THE COMPLETE
TECHNICAL PAPER PROCEEDINGS
FROM:

NCTA
Technical
Papers

4096-OFDM Implementation on the HFC Plant with Fiber Deep and Distributed Access Architecture

Maxwell Huang
Cisco Systems

Abstract

The paper will introduce what are the limitations existed in the current HFC architecture, why it is wise to continue the investment on the HFC infrastructure evolution instead of reconstructing to an all - fiber infrastructure, and why the new HFC architecture with the Distributed Access Architecture (DAA) in conjunction with the Fiber Deep (FD) can resolve the limitations and helps make the 4096-OFDM application on HFC plant a reality.

The paper will also introduce the parameters that impact the MER and BER performance such that AWGN, IQ Gain, IQ Phase, Phase noise, CW interfere-nce, Group delay, Micro-reflection, distortion and PAPR; and addresses why the PAPR becomes a major concern in the DAA + FD architecture where the 4096-OFDM application and the higher output RF power of node are both being expected.

In the fiber deep deployment a much higher output RF power is expected for the node that, together with the R-PHY module installed in the node, could cause a significant power increase therefore leading to serious concern at the thermal dissipation. The paper will finally introduce some possible solutions to reduce the PAPR and the power. This is to accomplish the goal to make the HFC a 10Gbps access network while maintaining the good and stable BER performance on top of the 4096-OFDM application.

DOCSIS 3.1 Evolution: A Journey To Continue

The DOCSIS 3.1 developments took place a couple of years ago, and now the 1.2GHz HFC transportation products, CCAP with the DOCSIS 3.1 PHY and MAC implemented, DOCSIS 3.1 cable modem and test equipment become available for demonstrations, applications and deployments.

In the earlier roadmap the DOCSIS 3.1 evolution was planned dividing into two phases. Phase 1 aims for 7Gbps downstream (DS) data throughput and 1Gbps upstream (US) data throughput, and phase 2 shoots for the 10Gbps DS throughput and 2.5Gbps US throughput. To get such the robust increase in throughput, some major changes were planned to introduce into the DOCSIS 3.1 evolution as follows.

- Involve Orthogonal Frequency Division Multiplexing (OFDM).
- Involve the Low Density Parity Check Code (LDPC) for the OFDM FEC.
- Expand the bandwidth to the 1.218GHz in phase 1 and to 1.794GHz in phase 2 for purpose of adding more QAM / OFDM channels.

OFDM has the 1dB improvement in the Signal Noise Ratio (SNR) for its filter edge vs. that of SC-QAM; LDPC with the 88% of coding rate can lower the requirement for SNR by 5dB compared to the Reed Solomon FEC used for SC-QAM. Hereby we can account for the 6dB reduction in the requirement of the SNR by the DOCSIS3.1 PHY.

Phase 1 and phase 2 both aim to keep the HFC plant architecture shown in the Figure 1 unchanged, e.g. with analog fiber for DS; analog or digital fiber for US; and with a long amplifier cascade for cable transportation. In this architecture a node normally connects to about 25 - 30 trunk or line extender amplifiers and may still serve 200 to 500 households (HHP).

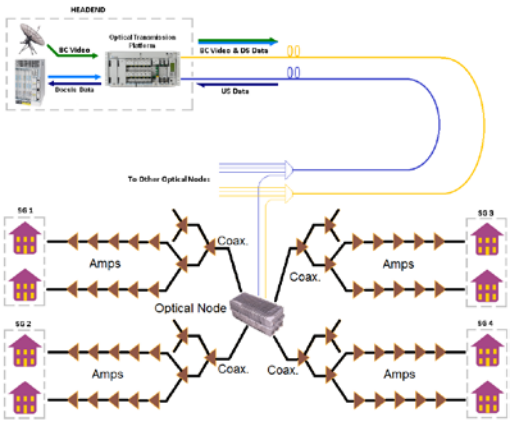


Figure 1: Current HFC Architecture

In such an evolution path, only will those 1GHz or 870MHz or 750MHz active devices such as forward transmitters, nodes and amplifiers be replaced with 1.2GHz devices, but the previously deployed fibers and cables are primarily kept untouched. This is called Active Drop In (ADI) upgrade whose primary benefit is to avoid a "forklift" upgrade to allow MSO's to gradually refresh their network infrastructure depending on the Compound Annual Growth Rate (CAGR). This could be the most economical solution compared to other new access technologies like EPOC and Google fiber. However, the MER performance is unexpected to be improved since analog fiber and long amplifier cascade essentially limit the Signal to Noise Ratio (SNR) performance that is often considered to be one of the major contributors among the impairments to Modulation Error Ratio (MER). Figure 2 shows an assessment result for the SNR performance over the millions cable modem in an existing "good" HFC plant.

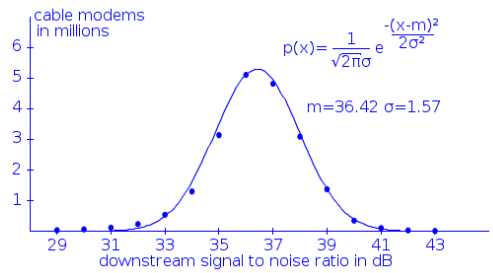


Figure 2: SNR Distribution over CM in millions

In the standards of DOCSIS 3.1 MAC and PHY, both CMTS and CM are required to support the 4096-QAM, and optionally support 8192-QAM and 16384-QAM. But typically the current HFC plants could not support beyond 1024-QAM (or even 512-QAM) order due to the limitation in the SNR performance.

DS	Band	Assumed MER	Typ. QAM Mod.	Date Rate / 6MHz	Total Data Rate
	MHz	dB		Mbps	Mbps
D3.0, All QAM,	105 - 1002	33	256	38	5681
D3.1, All OFDM	105 - 1002	33	1024	49	7326
	258 - 1218	33	1024	49	7840
	500 - 1794	33	1024	49	10568

Table 1: Planned DS Throughput Increase

US	Band	Assumed MER	Typ. QAM Mod.	Date Rate / 6MHz	Total Data Rate
	MHz	dB		Mbps	Mbps
D3.0, All QAM,	5 - 85	27	64	28	373
D3.1, All OFDM	5 - 85	27	256	38	507
	5 - 204	27	256	38	1260
	5 - 400	27	256	38	2502

Table 2: Planned US Throughput Increase

The Table 1 shows the planned DS data rate increase by the OFDM w/ LDPC and the spectrum expansion based on assumed MER performance. Though OFDM w/ LDPC can lower

the requirement in SNR, but the 3dB margin is reserved to handle the variation over time and temperature, and interference. The 1024-QAM DS and 256-QAM US account for an improvement in the data rate by approximately 29% compared to the 256-QAM DS and 64-QAM US typically used today. However, the 33dB SNR DS and 27dB SNR US would only be the case for those a few "good" HFC plant. The lower SNR distribution could be expected for other "not so good" HFC plants where 512-QAM DS and 128-QAM US could be more realistic case.

An upstream spectrum upgrade to 204MHz was planned, but due to an impact to the Out Of Band (OOB) signalling needed for controlling millions of cable modems, this upgrade to 204MHz is still on hold, now the situation is that MSOs take upgrade to 85MHz as an immediate upgrade.

Data Throughput	Start Frequency	Stop Frequency	MER	Typical QAM Modulation	Data Rate / 6MHz CH	Total Data Throughput	HHP	Data Throughput / Household
	(MHz)	(MHz)	(dB)		(Mbps)	(Mbps)		(Mbps)
DOCSIS 3.1 Downstream, All OFDM	105	1218	30-33	512	43.8	8125	300	27.1
DOCSIS 3.1 Upstream, All OFDMA (long code)	5	85	24-27	128	34.5	460	300	1.53
DOCSIS 3.1 Upstream, All OFDMA (middle code)	5	85	24-27	128	32.9	439	300	1.46
DOCSIS 3.1 Upstream, All OFDMA (short code)	5	85	24-27	128	29.1	388	300	1.29

Table 3: DOCSIS 3.1 Throughput Based on 85/105MHz Split

Table 3 shows a more realistic throughput data based on this situation. We can see that even with the OFDM with LDPC there is still a big gap to accomplish the 10Gbps goal of DOCSIS 3.1 downstream and the 2.5Gbps upstream. From a data rate per household perspective, we see an increase by the OFDM with LDPC compared to DOCSIS 3.0 but not much. The 27.1Mbps DS and 1.46Mbps US (middle code) could potentially limit those applications such as 4k video, remote education / health care, and video game etc. which require much higher data speed to run.

In the table 4 John Ulm and his co-authors revealed the demands in the data speed across different tiers and the compound annual growth rate (CAGR) projection by the researches to the several large North American MSOs.

Subscriber Distribution	MSO #A	MSO #B	DS Speed (Mbps)	CAGR
Top Tier, Billboard speed	0.20%	0.1 - 1%	300	50%
Performance Tiers	24%	5 - 18%	75	32%
Common, Popular Tiers	64%	50 - 72%	25	26%
Economy Tiers	12%	15 -40%	5	15%

Table 4: Subscriber Mixes across the Service Tier

If we take MSO #A for its distribution percentages, and assume a node serving 300 of Households (HH) where the four different service tiers coexist, we can calculate the data speed needed for different tiers and the total

throughput needed. Results are shown in the Table 5. Recall that the 8125Mbps calculated for DOCSIS 3.1 DS throughput in the Table 3, it doesn't fully satisfy the demand in the data throughput even for that will come soon, not to mention the needs projected by the CAGR in the Table 4 that will come a few years later.

Subscriber Distribution	MSO #A	DS Speed (Mbps)	CAGR	Number of HH (300 HHP based)	Current Demand (Mbps)	Demand in 5 years later (Mbps)
Top Tier, Billboard speed	0.50%	300	50%	1.5	450	3417
Performance Tiers	24%	75	32%	72	5400	21640
Common, Popular Tiers	64%	25	26%	192	4800	15244
Economy Tiers	13.50%	5	15%	40.5	203	408
Total Throughput Required (MSO #A based)					10853	40710

Table 5: Total Throughput Needed to Satisfy Different Tiers

Spectrum expansion to 1794MHz DS and 400MHz US was ever an option to further increase the data throughput capacity in the DOCSIS 3.1 standards. But this could be too challenge to implement and also could miss the original intent of the ADI upgrade for the following reasons.

- Taps must be replaced. Otherwise the loss at 1.8GHz could be up to 70dB! This suggests that the ADI strategy is not so worthwhile.
- Take CommScope P-III cable and 1705 feet as an example, the loss at 1794MHz is 12.5dB higher than that at 1218MHz. This means that node's output level and amplifier's gain must have another significant increase, therefore leaving tremendous challenges for gain-block, filter designs and the reliability associated with increased power / gain.

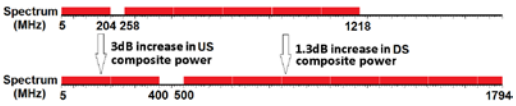


Figure 3: Increases in composite power

- The OMI to analog transmitter would have to reduce by 1.3dB according to the composite power increase shown on Figure 3 and therefore degrade the CNR in the DS.
- The 3dB increase in US composite power would cause 4 to 5dB reduction in the Noise Power Ratio (NPR) range. This appears a big concern against the current HFC architecture where sufficient operational headroom is needed against the variations over temperature from cables and cascaded amplifiers and possibly the analog fiber transportation if digital return is not being used.

Move to All Fiber or Continue HFC

As the current HFC architecture appears to have reached a point lacking the momentum to drive the continuous capacity increase, quite a few MSOs considered transitioning to the all fiber network to be a move for the access network evolution. However, we need to take investment, time, current demand & CAGR and available PON technologies into account.

To date USA has approximately 130 million households where 120 million, the great majority are reached by cable. To a tune of approximately 400 HHP per node, there are a total 300,000 node in the country. While in a typical coax architecture of N+6

there are around 25 actives per node. Comcast has made an estimate that 60 billion U.S. dollar would have to be invested to build up the infrastructure needed for the all fiber access network. Actually, Google spent about 850 U.S. dollar per household to spread its Google Fiber service to several cities in the US.

Verizon FiOS deployment for example took about 8 years to build fiber to roughly 18 million homes. So if the Verizon FiOS build-out rate were projected onto the whole nation, it would take half century to fiberize 120 millions of households in the nation.

Assume the GPON can deliver the 1Gbps per subscriber, then it is far ahead of the needs in the throughput indicated in the table 5. So doing a massive full FTTP migration using GPON would be an unwise move since MSOs may then have to take decades to recoup their tremendous investment.

For a move to full FTTP network, there also are several technical factors to be well considered. First is that PON architectures usually cover 32 or 64 ONUs while a node in a typical HFC network serves eight times of CPE device that implies 8 times of OLT ports be required for CCAP. Those key parameters such as ONU count, LLID per ONU and Grant Cycle time etc. were optimized for 32 ONUs or 64 ONUs on a PON network basis. Thus the impact must be well studied to apply these key parameters in a much larger service group. Especially during the transition to full FTTP network, the legacy HFC network has to be side-by-side operated, thus operators should encounter double challenges in both deploying PON to the serving group whose size is much bigger than it used to cover, and also taking care of maintaining and provisioning two networks whose protocols can not be coexisted to each other.

Radio Frequency over Glass (RFoG) would be an economical version of the fiber access network. Optical Beat Interference (OBI) has been a serious problem in the RFoG upstream. CMTS scheduler allows multiple CPEs to burst at the same time at different RF frequencies. This upstream multiple access has been a beauty of the traditional HFC system and never causes the collision. But it is not the case when transition to an RFoG system.

In a traditional HFC system depicted in the Figure 1, CPEs in households send their own burst signals via the different RF frequencies, these burst signals are firstly combined in the RF domain and then sent back via transmitter in node to headend. Transmitters with the same wavelength use the separate fibers, or share the same fiber but with the different CWDM or DWDM wavelengths. Thus in principle there should have no chance to cause the collision, and in practice it is.

In an RFoG system, CPEs in households send the burst signal via their own Optical Network Units (ONU) to headend. The ONUs turn on their lasers when the burst is detected. To date most of RFoG system used to use the same optical wavelength, though in reality the wavelength could be slightly different due to manufacturing tolerance.

Optical Beat interference (OBI) is the hyterodyning effect resulted form two or more wavelengths that are close enough (less than 0.0125nm) present on the same detector of receiver. OBI appears as the wideband noise and often severely degrading the SNR and

BER performance. When OBI effect occurs it affects no only the ones that send bursts simultaneously but also all the others within the same service group!

The elimination to OBI can not be solved in an opportunistic manner relying on the manufacturing tolerance of laser wavelength since nowadays laser manufacturers normally specify the wavelength tolerance less than 2.5nm. Thus the probability of two or more lasers in a 32 ONUs or 64 ONUs RFoG system falling into the OBI region of the 0.025 nm is high.

In reality the probability gets much higher because of the start up drift of the laser. The lasers in ONUs are turned off until there is a burst detected. So every time in the transition of laser tuning on, the laser starts absorbing bias current and output the optical power and drifts the wavelength through the initial start until reaching the steady state because of changes in the bias current and the temperature. The wavelength drift is about 0.5nm from initial start to a steady state. The drift usually lasts about 100 to 200 us that is being too fast to be possible for any correction.

Eliminating OBI therefore requires the ONU wavelengths to be separated by at least 0.5nm, preferably around 1nm to be thoroughly getting rid of the OBI even in a start-up event. Using DWDM is the only way to make it a reality. Such a solution is apparently expensive, and also operationally impractical in allocating the different wavelengths to households.

Using the Single Tx CMTS scheduler to allow only one cable modems to transmit at a time is simple and common in the DOCSIS 2.0 system. However, it ultimately limits the capacity if using the Single Tx CMTS Scheduler in the DOCSIS 3.0 and 3.1 systems. The capacity of a Single Tx CMTS scheduler would be merely half of that a 3.0 scheduler does, and gets even much less than that of a 3.1 scheduler. Thus attempting solving the OBI problem in a simplistic manner using Single Tx CMTS Schedule is totally against the purpose of moving to the all fiber network.

Fiber Deep and Distributed Access Architecture

Distributed digital architecture (DDA) can bring the significant benefits to the HFC network evolution such as enabling the higher PHY performance of DOCSIS 3.1, reducing the needs of space and power at the headends, hence the trend has become quite clear to evolve towards a distributed architecture rather than continuing investing the centralized architecture.

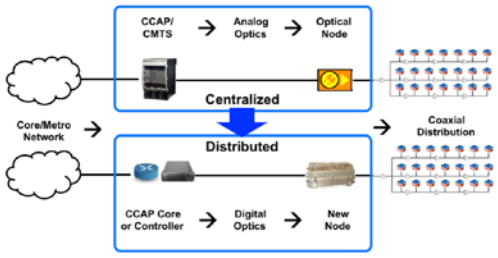


Figure 4: Centralized Vs. Distributed Architecture

Above figure illustrates a transition from the Centralized architecture to the Distributed architecture. There are two significant changes in this transition i.e. one is to move the PHY to a remote entity e.g. in HFC node, second is to use the digital fiber.

There are quite a few distribution technologies as to Baseband Digital Forward (BDF) / Baseband Digital Reverse (BDR), Remote PHY (R-

PHY), Remote PHY & MAC and Remote CMTS. Among these technologies, now Remote PHY has stood out as being accepted by almost all the major cable operators, system vendors, chip vendors, test equipment vendors. In the context of this paper the R-PHY will refer to DDA.

Using digital fiber can bring many benefits. We can leverage the existing WAN / LAN fiber infrastructures. The on/off keyed digital demodulation would only require about 20dB SNR to recover distorted digital signal to ideal 1 and 0, thus the degradation in SNR caused by fiber cross-talk effects, fiber non-linear effect do not play a role in limiting the application of multiplexing and de-multiplexing and long haul transportation. This is quite beneficial to support the fiber deep (FD) architecture where a significant increase in the number of optical node is expected.

In contrast, the link CNR performance of an analog optical transportation system is limited by a number of factors in terms of Relative Intensive Noise (RIN) of laser, shot noise, Optical Modulation Index (OMI) thermal noise, optical input power to Rx, photodiode's efficiency, noise current of receiver (EIN), and Amplified Spontaneous Emission (ASE) if EDFA is being used.

Let's take the following assumption to see how the link CNR performance looks like and varies vs. the change in the input power to receiver.

Assumptions:

- Wavelength: 1550nm
- Photodiode Efficiency: 0.9
- OMI / QAM channel: 1.5%
- EIN of Receiver: 6.5 pA / Hz
- Noise Figure of EDFA: 5.5

Optical Input Power to Receiver (dBm)	-4	-2	0
CNR (dB)	39.3	41.3	43.3

Table 6: CNR Variation over Optical Input

We can see that for the "best case" with 0dBm optical input, the link CNR performance is 43.3dB which could merely support the 4096-OFDM, and it varies 1dB as a result of 1dB change in the optical input power to Rx. In this regard the analog link CNR performance is quite optical link loss dependant. Other factors such as Optical Modulation Index (OMI), EIN of receiver, noise figure of EDFA that are varied from equipment to equipment, and operation to operation alter the analog link CNR performance, and in some cases make dramatic variation.

As CNR, SNR and MER are being referred in different places of this paper for the specific intents, so let's see the difference and the relationship between them for the benefit of addressing the following sections. Both CNR and SNR refer to the Adaptive White Gaussian Noise (AWGN), but as their terms imply: CNR is for the ratio expressed in decibels between the RF carrier and the noise; SNR is for the ratio expressed in decibels between the baseband signal and the noise.

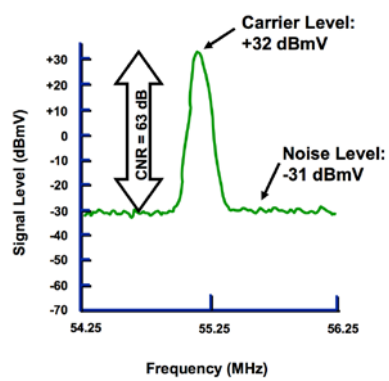


Figure 5: CNR definition

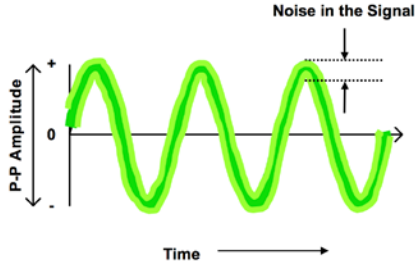


Figure 6: SNR definition

CNR is more often measured with the unmodulated carrier to yield the better measurement accuracy. SNR is the baseband measurement that has to be measured at either pre-modulation or post-detection and the noise. In other words CNR measurement is at the RF domain and ideal for characterizing the impact from those impairments specifically generated from network transportation including transmitter, optical amplifier, RF amplifier, transport path, and receiver; SNR measurement is at the baseband domain and ideal for characterizing the end–to–end signal quality which is impacted by signal source and demodulator in addition to those impairments that matter the CNR.

MER is the ratio between average signal power and average impairment power across all the constellation points in a QAM or an OFDM format used, and is a composite index covering all kinds of impairments such as phase noise, AWGN, I/Q gain error, I/Q phase error, distortion, clipping, ingress / impulse noise, compression, micro-reflection, and group delay etc. MER is considered to be a merit of the composite index representing the QAM / OFDM signal quality. Bit Error Ratio (BER) is defined as the ratio between the error data and the total data sent. MER typically has the certain relationship with BER for a certain QAM modulation order.

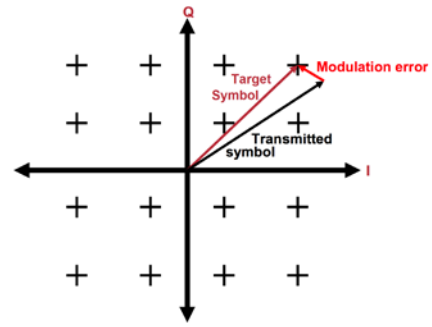


Figure 7: MER definition

In the most of communication systems including DOCSIS, the AWGN usually makes the dominant contribution to MER degradation, so CNR, SNR and MER are all used in the context to address the analysis needs depending on whether it is of interest in the RF domain, or in the baseband domain, or for the QAM / OFDM signal quality.

Since we are focusing to implement the 4096-OFDM, so let's see what the MER, SNR and CNR performance be needed to support this high order modulation of OFDM. A sub-carrier of an OFDM or a QAM has a number of the constellation points depending on the order of QAM modulation.

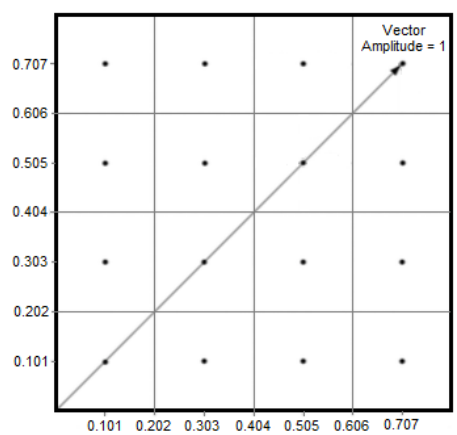


Figure 8: Quarter of Constellation diagram of 64QAM

Above is a quarter of the constellation diagram of a 64QAM. The right-top corner symbols has the maximum

vector amplitude. Assign a normalized amplitude of "1" to this symbol, then its I and Q components have 0.7071 of amplitude $1^2=0.7071^2+0.7071^2$. From 0.7071 to x (or y) axis, there are three constellation points. We can calculate out the boundary spacing for 64-QAM: $0.7071 / (n^{0.5}-1) = 0.1010$ (n=64) Accordingly, the boundary spacing for 256-QAM, 1024-QAM and 4096-QAM are 0.0471, 0.0228 and 0.0112 respectively.

For a pure AWGN channel, the BER can be calculated with the following formula.

$$f_x(x) = e^{-x\wedge 2 \,/\, 2*\sigma\wedge 2} / (\sigma * (2\pi)\wedge 0.5)$$

Where σ is the ratio between the boundary distance and the half average noise power (rms). Half average noise power (rms) is the half of noise power (corresponding to 3dB bandwidth relative to the symbol rate e.g. 5.0569MHz or 5.03605MHz in Annex B) derived from the SNR and average power across the constellation points. 64-QAM, 256-QAM, 1024-QAM and 4096-QAM have the respective average powers -3.68dB, -4.23dB, -4.50dB, and -4.64dB.

QAM Order	64	256	1024	4096
Bits	6	8	10	12
Boundary Distance	0.1010	0.0471	0.0228	0.0112
SNR (dB)	24	30	36	42
Half Noise Power (dB)	30.68	37.23	43.50	49.64
Half Average Noise Power (rms)	0.0292	0.0138	0.0067	0.0033
Distance in σ	3.45	3.43	3.41	3.40

Table 7: Distance in σ calculations for different QAMs

3.4 of distance in σ yields 6e-4 BER, this pre-FEC BER should work well since the error rate is low and usually be corrected to 10^{-8} by Reed-Solomon FEC. To get the 3.4 of distance in σ, the higher order QAMs / OFDMs that have the denser constellation points and the smaller boundary distance have to have the better SNR, usually 6dB penalty per 2 bits increase.

The 42dB of SNR requirement appears too severe. OFDM uses the LDPC FEC by which we can lower the SNR requirement to 36dB to support the 4096-OFDM applications when 87.85% of code rate is applied. The 87.85% of code rate means a spectral efficiency of 10.5 bps / Hz which is half bit higher than 1024-QAM. 4096-OFDM with the 36dB SNR works only in a pure AWGN system, but in reality there are many more impairments involved, the over temperature stability also needs to be considered. So if to leave the 3dB to 5dB margin for the degradations introduced by other impairments and the 2dB for the variation over temperature, we would need about 41dB to 43dB of the SNR in an end-to-end system.

Per the results shown in the table 6, the CNR performance on the analog optical link for those scenarios where the optical input power to receiver is below -2dBm apparently doesn't meet the requirement in supporting 4096-QAM. In contrast, the link performance of a digital transportation system is distance independent that the same MER can be maintained from 0 km to 100 km. The optical input power to Rx is no longer the key factor in determining the system CNR.

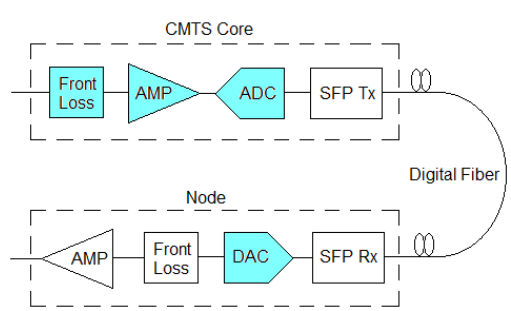


Figure 9: Noise Contributors in BDF System

Figure 8 shows the major noise contributors (blocks in cyan color) involved in the baseband digital forward (BDF) where the front loss and the noise performance of driving amplifier before ADC, the SNR performance of ADC, and the SNR performance of the digital to analog converter (DAC) do matter the link SNR performance.

The BDF solution has been considered for years but not been implemented because: it doesn't eliminate the noise contributions from front loss, driving amplifier and ADC; An ADC with ultra-high sampling frequency is too costly; BDF continuously produces 100% of throughput depending on bandwidth, sampling frequency, and overhead etc. but regardless the actual traffic; it doesn't allow CMTS to traverse a generic IP network to remote entities.

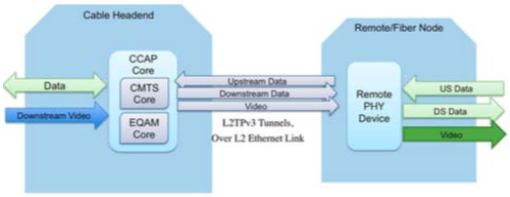


Figure 10: Remote PHY Architecture

The Remote PHY architecture shown in the Figure 9 uses a protocol with a concept of a pseudowire (PW). A pseudowire protocol used in the Remote PHY provides an IP tunnel and to move MPEG-TS packets and DOCSIS frames between CMTS core and Remote PHY. In this architecture the noise contribution only come from the DAC and the Noise Figure (NF) of node driving amplifier. Above the 50dB of SNR performance can be achieved by the DAC (or in an integrated chip) offered by major vendors.

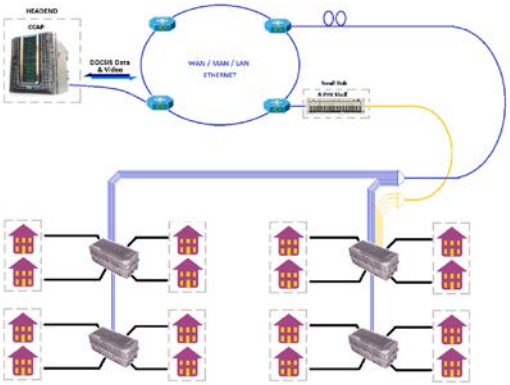


Figure 11: Fiber Deep Architecture

Figure 10 depicts a FD network where no RF amplifier is between nodes and households. We can get approximately 63dB of CNR for the driving amplifier in the node by the following assumptions.

Assumptions for CNR Analysis for All QAM loading:

- NF of Driving Amplifier: 15dB
- DAC's Output: 20dBmV
- Noise band: 5.0569MHz

In the FD architecture, there may still have a few active components along cable to households. Active components should add very little degradation to overall noise performance. The home gateway network would contribute some and may need to be improved if it limits supporting 4096-QAM applications. But generally, the new architecture with remote PHY in conjunction with fiber deep network eliminates the limitation in the CNR performance required for 4096-OFDM.

However, to support the 4096-QAM, we must consider a number of impairments generated from signal source, transportation network such as AWGN, CW interference, phase noise, I/Q gain error, I/Q phase error, distortion, compression, group delay micro-reflection etc.

Additive White Gaussian Noise (AWGN) is the most common and unavoidable impairment that could be the dominant factor for the overall SNR performance in an analog optical system as addressed in the previous sections. AWGN is White (flat power density function in frequency) and Gaussian (mathematically "normal" amplitude density function). It spreads the received symbols in a cluster around the ideal location.

Phase noise (PN) can be generated from any signal source with frequency synthesizer and oscillator. It is normally caused by thermal "agitation". PN is cumulated from the signal source through the network to the receiving instrument.

CW interference can be sourced from various place, some from the modulator / up-converter (residual Local Oscillator or beat), some from the transportation network such as ingress and distortion beats etc., and some from receiving equipment like very denser tuners. CW interference introduces vector error power that circles around the ideal symbol appearing like "doughnut", the radius of error power depends on the difference in frequency between carrier and interferer.

Distortions such as CTB and CSO are the narrow band interference but cannot be considered as CW interferer. Unlike CW interferers, CTB and CSO can have huge number of products that vary dramatically in amplitude, 15dB difference between peak and average can be observed. CTB / CSO looks similar to CW interference on the spectrum analyser display, but appear differently from a constellation view, rather than "doughnut" it looks like noise especially in the presence of so many CTB and CSO beats.

I/Q gain error is caused by imperfection of signal source and is a difference in gain between the In-phase and Quadrature components of the constellation shapes it into a rectangle instead of a square.

I/Q phase error is caused by imperfection of signal source and is an error in quadrature between the In-phase and Quadrature axes of the constellation shapes into a lozenge or diamond instead of square.

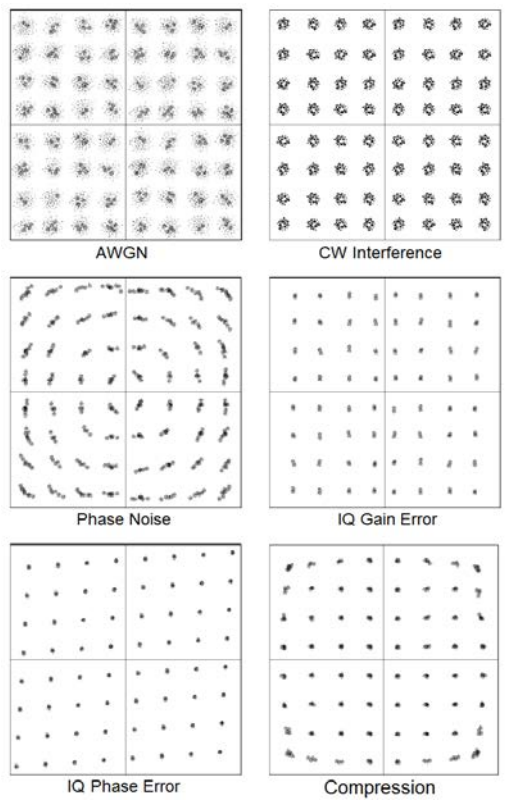


Figure 12: Distorted Constellation Diagram due to Impairments

Echo results from an imperfect impedance matching between express cables, nodes and amplifiers in the HFC plant. Nodes and amplifiers usually have a minimum 16dB return loss, assume a 1925-foot express cable, a 4.5µs echo could be expected. Echo would cause the Inter-symbol Interference that was ever of concern for OFDM. Now a simple method adding a single guard interval can

eliminate the inter-symbol interference without needing to know the channel response, so long as the guard interval is greater than the longest echo. The guard interval is configurable in the DOCSIS 3.1 PHY.

If the subcarrier frequency spacing is too small, then the phase noise generated from oscillators and frequency synthesizers could play a big role in degrading the performance. The OFDM sampling frequency must be larger than the channel width. DOCSIS 3.1 standards choose 204.8MHz as the sampling frequency, 192MHz as the biggest channel width, and the FFT size as 8192 or 4096. This keeps the complexity relatively low while providing a long enough useful time and wide enough subcarrier frequency spacing against the phase noise.

Subcarriers in an OFDM channel have a long time period, so that an impulse noise that is short in time can be averaged over a long symbol time to minimize its impact. OFDM has an ability to load bit differently, which means a single OFDM channel can contain different modulations such as 4096-QAM, 2048-QAM, 1024-QAM and so on of sub-carriers, to balance date rate vs. SNR over frequency. This ability can be utilized to handle the ingress noise. Ingress noise usually lasts long in time but short in frequency, having a few subcarrier adjusted to a lower bit modulation could be a good trade-off in handling the ingress noise.

The input level to the cable modem (CM) or set-top box (STB) primarily determines the SNR performance at the home gateway. Thermal noise is always present in any electronic device and it is flat over frequency. At room temperature, the thermal noise is -174dBm / Hz which can be converted to -125.25dBmV / Hz by adding 30 + 10 * log10 (75) for a 75ohm system. The noise floor is then added by the 192MHz bandwidth of an OFDM channel and the 10dB of noise figure of CM. If 42dB of SNR is required, we will at least need 43dB of CNR and therefore requiring 10.6dBmV level for a 192MHz of OFDM at CM (or STB) input. To compare to SC-QAM used today, the 10.6dBmV at 192MHz band can be extrapolated -4.5dBmV in the 6MHz band.

CM Input requirement		
Thermal noise	dBm/Hz	-174
Thermal noise	dBmV/Hz	-125.25
OFDM bandwidth	MHz	192
NF of Rx in CM	dB	10
CNR requirement	dB	43
CM Input requirement (192MHz)	dBmV	10.6
CM Input requirement (6MHz)	dBmV	-4.5

Table 8: CM Input Requirement

A statistics indicates that in the current HFC plants only 85% CM have an input level great than -6dBmV while the rest don't. As cable amplifier which usually includes the automatic gain control (AGC) to compensate for the change in cable loss due to the variation over temperature be eliminated in the FD network, the FD network should be designed with higher input level than current HFC plant, maybe -1.5dBmV to all the cable modems at room temperature to ensure 3dB margin to guarantee above the -4.5dBmV over temperature, for the 4096-OFDM implementation. A penalty is that FD node needs to output a higher level that could lead to an increase in power dissipation and therefore a thermal concern.

In the FD architecture, we need to count in the MER degradation from R-PHY node that includes the R-PHY module and the RF power amplifier

module. DOCSIS 3.1 specifies the 47dB MER at 1002MHz for complete OFDM, so it is reasonable to assume the R-PHY module can achieve the 47dB of MER performance as well. The RF power amplifier offered by major vendors could have above 41dB of MER performance before running into compression.

Based on the currently achievable MER performance for R-PHY module and RF power amplifier, and the target MER performance at CPE (CM or STB), we can see the EOL MER performance in the following table.

Device	MER (dB)	
	Room Temp.	Over Temp.
R-PHY	47	46
Power Amplifier	41	39
CPE	42	41
EOL	37.9	36.4

Table 9: EOL MER Estimation

As addressed in the previous sections we can lower the SNR requirement to 36dB to support the 4096-OFDM applications by applying the LDPC FEC with the 87.85% of code rate. The MER variation over temperature for a node should be within 2dB, thus the 37.9dB of EOL MER performance at room temperature should support the 4096-OFDM application in a wide temperature basis.

In summary,

- By the R-PHY that can guarantee 50dB above SNR performance at DAC's output, and by the FD architecture that eliminates the degradation in the noise performance from amplifier chain, the limitation in the SNR performance required for 4096-OFDM applications is primarily eliminated.
- Those impairments generated by the signal source such I/Q gain error, I/Q phase error, and phase noise can be controlled by the MER spec of R-PHY module.
- OFDM has the guard interval mechanism against the inter-symbol interference due to micro-reflection.
- The effect of an impulse noise short in time can be averaged by a long symbol time of OFDM.
- Adjusting a few sub-carriers to a lower modulation order can mitigate the impact by an ingress noise narrow in frequency.
- FD architecture eliminates the degradation in the distortion performance such as CTB, CSO, XMOD and HUM etc. So a FD node with the legacy distortion spec or even a relaxed spec shall ensure the distortion not be a limiting factor.
- A -4.5dBmV input to CM or STB can ensure the MER performance at household's home.

The above summary doesn't include the compression that is one of impairments caused by the Peak to Average Power Ratio (PAPR) related to QAM / OFDM modulation format and the non-linear characteristic of RF power amplifier (PA).

PAPR is often considered as the main disadvantage of the OFDM technique. An OFDM channel with 192MHz bandwidth and 8192 FFT has 7540 sub-carriers. These sub-carriers are located at the different frequencies but transmit simultaneously in the time domain. The envelop of an OFDM signal approximates to the Gaussian distribution and can have the large fluctuation. The RF power amplifiers have a limited linear region. In case the peak envelop exceeds the linear region of PA, as a result, in-band and out-band distortions are being created. In-band distortion would appear as the

noise degrading the OFDM channel's own MER performance, and out-band distortions degrade the adjacent channel leakage ratio (ACLR) performance and therefore interfering the adjacent channels.

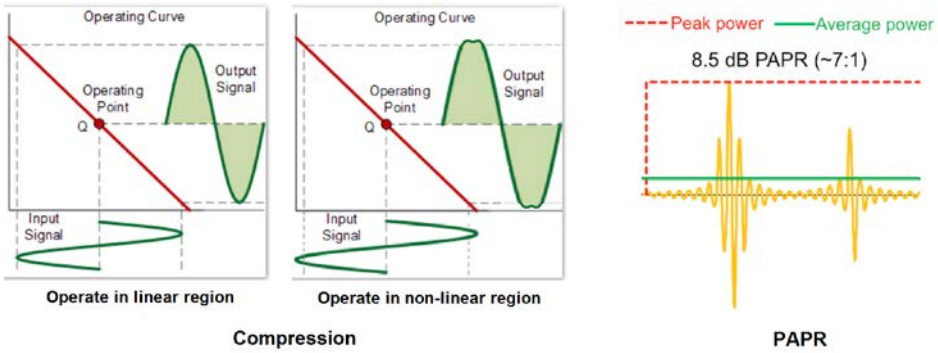


Figure 13: Compression and PAPR

The Complementary Cumulative Distribution Function (CCDF) is used to measure the probability of the time ratio spending on a given power. Table 10 shows results of the CCDF measured with PRBS for the respective loadings for all SC-QAMs and all 4096-OFDMs on 105 to 1218MHz band. The Power amplifiers used for the HFC system are usually sensitive with the crest power at the probabilities from 1% to 0.1%. According to the results in the following table, All 4096-OFDM loading has 0.39dB at 1% and 0.44dB at 0.1% higher crest power than all SC-QAM loading.

Probability	All SC-QAM	All OFDM
	Crest Power (dB)	
1.00%	6.33	6.72
0.10%	7.83	8.27
0.01%	8.83	9.55
0.001%	9.56	10.35
0.0001%	10.2	10.48
Peak	10.39	10.58

Table 10: CCDF SC-QAM vs OFDM

A 192MHz 4096-OFDM should have 1.87Gbps of the data rate that is 48 times bigger than the 38.7Mbps data rate delivered by a 256-QAM, so the impact must be well considered to avoid a significant data loss due to PAPR.

Output Power Back Off (OPBO) is a common but undesirable approach to handle the PAPR in both wired and wireless systems. The OPBO approaches setting the output signal power by ($OPBO_{(dB)}$) away from PA's saturated point. It is actually a trade-off between PA output capability and BER performance. DOCSIS 3.0 HFC plants use the 6dB OPBO to handle the PAPR of 256-QAMs. Increasing the $OPBO_{(dB)}$ for the DOCSIS 3.1 deployment in a DAA + FD architecture is unwise approach. The output power of a FD node is expected getting much higher but appearing very challenging for the following reasons.

The input level to CM or STB in household's home needs to be above the -1.5dBmV that is 4.5dB to 6dB higher than the current plant.

8 times or 10 times of FD node need to be deployed to cover an area which used to be served by a node plus about 25 to 30 pcs of trunk or line extender amplifiers. From operator's economic perspective, the higher output power that FD node can output, the less

number of new nodes are to be deployed.

HFC node used to use the class A power amplifier that has the best linearity but the least RF-to-DC power efficiency among all type of PAs. The best-case efficiency for class A PA would be around 13%, and for the worst case it is even lower than 4%. A 4-port FD node aiming to output 64dBmV of RF power per port would have an increase of about 25 watt of DC power and 29.4 watt (assume 85% of power supply efficiency) of AC power that is quite significant. A R-PHY module in replacement with 4 receivers and 4 transmitters previously used for node segmentation application would cause another increase of 20 watt of DC power and 23.5 watt of AC power. These bring an unprecedented thermal challenge for the node design. So adding more $OPBO_{(dB)}$ could make mission very difficult or impossible.

Device	MER (dB)	
	Room Temp.	Over Temp.
R-PHY	47	46
Power Amplifier (compression)	39	37
CPE	42	41
EOL	36.8	35.2

Table 11: EOL MER Degradation due to Compression

Assume power amplifier compressed due to the PAPR has 2dB degradeation in the MER. Indicated by Table 11 the EOL MER degrades to the 36.8dB at room temperature to have become marginal supporting the 4096-OFDM, and the 35.2dB of MER over temperature even can't support the 4096-OFDM.

Such a degradation in the MER can make the constellation diagram of 4096-OFDM indistinguishable and even could cause the OFDM constellations unlocked and therefore a significant data loss, this appears to be a big concern or say a severe problem.

Crest Factor Reduction (CFR) is the technique for the PAPR reduction. A number of the CFR techniques have been created and evolved, namely, Clipping and Filtering technique, Coding (C-OFDM) technique, Partial Transmit Sequence (PTS) technique, Active Constellation Extension (ACE) technique, Selective Mapping (SM) technique, and Pilot Reservation (PR) / Pilot Injection (PI) technique etc. Each technique has its own unique advantages and disadvantages. Among these techniques the Adaptive Baseband and Pilot reservation would be suitable for use in the DOCSIS 3.1. But their trade-offs in performance and computational complexity need to be taken into considerations.

Adaptive Baseband is one of Clipping and Filtering technique that is to implement the CFR at the baseband or say in the digital domain. Adaptive baseband is good for the ACLR improvement and can be implemented in the R-PHY module with a reasonable computational cost. Although the effect in the PAPR reduction by Adaptive baseband is not so deep, however, since the probability of exceeding 8dB PAPR for a wideband which may contain 6 x 192MHz OFDM channel is very low, so setting the 8dB as the threshold by the Adaptive Baseband would be good enough to prevent 4096-OFDM from running into a deep compression at which the OFDM is unlocked by receiver and all data lost.

Tone reservation is a sophisticated CFR technique that has the reserved tones allocated in the profile. These tones do not bear any information and are orthogonal to each other. The

amplitude of the peak reduction carriers is greater than those data carriers so that it can cancel out the peaks generated by the composite signal, removing them from the useful tones. Although tone reservation promises to accomplish CFR without degrading signal quality, it has several disadvantages that need to be considered, including a loss of spectral efficiency (due to pilot carriers added), loss of useful power, and increased computational overhead.

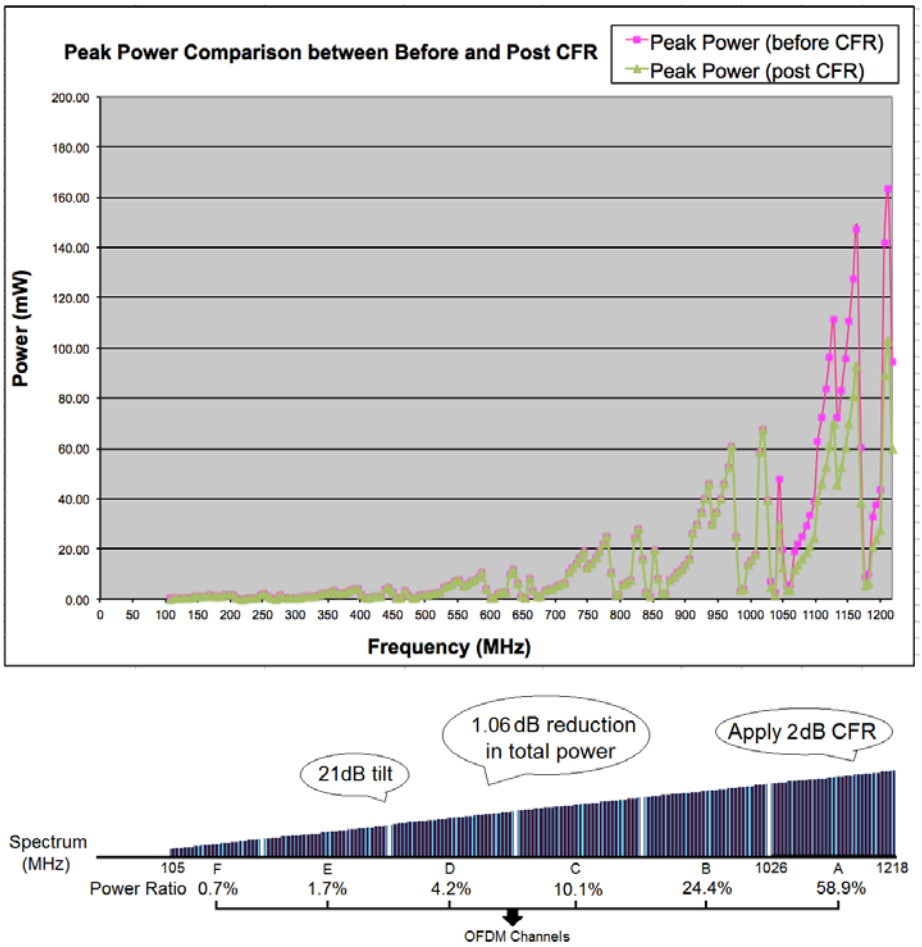


Figure 14: CFR implementation on the partial band

The cable spectrum has a unique character as having a tilted spectrum loading. As a consequence, OFDM or QAM channel at the higher frequencies have the much greater power than that of channels at the lower frequency. For instance, in the Figure 14, the channel A accounts for the 58.9% of power that is even greater than that all the rest channels combines. We can take advantage of this unique character to apply the CFR on the partial band to solve the concern at the increase in PAPR while keeping performance degradation and computational complexity relatively low. Given by an example illustrated in Figure 14, apply 2dB CFR on the OFDM channel A can reduce the total power in spectrum by 1.06dB.

As proven by simulation and test, the OFDM or QAM channels that are near to the upper edge of cable spectrum suffer more degradation in MER and BER than channels at lower

frequencies. To mitigate the risk in losing the data for an entire OFDM channel that may transmit up to 1.87Gbps of data, we may consider the another trade-off that is to lower the modulation order for the upper frequencies of OFDM sub-carriers, like from 1126 to 1218MHz. This is also in line with the partial CFR approach that would degrade the performance for the OFDM channel with CFR applied. We can still accomplish the more than the 10Gbps throughput as indicated by Table 12. Thus, this would be a more worthwhile trade-off than having the more $OPBO_{(dB)}$ for PA.

DS Data Throughput of DAA + FD Architecture	Start Freq.	Stop Freq.	MER	QAM Modulation	Date Rate / 6MHz CH	Total Data Throughput
	(MHz)	(MHz)	(dB)		(Mbps)	(Mbps)
4096-OFDM	105	1126	42	4096	58.5	9955
1024-OFDM	1126	1218	36	1024	48.7	747
Total						10701

Table 12

As addressed, moving the PHY to the HFC node and enabling the super high output significantly increase the power dissipation, and therefore bringing an unprecedented thermal challenge for HFC node.

In the wireless system, Digital pre-distortion (DPD) and envelope tracking (ET) are popular techniques for improving the RF to DC power efficiency for the PA.

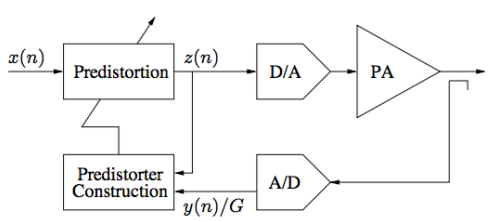


Figure 15: DPD System Diagram

Digital pre-distortion (DPD) adds a digital pre-distorter in the baseband to create an expanding nonlinearity that counteracts the compressing characteristic of the power amplifier. Ideally, the cascade of the pre-distorter and the power amplifier produces a constant gain for the original signal. With the pre-distorter, the power amplifier can be utilized more close to its saturation point while still maintaining a good linearity, thereby significantly increasing its efficiency. In reality, the power amplifier characteristics may change over time because of temperature drift, component aging, etc. Therefore, the pre-distorter should also have the ability to adapt to these changes.

