOCSIS control data can be significant from a packet perspective. Under these parameters, the ratio of control to data packets was 41%. Given that a CIN may have thousands of SGs, DOCSIS map traffic can represent an unexpectedly large portion of the packets traversing the CIN.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 3 – DOCSIS MMMs vs Data Traffic Ratios

		1 SG		
		PPS	Avg Pkt	BPS
Committed Information Rate (CIR)	UEPI Maps to RPD	3,000	120	2,880,000
	MAPs to primary DS	39,000	120	37,440,000
	MDD	25	520	101,920
	UCD	30	530	127,200
	OCD	7	176	9,856
	DPD	7	114	6,384
	RNG-SM	107	211	180,857
	Voice	0	280	0
	Control Sub Total	42,176	121	40,746,217
Data DS PSP Flow 0	Priority - Best Effort	104,034	668	555,762,196
	Total	146,210		596,508,413
	Control vs Data Ratio	40.54%		7.33%

Even before the root cause of the modem flapping was understood, rolling out mitigations was easily accomplished via automation. Ultimately, changing the number of downstream capable channels in a service group significantly reduced the impact of the control packets on the CIN. This analysis has also allowed us to improve our CIN network capacity sizing, demand monitoring, and switch vendor selection criteria, prioritizing for egress buffer capacity. Together, the improved ability to troubleshoot and rapidly deploy fixes is part of why automation significantly improves the overall reliability of the vCMTS.

2.3. Profile Management Application

While end-to-end automation is a hallmark of our vCMTS strategy, some automation features such as modulation profile management are utilized in both an iCMTS and a vCMTS. Modulation profile management involves maximizing bandwidth capacity while minimizing transmission errors. Comcast has hundreds of thousands of node segments, each with its own unique radio frequency (RF) characteristics. Attempting to manually manage the modulation profile configuration of each DOCSIS RF channel for every node segment is futile. For this reason, we developed a profile management application (PMA) that programmatically analyzes the DOCSIS RF channels and adjusts their modulation profile settings.

The DOCSIS radio frequency (RF) channel types that require their modulation profiles to be managed are the downstream orthogonal frequency division multiplexing (OFDM) channel, the upstream advanced time division multiple access (A-TDMA) channel, and the upstream orthogonal frequency division multiple access (OFDMA) channel. Figure 4 shows a cable modem's receive modulation error ratio (RxMER) heatmap. The recommended profile built by the PMA is represented by the solid yellow line showing the subcarriers (bit loading) assignment matching to the RxMER heatmap. Details of the profile management techniques, their performance, and benefits can be found in [5], [6], and [7] .

While modulation profile management is used in both the iCMTS and vCMTS platforms, its implementation varies between platforms. Typically, an iCMTS has a built-in PMA feature. In the vCMTS application framework, a PMA is one of the many applications that interact with vCMTS but which are logically separate from vCMTS itself.

As a cloud-native solution, the vCMTS architecture favors the microservice design pattern. This allows us to decompose a complex system into domains that minimize coupling. As independently deployable microservices communicating through versioned APIs, applications such as PMAs are deployed independently from vCMTS and can run in the public cloud to reduce cost via elastic scaling.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

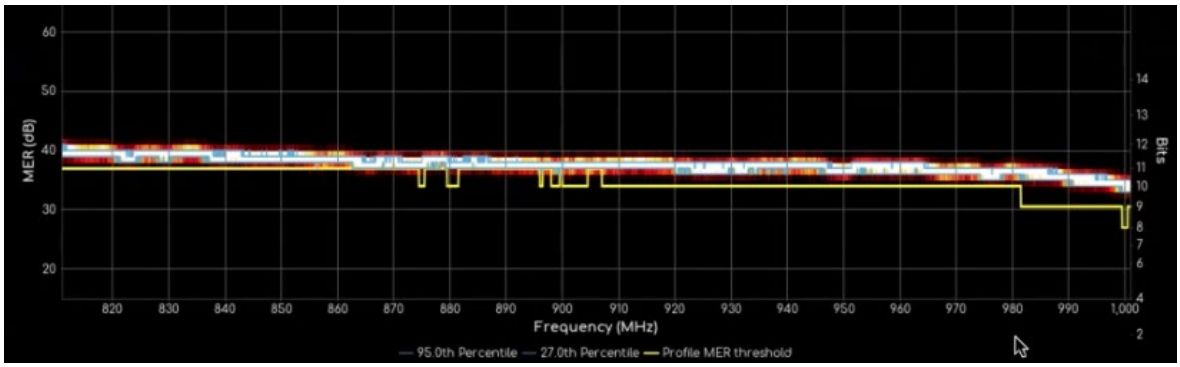


Figure 4 – RxMER Heatmap of a Cable Modem's OFDM Channel

Figure 5 illustrates how the vCMTS application architecture uses a hybrid public-private cloud. The Comcast Edge Cloud is a private cloud deployed as Kubernetes clusters that are physically located in head-ends. Instances of vCMTS are dynamically deployed as Kubernetes pods in the edge cloud and service one or more RPDs. There are currently more than 100,000 of these pods distributed across almost 1,000 Kubernetes clusters in the edge cloud. The PMA resides in the public cloud and ingests vCMTS telemetry data, performs analytics, and recommends profile updates as needed. The recommended profile is set via the service gateway (SGW), which routes the profile change request to the vCMTS pod's API endpoint, ultimately initiating the profile change procedure defined by DOCSIS.

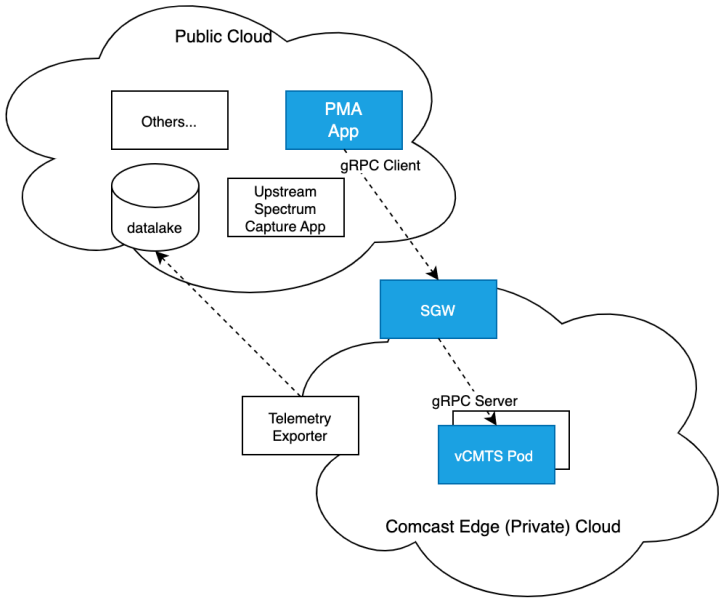


Figure 5 – vCMTS Application Architecture

As more and more nodes are being managed by the PMA, this architecture has scaled well. However, a PMA serving more nodes also means a diversity of customer and RPD devices and firmware versions, which has exposed several interoperability issues. To understand these issues, it is necessary to look a bit deeper into the profile update mechanisms.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

When an OFDM profile is updated, the vCMTS must initiate the DS profile descriptor (DPD) change procedure; when an US SC-QAM or OFDMA modulation profile changes, the vCMTS must initiate the US channel descriptor (UCD) change procedure. The change procedures are specified in both the MAC and upper layer protocols interface (MULPI) and the RPHY specifications, since the RPD also plays a role in RPHY.

The diagram below shows the DPD change procedures for the iCMTS and vCMTS. There is a minimum of 500 msec lead time in sending the new DPD MMM to the cable modems. The vCMTS sends the DPD over the generic control plane (GCP) to the RPD, and the RPD updates the "C" bit in the next codeword pointer (NCP). The "C" bit update is basically an odd-even bit toggle to match the least significant bit (LSB) of the change count field in the DPD message, which is used to signify a successful completion of the change procedure.

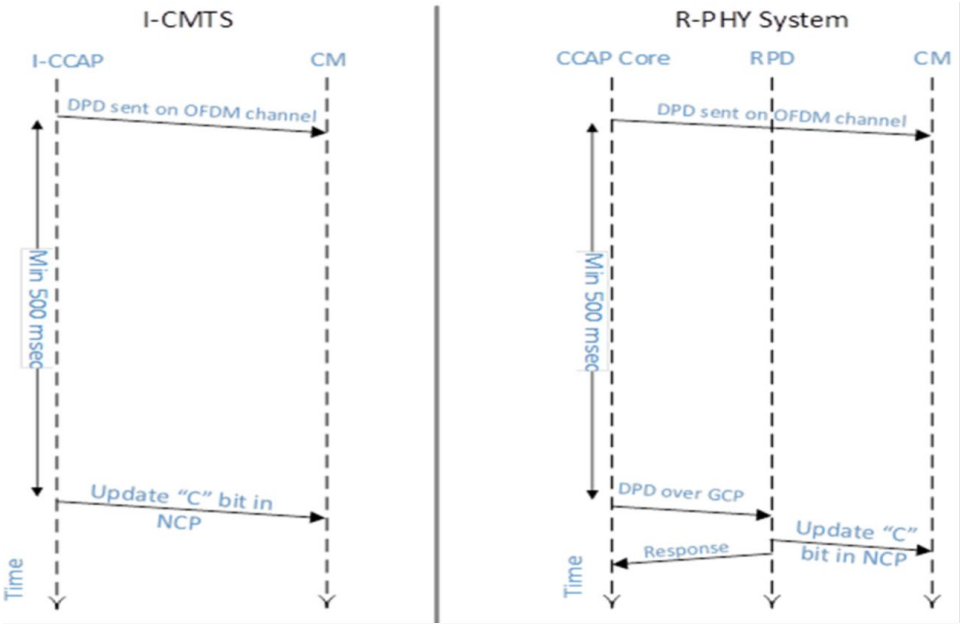


Figure 6 – OFDM profile update in iCMTS vs vCMTS (Source: [1])

When something goes wrong in the change procedure, cable modems utilize the DOCSIS protocol CM-STATUS to report error; this forms a closed loop and allows for recovery. The CM-STATUS reporting implementation varies among the modem chipset vendors. Similarly, RPDs from different vendors vary in implementation behaviors, such as timing sequence, concurrency of multiple DPD changes, timeout handling of "C" bit update.

The failure symptoms include modems going into OFDM partial mode or being unable to receive traffic on the OFDM channel. Figure 7 shows an incident of modems going into OFDM partial mode that was triggered by the PMA recommending a profile change. The time series data labeled "ofdmActual=1x0" (in blue) is the total modem population that acquired the OFDM channel. The other (in green) is the population in OFDM partial mode. As shown, many modems were stuck in OFDM partial mode after the profile change. Such behavior is not unique to the DS, as the UCD change procedures also have their interoperability challenges.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

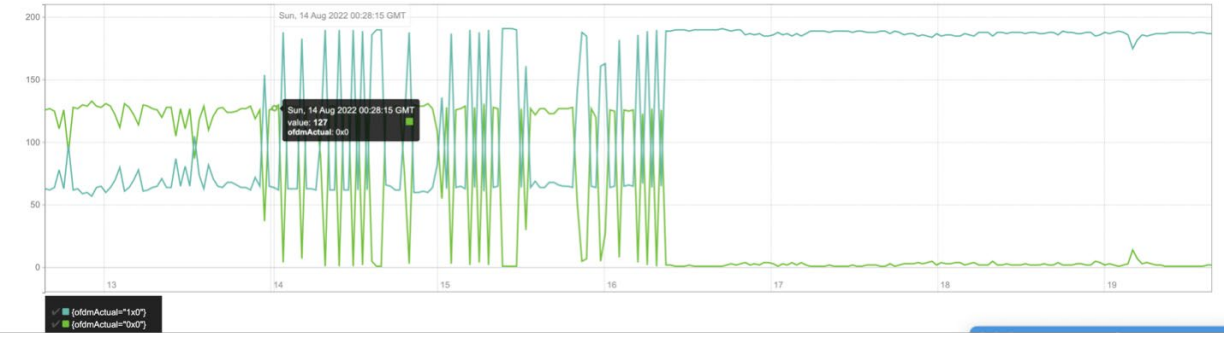


Figure 7 – Total Count of Modems in OFDM Partial vs Time for a Given RPD

We have been fortunate to have an incredible collaboration among both RPD vendors and modem chipset vendors with whom we partnered to identify and fix the root cause of these interoperability issues. Unfortunately, however, the release of new firmware to address issues discovered in the field tends to be a long process. The firmware release, verification, and deployment cycle can take weeks, if not several months.

This situation, where a root cause is known but the fix will take time to implement, is one instance where the end-to-end automation of a vCMTS has delivered enormous value. While a firmware issue may take months to resolve, in many cases, workarounds can be written in a matter of hours. As part of the automated remediations to alerts discussed in [3], the workarounds (often implemented as requests to the vCMTS API) can be triggered in response to an alert.

In iCMTS, a human must be paged to address an alert; in a vCMTS, machines page other machines (microservices) that mitigate the alert in seconds, significantly improving the customer experience by reducing downtime. While a PMA is a feature for both iCMTS and vCMTS, the ability to rapidly address issues encountered during profile management is a testament to the agility and resiliency of vCMTS.

2.4. Midsplit

RF plant capacity is an ever-pressing challenge for both iCMTS and vCMTS. The traditional hybrid fiber coax (HFC) plant has been in place for decades, utilizing a low-split (LS) with an upstream frequency range from 5 MHz to 42 MHz. More recently, mid-split (MS) HFC plants have been introduced with an upstream frequency range from 5 MHz to 85MHz. This additional spectral band can be configured for an OFDMA channel, increasing upstream capacity by more than two-fold.

The benefits and challenges of upgrading a LS plant to a MS one are well documented [8]. The in-Home Assessment Tool (iHAT) as proposed in Ref. [9] was introduced to address the challenges, including:

- With imperfect isolation, the MS-CPE's upstream OFDMA signal may leak into the LS-CPE's downstream.
- Active or passive components, such as amplifiers or splitters, can make MS-CPE inoperable on the OFDMA channel. Essentially, these components have a low-pass-filter, such as an LS diplexer. This issue was generalized in Ref. [8] as a 'drop-amp' issue.

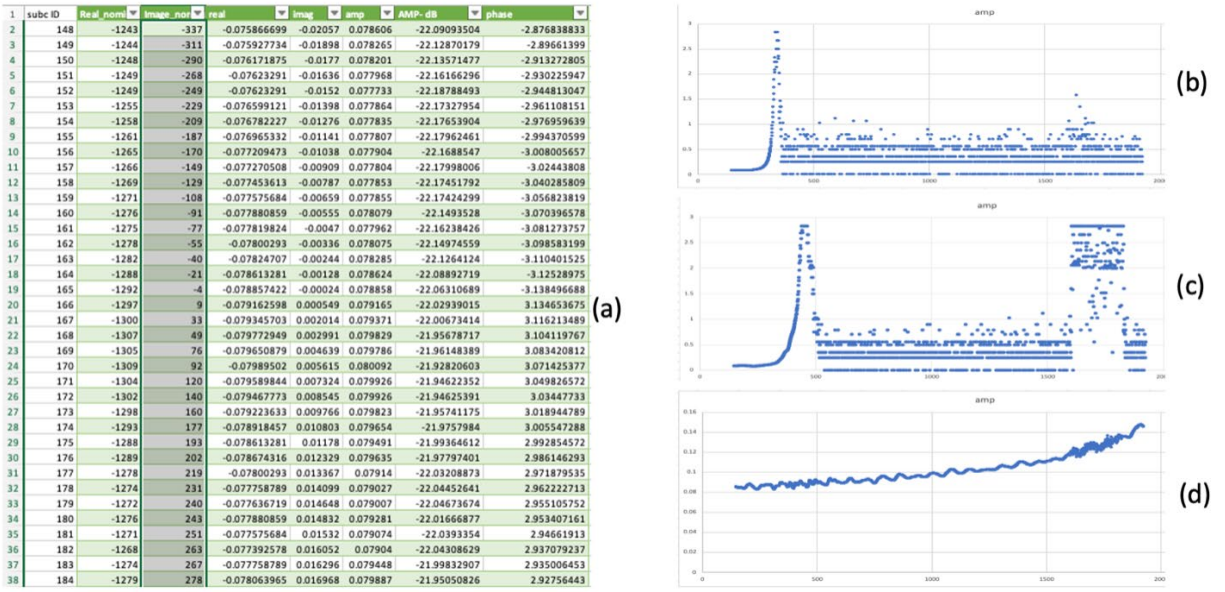
iHAT was based on the vCMTS application framework. Utilizing an API provided by the vCMTS, iHAT schedules OFDMA upstream data profile bursts to detect these issues and steer modems' configuration to mitigate their impacts.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

One challenge we have experienced with the transition to MS and iHAT was caused by a DOCSIS 3.1 interoperability issue between modems, RPDs, and the vCMTS. Central to this issue was the question: when a modem goes into OFDMA partial mode, how can one differentiate between a "drop-amp" and an actual interoperability issue?

The solution to this was to configure the OFDMA channel ranging zones below 42 MHz spectrum band, which allows the D3.1 modems to be detected. An unequalized probe is scheduled prior to the modem transitioning into station maintenance mode for the OFDMA channel ranging process. To achieve this, the vCMTS was modified to export the first sample of the unequalized probe data. Figure 8(a) shows a sample captured and converted real and imaginary samples into amplitude. Figure 8(b-c) show modems having a "drop-amp" issue. Figure 8(d) shows a cable modem's OFDMA channel response and successful operation on the OFDMA.

The ability to modify the vCMTS, deploy it via ISSU with no customer impact, and integrate it via API with iHAT allowed us to expose the unequalized probe data, which provided a mechanism to unambiguously diagnosis the drop-amp issue.



	subc ID	Real_nomi	Image_nor	real	imag	amp	AMP- dB	phase
2	148	-1243	-337	-0.075866699	-0.02057	0.078606	-22.09093504	-2.876838833
3	149	-1244	-311	-0.075927734	-0.01898	0.078265	-22.12870179	-2.89661399
4	150	-1248	-290	-0.076171875	-0.0177	0.078201	-22.13571477	-2.913272805
5	151	-1249	-268	-0.07623291	-0.01636	0.077968	-22.16166296	-2.930225947
6	152	-1249	-249	-0.07623291	-0.0152	0.077733	-22.18788493	-2.944813047
7	153	-1255	-229	-0.076599121	-0.01398	0.077864	-22.17327954	-2.961108151
8	154	-1258	-209	-0.076782227	-0.01276	0.077835	-22.17653904	-2.976959639
9	155	-1261	-187	-0.076965332	-0.01141	0.077807	-22.17962461	-2.994370599
10	156	-1265	-170	-0.077209473	-0.01038	0.077904	-22.1688547	-3.008005657
11	157	-1266	-149	-0.077270508	-0.00909	0.077804	-22.17998006	-3.02443808
12	158	-1269	-129	-0.077453613	-0.00787	0.077853	-22.17451792	-3.040285809
13	159	-1271	-108	-0.077575684	-0.00659	0.077855	-22.17424299	-3.056823819
14	160	-1276	-91	-0.077880859	-0.00555	0.078079	-22.1493528	-3.070396578
15	161	-1275	-77	-0.077819824	-0.0047	0.077962	-22.16238426	-3.081273757
16	162	-1278	-55	-0.07800293	-0.00336	0.078075	-22.14974559	-3.098583199
17	163	-1282	-40	-0.07824707	-0.00244	0.078285	-22.1264124	-3.110401525
18	164	-1288	-21	-0.078613281	-0.00128	0.078624	-22.08892719	-3.12528975
19	165	-1292	-4	-0.078857422	-0.00024	0.078858	-22.06310689	-3.138496688
20	166	-1297	9	-0.079162598	0.000549	0.079165	-22.02939015	3.134653675
21	167	-1300	33	-0.079345703	0.002014	0.079371	-22.00673414	3.116213489
22	168	-1307	49	-0.079772949	0.002991	0.079829	-21.95678717	3.104119767
23	169	-1305	76	-0.079650879	0.004639	0.079786	-21.96148389	3.083420812
24	170	-1309	92	-0.07989502	0.005615	0.080092	-21.92820603	3.071425377
25	171	-1304	120	-0.079589844	0.007324	0.079926	-21.94622352	3.049826572
26	172	-1302	140	-0.079467773	0.008545	0.079926	-21.94625391	3.03447733
27	173	-1298	160	-0.079223633	0.009766	0.079823	-21.95741175	3.018944789
28	174	-1293	177	-0.078918457	0.010803	0.079654	-21.9757984	3.005547288
29	175	-1288	193	-0.078613281	0.01178	0.079491	-21.99364612	2.992854572
30	176	-1289	202	-0.078674316	0.012329	0.079635	-21.97797401	2.986146293
31	177	-1278	219	-0.07800293	0.013367	0.07914	-22.03208873	2.971879535
32	178	-1274	231	-0.077758789	0.014099	0.079027	-22.04452641	2.962222713
33	179	-1272	240	-0.077636719	0.014648	0.079007	-22.04673674	2.955105752
34	180	-1276	243	-0.077880859	0.014832	0.079281	-22.01666877	2.953407161
35	181	-1271	251	-0.077575684	0.01532	0.079074	-22.0393354	2.94661913
36	182	-1268	263	-0.077392578	0.016052	0.07904	-22.04308629	2.937079237
37	183	-1274	267	-0.077758789	0.016296	0.079448	-21.99832907	2.935006453
38	184	-1279	278	-0.078063965	0.016968	0.079887	-21.95050826	2.92756443

(a)

Figure 8 – (a) Unequalized Probe Data; (b)-(d) Amplitude Response Samples from Three Different Modems

3. Digital Transformation of the Access Network Infrastructure

In the previous sections, we discussed how our end-to-end automation of a vCMTS has enabled Comcast to exponentially grow its vCMTS footprint, mitigate unexpected issues encountered during our path toward deploying more than 100,000 RPDs, and expand capacity over the existing HFC network. Automation, however, is a somewhat vague term since it could refer to anything from a shell script to simplify a repetitive task, to infrastructure as code. In this section, we would like to clarify that our usage of "automation" refers to a digital transformation of the access network infrastructure.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The term "digital transformation" seems to mean something different to every person. Indeed, research on organizations pursuing digital transformations has found at least "23 unique definitions" of the term [10], which was generalized into the following conceptual definition:

> *A digital transformation is "a process that aims to improve an entity by triggering significant changes to its properties through combinations of information, computing, communication, and connectivity technologies"* [10].

Given the ambiguity, it's helpful to contextualize digital transformations using frameworks such as Quali's "Infrastructure Automation Maturity Model" [11]. While we were unaware of this maturity model during our previous work, it nicely summarizes many of our desired end goals, for example:

- **Dynamic lifecycle**: from conception to destruction, infrastructure is created, modified, or retired according to demand. While maintaining a private, physical edge cloud significantly raises the cost of adding or removing resources relative to simply provisioning/destroying virtual resources in a public cloud, once the business decision to modify the physical resources is made, the process must be automated and seamless.
- **Single source of truth**: there should be no ambiguity surrounding which resources exist or their status, capacity, and configuration.
- **Intelligent predictive change**: while the edge cloud fleet is currently comprised of almost 1000 Kubernetes clusters distributed across the US, they are logically identical and manipulated at the fleet level. Changes are intelligently rolled out to the clusters in accordance with risk, characteristics of different sites (e.g., RPD vendor), utilization, and the results of automated feedback. Furthermore, workloads adaptively utilize available resources, e.g., latency-sensitive traffic is prioritized onto shorter optical transport legs.

We attribute the high maturity level of our CIN and PPOD automation to a few factors. One advantage is that the vCMTS was a greenfield project, with a very strong focus from day one on limiting complexity and unnecessary variation from site to site. Another factor that has contributed to our success was our investment in achieving a true digital transformation of the CIN and PPOD. As we mentioned in our previous paper:

> *After our early work defining our infrastructure in code, we naively thought we had solved the difficult part of the problem. In fact, someone in our organization famously predicted it should take "a few weeks" to whip up the automation to allow us to manage vCMTS at scale. Instead, what began with a few engineers has evolved over the past three years into multiple teams comprised of 15 people. Building capabilities such as self-recovery, complex decision-making logic, risk-based scheduling to name a few took months of hard work by some of our highest performing engineers* [3].

3.1. Concepts

3.1.1. Digital Twin

Crucial to the success of our PPOD and CIN automation has been the creation of a digital twin. A digital twin refers to the concept that alongside the physical infrastructure existing in the real world, a digital representation of the physical entity (its "twin") allows users to inspect and manipulate the physical object itself. While the term digital twin may be relatively new, the idea has been present for a long time. For example, (digital) CAD drawings of a desk can be consumed by a computer numerical control (CNC) machine to transform a (physical) piece of wood into a desk.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The concept of digital twins, however, significantly expands on the machine shop example; the digital entity's lifecycle is in-sync with the physical entity. While a desk may be created via a CAD drawing, deletion of the CAD drawing file has no impact on the physical desk. In contrast, for digital entities such as a CIN switch, manipulating the configuration of the switch's digital twin triggers automation that enacts the corresponding change on the physical switch. And just as the creation of a physical switch begins with the creation of its digital twin, the deletion of the digital twin triggers the deprovisioning of the switch itself.

3.1.2. Global Object Store for Digital Entities

Central to the idea of a digital twin is the concept that the digital world should match the physical world. Echoing domain-driven design [12], this implies that if there is one physical device, there should be only one digital device. Indeed, the automation system that manages the PPOD is the "owner," or single source of truth for the PPOD. In a sense, we benefit from the greenfield nature of the vCMTS, which has resulted in relatively few teams who need to know about the PPOD and CIN. When external systems need information about the PPOD or CIN, they simply integrate with our APIs following the usual microservices pattern.

Because the node is an aggregation point for many backend systems in Comcast, there are many more stakeholders in the RPD lifecycle. As we were investigating how to accomplish a digital transformation of the RPD, it quickly became clear that we needed to rethink how we facilitate easy integration with other teams.

Rather than integrating directly with other teams, we are leveraging the data mesh concept [13]. Information about the RPD will be published into a global object store, allowing any interested consumer to fetch the current (or historical) information about any given RPD. Additionally, consumers can subscribe to notifications, allowing them to receive events when, for example, an RPD's operational status changes.

3.1.3. Data Catalog

In addition to the global store of the digital objects, we briefly want to note that other data is required to maintain the access network. Raw telemetry, for example, is collected by Prometheus and stored for analysis. While the analyzed data may be stored directly on the digital entities, the raw data itself is not persisted in the global object store; instead, it can be linked to the digital entities via references to the external data catalog stores (e.g., data lakes). This ensures that only valuable and well-understood data is exposed to the larger organization.

3.2. RPD Digitization

3.2.1. Use Cases

As was mentioned in Section 3.1.2, there are many stakeholders that either supply the information necessary to provision an RPD or which consume information about the RPD for the purposes of customer support, incident management, etc. In Figure 9, we depict a highly simplified scenario where a new RPD is introduced into the field. A new (digital) RPD is created during the augment planning stage. This triggers procurement and construction to install the physical RPD at the requested location; in parallel, the DOCSIS and video configuration is created for the RPD.

When the device is connected, the provisioning process begins, during which it is authenticated, configured, and ultimately is ready to serve customers. Telemetry is emitted for the RPD and consumed

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

by many teams, including the capacity and impairment monitoring teams. In the event of an impairment, an incident is created which may result in the device being rebooted or reconfigured. Meanwhile, if the capacity team detects that the RPD is oversubscribed, a request can be made to the augment planning team to start the process of adding a new RPD.

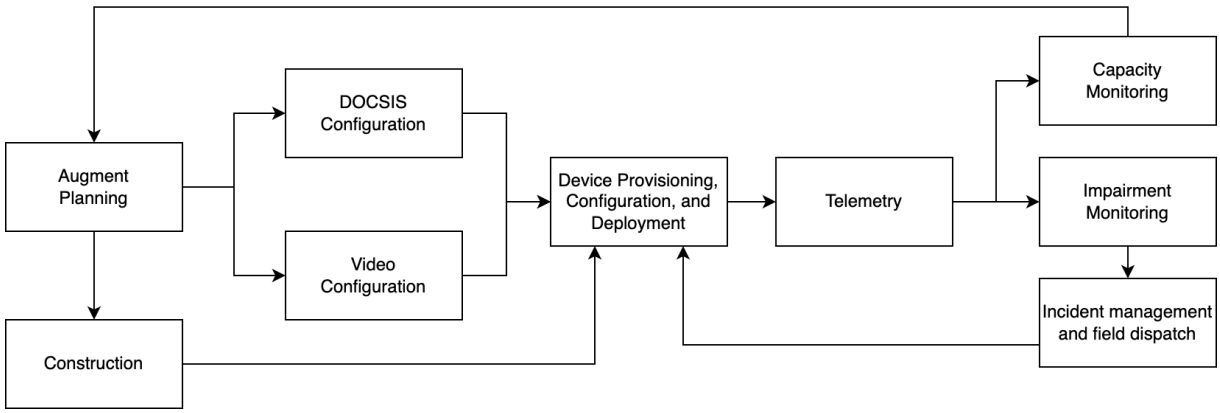


Figure 9 – Simplified Process for Provisioning a New RPD Based on Capacity Demands

Although simplified, this RPD lifecycle shows that many subject matter experts (SMEs) are required to manage an RPD. Teams are often organized around each of those domains. This leads to ambiguity around who is the "owner" of the RPD at any given phase of the lifecycle, since each team owns a different RPD subdomain. Since machine-to-machine communication is required in any highly scalable system, each team will typically have a microservice that exposes information about their subdomain.

However, each also depends on the other teams to build a complete picture of the RPD. The complexity of this model (where each subdomain needs to know about every subdomain to assemble the complete information about the RPD) grows as $O(N^2)$, where N is the number of subdomains or teams involved in the RPD lifecycle.

Phrased in the context of a digital transformation, while there is a single physical RPD, a process-driven framework naturally evolves into a multiverse of digital RPDs. Careful collaboration could in principal result in eventual consistency among these different teams, but achieving such consistency is a daunting challenge.

3.2.2. Model

In addition to employing microservices to separate domains, we employ a data mesh in the operational plane. This allows us to extend the microservices approach while addressing its shortcomings when it comes to scaling in the enterprise, where there can be thousands of microservices, and data consistency becomes a tremendous challenge. Here, the data mesh provides a federated data services framework; as a schema-driven service, the data mesh focuses application teams on defining their data as a product. That data can be linked to other entities via references in the global object store (Section 3.1.2).

For example, in Figure 10, a digital RPD is represented as a composition of device information (e.g., construction plan, its location, etc.), DOCSIS and video config, impairment status, and capacity information. While the provisioning process for an RPD is unchanged from Figure 9, the subdomain teams store their information in the data mesh. Because a complete RPD state can be obtained by traversing the connected objects, this eliminates the need for each team to maintain consistency with every other team. When the DOCSIS configuration changes, for example, the team that manages the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

DOCSIS config simply publishes the data to the data mesh for that RPD; teams that subscribe to the DOCSIS config events are notified of the change, and any subsequent query to the data mesh automatically reflects the updated configuration.

Due to its focus on lifecycle management, data consistency, and machine-to-machine communication, we anticipate that a full adoption of infrastructure digitization will significantly strengthen our ability to deliver the highest quality services to our subscribers.

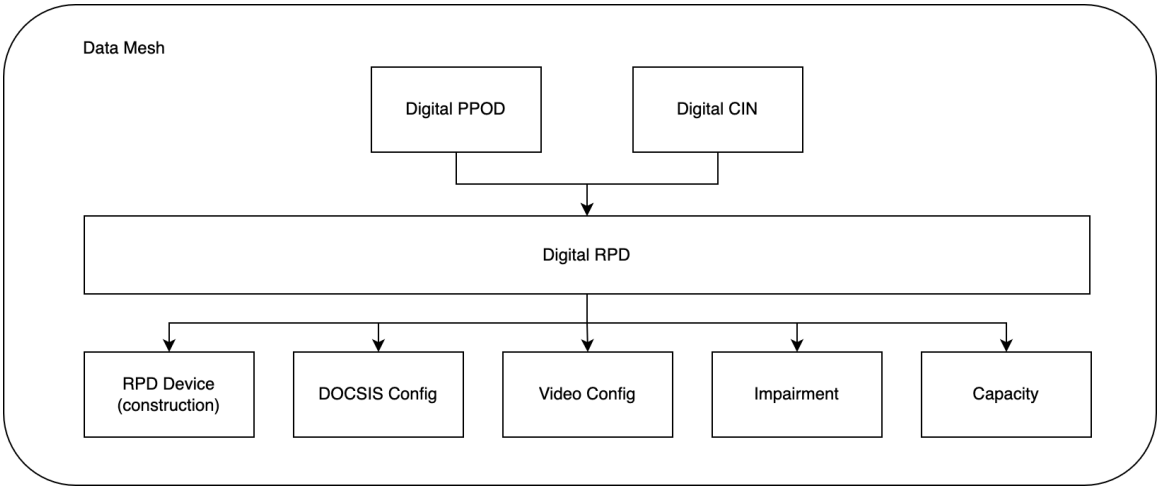


Figure 10 – Digitized RPD and its Subdomains in a Data Mesh

4. Conclusion

At Comcast, we are committed to providing the highest speeds on the most reliable network. We believe that virtualization of the CMTS is a necessary first step toward reducing cost and power consumption of internet delivery. However, the path toward providing symmetric, multi-gigabit services while maintaining an exceptional customer experience requires very tight coordination across the many different domains of the access network. A digital transformation of the access network infrastructure is critical to achieving the level of coordination and observability required to enable a smooth transition to DOCSIS 4.0 and FDX.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

ATDMA	Advanced time division multiple access
CIN	converged interconnect network
CLI	command line interface
CM	cable modem
CMTS	cable modem termination system
COTS	commercial off the shelf
CPE	customer premises equipment
DAA	distributed access architecture
DPD	downstream profile descriptor
DOCSIS	data over cable service interface specification
DS	downstream
EFK	Elasticsearch, Fluentbit, and Kibana
FDX	full-duplex DOCSIS
HFC	hybrid fiber-coax
iCMTS	integrated CMTS
iHAT	in-Home Assessment Tool
IP	internet protocol
L2TP	layer 2 tunneling protocol
LSB	least significant bit
MAC	medium access control layer
MAP	upstream bandwidth allocation map
MHAv2	modular headend architecture version 2
MMM	MAC management message
MS	mid-split
MSO	multiple system operator
MULPI	mac and upper layer protocols interface
NIC	network interface card
OFDM	downstream orthogonal frequency division multiplexing
OFDMA	upstream orthogonal frequency division multiple access
PHY	physical layer
PMA	profile management application
PPOD	primary pod (server rack containing vCMTS)
RF	radio frequency
RPD	remote PHY device
RPHY	remote PHY
RxMER	receive modulation error ratio
SC-QAM	single carrier quadrature amplitude modulation
SNMP	simple network management protocol
SG	service group
SGW	service gateway
LS	low-split
TSDB	time series database
UCD	upstream channel descriptor
US	upstream
vCMTS	virtualized CMTS

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

[1] Cable Labs, "Remote PHY Specification," in *CM-SP-R-PHY-I14-200323*, 2020.

[2] V. Mutalik, B. Gaydos, D. Rice and D. Combs, "Fifty Shades of Grey Optics: A Roadmap for Next Generation Access Networks," in *SCTE*, 2019.

[3] B. Krishnamurthy and G. Medders, "Humanoids Optional: Deploying vCMTS at Scale with Automation," in *SCTE*, 2021.

[4] G. Kim, J. Humble, P. Debois and J. Willis, The DevOps Handbook, Portland: IT Revolution Press, LLC, 2016.

[5] J. Leech and A. Martushev, "PMA Improvements - Strategies Employed for Faster Mitigation, Increased Capacity, and Cost Savings," in *SCTE*, 2022.

[6] M. Harb, J. Ferreira, D. Rice, B. Santangelo and R. Spanbauer, "A Machine Learning Pipeline for D3.1 Profile Management," in *SCTE*, 2019.

[7] K. Sundaresan, J. Zhu, M. Mishra and J. Lin, "Practical Lessons from D3.1 Deployments and a Profile Management Application (PMA)," in *SCTE*, 2019.

[8] L. Zhou, R. Thompson, R. Howald, J. Chrostowski and D. Rice, "A Proactive Network Management Scheme for Mid-split Deployment," in *SCTE*, 2020.

[9] R. Thompson, R. Howald, J. Chrostowski, D. Rice, A. Vieira, R. Vugumudi and L. Zhen, "Rapid and Automated Production Scale Activation of Expanded Upstream Bandwidth," in *SCTE*, 2021.

[10] G. Vial, "Understanding digital transformation: A review and a research agenda," *Journal of Strategic Information Systems,* vol. 28, pp. 118-144, 2019.

[11] Quali, "Infrastructure Automation Maturity Model," 2021.

[12] E. Evans, Domain-Driven Design: Tackling Complexity in the Heart of Software, Addison-Wesley Professional, 2003.

[13] Z. Dehghani, Data Mesh: Delivering Data-Driven Value at Scale, O'Reilly Media, 2022.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Towards Fully Automated HFC Spectrum Management

A Technical Paper prepared for SCTE by

Jay Zhu
Principal Software Engineer, Access Network Engineering
Comcast
4100 E Dry Creek Rd, Centennial, CO 80122
+1 (720) 955-6764
Jay_Zhu@comcast.com

Maher Harb
Distinguished Engineer, Machine Learning
Comcast
1800 Arch St, Philadelphia, PA 19103
+1 (215) 990-8376
Maher_Harb@comcast.com

Jeff Howe
Director, Access Network Engineering
Comcast
+1 (630) 988-9060
Jeff_Howe@comcast.com

Chad Humble
Senior Engineer, Access Network Engineering
Comcast
4100 E Dry Creek Rd, Centennial, CO 80122
+1 (720) 951-8092
Chad_Humble@comcast.com

Dan Rice
VP2, Next Generation Access Network
Comcast
4100 E Dry Creek Rd, Centennial, CO 80122
+1 (720) 512-3730
Daniel_Rice4@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Problem Formulation and Algorithm Design 4
   2.1. Real-World Constraints 5
   2.2. Algorithm 6
      2.2.1. Objective Function and Formulation of Linear Constraints 7
      2.2.2. Performance 9
      2.2.3. Testing 10
3. System Requirements and Architecture 10
   3.1. PoC Software 11
   3.2. Integration Context 13
4. Experiments and Results 13
   4.1. Using the Baseline Configuration 13
   4.2. Reducing the Minimum SC-QAM Allocation Requirement 15
   4.3. Increasing the SC-QAM Utilization Limit 16
5. Configuration Automation and Hit-less Changes 17
   5.1. Hit-less Spectrum Configuration Change Procedure 18
   5.2. Closed-loop Integration and Results 19
6. Conclusion 24

Abbreviations 25

Bibliography & References 25

List of Figures

Title **Page Number**

Figure 1 – The Objective Function 7
Figure 2 – Total Spectrum Segment Constraint 7
Figure 3 – SC-QAM Allocation Minimum Unit Constraint 7
Figure 4 – SC-QAM Allocation Frequency Limit Constraint 8
Figure 5 – OFDM Allocation Constraint 8
Figure 6 – SC-QAM Allocation Constraint 8
Figure 7 – Minimum Number of OFDM Channels Constraint 9
Figure 8 – Minimum Total SC-QAM Allocation Constraint 9
Figure 9 – SC-QAM and OFDM Maximum Utilization Constraint 9
Figure 10 – PoC Software Architecture 11
Figure 11 – PoC Software: Per Service Group Spectrum View 11
Figure 12 – PoC Software: Per Service Group Optimization Summary 12
Figure 13 – PoC Software: Automation and Integration Service 12
Figure 14 – Spectrum Manager: High-Level Deployment Integration 13
Figure 15 – Joint SC-QAM and OFDM Utilization Distributions (current) 14
Figure 16 – Joint SC-QAM and OFDM Utilization Distributions (optimized) 15
Figure 17 – Joint SC-QAM and OFDM Utilization Distributions (infeasible) 15
Figure 18 – Joint SC-QAM and OFDM Utilization Distributions (optimized with a minimum of 16 SC-QAM channels required) 16

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 19 – Joint SC-QAM and OFDM Utilization Distributions (optimized with a minimum of 16 SC-QAM channels required and 80% SC-QAM utilization limit) 17
Figure 20 – PoC Software: Automation and Integration Service 18
Figure 21 – Closed-loop Integration: Baseline Spectrum Configuration 19
Figure 22 – Closed-loop Integration: Baseline Spectrum Configuration (Spectrum Analyzer View) 20
Figure 23 – Closed-loop Integration: Optimized Spectrum Configuration 21
Figure 24 – Closed-loop Integration: Background Traffic 21
Figure 25 – Closed-loop Integration: Traffic Rates (After Moving to Intermediate Bonding Groups) 22
Figure 26 – Closed-loop Integration: Spectrum Analyzer View During the RF Change 22
Figure 27 – Closed-loop Integration: Spectrum Analyzer View After the RF Change 23
Figure 28 – Closed-loop Integration: Background Traffic Rates After the Configuration Change 23
Figure 29 – Closed-loop Integration: Downstream Speed Test After the Configuration Change 24

List of Tables

Title **Page Number**

Table 1 – Real-World Policy-Based Constraints 5
Table 2 – Fuzzing Conditions for the Algorithm 10
Table 3 – Spectrum Allocation Optimization: Baseline Parameters 14
Table 4 – Spectrum Allocation Optimization: Reduced Minimum SC-QAM Allocation 16
Table 5 – Spectrum Allocation Optimization: Increased SC-QAM Utilization Limit 17

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

In 2022, Comcast introduced a fresh perspective on optimizing the spectrum to achieve desired outcomes for product speeds, capacity, node segmentation, and cost-effectiveness by maximizing spectral efficiency via reallocation of data over cable service interface specification (DOCSIS®) 3.0 (D3.0) single carrier-quadrature amplitude modulation (SC-QAM) channels to DOCSIS 3.1 (D3.1) orthogonal frequency division multiplexing (OFDM) channels while accounting for the utilization distribution [1]. Comcast also presented a virtual network function (VNF) concept to manage the balance automatically and dynamically between the SC-QAM and OFDM spectrum.

As the penetration growth in D3.1 devices continues and the deployment of D4.0 devices comes to the horizon, such a VNF solution can ensure our network resources are utilized at their optimal efficiency and robustness during the transformation. At a high level, a requirement of this VNF solution is to consider each individual service group's characteristics and constraints at scale and produce tailor-made recommendations as inputs into the automation layer where configuration change procedures, error handling and failover logic are implemented. A unified source of truth for the current state of the spectrum configuration is also maintained to support the VNF operations in the larger picture of the architecture.

Based on these requirements, late last year, we continued to make progress on this initiative and designed and developed algorithms to effectively produce spectrum configuration recommendations based on the given real-world constraints and objectives. Using the analytical engine of the VNF service with baseline constraints, we estimated the average capacity gain that can be achieved across our distributed access architecture (DAA) footprint is 231 Mbps per service group at the time of measurement, accounting for strict constraints and without adding any new spectrum. This capacity increase is enabled by the ~44% capacity gain from our profile management application (PMA) VNF [2][3] while the SC-QAM spectrum is converted into OFDM spectrum when the constraints are satisfied.

To demonstrate and test our work, we built an automation layer with a configuration manager and a closed-loop system in our lab that supports dynamic radio frequency (RF) spectrum changes using our virtual cable modem termination system (vCMTS) – remote physical layer (PHY) device (RPD) configuration application programming interfaces (APIs) without disconnecting the cable modems (CMs) on the service group. This demonstrates the ability of the system to make hit-less RF spectrum changes in an integration context. We also measured the peak throughput changes of the DOCSIS 3.1 CMs after the hit-less spectrum changes.

2. Problem Formulation and Algorithm Design

In the last year's paper, converting a static number of SC-QAM channels into OFDM spectrum across the entire footprint was discussed and analyzed with 4 scenarios, where converting 4 SC-QAM channels could add ~66 Mbps of capacity to each service group in the network, and converting 12 SC-QAM channels could add ~200 Mbps capacity to each service group, with potential increases in SC-QAM utilization on a subset of service groups.

This one-size-fits-all experiment is an effective starting point and could achieve significant gains with system simplicity. As we progress with the VNF implementation, the solution can now be further improved and scaled to fit each individual service group's characteristics such as their spectrum constraints and the most recent 98th percentile utilization values on SC-QAM and OFDM spectrum, without having to search for an optimal configuration that could fit the entire footprint.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.1. Real-World Constraints

When designing the VNF, there are a few general real-world policy-imposed constraints the VNF's algorithm needs to account for as part of its design. For example,

Table 1 – Real-World Policy-Based Constraints

Attribute	Common Value	Description
SC-QAM/OFDM utilization upper limit	60%	The 98th percentile utilization on both SC-QAM and OFDM have their upper limits, such as 60% to ensure high service quality. This means that the algorithm must ensure the estimated SC-QAM utilization and OFDM utilization will not exceed the given limit after the optimization/rebalance. This threshold can account for some years of compound annual growth rate (CAGR) which could influence this threshold to avoid increases in consumption growth that would require adding back SCQAM spectrum. While these removals and re-additions have been proven out, more conservative thresholds could be chosen until a fully automated system exists.
RF tones		Existing tones such as automatic gain control (AGC) tones, local RF inserts, and video on demand (VoD) tones must be correctly handled for each service group. For example, analog AGC tones should not overlap with DOCSIS spectrum but may be "worked around" by adding exclusion zones in OFDM channels; digital AGC tones may require the presence of neighboring power within a certain frequency range to the tone for the best performance.
Exclusion zones		Known exclusion zones must be considered in the optimization. For example, if a frequency range is known to suffer from ingress impairments, the algorithm should ensure that the SC-QAM channels do not overlap with the ingress.
Minimum number of SC-QAM channels	24	A minimum number of SC-QAM channels, for example, 24 or 32 channels can be required to achieve maximum product offers on D3.0 modems even if the overall capacity would accommodate a channel reduction. This constraint may also vary on a per-service group basis.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Attribute	Common Value	Description
SC-QAM channel size	6 MHz	The allocation of SC-QAM channels must use a minimum unit size.
OFDM channel size limits	24 – 192 MHz	The OFDM channels have size limits. The DOCSIS 3.1 CM physical layer specification [4] allows an OFDM channel to be 24 megahertz (MHz) to 192 MHz wide, but this range could vary based on deployment requirements.
OFDM channel guard band size	1 MHz	The OFDM channels' guard band overhead should be accounted for by the algorithm. Such a requirement allows for optimization accuracy and more meaningful reasoning for the recommendations when necessary.
SC-QAM applicable frequency ranges		The SC-QAM channels can have preferred frequency ranges in deployment.
Contiguous spectrum allocation requirements		The SC-QAM channels can require minimum allocation sizes, for example, 4 contiguous channels per allocation block.
Spectrum ordering		The SC-QAM channels may be preferred to be allocated to the lower frequency end, relative to the OFDM spectrum.
Hardware limitations		Hardware limitations such as the RPD chipset's capabilities can affect the maximum supported number of certain types of channels.

The listed policy-based constraints support the definition of the problem space for the design of the algorithm.

2.2. Algorithm

To build an effective VNF for this optimization task, it is desired to design an algorithm that:

- Adjusts its operation based on simple policy attributes that allow for A/B testing to optimize without writing new code.
- Produces recommendations that are guaranteed to satisfy all given policy-based constraints if feasible solutions exist.
- Is simple to maintain and add/remove policies for different use case requirements.
- Produces optimal spectrum configurations based on the given constraints if feasible solutions exist.
- Is performant, can recompute optimal spectrum allocation plans for our entire footprint efficiently for performance and elastic cloud costs.

With the given complexity of the listed constraints, there are multiple options on how the algorithm can be implemented. In our design, this optimization problem is formulated as a mixed-integer linear programming (MILP) problem, and the constraints are modeled as linear inequality constraints. Using MILP results in an explainable model where the descriptive objective function and constraints can be used as the immediate representation of the algorithm, reducing the gap between a discrete, programmatic

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

implementation and a theoretical design. In addition, such a model can have maintainability benefits when the policy, constraints and rules are modified, added, or removed.

2.2.1. Objective Function and Formulation of Linear Constraints

First, a definition of contiguous spectrum segment is created to represent a frequency range that can be freely used by DOCSIS downstream channels. With this definition, constraints can be contained within each contiguous spectrum segment and be simplified. Constraints 2 and 3 in Table 1 become preprocessing rules for identifying contiguous spectrum segments.

Because the objective of this optimization can be defined as maximizing the total spectral efficiency (linearly related to the total capacity). The objective function is defined as:

$$\max \sum_{i=1}^{N} x_i + \alpha \times y_i'$$

Figure 1 – The Objective Function

where N is the number of contiguous spectrum segments that can be reallocated and each of the segments can vary in their frequency ranges, x_i is the target allocation of SC-QAM spectrum, y_i is the target allocation of OFDM spectrum (with guard bands subtracted), α is the spectral efficiency ratio of OFDM / SC-QAM which suggested by the overall downstream capacity gain (~44%) observed on OFDM channels.

For each individual contiguous spectrum segment, the total allocated spectrum must be smaller than or equal to the size (W) of the segment:

$$x_i + y_i + K \times n_i \leq W_i$$

Figure 2 – Total Spectrum Segment Constraint

where x_i is the target allocation of SC-QAM, y_i is the target allocation of OFDM (with guard bands subtracted), K is the total OFDM guard band size per OFDM channel, n_i is the number of OFDM channels (to take OFDM's guard band overhead into consideration), and W_i is the size of the i_{th} contiguous spectrum segment.

The allocated SC-QAM within each continuous spectrum segment should be divisible by the size of an SC-QAM channel, hence:

$$x_i = p_i \times R$$

Figure 3 – SC-QAM Allocation Minimum Unit Constraint

where p_i is an integer variable that represents the number of SC-QAM channels, and R is the size of an SC-QAM channel.

Considering constraint 8 in Table 1, the allocated SC-QAM within a continuous spectrum segment may also subject to an upper limit due to the maximum frequency an SC-QAM channel can be allocated:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

$$x_i \leq \eta_i$$

Figure 4 – SC-QAM Allocation Frequency Limit Constraint

where η_i is calculated by subtracting the continuous spectrum segment's starting frequency from the maximum frequency that an SC-QAM channel can be allocated.

We use a binary constraint to represent the presence of OFDM and take OFDM's guard band overheads into account:

$$c_i \times M \geq y_i$$

$$y_i + (1 - c_i) \times M \geq P - K$$

Figure 5 – OFDM Allocation Constraint

where c_i is a binary variable that indicates the OFDM presence in a contiguous spectrum segment, y_i is the target OFDM allocation (with guard bands subtracted), P is the minimum size of an OFDM channel (with guard bands included, e.g.: 24 MHz), K is the OFDM guard band overhead (both guard bands), M is a large enough number to ensure:

- When OFDM is present, y_i is greater than 0, c_i must be 1 to satisfy the first inequality, and in the second constraint the OFDM size must meet its minimum size requirement.
- When OFDM is not present, y_i equals to 0, the first constraint is always satisfied, c_i becomes 0 because of minimization of overheads, and because M is large, the second constraint is always satisfied and the OFDM size variable of the contiguous spectrum segment does not need to meet the minimum size requirement.

Similarly, we use a binary constraint to represent the presence of SC-QAM and ensure that enough number of SC-QAM channels is allocated sequentially within a contiguous spectrum segment:

$$b_i \times M \geq x_i$$

$$x_i + (1 - b_i) \times M \geq B$$

Figure 6 – SC-QAM Allocation Constraint

where b_i is a binary variable that indicates the SC-QAM presence in the contiguous spectrum segment, x_i is the target SC-QAM allocation, B is the minimum required allocation for SC-QAM channels within a contiguous segment, M is a large enough number to ensure:

- When SC-QAM is present, x_i is greater than 0, b_i must be 1 to satisfy the first inequality, and in the second constraint the SC-QAM allocation must satisfy the minimum SC-QAM allocation constraint within a contiguous spectrum segment.
- When SC-QAM is not present, x_i equals to 0, the first constraint is always satisfied, and because M is large, the second constraint is always satisfied.

In addition to the previous constraints, the number (minimum) of OFDM channels must be associated with the target OFDM allocation:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

$$n_i \times (Q - K) \geq y_i$$

Figure 7 – Minimum Number of OFDM Channels Constraint

where y_i is the target allocation of OFDM (with guard-bands subtracted), K is the total OFDM guard band size used by each OFDM channel, n_i is the number of OFDM channels, and Q is the maximum size of an OFDM channel (with guard bands included, e.g.: 192 MHz). This constraint allows the minimum number of OFDM channels to be known as n_i and used in the constraint described by Figure 2 – Total Spectrum Segment Constraint to account for OFDM guard band overhead in the optimization.

For the minimum total SC-QAM allocation requirement, such as 24 or 32 SC-QAM channels, the following constraint is defined:

$$\sum_{i=1}^{N} x_i \geq \Gamma$$

Figure 8 – Minimum Total SC-QAM Allocation Constraint

where **Γ** is the minimum required SC-QAM allocation.

Finally, utilization constraints are specified to provide the upper limits of estimated utilization values after optimization. Respectively, there are SC-QAM utilization constraint and OFDM utilization constraint (i.e.: the utilization of SC-QAM/OFDM after optimization must not exceed 60% allowing 2 years of growth at a conservative CAGR before service group augmentation. This policy can be made more aggressive as full automation is completed):

$$\sum_{i=1}^{N} x_i * \Theta \geq L * \beta$$

$$\sum_{i=1}^{N} y_i * \Lambda \geq H * \gamma$$

Figure 9 – SC-QAM and OFDM Maximum Utilization Constraint

where L is the original SC-QAM allocation, β is the original SC-QAM utilization, Θ is the upper limit of SC-QAM utilization after optimization, H is the original OFDM allocation, γ is the original OFDM utilization, and λ is the upper limit of the OFDM utilization after optimization.

2.2.2. Performance

It is known that complex MILP problems can consume a significant amount of time to solve. A common implementation of solving a MILP problem is to programmatically generate the objective function and constraint representations of the problem and pass them to a dedicated MILP solver for high performance computing. Many of the solvers implement various heuristic algorithms to reduce the problem-solving time. The performance of a solver depends on several factors, including but not limited to:

- The number of variables in the constraints. Usually, the solving time increases super linearly as the number of variables grows.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- The objective function and constraints. Different solver implementations can perform differently on the same problem. Often, a solver is not consistently slow or fast on all problems.
- The implementation of the solver. There are open-source solvers and commercial solvers, either could perform well depending on the optimization problem and the heuristics and tuning in place. Commercial solvers often have advantages in scalability and solving time, but it is not a guarantee.

Because one of the requirements of the spectrum management VNF is to process the service group data and produce spectrum allocation recommendations for the entire footprint efficiently, there is high expectation on its algorithm's performance. As a result, given the current design of the constraints, the algorithm spends ~5ms on average to compute an optimal DOCSIS downstream spectrum plan for a service group on a single CPU core, which infers that it is possible to compute spectrum plans for the entire footprint within minutes using limited resource. Note that this performance result was evaluated using an open-source MILP solver library. Given this performance result, we have not investigated the performance of commercial solvers.

2.2.3. Testing

In addition to unit tests, manual validations, and lab tests, we developed fuzzing tests to reinforce the robustness of the algorithm and ensure that expected outputs are produced with a wide range of random given inputs and constraints.

The fuzzing tests cover a 1 GHz spectrum reserved for random initial SC-QAM and OFDM allocations and utilizations. And the tests check the optimization output under the following combinations of conditions:

Table 2 – Fuzzing Conditions for the Algorithm

Condition	Value
Randomized minimum SC-QAM allocation requirement	true + false
With random insertion tones	true + false
Allow OFDM AGC overlapping	true + false
Randomized initial OFDM size	24 – 192 MHz (0.5 MHz steps)

For each fuzzing condition, 2000 random test cases are generated. This helps automatically cover potential edge cases of the constraints as well as the validity of the optimization output under all conditions.

3. System Requirements and Architecture

The spectrum management VNF requires interfacing capabilities with other systems and services to retrieve the existing spectrum configurations and information of constraints. The VNF should also generate recommendation tasks for the automation layer to perform configuration management.

The VNF's core algorithm is developed as a production-ready microservice that has optimized performance and ideal CPU and memory efficiency. It is not directly interacted with by the users/engineers. Instead, we built a proof-of-concept (PoC) application with a graphical user interface (GUI) as a playground for convenient experiments, testing and demonstrations during development. This PoC application implements all required APIs to communicate with the VNF and data APIs for retrieving

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the service group data and applying optimized spectrum configurations. It also provides a summary dashboard and data tables and visualizations for quick optimization result analysis.

3.1. PoC Software

The PoC software has a simple architecture to allow the users to interact with its GUI, parallelize optimization requests to the spectrum management VNF, and store constraint settings and spectrum allocation data and statistics for before and after optimization.

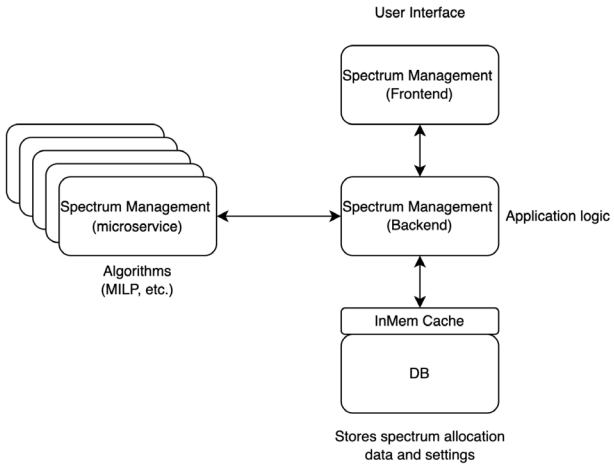


Figure 10 – PoC Software Architecture

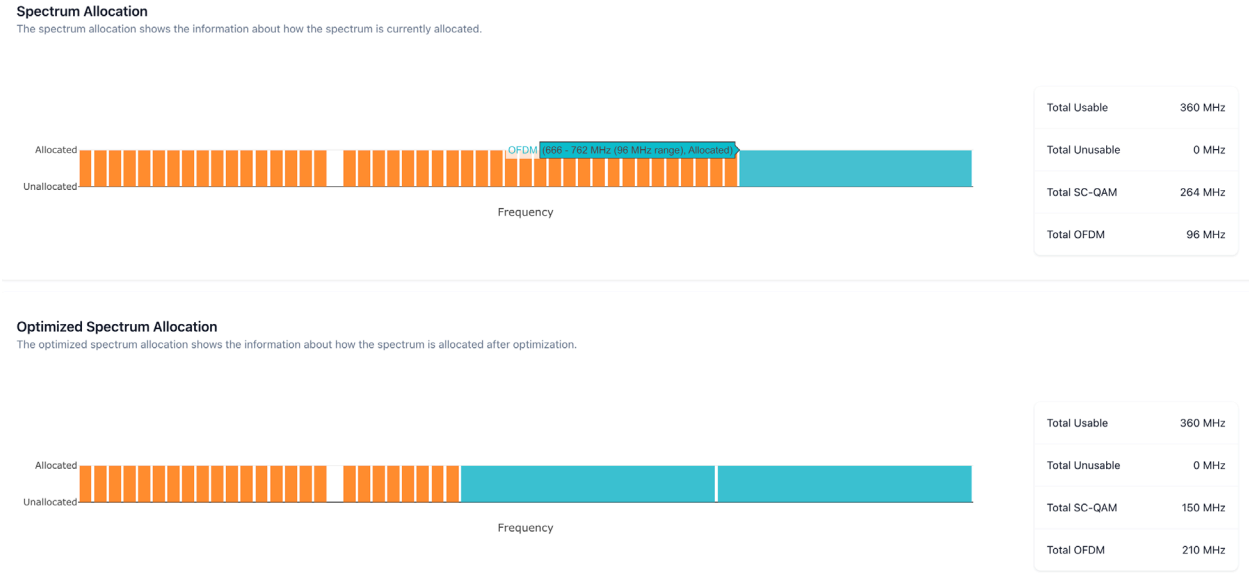


Figure 11 – PoC Software: Per Service Group Spectrum View

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

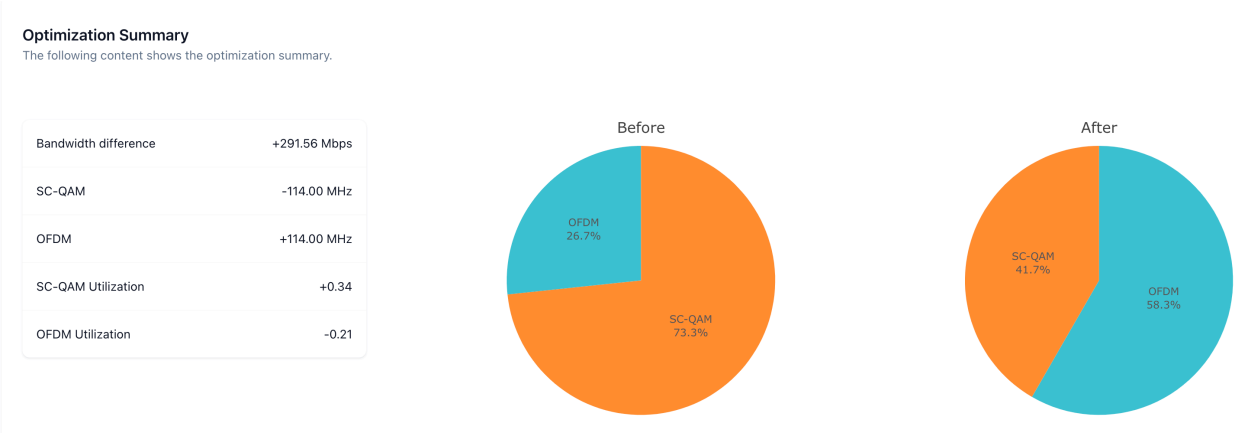


Figure 12 – PoC Software: Per Service Group Optimization Summary

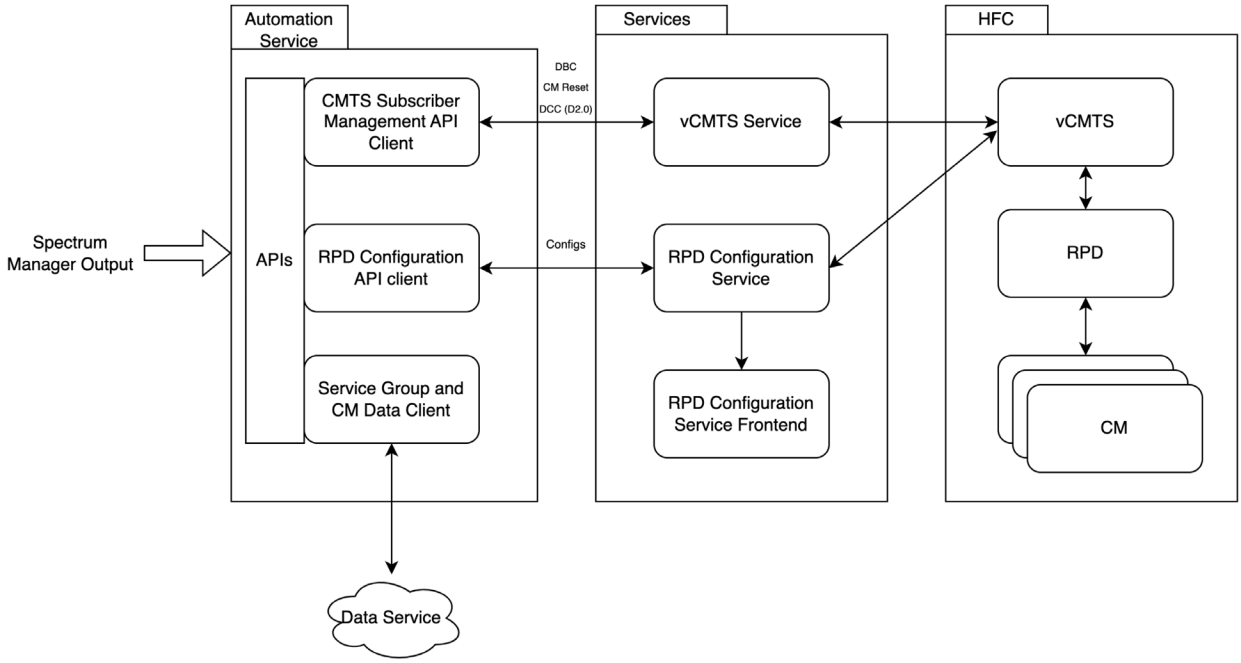


Figure 13 – PoC Software: Automation and Integration Service

A separate service illustrated in Figure 13 is developed as the automation layer for handling hit-less spectrum configuration changes, simple load balancing, and service group data retrieval for the PoC application. The PoC application's backend communicates with such service to extend its functionalities. More details of the automation service are discussed in section 5.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.2. Integration Context

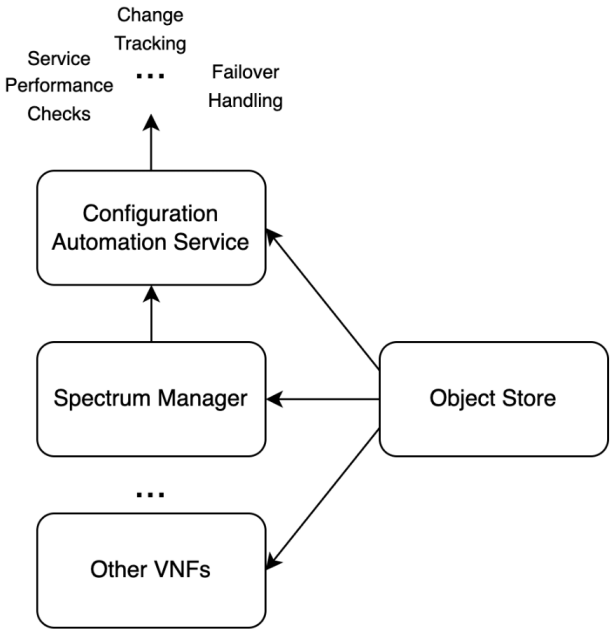


Figure 14 – Spectrum Manager: High-Level Deployment Integration

In the deployment integration context, the spectrum management VNF, along with other VNFs, are part of our event driven architecture centered around a unified object data store. The concept of this architecture is illustrated at a high-level in Figure 14. This architecture provides all required data for the computation performed by the spectrum management VNF and automation pipelines for scheduling and applying the configuration changes. The automation service illustrated in Figure 13 can be used as the initial implementation of part of the automation pipeline for spectrum management in this integration context.

4. Experiments and Results

To evaluate the performance and effectiveness of the VNF, we took a snapshot of the service group data of our entire DAA footprint and experimented with different constraint parameters. In all our experiments, the OFDM's spectral efficiency factor was configured to 1.44 which means 44% capacity gain based on RF compared to SC-QAM and OFDM channels that use 256 quadrature amplitude modulation (QAM) order. In addition, the experiments focused on rebalancing the existing configured SC-QAM and OFDM spectrum, which means no additional DOCSIS spectrum was added to the configurations after the optimization.

4.1. Using the Baseline Configuration

The baseline configuration is a set of conservative parameters. The upper limit of the SC-QAM and OFDM spectrum are set to 60% respectively. The minimum required SC-QAM channels is 24, which allows for up to 20 SC-QAM channels worth of spectrum to be converted into OFDM spectrum if the starting point is 44 SC-QAM channels. The minimum allocation block size of SC-QAM channels is 24 MHz (4 x 6 MHz channels). And finally, the minimum required OFDM channel size is 48 MHz.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 3 – Spectrum Allocation Optimization: Baseline Parameters

Parameter Name	Value
Maximum allowed SC-QAM utilization	60%
Maximum allowed OFDM utilization	60%
SC-QAM channel size	6 MHz
Base bandwidth per 6 MHz	37 Mbps
Minimum number of SC-QAM channels	24
Minimum SC-QAM allocation block	4 x 6 MHz = 24 MHz
OFDM guard band size (1 guard band)	1 MHz
OFDM size limit	48 – 192 MHz

Using the baseline parameters as constraints in the computation, it was estimated to improve each service group's capacity by 231 Mbps on average. Among all the service groups that were computed by the optimization algorithm, 94% of which received feasible solutions with improvements. The remaining 6% of the service groups were labeled "infeasible" because the solver could not find a solution that satisfies all constraints (such as the utilization limit constraints). The majority of the infeasible cases resulted from the high utilization on both SC-QAM and OFDM and the minimum SC-QAM allocation requirement (24 channels).

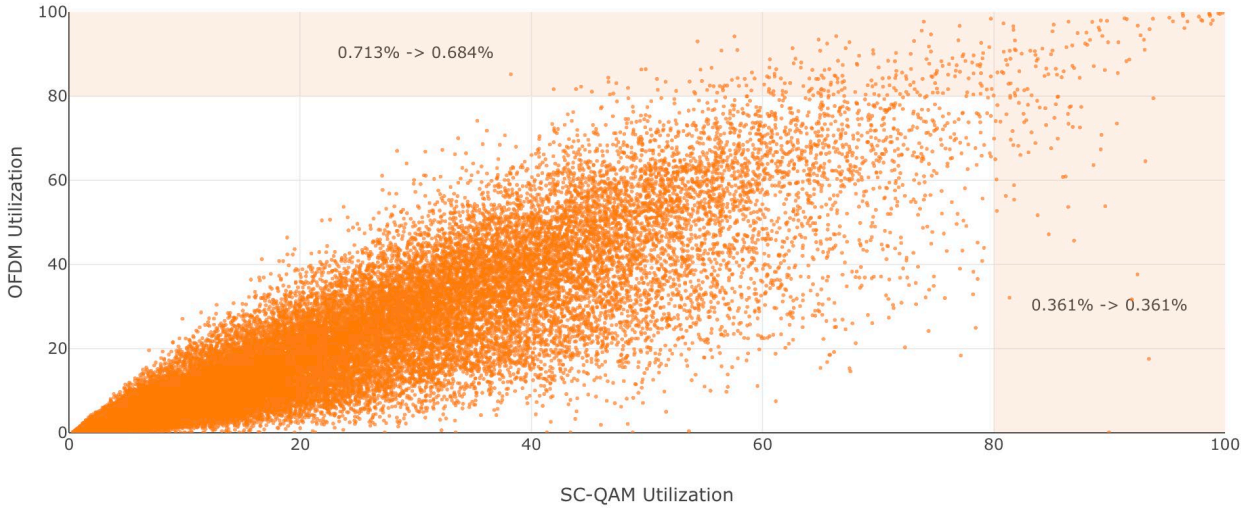


Figure 15 – Joint SC-QAM and OFDM Utilization Distributions (current)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

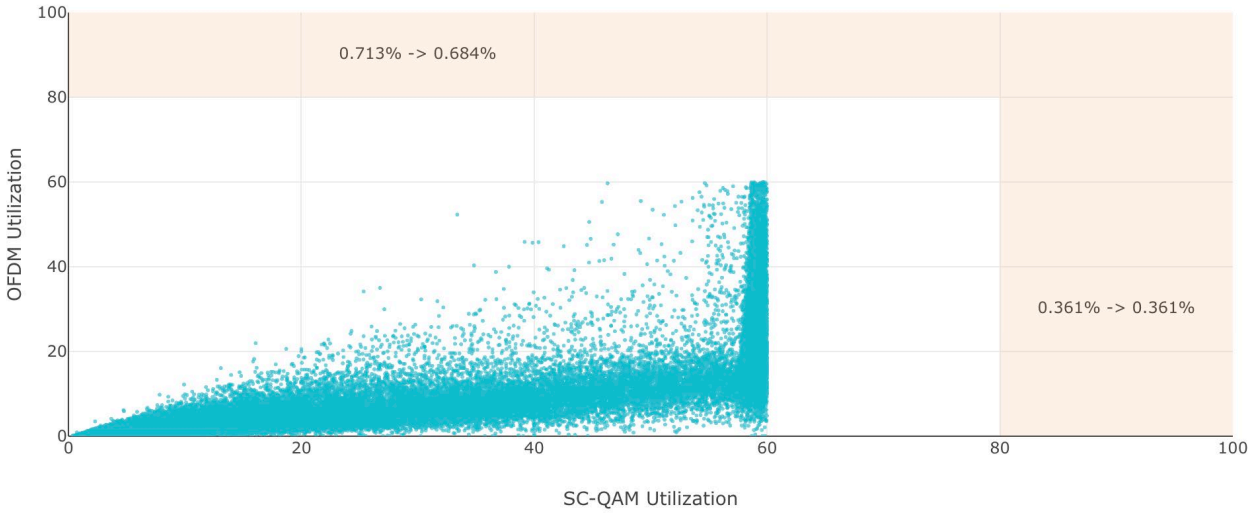


Figure 16 – Joint SC-QAM and OFDM Utilization Distributions (optimized)

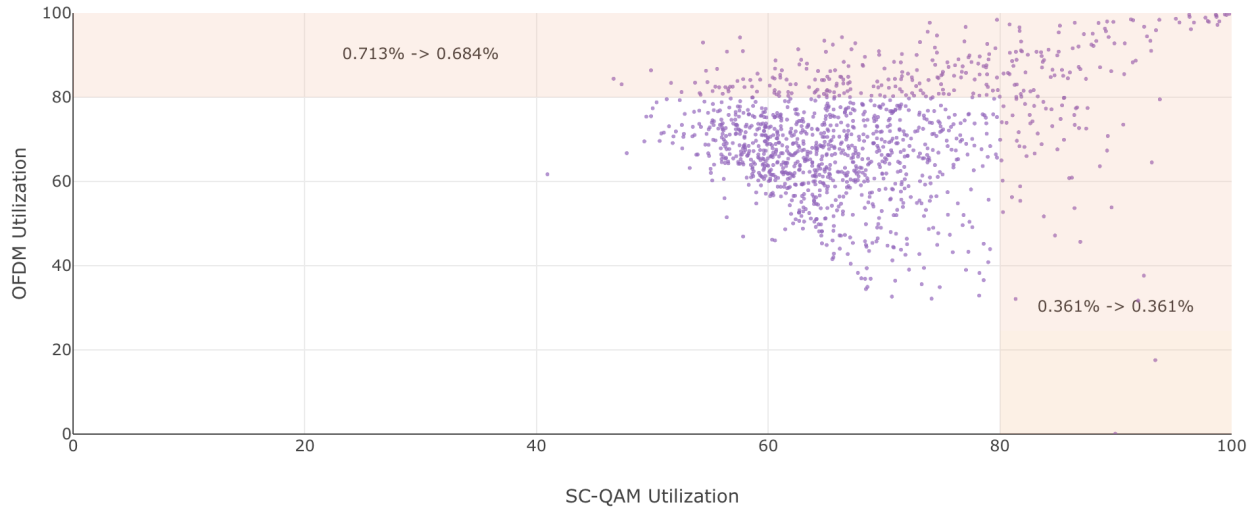


Figure 17 – Joint SC-QAM and OFDM Utilization Distributions (infeasible)

After the optimization, 0.029% of the service groups were relieved from high OFDM utilization (>= 80%) and had both SC-QAM and OFDM utilization contained within 60% by estimation. In Figure 16, the distribution shifted clockwise due to the lower OFDM utilization values, and the estimated utilization values on both SC-QAM and OFDM spectrum were capped at 60%. For most of the "infeasible" service groups, Figure 17 indicates that the existing high utilization is the cause.

4.2. Reducing the Minimum SC-QAM Allocation Requirement

The optimization using baseline constraint parameters produced promising results. However, the minimum number of SC-QAM channels (24) could be a limiting factor for the overall capacity gain. As the network transformation continues and newer technologies grow in adoption, we could expect this requirement to reduce over time.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In this experiment, all parameters were the same as the previous experiment, except the minimum number of SC-QAM channels was reduced from 24 to 16. This policy could be reduced as a result of a reduction in the maximum product speed offered to a D3.0 modem.

Table 4 – Spectrum Allocation Optimization: Reduced Minimum SC-QAM Allocation

Parameter Name	Value
Maximum allowed SC-QAM utilization	60%
Maximum allowed OFDM utilization	60%
SC-QAM channel size	6 MHz
Base bandwidth per 6 MHz	37 Mbps
Minimum number of SC-QAM channels	**16**
Minimum SC-QAM allocation block	4 x 6 MHz = 24 MHz
OFDM guard band size (1 guard band)	1 MHz
OFDM size limit	48 – 192 MHz

This constraint change increased the estimated average capacity gain from 231 Mbps to 321 Mbps.

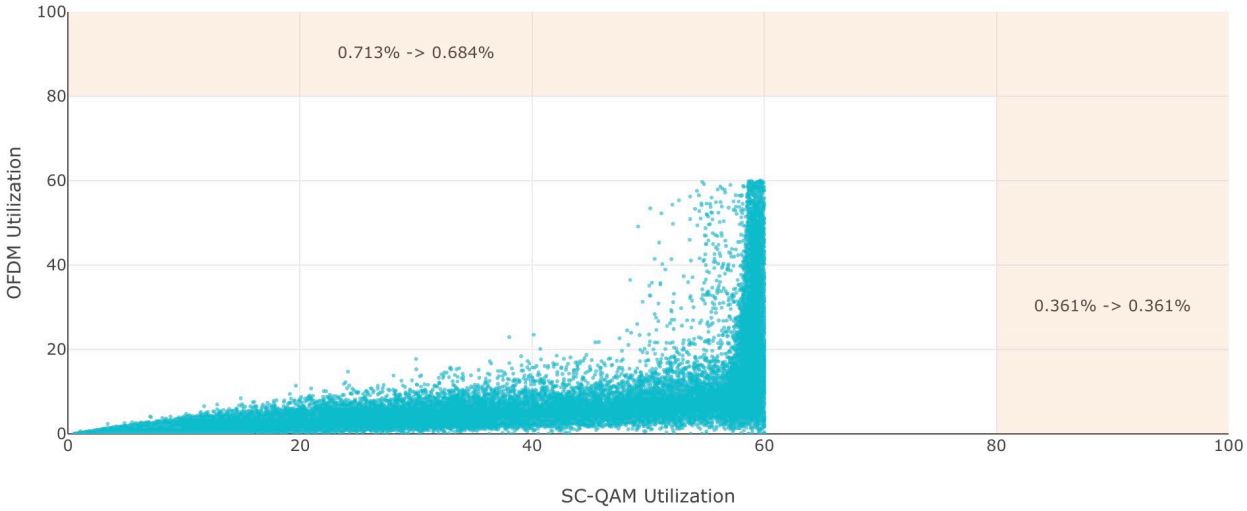


Figure 18 – Joint SC-QAM and OFDM Utilization Distributions (optimized with a minimum of 16 SC-QAM channels required)

From the optimization summary, there are still 6% of service groups that did not have feasible optimized spectrum configurations, but a significant amount of service groups could further reduce the SC-QAM spectrum within the allowed utilization limit to bring higher spectral efficiency and capacity gain.

4.3. Increasing the SC-QAM Utilization Limit

From a capacity management perspective, the full-scale deployment of the spectrum management VNF can potentially combine the SC-QAM utilization metric with the OFDM utilization metric and allow us to consider the DOCSIS spectrum as an integral spectrum resource. With this assumption, it is possible to reconsider the utilization constraints. The 60% SC-QAM utilization limit in the baseline configuration can be increased without risking overly high utilization on the SC-QAM spectrum. Meanwhile, this provides more room for SC-QAM to OFDM conversion for improved capacity and robustness.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In the next experiment, the SC-QAM utilization limit (after optimization) was increased from 60% to 80%.

Table 5 – Spectrum Allocation Optimization: Increased SC-QAM Utilization Limit

Parameter Name	Value
Maximum allowed SC-QAM utilization	**80%**
Maximum allowed OFDM utilization	60%
SC-QAM channel size	6 MHz
Base bandwidth per 6 MHz	37 Mbps
Minimum number of SC-QAM channels	**16**
Minimum SC-QAM allocation block	4 x 6 MHz = 24 MHz
OFDM guard band size (1 guard band)	1 MHz
OFDM size limit	48 – 192 MHz

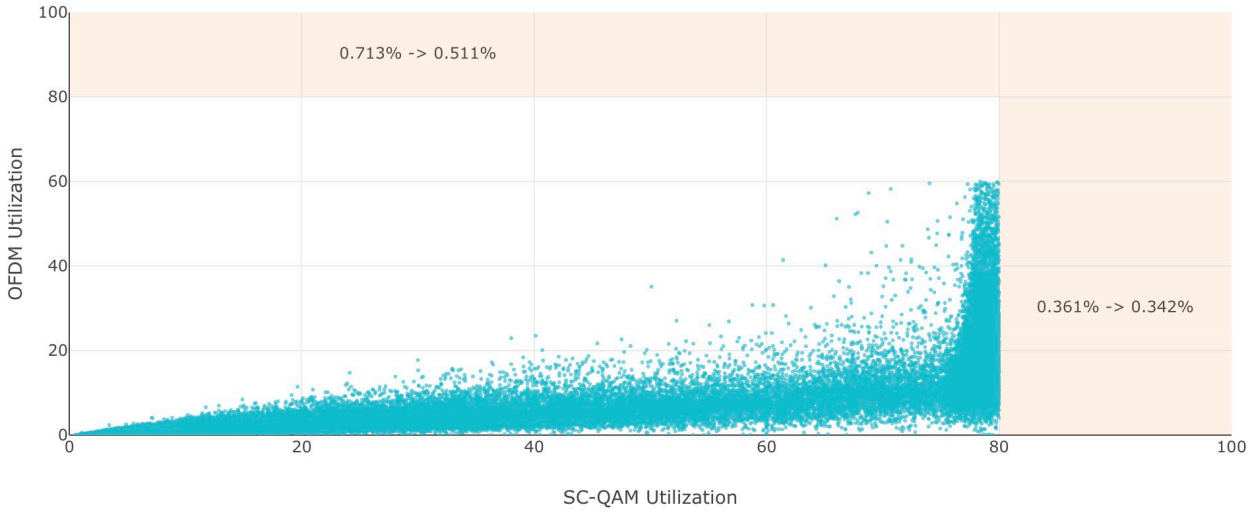


Figure 19 – Joint SC-QAM and OFDM Utilization Distributions (optimized with a minimum of 16 SC-QAM channels required and 80% SC-QAM utilization limit)

This constraint change increased the estimated average capacity gain from 321 Mbps to 347 Mbps. Meanwhile, the percentage of service groups that can be optimized increased from 94% to 98.3%, 0.102% of service groups were relieved from high OFDM utilization (>= 80%), and a small number of service groups (0.019%) were relieved from high SC-QAM utilization as a side effect from converting SC-QAM spectrum into OFDM spectrum.

5. Configuration Automation and Hit-less Changes

To apply the optimized spectrum configurations, we developed an automation service that is responsible for the translation and application of the updated spectrum configurations. This automation service coordinates with our other internal services to support seamless RF channel configuration changes without service disruption.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

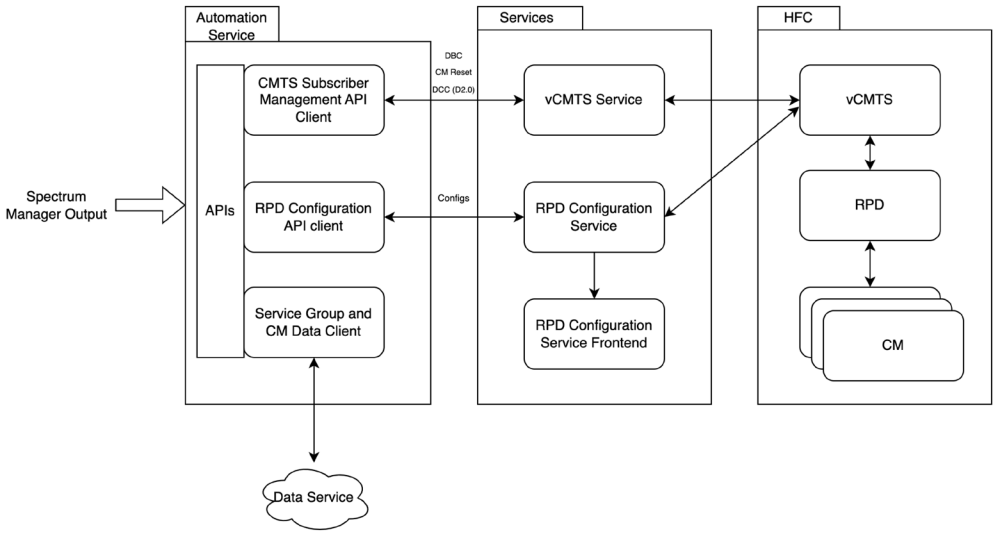


Figure 20 – PoC Software: Automation and Integration Service

5.1. Hit-less Spectrum Configuration Change Procedure

DOCSIS media access control (MAC) and upper layer interface specification [5] requires the CMs to support dynamic bonding change (DBC) and dynamic channel change (DCC) using DOCSIS MAC management messages. The DBC messages are supported by DOCSIS 3.0 and newer versions, and the DCC messages can be used for older versions of DOCSIS devices that exist in the network. In the downstream, DBC and DCC allow the CMs to dynamically change their receive channel set (RCS) without introducing significant service disruption (depending on the initialization technique used). This means that the CMs can dynamically change the frequencies they receive and can change their primary channel's frequency. For DOCSIS 3.0 and newer CMs, channel bonding is supported, and the cable modem termination system (CMTS) maintains the information of logical bonding groups that refer to different sets of channels.

In our lab testing environment, the automation service is programmed to execute the following procedure to ensure hit-less configuration changes for each individual service group at scale:

1. Compute the differences in the spectrum configuration change, create or reuse a set of intermediate bonding groups that make references to SC-QAM channels that are not being changed. These bonding groups are used by the CMs temporarily during the configuration change.
2. Use DBC messages to move all the service group's CMs with different capabilities, such as 4, 8, 16, 24, 32 SC-QAM CMs, to the intermediate bonding groups in parallel, with CM count-based load balancing by bonding capabilities. This step ensures that during the configuration change, the SC-QAM channels used by the CMs, including the primary channels, are not changed. In this step, the "direct" initialization technique is used. The parallel DBC messages can significantly reduce the wait time in a service group with many CMs. And the load balancing implemented by the automation service further reduces potential service impact due to the temporarily reduced total usable spectrum during the configuration change.
3. Apply RF spectrum configuration changes, respectively. These changes include SC-QAM and OFDM channel configurations, MAC domain configurations (primary channels, bonding groups etc.), and synchronization of any other associated configurations. This is also where vendor-specific configurations are handled.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. Once the new RF spectrum configurations are accepted and stable, use DBC messages in parallel to move the CMs to their best-fit bonding groups, with load balancing.
5. As a cleanup step, once step 4 is complete, remove all redundant bonding groups created during the configuration change.

This procedure allows for RF channel changes without service disruption and with minimal and temporary capacity reduction during the change. Each individual configuration change procedure is executed sequentially and changes for different service groups are handled in parallel for scalability.

5.2. Closed-loop Integration and Results

Using this automation service, we conducted experiments to test the spectrum management VNF and the automation service with various CMs in our lab. These CMs include a mix of DOCSIS 3.0 devices with 4, 8, 16, 24, 32 bonding capabilities, and DOCSIS 3.1 devices. The total number of CMs is 32.

In the following test, the existing spectrum configuration has 44 SC-QAM channels and a 96 MHz OFDM channel, as shown in Figure 21.

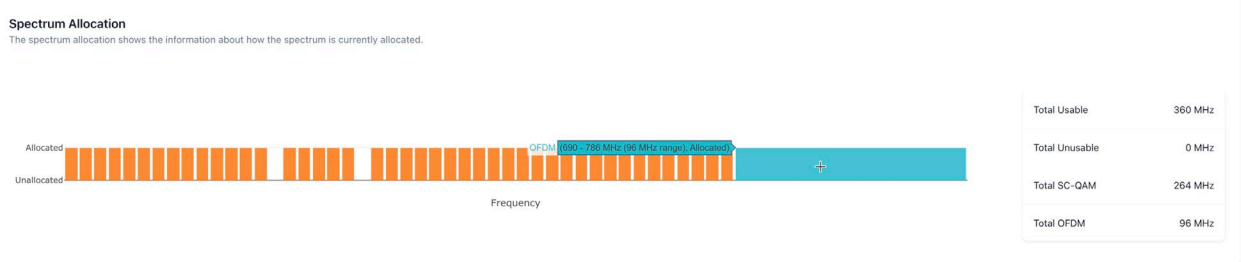


Figure 21 – Closed-loop Integration: Baseline Spectrum Configuration

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

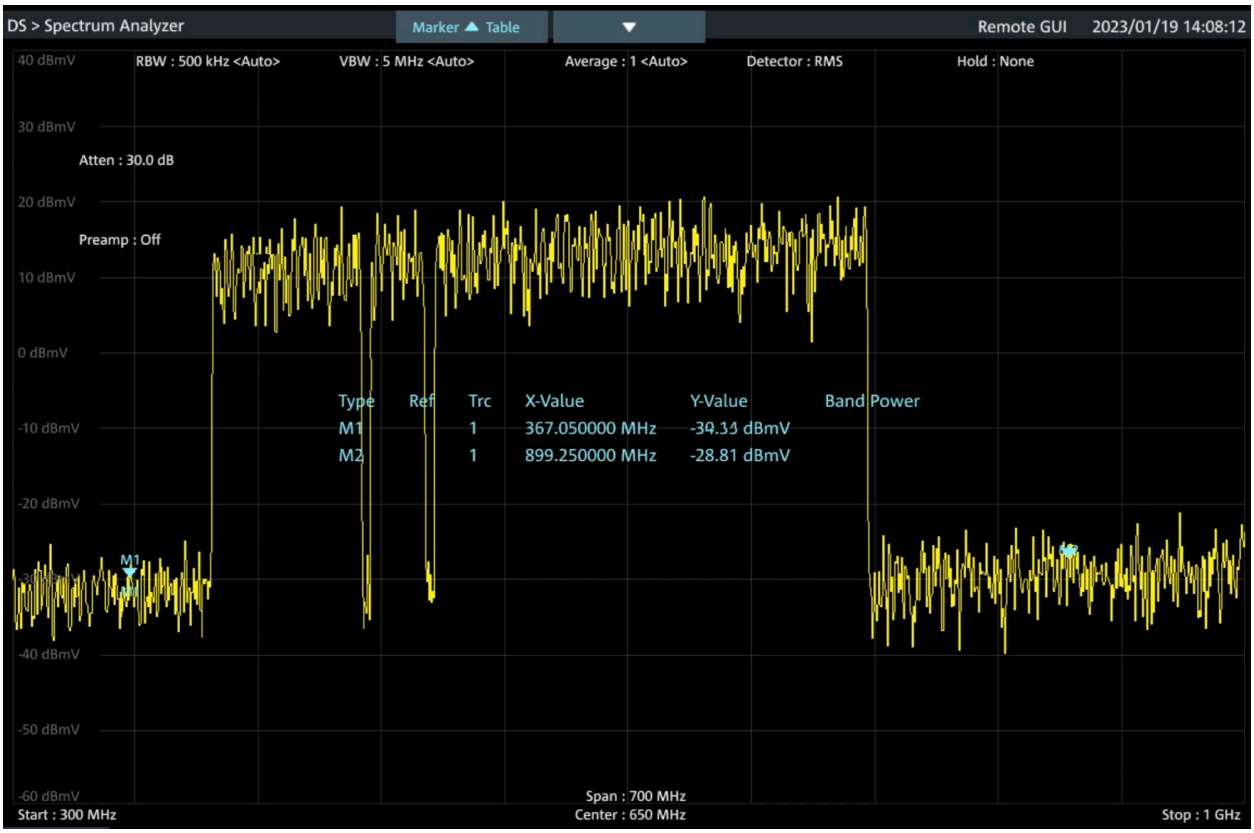


Figure 22 – Closed-loop Integration: Baseline Spectrum Configuration (Spectrum Analyzer View)

The peak throughput tested from a DOCSIS 3.1 device with a 2.5 Gbps Ethernet port is ~1.8 Gbps. The total capacity of this spectrum configuration is greater than 1.8 Gbps. However, because the DOCSIS 3.1 CM can only bond to up to 32 SC-QAM channels, 12 of the SC-QAM channels were not used by the CM during the speed test, reducing the peak throughput by 440 – 450 Mbps.

The optimized spectrum configuration from the VNF is:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

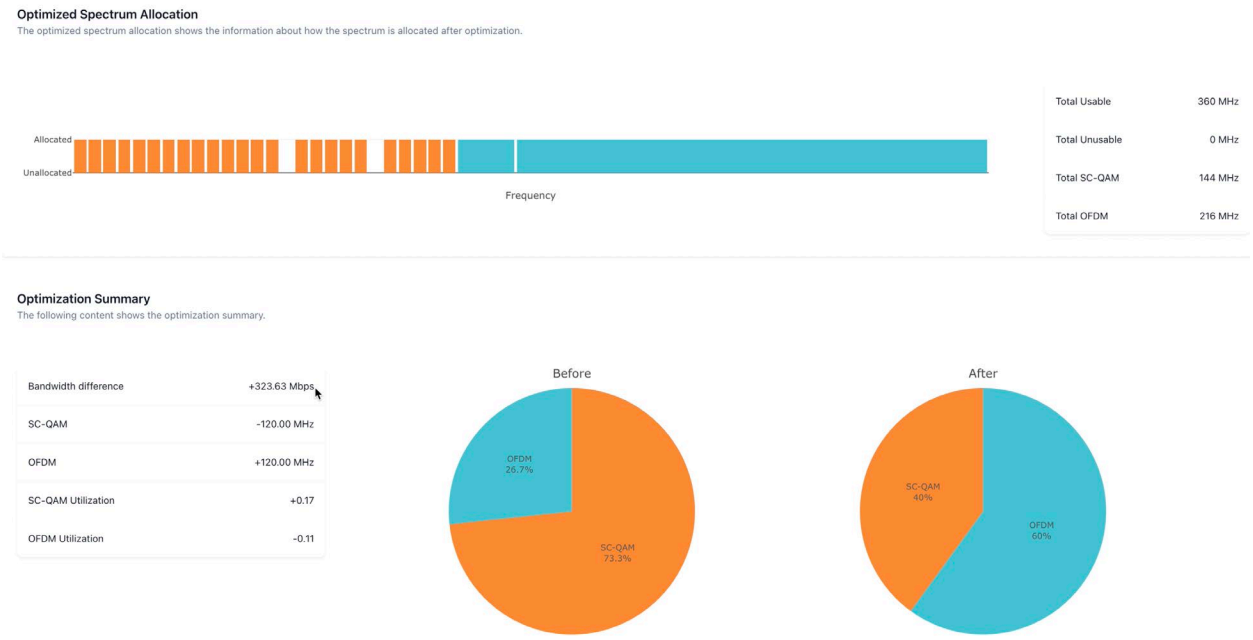


Figure 23 – Closed-loop Integration: Optimized Spectrum Configuration

To test the configuration impact on connectivity and total capacity, background traffic in both downstream and upstream to/from all CMs was scheduled in combination with periodic speed tests. The real-time traffic is displayed on a dashboard shown in Figure 24.



Figure 24 – Closed-loop Integration: Background Traffic

Once the optimized spectrum configuration was sent to the automation service, steps 1 and 2 in the procedure were executed and the CMs were moved to intermediate bonding groups. The DBCs caused a brief 1-2 second traffic rate drop as the CMs had to temporarily use the limited downstream spectrum during the bonding group change. This can be found in the traffic rates history shown in Figure 25 between 7:13 AM and 7:14 AM. After the bonding group changes, the traffic rates stabilized as shown in Figure 25.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 25 – Closed-loop Integration: Traffic Rates (After Moving to Intermediate Bonding Groups)

The automation service then sent the new RF channel configuration to the vCMTS API. As the first step, the RPD turned off the OFDM channel and removed 20 SC-QAM channels, as shown in Figure 26.

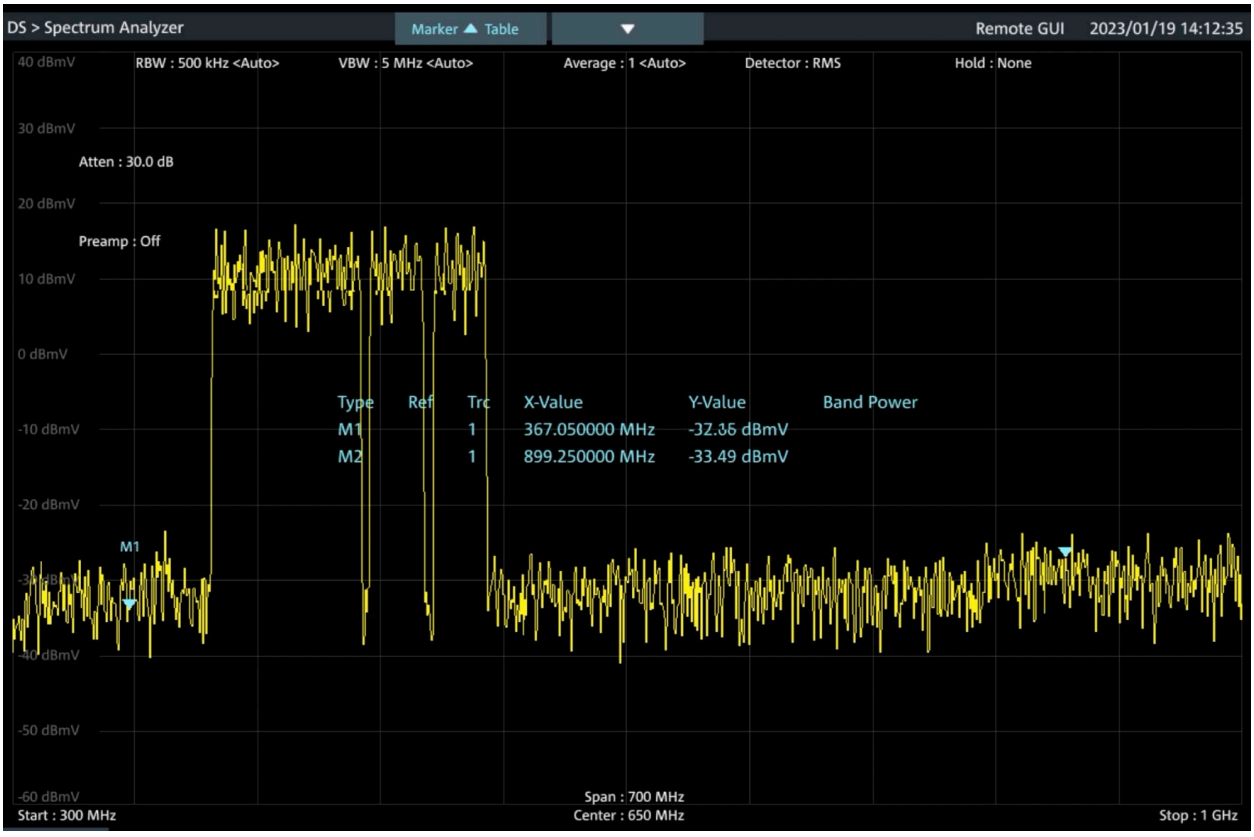


Figure 26 – Closed-loop Integration: Spectrum Analyzer View During the RF Change

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The RPD then turned on the 2 new OFDM channels configured as the VNF specified: a 24 MHz OFDM channel and a 192 MHz OFDM channel. As a result, this configuration change converted 120 MHz of SC-QAM spectrum into OFDM, as shown in Figure 27.