There are several challenges in implementing the DPD into DOCSIS 3.1 as follows.

DPD requires 5 times of over-sampling to be able to detect the distortion effects. For the current 1218MHz cable spectrum, an Analog to Digital Converter (ADC) with the approximate 6GHz sampling frequency is way too expensive and could not be economically justifiable by the DPD itself regarding its gain at the power reduction.

Power amplifier exhibits the memory effects when working for the wideband application and being deeply compressed. DOCSIS is the wideband application, and power amplifiers need to be deeply compressed in order to yield a compelling power reduction. So very likely, the DPD algorithm has to be based on the memory effect that is to deal with the instantaneous and past

signal, which certainly adds more challenge.

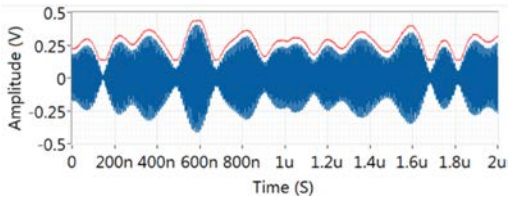


Figure 16: Envelope Tracking

Envelope tracking (ET) technique was created based on a fact that the actual envelope of OFDM or QAM signal fluctuate greatly, but because PA is supplied with a constant voltage, thus a big amount of energy becomes heat instead of RF output power particularly for those valley envelopes.

There are several challenges and concerns in implementing the ET into DOCSIS 3.1 as follows.

One of the main implementation challenges for an envelope-tracking IC concerns the bandwidth of the power supply that must accurately track the signal amplitude of the RF signal without clipping or introducing distortion. This requires a power supply with a control bandwidth of 1.5 to 3.0 times the RF channel bandwidth, i.e. 2GHz to 4GHz for a 1218MHz of DOCSIS 3.1 band.

The power supply needed in the ET system is also required to have the constant high efficiency simultaneously adapting the change in the supply voltage, very low noise, and high peak output currents. These don't make it technically infeasible but certainly ending up with a costly power supply.

With up to 40k sub-carriers simultaneously transmitting in the system, the probability of getting envelope peaks becomes relatively low, so does the envelope valleys! This suggests that the effect of power reduction by the ET becomes less significant.

As being discussed, the implementations of DPD and ET on the DOCSIS 3.1 system would face some challenges that may not be overcome in the short term. They could still have long way to go to be practical and cost-effective for the implementation in DOCSIS 3.1. Hence we must find a solution that can be immediately implemented to address the thermal challenge brought by R-PHY and super high output PAs in the HFC node.

The power amplifier devices are usually designed with the current adjustment feature to allow designers setting the PA current to an optimal performance associated with the output level. But in the past, it was always set to the maximum bias current to enable the maximum output capability of PA. But as illustrated in the Figure 16, in fact there is always slight or moderate difference in the cable loss between each of ports.

A power saving solution called Bias Adjustment in Field (BAiF) is to have an ability in the node to adjust the bias current for the PA in field according to the actual cable loss on the port of interest. Given by an example illustrated by the left part in Figure 17, assume node can compensate 54dB cable loss as a maximum, the bias current of PAs can be reduced according to the reduction in the cable loss. Three ports that are with 1dB, 2dB and 3dB reduction can save 19% - 27% of power on the power amplifier. The BAiF can be utilized when some channels are shutdown in the low traffic event. Given by an example illustrated by the right part in Figure 17, the shutdown to an OFDM channel near to upper edge with 192MHz bandwidth can lead to 30% to 40%

power saving by the BAiF on the power amplifier.

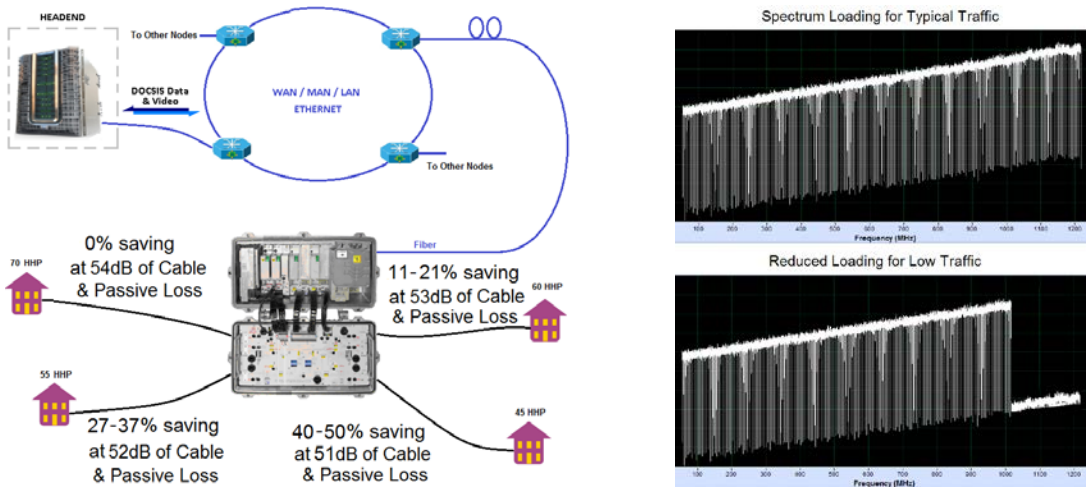


Figure 17: Cable Loss Based and Spectrum Loading Based Power Saving

Summary and Conclusion

Current HFC architecture becomes a limiting factor for the 4096-OFDM applications in regard to analog fiber and long amplifier cascade that are used. As a result, DOCSIS 3.1 can not meet the 10Gbps goal, and hereby may not keep pace with the CAGR in the data speed demand.

A move to an all-fiber access network using GPON is too costly and far ahead of CAGR that a tremendous investment spent would take decades to recoup. The OBI issue concerning the RFoG architecture, an economical version of fiber access network, and the operation challenge in managing two different networks side by side during transition could be the showstopper against this move.

The architecture with the Remote PHY plus the fiber deep primarily resolve the limitation in the link SNR performance and hereby making the 4096-OFDM implementation feasible. This new architecture is also good for the further spectrum expansion. However, the concerns at the increases in the PAPR from OFDM and in the power dissipation need to be taken care.

The increase in the PAPR could degrade the MER performance to an insufficient level to support the 4096-OFDM. CFR techniques such as the Adaptive Baseband and the Tone Reservation are the effective ways for the PAPR reduction but also resulting in some trade-offs. To minimize the negative effects we could apply the CFR for the partial band at the higher frequency where has the dominant power distribution over the entire spectrum.

The power increase brought by introducing the R-PHY module and enabling the super high output in the HFC node brings an unprecedented thermal challenge. A power reduction solution called the Bias Adjustment in Field that is readily to implement can significantly save the power based on the actual cable loss and the spectrum loading.

References

(1) The Yin And The Yang Of A Move To All Fiber: Transforming HFC To An All Fiber Network While Leveraging The Deployed HFC Assets Venk Mutalik, Marcel Schemmann, Zoran Maricevic, John Ulm

(2) Application of 4K-QAM, LDPC and OFDM for Gbps Data Rates over HFC Plant David John Urban

(3) Distributed CCAP Overview Technical Report CableLabs

(4) QAM Signal Impairments and Their Effects on MER and BER Gerard Terreault

(5) Remote PHY for Converged DOCSIS, Video and OOB John Chapman

(6) The Next Evolution in Cable: Converged, Distributed and Virtualized Access Network Jorge D. Salinger

(7) Digital Predistortion of Power Amplifiers for Wireless Applications Lei Ding

(8) Synergetic Crest Factor Reduction and Baseband Digital Predistortion for Adaptive 3G Doherty Power Amplifier Design Oualid Hammi, Scott Carichner, Bill Vassilakis, Fadhel M. Ghannouchi

(9) SCTE CNR vs SNR Ron Hranac

(10) HFC Improvement for DOCSIS 3.1 Evolution Maxwell Huang

A COMPARISON OF CENTRALIZED VS. DISTRIBUTED ACCESS ARCHITECTURES FOR PON

Michael Emmendorfer and Sebnem ZorluOzer
ARRIS

Abstract

This paper will define and compare two classes of access architectures that will emerge this decade for Passive Optical Network (PON). These two classes of access architecture may be referred to as centralized access architecture (CAA) and distributed access architecture (DAA) for PON. The CAA and DAA analysis is well underway for DOCSIS, but 10G EPON may be architected in a similar manner, in terms of splitting the PON access system components between the facility and a remote location, or placing all of the network device functions in a remote location. The DOCSIS use of distributed access architecture can place the MAC and PHY or just the PHY layer in the remote device, such as a node, cabinet, or multiple dwelling unit (MDU) location. The DAA for PON places both the MAC and PHY layers in the remote device. The DAA for PON defines how much of the upper layer functions are placed in the remote device and how much if any of the PON systems remain in the provider facility.

The optical line terminal (OLT) has evolved from a simple layer 2 device into some systems supporting higher layer networking functions, including advanced layer 2, layer 3 routing, and Multiprotocol Label Switching (MPLS). In addition, the cable industry defined a software mediation layer subsystem for Ethernet PON (EPON) and 10G EPON that enabled them to leverage their existing provisioning, network management, and network accounting systems used for their DOCSIS network to be used for EPON, this effort was the basis of CableLabs DOCSIS Provisioning of EPON (DPoE) standards.

The DPoE specifications defined advanced routing, switching, optical line terminal (OLT), and mediation layer subsystems to be part of a DPoE system. The DPoE system refers to the set of subsystems within the hub site that provides the functions necessary to meet the DPoE specification requirements [CableLabs]. This paper defines and examines next generation DPoE systems, which functions may not reside completely at a hub site, but rather can be in the cloud, hub site, and/or node locations. The next generation DPoE Systems will include additional features to facility based CAA DPoE systems and DAA DPoE systems.

The use of centralized access architecture (CAA) retains the MAC and PHY layer functions of the OLT subsystem in the headend or hub location. This paper will examine several current centralized access architectures for OLT subsystems. This paper will also propose the adoption of multiple wavelength technologies for 10G EPON DPoE systems. Three types of CAA DPoE systems will be examined in this paper.

The distributed access architecture (DAA) for PON began in the early 2000's with the placement of the entire OLT functions in a node or cabinet, this was referred to as remote OLT or OLT in the node. At the core of the next generation DAA DPoE systems there are three different distributed access architecture types that will be defined and evaluated in this paper. These define the functions that will be placed in the node or cabinet locations as well as those that may be placed at the facility; the facility may be a hub site or the cloud like a regional data center (RDC).

OVERVIEW OF CAA AND DAA FOR PON

This paper defines the centralized and distributed access architectures for PON. We have compiled an evaluation criterion for a side-by-side comparison of these different types of PON access architectures, so that a multiple system operator (MSO) can be aware of the next generation OLT or DPoE System possibilities.

The terms centralized access architecture (CAA) and distributed access architecture (DAA) are being used in the cable networking space. The term CAA means that the network media access control (MAC) and physical (PHY) layers are placed into systems in the service provider's facility, shown in Figure 1. The distributed access architecture places MAC and PHY layers or just the PHY layer in the outside plant node, cabinet or MDU location. If we consider cable deployments for DOCSIS, these utilize the centralized access architecture, which keeps the access MAC and PHY layers for DOCSIS in the service provider facility, while maintaining a transparent outside plant network. This is also the case with many PON deployments in the cable and telco space: the PON MAC and PHY are located in the facility enabling the outside plant (OSP) or optical distribution network (ODN) to remain passive and transparent.

The development of remote OLTs, since as early as 2002, placed the OLT MAC and PHY in a node or cabinet enclosure, but this did not garner significant market share. Most service providers preferred a centralized OLT that used a passive optical distribution network (ODN). However, placing the OLTs in the Headend/Central Office (HE/CO) facility may restrict the reach between the HE/CO and end users and/or reduce the split ratio and the number of subscribers served per OLT port. In recent years, there have been innovations to extend the reach or amount of subscribers per port of the centralized OLT architecture by placing an active networking device in the node or cabinet, called a PON Extender, while keeping the PON MAC and PHY in the facility.

The PON Extender is a powered device that could be placed in a carrier facility, cabinet, strand mounted node enclosure, or MDU building. The PON extender does not contain the OLT MAC and PHY, making this a centralized access architecture. The PON extender architecture may allow the OLT at the facility to use Coarse Wavelength Division Multiplexing (CWDM) or Dense Wavelength Division Multiplexing (DWDM) optics to the PON Extender. Then the PON extender may perform an optical-to-electrical-to-optical (O-E-O) conversion and then standard PON wavelengths are used at the customer facing side of the PON Extender, as shown in Figure 2.

Some of the drivers and benefits found in deploying the remote OLT architecture can be achieved with the PON extender architecture, but without placing the OLT PON MAC and PHY and even higher layers in the outside plant. There are architecture and performance differences between PON extender and remote PON MAC and PHY architectures. A paper examined some of these differences between PON extender and remote OLT architectures titled *"Remote PON Network Performance"*, by Edward Boyd, et.al., published at the 2015 NCTA / INTX conference. Our paper also includes some analysis work of the PON extender architecture and more work is currently underway to improve the capabilities.

There are two types of centralized access architectures for PON. The first is the traditional approach of placing the PON OLTs in the service provider's facility and having a completely passive OSP / ODN, as seen in Figure 1. In Figure 2, the PON MAC remains in the facility, like the headend, and a PON extender is used in the optical distribution

network (ODN). The PON extender may be used to increase the subscribers per PON OLT port while also saving space in the facility, increase the serving area distance between OLT and customers, and also to increase the number of PON links per fiber, from 1 to perhaps 16 or more.

Centralized Access Architecture (CAA) – PON attributes are listed below:

- OLT MAC and PHY are located at the Facility
- Complex Systems are not in the ODN / OSP
- ODN may be completely passive (no actives)
- ODN may have an active for a transparent O-E-O conversion of WDM optics to PON optics called a PON Extender
    - PON Extender creates an Active Optical Network (AON) not a Passive Optical Network

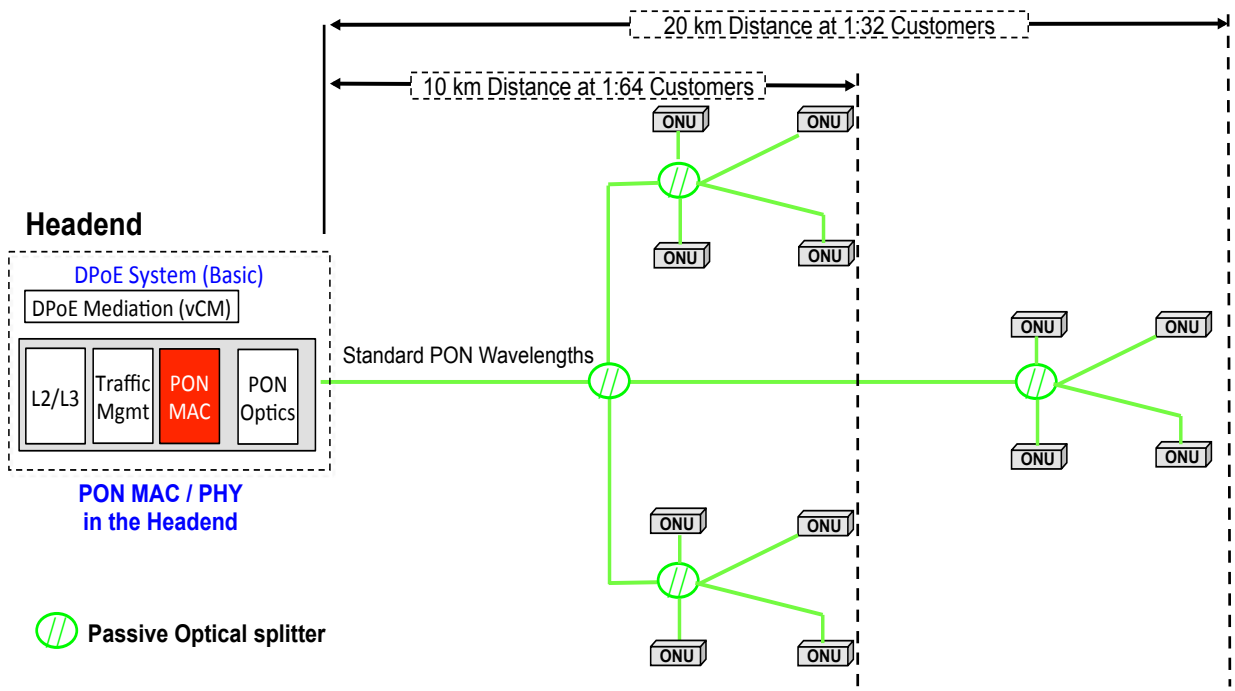


Figure 1 – High-Level – CAA DPoE System

The second CAA for PON keeps the PON OLT (MAC and PHY) in the service provider facility, but uses a PON extender device in the OSP / ODN, as seen in Figure 2. The PON extender architecture estimates a 128 way split at 2 km, however 5 km has also been tested. This illustrates support for multiple 10G EPON service groups or 10G EPON links served from a PON extender. The PON extender uses CDWM or DWDM to carry the equivalent of multiple 10G EPON links to the serving area. The PON extender then connects to different fibers serving different groups of homes or neighborhoods using standard PON optics, this allows all Optical Network Units (ONUs) to use standards 10G EPON optics.

In many ways, the PON extender solves the stacking of multiple 10G PON links using different wavelengths defined in NG-PON2. The NG-PON2 standard currently defines 10G (9.95G) as the maximum upstream and downstream PHY rate between an ONU and

OLT. NG-PON2 requires tunable optics in the ONU to select one of the 4 to 8 wavelengths across the PON.

Currently, a PON extender may support eight (8) 10G EPON links per node and sixteen (16) 10G EPON links per fiber, this is 80 Gbps and 160 Gbps of PON capacity per serving area. The use of different fiber paths to different neighborhoods allows the same color or standard 10G wavelength to be used, thus not requiring tunable optics on the ONU.

The use of a PON Extender (or any of the DAA architectures described in this paper) supports multiple 10G wavelengths to the serving area and then multiple 10G EPON links from the node to the customers using different fibers to different neighborhoods, as seen in Figure 3. Placing an active in the outside plant and fiber topologies to the neighborhood allows 10G EPON to have even more capacities than NG-PON2 at less costs due to the avoidance of tunable optics on the ONU.

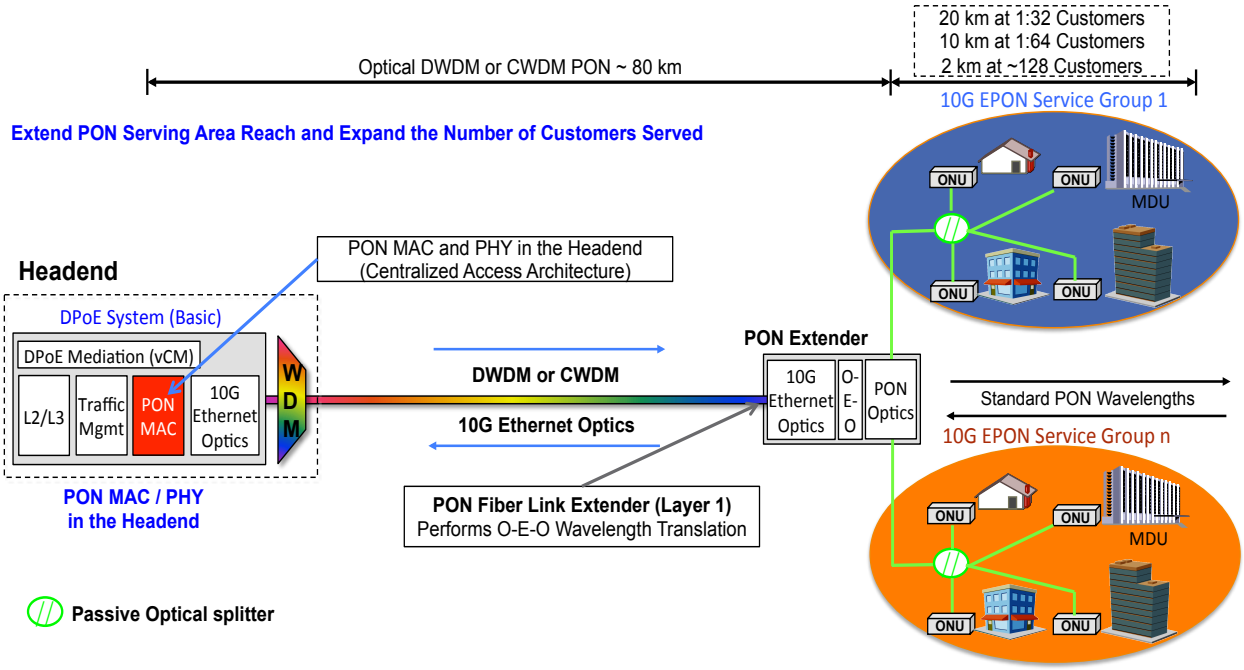


Figure 2 – High-Level – CAA PON Extender Network Architecture

The distributed access architecture for PON places the MAC and PHY of the OLT in the outside plant (OSP) / optical distribution network (ODN) as shown in Figure 3. The attributes of distributed access architecture are listed here.

Distributed Access Architecture (DAA) – PON

- The OLT MAC and PHY are always distributed and located outside an MSO facility in the ODN (node or cabinet)
- ODN definitely has an active device
    - Active Optical Network (AON) is not a Passive Optical Network
- Layer 2 / 3 architecture

- Requires Two (2) Data Transport Networks:
    - Trunk Link: Data Link Between Facility and Node (likely P2P)
    - Line Network: Node to Subscribers uses the PON MAC & PHY as well as PON optics
- Generally Called a "Remote OLT" (R-OLT)
    - This paper defines several system architectures when placing the PON MAC and PHY in the OSP/ODN
    - There may be several product form factors such as a strand-mounted node, field hardened cabinet device, or OLT for the MDU

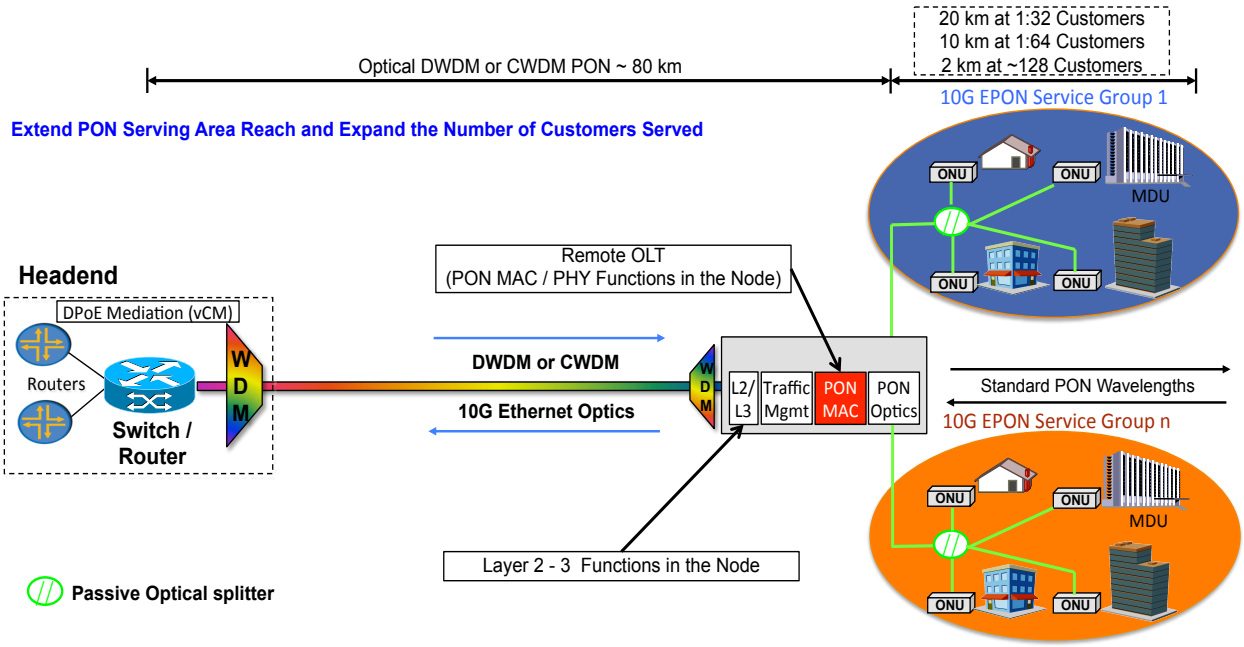


Figure 3 – High-Level – DAA - Remote OLT Network Architecture

The figures above highlight in red the location of the PON MAC and include other system functions in each architecture. In Figure 1, notice that the centralized OLT with a passive optical network, uses an optical wavelength plan defined by the IEEE or the ITU-T. The use of an IEEE defined wavelength plan will allow a fiber to carry a single PON. This means that if a service provider wanted to support more customers than that carried on the single fiber the operator would need another fiber between the facility and serving location. The number of fibers in a given serving area, like a neighborhood, is finite and it is expensive to build additional fibers from the facility to a serving area, thus service providers want to leverage the fibers they have in place. Some service provides may have enough fibers to support the desired number of customers without building new fiber. Many service providers will require the PON network to carry multiple wavelengths between the facility and the serving area to maximize the fiber utilization. Figures 2 and 3 use CWDM or DWDM enabling the existing fibers in the service area to support the customer needs.

This paper will define another architecture similar to Figure 1 that allows 10G EPON to use different wavelengths without requiring a PON extender in the field. This approach may adopt the NG-PON2 wavelength plan for use in 10G EPON OLTs and ONUs. This

paper will discuss several approaches to enable this function. Additionally, this paper will define new system architectures for what was known as remote OLT, this paper proposes that by placing different types of functions in the node these should be classified as different devices, not simply remote OLT.

DRIVERS AND ASSESSMENT FOR NEXT GENERATION PON ARCHITECTURES

The service providers are considering placing active, powered devices, in the optical distribution network to solve several challenges. There is also interest in the future evolution of the network to support a migration to a software defined network (SDN) and also network function virtualization (NFV). This paper also proposes several next generation DPoE system approaches that can support the business drivers listed below. This paper examines some methods that could be adopted by 10G EPON to solve some of these challenges and avoid placing actives in the ODN, referred to as multiple wavelength "Multi-wave" 10G EPON. This paper introduces several next generation DPoE system approaches in the distributed access architecture class.

Business Drivers for the Next Generation DPoE Systems

1. Improve Cost end-to-end (E2E) (OPEX and CAPEX)
2. Maximize Fiber Utilization
3. Increase Serving Area Distance
4. Improve Space and Power Utilization
5. Support Software Defined Network
6. Support Network Function Virtualization

In this paper, we have expanded the business drivers to be part of an overall assessment criteria, as seen in the list below. We have added overall space, power and scaling of the network end-to-end. We have added some performance difference for capacity and latency. The system needs to have support desired availability thresholds, as the system may need to have internal system protection schemes as well as network protection architectures. The complexity tolerance of the system varies in large part by the location of the subsystem functions but also by the number of the devices required to enable complete DPoE system operations. Placing all of the DPoE subsystem functions in the node may reduce the reliability, as there are more hardware components in the node. However, splitting subsystems functions between the facility and node location may increase complexity of the next generation DPoE system. If a network service can be delivered by a completely passive optical network (PON), with no active or powered devices in the ODN or outside plant (OSP), then this is generally preferred by most service providers. If the service provider moves to an active optical network (AON) this is generally done to solve some of the business drivers listed earlier. As with most situations there are tradeoffs as well as other drivers that are considered when selecting a system and network architecture. The assessment criteria define how we approach assessing the current and next generation DPoE system architectures.

Assessment Criteria of Access Architecture

1. Improve Cost end-to-end (E2E) (OPEX and CAPEX)
2. Maximize Fiber Utilization
3. Increase Serving Area Distance
4. Space/Power/Scaling Utilization in HE
5. Space/Power/Scaling Utilization in ODN
6. Support Software Defined Network
7. Support Network Function Virtualization
8. Assess Capacity and Latency Performance Differences
9. Assess High Availability System / Network Differences

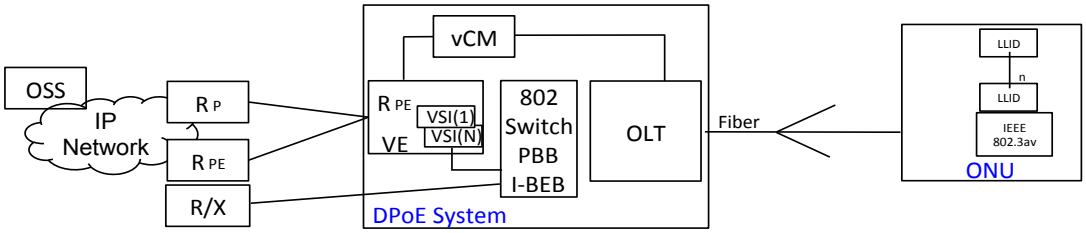
10. Assess System and Operational Complexity Differences

DETAILED DPOE SYSTEM FUNCTIONS

The term DPoE system refers to the set of subsystems within the hub site that provides the functions necessary to meet DPoE specification requirements, according to the DPoE-SP-ARCHv2.0-I04-150611 specifications. The DPoE specifications and perhaps others would certainly need to evolve to support some of the proposed solutions found in this paper. Figure 4 illustrates the high-level set of subsystems that comprise the DPoE system within an MSO hub site or facilities. Figure 5 introduces the functions within the OLT subsystem and renames vCM to DPoE Mediation Layer. These subsystem blocks are important in understanding the types of remote PON architectures, because the location of these subsystems in many way define the type of DAA for PON.

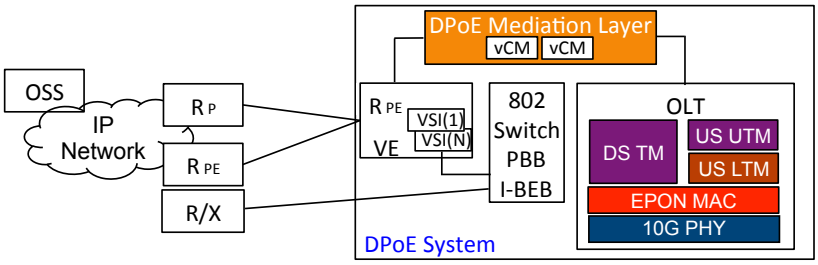
The DPoE specification, referenced previously, defines the Virtual Switching Instance (VSI) functional block shown in Figure 4, Figure 5 and Figure 6. The purpose of the VSI is to support bridging, learning, and forwarding within a layer 2 service instance. The VSI would forward the frames based on standard [802.1d] bridging (unqualified learning via [802.1d]; or qualified learning via [802.1d] + [802.1Q]) behavior to the appropriate pseudowire spoke, VPWS, or local attachment circuit [CableLabs]. The DPoE specification allows the VSI function to be "distributed", contained within the DPoE system or maybe "centralized" not in the DPoE system where the service interface is provisioned. In the centralized bridging and forwarding model the [802.1] bridging/learning occurs on some remote system (i.e., not on the DPoE System where the service interface is provisioned) [CableLabs]. The VSI block and the supporting encapsulation methods like MPLS, shown as Provider Edge (PE) or VPLS Edge (VE) in Figure 4, 5 and 6 may not be part of the DPoE system if a "centralized" forwarding model is used.

It is important to note that the use of "distributed" or "centralized" for purposes of bridge or Virtual Switching Instance (VSI) for the forwarding models of layer 2 traffic should not be confused with the CAA and DAA for PON described in this paper. In other words, the use of the terms centralized and distributed for the forwarding models of layer 2 bridging and learning for a VSI are to be treated separately. The VSI functional block in a "fully distributed model" may be part of any CAA or DAA for PON class or subclass. Later in this paper, the use of term Remote DPoE System (RDS) will be applied to a device outside the facility, such as node, cabinet, or MDU basement, and this device may use the centralized or distributed VSI forwarding models. The Remote DPoE System (RDS) supports either forwarding model just like the hub based DPoE System architecture.



vCM: Virtual Cable Mode
R: Router
PE: Provider Edge
VE: VPLS Edge (Virtual Private LAN Service Edge)
VSI: Virtual Switching Instance
802 Switch: Switch based on Layer 2
PBB: Provider Backbone Bridging [802.1ah]
I-BEB: I-component backbone edge bridge
OLT: Optical Line Terminal

Figure 4 – CableLabs DPoE System Based on DPoE-SP-ARCHv2.0



vCM: Virtual Cable Mode
R: Router
PE: Provider Edge
VE: VPLS Edge (Virtual Private LAN Service Edge)
VSI: Virtual Switching Instance
802 Switch: Switch based on Layer 2
PBB: Provider Backbone Bridging [802.1ah]
I-BEB: I-component backbone edge bridge
OLT: Optical Line Terminal
DS TM: Downstream Traffic Management
US UTM: Upstream Upper Traffic Management
US LTM: Upstream Lower Traffic Management Scheduling / Shaping Algorithms / QoS

Figure 5 – OLT and DPoE Mediation Layer Subsystems

This paper illustrates and defines the DPoE subsystems in more depth and detail. This may include some functions that could be considered out of scope for DPoE 2.0 and also those that others may consider not to be a part of a DPoE subsystem. This detailed view of the subsystems for DPoE will be critical in understanding the functions that could be placed in the remote node or cabinet location. Additionally, the DPoE subsystems can become part of the Software Defined Network (SDN) and also Network Function Virtualization (NFV). Figure 6 provides a detailed view of the subsystems that comprise a DPoE System. After the figure roles and functions of the some the subsystem blocks.

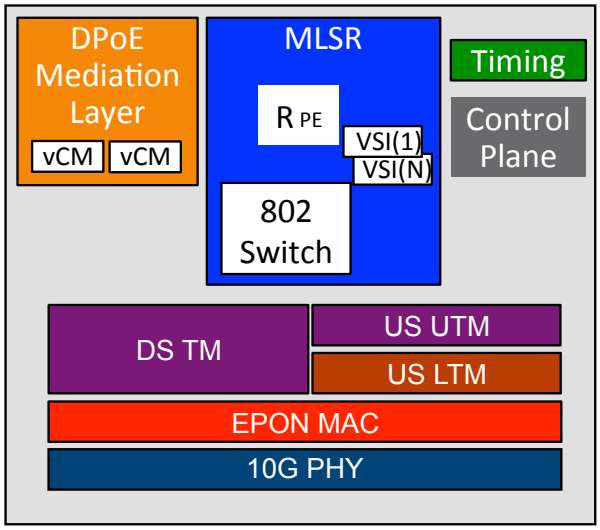


Figure 6 – Detailed Illustration of the DPoE Sub System Functions

Multilayer Switch Router (Control & Data Plane)

- Router (R)
- Address Resolution Protocol (ARP)
- Neighbor Discovery Protocol (NDP)
- Intermediate System to Intermediate System (IS-IS)
- Open Shortest Path First (OSPF)
- BGP (Border Gateway Protocol) only supported IPv4 unicast prefixes
- MP-BGP (Multiprotocol BGP) supports different addresses: IPv4/IPv6 unicast, IPv4/IPv6 multicast, and MPLS VPN where we use MP-BGP to exchange the VPN labels.
- Multiprotocol Label Switching (MPLS)
- Provider Edge (PE)
- Virtual Private Wire Service (VPWS)
- Virtual Private LAN Service (VPLS)
- Virtual Switching Instance (VSI)
- 802 Switch based on Layer 2
- Provider Bridging [802.1ad]
- Provider Backbone Bridging [802.1ah]
- I-component backbone edge bridge
- DHCP Relay
- Interface Access Control Lists

Control Plane Functions

- Centralized Control for all required configurations (e.g. port, channel and MC domain etc.)
- Centralized Control for sending all the collected stats (e.g. port, channel and mac domain etc.).
- EPON MAC for programming all required functionality (e.g. port, channel, LLID and upstream QoS etc.).
- DS TM for programming all required functionality (e.g. service flow, classifier, and downstream QoS etc.).
- Implement the control plane for multicast forwarding

DPoE Mediation Layer Functions

- Mediation between OSS, NMS, and EPON Layers
- Virtual Cable Modem (vCM) for registered D-ONU
- vCM handles all the OAMP functions for DOCSIS
- vCM can proxy requests, signaling, and messages to the D-ONU using EPON OAM messages
- Communicate with D-ONU for provisioning all required services (e.g. IP HSD, MEF etc.)

- vCM Interfaces with System Control Plane for configuration (e.g. service flow, classifier, and downstream QoS etc.)
- Platform Management Interfaces (CLI, SNMP, etc.)

Downstream Traffic Management (DS TM):

- Subscriber Management Filtering (drop)
- Classification & Policing to traffic toward a service flow
- Multicast / Packet Replication
- Scheduling / Shaping Algorithms / QoS Priority of Service Flows (LLIDs)
    - DOCSIS QoS
    - MESP Bandwidth Profile (MEF Metering / Color Marking)
- Packet Counters and Statics (IPDR Raw Data)
- DSCP / ToS / TPID (MEF) Overwrite and Translation
- CoS Mapping
- CALEA / Lawful Intercept
- SF / ASF Encapsulation to VLAN (tunnel ID)

Upstream Upper Traffic Management (US UTM)

- CALEA / LI
- DSCP / ToS / TPID (MEF) Overwrite
- CoS Mapping
- Packet Counters and Statics (IPDR Raw Data)
- Subscriber Management Filtering (drop)
- Cable Source Verify / MAC Learning /Protocol Throttling
- Classification & Policing for forwarding toward the NNI or backplane (aggregate rate limiting)

Upstream Lower Traffic Management (US LTM)

- Scheduling / Shaping Algorithms / QoS Priority of LLIDs (Service Flows)
    - Dynamic Bandwidth Allocation (DBA) – solicited scheduling
        - Token Size per LLID
        - Polling Interval per LLID
        - Scheduling / Queuing Algorithm
    - Unsolicited Scheduling
    - MPCP Processing
    - The figure shows two blocks called US TM, because it is common that this lowest layer function is found with the EPON MAC subsystems but this is an upstream TM function)

EPON MAC Layer

- LLID to VLAN (tunnel ID)
- Operation, Administration, and Management (OAM)
- Multipoint Control Protocol (MPCP)
    - Discovery & Registration
    - GATE Generation
    - REPORT Processing
    - Round Trip Time
    - LLID / Preamble (Frame Formation)
- Encryption DPoE 10G - AES 128 Counter Mode (CTR) is bidirectional
- Encryption SIEPON - GCM-AES 128 is bidirectional

EPON PHY Layer

- FEC
- PR-type PMD (PON Optics)

This is an illustration of the subsystems and possible subsystems that comprise a modern 10G EPON DPoE System. These functions are important to understand as this paper examines future system architectures that may change or split the subsystem functions between the hub site and the node location as well as to the cloud, hub site, and node location. Figure 6 is used to illustrate the subsystem block location of the next generation DPoE system architecture. This paper also illustrates and defines new subsystem functionality blocks needed to support the next generation DPoE systems.

These may include wavelength management for the Multi-wave DPoE 10G System, MACsec for encrypted communications with remote PON nodes and cabinets, new communication methods for provisioning and managing remote subsystems, timing, and more. In this paper, we assume extended DPoE System includes functionalities for DAA architectures such as discovering, authenticating and secure provisioning, maintaining and managing Remote Devices, vCM creation and management, OSS/BSS interfacing. Some of these functionalities may also reside in a separate controller.

OVERVIEW OF THE NEXT GENERATION CAA AND DAA FOR PON OPTIONS

CAA for PON Overview

The centralized access architecture (CAA) is an architecture where the MAC and PHY functions for the access layer are located within a carrier facility. The CAA for DOCSIS may have the MAC and PHY layers in same facility or the MAC and PHY layers could be in separate MSO facilities, for example the MAC core could be in the headend and a PHY shelf in the hub location. The CAA for PON may have all of the DPoE system functions in one facility or may be separated in between headend and hub site as well. The CAA for PON architecture has the OLT subsystem that contains the PON MAC and PHY in the service provider's facility and when the OLT subsystem is placed in the outside plant or non-carrier facility like an MDU this is referred to as distributed access architecture (DAA).

Integrated DPoE System (IDS)

The DPoE specifications did not define the name of the device that includes all of the subsystems in a single platform. In this paper, this is referred to as an Integrated DPoE System (IDS), as shown on the top of Figure 7.

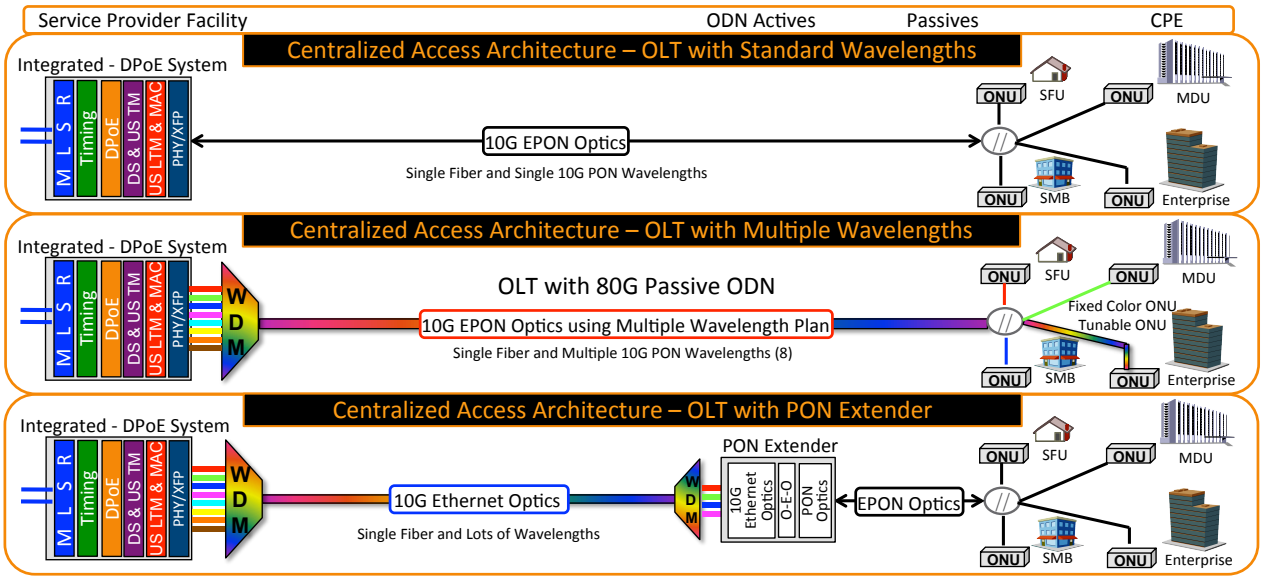


Figure 7 – Centralized Access Architecture (CAA) for PON Options

If the subsystems remained separate the existing platform names would be used, such as OLT, switch, multilayer switch (MLS), router, P-router, PE router, and more. The DPoE mediation layer, that contained the virtual cable modem (vCM) functions, was either integrated into one of the subsystems or a standalone server model. The earliest deployments of DPoE subsystems used the standalone server model. A modern DPoE subsystem architecture has integrated the DPoE mediation layer functions into one or more of the subsystems. As this paper is being published, there is interest in the industry to adopt the server model approach again, this time called, DPoE Mediation in the cloud as part of a SDN architecture.

A 10G DPoE system with MAC and PHY functions using IEEE standard PON wavelength is shown on the top of Figure 7 and implemented as an Integrated DPoE System (IDS), whereby all of the DPoE subsystems are placed in a single platform. This deployment architecture and use of standard 10G EPON wavelengths allow a single 10G PON per fiber, the service provider must have additional fibers to a serving area to provide additional 10G capacity.

Multi-wave Integrated DPoE System (MIDS)

This paper also proposes a Multi-wave Integrated DPoE System (MIDS) as part of the CAA architecture, as shown in the middle of Figure 7. The Multi-wave Integrated DPoE System does not use the standard IEEE 10G wavelengths, but rather suggests the adoption of the NG-PON2 wavelength plan to be use in the 10G EPON OLTs or DPoE Systems. The adoption of the NG-PON2 wavelength plan does not require the use of tunable optics at the ONU, but this is optional in this proposal. This proposal would adopt the NG-PON2 wavelength plan, but not all of the optical requirements for use in the 10G EPON system.

The 10G DPoE system may use a single wave or fixed wavelength solution to the ONUs. There are operational challenges with fixed optics solutions and there are cost improvements under development for tunable optics. There are also operational benefits in having an all passive network, this avoids actives in the outside plant and associated equipment like battery supplies and ongoing plant power and maintenance. The pros and cons of this proposal as well as others are examined later in this paper. The 10G DPoE System that supports multiple wavelengths is called Multi-wave DPoE System. In this paper, because we illustrate all of the DPoE subsystems in a single system, the multiple wavelengths are called the Multi-wave Integrated DPoE System or MIDS.

PON Extender

The PON extender has been used for 1G EPON and GPON systems for many years, and most recently 10G EPON systems, as shown on the bottom of Figure 7. The PON extender increases the serving area distance from the service provider's facility to the customers. The PON extender likely increases the number of customers served per PON port at the facility. The use of the PON extender allows for multiple CWDM or DWDM wavelengths between the OLT in the facility and the PON extender node link and then uses standard PON optics from the PON extender to the ONUs. The PON extender introduces an active component in the ODN, but this is not considered a distributed access architecture, because all of the DPoE subsystems remain in the facility.

The PON extender's use of an optical-to-electrical-to-optical (O-E-O) conversion process places the lowest layer of the PHY in the remote the device, called the Physical Medium Dependent (PMD) sublayer PRX-type PMD (clause 75) [IEEE]. This is the PON optics transmitting at 10.3125 GBd

continuous mode and receiving at 1.25 GBd burst mode, for asymmetric-rate operations (referred to as 10/1GBASE–PRX–D) [IEEE]. A symmetric-rate PMDs (referred to as 10GBASE–PR–D) is possible transmitting at 10.3125 GBd continuous mode and receiving at 10.3125 GBd burst mode [IEEE]. The authors considered referring to the PON extender using O-E-O as a remote PHY option for PON, but since all other PHY functions like the FEC remains in the OLT at the facility we decided to place PON extender in the CAA class.

The PON extender is active component requiring minimal power and no intelligence compared to distributed DPoE subsystem architectures as shown in Figure 7 and described later in this paper. The PON extender is most analogous to the HFC node in the cable world; performing media conversion and minimal processing with no software.

CAA for PON using SDN or NFV

In this architecture higher layer DPoE system functionality may be integrated in the same physical headend shelf or some functionality may be moved to the MSO cloud.

Though in Figure 7, the DPoE subsystems are shown in a single system, we term as an Integrated DPoE System. In the future the DPoE subsystems may use a software defined network (SDN) architecture design and the control and data plane processes may then be separated. The control plane processes may be placed on servers and the control traffic through spine switches. As a next step to evolve to a programmable network and service architecture, the management and configuration functionality may be abstracted and implemented in the SDN controller.

Additionally, using a Network Function Virtualization (NFV) architecture design, both control and data plane processes may be placed on servers and traffic through spine switches. Furthermore, functionalities such as routing/switching, traffic management, and DOCSIS mediation may be virtualized and implemented on the commercial off-the shelf (COTS) hardware and eventually moved to the MSO cloud. An orchestrator is then implemented to manage transactions between these functions and controller. The goal of the SDN and NFV architectures is to reduce capital and operational costs. Additionally, there is a belief that faster delivery of new services through SDN and NFV evolution is possible for all these architectures.

DAA for PON Overview

This paper defines a new class of OLT or DPoE System architecture called distributed access architecture for PON or DAA for DPoE System. The distributed access architecture has defined three subclasses, the Remote PON MAC (RPM), Remote OLT (R-OLT), and Remote DPoE System (RDS). There may be variants within these subclasses of DAA for PON, but in general these define the key subclasses.

As described above, an architecture that uses SDN and NFV may be applicable to these architectures as well, and this is described further in this section. The section above called Detailed DPoE System Functions may be helpful in understanding the location of the subsystem functions as the DAA for PON places some or all of the DPoE subsystems outside the hub and into the ODN. The following describes the three subclasses for the distributed access architecture for PON, shown in Figure 8. As stated above there can be variants within these three classes.

Additionally, the SDN architecture for control plane and NFV for control and data plane can place these functions in the cloud on COTS hardware and software. We have chosen not to show the control plane on

external devices and we have also not shown the systems in the facility that could be part of a cloud based NFV architecture as well. We have separated the functions and these functions may be placed in an SDN or NFV architecture as desired.

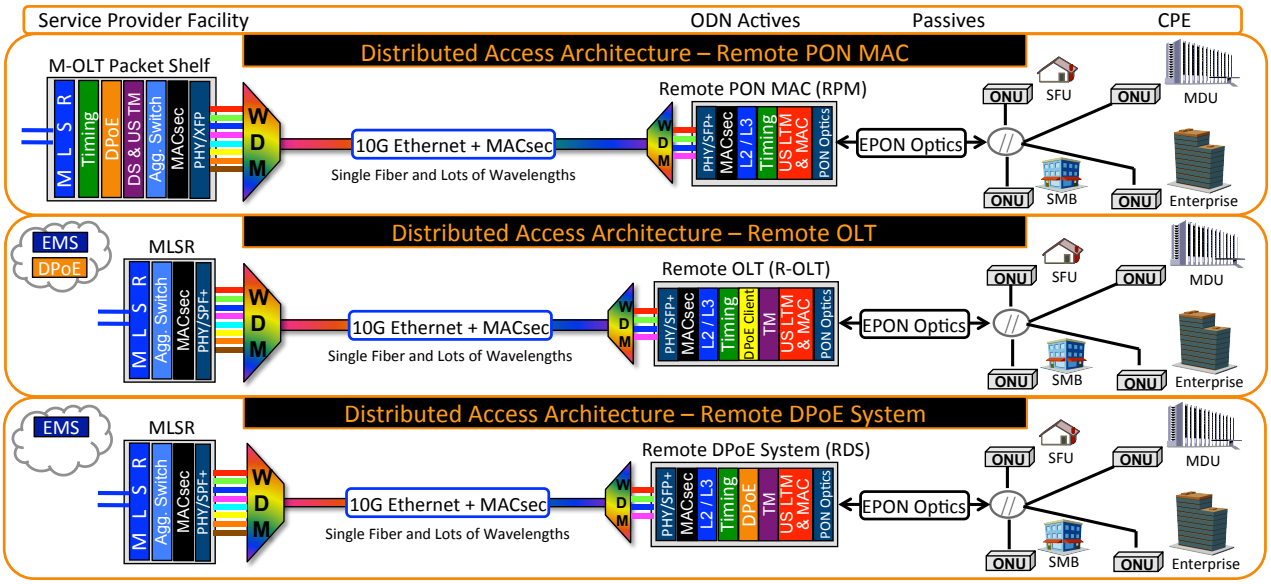


Figure 8 – Distributed Access Architecture (DAA) for PON Options

This paper defines the location of each subsystem functional block from a data plane perspective at minimum. Placing more subsystem functions in the node assumes the carrier is not likely planning to place those functions as part of their NFV architecture today, yet they could be part of a SDN, control plane architecture in the future. This assumes that a service provider would not desire to pay for the hardware, space, and power in the outside plant equipment, node or cabinet, if some functional blocks would later be removed and placed in the NFV cloud. In other words, the carrier would likely use remote subsystem functions that it believes would not be part of NFV architecture.