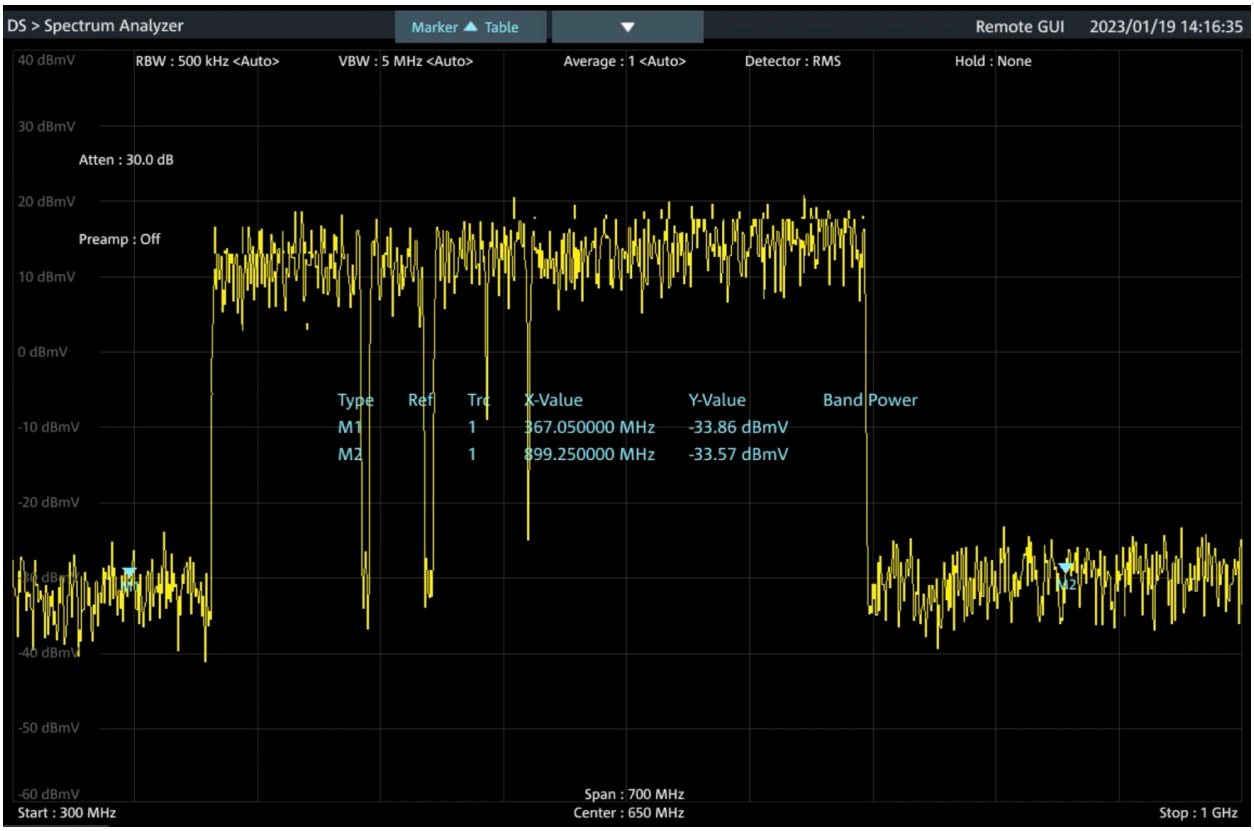


Figure 27 – Closed-loop Integration: Spectrum Analyzer View After the RF Change

Once the RF change was complete and stabilized, the automation service started to move the CMs back to their best-fit bonding groups based on their capabilities. No CM indicated partial service after the bonding change and the background traffic rates recovered to the starting point, as shown in Figure 28.



Figure 28 – Closed-loop Integration: Background Traffic Rates After the Configuration Change

After the configuration change was complete, the DOCSIS 3.1 CMs were fully bonded to the 24 SC-QAM channels and 2 OFDM channels (216 MHz total spectrum) on the downstream. This means it is

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

expected to see a peak downstream throughput increase beyond the estimated 323.63 Mbps capacity gain coming from spectral efficiency improvements. The speed test measured ~2.38 Gbps application traffic throughput on a single CM with 2.5 Gbps Ethernet ports when the channels were under 100% utilization, as shown in Figure 29. As a result of re-allocating the DOCSIS technology across spectrum, the device was able to achieve almost 600 Mbps higher speeds above the baseline configuration. These higher speed capabilities come along with additional benefits to the total capacity and lower utilization of the service group. This was all accomplished with non-service affecting customer impact.



Figure 29 – Closed-loop Integration: Downstream Speed Test After the Configuration Change

6. Conclusion

In this paper, we presented our design of an efficient and effective algorithm for the spectrum management VNF, and various experiments and their results from applying this VNF to production service group data. We also introduced our PoC software and the automation service that we have developed to support closed-loop testing and experiments and a lab demonstration with real-time information captured during the automated, non-service-affecting spectrum change, and shared the peak downstream throughput improvement after the automated configuration change. We conclude that:

- The spectrum management algorithm can effectively account for various real-world constraints that have been identified so far, and it is performant.
- The experiments around applying the VNF to production service group data showed promising average capacity gain and peak speed improvements from automatically converting SC-QAM to OFDM based on each service group's characteristics, without adding new spectrum.
- The PoC software and the automation service helped envision and demonstrate the closed-loop integration for the spectrum management VNFs. And it is possible to perform DOCSIS spectrum changes without disrupting service.

As part of this effort, we also discovered improvements to the load-balancing algorithms that are now being implemented as a VNF outside of the core vCMTS microservice. As the load balancing VNF becomes mature, production trials of this spectrum management technique will be a key element associated with our FDX deployment scale for DOCSIS 4.0.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AGC	automatic gain control
API	application programming interface
bps	bits per second
CAGR	compound annual growth rate
CM	cable modem
CMTS	cable modem termination system
DAA	distributed access architecture
DBC	dynamic bonding change
DCC	dynamic channel change
DOCSIS	Data Over Cable Service Interface Specifications
Gbps	gigabits per second
GHz	giga hertz
GUI	graphical user interface
HFC	hybrid fiber coax
MAC	media access control
Mbps	megabits per second
MHz	mega hertz
MILP	mixed-integer linear programming
OFDM	orthogonal frequency division multiplexing
PHY	physical layer
PMA	profile management application
PoC	proof of concept
QAM	quadrature amplitude modulation
RCS	receive channel set
RF	radio frequency
RPD	remote PHY device
SC-QAM	single carrier-quadrature amplitude modulation
vCMTS	virtual cable modem termination system
VNF	virtual network function
VoD	video on demand

Bibliography & References

1. *Time is Ripe for D3.0 Farming: Achieving Optimal Spectral Efficiency by Allocating D3.0 Spectrum to OFDM*, M. Harb, C. Humble, S. Ozer, and D. Rice, NCTA technical paper, 2022.
2. *A Machine Learning Pipeline for D3.1 Profile Management*, M. Harb, J. Ferreira, D. Rice, B. Santangelo, and R. Spanbauer, NCTA technical paper, 2019.
3. *Full Scale Deployment of PMA*, M. Harb, J. Ferreira, D. Rice, B. Santangelo, NCTA technical paper, 2020.
4. *DOCSIS 3.1 Physical Layer Specification*, CM-SP-PHYv3.1-I19-211110
5. *DOCSIS 4.0 MAC and Upper Layer Protocols Interface Specification*, CM-SP-MULPIv4.0-I05-220328

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Turning Backup Power Reserves Into Green Revenue

A Technical Paper prepared for SCTE by

Douglas Ridley
Director of Engineering
ATX Networks
Newport, NC
(866) 968-7289, Ext. 1315
dridley@atx.com

Todd Pistorese
Chief Technology Officer
Supercapacitor Energy, LLC
Seattle, WA
(206) 399-3815
todd@supercap-energy.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Defining Hybrid Supercapacitors ... 4
3. Understanding EIMs And Energy Buybacks ... 6
4. Telecommunications Providers And The EIM: A Near-Perfect Match ... 8
5. How Energy Arbitrage Works ... 8
6. How Hybrid Supercapacitors Can Generate Revenue ... 9
7. Energy Storage Models ... 10
8. Investment Required for EIM Participation ... 11
9. Conclusion ... 11

Abbreviations ... 12

Glossary Of Terms ... 12

Bibliography & References ... 13

List of Figures

Title **Page Number**

Figure 1 - Electrochemical Energy Storage Dates Back To Ancient Babylon ... 3
Figure 2 - Anatomy Of A Hybrid Supercapacitor ... 5
Figure 3 - Mapping Today's Energy Grid ... 7
Figure 4 - How Energy Rates Vary ... 9

List of Tables

Title **Page Number**

Table 1 - Comparing Energy Storage Technologies ... 6
Table 2 - Energy Sellback Model ... 10
Table 3 - TOU Model ... 11

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Hybrid supercapacitors have several unique technical properties, such as a significantly greater number of charging/discharging cycles than other energy storage devices. As a result, communications service providers (CSPs) adopting this technology for standby power operations can generate revenue by participating in regional energy imbalance markets (EIMs) and pursuing energy arbitrage options offered by their local utilities. The utility industry created the EIM concept a few years ago to establish an energy bid market for providing flexible power reserves that can be tapped as needed to meet customers' energy demands. This overview paper will explain how cable operators and other CSPs can adopt hybrid supercapacitors to create new revenue sources and curb carbon emissions by selling their idle energy back to utility companies — without compromising or interrupting their own backup power operations.

For the most part, battery-based energy storage has not fundamentally changed over the past 2,000 years. In 1938, an excavation team working in southern Iraq unearthed what is now commonly known as the Babylon Battery, an ancient energy source that was essentially an archetype for the batteries that now provide backup power in telecommunications facilities. Like today's lead-acid batteries, the roughly five-inch clay pots found in Iraq featured copper cylinders with encased iron rods, as well as liquid that apparently once acted as an electrolyte.



Figure 1 - Electrochemical Energy Storage Dates Back To Ancient Babylon

Although legacy electrochemical batteries will likely extend their millennium-plus reign for decades to come, the recent introduction of energy storage devices that also incorporate an electrostatic, or non-electrochemical, design offers telecommunications providers, datacenters and other companies the ability to significantly improve the performance and safety of their backup electric power operations. Known as hybrid supercapacitors, these new storage devices also offer telecom providers the ability to create a promising revenue source via participation in energy arbitrage applications and the EIM, the focus of this paper.

The utility industry created regional energy market mechanisms such as the EIM to take advantage of reserve power that can be tapped as needed to meet their customers' ever-changing needs. Given the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

distributed nature of their facilities, telecom providers that adopt hybrid supercapacitors as an energy source could be prime contributors to the EIM, enabling them to generate new revenue.

This paper seeks to provide readers with an introduction to hybrid supercapacitor technology, an understanding of how hybrid supercapacitors differ from traditional energy storage media and an appreciation of how those distinguishing characteristics enhance the technology's ability to generate revenue for service providers through the sale of their idle stored energy. The paper also seeks to educate readers about today's electric power grid and how it functions to meet the nation's energy demands, as well as the critical role that new energy storage media such as hybrid supercapacitors can play in meeting those demands.

2. Defining Hybrid Supercapacitors

Standard supercapacitors are a non-chemical type of energy storage device. Known technically as electric double-layer capacitors (EDLCs), supercapacitors leverage the electrostatic effects that occur between two activated-carbon electrodes with high specific surface areas per volume. The electrodes are immersed in an electrolyte, and a separator between the electrodes is used.

Hybrid supercapacitors are next-generation variants of these devices. Unlike standard supercapacitors, hybrid supercapacitors are partly electrostatic and partly electrochemical in nature and store and release energy through the reversible adsorption and desorption of chemical ions at the interface between electrode materials and electrolytes.

In other words, hybrid supercapacitors leverage both chemical and electrostatic processes to store electric power. Thanks to this ability, they blend the best attributes of both battery and capacitor to create a novel storage medium particularly suited for energy arbitrage applications.

Unlike purely electrochemical devices, hybrid supercapacitors are asymmetrical devices, constructed of a lithium-doped graphite anode and an EDLC-activated carbon cathode. As a result, their charging and discharging cycles are electrostatic processes, free of chemical reactions. This difference eliminates the potential for leakage of harmful chemicals or for thermal runaway, which can cause dangerous, difficult-to-extinguish fires.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

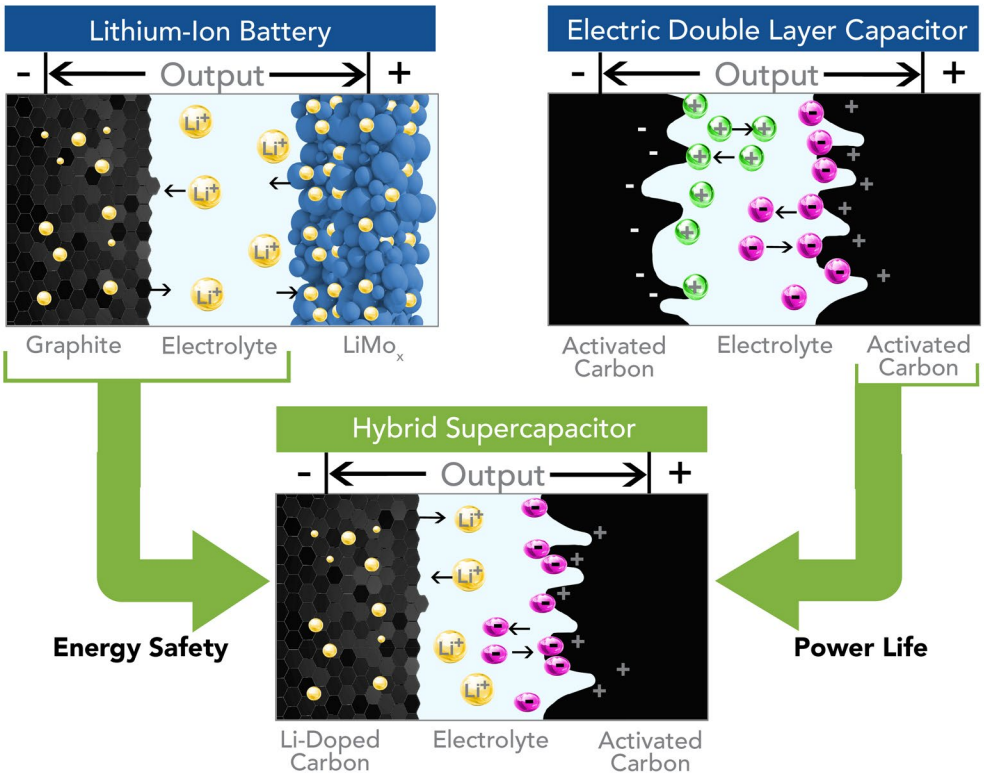


Figure 2 - Anatomy Of A Hybrid Supercapacitor

What makes hybrid supercapacitors different from both standard supercapacitors and older, lead-acid and lithium-ion batteries is that they link two energy-storage mechanisms in parallel, combining lithium-ion technology and EDLC construction in one neat package. Due to this unique combination, hybrid supercapacitors bring together the high-power capabilities of capacitors with the high-energy storage capabilities of batteries. This means that hybrid supercapacitors can safely and efficiently deliver improved performance over chemical-based storage media, as well as offer energy arbitrage applications that can generate revenue in an environmentally friendly way.

Delving into the specifics, hybrid supercapacitors feature a vastly greater number of charging/discharging cycles than conventional batteries. Hybrid supercapacitors have an expected lifespan of roughly 20,000 to 50,000 charging/discharging cycles, based on full depth of discharge (DoD), a key energy storage metric, at 25 degrees Celsius. That's significantly more than the roughly 300 lifecycles of lead-acid batteries, based on less than 80% DoD at 25 degrees Celsius.

The relatively low charging rate of lead-acid batteries renders them essentially ineligible for energy arbitrage. Looking at the middle range, lithium-ion batteries, with a lifespan of no more than 5,000 cycles for a 90% DoD at 25 degrees Celsius, are currently deployed in some energy arbitrage environments. However, their lifespan pales in comparison to that of hybrid supercapacitors and they are prone to thermal runaway.

Furthermore, nearly 100% of a hybrid supercapacitor's storage capacity is usable energy because its self-discharge is mitigated by the lithium doping of the device's anode and cathode. That's due to the addition of the electrochemical component to the EDLC, which has notorious self-discharge issues because it is

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

purely electrostatic. As a result, hybrid supercapacitors can maintain their voltage over much longer stretches of time than other energy storage media without the need to connect to a charging source.

In addition, hybrid supercapacitors are significantly safer to operate and maintain than lithium-ion or lead-acid batteries for two main reasons. First, hybrid supercapacitors do not overheat like the other energy storage devices because they are very efficient, achieving greater than 97% energy efficiency, because of their electrostatic component.

Second, hybrid supercapacitors do not run the same risk of "thermal runaway" as the other main energy storage media. Although they do contain electrolytes that could leak if damaged and are flammable, a puncture or other damage would not generate enough internal heat to generate the chemical process of thermal runaway. That's because hybrid supercapacitors have a different chemical composition than lead-acid and lithium-ion batteries, the chemical composition does not contain oxygen or metal oxide. A hybrid supercapacitor efficiency cannot generate enough heat to break down the internal chemical composition. Hybrid supercapacitor efficiency combined with the lack of metal oxides eliminates its ability to feed fires or propagate combustion associated with "thermal runaway" that's inherent with traditional lithium-ion batteries available today.

Table 1 - Comparing Energy Storage Technologies

	Hybrid Supercapacitors	Lithium-Based Batteries	Lead-Acid Batteries
Type	Electrostatic & Electrochemical	Electrochemical	Electrochemical
Lifecycles	20,000–50,000	5,000	300
Depth of Discharge	100%	80-90%	60-80%

3. Understanding EIMs And Energy Buybacks

As noted in the paper's introduction, EIMs are regional energy marketplaces that the utility industry created to supply electricity generation and transmission services whenever and wherever they are required. The idea is to provide flexible energy reserves to power companies, reserves that can be tapped at any time to meet customers' ever-changing energy needs. The figure below depicts the common components in today's energy grid, as well as its distribution infrastructure.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

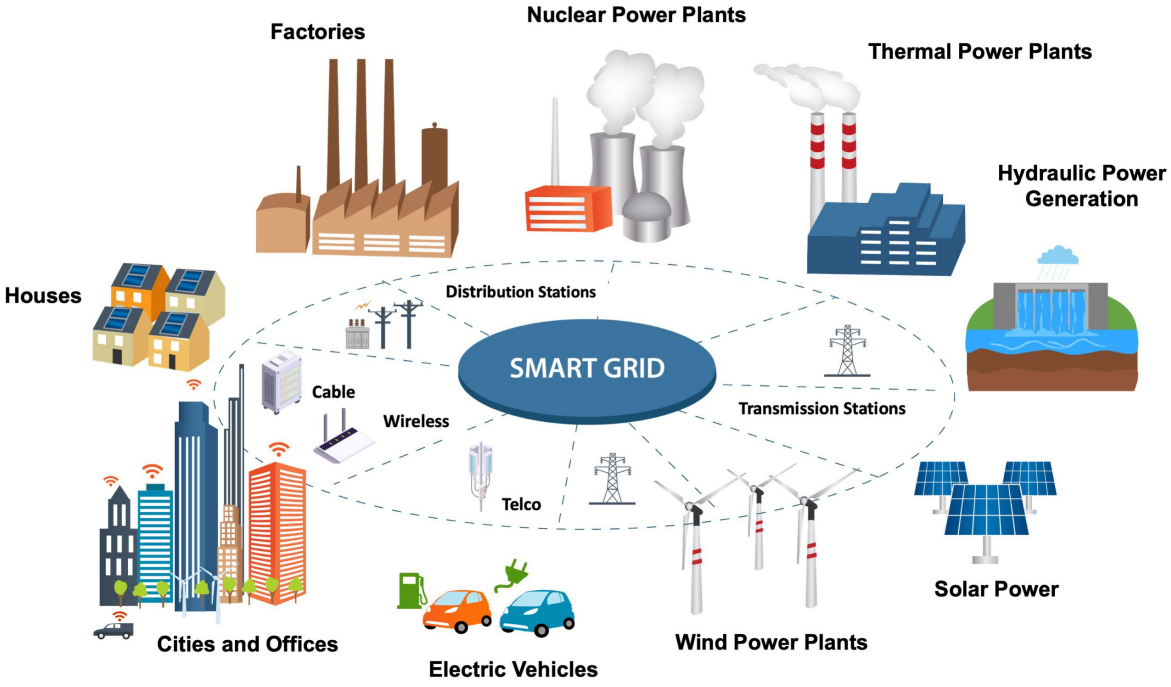


Figure 3 - Mapping Today's Energy Grid

EIM participants may come from multiple public authority areas and utility territories. Combining in this way increases the economic efficiency of the nation's power system and enables centralized, automated and region-wide power generation and distribution. As more utilities join an EIM and boost its size, range and diversity, the benefits increase for all customers.

Currently, most utilities rely predominantly on legacy power plants, which burn heavily polluting fossil fuels, to meet peak customer demand for electricity because they have few other energy sources available to tap. These reserve resources, known as spinning reserves, are significant contributors to the planet's growing carbon footprint and accelerating global warming. As more of the planet experiences extra-hot temperatures of the type seen in summer 2023, what was "peak" demand yesterday becomes tomorrow's "normal" high level of demand.

But hybrid supercapacitors could start addressing that problem. Thanks to their unique features, as detailed earlier, hybrid supercapacitors could help utilities meet their customers' energy needs by filling the gap more efficiently and economically, and in a more environmentally sustainable way, than current energy storage solutions, many of which are still reliant on fossil fuels. Hybrid supercapacitors can also be managed to contribute energy reserves to the EIM without interfering with their own backup power responsibilities. Thus, they could enable CSPs to engage in energy arbitrage without compromising their ability to maintain continuous operations.

To qualify as ancillary energy suppliers, entities must have a megawatt-hour or more of electricity available to dispatch to the power grid to meet the needs of the regional energy imbalance market. Telecom providers can easily satisfy this requirement by taking advantage of the power of hybrid supercapacitors.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Benefits like these make hybrid supercapacitors strong candidates for enabling the sale of excess, renewable electrical power back to an EIM and the overall energy grid. In other words, hybrid supercapacitors can leverage their unique electricity supply capabilities to produce potentially substantial new revenues for telecom providers — while also lowering carbon emissions to preserve the earth's health.

4. Telecommunications Providers And The EIM: A Near-Perfect Match

Telecommunications providers can generate more revenue by taking part in energy arbitrage markets and the EIM. Apart from the other factors already cited, the revenue potential is due in large part to the highly distributed nature of telecommunications facilities, which typically are scattered throughout regions, cities and even neighborhoods. Such a pattern lines up neatly with the distributed nature of the nation's energy distribution grid.

A national wireless provider like AT&T®, for example, might operate thousands or even tens of thousands of cell towers across the country. As a result, an operator of that size is likely to have thousands of megawatt-hours of untapped energy capacity waiting to be used across its various sites.

Because most major telecom operators rely heavily on conventional lead-acid batteries in their transmission facilities, all that stored electricity can only be used to provide backup power in the event of a utility blackout or service interruption. As noted earlier, lead-acid batteries have limited charging/discharging cycles and thus are not able to support energy arbitrage applications.

Even if facilities can leverage the newer lithium-ion batteries for power, that's still far from an ideal solution for participating in the EIM. That's because lithium-ion batteries also lack the extended charging/discharging cycle lives of hybrid supercapacitors. Thus, lithium-ion batteries degrade much faster, which means they would need to be replaced much sooner than hybrid supercapacitors.

Hybrid supercapacitors offer one other key benefit. Unlike other energy storage solutions, they can ramp up quickly to respond to sudden spikes in electricity demand. That attribute makes them well suited to take advantage of ongoing price changes in the real-time energy markets as electricity demand fluctuates throughout the day, week, month or season.

Note: AT&T is the registered trademark of AT&T Properties, L.P.

5. How Energy Arbitrage Works

How can communications service providers make money by having more energy than they need to power their own facilities? They can take a page out of the existing energy arbitrage business, building up their energy reserves and then selling them to power companies at the right time, netting a healthy profit on their stored electricity.

Fundamentally, energy arbitrage is defined by the stock market's tried-and-true maxim of buying low and selling high. In the case of power, this can be carried out by leveraging a time-of-use (TOU) rate structure, wherein the utility penalizes the use of power during peak demand hours and rewards power consumption during off-peak hours through the prices it sets for each. With the help of hybrid supercapacitors, telecom providers can continue operations as usual but choose when to dispatch and recharge their energy capacity. Telecoms who participate in TOU can also take advantage of significant savings reducing their overall energy cost while playing a productive role aiding the utilities during high-demand hours.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Not surprisingly, providers should generally seek to sell their stored energy when electricity rates rise to their highest levels and replenish their supplies when energy rates sink to their lowest levels. Because energy rates can vary widely over the course of a day or even just a few hours, such timing is critical to success. For example, as depicted in the figure below, wholesale energy rates may drop to as low as 5 cents per kilowatt-hour (kWh) during the super off-peak hours late at night and then rise as high as 25 cents per kWh during the peak demand periods of late afternoon and early evening.

Fortunately, no elaborate calculations are required to carry out this strategy. That's because energy arbitrage is based on published utility TOU rates so it can be calculated quite easily in advance and requires no additional equipment to participate. It is simply a matter of setting the rectifiers to operate on the power grid during low-rate periods and then off the grid during high-rate periods, thereby taking advantage of the excess energy that was stored internally during the off-peak hours.

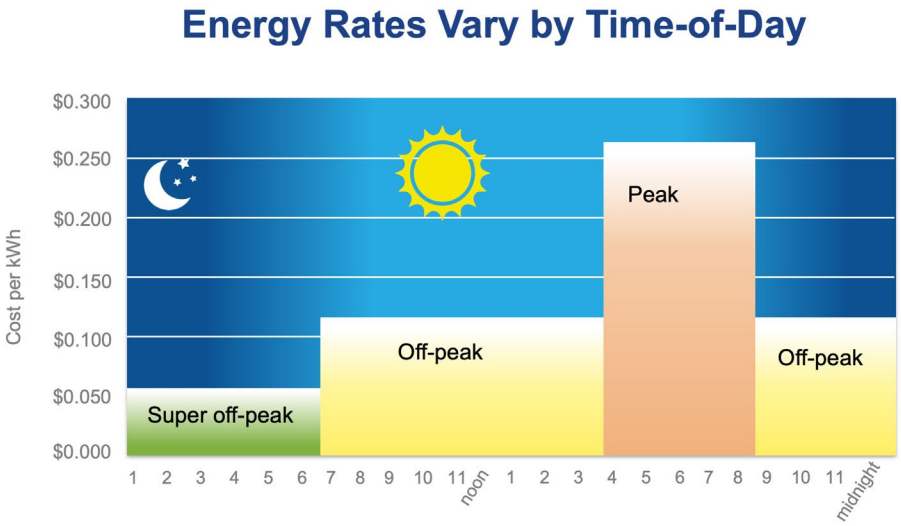


Figure 4 - How Energy Rates Vary

By engaging in energy arbitrage, cable and other telecom providers could generate substantial revenue from an otherwise idle asset in their facilities. Plus, by replacing heavily polluting energy from fossil fuel-powered plants with clean energy stored by hybrid supercapacitors, operators could help rid the planet of some of its biggest climate-change accelerators.

To illustrate this point, the following section considers several basic energy arbitrage revenue models.

6. How Hybrid Supercapacitors Can Generate Revenue

As noted in the previous section, wholesale energy rates at different market levels vary greatly throughout the day and, at their peak levels, can greatly exceed the cost of electricity sold to consumers. Although hourly energy prices vary by both season and location, prices generally increase as the energy load increases.

So how much revenue could telecom providers generate by engaging in energy arbitrage and/or participating in the EIM with the aid of hybrid supercapacitors? Although the initial financial projections are rough, they indicate that operators could generate a return on investment (ROI) in as few as six or

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

seven years, depending on the market in which they are participating and their supply/demand circumstances, by selling their idle stored energy back to the power grid and electric utilities.

As explained by Green Rhino Energy (www.greenrhinoenergy.com), a green-energy project developer and consultant, the greatest revenue stream would come from selling electricity back to the power grid at either a fixed rate (the TOU use case) or a variable market price (the EIM use case). Even if the telecom facility was not connected to the greater power grid, the operator would still benefit from reduced generator run-time. In addition, depending on the state or region, the service provider might be able to generate and sell renewable energy certificates or carbon emission reduction certificates to others.

If they make use of solar power to produce electricity, cable and other telecom operators could also reap revenue from a second income stream: federal and state tax benefits. These benefits include both *production* tax credits, which offer set amounts for every kilowatt-hour of solar power produced over a fixed-time period, and *investment* tax credits, which enable companies to offset some or all of their initial investment over a certain number of years against pre-tax profits. The value of the tax credits depends on the investor's overall tax capacity.

Previously, CSPs would need to install solar to realize significant tax benefits. Federal investment tax credit (ITC) has been extended to energy storage only.

7. Energy Storage Models

As noted in the previous section, wholesale energy rates at different market levels vary greatly throughout the day and, at their peak levels, can greatly exceed the cost of electricity sold to consumers. Although hourly energy prices vary by both season and location, those prices generally increase as the energy load increases.

How much revenue telecom providers generate by engaging in energy sellbacks and participating in the EIM with the aid of hybrid supercapacitors is dependent on several variables, including the available energy storage, the frequency of participation and the rate at which the energy is sold back to the utility.

The following chart provides an example of an energy sellback model that is based on a telecommunications company with 100kWh of available storage selling off all available energy storage for an hour a day, over 91 days of the year. The rate per hour is based on public information from Southern California Edison.

Table 2 - Energy Sellback Model

	Rate Per Hour	Rate Per Hour	Rate Per Hour
	$3.79706	$3.65744	$0.91592
Revenue Per Day (1 hour)	$379.70600	$365.74400	$91.59200
Revenue Per Year (91 days)	$34,553.24600	$33,282.70400	$8,334.87200

As the chart indicates, communications providers that participate in energy sellback when rates are at their highest ($3.80) can generate nearly $35,000 annually by discharging energy storage for just one hour over three months of the year. Even at a rate of less than $1, annual revenue can exceed $8,000.

Though not the subject of this paper, telecommunications providers can also reduce energy costs by using idle energy storage to power operations when rates are at their highest and then recharging hybrid

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

supercapacitors when energy storage costs are at their lowest. This energy savings model, known as peak shaving or Time of Use (TOU), can result in significant annual savings for telecommunications providers with a similar energy storage profile as the provider in the sellback example.

Table 3 - TOU Model

	Summer Weekdays (87)	Summer Weekends (34)	Winter Days (244)	
Peak Rate (kWh)	$0.66	$0.38	$0.60	
Non-Peak Rate (kWh)	$0.26	$0.26	$0.23	Total
Gross	$28,710.00	$16,530.00	$26,100.00	$71,340.00
Grid Recharge Cost	-$2,262.00	-$2,262.00	-$2,001.00	-$6,525.00
Net	$26,448.00	$14,268.00	$24,099.00	$64,815.00

*Peak hours per day: 5

*Non-Peak hours per day: 19

Minus the costs to recharge the energy storage modules, the telecommunications provider depicted by the TOU chart, which is also based on typical residential rate charges associated with Southern California Edison, can save nearly $65,000 annually.

8. Investment Required for EIM Participation

To engage in energy arbitrage and/or participate in the EIM, communications providers will need to invest in new equipment, ranging from the hybrid supercapacitors themselves to the supporting gear, components and software.

In addition to the hybrid supercapacitors, CSPs will need to review their current electrical solutions to support a demand response architecture.

Demand Response - Demand response is a process where the customer plays a role in the operation of the electric grid by reducing or shifting their electricity usage during peak periods in response to time-based rates or other forms of financial incentives. Alternatively, demand response can be used for balancing supply and demand. It also includes direct load control programs which provide the ability for power companies to cycle electrical loads during peak demand in exchange for a financial incentive.

Demand Response components will generally consist of the following components: metering/net metering, inverters, rectifiers, software, etc.

9. Conclusion

The energy imbalance market (EIM) offers a promising new revenue opportunity for cable operators and other telecom providers, especially as they transition to more renewable energy sources. Just as a few pioneering utilities are already doing, telecom players could draw on the idle stored energy sitting in their facilities and feed that excess power to the nation's electric power grid — all the while generating new income.

As this paper has discussed, hybrid supercapacitors offer a safe, efficient, cost-effective and environmentally sustainable way for service providers to store electric power and deliver it to the grid. By

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

leveraging the unique qualities of hybrid supercapacitors, operators can meet their facilities' backup energy requirements AND sell their excess electricity to other players in the energy market.

Although it is difficult to project right now exactly how high the proceeds of energy arbitrage might be, the market is clearly wide open for cable operators and other telecom providers to participate by adopting hybrid supercapacitors and taking advantage of their unique energy storage capabilities. For the sake of both their business and the world's climate, the sooner they can start taking these steps, the better.

Abbreviations

CSP	communication service provider
DoD	depth of discharge
EDLC	electric double layer capacitor
EIM	energy imbalance market
ITC	Investment tax credit
kWh	kilowatt-hour
ROI	Return on Investment
SCTE	Society of Cable Telecommunications Engineers
TOU	time of use

Glossary Of Terms

Term	Definition
Depth of discharge	The percentage of an energy storage device's capacity that has been discharged relative to the device's total storage capacity.
Demand response	A process where the customer plays a role in the operation of the electric grid by reducing or shifting their electricity usage during peak periods in response to time-based rates or other forms of financial incentives.
Energy arbitrage	A system of buying power during off-peak hours when electricity grid prices are the lowest and then storing the power to sell back to the grid during peak hours when electricity prices are the highest.
Energy buybacks	A way to leverage the energy arbitrage markets by purchasing power when grid electricity prices are low for either storage or sale at a later time.
Energy imbalance market	A voluntary market that economically delivers power from participating utilities and other energy producers to balance regional energy supply and demand on a frequent, timely basis.
Hybrid supercapacitor	A variant of standard supercapacitors that combines lithium-ion technology and electric double layer capacitor construction for improved performance.
Lead-acid battery	A type of rechargeable battery that uses lead and sulfuric acid to function. The lead is submerged into the sulfuric acid to generate a controlled chemical reaction, which causes the battery to produce electricity.
Lithium-ion battery	A type of rechargeable battery that uses the reversible reduction of lithium ions to store energy. The negative electrode of a conventional

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

	lithium-ion cell, sometimes called the anode, is typically graphite, a form of carbon. The positive electrode, sometimes called the cathode, is typically a metal oxide.
Peak shaving	A process that proactively manages overall demand to handle short-term spikes in energy demand by lowering and smoothing out peak loads, thereby reducing the overall cost of electricity.
Supercapacitor	A type of electrochemical energy storage system that has great power density and specific capacitance, enabling the efficient release of large amounts of energy over a relatively short period of time.

Bibliography & References

California ISO. March 2, 2020. *Energy Storage and Distributed Energy Resources Initiative: Second Revised Straw Proposal, Day 1*. Public Presentation

California ISO. February 14, 2018. *Meter Data Acquisition and Processing Procedure*. Operating Procedure 5740, Version 1.3.

Green Rhino Energy. Retrieved August 10, 2023, from http://www.greenrhinoenergy.com/

Smart Electric Power Alliance. September 2017. *DER Aggregations in Wholesale Markets*. Technical Paper.

Southern California Edison Company. December 13, 2019. *The Interconnection Handbook Southern California Edison Company*. Handbook, Version 10.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Unlocking the Power of Big Data: “The Key Workforce Skills Operator’s Need for the Future of Analytics”

A Technical Paper prepared for SCTE by

Steven R. Harris
VP Global Market Development | Office of the CEO
Society of Cable Telecommunications Engineers, Inc.
140 Philips Rd, Exton, PA 19341
1-610-594-7324
sharris@scte.org

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
   1.1. Importance of Big Data in the Broadband and Telecommunication's Industry ... 5
   1.2. The Need for Advanced Workforce Skills ... 7
   1.3. Overview of the Paper's Objective ... 7
2. Understanding Big Data and its Relevance to Operators ... 8
3. Fundamentals of Data Science and Engineering for Operators ... 9
   3.1. Personalization of the Customer QoE and Business Decisions through Data Utilization ... 11
4. Applications of AI/ML in Telecom Networks ... 12
   4.1. Real-World Examples of Supervised AI/ML Applications ... 12
   4.2. Practical Illustrations of Unsupervised AI/ML in the Telecom Industry ... 13
   4.3. Applying Big Data Analytics to Enhance Telecom Networks ... 14
5. Managing Big Data in Telecom ... 15
   5.1. Beyond SQL: Benefits of Unstructured Big Data Databases ... 17
   5.2. Summarizing and Visualizing Telecom Data for Better Decision-Making ... 17
6. Key Workforce Skills for Handling Big Data ... 18
7. SCTE's New Programs for Data Analytics and Python ... 19
8. Conclusion ... 19

Abbreviations ... 21

Bibliography & References ... 22

List of Figures

Title **Page Number**

Figure 1 - Digital Transformation by Liberty Global ... 3
Figure 2 - Global Data Generated Annually, in 2025, hitting 181 zettabytes (Source: Statista) ... 4
Figure 3 - The Data Science Process ... 4
Figure 4 - Tableau, SCTE 2023 Members by Country ... 5
Figure 5 - Kirkpatrick New World Model and Phillips ROI Model ... 6
Figure 6 - Absorb Analyze ... 6
Figure 7 - Types of Data Analytics ... 9
Figure 8 - Natural Language Processing ... 10
Figure 9 - Differences between Statistics, Data Science, AI, ML, and Automation ... 11
Figure 10 - Data Pipeline ... 15
Figure 11 - Cloud Models ... 16
Figure 12 - Visualization Examples: Heatmap, Geospatial and Dashboards ... 17
Figure 13 - Moodle LMS Visualization Heatmap, Identification of Popular Learning Assets ... 18
Figure 14 - SCTE Big Data Courses ... 19

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Today's global broadband and telecommunications companies are deep into their digital transformation strategy, compelling operators to be more analytics and data-driven shortly to remain highly competitive in our industry. Liberty Global and other organizations like Comcast, Cox, Rogers, Orange, Plume, and Vodafone all are spearheading digital transformation in broadband and telecommunications.



Figure 1 - Digital Transformation by Liberty Global

Operators who have a handle on vast amounts of data allow them to have a deeper understanding of the business, as well as a competitive edge. For example, with the advancement of data collection and analytical capabilities, all the analytics teams across Cox's technology and operations were centralized [14]. The vast amounts of data paradigm will be referred to herein as "big data". Operators that leverage workforce-related big data, along with data analysis will enhance their operational goals, as well as predict behaviors of their workforce and forecast productivity improvements to drive customer quality of experience (QoE). Moreover, a proficient workforce in big data, analytics, and associated technologies will play a pivotal role in enabling their organizations to succeed in their endeavors toward digital transformation strategies. Contemporary broadband and telecommunications data processing systems need to exhibit greater agility compared to their predecessors as data is ubiquitous, and data ingest is growing exponentially each year. Every day, Facebook produces an immense volume of data, totaling 4 petabytes (PB), equivalent to one million gigabytes. This vast dataset finds its home in a storage system referred to as the "Hive", boasting a staggering storage capacity of approximately 300 PB [9].

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

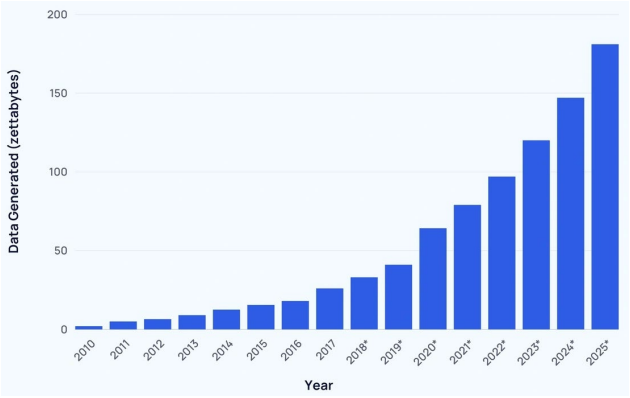


Figure 2 - Global Data Generated Annually, in 2025, hitting 181 zettabytes (Source: Statista)

Furthermore, identifying the significance of ingesting, navigating, storing, transforming, and organizing big data pipelines will lead to workforce efficiencies in the business. The significance of big data pipelines is that they help organizations make sense of the vast quantities of big data they create and stream at unprecedented rates. A workforce that recognizes the key fundamentals of data science and data analysis allows those who work within our industry to be more efficient and streamlined in the areas of big data transformation. For example, correlating data from a data lake, a central repository of structured, semi-structured, and unstructured data, offers the association of big data to provide a deeper insight into workforce and customer key performance indicators (KPI). Using enhanced KPIs allows operators to track trends when a new product is launched, a network outage occurs, or when the school returns to session. In addition, utilizing statistics, artificial intelligence (AI), machine learning (ML), deep learning, and automation will drive big data-based decision-making within the organization.

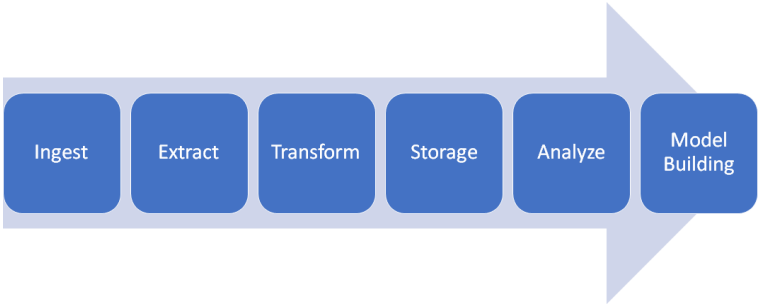


Figure 3 - The Data Science Process

Using structured query language (SQL) and complex pivot tables for today's data visualization is not always a feasible option. In light of ongoing digital transformation efforts within our networks, operators must extend their focus beyond SQL and delve into unstructured big data systems to fully harness the advantages they offer to our industry. This shift will be essential for capitalizing on the potential benefits of these advanced big data management approaches. Summarizing and visualizing broadband and telecommunications big data aids in its accessibility and comprehension for decision-makers, the workforce as a whole, as well as learning and development (L&D) teams. Workforce proficiency in data science (e.g., Kafka, Hadoop), data sources (e.g., Absorb LMS) visualization tools (e.g., Tableau, MS Power BI), data transformation and programming languages (e.g., Python, Scala) are essential for working with big data and data lakes, as well as familiarity with its architecture.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

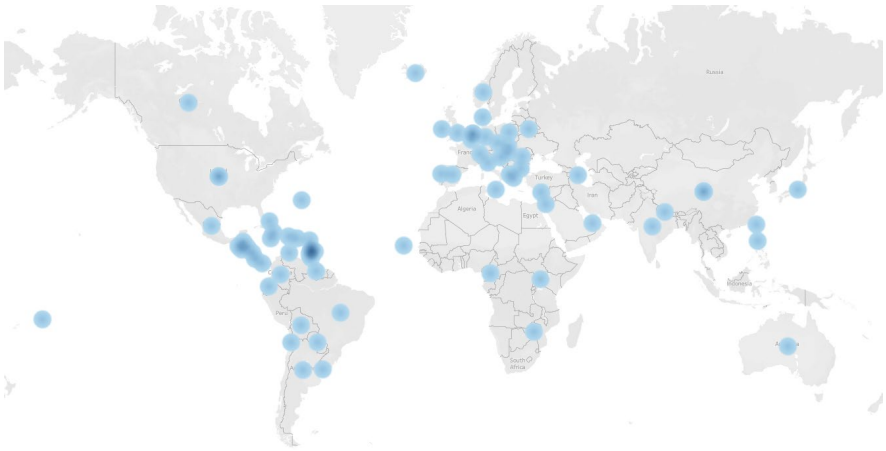


Figure 4 - Tableau, SCTE 2023 Members by Country

1.1. Importance of Big Data in the Broadband and Telecommunication's Industry

According to Gartner [10], by 2024, organizations that fail to build sustainable and scalable big data analytics frameworks will likely experience performance setbacks of at least two years. As shared, big data refers to extremely large and complex sets of data that cannot be effectively processed, managed, analyzed, and visualized using traditional data processing tools and methods such as a relational database (e.g., SQL). The term big data encompasses three primary characteristics, often referred to as the three Vs known as volume, velocity, and variety [1]. Big data involves vast amounts of data generated from various sources, such as social media sites, our broadband CPE, network sensors, mobile device data (images, texts, videos), website-based data, employee key KPIs, eLearning interactions, psychometric data analysis, and more. The sheer volume of data in the broadband and telecommunication industry is typically in the terabytes (TB), PB, or even exabytes (EB). Today with operator digital transformations of their networks, big data is generated at an incredibly high speed and velocity. Often data like network performance or a technician install is produced in real-time as well as near-real-time, and its accumulation can be rapid, requiring immediate pipeline processing and data analysis. Big data comes in various formats and types, including structured data (e.g., SQL-based), semi-structured data (e.g., XML, JSON), and unstructured data (e.g., text, PDF, multimedia, spatial). Data may reside in different vendor-based systems (e.g., DOCSIS, PON) and require aggregation into a data lake for processing. Furthermore, the diversity of data formats adds complexity to the data analysis process. Beyond the three Vs, big data is also characterized by a fourth and fifth "V" known as veracity and value, referring to the trustworthiness, reliability, and worth of the data collected [14]. Big data may include field inconsistencies, errors, and noise due to its diverse sources and unstructured nature, making data quality an important consideration. According to Tractica, companies from the telecommunications industry are expected to invest $36.7 billion per year in AI-related solutions like software, hardware, and services by 2025 [4].

Big data analytics has become increasingly important across various industries besides broadband and telecommunications, including finance, healthcare, retail, and marketing. In the L&D space, data analytics has become crucial and will reign supreme in the future. As budgets tighten, big data analysis becomes even more important to keep a pulse on field and network activities, as well as aligning business stakeholders and showing return on investment, or ROI, for training events. L&D folks refer to the Kirkpatrick New World Model or the Phillips ROI Model that builds on Kirkpatrick to show ROI. Big data and analytics allow operators to be proactive instead of reactive. According to

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Knowledge Base Value Research, the global telecommunications analytics market is expected to grow to $8.7 billion by 2025 [4].

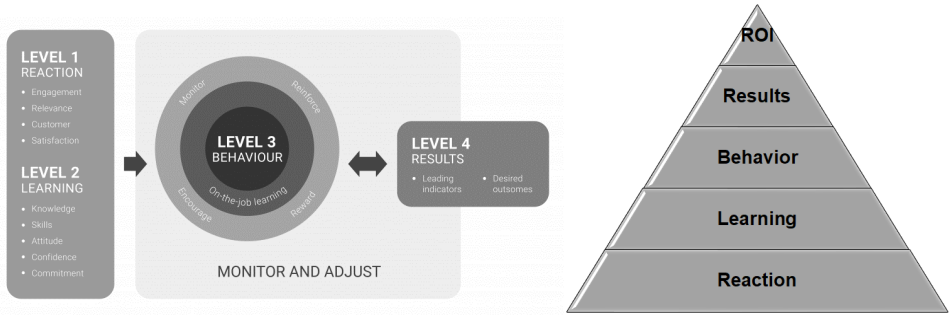


Figure 5 - Kirkpatrick New World Model and Phillips ROI Model

"Data observability has become a necessity, not a luxury, for us. As the business has become more data-driven, nothing is worse than allowing leadership to make a decision based upon data that you don't have trust in. That has tremendous costs and repercussions." - Alex Tverdohleb, VP of Data Services, Fox Networks

By harnessing the potential of big data, operators can gain valuable insights, make data-driven decisions, improve calculated risk, identify patterns, understand trends, develop correlations, and ultimately improve efficiency, productivity, and competitiveness, and drive down cloud costs. Reducing cloud expenses holds significance, given that 81% of IT leaders indicate that their c-suite has directed either a halt in extra expenditures or a decrease in cloud-related costs [3]. Operator L&D organizations and those utilizing SCTE L&D can harness big data throughout their operations to analyze their workforce, including conducting productivity trend analyses.



Figure 6 - Absorb Analyze

Analytics solutions that mine structured and unstructured data are important as they will help operators gain insights not only from their privately acquired data but also from large amounts of data publicly available on the Internet [2]. Such data can be used to improve employee and customer QoE. For example, Vodafone analytics was created to improve business intelligence using big data, using location and mobility insights to help their business customers. Handling broadband and telecommunication big data requires advanced technologies, tools, and techniques that can efficiently process, store, and analyze large volumes of data. Traditional relational databases and data processing methods are often inadequate to cope with our big data challenges. As a result, specialized technologies like distributed computing frameworks (e.g., Hadoop, Azure), AI models, and ML models are used to derive insights and value from our big data sets. Comcast's big data initiative has grown from a small Hadoop cluster to over 10 PB and over 430 nodes in less than two years, allowing

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

them to understand how customers are using their products [5]. The Hadoop framework consists of the Apache Spark (formerly MapReduce) data processing framework and the Hadoop distributed file system (HDFS) serving as the data storage framework.

In the education sector, these technologies allow higher education to analyze learner behavior, assessment data, learner preferences (e.g., modality), learner training time, asset likes, career pathing of learners, and society educational requirements. Data from L&D and higher education will revolutionize the way our workforce learns and instructors present materials at in-person training-led events. The goal is to create a better learner or workforce, meeting them where and how they learn. The facilitator can monitor all learners in real-time and have a deeper conversation on the topics with the learner, providing a stronger understanding of the desired learning behavior [6].

1.2. The Need for Advanced Workforce Skills

As our industry changes at a rapid pace, technology is evolving quickly, and operators must pivot faster than ever. Broadband and telecommunications rely on a knowledgeable and well-skilled workforce with high competency, delivering the latest connectivity options to our customers. Our future workforce must maintain complex systems, implement technological advancements, apply network security, optimize our network, connect a proliferation of devices, make informed decisions with analytics, be well-versed in cloud technologies, collaborate, have strong interdisciplinary skills, and leverage emerging technologies! Advanced workforce skills are essential to meeting the future challenges and opportunities of our industry, furthering our growth and innovation. An employee base that ensures an exceptional customer QoE will result in cost savings for operators through churn reduction and lowered expenses related to acquiring new customers [4]. Leveraging big data additionally empowers our learners to deliver a personalized customer QoE, in the case of Comcast and Vodafone, understanding how customers are using their products and how employees interact with their customers.

It was cited by Pearson Education, that roughly seven (7) in ten (10) employees are currently in job roles where the future is uncertain. It was also mentioned by Pearson Education that occupation redesign coupled with workforce training programs could promote growth in occupational areas that will shrink [7]. Occupation redesign increases productivity, reduces employee churn, increases employee morale, develops new skills (e.g., data science), increases motivation (e.g., less monotone), and encourages innovation. Occupation or job role redesign typically involves a comprehensive analysis of job roles, tasks, skills, behaviors, responsibilities, and interactions within the broader organizational context. Employers are seeking more soft skills such as critical thinking, analysis, complex problem-solving, self-management, and active learning skills states the World Economic Forum (WEF). From a technical perspective, SCTE has collected data on ten trending technical skills: fiber optics, wireless (e.g., 5G, IoT), Wi-Fi, Cybersecurity, cloud computing, DOCSIS, construction practices, network design, network connectivity, and RF theory. All of these areas will require soft skills, as well as technical skills.

1.3. Overview of the Paper's Objective

With broadband and telecommunication data being ubiquitous and its collection growing exponentially, operators and their workforce need to recognize the significance of ingesting, navigating, transforming, storing (e.g., cloud), organizing, and harvesting big data. Understanding the core principles of big data, data science, and data analytics will enable our workforce and networks to operate with enhanced efficiency and seamless coordination.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The paper intends to provide an understanding of how operators can utilize statistics, AI, ML, and automation to drive data-based decision-making as well as call to attention the skills needed by our workforce. Operator data can be used to personalize a customer's experience as well as drive business decisions more effectively. The paper provides definitions and illustrations of practical real-world applications of supervised and unsupervised AI/ML big data, along with how big data applies to broadband and telecommunication networks.

The use of complex data sets or big data requires advanced workforce skills, innovation in operator tools, and updated techniques to process and analyze the information. Managing big data involves the use of data pipelines to ingest, organize, and store large amounts of data from various sources. Using SQL for data is an option today, looking beyond at unstructured big data databases and their benefits to our industry and workforce. Summarizing and visualizing broadband and telecommunication data helps make it more accessible and understandable for decision-makers, as well as our workforce. Proficiency in data science, data transformation with Alteryx, visualization tools like Tableau, and programming languages such as Python and Scala are crucial for the effective handling of big data. Additionally, a working knowledge of architectures like Hadoop or Microsoft Azure is essential for managing unstructured data.

Finally, for immediate action and takeaway, this session will share a view of SCTE's new programs for Python and data analytics, and how attendees can start the occupation redesign process while earning skill-based badges and professional credentials.

2. Understanding Big Data and its Relevance to Operators

Data is created every day in broadband telecommunications, through daily activities servicing our networks or our customer premises. All this data represents big data in the context of broadband and telecommunications. The advancement of data sharing between systems has led to data proliferation across an operator's ecosystem. Here are a few examples of various data sources that are relevant in the context of improving the workforce:

1. Customer care detail records are information about call center calls: such as call duration, time of call, origination of the call, location data, reason for the call, and more.
2. L&D KPIs include courses/journey completion, enrollment, and engagement rate, as well as time spent, assessment performance/patterns, retention rate, dropout rate, level 1 evaluation, skills acquisition, course ratings, and certification achievement.
3. Network device and sensor logs include cable modems, gateways, Wi-Fi extenders, routers, switches, optical nodes, headend equipment, and cell towers providing insights into network performance, traffic patterns, and potential issues caused by craft or failures.
4. Device data encompasses CPE, such as mobile devices, TV displays, and Internet of Things (IoT) devices, along with details about their functionalities and software revisions.
5. Customer data includes all the information about our subscribers, in the case of Comcast, and insights into how customer products are used and connected.
6. Operator social media and online interactions include qualitative data about our customer sentiment and behavior.
7. Quantitative installation data metrics such as discrete service calls on installs, trouble calls, etc.

Operators who effectively gather and correlate big data can make informed decisions about workforce development opportunities. Operators can gather data that has been observed, data that has been volunteered, and data that can be inferred. Adding a layer of visualization, operators can present complex workforce data patterns and trends.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The significance of data collection in the telecommunications industry has been growing rapidly in recent years due to technological advancements, changing consumer behavior, and the increasing demand for better services and experiences. There are examples of Comcast and Vodafone developing their own analytics and big data programs. Big data becomes more significant as our customers are looking for personalized services, personalized customer QoE with our workforce, tailored recommendations, and targeted marketing campaigns. Operators may use big data with data analytics to build an employee profile or persona to align their workforce personnel to particular customer requests. In addition, the network data analysis allows operators to optimize network performance, predict potential issues (e.g., DOCSIS PNM), and ensure seamless connectivity for their customers. Predicted maintenance may be conducted by our workforce from data analysis of various network and customer components, ensuring seamless connectivity. Our workforce can be on top of QoS issues affecting Internet services such as traffic latency, jitter, and delay. A workforce that is rooted in big data provides a significant advantage over its competition. Operators can be more agile in their understanding of market trends, emerging technologies, consumer preferences, and workforce resources.

With the sheer volume and velocity of big data, operators require a place to store the variety of collected data. The cloud computing model plays a crucial role in handling big data due to its scalability, accessibility, cost-effectiveness, and flexibility. Data lake cloud storage addresses many of the operator challenges posed by big data such as scalability, cost-effectiveness, accessibility, flexibility, redundancy, availability, analysis, elastic computing, parallel processing, disaster recovery, diverse data sets, and real-time streaming. In essence, cloud storage empowers operators to handle the challenges of big data. Cloud allows operators to focus on deriving insights from their data without the complexities of managing an on-premises infrastructure like a headend or data center.

3. Fundamentals of Data Science and Engineering for Operators

Several types of data analytics are important to workforce and network efficiency. The first type is known as "descriptive analytics" describing what has happened over a given period. Has the number of truck rolls gone up over the month? Are service calls more this month than last? The second type is known as "diagnostic analytics" which focuses more on why something happened. This involves more diverse data inputs and a bit of hypothesizing. Did the network affect service installs? Did that latest outage impact customer churn? The third type is known as "predictive analytics" and looks at what is likely going to happen in the near term. What happened to service calls the last time we switched Wi-Fi versions at the premises? How many cable modem models predicted network issues? Finally, the "prescriptive analytics" suggests a course of action. If the likelihood of a hot summer week is measured as two days over 100 degrees F, we should add additional technicians to the on-call list as the temperature changes network attenuation [1].

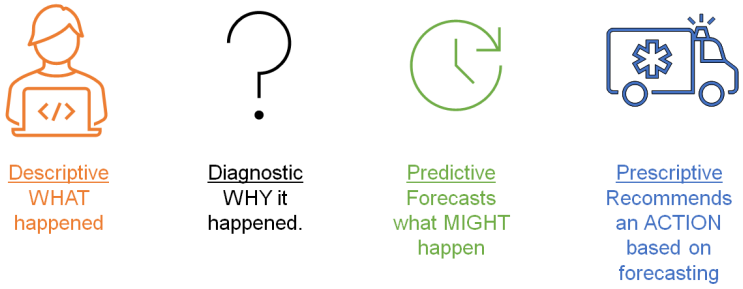


Figure 7 - Types of Data Analytics

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

As mentioned, statistics, data science, AI, ML, and automation are all key pillars for operators to make data-driven decision making. These pillars collectively enable organizations to analyze big data, extract meaningful insights, and make informed business decisions that drive results in growth and innovation. Let's explore how each of these elements contributes to the process.

Statistics involves the ingestion, analysis, transformation, organization, and presentation of big data. Statistics provides techniques for summarizing and understanding data patterns, relationships, and trends. For data-driven decision-making, statistics help in descriptive statistics, summarizing and describing data through measures like mean, median, and standard deviation. Inferential statistics serve as a valuable tool for deriving conclusions and making predictions about populations, such as the workforce, by analyzing sample data. Hypothesis testing involves evaluating the accuracy of hypotheses and making well-informed data-driven decisions. For example, is a skill like optical power budget measurements in fiber networks missing in the workforce? Is this missing skill increasing service calls on installs?

AI encompasses the development of intelligent systems that simulate human intelligence and decision-making processes. AI technologies contribute to data-driven decision-making by natural language processing (NLP) to extract insights from unstructured text data, enabling sentiment analysis and language understanding. For example, the Vodafone Remedy ticket system employs an AI decision engine powered by NLP. This engine predicts the resolution of raised incidents and initiates existing automation to address and close them, all without requiring human intervention [15]. These AI expert systems automate decision-making processes by encoding domain expertise into software. Another example is DOCSIS PNM, which can automate the decision-making process at the cable modem termination system (CMTS), allowing a CMTS to compensate for poor network conditions by sending data to a cable modem to adapt to conditions. Furthermore, AI expert systems are used to provide real-time workforce information to improve/re-enforce skills or deliver a more personalized customer QoE. A future area of AI algorithms will be cognitive computing or the notion of emulating human thinking and decision-making.

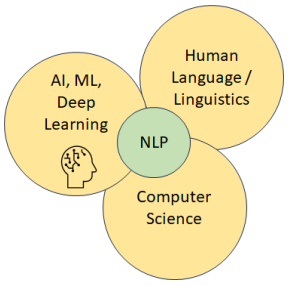


Figure 8 - Natural Language Processing

ML involves the development of algorithms and models that enable computers to learn patterns from data and make predictions or decisions. In the case of DOCSIS PNM, operators can teach the software to recognize network events that may impact the customer QoE. ML techniques enhance operator data-driven decisions by allowing for predictive analytics to forecast future outcomes based on historical data patterns and trends. ML provides classification, categorizing data into predefined classes or categories. Finally, ML can cluster or correlate, identifying groups within data to reveal patterns.

Automation involves the use of technology to perform tasks or processes with minimal human intervention. Operator Internet service self-installs are a good example of where truck rolls and

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

technicians are typically not required. In the data-driven decision-making model, automation contributes by providing data preprocessing or automatically cleaning, transforming, and preparing data for analysis. Automation provides mechanized reporting and visualization to communicate insights. Automation can assist decision-makers by providing relevant data and insights in real time.

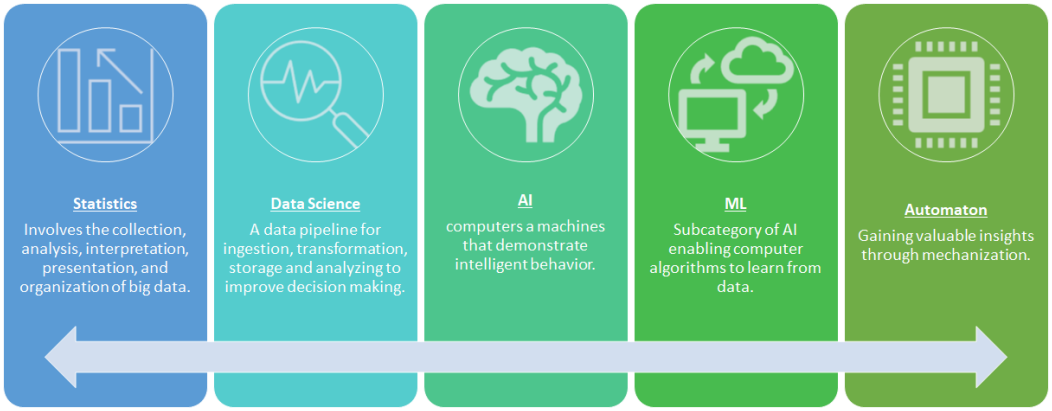


Figure 9 - Differences between Statistics, Data Science, AI, ML, and Automation

There are numerous advantages to utilizing statistics, data science, AI, ML, and automation in the context of data-driven decision-making with big data. Their application will improve an operator's efficiency, and accuracy, reveal insights, expose correlations, be scalable with cloud-based solutions, drive innovation into future solutions, and offer a competitive advantage. In summary, integrating statistics, data science, AI, ML, and automation into data-driven decision-making processes empowers organizations to harness the full potential of their big data, make strategic choices, and achieve their goals in a rapidly evolving business.

3.1. Personalization of the Customer QoE and Business Decisions through Data Utilization

Personalization of customer experiences and business decisions through data utilization is a powerful strategy that leverages big data to create tailored interactions, services, products, and strategies for individual customers. Personalization can also be used with an operator's workforce. For customers, social media is one place that provides valuable customer data that can be harnessed to create a personalized experience. Such an approach enhances customer satisfaction, loyalty, and engagement while also driving more effective and efficient business outcomes [8].

Big data utilization enables personalization in both customer QoE and business decisions in customer profiling by collecting and analyzing customer data. Operators can create detailed personas to understand unique needs and interests that include preferences, behaviors, demographics, and past operator interactions. Furthermore, big data drives segmentation, dividing customers into groups based on shared characteristics. This allows operators to target marketing efforts, product recommendations and offers more effectively. Operators monitor customer behavior, or perform behavior tracking, at various touchpoints (e.g., website, mobile app, set-top box, social media, chatbot) helping in understanding the customer journey, preferences, and pain points. Future operator interactions can then be tailored to meet the customer's specific needs.

Utilizing customer data, operators can deliver personalized marketing messages, recommendations, and promotions that resonate with each individual, leading to higher engagement and conversion rates.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Websites and apps can use data to dynamically display content, such as product recommendations or relevant articles, based on the user's preferences and browsing history. Analyzing big data allows operators to gain insights into market trends, customer preferences, and operational performance. These insights guide strategic decisions, product development, optimized pricing, and resource allocation (e.g., workforce).

By employing predictive models, operators can forecast future trends and customer behaviors. This informs decisions related to inventory management, optimizing supply chain, sales forecasting, enhanced risk assessment, workforce re-skilling, and resource planning. Operators can use big data-driven dashboards and analytics platforms to perform real-time monitoring of KPIs, enabling swift and informed decisions.

The benefits of personalization through data utilization not only enhance customer engagement but allow for improved conversion rates with personalized marketing, increased revenue, cost efficiency with optimized decisions, competitive advantages, and a more innovative organization.

In conclusion, personalization of customer QoE and operator decisions through big data utilization and data analytics is a transformative strategy that fosters stronger customer relationships, better decision-making, and sustainable business growth.

4. Applications of AI/ML in Telecom Networks

4.1. Real-World Examples of Supervised AI/ML Applications

Supervised statistical, data science, AI, ML, and automation applications involve training a model on labeled data (input-output pairs) to make predictions or decisions based on new, unseen data. There are several real-world examples of supervised AI/ML applications:

1. An LMS like SCTE's Absorb supports supervised AI/ML that enables personalized experiences, collaboration through social learning, badging, and content recommendations like Netflix through predictive AI and NLP.
2. Operator email spam filtering uses supervised learning to categorize incoming emails as spam or pass them as a good emails. The spam model is trained on a labeled dataset of emails, with features like words and/or phrases, and their corresponding labels (spam or not spam). The model then predicts the likelihood of new emails being spam based on these features.
3. Comcast Xfinity deploys a supervised cybersecurity system for their Internet gateway devices providing a continuous AI/ML learning system that employs its cumulative knowledge to recognize and predict cyber-attacks.
4. Wi-Fi uses adaptive Wi-Fi services and optimization using an application learning model. In the case of Plume, the model provides application prioritization or fine-tunes application performance. Software automatically detects and optimizes the performance of each application traffic flow, improving latency, and ensuring an optimal customer experience [11].
5. Amazon deploys an image classification and recognition model. The model can identify objects in photos, using supervised learning. The model uses AI/ML to learn from a dataset of images labeled with object categories, enabling Amazon to classify new images into those categories.
6. Speech recognition is offered by voice assistants like Siri, Amazon, and Google Assistants using supervised learning to understand and interpret spoken languages. The NLP model is trained on spoken words or phrases and their corresponding text transcriptions. Another example is the Xfinity voice remote, an NLP speech recognition model.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

7. Language translation services employ supervised learning to translate text from one language to another. The model learns from parallel corpora, where sentences in one language are paired with their translations in another language. SCTE's Absorb LMS allows translation to over 30 different languages.
8. Broadband and telecommunications organizations can use supervised learning to predict churn by training ML models in customer or employee behavior.

In conclusion, the application of supervised AI/ML models across broadband and telecommunications is evident in the examples mentioned. From personalized learning experiences in education through SCTE's VirtuLearn™ platform to efficient email filtering systems utilizing predictive algorithms, the power of supervised learning is harnessed to enhance user experiences and streamline processes. In the realm of cybersecurity, organizations like Comcast Xfinity leverage continuous learning systems to fortify their defenses against cyber-attacks. The optimization of Wi-Fi services with adaptive learning models, exemplified by Plume, showcases the adaptability and performance improvements achieved through this approach. Moreover, the wide-ranging applications extend to image recognition, speech interpretation, language translation, and even predictive analysis in telecommunications. The proficiency of supervised learning, drawing insights from labeled datasets, not only facilitates accurate predictions and classifications but also underscores the transformative potential of AI/ML technologies in shaping our modern-day functionalities. As our industry continues to integrate these models, the landscape of possibilities expands, promising further advancements in efficiency, accuracy, and user satisfaction.

4.2. Practical Illustrations of Unsupervised AI/ML in the Telecom Industry

Unsupervised AI/ML techniques are valuable in the broadband telecommunication industry for uncovering patterns, identifying anomalies, and gaining insights from large and complex datasets. There are several real-world examples of unsupervised AI/ML applications:

1. Network anomaly detection by unsupervised learning can help identify abnormal patterns in network traffic, indicating potential security breaches, outside plant issues, irregular CPE operation, or technical issues in the facility. Analyzing our network behavior, a system can detect unusual activities and trigger alerts for further investigation. A good example is DOCSIS proactive network maintenance (PNM) or an optical vendor that uses an optical time domain reflectometer (OTDR) test across a fiber network to detect patterns.
2. Our industry can segment its customer base using unsupervised clustering algorithms. By analyzing customer data such as usage patterns, location, and preferences, the company can identify distinct customer groups, allowing for more targeted marketing and service offerings.
3. Analyzing workforce development in various ways allows an organization to extract valuable insights from employee data, identify patterns, and enhance strategic decision-making. A few examples of unsupervised workforce clustering algorithms include talent segmentation, skill gap analysis, personalized career learning journeys, engagement analysis, ID emerging leaders, team formation, etc.
4. Unsupervised learning can analyze large amounts of network data to optimize resource allocation, improve coverage, and enhance network performance. It can identify areas with high demand and potential congestion, leading to more efficient network planning.
5. Prediction of customer or employee churn by analyzing historical usage data. Unsupervised learning can identify patterns and correlations in behavior, indicating potential churn, and allowing organizations to take proactive measures to retain customers and employees.
6. Optimizing customer support by analyzing customer interactions and feedback, unsupervised learning can identify common issues and trends. This insight can be used to improve self-

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

install options, and technician interactions during install/repair, refine support processes, and enhance the overall customer experience.

7. Unsupervised NLP techniques can analyze customer call recordings and text interactions to identify sentiment, common concerns, and emerging issues. This information can guide service improvements and strategic decisions.
8. Network traffic analysis can be performed by unsupervised learning helping analyze network traffic patterns to optimize data routing, reduce latency/jitter/packet loss, and ensure efficient Internet data transmission.
9. Unsupervised learning can analyze data on resource usage such as bandwidth and data consumption to optimize resource allocation and pricing plans, ensuring fair distribution and profitability.
10. QoS enhancements by monitoring and analyzing QoS metrics to identify areas of improvement in network performance and service delivery.

In conclusion, the application of unsupervised learning methods across broadband and telecommunications has showcased their transformative potential in uncovering valuable insights, enhancing decision-making, and optimizing various aspects of operations. Whether it's network anomaly detection for technical issue identification, customer segmentation for targeted marketing, or the intricate analysis of workforce development, unsupervised algorithms offer a powerful tool for uncovering hidden patterns and trends within complex datasets. These techniques contribute to efficient resource allocation, personalized services, and improved customer experiences. Furthermore, the ability of unsupervised learning to discern patterns in historical data extends to predicting churn, optimizing customer support, and refining network performance. Leveraging unsupervised natural language processing aids in sentiment analysis and issue identification from customer interactions, while network traffic and quality of service improvements bolster data transmission efficiency. Ultimately, the versatility of unsupervised learning serves as a catalyst for innovation and optimization across a spectrum of industries, underscoring its role in shaping more data-driven, efficient, and customer-centric landscapes.

4.3. Applying Big Data Analytics to Enhance Telecom Networks

Applying big data analytics to enhance our broadband and telecommunication networks holds the potential to usher in a new era of improved network performance, elevated customer experiences, enhanced workforce benefits, and streamlined operations. This innovative approach can be harnessed across various facets of the telecom industry.

Big data analytics offers a range of benefits in network performance optimization. By scrutinizing massive volumes of network traffic data, it identifies congestion points and bottlenecks, enabling targeted optimization efforts. Furthermore, predictive maintenance (e.g., DOCSIS PNM) based on equipment and sensor data foresees potential hardware failures, allowing for proactive intervention. Real-time monitoring of QoS metrics facilitates swift issue detection and resolution, preserving user satisfaction.

Understanding customer and employee behavior through data analysis facilitates personalized services and strategic customer marketing. Historical data assessment and predictive analytics empower companies to anticipate and mitigate customer/employee churn, enhancing retention efforts. Personalized service packages and learning plans aligning with individual preferences contribute to enriched interactions.

All operators are looking to expand and evolve their networks for the next technology, whether it is a DOCSIS, wireless, or fiber roadmap. Anticipating future demand, optimizing network resources, and re-skilling our workforce are central to expansion strategies. Big data analytics offers demand forecasting

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

insights to support capacity planning, labor, and infrastructure growth. Geographic analysis aids in identifying optimal locations for new networks, ensuring efficient coverage in a footprint. Furthermore, optimizing resource allocation, including bandwidth and spectrum, improves efficiency while minimizing waste. Additionally, analyzing energy consumption data supports eco-friendly practices in network operations, infrastructure, and future footprint expansions.

Big data analytics bolsters operator cybersecurity efforts like Rogers' Ignite gateway by detecting anomalous patterns in network traffic indicative of fraudulent activities or security breaches [13]. Timely identification and response mitigate potential risks, safeguarding network integrity. Customers also value such programs and may choose operators because of their cybersecurity value. Couple efforts in cybersecurity with real-time dashboards and visualizations for network health, performance, and usage to empower rapid decision-making by operators. Predictive insights derived from trend and pattern analysis offer early warnings and guide proactive measures.

With the proliferation of IoT and mobile connectivity in our industry, leveraging big data analytics for data processing allows insights into connected device behaviors and facilitates optimization for diverse applications. The management of such devices' intricate connectivity requirements ensures seamless communication and efficient data handling.

Incorporating big data analytics into broadband and telecommunication networks and their workforce not only empowers companies to make informed decisions but also amplifies network efficiency and elevates customer satisfaction. This data-driven approach ushers in proactive responses to challenges, fosters innovation, and cements telecom providers at the forefront of technological advancement.

5. Managing Big Data in Telecom

Data pipelines play a critical role in collecting, organizing, and storing data in the telecom industry. They facilitate the efficient and automated flow of data from various sources to storage systems like a data warehouse, enabling telecommunication organizations to process and analyze data for insights and decision-making. Here's how data pipelines are used in the telecom sector:

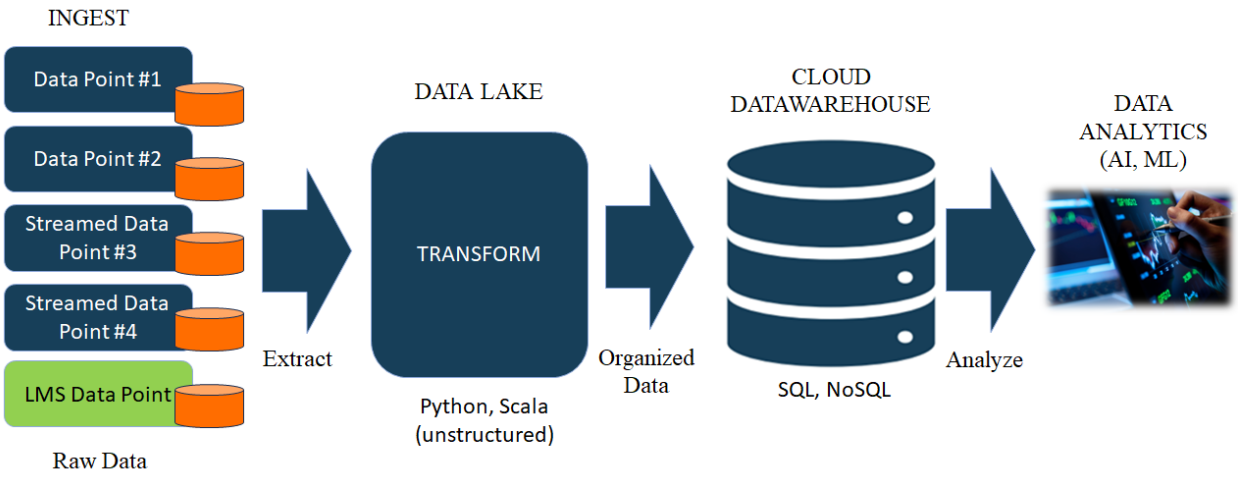


Figure 10 - Data Pipeline

Data pipelines ingest real-time data from network and system equipment, such as routers, gateways, modems, servers, LMSs, call centers, cell towers, etc. This data includes KPIs, traffic patterns, and device

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

status. Data pipelines capture customer interaction data, including call logs, text messages, and app usage. This information helps analyze customer behavior and preferences. Data pipelines can be used to ingest workforce data, including learning asset completion, credentials tracking, learner preferences, predicting learner success/churn, personalization KPIs, optimized learning journeys, etc. [11]. Additional data includes services supporting IoT devices, collecting data generated by sensors and connected devices, and enabling monitoring and control of operator IoT deployments. Ingest billing and financial data to track revenue, expenses, and customer billing information.

After the ingestion phase, data is transformed and organized. The data pipelines clean and transform raw ingested data into usable formats, removing duplicates, correcting errors, and enriching data with additional context. Data from various sources is aggregated to provide a unified view of performance, interactions, and other relevant metrics. Data pipelines ensure that data is consistent and standardized, enabling accurate comparisons and analysis.

The next phase of the data pipeline is storing the transformed and organized data in the data warehouse. The warehouse provides a centralized repository for historical and real-time data. Broadband and telecommunication companies may also use data lakes to store raw and unstructured data, enabling flexible and scalable storage for future analysis. Frequently, storage entails utilizing cloud services, offering a flexible solution for managing vast amounts of data in the order of PBs or EBs. Cloud-based storage solutions offer scalability and cost-effectiveness, allowing telecom companies to store large volumes of data without the need for extensive on-premises infrastructure. Several cloud models exist, allowing flexibility with analysis, public, private, hybrid (public/private), and community.

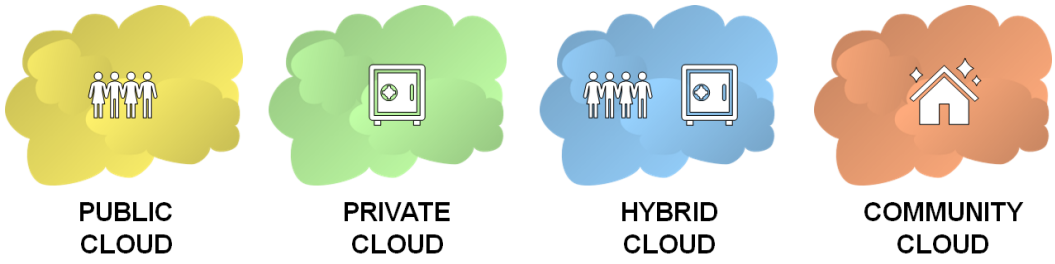


Figure 11 - Cloud Models

The final phase is the processing of the data, as well as data analysis. Processed data is made available for analysis using analytics platforms and tools. This enables operators to gain insights into performance, behavior, and other relevant metrics. Some data pipelines support real-time processing, allowing for immediate analysis and response to events as they occur. For example, operators could be notified when a technician makes a mistake in the network caused by human error factor (HEF). This real-time data event can be analyzed and correlated with network sensors and a technician's learning journey.

Data pipelines are monitored for data quality, performance, and potential issues. Monitoring ensures that data flows smoothly and accurately. Pipelines include error handling mechanisms to address issues such as data source failures or data format inconsistencies. Data pipelines are designed to adhere to regulatory requirements and data governance standards, ensuring data privacy and security. Data is often encrypted during transmission and storage to protect sensitive or private information.

There are a few solutions that operators can leverage in the marketplace. The Hadoop family is an open-source framework used for processing large data sets, as well as updated methods for data ingestion, transformation, storage, and analysis. Hadoop offers a fault-tolerant and parallel processing solution that will be important for our success in the future. Another popular solution is Microsoft Azure is a

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

comprehensive cloud platform that includes a vast range of services, including those for big data processing. this sample conclusion text with your text. Rogers will leverage Azure's enhanced data and AI capabilities across its business to enable use cases and operational improvements through Microsoft's new Data Centre of Excellence in Toronto [13].

In summary, data pipelines are a fundamental component of telecom operations, enabling the seamless collection, organization, and storage of data. They support data-driven decision-making, enhance customer experiences, and drive operational efficiency in the dynamic and data-intensive telecom industry.

5.1. Beyond SQL: Benefits of Unstructured Big Data Databases

Unstructured big data databases offer a range of significant benefits to our industry in the realm of data management and analysis. Unlike traditional structured databases, which rely on rigid schemas (SQL), these databases excel at handling vast amounts of diverse and unstructured data, such as text, images, videos, and social media posts. This adaptability allows broadband and telecommunication organizations to capture and store data in its raw, original form, eliminating the need for predefined structures and enabling flexibility in data ingestion. Unstructured big data databases empower businesses to gain deeper insights by processing and analyzing this heterogeneous data, uncovering hidden patterns and trends that might otherwise be overlooked. Additionally, these databases are well-suited for applications in AL, NLP, and ML, as they provide the foundation for training advanced models on diverse data types. Overall, unstructured big data databases empower enterprises to harness the full potential of their data, fostering innovation, informed decision-making, and enhanced customer experiences. Several examples of big data databases are MongoDB, Apache Cassandra, Amazon simple storage service (S3), Amazon DynamoDB, and Hadoop HDFS. Shaw, now Rogers, chose to use DynamoDB to manage customer data in JSON format [16], while Comcast and Cox are using Hadoop HDFC [5, 14].

5.2. Summarizing and Visualizing Telecom Data for Better Decision-Making

Summarizing and visualizing broadband and telecommunication big data is a powerful approach that enables better decision-making by providing insights, trends, and patterns clearly and intuitively. Summarization and visualization can enhance decision-making in our industry. As shared Cox centralized all their analytics teams to drive best decision making [14]. Summarization includes customer and workforce KPIs, usage patterns, segmentation (demographics, behavior, and preferences), network KPIs, and churn analysis. Visualization includes interactive dashboards, geospatial maps, trend analysis, heatmaps, network topologies (OSP/ISP), customer/employee flowchart journeys, predictive analysis, and social media sentiment.

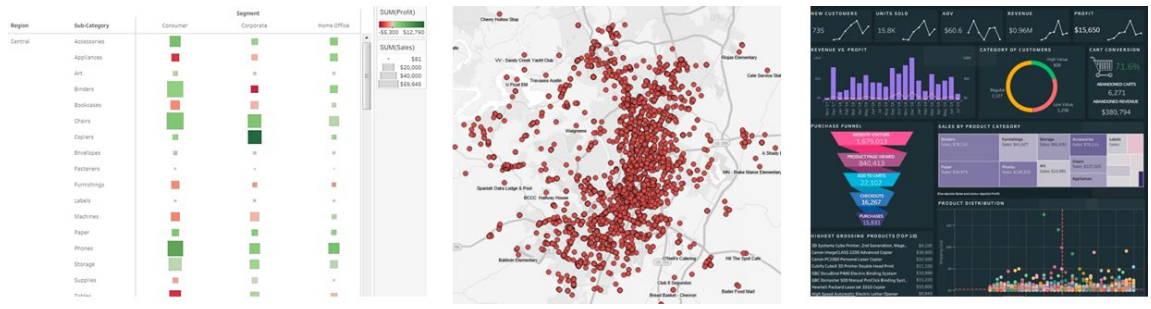


Figure 12 - Visualization Examples: Heatmap, Geospatial and Dashboards

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

There are a host of benefits of summarization and visualization. One of the big benefits is allowing complex data to be more accessible, enabling decision-makers to quickly grasp the key insights without delving into raw data. Visual representations help identify recurring patterns, correlations, trends, and anomalies that may not be evident in tabular or raw data. Interactive dashboards and visualizations provide real-time information, enabling timely and informed decision-making. For example, summarized data and visualizations facilitate effective communication between technical operation folks and non-technical L&D stakeholders by presenting information in an easily digestible format. Insights gained from summarization and visualization support strategic planning, resource allocation, workforce evolution, and investment decisions for network upgrades and expansion.

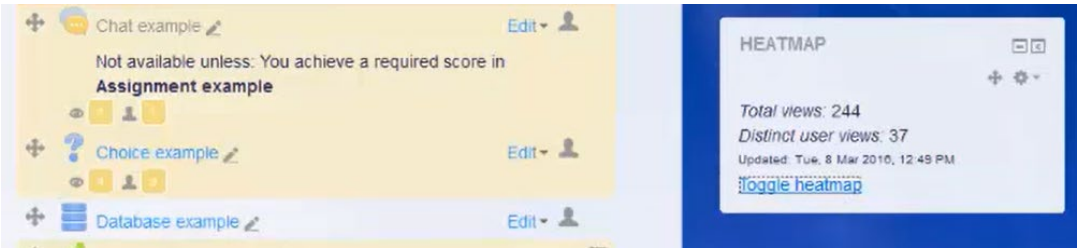


Figure 13 - Moodle LMS Visualization Heatmap, Identification of Popular Learning Assets

Incorporating summarization and visualization techniques into broadband and telecommunication data analysis empowers decision-makers to make data-driven choices, optimize network performance, enhance customer experiences, and drive overall business growth.

6. Key Workforce Skills for Handling Big Data

As we transform our industry into the digital age, our workforce must evolve in the same direction as the business. Proficiency in data science is highly relevant to our industry, as it empowers us to extract valuable insights from the vast amount of data generated by our networks, services, workforce, educational programs, and customer interactions.

Proficiency in data science has many benefits for broadband and telecommunication operators such as network optimization, capacity planning, proactive maintenance, customer/employee insights, customer/employee churn prediction, segmentation, service/product improvement, QoS/QoE, personalization, fraud detection, cybersecurity, anomaly detection, optimizing resources, energy management, resource allocation, footprint expansion, real-time insights, data governance, CPE management, strategic planning, market insights, and data-driven innovation.

The workforce's skill set must align with the entire data pipeline ecosystem. Nurturing a data science mindset is essential, beginning with a comprehensive grasp of data science operations. It is recommended to have a workforce knowledge of data storage frameworks like SQL and NoSQL. Our workforce must be familiar with big data tools like Hadoop with Apache Spark, Azure, and Kafka. Alteryx for the structural data transformations and scripting and programming skills like Python and Scala. As well as cloud-based AWS, Azure, and Google Cloud [14]. MongoDB, Apache Cassandra, Amazon S3, Amazon DynamoDB, and Hadoop HDFS big data databases are growing workforce skills. Soft skills are also required program solving, collaboration, project management, adaptability to learn new tools, and troubleshooting.

In conclusion, proficiency in data science is crucial for telecom operators to effectively manage their networks, improve customer experiences, optimize resources, enhance security, and drive business

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

growth. By leveraging data science techniques, operators can stay competitive, provide high-quality services, and adapt to the evolving demands of the telecommunications industry.

7. SCTE's New Programs for Data Analytics and Python

Through the SCTE education partnership with Cisco Network Academy, we can offer three new courses related to big data and data analytics.

SCTE provides an innovative online course titled "Introduction to Data Science", offering a comprehensive overview of how data analytics adds value to organizations, businesses, governments, and industries. Learners understand how data is being utilized, collected, and visualized, as well as the role of AI. The course helps learners understand the roles and responsibilities of a data analyst and how data scientists and engineers play a role in analytics.

Besides the "Introduction to Data Science", SCTE offers a “Data Analytics Essentials” course teaching learners the fundamental tools of a data analyst. The program was designed for learners to understand how data is collected, transformed, organized, and visualized with spreadsheet tools such as Excel. The program also allows learners to query data from a relational database using SQL and to improve data presentations using powerful business intelligence tools like Tableau. The course allows learners to create an analytics portfolio complete with an analysis of popular datasets. The portfolio will allow learners to showcase their skills in Excel, SQL, and Tableau.

Finally, SCTE offers Python, a general-purpose programming language adopted by the telecommunication industry and used to build just about anything. For this course, no prior programming knowledge is required! Python is key for backend web development, data transformation, data analysis, AI, and scientific computing, all of which are key for expanding a career in broadband and telecommunications. Learners will design, write, debug, and run programs encoded in the Python language using an integrated hands-on Sandbox. The course begins with the very basics guiding learners step by step until they become adept at solving more complex problems. By the end of this course, the learner will have a working knowledge of how computer programs are executed, improve critical thinking, and problem-solving skills; and translate real-world issues into computer-solvable problems.

All three of these courses all learners to earn badges and course certificates, as well as a certification as a Python-certified professional with the Python institute.

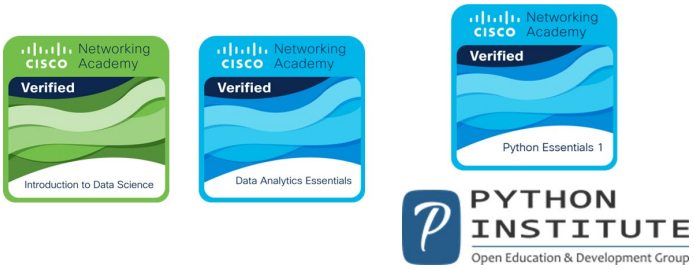


Figure 14 - SCTE Big Data Courses

8. Conclusion

In conclusion, this paper has provided a comprehensive exploration of the significance of big data in the broadband and telecommunications industry, highlighting the essential connection between advanced workforce skills and the successful utilization of this data-driven landscape. L&D teams can start to

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

explore data's value beyond the LMS, correlating across the business and with educational partners like SCTE. Throughout the paper, we delved into the fundamentals of big data, its relevance to broadband and telecommunications operators, and the pivotal role of data science and engineering in transforming customer experiences, employee experiences, and business strategies. The applications of AI/ML were examined through real-world examples, showcasing both supervised and unsupervised approaches.

Moreover, the paper underscored the importance of managing big data in the broadband and telecommunications sector, emphasizing the benefits of unstructured databases and data visualization techniques in making informed decisions. Recognizing that the potential of big data can only be harnessed by skilled professionals, the discussion of key workforce skills highlighted the need for expertise in handling and interpreting vast datasets.

Additionally, the insight into SCTE's new programs for data analytics and Python underscored the industry's proactive approach to equipping professionals with the necessary skills to navigate the evolving landscape of big data in broadband and telecommunications.

As the broadband and telecommunications industry continues to evolve with its digital transformations, this paper serves as a valuable resource for understanding the interconnectedness of data-driven strategies, workforce readiness, and technological advancements. By embracing the insights and strategies discussed here, stakeholders in the broadband and telecommunications sector can position themselves to thrive in an era where data is the cornerstone of innovation and progress.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AI	artificial intelligence
CMTS	cable modem termination system
CPE	customer premises equipment
DOCSIS	data over cable service interface specification
EB	exabyte
HDFS	Hadoop distributed file system
HEF	human error factor
IoT	Internet of things
ISP	inside plant
JSON	JavaScript object notation
KPI	key performance indicator
L&D	learning and development
LMS	learning management system
ML	machine learning
MS	Microsoft
NLP	natural language processing
NoSQL	not only structured query language
OSP	outside plant
OTDR	optical time domain reflectometer
PB	petabyte
PDF	portable document format
PNM	proactive network maintenance
PON	passive optical network
QoE	quality of experience
QoS	quality of service
ROI	return on investment
S3	simple storage service
SCTE	Society of Cable Telecommunications Engineers
SQL	structured query language
TB	terabyte
Wi-Fi	wireless fidelity
WWW	world wide web
XML	extensible markup language
ZB	zettabyte

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

1. Johnson, Jeff; Friend, Scott; Lee, Hannah (2017). "Big Data Facilitation, Utilization, and Monetization: Exploring the 3Vs in a New Product Development Process: BIG DATA VOLUME, VARIETY, AND VELOCITY." Journal of Product Innovation Management, 34. DOI: 10.1111/jpim.12397.
2. Schomm, F., Stahl, F., Vossen, G. (2013). Marketplaces for data: An initial survey. SIGMOD Record, 42(1), 15-26.
3. Monte Carolo. (2023). Data Engineering Trends & Predictions. Retrieved July 15, 2023, from, https://cdn.pathfactory.com/assets/10854/contents/468540/232de070-4c94-40b5-a1a3-34d885e8313a.pdf
4. KBV Research. (2019). Telecom Analytics Market Size, Share, Trends & Forecast 2019-2025. Retrieved from https://anywhere.epam.com/business/big-data-analytics-for-telecommunications
5. Informatica. (2023). CUSTOMER SUCCESS STORIES COMCAST. Retrieved August 1, 2023, from https://www.informatica.com/about-us/customers/customer-success-stories/comcast.html
6. Happiest Minds Technologies. (2014). Big Data: Why should enterprises adopt it? Retrieved August 15, 2023, from https://www.happiestminds.com/whitepapers/Big-Data-Why-should-enterprises-adopt-it.pdf
7. Hasan Bakhshi, H., Downing, J., Osborne, M., Schneider, P., The Future of Skills Employment 2030. (2017). Retrieved from https://ora.ox.ac.uk/objects/uuid:86577437-1353-4743-8520-401c1f99ad1b/download_file?file_format=application%2Fpdf&safe_filename=Future%2Bof%2Bskills%2BVoR.pdf&type_of_work=Report
8. Rane, Nitin & Achari, Anand & Choudhary, Saurabh. (2023). Enhancing customer loyalty through quality of service: Effective strategies to improve customer satisfaction, experience, relationship, and engagement. International Research Journal of Modernization in Engineering Technology and Science. 5. 427-452. 10.56726/IRJMETS38104.
9. Osman, M. (2023). Wild and Interesting Facebook Statistics and Facts. Retrieved from https://kinsta.com/blog/facebook-statistics/
10. Top Trends in Data and Analytics, 2022
11. Plume leverages AI to introduce 'Full Stack Optimization'--a set of game-changing network features to drive application performance to the next level. (2023). Retrieved from https://www.prnewswire.com/news-releases/plume-leverages-ai-to-introduce-full-stack-optimizationa-set-of-game-changing-network-features-to-drive-application-performance-to-the-next-level-301890384.html
12. Is Big Data Better? LMS Data and Predictive Analytic Performance in Postsecondary Education. https://www.edworkingpapers.com/sites/default/files/ai22-647.pdf
13. Rogers and Microsoft announced a strategic alliance to revolutionize hybrid workplace communications and power 5G innovation across Canada with Azure. Retrieved from https://news.microsoft.com/en-ca/2022/03/16/rogers-and-microsoft-announce-strategic-alliance-to-revolutionize-hybrid-workplace-communications-and-power-5g-innovation-across-canada-with-azure/
14. Kuruppath, R. (2023). Emerging Trends in Continuous Integration and Continuous Deployment.
15. Dr. Hou, J. (2023). Zero Touch Operations: How we're using AI to enable IT systems/process automation. https://careers.vodafone.com/life-at-vodafone/projects-stories/how-were-using-ai-to-enable-it-systems/
16. Bone, G., Agarwal, G. (2022). Translating Customer & Employee Experience with Shaw's Data Journey

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Upstream Triggered Spectrum Capture

The Feature That Makes Capturing Transient Noise A Reality

An Operational Practice prepared for SCTE by

Allen Maharaj
Manager – HFC Operations
Rogers Communications
8200 Dixie Rd, Brampton, Ontario, L6T 0C1
416-450-4820
allen.maharaj@rci.rogers.com

James Medlock
Founder & CEO
Akleza, INC
18695 Pony Express Dr, 2350, Parker, CO, 80134
303-670-7951
jmedlock@akleza.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. What is Upstream Triggered Spectrum Capture? ... 3
3. Why use Upstream Triggered Spectrum Capture? ... 4
4. Operational Theory and Requirements ... 5
5. Operationalization Options ... 6
   - 5.1. Implementation Table ... 6
   - 5.2. iCCAP Implementation ... 6
   - 5.3. Remote Phy Device UTSC Implementation ... 7
6. Triggers ... 9
7. Considerations and Limitations ... 9
   - 7.1. Considerations: ... 9
   - 7.2. Limitations: ... 10
8. Operational Process, Implementation and Use Cases ... 11
   - 8.1. Operational Process ... 11
   - 8.2. General Implementation requirements ... 11
   - 8.3. Implementing a Periodic Time-Based Trigger ... 12
   - 8.4. Implementing Quality of Service Triggers ... 12
     - 8.4.1. Quality of Service Parameter Triggers ... 12
     - 8.4.2. Implementation of Triggers: ... 13
   - 8.5. Implementing a User-Requested Trigger ... 13
   - 8.6. Implementing a Dynamic Trigger based on Utilization or Congestion ... 14
     - 8.6.1. Implementing a Dynamic Trigger based on Congestion ... 14
     - 8.6.2. Implementing a Dynamic trigger based on Utilization ... 14
     - 8.6.3. Congestion vs Utilization, which provides more value? ... 15
   - 8.7. Intermittent Impairments ... 16
     - 8.7.1. Impairment Types ... 16
     - 8.7.2. Implementing a Trigger for Intermittent Impairments ... 18
9. Future Scope ... 19
10. Conclusion ... 20

Abbreviations ... 21

Bibliography & References ... 22

List of Figures

Title **Page Number**

Figure 1 – Functions, Connections, and Protocols ... 5
Figure 2 – UTSC spectrum display (courtesy Akleza) ... 8
Figure 3- Operational Process Map ... 11

List of Tables

Title **Page Number**

Table 1 – UTSC Implementation Methods ... 6

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Upstream triggered spectrum capture (UTSC) is a proactive network maintenance (PNM) test defined in the DOCSIS 3.1 CCAP Operations Support System Interface specification (OSSI-CCAPv31. This PNM test is supported for both integrated CCAPs with local radio frequency (RF) ports as well as for distributed access architectures (DAA) with a shift of RF generation and reception in remote PHY (R-PHY) to the outside plant.

This functionality facilitates the capture of upstream spectrum data from the CCAP or remote PHY device (RPD) remotely and allows operators to move from using that information not only as a real time diagnostic tool, but to a more effective automated triggered capture and analysis function applied to transient events.

As game changing as this feature is for troubleshooting intermittent impairments, it is not without its deployment and operational challenges.

This paper will endeavor to explain the operational theory of UTSC, and include:

- Operationalization methods like simple network management protocol (SNMP), trivial file transfer protocol (TFTP), and layer two tunneling protocol (L2TP)
- Challenges that come along with each of those options
- Benefits of each option including factors like scalability and data refresh rates
- Caveats of each option including cost and implementation requirements like hardware and routing, as well as link utilization
- Considerations that need to be evaluated which include security.
- Use Cases

2. What is Upstream Triggered Spectrum Capture?

In cable broadband networks, data is transmitted in two directions: downstream, from the CCAP core to the end-user, and upstream, from the end-user back to the CCAP core. To ensure efficient use of available bandwidth and to monitor for interference, cable modems and CCAP's can use a mechanism called spectrum capture to identify a variety of impairment conditions.

In traditional cable systems, spectrum capture is done manually, either using a spectrum analyzer connected directly to a network element within the plant, or remotely using a dedicated hardware platform in the headend monitoring each RF link. Either way it is utilized, it requires eyes on glass when an impairment is present and is not the most efficient when dealing with dynamic network conditions or sudden bursts of data.

PNM's spectrum capture feature allows downstream spectrum to be captured by the cable modem and provides many more visibility points than are possible using field test equipment.

UTSC provides an alternative, software-based approach to monitoring the upstream spectrum. It allows for upstream spectrum data to be captured in an automated fashion either in real-time or based on a predetermined trigger. The trigger can be based on numerous factors, such as network congestion, Quality of Service (QoS) requirements, or simply a time-based schedule. When triggered, it captures the spectrum data for upstream channels of interest which can be displayed in real-time, analyzed via an algorithm, or stored for future analysis by a technician or engineer.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

By employing UTSC, cable broadband networks can adapt to changing conditions more efficiently, leading to better overall performance and reduced interference. This technique allows for more intelligent and dynamic management of the available spectrum, leading to a more reliable and robust internet experience for users.

3. Why use Upstream Triggered Spectrum Capture?

UTSC offers several significant benefits for cable and network operators. These benefits contribute to improved network performance, enhanced troubleshooting capabilities, and better customer satisfaction. Here are the key advantages that UTSC offers:

- **Proactive Issue Detection**: UTSC enables operators to proactively detect intermittent impairments and issues in upstream data transmission. By continuously monitoring QoS parameters and capturing spectrum data when thresholds are crossed, operators can identify problems before they escalate, allowing for timely intervention.
- **Faster Troubleshooting**: Capturing spectrum data during critical events provides valuable insights into the root causes of intermittent impairments. Network engineers can analyze the captured data to pinpoint the exact time and location of the issue, streamlining the troubleshooting process and reducing the Mean Time to Repair (MTTR).
- **Optimized Resource Allocation**: By identifying areas with intermittent impairments, operators can optimize their resource allocation and prioritize network upgrades or maintenance efforts where they are most needed. This targeted approach ensures that resources are used efficiently and effectively.
- **Improved Network Performance**: Timely detection and resolution of intermittent impairments leads to improved network performance and stability. This translates to better service quality for end-users, reduced service disruptions, and higher customer satisfaction.
- **Enhanced Customer Experience**: With fewer service disruptions and improved network performance, UTSC contributes to a positive customer experience. Customers are less likely to experience issues like slow speeds, packet loss, or high latency.
- **Proactive Network Maintenance**: Operators can use captured spectrum data to identify potential issues and trends in the network. This insight allows them to proactively address network maintenance requirements, reducing the likelihood of major service outages.
- **Quality Assurance and Benchmarking**: The captured spectrum data serves as a benchmark for network performance. operators can compare historical data to current performance, identifying improvements and ensuring that the network meets or exceeds quality standards.
- **Data-Driven Decision Making**: UTSC provides operators with valuable data for informed decision making. Data analysis helps prioritize investments, plan upgrades, and optimize network capacity.
- **Supports Remote Monitoring**: Remote monitoring of upstream channels enables operators to monitor network performance from a centralized location, reducing the need for physical on-site interventions.
- **Reduced Capital and Maintenance Costs**: UTSC offers a pure software-based return path monitoring, and does not require the expensive hardware utilized by traditional return path monitoring system.
- **Competitive Advantage**: Implementing UTSC sets an operator apart from competitors, demonstrating a commitment to network excellence and customer satisfaction.
- **Regulatory Compliance**: In some regions, regulatory authorities may require operators to demonstrate their ability to monitor and troubleshoot network issues. UTSC can help meet these compliance requirements.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

UTSC provides a proactive approach to network monitoring, troubleshooting, and maintenance. By capturing intermittent impairments and identifying their root causes, operators can optimize network performance, enhance customer satisfaction, and maintain a competitive edge in the market.

4. Operational Theory and Requirements

UTSC as defined in the PNM section of the DOCSIS 3.1 OSSI CCAP specification evolved out of proprietary upstream spectrum capture capabilities offered by several CCAP vendors. This functionality leverages the ability of the RF port burst receiver, either on an integrated line card in an iCCAP, R-PHY device (RPD), and Remote MACPHY (R-MACPHY/RMD) to capture the upstream spectrum. The RF data is sampled in the time domain with an analog-to-digital converter (ADC) and then converted to the frequency domain using a fast Fourier transform (FFT) to produce a digital spectrum sample. This data is then transmitted to, or collected by, a PNM server for display, storage, and analysis. Legacy CCAP spectrum capture used SNMP to both request a spectrum capture, and read the results.

In DOCSIS 3.1, a PNM bulk data retrieval mechanism was defined where the device under test (DUT) being either a cable modem or CCAP would transmit large data sets like spectrum capture more efficiently. The request for data, or test, is still configured and initiated using SNMP but the data is then pushed from the DUT rather than being read via SNMP. The specification requires that the DUT pushes the data using the trivial file transfer protocol (TFTP), copying the results to a remote destination, a PNM server. The file format for each of the PNM tests is defined in the DOCSIS 3.1 OSSI specifications.

With the introduction of DAA, an additional mechanism was defined for UTSC where the captured data is pushed directly from the RPD over an L2TP pseudowire. This bypasses the CCAP and delivers the data as a raw binary data stream.

To fully leverage the capabilities provided by UTSC, several back-office functions are required, including a PNM server to configure and receive the spectrum data via TFTP and L2TP pseudowires, a monitoring server to collect additional burst receiver statistics, an analysis server to process and analyze the received data, and various display, alarm, and reporting servers providing user interface functions.

Figure 1 shows the different required functions, their connections, and communication protocols.

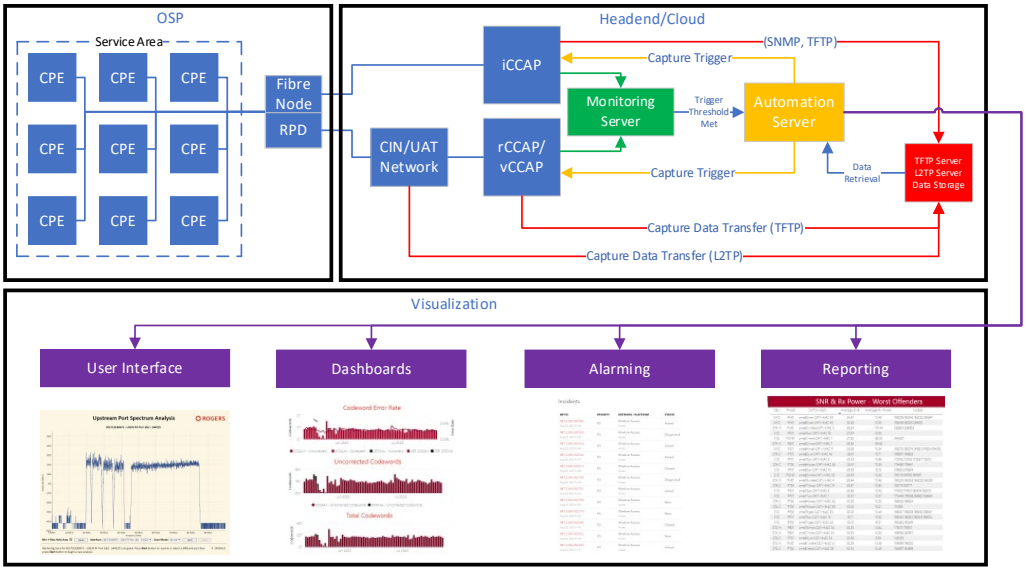


Figure 1 – Functions, Connections, and Protocols

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Operationalization Options

5.1. Implementation Table

Table 1 provides an at-a-glance overview of the implementation methods available for upstream spectrum capture. It includes both vendor proprietary and specification-defined options and some of the pertinent information required to make an informed decision.

Table 1 – UTSC Implementation Methods

Technology	Configuration Method	Capture Method	Pros	Cons	Data Streaming Rates	Display Refresh Rates	Spectrum Resolution	Comments
iCCAP	SNMP	SNMP	- Simple SNMP request-response implementation - Provides upstream spectrum visibility at lower cost - Supports polling during active or quiet probes	- Vendor proprietary implementations - Slow refresh rate - Unable to capture fast noise bursts - Resource heavy on CCAP limiting concurrent captures	< 50 kbps	2 s–10 s	50 kHz–300 kHz; typically, 150 kHz	- Vendor proprietary implementation (for more information, speak with the CCAP vendor). - Spectrum resolution is configurable based on vendor implementations. The finer the resolution, the more data that must be returned and, therefore, the slower the data refresh rate will be. - Note that the data streaming/refresh rate is dependent upon the SNMP response time of the CCAP.
iCCAP	SNMP	TFTP	- Standards-based PNM UTSC solution - Provides upstream spectrum visibility at a higher refresh rate - Requires less CCAP CPU resources while polling - May support polling during active or quiet probes	- Not all PNM UTSC features supported across all CCAP vendors - More complex implementation to support asynchronous data push and analysis	~ 1 Mbps	200 ms–1 s	50 kHz for sub-split and mid-split; 100 kHz for high-split and up	- A TFTP file is generated each second and may contain one or more captures, resulting in a file size of 7 kB to 70 kB.
rCCAP/ vCCAP	SNMP	SNMP	- Simple SNMP request-response implementation - Provides upstream spectrum visibility at lower cost - Supports polling during active or quiet probes	- Vendor proprietary implementation - Slow refresh rate - Unable to capture fast noise bursts - Resource heavy on CCAP limiting concurrent captures	< 50 kbps	2 s–10 s	50 kHz–300 kHz; typically, 150 kHz	- Vendor proprietary implementation (for more information, speak with the CCAP vendor). - Spectrum resolution is configurable based on vendor implementations. The finer the resolution, the more data that must be returned and, therefore, the slower the data refresh rate will be. - Note that the data streaming/refresh rate is dependent upon the SNMP response time of the CCAP.
rCCAP/ vCCAP	SNMP	TFTP	- Standards-based PNM UTSC solution - Provides upstream spectrum visibility at a higher refresh rate - Requires less CCAP CPU resources while polling - May support polling during active or quiet probes	- Not all PNM UTSC features supported across all CCAP vendors - More complex implementation to support asynchronous data push and analysis - Not supported by all CCAP vendors	~ 1 Mbps	200 ms–1 s	50 kHz for sub-split and mid-split; 100 kHz for high-split and up	- A TFTP file is generated each second and may contain one or more captures, resulting in a file size of 7 kB to 70 kB.
rCCAP/ vCCAP	SNMP	L2TP	- Standards-based PNM UTSC solution - Constant data stream being transmitted direct from the RPD - Fast capture granularity allows for detection of noticeably short noise bursts	- Configuration mechanism not specified, so vendor proprietary - Not all PNM UTSC features supported across all CCAP vendors - More complex implementation to support asynchronous data push and analysis - High data streaming rate may limit number of concurrent captures	80 Mbps–100 Mbps	100 ms–200 ms	50 kHz for sub-split and mid-split; 100 kHz for high-split and up	- Identified in CM-SP-R-PHY that it is to be 100 kHz or finer.

5.2. iCCAP Implementation

Upstream spectrum capture functionality is supported across several implementation methods and varies by CCAP vendor and technologies such as integrated line card ports versus DAA ports.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Vendor proprietary implementations using SNMP to configure and request spectrum capture data are available from several CCAP vendors. Operators and implementors should consult with the equipment manufacturer for specific implementation details. As this method uses SNMP to request and retrieve data, spectrum display refresh rates are constrained by the SNMP response time of the CCAP and network. Additionally, there may be limitations on the number of concurrent spectrum captures that can be supported either due to line card limitations and/or system processor card resources. A typical spectrum display using this method can produce spectrum data at 100 kHz resolution being updated every 1-10 seconds. Finer resolution may be possible, but this will have an impact on the refresh rates, as more data must be processed by the CCAP and returned to the PNM application.

The DOCSIS 3.1 specifications introduced UTSC to standardize the configuration interface for spectrum capture, as well as introduce additional capture modes based on triggered events. Rather than using SNMP to deliver results, TFTP is used to push data in a more concise format from the CCAP to the PNM application by most vendors. To date, UTSC functionality has been implemented by several CCAP vendors; however, not all capture modes are supported, and there are also differences in the configuration method based on CCAP vendor and software release. There are also some differences in the data from the CCAP to the PNM application, some containing multiple samples per file with new files being generated each second, whereas others produce just one sample per file, per second. Using this method, it is possible to generate spectrum displays with 50-kHz resolution updating every 100 ms to 200 ms if multiple captures per file are supported or updating every second if just a single sample is captured. This method also significantly reduces the load on the CCAP system processor and network, as the raw spectrum data is not being encapsulated in an SNMP payload.

Implementations of UTSC vary in cost and complexity, dependent on whether historical data is required. Data storage and the associated operational costs will then become a factor requiring consideration. As an example, in the TFTP implementation, the capture files transferred from the CCAP vary because of the difference in the number of capture samples they contain. Usage and visualization of the data in all three methods defined also incur higher development costs based on how the implementer decides to present the data and the number of users they wish to support.

5.3. Remote Phy Device UTSC Implementation

Section 15.3 of the CableLabs Remote PHY Specification, CM-SP-R-PHY-I17-220531, describes the operation of UTSC in an R-PHY environment and how the spectrum is captured by the RPD and delivered to the CCAP core or a PNM server via pseudowires. It also states that configuration and control is via the CCAP core and modeled after the DOCSIS 3.1 specifications. The Remote PHY OSS Interface Specification, CM-SP-R-OSSI-I19-220930, states that there are no additional PNM requirements to support UTSC.

Unlike the iCCAP TFTP transfer mechanism, the L2TP implementation produces a steady data stream, with much larger data streaming rates that will push significantly more data to the users. This higher stream rate also limits the number of concurrent streams, depending on how the network is built. Using this method, it is possible to generate spectrum displays with similar resolution to the iCCAP TFTP method described above and even faster refresh rates; though this introduces higher load on the network and PNM application as more data is transmitted and processed.

Though the DOCSIS 3.1 CCAP OSSI describes the configuration mechanism to allow UTSC data to be delivered to a PNM server via TFTP or future streaming protocols, it does not support the use case where spectrum capture data is delivered to a PNM server via a pseudowire directly from an RPD. As such, vendors have implemented differing proprietary configuration mechanisms to support this scenario.

For specific configuration steps and capabilities, implementers should consult with the specific CCAP core equipment vendors. As the configuration is via the CCAP core, which then configures the RPD

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

spectrum capture function via generic control plane (GCP), the functionality supported may also be impacted by this configuration interface and RPD capabilities, requiring further consult with specific RPD vendors. The current proprietary configuration methods include custom configuration of docsPnmCmtsUtscCfgTable parameters or use of a vendor specific MIBs. These configuration options are meant to allow the CCAP core to direct the RPD where to send its spectrum capture data and the destination end point for required pseudowires.

Once configured, however, and as described in the remote PHY specification, spectrum capture data is streamed to a PNM server over an L2TP pseudowire. Depending upon the configured frequency span and output format, the spectrum data may span one or more PSP segments and PSP packets, as described in the CableLabs Remote Upstream External PHY Interface Specification, CM-SP-R-UEPI-I13-201207.

It should be noted that, to support spectrum capture data delivered via an L2TP pseudowire, the network between the RPD and the PNM server and any firewalls therein must be configured to allow LT2P traffic, IP Protocol 115, to traverse. Additionally, the data streamed from the RPD is typically delivered at the capture and process rate of the RPD. In real-world testing, this may be in the order of 80+ Mbps per port for a mid-split configuration, so adequate network resources must be available to support these rates.

Figure 2 shows a screenshot of a spectrum capture from an RPD port at 50-kHz resolution and refreshing every 100 milliseconds using UTSC data streamed directly from the RPD over a pseudowire. Note the waterfall plot at the bottom, which also allows some discovery of intermittent issues.



Figure 2 – UTSC spectrum display (courtesy Akleza)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Triggers

Triggering the capture of upstream spectrum data in cable broadband networks can be achieved using various methods. UTSC defines several automatic triggers that can be implemented by the CCAP, and RPD/RMD to initiate spectrum capture. To date not all of these have been implemented universally across CCAP vendors. A spectrum capture request may also be triggered by an external application based on other conditions or monitored statistics. The selection of a specific triggering method depends on the network's requirements and the capabilities of the headend equipment. Here are some options for triggering the capture:

- **Periodic Time-Based Trigger**: Scheduled periodic triggers at specific intervals. For example, it might send a trigger command to all upstream ports in the network every few minutes, prompting them to capture spectrum data and report back. Note that CCAP resource limitations may require spectrum captures to be staggered across RF ports based on the line card, RPD, and CCAP limits on the number of concurrent spectrum captures.
- **Dynamic Trigger Based on Congestion or Utilization**: Network teams regularly monitor network congestion levels. If automated, congestion is detected exceeding a certain threshold, it can trigger spectrum capture for specific upstream ports that are potentially causing the congestion or experiencing it. They can automate analysis of channel utilization patterns and trigger the capture for upstream ports that are experiencing high utilization.
- **Trigger Based on Quality-of-Service Metrics**: QoS parameters, such as latency, packet loss, or jitter, can be monitored by the CCAP. If QoS metrics fall below acceptable levels, an automation can initiate the spectrum capture process to investigate potential causes.
- **User-Requested Trigger**: Technicians will require the ability to initiate the spectrum capture process themselves. This could be done through an application or user interface provided by operational support systems, allowing them to investigate and review the spectrum as changes are made in real time.

It is worth noting that the choice of triggering method and the frequency of capturing spectrum data should strike a balance between the need for real-time network adjustments and the overhead involved in the capture process. Additionally, the triggers should be implemented securely to prevent any potential abuse or unauthorized access to the spectrum data. The specific implementation may vary based on the equipment used and the specific requirements of the cable broadband network.

7. Considerations and Limitations

Implementing UTSC comes with various considerations and limitations that operators and network administrators should be aware of.

7.1. Considerations:

- **Hardware and Software Compatibility**: Ensure that the network devices support upstream spectrum capture and are compatible with the trigger commands. Upgrading or replacing legacy hardware may be necessary to implement the feature across the network.
- **Real-Time Monitoring Infrastructure**: Implementing UTSC requires a robust real-time monitoring infrastructure capable of continuously monitoring and analyzing various QoS parameters.
- **Threshold Setting**: Carefully determine appropriate thresholds for triggering spectrum capture. Setting them too low may lead to excessive captures, while setting them too high may cause major events to be missed.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- **Data Handling and Storage**: Spectrum capture generates significant data volumes, and operators need adequate storage and data management solutions to manage the captured data efficiently.
- **Data Privacy and Security**: Capturing spectrum data may involve sensitive information. Ensure that appropriate data privacy and security measures are in place to protect user data.
- **Data Analysis and Expertise**: Analyzing captured spectrum data requires expertise in interpreting the information and diagnosing network issues effectively.
- **Impact on Network Performance**: Triggering and capturing spectrum data can put additional load on the network, affecting customer traffic and service quality. Minimize the impact on network performance during captures.
- **Data link utilization**: Spectrum data captures utilize the same links which also support customer data and device control. Depending on the implementation, the additional overhead required can create an impact to customer services if oversubscribed.
- **Network Scale**: For large operator networks with numerous subscribers and network devices, implementing UTSC across the entire infrastructure can be a complex task that requires careful planning and coordination.
- **Trigger Accuracy**: Ensuring accurate and precise triggering of spectrum capture is crucial. False positives (capturing when no real issue exists) or false negatives (failing to capture during actual issues) can lead to inefficiencies in troubleshooting and diagnosing network problems.

7.2. Limitations:

- **Intermittent Issues**: Capturing intermittent problems can be challenging, as they may not occur frequently or predictably. Some issues may be missed if they do not align with the trigger criteria.
- **Capture Window**: The duration of spectrum capture may be limited, and capturing long-duration intermittent issues may require fine-tuning the capture duration.
- **Resource Constraints**: Implementing UTSC requires additional network resources for monitoring and data storage, which may be limited in some environments.
- **Device Support**: Not all network devices may support upstream spectrum capture, limiting the scope of implementation.
- **Costs**: Implementing and maintaining the necessary infrastructure for UTSC can involve significant costs, including hardware, software, and personnel training.
- **Complexity**: The implementation and management of UTSC can be complex, requiring careful planning and coordination.
- **Testing and Validation**: Thorough testing and validation are essential to ensure the system functions as intended and does not introduce novel issues.
- **Frequency of Capture**: Frequent triggering and capturing can generate a large amount of data, making it challenging to manage and analyze effectively.

It is crucial for operators to thoroughly evaluate these considerations and limitations before deciding to implement UTSC. Proper planning, testing, and optimization can help maximize the benefits of the feature while minimizing its impact on the network and ensuring a positive user experience.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

8. Operational Process, Implementation and Use Cases

8.1. Operational Process

Figure 3 shows the operational process required for effective use of UTSC

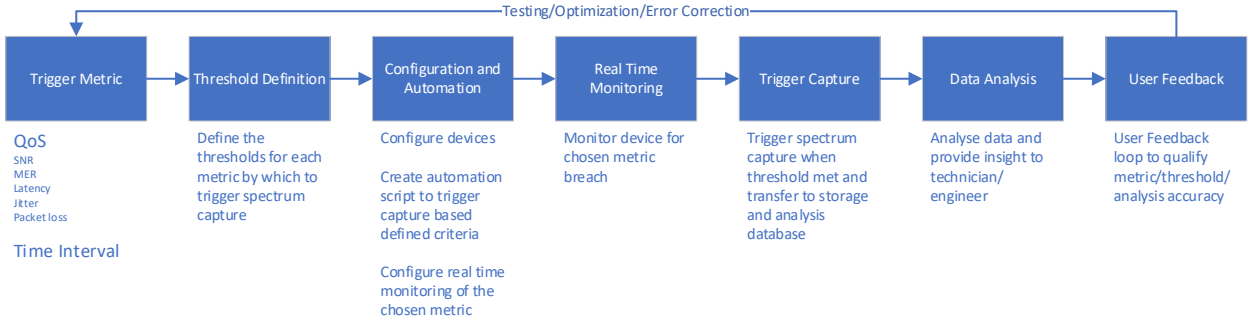


Figure 3- Operational Process Map

8.2. General Implementation requirements

The following list of requirements apply to all potential use cases for triggering and capture of upstream spectrum data. They are necessary to ensure that the captures are generated and transmitted accurately and efficiently and to mitigate any potential impact to the network and its customers.

I. Configuration: Automation will be necessary to conduct the necessary commands and capture the spectrum data. The necessary network infrastructure will be necessary to support this and must be appropriately configured with data routes, access permissions, and the necessary compute power and storage capabilities required.

II. CCAP Communication Protocol: Ensure that the CCAP can communicate using a standardized communication protocol that allows the transmission of the trigger command and spectrum data. In the cable network this include SNMP (simple network management protocol), TFTP (trivial file transfer protocol), or L2TP (layer two tunneling protocol).

III. CCAP Compatibility: Verify that all the CCAP's in the network support the periodic time-based trigger functionality. If any CCAP does not support it, you may need to consider alternative methods for triggering spectrum capture for those devices.

IV. Time Synchronization: To maintain accurate timing for triggers, make sure that all devices in the network, including the headend and cable modems, are time-synchronized using a reliable time source (e.g., NTP - network time protocol).

V. Security and Authorization: Ensure that the triggering mechanism is secure and properly authorized to prevent unauthorized access to the spectrum data or manipulation of the triggering process.

VI. Privacy Considerations: Ensure that the user-requested trigger process complies with data privacy regulations and that user data is managed securely.

VII. Error Handling: Implement appropriate error handling and logging mechanisms in the headend to deal with any issues that may arise during the trigger process.

VIII. Testing and Optimization: Thoroughly evaluate the periodic time-based trigger implementation in a controlled environment before deploying it in the live network. Fine-tune the trigger intervals and settings based on the network's performance and requirements.

IX. User Feedback: Consider providing feedback to the user regarding the success of the spectrum capture request or any potential issues detected during the process.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

8.3. Implementing a Periodic Time-Based Trigger

Implementing a periodic time-based trigger for upstream spectrum capture involves setting up a schedule that triggers the capture process at specific intervals. To implement a periodic time-based trigger:

I. Define the Trigger Interval: Determine the time interval at which you want to trigger the spectrum capture process. For example, one may decide to trigger the capture every 5 minutes, every hour, or any other desired frequency.
II. Create a Trigger Command: Design a specific trigger command or message that will be sent from the headend to the CCAP. This command will instruct the CCAP to initiate the spectrum capture process.
III. Implement Trigger Scheduler: Within the headend software, implement a trigger scheduler that will initiate the trigger command at the defined time intervals. This scheduler should send the trigger command to the CCAPs in the network.
IV. Data Analysis: Once the spectrum data is collected, the headend should process and analyze the captured information to assess the performance of different channels and identify potential issues.

Using a periodic time-based trigger, operator's can systematically and regularly capture spectrum data, enabling real-time monitoring and optimization to provide a better user experience and ensure efficient network performance.

8.4. Implementing Quality of Service Triggers

QoS parameters are essential for ensuring a certain level of performance and user experience in a network. When implementing triggers for upstream spectrum capture, QoS parameters can be used as a basis to determine when spectrum data should be captured.

8.4.1. Quality of Service Parameter Triggers

Here are some common QoS parameters that can be used for triggers and how they can be implemented:

- **Latency**: Latency refers to the time it takes for a data packet to travel from the source to the destination. High latency can result in delays and sluggish network performance. Triggers can be set to activate spectrum capture if the latency exceeds a predefined threshold, indicating potential network congestion or issues.
- **Packet Loss**: Packet loss occurs when data packets are dropped or fail to reach their destination. Excessive packet loss can degrade the user experience, especially in real-time applications like video conferencing or online gaming. Triggers can be designed to capture spectrum data when packet loss rates cross a certain limit.
- **Jitter**: Jitter is the variation in the delay of received packets. It can lead to disruptions in audio and video streams. If the jitter exceeds a specified level, it can trigger the capture of spectrum data to investigate the cause.
- **Bandwidth Utilization**: Monitoring the bandwidth utilization of individual cable modems or specific channels can help identify congestion or heavy data usage. Triggers can be set to capture spectrum data for modems or channels experiencing high bandwidth utilization.
- **Signal-to-Noise Ratio (SNR)**: SNR measures the strength of the signal compared to the background noise. Low SNR values can lead to poor data transmission and reception. Triggers can be implemented to capture spectrum data when SNR falls below a certain threshold.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- **Modulation Errors**: Modulation errors occur when the data encoding and decoding process experiences issues, leading to corrupted data. Triggers can be set to capture spectrum data when the number of modulation errors exceeds a specified limit.
- **Service-Level Agreement (SLA) Compliance**: Operators often offer SLAs to customers, specifying certain performance guarantees. Triggers can be implemented to capture spectrum data when the network is not meeting the SLA commitments.

8.4.2. Implementation of Triggers:

I. Monitoring System: Implement a monitoring system at the cable headend that continuously tracks the QoS parameters for each cable modem or upstream channel. The system should be capable of analyzing the real-time data to detect any deviations from the predefined thresholds.
II. Threshold Configuration: Set appropriate thresholds for each QoS parameter based on the desired network performance and user experience. Thresholds should be carefully chosen to avoid unnecessary triggering due to normal fluctuations.
III. Triggering Mechanism: When the monitoring system detects that a QoS parameter has exceeded its threshold, it should trigger the spectrum capture process for the relevant cable modems or channels. This can be achieved by sending commands to the CCAP to initiate the capture.
IV. Data Analysis and Response: Once the spectrum data is captured, the headend should analyze it to identify the cause of the QoS degradation. Based on the analysis, appropriate measures should be taken to address the issue, such as optimizing channel selection, adjusting modulation profiles, or resolving network congestion.

By implementing triggers based on QoS parameters, customer service health can be proactively monitored and their performance managed, leading to a better user experience and more efficient use of available network resources.

8.5. Implementing a User-Requested Trigger

User-requested triggers allow for remote monitoring while troubleshooting events in real time. It allows the technician or engineer to view the impact of any changes being made to the network as they are made, allowing them to determine whether the impairment has been mitigated or resolved. This proves especially useful as RF board within fibre nodes and amplifiers are being converted to use digital attenuation and equalization.

To implement a user-requested trigger:

I. User Interface Development: Develop a user-friendly interface that allows users to request the spectrum capture. This interface could be a web portal, a mobile app, or any other user-accessible platform.
II. Authentication and Authorization: Implement a secure authentication mechanism to ensure that only authorized users can access the trigger functionality. Users may need to log in with their credentials before they can use the feature.
III. User Permissions: Determine which users or customer groups should have access to the trigger feature. Depending on the network's configuration, not all users may be allowed to request spectrum capture.
IV. Trigger Request Form: Create a form or button within the user interface that enables users to request the spectrum capture. This form should include relevant information, such as the user's account details and any additional notes or comments.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

V. Command Generation: When a user submits the trigger request form, the system should generate a specific command or message that instructs the CCAP to initiate the spectrum capture process.
VI. Data Analysis and Response: Once the spectrum data is captured, the headend should analyze it to identify the cause of the QoS degradation. Based on the analysis, appropriate measures should be taken to address the issue, such as optimizing channel selection, adjusting modulation profiles, or resolving network congestion.

8.6. Implementing a Dynamic Trigger based on Utilization or Congestion

8.6.1. Implementing a Dynamic Trigger based on Congestion

Implementing a dynamic trigger based on congestion for upstream spectrum capture involves continuously monitoring network congestion levels and triggering the capture process for specific cable modems or channels when congestion exceeds a predefined threshold. To implement a dynamic trigger based on congestion:

I. Congestion Monitoring: Set up a congestion monitoring system at the cable headend that continuously tracks network congestion levels in real-time. This system should analyze various performance metrics to detect congestion, such as packet loss rates, latency, and bandwidth utilization.
II. Define Congestion Threshold: Determine the threshold value beyond which the network is considered congested. This threshold could be based on a combination of different congestion metrics or a single critical metric that indicates congestion severity.
III. Congestion Evaluation Interval: Decide on the frequency at which the congestion monitoring system evaluates congestion levels. It could be on a continuous basis or at specific intervals (e.g., every few seconds).
IV. Trigger Command Generation: When the congestion monitoring system detects that the network congestion has exceeded the predefined threshold, it should generate a specific trigger command or message.
V. Identify Affected Modems or Channels: Based on the congestion analysis, the system should identify the cable modems or upstream channels that are experiencing the highest levels of congestion.
VI. Data Analysis: Once the data is collected, the operator's system can process and analyze the spectrum information to understand the congestion patterns and potential causes.
VII. Optimization and Load Balancing: Analyze the captured data to identify channels or areas that are consistently experiencing congestion. Based on the analysis, the operator can optimize channel assignments and implement load balancing strategies to alleviate congestion issues.

8.6.2. Implementing a Dynamic trigger based on Utilization

Implementing a dynamic trigger based on utilization for upstream spectrum capture involves continuously monitoring the utilization levels of different channels or cable modems in real-time and triggering the capture process when utilization exceeds a predefined threshold. To implement a dynamic trigger based on utilization:

I. Utilization Monitoring: Set up a utilization monitoring system at the cable headend that continuously tracks the utilization levels of upstream channels or individual cable modems in real-time.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

II. Define Utilization Threshold: Determine the threshold value beyond which a channel or cable modem is experiencing high utilization. This threshold could be a percentage of the maximum available bandwidth or a specific data rate.
III. Utilization Evaluation Interval: Decide on the frequency at which the utilization monitoring system evaluates the utilization levels. It could be on a continuous basis or at specific intervals (e.g., every few seconds).
IV. Trigger Command Generation: When the utilization monitoring system detects that the utilization of a channel or cable modem has exceeded the predefined threshold, it should generate a specific trigger command or message.
V. Identify Affected Modems or Channels: Based on the utilization analysis, the system should identify the cable modems or upstream channels that are experiencing high utilization levels.
VI. Data Analysis: Once the data is collected, the operator's system can process and analyze the spectrum information to understand the utilization patterns and potential causes of high utilization.
VII. Optimization and Load Balancing: Analyze the captured data to identify channels or cable modems that are consistently experiencing high utilization. Based on the analysis, the operator can optimize channel assignments and implement load balancing strategies to distribute the traffic more evenly.

8.6.3. Congestion vs Utilization, which provides more value?

Both triggering based on congestion and triggering based on utilization are valuable techniques, but their value depends on the specific goals and requirements of the cable broadband network. Let us explore the benefits of each method:

8.6.3.1. Triggering Based on Congestion:

I. Real-Time Issue Detection: Congestion-triggered spectrum capture allows for real-time detection of network congestion. When congestion exceeds a predefined threshold, the capture process is initiated, enabling quick identification of congestion-related issues.
II. Proactive Network Management: With congestion-triggered capture, operators can proactively address congestion issues and take corrective measures before they significantly impact user experience.
III. Insights into Network Performance: Analyzing captured spectrum data from congested areas helps identify the root causes of congestion, such as interference, faulty equipment, or bandwidth-hungry applications.
IV. Dynamic and Adaptive: The trigger can be adjusted to respond to varying network conditions, allowing the system to adapt to changing traffic patterns and optimize performance accordingly.

8.6.3.2. Triggering Based on Utilization:

- Efficient Resource Allocation: Utilization-triggered spectrum capture enables operators to monitor the efficiency of resource allocation and identify areas where resources are underutilized or overloaded.
- Load Balancing: By capturing spectrum data based on utilization levels, operators can implement load balancing strategies to evenly distribute traffic across available channels, optimizing the use of network resources.
- Preventive Maintenance: High utilization levels can be indicative of potential issues. Triggering spectrum capture based on utilization helps identify areas where network capacity might need to be expanded before congestion becomes a significant problem.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Identifying Traffic Patterns: Analyzing utilization data provides insights into user behavior and traffic patterns, which can inform network planning and optimization.

8.6.3.3. Choosing the Right Method:

Selecting the more valuable triggering method depends on the operator's priorities and network conditions. If the primary concern is real-time detection and resolution of congestion-related issues, triggering based on congestion might be preferred. On the other hand, if efficient resource allocation and load balancing are top priorities, triggering based on utilization could be more valuable.

In many cases, a combination of both methods may be beneficial. Operators can use a dynamic trigger based on congestion for real-time issue detection and resolution, while utilizing a utilization-based trigger for longer-term network planning and optimization.

The value of each method comes down to how effectively they help the ISP manage network performance, improve user experience, and optimize resource allocation in line with their specific goals and objectives.

8.7. Intermittent Impairments

The threshold crossing trigger method can be used to capture distinct types of intermittent impairments in a cable broadband network. Here are some examples of intermittent impairments that can be captured using this method:

8.7.1. Impairment Types

8.7.1.1. Packet Loss

- Threshold crossing can detect instances where the packet loss rate exceeds the predefined threshold, indicating intermittent issues with data transmission.
- To capture intermittent packet loss, the QoS parameter you would typically use is the Packet Loss Rate. The Packet Loss Rate measures the percentage of data packets that are lost or not delivered successfully from the source to the destination over a specific period.

8.7.1.2. Latency Spikes

- Intermittent latency spikes can cause delays in data transmission. By monitoring latency and using threshold crossing, network administrators can capture such spikes for further analysis.
- To capture intermittent latency spikes, the QoS parameter you would typically use is the Round-Trip Time (RTT) or Ping Time. RTT measures the time it takes for a data packet to travel from the source to the destination and back again, and it is commonly used to assess network latency.

8.7.1.3. Jitter Variations

- Intermittent variations in jitter can cause disruptions in audio and video streams. The threshold crossing method can identify periods when jitter exceeds normal levels.
- To capture intermittent jitter variations, the QoS parameter you would typically use is the Jitter. Jitter measures the variation in the delay between data packets as they traverse the network from the source to the destination. It quantifies the irregularities in packet arrival

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

times and is an essential metric for assessing the stability and quality of real-time communication, such as VoIP (Voice over Internet Protocol) and video conferencing.

8.7.1.4. Signal-to-Noise Ratio (SNR) Fluctuations

- Intermittent SNR fluctuations can result from external interference or other factors. Threshold crossing can detect such variations, pointing to potential issues affecting signal quality.
- To capture intermittent signal-to-noise fluctuations, the QoS parameter you would typically use is the Signal-to-Noise Ratio (SNR). SNR measures the strength of the signal compared to the background noise present in the communication channel.
- Two intermittent impairment types that can be captured using the SNR Quality of service metric are
  - Common Path Distortions (CPD)
  - Interference.

8.7.1.5. Channel Noise

- Intermittent channel noise, such as impulse noise, can lead to momentary disruptions. This method can capture periods when noise levels exceed the predefined threshold.
- To capture intermittent channel noise, the QoS parameter you would typically use is the SNR. SNR measures the strength of the desired signal compared to the background noise present in the communication channel.

8.7.1.6. Channel Utilization Spikes

- Threshold crossing can identify instances when upstream channel utilization experiences sudden spikes, which may lead to congestion and degraded performance.
- To capture intermittent channel utilization spikes, the QoS parameter you would typically use is the Channel Utilization or Network Traffic Load. Channel utilization refers to the percentage of time that a communication channel is occupied by data traffic or transmissions.

8.7.1.7. Modulation Errors

- Modulation errors can occur intermittently due to numerous factors. Threshold crossing can capture periods when the number of errors exceeds the predefined threshold.
- To capture intermittent modulation errors, the QoS parameter you would typically use is the Forward Error Correction (FEC). The uncorrectable codeword error rate measures the rate of errors in the received data compared to the original transmitted data that cannot be repaired. It quantifies the accuracy of data transmission and reflects the quality of the modulation and demodulation processes in the communication channel.

8.7.1.8. Service Unavailability

- Intermittent service outages or unavailability can be captured using the threshold crossing method, triggering spectrum capture for analysis during such occurrences.
- To capture intermittent service unavailability, the QoS parameter you would typically use is the Service Availability or Uptime. Service availability measures the percentage of time that a service or network is operational and accessible to users.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

8.7.1.9. Bandwidth Degradation

- This method can detect instances of intermittent bandwidth degradation, which may impact data transfer rates and overall network performance.
- To capture intermittent bandwidth degradation, the QoS parameter you would typically use is the Throughput or Bandwidth Utilization. Throughput measures the amount of data transmitted over a network or link in a given period, and bandwidth utilization represents the percentage of available bandwidth being used at any given time.

It is important to note that the effectiveness of the threshold crossing trigger in capturing intermittent impairments depends on the selection of appropriate QoS parameters and the proper setting of thresholds. Careful consideration of the network's characteristics and the specific impairments to be detected is essential to ensure accurate and meaningful captures of intermittent issues.

Capturing intermittent impairments using the threshold crossing trigger allows network administrators to investigate and address these issues promptly, leading to improved network stability and better overall user experience.

8.7.2. Implementing a Trigger for Intermittent Impairments

To capture intermittent impairments, a trigger based on Threshold Crossing could be used. The threshold crossing trigger is designed to initiate the spectrum capture process when specific QoS parameters or performance metrics cross predefined thresholds.

Implementing a threshold crossing trigger to capture intermittent impairments involves the following steps:

I. Selecting Relevant QoS Parameters: Identify the QoS parameters that are indicative of intermittent impairments.
II. Defining Thresholds: Set appropriate thresholds for each selected QoS parameter. The thresholds should be carefully chosen based on the network's normal operating range and the severity of impairments that need to be captured.
III. Monitoring System: Implement a monitoring system at the cable headend that continuously tracks the selected QoS parameters in real-time.
IV. Threshold Crossing Detection: Configure the monitoring system to detect when any of the selected QoS parameters cross their predefined thresholds. This indicates the occurrence of intermittent impairments.
V. Trigger Command Generation: When the monitoring system detects a threshold crossing event, it should generate a specific trigger command or message.
VI. Identifying Affected Modems or Channels: Based on the threshold crossing analysis, the system should identify the specific cable modems or upstream channels that experienced the intermittent impairments.
VII. Data Analysis: Once the data is collected, the ISP's system can process and analyze the spectrum information to investigate the intermittent impairments and their potential causes.

By implementing a threshold crossing trigger, the cable broadband network can capture intermittent impairments as they occur, allowing for timely investigation and resolution of issues, improving network performance and user experience.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

9. Applying UTSC to PHM and PNM frameworks

UTSC has further applications in network health prognosis and fault prediction when integrated into both PHM framework and PNM guidance. Effective utilization of spectrum capture data rely on several steps.

Prognostic Health Management (PHM)

I. Capture spectrum data to continuously monitor the quality of upstream data transmission using the triggers outlined in Section 6 of this paper.

II. Preprocessing of the captured data handling missing values or outliers and alignment of timestamps with other datasets

III. Feature engineering, extracting meaningful features from the captured spectrum data, such as statistical measures, frequency domain characteristics, and patterns of degradation. Combine these features with other relevant features to create a comprehensive dataset for analysis.

IV. Health indicator development using the extracted features to develop health indicators that quantitatively represent the condition of the network components. These indicators can be designed to capture signs of degradation, deviations from normal behavior, and early warning signs of potential issues.

V. Model development utilizing machine learning or statistical modeling techniques to develop predictive models based on the health indicators. Train these models using historical data, where failures or degradation events are labeled.

VI. Prediction and prognosis applying the trained models to real-time or near-real-time data to predict the likelihood of future failures or performance degradation. Generate forecasts of remaining useful life, estimating how much time remains before a component might fail or degrade significantly.

Proactive Network Maintenance (PNM)

VII. Anomaly detection by deploying the trained predictive models to analyze real-time spectrum data. The models can identify anomalies or deviations from normal behavior, indicating potential network degradation or imminent issues.

VIII. Alert generation when an anomaly or degradation is detected and crosses a predefined threshold, generate alerts or notifications for network operators or maintenance teams. The alerts should provide details about the issue and its severity.

IX. Root cause analysis upon receiving an alert, using the captured spectrum data. The data can help pinpoint the location and potential cause of the issue.

X. Decision support to network operators by presenting them with actionable insights derived from the predictive models. This enables operators to make informed decisions about maintenance schedules and resource allocation.

XI. Proactive actions based on the analysis, decide on the appropriate proactive actions. These might include adjusting network parameters, redistributing traffic, optimizing routing, or scheduling maintenance.

XII. Maintenance scheduling using the insights from the analysis to schedule maintenance activities. Proactively address the identified issues during planned maintenance windows to minimize service disruptions.

XIII. Verification and validation after performing maintenance or optimization actions, verify their effectiveness by monitoring the spectrum data and observing changes in network behavior.

XIV. Performance evaluation of the predictive models and the accuracy of the prognostic predictions

XV. Continuous learning, updating, and refining the predictive models using new data as it becomes available. Incorporate feedback from maintenance actions and outcomes to improve the accuracy of predictions over time.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

By incorporating UTSC into your PHM framework and driving PNM activities based on its data, operators can enhance their ability to predict and prevent network issues. Leveraging captured spectrum data alongside other relevant information, can detect early signs of degradation, predict failures, and take proactive measures to maintain network health and optimize performance. This allows operators to assess the condition of network components, minimize downtime, improve customer satisfaction, reduce the operational costs associated with reactive troubleshooting, improve resource utilization, and enhance the overall reliability of the network.

10. Future Scope

The future scope of UTSC is promising, and the technology holds potential for various applications beyond its current use cases. As cable broadband networks continue to evolve, UTSC can play a crucial role in optimizing network performance, enhancing user experiences, and supporting innovative technologies. Some potential future applications and scopes for UTSC:

- **Integration with AI and Machine Learning**: The integration of AI and machine learning algorithms with UTSC can enhance data analysis capabilities. AI can help in automating issue detection, predicting potential network problems, and suggesting optimal solutions, leading to more efficient and proactive network management.
- **Support for New Services and Technologies**: As new services and technologies emerge; UTSC can adapt to monitor and optimize their performance. Whether it is supporting small cell over DOCSIS, smart city applications, IoT devices, or other bandwidth-intensive applications, the technology can provide critical insights to maintain high-quality service delivery. UTSC can contribute to reliable data transmission and network performance, supporting their growth.
- **Network Automation and Self-Healing Networks**: UTSC can be part of an overall network automation framework. By combining real-time monitoring and automated issue resolution, the technology can contribute to the creation of self-healing networks that can detect and fix problems without human intervention.
- **Security and Anomaly Detection**: The captured spectrum data can also be analyzed for security purposes. Anomalies in the spectrum can indicate potential security threats, and UTSC can be integrated with cybersecurity tools to detect and mitigate such threats.
- **Energy Efficiency and Green Networking**: Operators can use UTSC to optimize network resources and reduce energy consumption. By identifying and resolving network issues promptly, unnecessary energy expenditure can be minimized, contributing to green networking initiatives.
- **Support for Edge Computing**: As edge computing becomes more prevalent, UTSC can be extended to monitor and optimize performance at the network edge, ensuring efficient data processing and reduced latency for edge applications.
- **Interoperability and Standards**: As UTSC becomes more widely adopted, there may be a push towards standardization and interoperability across different equipment and vendors, allowing for seamless integration into diverse network environments.

As technology evolves, UTSC can play a vital role in ensuring network efficiency, supporting new services, and enhancing user experiences. With advancements in data analysis, automation, and network intelligence, the technology has the potential to revolutionize network management and pave the way for more advanced and efficient cable broadband networks in the future.

11. Conclusion

UTSC holds significant importance in the realm of cable broadband networks and network management. Its key significance lies in the proactive and efficient approach it brings to monitoring, troubleshooting,

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

and optimizing the upstream data transmission. Some of its major benefits include proactive issue detection, faster troubleshooting and resolution, data-driven decision making leading to improved network stability and performance and an enhanced customer experience, reduced operational costs leveraging automation, and support for future network evolution.

While its implementation is not without challenges and limitations, it represents a major leap in network troubleshooting and management. It allows for investigation of upstream events to be automated, becoming more efficient and effective, setting it up to support upstream spectral impairment detection. UTSC provides further benefits improving network reliability and optimizing resource allocation, which will lead to improved customer satisfaction and cost savings. The implications of UTSC cannot be overstated. This innovative technology empowers operators with a proactive and data-driven approach to network monitoring and troubleshooting. By capturing intermittent impairments, operators can optimize network performance, reduce downtime, and provide a superior internet experience to their customers, enhancing their reputation and competitiveness in the industry.

Abbreviations

UTSC	Upstream Triggered Spectrum Capture
PNM	Proactive Network Maintenance
RF	Radio Frequency
DAA	Distributed Access Architecture
R-PHY	Remote Phy
CCAP	Converged Cable Access Platform
RPD	Remote Phy Device
SNMP	Simple Network Management Protocol
TFTP	Trivial File Transfer Protocol
L2TP	Layer Two Tunneling Protocol
ISP	Cable operators
FBC	Full Band Capture
QoS	Quality of Service
MTTR	Mean Time to Restore
ADC	Analog to Digital Converter
FFT	Fast Fourier Transform
DUT	Device Under Test
iCCAP	Integrated Converged Cable Access Platform
rCCAP	Remote Phy Converged Cable Access Platform
vCCAP	Virtual Converged Cable Access Platform
kHz	Kilohertz
ms	Millisecond
RCP	Remote Phy Control Protocol
GCP	Generic Control Plane
MIB	Management Information Base
SLA	Service Level Agreement

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SNR	Signal-to-noise
MER	Modulation Error Rate
CPD	Common Path Distortion
BER	Bit Error Rate
AI	Artificial Intelligence
IoT	Internet of Things

Bibliography & References

DOCSIS 3.1 CCAP Operations Support System Interface specification, OSSI-CCAPv31

CableLabs Remote PHY Specification, CM-SP-R-PHY

Remote PHY OSS Interface Specification, CM-SP-R-OSSI

CableLabs Remote Upstream External PHY Interface Specification, CM-SP-R-UEPI.

PNM Current Methods and Practices in HFC Networks (DOCSIS® 3.1) CM-GL-PNM-3.1

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Using Automated Live QC to Reduce Construction Costs and Improve Readiness

An Operational Practice prepared for SCTE by

Scott Johnston
Director, Network Engineering
Comcast
183 Inverness Drive West, Englewood, CO
bryan_johnston@cable.comcast.com

Anthony Curran
Principal Eng II, Software Dev & Eng
Comcast
Virtual Location, Virtual, MA
anthony_curran@cable.comcast.com

Michael O'Dell
Distinguished Engineer
Comcast
Virtual Location, Virtual, OH
michael_odell@cable.comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Using Cloud-Based Meter Technology to Increase Consistency and Quality ... 3
   - 2.1. Sweep and Balance ... 4
   - 2.2. Leveraging Automation in the Cloud Environment ... 4
   - 2.3. Change is a Constant ... 6
   - 2.4. Database Connections ... 6
   - 2.5. Reporting and End of Job Closeout ... 7
3. Future Enhancements ... 7
4. Anticipated and Observed (Trial) Outcomes ... 9
5. Conclusion ... 9

Abbreviations ... 10

Bibliography & References ... 10

List of Figures

Title **Page Number**

Figure 1 - A clean, linear frequency full band capture response ... 4
Figure 2 - Typical amplifier upgrade order of operations ... 5
Figure 3 - DS Sweep/Alignment Low/High/Peak to Valley Values ... 5
Figure 4 - Reference and subsequent sweeps ... 6
Figure 5 - Network architecture showing a group of modems not bonding to orthogonal frequency division multiple access (OFDMA) ... 8

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

1. Introduction

Multiple systems operators (MSO) are deploying a remarkable number of new amplifiers in modern hybrid fiber coaxial (HFC) systems to facilitate mid-split, high split, and data over cable system interface specifications (DOCSIS) 4.0 full duplex (FDX) and frequency division duplex (FDD) deployments. This capital investment is focused entirely on our industry's ability to increase the bandwidth needed to satisfy our expectations around customer expansion. This capital investment doesn't come cheap, and all operators are interested in a rapid return on investment. Network upgrades should yield the desired result on the day of the cut. Remediations in the form of partner go-backs or internal investigations are very costly. Lost partner time result in construction slowdowns as crews are required to go back to areas they've previously worked. Network investigations performed by the local plant maintenance team take away from demand and preventive maintenance activities. The customer experience is lackluster due to additional, unwanted service interruptions. Regardless of who is going back, the cost is high, and the impact is an overall reduction in value.

If we look at today's construction environment, we find ourselves in a fast-paced culture of cut, swap, and align. The desired end state in terms of network architecture dictates whether business partners, generally, will only cut the amplifiers in the HFC node or the taps and passives as well. As one can imagine, the "amplifier only" solution is the most affordable and straightforward option. Typically, a single node with four bus legs can be cut in as little as two days depending on the number of homes passed and the workforce doing the job. Resources, time, and peoplepower are in limited quantity in today's world. This upgrade activity requires a high level of trust between both the MSO and the business partner. Expectations must be clearly delivered in the statement of work and both parties must be made aware of the overall objective and the focus on quality. In most upgrade situations, we're trusting the splicers, technicians, and coordinators to manage the quality and performance of the amplifier setup with limited interaction.

This focus on quality is generally assumed as both parties work with each other in good faith. But, as inferred, the upgrade process must be completed with limited error. MSOs expect to turn up more advanced modulation methods in larger sections of bandwidth in both the forward and return paths. Those turn-ups are the foundation for increased speeds, lower latency, and additional capacity for the customer to utilize and enjoy. Many of our splicers and technicians are extremely capable, talented, and trusted, while some need more training. Given the varying levels of knowledge, skill, and ability, we expect inconsistent outcomes at the end of the day. Since skill variation does exist, the quality assurance (QA)/quality control (QC) process continues to be a valuable function in the construction process often constrained to the lowest common denominator. Our biggest problem is that Quality Assurance/Control coordinators simply cannot be in all places at all times to evaluate network performance.

2. Using Cloud-Based Meter Technology to Increase Consistency and Quality

Today's modern signal meters are more than a standalone testing tool. Several of the industry-available meters currently operate as part of an ecosystem. The platform typically consists of a hand-held field unit, a communications device (either embedded in the unit or via a mobile device), and a back-end cloud data storage system. This meter platform is extremely powerful as it collects an incredible amount of data with unique identifying properties. Using these platforms, we immediately know the following:

1. Who's logged into the meter platform taking the measurements
2. Where these measurements were taken
3. The time the measurement was taken

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

4. Detailed signaling information inside the scan payload (signal level, modulation error ratio (MER), bit error rate (BER), in channel frequency response (ICFR), ping, throughput, etc.) (Curran/Martushev SCTE 2020)

With this detail, we know exactly who is doing what, where, when, and how. Managing performance, outcome, and accountability is clear. The power of the platform is self-evident and should be applied in a professional and ethical manner.

2.1. Sweep and Balance

The replacement of any HFC amplifier necessitates the use of hand-held test equipment in order to properly balance and sweep. If you're not already familiar with sweep and balance, the process is used to achieve the proper inputs and outputs to the amplifiers in the cascade. This ensures that the last amplifier in sequence performs similarly to the optical node at the head of the cascade. This "unity gain" is what drives a consistent experience for any of our customers whether they are connected to a tap the node or a tap at a remote line-end in the network; a technical advantage over DSL systems and an equalizer when looking at a fiber competitor.

Properly setting this frequency response and unity gain in the outside plant yields a layer-one environment suitable for anything that network engineering (or marketing) chooses to place on it. A linear response set to the correct signal level ensures packet performance. When networks are constructed and swept to design specifications the system will provide maximum throughput and a full return on capital investment.

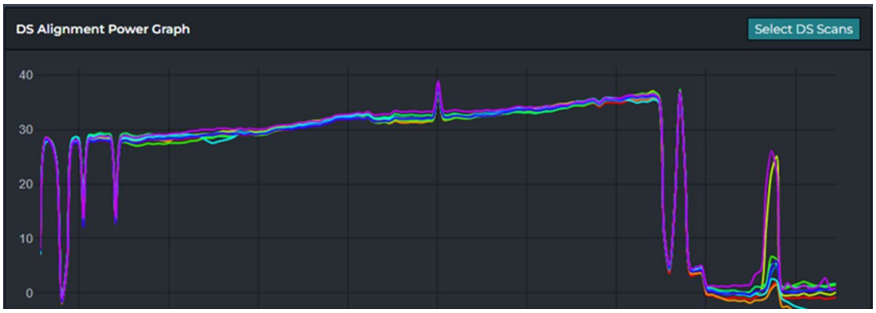


Figure 1 - A clean, linear frequency full band capture response

2.2. Leveraging Automation in the Cloud Environment

In times past, sweep technicians were required to store images of sweep; either in the form of polaroid photos or dot matrix printouts placed in binders for each node worked. Today technicians are using a cloud-based meter platform. The data collected can be used to automatically grade the outcome of a sweep and balance evolution.

In order to setup automated Live QC, you will need to have the following components in place:

1. Cloud-based meter/platform
2. Back-end software that can record/grade the work
3. Statement of work requirements for your business partners and internal technicians
    a. User guides and technical articles
    b. System parameters and performance expectations including pass/fail ranges
4. Scans taken and uploaded to the cloud platform

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Next, we will define the basic requirements for performing this task in automation:

Your back-end cloud must be able to accommodate multiple users sending scans to one job/record. Historically most sweep and balance tasks are done by one technician. In today's fast paced construction environment, many work crews consist of four or more splicers/sweep techs that cut amps very quickly in sequence.

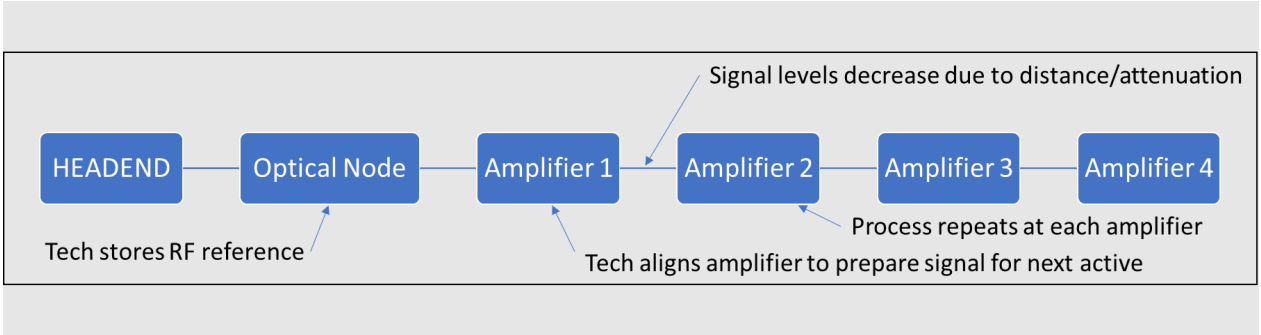


Figure 2 - Typical amplifier upgrade order of operations

Since there are component level differences and tolerances in each field signal meter, the technicians must store their own reference at the optical node to ensure that their individual amplifier sweeps are as close to "true" to the node as possible. Running channel scan comparisons between meters is a recommended step in order to ensure that each meter is within a reasonable measurement tolerance to each other. Meters that are significantly disparate should be factory/bench calibrated as soon as possible to improve the overall outcome of a group sweep and balance effort.

Thresholding must be set via variables in the back-end cloud to grade the activity. Typical amplifier sweep and balance consists of setting the amplifier to the low-frequency level and the high frequency level. Example: a standard level set in recent years would be 31 dBmV at 54 MHz and 41 dBmV at 750 MHz (10 dB tilt) using digital levels. The quality of the sweep is dictated by the Peak to Valley measurement between 54 and 750 MHz in this example. In a typical six-amplifier deep cascade a 3.5 dB peak to valley measurement would be acceptable to most technicians and field leaders. (Hranac, Broadband Library Spring 2022)

DS Alignment Info

Sweep	Low Ch. 3 31	High Ch. 120 41.5	Tilt	Peak to Valley	Test Point Compensation
Reference 3/29/19 11:02 AM	30.8dB	41dB	10.2dB	---	20dB
3/29/19 11:40 AM	31.1dB	41.3dB	10.2dB	1.1dB	20dB
3/29/19 12:30 PM	31.4dB	41.5dB	10.1dB	1.3dB	20dB

Figure 3 - DS Sweep/Alignment Low/High/Peak to Valley Values

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

In the example shown below (Figure 4), the red line is the "normalized" reference. Although the Y axis clearly shows a level of zero the actual RF spectrum levels have a value other than zero. The reference data is flattened visually while maintaining its true level information. Subsequent scans are contrasted against the reference (each color is a physical location) to show the end user the extent of the frequency response variations in the band. As you can see most of the individual amplifier sweeps are within approximately 2dB peak to valley until approximately the 600 MHz range. Two amplifier sweeps instantly stand out to the viewer as a notch/suck-out is seen at approximately 690 MHz (red oval). This notch is roughly 10-12dB down from the reference. Delivering the highest bit rate and modulation in this impaired frequency spectrum will be improbable (Leech/Martushev SCTE 2022). Additional analysis of Figure 3 shows that the high balance frequency is not being set properly (yellow oval). There are several real-world reasons for this inadequacy or it's possible that the technician doing the balancing is not aligning to ~750 MHz but instead choosing a high level/channel closer to 600 MHz.

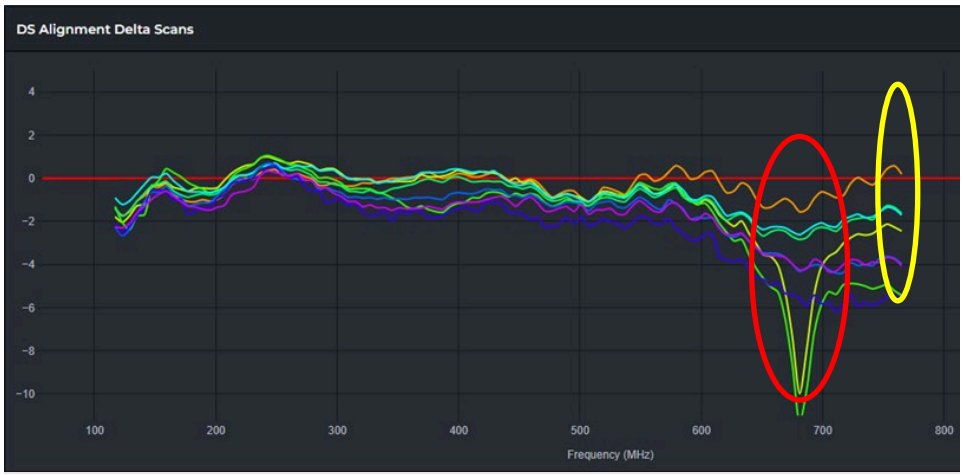


Figure 4 - Reference and subsequent sweeps

2.3. Change is a Constant

One anticipated behavioral change is regarding pass/fail implications and technician reaction. If splicers/sweep techs are held accountable via real-time QC, compliance becomes crucial to move to the next amplifier in cascade and complete the job. Some people have anticipated slowdowns in work completion and overall productivity. Internal trials have shown that cloud-monitored sweep and balance processes set clear expectations while driving quality and training efforts. Most technicians will quickly learn or discover what actions are necessary at the amplifier to pass a quality control check. Immediate notification can alert the technician if the alignment widget in the meter application is programmed to advise the technician when the optimum levels window is achieved. They won't have to wait for end-of-day reporting nor should they as the preceding amplifier's performance partially dictates the proceeding.

2.4. Database Connections

This automation process requires several different database connections. When the automation platform is properly synchronized with the construction management platform, moving a job number from one to the other should be a fairly easy task. Multiple technicians will need to work from the same job number in order to properly organize all of the sweep and balance records and construction coordinators will need to be able to reference the construction job to the appropriate scans and grading reports.

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

A connection to the MSOs monitoring and maintenance platform could be made in order to automate customer-facing communications before, during, and after the day of cut so that customers who need all-day connectivity can make arrangements outside of their normal subscriber location.

2.5. Reporting and End of Job Closeout

Reports are implied in any data collecting and analysis project. For this automated QC process there are several significant items that can be analyzed. The obvious checkpoint to report would be the pass/failure percentage for the overall job. But since amplifier work relies on success at the previous station this metric is redundant. Any given job will have a preset number of amplifiers to be worked. End of job reporting will tell us if the crew touched all of the amplifiers or may have forgotten one that was on the initial job list. In short, how many amplifiers did we ask you to cut, how many did you get to in the day and how well did you do on the amplifiers that you completed?

Since time stamping is built-in to the process, reporting on job times is simple down to the individual amplifier. Making recommendations on the number of people needed to work a node in order to meet a specific timetable would be easy to calculate using the average time per amplifier cut/sweep metric.

An overall quality/experience metric could be achieved per technician. Since pass/fail metrics are graded, applying a real-time score to the individual could be a realistic option. That sort of metric would provide data on who is a quick achiever versus someone who might need some additional training or help.

At the end of the job, a full report could be created to include the personnel doing the work, the coordinator facilitating the work, the locations, and the overall outcome of the activity. Actively reporting exceptions to any of the expected requirements would allow for quick investigation and same-day or next-day repair before work crew move farther away from the node in question.

3. Future Enhancements

Currently, we trust the technician to choose the correct low and high channels and their RF levels. In the future, we hope to auto-scrape the amplifier details from digital tombstones and place them in the sweep/balance targets inside our iOS application via GPS automation. Imagine opening a pedestal and having the application ask you if you're at HR10D001. A quick verification of the system prints, and the technician knows exactly what pads and equalizers to start with to achieve the designed system levels.

Another innovation we hope to add is the use of both RF cables on our meter platform. If we can measure the input and the output on the amplifier test points at the same time it is highly likely we can advise the technician on any necessary changes to the input, mid-stage, or output stage pads and equalizers. Technicians can unfortunately achieve proper output levels and sweep response while placing the incorrect pads/equalizers in the station. Improper padding at any stage of the amplifier can result in poor RF gain/quality performance (D. Linton, Broadband Library Spring 2022).

On a related note, colleagues have discovered a tell-tale method to know if you're operating an amplifier in (low) noise range or in distortion. It's related to the optical noise power ratio curve in an optical circuit and the concept is discussed in the Diane Linton article referenced in the previous paragraph. (O'Dell and Darby, SCTE 2023) go to greater lengths to describe how observing the noise floor on any RF scan at both the low, mid, and high sections of a Gigahertz wide spectrum will tell you if you're operating in noise or in distortion. We hope to apply these concepts in our meter platform application to advise a technician that their amplifier station is at risk for self-inflicted MER degradation.

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

And finally, a note on next-gen profile enablement and tooling to aid in understanding proper amplifier setup. Up until now, we have avoided talking about next-generation bonding profiles other than to mention that they offer additional bandwidth. One thing to note about any sweep process is that, ideally, the channel plan doesn't matter as long as it's loaded correctly (for total composite power) and it is swept out to each last amplifier in the cascade. One advantage to pre-loading the next-gen profile is the ability to see and overcome any RF impairments in both the linear and non-linear space. Occupied bandwidth is measured bandwidth. In addition, turning up full OFMDA capabilities allows you to query all of the modems in the service group to view their bonded state. Populations of modems that fail to bond when they should be an easy indicator of an improper amplifier setup or a missing amp altogether.

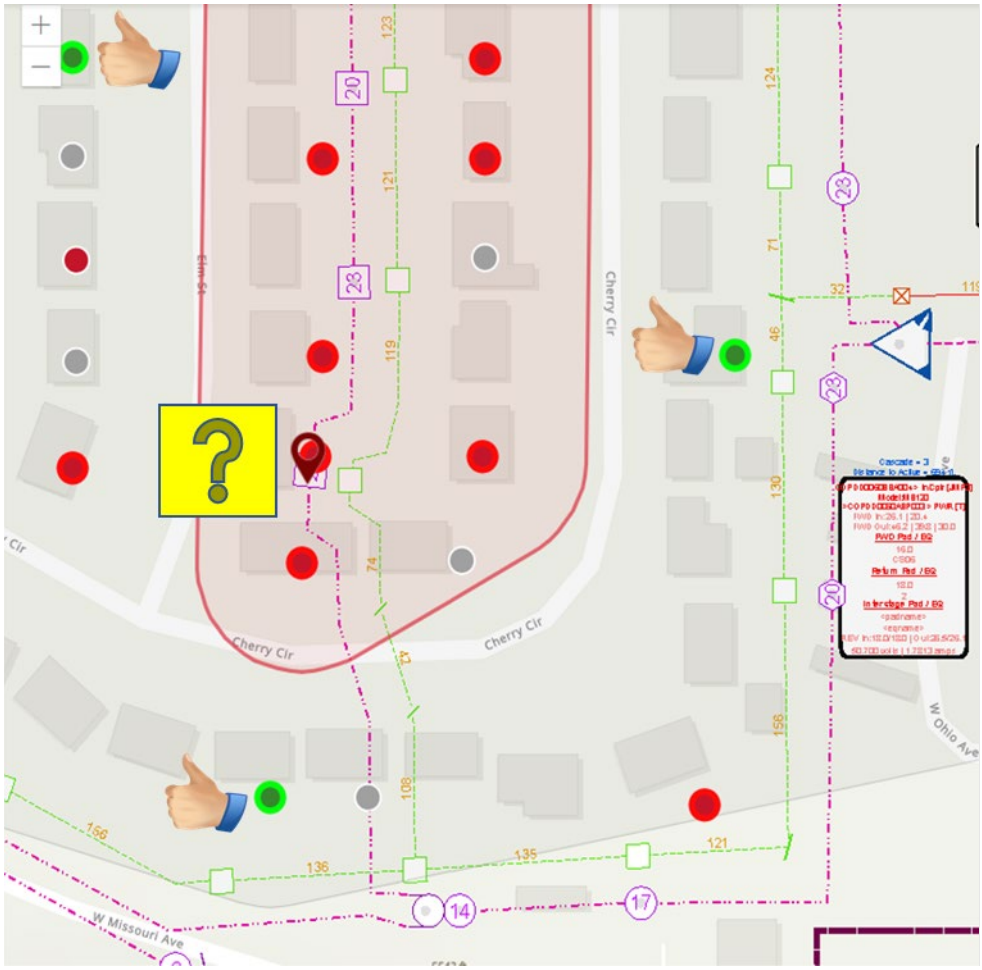


Figure 5 - Network architecture showing a group of modems not bonding to orthogonal frequency division multiple access (OFDMA)

In Figure 5, notice how the design goes from a 14 two-way tap to a splitter. The northbound leg heads to a 23 four-way tap. The amplifier icon and tombstone are missing. In this example, our tools told us that the construction team was never made aware of this amplifier. Since the system was cut with the previous generation low-split 5-42 return path configuration the amplifier was missed until the time that the next-gen bonding profiles were added to the system. This is an obvious go-back situation that resulted in additional customer downtime and pain. This would have been avoided with a comprehensive automated QC approach.

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

4. Anticipated and Observed (Trial) Outcomes

As stated earlier, in the short term we believe that technicians might be uncomfortable with the automation and change in expectations. Many splicers and line techs are pride oriented and don't feel the need to be observed or graded. We imagine this may be a limited sentiment as the automation allows for real-time documentation, quality control, and an unquestionable form of accountability. When one observes a sweep trace there is little to no doubt about the ability of the technician given new, high-quality materials to install. Our internal trials followed this pattern, slight apprehension, some doubt, review of expectations, a short learning curve and then adoption of the process. Nine months into our field trial the technicians under trial prefer the upload and grade method over previous methods as all parties involved have very few questions about work quality and performance.

In our forementioned ongoing field trial the results have been extremely positive. Go-backs and remediations in the trial area have dropped significantly over the control group. Pre-enablement has been in place as well and is driving down missed amplifiers and OFDMA bonding failures as well. There have been no significant increases in customer call-in rates or trouble calls in these areas.

5. Conclusion

As HFC technology has matured, the depth of the amplifier cascades has shortened. This reduced amplifier count has allowed us to slowly lose our empirical knowledge when it comes to amplifier setup; chasing noise became paramount. As the amplifiers have become more efficient along with better error correction we started walking away from sweep and balance. Technologies like profile management applications (PMA) have made it easy to disregard a physical layer issue up until the point of total failure and in some instances, we're relying on signal levels and MER at the customer premise equipment to gauge performance.

With the advent of Remote PHY, it's no longer acceptable to see MER values in the mid-thirties. Simple binary online/offline indications only tell us if we're "working." Before the widespread use of digital modulation in cable networks that may have been a sufficient metric for performance. In today's networks, we need fully quantitative measurements that nearly guarantee the maximum level of performance, availability, and reliability. Without observing the performance of every one of the new amplifiers you're putting in the network how will you know if you're truly getting your money's worth?

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Abbreviations

BER	Bit Error Rate
CPE	Customer Premise Equipment
DOCSIS	Data Over Cable System Interface Specifications
FDD	Frequency Division Duplex
FDX	Full Duplex (DOCSIS)
HFC	Hybrid Fiber Coaxial (network architecture)
ICFR	In Channel Frequency Response
MER	Modulation Error Ratio
MSO	Multiple Systems Operator
OFDMA	Orthogonal Frequency Division Multiple Access
PMA	Profile Management Application
QA/QC	Quality Assurance/Quality Control
RPHY	Remote Physical Layer technology
SCTE	Society of Cable Telecommunications Engineers

Bibliography & References

Curran, A., & Martushev, A. (2020): *Software Revolution of Field Meters Using a Field-Capable Measurement Device*; SCTE Cable-Tec EXPO 2022, A Virtual Experience

Frequency Response Testing, A Peek Under the Hood, Ron Hranac; Broadband Library Spring 2022

Leech, J., & Martushev, A. (2022): *PMA Improvements – Strategies Employed for Faster Mitigation, Increased Capacity, and Cost Savings*; SCTE Cable-Tec EXPO 2023, Philadelphia, PA

How Can NPR Improve Future 1.8 GHz Extended Spectrum Amplifier Characterization?; Diane Linton; Broadband Library Spring 2022

O'Dell, M., & Darby, W. (2023): *Using PNM Tools to Identify Misaligned Amplifiers*; SCTE Cable-Tec EXPO 2023, Denver, CO

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23 CELEBRATING 40 YEARS

2023 Fall Technical Forum SCTE • CABLELABS • NCTA

Using ODM 3.0 to Fix Customer Impairments

Overall DOCSIS Metric

A Technical Paper prepared for SCTE by

Andrew Hennebury
Sr. Cable Access Network Specialist
Rogers Communications
8200 Dixie Rd, Brampton, On L6T 0C1
437-779-0663
Andrew.Hennebury@rci.rogers.ca

Ray Stevens
Manager Service Reliability Engineering
Rogers Communications
8200 Dixie Rd, Brampton, On L6T 0C1
647-204-2493
Ray.Stevens@rci.rogers.ca

Iqra Syed
Sr. Cable Access Network Specialist
Rogers Communications
8200 Dixie Rd, Brampton, On L6T 0C1
647-278-2632
Iqra.Syed@rci.rogers.ca

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
   1.1. Background ... 3
2. Scope and Limitations ... 3
3. Overview of HFC Networks ... 4
   3.1. SC-QAM Channels and Signal Characteristics ... 4
   3.2. Orthogonal Frequencies and Technical Visibility ... 4
4. Initial Iterations of ODM ... 4
   4.1. Overall DOCSIS Metric Inception and Execution ... 4
5. Network Lab Testing ... 6
   5.1. Testing Tools and Methodology ... 6
   5.2. Characterization Method and Data Capture ... 6
6. Analyzing the Algorithm ... 7
   6.1. Developing a Modular Solution ... 7
7. Implementation of Algorithm and Cloud-Based Systems ... 8
   7.1. Proof-of-Concept (PoC) Development ... 8
   7.2. Cloud Solution ... 8
8. Measure of Success ... 9
   8.1. Measuring Success of ODM ... 9
   8.2. Production Example ... 9
9. Conclusion ... 10
   9.1. Summary of Key Findings ... 10
   9.2. Future Improvements and Areas of Further Research ... 11

Abbreviations ... 12

Bibliography & References ... 12

List of Figures

Title **Page Number**

Figure 1- ODM Confusion Matrix ... 4
Figure 2- Decrease in false-positives with optimized weights ... 5
Figure 3- Decrease in false-negatives with optimized weights ... 5
Figure 4- Automation Lab ... 6
Figure 5 - This is a sample of US Score ... 9
Figure 6- Sample Incident ticket with hardware failure ... 10
Figure 7 - Before/After ODM Score Improvement ... 10

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

1.1. Background

Detecting network issues and solutioning them has served as a complex challenge for telecommunication firms across the globe for decades. In a post-pandemic world, with connectivity at the forefront, reliability becomes the key focus. Overall DOCSIS® metric (ODM) helps build a reliable network with its rapid detection of network impairments and added clarity of where and what action is required.

ODM's first version was developed for upstream (US) SC-QAM 64-QAM deployment and proactive network maintenance (PNM). The algorithm was effective but biased towards upstream metrics. ODM+, the second version of the algorithm, was developed to enhance detection by integrating downstream KPIs (key performance indicators) and DOCSIS (Data Over Cable Service Interface Specification) 3.1 spectrum. The third iteration of the algorithm, ODM 3, focuses on connectivity and stability. ODM 3 combines DOCSIS 3.0/3.1 HFC (hybrid fiber-coaxial) parameters and applies a scoring to each weighted KPI by service impact. These scorings are utilized to detect connectivity impairments and grade overall node health within the network.

Each KPI is polled via SNMP (Simple Network Management Protocol) and IPDR (Internet Protocol Detail Record). KPI's include Rx (receive), Tx (transmit), SNR (signal-to-noise ratio), CER (correctable rrror ratio) and MER (modulation error ratio) 2nd percentile. Each KPI carries an individual scoring based on defined thresholds. A single KPI can have multiple thresholds to capture severity of impact.

A modem will be considered failing if its CM (cable modem) score is equal to or above 75. Once the pass/fail scoring for each modem on the network is completed, the node is scored based on pass/fail percentage of customers, nodes below a 90% passing customer threshold are considered actionable for PNM activities.

1.2. Purpose of the White Paper

The purpose of this white paper is to explore the ODM algorithm, its history, development and benefits. ODM was developed to identify both customers and areas where network impairments are present. With assigned scoring, technicians are able to identify which customers must be prioritized. Equipped with this prioritization and the clear identification of which KPI's within ODM are impaired, technicians are able to quickly resolve issues. With over 600 tickets generated, ODM has observed a 96% success rate.

2. Scope and Limitations

The ODM 3 algorithm is reliant on modem data polled every 15-60 mins. Multiple data sets exist to remove discrepancies and provide back-up data if primary data feeds fail.

Self-corrective and weather-related incidents may cause an increase in "no fault found" closure of IMTs which can impact the accuracy of ODM 3. Further development is currently underway to remove such cases.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Overview of HFC Networks

3.1. SC-QAM Channels and Signal Characteristics

DOCSIS 3.0 and 3.1 allow for visibility of many KPIs via MIBs (management information base). Many are utilized by ODM: downstream Rx, upstream Tx, downstream/upstream SNR, downstream/upstream CER. Each of these are polled hourly via an SNMP platform and the data is stored within a cloud service.

3.2. Orthogonal Frequencies and Technical Visibility

DOCSIS 3.1 allows the use of additional frequencies and better impairment visibility. ODM utilizes: orthogonal frequency-division multiplexing (OFDM) Rx power, (orthogonal frequency-division multiple access) OFDMA Tx power, OFDM Rx MER 2nd percentile and OFDMA MER 2nd percentile. The transmit levels at the time of polling will be obtained as is, as such OFDMA mini slots will not be utilized within the ODM model. Each of these are polled hourly via an SNMP platform.

4. Initial Iterations of ODM

4.1. Overall DOCSIS Metric Inception and Execution

With various activities of service monitoring and PNM in place focused on individual service impacting KPI's, the requirement for an overall metric became necessary. ODM's purpose was to qualify nodes for US SC-QAM 64-QAM activation & create a basic PNM process while ensuring service impacting situations may be identified and addressed.

The 2nd iteration, ODM+, expanded to include DS SC-QAM and OFDM KPI measurements including MER 2nd percentile, Rx power and SNR; with an objective to stabilize connectivity for modems and to score the health of a node. During testing several versions were evaluated to determine the accuracy by utilizing accessibility within a confusion matrix (Figure 1), each model was given different weights and total score thresholds.

		ODM CM Score	
		Good	Bad
Accessibility	Good ≥98%	**ODM CM Score aligns well** High accessibility with Low ODM CM score	**False-Positive** High accessibility with High ODM CM score
	Bad <98%	**False-Negative** Low accessibility with Low ODM CM score	**ODM CM Score aligns well** Low accessibility with Low ODM CM score

Figure 1- ODM Confusion Matrix

Through testing, KPI weights were adjusted to ensure ODM was able to capture actionable issues. Testing found that when overall score threshold is low and optimized weights are used (<75 score per modem), a decrease in both false-positives and false-negatives is observed. Figure 2 shows this decrease in false-positives with optimized weights.

A similar analysis was conducted for only false-negatives. False-negatives showed a lower percentage when the optimized weights were used. Figure 3 shows the decrease in false-negatives with optimized

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

weights. It is noted in both scenarios that the dramatic decrease/increase in accessibility as ODM score lowers indicates potentially other factors impairing accessibility in addition to the existing KPIs.

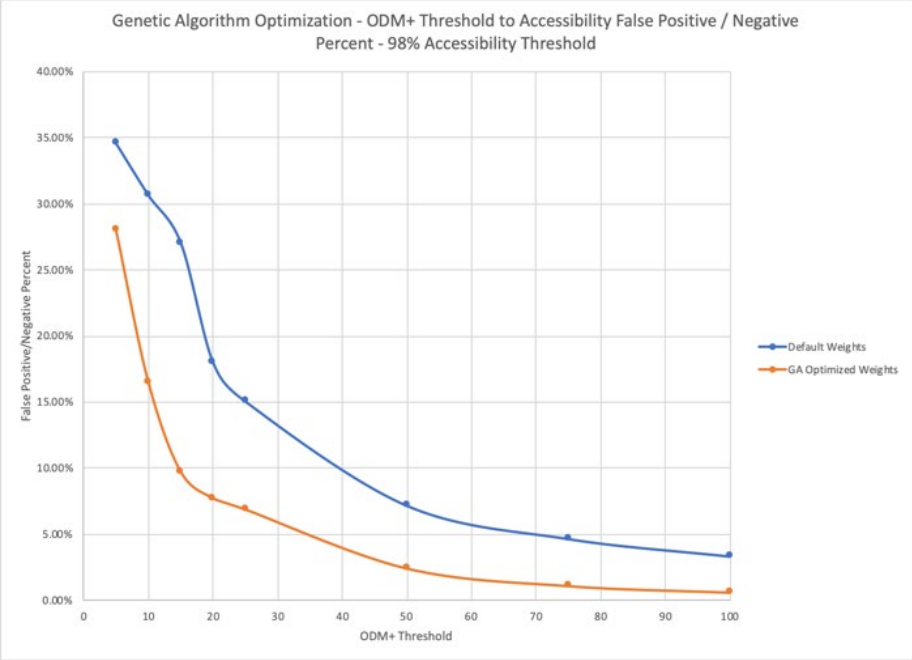


Figure 2- Decrease in false-positives with optimized weights

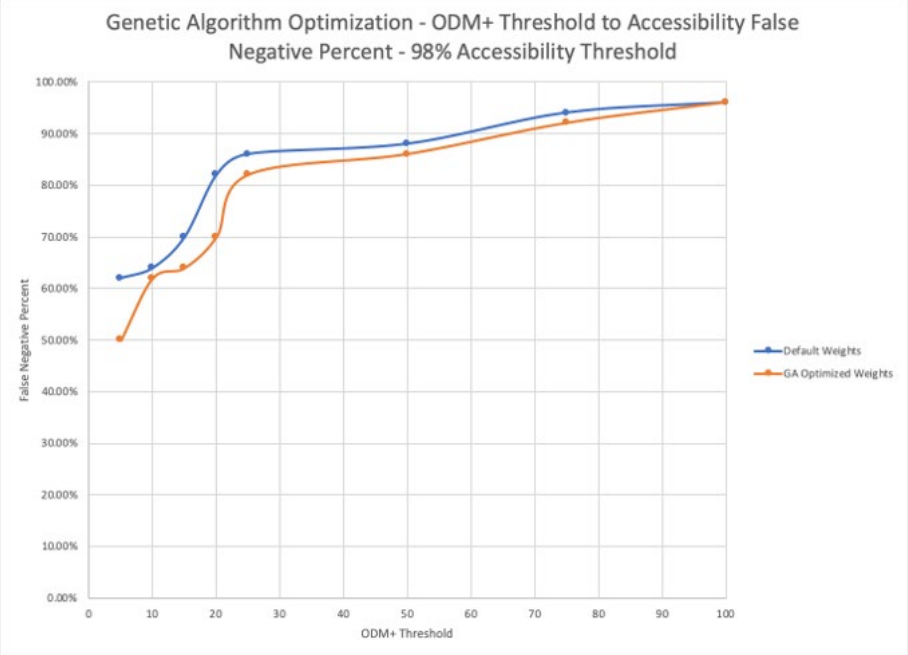


Figure 3- Decrease in false-negatives with optimized weights

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Network Lab Testing

5.1. Testing Tools and Methodology

A lab environment encompassing 24 DOCSIS 3.1 modems of various chipsets was utilized, each with 4 Internet Protocol television (IPTV) set top boxes to mimic production as closely as possible. At the CMTS, a noise generator was placed along with attenuator pads for adjusting transmit and receive levels. Separate testing was done for SNR, Tx power and Rx power levels. Each test was set for 15-minute intervals with increasing impairments until complete failure was observed across all modems. This lab testing generated the first set of thresholds used in ODM 3 development; these thresholds were further validated with production trials.

Production trials were performed with field technician assistance, these trials verified each impaired KPI within the algorithm denoted the point of corrective action required. Through continuous feedback, multiple sessions with field technicians were completed to improve the weighting and prioritization method.



Figure 4- Automation Lab

5.2. Characterization Method and Data Capture

Within the lab, multiple data sources were utilized to correlate customer experience to ODM3. Visual inspection was performed on video feeds during impairment threshold testing to replicate customer experience when using the service. To ensure accuracy of data and efficient collection, automation scripts were built to modify KPI levels and collect service metrics, this was visualized on internal reporting tools for analysis. Key infliction points observed in the visualizations were used to adjust weightings to further increase accuracy.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Analyzing the Algorithm

6.1. Developing a Modular Solution

Each individual KPI is collected throughout the day. The final days polling is then run through the ODM 3 algorithm to provide each nodes scoring. Currently, the daily averages of these metrics are being utilized, however, a methodology to use data with min/max values will be released by next year.

When a given metric is pulled into the algorithm, depending on its severity of impact, the ODM 3 algorithm will apply a score. The score is higher, depending on how severely the KPI breaches the thresholds and in some cases which frequency is being impacted.

Each KPI CALL (ex: US_TX0, US_TX1) scoring is aggregated into their respective KPI buckets: US SCORE, DS SCORE, OFDM SCORE and OFDMA SCORE. As visually indicated by the US Score sample below this summation provides the modems individual score and determines if a modem is passing or failing.

Sample:

$$US\ SCORE = \sum_{i=1}^{n} \frac{tx_score}{i} + \sum_{i=1}^{n} \frac{snr_score}{i} + \sum_{i=1}^{n} \frac{cer_score}{i}$$

Where:

tx_score = US TX = $\begin{cases} x, if\ TX < 30.5\ dBmV\ or\ TX > 52.5 \\ 0\ dBmV, otherwise \end{cases}$

snr_score = (US SNR0 = $\begin{cases} x, if\ SNR \geq 25\ dB\ or\ SNR \leq 28 \\ 0\ dB, otherwise \end{cases}$) +
(US SNR1 = $\begin{cases} x, if\ SNR < 25 \\ 0\ dB, otherwise \end{cases}$)

cer_score = (US CER0 = $\begin{cases} x, if\ CER \geq 0.0057\ AND\ CER \leq 0.01 \\ 0, otherwise \end{cases}$) +
(US CER1 = $\begin{cases} x,\ if\ CER > 0.01\ AND\ CER \leq 0.03 \\ 0, otherwise \end{cases}$) +
(US CER2 = $\begin{cases} x\ , if\ CER > 0.03\ AND\ CER \leq 0.1 \\ 0, otherwise \end{cases}$) +
(US CER3 = $\begin{cases} x,\ if\ CER \geq 0.1\ AND\ \leq 1.0 \\ 0, otherwise \end{cases}$)

This is a sample of one of the modular sections of the algorithm, the SC-QAM upstream component, where x is the proprietary variable for the weighted scoring per threshold (KPI CALL). Each modular component is calculated as the data is iterated.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

$$cm\ score = \sum_{i=1}^{n} \frac{us_score}{i} + \sum_{i=1}^{n} \frac{ds_score}{i} + \sum_{i=1}^{n} \frac{ofdm_score}{i} + \sum_{i=1}^{n} \frac{ofdma_score}{i}$$

Based on the above sample set, each bucket of calculations is aggregated to derive the CM scoring.

$$ODM = \left(\frac{\sum_{i=1}^{n} pass}{\sum_{i=1}^{n} pass + \sum_{i=1}^{n} fail} \right) x\ 100$$

The final ODM 3 score is a percentage of modems passing their individual modem score. Wherein 90% or greater is a passing node.

To summarise, each upstream and downstream channel is scored for each KPI, and the sum of the KPI scores represents the customer's overall CM score. The CM score for the customer can be used to determine single customer impairments, direct service trucks as required and inform frontline of any area issues detected by the ODM algorithm.

If the CM score is greater than the 75 threshold, that individual modem is failing. If passing customers is greater than 90%, the node is classified as passing and no action is required.

7. Implementation of Algorithm and Cloud-Based Systems

7.1. Proof-of-Concept (PoC) Development

A dashboard was built using production modem IPDR data within the company's internal visualization and automation platform. This dashboard assesses and validates each metric against the defined scoring values determined by lab testing for each KPI CALL. This PoC was shared with maintenance team members to ensure results as indicated were in line with their findings and that each indication of failure noted was actionable in the field.

7.2. Cloud Solution

Previous versions of ODM utilized an on prem system of automation. As cloud services became readily available, its efficient data processing and automation capabilities were utilized.

Applying the standards set for each KPI scoring for thousands of modems must be done in an efficient manner to ensure optimal load on the cluster, as such the code was written in modules to ensure efficiency, ease of updating and rapid validation. Sections were separated into four modules, downstream SC-QAM, upstream SC-QAM, OFDM and OFDMA. Each can be run at the same time, separately or in any order necessary. After the modules are completed the final merge and calculation of an ODM score can begin. This entire process takes on average ~130 seconds for a full day's data.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table +

	MAC_ADDRESS	US_SCORE	TX	US_SNR	US_CER	PTIME	CHANNEL_NAME	FREQUENCY	CER_AVG	SNR_AVG	TX_AVG
1	[redacted]	0	0	0	0	2023-07-11	Logical Upstream Channel 10/11.2/0	32300	0	36.4833	43.4917
2	[redacted]	0	0	0	0	2023-07-11	Logical Upstream Channel 11/10.1/0	25900	0.000508696	33.3043	42.787
3	[redacted]	0	0	0	0	2023-07-11	Logical Upstream Channel 9/9.0/0	21100	0	40.7458	46.9
4	[redacted]	0	0	0	0	2023-07-11	Logical Upstream Channel 11/3.3/0	38700	0	37.6833	44.575
5	[redacted]	0	0	0	0	2023-07-11	Logical Upstream Channel 10/15.1/0	25900	0	42.1	42
6	[redacted]	0	0	0	0	2023-07-11	Logical Upstream Channel 10/11.2/0	32300	0	34.7417	39
7	[redacted]	0	0	0	0	2023-07-11	Logical Upstream Channel 12/8.3/0	38700	0	42.1	46.8

8,393 rows | Truncated data | 13.11 seconds runtime Refreshed 12 days ago

Figure 5 - This is a sample of US Score

8. Measure of Success

8.1. Measuring Success of ODM

Measuring ODM's success is both qualitative feedback from technicians and quantitively assessing impairments found in field. Since Jan 1, 2023, approximately 600 IMTs were generated. The IMTs have had a 96% success rate in finding impairments out in field. The incident tickets generated return a variety of impairments ranging in nature from noise-related to hardware failure.

8.2. Production Example

Provided is an example of a submitted ticket for ODM 3, the initial failing score was 62% (failing), within field a hardware fault was identified and resolved increasing the ODM score to 96% (passing). After the impact is resolved the node is monitored for additional days to ensure issues are corrected and the node is stable. In addition to the example provided below, other ODM 3 tickets have identified and resolved: noise floor rolling impedance, power supply impairments, faulty drop impairments as well as high transmit/receive impacts requiring additional filters all within a single day.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

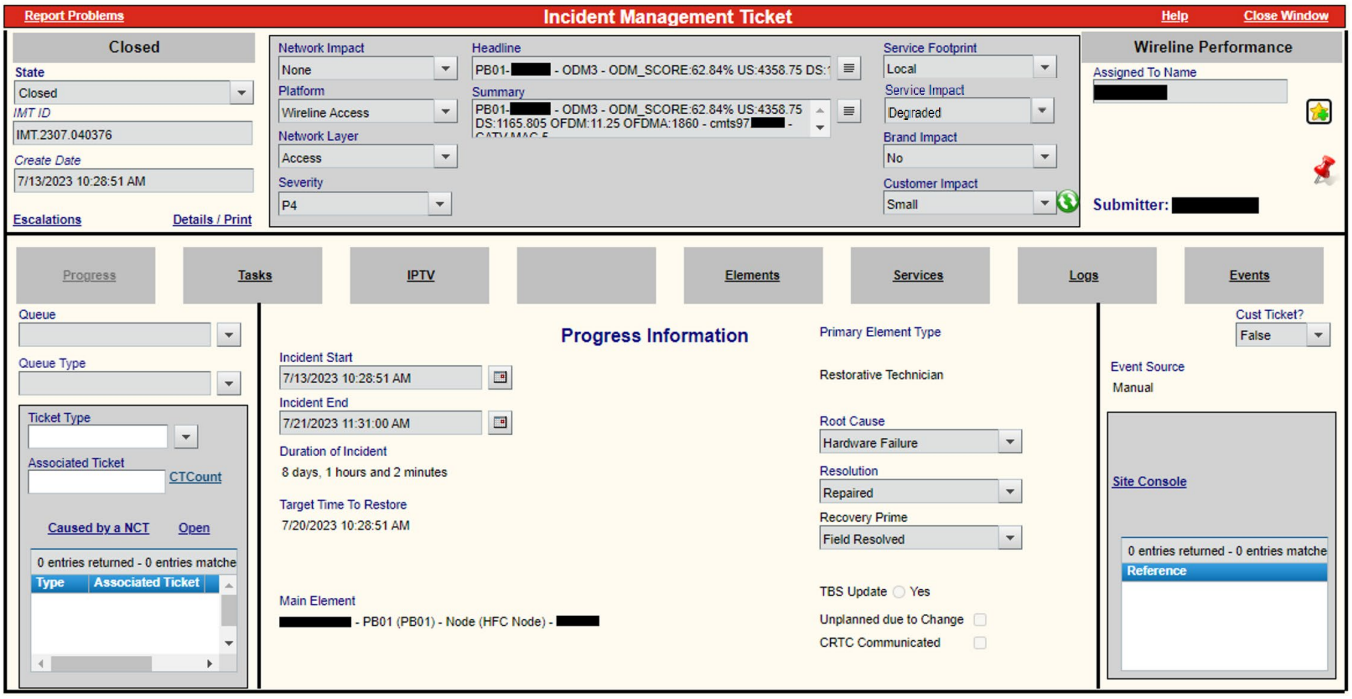


Figure 6- Sample Incident ticket with hardware failure

	PTIME	ODM_SCORE	MDMs_PASSING	MDMs_FAILING	US_SCORE	DS_SCORE	OFDM_SCORE	OFDMA_SCORE
1	2023-07-10	62.99212598425197	80	47	4381.25	1189.77	22.5	1740
2	2023-07-11	62.698412698412696	79	47	4336.25	1141.8400000000001	0	1980
3	2023-07-12	66.66666666666666	82	41	4050	1160.47	0	1548.75
4	2023-07-13	62.295081967213115	76	46	4212.5	1037.65	0	1822.5
5	2023-07-14	93.7007874015748	119	8	1455	884.2100000000002	0	858.75
6	2023-07-15	96.8503937007874	123	4	56.25	808.7299999999999	0	75
7	2023-07-16	96.8503937007874	123	4	75	694.4399999999999	7.01	75
8	2023-07-17	96.06299212598425	122	5	75	747.55	0	150
9	2023-07-18	95.2755905511811	121	6	56.25	680.4000000000001	0	225
10	2023-07-19	95.23809523809523	120	6	93.75	702.2400000000001	0	225
11	2023-07-20	96.03174603174604	121	5	131.25	611.2900000000001	0	150

Figure 7 - Before/After ODM Score Improvement

9. Conclusion

9.1. Summary of Key Findings

ODM 3 has proven its success by consistently identifying actionable network pain points. Since the beginning of 2023, ODM has had a 96% success rate across the almost 600 IMTs generated. Technicians can consistently find, triage and repair impairments. ODM's unique ability to find and determine key impairments has offered technicians ease in solutioning problems found in the field.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

9.2. Future Improvements and Areas of Further Research

Weather data integration into ODM3+ (next version of ODM) is one of the next steps to avoid degraded service due to weather fluctuations. Continued testing and analysis on specific weather patterns and impacts to the cable plant will allow for better issue identification in the ODM3+ model, which will further increase accuracy and reduce the 4% "no fault found" tickets. Additional KPI inclusion such as partial service and in-channel frequency response (ICFR) will be integrated into future releases. The process will include continuous feedback from maintenance technicians for new enhancements and accuracy improvement.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

CER	correctable error ratio
CM	cable modem
CMTS	cable modem termination system
dB	decibel
dBmV	decibel millivolt
DOCSIS	Data-Over-Cable Service Interface Specification
DS	downstream
HFC	hybrid fiber-coaxial
ICFR	in-channel frequency response
IMT	incident management ticket
IPTV	Internet Protocol television
KPI	key performance indicator
KPI CALL	The definition for each individual KPI (includes multi-threshold KPIs) to apply a scoring based on each channels value
MER	modulation error ratio
MIB	management information base
ODM	overall DOCSIS metric
OFDM	orthogonal frequency-division multiplexing
OFDMA	orthogonal frequency-division multiple access
PoC	proof-of-concept
Rx	receive
RDK-V	reference design kit for video
SC-QAM	single channel quadrature amplitude modulation
SCTE	Society of Cable Telecommunications Engineers
SNR	signal-to-noise ratio
SNMP	Simple Network Management Protocol
Tx	transmit
US	tpstream

Bibliography & References

1. CableLabs. “DOCSIS Proactive Network Maintenance Using Pre-Equalization ... - Cablelabs.” Https://Community.Cablelabs.Com, 23 June 2011, community.cablelabs.com/wiki/plugins/servlet/cablelabs/alfresco/download?id=fe685976-3dd1-44c6-91c3-a16d7031d79f.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Using PNM Tools to Identify Misaligned Amplifiers

A Technical Paper prepared for SCTE by

Mike O'Dell
Distinguished Engineer
Comcast
michael_odell@cable.comcast.com

Westley Darby
Engineer, Quality Assurance
Comcast
westley_darby@cable.comcast.com

Jason Rupe
Distinguished Technologist
CableLabs
j.rupe@cablelabs.com

Andy Martushev
Distinguished Engineer
Comcast
andy_martushev@cable.comcast.com

Alex Zayvelev
Senior Access Engineer
Comcast
alexander_zayvelev@comcast.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Background 4
   - 2.1. Inspiration 4
   - 2.2. PNM 5
   - 2.3. Amplifier Characterization 5
   - 2.4. NPR 6
3. MER Curve 7
   - 3.1. Building the Curve 7
   - 3.2. Symptoms of Amplifier Non-Linearity 9
4. Lab Testing Setup 11
5. PNM Tooling 14
   - 5.1. Tooling Considerations 14
   - 5.2. Visualizing the Data 15
   - 5.3. Analyzing the Data 16
   - 5.4. Practical Application 20
     - 5.4.1. Pre-Enablement 20
     - 5.4.2. Post Enablement 21
6. Conclusion 21

Abbreviations 22

Bibliography & References 23

List of Figures

Title **Page Number**

Figure 1- PNM Tools for Amplifier Alignment 4
Figure 2- Typical NPR Curve Format 6
Figure 3- MER curve of Amp 1, using Channel Plan A 8
Figure 4- MER curve of Amp 1, using Channel Plan B 9
Figure 5- Typical Amplifier Schematic (Two Downstream Gain Stages Shown) 9
Figure 6- Misaligned Amplifier 10
Figure 7- Full Band Capture from CPE Connected to a Non-Linear System 11
Figure 8- Same Amplifier When Properly Aligned 11
Figure 9- Diagram of Test System 12
Figure 10- 750 MHz and 2G Channel Plans Used During Testing 13
Figure 11- Total Composite Power Measurement of Channel Plan A 13
Figure 12- Total Composite Power Measurement of Channel Plan B 14
Figure 13- Sample Spectrum (FBC) Bin Data With Averaging 15
Figure 14- Identifying Locations (Red Devices) Experiencing Elevated Noise Floor 15
Figure 15- Diplex Filter Transition Region of a Device in Optimal Conditions 16
Figure 16- FBC Example of a Device With Optimal Amplifier Setup 17
Figure 17- Example of Noise Delta (dn) Calculation 17
Figure 18- Diplex Filter Transition Region of a Device in the Noise Region of the MER Curve 18
Figure 19- Example of a Misaligned Amplifier Operating in the Noise Region 18

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 20- Diplex Filter Transition Region of a Device in the Clipping Region of the MER Curve 19
Figure 21- Example of a Misaligned Amplifier in Compression 19
Figure 22- Example of Distortion Noise in the Diplex Filter Transition Region 20

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Broadband networks widely use Proactive Network Maintenance (PNM) tools to identify impairments actively. While PNM has successfully identified upstream cable and common plant impairments, this paper explores whether PNM telemetries can also identify improperly aligned amplifiers that cause low Modulation Error Ratio (MER) and other performance issues. We must also identify some technical considerations necessary to create PNM tooling that can correlate plant impairments to amplifier misalignment, **Figure 1**.