Remote PON MAC (RPM) and Modular OLT

The Remote PON MAC (RPM) access architecture splits the OLT functions in two parts, with one part in the facility containing the downstream and upstream upper traffic management subsystem and the part in the remote device containing the upstream lower traffic management, PON MAC, and PON PHY subsystem functions. A key concept is that the OLT block, shown in Figure 5, assumes that the PON MAC and PHY as well as the traffic management functions were combined to form the OLT subsystem block as part of the DPoE system. The top of Figure 8 illustrates a device called a Modular Optical Line Terminal (M-OLT) containing all of the downstream traffic management functions and upstream upper traffic management functions in the facility, while the Remote PON MAC (RPM) device contains the upstream lower traffic management, PON MAC, and PON PHY subsystem functions. The Remote PON MAC will also include functions found in the routing and switching block dependent on service provider requirements, but most likely will be configured with a basic IEEE 802 switch.

A remote device defined using the Remote PON MAC (RPM) architecture MUST contain the upstream lower traffic management, PON MAC, and PON PHY subsystem functions. The RPM devices MUST NOT place the downstream traffic management, upstream upper traffic management, and DPoE mediation layer functions in the remote device; these functions are placed in the facility.

The upstream traffic management (US TM) function in the RPM is responsible for upstream scheduling, QoS, shaping, REPORT, GATE, and LLID generation. This small portion of the upstream traffic management is typically part of the system on a chip (SoC), which includes the PON MAC and PHY. The remaining upstream upper traffic management functions as well as all of the downstream traffic management functions that comprise the Modular OLT (M-OLT) subsystem were covered in the sections titled, detailed DPoE system functions.

If we consider that the DPoE subsystem, shown in Figure 5 contained a box called OLT, then the M-OLT and RPM architecture splits the OLT function in two pieces. The Remote PON MAC system architecture, shown in Figure 8, combines the Modular Optical Line Terminal (M-OLT), which has the downstream, and upstream traffic management subsystem with all of the other subsystem blocks to be called a Modular Optical Line Terminal (M-OLT) Packet Shelf.

The EPON MAC functions for the downstream and upstream as well as a portion of the upstream traffic management functions are placed in the remote PON MAC (RPM), this could be a node or cabinet form factor. In this architecture the majority of the software and packet processing are handled at the facility and in this case the M-OLT Packet Shelf, but this could be cloud-based system in part or in whole. The M-OLT packet shelf in this system architecture contains the DPoE mediation subsystem as well as element management of perhaps hundreds of RPMs that may be served in the field. In many ways the RPM is an extension of the I-DPoE System (IDS) in that it is extending the PON MAC and PON port from the line card to the node location.

The concept of a modular OLT (M-OLT) subsystem and a Remote PON MAC (RPM) device subsystem, or RPMD is an illustration of a next generation DPoE System. In this case the traffic management functions of the OLT are retained in the service provider's facility (hub, headend, central office, data center, or cloud) and the PON MAC and PHY are placed in the field (node or cabinet) or MDU location. Additionally, new blocks for link encryption are added in the NG DPoE system architecture as well as other functions described later in this paper that are shown at the top of Figure 8.

Remote OLT

The remote OLT access architecture places the entire OLT functions in the remote device, including the PON MAC as well as all of the PON upstream and downstream traffic management functions. The R-OLT architecture does not place the DPoE Mediation Layer subsystem functions in the remote device. The Remote OLT architecture places the DPoE mediation layer functions handling the northbound provisioning system interfaces and vCM in the facility, on a standalone server model or integrated into a network element. The DPoE mediation layer sends control plane communication messages southbound to the R-OLT subsystems. This may require a thin client for processing and parsing of these control messages in the R-OLT as seen in the middle of Figure 8. The Remote OLT will also include functions found in the routing and switching block dependent on service provider requirements, but most likely will be configured with a basic 802 switch.

The Remote OLT (R-OLT) MUST place downstream and upstream traffic management functions, PON MAC, and PON PHY in the remote device. The R-OLT MUST NOT place the DPoE mediation layer functions, such as Virtual Cable Modem in the remote device; this function is placed in the facility.

This architecture also introduces an element management system (EMS) to manage the five core areas of the Telecommunications Management Network (TMN) model key functions of fault, configuration, accounting, performance and security (FCAPS). The northbound EMS interfaces to network management systems SNMP, CLI, IPDR, authentication and access, and inventory systems. There is ongoing work to define YANG models with NETCONF protocols to configure/manage remote devices. Although the parts that are configured change for the three DAA architectures, YANG models may be applied for different options. The southbound interfaces communicate with remote OLT network elements.

Remote DPoE System (RDS)

The Remote DPoE System (RDS) access architecture MUST place the entire OLT subsystem as well as DPoE Mediation layer subsystem in the remote device. These are the two key building block that separate Remote OLT and Remote DPoE System. The Remote DPoE System will also include functions found in the routing and switching block dependent on service provider requirements. This architecture suggests an element management system (EMS) to manage the RDS devices in the same manner as described in the R-OLT description immediately proceeding.

DAA for PON using SDN or NFV

The Modular Optical Line Terminal (M-OLT) Packet Shelf is shown as a single shelf. However, using a SDN architecture design the control plane processes may be separated. The control plane processes placed on servers and the control traffic through spine switches. The data plane may be on purpose built systems. Additionally, using a Network Function Virtualization (NFV) architecture design, both control and data plane processes may be placed on servers and traffic through spine switches. In regards to R-OLT is shown using NFV for the DML only and SDN may be used as well. The PON MAC and a portion of the upstream traffic management functions of the data plane are placed in the RPM node and these cannot be moved to the cloud in this configuration. The entire PON MAC and traffic management functions of the data plane are placed in the R-OLT node and these cannot be moved to the cloud in this configuration.

Finally, the Remote DPoE system (RDS) places all of the data plane functions in the node and may use SDN for control plane functions for routing, switch, and other functions. In the R-OLT and RDS architecture Network Function Virtualization to the cloud are simply very limited or not at all because the service providers decided to place the data plane functions in the node and may only implement SDN for the control plan functions to the data center.

In examining the DAA options for PON if NFV is planned in a significant way, then the top approach in Figure 8, the RPM access architecture would be desirable. If there is limited NFV planned, then investment in R-OLT and RDS would be reasonable. If the lowest amount of software and configuration is desirable for the remote PON device, while staying in the DAA class, then the Remote PON MAC (RPM) would be choice.

EVALUATION OF THE NEXT GENERATION CAA AND DAA FOR PON OPTIONS

CAA – OLT with Standard Wavelengths

In this CAA for PON, standard 10G EPON wavelengths are used per IEEE 802.3bk specifications as shown in Figure 9. Table 1 summarizes supported splitting ratio and reach lengths for different power budget classes that meet channel insertion loss and dispersion requirements. The maximum nominal distance is 20 km with maximum split ratio 1:64. Depending on the ODN implementation, for a given split ratio, longer or shorter distances may be achieved. Typical deployments are based on PR30 class while PR40 requires new development for the optics.

The OLT PHY and MAC functionality is implemented in special purpose hardware in the headend or hub. This architecture is an extension of first generation 1G EPON technology with proven deployment. Incrementally SDN and NFV may be adopted to dynamically program network devices and services, deploy COTS hardware and move network function workloads between hardware platforms.

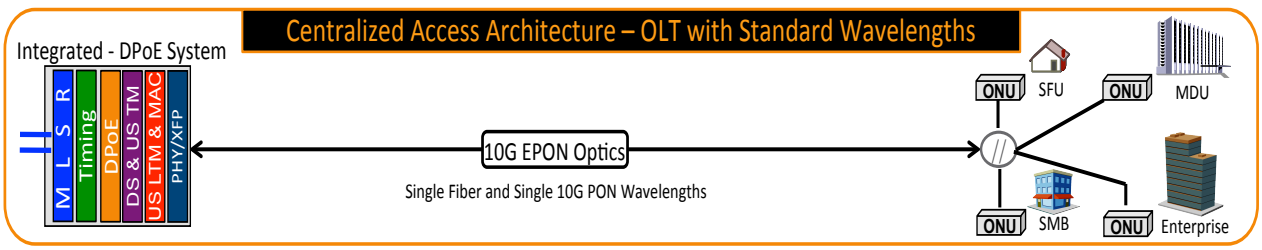


Figure 9 – CAA – OLT with Standard Wavelengths

	10G/1G-EPON	***10G/10G-EPON***
PON Rate (Gbps)	10.3125G/ 1.25G	10.3125G/ 10.3125G
DS λ (nm)	1575-1580	1575-1580
US λ (nm)	1310 ±50 Or 1310 ±20	1270 ±10
Split Ratio	1:64	1:64
Max Reach (km)	PRX10: ≥10 (1:16) PRX20: ≥10 (1:32), ≥20 (1:16) PRX30: 20 (1:32) PRX40: 20 (1:64)	PR10: ≥10 (1:16) PR20: ≥10 (1:32), ≥20 (1:16) PR30: 20 (1:32) PR40: 20 (1:64)

Table 1 – System Parameters for CAA – OLT with Standard Wavelengths

The major benefits with CAA – OLT with Standard Wavelengths are its technology maturity, lower MTTD/MTTR and operational simplicity thanks to the passive ODN with no space/power requirement in the OSP. Integrated DPoE System functionality can be implemented and operated with current standards and operations practice without requiring an extension to manage distributed architecture devices. However, this architecture has market limitation due to its shorter serving area distance (20 km) and has a higher cost per customer due to its smaller serving group size (1:32/64). Single EPON wavelength per fiber requires additional fiber installation in the ODN to each serving area.

For example, if an MSO built a 500 HHP node architecture during a rebuild, this may have been built with a 2 fiber or 6 fiber node architecture. This meant that the number of fibers allocated to that serving group between the headend and node were designed for between 2 and 6, some MSOs may have more. If the MSO would like to use a CAA with standard 10G EPON optics the fiber constraint and currently occupied wavelength can limit the number of PON subscribes served with exiting fiber. Limited facility consolidation and higher space/power requirements of an integrated headend system are other disadvantages as listed below.

Pros:

- **Location of DPoE subsystem hardware:** 100% in the conditioned facility
- **Location of DPoE subsystem software:** 100% in the conditioned facility
- **SDN Architecture Support:** Yes Completely
- **NFV Architecture Support:** Yes Completely
- **Operations MTTD/MTTR:** Lowest
- **Plant Actives:** None
- **Plant Space/Power:** None (no actives, no batteries, no electricity, no maintenance ever)
- **Optical Impact:** All CPE use same optics
- **Number of Optical Ports in the Access Layer:** 1

Cons:

- **Fiber Utilization:** single wavelength per fiber may have no impact when service take rates are low. Some MSOs may not have enough fiber strands between headend and node serving areas for full scale PON deployment.
- **Serving Area Distance:** up to 20 km and will not cover entire MSO market and will rely on PON Extender for full coverage
- **Facility Consolidation:** precludes massive headend consolidation (20 km limit)
- **Facility Space/Power:**
    - 2 or 4 times more ports (assuming 64 or 32 subs per port) than PON Extender or Non-blocking RPM assuming 128 subs per port.
    - 4, 8 or 16 times more ports than RPM, R-OLT, and RDS if blocking / oversubscription of the node uplink is allowed
- **Cost Per Customer:** CAA PON reduces PON subs to 32-64 per port increasing costs per sub

Assessment

MSOs can deploy and operate CAA - OLT with Standard Wavelengths effectively as both the ODN optics and headend OLT with DPoE functionality are well-established and mature technologies. All passive OSP with high reliability, easy troubleshooting, and minimal technician training makes this architecture attractive to MSOs. However limited fiber utilization, shorter serving area distance, smaller serving groups, high space, and power requirements in the facility all limit

the market deployment cases severely for this architecture. Fiber utilization, reach, and split levels may be extended by using NG-PON2 wavelengths and/or a PON extender as described in the next sections.

CAA – OLT with Multiple Wavelengths

In the CAA – OLT with Multiple Wavelengths, we are proposing that 10G EPON system could use the NG-PON2 wavelength plan, as shown in Figure 7. Similar to CAA – OLT with Standard Wavelengths, OLT PHY and MAC functionality is implemented in the facility and there is no active component in the OSP.

Table 2 summarizes supported splitting ratio and reach lengths for different power budget classes that meet channel insertion loss and dispersion requirements. The maximum nominal distance is 20 km with maximum split ratio 1:64 for 40 Gbps DS and 10 Gbps US and may support 40 km with maximum split ratio 1:32 for 10 Gbps per DS and 10 Gbps per US channel. The standards requirements include up to 1:256 split ratio. Depending on the ODN implementation, for a given split ratio, longer or shorter distances may be achieved. Similar to other PON technologies, NG-PON2 standards include higher reach lengths (i.e. reaching 60 km) with reach extenders if needed. The standards suggest that longer reach (e.g., 60-100 km) NG-PON2 systems could facilitate CO consolidation and other network architectures and capabilities.

NG-PON2 (ITU G.989) is based on multiple wavelength channel TWDM architecture over a single fiber. 4-8 TWDM channel pairs may be deployed with each channel pair comprising one DS and one US wavelength. In this section, we consider only 10 Gbps per channel option. The ONU lasers and receivers are tunable and assigned a wavelength based on a dynamic wavelength assignment and management functionality. Although static assignment with cheaper non-tunable ONU optics (i.e. fixed color) is possible, large number of ONUs may be impractical to manage. The likely solution for fixed optics to work is to define the area with specific wavelength upfront and use the BSS or OSS to inventory service ability and the wavelength assignment. The next solution would likely scale best with pluggable so that a standard ONU CPE device could be used and the pluggable with the fixed wavelength is selected for the inventory of parts on the truck.

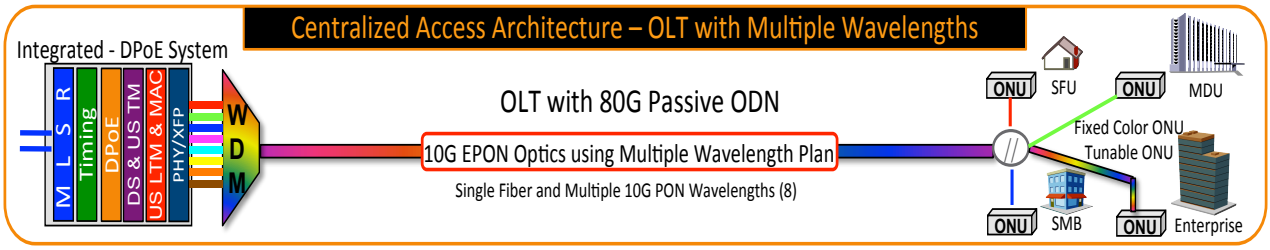


Figure 10 – CAA – OLT with Multiple Wavelengths

	NG-PON2
PON Rate (Gbps)	10.3125G/ 10.3125G
DS λ (nm)	1596 - 1603
US λ (nm)	wide: 1524-1544 reduced: 1528-1540 narrow: 1532-1540
Split Ratio	1:256
Max Reach (km)	20km (1:64 and 40G DS/10G US) 40km (1:32 and 10G DS/10G US) Up to 1:256 N1:PR30; N2; E1=PR40; E2

Table 2 – System Parameters for CAA – OLT with Multiple Wavelengths

This architecture has similar advantages with CAA – OLT with Standards Wavelengths option. In addition, multiple wavelength options using Time and Wavelength Division Multiplexing (TWDM) architecture over a single fiber improves fiber utilization without the need of a PON extender or remote OLT device. In table 2, the NG-PON2 standards support 9.953G/9.953G, and when NG-PON2 optics are used for 10G EPON networks, the rate is 10.3125G.

NG-PON2 is a new technology and very few optics vendors have products today. The power budget support of all 4 classes (N1, N2, E1, and E2) require high transmit power at the ONU and high receiver sensitivity at the OLT for upstream. The use of a SOA in the ONU and OLT optics increases the cost significantly. To expedite the use of NG-PON2 core technologies for 10G/10G, a new standard, XGS-PON (ITU-T G9807.1) was developed using fixed wavelength defined by the IEEE 10G EPON standard in 2009 (i.e. 1577 nm DS/1270 nm US). In the meantime, optics suppliers are examining cheaper NG-PON2 optics designs.

There are operational challenges with fixed optics solutions because these will have to be deployed so that operations can manage different wavelength or color ONUs. There are also operational benefits in having an all passive network. It avoids actives in the outside plant and associated equipment like battery supplies and ongoing plant power and maintenance. The use of tunable optics shares the benefits of an all passive network and avoids the operational management of tracking and allocating specific colored ONUs. Tunable optics come at a steep cost to support a tunable range of four wavelengths or more. The use of NG-PON2 wavelengths may increase the maximum reach and split support while keeping OSP all passive; this will be examined later in this paper.

Due to current limitation on the optics technology to support higher split ratios and reach lengths, limited serving area distance and facility consolidation remain as disadvantages. Current NG-PON2 optics prices increase the cost per customer at the ONU and OLT. The use of NG-PON2 optics can support greater reach and greater splits, thus increasing the number of subscribers per port when compared with traditional 10G EPON wavelengths. In most cases PON

extender or DAA for PON architectures that place the PON optics in the neighborhoods and close to the customers will enable more customers per port. The Multi-wave DPoE System does not require actives in the outside plant to enable more links of 10G EPON on the fiber, which is an advantage of this architecture. The pros and cons are list below:

Pros:

- **Location of DPoE subsystem hardware:** 100% in the conditioned facility
- **Location of DPoE subsystem software:** 100% in the conditioned facility
- **SDN Architecture Support:** Yes Completely
- **NFV Architecture Support:** Yes Completely
- **Operations MTTD/MTTR:** Lowest
- **Plant Actives:** None
- **Plant Space/Power:** None (no actives, no batteries, no electricity, no maintenance ever)
- **Number of Optical Ports in the Access Layer:** 1
- **Solves Fiber Utilization:** using WDM at the OLT solves fiber utilization challenge of traditional PON and the driver to place actives in the OSP/ODN (PON Extender or remote OLT) enabling service providers to leverage the fiber they already have between facility and node serving area.
- **Operational Benefits:** supporting multiple wavelengths per fiber and having an all passive network save operational costs

Cons:

- **Serving Area Distance:** up to 20 km and will not cover the entire MSO market and will rely on PON Extender for full coverage; 40 km may be supported but the split ratio is reduced by half to 1:32 and the optics' price increase
- **Facility Consolidation:** precludes massive headend consolidation (20/40 km limit)
- **Facility Space/Power:**
    - 2 or 4 times more ports (assuming 64 or 32 subs per port) than PON Extender or Non-blocking RPM assuming 128 subs per port.
    - 4, 8 or 16 times more ports than RPM, R-OLT, and RDS if blocking / oversubscription of the node uplink is allowed
- **Cost Per Customer:** CAA PON reduces PON subs to 32-64 per port increasing costs per sub. Optimizing fiber utilization and keeping the ODN a "PON" adds costs to optics OLT/ONU.
- **ONU optics:** Will increase ONU (and OLT) optics cost, volumes will reduce cost but not to the 10G EPON level, tunable ONU enable flexibility at high capital cost or fixed color are lower in cost but will impact operations and require management of colored devices
- **Operational Challenges:** management of different wavelength or color ONUs can be costly

Assessment

CAA with NG-PON2 wavelengths improves the fiber utilization by using hybrid time and wavelength division multiplexing channels on a single fiber between the Facility and Splitter node. Therefore, this idea may reduce the market size for PON Extender and remote OLT. However, the biggest challenge today is NG-PON2 optics cost at both the ONU and OLT compared to nearly all other solutions. Aligning 10G EPON TWDM and NG-PON2 TWDM will drive volumes and will reduces cost but the current SOA design will likely not reach the 10G EPON single wavelength level.

We are exploring many different CPE optical wavelength plans and non-tunable

solutions to reduce cost. Customer premise equipment with PON tunable optics increases one-time capital cost of the solution, and also yields an all-passive network for operations. The single wave may reduce capital but will impact operations due to management of different colored pluggables or CPE.

CAA – OLT with PON Extender

The PON Extender may be designed with optical-to-electrical-to-optical (O-E-O) conversion and/or Optical Amplification (OA) based wavelength conversion schemes. An example PON extender with transparent O-E-O conversion of WDM optics to PON optics is shown in Figure 11. The CAA-OLT with a PON Extender expands the split ratio, hence serving group area and the reach extends as shown in Table 3. The high optical transport layer reach enables massive facility consolidation. The PON extender has a one-to-one physical interface mapping from the facility DPoE System across the 10G Ethernet optical link to the PON extender 10G Ethernet link to the PON optic port. The customers served on a PON extender PON port are mapped to a specific 10G Ethernet optical up link port and to a port in the DPoE system. Having a single 10G Ethernet up link, while supporting more than one PON link, is not possible.

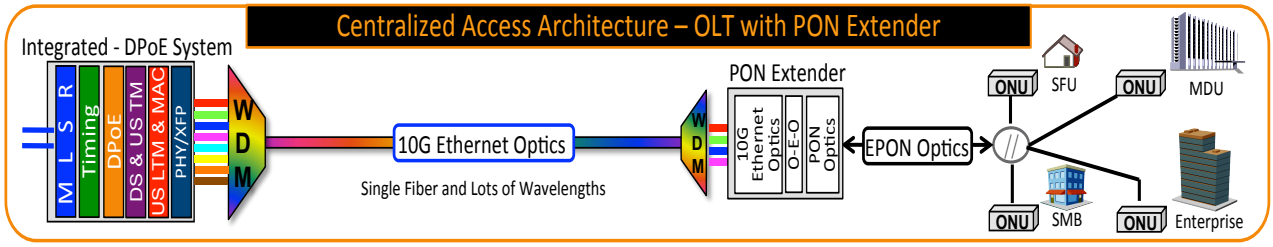


Figure 11 – CAA – OLT with PON Extender

	PON EXTENDER	
	10G/1G-EPON	**10G/10G-EPON**
PON Rate (Gbps)	10.3125G/ 1.25G	10.3125G/ 10.3125G
DS λ (nm)	1575-1580	1575-1580
US λ (nm)	1310 ±50 Or 1310 ±20	1270 ±10
Split Ratio	1:128	1:128
Max Reach (km)	OTL: 80km[1] ODN: 20km[2]	OTL: 80km[1] ODN: 20km[2]
Note 1: OTL reach depends on the design that affect loss (e.g. MUX/DEMUX) and maximum delay Note 2: ODN reach depends on the split ratio;		

Table 3 – System Parameters for CAA – OLT with PON Extender

The trunk fiber between the OLT in the headend and the PON extender in the node is maximized with the use of a PON extender because each PON port may use a CWDM or DWDM band enabling a single serving area to support multiple 10G PON ports on a single fiber. Due to the PON optical ports proximity to the customers the subscriber count per port is high, this reduces the number of PON port needed and 10G Ethernet wavelengths.

The PON extender may support 10G/10G or 10G/1G EPON. Many MSOs do not prefer co-existence of 1G and 10G on the same PON port due to the throughput degradation and lack of support of 10G and 1G on the same PON extender port is acceptable.

The PON Extender is installed and powered from the plant (e.g. in a cabinet, strand mounted node enclosure, or a facility). Power consumption is small compared to distributed architectures. Although additional optics is introduced, lowest cost Ethernet optics may be used at the OLT and the PON Extender NNI when no fan out fiber is used. The extender relaxes the transmitter optical power and receiver sensitivity requirements resulting in lower ONU and OLT optics cost.

The extender is interoperable with multi-vendor OLT and ONUs. Although it is an active component in the OSP, increasing operations MTTD/MTTR and requires maintenance, the PON Extender is not as complex as distributed OLT nodes. Minimal technician training and troubleshooting are required compared to remote architectures. The list of pros and cons including facility power/space saving are listed at the conclusion of this section.

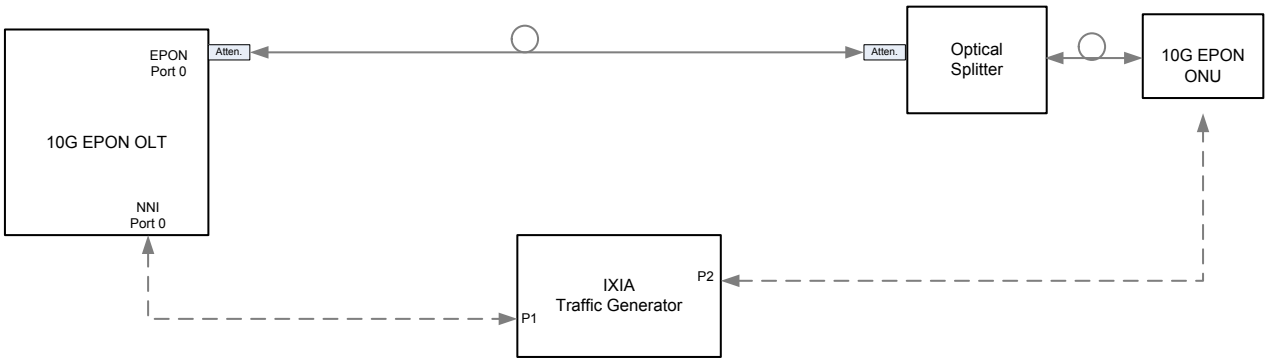
PON Extender Analysis

The Dynamic Bandwidth Allocation in 10G EPON standards was originally designed and tested for reach length < 20 km. The PON extender trunk distance increases the round

trip time of the link. The extra delay may have a negative impact in the achievable throughput depending on the number of LLIDs and traffic demand and requirements for each LLID. Since ONUs have fixed link queue sizes, if the corresponding LLID has a very high traffic load, the extra delay may cause packet drops.

There are different proposed algorithms to improve the efficiency of long reach PON networks such as multi-thread and real time polling schemes. We are in the process of identifying the delay and jitter issues for different use cases in realistic environments and testing DBA modifications. The key parameters are algorithm efficiency, complexity, reuse of value-added features of widely available chips and support of standard ONUs.

As an example, we performed tests per the setups shown in Figure 12. The maximum achievable US throughput (~8.7 Gbps) depends on total overhead from FEC, burst mode transmission, guard band, idle prefix, codeword quantization, time quantum rounding, and control messages. First we tested a single LLID case with US and DS traffic for both systems with 0km to show that ~8.7 Gbps (at ODN link level) is achievable in both systems with appropriate SLA parameters, optical overhead, report thresholds and traffic load.



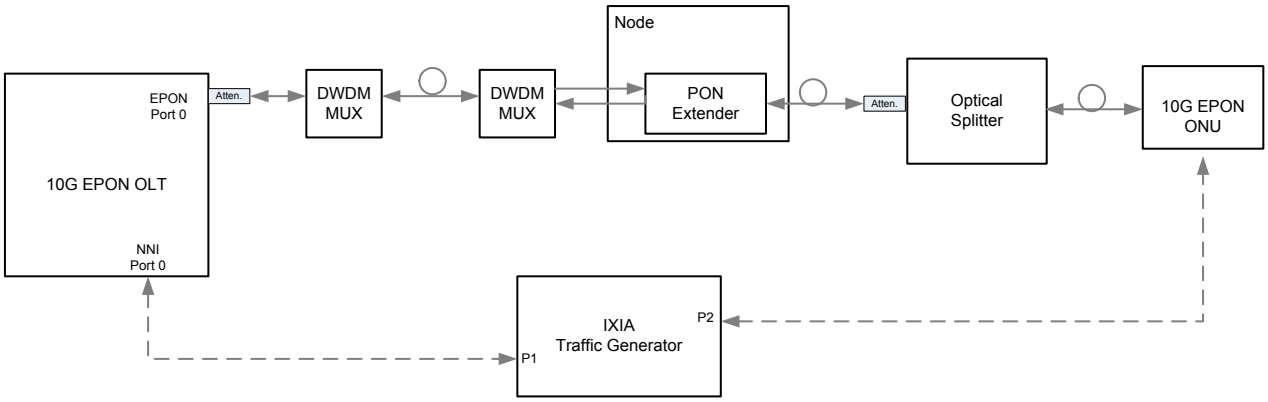


Figure 12 – Top: CAA – OLT with Standard Wavelengths
Bottom: CAA – OLT with PON Extender

We then kept the ODN distance to 2 km and increased the OLT trunk distance from 20 to 80 km for different traffic loads to measure the throughput, delay, and jitter performance. We also tested the system with unsolicited scheduler (e.g. UGS), which is available today in widely used EPON chips for both ONUs and OLTs. The DPoE system does not prevent the use of UGS schemes, but the current standard does not define its configuration and dynamic UGS SF provisioning. As also stated in SIEPON specifications unsolicited scheduling is suitable for mobile backhaul type applications; this may depend on the PON reach and split architecture. It may be also used efficiently for other traffic cases if the appropriate configuration can be supported. For all distances we tested, maximum achievable UGS throughput was measured with no degradation due to PON Extender to show the impact of increased RTT on the DBA.

For the 40 km OLT distance and 2 km ODN distance, the packet captures from an Ixia traffic generator and the OLT and ONU statistics are analyzed. When UGS scheduler is applied (with length 30720/35412 TQ and interval of 30720 TQ for a fixed packet size of 1460 bytes IP packet and 20 bytes of IPG and preamble overhead), we can observe that 30720 *20 / 1480 = 415 packet bursts are transmitted every ~0.000566 seconds which is the transmission duration of the burst over the link with 8.7Gbps effective throughput). Higher inter-packet delays in Figure 13 define the unsolicited scheduler burst boundaries. This test and other tests done at different distances and system parameters show that PON extender itself does not introduce any degrading delay and/or jitter (Figure 13).

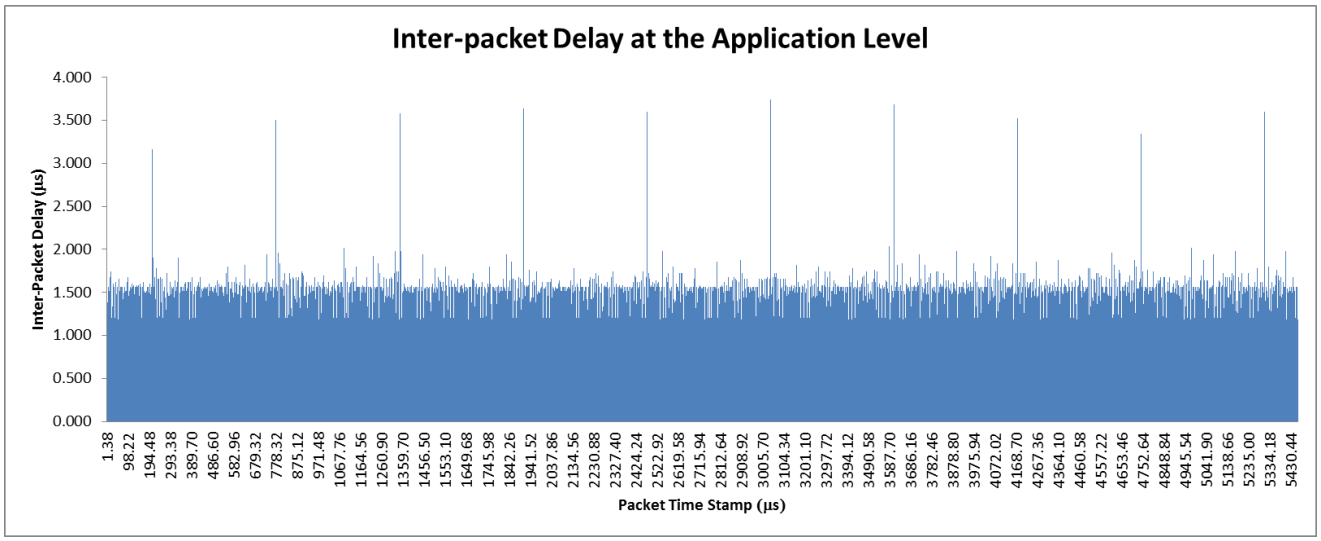


Figure 13 – Inter-packet Delay measured at the traffic generator receive port for UGS scheduler

When we tested the system with original DBA algorithm, we observed ~6.8 Gbps for the given system parameters (i.e. token size and report threshold of 16384 TQ and no rate shaping). Statistics at the traffic generator, ONU and OLT show that packets are dropped at the ONU link queue with maximum size of 1020kB due to increased waiting time (Figure 14, maximum inter-packet delay measured was ~250µs). All other packets are transmitted correctly over the EPON link. We then tested an algorithm that increases the accuracy of grant by taking into account the delay introduced by measured RTT and were able to increase the throughput by more than 1 Gbps (Figure 15). Our preliminary test configurations are made to achieve maximum efficiency levels to isolate the issues while depending on the number of LLIDs and traffic characteristics the performance of different schedulers may change. Our ongoing work will continue to focus on increasing the efficiency of scheduler over the long reach PON cases by considering different distance and split architectures and traffic requirements.

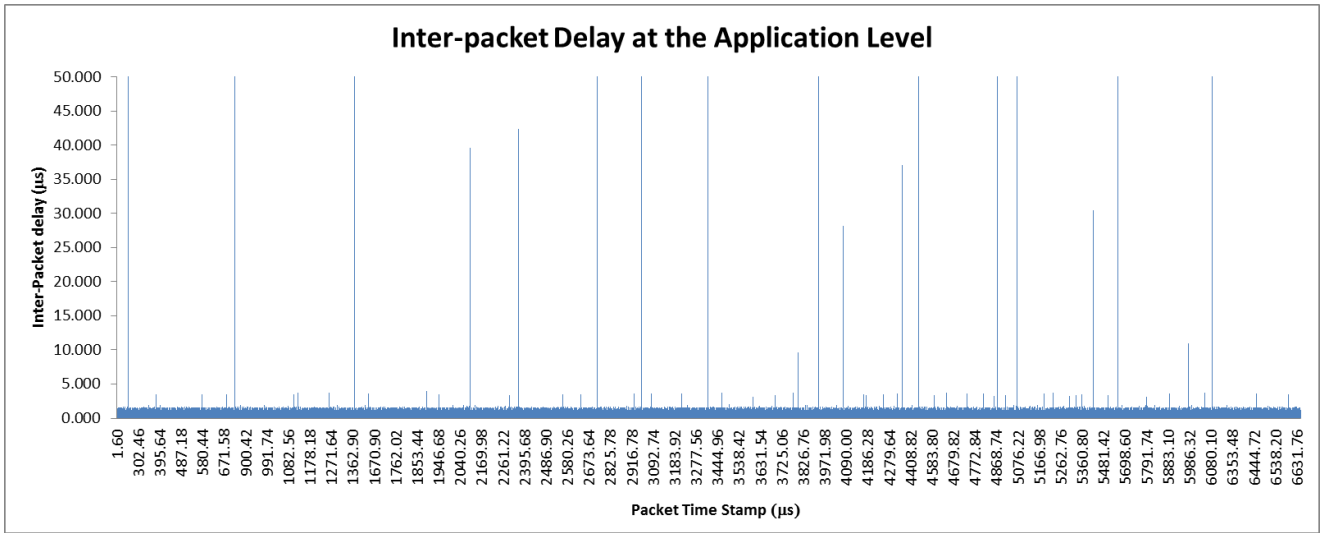


Figure 14 – Inter-packet Delay Measured at the Traffic Generator Receive Port for the Solicited DBA Scheduler

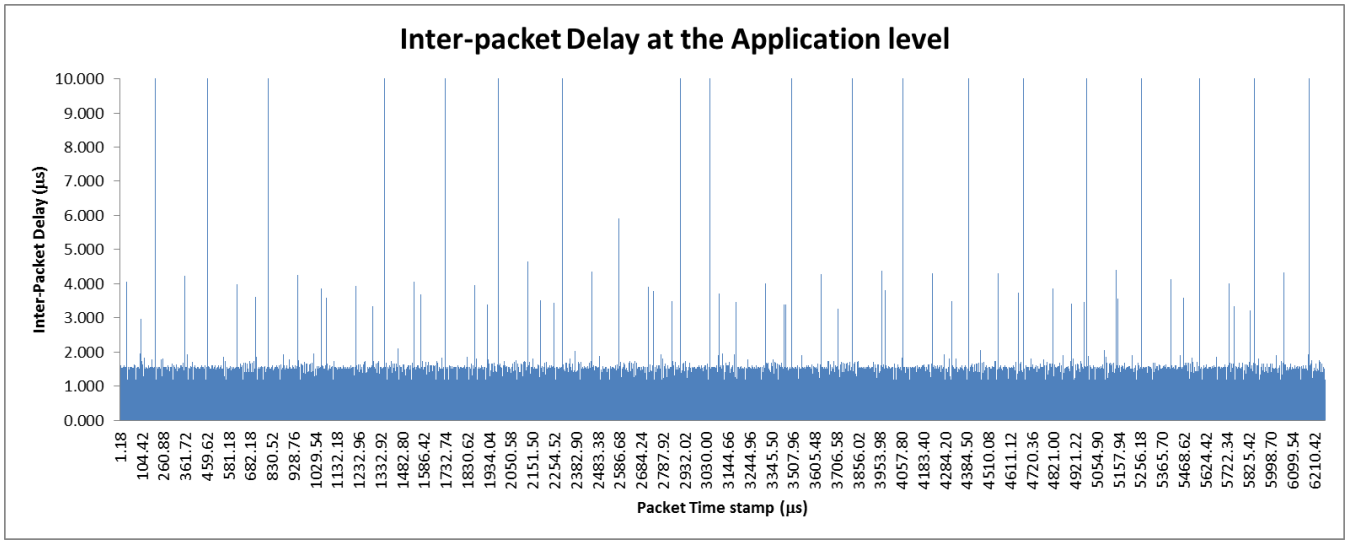


Figure 15 – Inter-packet Delay Measured at the Traffic Generator Receive Port for the Modified Solicited DBA Scheduler

Pros:

- **Location of DPoE subsystem hardware:** 95% in the conditioned facility (PON optics in remote device)
- **Location of DPoE subsystem software:** 100% in the conditioned facility
- **SDN Architecture Support:** Yes Completely
- **NFV Architecture Support:** Yes Completely
- **Plant Actives:** less complex than all DAA solutions
- **Plant Power:** less power than all DAA solutions
- **Plant Space:** less space than all DAA solutions
- **Solves Fiber Utilization:** using WDM at the OLT solves fiber utilization challenge of traditional PON and enabling service providers to leverage the fiber they already have between facility and node serving area.
- **Solves Serving Area Distance:** will cover nearly the entire MSO market up to about 80 km limit (Depending on the subscriber and traffic requirements)
- **Solves Facility Consolidation:** allows massive headend consolidation perhaps distances of 80 km could be reached
- **Solves Facility Space/Power:**
  - Same space is needed in Non-blocking architecture vs. RPM
  - Little more space is needed in Non-blocking architecture vs. R-OLT and RDS
  - Overall there may not be a massive space saving of PON Extender vs. the DAA approaches
- **Solves Cost Per Customer:** CAA PON Extender increases subs to 64-128 per port thus decreasing costs per sub in fiber spans 5-10km from the PON extender
- **Solves ONU Optics:** All CPE use same optics and can be least costly version but depend on split ratio
- **Operations MTTD/MTTR:** higher than passive ODN but less than any of the distributed access architecture.

Cons:

- **Plant Space/Power:** Actives require batteries, electricity, and maintenance forever
- **Facility Space/Power:**
  - PON Extender cannot oversubscribe the 10G link between facility and node
  - Depending on oversubscription ratio there could be a little space increase
- **Capacity and Latency Performance Differences:** The PON extender allows the distance between the OLT and ONUs to expand longer than typical PON networks of less than 20km. The "current" upstream OLT scheduler was not designed for long distances between PON upstream scheduler and the ONUs. Currently as the distance expands between the OLT scheduler and the ONU link capacity will decline, this is not a PON extender architecture issue. The industry is developing methods to improve capacity levels. Some service providers may find sufficient upstream capacity exists and the reduction is worth not placing PON OLT hardware and software in the remote device (node or cabinet)
- **Cost Per Customer:** The PON extender has a 1:1 mapping from the OLT port in the facility across the trunk fiber to the PON extender node (known as the 10G Ethernet trunk link) and this is mapped to a 10G EPON Optic line link to the customers. The trunk link between the facility and PON extender many not be shared with multiple 10G EPON Optic line links, known as oversubscription of the trunk link. The PON extender improves the economics in some CAA architectures, but when compared to the DAA approaches which all can oversubscribe the 10G Ethernet trunk link, such as having more 10G EPON line link capacity then 10G Ethernet trunk line capacity the cost per customer can be a bit higher "if" oversubscription of the trunk link is used by the DAA approaches. Oversubscribing the trunk link is like oversubscribing the backplane lanes or having a blocking architecture. Generally most service providers prefer to have non-blocking backplane architecture between 10G EPON client cards and the fabric / WAN modules.
- **Number of Optical Ports in the Access Layer:** 2 additional Optics (10G Ethernet OLT & NNI at PON Extender Node)

Assessment

PON Extender is an active component in the outside plant making the solution Active Optical Network (PON moves to AON). The distance may support up to 80 km between DPoE System and PON Extender and another 20 km to the customer, depending on the split ratio, wavelength usage (MUX/DEMUX loss) and throughput/delay requirements. The use of WDM optics between the headend and the node improves the fiber utilization significantly. Furthermore, higher split ratio maximizes the PON port utilization. CPE uses the same 10G EPON 802.3av optical wavelengths but with a more relaxed requirement set compared to all passive CAA solutions. The throughput and delay improvements are ongoing work.

DAA – Remote PON MAC (RPM)

The Remote PON MAC (RPM) access architecture splits the OLT functions in two parts. The Modular Optical Line Terminal (M-OLT) Packet Shelf contains all of the downstream traffic management functions and upstream upper traffic management functions.

If we consider that the DPoE subsystem, shown in Figure 5 contained a box called OLT, this architecture splits the OLT function in two pieces. The EPON MAC functions for the downstream and upstream as well as a

portion of the upstream traffic management functions are placed in the Remote PON MAC (RPM), this could be a node or cabinet form factor. In this architecture the majority of the software and packet processing are handled at the facility and in this case the M-OLT Packet Shelf, but this could be cloud based systems. The M-OLT in this system architecture contains the functions of DPoE mediation as well as management of the perhaps hundreds of RPMs in the field. In many ways the RPM is an extension of the I-DPoE Systems in that is extending the PON MAC and port from the line card to the node location.

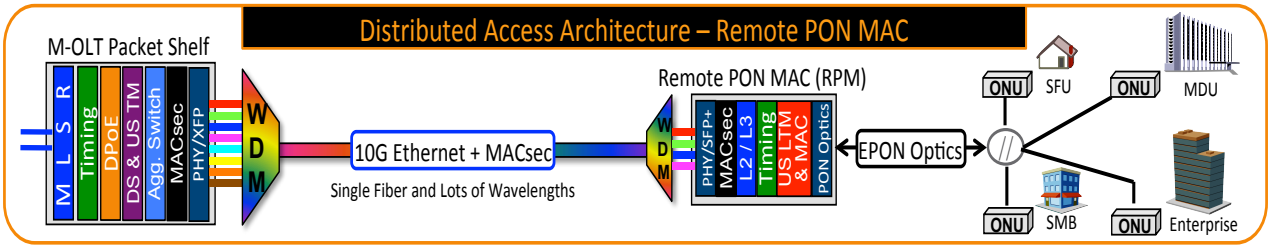


Figure 16 – DAA – Remote PON MAC (RPM)

Pros:

- **Location of DPoE subsystem hardware:** 90% in the conditioned facility (PON MAC, Optics, and basic L2 in then remote device)
- **Location of DPoE subsystem software:** 95% in the conditioned facility
- **SDN Architecture Support:** Yes
- **NFV Architecture Support:** Yes (except for MAC and Upstream lower Traffic Management)
- **Plant Actives:** less complex than R-OLT and RDS solutions
- **Plant Power:** less power than R-OLT and RDS solutions
- **Plant Space:** less space than R-OLT and RDS solutions
- **Solves Fiber Utilization:** using WDM at the OLT solves fiber utilization challenge of traditional PON and enabling service providers to leverage the fiber they already have between facility and node serving area
- **Solves Serving Area Distance:** There is no distance limit because of the DPoE subsystems in the facility and the remote device.
- **Solves Facility Consolidation:** allows massive headend consolidation
- **Solves Facility Space/Power:**
  - Supports oversubscription of the 10G node trunk uplink allowing for greater sub per link and less headend space
  - Requires a little more space in Non-blocking architecture vs. R-OLT and RDS
- **Solves Cost Per Customer:**
  - Increases subs to 64-128 per port non-blocking
  - Costs improve with oversubscribed uplink increasing subs per facility port
- **Solves ONU Optics:** All CPE use same optics and can be least costly version but depend on split ratio.

Cons:

- **Operations MTTD/MTTR:** higher than passive ODN and higher that PON Extender (likely lower than R-OLT and RDS)
- **Number of Optical Ports in the Access Layer:** 2 additional Optics (10G Ethernet OLT & NNI at PON Node)

- **Facility Space/Power:**
  - No or little space savings for Non-blocking RPM vs. PON Extender
  - R-OLT and RDS architectures will enable slightly increased space savings if oversubscription is used

Assessment

The RPM architecture allows the PON MAC and scheduler to be placed in the node or cabinet location. This allows for low latency and high capacity because the MAC is near the subscriber location. Additionally, this architecture will allow the service provider to enable several 10G PONs to different service areas, perhaps 4 groups of 128 subscribers could be supported. The RPM architecture like any of the DAA for PON solution has the flexibility to have the PON customer share the 10G Ethernet node uplinks ports.

This is a differentiation between PON extender and the DAA approaches. A service provider may elect not share the node up links ports between several PON port and just configure a non-blocking architecture, ensuring that a 10G PON port will have a 10G Ethernet uplink port. Many service providers around the world required non-blocking client card capacity across the backplane of a system. The remote PON architectures are really just remote port or remote line cards, and some carriers do not want to manage capacity and blocking. This approach places the least amount of DPoE subsystems functions in the node in terms of hardware and software.

DAA – Remote OLT (R-OLT)

The remote OLT access architecture places the entire OLT functions in the remote device, including the PON MAC as well as most of the PON upstream and downstream traffic management functions. The R-OLT architecture does not place all of the DPoE subsystem functions in the node, most particularly the DPoE mediation layer functions handling the northbound provisioning system interfaces and vCM. The southbound interface we are suggesting a thin DPoE management client for control plane communications with the local functional blocks.

This architecture also introduces an element management system (EMS) to manage the five core areas of the Telecommunications Management Network (TMN) model key functions of fault, configuration, accounting, performance and security (FCAPS). The northbound EMS interfaces to network management systems SNMP, CLI, IPDR, authentication and access, and inventory systems. There is an ongoing work to define YANG models with NETCONF protocols to configure/manage remote devices. The southbound interfaces communicate with remote OLT network elements. This may also apply to RPM architecture for MAC and US LTM and other functionality in the node.

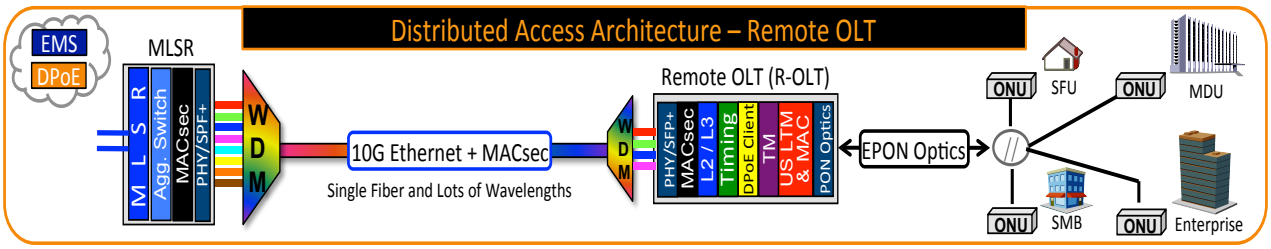


Figure 17 – DAA – Remote Optical Line Terminal (Remote - OLT)

Pros:

- **SDN Architecture Support:** Yes
- **NFV Architecture Support:** Yes (Routing and advanced Layer 2 subsystem only)
- **Solves Fiber Utilization:** using WDM at the OLT solves fiber utilization challenge of traditional PON and enabling service providers to leverage the fiber they already have between facility and node serving area
- **Solves Serving Area Distance:** There is no distance limit because of the DPoE subsystems in the facility and the remote device.
- **Solves Facility Consolidation:** allows massive headend consolidation
- **Solves Cost Per Customer:**
    - Increases subs to 64-128 per port non-blocking
    - Costs improve with oversubscribed trunk uplink increasing subs per facility port
- **Solves ONU Optics:** All CPE use same optics and can be least costly version but depend on split ratio
- **Solves Facility Space and Power:** The requirement for CWDM and DWDM optics in the MLSR aggregation device as shown in the architecture may not yield significant space savings compared to Remote PON MAC

Cons:

- **Location of DPoE subsystem hardware:** 50% in the field unconditioned environment
- **Location of DPoE subsystem software:** 50% in the field unconditioned environment
- **NFV Architecture Support:** The PON MAC, Upstream Lower Traffic Management, Upstream Upper Traffic Management, and Downstream Traffic Management are in the remote device and cannot be part of NFV.
- **Operations MTTD/MTTR:** higher than passive ODN and higher that PON Extender
- **Plant Actives:** more complex than PON Extender and Remote PON MAC
- **Plant Power:** more power than PON Extender and Remote PON MAC
- **Plant Space:** likely more than PON Extender and Remote PON MAC
- **Number of Optical Ports in the Access Layer:** 2 additional Optics (10G Ethernet OLT & NNI at PON Node)

Assessment

The remote OLT is often used to describe the placement of PON functions in the node and as this paper illustrates there can be a vast range of functions that could be placed in the node. The reason why this architecture, and not any other, is referred to as remote OLT is because the entire OLT subsystem in the DPoE reference architecture is place in the node location. This includes the PON MAC, upstream traffic management and downstream traffic management functions.

This does not include the full DPoE mediation layer software functions, so this could not be called a Remote DPoE system as show below. The placement of the PON MAC and traffic management layer functions requires more software in the remote location. Additionally the DPoE mediation layer systems are split to manage the southbound interfaces to the MAC, TM, and switching functions and the northbound and virtual cable modem functions are located in the facility. Additionally, to scale the remote OLT will need an element management system (EMS). The functions of hardware and software are split.

DAA – Remote DPoE System (RDS)

The Remote DPoE System (RDS) access architecture places the entire DPoE system functions in the remote device. This architecture also introduces an element management system (EMS) to manage the five core areas of the Telecommunications Management Network (TMN) model key functions of fault, configuration, accounting, performance, and security (FCAPS). The northbound EMS interfaces to network management systems SNMP, CLI, IPDR, authentication and access, and inventory systems. There is an ongoing work to define YANG models with NETCONF protocols to configure/manage remote devices.

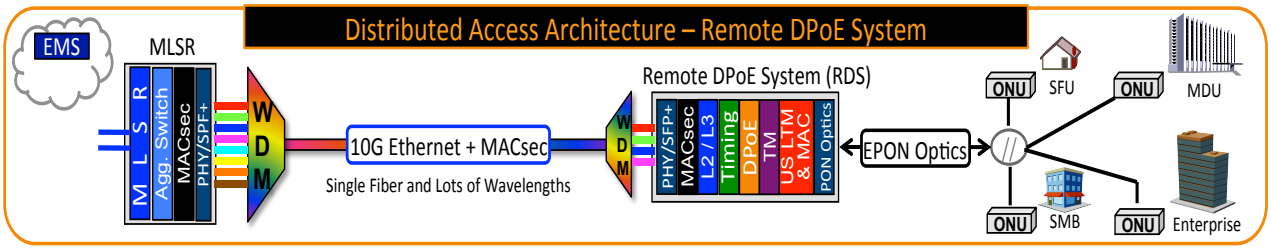


Figure 18 – DAA – Remote DPoE System (RDS)

Pros:

- **Location of DPoE subsystem software:** All of the DPoE subsystems software are together thus reducing the system complexity and trouble shooting
- **SDN/NFV Architecture Support:** Yes
- **Solves Fiber Utilization:** using WDM at the OLT solves fiber utilization challenge of traditional PON and enabling service providers to leverage the fiber they already have between facility and node serving area
- **Solves Serving Area Distance:** There is no distance limit because of the DPoE subsystems in the facility and the remote device.
- **Solves Facility Consolidation:** allows massive headend consolidation
- **Solves Cost Per Customer:**
  - Increases subs to 64-128 per port non-blocking
  - Costs improve with oversubscribed trunk uplink increasing subs per facility port
- **Solves ONU Optics:** All CPE use same optics and can be least costly version but depend on split ratio
- **Solves Facility Space and Power:** The requirement for CWDM and DWDM optics in the MLSR aggregation device as shown in the architecture may not yield significant space savings compared to Remote PON MAC

Cons:

- **Location of DPoE subsystem hardware:** As high as 100% in the field unconditioned environment, but at minimum PON MAC/PHY, upstream and downstream traffic management, DPoE mediation layer, and Layer 2 switching.
- **Location of DPoE subsystem software:** As high as 100% in the field unconditioned environment (assuming all of the routing, advanced L2 functions, VSI, and VPN encapsulation function take place in the remote device) At minimum PON MAC/PHY, upstream and downstream traffic management, DPoE mediation layer, and Layer 2 switching.
- **NFV Architecture Support:** In some Remote DPoE Systems architectures none of the subsystem could be part of NFV. At minimum the PON MAC, Upstream

Lower Traffic Management, Upstream Upper Traffic Management, and Downstream Traffic Management are in the remote device and cannot be part of NFV.

- **Operations MTTD/MTTR:** higher than passive ODN and higher that PON Extender
- **Plant Actives:** more complex than PON Extender and Remote PON MAC
- **Plant Power:** more power than PON Extender and Remote PON MAC
- **Plant Space:** likely more than PON Extender and Remote PON MAC
- **Number of Optical Ports in the Access Layer:** 2 additional Optics (10G Ethernet OLT & NNI at PON Node)

Assessment

The integrated Remote DPoE system is a self-contained platform with all or nearly all of the subsystem supported in the device. The major DPoE subsystems including the MAC, traffic managements, DPoE mediation layer software, vCM, layer 2, layer 3, VSI, and MPLS PE functions. The integrated remote DPoE system contains the mediation layer software. The standalone systems reduces the software complexity because like the I-DPoE System in the hub that contains all of the subsystem functions this can be easier to troubleshoot.

DATA LINK ENCRYPTION BETWEEN SERVICE PROVIDER FACILITY AND THE REMOTE PON DEVICE

We are proposing a new data link encryption requirement between access network component at the edge of the trusted network, such as the M-OLT or switch/router, and the remote PON device, such as the remote PON MAC, remote OLT, or remote DPoE System. This paper suggests that the DAA for PON approach should align with the CableLabs Remote PHY and Remote MAC-PHY specifications for data link encryption and authentication between the remote devices, which is based on MACsec, as defined in IEEE standard 802.1AE MAC Security. The link between the service provider's facility and the remote PON device is referred to as the trunk link or the uplink. This link is where our newly proposed data link encryption and authentication needs to be defined and accepted by the industry.

The remote PON device must be authenticated and authorized, while the control/management plane must be secured. The 802.1X authentication using EAP between the Remote Device (supplicant) and network access port of the switch/router or M-OLT (authenticator) may be implemented. In this case, Authentication Server has the security credentials and communicates with EAP messages with the authenticator. MACsec may provide additional security that defines a virtual port for each connected physical port of remote device. The MKA (802.1x) key management algorithm is used for the exchange of encryption keys (e.g. AES128 based encryption). MACsec is a Layer 2 and hop-by-hop security protocol.

The implementation simplicity and high throughput capability makes MACsec preferable for data link security. The DAA for PON figures used in this paper illustrate a MACsec subsystem, which is part of the M-OLT PS and also part of a multilayer switch router (MLSR) both in the trusted network. In all cases the remote PON devices in the untrusted network also has a MACsec subsystem. If the M-OLT does not have the MACsec subsystem or if it was not directly connected to the Remote PON MAC device, a switch/router enabled with MACsec would be used to enable secure communications.

The PON link security between the OLT subsystem and ONUs is defined in the DPoEv2 and/or the SIEPON security specifications and remains unchanged in this proposal. These include DPoE 10G - AES 128 Counter Mode (CTR) in downstream only or bidirectional methods are available. MKA (802.1X) is used for key exchange when MACsec is implemented for bidirectional encryption. Due to data link encryption and authentication already being defined between the OLT and the ONU, none CAA for PON approaches need additional requirements; this applies to PON extender as well. In addition, the DPoE System/controller must provide secure provisioning and management of the Remote Device.

SUMMARIES AND CONSIDERATIONS

The architectures defined and reviewed in this paper have many trade-offs and considerations for the service provider. The service provider will have specific technology constraints that may limit the selection of the architecture. These architectures represent the next generation of DPoE systems and subsystems. There are several new subsystems as well as communication methods that will need to be determined and defined by the industry. These may include wavelength management for the Multi-wave DPoE 10G System, MACsec for communications with remote PON nodes and cabinets, new communication methods for provision and managing remote subsystems, timing, and more.

In Figure 19, this is meant to summarize the location of the hardware and software functions. The placement of the hardware and software functions may also determine the future SDN and NFV possibilities. In the SDN architecture design the control plane processes may be separated and placed on servers and the data plane may be on purpose built systems, in the table below this describes the possible locations. Additionally, using Network Function Virtualization (NFV) architecture design both control and data plane processes may be placed on servers and traffic through spine switches.

If NFV is planned, then the service provider could examine the table below to determine which architecture aligns with their vision. If the service provider believes that over time all of the DPoE subsystems could be virtualized, then staying to the left side of the table is best, because these systems are not placed in the field but rather facilities to be far easily swapped out. If little or none of the data plane functions are planned to be virtualized then investing in any of the approaches is fine.

The final figure listed at the bottom of this paper is a summary assessment of the different CAA and DAA options. The figure shows that the Remote DPoE System (RDS) has all of the major subsystem functions in the remote device, including the DPoE mediation layer software. As with the hub based DPoE subsystems some Layer 2 and advanced routing and MPLS may be inside the DPoE system or in an external device. Figure 19, uses the term optional with regards to the location may be at the facility or the node/remote device.

Location of Hardware and Software Functions	10G EPON OLT with Standard Wavelengths	10G EPON OLT with NG-PON2 Wavelengths	PON Extender	Remote PON MAC (RPM)	Remote OLT (R-OLT)	Remote DPoE System (RDS)
Element Mgmt. Sys (EMS), Controller, AGM	Facility	Facility	Facility	Facility	Facility	Facility
L3VPNs MPLS L3	Facility	Facility	Facility	Facility	Facility	Optional
L2VPNs VPWS/VPLS with MPLS Encapsulation	Facility	Facility	Facility	Facility	Facility	Optional
Advanced L2 Virtual Switching Instance (VSI)	Facility	Facility	Facility	Facility	Facility	Optional
Advanced L2 802.1ah (MAC-in-MAC)	Facility	Facility	Facility	Facility	Facility	Optional
Routing Protocols (BGP, OSPF, IS-IS, BFD)	Facility	Facility	Facility	Facility	Facility	Optional
DPoE Mediation Layer Functions	Facility	Facility	Facility	Facility	Facility	Node
MACsec (802.1ae) facility & remote	N/A	N/A	N/A	Facility & Node	Facility & Node	Facility & Node
NNI and PON Interface ACL (ingress & egress)	Facility	Facility	Facility	Facility & Node	Facility & Node	Node
Layer 3 Aware (DHCP Relay)	Facility	Facility	Facility	Facility & Node	Facility & Node	Node
Layer 2 Switch with Multicast & Link Agg	Facility	Facility	Facility	Facility & Node	Facility & Node	Node
Downstream Traffic Management (DS TM)	Facility	Facility	Facility	Facility	Node	Node
Upstream Upper Traffic Management (US UTM)	Facility	Facility	Facility	Facility	Node	Node
Upstream Lower Traffic Management (US LTM) Scheduling / Shaping Algorithms / QoS	Facility	Facility	Facility	Node	Node	Node
PON MAC/PHY (Report/Gate/LLID/FEC)	Facility	Facility	Facility	Node	Node	Node
10G EPON Optics (PMD layer)	Facility	Facility	Node	Node	Node	Node
PON MACsec (802.1ae) OLT & ONU	Facility & CPE	Facility & CPE	Facility & CPE	Node & CPE	Node & CPE	Node & CPE

Figure 19 – Locations of CAA and DAA DPoE System Functions

Figure 20, provides a summary comparison of the different architectures. In the area of reliability and operational complexity, the CAA options have no active or an active with minimal complexity (PON Extender), therefore reliability of the OSP is high.

The DAA options introduce new software and hardware in the plant, which introduces new reliability issues that require new redundancy features. Note that the failure of a remote device will impact only a small base of subscribers, and both the failure detection and recovery may be faster, if these systems have diverse fiber paths to redundant aggregation systems. The Remote DPoE System (RDS) has the minimum dependency on the other system components. However to scale the systems an EMS is likely needed as well as development for field support engineers to troubleshoot.

Areas	10G EPON OLT with Standard Wavelengths	10G EPON OLT with NG-PON2 Wavelengths	PON Extender	Remote PON MAC (RPM)	Remote OLT (R-OLT)	Remote DPoE System (RDS)
Location of DPoE subsystem hardware	Facility Only	Facility Only	Facility except PON optics	Facility except Basic L2, PON MAC, & PON optics	90% in the field	100% in the field
Location of DPoE subsystem software:	Facility Only	Facility Only	Facility Only	95% in the Facility	90% in the field	100% in the field
SDN Future Architecture Support:	Yes	Yes	Yes	Yes	Yes	Yes
NFV Future Architecture Support:	Yes - Completely	Yes - Completely	Yes - Completely	Yes - Except MAC & Some Upstream TM	Only Routing subsystem only	No DPoE Subsystems or limited
Plant Actives:	None	None	Low	Medium	High	High
Plant Power:	None	None	Low	Medium	High	High
Plant Space:	None	None	Low	Medium	High	High
Fiber Utilization:	1 (Poor)	8 or more	16+	16+	16+	16+
Serving Area Distance:	10km – 20km	20km – 40km	Up to 80km + 20km	No Limit to node 20km to ONUs	No Limit to node 20km to ONUs	No Limit to node 20km to ONUs
Facility Consolidation:	Limit to 20 km	Limit to 40 km	Up to 80km + 20km	No Limit to node 20km to ONUs	No Limit to node 20km to ONUs	No Limit to node 20km to ONUs
Facility Space / Power (assuming non blocking architecture and 128 split is possible for PON Extender and 10G Ethernet RPM, R-OLT & RDS):	64 subs 10km (2X) 32 subs 20km (4X)	128 subs 10km (same) 64 subs 20km (2X) 32 subs 40km (4X)	128 per PON port - No oversubscription of uplink possible	128, 256, or 512 Subs Possible per 10G Ethernet port	128, 256, or 512 Subs Possible per 10G Ethernet port	128, 256, or 512 Subs Possible per 10G Ethernet port
Cost Per Customer:	64 subs competitive vs. 128 PON Ext	Highest Capital Cost	128 subs per port is best case	Lowest cost with oversubscription	Lowest cost with oversubscription	Lowest cost with oversubscription
ONU optics:	Low Cost but same wave	Highest Cost but different waves	Lowest Cost but the same wave	Lowest Cost but the same wave	Lowest Cost but the same wave	Lowest Cost but the same wave
Number of Optical Ports in the Access Layer:	1	1	3	3	3	3
Data Traffic Encryption 10G Ethernet Segment	N/A	N/A	Uses existing OLT to ONU security	Proposed MACsec	Proposed MACsec	Proposed MACsec
Latency Performance Delta (vs. Standard OLT)	N/A	None	Variable Impact	None	None	None
Capacity Performance Delta (vs. Standard OLT)	N/A	None	Variable Impact	None	None	None
System Complexity (Software Architecture)	Low	Low	Low	Medium	High	Low
Operational Complexity (MTTD/MTTR)	Low	Low	Low	Medium	Highest	High
Reliability or MTBF Assessment	Highest	Highest	High	Medium	Low	Low
Supported by Existing OSP System Technicians	Yes	Yes	Yes	Marginal Increase	Different Skill Set	Different Skill Set

Figure 20 – Summary Assessment of CAA and DAA DPoE System

Centralized Access Architecture – OLT with NG-PON2 Wavelengths

- This may reduce the application for PON Extender and remote OLT due to:
  - The main driver for PON Extenders or remote OLT is fiber conservation (use of WDM between Facility and Node this approach eliminates that driver)
- Major Challenges:
  - Costs will increase at the OLT and ONU compared to nearly all other solutions
    - Aligning 10G EPON TWDM and NG-PON2 TWDM will drive volumes and will reduce cost
    - Cost will likely not reach the 10G EPON single wave level
    - We are exploring many different CPE optical wavelength plans and non-tunable solutions (should reduce costs)
  - Full Band Tunable Optics increase the one-time capital cost of the solution but yields an all passive network for operations
  - Partial band Tunable or Single Wave may reduce capital but will impact operations due to management of different colored pluggables or CPE
  - Capital costs are the major problem of this solution

Summary Assessment of CAA for PON Extender:

- Requires active in the outside plant (PON moves to AON – Active Optical Network)
- Use of WDM optics to the PON extender node for fiber utilization
- Extends distance to ~80 km
- Maximizes PON port utilization
- Maintains the use of the same 10G EPON 802.3av PON optical wavelengths to the CPE
- Costs compared to CAA PON and DAA need to be examined closely

Summary Assessment of DAA for PON:

- The functions required/desired in the Remote PON Device shifts complexity and costs
- Requires actives in the outside plant (PON moves to AON – Active Optical Network)
- Maximizes fiber utilization with use of WDM optics to the DAA for PON device
- Maximizes fiber distance between facility and customers served to ~80 km or more
- Maximizes PON port utilization because PON closer to subs (parity with CAA PON Extender)
- Space Savings in the headend (depends and may not be significant)
- Depends on customers per headend aggregation router port (within 2 km allows 128 subs per port)
- If non-blocking between headend and node is desired this may have marginal headend space savings (vs. PON Extender)
- Preference for blocking between headend and node may have significant space savings but with required intelligence in the ODN and potential blocking during peak periods
- Maintains the use of the same 10G EPON 802.3av PON optical wavelengths to the CPE
- Costs compared to CAA PON and PON Extender needs to be examined closely

ACKNOWLEDGEMENTS

The authors would like to thank the following individuals and test teams: Janet Bean, Allen Hui, Anand KumarGoenka, Wichai Wongchaisuwat, Tony Cotter, Greg Cyr, Tom Ferreira and Dave Hartrum.

REFERENCES

[CableLabs] DOCSIS® Provisioning of EPON Specifications DPoEv2.0 DPoE Architecture Specification DPoE-SP-ARCHv2.0-I04-150611

[IEEE] IEEE 802.3av-2009 Standard

A Converged Content Delivery Platform for IP and QAM Video

James Barkley, Weidong Mao
Comcast Cable

Abstract

HTTP based Adaptive Bit Rate (ABR) video delivery to IP enabled CPEs via Content Delivery Network (CDN) for both primary and secondary screens (IP Set-Top Box, mobile devices, etc.) is becoming a common feature of service providers. However, the large installed base of QAM STBs will require QAM based linear and Video On Demand (VOD) delivery for the foreseeable future. Traditional approaches to QAM linear delivery include statistical multiplexing and Multiple Program Transport Stream (MPTS) multicasts. A dedicated content library and purposely-built proprietary CDN have been used for QAM VOD delivery. As video migrates to H.264, separate encoding systems are dedicated to produce non-ABR assets (for QAM delivery), sharing little with the production of ABR assets (for IP delivery). Other video features such as advertising insertion, content encryption, data collection and reporting are also enabled through separate silos for IP and QAM delivery. This will not only increase the overall infrastructure cost, but also require separate operational resources.

A converged CDN platform along with common encoding, packaging and content origin are proposed to enable a single common content delivery platform for both IP and QAM video delivery, supporting all use cases, including linear and VOD. The proposed platform uses an MPEG DASH-TS based Common Intermediate Format (CIF) for packaging, ingest/storage, and delivery. It leverages the common ABR video and audio encoding system for H.264 video and various audio codecs. Legacy QAM video encoding is transitioned from MPTS, statistically multiplexing and multicasting to Single Program Transport Stream (SPTS) CIF segments with HTTP delivery through CDN. A CIF Media Presentation Document (MPD) is developed to describe all the codec and bitrate / resolution profiles for the CIF segments. Common encryption on CIF can be used to enable content security at rest in the storage and through the CDN delivery path. Common reporting and data collection schemes can be applied to the infrastructure. A common advertising system can be enabled for both IP and QAM video. At the edge of the network, the CIF MPD and segments can be converted to the appropriate MPEG-2 MPTS format for QAM and player ABR format (HLS, DASH ISO-BMFF, etc.) in conjunction with Digital Rights Management (DRM) for IP.

This converged content delivery infrastructure allows operators to increase the number of IP video clients while still supporting existing QAM STBs. This approach will result in lower overall capital and operational expenses. The approach will also enable seamless incorporation of the latest generation of ABR and codec technologies into the content delivery platform.

INTRODUCTION

With a large amount of QAM based digital set-top boxes deployed in the field, cable operators have been delivering linear video and services over backbone and regional IP networks through Gigabit Ethernet interface edge QAMs.

Traditional linear video delivery has been using techniques such as MPEG-2 Transport Stream over IP multicast, statistical multiplexing, and local Ad splicing. In particular, dedicated MPEG-2 encoders encode linear channels as SPTSs. Statistical multiplexers combine multiple SPTSs into an MPTS by taking advantages of dynamic allocation of bits to each SPTS, depending on the variability of the content. The statistical multiplexing technique enables bandwidth efficiency by packing more SPTSs into a single 6MHz QAM channel than would be supported without it. Local Ad insertion can be performed on each resulting MPTS by using a dedicated Ad Splicer with Ad Servers before sending to the edge QAMs.

Traditional video on demand delivery has been using techniques such as MPEG-2 SPTS over UDP/IP unicast. Typically, each VOD asset is encoded as SPTS using CableLabs ADI (Asset Distribution Interface) and CEP (Content Encoding

Profile) specifications. A purposely-built proprietary CDN is then used to deliver VOD assets from the content library to the headend locations, where streaming servers are typically located.

Given the large installed base of H.264-capable HD QAM digital set-top boxes in the field, significant bandwidth and storage savings can be achieved by transitioning linear and VOD delivery from MPEG-2 to H.264. On the other hand, IP video for primary and secondary screen is becoming a reality. The ABR-encoded H.264 streams are delivered through HTTP based CDNs and various IP enabled access networks to the IP players. This includes all the linear, VOD, and cloud DVR services.

Due to the large installed base of QAM STBs, migration to IP STBs and devices is slow. Having completely separate platform/component infrastructures dedicated to QAM and IP (encoder, packager, origin, delivery, security, ad insertion, and operation management tools) is not cost effective.

This paper describes a converged delivery platform for both IP and QAM video, supporting all use cases, including linear and VOD.

- Common encoding: It leverages the common ABR video and audio encoding system for H.264 video and various audio codecs.
- Common packaging: The proposed platform uses the MPEG DASH-TS based CIF.
- Common CDN delivery: In addition to IP video ABR HTTP delivery via CDN, the legacy QAM linear video delivery is also transitioned from MPTS, statistically multiplexing and multicasting to the SPTS CIF segment with HTTP delivery through the same CDN. The CDN also delivers CIF segments for QAM VOD.
- Common encryption: Common encryption on CIF can be used to enable content security at rest in storage and through the CDN delivery path.
- Common advertising systems can be enabled for both IP and QAM video. At the edge of the network, the CIF MPD and segments will be converted to the appropriate MPEG-2 MPTS format for QAM and player ABR format (HLS, DASH ISO-BMFF, etc.), in addition to with DRM for IP.
- Common reporting and data collection schemes can be applied to the infrastructure.

This converged content delivery platform allows operators to migrate to an all IP infrastructure while still serving millions of QAM STBs in the field. This approach will result in lower overall capital and operational expenses. The approach will also enable efficiency in storage and network bandwidth resources in the evolution of content delivery platforms.

COMMON CDN DELIVERY

When QAM VOD was first introduced, Streaming Servers were deployed regionally. Each region was loaded individually with the titles in the library. As the number of titles and usage grew, CDNs were introduced allowing the use of central libraries to hold the library. These CDNs were proprietary systems built around the specific characteristics necessary to provide assets to the Streaming Servers. The QAM VOD delivery is shown in **Figure 1**.

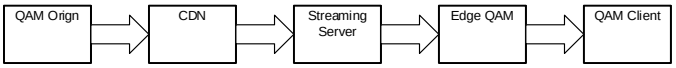


Figure 1 - QAM VOD Delivery

When IP VOD was introduced, it used a central library. However, the new IP library differed greatly from the QAM library because the QAM library contained MPEG-2 titles and the IP library contained H.264 titles. Furthermore, the CDN was different between QAM and IP. For IP VOD a title is transferred as a series of small files. QAM VOD views a title as a single, much larger file and is very sensitive to latency in the delivery. The IP VOD delivery is shown in **Figure 2**.

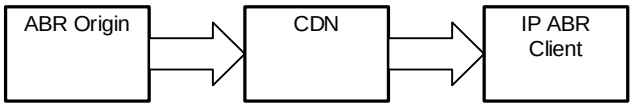


Figure 2 - IP VOD Delivery

In their 2015 paper, “Leveraging IP Video Delivery for Linear QAM TV” Weidong Mao and David Brouda introduced a HTTP ABR to UDP Gateway (QAM Linear Gateway) as an IP ABR

client [1]. The gateway retrieves designated ABR profile(s) from a CDN, performs the necessary HTTP ad insertion and streams the result to the edge QAM over UDP/IP as a transport stream. Using the gateway allows a CDN to be shared in the distribution of linear content with both IP and QAM clients. A converged linear architecture is shown in **Figure 3**.

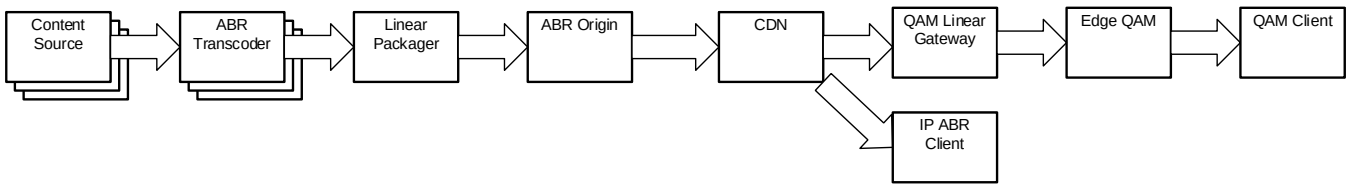


Figure 3 - Converged CDN for QAM and IP Linear

Taking the general concept of an HTTP ABR to UDP Gateway, we create a QAM VOD Gateway that converts from the large file based requests that are made from existing Streaming Servers into the ABR segment requests of the CDN and origin that are deployed for IP VOD. The QAM VOD Gateway satisfies the latency requirements of the Streaming Servers. This makes any H.264 title available to all QAM clients that are capable of decoding H.264.

All ABR delivery begins with reading a manifest that identifies the profiles and the files (segments) that make up the profiles. Manifest are one of two types dynamic, or static. The QAM Linear Gateway works with dynamic manifests, in which segments are constantly being added and removed. The QAM VOD Gateway, however, uses static manifests. Working with static manifests is, at the same time, easier and harder. A static manifest needs only to be read once, rather than at each update. Once read, it is only the most recent segment that requires processing for a dynamic manifest. Conversely, for a static manifest all segments require processing.

In contrast to QAM Linear, which traditionally supports statistical multiplexing, QAM VOD is constant bit rate delivery. This means that unlike other ABR clients a QAM VOD Gateway will select a single bit rate and never change. A gateway could change bit rate if there are issues with the network through the CDN back to the origin, however, there is no way to reclaim the unused QAM capacity.

Caching has long been use used in QAM VOD Streaming Servers to manage the impact of network through the CDN back to the origin. Having a gateway change the bit rate would not combine well with this caching. The change of bit rate would affect multiple sessions (until the stream ages out of the cache). Transition from caching in the Streaming Server to adjusting bit rate in the gateway would significant increase the network traffic into the Streaming Server.

A converged VOD architecture for H.264 delivery utilizing the QAM VOD Gateway is shown in **Figure 4**.

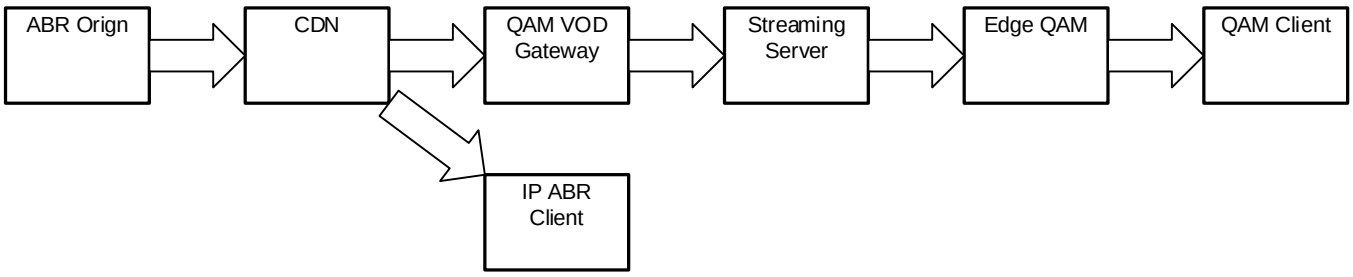


Figure 4 – Converged CDN for QAM and IP VOD

This converged CDN provides H.264 delivery to QAM VOD clients without incurring additional origin costs by reusing ABR origin. However, this does not reduce the number of CDNs, as there is still a separate CDN for MPEG-2. Additional savings is possible by transition to a single CDN for

all VOD. This does require that the existing MPEG-2 library be transformed from large files to ABR files. It is not necessary to create multiple different bit rate profiles since the network has already been scaled to support the single MPEG-2 bit rate.

Thus far we have discussed linear, or VOD, separately. However, when looking at the CDN, the capabilities are the same, whether the files being delivered are linear or VOD. Files used for linear can have a much shorter relevance, but this difference is straightforward to express within the HTTP standard. Rather than having two CDNs, (one for VOD and one for linear), additional savings are possible by sharing the CDN between linear and VOD.

COMMON ENCODING

The introduction of the gateways allows a single encoding platform to be used without consideration of the IP or QAM delivery format. An asset (linear channel or VOD title) is encoded and packaged into ABR segments. IP clients, using the manifest, will select the specific encodings that they are capable of decoding. QAM clients will have access to the same assets via gateways.

Common video encoding is H.264. Some QAM devices are only capable of supporting MPEG-2 video encoding. However, that encoding will not be destined for IP devices. Similarly, there are experiments in delivering H.265 video encodings to IP devices, but little interest in creating QAM devices capable of such a video decode. Since we are using ABR, separate segments could be selectively produced for those assets that require alternative video encodings.

It may not be necessary to have multiple video encodings on the origin. Gateways already are transforming from ABR to MPEG-2 transport streams. Depending on the tradeoff of origin storage, CDN cache and gateway computing performance, it is possible that the gateway may transcode from H.264 to MPEG-2 on the fly.

COMMON PACKAGING

DASH-TS based Common Intermediate Format is introduced as a common packaging format for both linear and VOD content at the origin and CDN. This introduces a CIF conversion between the CDN and IP client, but allows for more efficient use of the origin and CDN cache. The savings of resources in the origin and CDN is partially offset by the resources at the edge, converting on-demand to the IP streaming format. However, having the conversion at the edge and not the origin further reduces the burden when multiple streaming formats are to be supported.

In order to support ABR delivery multiple encodings are produced at differing bit rates. Video not only requires more bits to encode than audio, but also is easier to change the bit rate through changes in resolution. The result is that only one bit rate will be created for audio. Rather than multiplex the audio into all of the video profiles the audio is packaged separately from video.

Although QAM has traditionally relied on a single audio, encoding IP has not. It is common to have some devices capable of decoding Dolby Digital Plus (DD+ also known as EAC-3), others limited to Dolby Digital (DD also known as AC-3 and still others limited to AAC. Creating unique segments for each of these audio encodings and identifying them in the manifest allows the client to select the encoding they are capable of supporting.

Separate audio segments also simplify accessibility support within VOD. QAM settled on a limit of two audios minimizing the fixed bandwidth cost for audio. As a result, QAM VOD selects either a Secondary Audio Program (SAP) or Descriptive Video Service (DVS) for the second audio – never both. With separate audio segments, a title can be created with any number of audio tracks. The gateway will select the two that are to be delivered.

COMMON ENCYPTION

For QAM delivery, encryption is applied at the QAM device itself. Traditionally the distribution network has not required encryption because it has been a closed system. Only specific, qualified devices are allowed to access the distribution network. Careful monitoring is required to ensure that no unqualified device is granted access and periodic audits ensure that qualified devices are not compromised.

For IP delivery, encryption is tied with the DRM system associated with adaptive streaming format.

Thus far, it has been stated that the gateways function as IP clients for the converged system. Therefore, common encryption can be achieved by when gateways utilize the same streaming format as all other IP clients. Unfortunately, not only is there no single streaming format, innovation changes the existing protocols while introducing new ones.

Instead, the aforementioned CIF is used. This format supports encryption of the data as it is loaded onto the origin. Data stays encrypted as it passes through the CDN.

For QAM, the data is decrypted in the gateway. Depending on the exact security desired, data can be encrypted leaving the gateway, or it can be encrypted at the QAM.

For IP, a device is necessary to map from CIF to the client's streaming format. If this were done prior to loading on the origin, there would be no sharing of storage, or cache, between QAM and IP delivery. In addition, new copies would be required for every new protocol. If CIF conversion were performed between the origin and the CDN, storage would be shared, but not cache. In order to share storage and cache, the conversion must take place between the CDN and the IP client.

COMMON ADVERTISING SYSTEMS

It is reasonable to start with the understanding that there are two advertising systems in content delivery; one for linear and a second for VOD. Linear delivery, both IP and QAM, requires selection of the advertisements in real time. In contrast, QAM VOD selects advertisements prior to starting any delivery. IP VOD can leverage either technique by selecting advertisements either prior to starting any delivery, or in line with the delivery of segments.

VOD advertisement selection is personalized to both the viewer and the title. The effectiveness of advertisement campaigns can be further increased by considering all advertisements presented in a single VOD session.

Linear advertisement selection is based on station, rough-time boundaries and broadcast area. Unicasting reduces the broadcast area to a single client, allowing personalization similar to that achieved in VOD. This degree of personalization can even be achieved during broadcasting when the client orchestrates the advertisement selection.

Given the progression toward personalized advertisement selection, the differences between linear and VOD advertisement selection become a lot less significant. Selecting an advertisement requires three items: identification of the advertisement opportunity; identification of the client; and identification of the entertainment.

Previous sections discussed common encoding and common packaging. The common encoding includes conditioning content for clean transitions into and out of advertisements. The manifest generated by the common packaging includes signaling of the advertisement opportunities. The packaging also includes the creation of package boundaries at these transition points. Any advertising system only has to work with this one format for advertisement opportunities.

When the client makes the selection of the advertisement, identification of the client is straightforward. For QAM Linear delivery, there exist techniques that allow the client to select the advertisements, but it is more common for the selection to be made during the broadcast. In this case, the identification is via the broadcast ad zone and all clients within that ad zone receive the same advertisement.

Linear and VOD assets have different identification strategies (station versus title). The manifest generated by the common packaging manages both strategies in a single structure.

Having aligned the inputs for selecting linear and VOD advertisements, a common advertisement system is possible. The common advertising system is called individually for each advertisement opportunity. Title or program context is provided as another input, tracking with any previous decisions that have been made.

For QAM VOD, the smallest change is to continue to select all advertisements prior to session

start. Moving the selection of advertisements to the QAM VOD Gateway is possible, but has some expense; either removing caching from the Streaming Server, or additional signaling between the Streaming Server and the gateway. In addition, any support of fast-forward or fast-rewind in VOD presents challenges, predicting the timing of a request for an advertising selection to ensure all opportunities are resolved before the data for the advertisement needs to be fetched.

COMMON REPORTING AND DATA COLLECTION

The first content delivery by most service providers was QAM linear. When QAM VOD was added, the use case was sufficiently different there was no serious attempt to share any structure. Eventually IP delivery was added. However, the status quo had been formed, and there was little shared between IP Linear and IP VOD.

This has tended to result in large *diverged* content delivery systems as shown in **Figure 5**

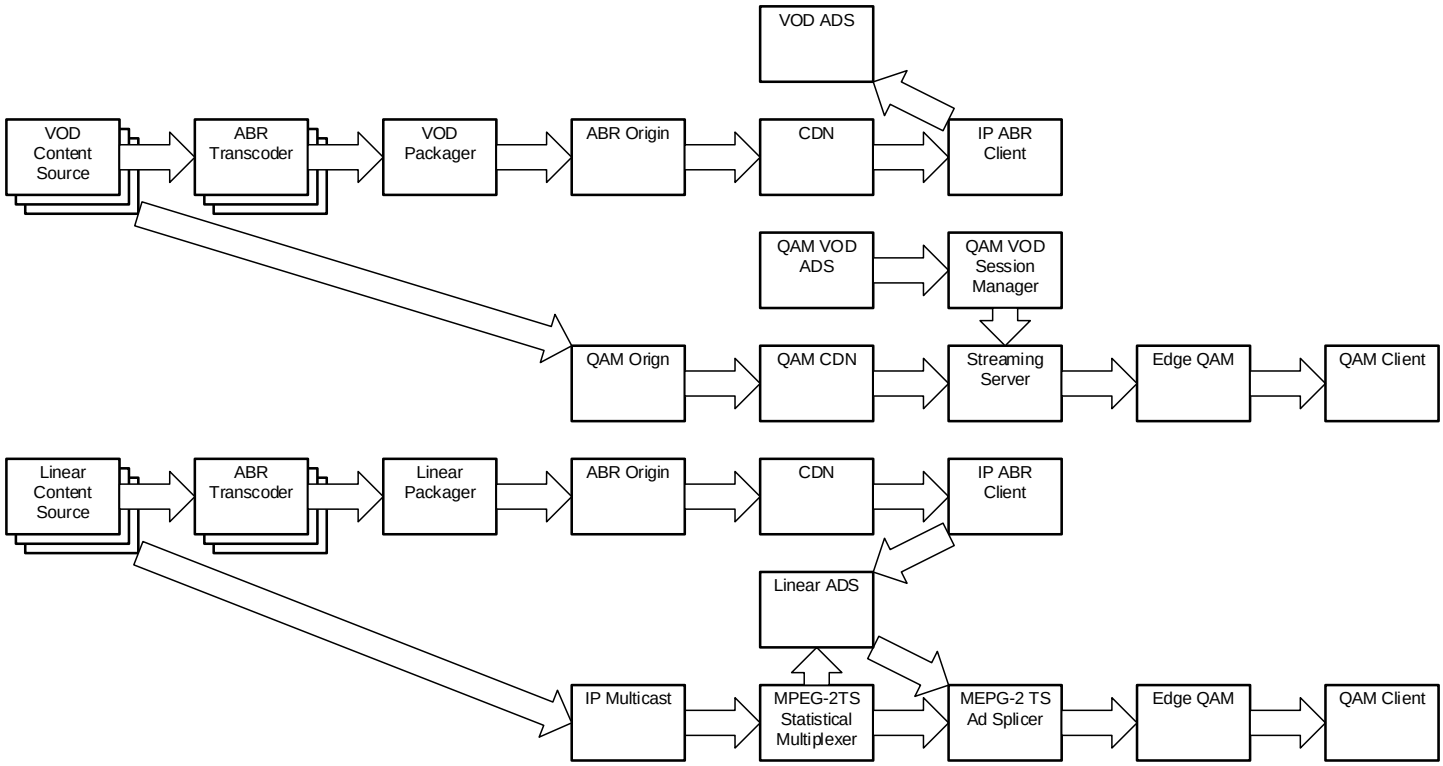


Figure 5 - *Diverged* Content Delivery

The large number of different components involved complicates Reporting and data collection.

Legacy (non-ABR) delivery makes significant use of the UDP protocol paced at the source to match the rate of video decoding. ABR delivery makes significant use of the TCP protocol. TCP allows the delivery rate to vary for short time intervals while trending to the video decoding rate over the long term. These differences further discourage any desire to converge reporting and data collection.

However, as we have shown it is possible to reduce the number of unique components in a converged content delivery system as shown in **Figure 6**.

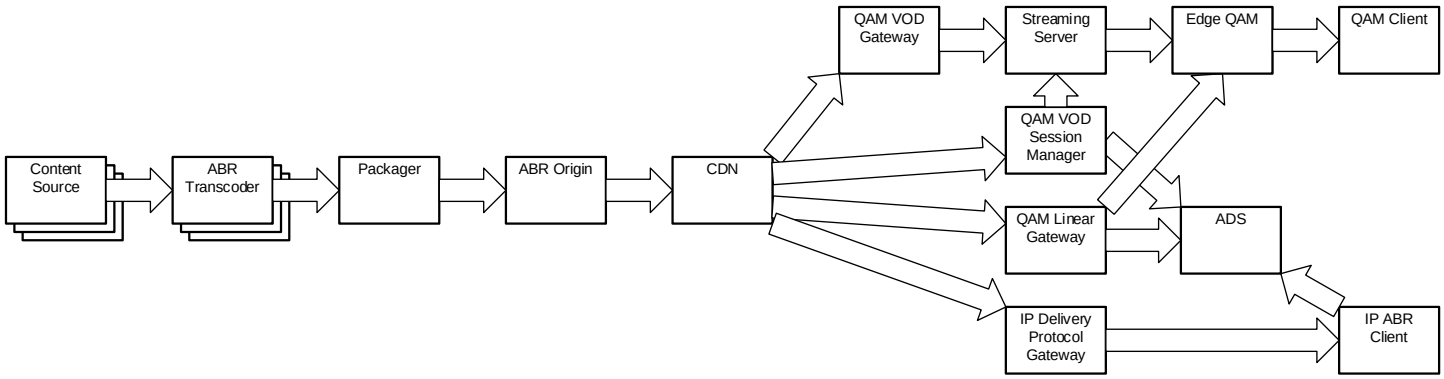


Figure 6 - Converged Content Delivery

With converged content delivery, Linear and VOD can share contents from the content source through the CDN. There are different performance characteristics for Linear and VOD, and there may be partitioning of the components to accommodate these differences. However, the default as common components is to have common reporting and data collection.

Although the components diverge after the CDN, the functionality of retrieving the manifest and segments in addition to selecting the advertisements is still common. This will continue to drive common reporting and data collection.

CONCLUSION

Content delivery has developed in a slow fashion. Different services and delivery formats (e.g., linear or VOD, QAM or IP) have been added over time. Taking a pause from the work of delivering the content allows the time to reflect and to begin to see the similarities are greater than the differences.

Although this discussion has been primarily focused on MPEG-4 delivery, as it is common to both IP and QAM content delivery, this can also be used for MPEG-2 delivery. Obviously, with MPEG-2 there is no sharing of content between QAM and IP. However, as non-MPEG-2 content becomes predominant, operations teams will be less familiar with the structure and issues of non-ABR delivery. Transitioning MPEG-2 into ABR will reduce the operations cost of continuing to carry MPEG-2.

This converged content delivery infrastructure allows operators to increase their IP video clients without needing to replace their QAM STBs. This approach will result in lower overall capital and operational expenses while enabling seamless incorporation of the latest generation of ABR and codec technologies into the content delivery platform.

REFERENCES

[1] Leveraging IP ABR Video Delivery for Linear QAM TV, Weidong Mao, David Brouda, 2015 Spring Technical Forum.

[2] R. Fielding et al., "Hypertext Transfer Protocol—HTTP/1.1: Message Syntax and Routing," Request for Comments (RFC) 7230, Internet Engineering Task Force, 2014.

[3] R. Fielding et al., "Hypertext Transfer Protocol—HTTP/1.1: Semantics and Content," Request for Comments (RFC) 7231, Internet Engineering Task Force, 2014.

[4] R. Fielding et al., "Hypertext Transfer Protocol—HTTP/1.1: Range Requests," Request for Comments (RFC) 7233, Internet Engineering Task Force, 2014.

[5] W. Mao, "Building Large VOD Libraries with Next Generation On Demand Architecture," NCTA Technical Papers, 2008.

[6] W. Mao, "Key Architecture and Interface Options for IP Video Over Cable," SCTE Conference on Emerging Technologies, 2009.

[7] S. Krishnan, W. Mao, "Open Content Delivery Networks for Managed Video Delivery to Multiple Screens," SCTE Cable-Tec Expo, 2010.

[8] ISO/IEC Draft International Standard 23009-1.2, "Information Technology – Dynamic Adaptive Streaming over HTTP (DASH) Part 1: Media presentation description and segment formats," 2011.

A STATISTICAL-ANALYTICS APPROACH TO MAKING VIDEO BANDWIDTH AND QOE DECISIONS WITH CONFIDENCE

Sean T. McCarthy, Ph.D.
ARRIS

Abstract

This paper presents a new statistical-analytic approach intended to enable operators to make video bandwidth and QoE decisions with confidence. The method we present is based on a new way of describing video-quality and bandwidth efficiency in terms of statistical probabilities. It can be applied to any video distribution method including ABR, CBR, and statmux for any format. A significant aspect of our method is that is does not require explicit traditional measurement of video quality in terms of PSNR, SSIM, or MOS values. Instead, we show that a metric derived from program complexity can be used as a statistical indicator of quality. Finally, this paper will show how real world data from in-service operations can be used to address key performance questions such as: What is the probability that video quality drops below any given level? Which programs are not receiving enough bandwidth? What is the efficiency of dynamic bandwidth allocation (VBR or ABR) compared to CBR allocation to each video program? Which operational parameters could be changed to improve overall video quality and efficiency? How would introduction of a new service impact existing services?

INTRODUCTION

Television has become more diverse and complex; but one thing remains the same -- operators want to deliver great video experiences to subscribers with optimal use of bandwidth. Yet, as the number of formats, displays, codecs, distribution protocols, and access technologies proliferate, it becomes increasingly more difficult for an operator to know with confidence that performance targets are being met. There are simply too many permutations to test using traditional methods. The challenge is compounded when new technologies such as DOCSIC 3.1 and HEVC enable 10's or even 100's of individual video programs to share bandwidth.

The motivation of the work reported in this paper was to find a way of assessing the performance of video distribution without needing to measure video quality for each individual program directly using methods such as PSNR, SSIM, JND, or MOS subjective testing. None of those tried-and-true methods scale very well in a mixed world full of adaptive bit rate protocols (ABR), CBR, statistical multiplexing, and a variety of core codecs (MPEG-2, AVC, HEVC, etc.) and numerous distribution models (IPTV, DTH, OTT, QAM, DOCSIS, etc.)

The new alternative method we have developed and have been exploring is based on the concept of "video-quality stress," which we define as the moment-to-moment ratio of 1) the varying bitrate that would be needed to achieve a target constant video quality and 2) the bitrate that is actually allocated to a program.

Moreover, we have found it useful to think of video-quality stress in terms of probability distributions rather than as a single value. The probability distributions provide insights into how stress changes frame-to-frame, scene-to-scene, and across different programs. Thinking in terms of probabilities opens up new ways of setting operational performance targets.

For example, instead of setting a bitrate to achieve or exceed a particular MOS value, the use of probability distributions allows a more

nuanced approach in which bitrates can be allocated so that programs meet or exceed a particular video-quality stress target for a predetermined percentage of time. In other words, the use of video-quality stress probabilities can enable an operator to set expected video-quality performance targets with foreknowledge of how often actual programming might deviate from the target. We believe that both the target video-quality stress and the probability of deviation are useful components in understanding a consumer's overall quality of experience (QoE), and could thus prove to be very useful to operators when designing video programming product offerings.

VIDEO-QUALITY STRESS

We define "video-quality stress" as the ratio of the bitrate that would be needed to produce constant video quality to the actual bitrate of the video program.

Video-quality stress is illustrated in Figure 1 for three video-distribution use cases: constant bitrate encoding (CBR); variable bitrate statistical multiplexing; and adaptive bitrate (ABR).

Figure 1 panel ***A*** illustrates the hypothetical variable bitrate (VBR) that would be required to maintain constant video quality over time for a particular program. We call this hypothetical variable bitrate a "need parameter." The need parameter is thus a measure of the moment-to-moment complexity of the particular video program with respect to compression. (The manner by which need parameter values may be obtained is discussed below.)