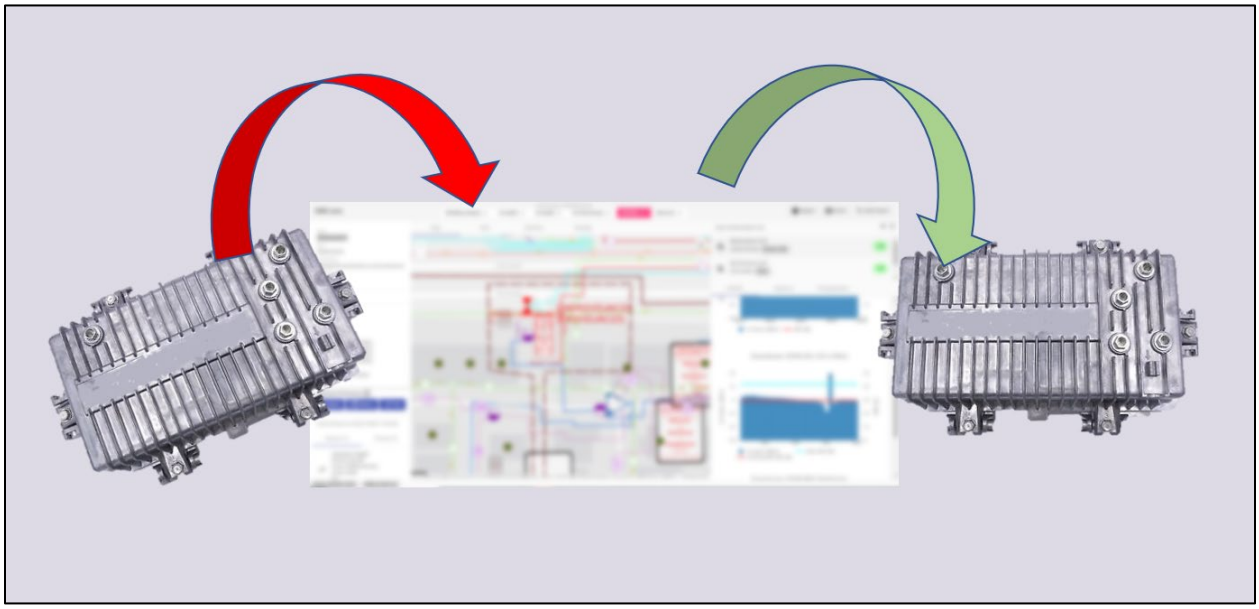


Figure 1- PNM Tools for Amplifier Alignment

2. Background

2.1. Inspiration

There are challenges to constructing and enabling a modern CATV network at scale. It is an effort that operators worldwide are undertaking and facing many of the same issues. The evolution to a DOCSIS 4.0, or 10G Network, requires a transformative process that touches many of the core components of the network. It may require Cable Modem Termination System (CMTS) replacements, core network upgrades, optical node replacements, and other plant components like amplifiers and passives. This work requires a workforce that can be less tenured and experienced in the nuances of amplifier setup and operation but, by necessity, is focused on splicing and activation volume. It also requires generating and receiving new signal types and frequencies with which we don't have history or expertise.

Some timing considerations became apparent in our 10G Network rollout process. As mentioned, the primary stages of network development and deployment have been ongoing for a year or more. CMTS transitions that happened a year ago, or more have yet to generate a portion of their full potential where Orthogonal Frequency Division Multiplexing (OFDM) and Orthogonal Frequency Division Multiple Access (OFDMA) signals are concerned. Likewise, amplifiers that are mid-split capable and have been in place for a year or more still carry sub-split, 750 MHz channel plans. It was this exact scenario that served as the inspiration for this effort. With CMTS upgrades, Remote-Phy node installations, and amplifier

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

upgrades and activations, the plan was to enable the spectrum. When the fateful event happened, it wasn't rainbows and confetti. We created a service outage.

Technicians scrambled, and difficult conversations were had. Amplifier misalignment during installation and activation was ultimately determined. Why didn't we realize it at the time? Why did all of our monitoring tools work fine? Why weren't there service calls or other indicators to identify the problem? How were the customer services working before? All valid questions. This was the beginning of a long journey of discovery into amplifier characterization and setup, training, communications, business partner relationships, testing, tool development, and, hopefully, a smoother path to future spectrum enablements.

This paper recounts many things considered part of that learning journey- from developing, deploying, and using PNM tools, testing, measuring, and characterizing amplifiers to approaching future spectrum enablements.

2.2. PNM

It has been over a decade since we gained access to spectrum data from cable modems. In PNM, the focus is often on the signal in the spectrum capture and the search for spectrum impairments impacting those signals. We collect downstream statistics based on the RF signal, and our analysis of the various PNM tests also focuses on the RF signal quality. Utilizing MER values from cable modems to analyze the noise floor and examining that against the noise floor of the device, we can infer much about the non-linearity of the noise floor, which could indicate an amplifier misalignment issue.

Further, by examining the Modulation Error Ratio per carrier or Receive MER (RxMER) per subcarrier, we can utilize these powerful downstream measurements to determine if that noise is uniform across the spectrum range, potentially indicating more precisely what amplifier alignment issue is present.

By adding amplifier alignment troubleshooting to the PNM set of tools, we can make maintenance operations more efficient. Managing PNM as an opportunity to manage network health ahead of service impact, planned maintenance can be optimized better by combining issues and optimizing travel time and technician effort.

As described by Brady Volpe: "No longer is it [PNM] seen as a shiny gimmick or a novelty to detect the poltergeist in the network before it goes bump in the night. It is a go-to network maintenance tool with the added benefits of workforce optimization" [1].

2.3. Amplifier Characterization

While we desire our signal distribution networks to operate as linearly as possible, we expect and have known instances where non-linearities can occur. CATV amplifiers, for example, while an essential component in defining the capacity and performance of a Hybrid Fiber Coax (HFC) network, can be a source of such non-linearities. Understanding the behaviors and performance characteristics of these necessary devices can help determine when and where their misalignment could result in impairments or, at the very least, act as inhibitors to the optimal performance of the amplifier chain or cascade.

Historically, characterizing or measuring amplifier performance used categories like Carrier-to-Noise Ratio (CNR), Composite Second Order (CSO) distortions, or Composite Triple Beat (CTB). While these have been useful and meaningful when delivering analog video signals over long cascades of amplifiers as a measure of video signal quality, many of today's networks strictly carry digitally modulated data channels, with modulation orders of up to 4096 Quadrature Amplitude Modulation (QAM). Those historical measures aren't invalidated, but the evolution of DOCSIS and data delivery network technology have resulted in

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

much wider broadband carriage systems. Since coaxial cables are still a fundamental and primary transmission medium and have a strong frequency-dependent loss [2], the resulting increased gain of amplifiers can present instances to significantly and negatively impact the designed signals that those amplifiers carry. Understanding amplifier performance using other measures, including, but not limited to, Carrier-to-Thermal noise (CTN), Carrier-to-Composite noise (CCN), Carrier-to-Intermodulation noise (CIN), Bit Error Rate (BER), and MER can be helpful in understanding and even predicting amplifier behavior across known Total Composite Power (TCP) ranges [3].

2.4. NPR

Using Noise Power Ratio (NPR) measurements to determine network performance is not new. NPR has been the standard performance measurement for characterizing upstream analog optical transport links for many years.

The ANSI/SCTE 119 (2018) Standard defines that "NPR is a test method that examines the amount of noise and intermodulation distortion in a channel. A test signal, comprised of flat Gaussian noise band limited to the frequency range of interest and with a narrow band (channel) of the noise deleted by a notch filter or other means, is injected into the Device Under Test (DUT). The NPR is measured at the output of the DUT as the test signal is swept across a power range". [4]

When charting NPR curves, **Figure 2**, typically, the total input or composite power is on the horizontal axis, and the noise power ratio is on the vertical axis:

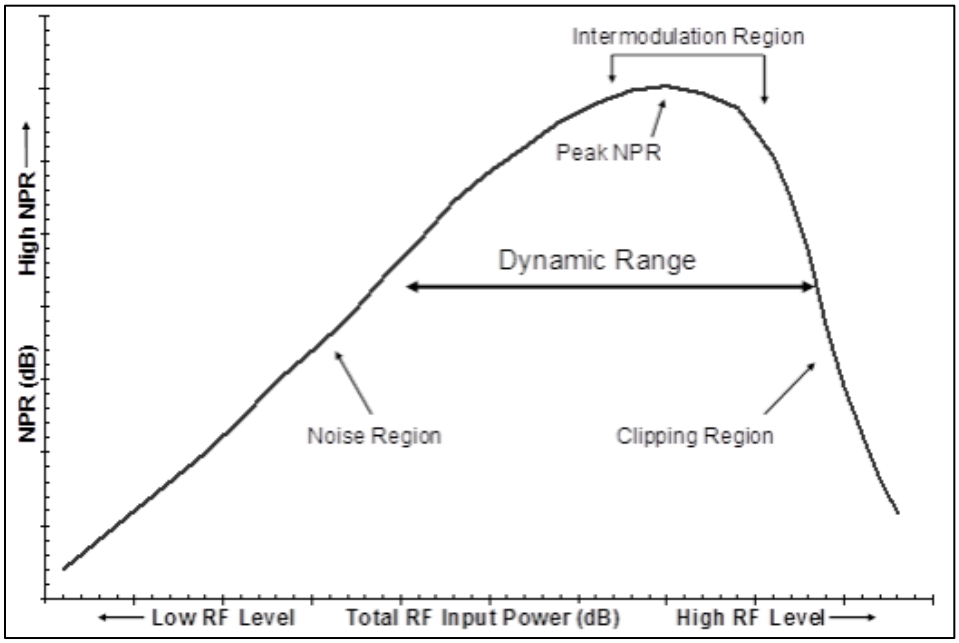


Figure 2- Typical NPR Curve Format

Another way to think of this chart is with a signal quality measure on the vertical axis, like MER. The signal quality improves as the total RF input power increases through the noise region. The peak signal quality potential exists in the intermodulation or transition region at the curve's apex. As the total RF input power increases, the signal quality declines rapidly through the clipping region, where the amplifier's performance becomes non-linear.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

At the extreme lower end of the noise region, the system's power is low and has very little differentiation from the noise. Signals in this region, when subject to high gain, result in high amplitude signals of poor quality because the amplified noise also has high amplitude. Signals in the intermodulation or transition region continue to have higher quality as their relative relationship to the system's noise increases. Past the peak NPR point, continuing to increase the signal into the clipping region while still "separating" from the noise floor results in the system becoming non-linear and generating intermodulation distortions. In a system deploying only digitally modulated channels, these distortions can appear as an elevation in the system's noise, even though they may be second or third-order distortion products.

3. MER Curve

Similarly, specific amplifiers can be tested and characterized for performance by introducing high-quality QAM signals across a broad range of frequencies and input powers to create an MER curve. This is a distinctly different methodology than that used in creating an NPR curve, but the results are similarly informative of the relationship between power, noise, and signal quality.

This method of characterizing amplifiers can help quantify how a particular amplifier behaves when aligned according to regionally standardized input and output design levels and relative to the output signal quality of the optical nodes deployed. Varying numbers of channels can create this MER curve.

Using a power per 6 MHz channel (amplifier input) power measurement on the horizontal axis to graph the (amplifier output) MER curve according to the desired input and output channel or channels directly relates to the designed output levels of the amplifier determined by the tap design deployed. This method makes it easy to determine the optimal amplifier input or drive level required to achieve the best amplifier output signal quality. That peak MER point is also the clipping point, or the point past which any additional input power results in a degraded performance in the output signals of the amplifier. Refer to this point as the *Peak*.

This method is easily explainable to field personnel; however, it must be understood that these results are directly related to the total composite power or the number of active channels present in the lineup at the time of measurement.

It is important to note that the channels tested in an MER curve may reach their peak or clipping point at different power levels. A singular clipping point value would be the lower of the available values.

The data derived from developing an MER curve this way is fundamental to our hypothesis that new configurations and uses of PNM tools can identify amplifiers operating in a non-linear state.

3.1. Building the Curve

Having established a test system where there is a known and common channel plan, Channel Plan A, used to create an MER curve for a specific amplifier model, Amp 1, **Figure 3**, we can determine the maximum input level on our input and output reference channels at Amp 1 to achieve the peak MER figure in the transition region of the curve.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

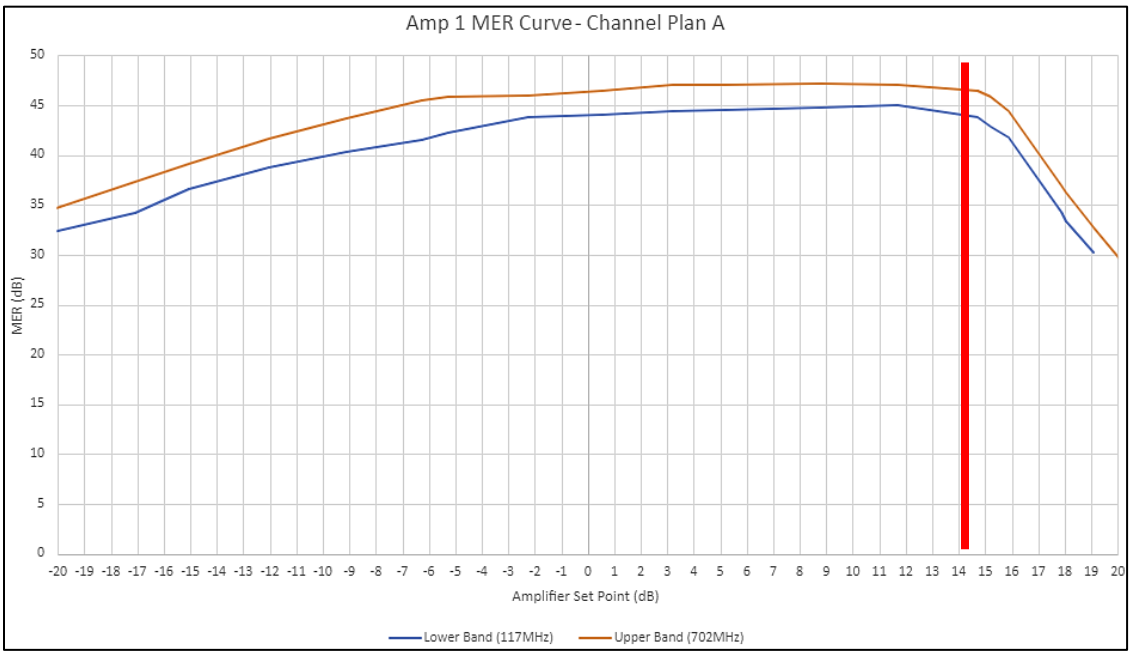


Figure 3- MER curve of Amp 1, using Channel Plan A

As the constant need for operational bandwidth dictates that operators continue to expand their downstream channel plans to add data capacity, it is reasonable to test the performance of the amplifier deployed into that system with a representative channel load. For this paper, we'll call this Channel Plan B. By measuring each channel plan's TCP at the input to the amplifier's first gain stage across the wide range of test values, we can calculate the TCP delta between the two plans. This additional power introduced during spectrum turnup (enablement) is useful in practical applications.

With the additional channels on Channel lineup B, the peak of this MER curve changes due to the additional composite power introduced into the input of the amplifier, **Figure 4**.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

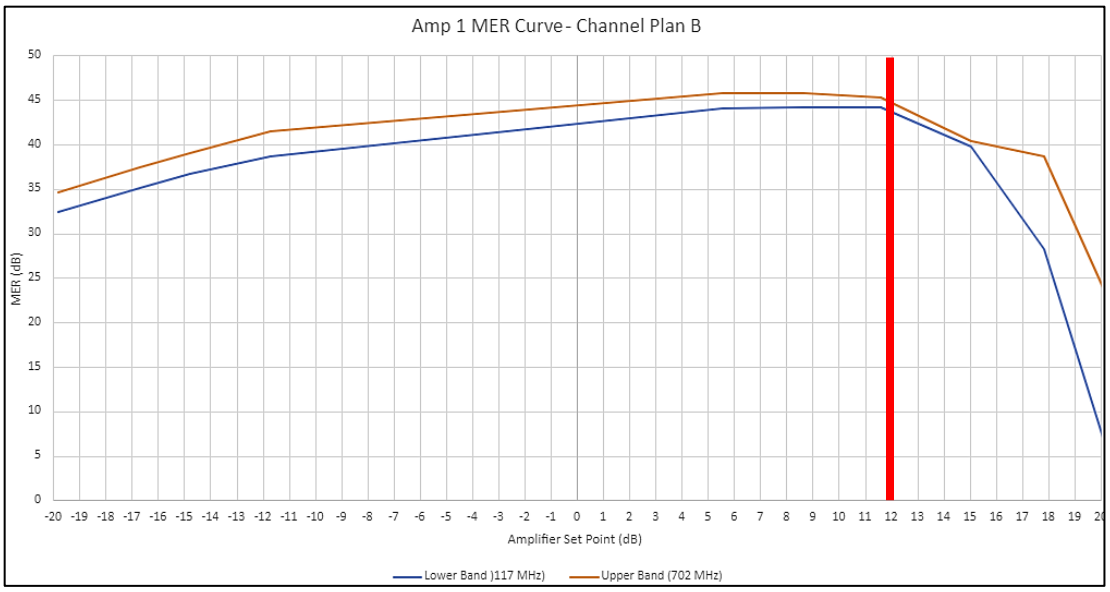


Figure 4- MER curve of Amp 1, using Channel Plan B

This delta would also indicate the amount of headroom necessary in the 750 MHz MER curve to deploy additional channels without introducing distortions into the system. In essence, how far left of the peak do we need to be to enable more spectrum successfully?

3.2. Symptoms of Amplifier Non-Linearity

Amplifiers in today's networks are of high quality and reliability. However, they require higher gain to support today's wider bandwidths and the frequency-specific loss figures inherent in coaxial cables at higher frequencies. Some amplifiers may have as many as four (4) distinct gain "stages" in the downstream amplification path. With this increased gain comes the risk of improper alignment. There are typically multiple points in the forward path to attenuate and equalize the input signals to ensure proper power or drive level at the various amplifier gain stages, **Figure 5**.

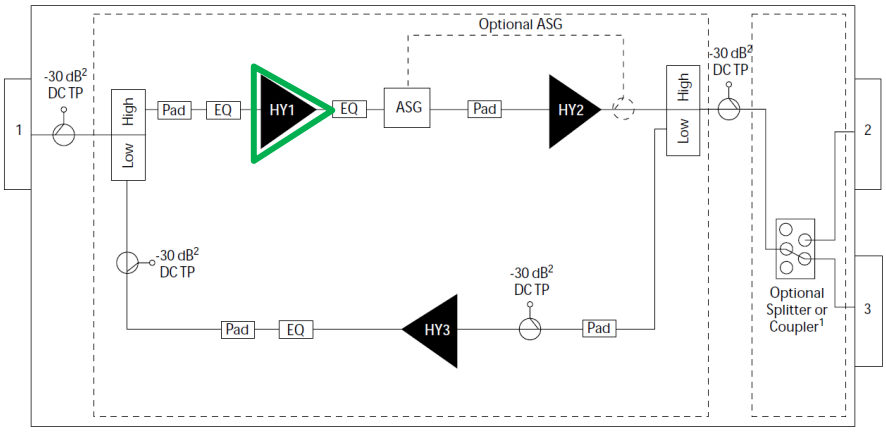


Figure 5- Typical Amplifier Schematic (Two Downstream Gain Stages Shown)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Arguably, the most important of these points in the amplification path is the drive level into the first gain stage of the amplifier, often located after the input attenuator and input equalizer locations in the amplifier station. Today's complex outside plant designs can result in an extensive range of signal levels into the input of an amplifier station. There are several common scenarios where an amplifier is aligned improperly, resulting in an input drive level in the noise region of the MER curve or the clipping region. Both extremes can result in poor overall system performance and generate symptoms in the signal and the noise. These symptoms can include poor MER and or RxMER, high BER, or indications of elevation in the noise floor. However, understanding the symptoms available from the Customer Premise Equipment (CPE) connected to these systems can provide some clues as to the precise cause of the poor performance.

In the following field example, customers in the network were experiencing poor service because of very low signal quality, although the RF levels were consistent with design specifications. An investigation by a network maintenance technician determined that the input attenuator in a high-gain amplifier was too low, resulting in a high drive level at the amplifier's first gain stage. This placed the first amplifier gain stage in the clipping region and created high intermodulation distortions in the amplifier station, **Figure 6**. As noticed in the channel scan taken by the technician with a field strength meter at the amplifier's output test point, not only was there low MER (31 dB), but the noise floor in the vacant portions of the spectrum was extremely elevated.



Figure 6- Misaligned Amplifier

When we observed the full band capture (FBC) data from CPE fed by this amplifier, **Figure 7**, we also observed the elevation in the vacant portions of the spectrum, similar to the example below. PNM tools' FBC clearly illustrates the apparent noise of the system created by the amplifier operating in a non-linear state.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

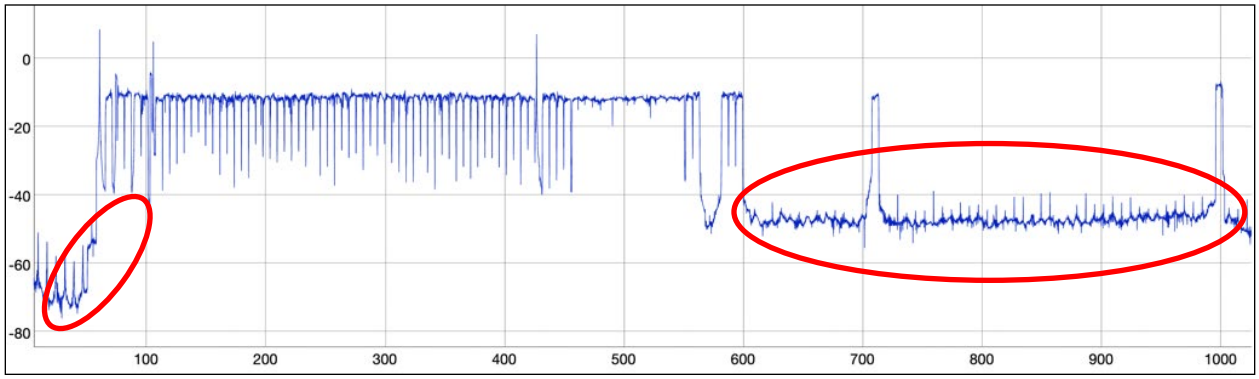


Figure 7- Full Band Capture from CPE Connected to a Non-Linear System

When the technician re-aligned the amplifier by moving 10 dB of attenuation from the output attenuator location to the input attenuator location, the drive level of the station moved from the clipping region of the MER curve into the more linear transition portion of the curve. This allowed the amplifier to operate at a much higher level of performance. While the total station attenuation remained the same, the 10 dB "shift" of the attenuation resulted in a 13 dB increase in MER, illustrating the non-linear impact of amplifiers operating in the clipping region. We also saw greater than 10 dB of reduction in the elevation of noise in the vacant portions of the channel plan. **Figure 8**.



Figure 8- Same Amplifier When Properly Aligned

These were the initial indications that PNM tools could identify impairments caused by amplifier misalignment.

4. Lab Testing Setup

A critical part of this effort is constructing a testing environment that closely aligns with the architecture deployed into the network, has the flexibility to deploy various channel plans, and has access to highly precise test instruments to measure and record results. Comcast is fortunate to have such a facility, staffed with talented engineers, in its Downingtown, Pennsylvania facility.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The network created for this initiative, **Figure 9**, required access to a video core, Remote-Phy nodes, amplifiers, and lab-grade test equipment.

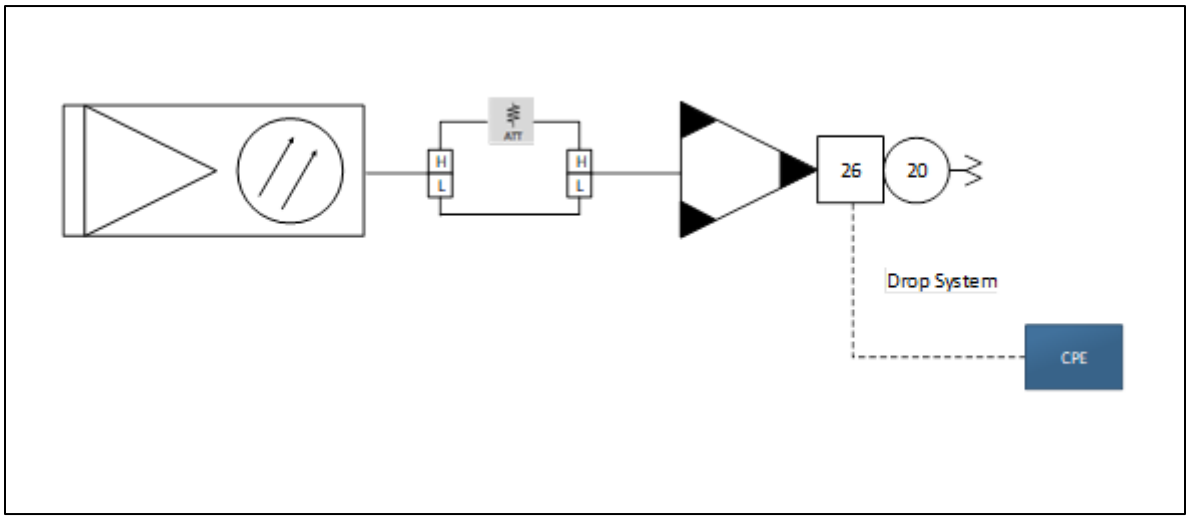


Figure 9- Diagram of Test System

A Remote-Phy node was activated and fed into a mid-split diplex filter. The filter allowed us to vary the input levels to the amplifier without impacting the upstream transmit levels of the CPE connected to the system. The high side of the diplex filter fed through a variable attenuator before being recombined with the upstream in a second diplex filter. The common port of the second diplexer connected to the input of a mid-split 1.2 GHz bridger amplifier. The main output of the bridger amplifier fed a series of tap ports that distributed a signal to the CPE used to collect the PNM data for the trial.

The drop system's design matched the upstream and downstream signals consistent with levels typical to our subscribers across the enterprise. The targets at the CPE were -5 dBmV downstream receive level on the DOCSIS channels and an upstream transmit value of 50 dBmV. The customer premise equipment chosen for the test is a cross-section of the equipment actively deployed throughout our network- a representative sample of video set-top boxes and a selection of DOCSIS gateways. All CPE had current production versions of firmware installed. In the test system, we accomplished SNMP polling by logging into the devices using the IPv6 address to acquire real-time full-band captures. Data files from the devices were polled, cataloged, and stored for visualization and analysis.

The test setup generated two distinct channel plans for the exercise, **Figure 10**.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

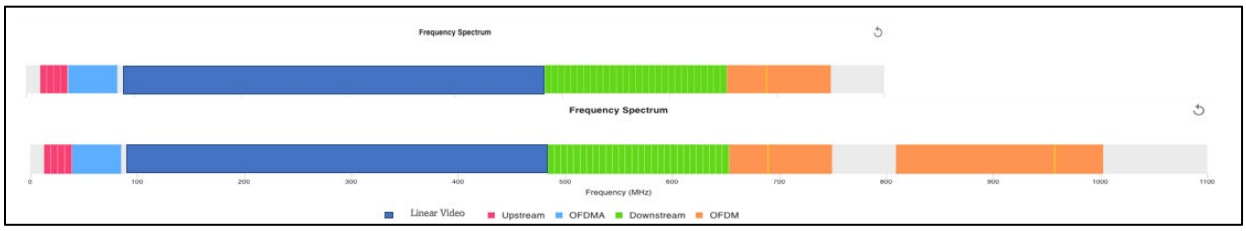


Figure 10- 750 MHz and 2G Channel Plans Used During Testing

The first plan is categorized as a 750 MHz channel plan consistent with what is currently deployed in production in our 750 MHz band limited systems. That plan is recreated for our Remote-Phy nodes' consistency before replacing all the sub-split amplifiers and passives in the network with mid-split capable amplifiers and passives. The second channel plan deployed in this test is consistent with production plans capable of delivering 2 Gbps and is loosely identified as the 2G downstream channel plan. The amplifier selected for this exercise is the default 1.2 GHz mid-split amplifier platform deployed across Comcast. The upstream and downstream levels of the amplifier are representative of a large number of systems across Comcast. Slight variations exist in production based on the tap designs used during the 750 MHz rebuilds, but the levels used are generally accepted as standard.

The test environment was provisioned and set to levels accepted as Comcast's optimal alignment and configuration using the 750 MHz channel plan. Levels and TCP were measured (**Figure 11**, **Figure 12**), validated, and recorded as a baseline configuration. CPE were queried for telemetry, cataloged, and stored at the baseline configuration.

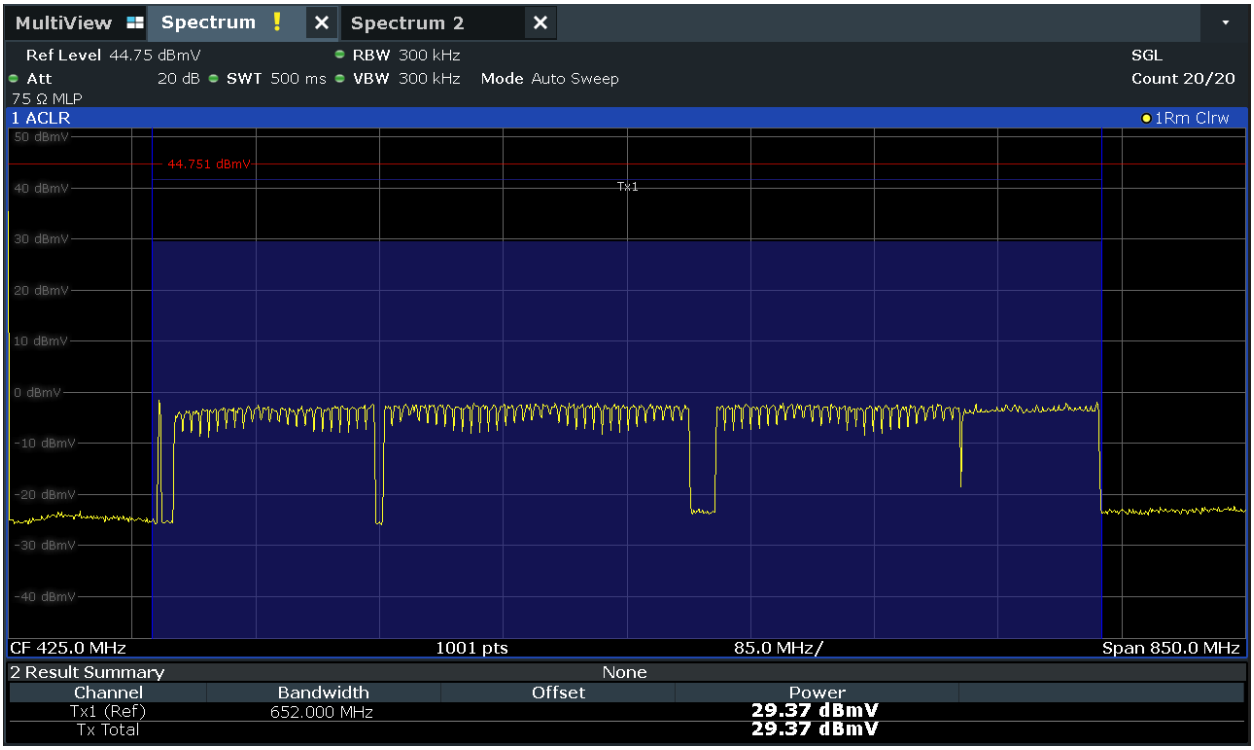


Figure 11- Total Composite Power Measurement of Channel Plan A

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

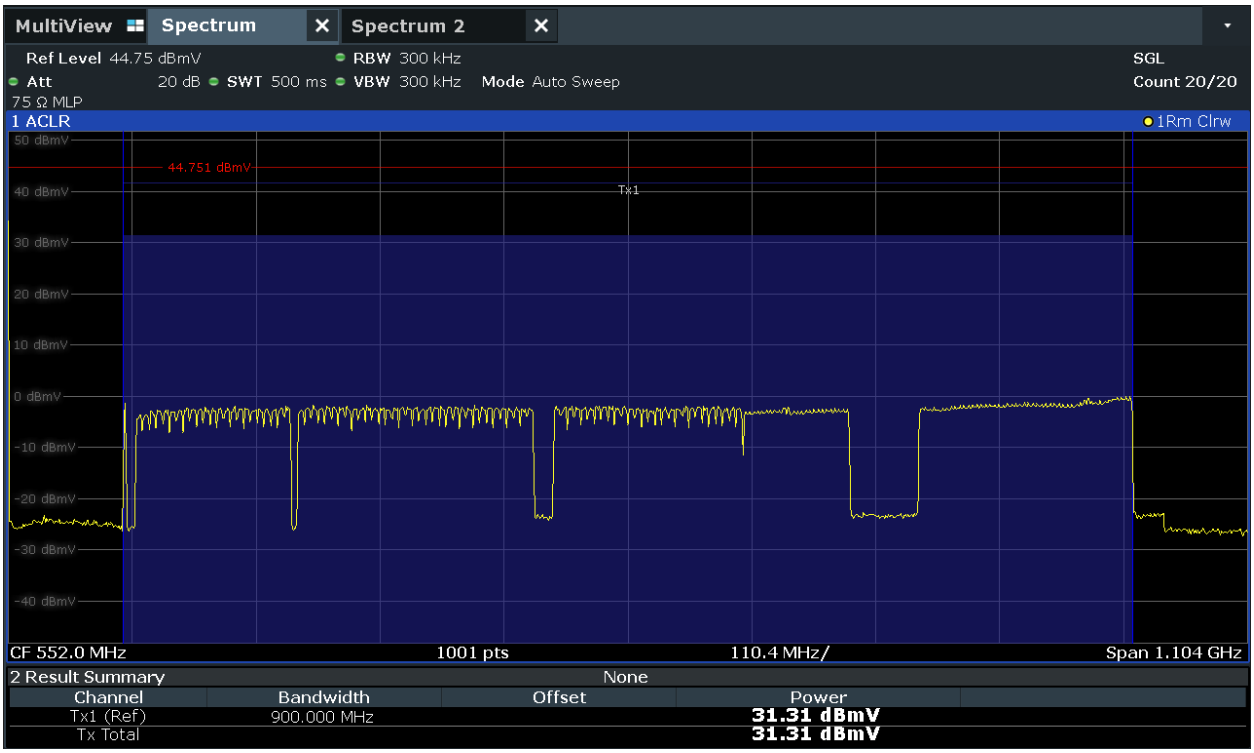


Figure 12- Total Composite Power Measurement of Channel Plan B

The forward input levels to the amplifier were then adjusted using the in-line variable attenuator described in the test setup description across a wide range of values. At each adjustment point, the output of the amplifier was adjusted to the design output level and tilt using the output attenuator locations. The dynamic gain value of the amplifier resulted in a wide range of test scenarios, ending when the amplifier output levels could no longer be achieved or the devices could not achieve a DOCSIS lock. In this fashion, the amplifier's output level and the CPE's upstream and downstream levels remained consistent across the entire range of test values. The variable in each instance was the input or drive level to the first gain stage of the amplifier station. Various data points and telemetry measurements were made, cataloged, and stored at each incremental adjustment level. When all plausible test scenarios were exhausted, the node was configured with the 2G channel plan, the equipment was re-baselined and verified, and the process was repeated. The stored data can then be used to populate tools to visualize the network's performance at various points in the architecture.

5. PNM Tooling

5.1. Tooling Considerations

Building a field tech-facing tooling system that helps alarm and identify distorted amplifiers starts with data. The crucial first step is collecting and storing PNM full band spectrum bin data, **Figure 13**. When collecting FBC data from CPE equipment, use on-chip averaging if possible; otherwise, collect multiple samples, choose a number between 4 and 16, and average the data server side before storing. When comparing samples and determining severity, capturing device MER and RxMER from DOCSIS SC-QAMs and OFDM channels is useful. Storing the data by node segment simplifies data storage and access. Consider storing metadata such as timestamp, mac address, node, CMTS, latitude, and longitude.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

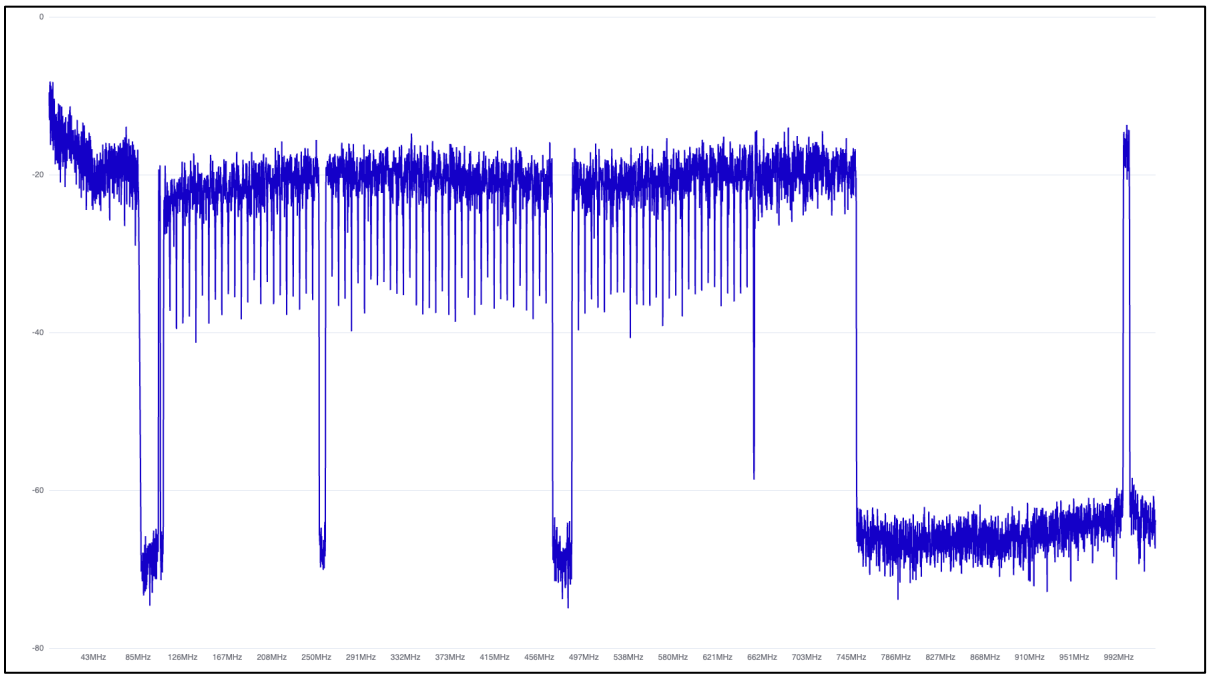


Figure 13- Sample Spectrum (FBC) Bin Data With Averaging

Mapping spectrum data with color-coordinated locations helps identify at which amplifier to start troubleshooting, **Figure 14**.

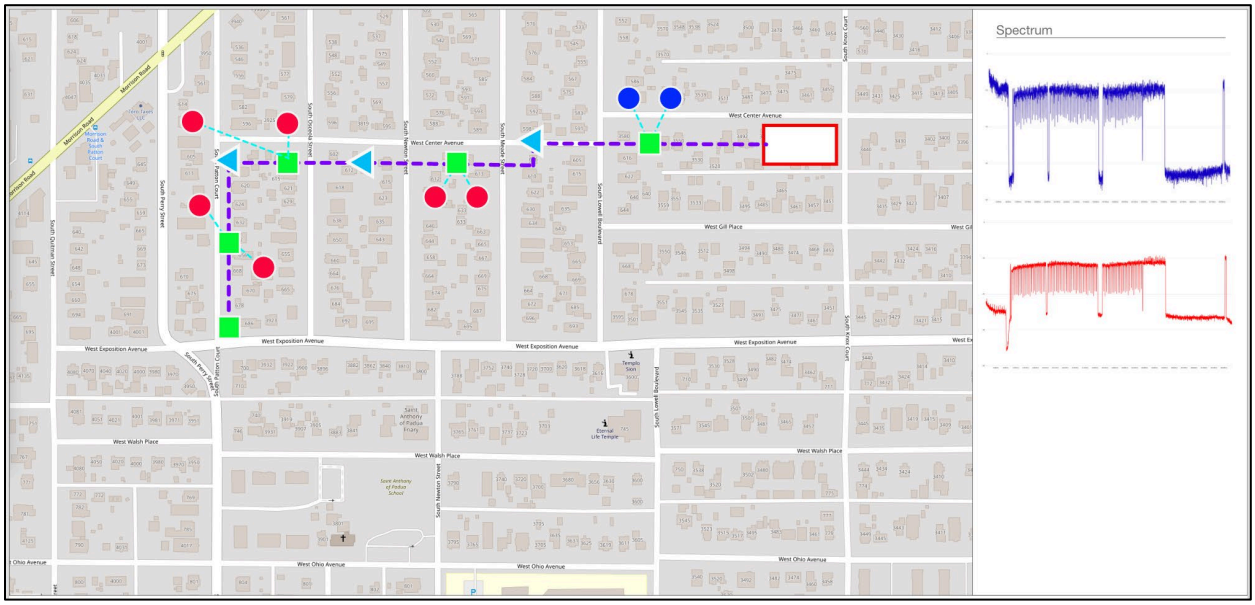


Figure 14- Identifying Locations (Red Devices) Experiencing Elevated Noise Floor

5.2. Visualizing the Data

Creating FBC images from bin data to represent the behavior of the signal is very common and well-understood. In this context, however, we are as interested in visualizing the noise *and* the signal. To that

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

end, when we create the device representations, we must carefully consider the frequency bands of interest and the reference axes if we are to apply the data practically. In our test scenario, there is value in closely examining the spectrum between 87 MHz and 107 MHz. We consider this region to be the diplex filter transition region. Both of the channel plans tested were mid-split. As such, the first linear video channel began at 108 MHz, and the diplex filters of the DOCSIS 3.1 devices under test provided a low internal noise figure at these frequencies for us to analyze against. There is an Out-of-Band (OOB) data carrier at approximately 104 MHz; however, its width and the guard band spacing on either side didn't make it a useful reference point.

There is also great value in visualizing the entire catalog of available bin data to create a representation of the signals and the noise floor at both low and high frequencies. This is the more typical and accepted representation of FBC data.

5.3. Analyzing the Data

With the considerations described above in mind, the visualization of the test devices connected to an amplifier set to optimal input drive levels can be considered the baseline for good, or linear performance, at both the full spectrum representations and in the diplex filter transition region.

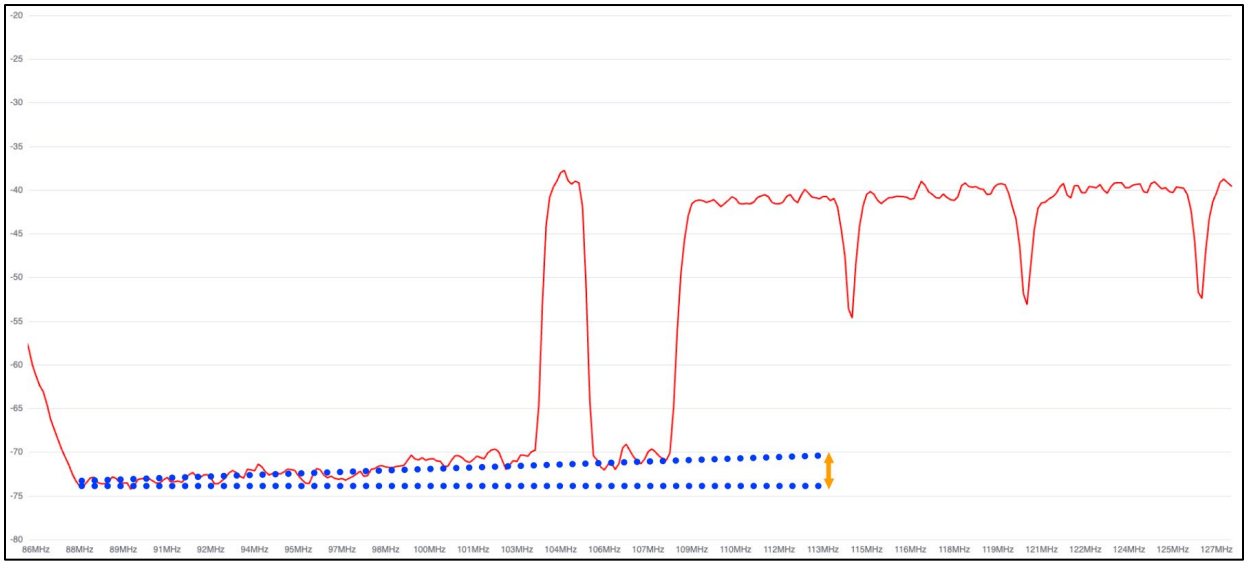


Figure 15- Diplex Filter Transition Region of a Device in Optimal Conditions

We added two lines to the bin visualization, beginning at a transition frequency of 87 MHz: one at a parallel amplitude to the noise figure at 87 MHz and another that intersects with the noise floor where the noise and the signal intersect at the leading edge of the linear video channel at 107 MHz, **Figure 15**. This created what we call a "noise angle." The diplex filter transition region exhibits an acute noise angle for devices connected to an amplifier operating in a linear or optimal fashion. The precise angle has not been calculated for this paper but is a consideration for future work on this topic.

Ingress can impact the low-frequency noise in an individual device within a premise, and some localization logic may have to be applied to discern the impact of the plant alignment, exclusive of premise wiring or ingress that may be present.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

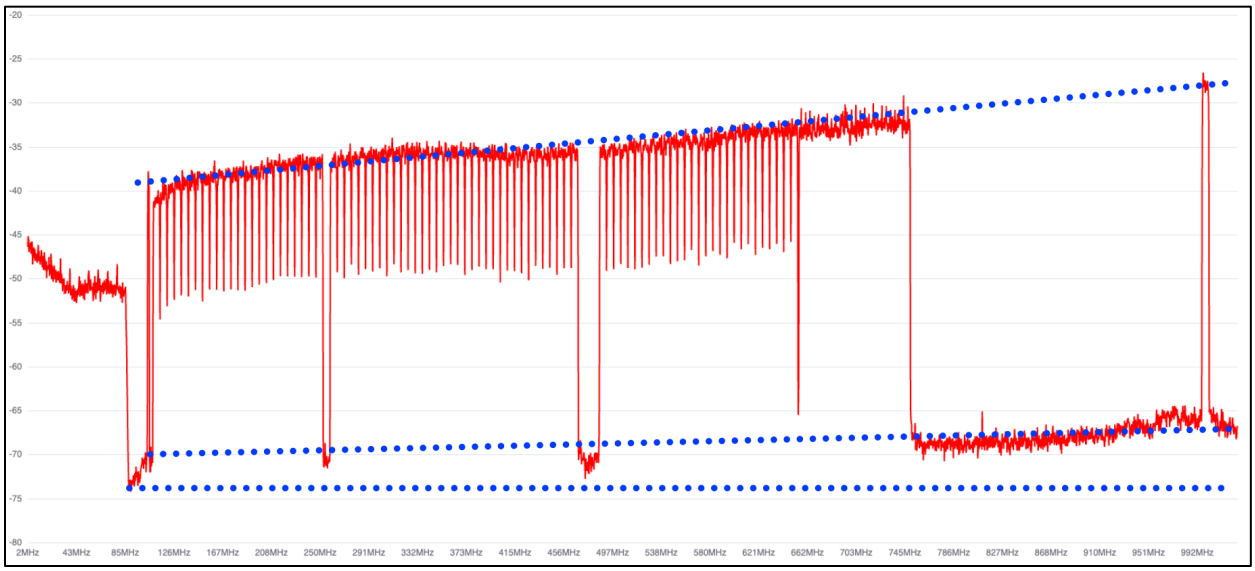


Figure 16- FBC Example of a Device With Optimal Amplifier Setup

We also observed other angles where the signal and noise relationship in the FBC representation of devices appear informative, **Figure 16**. A parallel line was constructed from the 87 MHz noise figure point extending to the highest frequency bin data point available from the device. This line indicates the noise figure of the measurement device. Creating another line that begins at the intersection of the noise floor, and the leading edge of our first video channel, in this instance 107 MHz, and extending that line to the highest frequency noise bin data point, we create another angle, which has a comparative amplitude at high frequency, which can be considered the noise delta (dn, **Figure 17**). The noise delta calculation can is expressed as: (dn) = high-frequency noise floor (hfnf) – low-frequency noise floor (lfnf). While this noise delta does not indicate the approximate position of the device under test (DUT) on the NPR or MER curves, it can be informative when combined with other pieces of information, like CPE receive level, MER, or RxMER. We also observed the relative relationship between the slope or tilt of the signal, with the slope or the tilt of the noise floor itself. These could be described as the signal slope and the noise slope.

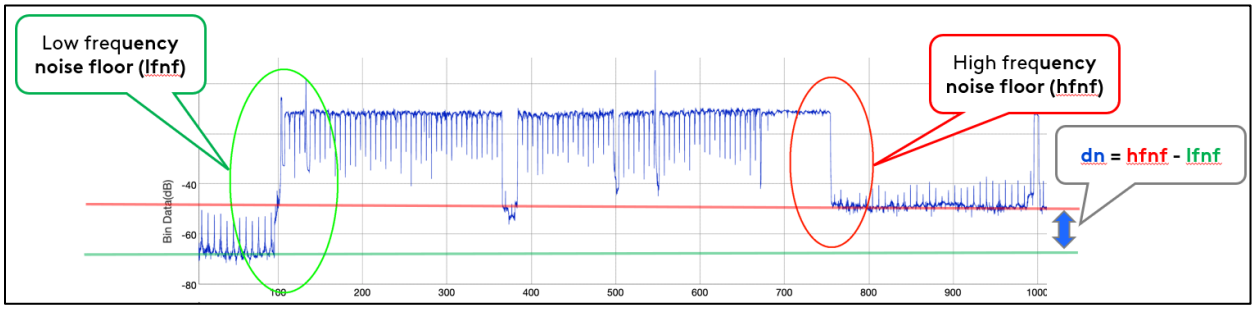


Figure 17- Example of Noise Delta (dn) Calculation

Using the data and visualizations from devices in ideal conditions, we compared data captured at other defined positions on the MER curve.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

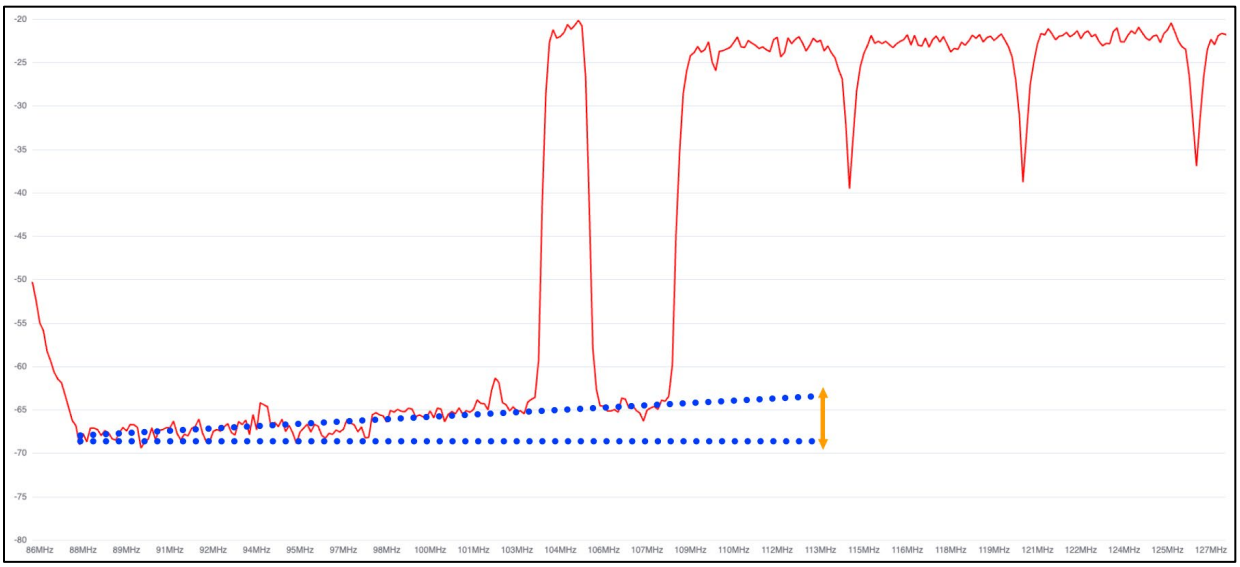


Figure 18- Diplex Filter Transition Region of a Device in the Noise Region of the MER Curve

Examining the noise angle, **Figure 18**, of a device connected to a system operating in the noise region of the MER curve, we found negligible difference in the noise angle when visually compared to a device operating in ideal conditions.

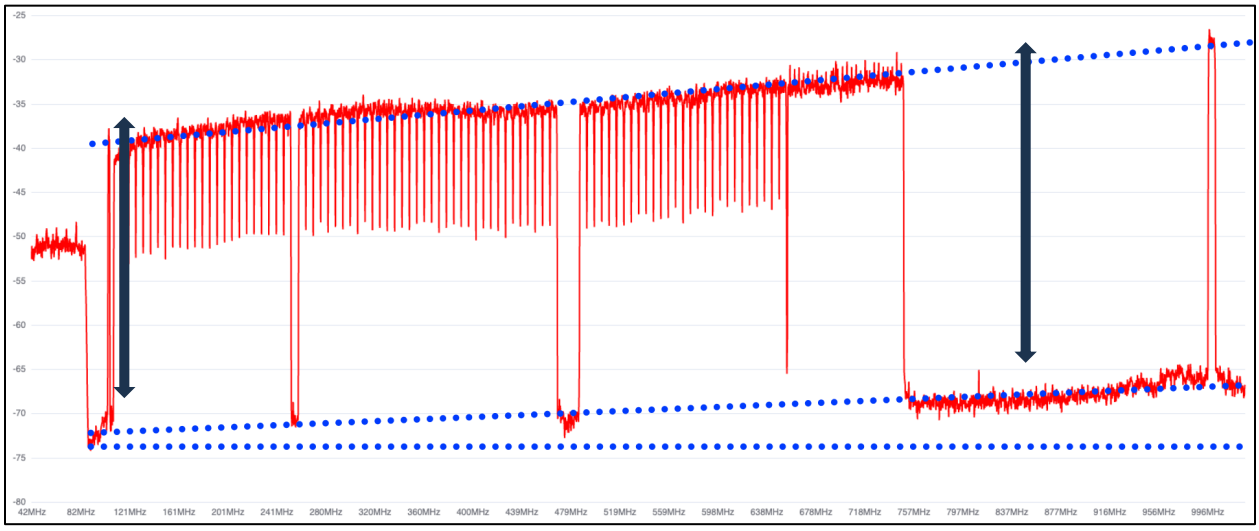


Figure 19- Example of a Misaligned Amplifier Operating in the Noise Region

When we observe the entire spectrum plot, some differences begin to emerge. Compared to the signal slope, the noise slope appears parallel, and the intercept point where the noise and signal converge at 107 MHz appears closer than the optimal setup condition, **Figure 19**. The noise delta calculated using the hfnf – lfnf equation is similar to optimal. However, the slope relationships are slightly different.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

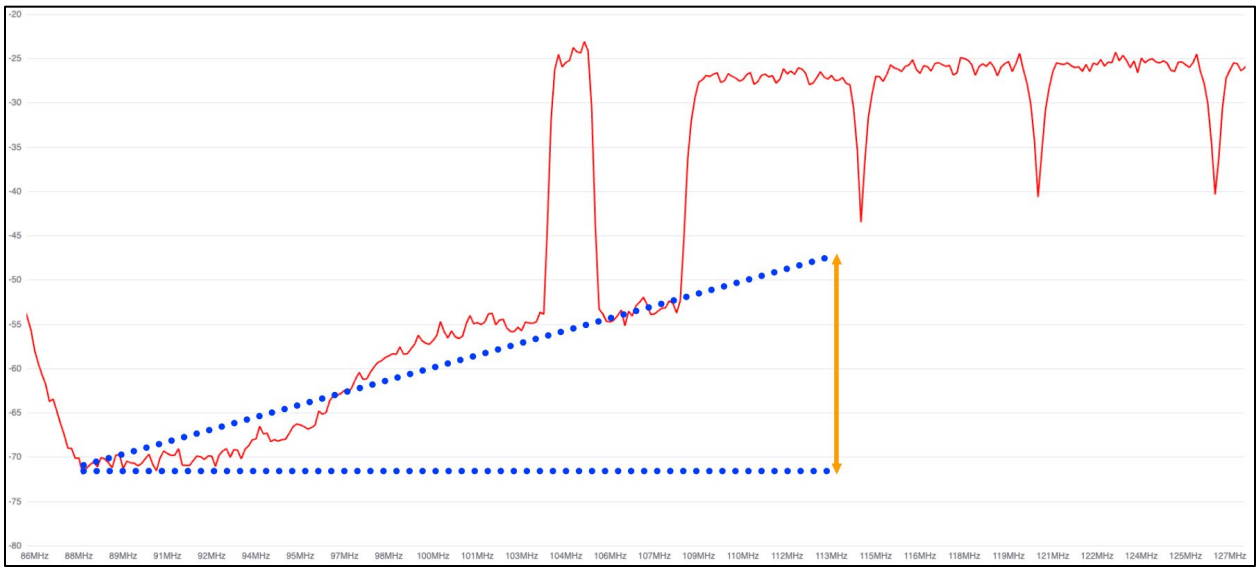


Figure 20- Diplex Filter Transition Region of a Device in the Clipping Region of the MER Curve

Moving to the device data from a system operating in the clipping portion of the MER curve, the impacts on the noise and signal relationship become immediately apparent. Examining the noise angle in the diplex filter transition region of a device connected to a system operating in a non-linear state shows significant impact using this noise angle method, **Figure 20**. This may also appear to be premise-related ingress or noise until we examine the entire spectrum plot and the impact of the noise floor at higher frequencies.

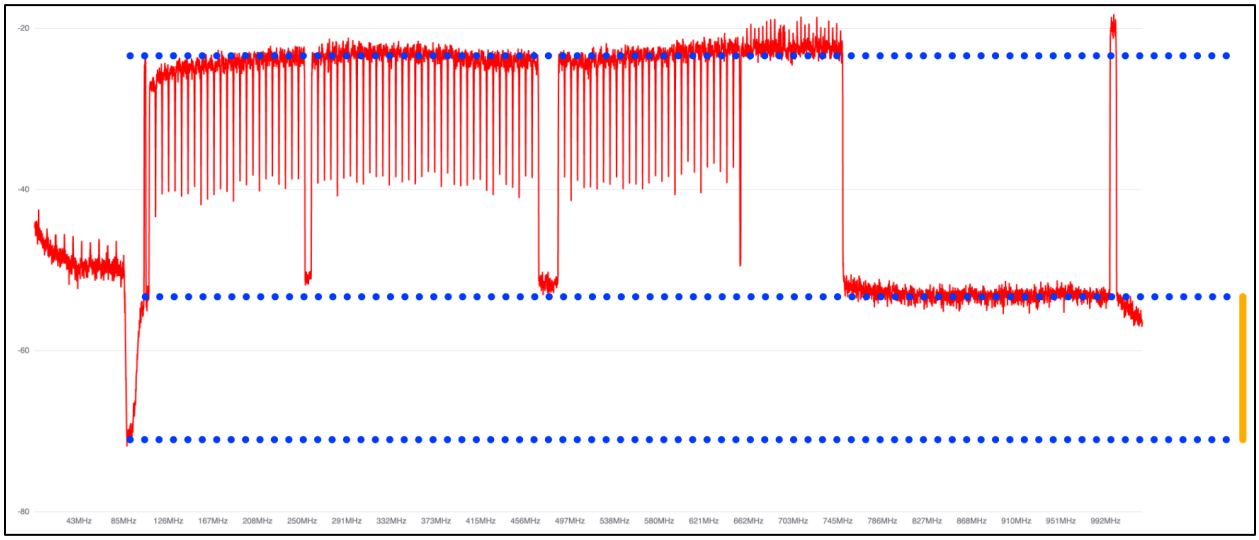


Figure 21- Example of a Misaligned Amplifier in Compression

Devices connected to this system were operating at a point 6 dB past the peak. The impact of non-linear performance becomes evident when evaluating the signal, noise amplitude and angle relationships. The amplitude and slope of the noise change with respect to the amplitude and slope of the signal, **Figure 21**. This appears to be the equalized response of the second and third-order (CSO and CTB) distortions of our digital channel plan generated by the amplifier.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Compared to the optimal condition, we are displaying the visual representations of the extreme cases at either end of the MER curve. The test data included many additional data points between the examples presented in this paper. Initial analysis of those other data sets supports the behaviors indicated by both NPR and MER curves. As devices transition through the curves, beginning in the noise region of the curve, as power is applied, there is a relatively linear relationship between the signals and the noise until you reach the clipping region, where the power-to-noise relationship is affected at a greater than 1:1 relationship.

Using the methods we describe for this test, it is possible to estimate the position of a system or amplifier on an MER curve and predict the impact of adding power to the system. Additionally, by understanding the amount of power added to the system, the predictions could include the severity of the impact if the system is in a non-linear state.

5.4. Practical Application

5.4.1. Pre-Enablement

The initial development for practical application of this effort was to predict the success of additional spectrum enablement. Using the data derived in our test environment directly informs this spectrum enablement process. A PNM-informed process could look something like this:

In the target node, acquire and store pre-enablement FBC bin data for the service group with the additional spectrum activated. Using PNM tooling and thresholds, evaluate the device pre-enablement telemetry for key indicators of non-linearity. Knowing that our spectrum enablement increases the system's TCP by 2 dB, any indications of non-linearity in any device in the node would predict that the system would deteriorate by a value greater than 2 dB, and enablement could end in catastrophic system failure.

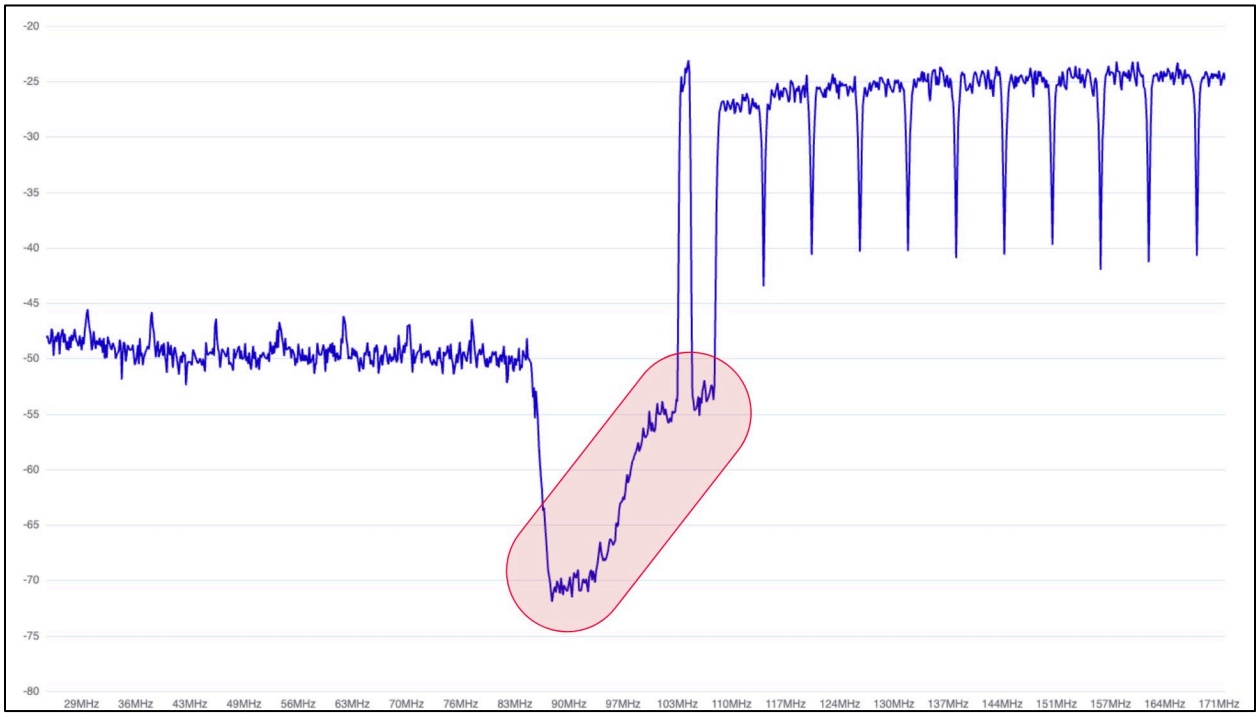


Figure 22- Example of Distortion Noise in the Diplex Filter Transition Region

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

If the target system does not detect impaired devices, the system could be considered a candidate for full deployment or a multi-phase approach to enablement. This multi-phase approach is the "canary" approach. This involves introducing a portion of the desired power into the system, 1 dB of TCP, for instance, and then requery devices for evidence of non-linear impairments, **Figure 22**. If the node devices show no or minimal indication of impairment, then the risk of degradation or failure due to the additional TCP can be considered negligible.

5.4.2. Post Enablement

Are there practical applications to this approach once your downstream payload is fully deployed? If we consider the dynamic conditions of the outside plant networks, and the inherent variability of weather, then this logic could have continued applications well beyond spectrum enablement. Instances of Automatic Level Control (ALC) failures in amplifiers, resulting in increased power into the successive amplifiers in cascade, are not uncommon. These changes in amplifier drive level could result in amplifiers becoming non-linear and would be in scope for tooling efforts such as this.

6. Conclusion

As outlined in this paper, there are many things to consider when evolving a CATV network, including the monitoring tools. PNM has become an indispensable platform for an operator to measure and qualify the network's performance. We believe the data collected, analyzed and shared in this paper support the position that amplifier alignment analysis using PNM tooling is one more tool that can be successfully developed and added to the toolbox. Operators desiring to enable additional spectrum into the downstream of their networks should be able to predict whether their system is operating in a condition that would allow them to do so successfully using PNM tools.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

ALC	Automatic Level Control
BER	Bit Error Rate
CATV	Community Antenna Television
CCN	Carrier-to-Composite Noise
CIN	Carrier-to-Intermodulation Noise
CMTS	Cable Modem Termination Systems
CNR	Carrier-to-Noise Ratio
CPE	Customer Pr
CSO	Composite Second Order
CTB	Composite Triple Beat
CTN	Carrier-to-Thermal Noise
DUT	Device Under Test
FBC	Full Band Capture
HFC	Hybrid Fiber Coax
MER	Modulation Error Ratio
MHz	Megahertz
NPR	Noise-to-Power Ratio
OOB	Out-of-Band
PNM	Proactive Network Maintenance
QAM	Quadrature Amplitude Modulation
RF	Radio Frequencies
RxMER	Receive MER
SNMP	Simple Network Management Protocol
TCP	Total Composite Power

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

1. Volpe, Bradley. "Practical Implementation of Profile Management Application (PMA) to Improve Data Throughput in the Presence of Impairments." *SCTE Technical Paper*, 2021, https://volpefirm.com/wp-content/uploads/2021/11/SCTE-2021-PMA.pdf
2. "Basic Cable TV Background: Modulation Signal Formats and Coaxial Cable Systems." *Pearson*, https://www.pearsonhighered.com/assets/samplechapter/0/1/3/0/0130864218.pdf
3. Linton, Diana. "How Can NPR Improve Future 1.8 GHz Extended Spectrum Amplifier Characterization?" *Broadband Library*, Spring 2022, pp 24-25.
4. "ANSI/SCTE 119 2018." *SCTE*ISBE Standards*, 2018, page 5.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Utilizing OFDM for Measuring Spectrum Performance Above 1 GHz

A Technical Paper prepared for SCTE by

Matthew Olfert
Principal Network Architect - Access Technology
Rogers Communications
2400 32 Ave NE Calgary, AB T2E 9A7
403.538.5210
matthew.olfert@sjrb.ca

Justin Stiles, Charter Communications

Robert Wittmeier, Rogers Communications

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Life Cycle of Plant Development 4
   2.1. Plant/Spectrum Upgrade Plan 4
   2.2. Plant Upgrade 5
   2.3. Post-upgrade Assessment 5
   2.4. Initial Customer Enablement 5
   2.5. Measure Spectrum Quality Over Time 5
   2.6. Plant Maintenance 5
3. Carrier Limitation Above 1 GHz 5
4. Restricting Cable Modems from Active DOCSIS Channels 5
   4.1. Determine Attribute Mask Value 6
   4.2. Build Bonding Groups 6
   4.3. Update Downstream Service Flows 6
   4.4. Update Cable Modems 6
5. Utilizing PNM Spectrum Data to Identify Plant Impairments 6
   5.1. Case Study: Roll-off Impairment 7
6. Spectrum Efficiency 10
   6.1. Why is Spectrum Efficiency Valuable? 10
   6.2. Data Collection Methodology 11
   6.3. Spectrum Efficiency Aggregation 13
   6.4. Field Applications/Tool Development 13
   6.5. Further Analysis 14
7. Conclusion 14

Abbreviations 15

List of Figures

Title **Page Number**

Figure 1 – Life Cycle of Plant Development 4
Figure 2 – Proposed 1.2 GHz High-Split Downstream Carrier Loading Plan with Level Reference 7
Figure 3 – Example of Customer Drop with Severe Roll-Off Caused by Bandwidth Limited Splitter – Spectrum Analyzer View 8
Figure 4 – Example of Customer Drop with Severe Roll-Off Caused by Bandwidth Limited Splitter – OFDM Power Spectrum View 8
Figure 5 – Example of RxMER at Customer Drop With (Red) and Without Bandwidth Limited Splitter (Blue) 9
Figure 6 – Example of Customer Drop with Bandwidth Limited Splitter Removed – Spectrum Analyzer View 9
Figure 7 – Example of Customer Drop with Bandwidth Limited Splitter Removed – OFDM Power Spectrum View 10
Figure 8 – Poor Upstream Spectrum Efficiency Limiting Upstream Traffic 11
Figure 9 - Spectrum Efficiency Tableau Dashboard 13
Figure 10 - Traffic Impact Indicators 14

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Tables

Title **Page Number**

Table 1 - Spectrum Efficiency Data Fields 12
Table 2 – OFDM Downstream Profile Design 12

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Our industry is going through a development phase of our hybrid fiber/coax (HFC) networks. This next phase will rely on orthogonal frequency division multiplexing (OFDM) and orthogonal frequency division multiple access (OFDMA) channels. With new spectrum being utilized for the first time, we can introduce a new activation process to maximize the quality of the spectrum prior to customer utilization. This paper covers the life cycle of plant development that can avoid risks. This new life cycle will utilize multiple data sources to identify potential impacts and provide a key performance indicator (KPI) that can monitor spectrum after service enablement.

2. Life Cycle of Plant Development

As Data-Over-Cable Service Interface Specifications (DOCSIS®) services increase the need for reliability and the HFC plant evolves, operators need to develop a strong spectrum development life cycle to maintain a high level of service. For existing customers, taking the time to assess post-construction state prior to service enablement can mitigate risks. It qualifies the success of the construction and provides data that demonstrates the readiness for service enablement.

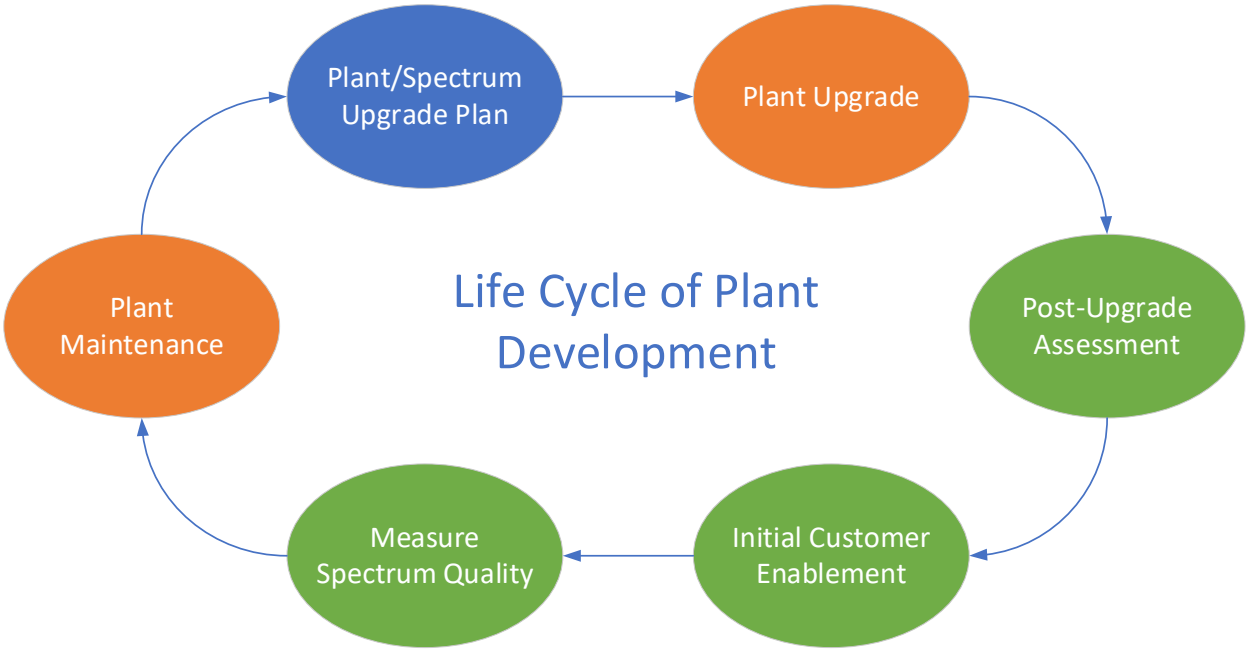


Figure 1 – Life Cycle of Plant Development

2.1. Plant/Spectrum Upgrade Plan

The first step in the process is to build the plan for the plant and/or spectrum upgrade. If spectrum changes are going to occur, it is key to identify the DOCSIS spectrum plan to support the construction process. For example, if a plant upgrade is occurring that will upgrade the HFC plant from 1 GHz to 1.2 GHz on the downstream, the plan for an additional OFDM channel above 1 GHz should be expected.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2.2. Plant Upgrade

The plant upgrade process typically starts with the node upgrade. As the node is upgraded, this is the ideal time to turn up the new DOCSIS carriers. These new carriers should be restricted from services on cable modems to avoid impacts. Every active and passive module upgrade after the node down the cascade will have this energy to provide a reference for setup and a validation of the success of the upgrade.

2.3. Post-upgrade Assessment

Once the plant upgrade has been completed all capable cable modems are able to collect proactive network maintenance (PNM) spectrum data. This data can provide information on frequency response roll-off, suck-outs, and other known spectrum impairment profiles. This can validate if additional maintenance is needed prior to activating the new spectrum for services.

2.4. Initial Customer Enablement

After successful assessment of the new DOCSIS carriers, cable modem restrictions can be removed by an update to the cable modem configuration files. By removing the restrictions, cable modems will be cleared for the new spectrum.

2.5. Measure Spectrum Quality Over Time

As the new spectrum is utilized it is important to measure the cable modems' performance within the new spectrum. Operators have their standard KPIs that can be utilized for this function. At Rogers Communications we use KPI called Spectrum Efficiency, which is covered in detail below.

2.6. Plant Maintenance

If your measurement crosses a trigger threshold, moving forward with plant maintenance or in-home service call(s) could be initiated. This workflow is about recovering from the issue and restoring services for your customer. This process could also recommend additional plant upgrades to correct issues.

3. Carrier Limitation Above 1 GHz

When utilizing spectrum above 1 GHz for DOCSIS technology, operators will be limited to OFDM-based channels for cable modems. Since this is the planned use of this spectrum it should not be looked at as a limitation, but as opportunity to assess performance. This information can predict how it will be utilized in the future by the customers' cable modems. Thus, the challenge is the limitation of cable modems from this assessment channel prior to activation. As an example, if you enable OFDM carriers between 1.2 GHz and 1.8 GHz, only DOCSIS 4.0 cable modems can be utilized for the assessment.

4. Restricting Cable Modems from Active DOCSIS Channels

The most effective way to restrict DOCSIS channels for cable modems under the DOCSIS specification is by utilizing the service flow attribute mask feature. This attribute mask can be utilized to require or forbid certain channels or bonding groups. From working with this DOCSIS feature, we find with bonded cable modems the required attribute mask is the easiest to work with.

In order to build a system that will restrict all DOCSIS 3.1 and 4.0 cable modems from a OFDM channel above 1 GHz you will need to do the following:

1. Determine attribute mask value

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2. Build bonding groups
3. Update downstream service flows
4. Update cable modems

4.1. Determine Attribute Mask Value

For the purposes of this use case, we only require an attribute mask for bonding since non-bonded cable modems cannot access OFDM channels or spectrum above 1 GHz. One example cable modem termination system (CMTS) has the default attribute mask value for a bonding group configured to 0x80000000. The CMTS also requires a value greater than the default value to be used for a required attribute mask.

4.2. Build Bonding Groups

Typically, the CMTS utilizes dynamic bonding groups. Since we want to restrict the new OFDM channel we need to create static bonding groups for the modems that can access that new OFDM channel so a dynamic group can no longer be utilized. This mean that the created static bonding groups will all have the new attribute mask value applied to them. The dynamic bonding groups can remain at default.

4.3. Update Downstream Service Flows

Each cable modem that requires this restriction will need to have all of its downstream service flows updated with the required attribute mask and attribute aggregate rule mask. The value must match the value selected above.

4.4. Update Cable Modems

With the CMTS and cable modem configuration file updates in place a reset to the cable modem is needed. This update applies the attribute mask to the cable modem bonding groups upon registration. You can time these updates to occur at the same time of the OFDM channel configuration change to avoid multiple cable modem resets.

Now with all the required cable modems having the attribute mask in place they will be restricted from accessing the new DOCSIS assessment channel. Once the assessment channel has been cleared for service activation just revert the cable modems back to the default attribute mask with another update and they will start accessing the new capacity.

5. Utilizing PNM Spectrum Data to Identify Plant Impairments

Utilizing OFDM above 1 GHz, even without leveraging the bit-loading for throughput, can offer cable operators many advantages. Downstream DOCSIS channels can assist with new installations and cutovers serving as a carrier amplitude reference for sweep and alignment operations. Furthermore, it enables the operator to collect PNM spectrum data and receive modulation error ratio (RxMER) data per subcarrier from cable modems and digital hand-held meters. This provides insights into the noise floor for detecting potential ingress issues and other impairments that may be lurking in the coaxial plant and premises network above 1 GHz. Downstream DOCSIS channels also have the potential to aid in leakage monitoring that could reveal previously undetected shielding integrity problems, as well as help discover new measurement techniques and strategies to gauge the performance of the HFC network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1.2 GHz frequency division duplex (FDD) deployments can leverage both legacy HFC nodes and distributed access architecture (DAA) nodes in the network. However, with OFDM deployed above 1 GHz traditional downstream active sweep systems face hurdles with limitations for the number of active sweep points that reduce the resolution of the active sweep tool, making it more challenging to identify frequency response issues. Sweep views are limited to a top-of-carrier amplitude reference called sweepless sweep. A sweepless sweep reference can be beneficial in networks that deploy additional OFDM carriers spread across all possible spectrum in the system. Sweepless sweep provides a good representation for the network conditions; however, it is important to note that it only works when carriers occupy the entire bandwidth through nodes, amplifiers and passive devices. Populating the entire radio frequency (RF) spectrum with OFDM also provides visibility into any alignment issues that may exist after upgrading to new 1.2 GHz and 1.8 GHz amplifiers. Figure 2 shows a proposed 1.2 GHz high-split downstream carrier loading plan with level reference.

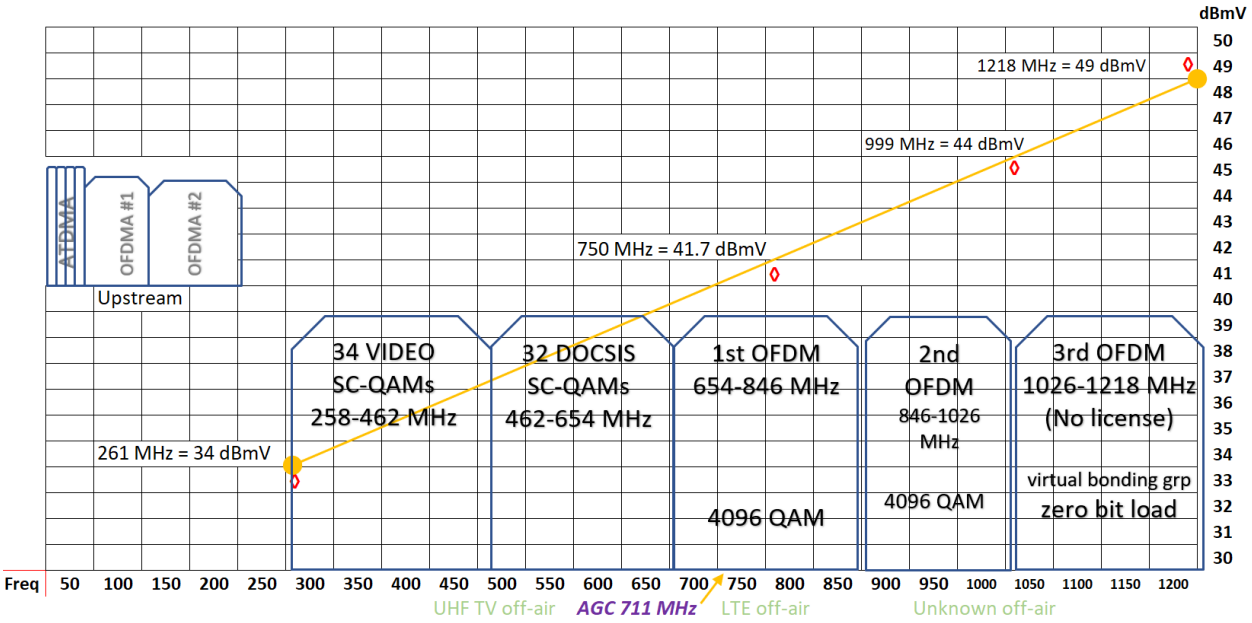


Figure 2 – Proposed 1.2 GHz High-Split Downstream Carrier Loading Plan with Level Reference

Besides having a downstream RF level and RxMER data, the additional loading also allows for over-the-air leakage monitoring, utilizing OFDM pilot carriers as leakage tones. One of the greatest realizations is that this newfound OFDM derived data collection is useful for RxMER evaluation with no impact to customer service. The operator can determine OFDM channel settings such as physical layer link channel (PLC) placement, next codeword pointer (NCP) location, data profiles, exception zones, occupied bandwidth, and more for a particular service group as part of a quality control exercise.

5.1. Case Study: Roll-off Impairment

One common impairment operators will see when deploying OFDM above 1 GHz is a roll-off frequency response characteristic related to many different network limitations. Sweeping and aligning each amplifier is good preventive practice, but what if an operator missed an amplifier or line passive that was supposed to be upgraded? What if there is an issue that only exists inside of the customer's drop network? We can utilize PNM spectrum data and RxMER tools that work with cable modems to help identify these issues in both the premises (single modem issue), and the plant (multiple modem issues). Figure 3 shows

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

a spectrum analyzer view of a customer's drop that had a bandwidth-limited splitter in the home that caused severe 20 dB attenuated roll-off above 1 GHz.

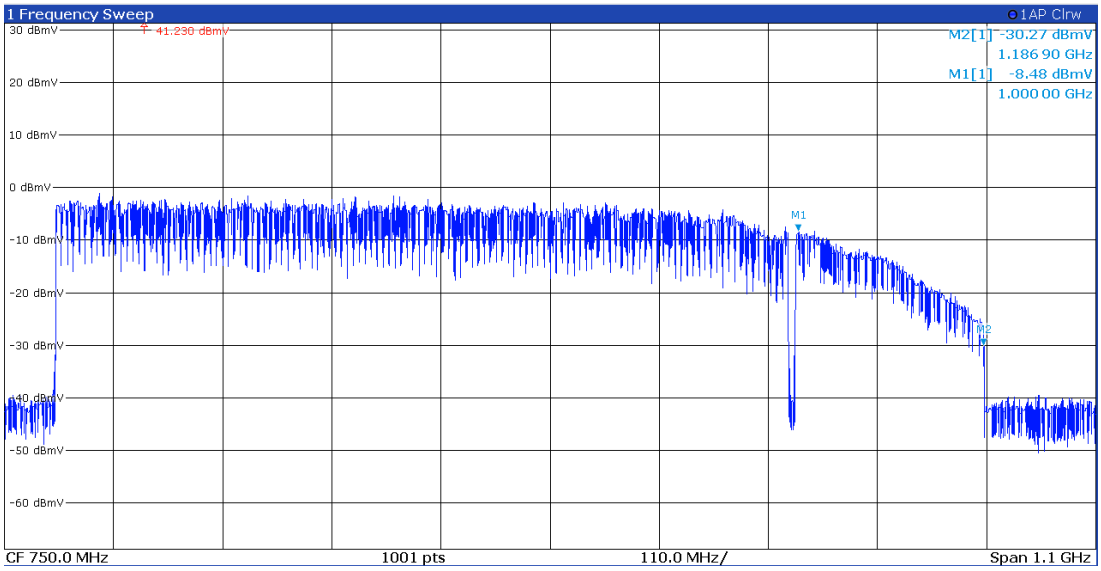


Figure 3 – Example of Customer Drop with Severe Roll-Off Caused by Bandwidth Limited Splitter – Spectrum Analyzer View

Figure 4 shows a closeup of the OFDM channel power spectrum. Again, the roll-off characteristic is very severe and will affect RxMER as shown in Figure 5. The red line represents the RxMER of the OFDM channel above 1 GHz at the customer drop with a bandwidth limited splitter installed, and the blue line represents the same customer drop with the splitter removed.

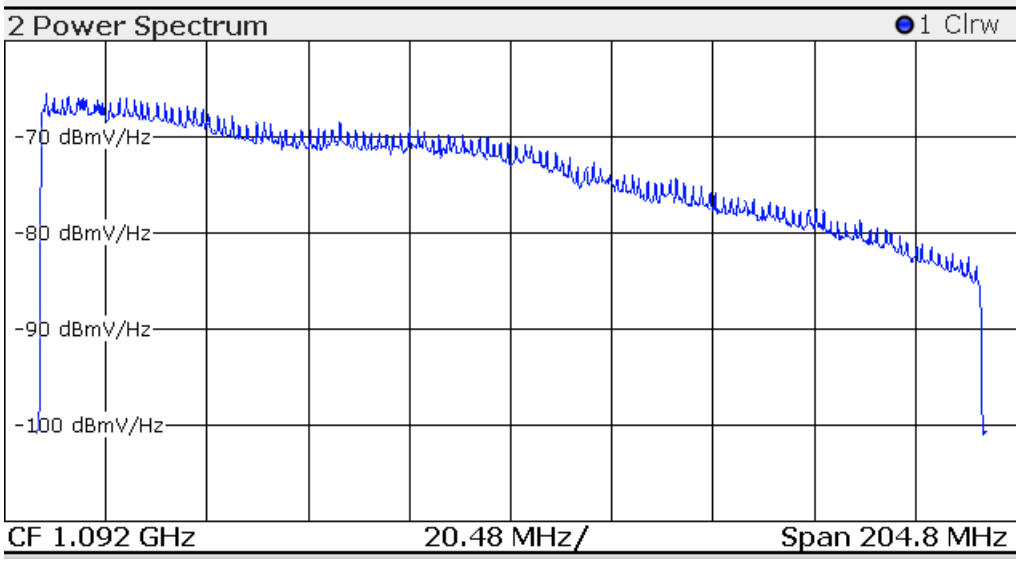


Figure 4 – Example of Customer Drop with Severe Roll-Off Caused by Bandwidth Limited Splitter – OFDM Power Spectrum View

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

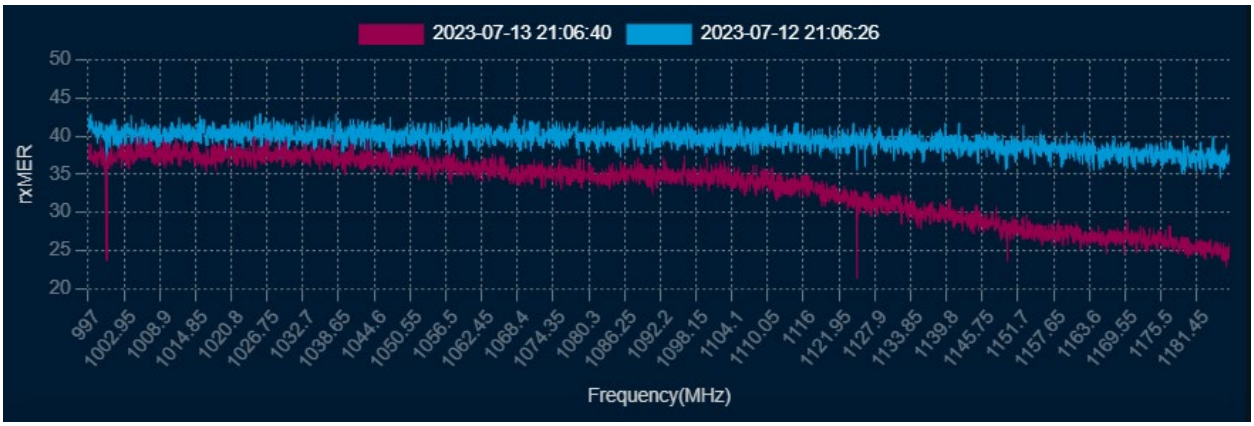


Figure 5 – Example of RxMER at Customer Drop With (Red) and Without Bandwidth Limited Splitter (Blue)

By removing the bandwidth limited splitter from the customer premises, the RxMER of the OFDM channel has improved greatly, and with an average RxMER of >38 dB, the modem should be able to receive a 4096-QAM data profile. In this customer drop simulation, there will always be some amount of roll-off seen above 1 GHz because the plant passive devices and taps were not upgraded to 1.2 GHz. This modem has four 1 GHz passive devices between the node and the customer drop. Figure 6 and Figure 7 show the spectrum and OFDM power spectrum view after the bad splitter has been removed from the customer's premise. The roll-off on the OFDM channel improved from 20 dB negative tilt to 10 dB negative tilt, and the level at 1 GHz has increased from -8.5 dBmV to 1.6 dBmV.

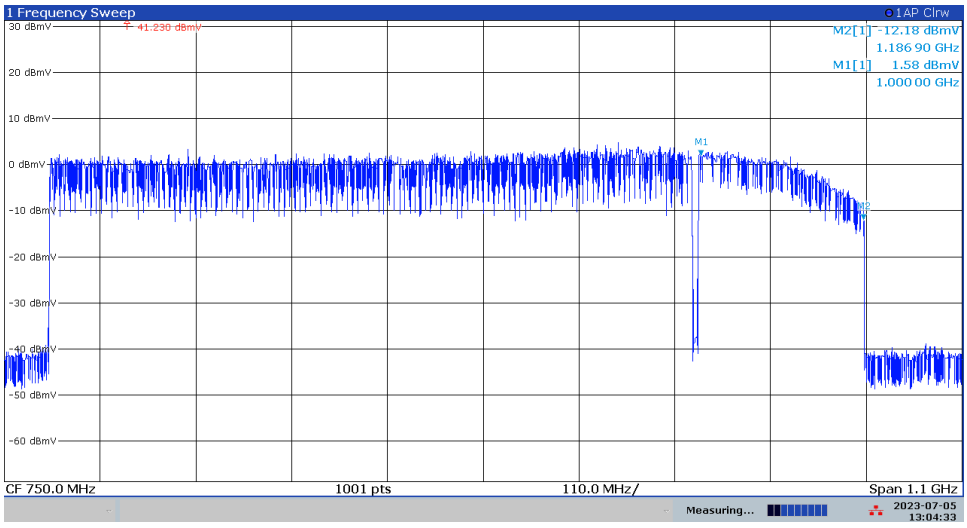


Figure 6 – Example of Customer Drop with Bandwidth Limited Splitter Removed – Spectrum Analyzer View

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

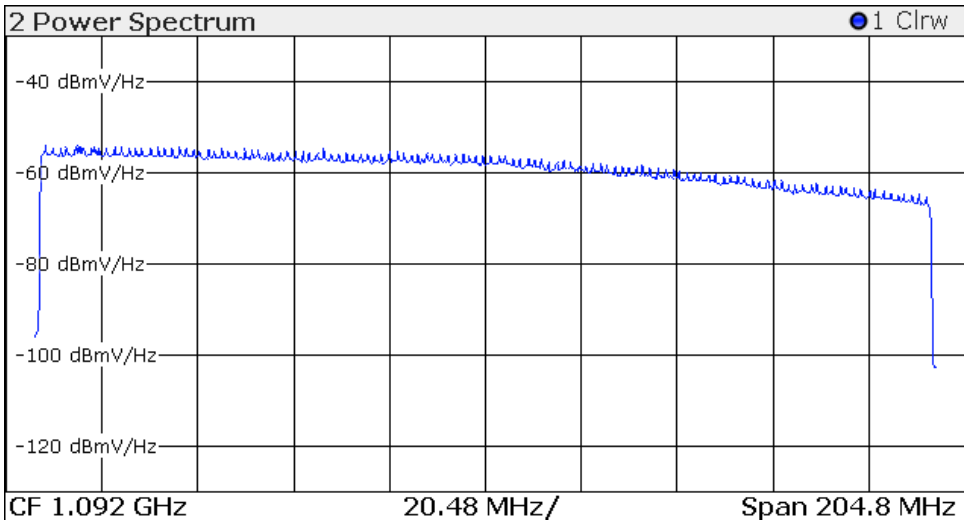


Figure 7 – Example of Customer Drop with Bandwidth Limited Splitter Removed – OFDM Power Spectrum View

6. Spectrum Efficiency

Spectrum Efficiency is a new KPI to give operators insight into how well the spectrum plan designed DOCSIS capacity is utilized by each modem. By combining (aggregating) this metric across multiple modems over time, this metric can highlight chronic plant issues on specific frequencies, groups of modems fed by an amplifier, nodes, or even larger areas (hubs, regions, or even nationally).

6.1. Why is Spectrum Efficiency Valuable?

Spectrum Efficiency provides a complementary measure to congestion. In some cases, poor plant performance can inhibit modems from achieving their capable speeds. This case will never be flagged in a node congestion report because the traffic levels can never reach a point of saturating the node. Spectrum Efficiency on the other hand will highlight such an issue.

Figure 8 shows how poor upstream Spectrum Efficiency on a node is acting as a “soft ceiling” that limits the upstream traffic.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

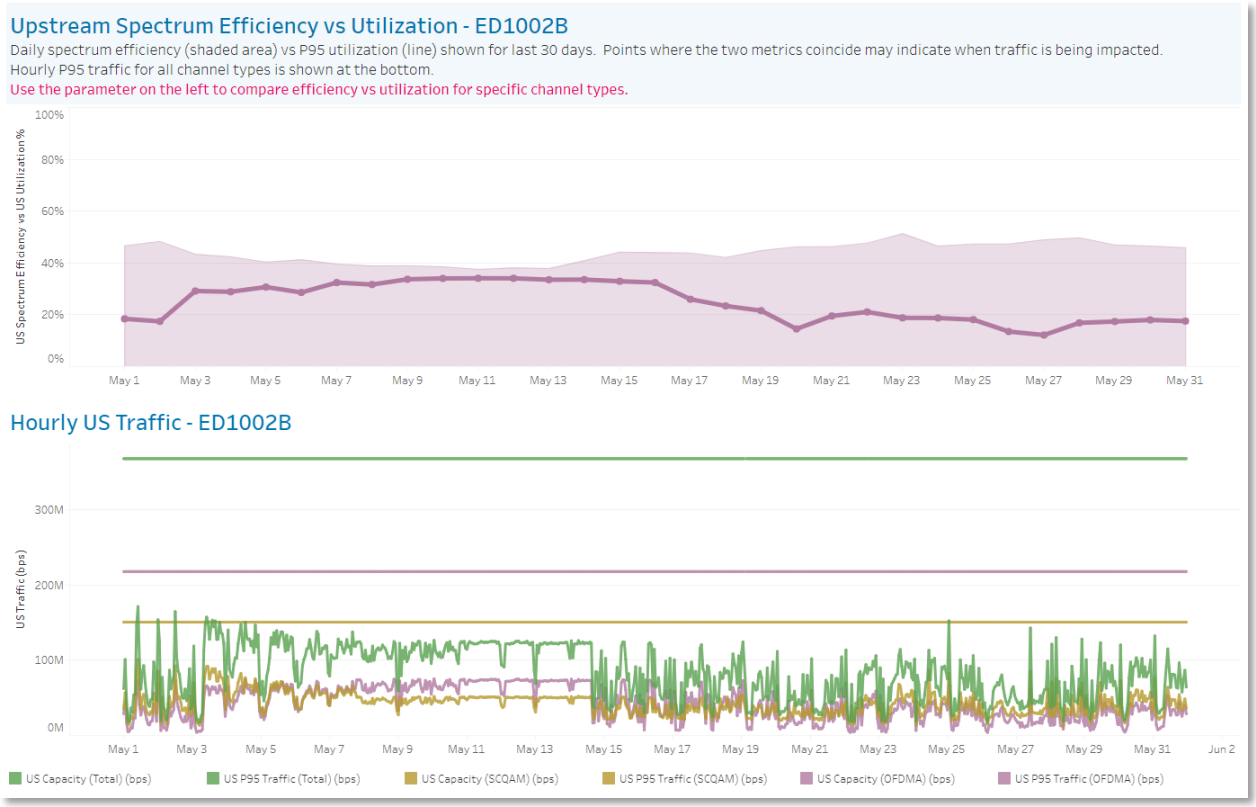


Figure 8 – Poor Upstream Spectrum Efficiency Limiting Upstream Traffic

In the upper half of Figure 8, the purple shaded area shows the daily upstream Spectrum Efficiency of modems using the OFDMA channel. In this example the OFDMA Spectrum Efficiency is close to 50% on average. The thick purple line shows the daily traffic demand as a percentage of engineered channel capacity. When the traffic demand pushes up against the available Spectrum Efficiency, it is being effectively limited from climbing higher. We refer to this condition as "traffic impact."