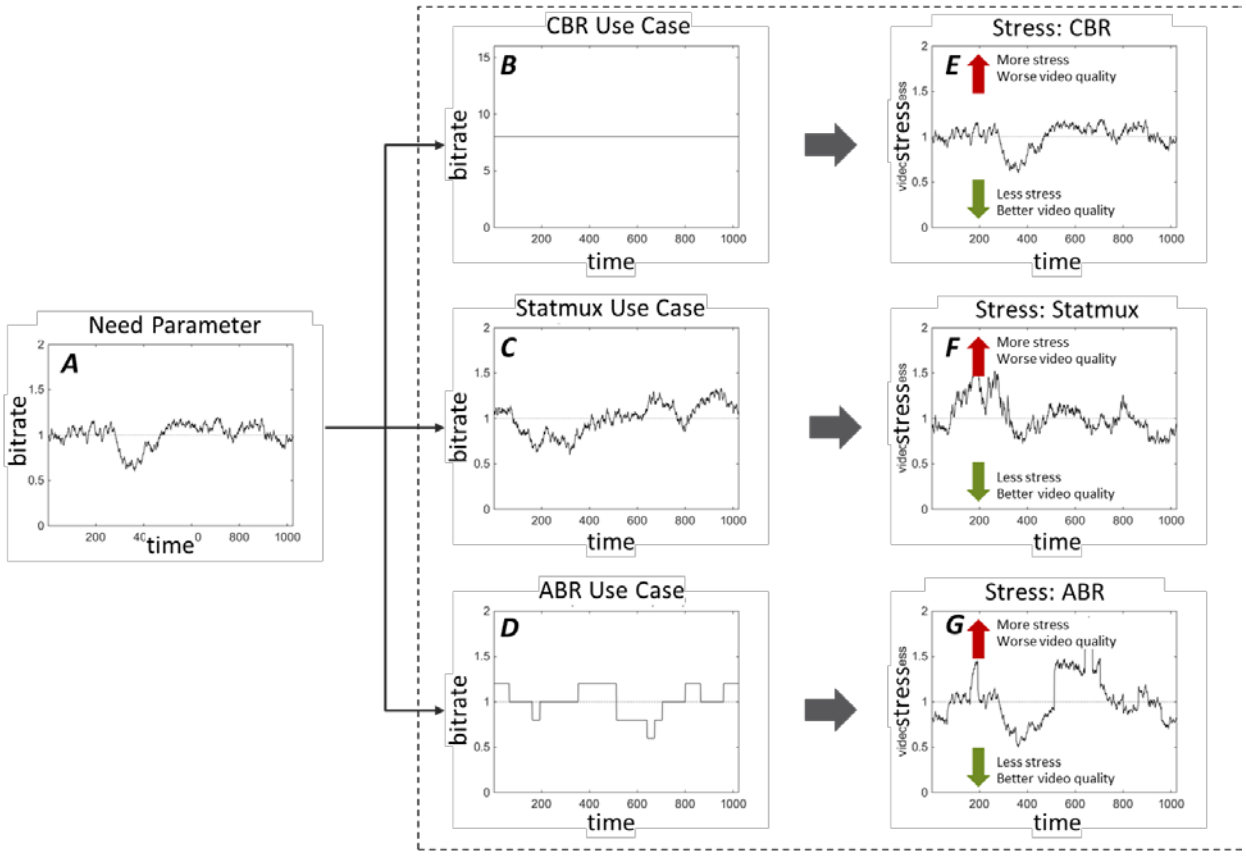


Figure 1 – Illustration of video quality stress

CBR encoding of a video program having the need parameter shown (***A***) is illustrated in Figure 1 panel ***B***. In the example shown, the CBR bitrate is equal to the average hypothetical VBR bitrate (the average value of the need parameter).

An example of the same video program being encoded as part of a statistical multiplexed (statmux) group of programs is illustrated in Figure 1 panel ***C***. In statmux, several programs are grouped into a pool that shares a typically fixed total bandwidth (see Figure 2). The objective of statistical multiplexing is to equalize video-quality across programs. This is achieved by dynamically allocating a portion of the total bandwidth to each individual program based on its complexity (as represented by a "need parameter", for example) compared to the complexity of the other programs in the pool. When a program has a high relative complexity the statmux controller will allocate it a greater share of total bandwidth, and, conversely, a lesser share when its relative complexity is low. The result is a variable-bitrate video bitstream as illustrated in panel C. (Note that only one program from the statmux pool is shown.)

Finally, an adaptive bitrate (ABR) use case is illustrated in Figure 1 panel ***D***. In the example shown, the video program is delivered as a series of contiguous piecewise-continuous CBR segments. The bitrate of each segment is allowed to change from segment to segment. The bitrate changes are typically driven by requests from client devices in response to local bandwidth availability.

The video-quality stress corresponding to each use case is shown in Figure 1 panels ***E*** (CBR), ***F*** (statmux), and ***G*** (ABR). In the simple CBR example shown in Figure 1***E***, the video-quality stress is the need parameter value, ***A***, divided by the CBR bitrate, ***B***, for each point in time. Similarly, the video-quality stress for the statmux example is mathematically the need parameter value divided by the actual statmux bitrate, ***C***. Video-quality stress for the ABR is obtained in the same manner for the data illustrated in ***A*** and ***G***.

Video-quality stress is not an absolute measure of video quality in the same way as PSNR, SSIM, JND, and other traditional video-quality metrics. Rather, video-quality stress is an indicator of the deviation of quality from a target.

In the CBR example, illustrated in Figure 1, ***B*** and ***E***, the CBR bitrate is deliberately set equal to the average value of hypothetical VBR bitrate (the need parameter) to make a point. One should expect that about half the time the CBR video quality would be better than the hypothetical VBR video quality because about half the time the hypothetical VBR bitrate – the bitrate needed for constant quality – is less than the average VBR rate (the CBR rate). Likewise, about half the time, CBR video quality would be worse than that for VBR.

Video-quality stress is thus an indicator of the deviation of video-quality from a predetermined benchmark. A value of video-quality stress equal to 1 indicates that a program is matching the video-quality benchmark. A value greater than 1 indicates that the program is stressed and video-quality is likely worse than the benchmark. A video-quality stress value less than 1 indicates that the program is less stressed and video quality is likely to be greater than the benchmark. (Below, we describe a method of calibrating video-quality stress to the video-quality benchmarks chosen by a video-service provider.)

The variation in video-quality stress for the statmux and ABR use cases are illustrated in Figure 1 ***F*** and ***G***, respectively. In each case,

the dashed horizontal line corresponds to a video-quality stress factor equal to 1. Deviations above the line indicate better-than-benchmark video quality, and deviations below indicate worse-than-benchmark video quality.

Video-Quality Stress to Validate Performance

Figure 3 illustrates the use of video-quality stress to validate the performance of a statmux system. In this example, 3 commercial video programs were fed into a 3-program statmux system in our lab, as illustrated in Figure 2. The system was made up of 3 encoding modules and 1 controller module that managed the allocation of bandwidth between the programs by sending bitrate allocation messages to each encoding module. In addition to producing a compressed program bitstream, each encoding module also produced a need parameter time series for the corresponding input program.

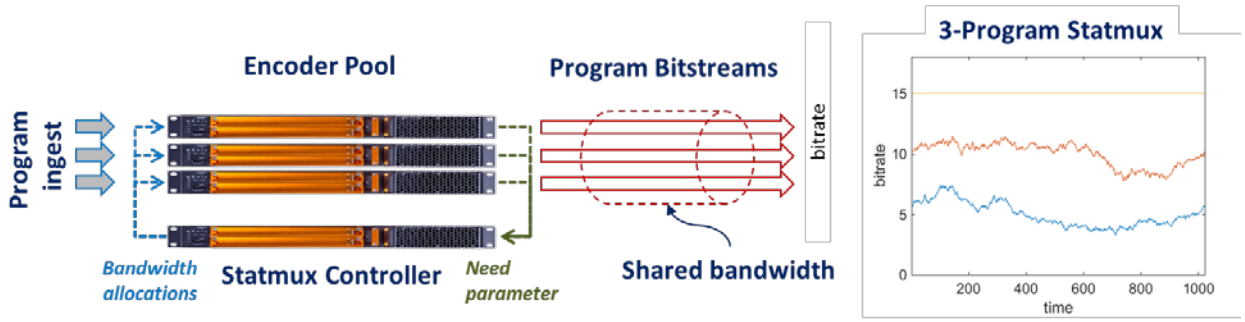


Figure 2 – Statistical Multiplexing

In Figure 3, examples of the resulting statmuxed video bitrates (***A***, ***D***, and ***G***) and the corresponding need parameter time series (***B***, ***E***, and ***H***) are shown. The corresponding video-quality stress for each program is shown individually in ***C***, ***F,*** and ***I***. The video-quality stress time series are also plotted together in ***J***. Note that the video-quality stress is very similar for all programs thus indicating that the statmux system is correctly doing its job of equalizing video quality across the programs.

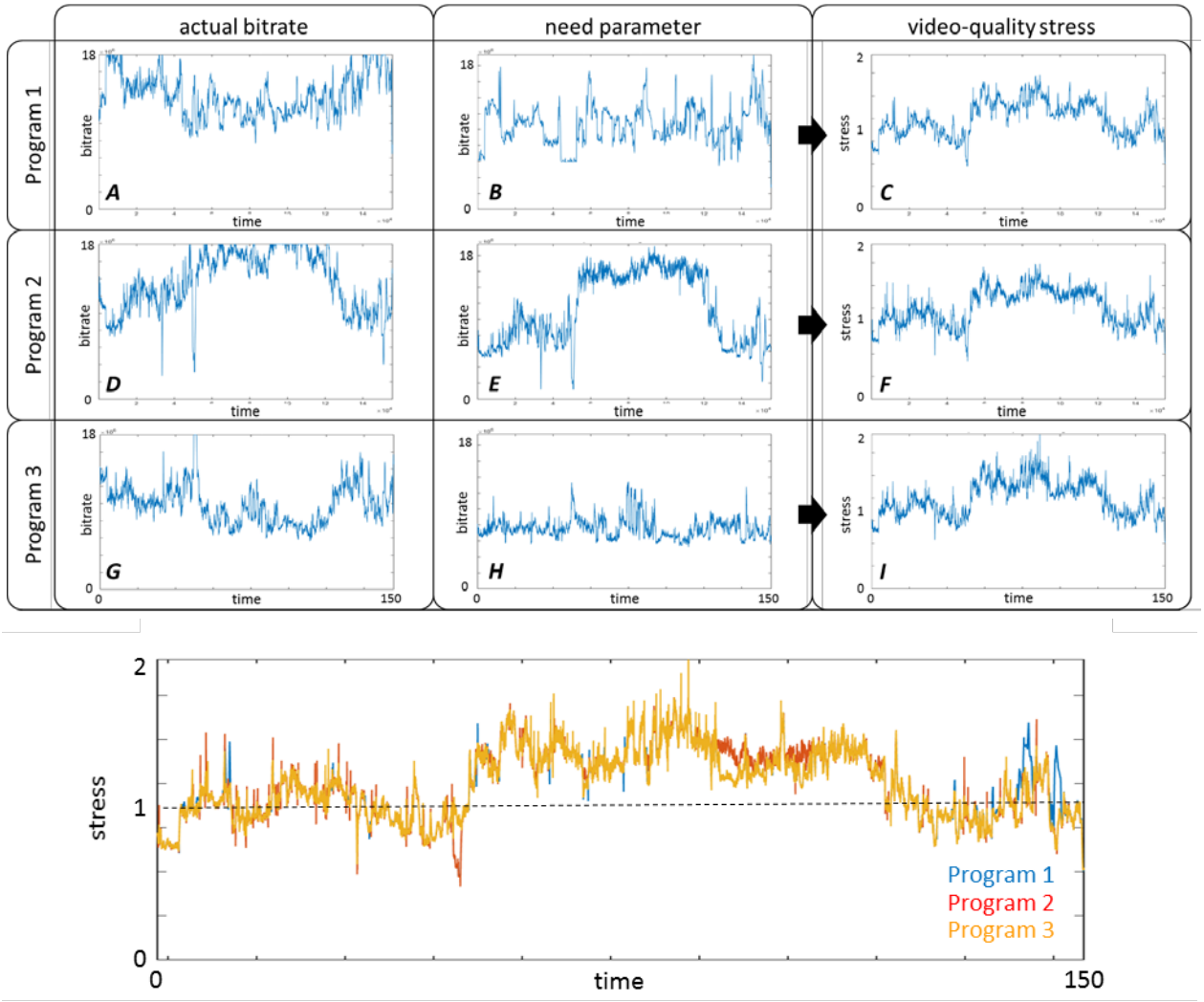


Figure 3 – Illustration of Video-Quality Stress Equalization in Statistical Multiplexing

Generating Need Parameters

It is worth noting that over the years we have developed and optimized algorithms for calculating need parameters for common video formats and codecs. Most of this work came from our work on statistical multiplexing where our approach has been to analyze video as part of the encoding process and thereby generate a need parameter that is provided to statmux controllers which then allocates bandwidth accordingly. Though born of statmux, need parameters can be calculated separate from the encoding process for use in any application (ABR, CBR, etc.). Other vendors working in this field have developed similar metrics, though perhaps under different names. What is important for the purpose of this paper is that the need parameter be as proportional as possible to the hypothetical VBR bitrate that would be needed to achieve constant video quality.

Calibrating Video-Quality Stress

Video-quality stress is a relative rather than absolute metric. It is a quantified indicator of video quality relative to an implicit video quality benchmark.

Setting video quality performance targets is something that video providers already do.

Indeed, setting appropriate video quality targets is a critical part of setting overall QoE targets that drive successful commercial services. The absolute video quality (as measured by subjective testing, for example) of a successful service depends on many factors, and can vary between free and premium services.

Once video-quality benchmarks are selected for a particular service, that absolute video quality level can be used to calculate a *video-quality-stress calibration factor* using steps equivalent to the following:

1. Collect bitrate data for a set of programs and times that is representative of a service offering.

2. Collect need parameter data for the same set of programs and times as in step 1.

3. Calculate the average value of the bitrate data collected in step 1.

4. Calculate the average value of the need parameter data collected in step 2.

5. Calculate a *video-quality-stress calibration factor* by dividing the average value of the need parameter data by the average value of the bitrate data.

With a *video-quality-stress calibration factor* in hand, a video provider can determine video-quality stress on an on-going basis. Calibrated video-quality stress is calculated by multiplying any need parameter time series by the *video-quality-stress calibration factor* and then dividing by the corresponding bitrate time series. When calibrated correctly, the expectation of the video provider should be that video-quality stress thus determined should fluctuate around 1 on an on-going basis; i.e., the average video quality of the service should be the equivalent to the video quality of the sample data set used in calibration.

Comparing Services

It is worth noting that a video provider could calibrate a premium flagship service, for example, and then use the resulting *video-quality-stress calibration factor* to gain insight into the relative quality of other services. For example, a video provider could use the *video-quality-stress calibration factor* obtained for the premium service as the calibration factor for other services. For example, this technique could be used to gage the relative quality of a new service under consideration, or to quantify the relative quality of premium services compared to non-premium services.

VIDEO-QUALITY STRESS PROBABILITY

The notion of video-quality stress by itself could lead to new insights; but we believe its real power comes when it is considered in terms of probabilities.

In Figure 4***A***, we show video-quality stress for a dozen simulated programs. (The simulated set of programs is shown as an illustration on an operator collecting a sample of programs from actual operations or by use of a standard video library used for internal testing and planning.) In the example shown, each program has a video-quality stress that tends to deviate from the calibration point (1, in this case) but by different amounts.

The ensemble likelihood that video-quality stress would exceed a particular value can be described with a cumulative probability, as shown in Figure 4***B***.

Cumulative probability provides a quantitative basis for answering questions such as, “how often does video-quality stress exceed a certain value?” For example, the lower-left crosshair in Figure 4B indicates that approximately 20% of the time video-

quality stress would be less than ~0.8. The upper-right crosshair indicates that video-quality stress would be more than ~1.4 only approximately 10% of the time.

Recall that video-quality stress is the ratio of a hypothetical bitrate needed to achieve a predetermined video-quality benchmark relative to the actual bitrate. Thus, a video-quality stress of 0.8 corresponds to a situation in which 125% of the hypothetical constant-quality bitrate is actually being consumed by a program. The lower-left crosshair in Figure 4B therefore indicates that more than 125% of the benchmark content-quality bitrate is being used by programs approximately 20% of the time.

Similarly, the upper-right crosshair indicates that approximately 10% of the time programs are getting less than ~70% (1/1.4) of the bitrate they would need to achieve the benchmark constant-quality target.

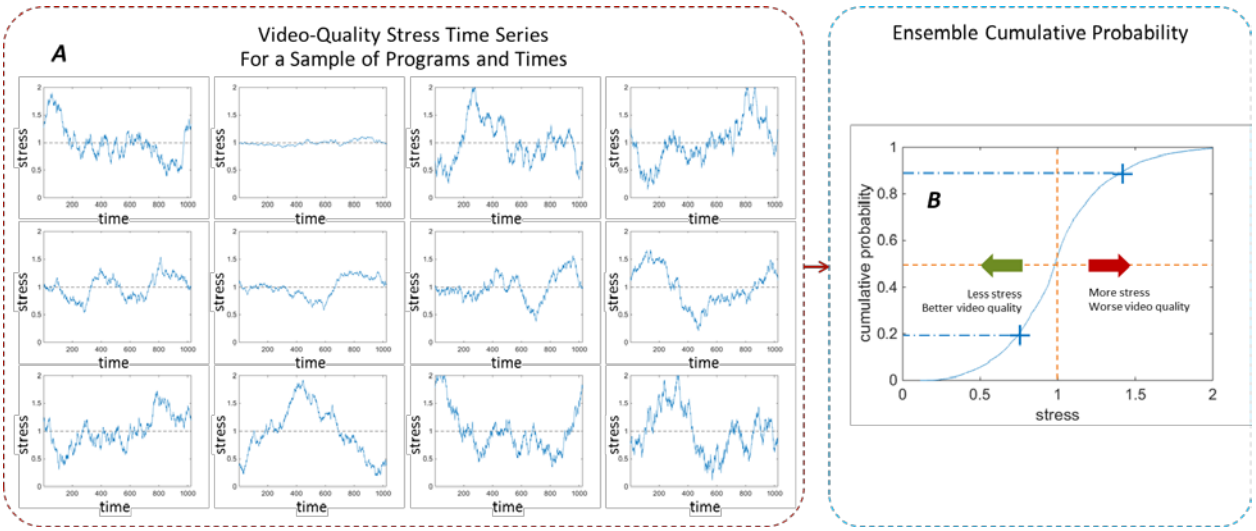


Figure 4 – Illustration of a determining Video-Quality Stress Cumulative Probability

Tuning Video-Quality Performance

In Figure 5, we show how video-quality stress can be used to fine tune video-quality performance. The key idea is that a service operator can determine how much bandwidth to allocate to a particular program in order to achieve performance defined by how often a program's video quality is likely to meet or exceed a benchmark video-quality target.

For the example shown in Figure 5, the cumulative probability of video-quality stress for a program (or group of programs) is shown as the right-most curve. Allocating more bandwidth to a program (or group of programs) would have the effect of moving the cumulative probability to the left. The amount of shift desired quantifies the amount of bandwidth that should be allocated.

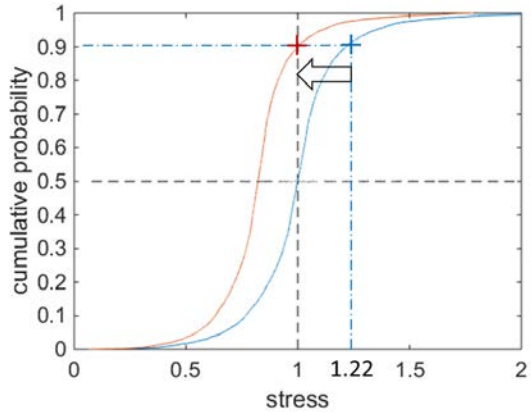


Figure 5 – Calculating Bandwidth to Achieve Video Quality Performance Targets

If for example an operator wished to provide a program having video quality equal or better than a benchmark target for 90% of the time, the operator could take the following steps:

1) Calculate the cumulative probability for an original bandwidth.

2) Find a first video-quality stress along the horizontal axis that corresponds to 50% along the vertical axis. (In this example, that video-quality stress is 1).

3) Find a second video-quality stress along the horizontal axis that corresponds to 90% (in this example) along the vertical axis.

4) Divide the second video-quality stress value by the first video-quality stress value. (The result is 1.22 in the current example.)

5) Multiply the result into the original bandwidth. The result is the bandwidth that if it were allocated to the program would achieve the desired performance target. (The tuned bandwidth would be 122% of the original bandwidth in this example.)

It is worth noting that an operator could use the method described above to explore several candidate performance targets to arrive and the best balance of video-quality and bandwidth utilization. Such steps can be performed for individual program or for any group of programs or even an operator's entire programming line-up.

Bandwidth Sharing

Video-quality stress can be calculated for any group of channels sharing bandwidth, as illustrated in Figure 6 for a statmux use case. The moment-by-moment sum of the need parameters for each program in the statmux pool yields an aggregate need parameter, which in turn yields an aggregate video-quality stress for the entire pool of programs. The same principles could be applied to gain insight into the aggregate performance of managed and unmanaged ABR distribution.

Statistically, the total accumulated neediness of the programs varies over time. Consequently, the total video-quality stress varies over time. The amount of that variability depends on the number of programs sharing bandwidth.

Figure 6 illustrates the dependence of aggregate video-quality stress on program count for a statmux use case. The left-hand graphs represent statmux pools of 3, 30, and 300 programs, respectively. The middle graphs illustrate the corresponding aggregate video-quality stress over time. Note that the variability of aggregate video-quality stress decreases with increasing program density (from top to bottom in the figure). The rightmost graph shows that the standard deviation of the aggregate video-quality stress decreases in inverse proportion to the square-root of the number of programs in the statmux pool.

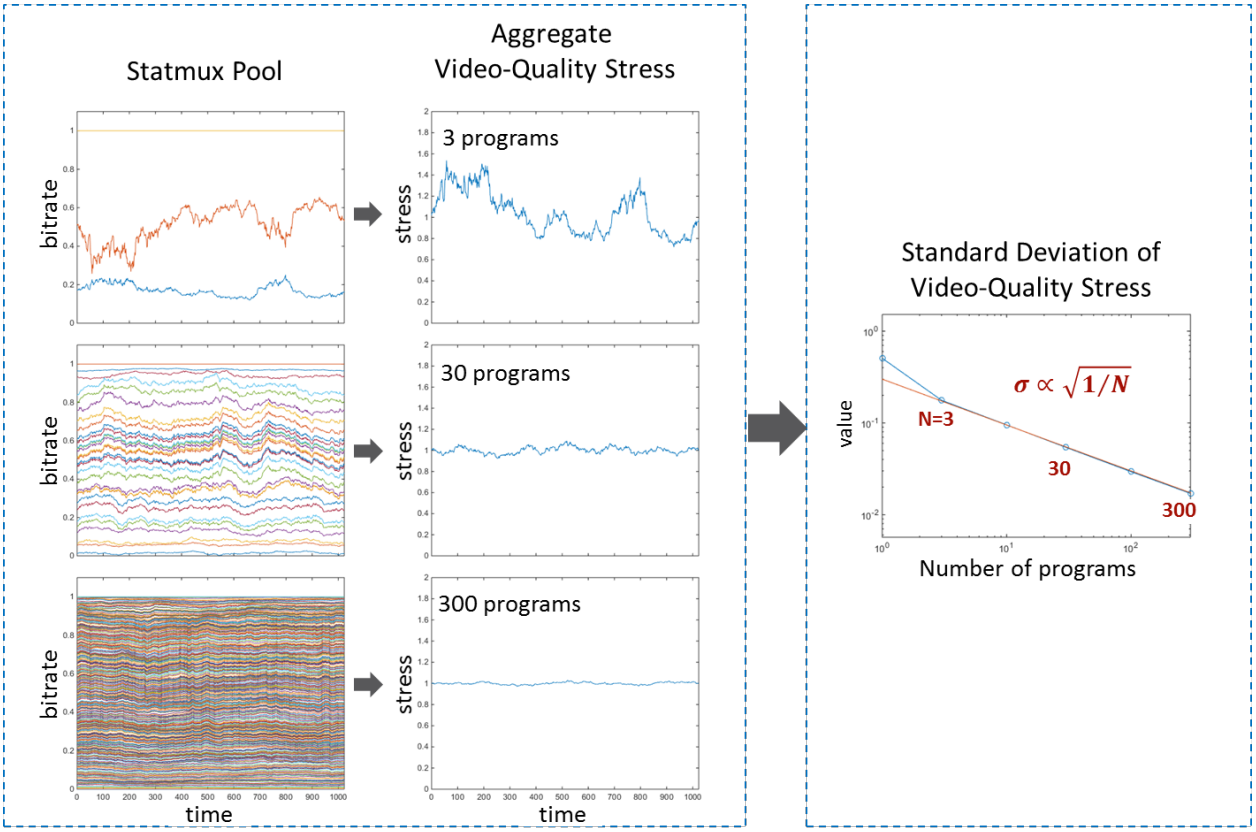


Figure 6 – The variability of the video-quality stress decreases in a predictable manner with program density

Another way to visualize the impact of program density on variability is illustrated in Figure 7. Panel ***A*** shows cumulative probability distributions for statmux systems having 1, 3, 10, 30, 100, and 300 programs, respectively. As the number of programs in the statmux pool increases, the cumulative probability curve narrows and becomes steeper. Panels ***B*** though ***G*** are the statmux equivalents of the shifted cumulative probability curves that were shown in Figure 4.

As the number of programs in the statmux pool increases, the amount by which the cumulative probability curve needs to shift decreases. In other words, operators having a large number of programs sharing bandwidth might find opportunities to optimize video-quality performance with less investment in additional overall bandwidth per program.

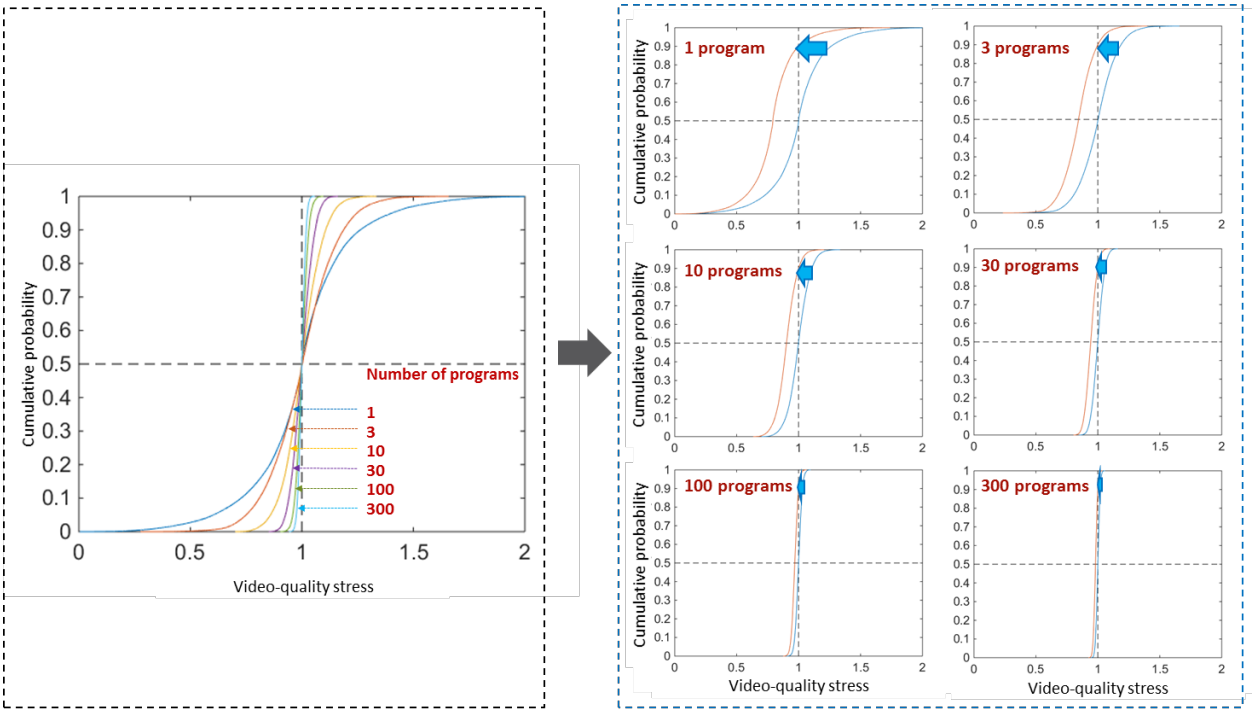


Figure 7 – The cumulative probability distribution of the video-quality stress narrows and steepens with increaseing program density.

CONCLUSION

The objective of the work reported in this paper was to find a way of assessing the performance of video distribution without needing to measure video quality of each individual program directly using methods such as PSNR, SSIM, JND, or MOS subjective testing. Such an alternative would provide advantages of scalability that are not otherwise possible, which is of particular importance as the number of programs increase and the ways in which those programs are distributed proliferate.

A key concept we introduced and explored in this paper is "video-quality stress," which we define as a relative indicator of video quality with respect to a video-quality benchmark. Video-quality stress is derived from a "need parameter" and the actual bandwidth consumed by a video program. The need parameter is a metric that is extracted from decades of statistical multiplexing experiences, and is often available as metadata (or closely related metadata) from professional encoders and transcoders. For the purposes of this paper, the need parameter is a moment-by-moment indicator of the hypothetical variable bitrate that would be needed to achieve constant video quality over time. In this sense, the need parameter is a measure of the complexity of a program.

We provided a method that operators can use to calibrate video-quality stress to the particular needs of their own video services. We also showed how a *video-quality stress calibration factor* could be used to compare the relative video quality of difference services.

Another key concept presented is the notion of describing video-quality stress in terms of probabilities. Doing so enables more nuanced consideration of video-quality performance and benchmarks. Calculation of

cumulative probabilities from programs and/or groups of programs enables operators to gain insight into how often the video quality of individual program or groups of programs exceed or drop below any particular level. Perhaps as important, we showed how video-quality stress probability could be used to tune bandwidth allocation to programs so as to achieve quantifiable performance targets.

Lastly, we explored the behavior of video-quality stress when a large number of programs share bandwidth and were able to quantify the dependence of the variability of video-quality stress on program count.

We plan to continue to investigate areas in which the notion of video-quality stress and associated probabilities could provide new actionable insights. In the meantime, we hope that the information we present here will prove useful to operators as they explore the optimal balance between video quality and bandwidth utilization. We welcome the opportunity to explore these concepts more fully with our colleagues and partners in the industry.

A Stream Data Platform For Video Delivery Telemetry

Michael Bevilacqua-Linn
Comcast

Abstract

As video delivery systems transition to all IP, they are becoming massively scaled distributed systems. Comcast's video delivery systems span hundreds of sites, and failures in one site can have unexpected negative imapcts on video delivery. When troubleshooting these sorts of systems, it's useful to have as broad, but often shallow, operational view of the system as a whole.

In addition, these systems generate data essential for business intelligence, capacity planning, reccomendations, and a whole variety of other essential functions. Without a system to methodically collect data from across this infrastructure, data collection is usually done via a set of ad-hoc integrations, usually with log files, or, at best, custom telemetry collection schemes. This leads to a stew of ever-shifting data formats, which much be parsed and reconciled to make sense of the system as a whole

In this paper, we present an architecture for a stream data platform which allows us to comphrehensively collect high value data, for both operational and other business purpouses, at large scale. In addition, we present a method of defining and evolving schemas for this data.

INTRODUCTION

Large scale distributed systems are notoriously tricky to reason about and debug. Our current methods generally assume that only the engineering or operations team responsible for an individual component in a larger distributed system has visibility into that component.

When an incident occurs, those engineering and operations teams are often brought together in a conference bridge, chat room, or email chain of doom. Using parochial tools and data collection, the individuals responsible for the troubleshooting effort attempt to communicate as best they can about complex issues, often without the ability to effectively share data.

The data that's collected for these parochial toolsets often comes in the form of unstructured logs, which are under the control of an individual engineering team. They may change the format of those logs at any time to better meet their needs. As this is the often only data collection source for a given component, point-to-point integrations between log collection systems and other systems which produce business value, such as business intelligence clusters, customer care tools, and so on.

As the primary motivation for the initial data collection is parochial, and the logs are not considered part of the contract that the component has with the outside world, this often causes much strife. A development team may add to, modify the semantics of, or remove elements from the log file at will.

Additionally, the point-to-point integrations between systems are expensive, and involve time consuming, and error prone, ETL from one system to another.

These issues can be overcome with the addition of a stream data platform, which is intented to summplement existing parochial data collection systems. The stream data platform has several main goals.

First and foremost, it serves as an extremely high volume, resilient, real time data bus for data which needs to leave the parochial confines of an individual component. When a stream goes into the stream data platform once, it can be used by many other systems which provide business value. For instance, a stream of video starts from an IP player may be used for business intelligence reporting, operational tooling, customer care tools, reccomendations, and so on.

Second, it provides a means to associate a schema with streams of data, and a means to evolve the schema for that data without breaking existing consumers of it. A schema is essential to aid consumers of the data in their data integration projects, however, in recent years, schemas have fallen somewhat out of style in the big data ecosystem.

As the ecosystem matures, this is changing. However, the schema technology needs to be more flexibile than traditional RDBMS schemas. In particular, a single stream may be consumed by many consumers, and not all of them will be willing or able to update their schema when a new version of it is created. Through a combination of technology choices and process, we enable schemas to be evolved by a data producer independent of any data consumer.

Third, the stream data platform provides ETL tools which can take raw streams of data, and transform them into clean streams of data once at ingest time. While we prefer, and recommend, that a source system directly emit clean data conforming to a schema, it's often impossible or impracticle to modify an existing system to do so. Doing this ETL once at ingest into the platform is still a win over doing it in an ad-hoc, case by case bases for point-to-point integration projects.

Fourth, the stream data platform will provide tools for data governance, security and discovery. Through the platform we must be able to secure an individual stream of data, through authorization before the stream can be accessed, as well as data security methods such as encryption of the data while it's at rest in the platform, and tokenization of individual fields. This security must be tied in with reasonable data governance policies, which specify what types of data need which security policies.

For example, a data stream which has elements in it which, when combined with other data sets, may constitue PII may need to both be encrypted while at rest, and have the the sensitive fields tokenized, to limit the risk of a data breach and ensure regulatory compliance.

Subject to data governance and security concerns, any user of the system should be able to discover the data sets, associated schemas and other metadata. This will ensure that users can find existing data streams and have a holistic view of the data available.

As of the writing of this paper, we mainly address the first three concerns of the stream data platform, and have done experimental work on the fourth.

THE ELEMENTS OF A STREAM DATA PLATFORM

The stream data platform has several elements. We have chosen to build the platform on top of open source software, primarily from the Apache foundation, along with some home grown components.

First, a system for high capacity data transfer. For this, we have chosen Apache's Kafka, a distributed messaging system which is highly optimized to ensure that all messages can be persisted to disk. This is essential to ensure that many consumers, all of which may not be consuming at the same time, can access

the datt streams. We have been scaled Kafka to a point where an individual cluster can handle over one million messages per second, without much tuning or effort, and are confident we will be able to scale it to meet our throughput targets as we grow the system.

A key element in Kafka is the topic, which represenents an individual data stream. We have tooling to create topics conforming to a naming convention, so that we can separate out different types of data streams

Our Kafkas are organized into ingest clusters, which ingest some subset of data from across our entire footprint, and larger aggregate clusters, which aggregate data from ingest clusters. Producers produce data to an ingest cluster, and consumers consume from an aggregate cluster. For resillency purpouses, we provide both ingest and aggregates in multiple, geographically isolated datacenters

Second, we have chosen Apache's Avro as our schema language. Avro allows us to create a schema for a given event, and also provides a set of libraries which allow us to serialize and deserialize data conforming to that schema in a variety of languages.

We have built a home-grown schema mananger that allows us to associate a schema with a topic, and evolve that schema. In the next section, we will describe the evolution process in more detail.

Third, we have chosen Apache's NiFi for basic data cleansing and ETL. While there will always be a need for advanced ETL, both at the batch and stream level, outside of this system, we decided that the capability to provide simple ETL was essential.

We use NiFi only to take streams of data that are in some non-standard format, such as raw logs, ad-hoc JSON, XML, etc, and put them into a standard format with a schema. NiFi was the right tool for this job because the vast majority of these transformations can be done by stringing together and configuring existing NiFi processors, without the need to write custom code. This makes it possible for analysts to do this ETL without needing engineering assistance.

Fourth, we provide a home grown HTTP ingest system. This is stood up alongside the ingest Kafka clusters, and allows us to work with systems which would be difficult or impractical to ingest a Kafka client into. This includes CPE and COAM devices.

The HTTP ingest system allows for two rough modes of operation. First, we allow data that comes into it in a clean format. This data is placed directly into a clean Kafka topic.

Second, we allow data that comes into the system in a raw format. This could be some legacy CSV, XML, JSON, etc. This data is placed into a raw, staging topic. From there, we use our data cleansing tool, NiFi, to clean and wrap a schema around the data. This allows us to avoid the need to duplicate data cleansing functionality in the HTTP ingest layer, as well as NiFi.

The final important platform we will provide in the stream data platform are tools for data security, discover, and governance. We are still early on in the selection of this toolset, but we have a good idea of some of the properties that we'll need.

Our data security tooling will need to allow us to enforce and configure two types of security policies, both of which we want to manage centrally.

First, we'll need to be able to configure access and authorization policies. This will allow access to topics on a per topic basis. Second, we'll need to configure data security policies. This is much more difficult, but at a

high level, these policies will allow us to encrypt streams of data, both at rest and in flight, and allow us to tokenize sensitive fields in the data.

Note that encryption at rest is not sufficient data protection. While at rest encryption will protect the data from filesystem access, assuming that the key management system has not been compromised, doing any analysis of large amounts of data will require it to be decrypted.

This means that many analysts, and analysis systems, will be authorized to access the data. If any of those systems, or people, are compromised, they will be able to decrypt and acquire large amounts of sensitive data.

The solution to this problem is to tokenize the most sensitive pieces of data, such as account identifiers, and other pieces of data which, when combined, could constitue PII. This must be done in a consistent way, so that analysts can still do joins, between datasets.

When done consistently, this level of tokenization will protect from de-ananomyization attacks associated with large scale data breaches.

Hand in hand with these security tools are our data governance and discovery tools. These tools will allow us to view the schemas for different data streams, view sample data for data streams (subject to security policy).

Basic, lightweight data governance, such as ensuring that clean data streams have a schema that conforms to basic guidelines, and so on, are essential to the ensuring that the stream data platform serves as a landing ground for high quality data, and not just a dumping ground of data streams.

SCHEMA DEFINITION AND EVOLUTION.

Our schema definition and evolution processes form the backbone of our data governance policies. We'll cover both in detail here, starting with schema definition.

We define certain core schemas, including distributed trace metadata, a common header format, common data types, such as account and device identifiers, and so on.

Schemas for a specific data stream are composed of these core schemas, in addition to data fields which are specific to that stream. Our data governance policy is centered around ensuring that we're making appropriate use of these schemas.

Wherever possible, we will automate data governance checks. For instance, we'll automatically validate schemas, check for required fields, such as docstrings, and so on. All of this will be done in a self-service manner, and be backed by a lightweight manual check.

Once a schema is created, it will be placed in the runtime schema management system we described in the last section. From there, a single producer system will write data into the stream data platform, conforming to the schema. Many consumers may then read and work with the data.

Producers must have the ability to evolve their schema independent of any consumer, as we cannot force new shcemas on highly distributed set of consumers. In order to do so, we take advantage of Avro's schema compatability feature. Avro allows for a reader schema, which is the schema used by the data consumer, and a writer schema, the schema used by the data producer.

Avro strictly specifies, and Avro libraries enforce, what it means for a reader schema to be compatible with a writer schema. The exact semantics can be found in the Avro

specification, but as a motivating example: a writer schema may have a field added to it, and remain compatible with an existing reader schema. However, a field can be removed if and only if a default was provided for it in the original schema.

While Avro's schema compatability feature strictly specifies what it means for a reader and writer schema to be compatibile, it does not have semantics to deal with changing a schema over time.

For instance, as mentioned earlier, it's "safe" to remove a field with a default from a writer schema. In this case, existing consumers will see the default value for the field.

However, it's possible to make a series of safe modifications that result in incompatible schemas. For instance, a field with a given name could be removed in one revision, then added back in in a subsequent revision, with a different type.

Consumers with the original schema would no longer be able to parse data produced by the new schema, as they would see data produced with a different type than they were expecting. At the time of this writing, we rely only on Avro's schema compatability, but we expect to extend this through our schema management system to enforce compatability through multiple schema versions.

SUMMARY

Comcast's stream data platform allows for the collection of telemetry data, at scale, and for a wide variety of purpouses. To do so, we provide a high capacity data bus, using Apache Kafka, as well as schema management and data ingest systems. Furture work will center around advanced schema management, data governance and security

Accelerating big data applications with multi-Gigabit-per-second file transfers

Charles Shiflett
Aspera, an IBM company

Aspera, an IBM company, provides a set of collaboration and file transfer technologies that leverage the Aspera FASP technology which is a distance neutral transfer technology.

This paper will describe Aspera's FASP transfer technology, both next generation FASP which is designed for 100gbit/s transfers as well as our standard FASP technology. This paper will aslo describe how the FASP protocol integrates into high performance / high scalability environments.

A Brief Introduction to Aspera FASP

Aspera FASP is a file transfer protocol designed to give users a high performance solution to sending data across the world limited only by the performance of your internet connection. Aspera FASP is a full stack transfer solution, and as such it provides congestion control, encryption, reliability, file checksums, and direct-to-cloud functionality.

Aspera FASP is a transfer solution designed around real networks meant to just work. It's a point-to-point solution which works on all platforms and within most use cases; Desktop, Command Line, Mobile, Web, REST, and as a library for most programming languages. The FASP protocol is typically layered on top of UDP with reliability, congestion control, and data transmission algorithms that are tailored to real networks, long haul fiber, satellite, wireless, firewalls, and VPNs.

As compared to TCP/IP, which is the standard IP congestion control protocol and used in most data and file transfer scenarios, TCP/IP doesn't effectively utilize available bandwidth in a number of scenario's.

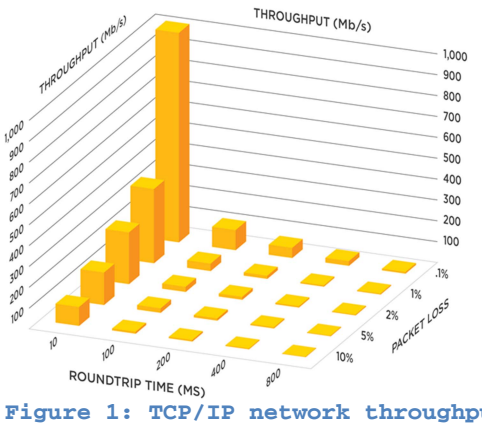


Figure 1: TCP/IP network throughput

As an example of why you wouldn't want to use TCP/IP for everything, *Figure 1* shows typical TCP/IP performance as related to round trip time and/or packet loss on a 1 gig link. In contrast, *Figure 2* shows Aspera FASP performance across the same link. It is exactly due to the relative increase in performance that Aspera FASP is widely used in a number of industries to speed content delivery around the world. For use cases see *asperasoft.com*.

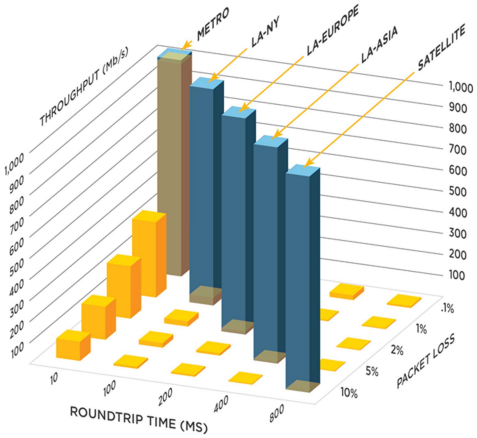


Figure 2: FASP network throughput

Aspera's goal has constantly been to remove bottle necks associated with high throughput file transfers. In 2013 Aspera showed that next-generation FASP was able to utilize 4x10gbit interfaces for a total transfer throughput of about 40 gbit/s in *Big Data Technologies for Ultra-High-Speed Data Transfer in Life Sciences* [http://asperasoft.com/fileadmin/media/Asperasoft.com/Resources/White_Papers/Big_Data_Life_Sciences_AsperaWP.pdf]. In 2014 Aspera showed that FASP was able to operate at near to 80 gbit/s with 2 x 40gbit network cards [https://communities.intel.com/community/itpeernetwork/healthcare/blog/2014/11/12/sc14-accelerating-life-sciences-at-80-gbits]. Our next whitepaper scheduled to be released contemporaneously with this paper will be showing Aspera FASP utilizing a 100 gbit network link using 3 x 40gbit network cards to transfer scientific data sets between HPC centers across the united states.

As speeds increase into the hundred gigabit range it becomes increasingly difficult to expect the storage and operating systems to keep up with network performance when using traditional I/O interfaces. This paper will describe some of the new and emerging technologies which Aspera utilized in scaling into and past 100gbit/s as well as how the FASP framework plugs into those solutions to provide a transfer framework that scales both into the use case of meta data intensive transfers (millions/billions of small files), as well as high throughput transfers with petabyte sized data sets.

A Design for 100 Gigabit Ethernet

Traditional POSIX I/O interfaces, which form the backbone of all I/O on every major platform suffer from designs that assume memory is fast and I/O is slow. As such they are designed to arbitrate between software that might be using hardware inefficiently and hardware that needs to be coddled to achieve maximum performance.

In practice this means massive caching between both storage and networking devices where data sits staged, waiting to be transformed and consumed by it's respective device. This waiting data obviously must be written to memory, and then re-read when a kernel device wakes up and begins to handle the associated queues. This typically means a complete flush of cachelines (The data in a CPU cache associated with an I/O operation), along with increased pressure put on the memory controller as it must write then re-fetch data.

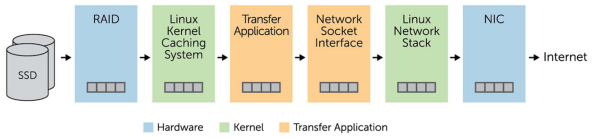


Figure 3: Memory Copies in POSIX Model

Figure 3 shows roughly what this looks like for a traditional transfer application where each of the grey bars represents a copy from one I/O queue/domain to another I/O queue/domain. The result is increased consumption of memory bandwidth and increased latencies when waiting on memory which in turn causes other processes waiting on memory to slow down as individual cores on the same physical piece of silicon compete for the same resources. This is the principal reason that transfers are slow when using traditional I/O stacks and it is a problem that compounds the more you are trying to do on the transfer node.

Exactly how slow this is depends on exactly how things are configured and what else is going on in the system. As an example, using the reference system Aspera used to show 100gbit/s transfers (XFS filesystem, 4x DC P3608, Intel® Xeon® E5-2699 v3), results in about 6 GBytes/s read with 2 or more I/O

threads, where as in direct mode (which bypasses cache), or when using Intel® SPDK (A userland storage framework, as opposed to a kernel framework) it is trivial to show results well past 12 GBytes/s on read.

Using a more traditional or featureful filesystem like EXT4 or ZFS yields reduced performance and in these cases one would expect performance closer to 2.5 GBytes/s (20 gbit/s). As one puts more pressure on the Memory Controller, the performance on the file system drops and throughput decreases. As an example using our classic FASP transfer framework in multi-threaded mode, we would see results on the same system of a little over 1GB/s (9-12 gbit/s) and this is largely due to the additional overhead caused by using the in kernel networking stack (Berkeley Sockets).

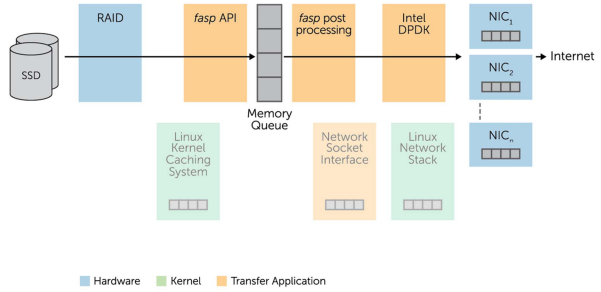


Figure 4: Zero Copy with FASP NX

In Figure 4, an Ideal I/O stack is shown. While this model doesn't eliminate the need to use memory, it fully eliminates the need to perform memory copies and with that significantly reduces the pressure put on the Memory controllers which allows the application to scale in proportion to available memory and I/O bandwidth.

We still retain the important POSIX principles such that we give good performance when doing a transfer. For instance, we read from disk into memory, and then transmit from memory which minimizes latencies when it comes time to transmit a block or in the event that we need to re-transmit a packet due to packet loss or corruption. This is almost the same as reading from disk to cache, and then from cache to application, application to NIC, only we have removed the process domains (kernel->user and vice/versa) as well as prevented double caching of data. Of course a simpler design is not to have an application Cache/queue and just send all of our data to the file system cache as soon as possible, and that is exactly what we do with standard ASCP. The only downside to that approach is that we lose the ability to control fine-grained I/O and memory performance which is important past about 10 gbit/s.

Of course it isn't enough to just eliminate memory copies and claim victory. Memory copies are just part of the problem, and as we consider that portion of the problem to be solved other issues begin to become evident such as how to handle processor intensive tasks (like compression or encryption), per-core limits to how much memory can be utilized (each core is limited to how quickly it can consume memory based on the number of hardware memory prefetchers), and how quickly we can write to I/O and NIC hardware queues.

Aspera FASP Next-generation attempts to address each of these issues through things like lockless multi-threaded architectures which avoid contention for any single object, in better utilizing Hardware offload such that CPU cores can be pinned to specific RX or TX queues on the hardware, and in how threading and encryption is handled such that we are able to fully utilize the hardware available on modern processor architectures.

One of the most interesting things about next-generation FASP is that the architecture had to change to accommodate these features. For instance, it turns out that if you want to design your own highly optimized network stack (which Aspera did for FASP Next Generation), you really can't also coexist with

the Kernel, and you almost certainly don't want to plumb in those changes into a new framework. To provide a high performance path forward, Aspera utilized Intel® DPDK to provide a userland transfer solution which eliminates the need to depend on the Kernel to provide network services. In this model, clients are able to connect to a FASP service via shared memory and initiate transfers between systems.

Intel® DPDK provides other benefits to enable applications to maximize throughput such as pinning a thread to a particular core (which eliminates losing L1/L2 caches), providing a framework to enable NUMA aware applications, and providing a burst oriented framework which enables efficient packet processing.

Optimizing for faster Storage I/O

Storage has undergone a radical transformation as file systems transition from single large systems to clustered and cloud based object stores. In these configurations, metadata is typically separated from the file payload and specific access conditions are typically required to fully utilize the throughput offered by clustered parallel file system and or cloud based object storage.

In both FASP and next-generation FASP, a filesystem abstraction layer is provided which provides flexibility in both how metadata is written and in how file data is written to disk. The two are handled independently and each stage of the I/O chain can be customized to provide the integration you need to fully maximize the potential of your storage solution.

For instance, if your local storage solution requires that you read/write with specific block sizes and you have slow meta-data retrieval you can pass Aspera FASP exactly those hints and it will eliminate all un-necessary metadata queries and perform disk I/O with fixed block sizes.

In the same vein, if you are transferring to or from cloud that same file system abstraction is available to translate from traditional file systems to object based file systems where data is stored through REST API calls.

The same holds true if you are attempting to transfer millions of very small files. In this use case, you can have multiple readers which do nothing but read metadata information along with the file contents and once they have packed enough data into a block they then send the entire block across the internet where the files are written as quickly as possible.

Even in use cases where we are doing transfers on behalf of a system agent where it is expected that we transfer files and then transform that data and interact with system services (as in Aspera's Avid Integration: http://asperasoft.com/partners/joint-partner-solutions/fasp-plug-in-for-avid), Aspera provides a flexible enough framework that the integration point is just going to be mapping API's from one system to another.

In most cases, the only optimization needed for good performance is to use the storage system as designed. However in some cases it isn't enough to just go fast, you need to maximize your storage utilization and the POSIX overhead is creating a barrier for efficient utilization. In these cases we are able to take advantage of Direct I/O when possible to eliminate a copy to cache in use cases where it is not expected that the data being cached will be used again. Aspera can utilize Direct I/O in a number of use cases such as, when we are using RDMA or Intel® Omni-Path, When the file system supports it (i.e. XFS), or when we can take advantage of Intel® QuickData or similar userland transfer technologies.

Aspera provides solutions which optimize both network I/O and storage I/O in a way that is vendor neutral and high performance. Aspera's goal is to provide solutions which solve real problems in both sending data over the network and in efficiently reading and writing that data to disk. In addition Aspera provides the safeguards you need to be sure that the data you created is securely sent across the Internet using industry standards as well as safeguarding the integrity of your data with end-to-end checksums.

A Flexible High Performance FASP Solution

The goal of Aspera FASP is to provide a transfer solution which efficiently utilizes your unused bandwidth to move content as quickly as possible. In the end, be it sending 8k uncompressed video @ 60 FPS (~26 Gbps) in near real time, or replicating your data center complete with disk images which need to be compressed before being sent, Aspera has a solution for you.

All of our solutions are built around the same core FASP framework, they are built to be fast, and they are meant to be a solution to your problems. Since 2004 we have been approaching the problem of how can we send faster, how can we provide more flexibility, and how can we better meet the needs of prospective customers. In addition to our core FASP protocol we have a number of tools to try to make our software easier to use. These include things like web apps for Collaboration, Management, Sharing, and Distribution of content. Tools for synchronizing massive amounts of data, along with automation and policy tools.

Aspera optimizes the FASP protocol across the entire system domain, starting from Storage and ending with people consuming data. At almost every level of the transfer chain, you can tweak things to provide the policy you need within your organization. This includes things like different transfer priorities for different users, virtual links to shape traffic how you want, API's and file notification hooks.

At Aspera our goal is to provide the transfer solution you want and we are more than happy to work with you until things are exactly what you are looking for. Our solutions currently scale to hundreds of gigabits/s per system when utilizing our in development Next-Generation FASP, and they scale out to utilize as many systems as you have when performing parallel transfers.

Cloud, Hybrid, or On Premise Aspera FASP provides the performance and throughput to minimize the time you spend waiting on content and maximize your ability to consume and provide content.

ADVANCED WIRELESS POSSIBILITIES

Carol Ansley
ARRIS

Abstract

Wireless connectivity is a constant expectation from modern consumers. They expect their devices to work seamlessly everywhere: at their workplace; on their commute; and in their home, even as they might be migrating from an enterprise Wi-Fi system at work to a cellular system on the road to a consumer class Wi-Fi solution at home. Broadband providers must be ready to meet these needs today and tomorrow.

This paper considers how consumers are using Wi-Fi today at home and extrapolates about how Wi-Fi services will be used in the future. Alternative home networking architectures are discussed with attention to how they can meet the needs of new services and new devices. A closer look is given to the capabilities of 60 GHz Wi-Fi® and how it may become the centerpiece of a new wireless home strategy.

INTRODUCTION

Wireless connections are ubiquitous in the consumer world, particularly Wi-Fi. A consumer expects to access any entertainment content on any device without thought to where they are or what the device is. Outside of entertainment, most general computing devices offer Wi-Fi connectivity options, such as printers or laptops. Even more generic consumer goods, such as refrigerators or dryers, are becoming a part of the Wi-Fi universe within the world of use cases generally called the Internet of Things. Home security systems are also moving into Wi-Fi-enabled devices to reduce the installation complexity and provide superior services, such as full-motion video recording coupled with motion detection sensors.

We will review recent Wi-Fi data for insights into how Wi-Fi is being deployed within the home in 2016. These results will be compared with results from similar studies a year earlier, where relevant.

Next, we will consider how consumers may want their Wi-Fi and wireless services to evolve within their homes over coming years. As a part of examining those services, we will also examine alternative architectures for future Wi-Fi and wireless deployments.

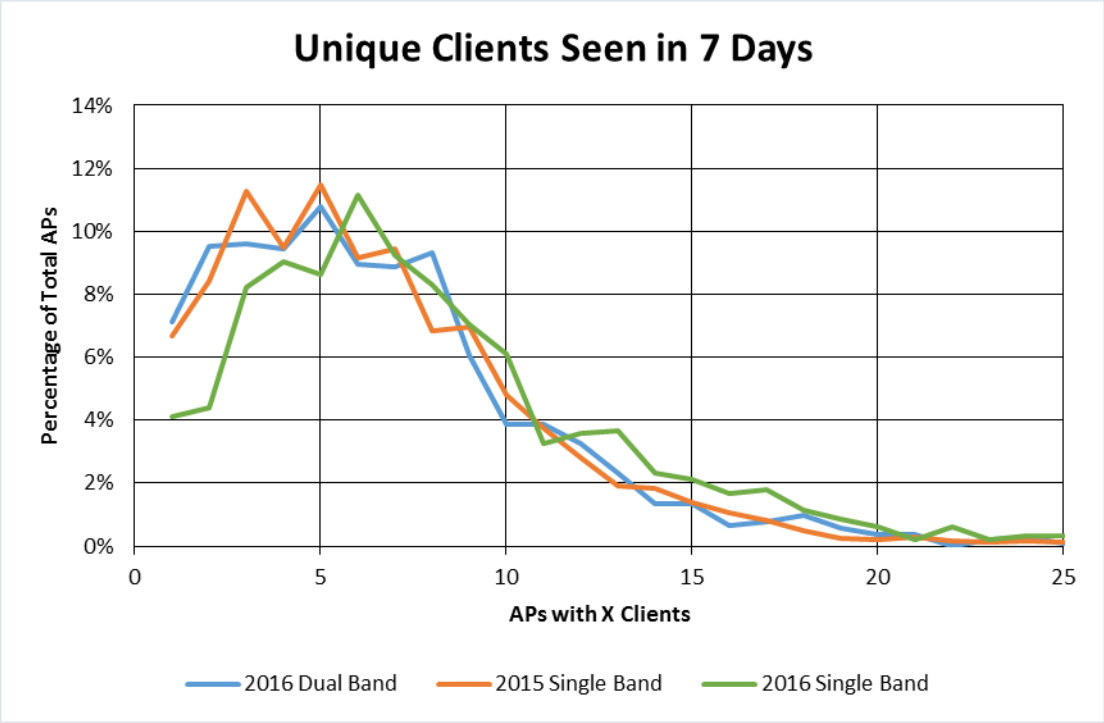
The potential of 60 GHz Wi-Fi, or WiGig, in future wireless architectures deserves a closer look. WiGig has been lurking in the background as a potentially significant technology for several years and finally had some significant announcements at CES 2016. We will provide a brief overview of the advantages and tradeoffs inherent to 60 GHz Wi-Fi networking.

TODAY'S RESIDENTIAL WI-FI USAGE

Many operators around the world are deploying DOCSIS gateways with embedded Wi-Fi Access Point (AP) technology. Most new deployments are using dual band concurrent APs with a 2.4 GHz radio and a 5 GHz radio, but a substantial population of single band AP gateways, usually 2.4 GHz, is also still deployed. In the larger consumer marketplace, stand-alone APs with two 5 GHz radios have also appeared on the market.

In a population of operator-provided APs studied recently, the presence of dual band radios versus single band radios in APs did not impact the number of clients used by a home, though it did appear to improve the quality of service seen by those clients since splitting traffic across two bands significantly reduces the overall congestion.

Each AP typically supported an average of just under 7 clients in both surveys, but the number of clients on any one AP varied widely as seen in the following graphic with some households seeing 20 clients or more. Also, interestingly, the distribution of clients per AP was very similar to that seen a year ago, suggesting the constant year over year growth in Wi-Fi clients may be slowing. The 2016 Single Band client count did show a higher mean. This may suggest that when users move to a new AP, they may not carry over all of their older devices, but while they remain on an AP, they do not bother to explicitly remove little used devices.

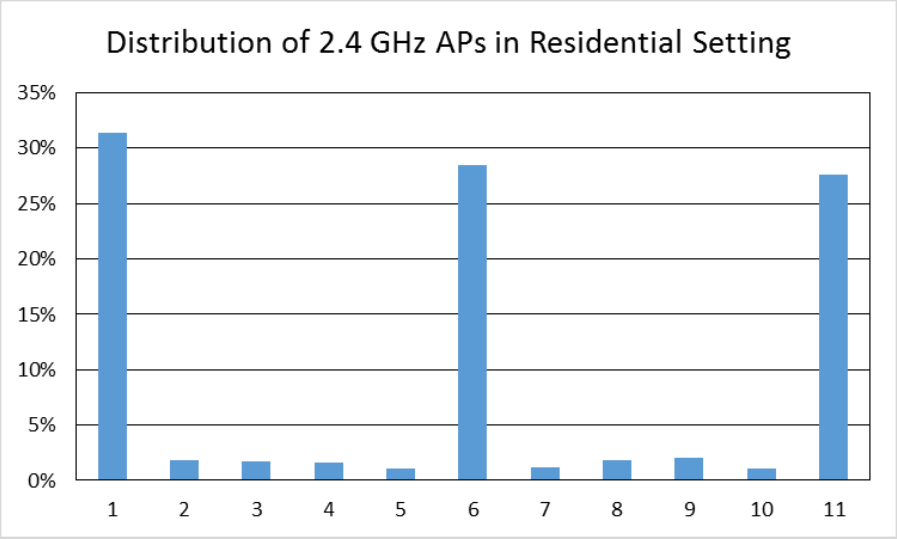


Present Wi-Fi deployments are primarily in support of high speed data deployments, with operator-deployed video set-top boxes just beginning to find widespread use. On the consumer side, however, video is already the dominant use for all Internet traffic, and in-home wireless video distribution to mobile devices and even some stationary devices within the homes. Returning to the 2016 survey of Wi-Fi usage mentioned earlier, over 55% of the devices connected to home APs were mobile phones or tablets. These mobile devices also tended to connect preferentially to the 5 GHz band when they were new enough to have a 5 GHz radio available.

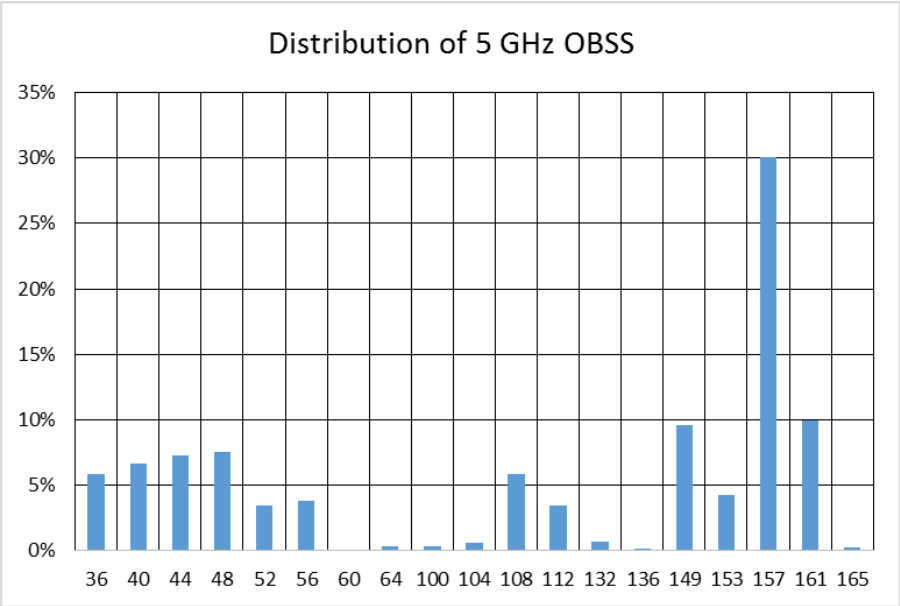
The other service quality consideration for current AP deployments is the amount of overlap between different APs and their associated clients. When two or more APs and their associated clients share a channel, all the clients experience increased congestion causing difficulty sending and receiving data. The 2.4 GHz band in North America has only enough bandwidth allocated to it to support 3 non-overlapping channels. Comparing 2015 to 2016, the 2.4 GHz band has only gotten more crowded. Data from 2015 found about 20% of GW APs with no overlapping 2.4 GHz traffic, but that dropped to only 3% in 2016. The 5 GHz band has fewer instances of overlapping APs in the residential environment, largely thanks to its inherent

propagation limitations. The 5 GHz band is also significantly wider, allowing more 5 GHz systems to coexist without overlapping. In 2016, 65% of the GW APs surveyed did not report any overlapping transmissions in the 5 GHz band.

This data suggests that in the future wider bands with more channels to spread out competing traffic is important for continuing reliability of Wi-Fi services. It also suggests that higher frequency bands with poor long-distance transmission characteristics may offer advantages to keep neighboring APs from causing congestion in the home. Channel usage data can also offer hints for future optimization. Generally, 2.4 GHz APs will perform at least a basic level of channel search trying to find the least loaded channel. This graphic shows that overall, 2.4 GHz APs are well distributed across the 3 non-overlapping channels.



Comparing the 2.4 GHz distribution to the 5 GHz distribution illustrates a potential problem. In the region where this data was gathered, one provider appeared to have a fixed frequency assignment for its APs, resulting in an uneven distribution of APs across the 5 GHz band.



A LOOK INTO THE FUTURE OF WIRELESS SERVICES

Many new services are rolling out into the market which could substantially affect consumer demand for Wireless connectivity. The wireless standards community is also moving along at full speed expanding the technology choices for future home architectures.

Future Use Cases

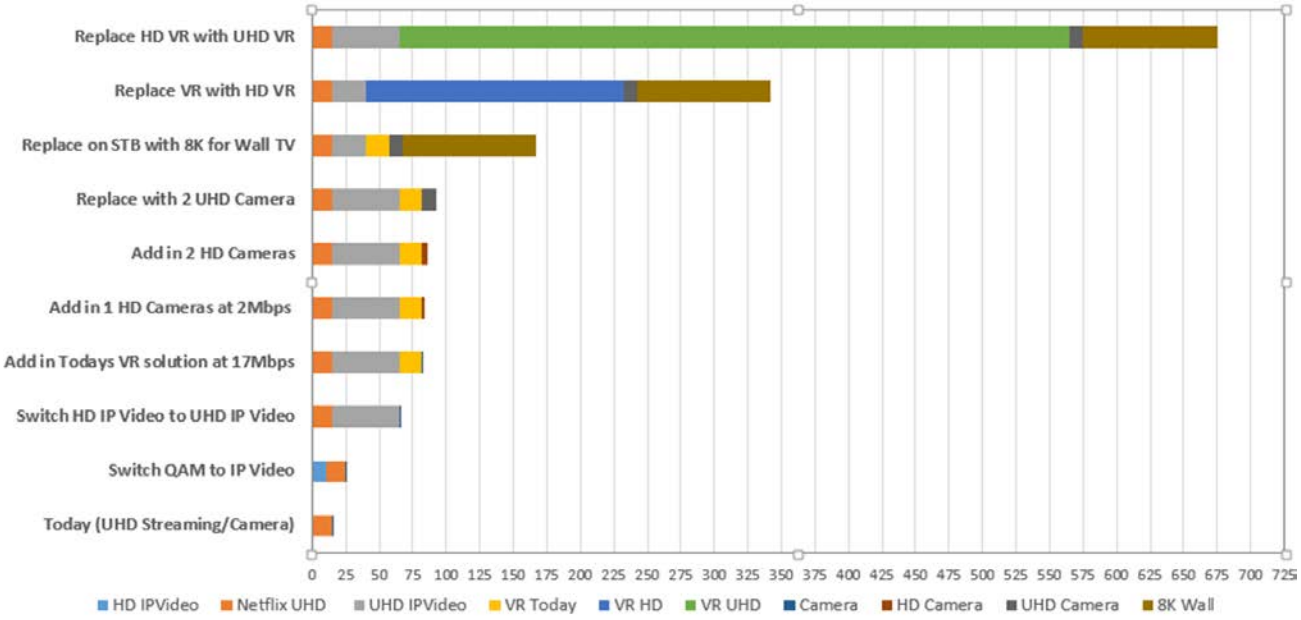
The future directions for wireless connectivity are moving in two separate directions: low power low bit rate, and high performance high bit rate. We will consider two areas, Internet of Things (IoT) and virtual reality, as examples.

The nascent IoT marketplace is driving a lot of innovation. One prominent IoT use case is home sensing and security. Some potential devices within this use case demand low power so that non-power line devices can have a long service life, such as door locks or window sensors. Some other devices might demand long range, but require only a nominal bit rate, such as a motion sensor placed on the perimeter of a home. Whenever video comes into play, such as for security cameras, higher bit rates and low latency are generally required.

To provide thorough coverage of IoT devices, multiple radio bands will probably be needed as well as potentially multiple wireless protocols. Within Wi-Fi, 802.11ah® can provide low power long range coverage at 900 MHz, as can 802.15.4 variants such as Zigbee® as 2.4 GHz. The crowded field at

2.4 GHz currently suggests that moving IoT devices to another band may be useful to ensure quality of service. IoT devices that require higher bit rates may be usefully served with the current Wi-Fi band choices of 2.4 GHz and 5 GHz.

Turning to future entertainment use cases, bandwidth growth is a virtual certainty (pun intended). The consumer electronics industry is already well into a switch-over to Ultra High Definition (UHD) televisions with 4K resolution. Television manufacturers have been showing 8K televisions for several years, though real volume market introduction is probably several years away. However, virtual reality (VR) shows signs of becoming popular in a way that other new viewing modes, such as 3D, have not. It differs from other video services because it requires a uniquely personal stream delivered with low latency to each user. Even if several people are sharing a VR experience, each one would require a different feed as they turn their head to different angles to see different parts of the VR experience. From a technology perspective, an ideal wireless distribution system must be capable of supplying multiple high bit rate streams with very low latency. Using the current 2.4 GHz and 5 GHz bands to support these new services, or others like them, may be problematic, since they are already challenged with the current traffic loads. The graphic below illustrates the potential growth in bandwidth related to VR and UHD usage.



New wireless options, like 60 GHz Wi-Fi, a.k.a. WiGig, may hold the answer to keeping entertainment traffic flowing smoothly in the home of the future. Transmission of 60 GHz signals is highly attenuated by anything solid, so transmissions tend to stay mostly within the space where they originated, and neighboring systems in other rooms do not tend to interfere. For rooms with devices that demand high bit rates, such as VR or 8K displays, going wireless by way of WiGig in-room hot spot/APs may be the future.

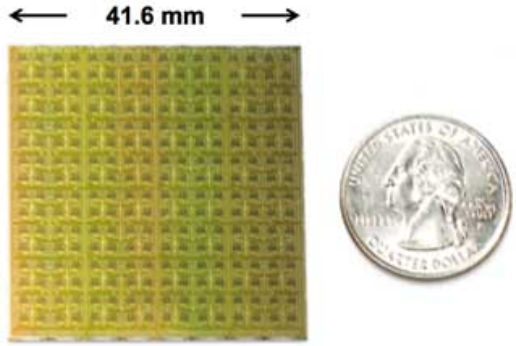
A Closer Look at 60 GHz WiGig

WiGig was mentioned earlier as offering a promising option for high bit rate wireless connectivity. The 60 GHz ISM band offers over 8 GHz of bandwidth, as opposed to 79 MHz in the 2.4 GHz band and about 600 MHz

in the 5 GHz band. The current 11ad products can use that bandwidth to provide high speed (over 6 Gb/s) mostly point-to-point links in indoor settings. Next generation products will provide higher throughput and more flexibility to support indoor multi-user scenarios as well as outdoor scenarios. As an example, the first 11ad router was announced at CES 2016 earlier this year. The Wi-Fi Alliance has also announced that they will begin certifying WiGig products for interoperability this year as well. The IEEE 802.11 standards committee is already working on the next generation of 60 GHz, currently known as amendment 11ay. It is projected for completion in 2020, but early pre-standard products will probably appear sooner.

On the technical side, using the 60 GHz band for high bit rate connections imposes interesting constraints. This high frequency band is blocked or heavily attenuated by most solid objects. A clear line of sight connection between the transmitter and the receiver provides the best throughput. Non-line-of-sight (NLOS) transmission is possible, but requires a high dynamic range from the transmitter/receiver pair. A mix of LOS and NLOS targets can be challenging because of the need for gain control or other means of enabling the high dynamic range needed.

Antenna selection and design is very important in WiGig. The optimal antenna elements are very small due to the high frequency/short wavelength. Because of the small antenna size, power transmission characteristics can be challenging. To overcome these disadvantages, antenna arrays are the favored implementation for 60 GHz. With the small individual element size at 60 GHz, antenna arrays can still be smaller than the current 2.4 or 5 GHz antennas. Shown below is a 256 element array that might be suitable for an AP. Smaller arrays may be more suitable for client devices.



256-Element Array Chip

Antenna arrays have the advantage of improving overall reception, but they also require more pre and post processing than current Wi-Fi solutions. Highly focused antenna patterns can provide high gain, but also narrow the useful angular window. If an AP supports clients scattered across a wide area, multiple arrays may be required to achieve uniform coverage, or the placement of the AP will need careful consideration. For example, a 60 GHz AP may be best placed in a corner of a room so that the AP sees clients distributed across 90° in the horizontal plane. It may be best placed near the ceiling, so that the clients are within a 90° spread in the vertical plane. These challenges imply that in the long term, WiGig APs may have a different form factor than today's Wi-Fi APs. On the next page is a graphic of a potential WiGig AP designed to light up a room with high bit rate services.

FUTURE WIRELESS HOME ARCHITECTURES

Putting these different threads together suggests that the home wireless networks of the future will have a complex network of wireless APs, serving different bands and potentially different use cases throughout the home. All the low power IoT devices may be served from one central IoT 900 MHz AP or if 2.4 GHz IoT devices are being used, then several APs may need to service IoT devices. Higher speed low latency services are best served with several distributed APs, perhaps even with one per room in the most frequented areas of the home.

Backbone Networking

The entertainment and Internet data we have discussed will enter the home from a broadband access network that could be wired, cable or fiber, or even wireless, but will probably be physically present at a single point at the edge of the home. That flood of data must be distributed around the home to the appropriate devices. While an IoT hub might be located near the broadband access point to serve the low power low data rate devices spread throughout the home, because of the challenges discussed earlier, high bit rate services would probably be more efficiently served by a network of APs. The APs could perhaps be controlled or directed to offer specific services within specific rooms. These APs would still need to be connected together by a backbone data network that reaches back to the broadband access entry point.

The backbone data network in the home is an important part of this distributed future wireless architecture. A wireless back bone might be used, if the transmission challenges can be met. A 60 GHz backbone with highly focused beams could be used, but to achieve the highest bit rates, line of sight transmission would be needed. Companies have also proposed using a wireless mesh to pass data around the home, when a single connection does not reach far enough, but the current 5 GHz and 2.4 GHz meshes have serious capacity problems when one tries to extend them to the data rates that will be required in the UHD VR future.

Wired backbones are also possible. MoCA and G.hn networks are flirting with Gigabit speeds today and will probably continue to improve with time. The biggest challenge that a wired backbone poses is how to get the data into the APs. A G.hn power line system might provide the data over the power line connection. A MoCA system might provide the MoCA data signal along with a center core power connection to minimize wiring for the distributed nodes.

Future Wireless Concerns

The unlicensed bands discussed in this paper have also attracted the attention of cellular operators. They offer the mobile operators available spectrum without the governmental regulations and high fees of their current licensed spectrum. The reason to note this alternative development is that it may affect the success of the deployment of the future wireless networks we have discussed. If a home user has an LTE-U home hotspot that operates in the 5 GHz band, then, the homeowner's use of the 5 GHz band for services over regular Wi-Fi may be affected and vice versa. As consumers use all wireless services more and more, this issue of shared spectrum may become challenging.

CONCLUSION

The wireless systems of the future may not resemble the systems of today since they will need to support radically increased amounts of data and new use cases. The consumer appeal of UHD and VR may determine how much the current systems have to change to ensure that there is enough bandwidth to satisfy the end users. New wireless technologies, like 60 GHz WiGig, offer interesting capabilities that can be used to meet those new challenges. Other new services, like LTE-U, may provide additional challenges to maintain and improve wireless data service within the home. Operators will need to monitor the growth of new use cases, and decide how best to keep the consumer's wireless connections on-line.

AN ARCHITECTURAL APPROACH TOWARDS ACHIEVING BANDWIDTH EFFICIENT CONTENT DELIVERY

Dinkar Bhat, Navneeth Kannan, and Wendell Sun
ARRIS

Abstract

Content delivery is a complex problem given the variety of services to be supported. Bandwidth availability and efficiency at high-end homes becomes key as high bitrate applications including 4K, 4K-HDR, and VR video become available. Performance is usually estimated in terms of latency and different types of latency measurement criteria could be considered. In order to achieve high performance across applications, we consider two aspects: modeling of viewership and efficient adaptive bitrate streaming delivery over IP. We delve into modeling viewer behavior using machine learning clustering techniques that use genre and demographics as features. The aim is to create effective content tranches to which bandwidth can be allocated more efficiently. We then examine the problem of adaptive bitrate streaming delivery of the content to the classes of viewers. In particular, we look at advanced encoding features at the video and transport level that can significantly improve user experience, while preserving video quality. We present the paper from an overall architectural perspective with deeper focus into the above described technical aspects.

INTRODUCTION

The number of services being deployed at homes continues to grow at a very rapid pace. With the introduction of bandwidth intensive applications requiring 4K, 4K-HDR, and now VR video, the architecture for content delivery has to change accordingly. Viewers regardless of underlying network complexity expect low latency, fast availability, and glitchless performance.

We take a look at two technical aspects that would significantly benefit content delivery. First, how should viewership in a network be modeled effectively in terms of their viewer profiles? We describe how they can be clustered based on viewing habits and demographic attributes. Once they are clustered, we argue how content sources can be sliced and linked to the clusters. This linking in turn aids in bandwidth allocation for a delivery mechanism like IP multicast. Second, what steps can be taken to deliver content sources (like live TV services) to viewers with minimal latency. We describe a hybrid network architecture that combines IP multicast with HyperText Transfer Protocol (HTTP)-based adaptive-bitrate (ABR) approach to provide low delay services to clients. We introduce a key component called the ABR-aware Packager/HTTP server that links the IP Multicast cloud with the ABR clients requesting content over HTTP.

VIEWER MODELING

Motivation

As the viewer population grows more diverse and fragmented, the type of content being consumed by different groups becomes quite varied and distinct [4]. The implication is that a program set, which could be a list of channels, can be organized into appropriate slices so that they can be targeted effectively. This would be more effective in terms of bandwidth allocation and developing user interfaces like guides [2]. Figure 1 shows an example of how channels of different genre may be classified based on viewer gender and age.

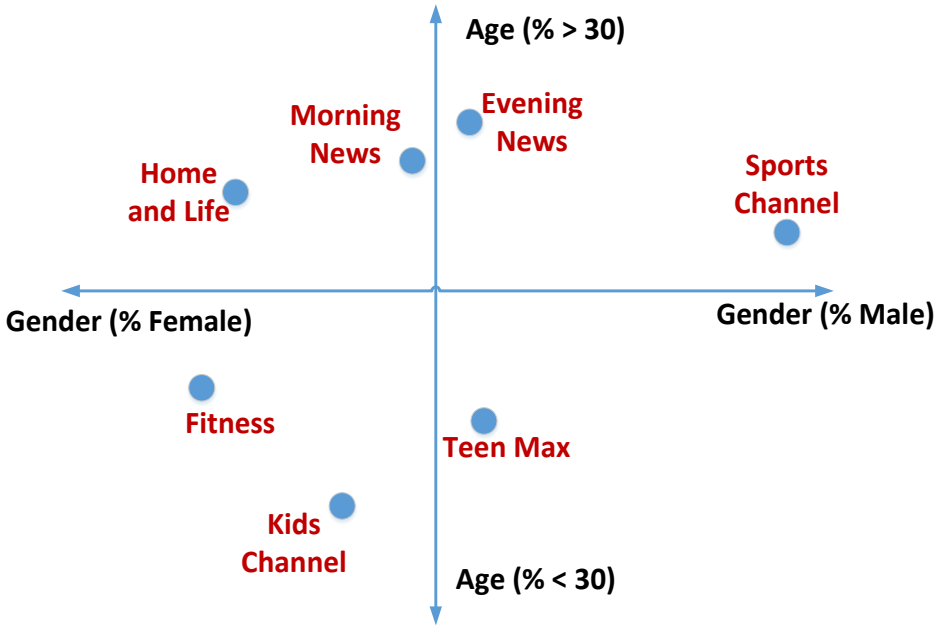


Figure 1: Depicts a sample classification of six different channels based on gender and age

Targeted advertising, which has been the topic of significant technological research due to its obvious commercial importance, is achieved by matching viewer profiles with a set of ads, in order to obtain addressable sets of viewers. Viewer profile attributes including age, buying habits, and income range are obtained from third-party market researchers. In addition, some approaches combine demographic profiles with viewing habits to target ads in specific channels or shows at specific times.

In this paper, we look at viewership modeling from a different perspective, namely the aim is to group viewers into clusters based on viewing habits and demographics attributes, then to use the resulting clusters to create content slices or *tranches.* Once the tranches have been created, they can be allocated appropriately to bandwidth groups (like IP Multicast groups) for transmission (see Figure 2). Note that the tranches can be modified dynamically based on temporal viewing characteristics.

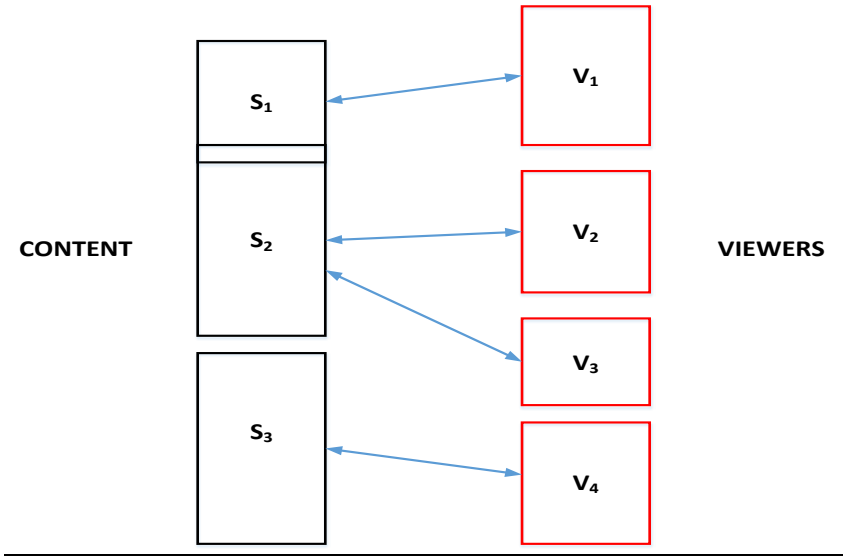


Figure 2: Illustrates a content set divided into tranches and viewer clusters assigned to the tranches

Machine Learning Approaches

Machine learning techniques applied to Big Data now offers significant opportunities into examining viewing behavior, which can then be used to provide better user interfaces, efficient bandwidth management, towards better user experience. In this section, we look at approaches to clustering of viewers based on demographic attributes and viewing behavior.

In general, the aim of clustering algorithms is to find groups within a data set, or clusters that are similar in some defined sense. Clustering algorithms fall in the domain of unsupervised learning techniques because the clusters that may be identified are not known a priori. There is no a priori knowledge of what kind of clusters are present in the data. Hence, a key input to any clustering algorithm is the number of clusters to be identified and often a lot of experimentation is required before an optimal value is used. Figure 3 shows a schematic of how clustering works in general.

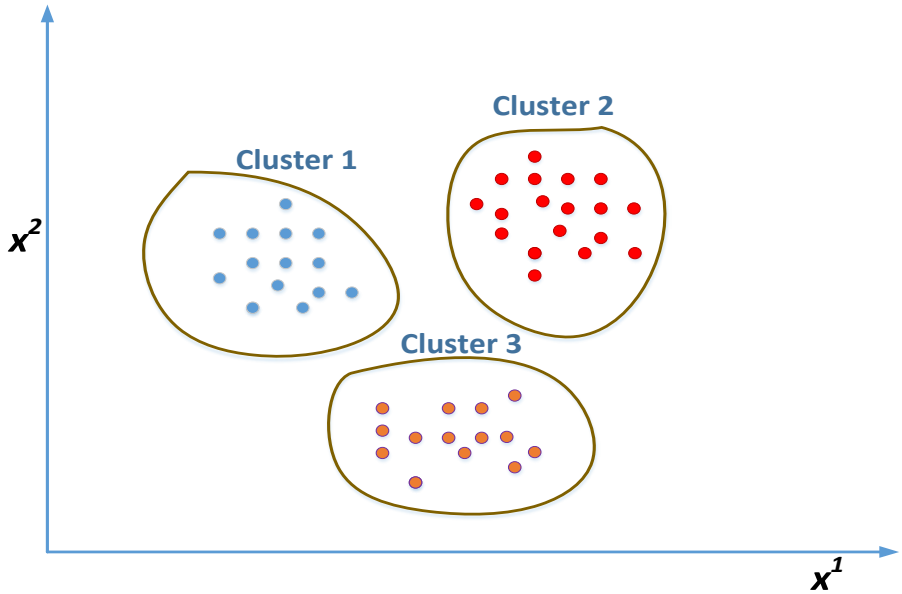


Figure 3: Depicts the possible output of a clustering algorithm (like k-means). The input to the algorithm is the set of data points, each represented by two attributes. Input also includes the number of clusters to be identified (k=3).

We briefly describe a popular clustering approach that could be used in our application, namely, the *k*-means algorithm. Formally, let the set of n data points in our data set be $D = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, .., \boldsymbol{x_n}\}$, where $\boldsymbol{x_i} = (x_i^1, x_i^2, \dots, x_i^r)$, $i = 1, .., n$ is a vector with r attributes. The *k*-means algorithm partitions the data set D into k clusters, given k. Each cluster has a cluster center, called a *centroid*. The algorithm works as follows: Randomly choose k data points as seeds to be the initial centroids. Repeat the following steps till convergence:

1) Assign each data point $\boldsymbol{x_i}$ to the closest centroid $\boldsymbol{m_j}$ of cluster j, using a measure which is usually the Euclidean distance.

2) Re-compute the centroids using the current cluster memberships.

3) If the centroids have not changed as per convergence criterion, break and output the set of N=k clusters.

During each iteration described above, the centroid $\boldsymbol{m_j}$ of cluster j containing C_j data points is computed as $\boldsymbol{m_j} = \frac{1}{|C_j|} \sum_{\boldsymbol{x_k} \in C_j} \boldsymbol{x_k}$. Selection of the convergence criterion is very important to generate the right set of clusters. The *k*-means algorithm was developed assuming data sets were small, but with large amounts of data its complexity increases prohibitively. Thus, it has seen many novel changes to adapt it to larger data sets (for example [3]).

In our application, each $\boldsymbol{x_i} = (x_i^1, x_i^2, \dots, x_i^r)$ denotes an instance of viewing. The components of the viewing vector could constitute both demographic variables like age, gender, income, and viewing variables like content genre components and time of day [2]. To prevent too much noise in the data, only instances of viewing where duration was greater than a threshold could be included. k would denote the number of viewer clusters to be output by the algorithm.

To create the content tranches and then match them with the viewer clusters, a simple

approach could be adopted as follows. Let's say the content set of *s* units is represented by

$E = \{\boldsymbol{p_1}, \boldsymbol{p_2}, \ldots \boldsymbol{p_s}\}$ where $\boldsymbol{p_j} = (p_j^1, p_j^2 \ldots, p_j^l), l \leq r,\ j = 1, .., s$ is a vector of size l that contains the same genre components used in viewer clustering. Cluster the content units using the *k*-means algorithm into a set of tranches of size *M*. Now link each viewer cluster (*N*) to one or more tranches (*M*) by comparing centroids using the Euclidean measure.

Bandwidth Allocation

Our approach of dividing content into tranches and linking them to viewer clusters fits well into an IP multicast delivery model [3]. IP multicast reduces traffic in a delivery network by simultaneously delivering a content stream to many recipients. Multicast routing establishes a tree that is rooted at the source with the receivers as the leaves. As opposed to unicast delivery, data in multicast is not copied at the source, but is copied inside the network at branch points of the multicast distribution tree. Thus, only a single copy of data is sent over links that lead to multiple receivers resulting in bandwidth gains. Given that several viewers in a cluster are likely to request content on the tranche they are linked to, the bandwidth savings can be quite significant.

The content tranches are assigned unique multicast addresses. Each receiver in a viewer cluster listens on multicast addresses assigned to the content tranches it has been linked to. Our model, in which several sources transmit on an IP multicast address, is more optimal for an IGMP v3 enabled network where source specific multicast (SSM) is available. In an SSM-enhanced network receivers can signal their intention to join a specific source within a multicast group, unlike a traditional network (i.e. before IGMP v3) where receivers will get traffic from all sources sending on a multicast address. Thus, in an older multicast network there could be unwanted network traffic and it is then up to the receivers to filter data.

Note that although we have assigned viewer clusters to content tranches, nothing prevents receivers from requesting content from other content tranches they are not linked to. This can be done through unicast. We expect that by modeling using content tranches and clusters, such unicast delivery will be more limited. However, the content guides on receivers would list all available content in an intuitive fashion [2] and it would be transparent to the viewer as to how it is delivered. In the next section, we describe a hybrid network architecture combining IP multicast with HTTP-based adaptive bitrate streaming.

ADAPTIVE BITRATE DELIVERY

Common Practice and Problems with Live TV Service

Internet-based video streaming has a long history. For example, IPTV services have been offered by telco operators to compete with cable operators for more than a decade. IPTV services provide subscribers with similar TV viewing experience, such as fast channel change time and low end to end transport latency, as traditional broadcast or linear pay-TV service offers. It uses IP multicast as primary protocol, which provides true data streaming, minimizes delivery latency, and supports content sharing among multiple clients.

However, IPTV services require guaranteed bitrate match bandwidth for smooth service delivering. As a result, an IPTV service is only provided in a managed IP network. Over-the-Top (OTT) services are relatively new to TV service. OTT uses HyperText Transfer Protocol (HTTP) protocol with introduction of multiple bitrate video

source, adaptive bit selection and seamless bitrate switching, which provide transport resiliency to network bandwidth changes, plus the usage of popularly deployed HTTP protocol. Consequently, OTT services are attractive and may replace IPTV as the primary video service in Internet-based video service delivery. However, the HTTP-based transport has major shortcomings, especially for live or linear video delivery.

The HTTP protocol is designed for transaction based file transfer. It may be perfect for a Video-on-Demand (VOD) type of video service, especially with progressive file download, where a user can watch the downloaded video even before the file download is complete. However, it is not quite suitable for live streaming. The current HTTP Adaptive Bitrate (ABR) streaming uses small segments to compromise HTTP file transfer request. Normally there is a manifest file that describes availability and URL of media segments. Media transport and playback process is initiated by HTTP ABR client via sending a HTTP request to get the manifest file, then the client selects a media segment in an adequate bitrate and starts download it. The client continues the media segment request one after another either in the same bitrate or in a different bitrate based on network bandwidth condition. Each media segment delivery is a HTTP transaction.

In the case of a live TV service, the manifest file will have to be updated to catch up with newly available segments while the media source moves forward in timeline. The client also needs to get the updated manifest file to view and select newly available media segments. When an HTTP ABR client asks for media segment, the HTTP server does not respond until the media segment is available. From a media transport perspective, the delay of live media transport is at least the length of segment besides other delays caused by media processing, such as encoding. The bigger the segment, the longer the delay latency is.

Taking a normal HTTP Live Streaming segment, for example, usually its length is 10 seconds, then the media transport end to end delay is about 15-20 seconds or more. If this forces us to create smaller media segment such as 1 or 2 seconds, to reduce the delay, it causes two other issues. For the reason of seamless switching between bitrate streams, the media segment is bordered by an IDR frame, which requires full decodable I frame. The smaller the media segment is, the more I frames are inserted. The more I frames exist, the more coding resource is required, thus the worse video encoding efficiency is achieved.

In addition, each segment delivery corresponds one HTTP get/reply transaction. The smaller the segment is, the more HTTP protocol traffic is in network.

This type of segmentation delay does not exist in an IPTV service that uses IP multicast technology. In this paper, we will propose a hybrid network architecture, which combines IP multicast in backbone network and HTTP in edge access network. With usage of HTTP Chunked Transfer Encoding [1], it minimizes the delivery latency for live/linear video service without sacrificing video encoding efficiency.

Hybrid Network Architecture

IP multicast has been used in IPTV services and is proven to be capable of delivering low delay media to TV viewers. Why cannot we continue to leverage its benefit and use it for ABR OTT service? At the same time, cannot we also preserve the popular usage and attractiveness of HTTP protocol, especially when it interacts with ABR client?

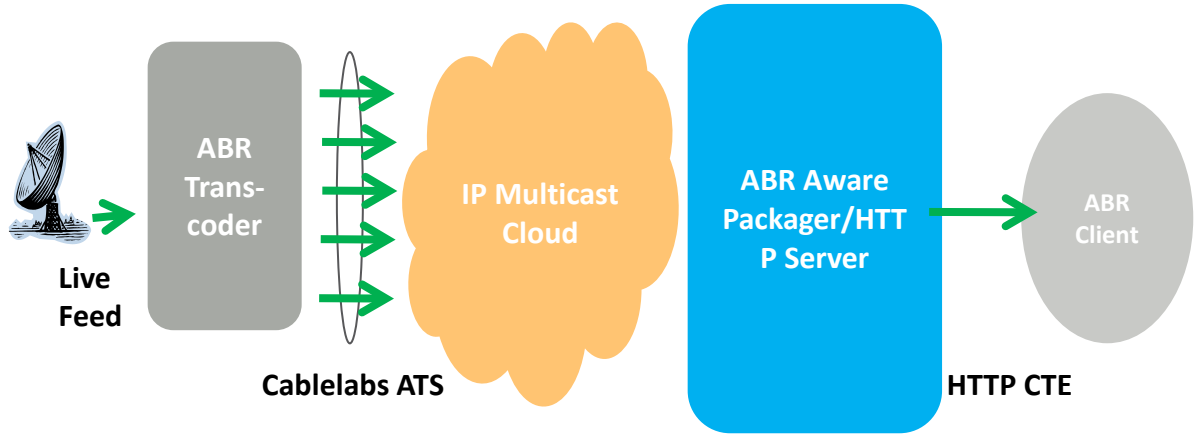


Figure 4: IP Multicast + HTTP Chunked Transfer Encoding

Figure 4 shows the hybrid network architecture, in which IP multicast is used in backbone network and HTTP is used in client oriented access network.

The live source is prepared by an ABR transcoder to create multiple bitrate streams per ABR OTT service profile. These streams are MPEG-2 Transport Stream (TS), just like legacy IPTV media streams carried in IP multicast network. The difference between the IPTV case and ABR OTT scenario is the number of streams. The IPTV has one single stream while the ABR OTT has a set of streams.

The IP multicast cloud is a backbone transport network. In general, the backbone network is well designed and engineered to meet media transport requirements. For example, the IP multicast cloud should be able to deliver all ABR streams in the case each of them is requested by a client. At the minimum, it should guarantee delivery of the stream in the lowest bitrate to keep smooth playback in client.

A new component, the ABR Aware Packager and HTTP Server, is introduced in Figure 4. It is different from a normal ABR HTTP server, which receives media segments from an ABR packager and simply serves ABR client via HTTP protocol. Instead, it is a Just In Time Packager (JITP) that can accept MPEG-2 TS carried in IP streams and generate client requested media segment **in real-time** when it responds to the client request for a media segment.

Similar to a normal ABR HTTP server, the JITP/HTTP server composes and publishes an ABR manifest file. For live/linear TV service, it just needs to tell what bit-rate/resolution streams are available with a virtual segment URL. When it receives a media segment request from ABR client, it checks if the requested bitrate stream is available or not. If the bitrate stream does not exist, it will join the IP multicast group of that bitrate stream. After the server receives the IP multicast stream, it starts to create the requested segment and prepare to serve the ABR client.

As we discussed in the previous section, if the JITP/HTTP server does not send the client requested media segment until the segment is ready, it still bears the extra segmentation delay to the media transportation. What we expect is to send the media segment incrementally in small chunks while the JITP/HTTP server receives and prepares it.

HTTP Chunked Transfer Encoding

The HTTP Chunked Transfer Encoding (CTE) is defined by HTTP/1.1: Message Syntax and Routing [1]. It is a data transfer mechanism in HTTP 1.1, in which data is sent in a series of "chunks". It uses the Transfer-Encoding header, instead of the Content-Length header. The HTTP sender does not need to wait for the total size of content and can start sending data as small chunk with any amount available while still receiving the content. When the chunk is sent, its size is indicated in the Transfer-Encoding header. If the chunk length is set to zero, then it is the end of content.

This is what exactly we want, isn't it? When the JITP/HTTP server receives media segment request from ABR client and the corresponding IP multicast stream is available, it starts the segmentation process, but it can start sending the media chunks in meaningful small size, such as small as one video frame, while it still receives the IP multicast stream. At the end of segment, it just needs to send a zero sized chunk to finish the segment transfer. And then the ABR client repeats sending segment request one after another, the JITP/HTTP server continues serving it with incrementally delivered chunks of segment. In this way, the delay caused by media segmentation is eliminated.

Combining with the usage of IP multicast in backbone network, this hybrid network approach, together with the help of HTTP CTE, can reduce the end to end transport latency to minimum and makes ABR OTT service be comparable with IPTV service.

CONCLUSION

We described an approach that combines viewer modeling with a hybrid network architecture to delivery content to ABR clients. The aim was to allocate network bandwidth effectively based on viewer clusters and then to deliver the content using a combination of IP multicast and HTTP-based ABR streaming to clients. Combining technologies is likely to provide much better performance with the advent of high bitrate services.

1. IETF RFC7230, Section 4.1, "Chunked Transfer Coding", http://tools.ietf.org/html/rfc7230#section-4.1

2. Bhat D., Kannan N., "Adaptive Television User Interface Using Machine Learning Concepts", IBC 2014, September 2014.

3. Legout, A., Nonnenmacher, J. and Biersack, E. W. , "Bandwidth Allocation Policies for Unicast and Multicast Flows", Infocomm 99, March 1999.

4. Cohen, J. "Television viewing preferences: Programs, schedules, and the structure of viewing choices made by Israeli adults", Journal of Broadcasting & Electronic Media, June 2002 204-221.

AN ECONOMIC ANALYSIS OF BROWNFIELD MIGRATION CTTH VS. FTTH

Michael Emmendorfer
ARRIS

Abstract

Cable service providers are interested in the network technology options and the economics to meet the highly competitive high-speed data and video delivery market needs. Many cable operators are facing fiber to the home (FTTH) competitors that are offering or capable of offering gigabit per second data services or higher. Many cable operators have determined that new build or greenfield areas, where coax to the home (CTTH) does not exist, will use FTTH. The majority of the cable network is brownfield, which already has Hybrid Fiber Coax (HFC) to the neighborhood and CTTH. Throughout the cable industry there is interest in understanding the options for the brownfield migration. Cable operators are wondering if they should continue investments in HFC and DOCSIS or forego the current brownfield coaxial network and build FTTH using passive optical network (PON) technologies.

This paper focuses on some of the network technologies choices and the economics of the brownfield migration. The paper expands the analysis that was published at the 2015 SCTE Cable Tech EXPO, which examined brownfield migration options for a specific set of network topology assumptions [EMM]. The SCTE paper examined systems that had high spectrum (750 MHz or higher), 75 homes per mile and one hundred percent aerial network topologies, though this is representative of some MSO deployments, like those found in suburban areas, there are many cable operators that have different network topologies. This paper addresses different brownfield migration networks like those found in urban areas. In urban areas for example, the cable operator will likely have much higher homes per mile, such as over 100 homes per mile, and have a mix of aerial and underground networks, these change the economics. The impacts of urban brownfield migrations may yield far different results than those found in typical suburban topologies. Typically, the approaches examined from brownfield migration are system wide build outs meaning that all customers in the serving area may have access to the upper tier services and capacity needs for the business. The paper will analyze another build out approach, called success based network builds. The success based approach does not build out high capacity networks across the entire serving area like the system wide approach, but rather is a targeted success based network build for addressing the competitive challenge of gigabit or multi-gigabit symmetrical service offering to residential consumers.