The lower half of Figure 8 shows an hourly traffic chart. During the periods of traffic impact, a noticeable plateau occurred which may have the potential to cause a poor customer experience, even though the traffic usage is far from the full capability of the channel.

6.2. Data Collection Methodology

Spectrum Efficiency is measured from the modem perspective. Every four hours, modem statistics are polled from all CMTSs across the country. This anonymized data is stored as structured data in a database that can be queried. Historical data is maintained, as well for trending information.

Table 1 shows the data columns from a single poll of a modem with two OFDM channels.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 1 - Spectrum Efficiency Data Fields

polltime	CMid	networkid	chfreq	chwidth
2023-05-21-01-59	00:cb:7a:04:c8:51	CGCHDX7-5	798	192
2023-05-21-01-59	00:cb:7a:04:c8:51	CGCHDX7-5	606	192

chdir	impairment	vblspeed	ofdmSecondProfileCap
DownstreamOFDM	0	1451592704	1448953344
DownstreamOFDM	0	1502616832	1499884800

In the table above, the most important columns are *networkid* (serving group, used to associate the modem with a node), *chfreq* (OFDM channel frequency in MHz), *impairment* (a flag to indicate if the channel is impaired for the current poll), *vblspeed*, and *ofdmSecondProfileCap*.

The *vblspeed* column is from the DOCSIS 3.1 downstream OFDM channel capacity object identifier (OID), docsIf31CmtsDsOfdmProfileStatsFullChannelSpeed. This value considers channel width, exclusion zones, and accounts for variable bit-rate downstream profiles. This is used as the *numerator* of the downstream Spectrum Efficiency calculation. Note that if the *impairment* flag is non-zero (indicating some type of channel impairment), then the numerator is 0 for the current poll.

The *ofdmSecondProfileCap* column is based on the OFDM channel bit rate provided by a designed capacity downstream profile of 1024-QAM, including exclusion zones. This is used as the *denominator* of the downstream Spectrum Efficiency calculation.

Spectrum Efficiency for a particular modem and channel is therefore the ratio:

$$OFDM\ Spectrum\ Efficiency = \frac{vblspeed\ (bits\ per\ second)}{ofdmSecondProfileCap\ (bits\ per\ second)} * 100\%$$

Table 2 summarized OFDM downstream profile numbers and default modulation orders.

Table 2 – OFDM Downstream Profile Design

OFDM DS Profile Number	Default Modulation
0	64-QAM
1	256-QAM
2 (Designed Capacity)	1024-QAM
3	4096-QAM

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6.3. Spectrum Efficiency Aggregation

Spectrum Efficiency can be calculated per modem, per node, per channel frequency, and over a specific period (for example every four hours, daily, weekly, or monthly).

To calculate the Spectrum Efficiency for a node over a two-week period we first sum all *numerators* from all modem samples on the node. Then we sum all *denominators*. We finally calculate the ratio of the sums to get the Spectrum Efficiency percentage.

Spectrum Efficiency can also be calculated per OFDM channel individually or across multiple OFDM channels (if dual OFDM channels have been deployed).

6.4. Field Applications/Tool Development

A two-week aggregation of Spectrum Efficiency of all modems on the node was found to be sufficient to bring chronic poorly performing nodes to the top of a list for action.

Figure 9 shows a Tableau dashboard that is used to rank Spectrum Efficiency so that the worst nodes are at the top of the list.

Optical Receiver	Serving Group	Spectrum Efficiency Average (%) (30-Day)	Peak DS Utilization (OFDM) (%) (30-Day)	Modem Total Average	4/15/2023 Spectrum Efficiency (OFDM) (%)	4/30/2023 Spectrum Efficiency (OFDM) (%)	5/15/2023 Spectrum Efficiency (OFDM) (%)	5/31/2023 Spectrum Efficiency (OFDM) (%)
VC993B	VCRIDX11-46	42%	31%	123	40%	40%	40%	44%
GV618A	GVCVDX16-20	47%	8%	21	55%	49%	48%	46%
TB249A	TBTBDX7-46	57%	20%	66	59%	58%	57%	57%
EK535E	EKFNDX4-35	57%	36%	114	98%	64%	57%	58%
WP2383C	WPSCDX19-35	59%	12%	49	61%	60%	60%	58%
SC15A	SCCHDX3-14	59%	18%	81	92%	75%	64%	54%
MJ125A	MJMJDX3-12	61%	29%	54	66%	64%	60%	63%
VA717B	VAAGDX3-5	62%	53%	190	59%	58%	60%	64%
TB610	TBFFDX3-6	63%	29%	30	64%	64%	63%	63%
ED3381A	EDCDDX5-34	65%	37%	104	66%	65%	65%	65%
GV385A	GVDTDX8-7	65%	15%	60	66%	65%	65%	66%
NB52C	NBDCDX3-13	66%	12%	17	66%	68%	67%	66%
CA110	CAMHDX3-11	67%	58%	245	87%	90%	84%	50%
VS147A	VSNEDX11-39	67%	17%	24	59%	66%	68%	67%
FM78B	FMFMDX8-3	68%	50%	156	89%	74%	67%	69%
ED652A	EDBFDX1-93	68%	4%	18	96%	75%	66%	70%
VC378C	VCSTDX15-48	69%	34%	99	67%	67%	68%	69%
VC391B	VCSTDX14-28	69%	20%	90	68%	67%	69%	69%
GV159A	GVMTDX6-55	69%	52%	106	68%	68%	69%	68%
VS1342D	VSWYDX12-45	69%	9%	11	87%	82%	68%	70%
RD3	RDHEDX11-41	69%	53%	150	72%	77%	74%	65%
VC4199B	VCBBDX14-46	69%	29%	50	71%	71%	69%	70%
VC1876A	VCSTDX15-11	69%	42%	163	70%	69%	69%	70%
VC1098A	VCANDX10-16	70%	25%	103	72%	71%	70%	70%
LB148B	LBTBDX2-5	70%	39%	81	74%	72%	71%	69%
GV1872A	GVEQDX6-23	70%	34%	68	74%	73%	71%	70%

Figure 9 - Spectrum Efficiency Tableau Dashboard

A star is shown in Figure 10 to indicate a node is suffering "traffic impact" from poor upstream Spectrum Efficiency. During the past month, if the median daily delta between traffic demand and Spectrum Efficiency is less than 10% then the traffic impact indicator is shown. This node will be chronically experiencing a traffic ceiling.

Incidents are then opened and tracked against traffic impacted nodes for further monitoring.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Traffic Impact Indicator:
(All)
-
☆
Modem Total:
10 1,711
Optical Receiver Search:

Optical Receiver	Serving Group	Spectrum Efficiency Average (%) (30-Day)	Traffic Impact Indicator	Modem Total Average	Incident Number	4/30/2023 Spectrum Efficiency (%)	5/15/2023 Spectrum Efficiency (%)	5/31/2023 Spectrum Efficiency (%)	6/15/2023 Spectrum Efficiency (%)
NO177	NOBBDX12-1	48%	-	188	-	48%	48%	48%	48%
ED1269A	EDWEDX6-19	49%	-	511	-	86%	58%	43%	55%
RD284C	RDHEDX14-49	50%	☆	997	-	90%	84%	48%	52%
ED1002B	EDCHDX7-44	51%	-	385	INC1240437	58%	42%	49%	53%
ED1398B	EDSHDX6-64	53%	-	773	-	80%	61%	53%	52%
CG543A	CGRUDX5-18	56%	-	199	-	92%	99%	60%	52%
WK80A	WKTRDX1-4	57%	-	1,069	-	90%	89%	67%	46%
ED23A	EDLODX6-48	58%	-	154	-	92%	93%	54%	61%
CG161B	CGDTDX5-48	58%	-	77	-	63%	95%	60%	56%
ED2672A	EDNIDX6-46	58%	-	467	-	83%	76%	57%	59%
LB513A	LBLBDX8-24	60%	-	645	-	90%	94%	59%	60%
WP123D	WPNVDX3-4	61%	-	219	INC1338198	79%	61%	61%	62%
BR8	BRBKDX2-9	62%	-	44	-	99%	86%	58%	66%
RD71	RDHEDX12-56	64%	-	643	-	82%	76%	64%	63%
LS402A	LSWRDX3-2	64%	-	140	-	63%	64%	64%	64%
ED2038A	EDNIDX7-32	64%	-	151	-	96%	90%	67%	61%
WP557B	WPCADX8-30	64%	-	649	-	63%	63%	65%	64%
WP2415C	WPSCDX16-19	65%	-	305	-	64%	62%	65%	65%
SU134A	SUHEDX2-86	67%	-	666	-	65%	67%	67%	67%
CG1013	CGMODX6-6	67%	-	24	-	75%	73%	69%	65%
VSSS11A	VSNWDX10-58	67%	-	725	INC1338169	66%	65%	68%	67%
WP2278B	WPTCDX5-34	68%	-	226	-	66%	67%	67%	68%
ED1869C	EDCHDX7-35	68%	-	476	-	67%	71%	67%	70%
WK240A	WKNLDX4-2	68%	-	481	-	42%	65%	66%	71%
VC815	VCRHDX11-29	69%	-	502	-	85%	77%	69%	69%
SS390A	SSHADX11-33	69%	-	604	-	85%	90%	66%	71%

Rev. 5f - February 3, 2023

Figure 10 - Traffic Impact Indicators

6.5. Further Analysis

Spectrum Efficiency can further be used by operators to spot issues in groups of modems that are spatially correlated (fed off a particular amplifier or tap), flag impairments on specific channels (one OFDM channel performing worse than the other) or even highlight intermittent impairments that occur over a span of time.

7. Conclusion

The life cycle of plant development builds a rigorous process that is focused on building a strong customer experience through an upgrade process. The development of proactive maintenance network data has provided this means to operators to improve our upgrade process. Utilizing the collaboration of DOCSIS technology with the RF plant construction brings together a powerful combination of data to improve the service activation process.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

CMTS	cable modem termination system
DAA	distributed access architecture
dB	decibel
dBmV	decibel millivolt
DOCSIS	Data-Over-Cable Service Interface Specifications
FDD	frequency division duplex
GHz	gigahertz
HFC	hybrid fiber/coax
KPI	key performance indicator
NCP	next codeword pointer
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
OID	object identifiers
QAM	quadrature amplitude modulation
PLC	physical layer link channel
PNM	proactive network maintenance
RF	radio frequency
RxMER	receive modulation error ratio
SCTE	Society of Cable Telecommunications Engineers

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Virtual BNG on Public Cloud for Disaster Recovery and Capacity on Demand

A Technical Paper prepared for SCTE by

Jim Huang
Principal Solutions Architect
Amazon Web Services
jimhuan@amazon.com

Juan Luis Esteban
Director, Broadband and Convergence Solutions
Ciena
jesteban@ciena.com

Christopher Adigun, Telco Solutions Architect, Amazon Web Services

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Today's DBNG – An Overview ... 4
   2.1. Multi-Service Disaggregated BNG ... 4
   2.2. DBNG On-premises Deployment and Workload Characteristics ... 5
3. DBNG Cloud Solution Design ... 6
   3.1. Design Approach and Goals ... 6
   3.2. Reference Architecture and Solution Options ... 8
4. Solution Prototyping and Implementation Methods ... 15
5. Conclusions ... 18

Acknowledgements ... 18

Abbreviations ... 19

Bibliography & References ... 19

List of Figures

Title **Page Number**

Figure 1 - BNG Cloudification Use Cases ... 3
Figure 2 - Disaggregated BNG with Control and User Plane Separation (CUPS) ... 5
Figure 3 - DBNG Deployment Model with Geo-redundancy ... 5
Figure 4 - Cloud Continuum: Region, Edge, and Far Edge ... 6
Figure 5 - DBNG Functions Allocation along Cloud Continuum ... 7
Figure 6 - Cloud DBNG Solution Architecture ... 8
Figure 7 - 1:1 Active-Standby DBNG-CP VM Redundancy in Cloud ... 11
Figure 8 - 1:1 Active-Standby DBNG-CP Pod and Node Redundancy in Cloud ... 12
Figure 9 - DBNG-UP Redundancy across On-premise Datacenter and Public Cloud ... 13
Figure 10 - Subscriber Groups across DBNG-UPs ... 13
Figure 11 – DBNG-UP Disaster Recovery Life Cycle ... 14
Figure 12 - Cost Comparison of DBNG-UP Standby Modes ... 15
Figure 13 - Prototyping Environment ... 15
Figure 14 - VXLAN Implementation Method Using VPP ... 17

List of Tables

Title **Page Number**

Table 1 - BNG Workload Characteristics ... 6
Table 2 - L2 Frame Transport Solution Options ... 10

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Broadband Network Gateway (BNG) is a critical network component in broadband and Internet Service Provider networks. Its primary function is to manage and control the delivery of broadband services, such as internet access, to subscribers. The BNG technology has been evolving over the past two decades from specialized hardware devices to virtual BNG (vBNG) and further to disaggregated BNG (DBNG) with control plane and user plane separation (CUPS) as specified by Broadband Forum Technical Report TR-459[1]. The vBNG technology was driven by the need for more flexible solutions and higher operation efficiency through virtualizing BNG functions and running them as software on industry-standard hardware or virtual machines. The DBNG technology allows service providers to select and deploy individual components, instead of relying on monolithic integrated solutions. It enables service providers to optimize their network architecture and scale each component independently based on demand, allocating resources where they are needed the most and avoiding overprovisioning.

This paper presents BNG cloudification – a broadband technology frontier built on vBNG and DBNG. Today, public cloud has been widely adopted by enterprises and industry verticals including telecommunications due to its economies of scale and capabilities for gaining operation efficiency, enabling business agility, as well as providing a wealth of latest technology services. The BNG cloudification presented in this paper is about running vBNG control plane and vBNG user plane in the public cloud.

The key use cases driving the BNG cloudification are the following, as illustrated in Figure 1.

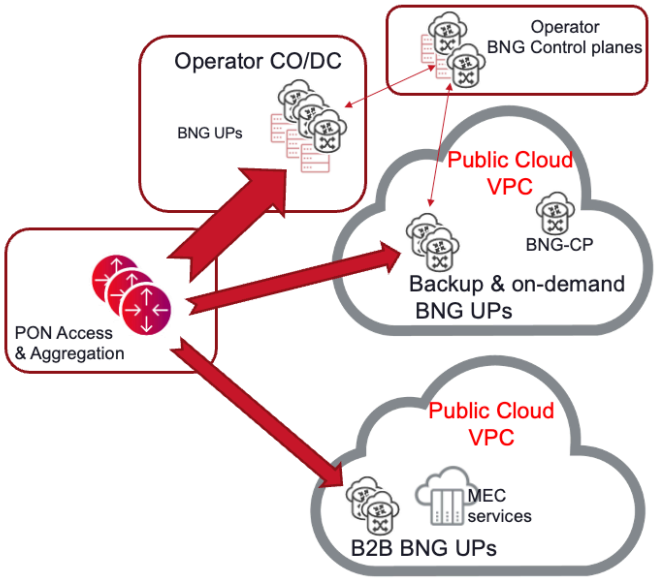


Figure 1 - BNG Cloudification Use Cases

- *Disaster Recovery* – ability to redirect subscriber traffic from DBNG-UPs in the operator network to backup DBNG-UPs in a public cloud. In current on-premise environments, DBNG user planes are deployed in geo-redundant locations to enable disaster recovery. This requires allocation of extra capacity, power, and rack space in each location for the failure case. Allocating disaster recovery capacity in the public cloud allows the operator to reduce CapEx investment.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- *Capacity on demand* – ability to steer traffic to DBNG-UP instances in public cloud based on traffic spike conditions, which may occur daily or seasonally. It leverages the public cloud elasticity to handling bursty workloads instead of building extra capacity in operator data centers, hence lowering CapEx investment.
- *Multi-access Edge Compute (MEC) colocation* – DBNG-UPs may be collocated with edge compute infrastructure for market segments where a high percentage of the traffic is sent to public cloud edge services. The DBNG-UP and MEC collocation let subscriber traffic reach MEC applications at the edge without traversing Internet to regional data centers or public cloud region, thus reducing traffic latency for time-sensitive applications.

This paper describes a cloud-native DBNG solution to support the BNG public cloud use cases. It employs a cloud continuum model to establish a reference architecture with DBNG-CP and DBNG-UP instances distributed across operator's data center, cloud edge, and cloud region. The paper provides design options and tradeoff analysis on several key architecture components – the DBNG connectivity and user plane traffic transport between layer-2 oriented broadband access network on premises and layer-3 oriented network in the cloud, cloud DBNG-UP activation, and cloud DBNG high availability (HA). The paper also shares first-hand experience in the solution prototyping with focus on the implementation methods. The paper provides a direction and solution schemes for using public cloud for BNG disaster recovery and capacity on demand.

In the following, we review today's DBNG technology and typical on-premise deployment scenarios in Section 2. Section 3 details the DBNG cloud solution design. In Section 4, we share our initial solution prototyping experience and lessons learned. We conclude the paper in Section 5.

2. Today's DBNG – An Overview

This section provides an overview of the Disaggregated BNG (DBNG) technology which is a foundation for our DBNG cloud architecture design. The section also describes typical DBNG deployment scenarios today and characterizes DBNG workloads expected for support in the cloud.

2.1. Multi-Service Disaggregated BNG

The concept of Control and User Plane Separation (CUPS) in the telecom industry started to gain attention and prominence over the past several years. It emerged as a result of the increasing demands for flexibility, scalability, and efficiency in network architectures. Broadband Forum released its first Technical Report TR-459 "Multi-Service Disaggregated BNG with CUPS" in 2020[1].

A high-level DBNG architecture with CUPS is shown in Figure 2, where the control plane and the user plane are separated with management, control packet redirect, and state control interfaces in-between. The State Control Interface, SCi, is based on the Packet Forwarding Control Protocol (PFCP). The Control Packet Redirection interface, CPRi, is based on GPRS Tunnelling Protocol User Plane (GTP-U) which encapsulates control plane packets. The Management Interface, Mi, uses Network Configuration Protocol (NETCONF).

The relationship between DBNG Control Plane (DBNG-CP) and DBNG User Plane (DBNG-UP) is 1-to-many with a single DBNG-CP instance controlling multiple DBNG-UP instances. Multiple DBNG-UP instances may be deployed to achieve the desired network scaling, to allow load balancing of sessions, to support different service levels, to provide Multi-Access Edge Compute (MEC) services to a subset of subscribers, to provide DBNG-UP functions within different network slices, or to provide data path redundancy for BNG resiliency. The Subscriber Session Steering (SSS) is a new network function being

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

specified by Broadband Forum[2][3][4]. It provides the capability to direct subscriber sessions to the most appropriate DBNG-UP per the multi-DBNG-UP use cases listed above.

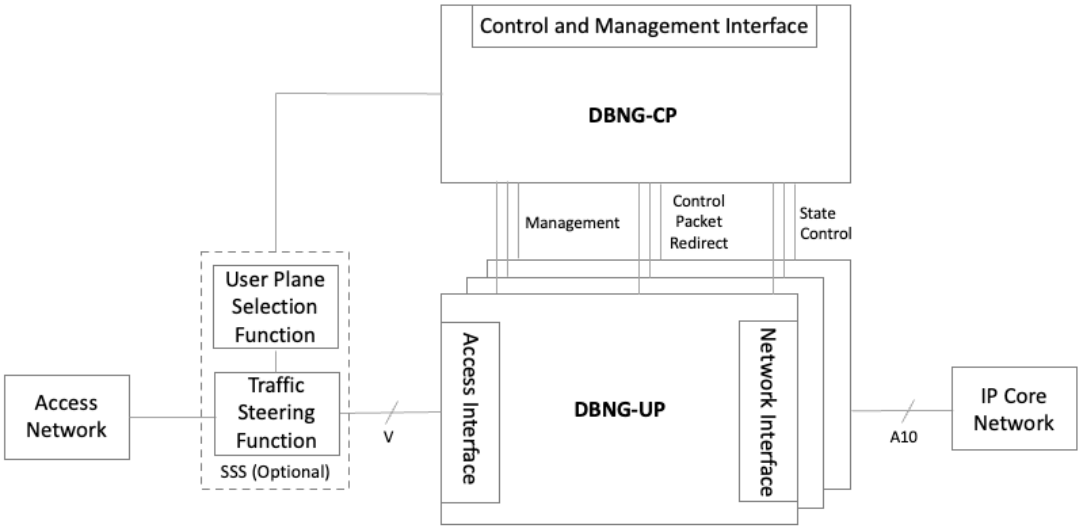


Figure 2 - Disaggregated BNG with Control and User Plane Separation (CUPS)

2.2. DBNG On-premises Deployment and Workload Characteristics

The DBNG technology has been productized by several vendors and deployed in broadband operators' data centers over the last few years. Figure 3 shows a typical DBNG deployment model with geo-redundancy where control planes are deployed as redundant instances managing multiple user planes per cluster which are connected to southbound optical line termination (OLTs). DBNG-UPs also include provider edge routing capabilities connected to the core network such as VRF support with BGP routing and different transport options such as VLANs, MPLS or SR-MPLS. In a typical broadband access network, OLTs communicate with BNG-UPs using L2 (802.1Q or QinQ) with different topology option, directly connected through local switching infrastructure or utilizing an L3 overlay (MPLS Pseudowires, EVPN over SR-MPLS or SRv6) to carry L2 frames. Backup paths are provisioned on the access network to protect against network connection or BNG-UP failures as well as between BNG-UPs and BNG-UPs with redundant CPs.

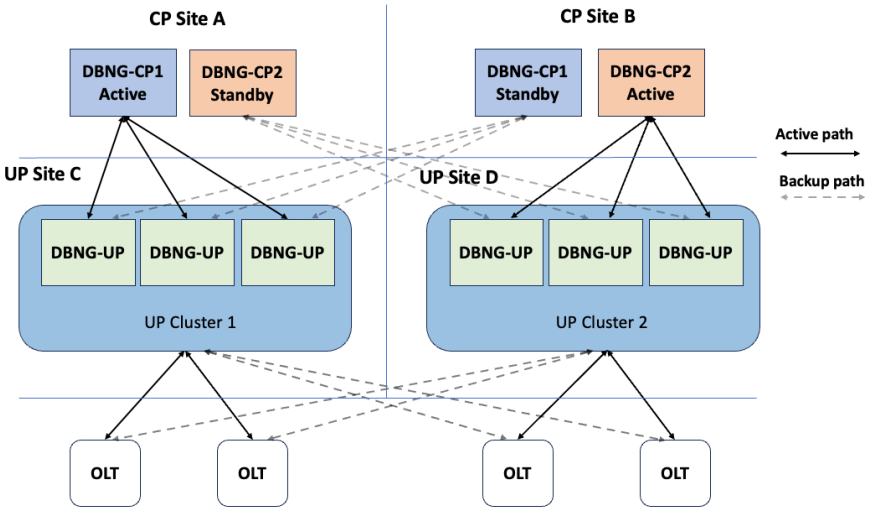


Figure 3 - DBNG Deployment Model with Geo-redundancy

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In most cases with distributed architectures, operators design geo-redundant sites for User Plane and Control Plane instances to protect against site failure. DBNG-UPs are deployed with N+1 redundancy within the same site or across sites. Figure 3 depicts the case with resilient User Plane clusters per site with Active and Standby Control Planes on different sites. Other architectures are possible.

The DBNG-CP redundancy is based on utilizing the same CP IP address by Active and Standby CPs as described further in this paper.

DBNG-CPs and DBNG-UPs are deployed on x86 servers as bare metal or virtual machines (VMs) or Kubernetes Pods as vBNG, scaling throughput from 40 Gbps to 400 Gbps per server or virtual instance.

Table 1 lists typical broadband access requirements for DBNG.

Table 1 - BNG Workload Characteristics

BNG Performance Metrics	Typical Requirement
BNG-UP throughput per instance	40 Gbps to 400 Gbps depending on network design involving user plane distribution level and subscriber capacity per user plane
BNG-CP throughput between BNG-UPs and BNG-CP	1 Gbps to 5 Gbps per DBNG-CP
OLT to BNG-UP latency	25 ms for standard services, < 10 ms for low latency services
BNG-CP to BNG-UP latency	Up to 100 ms
BNG-UP interfaces	QinQ encapsulation from OLTs IPoE (DHCP) and PPPoE

3. DBNG Cloud Solution Design

3.1. Design Approach and Goals

Our approach to DBNG cloud transformation is built on two technology pillars – the DBNG CUPS architecture described above and Cloud Continuum as illustrated in Figure 4. The cloud continuum extends Cloud infrastructure and services from Region to Edge and Far Edge to support applications with different timing constraints or local residency requirements. The region, edge, and far edge are inter-connected and have the same "look and feel" for cloud infrastructure services such as networking, compute, storage, command line interface (CLI) commands, and deployment automation.

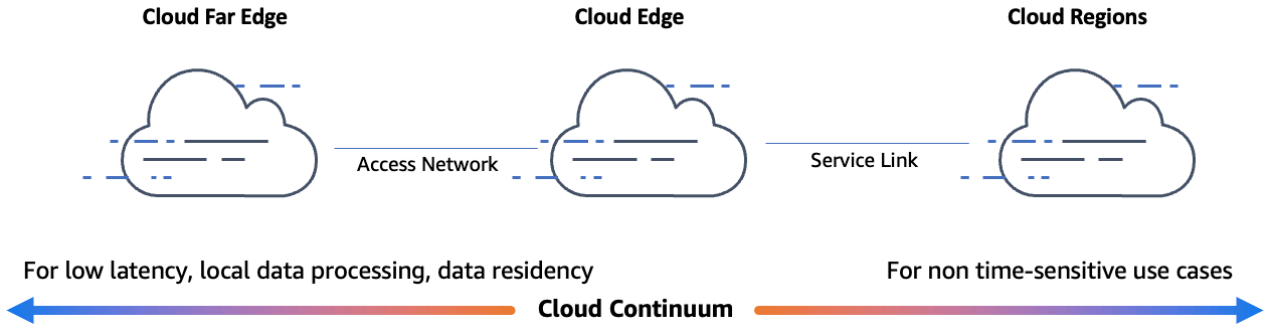


Figure 4 - Cloud Continuum: Region, Edge, and Far Edge

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

With the Cloud Continuum infrastructure and the CUPS architecture, DBNG control plane and user plane functions can be distributed across cloud region, edge, and far edge to achieve several goals:

- Higher DBNG resiliency – Cloud Edge or Region is used for Disaster Recovery (DR) of BNG storage or UP functions. In case of BNG equipment failure or natural disasters in operator's central office on premises, the BNG instances in the cloud are activated to process subscriber traffic.
- Dynamic capacity scaling – When subscriber traffic load surges due to, for example, a major sport event, Cloud Edge or Region provides extra DBNG-UP capacity to handle the bursty traffic on demand.
- Multi-access edge computing (MEC) support – Cloud Edge is used to host a variety of time-sensitive MEC applications, such as AR/VR 3D rendering, gaming, and video streaming. The BNG-UP co-located at the edge directs the subscriber traffic to/from the MEC applications locally without going out to Internet.
- Cost saving – It is the cloud elasticity that provides DBNG-UP capacity on demand. Operator doesn't need to invest in building central office facilities for DR or handling traffic load bursting, thus saving capital expenses.

Allocation of BNG-CP and BNG-UP functions along the cloud continuum can be of different schemes. Figure 5 illustrates a few options to achieve the goals listed above. Let us walk through the diagram by the order of label numbers:

1) It is a baseline to deploy the DBNG-CP in cloud, in a region or edge, whichever is geographically closer to the access network aggregation point(s).
2) One or more DBNG-UP instances may reside in the network operator's central office on premises as it is today. The DBNG-CP functions communicate with the DBNG-UP functions through IP network connections.
3) One or more DBNG-UP instances are deployed in the cloud region (or the cloud edge) to serve the purpose of DR or capacity on demand.
4) One or more DBNG-UP instances are deployed in the cloud edge to serve latency-sensitive MEC applications hosted at the cloud edge.

5-6) To support Subscriber Session Steering (SSS), the SSS Control Plane (SSS-CP) is co-located with the DBNG-CP in the cloud and the SSS User Plane (SSS-UP) performs the traffic steering function at the access network aggregation point on premises.

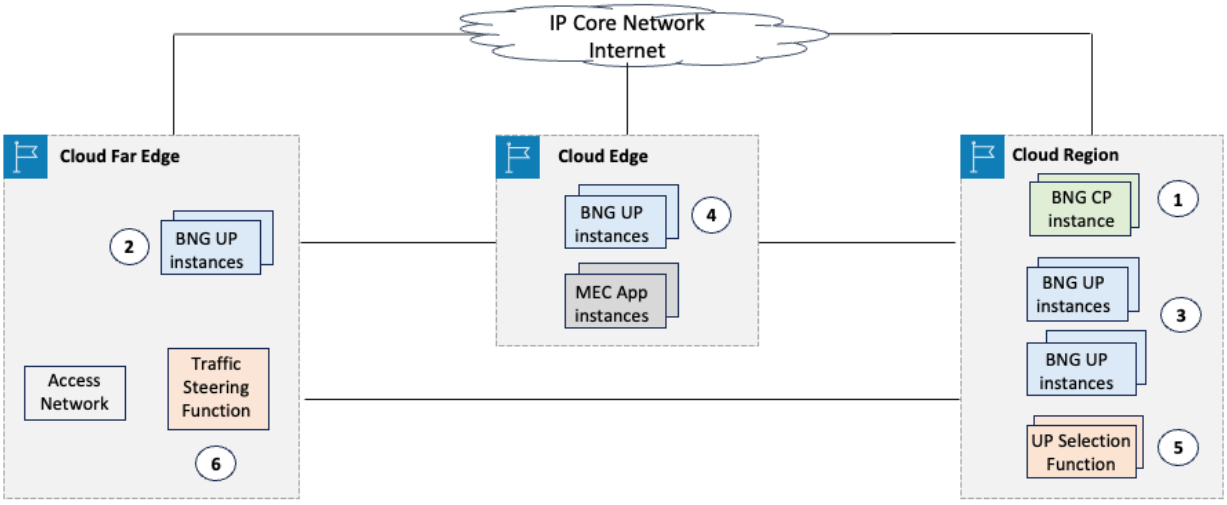


Figure 5 - DBNG Functions Allocation along Cloud Continuum

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Hybrid-cloud solution design is required to realize the DBNG cloud allocation scheme. The solution design must address a number of technical requirements, including (1) connectivity between the access network on premises which is layer-2 and the cloud infrastructure which is typically layer-3, (2) high bandwidth (400Gbps) of the network connections and cloud compute resources, (3) low-latency access from the Optical Line Termination (OLT) in the access network to the cloud DBNG-UP, (4) fast cloud DBNG-UP activation time (30sec.), and (5) high availability of DBNG-UP and DBNG-CP functions in the cloud.

3.2. Reference Architecture and Solution Options

This section provides a hybrid-cloud solution design for the DBNG cloudification. It shows how the solution meets the technical requirements identified above. To describe the solution design in sufficient details, this section uses certain public cloud constructs as examples of a cloud infrastructure.

Figure 6 depicts the solution architecture that reflects the DBNG functions allocation shown in Figure 5. The public cloud is comprised of a region and one or more cloud edge zones which are linked together through dedicated direct connection or site-to-site VPN connection. The region consists of two or more availability zones for cloud facility redundancy and application redundancy. Equivalent to traditional data center is Virtual Provide Cloud (VPC) which hosts virtual machines, storage, networking constructs such as subnets, routers, network address translation (NAT) gateway, virtual private gateway (VGW) for the dedicated connection with operator data center, and Internet gateway (IGW) for traffic to/from Internet. Note that with the cloud continuum model, the VPC is extended from its availability zones in the region to the edge zones as if it was one data center across geographically-distributed region and edge locations, as illustrated by the purple-color box in Figure 6.

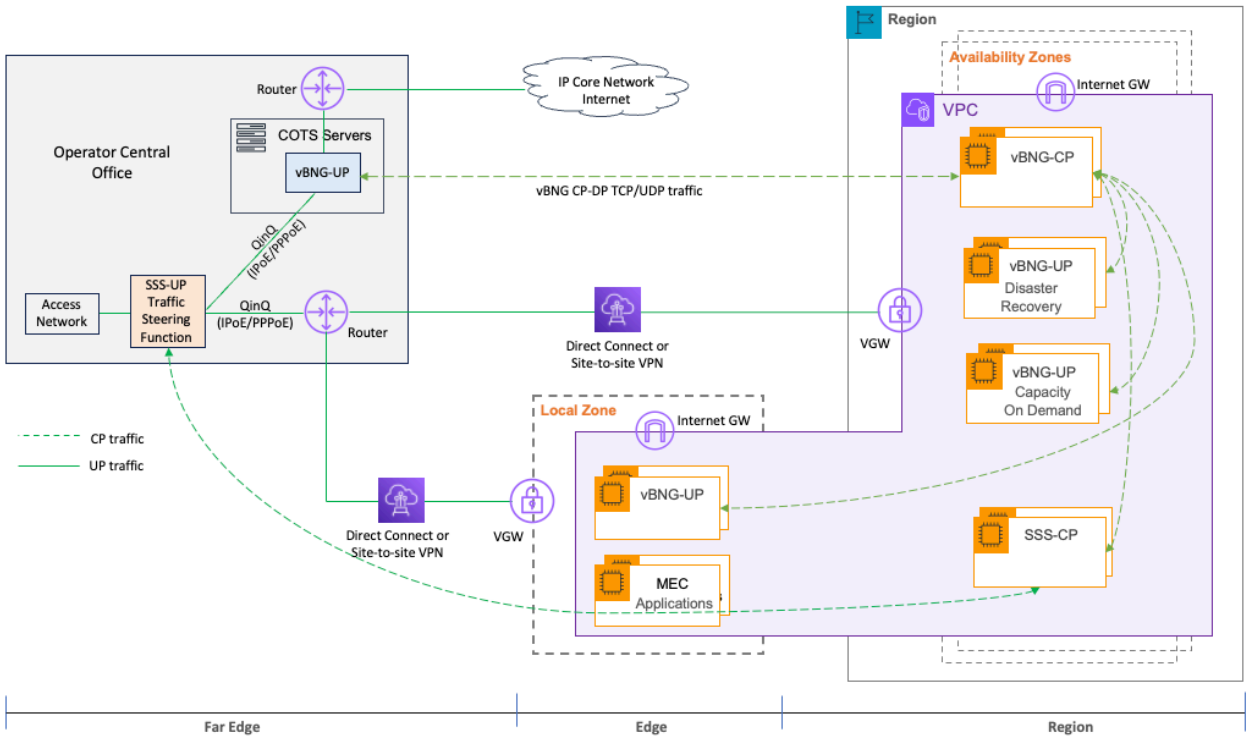


Figure 6 - Cloud DBNG Solution Architecture

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In the cloud, the DBNG-CP and DBNG-UP instances are deployed in the region's availability zones for the DR and capacity-on-demand purposes. The SSS-CP function resides in the region as well. DBNG-UP instances are also deployed at the edge Local Zones for MEC applications. The DBNG-CP, DBNG-UP, and SSS-UP instances in the cloud can be in the form of virtual machines or containers. The virtual machines and containers are in auto-scaling groups for cloud elasticity and reliability.

On the operator premises, DBNG-UP instances run on commercial-off-the-shelf (COTS) servers with standard Kubernetes or any commercial version Kubernetes. The egress subscriber traffic goes through the access network and the session steering function to (1) the local DBNG-UP at normal operation conditions, (2) the DBNG-UP instances in the region for DR, (3) the DBNG-UP instances for additional capacity in the region when the traffic load surges, or (4) the DBNG-UP at the edge for MEC services.

Several architecture components and their design options are described in detail below.

Connectivity between the access network and the cloud – In general, there are two types of connectivity between the access network on premises and the cloud VPC. One is a dedicated connection through a local carrier network and the cloud infrastructure backbone network and the other through Internet with VPN tunnels. Typically, the former can support 100 Gbps or more per connection link and the latter several Gbps per link. Assuming the DBNG-UP throughput is at 200 Gbps of data plane traffic or more from the Access network, multiple links of the direct connection can be used for DBNG-UP in the cloud. Network latency with the dedicated connection from the access network to a cloud edge zone in close proximity is in single digit of milliseconds and tens of milliseconds to a cloud region depending on the proximity among other factors.

DBNG-UP layer-2 support – Running DBNG-UP in the cloud requires support of layer-2 (L2) protocols such as QinQ between the access network and the DBNG-UP instances in the cloud. In general, there are two types of solutions for transporting L2 frames bidirectionally between the access network and cloud: L2 overlay network (e.g., VXLAN, L2TPv3, GRE) or layer-3 routing (e.g., SRv6). Each solution option has its pros and cons as highlighted in Table 2 - L2 Frame Transport Solution Options below. As an example, a VXLAN-based solution design and prototyping are described in Section 4.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 2 - L2 Frame Transport Solution Options

Transport	Feature	Pros	Cons
1. VXLAN	• Encapsulate L2 Ethernet frames within UDP packets	• Provide greater scalability (w.r.t. number of virtual networks) and multicast support compared to L2TPv3	• Has a larger header size compared to L2TPv3, resulting in less efficient use of network resources. • Management overhead of VTEPs
2. L2TPv3	• Encapsulate L2 frames for point-to-point links • Allow multiplexing of multiple PPP session between two IP-connected endpoints	• Relatively simple compared and SRv6 • Has built-in security features that provide encryption and authentication • Built-in redundancy in control plane	• Higher encapsulation overhead compared to GRE • Does not provide any routing functionality
3. BGP EVPN	• Use BGP as a control plane to distribute MAC and IP reachability information for layer-2 and layer-3 connectivity	• Efficient MAC address distribution • Scalability for large-scale deployments with many MAC addresses and endpoints.	• Configurations can be complex, especially for large deployments or using advanced features.
4. GRE	• Encapsulate L2 or L3 packet • Encapsulate a variety of network protocols	• Simpler point-to-point mechanism compared to L2TPv3 • More flexible, can transport a variety of protocols	• can be complex; service chains within a VPC (e.g., virtual router – virtual firewall – intrusion detection) will also need GRE tunneling between them. • Doesn't encrypt traffic. An additional layer, usually IPsec, to protect the traffic. • No redundancy in the protocol
5. SRv6	• Use the IPv6 routing infrastructure • Provide source routing capabilities, allowing packets to take specific paths for traffic flows	• Does not require additional encapsulation beyond the IPv6 packet format, more efficient use of network resources due to smaller header size compared to L2TPv3 and GRE • Good migration strategy to EVPN-VPLS over SRv6 • Offer fine-grained control over the network traffic, enabling more flexible traffic engineering, service chaining, and network programmability	• Not all networks or vBNG ready for IPv6, less versatile compared to GRE

Cloud DBNG-CP High Availability (HA) – High Availability (HA) is an integral part of DBNG-CP deployment in the public cloud. The cloud DBNG-CP must continue to function when a DBNG-CP instance fails. The DBNG control plane maintains the states of subscriber sessions, operational states from user

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

planes, etc. Since DBNG-CP is stateful, it typically requires 1:1 active-standby redundancy for fast stateful switchover in case of DBNG-CP instance failure or planned maintenance. The cloud infrastructure can support DBNG-CP active-standby redundancy with solution options for virtual machine (VM) based software and container-based software, respectively.

1:1 active-standby DBNG-CP VM redundancy is shown in Figure 7. The cloud VM redundancy solution is comprised of several HA mechanisms:

- VM cluster contains two VM instances, one for active DBNG-CP and the other standby DBNG-CP, deployed in two cloud availability zones. The VM cluster provides redundancy at two levels: VM instance and cloud availability zone.
- An HA software component provides HA logic and mechanism (e.g., CP-to-CP keepalive) for detecting DBNG-CP failure, VM instance failure, or availability zone failure and conducting DBNG-CP switchover.
- Session State consistency between the active and standby DBNG-CP VM instances either can be maintained through state replication from the active VM instance to the standby or sharing a persistent storage such as cache or database service in the cloud.
- Overlay IP serves the DBNG-CP interface as a virtual IP address over the IP addresses of the active DBNG-CP VM instance and the standby instance. The overlay IP stays the same no matter which of the DBNG-CP VM instances is active. The overlay IP can be realized by using a route table in the cloud VPC, where the overlay IP is the traffic DBNG-CP destination address and the target IP address is the active DBNG-CP instance's interface address. At the time of switchover, the HA cluster software replaces the route table's the target IP address with the IP address of the newly active DBNG-CP instance.
- BGP Anycast may be utilized in order to support redundant DBNG-CPs with the same IP address in different locations in Active-Standby mode.

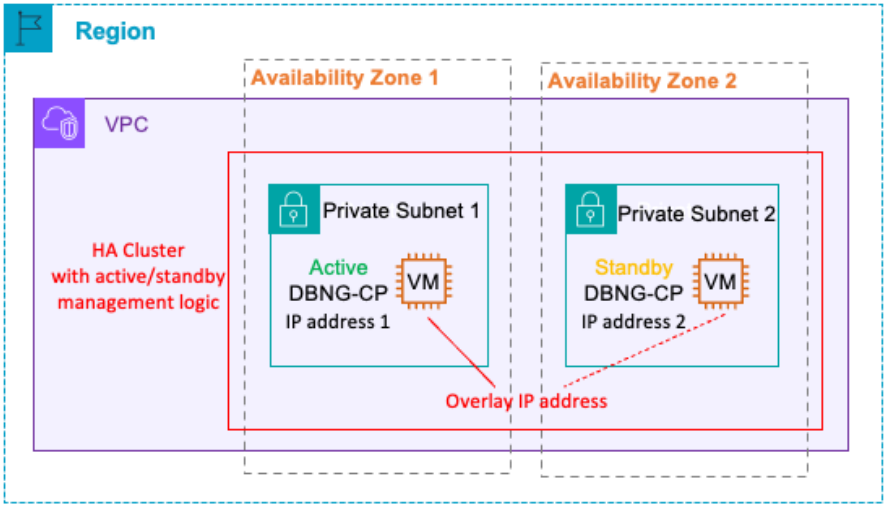


Figure 7 - 1:1 Active-Standby DBNG-CP VM Redundancy in Cloud

For examples of active-standby VM redundancy implementation in public cloud, please refer to references [7] and [8].

1:1 active-standby DBNG-CP Kubernetes Node & Pod redundancy is shown in Figure 8. The redundancy solution with Elastic Kubernetes Service (EKS) is comprised of several HA mechanisms:

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- EKS cluster contains two EKS nodes (EC2 instances) deployed in two cloud availability zones. An EKS pod in each of the nodes runs DBNG-CP with one in active mode and the other standby mode. The EKS cluster provides redundancy at three levels: EKS pod, EKS node, and cloud availability zone.
- HA software provides HA logic and mechanism for detecting DBNG-CP failure, pod failure, node failure, or availability zone failure. For example, control plane instances implement a keepalive mechanism to detect failures and trigger switchover.
- Session State consistency between the active and standby DBNG-CP pods can be maintained either through state replication from the active pod to the standby or sharing a persistent storage such as cache or database service in the cloud.
- Single IP address serves the DBNG-CP interface. There are different methods to implement the single IP. For example, the Kubernetes Service IP can be used over different DBNG-CP instances to guarantee that the active DBNG-CP utilizes the Service IP to communicate with DBNG-UPs. Alternatively, redundant DBNG-CP instances employs BGP Anycast configured with the same IP address and different preference for the active and standby nodes.

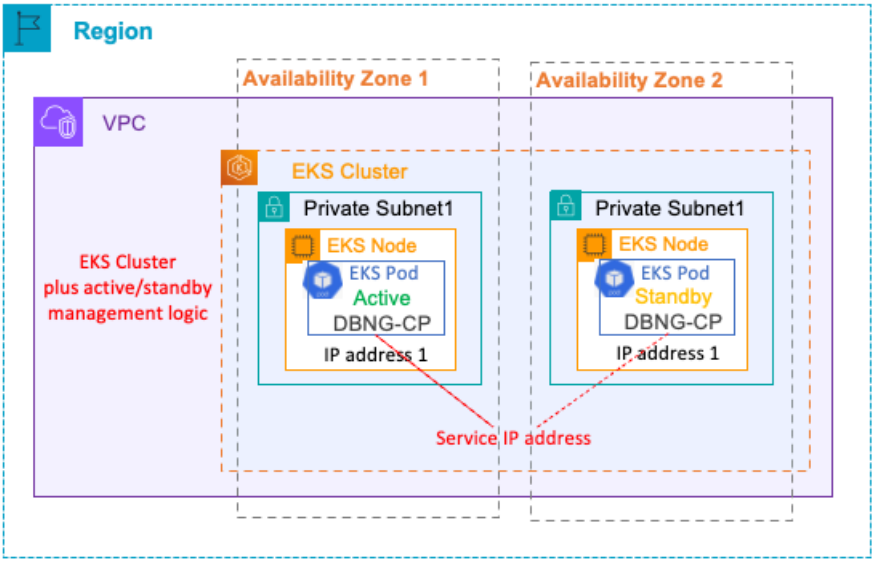


Figure 8 - 1:1 Active-Standby DBNG-CP Pod and Node Redundancy in Cloud

DBNG-UP Redundancy – To support the on-premise DBNG-UP disaster recovery, backup DBNG-UP instances are deployed in the public cloud as illustrated in Figure 9. This on-premise and in-cloud DBNG-UP deployment can be characterized by an active-standby redundancy model where the on-premise DBNG-UP is in active mode and the cloud DBNG-UP in standby in normal operation conditions; at the time of on-premise DBNG-UP failure, the cloud DBNG-UP becomes active.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

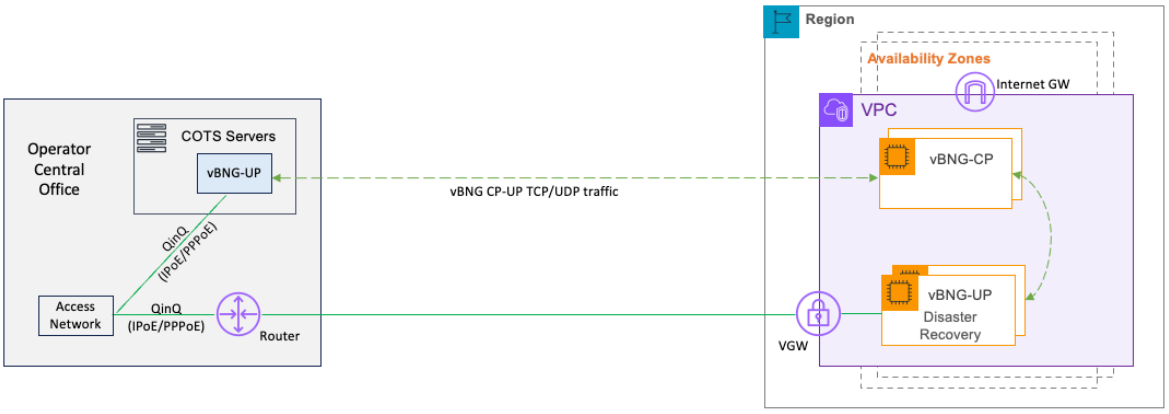


Figure 9 - DBNG-UP Redundancy across On-premise Datacenter and Public Cloud

The active-standby DBNG-UP redundancy can be achieved with Virtual Router Redundancy Protocol (VRRP). The VRRP software should be deployed in the DBNG-UP instances on premise and in the cloud. Note that the connection between Access Network OLT and DBNG-UP instances is layer-2. Hence, the VRRP configuration requires a unique MAC address on each of the DBNG-UP instances.

At the DBNG-UP level, resilience can be implemented with Subscriber Groups (SGRP) in different user planes effectively achieving N+1 redundancy. Figure 10 shows the DBNG-UP resiliency model. A subscriber group is a group of subscribers sharing the same policy and redundancy scheme. The subscriber groups are instantiated on Primary and Standby User Planes. When sessions are established on a primary subscriber group, the same session is programmed by the active control plane on the secondary user plane. For disaster recovery, Primary DBNG-UPs are on premises while Standby DBNG-UPs re on public cloud.

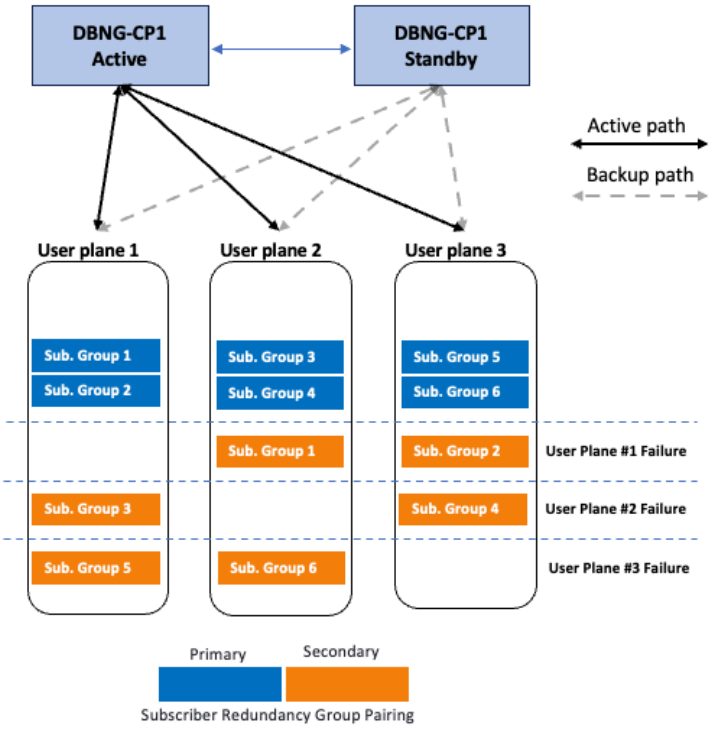


Figure 10 - Subscriber Groups across DBNG-UPs

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Cloud DBNG-UP activation and DR life cycle

Sessions are established on the primary Subscriber Group (SGRP) on the on-premise DBNG-UP and are programmed on the backup cloud DBNG-UP. Upon primary UP failure, the backup DBNG-UP takes over all SGRPs and existing sessions are steered to the new UP based on the virtual MAC move.

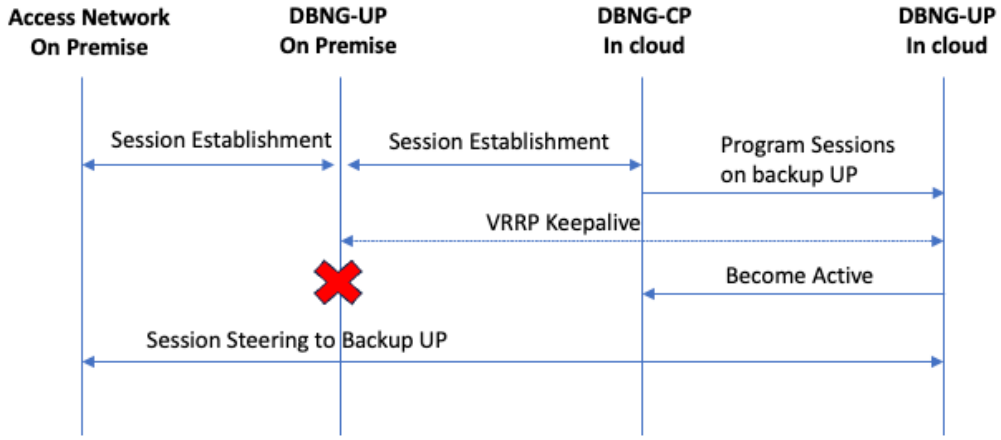


Figure 11 – DBNG-UP Disaster Recovery Life Cycle

Cloud DBNG cost analysis – The cloud DBNG consists of both DBNG-CP and DBNG-UP deployed in the public cloud with the former running in an active mode all the time and the latter in a standby mode. The standby DBNG-UP can be configured as Hot standby or Warm standby. In the hot standby mode, the DBNG-UP instances, i.e., VMs or Kubernetes Pods, are instantiated and in "running" mode. At the time of disaster recovery or traffic bursting from the on-premise data centers, the DBNG-UP instances start processing the subscriber sessions in near real time. In the warm standby mode, the DBNG-UP VMs or Kubernetes Pods are instantiated but in "stopped" mode. It may take a few minutes to get the stopped DBNG-UP instances up running and processing the subscriber traffic. With the cloud "pay-as-you-go" model, the DBNG-UP instances in the "running" mode incurs cloud service charges whereas the instances in the "stopped" mode does not.

Figure 12 shows a cost comparison between the DBNG-UP hot standby and warm standby. It is modeled with VM instances or EKS work nodes with aggregated 200 Gbps network bandwidth and 3 days processing time per month. The chart shows the cost of warm standby DBNG-UP is only 10% of the cost of hot standby DBNG-UP. Clearly, it is a tradeoff between the timeliness of cloud DBNG-UP activation and the cost of running DBNG-UP in the public cloud. It is up to broadband operators to make a tradeoff decision with cloud DBNG service cost and subscriber impact in mind.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

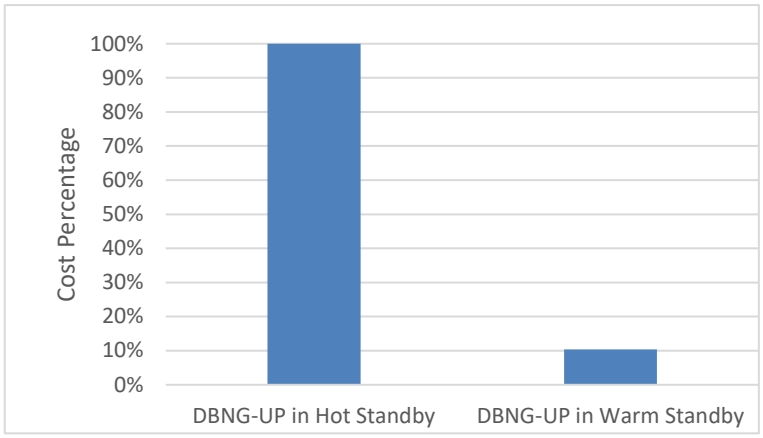


Figure 12 - Cost Comparison of DBNG-UP Standby Modes

4. Solution Prototyping and Implementation Methods

This section shares our first-hand experience in the solution prototyping using a public cloud and commercial disaggregated DBNG-CP and DBNG-UP.

The solution in prototyping is shown in Figure 13. It resembles the generic solution architecture pattern described in Section 3.2. In particular, a lab serves as an on-prem data center and a cloud region is used as the public cloud. A dedicated cloud network interconnects the lab on premises and the cloud region.

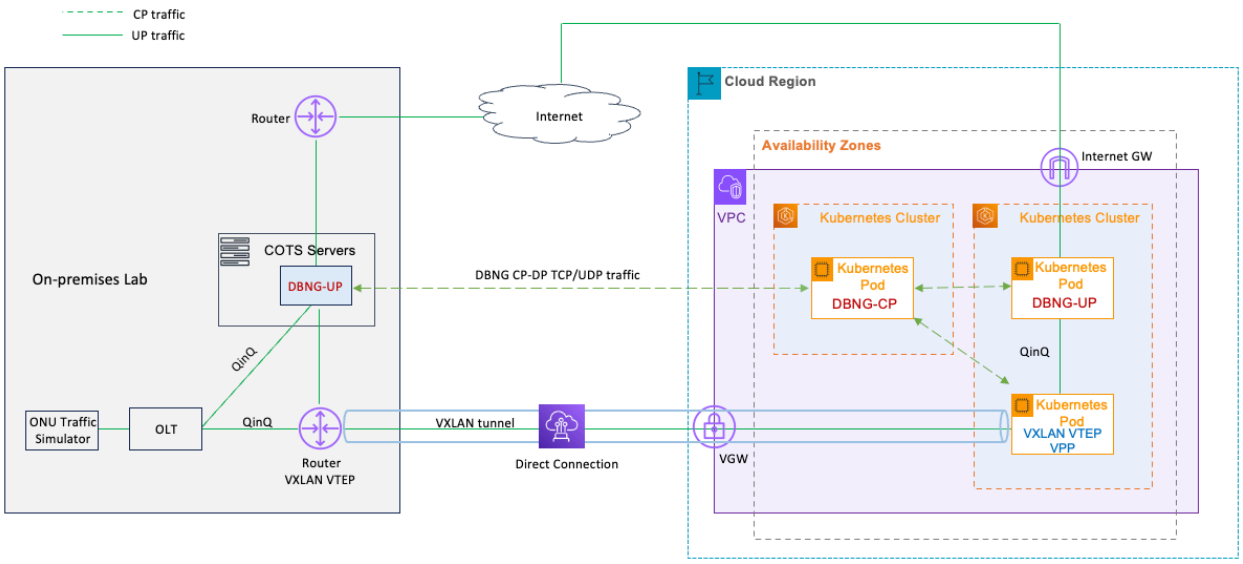


Figure 13 - Prototyping Environment

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In the cloud region, a Virtual Private Cloud (VPC), an equivalent of a data center, is created with Internet Gateway (IGW) and subnets associated with security rules and IP traffic routes. The subnets host the DBNG-CP container and DBNG-UP container in Kubernetes clusters, respectively. A Virtual Private Gateway (VGW) interfaces the direct connection link with the Lab. The DBNG-UP traffic flows between the DBNG-UP and the internet through IGW.

Realizing the layer-2 traffic transport between the layer-2 access network on premises and the layer-3 IP network in the public cloud is a major challenge in cloudifying DBNG-UP. Out of the layer-2 tunneling solution options listed in Section 3.2 Table 2, the VXLAN tunneling method is chosen for the prototyping. The VXLAN method is largely driven by existing vBNG deployment cases in operator environments. The VXLAN implementations in the cloud and on premises are described respectively in the following.

(1) In the public cloud, the VXLAN tunnelling is implemented in the Elastic Kubernetes Service (EKS) cluster as illustrated in Figure 14. It utilizes four software components: Data Plane Development Kit (DPDK), Vector Packet Processor (VPP), Layer-2 Bridge, and Tap interface.

- *DPDK* – is a set of open-source libraries and drivers designed to accelerate packet processing tasks in networking applications. DPDK provides a collection of optimized functions that allow software applications to efficiently interact with network interface cards (NICs) and perform packet processing operations directly on data plane hardware. By bypassing the traditional kernel-based networking stack, DPDK enables applications to achieve high throughput and low latency, making it particularly suitable for building high-performance networking solutions, such as routers, switches, and virtualized network functions like DBNG-UP.
- *VPP* – is a high-performance software framework designed for network data packet processing. It's often used in networking applications and network function virtualization environments. VPP is engineered to efficiently handle large volumes of network packets by optimizing packet processing tasks using vectorization techniques. This allows for improved throughput and reduced latency in network communication.
- *Layer-2 Bridge* – is a networking device or software function that operates at the data link layer (Layer 2) of the OSI model. Its primary purpose is to connect and forward Ethernet frames between two or more network segments or interfaces, effectively extending the same local area network (LAN) across multiple physical or virtual network segments.
- *Tap interface* – is a virtual network interface connected to a bridge device. This configuration is often used in virtualization environments to enable communication within VMs or between a container and the VM that is hosting it. In some instances, it can also facilitate communicating between VM internal and external networks.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

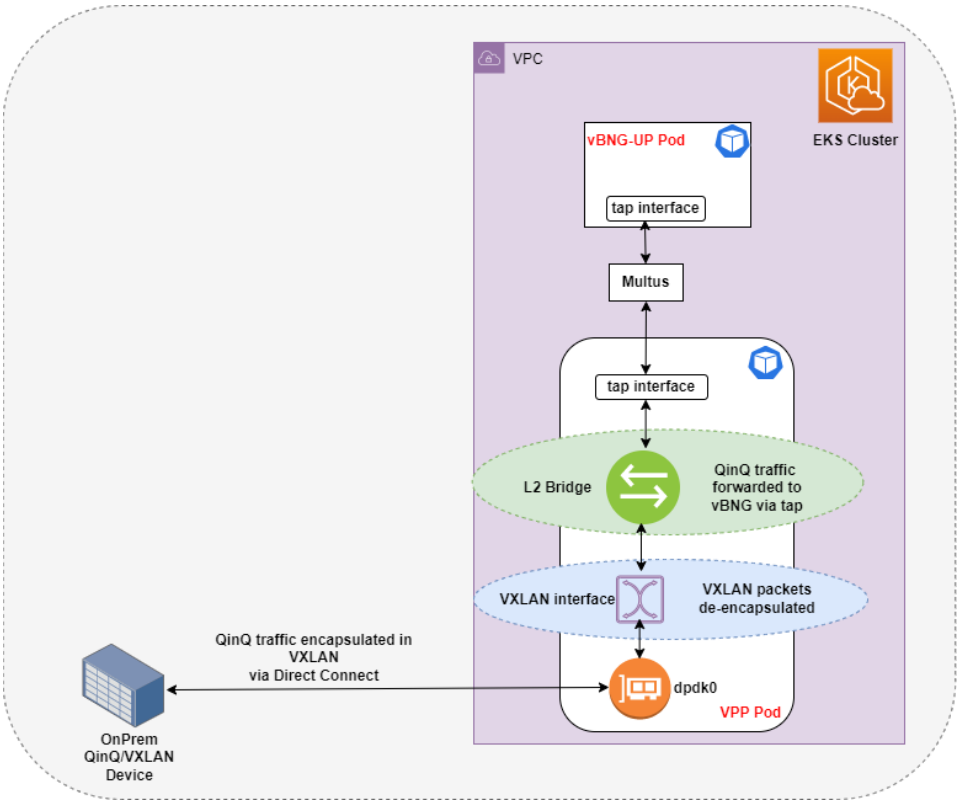


Figure 14 - VXLAN Implementation Method Using VPP

In the prototype, VPP is deployed in an EKS Pod. It provides the following functions:

- VXLAN Tunnel End Point (VTEP) for encapsulating or decapsulating layer-2 frames transported between the cloud DBNG-UP and the OLT on premises.
- A DPDK interface for the VXLAN traffic.
- A tap interface for transporting the QinQ frames between the VPP Pod and the DBNG-UP Pod in the EKS cluster.
- A bridge to connect the VXLAN VTEP and the tap interface

The VPP configuration is shown below with sample data:

```
set interface ip address dpdk0 10.0.0.115/24
set interface state dpdk0 up
ip route add 0.0.0.0/0 via 10.0.0.1 dpdk0
create vxlan tunnel src 10.0.0.115 dst 192.168.0.100 vni 1000 decap-next l2
set interface state vxlan_tunnel0 up
create bridge-domain 10
create tap id 1 host-if-name tap1
set interface state tap1 up
create sub-interface tap1 100 dot1ad 100
set interface state tap1.100 up
set interface l2 bridge vxlan_tunnel0 10
set interface l2 bridge tap1 10
set interface l2 bridge tap1.100 10
```

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The VPP pod configuration is defined by a Network Attachment Definition (NAD) file shown below. The configuration uses Kubernetes' host-device plugin. It specifies the device parameter with the tap interface created by the VPP.

```
apiVersion: "k8s.cni.cncf.io/v1"
kind: NetworkAttachmentDefinition
metadata:
  name: vpp-tap-net
  namespace: encaps
spec:
  config: '{
      "cniVersion": "0.3.1",
      "type": "host-device",
      "capabilities": { "ips": true },
      "device": "tap1"
    }'
```

The layer-2 traffic flow from the on-prem OLT to the DBNG-UP in the cloud is as follows. The on-prem router with VXLAN VTEP sends QinQ frames encapsulated in UDP packets to the VPP pod where the VPP VXLAN VTEP decapsulate the QinQ frames from the UDP packets. Through the internal bridge, the VPP sends the QinQ frames to the vBNG-UP pod via the tap interface. The tap interface is also created in the VPP pod, one leg of the tap interface is connected to the VPP internal bridge while the other leg is injected into the EKS worker node's kernel space, from which Multus is then used to pass this tap interface into the vBNG-UP pod.

(2) On premises, VXLAN VTEP is available in commercial products or open source. As an example, VyOS is an open-source network operating system that can be deployed in servers on premises to provide VXLAN VTEP as well as router, firewall, VPN, and other network functions. To test the VXLAN tunneling first in a cloud environment, the prototype uses VyOS Amazon Machine Image and creates a VyOS VM instance to mimic an on-prem VXLAN VTEP. The VXLAN tunneling was successfully tested between the VyOS VM and the EKS Pod with VPP end to end. For the actual VXLAN VTEP on premises, the prototype uses a commercial product.

5. Conclusions

Public cloud possesses the characteristics of resource scalability, on-demand provisioning, pay-as-you-go pricing, and global accessibility. Because of these capabilities, wireline broadband operators may leverage the public cloud for DBNG use cases such as disaster recovery and capacity on demand. Using the cloud instead of building additional data centers will reduce capital expense and increase operation flexibility. In this paper, we have presented a reference architecture for deploying DBNG control plane and user plane in public cloud region or at cloud edge. Our prototyping experience as well as design description indicate that it is potentially feasible to run DBNG-CP and DBNG-UP in the public cloud for DBNG disaster recovery and capacity on demand use cases. Our next steps are to (1) conduct performance scaling test under failover and workload bursting conditions and (2) implement SRv6-based layer-2 frame tunneling to further demonstrate the public cloud feasibility for operator's DBNG production deployment.

Acknowledgements

The authors would like to thank Dr. Jennifer Andreoli-Fang for her valuable comments and edits in the paper content. The authors also knowledge the Ciena engineering team for their prototyping work to gain the first-hand experience in the vBNG solution deployment on premises and in the cloud end to end.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

BNG	broadband network gateway
DBNG	disaggregated broadband network gateway
DBNG-CP	DBNG control plane
DBNG-UP	DBNG user plane
DPDK	data plane development kit
EKS	elastic Kubernetes service
SGRP	subscriber group
vBNG	virtual broadband network gateway
VM	virtual machine
VPC	virtual private cloud
VPP	vector packet processor
VRRP	virtual router redundancy protocol

Bibliography & References

[1] Broadband Forum TR-459 Issue 2, "Multi-Service Disaggregated BNG with CUPS: Reference Architecture, Deployment Models, Interface, and Protocol Specifications," April 2023, https://www.broadband-forum.org/pdfs/tr-459-2-0-0.pdf

[2] Broadband Forum WT-474, "Introducing Subscriber Session Steering and dynamic subscriber placement," https://www.broadband-forum.org/wt-474-introducing-subscriber-session-steering-and-dynamic-subscriber-placement

[3] Vodafone and Intel White Paper, "Transforming Fixed Access Using Traffic Steering on a Cloud Native Architecture," Nov., 2021, https://networkbuilders.intel.com/solutionslibrary/transforming-fixed-access-using-traffic-steering-on-a-cloud-native-architecture

[4] Broadband Forum WT-474, "Introducing Subscriber Session Steering and Dynamic Subscriber Placement," 2022, https://www.broadband-forum.org/wt-474-introducing-subscriber-session-steering-and-dynamic-subscriber-placement

[5] Networking Blog, "Connect a VXLAN-EVPN DC to the Public Cloud the right way," June 2020, https://ccie46985blog.wordpress.com/2020/06/21/connect-a-vxlan-evpn-dc-to-the-public-cloud-the-right-way/

[6] Networking Tutorial, "An introduction to the SRv6 (Segment Routing over IPv6 dataplane) technology," December 2017, https://www.segment-routing.net/tutorials/2017-12-05-srv6-introduction/

[7] Blog, "AWS Overlay IP In SAP Landscapes," last updated February, 2023, https://dzone.com/articles/aws-overlay-ip-in-sap-landscapes

[8] AWS Web Site, "Red Hat Enterprise Linux with High availability on Amazon EC2 ," last updated 2023, https://aws.amazon.com/partners/redhat/faqs/high-availability/

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Virtual RAN (vRAN) Over DOCSIS, DOCSIS as a Midhaul

A Technical Paper prepared for SCTE by

Ramneek Bali
Principal Engineer
Charter Communications
Ramneek.bali@charter.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 3
2. Lab Setup 5
3. Results Summary 6
4. Conclusion 7

Abbreviations 8

List of Figures

Title **Page Number**

Figure 1 – vRAN deployment options 4
Figure 2 – vRAN Lab Setup 5
Figure 3 – vRAN Performance Under Varying DOCSIS Load 6

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The objective of this paper is to present real-world results using commercial solutions that integrate Midhaul interface (F1) over DOCSIS and performance under varying traffic demands and user profiles in RAN and DOCSIS networks. The interface between virtualized distributed unit (vDU) and virtualized central unit (vCU) is called F1 or midhaul.

Virtual Radio Access Network (vRAN) enables the move to more disaggregated and virtualized architecture for deploying 5G networks. Three important elements, which are fully disaggregated in RAN, are the Centralized Unit (CU), the Distributed Unit (DU) and the Radio Unit (RU). Building elements of RAN have further been virtualized, vDU and vCU, and can be deployed on over-the-shelf hardware. To see the benefits of the disaggregated vDU and vCU deployment model, vDU with RU can be deployed at a site and vCU can be centralized at a data center or far-edge location.

Virtual Centralized Unit (vCU) provides non-real-time processing and access control. It manages higher layer protocols including Radio Resource Control (RRC) from the control plane, and Service Data Adaptation Protocol (SDAP) and Packed Data Convergence Protocol (PDCP) from the user plane. The vCU is connected between the 5G core network and the vDU. One vCU can be connected to multiple vDUs.

Virtual Distributed Unit (vDU) provides real-time processing and coordinates lower layer protocols including Physical Layer (PHY), Radio Link Control (RLC) and Media Access Control (MAC). Virtualization shifts the vCU and vDU from dedicated hardware to software components, allowing for flexible scaling, as well as rapid and continuous evolution. This virtualization allows the networks to easily meet the evolving demands of new and existing services with minimal impact on the deployment and operation costs. With vDU, all the baseband functions of real-time RLC/MAC/PHY layers are executed over the commercial-off-the-shelf (COTS) server.

Remote Radio Unit (RRU) provides the physical layer transmission and reception, supporting technologies such as Multiple Input Multiple Output (MIMO).

vRAN offers the flexibility of varied deployment options based on the latency requirements. The lowest latency is achieved by deploying the vDU and vCU at the edge cloud, along with the core in the mobile edge compute (MEC)server. Consequently, the user plane is now located nearest to the service available cell and with the core user plane terminated at the edge cloud itself.

A key asset in the cable industry is the hybrid fiber coax (HFC) network. The vRAN architecture split depends on the throughput and latency performance of the HFC network. These requirements will decide which interface (fronthaul, mid (haul or backhaul) is best suited to deploy over DOCSIS. Fronthaul has the most stringent bitrate and latency constraints and backhaul has the least.

Studies and vendor requirements have shown that the enhanced Common Public Radio Interface (eCPRI) based fronthaul transport needs significantly more bandwidth, overhead and one-way latency in the neighborhood of 100 microseconds between the RU and vDU. Hence, front haul is preferred to be on fiber exclusively, while mid-haul and backhaul can be supported by a transport network with 10-100ms latency. These limitations and cost considerations of deploying fiber makes it easier to deploy RU and vDU at cell site location and vCU/core at centralized location.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

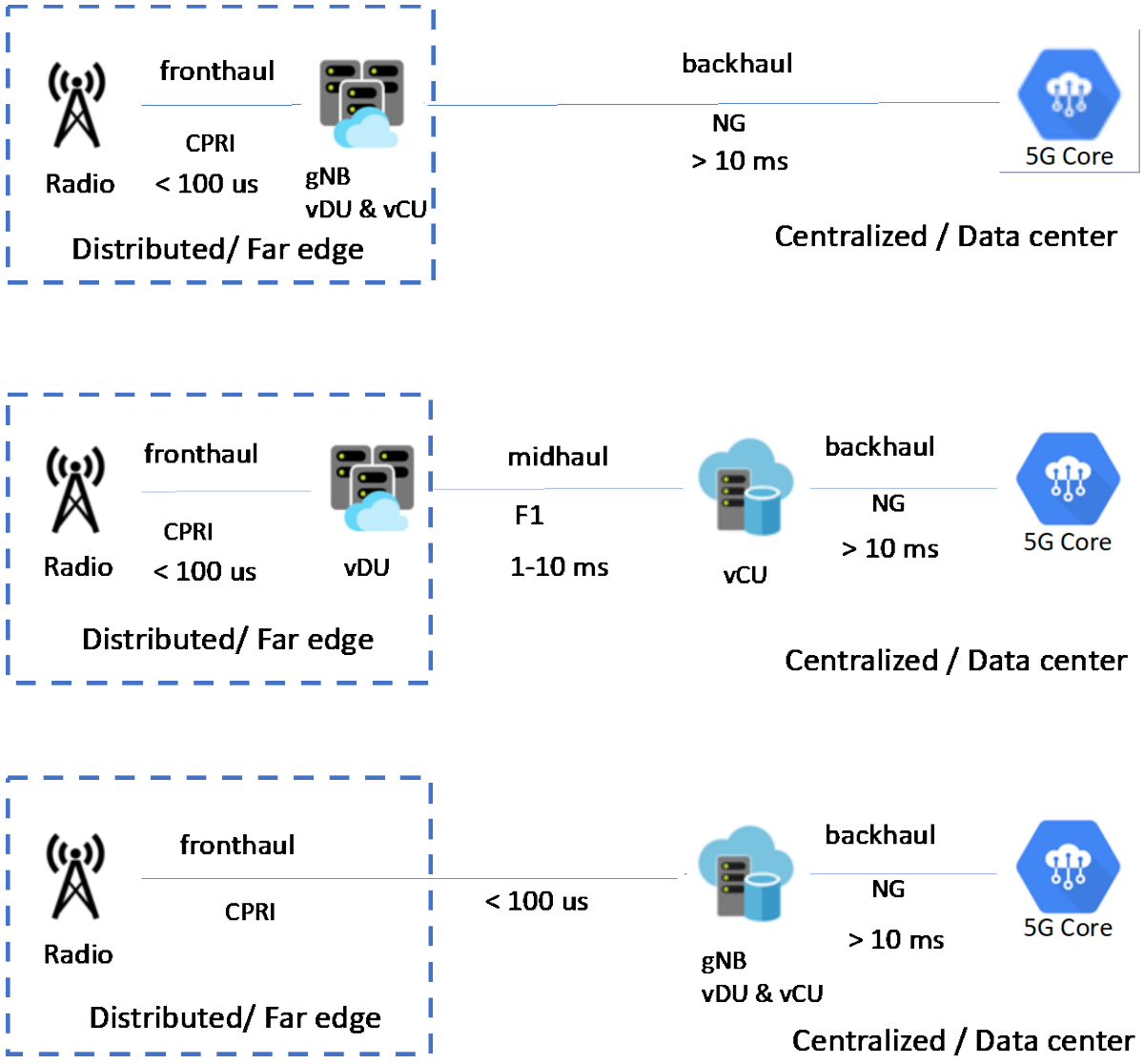


Figure 1 – vRAN deployment options

The interface between vDU and vCU is called F1 or midhaul and has a latency requirement of around 10ms. DOCSIS can play a key role in enabling disaggregated deployments while maintaining the key performance metrics. The latency performance of DOCSIS is best suited for this type of deployment with vDU at the edge and vCU centralized, or at the edge. For now, fiber is required for any fronthaul support and, depending on the availability of the fiber, vDU can be positioned either at the edge or centralized location.

Improvements in new DOCSIS technologies, such as Low Latency Xhaul (LLX), Low Latency DOCSIS (LLD), Active Queue Management (AQM), high split and Low Latency, Low Loss, Scalable Throughput (L4S) will be required to reduce latency to sub 1ms level, making HFC virtually indistinguishable from dedicated fiber for 5G and mid/fronthaul purposes.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

2. Lab Setup

Complete end-to-end vRAN lab was built to validate all the vRAN over DOCSIS, which included a 5G SA virtual core, vRAN, UE simulator and full DOCSIS system. vRAN was deployed on open platforms over COTS hardware and built using open platforms and common over-the-shelf hardware. Once the platform was built, RAN application from various vendors was deployed successfully on it.

The study was done to understand the hardware requirements, software deployments and integration effort required between different vendors. This platform was also used to study the RAN and DOCSIS performance under varying RAN and DOCSIS load.

The figure below shows the lab layout.

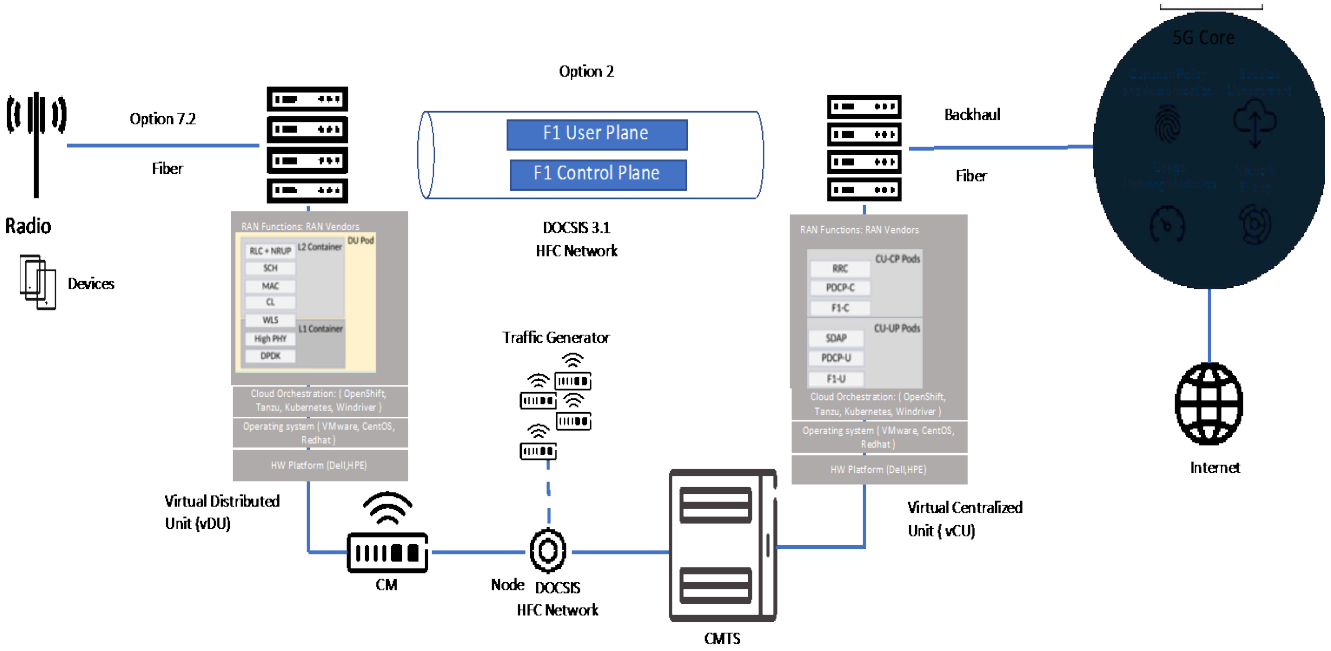


Figure 2 – vRAN Lab Setup

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Results Summary

vRAN system performance was measured for different key areas over DOCSIS for multi-UE, different radio conditions, mobility and varying traffic load.

The figures below show the results for multi-UE under varying DOCSIS load and different DOCSIS configurations.

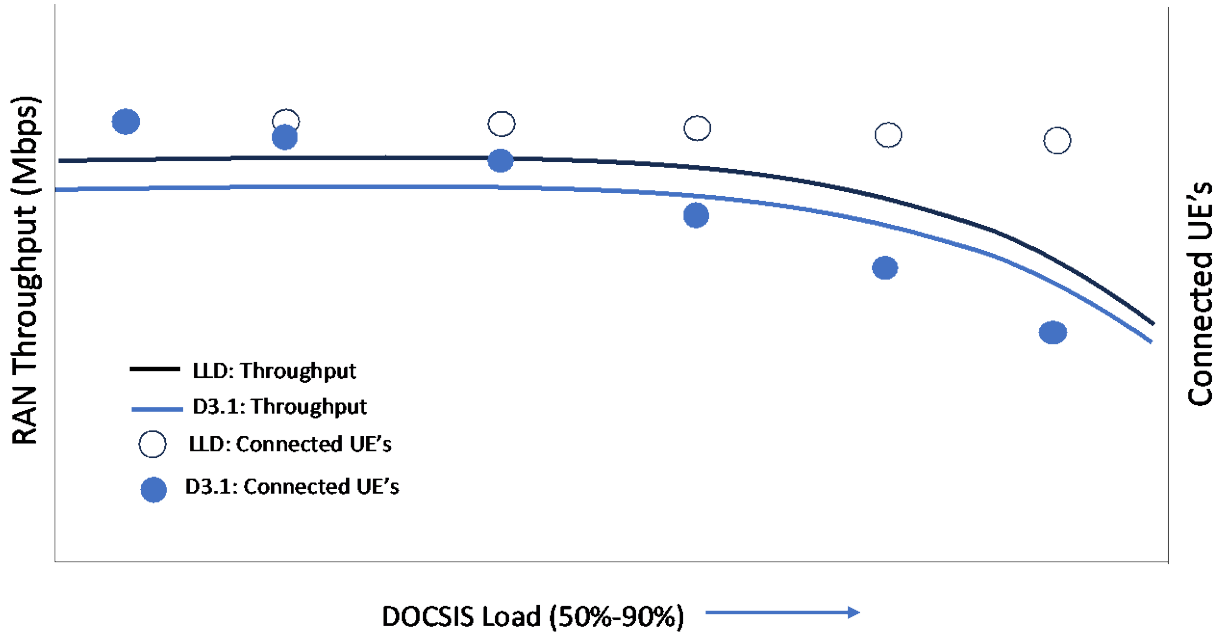


Figure 3 – vRAN Performance Under Varying DOCSIS Load

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

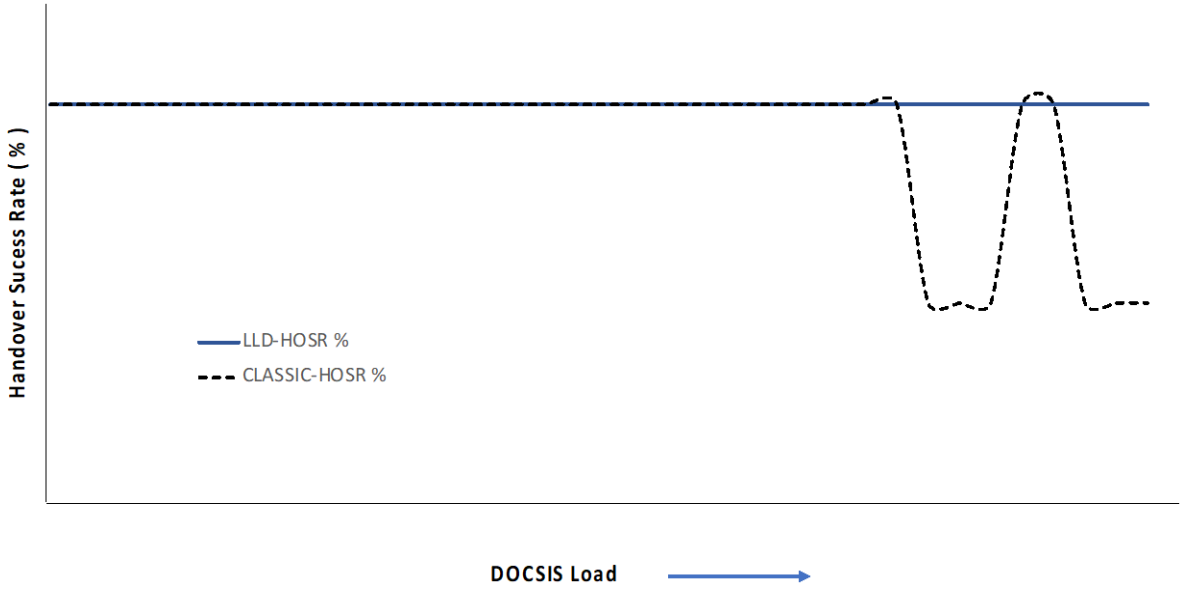


Figure 4 – Mobility vRAN Performance Under Varying DOCSIS Load

Key Observations

- RAN performance is limited by DOCSIS load. At higher DOCSIS load, RAN throughput degrades by more than 100%.
- Increase in latency triggered by DOCSIS load and RAN throughput will mutually degrade both DOCSIS and RAN performance.
- In DOCSIS 3.1, maximum number of simultaneously active UEs per small cell reduces with increasing DOCSIS load as compared to LLD. Contention is observed at CMTS, despite optimized user packet dataflow at CM.
- At high DOCSIS load, able to connect almost all the UEs in LLD as compared to reduction in the simultaneous call establishment for DOCSIS 3.1.
- Signaling traffic (SCTP) being routed via LLD service flow and user traffic is still on default service flow.
- LLD maintains a 100% handover success rate at higher loads as compared to classic DOCSIS, driven primarily by the higher priority given to signaling in LLD.

4. Conclusion

DOCSIS can play a key role in enabling disaggregated deployments while maintaining the key performance metrics. The latency performance of DOCSIS is best suited for this type of deployment with vDU at the edge and vCU centralized, or at the edge. Improvements in new DOCSIS technologies, such as Low Latency Xhaul (LLX), Low Latency DOCSIS (LLD), Active Queue Management (AQM), high split and Low Latency, Low Loss, Scalable Throughput (L4S) will be required to reduce latency to sub 1ms level, making HFC virtually indistinguishable from dedicated fiber for 5G and mid/fronthaul purposes.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

vDU	Virtual Distributed Unit
vCU	Virtual Centralized unit
RU	Radio Unit
MEC	Mobile edge compute
HD	high definition
Hz	hertz
K	kelvin
SCTE	Society of Cable Telecommunications Engineers

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Virtualization in Wi-Fi to Fix Many Long-Standing Customer Experience Issues

Using Virtual BSS (VBSS) To Address Complex and Time-Consuming Client Roaming in Multi-Access Point Networks

A Technical Paper prepared for SCTE by

Ian Wheelock
CTO Europe
CommScope Inc
Cork, Ireland
Ian.Wheelock@CommScope.com

With additional support from CableLabs technical personnel

Steve Arendt, CableLabs Advanced Technology Group

Steve Glennon, CableLabs Advanced Technology Group

John Bahr, CableLabs Advanced Technology Group

Tucker Polomik, CableLabs Advanced Technology Group

Zackary Foreman, CableLabs Advanced Technology Group

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Problem Space and Solution 4
2.1. Wi-FI Evolution And Roaming 4
2.2. Understanding The Problem 5
2.3. Wi-Fi Multi-Access Point Networks 5
2.4. Client Roaming In The Home 6
2.5. Client Roaming At The Edge Of The Network 6
2.6. RF Performance And Client Quality Of Experience (QoE) 7
2.7. Received Signal Strength Indicator (RSSI) & Signal to Noise Ratio (SNR) 7
2.8. Varying RSSI Impact On Maximum Performance 8
3. Client Roaming Between Access Points 9
3.1. Controller-led Client Roaming Between APs 9
3.2. Allow/Deny List Wi-Fi Client Steering 10
3.3. BSS Transition Management Wi-Fi Client Steering 10
3.4. Possible Consequences Of Steering Wi-Fi Clients 11
3.5. Wi-Fi Basics for Associating with an SSID 12
4. Virtualizing The BSSID, The Basis Of The Mobile Wi-Fi Solution 13
4.1. Previous Efforts for VBSS 14
4.2. BSSID Capacity In APs For VBSS 14
4.3. Mobile Wi-Fi Technology 15
4.4. Mobile Wi-Fi Basic Operation 15
4.5. VBSS Lifecycle 18
4.6. VBSS Controller Management 18
4.7. VBSS Transition Flow 18
5. How Effective is Mobile Wi-Fi 19
5.1. Mobile Wi-Fi Testing Environment 19
5.2. Normal Wi-Fi Association 20
5.3. VBSS Testing Approach 20
5.4. Wi-Fi Alliance EasyMesh VBSS Steering Extensions 21
5.5. Detailed VBSS Transition Analysis 22
5.6. ICMP Ping Testing Approach 23
6. Mobile Wi-Fi Results 24
6.1. Mobile Wi-Fi Steering Performance 24
6.2. What has been done and what is next 24
6.3. Challenges 25
6.4. Mobile Wi-Fi Opportunities Beyond Residential 25
7. Conclusion 26

Abbreviations 27

Bibliography & References 28

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

List of Figures

Title **Page Number**

Figure 1, Multi-AP Home With Internal Client Wi-Fi Roaming Issues 6
Figure 2, Multi-AP Home With External Client Wi-Fi Roaming Issues 7
Figure 3, Typical Associated Wi-Fi Performance Readout (Including RSSI) 8
Figure 4, Impact of Wi-Fi Client Movement From AP On RSSI And PHY Rate 8
Figure 5, Wi-Fi Client Roaming To A New AP, Ensuring Performance 9
Figure 6, Mesh Controller Forcing Allow/Deny List For Client Steering (to AP 3) 10
Figure 7, Comparison Of 802.11v And 802.11r FT Steering 11
Figure 8, Example Of SSID="Home" Mapped Across All The Radios Of Two APs, With 3 different BSS Formed 13
Figure 9, Typical BSSID Mapping Per Radio, With SSID And VBSSID Assignments 14
Figure 10, Client A Transmissions Received By Multiple APs, Only Decrypted By Associated BSS 16
Figure 11, VBSS Transition Requires The BSS Security Context To Be Copied From The Origin AP To The Target AP 16
Figure 12, VBSS/Security Context Operational On The Target AP, Deleted from the Source AP 17
Figure 13, VBSS Transitioning From Broadcast Beacons to Unicast Beacons 17
Figure 14, Main Test Network Description Using VBSS Running In prplMesh On GL.Inet B1300 Mesh Router 19
Figure 15, Standard Client Association To VBSS 20
Figure 16, View Of RSSI During Test Run, Including RSSI At Time of AP Transition 21
Figure 17, Wi-Fi EasyMesh VBSS Exchange To Support Steering Of VBSS between APs 22
Figure 18, Transparent Move Of Client from AP1 to AP2, Including Antenna Signal Strength Detail And BSSID 22
Figure 19, Simplified Test Environment To Deduce Performance Of VBSS Steering 23
Figure 20, Mobile Wi-Fi Key Milestones 26

List of Tables

Title **Page Number**

Table 1 – ICMP Test Example, 192.168.1.233 Is Client Being Steered 23

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Wi-Fi® has become the primary way that people and their many devices connect to the internet at home and on the go. As homes have more devices and are getting bigger, people are adding more Wi-Fi access points to their networks to improve coverage and capacity. Devices often switch from one access point to another as they move around, whether within the homes or when moving between home, work, and elsewhere.

However, how devices move around the home has been a long-held issue within Wi-Fi, even with a variety of solutions to the problem. Devices like smartphones, tablets, and wearables sometimes stay connected to an access point even when it cannot provide a good enough connection. This can be extremely frustrating for users who expect Wi-Fi to be as reliable as their cellphones.

Mobile Wi-Fi solves this Wi-Fi roaming issue by creating a virtual basic service set (VBSS) for each device that is moving around. This technology, developed by CableLabs and shared with groups like the Wi-Fi Alliance, forms the basis of the solution. A reference implementation, which was used to produce the results in this paper, has been designed and contributed to the prpl Foundation prplMesh project. The decision of which access point handles the VBSS for a device is made by an intelligent Controller function (with interfaces to the multiple access points in the Wi-Fi network), and not by the device. This allows devices to roam between access points smoothly, with the Controller making sure the device connects to the best access point and radio. The system can also disconnect the device if the Wi-Fi network cannot handle its needs anymore, such as when a device departs from home. What is important is that this solution works with older devices too, without needing changes or newer Wi-Fi features on the device. Mobile Wi-Fi will improve the user experience in homes, apartment complexes, businesses, and, eventually, in a city-wide manner to ensure completely seamless Wi-Fi mobility.

2. Problem Space and Solution

2.1. Wi-FI Evolution And Roaming

Back when Wi-Fi was first introduced as a wireless alternative to Ethernet, no one really thought about the idea of Wi-Fi roaming. The idea of smoothly transitioning between different Wi-Fi access points was not a consideration at that time. The challenges of making Wi-Fi roaming work well did not pop up first in homes. Instead, they showed up in places like offices and similar places where there were several Wi-Fi access points on the same network.

Fast forward to the last six to seven years, many homes have adopted the use of multiple Wi-Fi access points too, mostly to improve the performance of their networks, dealing with range and throughput issues. In such an environment, with multiple access points and large numbers of connected devices, with many mobile devices moving around while running video calls, video streaming or gaming Wi-Fi roaming problems have become an issue.

In the past, these issues were mostly a concern for businesses. They had to figure out how to make sure devices stayed connected as people moved around with those devices. But now, regular homes are dealing with similar challenges as they set up multiple access points to cover larger areas or deal with lots of devices all without a dedicated IT (Information Technology) team to make it work. The various business or enterprise solutions can work in a residential setting, but they are far more costly than the typical solutions provided by service providers to their customer. A lot of the cost is in the more advanced APs capable of dealing with much larger numbers of connected devices, as well as alternative security requirements, network management requirements and overall complexity of operating such devices in the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

home. A service provider solution can be based on a standalone in-home setup with an advanced Controller with several basic policies or a cloud-based solution, with a cloud Controller. Both are a lot less complex than an enterprise Wi-Fi solution.

To sum it up, the journey of Wi-Fi roaming, from being an overlooked detail when Wi-Fi was born to becoming a critical aspect of modern networks, shows how far technology has come. The shift from businesses to regular homes facing these challenges emphasizes its growing importance. With more home devices being connected to the Internet, more IoT (Internet of Things) devices in the home, then dealing with the complexities of Wi-Fi roaming will remain a big task, pushing the need to find new and better ways to make sure wireless connections stay strong and seamless.

Over the past ten years, there has been a significant increase in the number of devices that people carry around or use while moving around their homes or workplaces. It all started with adding Wi-Fi to smartphones and music players like the iPod Touch. This has now expanded to a whole plethora of devices like fitness gadgets, tablets such as the iPad or MS Surface, cheaper laptops like Chromebooks, pet trackers, robot vacuum cleaners, home security cameras, smart speakers, and even VR (Virtual Reality) headsets like the Meta Quest 2. Many of these devices are often used with demanding workloads while people are moving around their homes or offices. In the past, when the most demanding thing people did with Wi-Fi was checking emails on a Blackberry phone, there were few problems with switching between Wi-Fi points, and people hardly noticed any issues. But now, with things like streaming videos all around the house or using VR headsets in different rooms, these "smart" devices do not always work well with Wi-Fi, and it can cause problems.

2.2. Understanding The Problem

In the following paragraphs we describe the current situation relating to steering capability of existing networks and their connected clients and how it impacts on overall performance of the network and customer quality of experience.

2.3. Wi-Fi Multi-Access Point Networks

This paper assumes that the home Wi-Fi network already exists and is utilising some form of interconnectivity between multiple APs that allow for seamless connectivity between APs. Such a network could be based on an IEEE 802.11s mesh network or a Wi-Fi Alliance Certified Wi-Fi EasyMesh™ network [1], [2] or a proprietary meshing scheme. In any case, the network would already be established with APs connected (either via wireless or wired connections) and enable continuous connectivity throughout the home for clients. In this paper, the primary issue being tackled is when a client remains connected to an access point or frequency band that is not the optimal choice. For instance, when a device moves around a home with multiple access points, it should switch from one access point to another to maintain a connection with the strongest signal, the least crowded one, or the one offering the best throughput. Traditionally client devices decide on what AP to use, basing their decision on signal strength of available APs and some other historical data. The client device has no understanding of how the wireless network is connected, or how heavily loaded each potential AP might be (with other clients and/or network traffic). Connected Wi-Fi networks, based on a mesh approach, can propose to the client device that it switch to a specific AP (considering all the network conditions across the different APs in the network), however it is up to the client whether it will accept the suggestion and switch, or stay connected to the existing AP. In other words, the client has the final say whether it moves or not.

In real-life networks, this process varies among devices and does not always proceed smoothly for users. For example, a client device might persist in staying connected to a weaker access point even when a better one is available. Imagine being on a video call over Wi-Fi and moving from one room to another. If

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the client device continues to stay connected to the original access point, the call's quality could suffer due to slower speeds. Ideally, the client device should switch to the closer access point, but this often does not occur. Clients might hold onto a specific access point even when it is not delivering the best performance (commonly referred to as a "sticky client" situation). To the user, this results in applications or experiences that do not function well due to terribly slow data speeds.

2.4. Client Roaming In The Home

To explain this further, a typical multi-AP network in a home with two APs and a Controller are shown below. The performance of client A connected to AP 1 degrades as the client moves away from AP 1 in the Wi-Fi network. Even though the client ends up being adjacent to AP 2, it persists with its connection to AP 1, leading to a degraded experience. This is known as the “sticky client” problem, where the client persists with the poorer connection. This not only causes mediocre performance for the client but can also impact all other clients connected to the same AP, as well as clients in overlapping APs that use the same radio band and frequency.