OVERVIEW

The 2015 SCTE companion paper is not an apple-to-apple comparison to this paper for the reason cited in the section above where the SCTE paper used 75 homes per mile and this paper uses 114 homes per mile. The 2015 SCTE paper assumes 14.1 miles of total plant in the 1060 HHP serving area and this paper assumed 8.4 miles of total plant in a 960 HHP serving area, the difference in total miles and not much difference in total HHP serving area is because the variance in the HHP per mile between the markets. The SCTE paper may represent a suburban market and this paper may align with urban environments. The presumed reduction in costs for a market that has greater HHP density, fewer network elements, fewer plant miles, while maintaining a similar sized serving area should translate to lower costs if labor rates are the same. Though another difference between the SCTE paper and this paper is the aerial versus underground plant. The SCTE

paper used 100% aerial, which is far cheaper to build out than underground plant. This was selected in an effort to show that even under the most ideal environments for overlaying plant, especially for FTTH builds, the cost comparison between leveraging existing plant versus overbuilding new plant is vast. This paper assumes a 73% aerial / 27% underground mix, this could see higher costs if all the variables were the same against 100% aerial plant. This paper also assumes newer constructions and slicing techniques than the SCTE paper. Overall we observed marginal savings when considering the two outside plant builds. The DOCSIS capacity in the 2015 SCTE paper assumed 64 DOCSIS downstream channel equivalents for 2.4 Gbps of capacity and this paper used 48 channels, both papers continued to use the price per channel of $399.00, which is based on the Infonetics report from Q4/2014. The actual price per channel an MSO may pay could be less when assuming the licensing channel capacity for existing systems or perhaps even new systems. The papers are not meant to be an apples-to-apples comparison of suburban versus urban build out, because not all factors are consistent between both market samples.

The analyses are meant to show that under similar network assumptions we can illustrate relative difference for a particular set of use case migration options and should be used for discussion purposes only. All economic estimates may not represent actual build and equipment costs. The estimates expressed in this material are subject to change without notice. Reasonable people may disagree about the opinions expressed in our analysis and results. There are many other factors that were not considered that may impact the estimates and illustrate a significantly different result. It is strongly recommended that individuals and companies perform their own analysis based on their environment and costs.

Purpose of the Analysis

Should MSOs continue with HFC/DOCSIS investment?

Should MSOs overlay the existing HFC network with FTTH and PON?

What are the economics both from a financial and capacity perspective?

What are the Economics of CTTH & FTTH brownfield migration options to support:

- 1 Gbps downstream service to be available to 100% of the HHP in brownfield networks?
- 1 Gbps symmetrical service to be available to 100% of the HHP in brownfield networks?
- 1 Gbps symmetrical service to be available to 30-50% of the HHP in brownfield networks?

Scope of the Analysis

- Leverage existing brownfield investments
- Brownfield Capital Assessment from the headend, outside plant, and CPE
- Brownfield Migration of CTTH vs. FTTH in the single family unit (SFU) market
- Consider the service drivers of video and data services as well as the typical take rates

The scope of this paper is to provide an economic and capability analysis of several brownfield single-family unit (SFU) brownfield migration options, thus greenfield and multiple dwelling units are out of scope for this analysis. The analysis considers several HFC and DOCSIS migrations and several 10G EPON deployment architectures depending on the topology constraints of the MSO. The paper will summarize the

economics and capabilities of several brownfield CTTH and FTTH migration options. We will conclude with summary and a set of recommendations.

DATA CAPACITY COMPARISONS

In terms of the data capacity capabilities there can be vast differences depending on the media and technology evolution. Currently there is work underway in all the technology areas to expand the capacity over each media, copper, coax, and fiber. There are also different access architectures underway to maximize the reach and capacity as well. Please consider the image below a snapshot in time and under a given set of assumptions of the capacities of different technologies. The figure below provides a comparison from the current leading technologies, as with most technologies. These estimates account for physical layer (PHY layer) overheads after encoding and forward error correction (FEC) overhead if used and no other overhead is assumed.

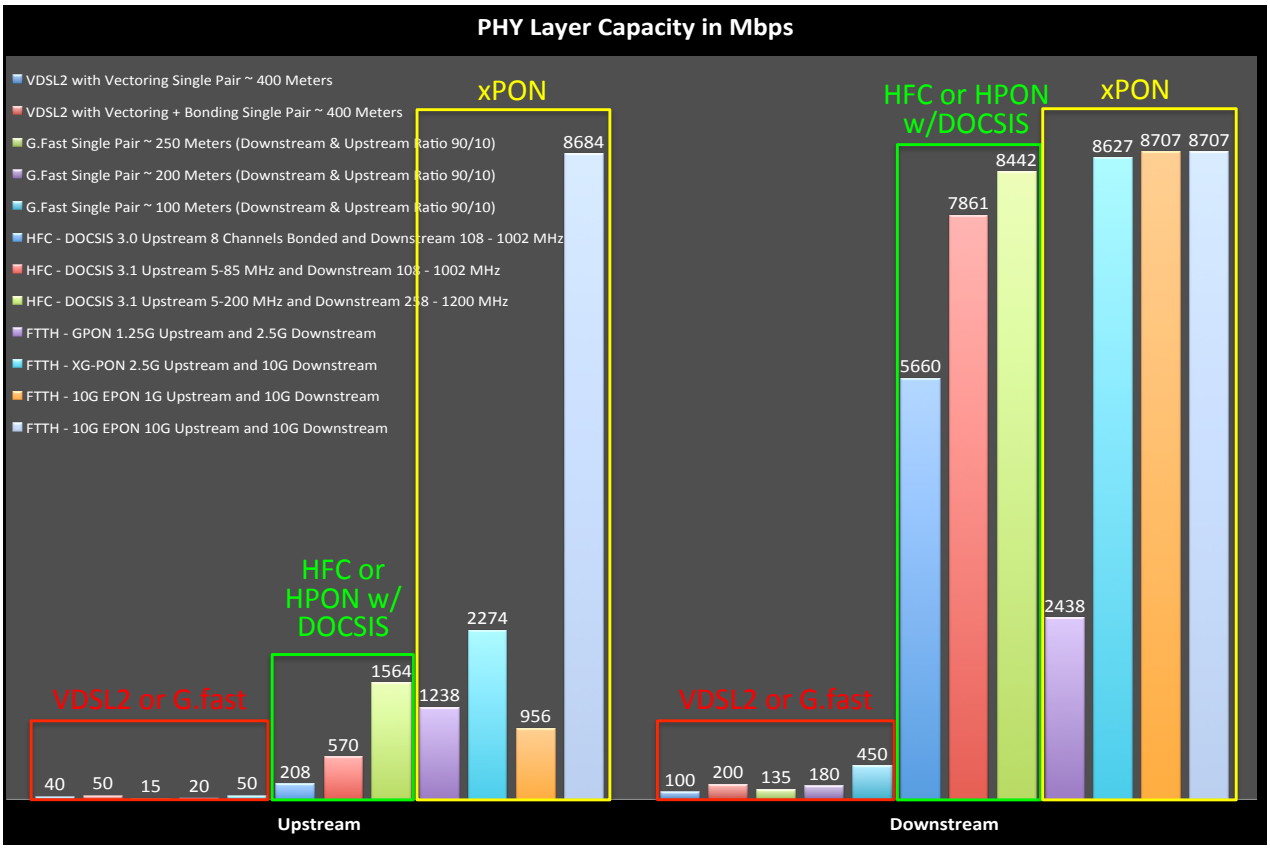


Figure 1 - Access Network PHY Capacities

The figure above estimates the PHY layer rates of DOCSIS 3.1 assumes an average of 2048-QAM in the downstream and 1024-QAM in the upstream.

GREENFIELD VS. BROWNFIELD

The investment in new build or greenfield represents a very small percent of the capital spend of a cable operator each year. Sometimes new build is referred to as line extension, because this is extending the network perhaps to the next neighborhood. New build is also tied to the new housing market when telcos and MSO build service to

new homes. Existing homes not yet wired with cable is also considered greenfield. These may represent single digital percent increases in total homes passes yearly and also a small portion of the capital spends. It is generally considered a foregone conclusion that greenfield is moving to FTTH, however some new build HFC is used for some line extensions depending on overall economic drivers. These days' the cable operators often select RFoG or HPON as the FTTH technology driven mainly by reuse of the existing data and video systems and the lack of readiness for IPTV. It is widely considered that when IPTV is ready for the full video service offering traditional PON technologies from the EPON or GPON family of standards will be deployed instead of DOCSIS over RFoG.

The largest network investments are directed to sustain and expand the capacity of the existing brownfield network. Most cable operators are leveraging their existing coax to the home (CTTH) network rather than overbuilding FTTH in the single-family unit (SFU) market. In highly competitive SFU markets some MSOs have built fiber to the home. The brownfield overbuild is mainly targeted at the highly competitive multiple dwelling units (MDUs) market, generally driven by the MDU property owner's desire to market Fiber to the Unit (FTTU). So MSOs overbuilding existing HFC in the SFU market are extremely rare. Most telcos worldwide want to leverage their existing copper network.

Many Telcos worldwide are continuing to invest in their brownfield copper network rather than building FTTH. They expand capacity in the twisted pair copper network by pushing fiber closer to the subscriber, which enables higher data rates to each subscriber. Some telephone operators decided to overbuild their existing copper network with fiber to the home, due to concerns over the capacity of the twisted pair network (refer to figure 1) supporting high bit rate video and very high speed internet service tiers. The telcos overbuilding their existing copper networks may have selected GPON technology capable of more than 2 Gbps downstream and 1 Gbps upstream, as seen in Figure 1. The evolution of GPON includes NG-PON2 with capabilities of 10 Gbps symmetrical (not accounting for PHY overheads) and this may co-exist with GPON on the same fiber separated by wavelength. The build out of FTTH by some telcos and overbuilders created competitive pressures for the MSOs. Some MSOs are wondering if continued investment in brownfield markets with DOCSIS and HFC is the best path or should they too overbuild their network with fiber to the home. The situation is certainly not the same for the telco with copper twisted pair faced with the capacities of cable competitor with fiber to the neighborhood and coaxial cable to the home. The capacities are very different between twisted pair and coax as seen in figure 1 and explored in detail in the companion paper [EMM].

NETWORK MIGRATIONS METHODS FOR CABLE BROWNFIELD

There are many different methods used to create capacity in the cable networks. Network optimization is a very popular approach because this allows the operator to place more capacity in the same amount of spectrum. Reducing the amount of customers sharing a data link also creates more capacity. Essentially this is just building another highway, and splitting the traffic, node splits, node segmentation, and fiber deep all reduce the number of subscribers sharing the link. Some methods increase the size of the data link by replacing network equipment to expand spectrum, called a system upgrade and in some cases moving to fiber deep is an opportunity to expand spectrum. Another approach is to overbuild the network with FTTH to increase capacity, as seen in the following sections this is very costly. The list

below captures the network migration options for cable brownfield.

- **Network Optimization**
  - Improve the efficiency of the network (encoding and more b/s/Hz)
  - Spectrum Reclamation to better network delivery technology Analog → Digital → IP/DOCSIS
- **Node Splits / Segmentation "Business as Usual (BAU)**
  - Maintain existing spectrum
  - Targeted Node Splits (smaller service groups to address capacity)
- **System Upgrade**
  - Spectrum Expansion (perhaps different spectrum split)
  - "Drop in Upgrade"
  - Maintains node and amplifier spacing
  - Approach typically good for an extra 200-250 MHz of spectrum from the originally designed system
  - "Rebalancing System Upgrade" new node and amplifier location and change in direction signals travel is possible
- **Fiber Deep**
  - Places fiber close enough to the customer often to remove all amplifiers in order to reduces the number of actives
  - Removing amplifiers drivers households passed per node way down (125 - 30 HHP per node)
  - May include spectrum expansion and perhaps spectrum split change
- **Fiber to the Home/Unit Overbuild**
  - System Wide Network Build to FTTH/U
  - Success Based Network Build to FTTH/U

NETWORK BUILD APPROACHES FOR CABLE BROWNFIELD

There is two types of network build outs examined in this paper, the first is known as system wide and the second is known as success based builds. The system wide approach is used for greenfield builds and may also be used from brownfield overbuilds.

This method builds out the network from the facility, headend or central office, to the curb or backyard of possible subscribers, perhaps within 150 feet of the home; this portion of the carrier build out is called enablement capital. The carrier is making a massive investment upfront in placing the network so close to the home that may not ever become a customer. The carrier makes this upfront enablement capital investment in the belief that a suitable number of homes will become customers. Additionally, having the network very close to the home will reduce the initial service installation times, to perhaps a day.

The paper also analyzes success based network builds. The success-based approach does not build out the network to pass all the homes but rather to actual customers. The success-based approach has been used to overlay an existing HFC network for the support of commercial services applications for nearly two decades. This is used when the HFC may not be able to sustain or economically justify the capacity allocation for a particular service, which may also have just a few customers. Though many commercial customers are supported on the HFC and DOCSIS network, when service speed tiers exceed the current capacity allocation of HFC/DOCSIS the MSO then builds fiber directly to those commercial customers, using a success based build approach. The success based network build can addresses the competitive challenge of gigabit or multi-gigabit symmetrical service offering to residential consumers. In an effort for MSOs to maintain a leadership position in the residential speed tier wars, some MSOs

have adopted this approach for the residential customers. This paper examines both the system wide based and success based network build approaches. The figures below illustrate both the system wide based builds and the success base builds, highlighting some the characteristics of each.

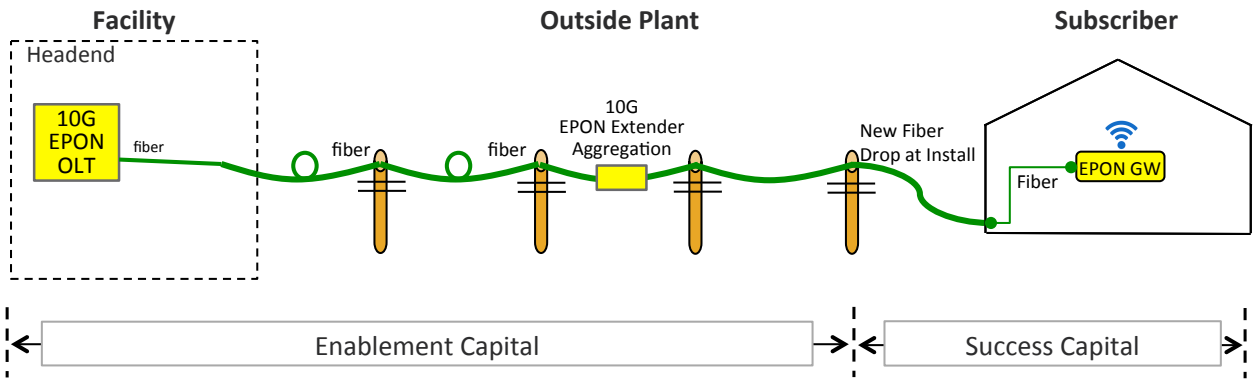


Figure 2 - System Wide Based Approach

System Wide Based Approach

- Proactive investment that is not targeted to any single customer
- New Build, Rebuild, or Upgrade that happen across a wide serving area
- Generally is assumed to be service that will have good take rate
- System wide approaches places the network to within 150 feet of all homes in the serving area
- Connecting an actual customer requires a drop installation for the remaining 150 feet to the home
- Service Turn up can be same day

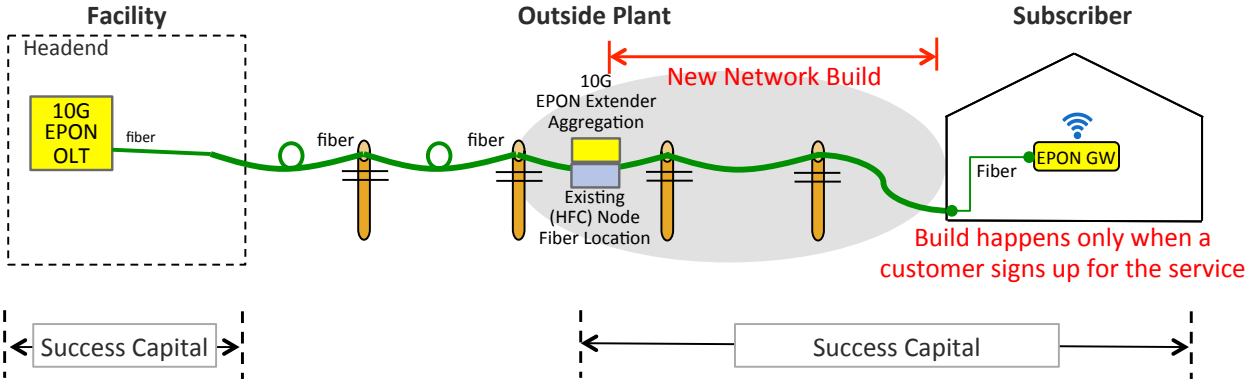


Figure 3 - Success Based "Targeted Capital" Approach

Success Based "Targeted Capital" Approach

- Reactive investment that is targeted to a single customer
- Network Build happens when a customer signs up for the service
- Success Based Capital often defines limit to the build from the "existing" fiber location

- Cost effective reachability depends on many factors (Aerial, Underground, other build cost impacts)
- Success Based approaches often requires very long service turn up window (plan, build, install, etc.)
- Service Turn up can be weeks to over a month
- Approach used for Business Services (Network Builds are targeted to a customer)
- Approach used by some MSOs to offer the very highest residential service tiers
- Some MSOs may have a HHP Serviceable range of 30% to 50% within 1/3 of a mile of a node, depending node location and the number of HHP per mile.

BROWNFIELD MIGRATION USE CASES

The analysis is a capital expenditure estimates and no operational costs have been included in this analysis. The capital estimates will assume leveraging the existing fiber to the node and HFC where applicable. This paper focuses completely on the existing "brownfield" network migration options for single family units (SFUs) "homes" and the only the access network build from the CCAP or OLT in the headend, outside plant to the customer premise equipment (CPE). The economics of the multiple dwelling Unit (MDU) market is very different than the SFU market. Competition in the MDU market may drive the business to replace coax to the unit (CTTU) with Fiber to the Unit (FTTU) and the costs are extremely different as well as the business relationships with the MDU owners and end users. The new build "Greenfield" is not the focus of this material but some data comparisons could be leveraged from this material to illustrate cost differences.

This paper will examine five use cases for using the system wide build approach and three use cases using the success based build approach.

System Wide Approach

1. HFC 2G DOCSIS
2. HFC 2G x 200M DOCSIS
3. HFC Mid-split 2G x 500M DOCSIS
4. FTTH 10G EPON
5. FTTH 10G EPON + PON Extender

Success Based "Targeted Capital" Approach

1. FTTH 10G EPON
2. FTTH 10G EPON + PON Extender
3. FTTH 1G/10G P2P Ethernet

Key Assumptions:

- Please read the disclaimer regarding the analysis
- The use cases examine a single area with the following characteristics:
- Total plant (miles): 8.4
- Total serving Area: 960 HHP
- Subscriber Penetration (% of HHP): 50%
- HFC optical serving area: 480 HHP or 240 HHP
- DOCSIS Service Group: 2 at 480 HHP with 240 Subs (use case 1 and 2)
- DOCSIS Service Group: 4 at 240 HHP with 120 subs (use case 3)
- 10G EPON Service Group: 8 at 120 HHP with 60 Subs (use case 4)
- 10G EPON PON Extender SG: 4 at 240 HHP with 120 subs (use case 5)
- Homes per mile: 114
- Aerial / Underground: 73% / 27%

The use cases requiring DOCSIS capacity assume 48 DOCSIS downstream channel equivalents for 2.0 Gbps of capacity and the SCTE 2015 paper used 64 channels, both papers continued to use the price per channel of $399.00, which is based on the Infonetics Q4/2014. The Infonetics price may not represent the actual price per channel an MSO may pay for a 48 downstream and mid-split upstream fully loaded system. The DOCSIS

use cases 1, 2 and 3 all assumed that 48 downstream channel equivalents include support for mid-split upstream. The 48 downstream channels are purchased for each DOCSIS Service Group. Use Cases 1 and 2 cannot utilize the mid-split spectrum, but both use cases assumes this is part of the 48 channel downstream purchase. Certainly some MSOs would explore ways of purchasing the upstream capacity that was needed since these DOCSIS use cases have different upstream DOCSIS needs.

This paper added more labor costs for the CPE installation due to the time for the wireless service activation. The 2015 SCTE paper and this paper are not meant to be an apples-to-apples comparison of suburban versus urban build out, because not all factors are consistent between both market samples as mentioned in the overview section. This is intended to show a comparison of each use cases against each other in a particular serving area.

SYSTEM WIDE USE CASES FOR BROWNFIELD MIGRATION

The system wide use cases enable the services to be available to 100% percent of the homes passed and the paper assumes a 50% take rate of the data service. The 50% take rate primarily applies to PON because optical budget and a certain threshold of users per port drive the OLT sizing. The enablement capital for the PON used cases we used the estimated 50% subscriber take rate. The DOCSIS CCAP sizing on the other hand can support many subscribers per port and cable operators manage the number of subscribers per port based on traffic load and service tiers supported. The use cases state the DOCSIS service group HHP served. In the use cases, the success capital is based on a per subscriber basis. An important point to understand about the models is that the enablement capital is shown on a per HHP basis and the success capital is shown on a per customer basis, this is a very important distinction.

The brownfield HFC and DOCSIS use cases, 1 through 3, provide the estimated enablement capital to allow the service to be available to all homes in the service area. The cable operator would not need to consider the replacement of all of the cable modems, but rather just the for their estimated take rate of the 1 Gbps service. The success capital number include estimates equipment and labor for DOCSIS 1 Gbps downstream and several upstream speeds. Additionally, the DOCSIS use cases 1 and 2 assumes that 48 six MHz channels are available and did not allocate capital to opening up spectrum if all 48 channels were not available. The HFC and DOCSIS use case 3 assumes 48 six MHz channels of DOCSIS capacity, however this did assume a spectrum expansion from 750 MHz to 1 GHz, thus creating 250 MHz of new spectrum that presumably is used to support DOCSIS service. The use case also assumed a node split since a portion of the downstream would be consumed for the mode to mid-split. In the sample 480 HHP were sharing a downstream spectrum range of 54 MHz -750 MHz and the enablement investment 240 HHP were sharing a downstream spectrum range of 108 MHz - 1002 MHz. This use case was an effort to show a path of finding room for the 48 six MHz channels of DOCSIS capacity needed for the service.

The PON use cases assume that the HFC network remains in service even though the fiber is built to within 150 feet of every home. We assume the HFC network remains in service as part of a transition strategy. The system wide build out approach for PON illustrates the enablement capital to offer 5 Gbps or higher symmetrical services across the entire market. The OLTs and PON extenders were sized for 50% penetration as part of the enablement capital. The success capital includes the 10G EPON symmetrical

equipment, drop builds, and labor. The existing HFC network and set-top boxes provide the video services for the 10G EPON customers as well. None of the models included the capital for the transition to IPTV.

Use Case 1: HFC 2G DOCSIS

Summary of Use Case 1:

- Existing HFC and Spectrum Split
- Total serving Area: 960 HHP
- HFC optical serving area: 480 HHP
- DOCSIS Service Group: 480 HHP
- DOCSIS Downstream: 32 D3.0 + 96 MHz OFDM
- DOCSIS Upstream: 4 ATDMA
- DOCSIS Cost:
  - Assume 48 Downstream DOCSIS channels equivalents and 5-80 MHz of SC-QAM and OFDMA upstream are assumed using the Infonetics report from Q4/2014 price per channel rate of $399.00
- Service Tier Support
  - 1 Gbps Downstream Service Tier Supported
  - 1 Gbps Upstream Service Tier (not possible for this use case)
  - 40 Mbps Upstream Service Tier Supported

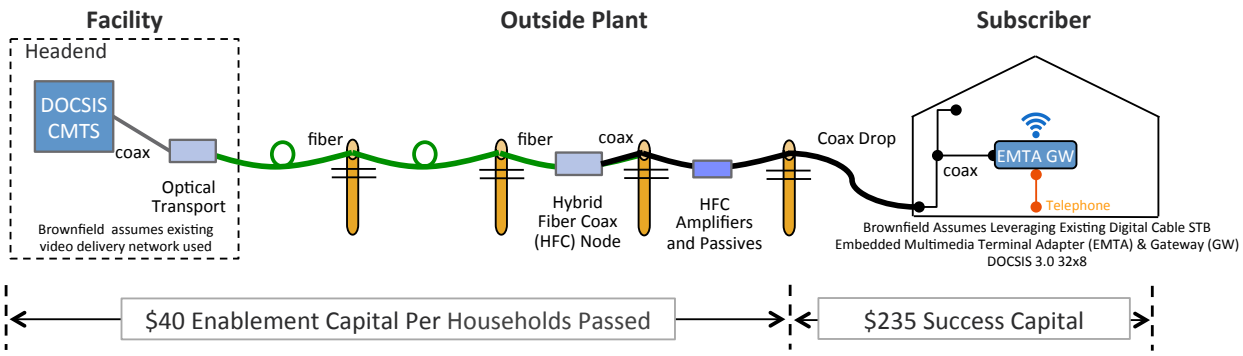


Figure 4 - Use Case 1: HFC 2G DOCSIS

Enablement Capital (Cost to Pass a Home)

- Capital Components
  - Assumes existing HFC network, spectrum split, and 480 HHP node
  - Assumed preexisting: fiber optical cable and coaxial cable costs, HFC nodes, coax amplifiers, coax passives, connectors
  - Assumes no existing CCAP chassis
  - The price per channel rate of $399.00, which is based on the Infonetics report from Q4/2014.
  - The actual price per channel an MSO may pay could be less when assuming all of the licensing channel capacity is activated.
  - Fully Load Chassis is added configured with D3.0/3.1 CCAP with 32 SC QAM + 96 MHz D3.1 (equivalent to 48 Channels)
  - It is assumed that the $399.00 Infonetics price assumed a fully loaded chassis and the channel pricing included a mid-split capable upstream configurations
  - Assumes 5-42 spectrum
  - Use of Four DOCSIS 3.0 upstream
  - DOCSIS 3.1 OFDMA upstream is not assumed
- Variables Impacting Capital
  - Enabling all channels per port and move to D3.1 will likely see lower price per channel & megabit
  - Excellent cost per sub

- Leveraging existing outside plant and spectrum thus there are no variables to consider on the OSP.

Success Capital (Cost to Connect a Home)

- Capital Components
  - $0 drop cable / labor (assumed existing)
  - $0 digital cable STB assumed 3 per home (assumed existing)
  - $0 cable modem for non 1 Gbps subs (assumed existing)
  - $235 Includes cost for 1 Gbps cable modem residential gateway and installation labor
- Variables Impacting Capital
  - CPE equipment type, selection & costs
  - Amplifier equipment needs
  - Variable installation time and labor rates

<u>Use Case 2: HFC 2G x 200M DOCSIS</u>

Summary of Use Case 2:

- Existing HFC and Spectrum Split
- Total serving Area: 960 HHP
- HFC optical serving area: 480 HHP
- DOCSIS Service Group: 480 HHP
- DOCSIS Downstream: 32 D3.0 + 96 MHz OFDM
- DOCSIS Cost:
  - Assume 48 Downstream DOCSIS channels equivalents and 5-80 MHz of SC-QAM and OFDMA upstream are assumed using the Infonetics report from Q4/2014 price per channel rate of $399.00
- Service Tier Support
  - 1 Gbps Downstream Service Tier Supported
  - 1 Gbps Upstream Service Tier (not possible for this use case)

Key Differences for Use Case 1:

- DOCSIS Upstream: 4 ATDMA and 5-42 MHz OFDMA
- Use of DOCSIS 3.0 and DOCSIS 3.1 upstream is assumed
- Frequency or Time and Frequency Multiplexing (TaFDM) is assumed
- Full 5-42 MHz upstream spectrum may be used
- 100 Mbps Upstream Service Tier Supported

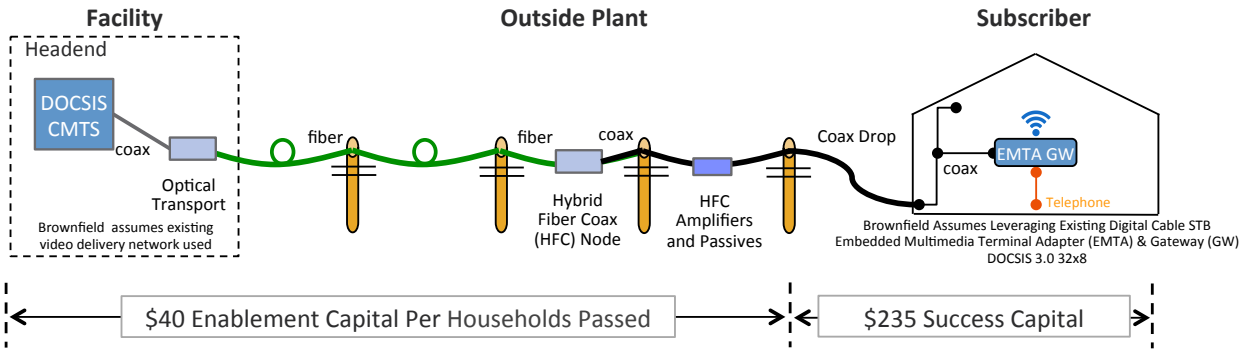


Figure 5 - Use Case 2: HFC 2G x 200M DOCSIS

Enablement Capital (Cost to Pass a Home)

- Capital Components
  - Same as use case 1 except
  - Use of Four DOCSIS 3.0 upstream "and" 5-42 DOCSIS 3.1 OFDMA upstream is supported
  - The enable capital is shown as the same as use case 1 and the cost per DOCSIS channel is the same for use case 3, this is due to the desire to use Infonetics channel price as stated earlier in the paper.
- Variables Impacting Capital
  - Same as use case 1

Success Capital (Cost to Connect a Home)

- Capital Components
- Same as use case 1
- Variables Impacting Capital
- Same as use case 1

Use Case 3: HFC Mid-split 2G x 500M DOCSIS

Summary of Use Case 3:

- Downstream Spectrum: Upgrade to 1 GHz
- Upstream Spectrum: Split Change to Mid-split 5-85 MHz
- Total serving Area: 960 HHP
- HFC optical serving area: Moves from 480 to 240 HHP
- DOCSIS Service Group: Moves from 480 to 240 HHP
- DOCSIS Downstream: 32 D3.0 + 96 MHz OFDM
- DOCSIS Upstream: 4 ATDMA and 5-80 MHz OFDMA
- DOCSIS Cost:
  - Assume 48 Downstream DOCSIS channels equivalents and 5-80 MHz of SC-QAM and OFDMA upstream are assumed using the Infonetics report from Q4/2014 price per channel rate of $399.00
- Service Tier Support
  - 1 Gbps Downstream Service Tier Supported
  - 1 Gbps Upstream Service Tier (not possible for this use case)
  - 400 Mbps Upstream Service Tier Supported

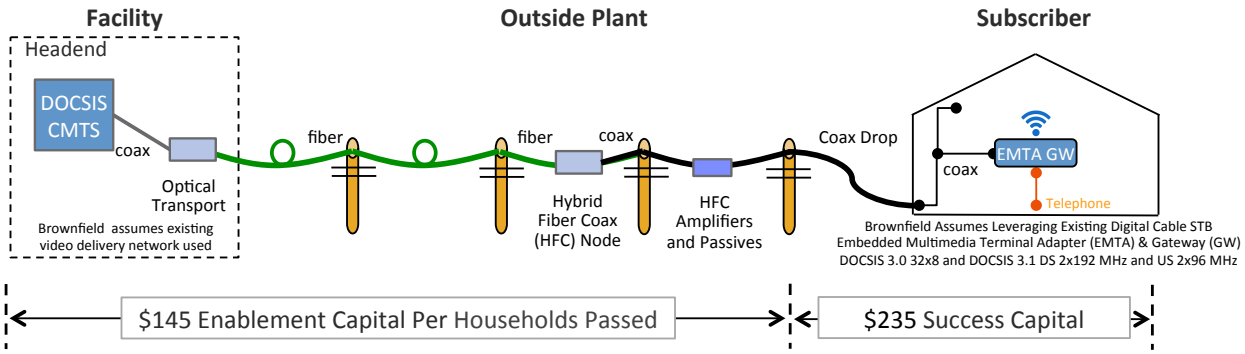


Figure 6 - Use Case 3: HFC Mid-split 2G x 500M DOCSIS

Enablement Capital (Cost to Pass a Home)

- Capital Components
  - Same as use case 1 except
  - Assumes spectrum expansion to 1 GHz and upstream Mid-split change
  - Assumed Node splits and fiber build to new node serving area
  - figurations
  - Assumes 5-85 spectrum
  - Use of Four DOCSIS 3.0 upstream "and" 5-85 DOCSIS 3.1 OFDMA upstream is supported

- Variables Impacting Capital
  - Same as use case 1

Success Capital (Cost to Connect a Home)

- Capital Components
  - Same as use case 1
- Variables Impacting Capital
  - Same as use case 1

<u>Use Case 4: FTTH 10G EPON</u>

Summary of Use Case 4:

- Assumes existing HFC remains in service to support video
- Overbuild FTTH using 10G EPON
- All passive ODN
- Serving Area 960 HHP (Tied to Enablement Capital)
- 8 PON Ports with 64 Subs
- 10 km is assumed to support 64 Subs per Port
- 8 fibers to the serving area exist with PON wavelengths available
- 8 Gbps Symmetrical Capacity
- Leverage Existing HFC for Video Service
- Service Tier Support
  - 5+ Gbps Downstream Service Tier Supported
  - 5+ Gbps Upstream Service Tier (not possible for this use case)

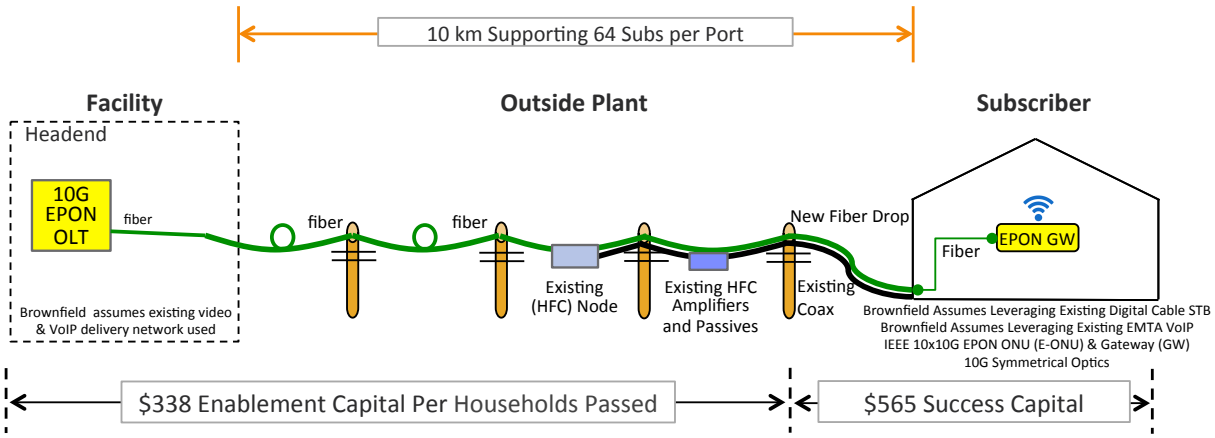


Figure 7 - Use Case 4: FTTH 10G EPON

Enablement Capital (Cost to Pass a Home)

- Capital Components
  - Includes 10G EPON OLT
  - Assumes 8 fibers per service group (if enough fibers do not exist at the service group the MSO should consider Use Case 5
  - Leverage existing DOCSIS back office
  - EPON is sized for 50% take rate
  - Assumed a fiber will be pre run to all existing tap location as part of enablement capital, this could be termed as fiber-to-the-curb
  - Running fiber from the tap location to the home is success capital

- Variables Impacting Capital
  - Migration to all IP video over 10G EPON
  - Some video equipment may be repurposed like bulk encryption
  - Aerial / Underground

Success Capital (Cost to Connect a Home)

- Capital Components
  - Existing digital cable set-top box used over HFC
  - Fiber drop cable
  - IEEE based 10 Gbps x 10G ONU
  - Installation for all components

- Variables Impacting Capital
  - Migration to IP set-tops will increase connect costs but reduces if IP-STB are in place
  - Material and equipment costs vary
  - Variable drop length (Assumed 150 feet)
  - Aerial / Underground
  - Variable installation time and labor rates

Use Case 5: FTTH 10G EPON + PON Extender

Summary of Use Case 5:

- Assumes existing HFC remains in service to support video
- Overbuild FTTH using 10G EPON
- All passive ODN
- Serving Area 960 HHP (Tied to Enablement Capital)
- 4 PON Ports with 128 Subs
- Up to 60 km is assumed to support 128 Subs per Port
- 1 fiber to the serving area exist with PON wavelengths available
- 8 Gbps Symmetrical Internet Services
- Leverage Existing HFC for Video Service
- Service Tier Support
  - 5+ Gbps Downstream Service Tier Supported
  - 5+ Gbps Upstream Service Tier

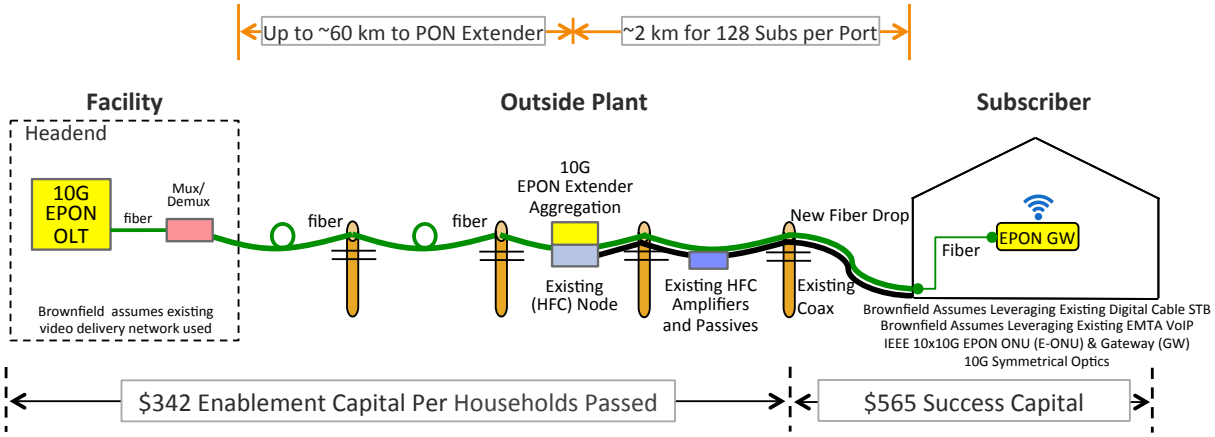


Figure 8 - Use Case 5: FTTH 10G EPON + PON Extender

Enablement Capital (Cost to Pass a Home)

- Capital Components
  - Same as use case 4 except
  - No requirement for 8 fibers per service group
  - 1 Fiber is required for the trunk between the facility and the headend
  - Four separate 10G PONs are required
  - 2 km distance between the four PON extender fiber nodes

- Variables Impacting Capital
  - Migration to all IP video over 10G EPON
  - Some video equipment may be repurposed like bulk encryption
  - Aerial / Underground

Success Capital (Cost to Connect a Home)

- Capital Components
  - Existing digital cable set-top box used over HFC
  - Fiber drop cable
  - IEEE based 10 Gbps x 10G ONU
  - Installation for all components

- Variables Impacting Capital
  - Migration to IP set-tops will increase connect costs but reduces if IP-STB are in place
  - Material and equipment costs vary
  - Variable drop length (Assumed 150 feet)
  - Aerial / Underground
  - Variable installation time and labor rates

SUCCESS BASED USE CASES FOR BROWNFIELD MIGRATION

The residential success based approach addresses the competitive challenges of gigabit or multi-gigabit symmetrical services differently than the system wide based approached shown in the section. The system wide approach built out the network across the entire serving area. The system wide FTTH, use cases 5 and 6, illustrate the high upfront enablement capital per HHP. The total enablement build out costs is the number of homes in the build out area times the dollars per HHP. If subscriber take rates are low then the actual cost per customer are high, so when there is high enablement capital, this should be coupled with high service penetration rates to justify the investment. The customer take rates for gigabit or multi-gigabit symmetrical services are expected to be very low, perhaps single digit percent.

Therefore, if the driver for FTTH is gigabit or multi-gigabit symmetrical services then high enablement capital build outs may be hard to justify.

The success based approach has been used for nearly two decades to overlay an existing HFC network with a fiber to the premise (FTTP) network in the support commercial services. This was used when the service speed tiers may have exceeded the capabilities of HFC and DOCSIS and there was also the likelihood that service take rates were low and therefore not economically justifiable to allocate HFC resources for just a few customers.

An advantage in the system wide approach enables the provider to market the service across 100% of the serving area, but this comes with high upfront cost, before a single customer signs up for the service. The success based approached may reduce the serving area to an acceptable range based on the economics and service turn up times.

Success based approaches extend the fiber for a single customers and this will also increase the service area coverage, because fiber was extended from the original location. Some cable operators deployed the HFC fiber deep architecture, which places a node to serve a smaller service group, this means that a larger percentage of homes passed are close to a fiber location. If we consider a serviceable range of 1/3 of a mile from an existing fiber location, likely a node.

If we consider 114 homes per mile used in this paper and a 1/3 of a mile build out range for the success capital approach the home passed serviceable is about 38 (114 * .333 = 37.9). However, if we assume that it may be possible that a north, south, east and west fiber builds could exist, serviceability going in four directions, this is 152 HHP (38 * 4 = 152) from a fiber location, assuming 114 homes per mile. This may be an aggressive assumption, because not all fiber locations would have an option to build fiber in all four directions and at the average homes per mile. If we assume that fiber is located at a node location serving 400 HHP and network builds are possible in all four directions, while also assuming 114 HHP this could be 38% of the HHP are serviceable (152/400 = .38).

If we use the 2015 SCTE paper values of 75 homes per mile and a 1/3 of a mile build out range for the success capital approach the home passed serviceable is about 25 (75 * .333 = 24.9). Assuming a north, south, east and west fiber builds could exist, serviceability going in four directions, this is 100 HHP (25 * 4 = 100) from a fiber location, assuming 75 homes per mile. If we assume that fiber is located at a node location serving 400 HHP and network builds are possible in all four directions, while also assuming 75 HHP this could be 25% of the HHP are serviceable (152/400 = .38).

The smaller the homes passed per node the greater the percentage of customers that are close to fiber, assuming similar HHP per mile. Many cable operators were at 500 HHP in the early 2000s but in past decade or more, through node splits and new build practices that HHP per node number has dropped.

The success based approach has to examine the costs at a larger scale to provide better visibility of the costs because these use case assume extremely low takes rates. The success based approach must also consider a time horizon they will eventually reach the 1% and 5% take rates in our models. So, we scaled the assumptions up to a headend serving area and then sized the network for the eventually reach the estimated take rates. We assumed we could reach the customers within 1/3 of a mile and that this represented the serving area. This assumes existing network power and not additional battery suppliers includes. Please refer to the disclaimer for additional considerations.

Success Based Use Cases Assumptions:

- Average Headend Serving Area
  80,352
- Average Number of Node per Facility
  279
- Average Serving Area:
  288 HHP
- Homes per mile:
  114
- Aerial / Underground:
  73% / 27%
- Fiber Construction Cost Per Mile
  $15,000
- 1 Gbps Symmetrical Take Rate
  5% or 1%

Use cases 6 through 8 illustrate the estimated costs a success based directed customer fiber build and equipment using 10G EPON, 10G EPON with PON Extender, and 1 Gbps point-to-point Ethernet.

Use Case 6: Success Based FTTH 10G EPON

- Assumes existing HFC remains in service to support video
- Overbuild FTTH using 10G EPON
- All passive ODN
- 10 km
- 8 fibers to the serving area exist with PON wavelengths available
- 8 Gbps Symmetrical Capacity
- Leverage Existing HFC for Video Service
- Service Tier Support
    - 5+ Gbps Downstream Service Tier Supported
    - 5+ Gbps Upstream Service Tier (not possible for this use case)

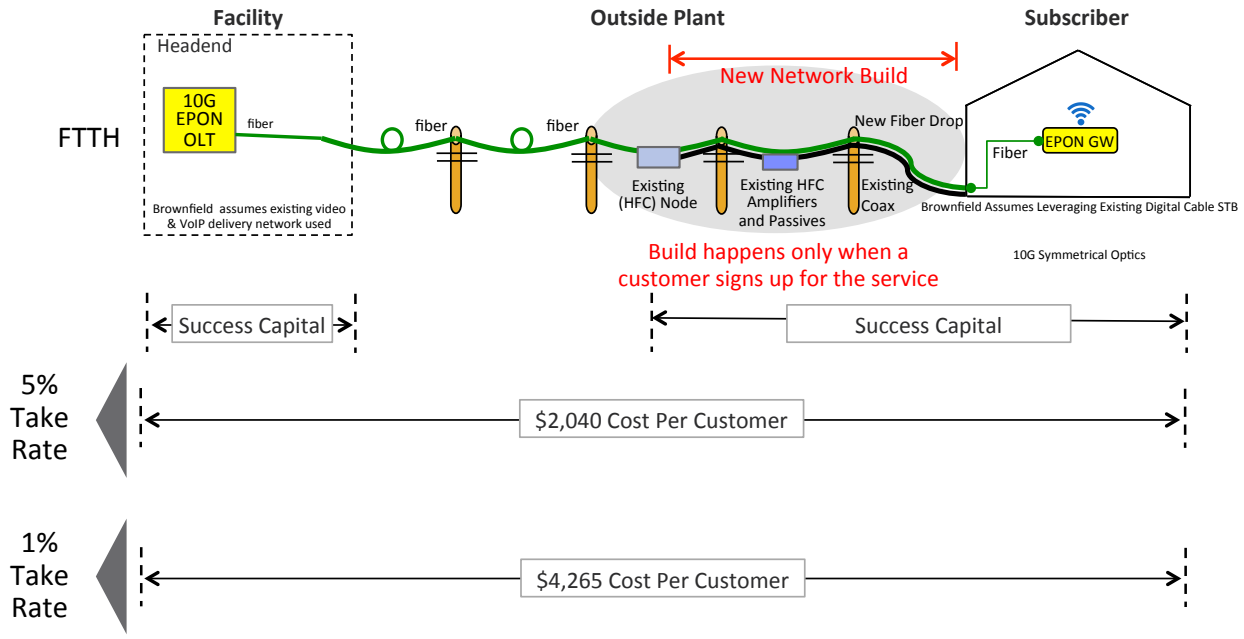


Figure 9 - Use Case 6: Success Based FTTH 10G EPON

Use Case 7: Success Based FTTH 10G EPON + PON Extender

- Assumes existing HFC remains in service to support video
- Overbuild FTTH using 10G EPON
- Up to 60 km is assumed
- 1 fiber to the serving area exist with PON wavelengths available
- 8 Gbps Symmetrical Internet Services
- Leverage Existing HFC for Video Service
- Service Tier Support
    - 5+ Gbps Downstream Service Tier Supported
    - 5+ Gbps Upstream Service Tier

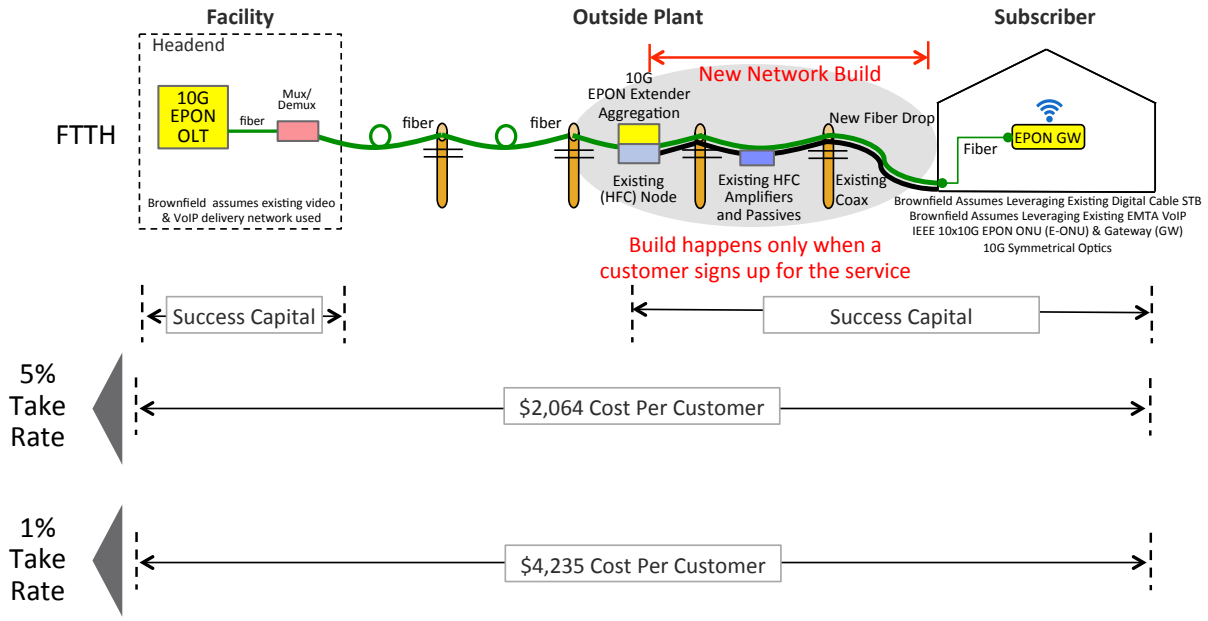


Figure 10 - Use Case 7: Success Based FTTH 10G EPON + PON Extender

Use Case 8: Success Based FTTH 1G P2P Ethernet

- Assumes existing HFC remains in service to support video
- Overbuild FTTH using Point-to-Point 1 Gbps Ethernet
- Up to 60 km is assumed
- 1 fiber to the serving area exist with wavelengths available
- Sub 1 Gbps Symmetrical Internet Services
- Leverage Existing HFC for Video Service
- Service Tier Support
    - Sub 1 Gbps Downstream Service Tier Supported
    - Sub 1 Gbps Upstream Service Tier

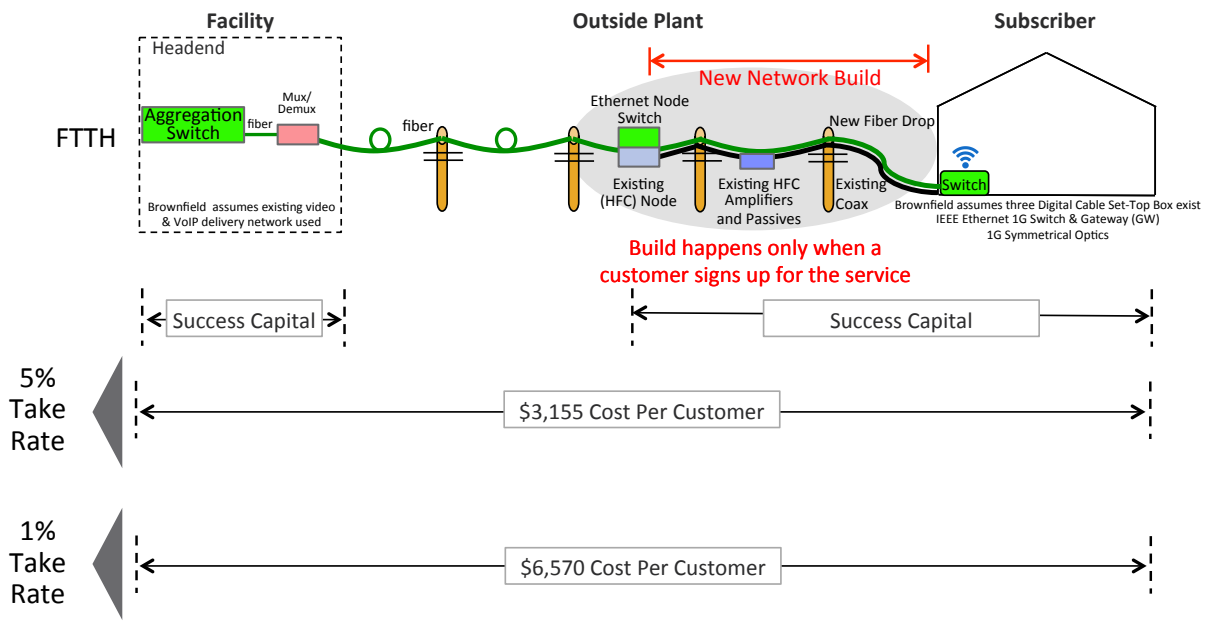


Figure 11 - Use Case 8: Success Based FTTH 1G P2P Ethernet

BROWNFIELD USE CASES SUMMARIES

The following figures illustrate cost breakouts for the enablement and success capital. Use cases 1 through 5 illustrate the estimated costs a success based directed customer fiber build and equipment using 10G EPON, 10G EPON with PON Extender, and 1 Gbps point-to-point Ethernet. Figures 12 and 13 show a breakout by the major categories of labor, materials, and equipment.