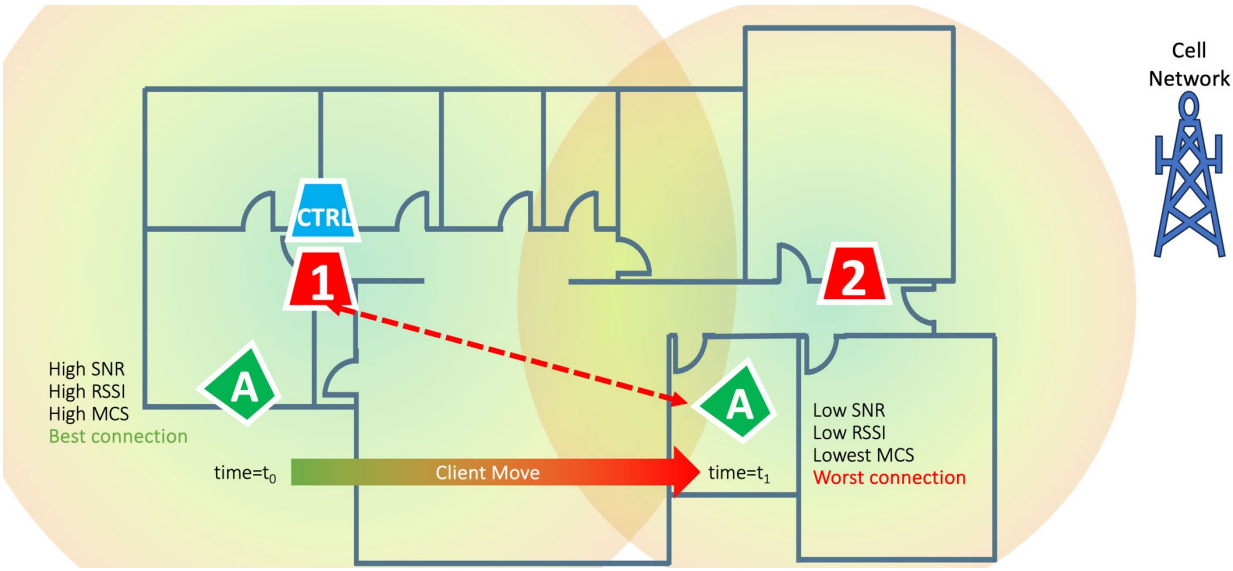


Figure 1, Multi-AP Home With Internal Client Wi-Fi Roaming Issues

2.5. Client Roaming At The Edge Of The Network

A similar problem occurs when there is no suitable AP to transition to when a client reaches the edge of coverage of the Wi-Fi network. In this case, the best course of action for the client would be to terminate the Wi-Fi connection altogether which would trigger the connection to switch over to a different interface such as cellular in the case of a mobile phone. However, clients often stubbornly cling to a Wi-Fi connection that no longer serves their needs. For example, a user gets in their car while still on the video or teleconference they start at home on their phone and starts to drive away. Instead of switching to the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

cellular interface, the phone clings to the fading Wi-Fi signal, resulting in a frozen conferencing application.

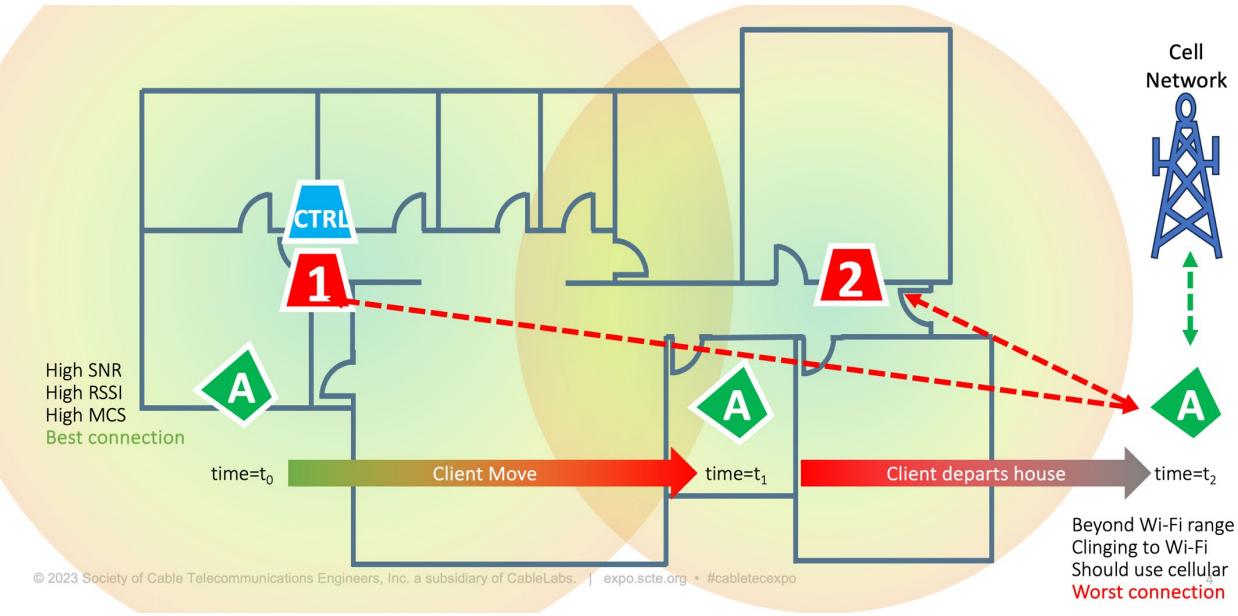


Figure 2, Multi-AP Home With External Client Wi-Fi Roaming Issues

2.6. RF Performance And Client Quality Of Experience (QoE)

The movement of the client throughout a home complicates Wi-Fi performance. RF signals used in the 2.4, 5 and 6 GHz frequency bands are subject to attenuation. This could be free space path loss, where there are no obstacles between an AP and a client, and the RF signal strength simply diminishes over distance due to physics. Another source of attenuation are obstacles such as walls (internal or external), insulation, windows, furniture, people, etc that all contribute to the attenuation of the RF signal – some obstacles are worse than others. The distance between the AP and client and any obstacles in between all impact the PHY rate that can be used – defined by a range of different Modulation and Coding Schemes (MCS). Depending on the version of Wi-Fi, the MCS options are numbered starting with MCS 0 that defines the lowest performance, ranging all the way up to the maximum performance available in Wi-Fi 7 using MCS 13. MCS0 offers a maximum PHY data rate of 72 Mbps for a 160 MHz wide Wi-Fi channel, while MCS11 gets up to a maximum of 1,200 Mbps. Higher Wi-Fi data rates are also possible when multiple spatial streams (SS) are used, which are enabled using Multi-Input/Multi-Output (MIMO) processing, with a typical AP supporting up to 4 SS per radio. In the case of a 160 MHz Wi-Fi 6 with 4 SS, the maximum PHY rate is 4,800 Mbps. These are maximum PHY rates for the Wi-Fi channel in use, and cover all communication, both uplink and downlink as Wi-Fi is half-duplex. This is different compared to 2.5 Gbps Ethernet which is full duplex and offers 2.5 Gbps for transmit and receive, for a total of 5 Gbps capacity.

2.7. Received Signal Strength Indicator (RSSI) & Signal to Noise Ratio (SNR)

The Received Signal Strength Indicator (RSSI) and the Signal to Noise Ratio (SNR) are used to determine what MCS can be used for data transmit/receive. While the RSSI value is a well-recognised metric that provides a simple measure of signal strength in Wi-Fi networks, most Wi-Fi mesh networks will use a more standardised metric, called the Received Channel Power Indicator (RCPI) instead. RCPI provides a method for measuring signal strength that also considers the noise and interference levels of

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the Wi-Fi channel. RCPI is also especially useful in dynamic conditions (such as devices moving swiftly through a multi-AP residential network), helping to make better decisions about which channels to use or what access points to connect to. Throughout the rest of this document for the sake of simplicity RSSI is used as the measure of signal strength, but RCPI is the actual measure that a mesh Controller expects to receive to optimise client steering.

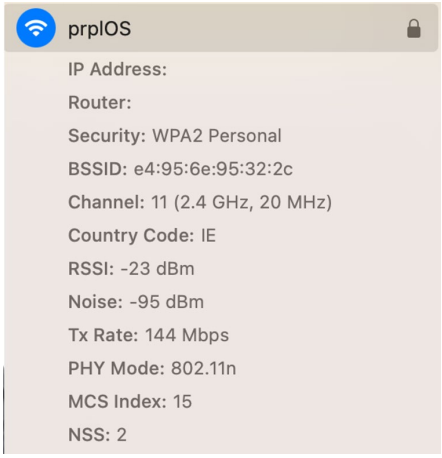


Figure 3, Typical Associated Wi-Fi Performance Readout (Including RSSI)

2.8. Varying RSSI Impact On Maximum Performance

The MCS defines what the maximum PHY rate is allowed. For static devices operating in a static environment, the RSSI/SNR will rarely change, allowing for the same MCS to be achieved; however, any Wi-Fi client that is moving around, or even held differently (e.g., landscape vs portrait viewing on a tablet), then the RSSI/SNR (and subsequently MCS) will change. The following diagram shows how the Received Signal Strength Indicator (RSSI) can change as a client moves away from an AP. This has a corresponding impact on the PHY rate (the diagram assumes that there is sufficient SNR for each MCS).

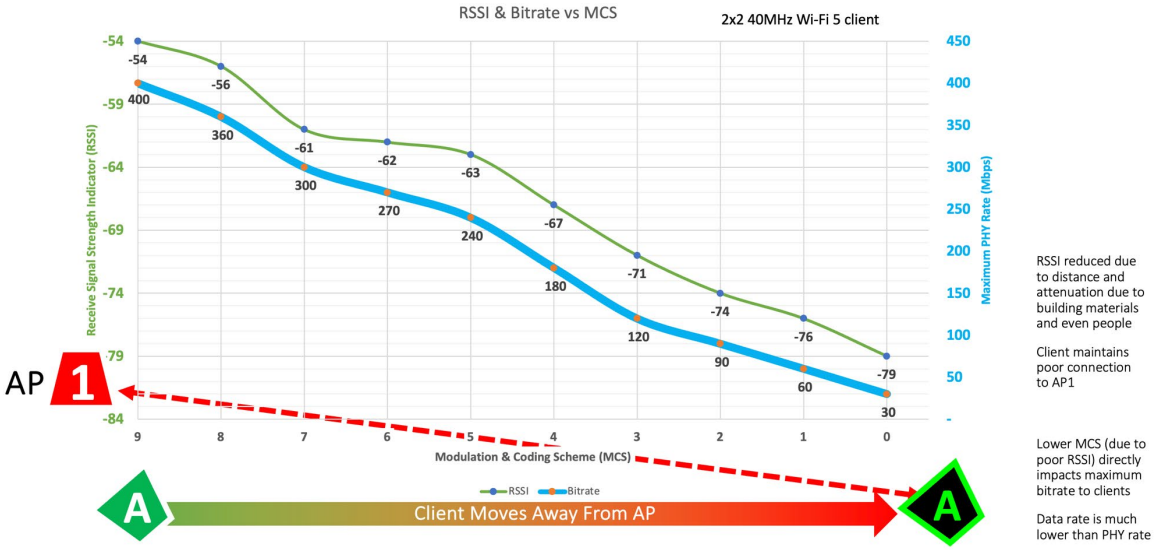


Figure 4, Impact of Wi-Fi Client Movement From AP On RSSI And PHY Rate

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The lower the RSSI (assuming sufficient SNR), the lower the achievable PHY rate is. It must also be understood that Wi-Fi is half-duplex, meaning the PHY rate is the total capacity of the link (both uplink and downlink). In addition, Wi-Fi is a contention based MAC (Media Access Control) (for the most part) meaning that there are delays before clients or APs can transmit data due to how clients must check for quiet time on the RF/air interface before attempting to transmit, meaning that the suggested PHY rate does not translate 100% into usable data rate.

3. Client Roaming Between Access Points

As the client moves away from the AP it is associated with, its constantly examining network conditions to determine what action to take to ensure best performance. One action might be to switch from 5GHz to 2.4GHz (it has longer range), or another action might be to connect to an AP with better RF performance.

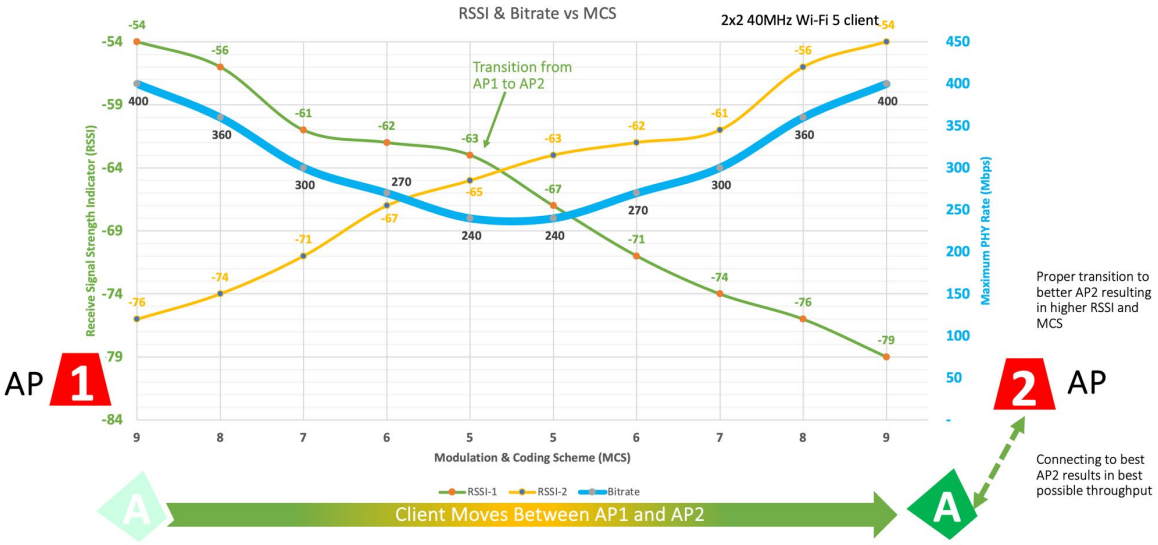


Figure 5, Wi-Fi Client Roaming To A New AP, Ensuring Performance

The above chart shows how an ideal client (maybe one that does not exist) can decide to roam to a different AP to maintain the highest performance levels. As the client moves, it identifies the available AP, and makes its own steering/roaming decision to disassociate from AP 1 and (re)associate with AP 2.

3.1. Controller-led Client Roaming Between APs

The Wi-Fi client has limited visibility into the organisation of the Wi-Fi network, what APs are connected where, if they use Wi-Fi backhaul or Ethernet backhaul, how loaded they are with other clients, etc. As a result, decisions it makes concerning roaming from one AP to another are primarily informed by the RSSI/SNR it can detect for different APs. In a lot of mesh networks, there are Controller entities that are used to help form the network of APs as well as deal with the operation of the network, with respect to optimising the performance of the connected APs and connected clients. The network provides detailed statistics from the APs and clients to the Controller to assist in forming a model of the network and clients. The Controller can use its model to identify opportunities for improving performance, particularly for clients that are mobile, or dealing with APs that have gone offline, or dealing with overlapping/ neighbouring Wi-Fi networks that use common RF frequencies.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In the case of clients that need attention (i.e., need to be moved to a new frequency band or a new AP), the Controller triggers a steering event that uses a variety of techniques to try to influence the client. Some of the primary methods that have been developed (there are others) include:

1. Allow-List/Deny-List
2. 802.11v BSS Transition Management (BTM)

3.2. Allow/Deny List Wi-Fi Client Steering

The Allow-List/Deny-List is one of the original approaches developed, relying on the ability of the Controller to tell certain APs in a network to not allow the nominated client from associating with them (as well as disassociating the client if it is associated to one of these APs). The Controller must also advise the target AP that it should allow the nominated client to associate. This brute force approach for steering does not require any extra Wi-Fi features with a client to work, and has a high chance of success, however some stubborn clients may refuse to re-associate, and drop from the network.

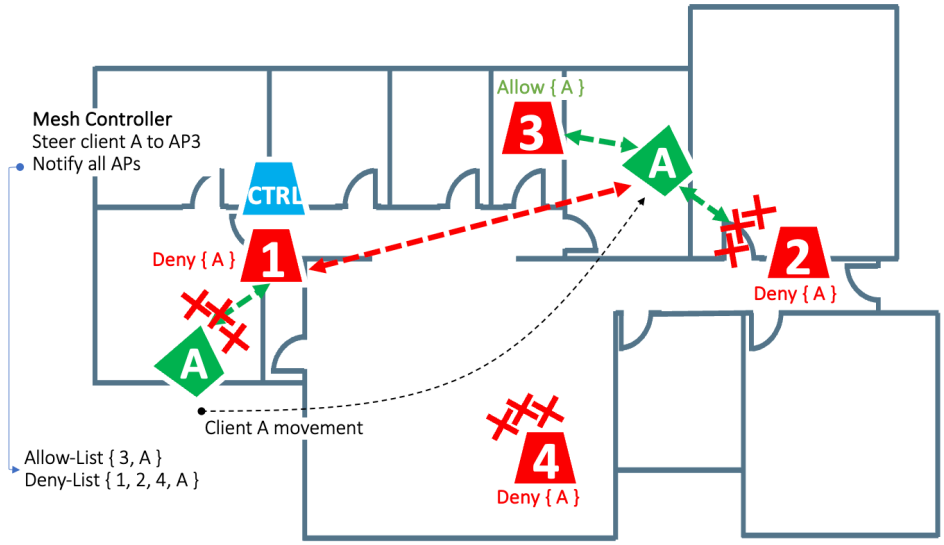


Figure 6, Mesh Controller Forcing Allow/Deny List For Client Steering (to AP 3)

3.3. BSS Transition Management Wi-Fi Client Steering

The BTM method takes a different approach, where the Controller requests the AP, the client is associated with, to send an 802.11 message with a list of other APs that it strongly suggests the client should consider associating with. This approach does require the client to support BTM, which is defined within the Wi-Fi Alliance Agile Multiband specification. There are two versions of BTM, 802.11v and 802.11r Fast Transition (FT). The main difference between these is the shortened time to complete the security establishment following client association. The overall transition time experienced using either of these approaches may be faster or slower depending on the combination of AP and clients involved.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

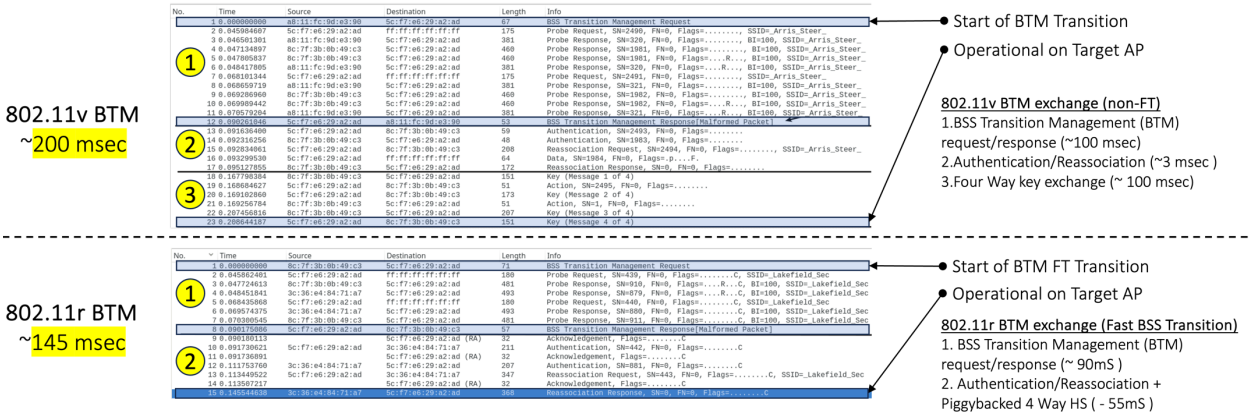


Figure 7, Comparison Of 802.11v And 802.11r FT Steering

3.4. Possible Consequences Of Steering Wi-Fi Clients

These approaches have some unintended consequences when moving clients. There is a tension between what the Controller and AP network request and what the client is willing to do. The client already has its own suite of roaming policies and approaches and receiving steering instructions from the network can often cause some conflict. This is compounded by having a variety of different Wi-Fi clients in the home that all have variations in what protocol support they have and how they react.

Common problems associated with this issue include:

Sticky Client – A client decides to stay associated with a particular AP even though a closer, much better performing AP is available. Due to operating at a very low MCS, this results in poorer performance for the individual client, but can also impact other clients in the same BSS due to the poor performing client consuming more "airtime" than is necessary (that is when a client continuously sends large amounts of data at a low MCS, is can take an unfair share of transmission time compared to other clients).

Ping-Pong Effect – a client may accept a steering instruction from the Controller/AP, stay on the new BSSID for a brief period, then switch back to the original BSSID. This has an obvious impact on client user experience, as the constant switching impacts on user traffic.

Interoperability issues – Some Wi-Fi Clients may have some or no support for BSS Transition Management mechanisms, as a result some may not handle steering requests as smoothly as others, leading to inconsistent results.

Unintended Disconnects – Sometimes clients get confused or give up on steering when they have been steered multiple consecutive times to different BSSIDs (Basic Service Set Identifier), partly because their steering algorithms believe their current connection is acceptable. This can result on clients becoming non-responsive, or worse, client disconnecting from the Wi-Fi network completely, requiring manual user intervention to reconnect.

Unfortunately, these problems are commonplace, and introduce considerable complexity for Controllers in terms of how they treat different clients differently, as well as having to maintain per-client information (device type, OS version, scorecard for different steering mechanisms, preferred steering method, etc). The possibility of the Controller negatively impacting a client (due to above problems) while trying to do good for the client presents a challenge as to how often devices should be steered.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

In addition to challenges with compliance to the steering instructions, excessive transition/steering times also impact on the Quality of Experience (QoE) of the users. If the time is short (<30msec) then the steer is not noticeable. Transition times of less than 100 msec are sufficient for most real-time traffic streams (i.e., the user will not notice the communication gap) however longer transition times disrupt any activity on the mobile client. In addition, highly latency-sensitive traffic (e.g., mobile gaming) can be heavily impacted by even 50 msec of disruption.

As Wi-Fi availability and connectivity are fundamental to a clients Quality of Experience (QoE), risking such connectivity due to steering is a real concern. Having a steering technique that achieves close to 100% steering success, without these problems that might impact on QoE, is unbelievably valuable. This is where Mobile Wi-Fi and Virtual BSS comes in and will be discussed in section 4.

3.5. Wi-Fi Basics for Associating with an SSID

Wi-Fi clients connect to an AP, into what is known as a Basic Service Set (BSS). A BSS consists of a single AP and the client devices connected to that AP. The BSS is identified by a BSS Identifier (BSSID) which is a unique MAC address associated with a radio on the AP. When clients connect to that BSSID they become part of that BSS.

The Service Set Identifier (SSID) is a human readable name assigned to a wireless network. All devices within a BSS share the same SSID. It is common to have the SSID configured on different radios in the same AP (i.e., the same SSID on the 2.4, 5 and 6 GHz bands but each with a different BSSID), enabling all the AP radios to offer access to a specific SSID. The SSID will also be configured on radios of other APs within the network. Using this model means a client has complete choice between different APs in the network and different radio bands when trying to connect to a specific SSID. SSIDs (Service Set Identifiers) are advertised in Wi-Fi Beacon messages, allowing devices to scan available RF channels and collect the SSID/BSSID details to decide which SSID and BSSID they should associate on which AP that are within the range of the device. Note that this constant scanning for other SSIDs and BSSIDs by the client can impact on energy saving options within the client, as it is always checking for a better Wi-Fi connection.

An AP supporting multiple BSSIDs per radio can operate multiple concurrent SSIDs. An AP could support up to 16 BSSID per radio, and potentially have between 2 and 4 radios (Dual Band Concurrent (DBC), Tri-Band Concurrent (TBC) and Quad-Band Concurrent (QBC)), resulting in anywhere from 32 to 64 BSSIDs in a single AP. The following diagram shows the mapping of SSID, BSSID and BSS configured across two APs. Clients (with appropriate radios) can roam to any of the “in-range” BSSIDs configured for the “Home” SSID.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

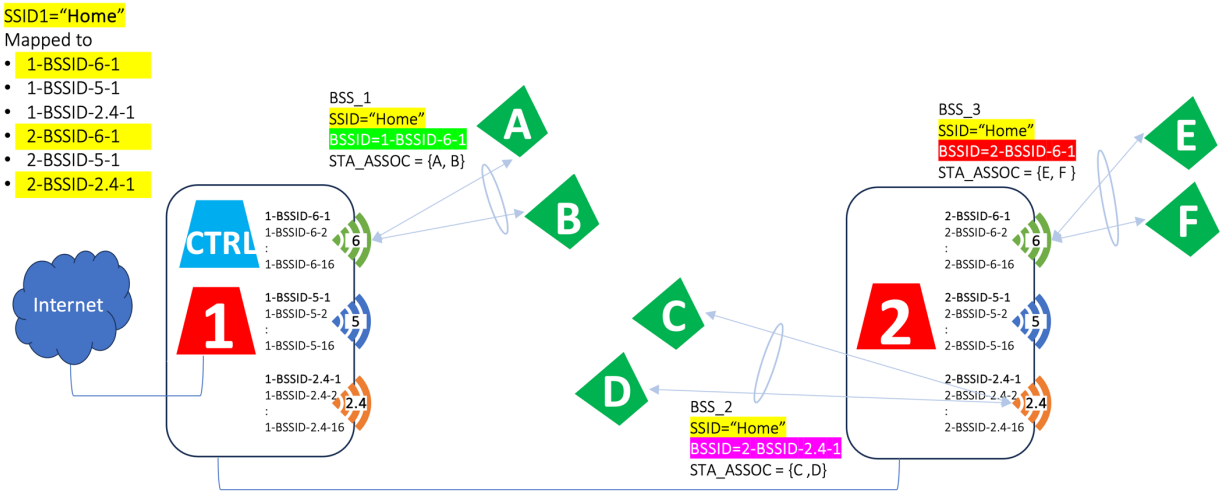


Figure 8, Example Of SSID="Home" Mapped Across All The Radios Of Two APs, With 3 different BSS Formed

When a client joins a BSS by connecting/associating to a specific BSSID, it conducts various MAC exchanges. A crucial exchange relates to establishing the security context for handling encrypted traffic sent between the client and the AP. This exchange results in a Pairwise Transient Key (PTK) for client<>AP transmissions and a Group Temporal Key (GTK) for multicast and broadcast traffic from the AP to clients. All encrypted traffic includes a sequence number, and each side (AP to client and client toAP) maintains separate packet numbers that are used in the creation of the security initialization vector (IV) when encrypting traffic (used to prevent replay attacks).

When Wi-Fi roaming is active, the client can pick any one of the available BSSIDs mapped to its current SSID and decide to roam to it (meaning it must complete Wi-Fi (re)association and repeat the 4-way security handshake for establishing security credentials before transmitting data traffic). Alternatively, a mesh Controller (maintaining details of every AP and the associated SSID/BSSID configuration) can direct a client to steer to an optimal BSSID to get maximum performance. Hopefully, the client complies with the direction. This whole roaming process, be it managed by the client or the Controller or both, continues the whole time the client is connected to the Wi-Fi network, constantly making efforts to ensure the best performance.

4. Virtualizing The BSSID, The Basis Of The Mobile Wi-Fi Solution

The details in the previous section show how standards and networks focus on ways to "request" a client to move from a BSSID on one AP to a BSSID on the same AP or a different AP, but never force the client to obey these directions. An alternative to this model is to move the actual "network" the client is connected to between APs. As mentioned already, when a client moves between BSSIDs, it must repeat the Wi-Fi association process. By focussing on moving the BSSID between APs, the client remains connected to the BSSID, never having to repeat the (re)association process. This is the basis of the "virtual BSS" concept – creating a per-client BSS with a unique virtual BSSID that is not tied to a particular AP, but rather a Virtual BSS that can be moved around a network with the client remaining connected the entire time.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

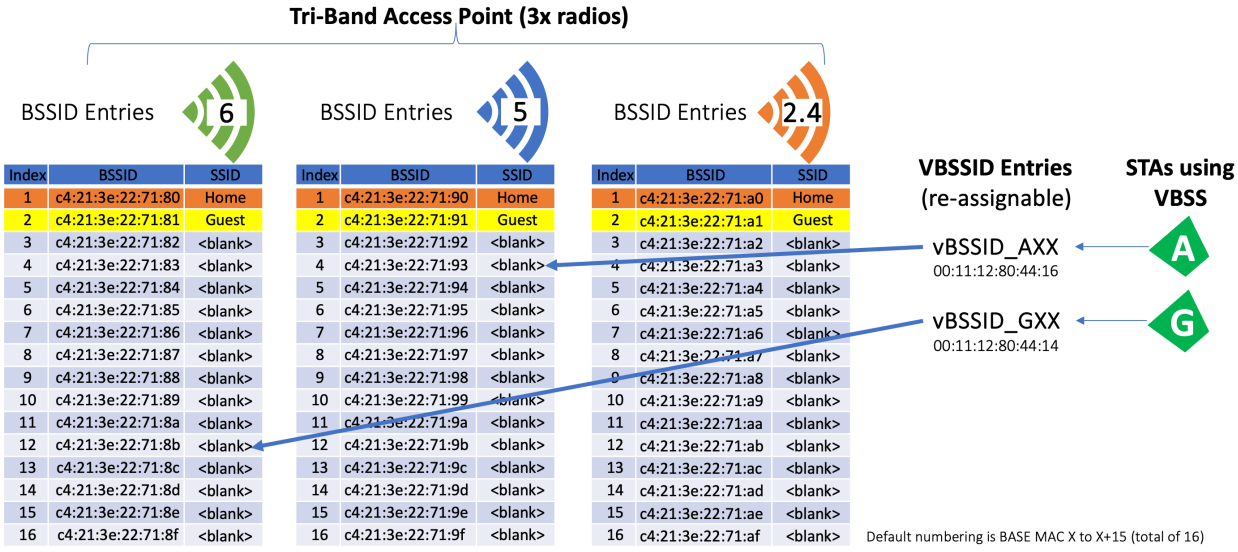
4.1. Previous Efforts for VBSS

The idea of a virtual BSS that follows a Wi-Fi client around actually has previously been explored. In 2016, the Wi-5 Project [3] worked on creating an open-source implementation for enterprise WLANs (Wireless LAN), but sources close to the project say that it suffered from extremely low throughput. In addition, there was a paper that focused on the personal virtual AP (PVAP) idea, [4], [5], [6] which presented the idea of moving a private BSS around in an enterprise WLAN (Wireless LAN) setup; however, CableLabs could not find any real-world implementations of it.

Both previous efforts focused on enterprise WLANs because (1) WLAN Controllers are present in enterprise networks to manage the movement of the BSSs (Basic Service Set), and (2) historically, enterprise networks were the only place where multiple APs on the same SSID were typically installed. However, in an enterprise environment, scalability becomes a problem because of the limited number of BSSs that current chipsets can support (e.g., currently 16 BSSs per radio) versus the number of active client devices moving around. Residential home networks are a more attractive target for virtual BSS implementations, given the lower number of BSSs required because of fewer mobile devices (out of all connected devices) in the home.

4.2. BSSID Capacity In APs For VBSS

For this paper, "Virtualising the BSS," the BSSID (and its related Wi-Fi MAC context) becomes "movable," and can be copied/transferred to a completely different radio on a different AP. This requires that the AP extract the entire operating context associated with the BSSID from the Origin AP Wi-Fi chipset/driver and the Controller communicate that information for installation in the Target AP when steering client devices. Obviously, there are limits to the number of these BSSIDs equipped per Wi-Fi radio. In most cases, a single BSS/BSSID is expected to manage 100's of connected devices dealing with all the per-client security and per client MAC context for each of these, while in the case of VBSS, the BSS/BSSID only deals with a single connected client. Referring to the following diagram, it is important to point out that in AP technology, most can operate between 8 and 16 different BSSID per radio – so in the example of a Tri-Band Concurrent (TBC) AP with three different radios, it would be expected to support up to 48 (3 x 16) BSSIDs.



Index	BSSID	SSID
1	c4:21:3e:22:71:80	Home
2	c4:21:3e:22:71:81	Guest
3	c4:21:3e:22:71:82	<blank>
4	c4:21:3e:22:71:83	<blank>
5	c4:21:3e:22:71:84	<blank>
6	c4:21:3e:22:71:85	<blank>
7	c4:21:3e:22:71:86	<blank>
8	c4:21:3e:22:71:87	<blank>
9	c4:21:3e:22:71:88	<blank>
10	c4:21:3e:22:71:89	<blank>
11	c4:21:3e:22:71:8a	<blank>
12	c4:21:3e:22:71:8b	<blank>
13	c4:21:3e:22:71:8c	<blank>
14	c4:21:3e:22:71:8d	<blank>
15	c4:21:3e:22:71:8e	<blank>
16	c4:21:3e:22:71:8f	<blank>

Index	BSSID	SSID
1	c4:21:3e:22:71:90	Home
2	c4:21:3e:22:71:91	Guest
3	c4:21:3e:22:71:92	<blank>
4	c4:21:3e:22:71:93	<blank>
5	c4:21:3e:22:71:94	<blank>
6	c4:21:3e:22:71:95	<blank>
7	c4:21:3e:22:71:96	<blank>
8	c4:21:3e:22:71:97	<blank>
9	c4:21:3e:22:71:98	<blank>
10	c4:21:3e:22:71:99	<blank>
11	c4:21:3e:22:71:9a	<blank>
12	c4:21:3e:22:71:9b	<blank>
13	c4:21:3e:22:71:9c	<blank>
14	c4:21:3e:22:71:9d	<blank>
15	c4:21:3e:22:71:9e	<blank>
16	c4:21:3e:22:71:9f	<blank>

Index	BSSID	SSID
1	c4:21:3e:22:71:a0	Home
2	c4:21:3e:22:71:a1	Guest
3	c4:21:3e:22:71:a2	<blank>
4	c4:21:3e:22:71:a3	<blank>
5	c4:21:3e:22:71:a4	<blank>
6	c4:21:3e:22:71:a5	<blank>
7	c4:21:3e:22:71:a6	<blank>
8	c4:21:3e:22:71:a7	<blank>
9	c4:21:3e:22:71:a8	<blank>
10	c4:21:3e:22:71:a9	<blank>
11	c4:21:3e:22:71:aa	<blank>
12	c4:21:3e:22:71:ab	<blank>
13	c4:21:3e:22:71:ac	<blank>
14	c4:21:3e:22:71:ad	<blank>
15	c4:21:3e:22:71:ae	<blank>
16	c4:21:3e:22:71:af	<blank>

Figure 9, Typical BSSID Mapping Per Radio, With SSID And VBSSID Assignments

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Traditionally, before multi-APs were commonplace, the different radios in an AP would be configured with different SSID names for 2.4 and 5 GHz radios, for example “home” and “home5g”. While these different Wi-Fi networks are fully connected at an IP (Internet Protocol) level, a user must manually switch their device to pick either the “home” or “home5g” SSIDs, introducing unnecessary steps to jump between bands. With multi-AP networks, it is important that devices can automatically switch between different radio bands on the same AP as well as switch to other APs. As a result, the “home” SSID is often configured across each of the different radio bands in each AP, providing the same Wi-Fi network for connected devices (in other words the user does not have to manually switch).

In a typical Dual-Band Concurrent (DBC) AP operating with 2.4 and 5 GHz radios, with 16 BSSID per radio, the “home” SSID occupies 1x BSSID per radio. If there is a “guest” network, then that also occupies 1x BSSID per radio. This leaves 14 BSSID per radio. These 14 BSSID are available for the Virtual BSS (VBSS) operation that this paper outlines. Future Wi-Fi silicon that adopts VBSS and Mobile Wi-Fi Technology may be able to increase this arbitrary limit of 16 BSSID to a much higher number of BSSIDs, where such BSSIDs would only deal with a single client device, rather than the 100’s it is currently designed for.

4.3. Mobile Wi-Fi Technology

CableLabs Mobile Wi-Fi technology is a VBSS solution that has been developed over the past number of years and has reached a certain level of maturity. Mobile Wi-Fi was included in the fifth release of Wi-Fi EasyMesh (R5), and a reference implementation has been submitted to the prplMesh codebase [7]. However, it must be noted that the technology is not specific to Wi-Fi EasyMesh but rather can be used in any multi-AP management system. In other words, this technology is more of a steering technique that can be considered an alternative to the Deny-List or BTM steering approaches. The technology has been proven to work and is now ready for commercial adoption.

Like other VBSS solutions, Mobile Wi-Fi creates a special VBSS for each client and moves that VBSS across APs and radios as the client moves to ensure the best radio is serving the client. A central Controller manages the transitions between APs. The central Controller is likely to operate on one of the APs, normally the primary Gateway for the home, but can also be operated from a cloud platform (given the necessary remote configuration protocols are in place).

4.4. Mobile Wi-Fi Basic Operation

A key principle of Wi-Fi exploited by Mobile Wi-Fi is that a client's transmissions are not just directed at the AP its associated with; any AP also operating or able to monitor, on the same radio band and channel will receive these same transmissions, as shown in the following diagram. APs without the VBSS security context cannot decrypt any of those client data transmissions.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

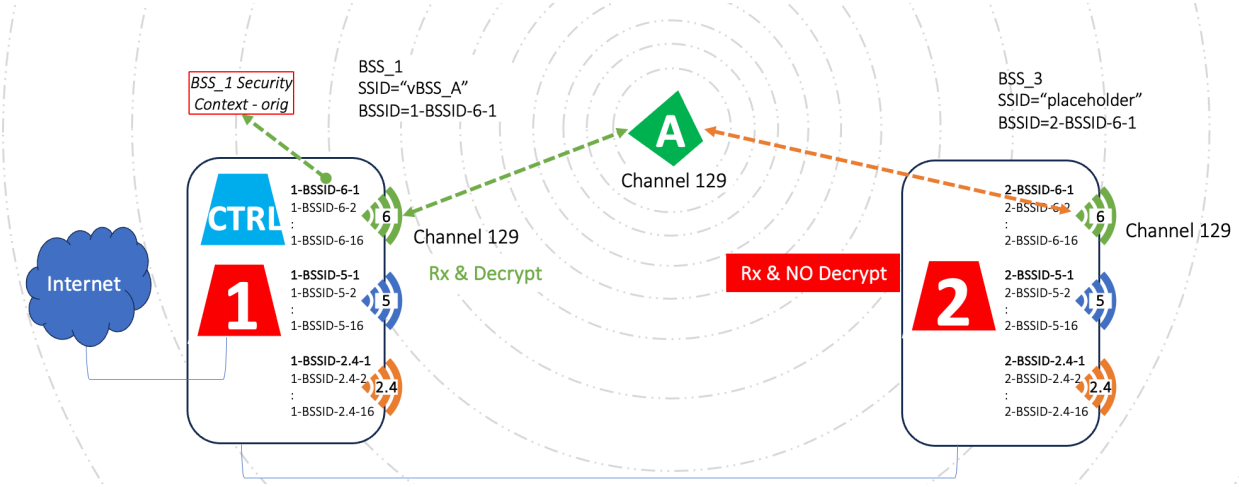


Figure 10, Client A Transmissions Received By Multiple APs, Only Decrypted By Associated BSS

This principle is taken advantage of when the VBSS is transitioned between AP 1 and AP 2, in that both APs will be able to receive the client's transmissions without having any interaction whatsoever with the client. In the case of a VBSS transition between two such APs, the client does not even know it is now talking to a completely different AP. The key here is to ensure the security context of the client BSS is copied to the target AP, enabling that AP to properly receive and decrypt the client data, as seen below. There is a brief period after the installation of the VBSS security context in the target AP (2) that traffic from the client (A) can be decrypted by both APs. This time must be minimised to prevent duplicate traffic entering the network from two different APs.

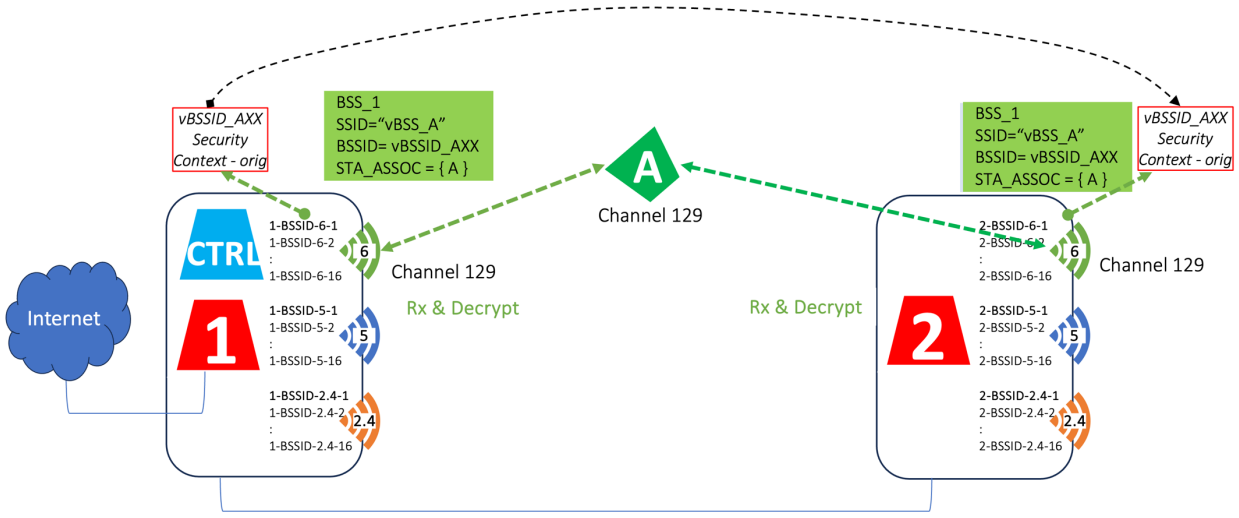


Figure 11, VBSS Transition Requires The BSS Security Context To Be Copied From The Origin AP To The Target AP

Once the VBSS has transitioned to the target AP and is confirmed to be operating, then the source AP (1) deletes the VBSS data, freeing up the AP resources for another VBSS client to use. The source AP can no longer decrypt data without the VBSS information so ceases forwarding data traffic received from the client.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

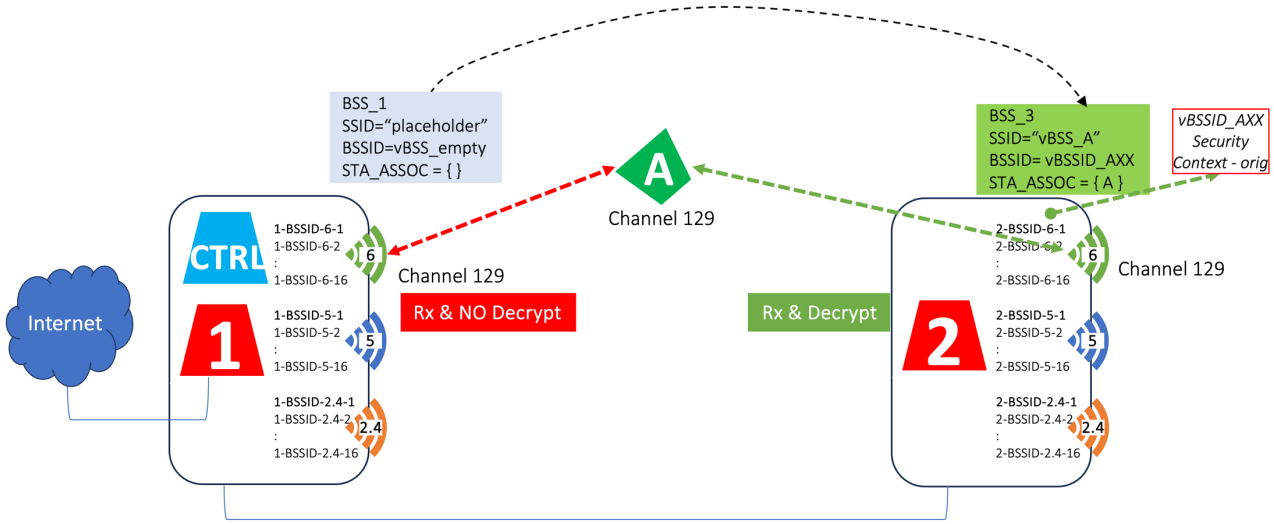


Figure 12, VBSS/Security Context Operational On The Target AP, Deleted from the Source AP

VBSS transitions related to moving a client to different channels in the same band do require some level of interaction between the current AP and the client, using a Channel Switch Announcement (CSA) message, where the Controller advises the Source AP to issue a CSA message to the client. The CSA message is normally used to advise a client that the AP is about to make a change of frequency – often in response to Dynamic Frequency Selection (DFS) when a radar event has triggered the evacuation of clients from a 5GHz channel. In the case of VBSS, because of having the same BSSID value being used on both Source and Target APs, it is possible to get the client to not only switch between different channels, but also to transparently switch between different channels on different APs. As there is a short delay in the handling of CSA messages, this has minor effect on the VBSS steering performance speed.

Unlike a standard Wi-Fi BSS that allows multiple clients to connect to an AP radio, a VBSS that moves between APs and radios based on client mobility demands can only serve a single connected client. To ensure this mode of operation a few steps are taken. Once a client is connected/associated with a VBSS, the solution stops responding to Wi-Fi probe requests from other clients. Additionally, to prevent other clients hearing about the VBSS, the solution adopts the use of unicast management and control frames (including beacons) that would typically be broadcast to all clients. By doing so, the VBSS can ensure only the intended (connected/associated) client hears the VBSS signalling.

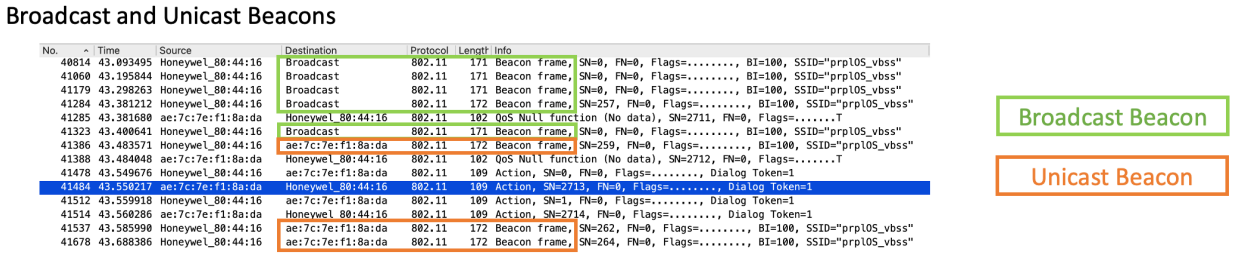


Figure 13, VBSS Transitioning From Broadcast Beacons to Unicast Beacons

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.5. VBSS Lifecycle

The life cycle of a VBSS is as follows:

1. A VBSS is created on an AP via a command from the Controller, with an assigned SSID name.
2. The VBSS initially operates using broadcast beacons and responds to all probe requests.
3. Any client can associate with the VBSS. Once associated to a client, the VBSS becomes dedicated to that client, meaning that beacons switch to being unicast and probe responses not from the connected client are suppressed.

4.6. VBSS Controller Management

As mentioned previously, to steer clients to the optimal AP/radio band, a Controller element must continually assess the quality of the Wi-Fi connectivity between the client and all other APs. This monitoring is no different to how other mesh networks work. Once the VBSS has a connected client, the Controller depends on continuous updates from all nearby AP, not just the connected AP, that are in range for the client. These updates include the RSSI value amongst other information about the client from the APs. The received data updates are evaluated by the Controller to determine the optimal AP, and if the client is not connected to it, the Controller initiates a transition not of the client, but of the VBSS/BSSID to the new/optimal AP. Care is taken to avoid repeatedly transitioning between APs with similar Wi-Fi performance, based on the use of hysteresis as well as evaluating other parameters such as latency, load, throughput, etc., of the candidate APs.

4.7. VBSS Transition Flow

Once a steering decision has been made by the Controller, then the VBSS transition is initiated via the following steps driven by the Controller:

1. Retrieves the security context from the source AP and sends it to the target AP,
2. Instructs the source AP to turn off Wi-Fi Block-ACKs on the connection,
3. Creates the VBSS on the target AP (with the client association already in place),
4. Instructs the source AP to send out a CSA pointing to the target channel/band (if the target channel/band is different than the source channel/band),
5. Deletes the VBSS on the source AP (without disconnecting the client),
6. Instructs the target AP to turn on Wi-Fi Block-ACK (if not already enabled) for the VBSS

At the end of these transition steps, the client is blissfully unaware that the VBSS is now being served from a different AP or radio, and any network activity that the client is engaged in is not interrupted. There is a small window of time between steps (3) and (5) where the VBSS is present on both APs, but in practice, that window of time is small enough to not cause any disruption. In experiments with a CableLabs VBSS implementation, this window of time was typically 5–10 msecs.

Occasionally, the source and target APs are on different channels or bands. In this case, a Channel Switch Announcement (CSA) is used to prompt the client to switch from the source channel or band to the target channel or band (this is step (4) above).

To complete the life cycle, when a client disassociates from a VBSS, the Controller instructs the AP to delete the VBSS and associated security context. Similarly, if the RSSI of the client on the connected AP goes below a useful minimum (indicating that the client cannot get usable throughput on the network), then the Controller instructs the AP to forcibly disconnect the client and delete the VBSS and associated security context. This covers the situation of devices leaving the home, where hard cutting the Wi-Fi

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

(again by deleting the VBSS and associated security context) means the device can cut over much quicker to the cellular network rather than persisting with an unusable Wi-Fi connecting that drives customer frustration. In normal mesh networks, where there are multiple client devices connected to the same network, having the ability to shut off the Wi-Fi network a poorly performing client is connected to is simply not possible, as ALL other clients would also be impacted, making the VBSS option more flexible.

Note that the above procedure does not need to associate a client using a fixed MAC address value; the sign-on process allows any client to sign on. In this way, Mobile Wi-Fi is robust in handling the challenge of random MAC implementations on clients. This is especially valuable because the clients using random MAC addresses (typically mobile phones, laptops, and tablets) are also those that will benefit the most from the mobility conferred by Mobile Wi-Fi.

Clients using Mobile Wi-Fi enjoy a seamless roaming experience—because the Controller controls the roaming, the user's experience is the same, regardless of which phone is being used, unlike previous efforts at Wi-Fi roaming. Furthermore, by keeping the mobile client on the best radio, the client rarely sees the need to go off channel and scan for a better AP, allowing the client to instead go into WMM (Wireless Multimedia) power save mode and save battery power. This requires no change to the client, as most clients only start scanning if their Wi-Fi experiences go bad in some way.

5. How Effective is Mobile Wi-Fi

5.1. Mobile Wi-Fi Testing Environment

The operation of Mobile Wi-Fi has been categorized by using the VBSS implementation contributed into the prpl Foundation prplMesh software component. This component implements the Wi-Fi Alliance Wi-Fi EasyMesh specifications, that includes the definition of VBSS operation. The software is integrated with prplOS and was run across several GL.inet B1300 dual band mesh routers [8] (used as a reference platform within the prpl Foundation development environment).

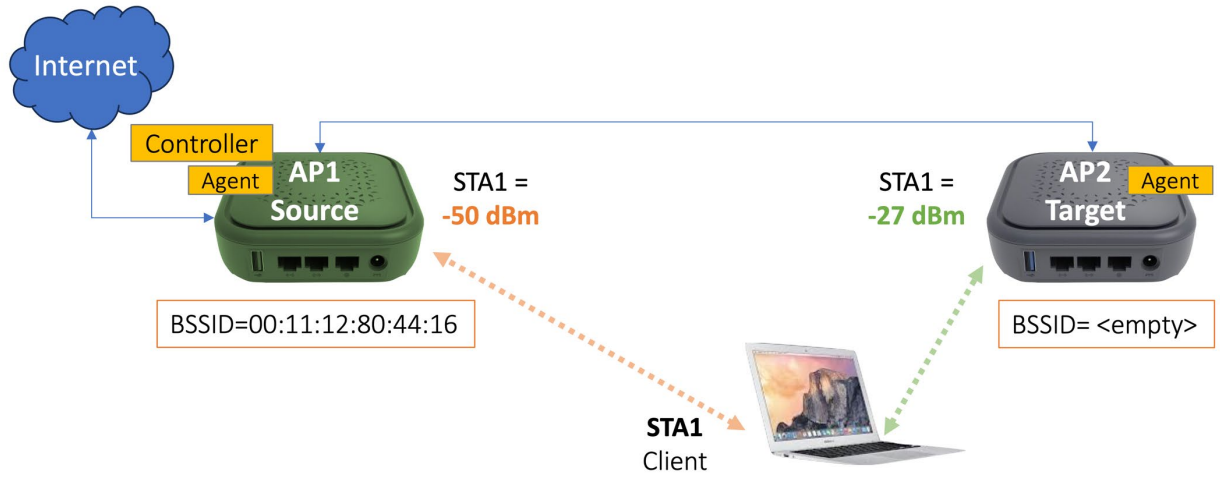


Figure 14, Main Test Network Description Using VBSS Running In prplMesh On GL.Inet B1300 Mesh Router

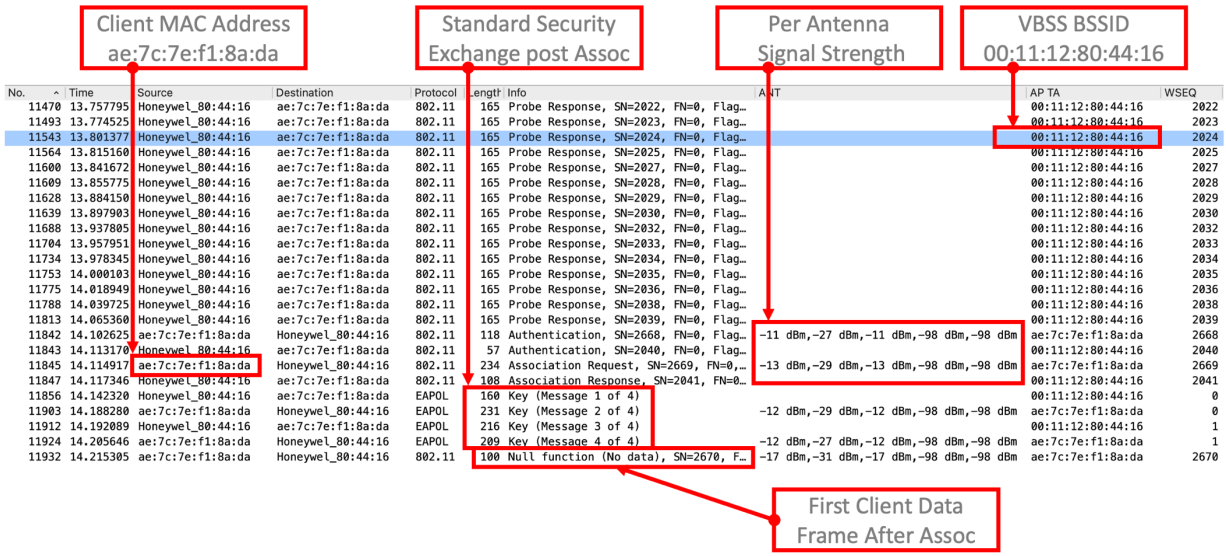
STA1, the client used in the testing is moved between the APs. A separate laptop in Wi-Fi monitor mode runs a continuous Wi-Fi capture that produces PCAP format files dissected over the following paragraphs. The two APs run prplOS with the prplMesh VBSS enabled software. AP1 includes the

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Controller functionality (again based on Wi-Fi EasyMesh), and both APs include the prplMesh agent implementing the low level VBSS functionality as a steering mode.

5.2. Normal Wi-Fi Association

The client associates with the Source AP following normal Wi-Fi Association rules, connecting to the BSSID of the VBSS. The client notices no difference compared to connecting to any other BSS.



No.	Time	Source	Destination	Protocol	Length	Info	ANT	AP TA	WSEQ
11470	13.757795	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2022, FN=0, Flag…		00:11:12:80:44:16	2022
11493	13.774525	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2023, FN=0, Flag…		00:11:12:80:44:16	2023
11543	13.801377	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2024, FN=0, Flag…		00:11:12:80:44:16	2024
11564	13.815160	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2025, FN=0, Flag…		00:11:12:80:44:16	2025
11600	13.841672	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2027, FN=0, Flag…		00:11:12:80:44:16	2027
11609	13.855775	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2028, FN=0, Flag…		00:11:12:80:44:16	2028
11628	13.884150	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2029, FN=0, Flag…		00:11:12:80:44:16	2029
11639	13.897903	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2030, FN=0, Flag…		00:11:12:80:44:16	2030
11688	13.937805	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2032, FN=0, Flag…		00:11:12:80:44:16	2032
11704	13.957951	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2033, FN=0, Flag…		00:11:12:80:44:16	2033
11734	13.978345	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2034, FN=0, Flag…		00:11:12:80:44:16	2034
11753	14.000103	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2035, FN=0, Flag…		00:11:12:80:44:16	2035
11775	14.018949	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2036, FN=0, Flag…		00:11:12:80:44:16	2036
11788	14.039725	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2038, FN=0, Flag…		00:11:12:80:44:16	2038
11813	14.065360	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	165	Probe Response, SN=2039, FN=0, Flag…		00:11:12:80:44:16	2039
11842	14.102625	ae:7c:7e:f1:8a:da	Honeywel_80:44:16	802.11	118	Authentication, SN=2668, FN=0, Flag…	-11 dBm,-27 dBm,-11 dBm,-98 dBm,-98 dBm	ae:7c:7e:f1:8a:da	2668
11843	14.113170	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	57	Authentication, SN=2040, FN=0, Flag…		00:11:12:80:44:16	2040
11845	14.114917	ae:7c:7e:f1:8a:da	Honeywel_80:44:16	802.11	234	Association Request, SN=2669, FN=0,…	-13 dBm,-29 dBm,-13 dBm,-98 dBm,-98 dBm	ae:7c:7e:f1:8a:da	2669
11847	14.117346	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	802.11	108	Association Response, SN=2041, FN=0…		00:11:12:80:44:16	2041
11856	14.142320	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	EAPOL	160	Key (Message 1 of 4)		00:11:12:80:44:16	0
11903	14.188280	ae:7c:7e:f1:8a:da	Honeywel_80:44:16	EAPOL	231	Key (Message 2 of 4)	-12 dBm,-29 dBm,-12 dBm,-98 dBm,-98 dBm	ae:7c:7e:f1:8a:da	0
11912	14.192089	Honeywel_80:44:16	ae:7c:7e:f1:8a:da	EAPOL	216	Key (Message 3 of 4)		00:11:12:80:44:16	1
11924	14.205646	ae:7c:7e:f1:8a:da	Honeywel_80:44:16	EAPOL	209	Key (Message 4 of 4)	-12 dBm,-27 dBm,-12 dBm,-98 dBm,-98 dBm	ae:7c:7e:f1:8a:da	1
11932	14.215305	ae:7c:7e:f1:8a:da	Honeywel_80:44:16	802.11	100	Null function (No data), SN=2670, F…	-17 dBm,-31 dBm,-17 dBm,-98 dBm,-98 dBm	ae:7c:7e:f1:8a:da	2670

Figure 15, Standard Client Association To VBSS

The client once connected to the VBSS/BSSID will start to receive unicast rather than broadcast Beacon frames. This is done to A) prevent other clients attempting to associate with the VBSS, and B) ensure that the client continues to receive beacons for the VBSS it is associated with. The VBSS also refuses to respond to any Probe requests issued by other clients that may have received the initial broadcast beacons, making sure only one client remains associated with the VBSS.

5.3. VBSS Testing Approach

The main approach to testing involved moving a client between the different APs and having the Controller determine based on received RSSI (and other data) reports when to move the client from one AP to another AP. Thresholds can be configured for RSSI that are used to determine at what RSSI level an AP becomes a candidate for steering to, and at what RSSI level a client must experience before steering to the suitable AP. Given RSSI reports can have some variability, the use of ranges of RSSI is ideal to prevent constant steering between devices. Other controls can also be factored in to reduce the possibility of repeated/unnecessary steering.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

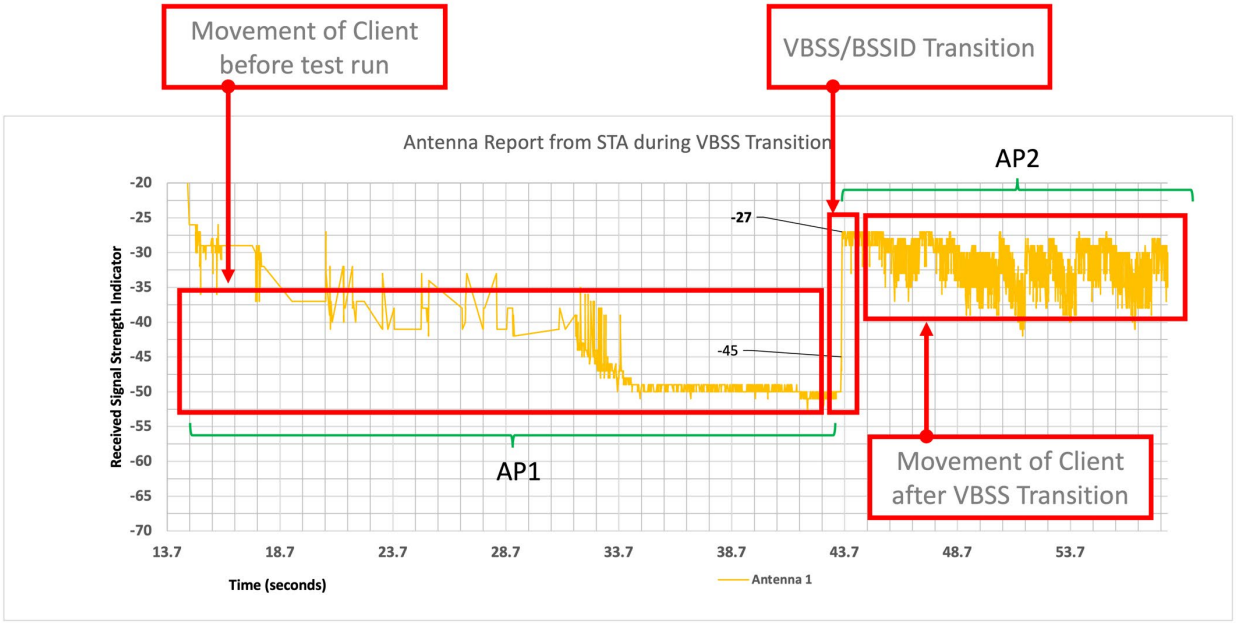


Figure 16, View Of RSSI During Test Run, Including RSSI At Time of AP Transition

5.4. Wi-Fi Alliance EasyMesh VBSS Steering Extensions

The testing relies on moving the client VBSS between the two APs, from AP1 to AP2, under the direction of the Controller, following the steering procedure defined in the EasyMesh Specification, outlined as follows. Each of these arrows represent protocol exchanges between the prplMesh Controller and the prplMesh agents that are used to coordinate the VBSS move between Source and Target APs. None of these messages interact with the client.

Please note the DELBA (Delete Block-ACK) optional step where the prplMesh agent requests the client to disable Block ACK (an optimization in Wi-Fi where multiple aggregated packets received by a client can be acknowledged back to the sender in one message rather than once for each packet). This reduces any packet acknowledgment issues during the brief interval when the VBSS can be active on both Source and Target APs.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

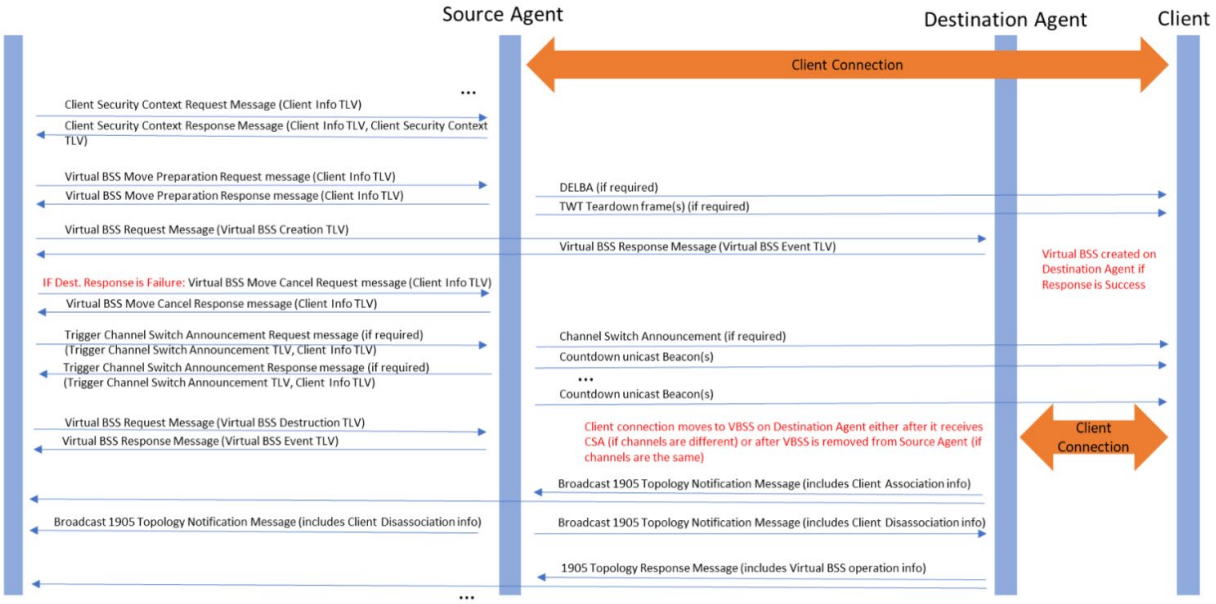


Figure 17, Wi-Fi EasyMesh VBSS Exchange To Support Steering Of VBSS between APs

5.5. Detailed VBSS Transition Analysis

The following capture taken during the VBSS transition between AP1 and AP2 shows no protocol interactions with the client, either before, during or after the transition. The BSSID that the client is connected to remains constant during the exchange. The only considerable evidence of the actual transition of the client to a new AP is the change in the reported antenna signal strength visible is the meta data collected by the wireless capture in the client. This should come as no surprise as the Controller is using RSSI to determine whether to steer a VBSS to a new AP.

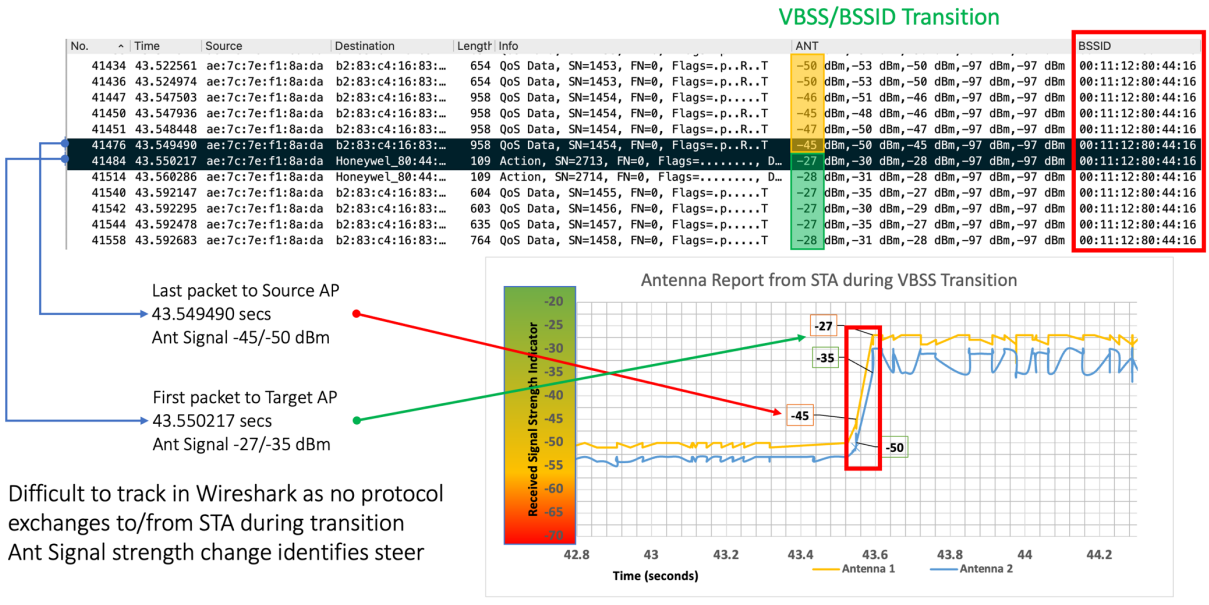


Figure 18, Transparent Move Of Client from AP1 to AP2, Including Antenna Signal Strength Detail And BSSID

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Like all other meshing systems managed by a Controller, the solution continues to monitor and evaluate the connected clients and their performance with a view to ensuring optimum operation within the network.

5.6. ICMP Ping Testing Approach

In addition to detailed PCAP file analysis, a more user-friendly approach for testing the VBSS steering performance was developed by utilizing the ICMP ping utility. A device connected to the network is configured to ping the client device under test for the test. The ping rate is set at a 10msec interval, and the client continues to be pinged during the VBSS steering.

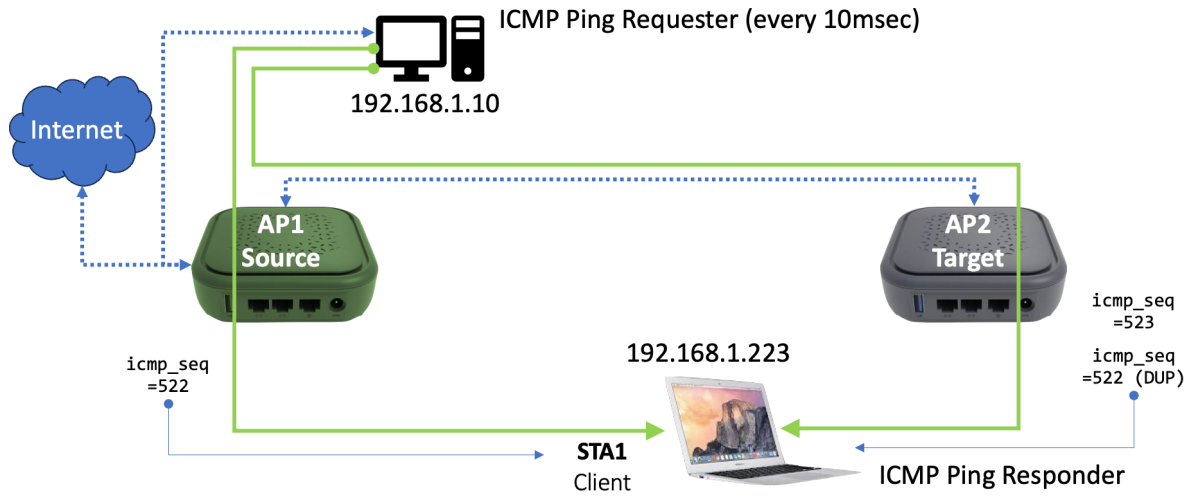


Figure 19, Simplified Test Environment To Deduce Performance Of VBSS Steering

The following is a snapshot of one test run during the VBSS transition.

Table 1 – ICMP Test Example, 192.168.1.233 Is Client Being Steered

```
64 bytes from 192.168.1.223: icmp_seq=522 ttl=64 time=6.949 ms

64 bytes from 192.168.1.223: icmp_seq=522 ttl=64 time=6.968 ms (DUP!)

64 bytes from 192.168.1.223: icmp_seq=523 ttl=64 time=6.224 ms
```

The first ICMP packet with a sequence #552 is received by the client from the source AP (AP1), while the second ICMP packet, shown as duplicate of sequence #552 is received by the client from the target AP (AP2). After the duplicate, all traffic is delivered by the Target AP to the client.

As the ping session was sending a packet every 10ms, and no packets were dropped, then it is possible to conclude a sub-10msec roaming time was achieved. Using this method makes testing VBSS a lot simpler and produces faster results for any device being submitted for testing.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. Mobile Wi-Fi Results

6.1. Mobile Wi-Fi Steering Performance

The testing details above show the performance of the solution with a specific client being steered between two APs. The time taken between the last packet on the Source AP to the first packet on the Target AP is calculated as just under 1ms when examining the PCAP captures in fine detail. This is achieved by not involving the client in any client<>AP protocol exchanges for steering or having to have the client reassociate and run through new security context for the new Target AP (like most other steering techniques expect).

In addition to the above tests, many more tests have been conducted with a mixture of different devices. The following is a list of these devices and their steering times based on the use of the less granular ICMP ping approach:

- iPhone XS: 100% success rate, 10 msec transition times typically
- Samsung S10: 100% success rate, 10 msec transition times typically
- MacBook Pro: 100% success rate, 10 msec transition times typically
- iPad Pro 2nd gen: 100% success rate, 10 msec transition times typically
- Pad Pro 4th gen: 100% success rate, 10 msec transition times typically

These results are extremely encouraging, given they cover a range of different device types (laptops, phones, and tablets – most likely to be roaming around a home) as well as each device type having a different version of operating system. The fact that the steering transitioned within 10ms points to minimal impact on the QoE of services operating on the device. Additionally, the 100% success rate points to the possibility of using VBSS steering for trouble free steering in residential networks, as well as increasing the confidence of being able to steer many different, if not all, such roaming devices, regardless of their Wi-Fi protocol or driver and without requiring the Controller to decipher each and every client to identify what will and will not work where steering is concerned.

6.2. What has been done and what is next

CableLabs built a complete proof of concept, which shows that the technology works, that has been integrated into the prplMesh (Wi-Fi Alliance EasyMesh compatible) software component that can run on prplOS (and potential to run on RDK-B). Mobile Wi-Fi shows how it removes the need for complex interactions between a mesh network/Controller and clients, which points to a future with far simpler Controller logic to steer heterogeneous Wi-Fi clients. This is available at the GitLab site listed in the references.

These are the various accomplishments achieved by the CableLabs team during the development of the Mobile Wi-Fi/VBSS software implementation. The results of the VBSS software are very impressive given that there was no work or expectation of the client devices, and they "just worked." The ability for the Controller to continuously make handover decisions, without losing the client devices is another testament to the robustness of the solution compared to other more fragile approaches for client steering.

- Seamless handoff of unmodified Apple and Samsung phones
  - iPhone 4 and beyond, Galaxy S6/S10, iPad Pro, MacBook Pro, and beyond
  - Both major silicon variants of each generation
- Implemented on Qualcomm 802.11ac Wave 2 AP silicon
  - Minor CableLabs modifications to firmware/driver

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Have proven seamless handoffs of WPA2 (Wi Fi Protected Access 2) secure connections
    - Without handset involvement
- Handoff decisions made every second
    - With no battery impact to phone

6.3. Challenges

The current support for Mobile Wi-Fi depends on the ability to extract certain security context information from the configured BSSID entries within a Wi-Fi Access Point, and deploy these in other APs. The current implementation work has been done with Qualcomm chipsets – where it was relatively easy to extract this information to enable the solution. To extend the solution to other Wi-Fi chipsets, more work is required with Wi-Fi SoC (System on Chip) companies to ensure the same types of interfaces are available and operate similarly where extracting security details are concerned. Work is happening in this space.

The current setup of most Wi-Fi chipsets is that they support up to 16 BSSIDs per radio. Since many devices in the home are NOT constantly moving around, this limitation of 16 is not a dealbreaker. These BSSIDs are created to scale to 100+ devices, each with their own independent set of security contexts. A VBSS BSSID is scaled to 1x device. There exists the potential in some Wi-Fi SoC to modify their internal architecture/organization of BSSIDs to accommodate more lightweight (e.g.,1x device) options, with a view to increasing to many more than 16x BSSIDs per radio.

One perceived challenge to the VBSS solution is the concern about an explosion in beacon related traffic if there are multiple BSSIDs in operation. The main drawback with beacon traffic is that it is normally broadcast at the lowest common PHY rate between connected devices, as well as requiring in most cases beacons being broadcast every 100ms. Additionally, clients tend to probe any broadcasted SSIDs they have a profile for, driving up more traffic. Under normal conditions, this would drive up the % of bandwidth capacity consumed to send beacons, however given that VBSS operation results in the use of unicast beacons to the single device connected to the BSS, then this broadcast overhead disappears, and without the beacon being sent to all devices, then only the connected device will probe the BSS.

6.4. Mobile Wi-Fi Opportunities Beyond Residential

As has been outlined, steering Wi-Fi devices around a Wi-Fi network is a major challenge, whether this is in a residential setting or beyond. A big challenge with pushing Wi-Fi into larger deployments has always been steering of clients, and the reliability and complexity of ensuring this happens flawlessly. Using Mobile Wi-Fi/VBSS enables many different applications previously thought too difficult to operate, including deployment in the MDU (multi dwelling units) or Enterprise space and consideration for Metropolitan Wi-Fi.

With devices constantly moving, be they within a building or outside in the open where they may travel at different speeds, the typical issue with Wi-Fi steering was the time to complete the steer. This paper points to some examples of 145/200ms for distinct types of BTM steering. Other steering options also exist. Such lengthy delays simply rule Wi-Fi out of the running for many different use-cases that are filled by cellular devices. Mobile Wi-Fi/VBSS, with its much shorter steering times, presents the opportunity to re-engage with Wi-Fi for several use-cases previously ruled out.

Beyond the residential setting described in this paper, Mobile Wi-Fi may also have utility in enterprise, MDU and Metropolitan Wi-Fi. Having a solution that is client device agnostic really helps when dealing

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

with such a large mix of device types encountered in these settings. There is the potential for Mobile Wi-Fi to provide truly seamless mobility in the case of Wi-Fi mobility.

In addition to enabling better roaming of clients in a residential setting, efforts are underway, driven by CableLabs to seek adoption of the Mobile Wi-Fi approach within the IEEE 802.11 standard. Several other companies are also interested in the concept of a VBSS within the next version of Wi-Fi [9] [10] [11].

The following is a timeline of the key milestones of the Mobile Wi-Fi project.

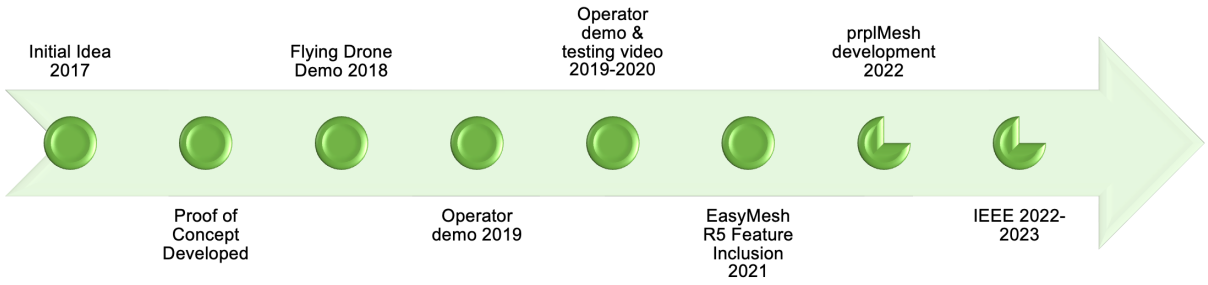


Figure 20, Mobile Wi-Fi Key Milestones

7. Conclusion

Mobile Wi-Fi/VBSS steering technology significantly enhances the reliability of Wi-Fi steering in mesh networks, while also delivering incredibly fast steering of clients between different APs ensuring maximum performance within the limit of available APs to the client. The solution is extremely robust, requiring no specific protocols, or software running on the client, making it transferrable to any Wi-Fi client. Mobile Wi-Fi has a positive impact on QoE, ensuring that the most mobile devices within a home network remain connected and deliver their services with minimal degradation The ability to deal with handover at the edge of Wi-Fi performance is also a key feature of Mobile Wi-Fi that enhances overall QoE.

Mobile Wi-Fi is in its infancy, and while it has achieved a lot in terms of a portable reference software implementation, a standalone reference demo (using a B1300 device) and impressive test results, more work is required to get Mobile Wi-Fi in front of operators, equipment manufacturers, and standards bodies for it to really get the momentum needed for it to become the de-facto approach for Wi-Fi device steering in all sorts of different applications.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AP	Access Point
BSS	Basic Service Set
BSSID	Basic Service Set Identifier
BTM	BSS Transition Management
CSA	Channel Switch Announcement
CTRL	Controller
DBC	Dual Band Concurrent
FT	Fast Transition
GHz	Gigahertz
GTK	Group Temporal Key
ICMP	Internet Control Message Protocol
IEEE	Institute Of Electrical and Electronic Engineers
IoT	Internet of Things
IT	Information Technology
LAN	Local Area Network
MCS	Modulation and Coding Scheme
MAC	Media Access Control
MDU	Multi Dwelling Unit
MHz	Megahertz
MIMO	Multi-Input, Multi-Output
OS	Operating System
PCAP	Packet Capture
PHY	Physical layer
PTK	Pairwise Transient Key
QBC	Quad Band Concurrent
QoE	Quality of Experience
QoS	Quality of Service
RCPI	Received Channel Power Indicator
RF	Radio Frequency
RSSI	Received Signal Strength Indicator
SCTE	Society of Cable Telecommunications Engineers
SNR	Signal To Noise Ratio
SS	Spatial Streams
SSID	Service Set Identifier
STA	Wi-Fi station
TBC	Tri Band Concurrent
VBSS	Virtual Basic Service Set
VR	Virtual Reality

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

WFA	Wi-Fi Alliance
WLAN	Wi-Fi LAN
WPA	Wi-Fi Protected Access
SCTE	Society of Cable Telecommunications Engineers

Bibliography & References

[1] Wi-Fi Alliance EasyMesh Release 5 Specification - https://www.wi-fi.org/system/files/Wi-Fi_EasyMesh_Specification_v5.0.pdf
[2] Download location for future Wi-Fi Alliance EasyMesh Specifications - https://www.wi-fi.org/file/wi-fi-easymesh-specification
[3] Wi-5 Project - https://github.com/Wi5/odin-wi5/wiki
[4] Programming the Enterprise WLAN: An SDN Approach - https://lalithsuresh.files.wordpress.com/2011/04/lalith-thesis.pdf
[5] Programmatic Orchestration of Wi-Fi Networks Video Overview - https://youtu.be/m8CEGuQSKQk
[6] Programmatic Orchestration of Wi-Fi Networks paper - https://www.usenix.org/system/files/conference/atc14/atc14-paper-schulz_zander.pdf
[7] Current release of VBSS incorporated into prplMesh - https://gitlab.com/prpl-foundation/prplmesh/prplMesh/-/tree/dev/vbss-integration-demo-staging?ref_type=heads
[8] GL.inet B1300 target hardware used for VBSS implementation - https://www.gl-inet.com/products/gl-b1300/
[9] Nokia Nov-2020 IEEE contribution on VBSS (https://mentor.ieee.org/802.11/dcn/20/11-20-1247-01-00be-virtual-bss-for-multi-ap-coordination.pptx)
[10] Samsung July-2019 IEEE contribution on VBSS (https://mentor.ieee.org/802.11/dcn/19/11-19-1019-00-00be-virtual-bss-for-multi-ap-coordination.pptx)
[11] Huawei Nov-2019 IEEE contribution on VBSS (https://mentor.ieee.org/802.11/dcn/19/11-19-1972-00-00be-operation-of-virtual-bss-architecture-for-multi-ap-coordination.pptx)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

SCTE CABLE-TEC EXPO 23
CELEBRATING 40 YEARS

2023 Fall Technical Forum
SCTE • CABLELABS • NCTA

Watts In An ID?

Utility-Tied, Energy Consuming Asset Life Cycle Management

A Technical Paper prepared for SCTE by

Richard Grivalsky
Principal Engineer
Comcast Cable
Vermont, USA
+1 (802) 316-6553
Richard_Grivalsky@cable.comcast.com

Anusha Nukala
Technical Project Manager
Comcast Cable
Pennsylvania, USA
+1 (610) 209-9514
Anusha_Nukala@comcast.com

Alexander Falcon, Comcast Cable

Shaju Kezakykanirakat, Comcast Cable

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Manual Data Alignment ... 3
3. Tools Ecosystem Integration ... 4
   - 3.1. EMD: Energy Management Database ... 4
   - 3.2. Machine Learning Integration ... 4
   - 3.3. Real Estate Database ... 6
   - 3.4. Power Supply Operational Database ... 6
   - 3.5. Design Database ... 6
   - 3.6. Visibility ... 6
4. Global ID ... 7
   - 4.1. G-ID Data Flow ... 7
   - 4.2. EMD and PSNB ... 8
   - 4.3. Global ID – Built for the Future ... 9
5. Conclusion ... 10

Abbreviations ... 10

Bibliography & References ... 10

List of Figures

Title **Page Number**

Figure 1 - SCTE Power Pyramid ... 3
Figure 2 - Example Map View: No Power Supply at Pole ... 5
Figure 3 - Example Map View: Metered, Aerial Power Supply at Pole ... 5
Figure 4 - Global ID Data Flow ... 8
Figure 5 - Global ID Structure ... 9

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Cable operators' energy-consuming asset portfolios are extensive. Tracking the energy consumption of each of these asset's lifecycle requires utility account alignment, which can present a unique challenge. Operators outside plant represents ~50% of total consumed power, according to the SCTE power pyramid, and these assets are distributed among thousands of utility grid connection points. Coupled with facility account management, maintaining alignment between asset, utility account, and meter numbers can become quite complex. To drive data accuracy and completeness for our energy consuming network devices, we are creating and deploying persistent, unique identifiers that will help us manage asset lifecycle and track data in our expanded in-house energy management tool. Using a combination of machine learning and heuristic matching, we can maintain asset alignment and issue new identifiers as well as retire identifiers when an asset is removed from the network—

To meet this challenge, we are integrating several tools to manage our diverse population of energy-consuming, utility-tied assets.

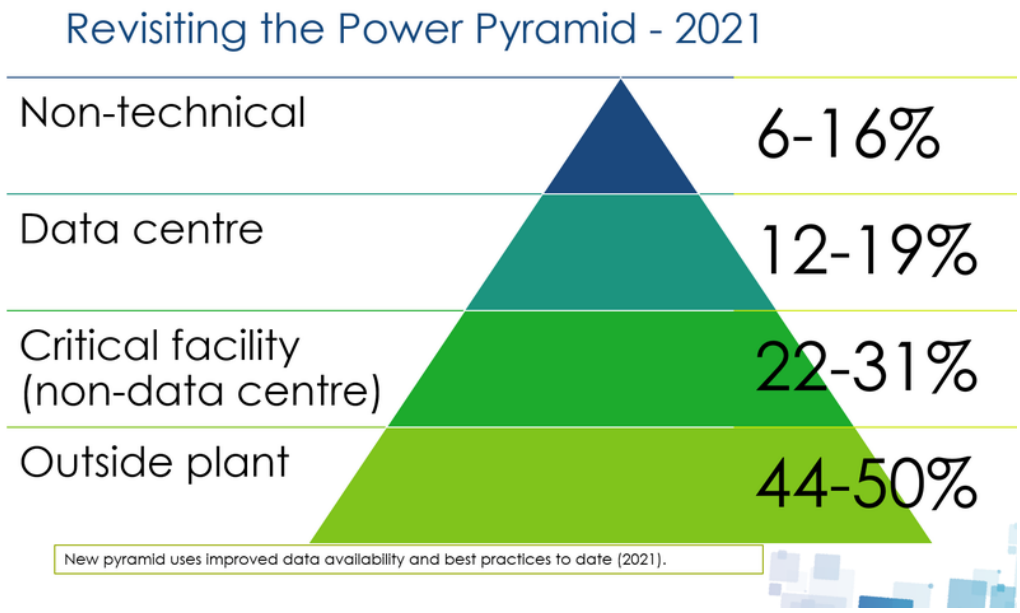


Figure 1 - SCTE Power Pyramid

2. Manual Data Alignment

Recognizing the need to align asset utility data with internal geospatial data we initiated an audit program. Early on, one problem was easily identified - three different data sources all referenced the same asset with disparate geospatial data. Utility billed asset location was sometimes nonexistent, latitude and longitude varied between internal data sources, and addresses were often not a textual match. Beyond the simple complexities of "Rd" or "Road", years of manual data entry meant misspelled street names, towns or street numbers were omitted from one of the data sets.

Patterns emerged when iterating through the manual audit data. Some utilities relayed pole numbers in billing data; other utilities were able to provide geospatial information with a join of their front-end billing and back-end geographic information systems (GIS).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

With each new utility dataset ingested, nuances were uncovered that further informed and enabled process improvement.

3. Tools Ecosystem Integration

Auditing each location to align utility account data proved to be operationally sound and effective but lacked one key item—permanence. Auditing thousands of locations' metadata is time-consuming; unless operators want to commit to auditing at some regular cadence, integration between tools is required and a persistent, unique identifier which all tools must reference is a fundamental necessity.

Internal conversations between tool owners and user communities began to shed light on opportunities for change. Application programming interfaces (API) were developed to allow tools to communicate key elements and join them with a persistent identifier.

3.1. EMD: Energy Management Database

Energy Management Database (EMD) is an internally developed tool that captures energy-related invoice data. Specific to operations and the overall network, electric utility invoice data is ingested, then a comprehensive overview of electricity consumption and spending across Comcast is provided. While the tool has many use cases and reports, this paper focuses on how the first step in aligning energy-consuming assets is foundational to creating a unique and persistent identifier for all utility-tied asset lifecycle management.

EMD ingests invoice data via utility electronic data interchange (EDI). As some utilities do not provide an EDI feed, a scraping bot service was developed to process paper invoices. The invoice details are ingested each billing cycle and compared to existing records for each location. This comparison allows for updates to data, such as smart meter upgrades; EMD is then able to write this data to appropriate tools via APIs.

3.2. Machine Learning Integration

Identification and confirmation of active or inactive power supply locations can be difficult; without dispatching network technicians to many locations, a solution was needed to confirm power supply asset location and utility connection type. We developed a tool we call LOCUS (Locating Outside Plant Comcast-to-Utility Services), which uses machine learning to identify and classify power supply locations. By identifying whether a location is metered or unmetered, EMD can ingest the data and confirm or dispute the assigned utility account or meter number.

Utilizing our multi-class model, LOCUS combs mapping software via open APIs to identify power supply cabinets and the associated utility connection methodology. Our training model consists of thousands of confirmed power supply locations with their associated latitude and longitude being referenced at a rate of approximately 10 per second. Additionally, utilizing geospatial radii, LOCUS "walks" the map until exhausting the maximum radius, returning null findings for any ingested unknown datasets. The model classes include aerial-metered, aerial-unmetered, underground-metered, and underground-unmetered.

By identifying both the installation methodology and the utility connection type, accounts can be more easily matched to the appropriate asset. Additional benefits arise when large areas undergo forced relocates, as occasionally the previous account is not disconnected; in these situations, the matching of account to asset leads to potential utility billing records clean-up.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 2 and Figure 3 show examples of LOCUS navigating to supplied geospatial location data and analyzing the connected asset. In Figure 2, there would have been billed usage, but low probability of it being an actual power supply. In Figure 3, the billing location data was directed to a PS location and was verified as a metered, aerial power supply.



Figure 2 - Example Map View: No Power Supply at Pole



Figure 3 - Example Map View: Metered, Aerial Power Supply at Pole

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

This system is also used to validate the removal of power supply locations for billing closure confirmations. If an account cannot be tied to an active asset, LOCUS will sleuth the surrounding area, based on predetermined radii, for any identifiable power supply locations, then is passed on to other teams for evaluation before investing person-hours to review locations manually.

3.3. Real Estate Database

Leveraging an existing real estate database, used to track owned and leased "brick and mortar" facilities, EMD ingests appropriate metadata for owned and leased sites. Within this dataset, the facility's primary use type, e.g., administrative, headend, datacenter, is conveyed and used to generate the asset type portion of the identifier. This site classification also aids in mapping each facility for the appropriate major bucket: technical, non-technical, and data center.

3.4. Power Supply Operational Database

Power Supply Notebook (PSNB) manages the inventory and tracks the status of Comcast's outside plant power supplies. More than 200,000 power supplies are managed using this software. Power supplies provide telemetry each minute, reporting the current status of the device and select parameters. When parameters reported through telemetry are out of tolerance and require maintenance, PSNB can initiate the process. Ensuring that all devices powering the Comcast network are captured and tracked through PSNB is imperative in understanding any potential impacts on customer services. Loss of commercial power or other network impairments are most effectively addressed when the plant topology is accurate.

Integrating PSNB with EMD will trigger an alert to validate whether the utility account is still required if a power supply's output power is zero for a set timeframe, ensuring utility accounts stay aligned with network changes.

3.5. Design Database

Geospatial tools provide designed location metadata and system-unique identifiers for all objects within the network. For our purposes, the locations and devices metadata for all power supply objects, along with all critical facilities and data centers, was required (non-technical facilities not represented). An API was developed, allowing EMD to ingest all relevant fields and house that data in the primary relationship table within the database.

3.6. Visibility

Energy Management Database, and the associated tools integrations, allowed insight into the changing of internal metadata as well as utility metadata rate of change (utility mergers/acquisitions, smart meter upgrades, etc.). This new level of visibility further demonstrated the need for a persistent identifier, leading to the development of Global ID (G-ID).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. Global ID

4.1. G-ID Data Flow

Global ID (G-ID) is a persistent identifier that will be used across all tools and databases for each individual physical asset. The G-ID will maintain inter- and intra-system relationships and store all individual tools or account identifiers as associated by G-ID in a primary table configuration to allow change management through the asset lifecycle.

The stages of the asset life cycle are design phase (GIS), construction, activation (PSNB), and invoice management (EMD). At the design phase, when an object is published for an asset (power supply or technical facility), the design database object ID is available via API. EMD then ingests and begins a heuristic matching process to all known assets in PSNB. Power supplies can share a location, be stand-alone, or exist several feet to several pole-spans apart, heuristic matching is needed to ensure proper G-ID assignment.

This same matching is used to align billing data. If a utility bill is ingested by EMD and is not already assigned to an asset, heuristic matching is used to locate the appropriate design and operational datasets. If none are found, an alert is issued to the appropriate team(s) for intervention and record creation.

As we have various system-level IDs, tracking the changes to any single ID back to tool or phase—is difficult—as these IDs can change due to factors such as location change, design schema change, decommission, or physical replacement of an asset. G-ID maintains the persistence of the unique identifier across all tools and phases of the asset lifecycle, which will allow changes to be tracked and documented.

This change control allows for more accurate inventory alignment and improves data quality for reporting and analysis.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

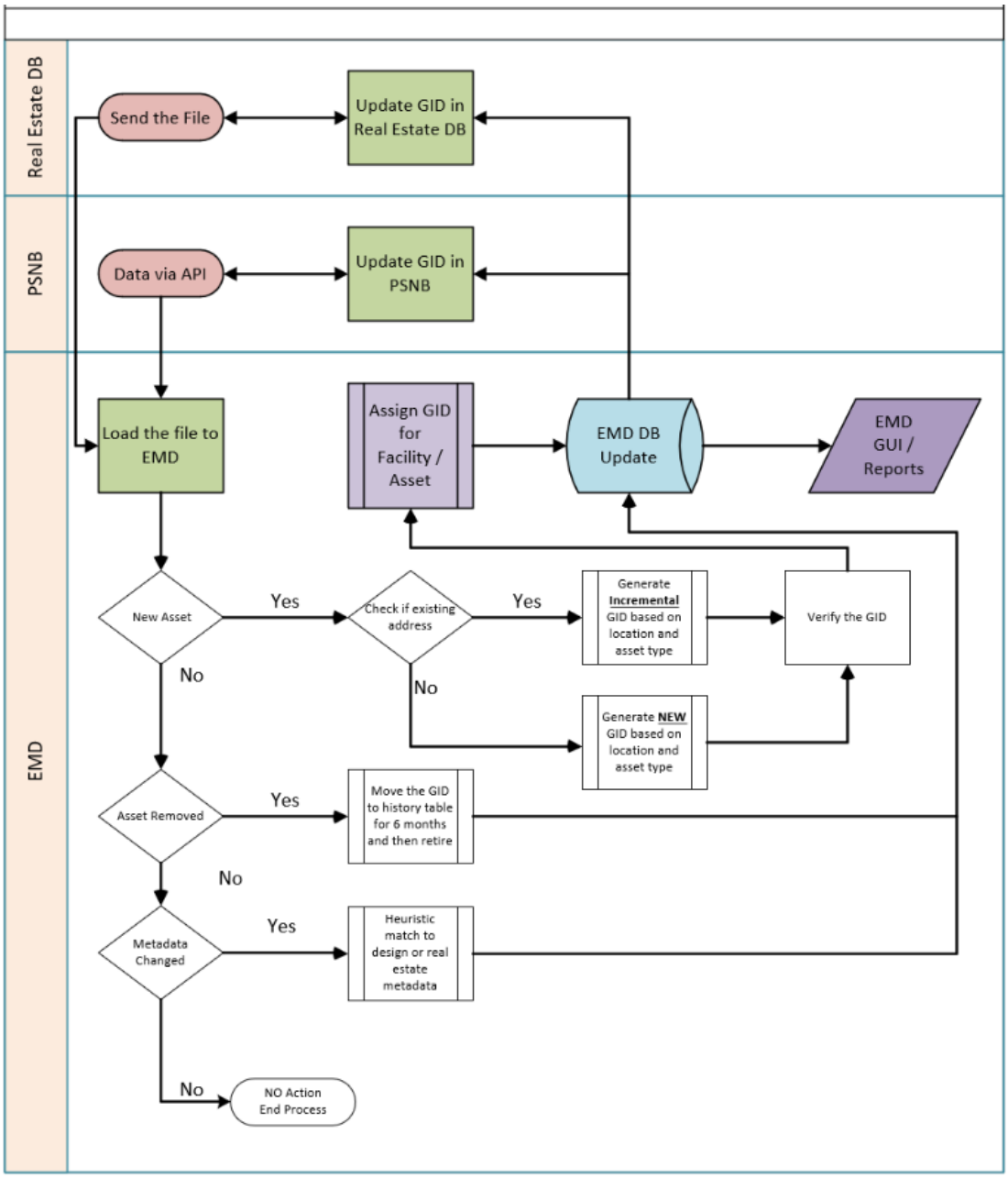


Figure 4 - Global ID Data Flow

4.2. EMD and PSNB

As power supplies are added to the network, moved to a new location, or removed, the associated record in PSNB is updated to ensure that the inventory remains accurate. Assigning a G-ID to each power supply record in PSNB ensures that the relationship between the record in the design database and the management application remains intact. When a new power supply is added to PSNB, EMD provides a new, unique G-ID and writes it to PSNB. The G-ID provides a durable relationship between the record in PSNB, the GIS system, and utility provider account information. When power supplies are moved, but

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

retain the same utility information, the PSNB identifier will remain unchanged, the G-ID will be updated according to the heuristic logic and, if within the specified radius, remain unchanged. If the move is outside the specified radius, the old G-ID will be retired, and a new G-ID will be issued for the asset. If the associated record in the design database changes, the relationship is managed within EMD to ensure the correlation between the GIS system and physical location remains intact.

When new power supplies are installed, a record is created in PSNB. The record does not contain the G-ID at the time of creation, as it has yet to be assigned. A scheduled process (delta report) compares the inventory changes in PSNB with the records in EMD and identifies the moves, additions, or deletions. When a new record is identified, EMD will issue a G-ID that is then added to the record in PSNB. This reconciliation process is key in ensuring that all power supplies in PSNB are accounted for in EMD. An API is used for the processing and updating of records in both PSNB and EMD. The G-ID, along with utility provider information stored in EMD, are provided, and updated in PSNB using this process. By conveying the utility data, such as account and meter number or utility contact information, network teams have the data readily available—should a utility-connected location be damaged or need utility response.

4.3. Global ID – Built for the Future

As we ideated and iterated through the G-ID schema, we found unique use cases and requirements from many of the tool owners. From character limitations to future use cases, all aspects were considered in the final identifier.

Ultimately, a 19-character alpha-numeric string was finalized. Utilizing characters for country and state codes, asset type identification, system initiator, location, and finally, asset identification are all components of the ID.

As several power supplies can exist in one location, there is a relational hierarchy to maintain the geospatial relationship and a unique, persistent identifier. The same applies to power supplies installed at a Comcast facility if the address and utility connection are the same, the location portions of the G-ID will be replicated. Maintaining separate location and asset strings within the ID allowed for associated records to be identified while still permitting a unique identifier. For reporting purposes, this allows the identification of multiple assets drawing power from the same utility meter.

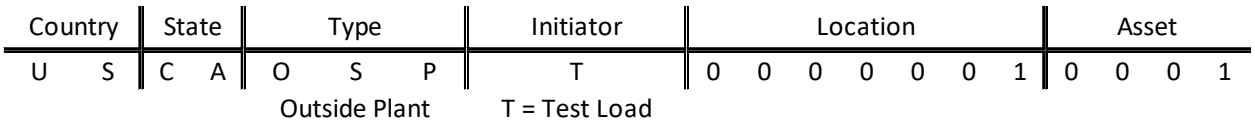


Figure 5 - Global ID Structure

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5. Conclusion

The need to align energy-consuming assets with their appropriate utility account(s) initiated an audit of Comcast's asset portfolio. Finding that data sources were not aligned, a need for a persistent identifier was recognized.

Energy Management Database was created to house and report on all assets and their respective energy consumption; as the data and the diverse identifiers from many systems were found to be dynamic, our internally generated Global ID was developed for all grid-tied, energy-consuming assets.

Global ID allows lifecycle management at a far more detailed level , provides the ability to track energy metrics, such as Electricity per Consumed Byte, at the facility level, and enables the creation and updating of dashboards for operational use to drive electricity consumption reduction.

Abbreviations

API	application programming interface
EDI	electronic data interchange
EMD	Energy Management Database
G-ID	Global ID
GIS	geographic information system
LOCUS	Locating Outside Plant Comcast-to-Utility Services
PSNB	Power Supply Notebook
SCTE	Society of Cable Telecommunications Engineers

Bibliography & References

Figure 1, SCTE Power Pyramid, https://www.scte.org/standards/development/energy-2020-powering-cables-success/
Figure 2, Figure 6 - Example Map View: No Power Supply at Pole, Richard Grivalsky
Figure 3, Figure 3 - Example Map View: Metered, Aerial Power Supply at Pole, Richard Grivalsky

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Weaving Real-Time Fiber Data Analysis Into a Reliable Network

A Technical Paper prepared for SCTE by

Matthew Wichman
Director
Comcast
+1 (425) 754-8388
matthew_wichman@Comcast.com

Venk Mutalik, Fellow, Comcast

Errol D'Souza, Vice President, Comcast

Josh Cook, Engineer 4, Comcast

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction ... 3
2. Remotely Detectable Networks ... 3
3. Fiber Monitoring and Telemetry ... 4
4. Weaving the Telemetry Together ... 4
   - 4.1. Inside Plant (ISP) Transmit Fiber Measurement ... 4
   - 4.2. ISP Receive Fiber Measurement ... 5
   - 4.3. ISP Common Fiber Measurement ... 5
   - 4.4. Bi-Directional Link Delta ... 6
   - 4.5. Multi-Metric Logic ... 6
   - 4.6. Next Steps ... 7
5. Operational Evolution ... 7
   - 5.1. Construction and Certification ... 7
   - 5.2. Proactive Monitoring ... 7
   - 5.3. Fiber Restorations ... 8
6. Conclusion ... 8

Abbreviations ... 9

List of Figures

Title **Page Number**

Figure 1 – Measuring TX Fiber Jumper Loss ... 5
Figure 2 - Measuring RX Fiber Jumper Loss ... 5
Figure 3 - ISP Common Fiber Link ... 6
Figure 4 - Bi-Directional Link Delta ... 6
Figure 5 - Multi-Metric Logic ... 7

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Fiber network problems have historically been difficult to diagnose, requiring dispatching multiple teams to locate and fix the issue. However, today's availability of real-time data from both sides of the link and common fibers has enabled faster and more accurate problem identification and resolution. Comcast, as the largest provider of multi-gig speeds to homes and businesses, has implemented a range of hardware and software telemetry tools to monitor its network.

Cable operators can leverage real-time data to improve the monitoring and reliability of fiber networks. By combining multiple metrics, the level of intelligence is amplified, enabling the early detection of problems that would have otherwise gone unnoticed. When the fiber path degrades, diagnosis is possible within minutes, and the impacted fiber of the fiber path is isolated for technician dispatch. Small Form-Factor Pluggable (SFP) metrics and the common fiber-monitoring platform are used to measure fiber loss in both directions, enabling the detection of failing fiber jumpers in hubs and single-wavelength fibers past the Optical Multiplexor (Mux) in the field. This approach has reduced repair time and can help avoid unplanned outages.

We will cover new insights into how real-time data analysis can enhance fiber network reliability and help providers avoid costly downtime for customers. The methodology and findings presented may be of interest to other providers looking to improve their network performance through data-driven insights.

2. Remotely Detectable Networks

Moving to an all-digital network has opened up a new world of devices sending us telemetry. To ensure effective remote detection capabilities, it is essential to incorporate them into network design from the onset. The ability to remotely detect complex issues on a network necessitates that operators have control over the variables within the network designs. By carefully considering and limiting the number of variables during the design stage, operators can create networks that are easier to troubleshoot, with predictable failure points.

The establishment of controllable variables requires alignment among design and construction teams through operational standards. It is important to evaluate each component of the network with the objective of standardizing unmeasurable components and measuring components that cannot be standardized. This means taking in to account the loss passive devices introduce and the way the impact the light need to be predictable. By achieving uniformity and consistency across the entire network, remote detection and prompt dispatch to address issues become feasible.

Optical components need to have common parameters and be compatible with accepted design standards. One-off designs that deviate from the established design principles are hard to quantify and remotely monitor.

Devices that should be standardized:

- **DWDM Mux** – Standard throughput loss, channel sets, and monitoring ports.
- **SFP/Optics** – Standardized optics with a narrow, accepted transmit level and minimum receive level.
- **Monitoring** – All Muxes and wavelengths need to be accounted for and maintained.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3. Fiber Monitoring and Telemetry

The fiber-monitoring system interfaces with a new evolution of C-Band Dense Wave Division Multiplexers (DWDM) that are equipped with a converged test point combining the forward and return C-Band spectrum into a single 20dB down test port as well as adding a low insertion-loss band-pass for a 1611nm Optical Time-Domain Reflectometer (OTDR). This converged test port configuration allows the fiber monitor to measure C-Band wavelength power levels bi-directionally via an Optical Spectrum Analyzer (OSA). Approximately every three minutes the fiber monitor samples both the C-Band spectrum via the OSA and OTDR of each fiber to monitor for new impairments or sudden impacts such as a fiber cut.

The OTDR has a clean insertion point to measure the fiber with minimal loss to dynamic range allowing for a single fiber connection per monitored Mux. Utilizing a 1611nm OTDR allows the fiber monitor to not interfere with existing C-Band customer traffic so that it can co-exist and be pervasive.

Standardized wavelength assignments (forward and return) are then enriched with design details assigning devices to each wavelength pair. This opens the door for new logic paths that can compare expected losses with observed levels to determine optical path health both end-to-end at scale and through automation.

4. Weaving the Telemetry Together

Fiber metrics have power in numbers. Fiber links have many metrics to evaluate performance. Traditionally we have minimum and maximum values and if a metric falls between these numbers it is considered a PASS. When we look at the link performance end to end, the numbers add context and meaning to each other and start to identify the performance characteristics of smaller segments of the larger optical link. In turn, contextual measurements can then be used in a larger logic algorithm that shows the entire end-to-end link. This new knowledge of organized measurements gives the team a clear view of the network.

When measuring from pluggable optics, like the transmit or receive level of light, it is very important that each model be evaluated. Measurements queries can be impacted by firmware or outside environmental factors like heat. When an SFP starts reading erroneous levels or returning no information, these need to be addressed as a loss of telemetry.

4.1. Inside Plant (ISP) Transmit Fiber Measurement

To calculate the integrity of the transmit light to the field, the SFP transmit level minus the monitoring port level should equal the known Mux loss value. When this number is below an acceptable threshold, we have a bad link between the SFP and the Mux, the Mux and the monitoring port, or a bad port on the Mux.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

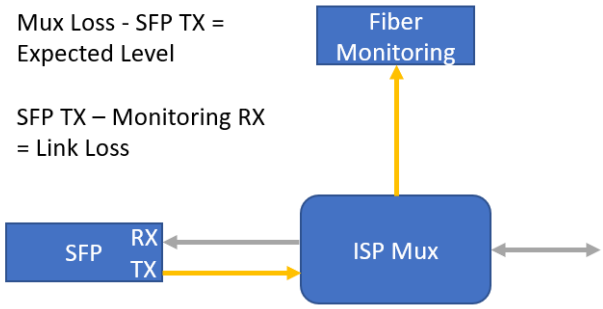


Figure 1 – Measuring TX Fiber Jumper Loss

4.2. ISP Receive Fiber Measurement

To calculate the integrity of the receive light from the field, the SFP receive level minus the monitoring receive should equal the difference of insertion loss of the two ports. If the loss is too small the fiber to the SFP is bad, if the loss is too great the fiber to the monitoring port is bad. In either situation a bad Mux port could also be the cause, but most times a dirty connection is most probable.

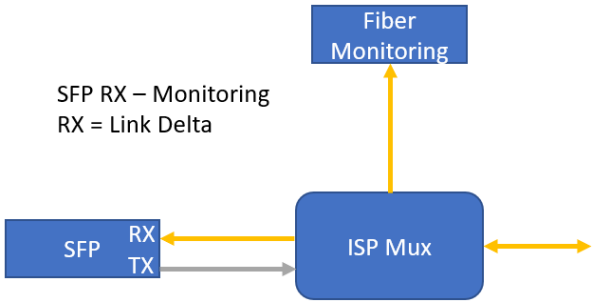


Figure 2 - Measuring RX Fiber Jumper Loss

4.3. ISP Common Fiber Measurement

When we look at the common fiber from the Mux to the field, dirty connections and bad jumpers can cause excessive loss and the OTDR of the fiber-monitoring system will fail to return a distance. This failure can be used to identify a bad common fiber. When we also look at the light levels received, it can be determined if a fiber is disconnected or if the OTDR failure is due to a poor connection.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

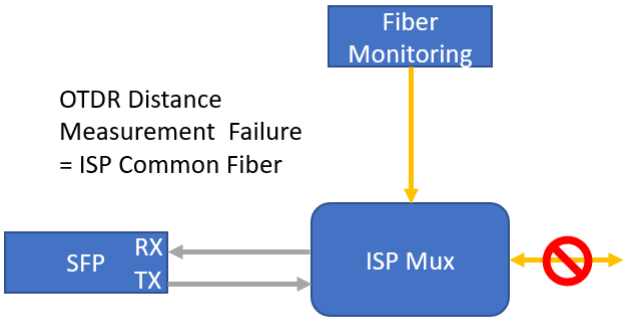


Figure 3 - ISP Common Fiber Link

4.4. Bi-Directional Link Delta

Utilizing end-to-end link loss calculations, we can identify single fiber impairments in the Outside Plant (OSP). To calculate a single direction loss, we use the SFP receive level minus the far side SFP receive level to determine the directional loss. Then take both link losses of the duplex connection to create a bi-directional link loss. Bi-directional link loss should be zero or close to it, within the tolerances of the SFPs. When one direction has more loss than the other a bad fiber between the SFP and the Mux is probable.

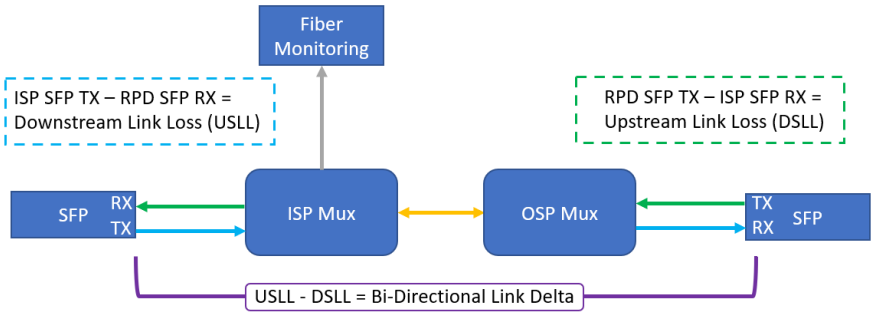


Figure 4 - Bi-Directional Link Delta

4.5. Multi-Metric Logic

Metrics alone get us halfway there; we are moving toward real-time remote divide and conquer of the optical network. However, logical assumptions can be made using the new metrics detailed above and existing metrics like fiber cut detection and low light. The individual readings are used to triangulate impairments in the optical network. The metrics together will validate the fiber or point teams to what does or does not need attention.

- When we can eliminate the ISP fibers as impacting light levels, but show a bi-directional delta impairment, this indicates an OSP impairment on a single fiber past the OSP Mux.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- When a fiber link is down and no light is received at the ISP SFP, we can eliminate or verify a common fiber cut impairment and confirm the SFP transmit fiber light to the Mux.
- When a link is down, but we are receiving light, we can identify the impacted fiber location ISP transmit fiber jumper, common fiber, or post OSP Mux transmit fiber.

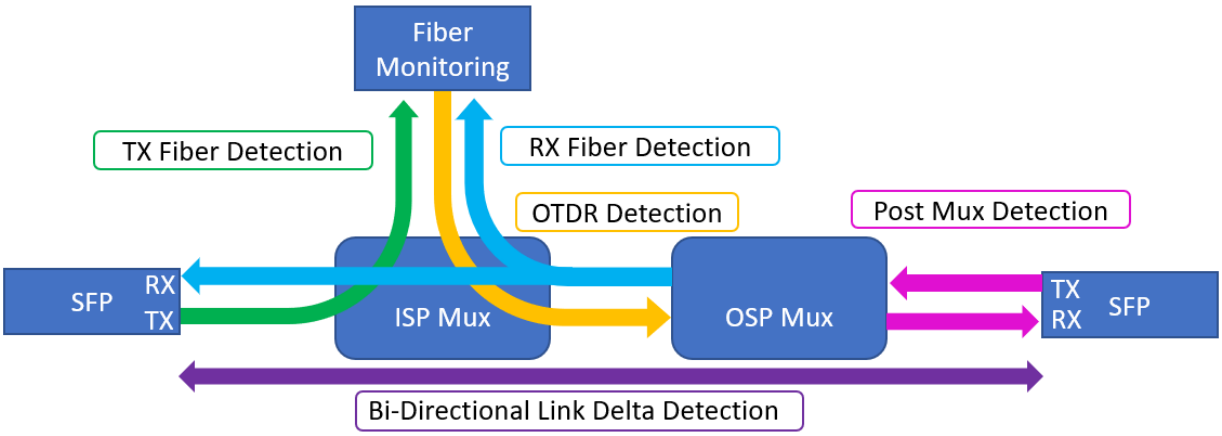


Figure 5 - Multi-Metric Logic

4.6. Next Steps

All the metrics mentioned previously can be measured in real-time with no historical context or baseline. When trending and deviations in levels are added with design information, the value is amplified to the next level and leads to component level reliability measurements.

5. Operational Evolution

As we move toward the future, teams that build and maintain optical networks will utilize intelligent tools. Real-time monitoring of the way we build, certify, and maintain fiber networks enhances the reliability and speed efficiency of how the network is maintained. With remote detection of individual fibers, we can focus the troubleshooting activity by quickly identifying the impacted fiber and dispatch a field technician to fix the issue. Currently, we troubleshoot the entire fiber link, which requires multiple teams. This is both resource-inefficient and time consuming.

5.1. Construction and Certification

The construction of new fiber links requires many individual components to work together properly. Traditional certification is limited to bi-directional OTDR traces and light-receive levels compared to the design link loss budget. When we add real-time measurements using monitoring and SFP data we can validate with a high degree of certainty that the link will work as expected prior to moving traffic to the fiber.

5.2. Proactive Monitoring

By leveraging real-time monitoring with enhanced detection logic, even slight variations outside of the expected thresholds can be identified and corrected, thereby improving the reliability of the link. Many

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

initially stable connections may become vulnerable to fiber-jumper impairments caused by bending, stretching, or disruptions to the ports, resulting in misalignment of the connections.

Prior to and following scheduled maintenance that involves fiber work, we have transitioned from merely validating network connectivity to remotely detecting issues such as dirty connections or faulty jumpers. This enables our network operations teams to gather valuable information and enhance reliability while our technicians are on-site, optimizing the scheduled maintenance process.

5.3. Fiber Restorations

When a fiber link is down, teams can quickly verify and isolate the state of ISP fibers. This removes steps where multiple teams would need to take measurements and combine to isolate the cause of the problem. In remote areas with unmanned locations this reduces the outage by eliminating drive times and troubleshooting steps. After restoration, links can be quickly verified that all components are operating as designed. This new insight and ability to direct teams to the fiber-impairment locations will improve device uptime and network reliability.

6. Conclusion

Fiber networks are becoming rich with data. Organizing and utilizing these new data streams will change the way we manage optical networks. Individual measurements have meaning when given context within the network design. The optical impairment detection methods serve as a map to isolate and identify fiber segments not operating correctly. This new methodology reveals where light is being lost, which is something that simple min-max and pass-fail criteria miss.

When teams are empowered with this new information, network reliability is increased and restoration times decreased. Bad connections, splices, optical ports, and stressed fibers can be identified by the team responsible for fixing them. Direct ticketing to teams with reduced troubleshooting steps and reduced referrals to other teams improves the customer experience.

Fiber networks are an integral part of connecting people, and a clear picture of this network will ensure we are always connected.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

DWDM	Dense Wave Division Multiplexer
ISP	Inside Plant
MUX	Optical Multiplexor
OSP	Outside Plant
OTDR	Optical Time-Domain Reflectometer
SFP	Small Form-factor Pluggable

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Why Location is Now a Thing in Wi-Fi:

Operational Considerations for Standard Power 6 GHz Wi-Fi

A Technical Paper prepared for SCTE by

Scott Voegele
Director, Product. Software And Services
CommScope
3871 Lakefield Drive, Suwanee, GA 30024
scott.voegele@commscope.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 3
2. Performance of Standard Power Wi-Fi 3
   2.1. Single-family Dwelling 3
   2.2. Multi-Dwelling Units 6
3. Automated Frequency Coordination 6
4. FCC Requirements for Use of Standard Power in 6 GHz 8
5. Device-to-AFC Messaging 10
   5.1. AFC Connection and Security 10
   5.2. AFC Messaging 11
6. Device Geolocation Acquisition 13
   6.1. Internal GPS Location Service 14
   6.1. Smart phone-based Location Service 14
   6.2. Location Validation on Power Cycle Event 19
   6.3. Height Determination 21
7. AFC Proxy Solutions 23
8. Expected Enhancements to Standard Power Rules 23
   8.1. Indoor-Only Standard Power Devices 24
   8.2. Extenders as Standard Power Subordinates 24
9. Conclusion 25

Abbreviations 27

Bibliography & References 28

List of Figures

Title **Page Number**

Figure 1 – Main Level Home Service Coverage with LPI AP 4
Figure 2 – Main Level Home Service Coverage with SP AP 5
Figure 3 – AFC System Architecture 7
Figure 4 – FCC 6 GHz Band Allocations 8
Figure 5 – FCC Part 15 Equipment Classes 9
Figure 6 – Sample Spectrum Inquiry Request Message 12
Figure 7 – Sample Spectrum Inquiry Response Message 13
Figure 8 – Use of Smart Phone To Determine SP Device Location 15
Figure 9 – Location Accuracy Reported by Mobile Phones 16
Figure 10 – Histogram of Distance between Phone Reported Location and Tester indicated Location 17
Figure 11 – Test Application Sample Result with Automated Dwelling Boundary Determination 18
Figure 12 – Logical Flow When Determining If An AP Location Has Changed 21
Figure 13 – US Housing Unit Percentage by Stories 22

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

The 6 GHz band has historically been licensed for point-to-point microwave communications as well as limited radio astronomy sites. In 2020, the Federal Communications Commission (FCC or "the Commission") enabled unlicensed use of the band allowing transmissions by Low Power Indoor-only (LPI) devices and indoor/outdoor Standard Power (SP) devices. Standard Power devices were given access to two portions of the band – 500 MHz in the Unlicensed National Information Infrastructure band Five (U-NII-5) and 350 MHz in the U-NII-7 band. Since licensed devices continue to operate in the band, the FCC requires that SP devices must operate under the control of an Automated Frequency Coordination (AFC) system to minimize interference with incumbents.

Given the relentless proliferation of Wi-Fi enabled devices as well as the steadily increasing demands on both Wi-Fi bandwidth and quality of service, access to this new spectrum is not only welcome but has the potential to be game changing. However, unlocking this new Wi-Fi resource places new requirements on providers of Wi-Fi services. This document provides details regarding the process of utilizing an AFC such that the benefits of Standard Power operations can be unlocked for US consumers. In addition, it provides guidance for how best to handle new requirements placed on the communications service provider. In particular, this paper addresses how to accurately determine the geolocation of installed SP devices and efficiently deliver this information to an AFC.

2. Performance of Standard Power Wi-Fi

As a starting point, it is important to reiterate the performance delivered by Standard Power over vs LPI within in the context of a typical single-family residence Wi-Fi deployment. Offering SP does add cost and new deployment requirements so these must be justified.

2.1. Single-family Dwelling

To demonstrate the substantial benefits of SP operation, a prior study was conducted in a typical, single-family dwelling.[1] An available 6E-compliant smartphone was tested against a multi-band 4x4 AP to evaluate rate/reach throughout the Wi-Fi house. In this study, the wireless router was placed in the central living room on the main floor. The mobile smartphone client with 2x2 antenna configuration was then moved room-to-room to measure data exchange rates and paint a coverage map for the house. The study was first undertaken with a low power indoor (LPI) effective isotropic radiating power (EIRP) spec access point (5 dBm/MHz power spectral density (PSD)) and the repeated with a Standard Power (SP) access point permitting up to 36dBm EIRP (23 dBm/MHz PSD). Per housing floor, the TCP bitrates were tabulated, and heat maps created. The results for the main floor for both LPI and SP are shown be :

[1] *Why 6 GHz Standard Power Wi-Fi is the Game Changer for Residential Use in the US*: Flesch, 2021

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

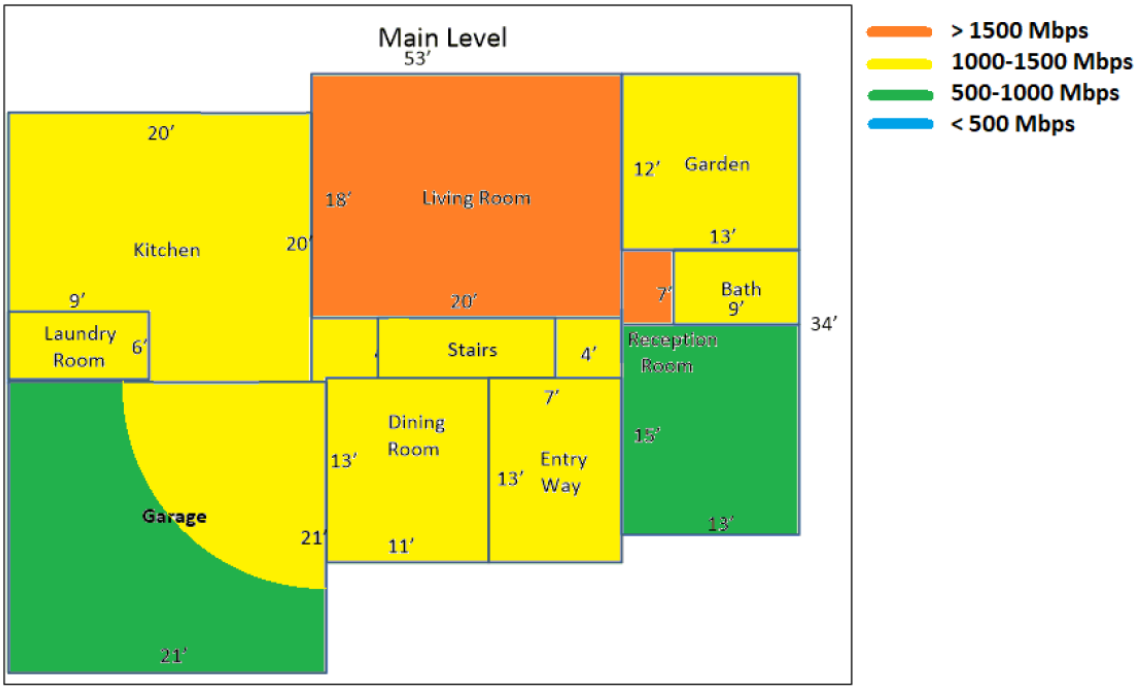


Figure 1 – Main Level Home Service Coverage with LPI AP

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

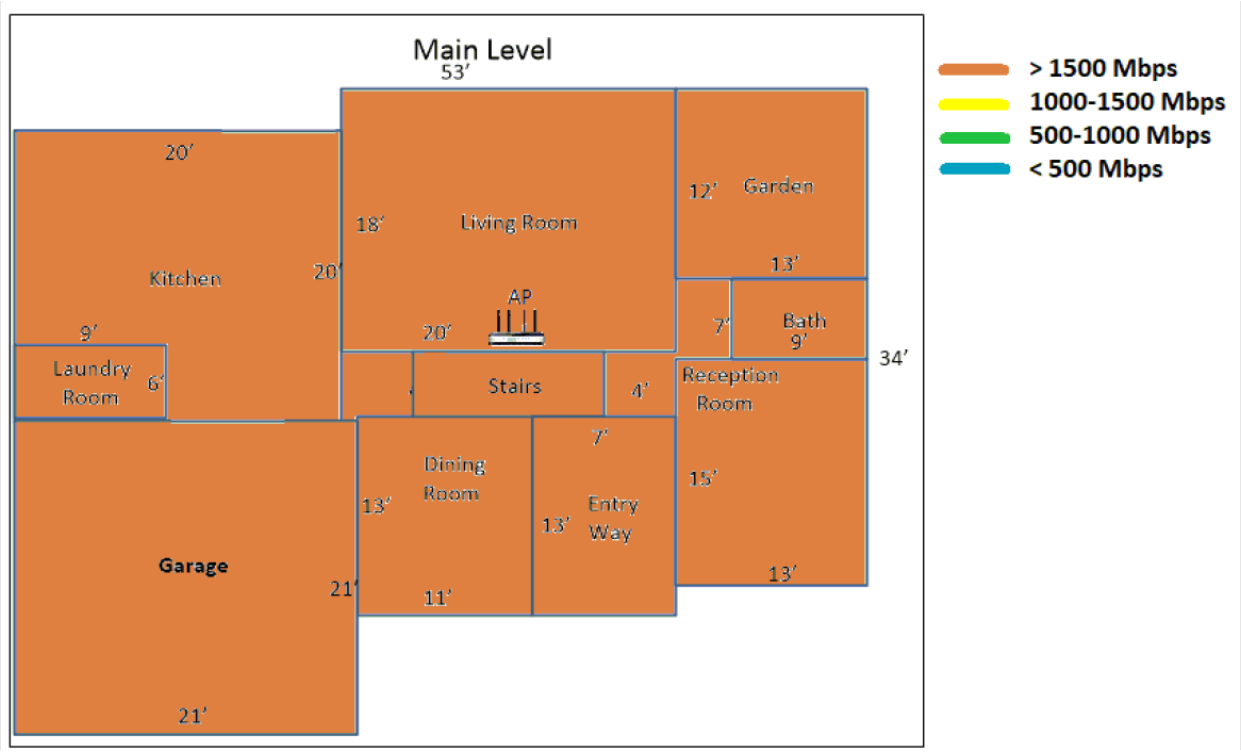


Figure 2 – Main Level Home Service Coverage with SP AP

As shown in the above figures, increasing to the higher output power (EIRP) of a single, centrally located AP significantly increased the downlink bit rate to clients. As such, when operating in standard power, a single AP can provide more than 1 Gbps everywhere, even in a larger home regardless of placement. This potentially reduces per home deployment costs by eliminating the need for multiple extender APs and allowing for a single installation point effectively anywhere within the home. In addition to increased raw bandwidth, standard power, 6 GHz client connections offer improved latency vs. what is possible in the lower Wi-Fi bands such as 5GHz. Leveraging standard power, "it should be possible to link large client populations throughout a spacious multi-story floorplan and still meet < 2 msec latency for all client links."[2]

When needed for very large or challenging home footprints, another important benefit of providing standard power indoors is to utilize that power to support wireless backhaul links to additional Wi-Fi Access Points (Aps). It should be noted that client devices operating in the 6 GHz band are required by FCC regulations to reduce their EIRP from that of the AP to which they are connected. This reduction is not insignificant (6dB) and must be applied regardless of whether the AP is operating using LPI or SP. The result is that coverage for a particular client in a home will be dictated more by the ability of a given AP to receive client transmissions vs. its ability to project downlink coverage to the same client.

Standard power 6 GHz can provide very high-speed connections between "mesh nodes". Adding one or more extenders will significantly impact coverage for client devices. Newer 6 GHz client devices supporting Wi-Fi 7 will not only take advantage of the raw increase in MCS/link speed. They will also

[2] *Why 6 GHz Standard Power Wi-Fi is the Game Changer for Residential Use in the US*: Flesch, 2021

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

benefit from full use of the range of Quality of Experience features available in the latest Wi-Fi standard including Multiple Resource Unit(s) (MRU).

Use of standard power in a home, while requiring the cost of AFC services and possibly requiring some added costs for SP hardware, can ultimately reduce the overall cost of deployment. Given the increased backhaul radius and higher downlink transmit power, it will cost less overall and create a more robust, less complicated network.

2.2. Multi-Dwelling Units

Unlike single-family homes, dense, multi-dwelling units (MDUs) may not benefit substantially from SP vs LPI operation. This may be somewhat expected given the typical MDU premise size and co-location of many access points in a single building. The US average apartment footprint is less than 1000 sq ft. Given this, offering standard power in these environments is not likely to substantially improve the customer Wi-Fi experience either in terms of delivered bitrate or latency vs. LPI operation.

In addition to providing minimal subscriber benefit, operation of SP within an MDU may have adverse consequences. For example, "use of 6 GHz could very well serve as a multi-dwelling source of OBSS energy which would tend to deny potential 6 GHz spectrum to very nearby alternate users. An additional aggravation would be that were standard power masks to be sought by several co-located users, the odds would be very high that any restrictive access to 6 GHz spectrum would be a shared restriction and so tend to funnel use into identical pieces of spectrum, compounding the interference potential."[3]

The exception to this blanket view lies in managed dwelling (enterprise) applications, where a single (or some very few connected) standard power Aps are used in concert with directional antenna coverage to supply stout trunk links to interior-placed extenders (each of the latter covering a single premises at LPI). The resultant subtended meshes are arranged for spatial and frequency division which provide the necessary keep outs to avoid the accidental interference (described above) which would otherwise occur. As far as the separate users in the MDU are concerned, each has his/her own AP – but it is actually a managed extender in a larger mesh network.

3. Automated Frequency Coordination

All kinds of wireless networks—those used by wireless operators, fixed wireless networks, enterprises and the ever-growing Internet of Things, for example—need access to high-capacity spectrum. Dynamic spectrum management methods such as the Automated Frequency Coordination (AFC) system unlock spectrum access by allowing these new users to share spectrum with incumbent licensees. 6 GHz AFC enables unlicensed access to the 6 GHz band by coordinating shared spectrum between Standard Power Access Points (SP) and incumbent point-to-point microwave.

Having established the benefits of operating using Standard Power, it is important to explore the core components of Automated Frequency Coordination. To access and utilize SP within the 6 GHz spectrum, an SP device must establish a connection with an approved AFC system and request what spectrum is available for use in a specified geographic location. The AFC uses FCC reference data on incumbent

[3] *Why 6 GHz Standard Power Wi-Fi is the Game Changer for Residential Use in the US*: Flesch, 2021. "Standard power masks" here referring to allowed channels (spectrum) as well as maximum output power per channel.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

systems, which may be combined with proprietary data to model the environment in the specified region and provide data on available spectrum.

Core functions of the AFC include:

- Allow unlicensed devices to securely register with the AFC individually or via a proxy
- Determine the available frequencies at a given geographic location or area and communicate this to devices
- Determine the maximum permissible radiated transmission power level for devices for given frequencies
- Enforce FCC Exclusion and Protection Zones
- Ensure that incumbent users of the 6 GHz band are protected from interference consistent with the FCC rules
- Collect incumbent data from the FCC Universal Licensing System (ULS) database and process it for use by the AFC
- Support interference identification and mitigation as needed identification

The figure below depicts the general functional architecture of the AFC managed 6 GHz system.

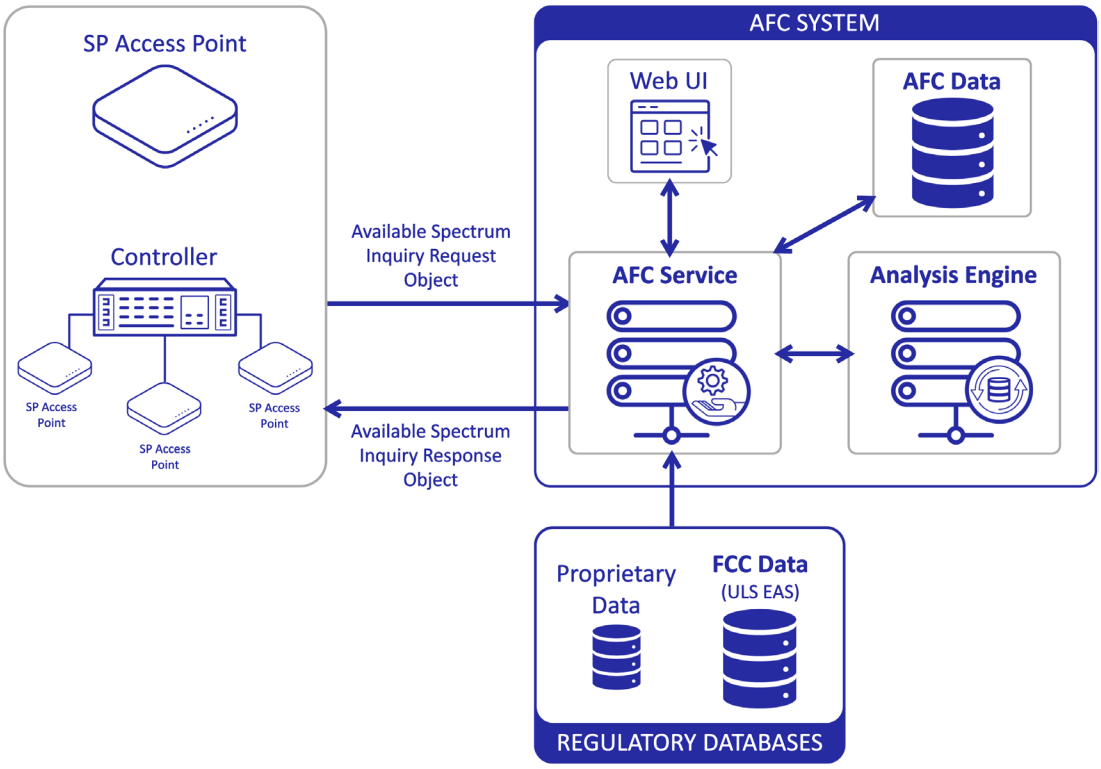


Figure 3 – AFC System Architecture

It is incumbent upon an AFC vendor to reliably provide a list of available frequencies and associated maximum transmit powers for standard power devices, so they don't interfere with incumbent microwave systems. Optimum performance of this task requires that the AFC maximize spectrum availability while thoroughly protecting incumbent microwave operations.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4. FCC Requirements for Use of Standard Power in 6 GHz

The FCC order regarding 6 GHz operation encompasses several spectral bands and different classes of products. Rather than reproduce the available FCC documents here, the following section focuses on critical requirements for a 6 GHz capable access point to operate in Standard Power. FCC product classes cover both Standard Power and LPI access points and clients.[4] Rather than reproduce the available FCC documents here, the following section and this paper more general will focus on critical requirements for a 6 GHz capable, non-fixed access point to operate in Standard Power. In the parlance of the FCC, this device is referred to as a Standard Power Access Point (6SD) and includes products like broadband Wi-Fi gateways and extenders.

Devices in this class operate in the 5.925-6.425 GHz (U-NII 5) and 6.525-6.875 GHz (U-NII 7) bands.

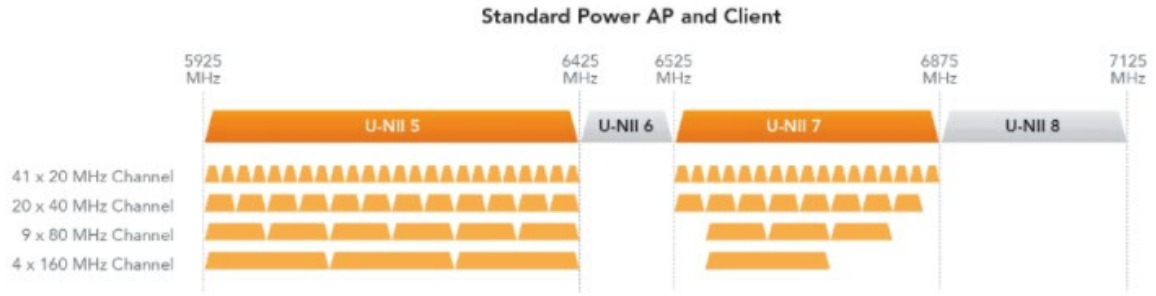


Figure 4 – FCC 6 GHz Band Allocations

Maximum power spectral density must not exceed 23 dBm EIRP in any 1-megahertz band. In addition, the maximum EIRP over the frequency band of operation must not exceed 36 dBm. For outdoor devices, the maximum EIRP at any elevation angle above 30 degrees as measured from the horizon must not exceed 125 mW (21 dBm).

SP access points are required to provide information to their connected clients such that those clients can adjust their EIRP to comply with the regulations. As noted in the prior section, even standard power clients associated with an SP must adjust their EIRP to no more than 6 dB lower than what is authorized by the AFC for the standard-power access point.

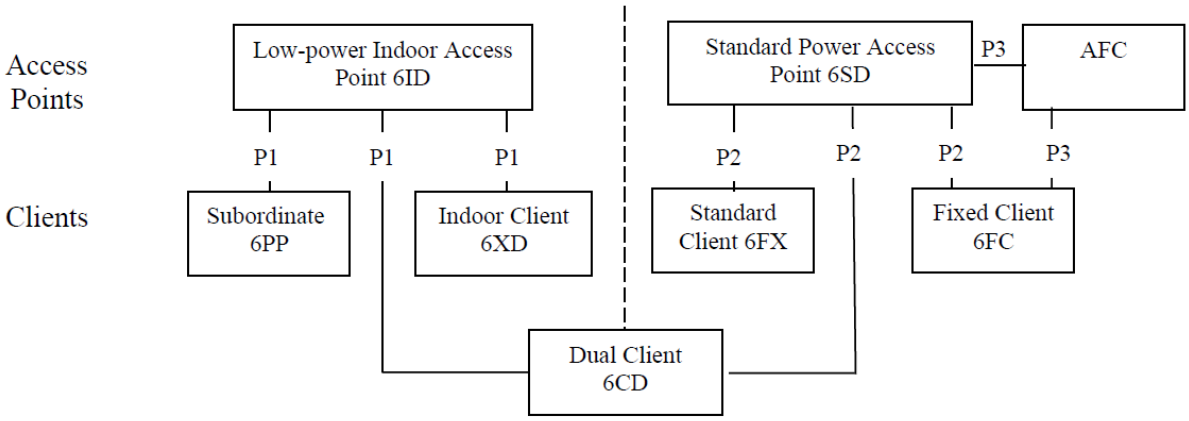
These devices must include specific labeling requirements and may operate as a bridge, peer-to-peer connection, connector between the wired and wireless segments of the network, or a relay between wireless network segments.

Importantly, devices in this class must access an AFC system to determine the available frequencies and the maximum permissible power in each frequency range at the device's geographic coordinates prior to

[4] Per the FCC, a client is a U–NII device whose transmissions are generally under the control of an access point and is not capable of initiating a network.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

transmitting and may transmit only on frequencies and at power levels that an AFC system indicates as available.



P1 Client and subordinate devices under control of low-power indoor access point.
P2 Client devices under control of standard access point.
P3 Standard Power Access Point and Fixed Client devices managed by the AFC.

Figure 5 – FCC Part 15 Equipment Classes[5]

To determine the appropriate channel masks, devices in this class must provide geolocation data to the AFC. To achieve this, the device must include either an internal geolocation capability or an integrated capability to securely connect to an external geolocation device or service to automatically determine the access point's geographic coordinates. The location uncertainty must be provided (in meters), with a confidence level of 95%. The device must report coordinates and location uncertainty to an AFC system at the time of activation from a power-off condition. Reporting of the device's height can be determined either automatically or by the device's installer/consumer and must include an associated height uncertainty. Additional details regarding this facet of SP operation are discussed at some length later in this document.

Registration of any SP device must be with, and operation be authorized, by an AFC system prior to initial service transmission. After a change in location, devices must obtain a list of available frequencies and the maximum permissible power in each frequency range at the device's location. Device's register with the AFC system by providing geographic coordinates (latitude and longitude referenced to North American Datum 1983 (NAD 83)), antenna height above ground level or sea level, FCC identification number, and unique manufacturer's serial number. If any of these parameters change, the device must provide updated parameters to the AFC system. All information provided by the device to the AFC system must be true, complete, correct, and made in good faith.

Consequently, all AP devices operating in standard power must contact an AFC system by rule at least once per day to obtain the latest list of available frequencies and the maximum permissible power the device may operate with on each frequency at the device's location. If the device fails to successfully contact the AFC system during any given day, the device may continue to operate until 11:59 p.m. of the

[5] FCC Knowledge Database 987594 D02 U-NII 6GHz General Requirements v0r02, August 22, 2023

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

following day at which time it must cease operations until it re-establishes contact with the AFC system and re-verifies its list of available frequencies and associated power levels.

Lastly, the device software must incorporate adequate security measures to prevent it from accessing AFC systems not approved by the FCC and to ensure that unauthorized parties cannot modify the device to operate in a manner inconsistent with the rules and protection criteria. Security mechanisms must also be in place to ensure that communications between devices and AFC systems are secure to prevent corruption or unauthorized interception of data. Additionally, the AFC system must incorporate security measures to protect against unauthorized data input or alteration of stored data, including establishing communications authentication procedures between client devices and standard power access points.

5. Device-to-AFC Messaging

In its ruling, the FCC refers to a document published by the Wi-Fi Alliance as the industry specification of the protocol between the AFC and a device.[6] It offers a technical specification that defines the messaging protocol and transport for the interface between an AFC System and an AFC Device. It includes the architecture, protocols, and functionality for entities that support the AFC System to AFC Device Interface.

5.1. AFC Connection and Security

The AP connects to the AFC via an HTTP POST method using HTTP version number 1.1 or later. In most cases, TLS is performed for AFC authentication by the SP device prior to any communication. This establishes an encrypted connection with the AFC system.

According to the FCC 47 CFR 15.407(k)(8)(v) requirement, standard power devices (SPDs) must incorporate adequate security measures to prevent unauthorized access to AFC Systems not approved by the FCC. Additionally, the AFC System may establish communications authentication procedures between the AFC System and SP devices or their Proxies.

The first mandatory rule requires that all SP devices must present valid, FCC-issued FCC IDs and must be validated by the AFC system when querying spectrum availability information. This is typically governed by the secured AFC system database preloaded with the FCC approved SP devices. SP devices establish a service relationship with a specific AFC system through registration.

The second highly recommended requirement pertains to the authentication between the devices and the AFC systems. This recommendation is outlined in the Wireless Innovation Forum (WInnForum) technical report WINNF-TR-2012. For the AFC System authentication, TR-2012 recommends employing a TLS Server Certificate issued by a Trusted Certificate Authority (WInnForum AFC System CA). This CA is attested using a CA certificate issued by a trusted root CA (WInnForum Root CA).

Regarding authentication, TR-2012 proposes two possible approaches. The first approach involves employing a TLS Client Certificate, mirroring the AFC System authentication through the server certificate. This method accomplishes mutual authentication of both the AFC System and SP device

[6] AFC System to AFC Device Interface Specification," Wi-Fi Alliance.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

during the phase of the TLS session establishment. In this approach, each device (or device Proxy) necessitates a distinct certificate. The SPD certificate may encompass information like the device's FCC ID and serial number, enabling the AFC System to conduct additional validation during SP device authentication. Similarly, the device client certificate might incorporate a manufacturer-provided unique identifier, facilitating additional validation during Proxy authentication by the AFC System. The second approach is to use Bearer Tokens issued to an SP device or Proxy by the AFC System. The AFC System will only generate tokens for trusted entities and thus they can be used to authenticate the holding entities. The Bearer Token is transmitted over an encrypted TLS session during communication. Bearer Tokens are part of OAuth 2.0, which is commonly used in industry.

The SP device/Proxy authentication based on a TLS Client Certificate offers a high level of security, constituting a robust authentication method albeit with relatively higher costs. In comparison, the Bearer Token approach is more cost-effective to implement and maintain, but it does not provide the same level of strength in security and authentication.

In summary, both the FCC and WInnForum stipulate that SP devices accessing AFC Systems must possess FCC-approved identities. The mutual authentications between AFC Systems and devices or their Proxies, while not mandated by the FCC, are recommended by WInnForum. It should be noted that diverse methods exist for implementing security and authentication within the AFC ecosystem and AFC operators have the flexibility to choose solutions that align with FCC Rules provided compliance can be substantiated. Nevertheless, it is likely that many operators will embrace WInnForum's recommendations to effectively address the growing concerns and demands for security. Adopting mutual authentication based on TLS certificates is a widely endorsed industry practice.

5.2. AFC Messaging

HTTP message payloads utilize JSON primitives. The AvailableSpectrumInquiryRequestMessage object is used by the SP device to request available spectrum from the AFC. The AFC responds with the AvailableSpectrumInquiryResponseMessage object. These two message objects represent the entirety of the AFC sequence flow.

The request object includes fields to describe SP operation including a device description, geographic location area for operation, a frequency range or list of channels associated with the request and optionally a minimum desired EIRP. The response object includes fields to describe the regulatory rules used by the AFC system to determine spectrum availability, the available frequency range(s) or channels including the maximum permissible EIRP for each, a time when the spectrum availability specified in the response expires, and information on the outcome of the inquiry. The following figures illustrate examples of AvailableSpectrumInquiryRequestMessage and AvailableSpectrumInquiryResponseMessage objects in JSON format.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

```
{
  "version": "1.3",
  "availableSpectrumInquiryRequests": [
    {
      "requestId": "2_132",
      "deviceDescriptor": {
        "serialNumber": "test1234",
        "certificationId": [
          {
            "nra": "FCC",
            "id": "OPS6GHZ"
          }
        ],
        "rulesetIds": ["US_47_CFR_PART_15_SUBPART_E"]
      },
      "location": {
        "ellipse": {
          "center": {
            "latitude": 36.36106,
            "longitude": -119.047
          },
          "majorAxis": 0,
          "minorAxis": 0,
          "orientation": 0
        },
        "elevation": {
          "height": 7.62,
          "heightType": "AGL",
          "verticalUncertainty": 0
        },
        "indoorDeployment": 2
      },
      "inquiredChannels": [
        {
          "channelCfi": [

          ],
          "globalOperatingClass": 132
        }
      ],
      "minDesiredPower": 24
    }
  ]
}
```


Figure 6 – Sample Spectrum Inquiry Request Message

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

```
{
  "version": "1.3",
  "availableSpectrumInquiryResponses": [
    {
      "requestId": "2_132",
      "rulesetIds":  "US_47_CFR_PART_15_SUBPART_E",
      "availableChannelInfo":
         {
            "globalOperatingClass": 132,
            "channelCfi": [7, 39, 55, 71, 135, 151, 167],
            "maxEirp": [27.8, 36, 36, 36, 36, 33.0, 36]

         },
       "availabilityExpireTime": "2023-07-08T14:35:05Z",
       "response":
         {
            "responseCode": 0,
            "shortDescription": "SUCCESS"
         }
    }
   ]
}
```


Figure 7 – Sample Spectrum Inquiry Response Message

6. Device Geolocation Acquisition

In support of Standard Power operation, a substantial new requirement is placed on the typical residential Wi-Fi device and by extension the service provider. This requirement is to accurately determine the installation coordinates and height of the device requesting SP operation. To determine the appropriate channel masks, devices in this class must provide geolocation data to the AFC. To achieve this, the device must include either an internal geolocation capability or an integrated capability to securely connect to an external geolocation device or service to automatically determine the access point's geographic coordinates. The location uncertainty must be provided (in meters), with a confidence level of 95%.

The device must report coordinates and location uncertainty to an AFC system at the time of activation from a power-off condition. Reporting of the device's height can be determined either automatically or by the device's installer/consumer and must include an associated height. In nearly all cases, the measured geolocation will be very similar (assuming the device is physically in the same location). It is likely that within the area of uncertainty, the geolocation coordinates will "match". Most AFC systems will, to improve efficiency, cache the previously calculated mask for that location/request. Masks will be flagged as invalid or removed from the system if new licensed 6 GHz operations are added to a specific locality in which case the spectral mask will be recalculated by the system.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6.1. Internal GPS Location Service

The most direct means to determine the location of a SP device is through the use of an embedded GPS radio. This allows the device to directly read its current location and include its current geolocation parameters in its daily request to the AFC.

Inclusion of the GPS system directly in the Standard Power AP does have some undesirable characteristics. One obvious consideration is cost. Each device is outfitted at the factory with a GPS device. Depending on the overall capabilities of the device (full gateway/router vs wireless AP only), this cost can be non-trivial. Also, if SP operation is only planned for a subset of the device population (for example, for customers in certain markets/regions or who purchase a specific type of service plan), the additional cost is unnecessarily invested for those customers or demands the use of multiple stocking units (SKUs), one for LPI and one for SP operation.

An additional consideration is that, without some type of visual indicator, it may be difficult for a technician or end user to determine if the correct geolocation is being presented. It is possible, especially in denser urban settings, that reflections in the GPS signal could occasionally result in incorrect coordinates being realized by the GPS system. To fully validate the location, these coordinates would need to be validated against another source such as the address of record where the device is installed. If the detected GPS location and address of record don't match, the device could be flagged, and manual intervention applied such as relocating the device slightly within the premise if possible. In the worst case, use of SP may have to be disabled if a valid reading cannot be obtained.

6.1. Smart phone-based Location Service

An alternative mechanism vs internal GPS is to use an external device to determine the geolocation of the SP capable device. The external GPS device by rule must be securely connected to the SP device. By using an external device, the SP device itself does not need to carry the burden of the GPS cost and, if the external device has a display, it can be used to visually confirm a proper location before submission to the AFC.

One obvious choice for the external device is the use of the ubiquitous smart phone to provide the needed geolocation information. The device can be securely connected to the SP device over Wi-Fi. To ensure that the smart phone is not too far from the SP device, a signal strength limit can be enforced to ensure that the smart phone is within close proximity to the SP device for which it is measuring the location. Any device geolocation mobile application components can be integrated with a service provider's mobile application to provide a more unified customer experience, though it is obviously possible to use an AFC Geolocation in a standalone form.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.



Figure 8 – Use of Smart Phone To Determine SP Device Location

Use of standard smart phone GPS is expected to be at least as accurate as onboard GPS radios. The performance of the GPS radios themselves are similar but, in the case of the smart phone, the operating system can also enhance the accuracy of the smart phone location using other available data.

To get a sense of the practicality of using a smart phone for measuring the location of a SP device, CommScope devised a brief study wherein random study participants used a test mobile application to sample their location as provided by the smart phone. Participants in most cases provided data when in proximity to their current broadband Wi-Fi device. The participant was then asked to pinpoint their actual location as closely as possible by touching the screen on a map overlay.

The data collected was quite positive in terms of reliably using the mobile phone to provide location services with the required 95% accuracy. As shown below in Figure 9 – Location Accuracy Reported by Mobile Phones, the vast majority of phones reported a GPS accuracy level within 20 meters and 96.9% reported accuracy within 50 meters.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

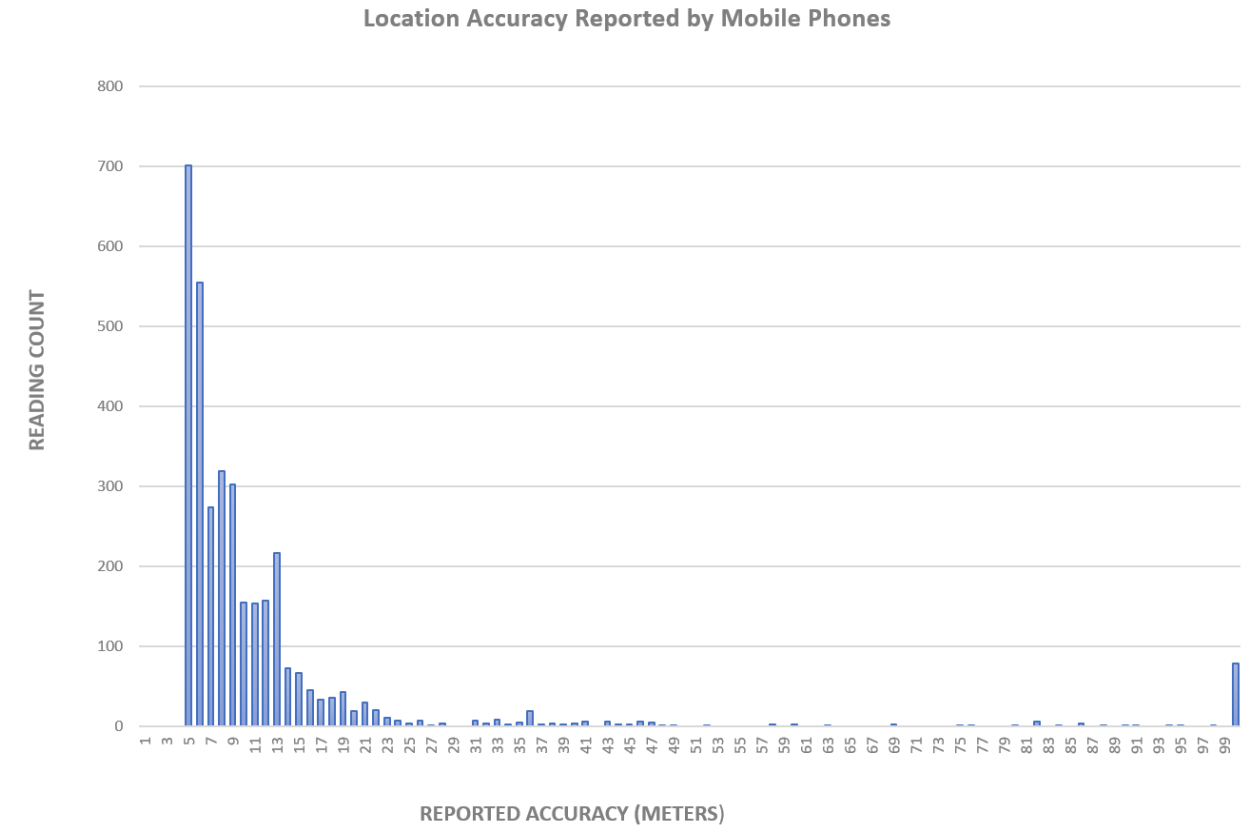


Figure 9 – Location Accuracy Reported by Mobile Phones

For the purposes of the test, users were instructed to indicate their actual location on the map overlay as precisely as possible by touch the screen. Doing this allows for further study as to how far the average center point of the GPS determined location was from the "correct" location as identified by the tester. Again, the results were promising. Of 32 recently completed tests with sufficient GPS reported accuracy, 32/32 (100%) were within a radius of 25.3m. 31/32 (96.9%) were within radius of 20.6m.

The normal distribution of these results are presented below. Note also that some error is introduced here by the resolution of a finger press on the screen. (The center pixel was used.) In a few cases, the underlying mapping software did not show a precise premise outline to help the tester more accurately touch their location.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

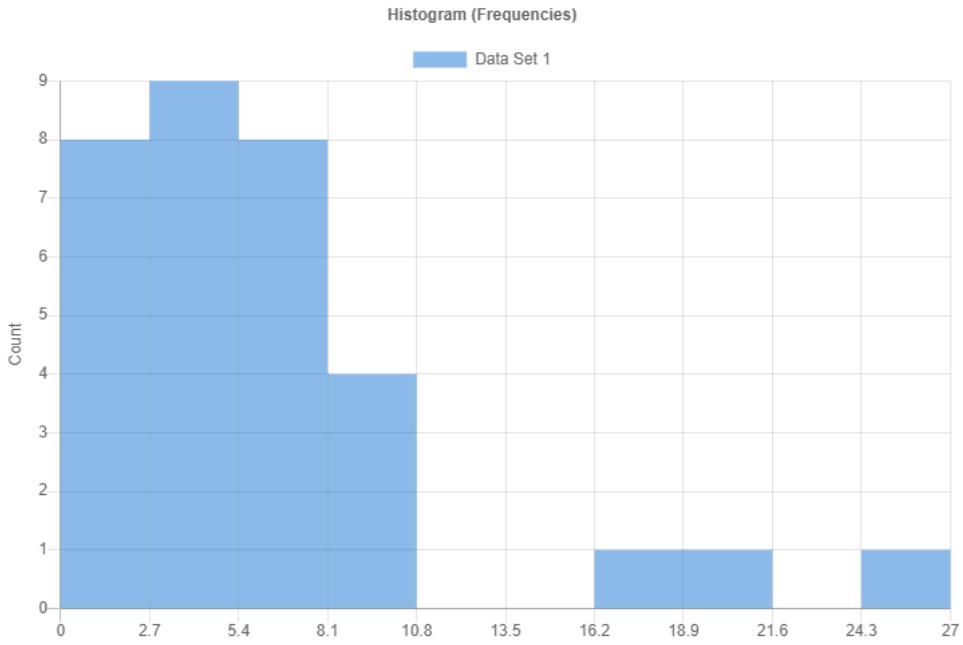


Figure 10 – Histogram of Distance between Phone Reported Location and Tester indicated Location

Using the same test application noted above, testers were shown dwelling structures and, based on the location of the indicated GPS coordinates, these structures were able to be automatically highlighted providing the precise outline of the premise in which the device was located. Using the boundaries of the premise as the area of uncertainty for AFC calculations offers a very precise means of determining the available spectrum mask and ensuring that the maximum number of channels and transmit power are made available.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

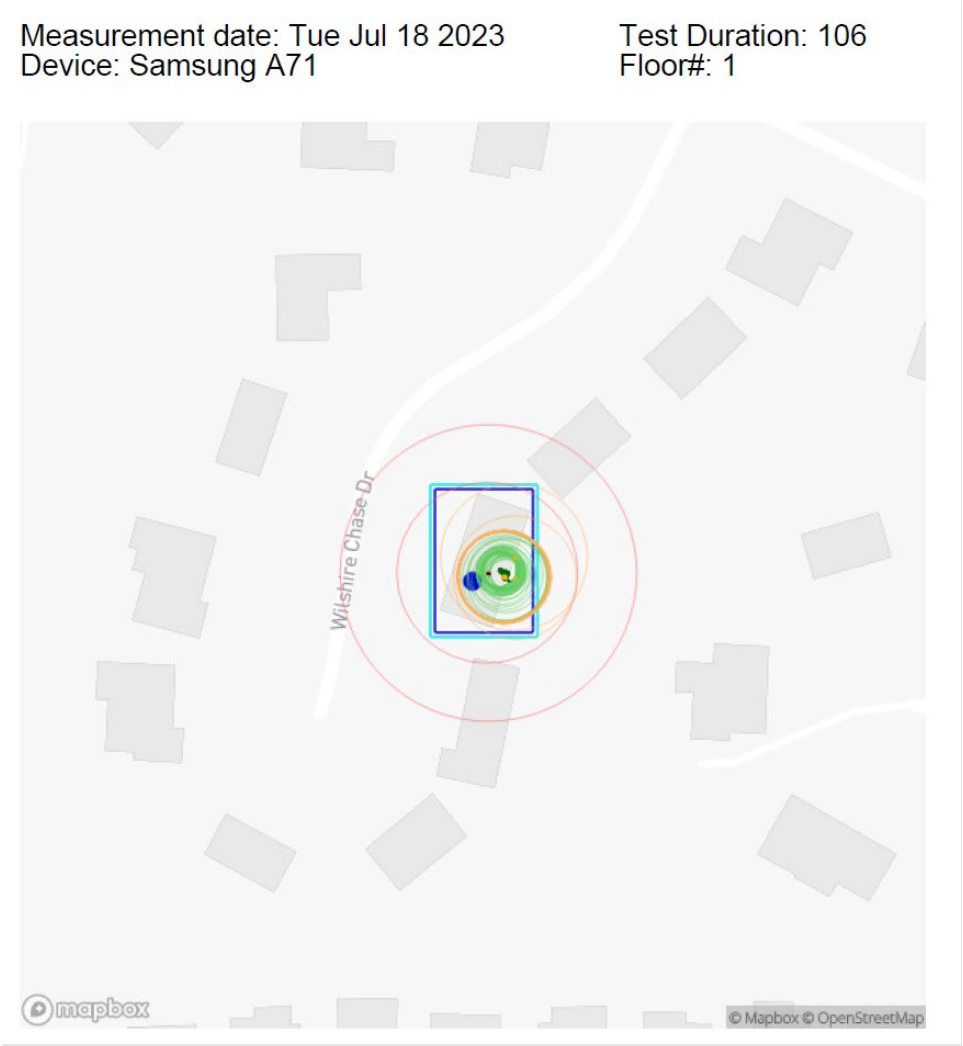


Figure 11 – Test Application Sample Result with Automated Dwelling Boundary Determination

An added benefit of the use of the smart phone for setting the geolocation of the SP is that mapping data, specifically structure related data can be used to further enhance location. This would allow the creation of a location and uncertainty that covered the entirety of the residence such that the channel mask would be valid anywhere inside the premise. In the consumer use case, the FCC is ultimately expected to allow for additional consideration for devices used inside of buildings. Thus "building loss" will be applied at the AFC when determining the proper spectral mask for the SP device given that the transmissions of the SP are attenuated by the external structure of the building itself. This is discussed in more detail in the section on" Indoor-Only Standard Power Devices" below.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6.2. Location Validation on Power Cycle Event

The current FCC specification does require the device to verify the device location after a reboot. If the device is in the same location/uncertainty area and the previously provided mask was obtained within the 24-hour window, no action is needed. If the location has changed, then the device must request a new set of mask information from the AFC. This requirement is straightforward in the case of integrated GPS, though given the infrequent nature and technical limitation of users relocating equipment from one premise to another, it again highlights that internal GPS may not be the most cost-effective means for SP geolocation determination.

In the case of an external location mechanism like a smart phone, it is possible to revert to Lower Power Indoor (LPI) operation after each reboot and then require revalidation of the location via the external device, using a smart phone notification as an example. However, this is impractical and would result in potentially extended periods of operation outside of SP. To avoid this, software-based validation mechanisms can be employed to determine if the SP device has changed location. These checks can be performed either with software running on the target platform or remotely using a cloud-based service which is remotely connected to the SP device.

As part of an SP device location and location error determination (initially or subsequently), baseline location "fingerprint" values can be determined for the SP device. The same fingerprint values can then be determined following SP device repowering and periodically (prior to an AFC Available Spectrum Inquiry Request) and compared with the baseline fingerprint values associated with its last location and location error determination.

If there is an adequately significant change in new and baseline fingerprint values such they there is a reasonable possibility the device has been relocated, its operation can be reverted to LPI operation and the smart phone or other external device/application is required to reconfirm the device location. Naturally, if there is no change or an insignificant change in new and baseline fingerprint values, the last measured location and associated error can be reused in a subsequent AFC Available Spectrum Inquiry Request.

Following are possible coarse "fingerprint" values to detect an inter-residence AP device move:

- Network access node (first-hop) address change detected from gateway
- Visible neighbor reporting (OBSS and signal strength)
    - Resulting from off or on-channel scans
    - Neighbors may add or remove devices which will change the BSSID information, however if there is a change, this can be ignored if other BSSIDs and beacon signal strength information remains unchanged (within threshold margin)
- WAN DHCP server response information
    - DHCP MAC address (presumably the access node MAC)
    - Router IP (option 6)
- Traceroute to given network server
    - Priority to the first few hops given that an PI route could be dynamically altered by the routing logic
- WLAN fixed client devices and their RSSI and/or PHY rate and gross number of known devices
    - Requires understanding of which client devices are fixed (e.g. video set-tops, printers, IOT devices, etc.)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

- Above may be discovered based on device type via DPI or host name or other means such as RSSI variation over time
- "Known" devices would be those retained over a period of time (e.g. weeks) to avoid the scenario of guest devices (e.g. gaming console) being seen over a relatively short period of time

The fingerprint values noted may be considered relatively variable, even without an SP device move. Consequently, they should be accompanied with a threshold margin to determine if a "change" in new and baseline values is significant enough to consider a location change. Also, to avoid false positive location move indications, some combination the multiple fingerprint values should be used. Moreover, fingerprint values with potentially high variability (e.g. static clients that are moved temporarily) should be omitted unless they persist over several measurement periods.

The choice of fingerprint values may also include a mix of both course and granular measurements. However, logic can be applied to look at course fingerprint value(s) before granular fingerprint value(s), since an inter-residence move rules out the need for granular value comparison. Also, based on the type of AFC frequency / power mask that was received previously, granular fingerprint measurements may not even be needed in a comparison.

An AP device's last measured location and associated error, its baseline fingerprint values, and new fingerprint values may be retained in the AP or in a device proxy that issues the AFC Available Spectrum Inquiry Request for the AP device. Naturally if used at a device proxy, this information must be accessible by proxy device.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The following block diagram shows the high-level logic for determining when an AP's location has changed and when a new location determination is needed.

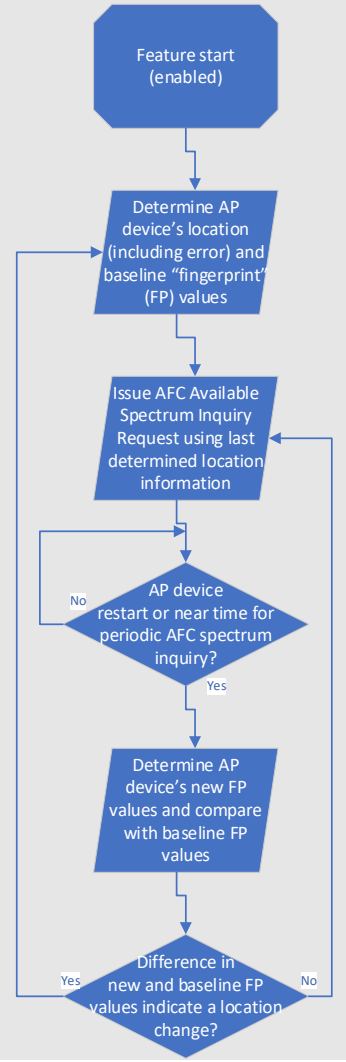


Figure 12 – Logical Flow When Determining If An AP Location Has Changed

6.3. Height Determination

To properly determine the channel mask for a given location, the AFC system must have SP height information in addition to the geolocation coordinates. To support this, as with the internal GPS system, one option is to embed an altimeter in the SP device. However, as with the inclusion of the GPS radio, this is a cost burden to the product. Also, unlike the GPS, the accuracy of these solutions is affected by real-time atmospheric conditions at the site. There are third party companies which, given a geolocation, altimeter value and time can use this information to accurately define height above ground to a specific structure floor. However, the geographic coverage of the specialized calibration sensors needed to make adjustments is somewhat limited and is primarily intended for urban locations with a greater number of

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

high-rise residential buildings. Use of these specialized systems which are more typically used for more mission critical applications like first responders will also add cost to the overall solution.

Correctly assessing height is further complicated by the practice of maintaining positive pressure in buildings, positive pressure meaning that the internal pressure inside the building is maintained slightly higher than the atmospheric pressure outside. Positive pressure is maintained for a number of reasons including lessening indoor humidity, preventing cold drafts in the winter and hot spots in the summer, and reducing energy costs due to untreated air entering the building. A positive pressure error would return a height lower than actual.

Given the above, it is likely that the height will be typical be manually provided by a technician or the Wi-Fi device owner. As before, use of a smart phone app to locate the SP can also be an effective tool for determining a reasonable height assessment. It is important to note that most residential structures in the United States are three stories or less. Therefore, a technician or end user can simply input the maximum number of stories of the structure. This value can be used to determine a maximum height using commercially available data for the specific geolocation and estimates of the typical height measurement of a single story plus a small amount of uncertainty. It should be noted that the FCC contemplated this in their order and, unlike geolocation, does allow for manual entry of the height assuming an appropriate uncertainty.

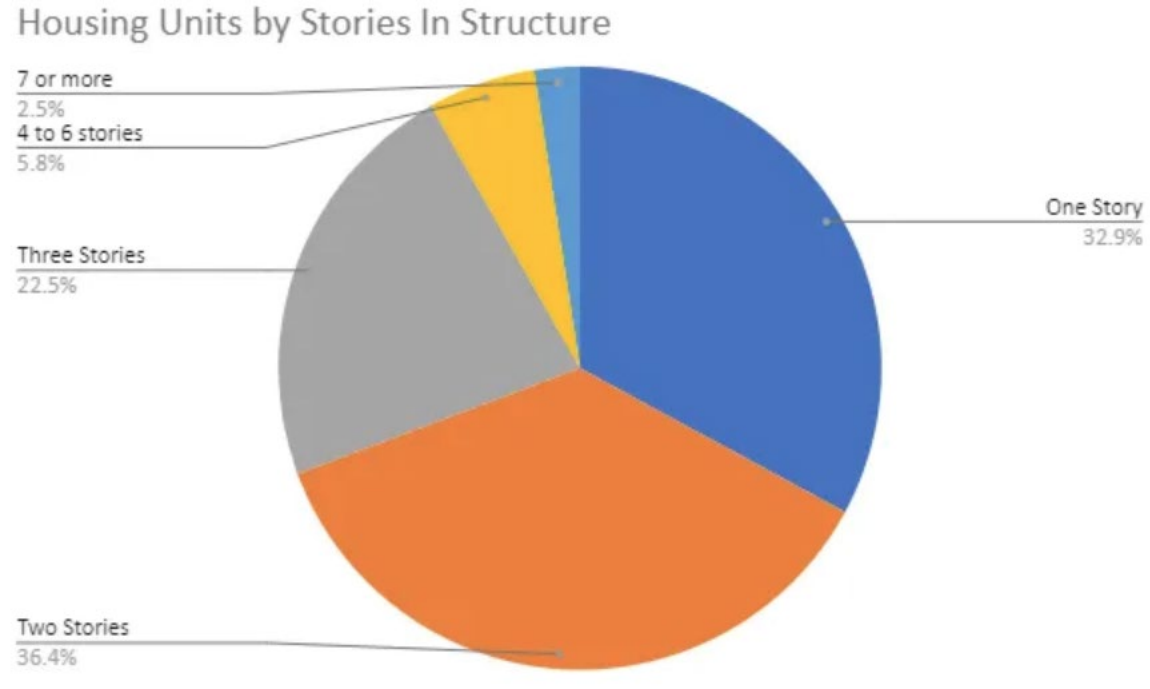


Figure 13 – US Housing Unit Percentage by Stories

For high rises, the entry level floor of the unit in which the SP device is installed can similarly be used to estimate the height. This value can be additionally validated by the height value provided by the mobile phone. Though not the primary focus of the above referenced smart phone study, height data captured does appear to provide an adequate level of accuracy to minimally validate manually entered data. This

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

would be in addition to some entry-based validation (e.g. disallow negative or excessively large numbers). For example, if the customer indicates a low-floor or one-story residence but the smart phone app is reporting significant height above ground, the entry app would require re-confirmation or disallow the entered value.

7. AFC Proxy Solutions

Leveraging a proxy between the SP device and the AFC system is permitted by the FCC and can provide several benefits. Service providers can utilize their existing gateway management solution including its well understood security and network-to-device communication elements such as the User Services Platform (USP) and BBF data models to facilitate 6 GHz operation. In addition, an AFC proxy can be used to provide additional insights into the performance of the 6 GHz network such as compliance reporting based on the actual channels in use in the event of an incumbent interference complaint.

Using an AFC proxy provides a means to reduce hardware costs by supporting software-based approaches for setting geolocation data in the primary residential gateway at the customer premise. It may also be possible in the future to leverage the GPS data generated by the main SP device in a premise to provide AFC support to connected extenders as well via the AFC Proxy. The AFC Proxy solution can gather location data from GPS systems embedded within the CPE with data inserted into the BBF TR-181 object model or directly from an external phone application. Once this information is gathered, the AFC Proxy can request updated Channel Mask data for updates from AFC database systems.

Employing the AFC Proxy allows for additional insights and usage reports to be generated around 6 GHz channel utilization. For example, the AFC Proxy can report devices that that been provisioned with a Standard Power mask but are still in LPI mode which may be the result of a software bug. In the event of an incumbent reporting interference on a licensed link, an AFC Proxy can generate a report confirming that all devices within the relevant geographic area of the link under dispute are operating on channels as authorized by the AFC. Generally, use of an AFC Proxy in conjunction with the AFC may reduce the likelihood of incumbent interference issues. Providing direct enforcement of the received channel mask, the proxy can force devices back into LPI or compliant channels when it is detected they are operating outside the prescribed channel mask.

Performance and compliance reports can be derived from the AFC Proxy database and audit updates done. Reports can be delivered direct to email based on an external utility which provides the reporting information.

8. Expected Enhancements to Standard Power Rules

As of the time of writing, there remain some outstanding topics which it is expected that the FCC will address. As is typically the case when the FCC makes substantial new rules, these tend to evolve over a period of time as new clarification requests and related public comments are considered. Below, we discuss some key subjects which have yet to be fully adjudicated but are important to the potential effectiveness of 6 GHz operation.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

8.1. Indoor-Only Standard Power Devices

In February 2023, the Wi-Fi Alliance filed a Request for Waiver with the FCC. This limited waiver request was related to the application of building loss when calculating the available 6 GHz channel mask for a given SP AP located indoors. Application of Building Entry Loss (BEL) can have a significant effect on the available channels as this loss is expected to be applied at -20.5dBm.

Taking into account the building entry loss is a simple "acknowledgment that interior wall and flooring transitions will serve to attenuate and scatter 6 GHz Wi-Fi energy on its outbound journey. Further, depending upon exterior construction, building entry loss (BEL) anywhere from 6 to 30+ dB will constrain the radiation footprint on its inside-to-outside propagation."[7]

As noted in the waiver request, the rules for unlicensed operations in the 6 GHz band authorize two distinct device categories: standard-power and low-power indoor ("LPI") devices. LPI access points ("6ID") are referenced earlier in this paper. LPI devices do not require a connection to an AFC but must operate in a limited number of bands and must operate using a relatively low power ceiling. These devices are limited to indoor locations, have an integrated antenna, and cannot use a weatherized enclosure. Low-power indoor access points must be powered by a wired connection and not by battery power. Low-power indoor access points may use battery backup only during power outages. In addition, they must include labeling indicating that the devices are for indoor use only.

In their waiver, the WFA argues that devices that meet the LPI requirements but also have the capability to transmit in Standard Power, should be given special consideration when compared to the standard SP device type. Since these devices meet the indoor-only constraint, the AFC should be able to take BEL into consideration when applying predictive propagation models to determine available spectrum. The WFA argues that this will provide greater access to the 6 GHz band and by extension improved bandwidth and coverage to the public, an assertion is borne out by empirical testing as noted earlier in this paper.

The mechanics of this process are still being worked out at the FCC, but will entail some method related to specific information a device's record in the FCC Equipment Authorization Database that will be available to the AFC to indicate that the device is an indoor-only SP AP.

It is expected that the FCC will be addressing this issue in the near future. As noted by the WFA in the conclusion of its waiver request, "Waiver of the rule will promote the Commission's goals, consistent with the public interest, by allowing … devices greater access to available spectrum, allowing further use by the American public of the critical connectivity capacity provided by the 6 GHz band."[8]

8.2. Extenders as Standard Power Subordinates

The use of Standard Power Wi-Fi extenders presents a unique challenge to the AFC managed Wi-Fi ecosystem. Since the FCC has indicated that standard power Wi-Fi extenders are not yet allowed per se, it is worth discussing how that may be accommodated if and when they are allowed.

---

[7] *Why 6 GHz Standard Power Wi-Fi is the Game Changer for Residential Use in the US*: Flesch, 2021
[8] *Wi-Fi Alliance Waiver Request in the matter of Authorizing 6 GHz Band Automated Frequency Coordination Systems,* Federal Communications Commission. ET Docket No. 21-352

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Registering Wi-Fi extenders will incur additional costs potentially in terms of including GPS systems in the extender as well as transactions against the AFC. An AFC system solution could address this by providing a mechanism in conjunction with an AFC Proxy offering to consider devices originating from the same household together, potentially reducing the overall deployment costs for households using extenders.

The fundamental issue to overcome here is that a co-located extender device in the same home mesh is permitted use of standard power (as is the main AP)–- but such should not produce an alternate operational spectrum mask if the same geolocation error footprint is used for both devices. The use of a proxy device to query the AFC system simply helps scale back the redundant query and calculation burdens of the AFC system in this regard and is easily achieved with a very slightly modified use of the original GPS location effort.

Recall that the main AP establishes its location and error footprint by leveraging either its internal GPS capability or that of a smartphone. In the smart phone case, proximity to the main AP is anchored via Wi-Fi RSSI value. In the extender case, geolocation registration can be performed but as an additional calculation step, this location and error can be compared with the main AP's and a virtual single device created with location midpoint of the two fixes and with a resultant error footprint the maximum of the two added together. This is supplied to the proxy which, on 24-hour mask lease renewal boundaries, requests a spectrum power mask on behalf of the virtual single device and then supplies this both to the main AP and its extender. Note that the above approach may preclude the use of extender-to-extender (multi-hop) mesh networks. Lack of multi-hop support (enforced by the devices themselves) is not a significant concern as noted earlier, a single SP access point can cover a large residential footprint allowing a "hub-and-spoke" only architecture with extenders to cover even the largest homes.

Note that this proposed methodology is not yet approved by the FCC but is intended shortly for PIA submission as the AFC mandate matures. Part of the FCC concern in this regard has to do with footprint growth which might permit obfuscation of the use of BEL (building entry loss). leading potentially to it being incorrectly used for an externally operating device (such resulting in an overly permissive mask being provided to an outdoor unit). But this can be quashed by the simple expedient of examining the FCC ID detail for the virtual device, which the proxy can validate based on its own copy of the FCC ID details for the actual devices installed. (Note that the proxy would have to supply a worst-case "outdoor" designation for the virtual device if ANY of the actual devices were anything but indoor units; it may be necessary to proxy the main AP details – so modified – in order for the AFC to acknowledge a known FCC ID).

9. Conclusion

Availability of the 6 GHz band is a welcome development, not only in the US and Canada but globally. Given the unrelenting growth of Wi-Fi connected devices and, in some cases, their need for much higher and reliable throughput, the availably of more spectrum, especially when combined with a new Wi-Fi generation (Wi-Fi 7), is an important development which is likely to have a substantive impact on both services providers and their customers. As noted in this paper, use of Standard Power can significantly extend high-bandwidth, downlink reach for US consumers vs LPI.

However, accessing this new spectrum using SP does require some added complexity. In particular, service providers must partner with an AFC system operator to coordinate spectrum with incumbent

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

licensees. The AFC determines both the available frequencies for a provided geographic location or area and the maximum permissible radiated transmission power level for devices for given frequencies. Further, the Standard Power devices themselves must meet specific requirements, the most important being to provide the AFC an accurate location and height such that the AFC can properly determine the above referenced channel and power mask.

Understanding device location represents the most significant and novel change for operators. Several options are possible, each with their own benefits and challenges. Including GPS radios and altimeters in Standard Power devices represents the most straightforward means of determining device location and height. However, it does add fixed cost to every device regardless of whether it is used in Standard Power operation. Use of an external solution such as a smart phone app is more flexible, does not add fixed costs, and allows for more technician and customer interaction to help confirm the proper location and height. However, use of an external location solution requires additional software either in a cloud proxy or the device itself to assess whether a device has been relocated in the event of a power-cycle. Automatically determining height involves additional complications due to the impact of variability in the local atmospheric conditions and the practice of maintaining positive building pressure.

Finally, continued innovation and evolution is expected with regard to the authorization and implementation of 6 GHz Standard Power operation. AFC Proxy systems will come to market which will provide additional benefits and features such as allowing for reuse of already deployed security and network-to-device communication elements to facilitate 6 GHz SP operation. In addition, AFC proxies can provide valuable 6 GHz performance insights and reporting and potentially allow for more efficient management of 6 GHz devices (gateways and extenders) operating within a single household. On the regulatory front, ongoing rule making is expected to allow the use of an allowance for building entry loss which should improve 6 GHz spectrum availability in many cases.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

6SD	FCC Standard Power AP Device Type
AFC	Automated Frequency Coordination
AP	Access Point
BBF	Broadband Forum
BEL	Building Entry Loss
CA	Certificate Authority
CFR	Code of Federal Regulations
dBm	Decibel-milliwatts
DHCP	Dynamic Host Configuration Protocol
DPI	Deep Packet Inspection
EIRP	Effective Isotropic Radiated Power
FCC	US Federal Communications Commission
GPS	Global Positioning System
HTTP	Hypertext Transfer Protocol
HTTPS	HTTP plus TLS
IoT	Internet of Things
IP	Internet Protocol address
JSON	JavaScript Object Notation
LPI	Low-Power Indoor
MAC	Media Access Control
MDU	Multi-Dwelling Unit
MRU	Multi Resource Unit
NAD 83	North American Datum 1983
OBSS	Overlapping Basic Service Set
PIA	Public Information Act
PSD	Power Spectral Density
SCTE	Society of Cable Telecommunications Engineers
SP	Standard Power
TLS	Transport Layer Security
ULS	Universal Licensing System
U-NII	Unlicensed National Information Infrastructure
URL	Uniform Resource Locator
USP	User Services Protocol
WAN	Wide Area Network
WFA	Wi-Fi Alliance
WInnForum	Wireless Innovation Forum

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

"Unlicensed Use of the 6 GHz Band; Expanding Flexible Use in Mid-Band Spectrum Between 3.7 and 24 GHz." Report and Order and Further Notice of Proposed Rulemaking, 35 FCC Rcd 3852 (5) (2020).

Federal Communications Commission, Office of Engineering and Technology. (August 22, 2023). *U-NII 6GHz General Requirements v02.* Publication 987594.

Wi-Fi Alliance. (2023). *AFC System to AFC Device Interface Specification*. Version 1.5.

Wi-Fi Alliance Waiver Request in the matter of Authorizing 6 GHz Band Automated Frequency Coordination Systems, Federal Communications Commission. ET Docket No. 21-352 (filed February 17, 2023). *https://www.fcc.gov/ecfs/document/10217816203286/1.*

WInnForum. 29 June 2023. *6 GHz Communication Security Technical Report.* Document WINNF-TR-2012. Version V1.0.0.

Flesh, J.R. (2021). *Why 6 GHz Standard Power Wi-Fi is the Game Changer for Residential Use in the US.* SCTE.

NGA.STND.0036_1.0.0_WGS84 (Version 1.0.0 - July 8, 2014): Department of Defense (DoD) World Geodetic System (WGS) 1984

IETF RFC-8259, "The JavaScript Object Notation (JSON) Data Interchange Format", December 2017

IETF RFC-8446, "The Transport Layer Security (TLS) Protocol Version 1.3", August 2018

Potter, Brian. "Every Building in America - an Analysis of the US Building Stock." *Construction Physics*. November 2020. *https://www.construction-physics.com/p/every-building-in-america-an-analysis* (accessed July 17, 2023).

Holland, Garrett. "Positive Building Pressurization – Why is it important?" *Working Buildings.* June 2019. *https://www.workingbuildings.com/blog_pressure.html* (accessed July 17, 2023).

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Wi-Fi Privacy and Security: Evolving to Keep up with Marketplace Demands

A Technical Paper prepared for SCTE by

Carol Ansley
Lead Architect
Cox Communications, Inc.
Carol.ansley@cox.com

Tushar Sharma
Lead Network Engineer
Cox Communications, Inc.
Tushar.sharma@cox.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 3
2. Terminology 3
2.1. Privacy Terminology 3
2.2. Wi-Fi Terminology 4
3. Wi-Fi Privacy and Security Standards Activity 4
3.1. Random and Changing MAC Addresses 5
3.1.1. What is a MAC address? 5
3.1.2. MAC Addresses and Wi-Fi 5
3.2. Additional Corelatable Information with Wi-Fi 6
3.3. Current 802.11 Task Group Activities 6
3.3.1. TGbh - Operation with Randomized and Changing MAC Addresses 6
3.3.2. TGbi – Enhanced Service with Data Privacy Protection 7
3.3.3. Overlap between TGbh and TGbi 8
4. Wi-Fi Privacy and Security Use Cases 8
4.1. Mobile device not recognized due to RCM 8
4.1.1. Parental controls 9
4.1.2. Help Desk interactions 9
4.1.3. Changes planned in TGbh 9
4.1.4. Changes planned in TGbi 9
4.2. User tracking by third party observers 10
4.2.1. Smart Home 10
4.2.2. Changes planned in TGbi 10
4.3. Open SSID and DMCA/CALEA 10
4.3.1. DMCA with Open SSIDs 10
4.3.2. CALEA/LI on Open SSIDs 11
5. Conclusion 12
Abbreviations 13
Bibliography & References 13

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

When Wi-Fi™ was first developed over twenty-five years ago, no one expected that it would become the dominant method of connecting consumer mobile devices into the larger network wherever people went. The past 10 years have seen Wi-Fi connected mobile devices go from a business convenience to a personal and professional necessity. More recently consumers using Wi-Fi mobile devices have become aware that their privacy was being impacted by these ubiquitous devices. A person could be tracked by the mobile devices they carried as they worked, shopped, and traveled. The privacy concerns associated with Wi-Fi have been written about in many papers. Information exposed in Wi-Fi headers and unencrypted control frames have been highlighted as providing a window for third parties to monitor and track users through their mobile devices.

IEEE 802.11™ first addressed these concerns with IEEE Std 802.11aq™ that recommended various steps such as randomizing a mobile device's MAC address to obfuscate the user's identity. The consumer electronics industry has enthusiastically adopted this recommendation which ended up having ripple effects in other areas. We discuss possible mechanisms to reduce the effect of randomized MAC address that are being considered by the Random and Changing MAC Address Operation Task Group in IEEE 802.11.

We will discuss ongoing work in 802.11 including proposed protocol changes that are being considered within IEEE 802.11 such as the work being done by the Enhanced Data Privacy Task Group in IEEE 802.11. The task group is developing new tools to reduce the ability of a third party observer to identify and track Wi-Fi mobile devices, and the users associated with them.

We also provide some references in the *Bibliography and References* section for further reading on similar efforts in the 3GPP Standards. Such references are not exhaustive but suggested literature.

2. Terminology

2.1. Privacy Terminology

Since we are covering different aspects of privacy in this paper, agreement on common terminology is useful. Privacy experts have developed categories for information to indicate its relevance to privacy concerns. Within the Recommended Practice for Privacy Considerations for IEEE 802 Technologies (IEEE Std. 802E-2020), these definitions below are useful to consider as we discuss potential improvements to Wi-Fi protocols.

- *Personally Identifiable Information (PII)* – Any data that directly or indirectly identifies a person or from which the identity or the contact information of a person can be derived.
- *Personally Correlated Information (PCI)* – Data gathered about an identified person or small group of people by observing activities (e.g., communications) or events associated with those people.
- *Personal device* – a device associated with a person or a group of users, such that identification of the device also allows identification of its user or group of users.

In the sections that follow, mobile devices will be discussed with the assumption that these devices are personal devices. For example, a smartphone is a common personal device. People rarely share smartphones with others, and commonly carry them wherever they go.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

The information discussed later in connection with Wi-Fi standards activities will be mentioned as PCI, where it allows data to be correlated with certain people or groups of people, and PII, when it allows data to identify, usually indirectly, a specific personal device and the device's user.

2.2. Wi-Fi Terminology

Wi-Fi also has a lot of terminology specific to its technology. A Wi-Fi system generally consists of an Access Point (AP) which is a unit with a connection into a wider network, usually a wired network, and a Wi-Fi radio interface. Connecting on that radio interface are stations, which are usually mobile devices using the AP to reach the wider network.

When a mobile device such as a smartphone establishes a connection to an AP, that process is called associating with the AP. At the same time the mobile device can also authenticate itself to the AP and the wider network in a security association. If a mobile device moves out of range of an AP, the AP generally will consider the device to have de-associated and will require the device to go through the authentication and association processes again to rejoin the network when the device returns to the AP's wireless range.

A mobile device determines which APs are in range by sending out Probe Request messages. These messages ask for APs to return Probe Response messages that contain information about their services. Similarly, Association Request and Association Response messages are used by mobile devices to request an association with an AP and by the AP to respond to that request, respectively.

3. Wi-Fi Privacy and Security Standards Activity

Wi-Fi standards originate from both the IEEE Standards Association and the Wi-Fi Alliance. In this section, we are discussing IEEE SA standards activity.

Due to a lack of a standardized device identifier, an IEEE MAC Address has become the *de facto* identifier for devices and by extension their users. However, this identifier lacks Security and Privacy constructs that have become increasingly important to users. For instance, a device can start using a MAC address that it created randomly. If it wants to impersonate another device at the MAC level, it just has to start using the MAC address of that other device.

Additionally, the 802.11 protocols did not initially focus on minimizing exposure of identifiable parameters, like the MAC address, to third party observers. Additional parameters shared openly in messages such as a Probe Request or Association Request message may also allow a mobile device's activities to be correlated with specific targets, and by extension may expose PCI or PII of the device's user.

A deeper dive into MAC randomization and other related Wi-Fi topics are discussed below. MAC Randomization enhances privacy and security but complicates the customer experience and network management since device OEMs have varied implementations. From a broader perspective, within the current generation of 802.11 protocols, a mobile device's MAC address is expected to remain stable while the device is associated with an AP. This continued use of a single MAC address allows tracking of a mobile device and its user. The discussion below will address this use case and ideas that are under discussion to reduce the exposure of PCI, such as a personal device's location. Finally, Open SSIDs typically use the MAC Address as a mobile device identifier and any subsequent changes to this identifier are treated as a new mobile device complicating Customer Experience. A discussion of potential impacts to existing operational functionality are covered in a discussion below of the privacy challenges associated with Open SSID.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

3.1. Random and Changing MAC Addresses

3.1.1. What is a MAC address?

For those who have spent a lot of time in this area, the authors invite you to skip ahead to the next section. But if you are curious, a Medium Access Control address, or MAC address, was originally developed to be able to distinguish one device from the next on a network. The MAC addresses in Ethernet headers allow a device to determine if an incoming message is addressed to it, for example, by matching the MAC address in the header of an incoming message with the device's MAC address. MAC addresses are also extensively used in the IP networking protocols for mapping IP Addresses to individual network interfaces on network infrastructure entities, such as DHCP servers.

The most common 48-bit MAC addresses consist of 12 hexadecimal digits with 2 bits indicating whether the MAC address is a unicast or multicast address and whether it can be expected to be globally unique or if it is locally administered. The IEEE Registration Authority developed a system of assignment for the first 3 octets of the MAC address. An equipment provider can purchase an allotment of MAC addresses so that its devices can be programmed with unique hardware MAC addresses.

3.1.2. MAC Addresses and Wi-Fi

The networks using MAC addresses were originally wired, of course, but the concept of using a MAC address to uniquely identify devices was brought into the development of wireless networking as well. Within the protocols of 802.11, MAC addresses are used extensively. For example, MAC addresses can identify the source and destination devices for each transmission in the header fields of that transmission.

As people began using more Wi-Fi enabled mobile devices, particularly smartphones, eventually industry began to equate the presence of a user with the presence of their mobile device. A personal device's MAC address became PCI and PII. Features assuming this linkage allowed simplifications in the user experience. For example, a user might log in once to an in-house system using their personal device, then the system would recognize that user by their mobile device MAC address when they returned, and not require the user to log in again. Over time, the use of MAC addresses to provide permanent identifiers was converted into a way for outside observers to track the movements of mobile device users, as well as to analyze their traffic. A user's path through a store could be mapped and advertisements adjusted to target their perceived interests. An employee's location in an office or warehouse could be tracked minute by minute. Privacy advocates began writing about this exposure and the industry began experimenting with options to increase people's privacy.

802.11 analyzed this problem in 802.11aq and added recommendations to the 802.11 standard. The potential use of randomized MAC addresses was discussed in that amendment along with other enhancements intended to increase the difficulty of tracking a particular mobile device. Packet sequence numbers and scrambler seeds can also offer an observer the ability to track a mobile device. Changing those values when the MAC address changes can make it more challenging for a remote observer to associate one packet with an old parameter set to another using a new random set of parameters.

The Wi-Fi mobile device industry began to implement MAC randomization to address consumer privacy concerns. The difficulty with MAC randomization was that if a mobile device presented its access point with a new MAC address after the authentication and association process was complete, the association with the access point would drop and the mobile device would need to go through the association process again. This re-association tended to interrupt traffic flows and could cause a user to be asked to reauthenticate with the system behind the access point since, to the access point, an entirely new mobile

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

device had suddenly appeared. Resulting from this observation, most devices utilizing MAC randomization did not change their MAC addresses while associated to an access point.

Because there was not a standardized overall solution, different vendors implemented MAC address randomization in different ways. One common implementation choice was to allow the mobile device to store a record of which MAC address it had last used with an access point and return to using that MAC address when the mobile device determined that it had returned to that access point. While this feature did reduce somewhat the ability of the third-party tracker to connect the mobile device's user across multiple unrelated access points, it did not help with a user who returned to the same coffee shop, for example. Another common implementation choice was to direct the mobile device to change its MAC address and associated parameters on a regular cadence of hours or days. A user would experience the need to log in to systems more often, but from one day to the next, a third party might have difficulty linking the user's activities.

3.2. Additional Corelatable Information with Wi-Fi

In addition to the MAC address, Wi-Fi devices can communicate a lot of information when they are searching for a new connection and while they are connected to an access point. In a Probe Request or an Association Request, a mobile device can include configuration information aside from just its MAC address. The purpose of the extra information was to allow the access point to potentially provide additional features to the mobile device, or more efficiently direct it to a radio band that might best suit its capabilities. But these assortments of device parameters and user configuration choices effectively have given an observer the ability to fingerprint a mobile device. That device fingerprint can provide at least PCI and, in some cases, PII about the user of a mobile device.

For example, one parameter a mobile device can include is the SSID of its preferred access point. Setting this parameter to a user's home access point's SSID allows an observer to know that that mobile device is at least connected to that user, if not a mobile device carried by that user as a personal device.

3.3. Current 802.11 Task Group Activities

802.11 has returned to the topic of MAC privacy enhancements to consider improvements in this area. A technical interest group or TIG was formed in May 2019 to consider whether the randomized and changing MAC addresses appearing in the field presented an issue that 802.11 should investigate in more detail. The TIG became a study group in 2020 that developed statements of work for two new task groups. 802.11 Task Group bh, TGbh, was formed to develop an amendment with updates to 802.11 that can address some of the problems created by RCM without significantly damaging the improved privacy that it offers users. 802.11 Task Group bi, TGbi, was formed to consider improvements to user data privacy within 802.11 more broadly in a separate amendment to the standard.

Within the IEEE standards development structure, a standard is managed by a Working Group. The 802.11 Working Group works on the standard for Wi-Fi. New amendments to the standard are developed in task groups at the behest of the working group. By inaugurating two new task groups, the 802.11 Working Group recognized that additional standards work in this area would have benefits for the standard and its user community.

3.3.1. TGbh - Operation with Randomized and Changing MAC Addresses

The identified scope for TGbh included preserving "the ability to provide customer support, conduct network diagnostics and troubleshooting, and detect mobile device arrival in a trusted environment." All

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

of these features had been degraded or rendered unreliable by the implementation of randomized MAC addresses on mobile devices.

When a user has been experiencing difficulties with a mobile device that has RCM active, a customer support agent might not be able to link any records of association difficulties or other problems to the problematic mobile device since the AP records might reflect an assortment of MAC addresses. Similarly, many networks track mobile devices by their MAC addresses. As RCM was deployed, often without a user even realizing it, the number of MAC addresses appeared to multiply exponentially. For access points that limited the number of allowed MAC addresses, they began to experience customer complaints when the mobile devices using RCM were no longer allowed to connect. Finally, complaints also came in from users who expected their mobile devices to be recognized by a trusted home or office environment. Instead of the seamless connection experience they had experienced previously, the users were required to log in every time they returned to associate to a well-known system.

TGbh has been developing mechanisms compatible with present networking methodologies that will help improve the performance of networks with mobile devices using RCM. The task group's challenge is how to enable a mobile device to share information securely with an access point that will allow the access point to recognize, to a lesser or greater extent, the mobile device without exposing the user to reduced privacy. An important part of the discussion has dealt with the tradeoffs facing a user. A straightforward decision must be made by the user whether to allow the network represented by the access point knowledge of the mobile device's identity or at least whether it has associated with this access point in the past. Emphasis has been placed on efforts to retain as much of the RCM privacy enhancements as possible to reduce the user's exposure to third party trackers while allowing a mobile device to expose additional information to the access point for improved user experiences. The information may be as simple as providing information from a mobile device to allow an access point to connect that newly associated mobile device with a mobile device that had a different MAC address when it associated with the access point in the past.

TGbh has also considered whether an access point should be able to recognize a returning mobile device during the association process, or if it is sufficient to recognize the device after it has associated. The ability to recognize a returning mobile device during the association process enables additional features, such as the ability to steer a mobile device to a favored band for association. On the other hand, for an access point to recognize the mobile device before association, some parameters must be passed before encryption is in place. The pre-association feature thus must be carefully designed to ensure that it does not increase the privacy risk.

For all of the possible TGbh features, similar to the current RCM features, Wi-Fi device manufacturers and software developers will have to decide the level of involvement of the end user. Will features default to enabled or disabled, for example.

The current timeline for TGbh predicts that its amendment may issue through IEEE SA in September of 2024.

3.3.2. TGbi – Enhanced Service with Data Privacy Protection

When 802.11 decided to investigate improvements in user privacy, TGbi was launched in parallel to TGbh with the scope of specifying "new mechanisms that address and improve user privacy." The task group was started in view of the need for standardized mechanisms to improve protection for users of Wi-Fi enabled mobile devices against user tracking and user profiling.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

TGbi began by considering different use cases and issues that related to user privacy. The group identified issues such as the ability of third party observers to track a user while the user's mobile device is associated, since the device cannot change its MAC address while associated without causing a full reassociation to be required. The task group also considered proposals for reducing information exposure in other actions that a mobile device might take, such as a location (PASN) session or a channel sounding. The suggestions of earlier amendments, such as TGaq, were also considered and expanded upon. For example, a mobile device seeking to find a new AP might send out a Probe Request that currently might include a lot of information that could be used to fingerprint the mobile device in later Probe Requests. TGbi is considering how to reduce messages such as a Probe Request to a bare minimum to also reduce the amount of information exposed to third parties.

TGbi has also been discussing possible improvements to AP privacy and security, particularly in connection to a mobile device used as a hot spot. A user might want to enable their mobile device as a hot spot to support other local mobile devices, but still want to minimize the information a third party could gather while the hot spot is active. Similar to minimizing content in a Probe Request, an AP using TGbi privacy mechanisms might restrict the amount of information it discloses to requesting mobile devices. The group is considering a new layer of encryption for some AP services that might further reduce the exposure of an AP to outside observers. Additionally, these features may also provide an additional layer of security to prevent AP spoofing, where a rogue AP is set up to capture traffic from mobile devices by using the SSID and other information from a trusted AP.

The current timeline for TGbi predicts that its amendment may issue through IEEE SA in March of 2026.

3.3.3. Overlap between TGbh and TGbi

The two task groups overlap in that they are both considering privacy related topics, but are distinct in that they have different focuses. TGbh is tightly focused on enabling mechanisms that allow a mobile device to be recognized by an AP when it returns to that AP, ideally with the continued use of randomized MAC addresses to maintain the user's privacy to third party observation. The mechanisms are all designed to allow the mobile device to implement them or not as instructed by a user or user provisioning. TGbi is more broadly considering methods of expanding the anonymity of a mobile device as it interacts with access points both before and after association.

Once both task groups have completed their work, a Wi-Fi device might implement mechanisms from both eventual amendments. It might use a mechanism from TGbh to allow an access point to recognize it when it returns to reassociate, and it might use a mechanism from TGbi to continue changing its over the air MAC address while associated to that same access point.

4. Wi-Fi Privacy and Security Use Cases

This section discusses particular use cases that provide examples of Wi-Fi Privacy and Security developments.

4.1. Mobile device not recognized due to RCM

The use of RCM by mobile devices has resulted a number of unintended side effects that in many cases resulted from a mobile device not being recognized by an AP or network of APs because the new MAC address is not familiar to the AP(s). These use cases illustrate a couple of key scenarios.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.1.1. Parental controls

Some parental control instantiations use mobile device MAC addresses to determine which rules to apply to a newly associated device. If the mobile device has RCM enabled, then it will probably not be recognized after its external MAC address changes. In some cases that may mean that the mobile device is only allowed to reach a walled garden of targets, in other versions, it might mean that the mobile device has no rules attached to its browsing. To avoid this issue before TGbh finishes its work, the user may be instructed to disable RCM on the mobile device. Alternatively, the user might be instructed to configure the mobile device to always use the same MAC address when associating with the local AP.

4.1.2. Help Desk interactions

In a commercial setting, a user may have a help desk that they can access when having connection problems with their mobile devices. The records the help desk can access may show several attempts of mobile devices to associate with the network, but with RCM enabled on a mobile device, each association attempt may use a different MAC address. The help desk technician can't tell which attempts are legitimate and which ones are suspect. To alleviate the issue currently, again the user must be instructed to disable RCM entirely, or to restrict it to a single MAC address for this AP.

It is important to note that the tactics noted above, disabling RCM or requiring reuse of MAC addresses for remembered APs, seem straightforward for technical professionals who are familiar with diving through technical menus and who already understand at least the basics of Wi-Fi operation. These directions can range from intimidating to extremely challenging for people who are non-technical or who are unfamiliar with the ins and outs of mobile device low level configuration. Because of this, most people tend to leave their mobile device settings on the device defaults. RCM spread most rapidly in the marketplace when mobile device vendors began to default their devices to enable RCM.

4.1.3. Changes planned in TGbh

The changes being discussed in TGbh are intended to allow a mobile device to identify itself to an AP with which it has previously associated to avoid the difficulties experienced currently.

In the parental controls use case, the mobile device might present an identifier to the AP after association. The AP could then compare that identifier to a previous record and determine the correct level of Internet access for that mobile device. The exchange of identifiers would be protected by the standard link level encryption of Wi-Fi, so a third part observer would not be able to glean the identifier through observation.

In the Help desk use case, the mobile device might use a random MAC address when it attempts association that has already be pre-shared with the AP in an earlier association. The pre-shared random MAC address would have been communicated in the previous association after a secure encrypted link was available to protect it from third party observation. Because the messaging request for association includes the declared MAC address of the requesting mobile device, the AP has the opportunity to identify the mobile device even before it has associated, thus allowing troubleshooting to occur even when or if the association fails.

4.1.4. Changes planned in TGbi

The changes under consideration for TGbi may present challenges for technical support in that the traditional fall back of capturing all the traffic on the air for later analysis may not be enough. As mentioned above, one of the TGbi's features is MAC address change while mobile devices are associated. For a complete analysis to be successful, a record of the MAC address transitions of the mobile devices

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

involved would be needed for the captured data to be useful. Alternatively, the AP and the mobile devices would need to be instructed to stop changing their MAC addresses for a time while the test is in progress.

4.2. User tracking by third party observers

The current behavior of Wi-Fi devices with or without RCM can expose user PII, such as location and activities. 802.11 currently does not allow a mobile device to change its MAC address over the air without a full reassociation. As mentioned earlier, current RCM implementations often use the same MAC address when returning to a known AP or network. As a result, with or without RCM the movements of a user though their mobile devices can be tracked while they are associated with an AP.

4.2.1. Smart Home

A smart home may have a mixture of fixed location and mobile Wi-Fi devices, where the fixed location Wi-Fi devices may be considered IoT devices. Generally, very few IoT devices use RCM features, though most do support encryption after association. Because of these fixed MAC addresses, even though an observer may not be able to decrypt the device transmissions, an observer may be able to determine that a particular MAC address is associated with lights or a door or a camera by simple observation. Once that pattern is known, an observer may be able to determine through remote monitoring of Wi-Fi frequencies when a person enters and leaves their home, for example. Similarly, if the person also carries a mobile device (a reasonable expectation), the observer may be able to associate the appearance of a particular MAC address with the presence of a particular person in the home.

4.2.2. Changes planned in TGbi

802.11 TGbi is considering new features that allow a device to change its over the air MAC address while associated with an AP. An external observer would record a set of MAC addresses that are changing and may not be able to determine without direct observation whether the changes represent actual new mobile devices or the existing mobile devices changing their MAC addresses. For a small set of devices, that problem might not be too complicated, but as the number of Wi-Fi enabled devices increases within the home and the number of channels available for those devices also increases, the problem of keeping track of how many devices are present rapidly becomes difficult to solve without statistical packet analysis or other more sophisticated efforts.

4.3. Open SSID and DMCA/CALEA

Open SSIDs have been used in the hospitality industry and other industries in conjunction with authentication services through Portals. The use of RCM can present issues if the user expects the portal behind the open SSID to recognize their mobile device and allow its services to be provided without incident. The use of RCM can also allow individuals who want to preserve their privacy for less savory means to hide their connection to particular locations or events.

4.3.1. DMCA with Open SSIDs

The DMCA was put in place to protect copyright owners from unlicensed use of their copyrighted content. The DMCA included specific provisions that copyright owners can use to enforce control over their content if it is used in an improper or unlicensed way.

Typical DMCA violations are captured using a public IP, date/time stamp and TCP/UDP port by the copyright owners. Based on the ownership information of an IP address, a Take Down Notice (TDN) is

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

generated and sent to the Service Provider associated with that IP address. The Service Provider maps this information back to a subscriber that was allocated the IP/Port combination. The mapping process uses the MAC address of the mobile device to ascertain the subscriber network information. Historical longevity and constancy of the MAC address allows this process to establish the Network identity using the MAC address as a correlation key. This relies on Usage Accounting information being provided by wireless gateways to analytic engines. Gateway Accounting information is not exposed to the AP and non-AP STA involved in RF communications.

Current MAC randomization methods are a compromise that may not easily comply with the need to support legal requirements under DMCA legislation. If a mobile device has RCM enabled without the feature of returning to the same MAC address at a given AP, it can be challenging to determine the actual device and to find the device's user to fulfill DMCA statutory requirements.

There are discussions of alternative identifiers that can be used to fulfill such requirements, which would allow MAC addresses to be no longer fixed – partially or fully. For example, RFC 4372 - Chargeable User Identity describes a solution capable of providing such capabilities, without the need for MAC address support.

Secure SSIDs can also leverage the same approach, or use other available identifiers which are PII and CPNI-friendly.

If TGbh features are enabled, then the access point may be able to associate repeat visits by the same device even if the over the air MAC address presented by the mobile device is changing between the visits. If the mobile device claims a different IP address on each visit, then the local AP or ESS network may be the only entity that can associate the different visits with a single device and user.

When TGbi privacy enhancements are rolled out, the privacy enhancements should only affect the ability of third party observers to track mobile devices. If the access point retains records of the devices associated, then the DMCA statutory compliance should be unaffected. If any system do rely upon third party observations of wireless traffic, they will be impeded by the changing of MAC addresses that will then occur even while the mobile device is associated.

4.3.2. CALEA/LI on Open SSIDs

The CALEA and Lawful Intercept legislation was put in place to provide support for law enforcement from telecommunications companies. CALEA/LI require support for wiretap capabilities on wireless networks. These capabilities typically use the MAC address of target mobile device in the Wi-Fi networks today.

On 3GPP networks, due to availability of other identifiers (IMSI – International Mobile Subscriber Identifier or its derivative identifiers), a MAC address is not necessary.

On a Wi-Fi network, the (3GPP) X1 interface (or equivalent) enables or disables the wiretap [typically] using a RADIUS Change-of-Authorization (CoA) request packet. The resulting control plane data is transferred to the LI Mediation Function over the (3GPP) X2 (or equivalent) interface. The tapped content is transferred to the LI Mediation Function via the (3GPP) X3 (or equivalent) interface.

If a device has RCM enabled, using the MAC address to identify a device and access its communications is no longer sufficient.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

As noted in the previous section, alternative identifiers can be used to fulfill such requirements, which would allow MAC addresses to be no longer fixed – partially or fully.

Secure SSIDs can also leverage the same approach or use other available identifiers which are PII and CPNI-friendly.

Again, if TGbh features are enabled on the AP and the mobile device, then the visibility into the activities of a specific mobile device is improved. Even if it uses a different MAC address, the TGbh feature can allow the AP to associate that device back to information from previous associations.

When TGbi features are active, they should not interfere with CALEA/LI as long as those systems are not dependent upon the over the air MAC address. The AP and larger network will still be able to identify the device and capture its communications.

5. Conclusion

This paper provides a glimpse into privacy related activities in IEEE 802.11. The updates under consideration may affect operator companies over the next 2-5 years as the features roll out. The experiences of the past suggest that the rollout of new Wi-Fi features can be slow, but RCM has been an exception to that rule. It was not even standardized formally, but spread quickly through the Wi-Fi world driven by the consumer interest in improved privacy. As TGbh and TGbi work through the standardization process, operators should monitor their development and consider how to prepare their networks for these new features.

The work of the IEEE 802.11 task groups will continue for at least two years, so interested parties have a chance to reach out to the group if they have additional ideas to contribute. If you are interested in becoming involved in the standards development process, please let the authors of this paper know and we can provide additional information.

Note that all comments in this paper are the professional opinions of the authors based on experiences and research. The authors do not represent 3GPP, IEEE, 802, 802.11 or any of the 802.11 task groups in any capacity.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AP	access point
CALEA	Communications Assistance for Law Enforcement Act
CPNI	customer private network information
DMCA	Digital Millennium Copyright Act
FEC	forward error correction
HD	high definition
IEEE	Institute of Electrical and Electronics Engineers
IOT	Internet of things
LI	legal intercept
MAC	medium access control
NE	Network Element
PII	personally identifiable information
RCM	random and changing MAC addresses
SCTE	Society of Cable Telecommunications Engineers
TIG	Technical Interest Group

Bibliography & References

Document Number	URL
Digital Millennium Copyright Act (DMCA)	The Digital Millennium Copyright Act
CALEA	Communications Assistance for Law Enforcement Act (CALEA)
Lawful Intercept (LI) – 3GPP	ETSI TS 103 221-1 V1.14.1 (2023-03) ETSI TS 103 221-2 V1.6.1 (2022-03) ETSI TS 101 671 V3.15.1 (2018-06)
WBA IMSI Privacy	WBA IMSI Privacy Standard v1.1
Android MAC Randomization	Android MAC Randomization Behavior
Apple MAC Randomization (iOS/iPadOS/watchOS)	Use private Wi-Fi addresses on iPhone, iPad, iPod touch, and Apple Watch
Customer Proprietary Network Information	Privacy/Data Security/Cybersecurity: Customer Proprietary Network Information
Personal Identifiable Information	Guidance on the Protection of Personal Identifiable Information
Task Group bh PAR	https://www.ieee802.org/11/PARs/P802.11bh.pdf
Task Group bi PAR	https://www.ieee802.org/11/PARs/P802.11bi.pdf
3GPP ITS Security	ETSI TS 102 941 - Intelligent Transport Systems (ITS); Security; Trust and Privacy Management; Release 2
3GPP Security Architecture	ETSI TS (1)33.501 – Security architecture and procedures for 5G System R17 (2022-05)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Wireline And Wireless Latency Improvements With AQM For Superior Quality Of Experience

A Technical Paper prepared for SCTE by

Gateek Fating
Manager, Wireless Product Connectivity
Charter Communications
6360 S Fiddler's Green Circle, CO - 80016
(818) 641-2827
Gateek.Fating@charter.com

© 2023, SCTE®, CableLabs® and, NCTA. All rights reserved.

Table of Contents

Title **Page Number**

1. Introduction 4
2. Background 4
3. Shifting the Focus to "Working Latency" 5
   3.1. Types of Latency (RTT) 5
4. Latency Improvements by Active Queue Management (AQM) 8
   4.1. Overview of Different Queue Management algorithms 9
      4.1.1. Passive Queue Management 9
      4.1.2. Active Queue Management 9
5. Cloud Gaming Latency Characterization Pre-AQM and Post-AQM 10
   5.1. Network KPIs For AQM Characterization 10
      5.1.1. Monitoring the DOCSIS link continuously 10
      5.1.2. Monitoring the Cloud Gaming QoE 11
   5.2. Pre-Requisites to AQM Enablement Verification on the CMTS 11
   5.3. Testing Methodology and Use Cases 12
   5.4. Network Topology for Private Gaming Characterization 12
   5.5. Results 14
      5.5.1. Network Monitoring Stats Pre-AQM and Post-AQM 14
      5.5.2. Gaming Characterization results Pre-AQM and Post-AQM 18
6. End-to-End Latency Management 24
   6.1. Latency Management Strategies on WiFi 24
      6.1.1. Current Practice 24
      6.1.2. Future Scope 25
7. Conclusion 26

Abbreviations 27

Bibliography & References 28

List of Figures

Title **Page Number**

Figure 1 - General Queuing Mechanism 6
Figure 2 - Various links on network introducing latency at different stages 7
Figure 3 - Coaxial and Fiber Transmission mediums 7
Figure 4 - Propagation Delay 7
Figure 5 - Utilizing AQM for Buffer Size Management 8
Figure 6 - AQM Enabled on the CMTS 11
Figure 7 - AQM Disabled on the CMTS 11
Figure 8 - Testing Methodology to Evaluate the Cloud Gaming Quality of Experience 12
Figure 9 - Network Topology to assess the Cloud gaming latency due to downstream congestion 12
Figure 10 - Network Monitoring for KPI's During the Game 13
Figure 11 - Idle ICMP Latency both Pre-AQM and Post-AQM 14
Figure 12 - CDF (Cumulative distribution function) plot of Idle UDP latency 15
Figure 13 - UDP latency pre-AQM and post-AQM 16
Figure 14 - ICMP Latency Under Downstream Load 17

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Figure 15 - Average Latency Trend for Wired Gaming Client with incremental traffic 18
Figure 16 - Average Jitter Trend for Wired Gaming Client with incremental traffic 19
Figure 17 - Network Latency trend for OTA Gaming Client with incremental traffic (in Near Field RF condition) 20
Figure 18 - Average Jitter trend for OTA Gaming Client with increment traffic (in Near Field RF condition) 21
Figure 19 - Impact of User Latency with incremental throughput 22
Figure 20 - Cloud gaming throughput impact at higher congestion levels 23
Figure 21 - Different Data-type classifications in priority order, highest to lowest 25

List of Tables

Title **Page Number**

Table 1 - Comparison between WRED and PIE 9
Table 2 - Median Idle ICMP Latency Pre-AQM vs Post-AQM 14
Table 3 - UDP Latency performance Pre-AQM and Post-AQM 15
Table 4 - Improvement in Packet Loss (%) Post-AQM 17
Table 5 - User Latency Comparison on Gaming Client in Pre-AQM and Post-AQM 22
Table 6 - Cloud Gaming throughput characterization (Pre-AQM vs Post AQM) 23
Table 7 - Fundamentals to Support Latency Management on WiFi 25

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

1. Introduction

Internet Engineering task Force (IETF) L4S and Low Latency DOCSIS (LLD) specifications enable cable ISPs to offer low latency and low jitter services over current DOCSIS 3.1 deployments. However, successful deployment of those services depends on the effective management of latency and jitter factors from source to destination, including WiFi and access and core networks. Although many tests have been conducted to evaluate LLD with IETF non-queue building per hop behavior (NQB-PHB) and L4S traffic across the access network, the performance evaluation on real production networks is very limited. This paper will demonstrate the benefits of access layer improvement in the form of Active Queue Management (AQM) for latency-sensitive applications in queue-building scenarios. It will also assess the advantages of AQM in a production environment with a mix of queue-building and non-queue-building traffic types to assess the quantitative and qualitative gaming performance of wired and wireless private client connectivity in the presence of passive features, which increase the end customers' network throughput. The production environment will demonstrate and characterize the network gaming experience over air interface on a private client in the presence of other mobile clients with this enabled speed increase, out-of-home WiFi connectivity, etc. Not only will this paper demonstrate performance gain for latency-sensitive traffic when implementing AQM, but it will also indicate the lack of performance degradation for latency-insensitive applications that are running simultaneously.

To determine the effectiveness of AQM in improving the client experience in the production environment, network monitoring tools were used to observe baseline latency and latency under incremental load in a real-world test-house serviced by a Cable Modem Termination System (CMTS) without AQM versus a CMTS with AQM. Different qualitative and quantitative network metrics – like network latency (RTT), jitter, user input latency (application responsiveness/lag) based on frame-per-second (fps) degradation – were calculated for a cloud gaming client in the presence of various congestion scenarios.

AQM provides a superior Quality of Experience (QoE) for queue-building traffic when multiple applications simultaneously contend for airtime on a user's network. In situations where the overall network utilization is higher than the actual cable modem's provisioned speed or link rate, efficient buffering of these packets at the CMTS will avoid excessive packet drops during network congestion. In these airtime bottleneck scenarios, AQM serves to efficiently process the packets at the CMTS to provide a better quality of experience by proactively dropping just enough packets to avoid queue build up from data bursts.

2. Background

Household network traffic is increasing, and it is helpful to understand that connection speed is not the only important factor in performance. To preserve the customer experience, end-to-end latency optimization has become increasingly critical.

The delay in network communications is referred to as **network latency**; it indicates the amount of time it takes for a data packet to traverse across the network. High latency networks have a longer delay or lag, whereas low latency networks offer quick reaction times. When cloud-based applications are used for performing basic day-to-day operations within a household or business, the lag time/delay in the network response can cause deficiencies in the system. Higher latency applications degrade the overall user experience and so, although all network devices favor low latency, it is crucial for certain streaming applications like online/cloud gaming, AR/VR, online betting, real-time auctions and video-enabled remote operations.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

To gauge the overall network experience from the end user's perspective, customers and industry reports also use **jitter, packet-loss and throughput** as crucial factors in determining the network performance.

- **Jitter** is the difference in time between data transmission and reception from source to destination. It has a greater influence on customer experience since it is the difference between the minimum and maximum delay. For a better user experience, a consistent delay is favored over delay changes.
- **Packet loss** is the measurement of packets that fail to reach their destination. End users may experience loss of network connectivity or slow service. Inadequate signal strength, network congestion or excessive system noise causes packet loss during data transmission. Real-time remote processing operations, like endoscopy cameras and drones for search-and-rescue operations, suffer the most when undergoing packet-loss.
- **Throughput** is the average amount of packets that can be sent/received through a network from source to destination in a given amount of time. It represents the number of data packets that successfully reach their destination despite network interference/congestion and packet loss.

This paper demonstrates the end-to-end latency improvements due to AQM deployment on the CMTS leveraging the DOCSIS-PIE algorithm. The analysis is performed in a production test house to characterize the network KPIs of a cloud gaming client via wired and wireless interface when there is an incremental load/traffic congestion. We will also discuss some of the opportunities to improve the ability to deliver improved overall quality of service on access and the WiFi network.

3. Shifting the Focus to "Working Latency"

A new way to measure real-world latency — referred to as working latency — is to look at what happens to latency when the link is utilized heavily by other clients on the network. However, the link's capacity need *not* be fully leveraged to see potential dramatic increases in working latency, as this can happen under normal usage conditions as well.

Network congestion is *not* vividly perceived by the end user unless there is an unexpected performance issue while using a latency-sensitive application that consumes significant bandwidth. Application providers use Adaptive Bitrate Streaming (ABS) by adjusting the streaming quality to detect the end client's CPU capacity and available bandwidth. End users often blame it on the network capacity in these conditions, but it is most likely the increased network latency resulting in a poor quality of streaming experience due to various factors in packet delay from the client all the way to the core network.

3.1. Types of Latency (RTT)

Idle Latency

Idle latency is the time it takes data to get from its source to destination *without any network congestion*. Idle latency measurements are beneficial to measure packet transmissions over a network path based on distance alone, but it doesn't help understand the latency under load.

Working Latency

Working latency is the time it takes data to get from its source to destination *when the network is congested.* It is a real-world measure of responsiveness when the home network connection is actively used and burdened with congestion from different clients on the network.

Working latency first determines if the network is active and then measures the time it takes for packets to reach the destination. It is a quantitative characterization of the delay for latency sensitive traffic, like video

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

conferencing or cloud gaming, due to the congestion/induced interference from “greedy” applications, like video streaming and file downloads.

In a nutshell, working latency assesses how strongly other internet traffic can interfere with your video conference or gaming experience. Most computer networks are inactive for most of the time. When measuring latency on an idle network, it is clear that the best-case latency is going to be delivered. This is analogous to using Google Maps travel times outside of peak hours and presuming that it represents the traffic conditions at other times of the day, which does *not* determine the network’s ability to handle congestion or reliably process the buffer bloat issue.

Network Congestion and Buffer Bloat

When a network is overburdened with more data packets than it can handle, it is said to be congested. When too many communication and data requests are made at the same time across a network that does not have the network bandwidth to transmit them, data traffic backup develops.

While **network congestion** is normally transient, it can generate annoying network issues that hinder performance, such as high levels of jitter, packet loss and delay, as well as a drop in throughput. A crowded network might indicate a broader problem in your network. As a result, it's critical to have network performance monitoring technologies in place that can detect network congestion both within and outside of your network.

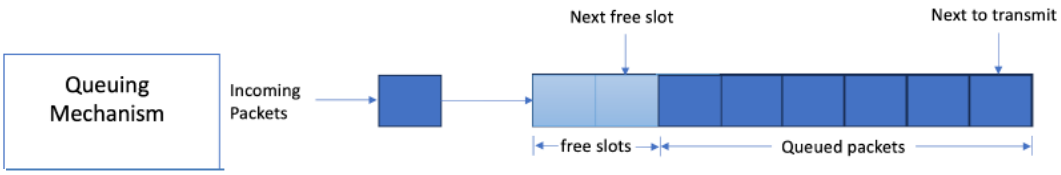


Figure 1 - General Queuing Mechanism

Buffer Bloat: Most modern network equipment, such as routers and switches, have several queues, each with a buffer to hold incoming packets. When high speed (classic TCP-based) applications put an excessive strain on the network, the queues begin to fill up and block real-time data. As a result, there is a significant delay, and perhaps fluctuating packet delays, also known as latency and jitter. This significantly degrades the experience of these interactive and latency-sensitive applications, if not totally breaking them ('buffer bloat'). Back in the days when memory constraint was not a consideration, buffer bloat was not a problem. Big in-transit buffers were designed to handle network congestion and that was not the most efficient way to tackle the buffer bloating issue. Increasing the buffer-size results in an increase in queuing delay and so Internet Service Providers (ISPs) have transitioned to an efficient way for queue management of incoming packets downstream to the CMTS.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

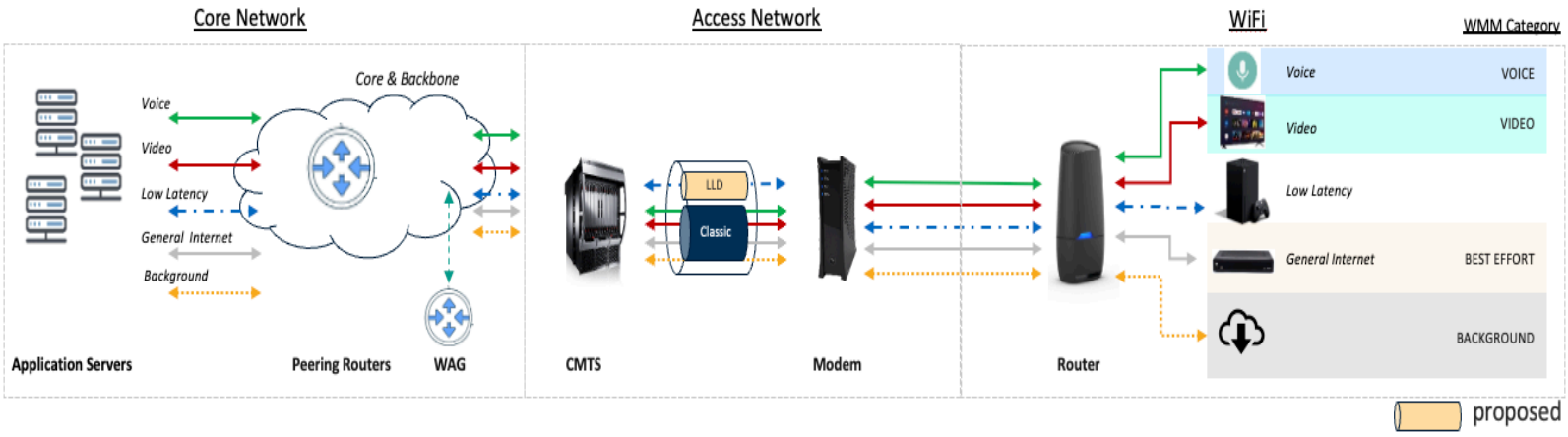


Figure 2 - Various links on network introducing latency at different stages

Figure 2 above shows the three segments within the network: **core network**, **access network** and the **in-home WiFi network**, where latency is introduced due to some of the major contributing factors like the transmission medium, propagation delay and queuing delay aspects in the network. As part of Low Latency, Low Loss and Scalable Throughput (L4S), a proposed LLD service flow would help isolate the low latency and high capacity traffic during peak network congestion.

Transmission Medium: As the packets move across the transmission medium, they introduce latency depending on the type of medium. Fiber and copper have different latency properties due to physics and so each time the network shifts from one media to another, a few milliseconds are added to the overall transmission time.



Figure 3 - Coaxial and Fiber Transmission mediums

Propagation Delay: All transmission components contribute to the propagation delay between the source and destination. All transmission components in the link affect the propagation delay of the packet. The amount of distance traveled in the medium results in propagation delay. Some types of Distributed Access Architecture (DAA) like Remote PHY (R-PHY), target to remove propagation delay by shifting/migrating the physical layer closer to the edge of the access network.



Figure 4 - Propagation Delay

Queuing Delay: This solely depends on the queuing mechanism at the cable modem as well as the CMTS. Packet congestion due to various traffic types that fill buffers within the service flow results in queuing delay. Section 4 of this paper explains queuing mechanism in detail.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Media Access Delay: In saturated network conditions, the probability of collisions and cross transmission increases resulting in high packet loss between two nodes. Asynchronization in data transmission on multiple frequency channels simultaneously, results in Media Access Control Delay.

4. Latency Improvements by Active Queue Management (AQM)

As more ISPs consider the idea of buffer bloat, interest in remedies to the problem grows. AQM is currently the most promising technology since it has the potential to produce significant network wide gains by focusing on a small number of bottleneck network components on the access network (e.g., cable modems and CMTS). The first key step in reducing latency and ensuring consistency from the CMTS to the cable modem can be achieved by implementing AQM and setting the right latency target.

Current AQM algorithms determine an estimated queueing delay of the CMTS queue and thereby calculate a packet drop probability. This drop probability is applied to incoming packets, followed by a recalculation of the drop probability. If a TCP packet is dropped, TCP congestion control takes affect and limits the throughput.

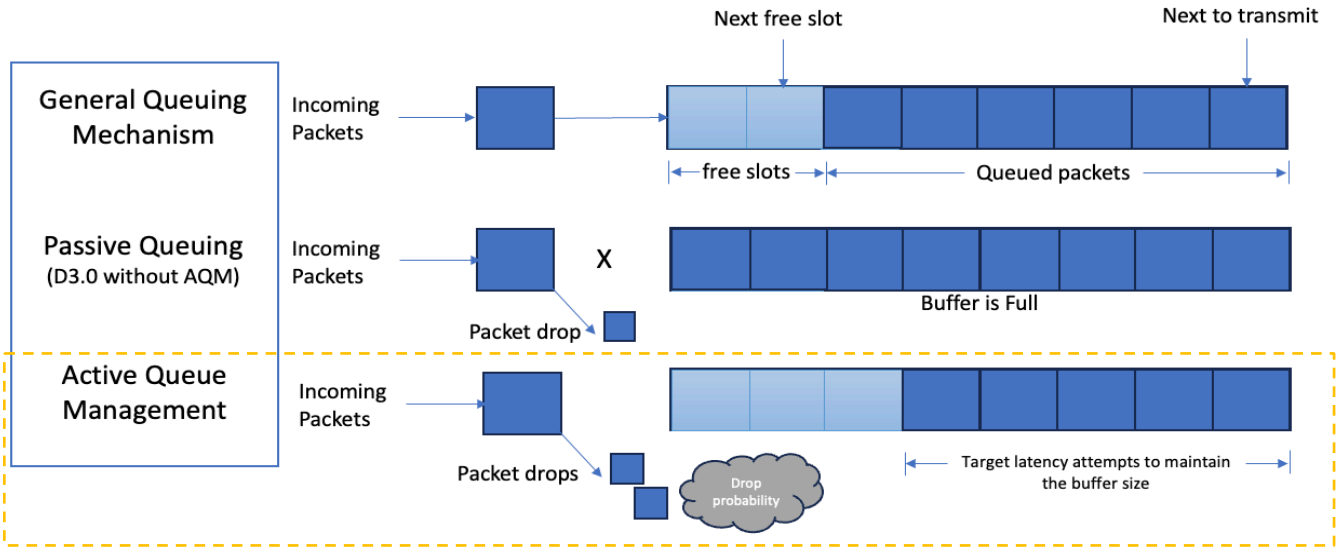


Figure 5 - Utilizing AQM for Buffer Size Management

Although an AQM technique that loses packets may be regarded unfavorable, it does reduce buffer bloat while having no major impact on any active applications. For example, a TCP session will slow down when a packet is lost, which reduces TCP traffic speed slightly, although this is typically not as essential as having high latencies and jitter for latency-sensitive real-time applications like video conferencing. When the technique employs marking instead of dropping, it is even better because it requests the sender to slow down without causing packet loss.

As part of this paper, implementation of AQM in a real-world environment at the test house is seen to radically improve the overall broadband user experience for an interactive application like cloud gaming in the presence of a congested environment. The need for AQM arises because of the presence of packet buffering in network elements and due to the mechanics of the TCP congestion avoidance algorithm.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

4.1. Overview of Different Queue Management algorithms

4.1.1. Passive Queue Management

Tail drop is a queue management algorithm that is an example of passive queue management. As per this algorithm, each packet is assigned the same priority and there is no distinction. In addition, the queue length and the basis of managing the queue is concerning, i.e., when the buffer capacity is reached, inbound packets are discarded/dropped indefinitely without any delineation between latency sensitive and latency tolerant applications. This is what the algorithm looks like:

```
while packets arrive
if (queue is not full)
then
    Enqueue the packet
else
    Drop the packet
End
```

4.1.2. Active Queue Management

In this type of queue management technique, the cable modem and CMTS will keep a close eye on the incoming packets, as well as the buffer size, to keep the queuing latency under check. As soon as CMTS detects that queue-building traffic has exceeded the target queuing latency, AQM will drop just enough packets randomly to maintain the target latency, allowing more appropriate buffer levels to be maintained as part of an efficient queue management.

Table 1 - Comparison between WRED and PIE

WRED algorithm (Weighted Random Early Detection)	DOCSIS – PIE algorithm (Proportional Integral controller Enhanced (PIE))
Initialization: Empty (queue) **while** packetsarrive **if** (queue is not full) **then** Enqueue the packet **else if** (Minimum WRED threshold has been exceeded AND queue is **not** full) Execute packetdrop(newpacket) Calculate new drop probability **else** Drop newpacket **End**	Initialization: Empty (queue) **while** packetsarrive **if** (queue is not full) **then** Enqueue the packet **else if** (AQM latency target has been exceeded AND queue is not full) Execute packetdrop(newpacket) Calculate new drop probability **else** Drop newpacket **End**

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Two well-known Active queue management algorithms are Weighted Random Early Detection (WRED) and DOCSIS Proportional Integral controller Enhanced (PIE). The major difference between WRED and PIE is that WRED has a concept of "drop probability" which is a function defined between *minimum threshold* and *maximum threshold*. As the queuing latency increases, the drop probability increases as well until the queuing delay reaches maximum threshold. Whereas DOCSIS-PIE defines threshold to maintain **target latency**, which in-turn sets the buffer queue size and drives the overall packet processing/queuing.

The two AQM algorithms above look similar, but "packet drop" as a function has a chance of dropping a packet, and the chance is dynamically calculated based on queue depth/buffer size. For example, if WRED, defined from 100 to 200ms, has a "packetdrop" function as per the above algorithm, it will have a 50% chance of dropping the "new packet" if the current queue depth is 150ms. In other words, we have the maximum drop rate set to 100% at 200ms, which means that 200ms is the full size of the buffer and anything beyond 200ms is tail dropped.

5. Cloud Gaming Latency Characterization Pre-AQM and Post-AQM

In this section, we will assess the network KPIs of a cloud gaming client, connected in a real-world environment without AQM and with AQM enabled on the CMTS. Different network metrics were measured as part of this analysis in a real-world environment at a test house to gauge the benefits/improvements for the latency sensitive applications under an idle environment and a congested environment with incremental load. This has also helped us understand the susceptibility of access network to congested traffic and the efficient response of the network to NOT degrade the queue building traffic due to implementation of AQM. Here are the network Key Performance Indicators (KPIs) calculated as part of this assessment.

5.1. Network KPIs For AQM Characterization

5.1.1. Monitoring the DOCSIS link continuously

This was performed using a SamKnows Whitebox tool hard-wired into the cable modem to monitor the pre-AQM and post-AQM measurements continuously. The below metrics were measured:

a) **Median Idle ICMP Latency:** This is the average Round Trip Time (RTT) Internet Control Message Protocol (ICMP) echo request of five 56 bytes packets sent from the Whitebox to the target node (on the access layer) every two hours.
b) **Median Idle UDP Latency:** This is the average RTT of a burst of User Datagram Protocol (UDP) packets transmitted to the target node every two hours.
c) **UDP Latency Under Downstream Load:** As part of this test, a 10-second downstream speed test was initiated and UDP datagrams were transmitted to the target server. Average RTT was measured for the packet delay from the Whitebox to the target server.
d) **Packet Loss (in %):** This was measured for both ICMP and UDP packets that were not received in response.

NOTE: All the above measurements were performed every two hours. Idle latency measurements were only recorded when there was no congestion detected on the network. If congestion was detected, that measurement was skipped and a retry was attempted in the next two hours.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.1.2. Monitoring the Cloud Gaming QoE

The below metrics were calculated to assess the cloud gaming QoE, both pre-AQM and post-AQM:

a) **Average ICMP Latency/RTT:** This was measured as the average delay for gaming packets to traverse through the network and back.
b) **Average RTT Jitter:** This was the variance in arrival of two packets back from different hops on the network.
c) **User Latency OR Input to Action Latency:** The delay between the end user stimulating an input/action in the game to shoot a bullet (e.g., left mouse click) and the user experiencing the corresponding action on screen. For example, a response on the screen of a muzzle flash to the input action of a bullet shot by the end user.
d) **Cloud Gaming Throughput:** Throughput on both the downlink and uplink generated by the gaming client playing a Destiny 2 game. This is around **40Mbps on the downlink** and around **200Kbps on the uplink**.

5.2. Pre-Requisites to AQM Enablement Verification on the CMTS

Pre-requisites: When validating AQM enablement on the CMTS, we ran the following command to receive the output, as per the two screenshots below:

```
<cmts-name>#show cable service-class 2 ver | i aqm
```

```
          #show cable service-class 2 ver | i aqm
 aqm-latency 32
 aqm-enable 1
 aqm-algorithm 1
 immediate-aqm-max-threshold 1000
 immediate-aqm-range-exponent 19
```

Figure 6 - AQM Enabled on the CMTS

```
#show cable service-class 2 ver | i aqm
 aqm-latency 10
 aqm-enable 0
 aqm-algorithm 0
 immediate-aqm-max-threshold 1000
 immediate-aqm-range-exponent 19
```

Figure 7 - AQM Disabled on the CMTS

For downstream traffic, one of our Converged Cable Access Platform (CCAP) providers used the DOCSIS-PIE AQM algorithm. However, it was disabled by default, thus when users overloaded their downstream service flows with downstream traffic, latency on the service flow rose resulting in buffer bloat. Additionally, for our validation at the test house, we temporarily set the DOCSIS-PIE AQM with a **latency target of 32ms** and **maximum buffer depth to 160ms**, after which the algorithm was designed to increase the drop probability to signal the sender to reduce their sending rate or bytes-in-flight, thereby reducing overall queuing delay on the access network at higher network congestion.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.3. Testing Methodology and Use Cases

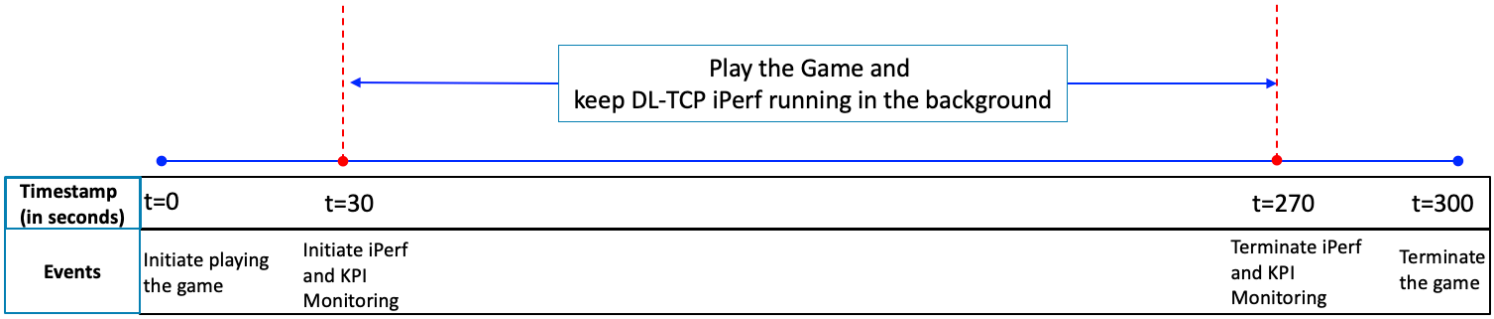


Figure 8 - Testing Methodology to Evaluate the Cloud Gaming Quality of Experience

The overall cloud gaming QoE was assessed by initiating the game and measuring the baseline latency. Following baseline measurement, incremental traffic from the other two mobile clients was induced to create contention for airtime for the latency sensitive cloud gaming traffic.

For each scenario, three iterations of four minutes (240 secs) each were performed to average the cloud gaming network performance without AQM and with AQM enabled on the CMTS. The gaming session was initiated 30 seconds before the actual measurement to ensure gaming traffic was prevalent on the link and healthy prior to introducing network congestion from the other mobile clients.

Gaming characterization was performed with the gaming PC connected via wired interface for some scenarios and wirelessly in near field RF conditions (without any lab constraints). To elaborate, when the gaming PC was connected wirelessly in near RF conditions, we maintained the RSSI signal levels between -33dBm to -36dBm and the interference between 2-3% during the entire test duration.

5.4. Network Topology for Private Gaming Characterization

The network topology below is the setup used for validation of the benefits of AQM enablement on the network. The results illustrate the improvements in section 5.5.

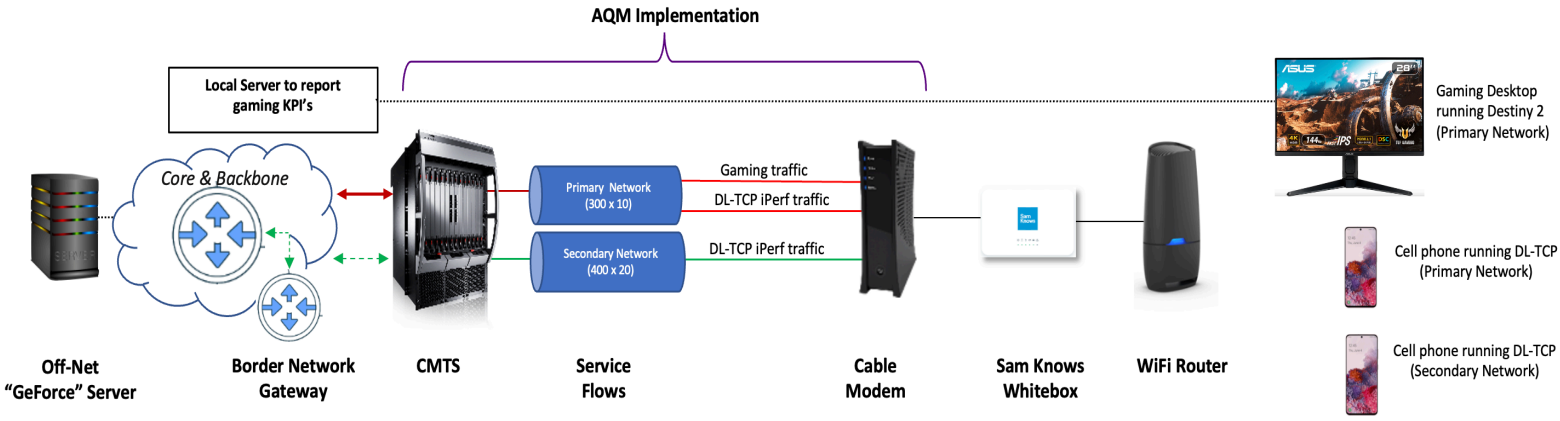


Figure 9 - Network Topology to assess the Cloud gaming latency due to downstream congestion

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

As per the above network topology in Figure 9, we used three clients in an environment that replicates a single-family residential household at our test house. Of these three clients, two clients (gaming PC and the cellphone 1) were always connected to the primary network provisioned at 300Mbps on the Downlink and 10 Mbps on the uplink, whereas the other phone (cellphone 2) was connected to the secondary network provisioned at 400 Mbps on the downlink and 20 Mbps on the uplink. Network KPI measurements were performed on the gaming PC to assess the impact of the cloud game with incremental throughput from two phones connected to the iPerf server.

The SamKnows Whitebox monitored the access layer for any changes to the link in case of incremental load on the private network. The tests were scheduled in the background every two hours and configured in a way to not affect the end client experience if they were already "busy." The gaming PC and the cell phone 1 were connected on the private network aka primary SSID, traversed through the "classic service flow" (i.e., the cable modem provisioned flow of 300 x 10) whereas the secondary network traffic injected by the cellphone 2 took the separate service flow (provisioned on the network for 400 x 20).

The cloud game used for this assessment was "Destiny 2" played on Nvidia GeForce Now. Additionally, the gaming KPIs were stored in our local server whereas all the SamKnows network KPIs were measured at the test target nodes located at major peering locations around the world. These peering locations are sometimes on-net or off-net, whereas the results recorded by the white boxes are stored at the data collection servers managed by SamKnows.



Figure 10 - Network Monitoring for KPI's During the Game

Latency monitoring for the gaming packets sent over the network with mean, max and last latency measurement for comparison during the game

Frame rate and fps fluctuation during the cloud game due to the airtime bottleneck at higher congestion

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.5. Results

5.5.1. Network Monitoring Stats Pre-AQM and Post-AQM

Median Idle ICMP Latency

For measuring the idle ICMP latency, we split this metric down to target nodes. The reason for this is SamKnows reaches the nearest peering location for an ICMP measurement, and we only filtered it down the target node on our network closest to the test location for a better understanding of the idle latency without any congestion. The below table indicates a 21% median idle ICMP latency improvement pre-AQM and post-AQM ICMP latency.

Table 2 - Median Idle ICMP Latency Pre-AQM vs Post-AQM

Pre-AQM (7/25 to 8/1)	Post-AQM (7/11 to 7/25)
24.85	19.65

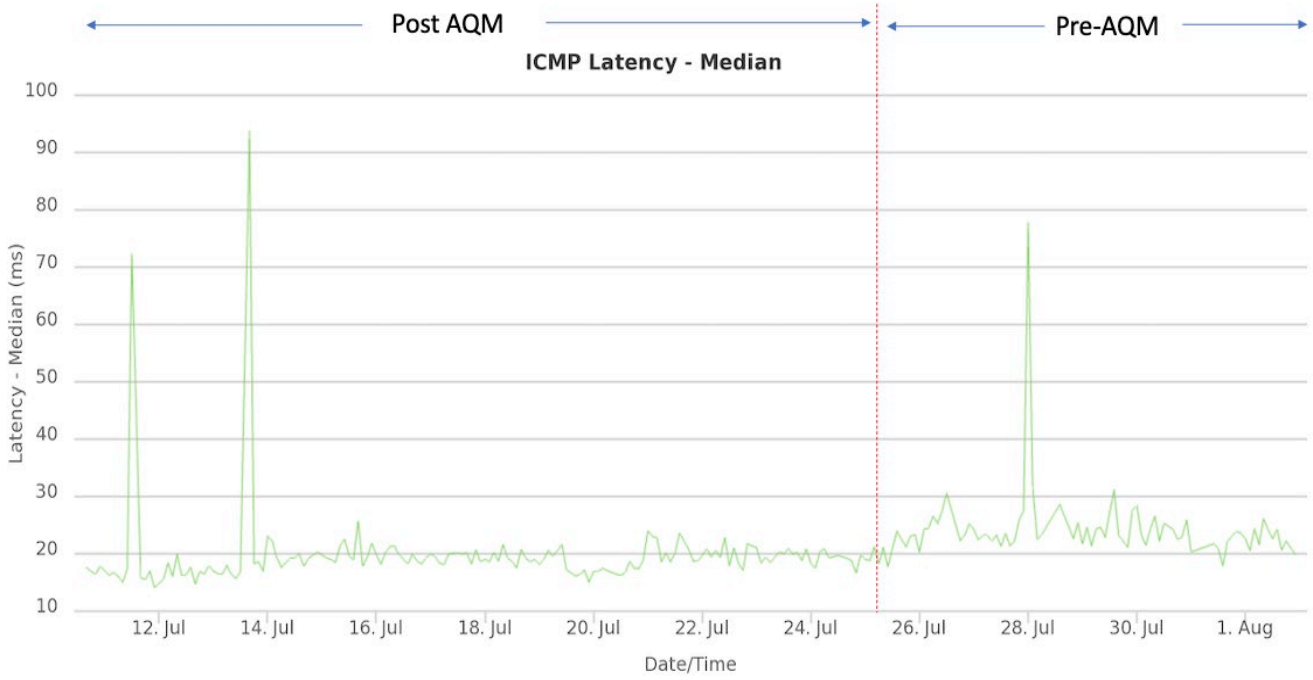


Figure 11 - Idle ICMP Latency both Pre-AQM and Post-AQM

Median Idle UDP Latency

The CDF plot of idle UDP latency in Figure 12 below indicates a clear improvement without much variance post-AQM deployment. Due to better processing ability post-AQM, idle UDP latency varies between 18.5ms to 19.5ms post-AQM deployment. Whereas UDP latency prior to AQM shows that there is a 95% probability for latency to stay between 19.7 to 25ms, but due to lack of queue management at the CMTS,

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

the latency numbers could also increase as high as 46ms without any load on the network. The standard deviation of UDP idle latency pre-AQM was found to be 1.96ms.

NOTE: Even though the measurement calls out Idle UDP latency, the measurement was performed by sending a burst of UDP datagrams from the whitebox to the target server.

On the other hand, post-AQM latency is seen to be remarkably improved. The idle UDP latency was confined from 18.5ms to 19.52ms (i.e., 1ms of standard variance) when the SamKnows Whitebox measured the latency over 14 days. The standard deviation of UDP idle latency post-AQM was found to be 0.18ms.

Table 3 - UDP Latency performance Pre-AQM and Post-AQM

	Range of Idle UDP Latency (in ms)	Standard Deviation (in ms)
Pre-AQM	19.7ms to 46ms	1.96ms
Post-AQM	18.48 to 19.5ms	0.18ms

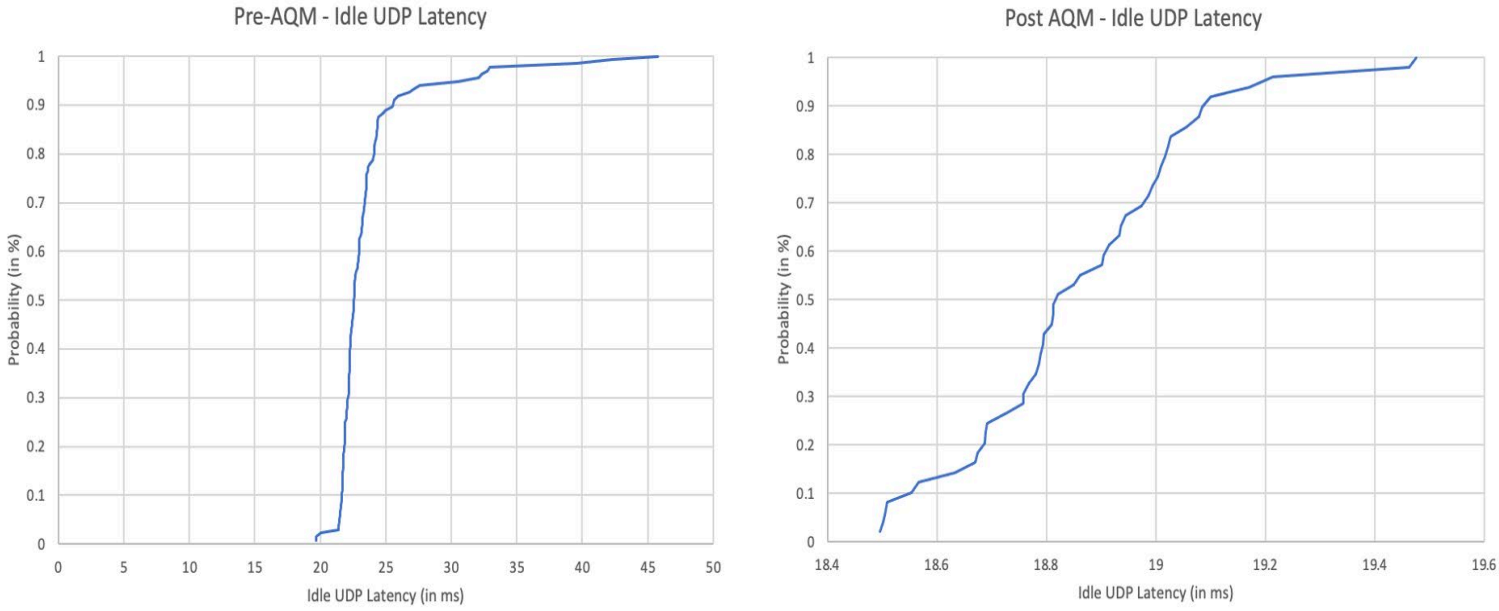


Figure 12 - CDF (Cumulative distribution function) plot of Idle UDP latency

In the below UDP latency graph, we see high jitter (variance in latency numbers in presence of queue-building traffic) prior to AQM enablement.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

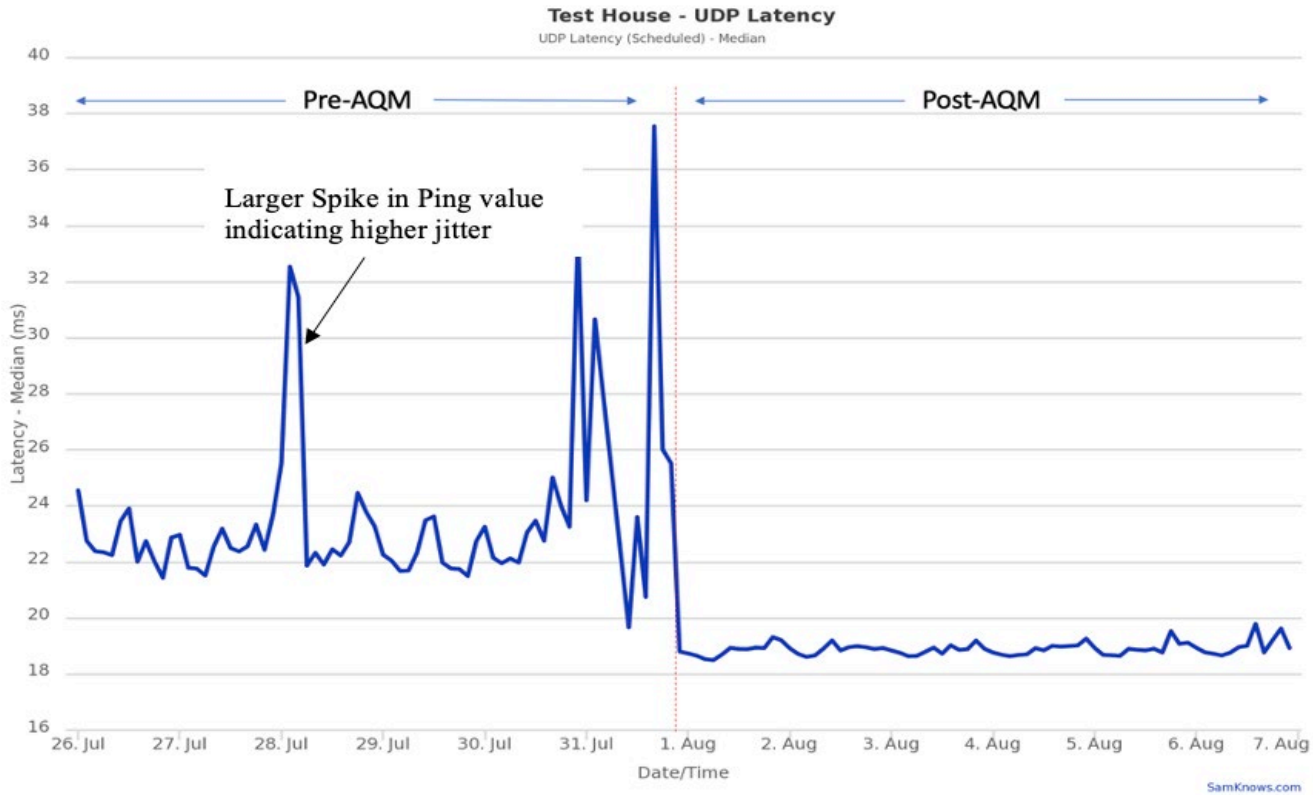


Figure 13 - UDP latency pre-AQM and post-AQM

Latency Under Downstream Load

Under a congested network, we see a clear improvement in latency by around 56% with AQM deployment on the network. Although, we see an improvement in idle conditions without much congestion. It is evident in Figure 14, below, which indicates that when TCP keeps the buffer full, AQM processes and queues the packets in the buffer as a way to slow down the TCP transmission by dropping a few packets to indicate congestion.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

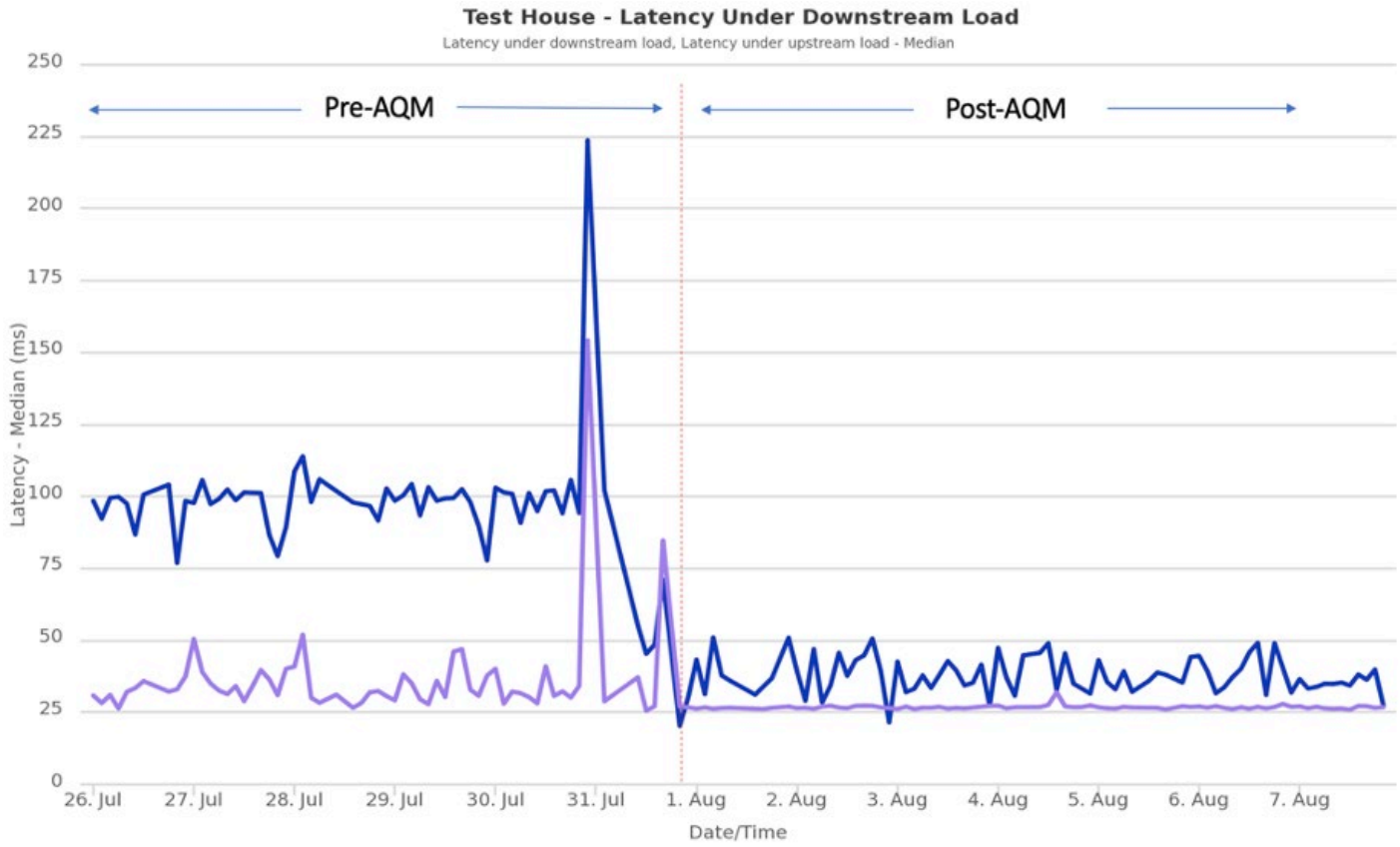


Figure 14 - ICMP Latency Under Downstream Load

ICMP Packet Loss (in %): This tracks the fraction of ICMP packets lost during the scheduled idle ICMP test every two hours to assess the health of the link.

Table 4 - Improvement in Packet Loss (%) Post-AQM

	Pre-AQM Aggregate *to 7/25 to 7/31*	Post AQM Aggregate *8/1 to 8/7*
Median	0.11%	0.02%
P95	0.36%	0.14%

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

5.5.2. Gaming Characterization results Pre-AQM and Post-AQM

Average Gaming Latency on Wired Interface (Pre-AQM vs Post-AQM)

With the gaming client hard-wired to the router and two mobile phones pushing incremental load on the network, after AQM enablement, gaming latency at higher network congestion is controlled due to better queue protection at the CMTS. Please note that gaming latency at higher network congestion is comparable with increased iPerf traffic from the two mobile clients.

Post-AQM Improvement (at higher congestion highlighted in the below graph) = 27%

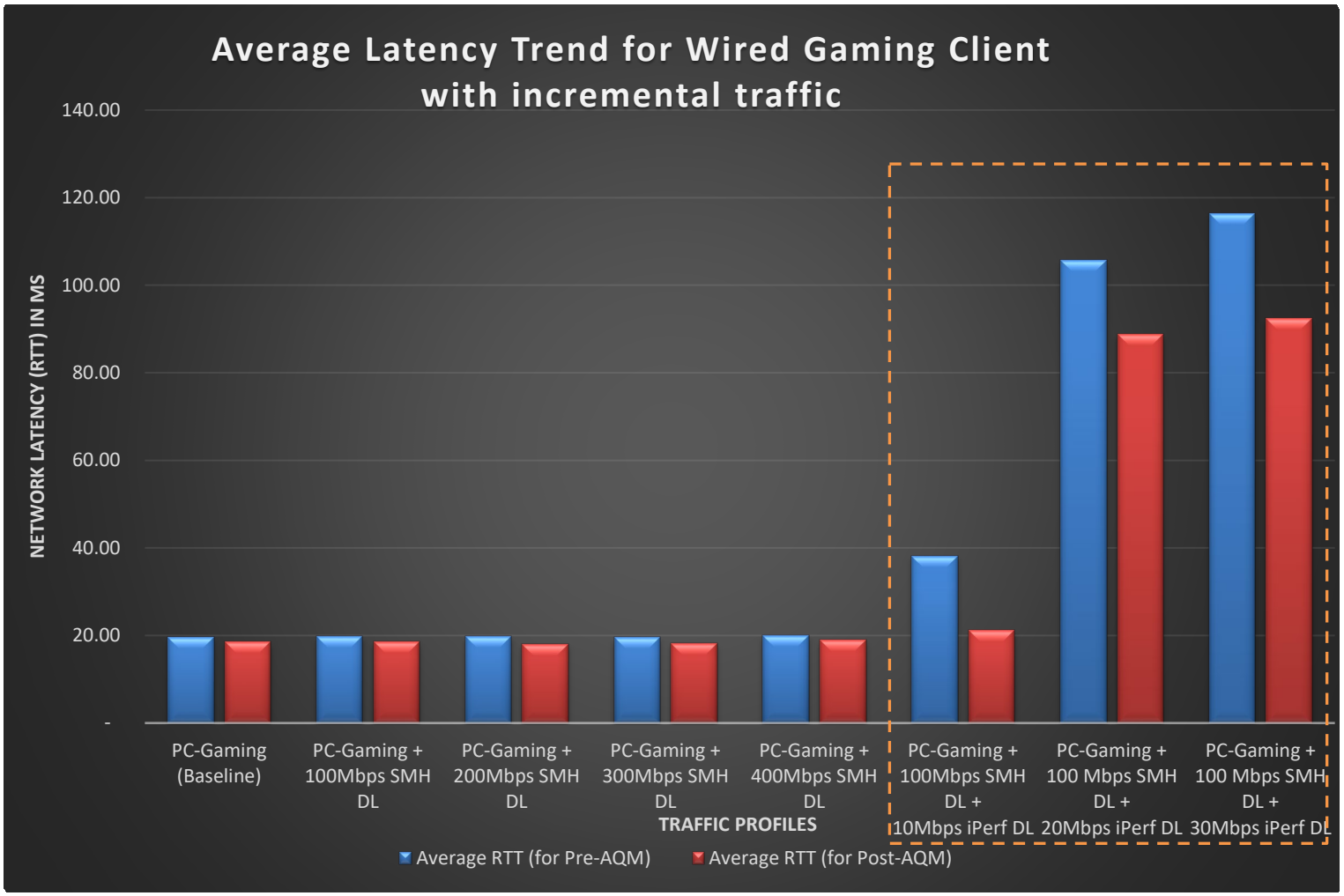


Figure 15 - Average Latency Trend for Wired Gaming Client with incremental traffic

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Average Gaming Jitter on Wired-Interface (Pre-AQM vs Post-AQM)

Average Jitter stats measured during the Destiny 2 gaming session at the test house indicated a significant improvement after AQM enablement. With higher congestion on the network, we saw no negative impacts like screen rendering on the gaming experience.

Post-AQM improvement = 39.3% (at higher congestion highlighted in the below chart)

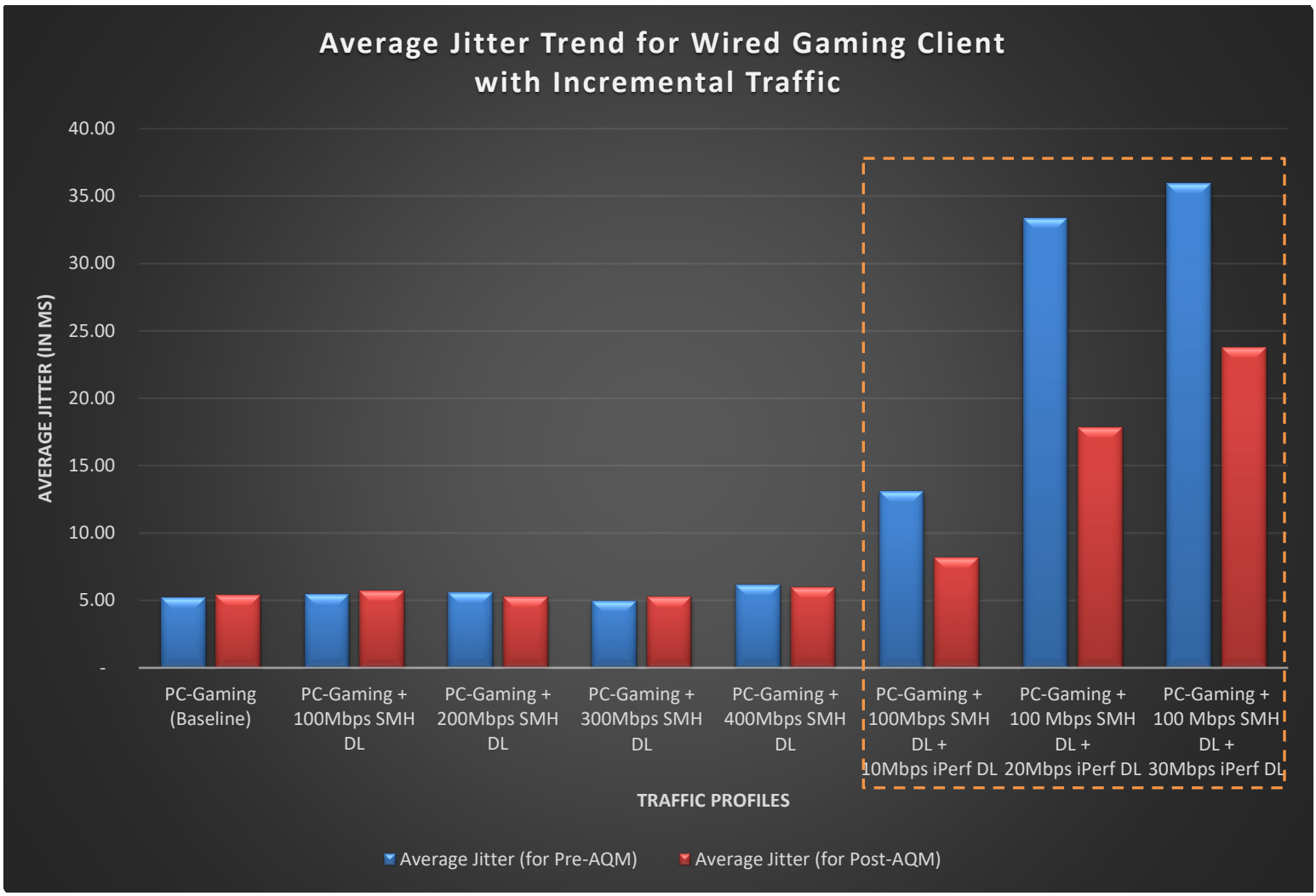


Figure 16 - Average Jitter Trend for Wired Gaming Client with incremental traffic

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Average Gaming Latency for Near Field OTA (Pre-AQM vs Post-AQM)

Post-AQM latency is slightly lower than pre-AQM for all the test scenarios where we introduced additional traffic for airtime contention. In this case, the overall gaming latency trend for pre-AQM and post-AQM do not experience a huge improvement after AQM enablement on the network.

Even in other scenarios without much interference on the network, we do see comparable results. Between pre-AQM and post-AQM when the gaming PC is connected wirelessly in near RF field from the router Post-AQM improvement = 12.8%

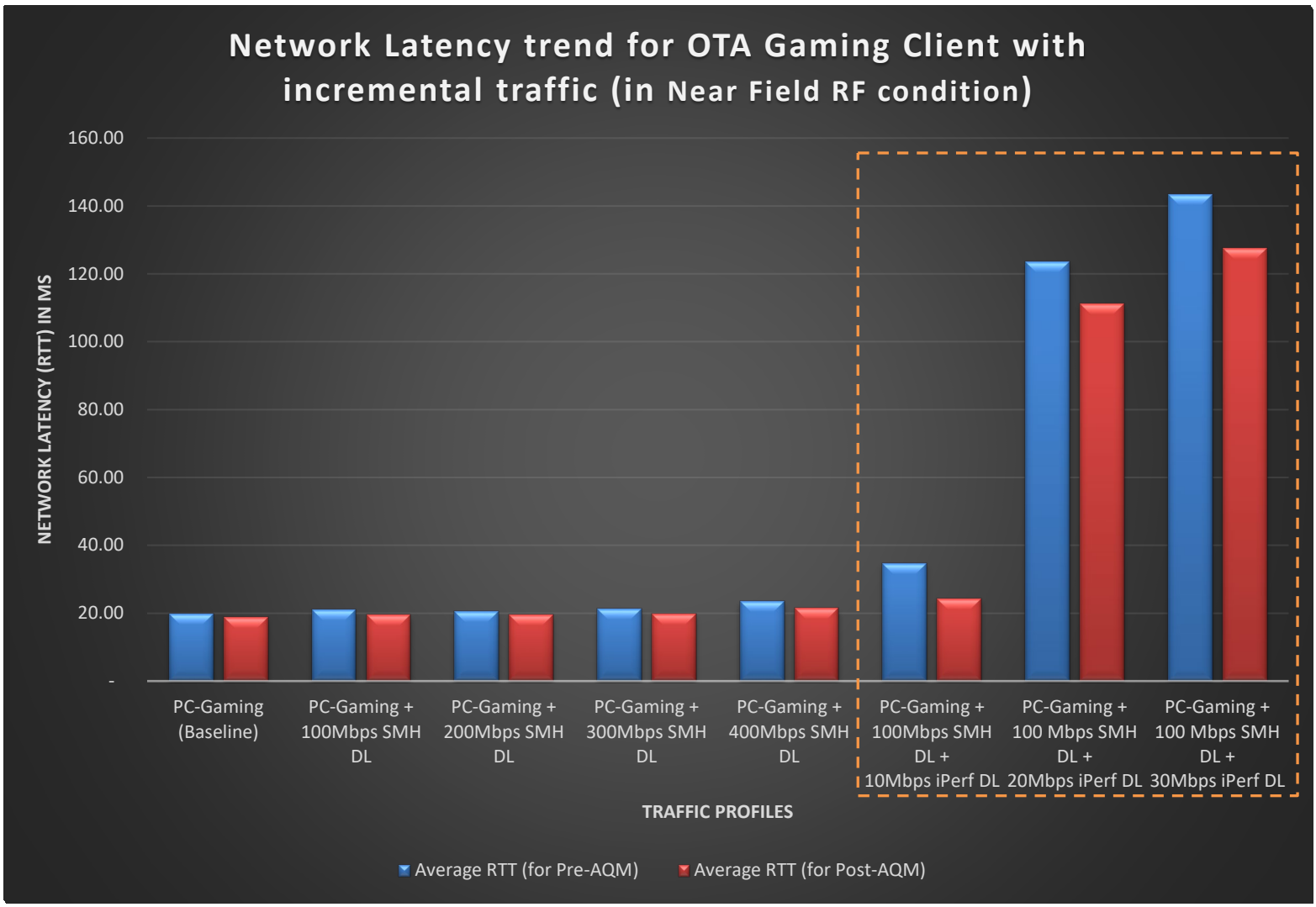


Figure 17 - Network Latency trend for OTA Gaming Client with incremental traffic (in Near Field RF condition)

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Average Gaming Jitter for Near Field OTA (Pre-AQM vs Post-AQM)

Jitter is very comparable both pre-AQM and post-AQM when the gaming client is connected wirelessly in near field RF environment.
Post-AQM improvement = 7.8 %

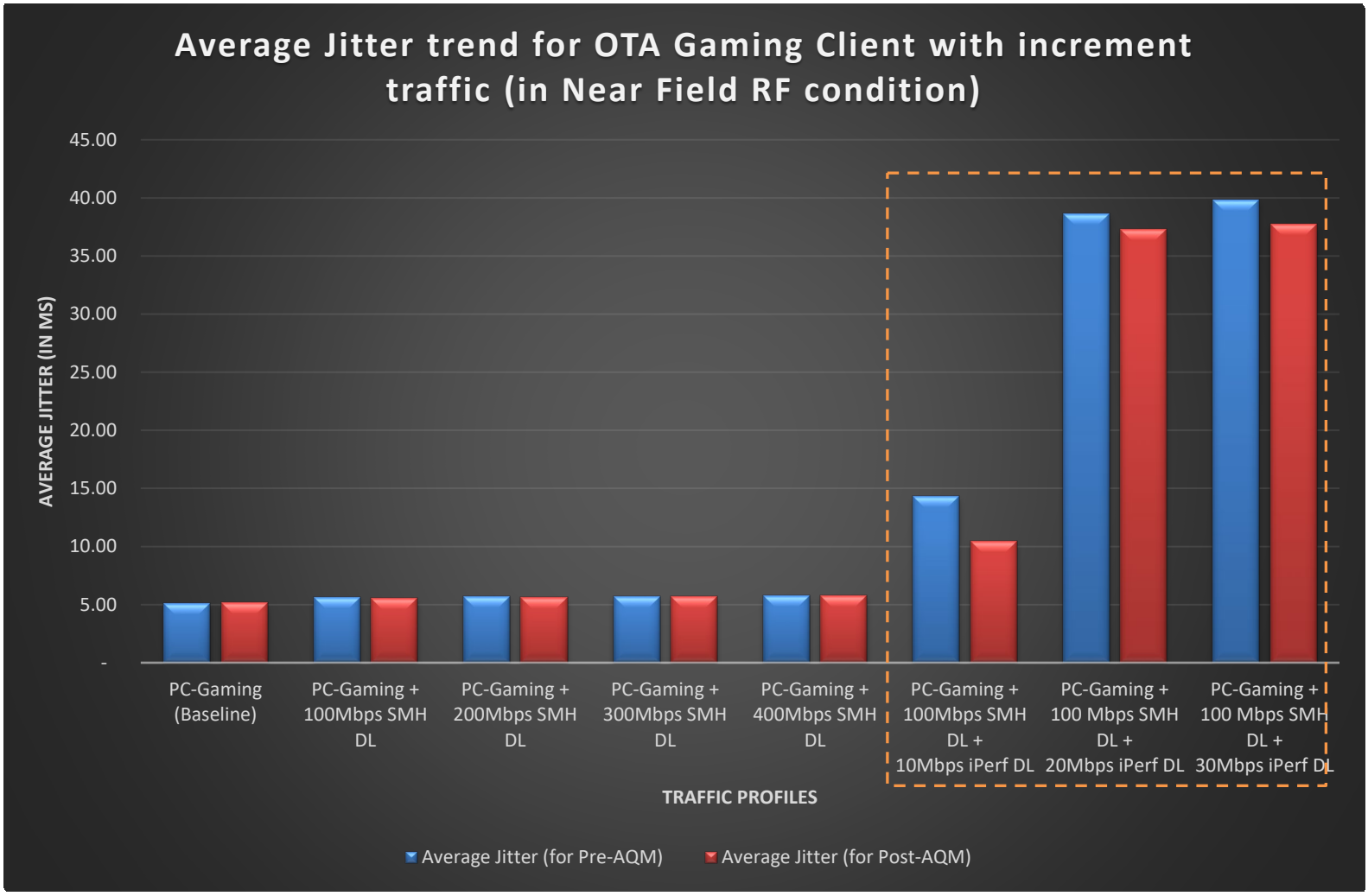


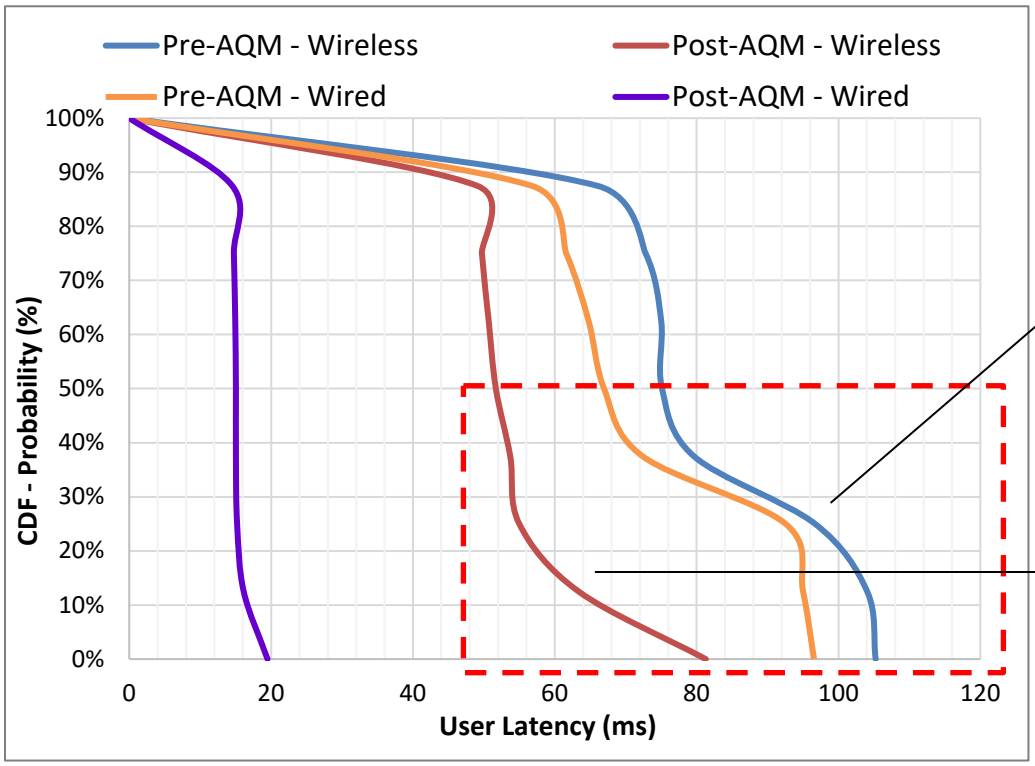
Figure 18 - Average Jitter trend for OTA Gaming Client with increment traffic (in Near Field RF condition)

User Latency: Significant improvement is seen in user latency (i.e., input to action latency). This is the reaction delay for the user to notice the fps (frame per second) change during the game. Higher user latency indicates delay in screen rendering. Also, higher congestion results in screen tearing phenomena for the end user without AQM enablement on the CMTS. The far right column in Table 5 illustrates the wired post-AQM user latency improvements by 7 to 11 times after AQM is enabled on the CMTS.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Table 5 - User Latency Comparison on Gaming Client in Pre-AQM and Post-AQM

Traffic Profile	User Latency			
	PRE-AQM		POST-AQM	
	Near OTA	WIRED	Near OTA	WIRED
PC-Gaming (Baseline)	65.99	49.05	56.73	14.51
PC-Gaming + 100Mbps SMH DL-TCP	72.80	49.72	61.61	14.75
PC-Gaming + 200Mbps SMH DL-TCP	75.01	50.68	64.77	14.95
PC-Gaming + 300Mbps SMH DL-TCP	75.12	51.69	66.94	15.04
PC-Gaming + 400Mbps SMH DL-TCP	79.61	53.72	72.38	15.04
PC-Gaming + 100Mbps SMH DL + 10Mbps iPerf DL-TCP	96.70	54.99	92.54	15.21
PC-Gaming + 100 Mbps SMH DL + 20Mbps iPerf DL-TCP	104.05	63.36	95.01	16.14
PC-Gaming + 100 Mbps SMH DL + 30Mbps iPerf DL-TCP	105.24	81.29	96.55	19.50



Convex curve indicates that at higher congestion, without AQM, queue building traffic will result in excessive packet drops and an overall worse quality of experience.

Concave curve indicates that at higher congestion, AQM monitors the buffer and adjusts its target latency to maintain the overall quality of experience

Figure 19 - Impact of User Latency with incremental throughput

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Cloud Gaming throughput: With incremental congestion, as shown in the table below, we notice that the private gaming throughput is deteriorated when AQM is disabled.

Table 6 - Cloud Gaming throughput characterization (Pre-AQM vs Post AQM)

Traffic Profile	Wired		Near Field OTA	
	Pre-AQM	Post-AQM	Pre-AQM	Post-AQM
Profile 1: PC-Gaming (Baseline)	41.4	40.87	41.14	41.49
Profile 2: PC-Gaming + 100Mbps SMH	41.63	41.61	41.55	41.67
Profile 3: PC-Gaming + 200Mbps SMH DL	41.59	41.51	41.65	41.92
Profile 4: PC-Gaming + 300Mbps SMH DL	41.57	41.64	41.52	41.4
Profile 5: PC-Gaming + 400 Mbps SMH DL	41.22	41.48	41.47	41.48
Profile 6: PC-Gaming + 100 Mbps SMH DL + 10Mbps iPerf DL	26.11	41.57	26.28	41.55
Profile 7: PC-Gaming + 100 Mbps SMH DL + 20Mbps iPerf DL	26.18	41.51	26.12	41.58
Profile 8: PC-Gaming + 100 Mbps SMH DL + 30Mbps iPerf DL	26.11	41.47	26.06	41.6

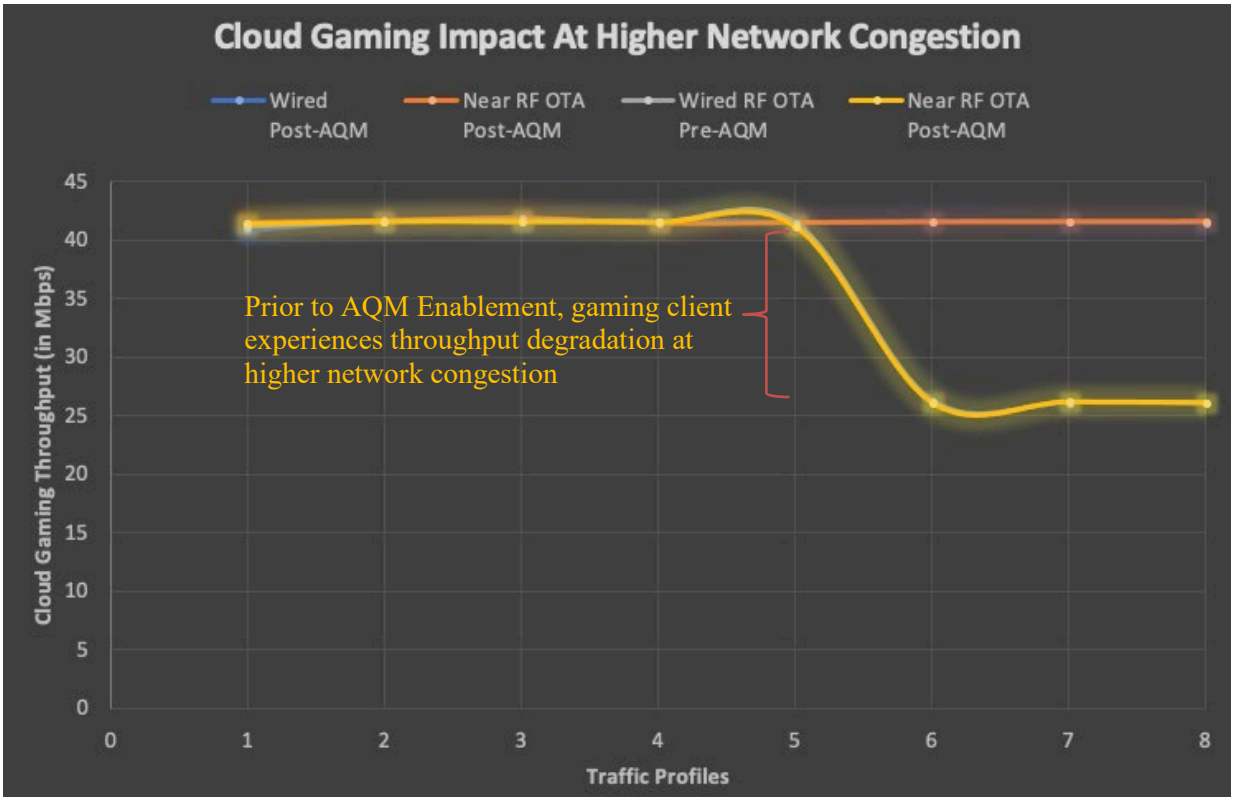


Figure 20 - Cloud gaming throughput impact at higher congestion levels

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

6. End-to-End Latency Management

The operational challenges of providing the best Quality of Service (QoS) solution for the user calls for an end-to-end latency improvement strategy for the client, including on WiFi so that congestion control notifications drive the prioritization and queue management of incoming packets based on the WMM access categories and implementation of L4S-AQM scheduler.

6.1. Latency Management Strategies on WiFi

With an increasing amount of mobile devices onboarding on the customer's network, WiFi routers are soon becoming a barrier/bottleneck to increasing the potential for buffer bloat on the customer's network. Although AQM implementation dramatically improves the end user's quality of experience on a wired interface, there are still opportunities for ISPs to efficiently manage working latencies on the WiFi link to deliver a superior quality of service. Latency optimization technologies exist today including Peering densification, AQM, LLD and WiFi WMM to address near to mid-term online experiences

6.1.1. Current Practice

Orthogonal Frequency Division Multiple Access (OFDMA)

OFDMA is part of the 802.11ax technology that demands very effective scheduling strategies to operate in highly dense and complicated network installations. Some of the most significant advantages of OFDMA include reduced network latencies. The entire Wi-Fi network latency in the context of WiFi networking is made up of the combination of downlink latency, or access point (AP) to client, and uplink latency, or client to AP.

Older WiFi versions allow any device in a network to commence transmission at any moment with little cooperation. While this method works well in less dense environments, it is inefficient in dense deployments owing to packet collisions. OFDMA is used in WiFi 6 to manage transmission and reception to and from non-AP devices through centralized coordination by an Access Point (AP). This enhances transmission efficiency and, as a result, lowers average latency.

WiFi Multimedia (WMM)

WMM is a WiFi specification built to prioritize voice and video traffic over best effort internet traffic while de-prioritizing background, non-latency sensitive packets on the network. In a congested environment, WMM guarantees to continue receiving the priority traffic it requires. This also ensures that latency sensitive packets spend less time in the queue, which reduces the overall latency and jitter.

According to the IEEE 802.11e wireless QoS standard, the WiFi Alliance specification allows network packets to be processed differently based on application or network tags that define the traffic type. WMM, which is based on tags sent over the network on the downstream and upstream, may be implemented efficiently in a hostile WiFi environment to prioritize latency sensitive traffic ahead and prevent congestion.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

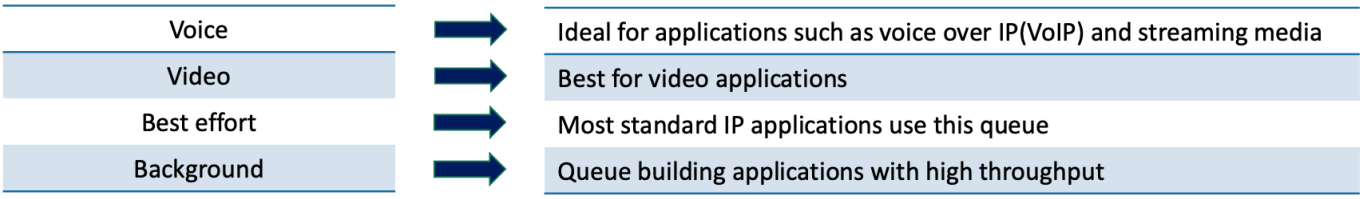


Figure 21 - Different Data-type classifications in priority order, highest to lowest

6.1.2. Future Scope

When clients are connected on the wired interface, the likelihood of latency degradation is greatly improved by implementing the right AQM algorithm. To enhance the experience further for clients on the WiFi link, other technologies like Low Latency, Low Loss, Scalable Throughput (L4S) would set the platform to achieve low latency and higher throughput for latency sensitive applications. Here are a few latency management initiatives that could greatly improve the overall quality of experience for the customer.

Low Latency, Low Loss Scalable Throughput (L4S)

The core basis of L4S is to resolve queuing delay by adopting a new/enhanced class of congestion control (assisted by a modified form of ECN) that seeks capacity with much less queuing delay. LLD's basic features include support for two queues in each direction:

a. "Classic" queue for standard congestion-controlled traffic with deep buffer (together with AQM), which enables classic congestion controllers to deliver high throughput while keeping latency under reasonable control.
b. "Low Latency" queue for traffic that does not cause delay, latency fluctuation or loss. The low latency queue contains an extremely thin buffer as well as additional features that allow "well behaved" applications to achieve ultra-low latency.

L4S extends ultra-low latency treatment to a wider range of applications and enables greater end-to-end latency management (beyond DOCSIS). L4S targets high data rate, ultra-low latency and low packet loss, which is good for cloud gaming, cloud AR/VR, etc. Implementation of L4S with AQM over WiFi should certainly be considered as a future scope on WiFi after LLD implementation on DOCSIS.

L4S uses application marking and for it to be used on WiFi, networks need to respect/carry Differentiated Services Code Point (DSCP) marking across the access and core networks. This requires industry adoption by external content providers as well as ISPs for low latency use cases.

Table 7 - Fundamentals to Support Latency Management on WiFi

Initiative	Description	Expected Action
Traffic / Application-Based Prioritization	Dynamic prioritization of traffic over WiFi based on congestion affecting user experience	Improved WiFi experience for customer dynamically based on congestion, traffic and application being used

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Device-Based Prioritization	Prioritization of specific devices that are connected to the network with or without user intervention	Potential use cases could include ISPs letting customers choose the traffic prioritization on a device based on their needs
WiFi Queue Management & Advanced Traffic Prioritization	Queue management using combination of traffic, application and device-based prioritization along with leveraging WMM access categories	Improved WiFi experience for customer dynamically based on congestion, traffic, device and application being used Note: This would need to have individual implementations completed for app, device and traffic

7. Conclusion

Network latency is often thought to affect only applications that leverage higher throughput, but persistent low latency results provide a better connectivity experience for the user. Access layer latency initiatives (like AQM, LLD, etc.) improve the experience for wired client connectivity and enable the access network infrastructure to more efficiently process packets. However, to deliver an end-to-end QoS to the customer on WiFi, classic Explicit Congestion Notifications (ECN) and "scalable" congestion control to prioritize and mark the packets at the router/access point on the airlink are critical.

The validation conducted at the production test house to assess the cloud gaming proved that AQM is much more effective for latency-sensitive traffic (like cloud gaming as discussed in this paper). When queue depth or buffer size approaches the threshold target latency set by the DOCSIS-PIE algorithm, we do see an efficient approach to calculate drop probability and some packet drops to signal the sender to slow down and thereby control congestion.

Additionally, although we do see improvements on the access layer KPIs enablement on the network, the end user improvements are perceived by wired clients only. Clients connected wirelessly to the router would not vividly experience a difference in their QoE unless latency management strategies are implemented on the router/access point to mark the traffic based on its classification type. These could be done through scalable and classic congestion notifications, WMM markings and low latency service flows on WiFi.

Standardization of how the QoS metrics are measured across core, access and in-home WiFi is critical, and so end-to-end QoE assessment will help the cable industry drive across the common goal of standardized latency monitoring and optimization across the transmission medium.

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Abbreviations

AP	Access Point
Mbps	Megabits per second
ms	Milliseconds
AQM	Active Queue Management
WRED	Weighted Random Early Detection
DOCSIS	Data over Cable Service Interface Specifications
PIE	Proportional Integral Controller Enhanced
L4S	Low Latency, Low Loss, Scalable Throughput
IETF	Internet Engineering Task Force
CCAP	Converged Cable Access Platform
CMTS	Cable Modem Termination System
PGS	Proactive Grant Service
LLD	Low Latency DOCSIS
DCTCP	Data Center Transport Control Protocol
WMM	WiFi Multimedia
OFDMA	Orthogonal Frequency Division Multiple Access
QoS	Quality of Service
ECN	Explicit Congestion Notification
SMH	Spectrum Mobile Home
DL-TCP	Downlink- Transport Control Protocol
fps	Frame Per Second
OTA	Over The Air
ICMP	Internet Control Message Protocol
SSID	Service Set Identifier
UDP	User Datagram Protocol
KPI	Key Performance Indicator
RTT	Round Trip Time
IEEE	Institute of Electrical and Electronics Engineers
QB	Queue Building
NQB	Non-Queue Building
PHB	Per Hop Behavior
QoE	Quality of Experience
QoS	Quality of Service

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

Bibliography & References

1 Improving The Latency Of An MSO Network For Gaming And Real Time Applications, by Colin Dearborn

2Low Latency DOCSIS Overview And Performance Characteristics by Greg White, Karthik Sundaresan and Bob Briscoe

3Fastest Path to Low Latency Services by Sebnem Ozer

4Configuring and Deploying Low Latency DOCSIS Networks by Greg White, Karthik Sundaresan

5Dual-Queue Coupled Active Queue Management (AQM) for Low Latency, Low Loss, and Scalable Throughput (L4S) RFC 9332 by K. De Schepper*, B. Briscoe and Greg White*

6Tracking Round Trip Time Latency in the MSO Network by Michael Overcash; Alan Skinner; Owen Parsons; Daniel Sciscoe

7Towards Predictable Low Latency DOCSIS Services by Dan Rice, Sebnem Ozer, Ph.D., James Martin, Ph.D.

8Webinar – Improving Latency with AQM June 2023 by Chris Comstock and Hamish Bala

9Webinar – Unleashing the Power of Low Latency: Enhancing WiFi Networks for Optimal User Experience by Annie George, Sebnem Ozerr and David Dickson

© 2023, SCTE® CableLabs® and NCTA. All rights reserved.

ISBN 0-940272-01-6; 0-940272-08-3; 0-940272-10-5; 0-940272-11-3; 0-940272-12-1; 0-940272-14-8; 0-940272-15-6; 0-940272-16-4; 0-940272-18-0; 0-940272-19-9; 0-940272-20-2; 0-940272-21-0; 0-940272-22-9; 0-940272-23-7; 0-940272-24-5; 0-940272-25-3; 0-940272-26-1; 0-940272-27-X; 0-940272-28-8; 0-940272-29-6; 0-940272-32-6; 0-940272-33-4; 0-940272-34-2; 0-940272-35-0; 0-940272-36-9; 0-940272-37-7; 0-940272-38-5; 0-940272-39-3; 0-940272-40-7; 0-940272-41-5; 0-940272-42-3; 0-940272-43-1; 0-940272-44-X; 0-940272-45-8; 0-940272-46-6; 0-940272-47-4; 0-940272-48-2; 0-940272-49-0; 0-940272-50-4; 0-940272-51-2; 0-940272-52-0; 0-940272-53-9; 0-940272-54-7

© 2015 National Cable and Telecommunications Association. All Rights Reserved.