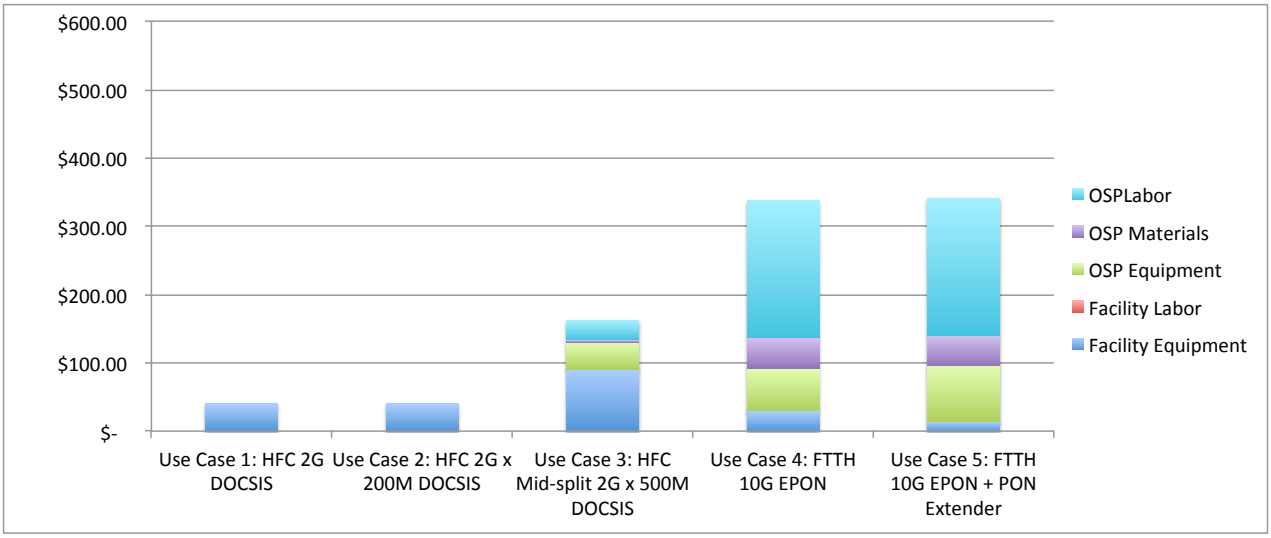


Figure 12 - Enablement Capital Composition (Per HHP)

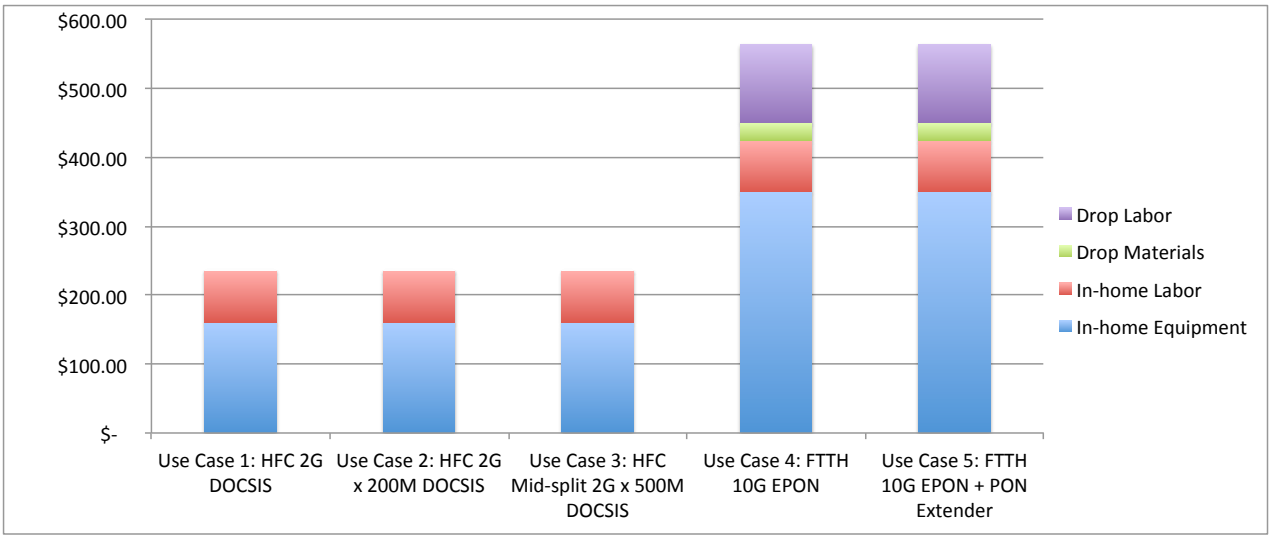


Figure 13 - Success Capital Composition (Per Customer)

Figures 13 and 14 examine the system wide network builds and separate the enablement and success cost among each use case. Figure 14, illustrates the impact of take rate, assumed at 50% and the real cost per customer served. Providers and/or services with low takes rates will have high cost per subscriber costs, especially when enablement costs are high. The key point is if a massive investment in enablement capital is required and high take rates are desirable to achieve acceptable return on investment.

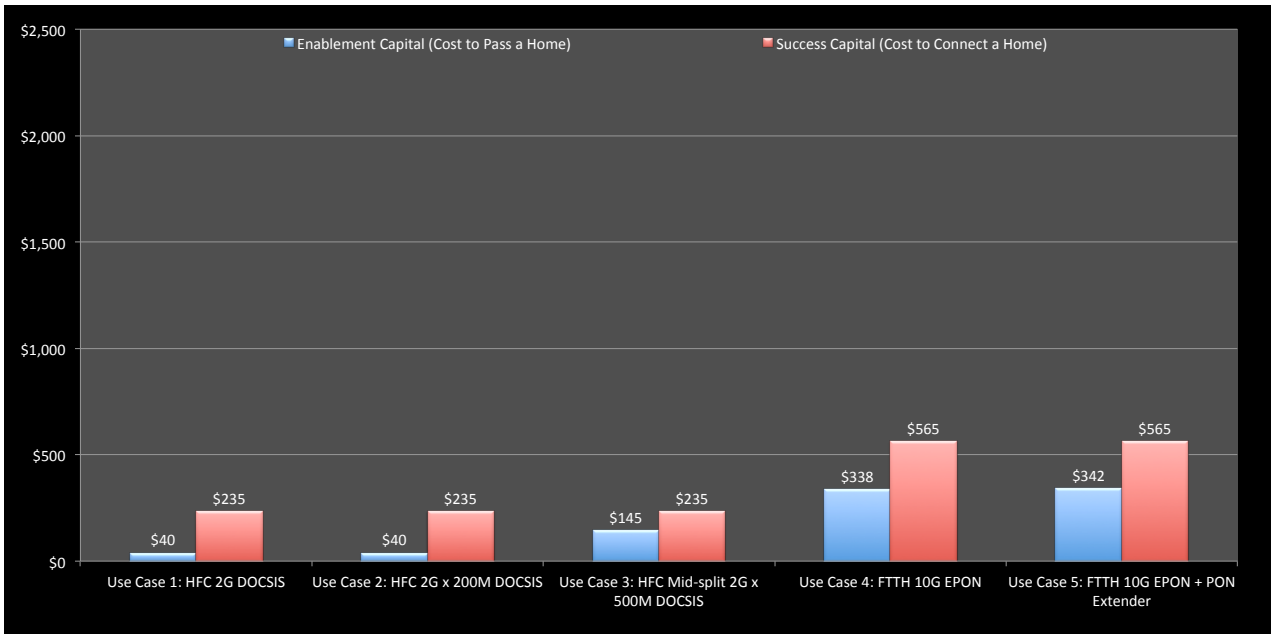


Figure 14 - Use Case Enablement & Success Comparison

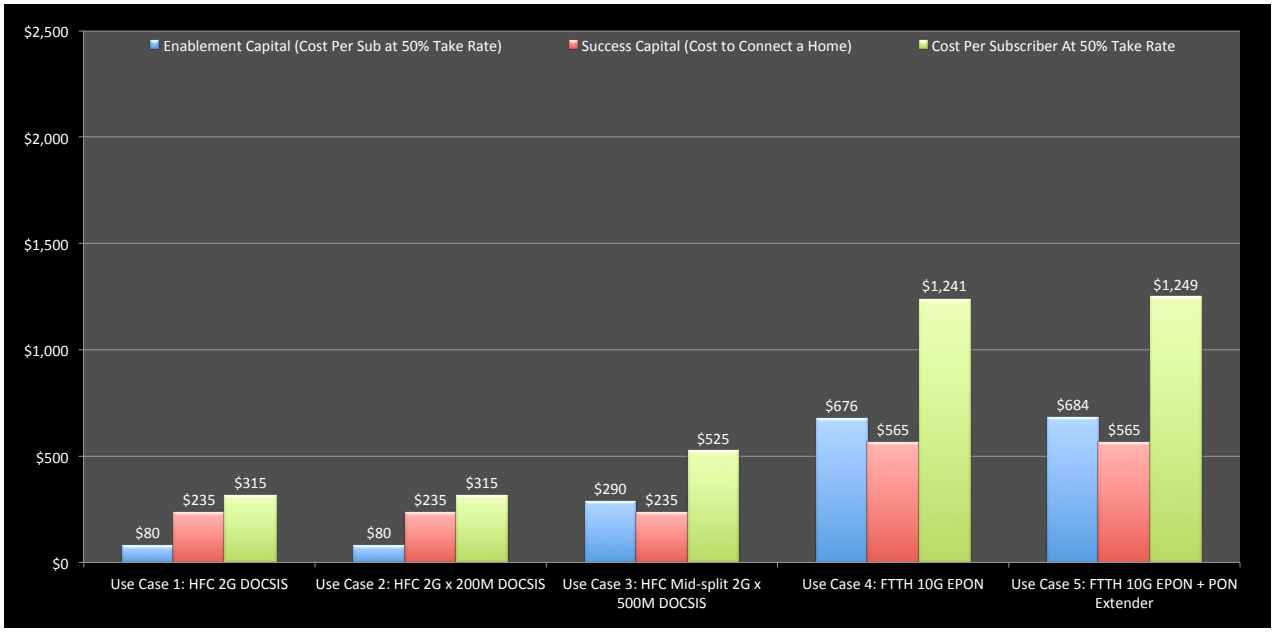


Figure 15 - Use Case Cost Per Sub at 50% Take Rate of HHP

The following figures illustrate cost breakouts for the enablement and success capital. Use cases 1 through 8 illustrate the estimated costs a system wide versus success based. Figures 16 and 17 show the impact of building FTTH to within 150 feet of every home as seen in use cases 4 and 5 and low penetration of the service. The HFC and DOCSIS use cases as shown in use case 1 through 3 show low cost for low take rates, this proves to be a good solutions for high bit rate downstream services and even with reasonable upstream speeds. The 1 Gbps symmetrical services are higher are best served with success capital builds found on use cases 6, 7, and 8.

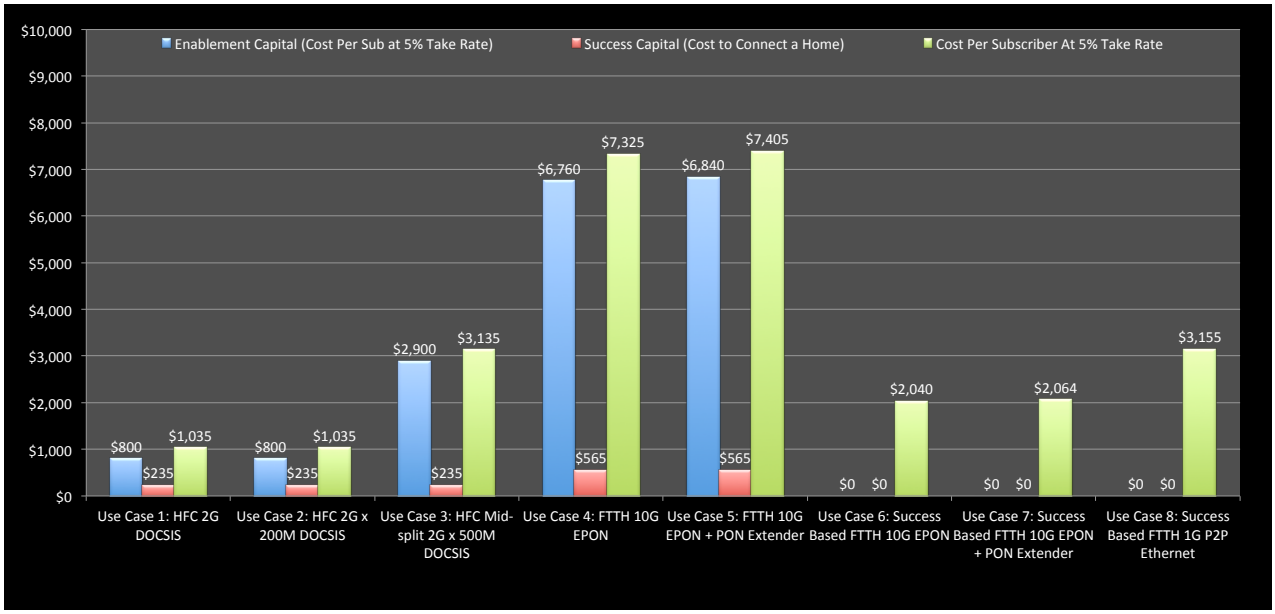


Figure 16 - Use Case Cost Per Sub at 5% Take Rate of HHP

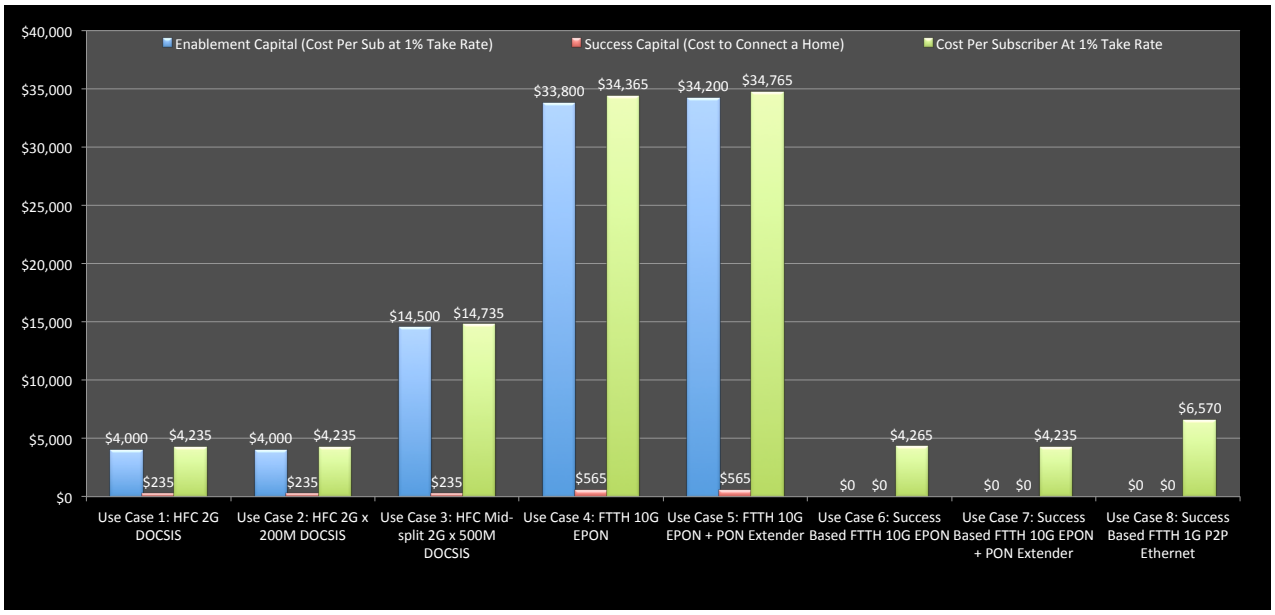


Figure 17 - Use Case Cost Per Sub at 1% Take Rate of HHP

Considering just 1% take rates for the service the HFC and DOCSIS use cases as shown in use case 1 through 3 still show relatively low costs and again proving to be a good solutions for high bit rate downstream services, even with reasonable upstream speeds. The 1 Gbps symmetrical services are higher are best served with success capital builds found on use cases 6, 7, and 8.

CONCLUSIONS

System-wide based HFC and DOCSIS are the most cost effective solutions to support very high bit rate downstream and sufficient upstream capacity to support Internet and IPTV services.

Success Based FTTH will have lower enablement capital than System Wide Rebuild FTTH Those uses cases drove fiber to within 150 feet of "all" homes increasing enablement capital Benefit is reduced service delivery times and smaller success capital

Lower enablement capital impacts A majority of the capital is success based (typically desirable approach for low take rate services) Often serviceability is not the entire footprint (50% or less is typical) Service delivery times are often weeks/months from the time of order (will vary between MSOs) Service delivery times can be 6-8 weeks (instead of days for the higher enablement capital approaches)

A major benefit addressing the billboard speed was is that since most MSOs are targeting 1 Gbps symmetrical and higher speeds for FTTH Selective Service Tiers, this saves HFC for Video Service and DOCSIS for the majority of Internet Users.

Key Conclusions:

1. HFC data capacity remains competitive against modern xPON (upstream remains a challenge)
2. HFC serves more homes and customers with far less capital
3. HFC is extremely cost effective when compared to any access layer technology
4. Upgrade 1 GHz+ / 85 MHz with 2.5G DOCSIS is extremely cost effective
5. Of the FTTH paths 10G EPON has lowest enablement capital
6. Enablement and success cost per bit favors FTTH 10G EPON solutions
7. Cable could push fiber deeper and even to the to the last active (amplifier) to position fiber closer to customers enabling faster service activation for 10G EPON for selective customers
8. Important factor like operational costs were not considered

RECOMMENDATIONS

As stated in the beginning of this paper the purpose was to examine these key questions below to determine the best path. The recommendations are based on the assessment of the capital expense of the alternative network build methods, like system wide and success based. This is also based capital assessment of the technology and architectures assess within each build option. There are other technology and architecture options that exist today, but these are likely the leading candidates. There will be new technologies and architectures that could improve HFC and DOCSIS as well as 10G EPON, there may also be other technologies that will emerge as well. This is a snapshot in time based on a set of network assumptions and networking approaches, as seen in the SCTE 2015 paper with different assumptions and network approaches the recommendation could change. Another key point, this examines the capital cost but not the total cost of ownership costs. Additionally, the ongoing operational costs such as utility costs, equipment and network maintenance costs, labor rate differences between each technologies and architecture, and many other drivers for OPEX. The OPEX costs will vary widely between operators and suppliers based on many factors.

Question:

Should MSOs continue with HFC/DOCSIS investment?

Recommendation:

Yes, based on use cases 1 through 3 these provide a cost effective approach to offer high bit rate data and eventually IPTV services at a

much lower cost point that other solutions. The capacity shown is 2 Gbps, which is very high capacity, but even much higher is possible over HFC and DOCSIS.

Question:

Should MSOs overlay the existing HFC network with FTTH and PON?

Recommendation:

No, in terms of performing a system wide investment because the existing HFC and DOCSIS network will meet the service needs for a majority of the customers. Considering spectrum and node split investment as shown in use case 3, additional capacity is achieved at a low cost and with node splits this can be a surgical investment made where and when needed.

Yes, in terms of performing a success based investment because MSOs can offer high symmetrical data rates, such as 2 Gbps x 2Gbps that will exceed the capacity of their competitors today and provide a solutions to meet the competitive needs in the future. Since take rates are extremely low for the highest Internet speed tiers success based approaches offer the most economical path.

Question:

What are the economics both from a financial and capacity perspective?

Recommendation:

The paper illustrates the capital comparison and HFC and DOCSIS has the lowest cost per customer however the lowest cost per megabit belongs to the 10G EPON solutions.

<u>Questions: What are the Economics of CTTH & FTTH brownfield migration options to support:</u>

Question:

1 Gbps downstream service to be available to 100% of the HHP in brownfield networks?

Recommendation:

The economic comparison to offer 1 Gbps downstream is shown in use cases 1 through 5; with HFC/DOCSIS (use cases 1-3) as the lowest cost solution.

Question:

1 Gbps symmetrical service to be available to 100% of the HHP in brownfield networks?

Recommendation:

The economic comparison to offer 1 Gbps symmetrical is shown in only use cases 4 through 5; the EPON uses cases. Currently there is not a DOCSIS based network solution to model 1 Gbps symmetrical.

Question:

1 Gbps symmetrical service to be available to 30-50% of the HHP in brownfield networks?

Recommendation:

The use of success based network builds to offer 1 Gbps symmetrical services is the most cost effective path for low take rate offering. The use of 10G EPON and 1 Gbps Ethernet are the solutions available today and are the most cost effective.

<u>DISCLAIMER</u>

All economic estimates may not represent actual build and equipment costs. The cable HFC and FTTH economic estimates contain inputs from a third party and ARRIS for construction, installation, labor, materials, current and future equipment estimates as well as our assumptions of network architecture and topology. These use cases were compiled independent from the cable operators. The use cases are based on an extremely small sample size and actual build costs may vary widely market by market. This material is for information and illustrative purposes only and should not be considered or derived as actual results. It is not, and should not be regarded as actual cost estimates. In addition, all of these estimates assume to varying degree different starting points or rather previous investments, which may not be represented in these costs as these are sunk costs. The core purpose of this data

is to show relative difference for a particular set of use case migration options and should be used for discussion purposes only. The estimates expressed in this material are subject to change without notice. Reasonable people may disagree about the opinions expressed in our analysis and results. There are many other factors that were not considered that may impact the estimates and illustrate a significantly different result. It is strongly recommended that individuals and companies perform their own analysis based on their environment and costs.

ACKNOWLEDGEMENTS

The author would like to thank the following individuals and test teams: Fred Slowik, Jon Bordner, Stuart Eastman, Steve Hersey, Charles Cheevers, Venk Mutalik, Zoran Maricevic, John Ulm and the technical publishing team.

REFERENCES

[EMM1] Emmendorfer, Michael J, "Comparing the DOCSIS® 3.1 and HFC Evolution to the FTTH Revolution", 2015 SCTE Cable-Tec Expo., October 13-16, 2015

Applications of Machine Learning in Cable Access Networks

Karthik Sundaresan, Nicolas Metts, Greg White; Albert Cabellos-Aparicio
CableLabs; UPC BarcelonaTech.

Abstract

Recent advances in Machine Learning algorithms have resulted in an explosion of real-world applications, including self-driving cars, face and emotion recognition, automatic image captioning, real-time speech translation and drug discovery.

Potential applications for Machine Learning abound in the areas of network technology, and there are various problem areas in the cable access network which can benefit from Machine Learning techniques. This paper provides an overview of Machine Learning algorithms, and how they could be applied to the applications of interest to cable operators and equipment suppliers.

INTRODUCTION

Recent advances in Machine Learning, and in particular, deep learning, have resulted in an explosion of real-world applications, including self-driving cars, face and emotion recognition, automatic image captioning, real-time speech translation and drug discovery. Powerful new learning algorithms, harnessed via open-source toolkits and running on general-purpose GPUs (even cloud instances) are the key enablers for this explosion.

Machine Learning technologies can learn from historical data, and make predictions or decisions, rather than following strictly static program instructions. They can dynamically adapt to a changing situation and enhance their own intelligence by learning from new data. Statistical pattern recognition, data mining, clustering and deep learning have proven to be powerful components in the Machine Learning space.

Potential applications for Machine Learning abound in the areas of network technology and applications. It can be used to intelligently learn the various environments of networks and react to dynamic situations better than a fixed algorithm. When it becomes mature, it would greatly accelerate the development of autonomic networking.

There are various problem areas in the cable access network which can benefit from Machine Learning techniques. For example, the Proactive Network Maintenance data which represents network conditions obtained from CMTS and CM can be processed through Machine Learning algorithms to enable automatic recognition of issues in the cable plant and initiating the needed corrective actions. Patterns in network traffic, equipment failures, device reboots or link failures can be used to identify the root cause of various network problems. Video channel usage data could be used to optimize the multicast channel lineups in a network, and be used to predict which programs to record automatically in a DVR.

This paper provides an overview of Machine Learning algorithms with focus on the applications of interest to cable operators and equipment suppliers. The paper discusses several learning toolkits and computational platforms, and hypothesizes some of the most compelling near-term uses. It will explore how to apply learning algorithms to various challenges faced by operators in managing network devices and services. In addition, the paper covers best practices for achieving optimal results with Machine Learning.

OVERVIEW OF MACHINE LEARNING

This section provides a very brief introduction to Machine Learning. The reader familiar with such techniques may safely skip this section. Machine Learning (ML) is a field of study of Artificial Intelligence that provides computers with learning techniques without being explicitly programmed. Traditional programming requires manually coding an algorithm and combining it with input data to produce the resulting output data. This model is fundamentally different in ML: a learning algorithm (e.g., a Neural Network) is trained with the input data and the corresponding output data, which is known as training set, to automatically produce the algorithm, which will be used to predict future results. But how do computers actually learn such an algorithm?

Systems can be always described by a mathematical function and in many cases such a function is multi-dimensional (i.e., has many input and output variables, maybe hundreds) and is non-linear. If we have such a function, we have a complete model of the system. For instance, the function that describes a network may have the form: function (topology, load, configuration) = performance. That is, the function links the topology of the network, the traffic (load) and configuration (e.g., routing) with the resulting performance (e.g., delay). Analytical modeling provides tools to obtain such a function (e.g., Queueing theory in networking). However, such techniques often use simplifications and cannot be applied when the network is too complex.

Machine Learning provides a different way to obtain such a function. As shown in Figure 1 below, a learning algorithm fed with a representative training set will fit the function, even if such function has hundreds of variables or it is complex and non-linear. The algorithm produced by a Machine Learning approach is the function itself and can also produce accurate answers for input points for which it has not been trained explicitly, provided that the initial training set was representative enough. This is accomplished by interpolation, that is estimating an unseen prediction point between two already seen test points and extrapolation, estimation of the point beyond the observed training set.

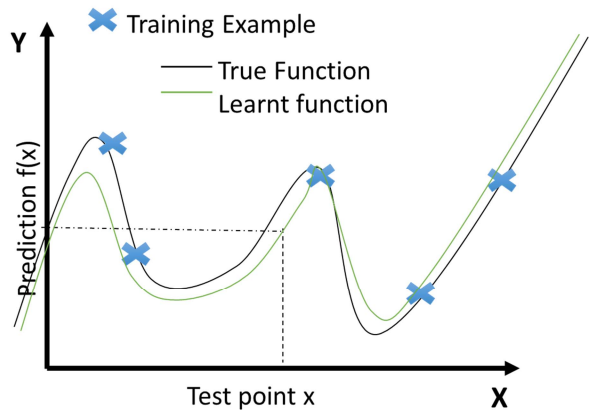


Figure 1. An overview of Machine learning

In what follows, we provide a summary of the two most relevant techniques in Machine Learning: supervised and unsupervised learning.

SUPERVISED LEARNING

In supervised learning, there is a given data set and the knowledge on what the correct output looks like. The idea is that there is a relationship between the input and the output.

The input data set is called training data and has a known label or result, such as cat/not-cat (for identifying cat photos) or a stock price at a time (for predicting stock prices). Network examples could include classifying traffic patterns for intrusion detection as harmful-traffic vs. normal-traffic, or predicting the usage of a certain network link based on various parameters. A model is prepared through a training process where it is required to make predictions and is corrected when those predictions are wrong. The

training process continues until the model achieves a desired level of accuracy on the training data.

Supervised learning problems are categorized into "regression" and "classification" problems. In a regression problem, the goal is to predict results within a continuous output, essentially trying to map the input variables to some continuous function. In a classification problem, the goal is to predict results in a discrete output, trying to map the input variables into discrete output categories.

Example classification algorithms include Logistic Regression and the Back Propagation Neural Network, SVMs, kNN, etc.

Support Vector Machine (SVM)

SVM is type of supervised Machine Learning algorithm. Given a set of training examples, each marked as belonging to one of two classes, an SVM training algorithm builds a model that assigns new examples into one class or the other. A hyperplane splits the input variable space. In SVM, a hyperplane is selected to best separate the points in the input variable space by the class to which they belong. The SVM learning algorithm finds the coefficients that results in the best separation of the classes by the hyperplane. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

The distance between the hyperplane and the closest data points is referred to as the margin. The best or optimal hyperplane that can separate the two classes is the line that has the largest margin. Only these points are relevant in defining the hyperplane and in the construction of the classifier. These points are called the support vectors as they support or define the hyperplane.

K-Nearest Neighbors Algorithm (kNN)

kNN makes predictions for a new data point by assigning it the value/label of its K nearest neighbors. For regression this might be the mean output variable, in classification this might be the mode (or most common) class value.

In kNN classification, an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. In kNN regression, the output value is the average of the values of its k nearest neighbors.

The simplest technique is to use the Euclidean distance between neighbors. kNN performs a calculation (or learns) when a prediction is needed, just in time. The training instances can be updated or curated over time to keep predictions accurate.

UNSUPERVISED LEARNING

Unsupervised learning attacks problems with little or no idea what the results should look like. The idea is to derive structure from data where there is no apriori knowledge of the effect of the variables. This structure is derived by clustering the data based on relationships among the variables in the data. With unsupervised learning there is no feedback based on the prediction results. The input data set is not labelled and does not have a known result.

A model is prepared by deducing structures present in the input data. This may be to extract general rules. It may be through a mathematical process to systematically reduce redundancy, or it may be to organize data by similarity.

Typical unsupervised algorithms are Clustering (K-Means, Fuzzy and Hierarchical

clustering, etc.), Principal Component Analysis and the Apriori Algorithm.

Clustering Algorithms (K-means, Fuzzy and Hierarchical Clustering)

Give a set of data points, clustering is the problem of grouping the points in such a way that points in the same cluster are more similar to each other than to those in other clusters. Usually, the points are in a high-dimensional space, and similarity is defined using a distance measure. As an example, let's assume that we want to understand content consumption among users, movies for instance. If we have access to a large dataset of movie consumption per user, K-Means will cluster and group users with similar movie preferences. With this we can analyze the profiles of such users and be able to predict which types of movies they will consume.

The K-means algorithm takes a set of points and attempts to identify the set of K clusters for which the average distance of each point to the centroid (mean) of its cluster is minimized. In an algorithm like K-means, data is divided into distinct clusters, where each data element belongs to exactly one cluster. In Fuzzy clustering, data elements can belong to more than one cluster, and each element belongs to each cluster to a certain degree (e.g., a likelihood of belonging to the cluster). Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters.

Principal Component Analysis (PCA)

Sometimes data is collected on a large number of variables from a single population. With a large number of variables, the data will be too large to study and interpret properly. There would be too many pairwise correlations between the variables to consider. To interpret the data in a more meaningful form, it is necessary to reduce the number of variables to a few, interpretable linear combinations of the data. Each linear combination will correspond to a principal component. Principal components are the underlying structure in the data. They are the directions where there is the most variance, the directions where the data is most spread out. PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. With PCA once these patterns in the data are found, one can compress the data, i.e., by reducing the number of dimensions, without much loss of information.

Apriori Algorithm

Mining for associations among items in a large database of transactions is an important database mining function. Apriori algorithm is a technique to find the sets of frequent items and learning the association rules between items in a dataset. The algorithm works by identifying frequent individual items in a dataset and extending them to larger item sets as long as those sets appear sufficiently often in the dataset. These frequent item sets can be used to determine association rules which highlight general trends in that data set.

The interested reader can find further information on Machine Learning in [4], [6].

APPLYING MACHINE LEARNING TO NETWORKING PROBLEMS

With the exponential growth in network traffic (over-the-top video, social networking, cloud services, etc.), there is just too much data to manually learn and understand the many patterns of traffic flow. Network equipment is getting more complex, e.g., bringing up a router or a CMTS can take thousands of lines of configuration. Along with this comes the increasing complexity of the network itself, e.g., various overlays, underlays and connection of services across disparate networks. Traffic modeling and analysis are becoming increasingly

challenging. With the proliferation of applications, it is difficult to predict and to generalize application behavior and the impact it has on network traffic. Simply put, there are too many sources of information for manual processing by humans. ML techniques are ideal solutions for these kind of problems.

One example in the access network is how an operator can effectively and quickly process all the PNM data which comes up from the access network devices. As another example, the problem of dynamically fitting optimal DOCSIS 3.1 OFDM Profiles to hundreds of modems on an OFDM channel with thousands of subcarriers with changing conditions is a complex challenge.

Also, to understand and predict network behavior, some problems in networks are so complex that one cannot express them in a simple fashion, in those cases it is easier to represent the problem as a black box with a set of inputs and outputs. That can now help frame the problem as a Machine Learning problem and start making progress towards solving it in a timely fashion.

Applying Machine Learning to networking has been proposed several times in the past, a notable example is D. Clark and the Knowledge Plane [5]. However and for many use-cases it is very challenging to apply learning techniques to an inherently distributed system. The rise of the Software-Defined Networking (SDN) paradigm and Network Telemetry are enabling technologies for applying Machine Learning. This provides all the data needed from the network elements at a centralized server; ML algorithms can learn from this centralized repository of data. Also since the SDN controller can configure and manage the network elements, it provides an avenue for the ML algorithm to take corrective action on the network based on the patterns the algorithm sees.

Many of these network problems naturally lend itself to be solved by Machine Learning algorithms, which can obtain the best possible knowledge from the large sets of data available on hand.

LEARNING TOOLKITS AND COMPUTATIONAL PLATFORMS

This section gives a brief overview of the tools and software platforms available which a practitioner of Machine Learning can use to solve various problems.

Scikit-learn [21] is a Python library for Machine Learning. It is an open-source product (BSD License) and is based on NumPy and SciPy, two mathematical and data analysis libraries known for their speed and efficiency. The scikit-learn library is somewhat of a Swiss Army Knife amongst Machine Learning packages. The functionality implemented in scikit-learn includes classification, regression, clustering, dimensionality reduction, model selection, preprocessing, and many other features such as performance metrics.

The scikit-learn library includes almost every well-known classification and regression technique (Naive Bayes, logistic regression, Support Vector Machines (SVMs), lasso regression, etc.), along with some more cutting edge techniques such as neural networks. Users considering scikit-learn should be familiar with the Python programming language and would benefit from some Machine Learning background as well as familiarity with NumPy and SciPy.

Since scikit-learn has been around for nearly a decade, there is a mature user community along with excellent API documentation and many helpful tutorials. If you have an issue with scikit-learn, you can likely find an answer on StackOverflow.

Matlab incorporates a Neural Network Toolbox and a Machine Learning Toolbox[7] that provides the most common features for automatic learning: classification, regression, clustering, etc. The tool is capable of programming neural networks and large networks with deep learning autoencoders. The tool can be accelerated using GPUs and parallel environments and, more importantly, provides out-of-the-box neural networks that are auto-configured by the toolbox. This represents an important advantage for people that are starting to with Machine Learning. Matlab also provides implementations of various ML algorithms, so it is a great starting point to apply many different techniques to a give data set, to quickly analyze the feasibility of various algorithms.

Theano [8] is an open-source deep learning framework in the form of a Python library that lets you define, optimize and evaluate mathematical expressions. Theano has its roots at the MILA lab at University of Montreal and allows you to define mathematical expressions (typically multi-dimensional matrix calculations) that are then compiled to run efficiently on either CPU or GPU architectures [15],[17]. Theano is deeply optimized and has many applications in Machine Learning; however, as opposed to Matlab, it does not provide any sort of auto-configuration.

Pylearn2 [9] is an open-source Machine Learning library toolbox for easy scientific experimentation. Pylearn2 is built on top of Theano. With Pylearn2, one can easily define learning algorithms (e.g., neural networks) that will be then computed and optimized by Theano. As such, Pylearn2 may be seen as a Machine Learning front-end for Theano that eases experimentation.

TensorFlow is a data flow based programming model/interface for expressing and implementing Machine Learning algorithms, recently open sourced by Google. A computation expressed in TensorFlow can be executed on a wide variety of platforms ranging from handheld devices to GPU cards to large distributed systems. The flexible system can be used to express a wide variety of algorithms, including training and inference algorithms for deep neural network models. TensorFlow provides a Python API. It has been used for conducting research and for deploying Machine Learning systems into production, mainly Google's products. [18]

Torch is an open source scientific computing framework that supports a wide variety of Machine Learning algorithms, with a focus on optimizing for GPU. Torch uses the Lua programming language in a scriptable form (by way of LuaJIT) and an underlying C/CUDA implementation. Torch is used by Google DeepMind and Facebook (FAIR) among others, and is supported on Linux, MacOS, Android and iOS. A large number of extensions, referred to as packages, have been developed and contributed to the community, providing support for solving a wide variety of Machine Learning problems. [16], [29]

Microsoft has created a toolkit to enable development of Machine Learning algorithms for big data, tasks that are too large for a single machine, and require coordinated processing across a cluster of machines. Their Distributed Machine Learning Toolkit is open source, and precompiled binaries are available for Windows and Linux. The goal of DMTK is to handle the complexities of interprocess communication, distributed storage, thread management, etc. so that the researcher can focus on building their model. [30]

The interested reader can find a performance comparison among several tools in [16], [17], [18].

APPLICATIONS OF INTEREST TO CABLE

This section describes various problems in the cable network space, each with large sets of data, to which various Machine Learning techniques can be used to solve various problems.

Proactive Network Maintenance

Proactive Network Maintenance (PNM) involves processing vast amounts of fine-grained information about the state of the network to look for patterns that indicate impending problems. Current PNM solutions have focused on leveraging the physical-layer channel measurements performed by DOCSIS cable modems and CMTSs, along with knowledge of the HFC plant topology and well-known digital signal processing algorithms in order to locate physical degradations of the plant. ML algorithms could open up this space to identify a much wider variety of network impairments, particularly when those impairments aren't predictable using a priori knowledge.

CMTS and cable modem features and capabilities can be leveraged to enable measurement and reporting of network conditions so that undesired impacts like plant equipment and cable faults, interference from other systems, and ingress can be detected and measured. With this information, cable network operations personnel can make modifications necessary to improve conditions and monitor network trends to detect when network improvements are needed [3]. DOCSIS Downstream PNM Measurements and Data include: Symbol Capture, Wideband Spectrum Analysis, Noise Power Ratio (NPR) Measurement, Channel Estimate Coefficients, Constellation Display, Receive Modulation Error Ratio (RxMER) Per Subcarrier, FEC Statistics, Histogram, and Received Power. DOCSIS Upstream PNM Measurements and Data include: Capture for Active and Quiet Probe, Triggered Spectrum Analysis, Impulse Noise Statistics, Equalizer Coefficients, FEC Statistics, Histogram, Channel Power, and Receive Modulation Error Ratio (RxMER) Per Subcarrier.

Some of the above measurements are very valuable in that they can reveal problems in the network. We illustrate a couple of examples below to which we can effectively apply Machine Learning techniques. This will help in quickly and automatically identifying problems, instead of a human operator looking through the data and flagging issues manually.

PNM : Upstream equalization

The upstream pre-equalization mechanism relies on the interactions of the DOCSIS ranging process in order to determine and adjust the CM pre-equalization coefficients. These coefficients can be read to reveal problems, such as echo tunnels and excessive group delay. Echo tunnels make reflections that create ripples in the frequency domain. Group delay is typically caused by filters.

The set of coefficients on each modem in the plant can be monitored over time, and a Machine Learning algorithms can spot patterns in the changes, for a modem or a group of modems. Support Vector Machine or other classification algorithms such as K-nearest neighbors can help classify in trouble CMs immediately and help flag the operator long before a customer actually sees and reports an issue.

PNM : Downstream spectral analysis

A CM with full-band tuner capability acts like a spectrum analyzer. The spectrum analysis reveals suck-outs, excessive tilt, frequency peaking, unwanted filters, FM (or other) radio ingress, early roll-off, etc.

Each of these issues (signal patterns) can be learnt by a Machine Learning application which can then monitor live signals to flag problems in the network.

Machine Learning algorithms can look for abnormalities from the data obtained from each CM and across all CMs in the network. Grouping of CMs with common problems can be used to identify problems affecting multiple subscribers and isolate the cause of an issue in the cable plant.

Predictive Video Channel Lineups/ DVR Recording

Applications around the video delivery product could enable more efficient network utilization as well as better engagement from customers. Using Machine Learning in order to understand user habits, and to provide recommendations, is almost a cliché in the ML world. Any industry that has both a large catalog of assets and a large user base is in possession of a huge resource that can be automatically mined by a ML algorithm in order to provide recommendations.

In the case of video distribution, the ability to accurately identify video assets that might appeal to a viewer or set of viewers could have multiple applications. Some of the more obvious applications are VOD suggestions to customers. This is similar to the Netflix recommendation engine, see the Netflix Prize [28]. This is essentially a clustering and classification Machine Learning problem.

One could take this a step further with predictive DVR recordings for a customer based on content which is trending across the user base and intersecting that with the individual custromers preferences. Another idea is to pre-position the most popular content in CDN caches or in the local DVR storage. The operator could push trending content ahead of time, during off peak hours to reduce congestion during peak times.

One could also use such a recommendation system to improve the network traffic load and utilization. This problem can be described as the selection of video programs/channels for IP multicasting for a linear TV lineup. In different service groups on a cable plant, the viewership of the most popular channels will differ based on the demographics. So instead of making canned decisions on the channel lineup, and statically deciding which programs/channels make the cut for IP multicasting, an operator could use Machine Learning algorithms to understand the usage patterns over time and make recommendations to the operators on which channels to add to the multicast lineup dynamically.

ML algorithms which learn from video viewer data can do a much better job at analyzing which are the top programs in an automatic and real-time fashion. This learning could also apply in a weekly/daily timeframe or even in real-time, where say certain programs or channels are popular at certain days or times and the ML engine could automatically push those programs to be part of the multicast lineup at the appropriate times (e.g., 'Saturday Night Live' on Saturday nights). The benefit of this is the conservation of access network bandwidth in a IPTV deployment by intelligently planning the content on the network.

Profile Management

DOCSIS 3.1 introduces the concept of modulation profiles for OFDM channels. A modulation profile defines the modulation order to be used for each subcarrier within a channel. The CMTS can define multiple modulation profiles for use on a channel, where the profiles differ in the modulation orders assigned to each subcarrier. The CMTS can then assign each CM to the modulation profile that best suits it, taking into account the characteristics of the channel between the CMTS and that CM. Determining the best set

of modulation profiles to use on a channel is complex, given the number of CMs (and subcarriers) and the differences in signal quality that they experience on the plant. A Profile Management Application can help operators determine what the best modulation profiles are for each channel, given the channel characteristics seen by each CM on the network. The goal of optimizing profiles is to increase throughput per CM, maximize network capacity, and minimize errors.

Given a set of CMs using a DOCSIS 3.1 OFDM channel, the problem is to find a specific set of profiles which maximize the network capacity by assigning each CM to the best profile it can handle with a limitation on the number of profiles to what a system can support.

The input data set is the MER data per subcarrier for all the CMs. This can be directly translated into a bit-loading number for each subcarrier. So for example, an OFDM downstream channel with a 4K FFT has 3800 active subcarriers, and 200 CMs are using that channel. That means the input data set is a 200x3800 matrix of bit loading capabilities. Machine Learning clustering algorithms (like K-means) could help identify clusters of CMs which have similar bit loading capabilities across the whole channel. For each of these clusters a profile definition can be easily built by using the minimum bit loading in all the subcarrier dimensions, for that cluster of CMs. The set of profiles thus produced can be optimized further, balancing the number of profiles with the cost of having too many profiles.

Network Health Patterns

The access networks deployed by cable operators often experience abnormal behavior. Machine Learning techniques help create anomaly detection algorithms that are adaptive to changes in the characteristics of normal behavior in the network. Patterns in network traffic, equipment failures, device reboots or link failures can be used to identify the root cause of various network problems. Some examples are:

- CM/STB reboots after certain uptime.
- CM performance degradation compared to neighbors.
- IPv6 session failures on a certain SW version of the CM.
- CM performance as it relates to the time of day.
- Denial-of-service attacks on the plant.
- Correlation of customer WiFi issues to channel settings, etc.

For the example of a cable modem or a TV Set-top box rebooting in the network, the input learning parameters/features could be device uptime, software/hardware version, number of packets processed through the device, or the type of services running on a device. Some useful indicators available in the network are CM status, Upstream Transmit Level, Upstream Receive Level, Upstream SNR (MER), Upstream Codeword Error Rate, Downstream Receive Level, Downstream SNR (MER), DS Codeword Error Rate, etc. Each of these data points for a CM could reveal health issues at a given time. Over time an application could gather all the relevant data and supply it to an ML algorithm to figure out any patterns. This would be an unsupervised learning problem, where the idea would be to identify which of the features contribute towards a catastrophic reboot of a device. Principal component analysis could help in identifying the factors which contribute most towards a failure.

A similar approach would work for another class of problems debugging connectivity issues, this example is around IPv6 connectivity in an operator's network. A large population of devices in the network were experiencing IPv6 connectivity issues, which the operator was not aware of and only came to realize based on some IPv6 testing by a

third party. After manually tracing the broken IPv6 prefixes to a set of CMs, they realized that most of the failing modems were from one manufacturer and a specific software version. This type of debugging would be an ideal unsupervised learning problem, where a Machine Learning application could be running various network health tests and tracking issues or problems over time to identify problems and bubble them up the operator, before they become a major issue with the consumers or with the press.

Internet Traffic Classification

Traffic classification is the ability to understand the types of traffic flows which pass through the network and being able to identify the normal and abnormal patterns in those flows. It enables network operators to better manage their network resources, e.g., apply QoS or route traffic appropriately. It also helps with network security for an IPS/IDS (intrusion prevention/detection system) by finding abnormal patterns of traffic, and flagging a denial of service (DoS) or other attacks.

For every set of flows in the network, an operator wants to understand the patterns represented, as it will help make informed decisions or answer a particular question. For example, observing and learning from 5-tuple IPs flow (IP source/destination, Port Source Destination, Protocol) will help answer the question of which applications are generating a specific flow: Skype vs YouTube vs bitTorrent? For security applications, another pattern to learn from could be the aggregated traffic directed to the same IP destination in a certain timeframe. Given some well-defined notion of normal patterns of traffic, the goal would be to predict if the patterns from a new flow are markedly different.

Various supervised and unsupervised Machine Learning techniques can be applied to classify internet traffic. Traditionally port-based techniques are used to classify traffic, this is based on knowing the port numbers for the applications, but many new applications use random port numbers which makes this method of classification hard. Payload-based techniques are used by Deep Packet inspection engines which match the payload of the traffic based on well-known signatures. This method becomes hard with encrypted traffic. Machine Learning techniques such as classification and clustering can help solve this in an elegant fashion.

See [25], [26], [27] for examples of how to use ML techniques to the classification problem.

Network Traffic Engineering

Reacting to changes in network load is an important task for a network operator, since it is not cost-effective to deploy network equipment in excess of demand, while on the other hand, it is unacceptable from a user-experience perspective to run networks well into saturation. Properly done, tracking and responding to network utilization both for long-term capacity planning and for short-term traffic engineering purposes can result in well-functioning networks that operate cost efficiently.

ML techniques can bring some new tools to bear on this problem, with the result being quicker reaction to sudden changes in traffic flows, and the possibility – longer term – of these reactions being put into place automatically to ensure that networking resources are put to their best use.

A ML application could track utilization of various links, the byte counts of data traversing an interface, dropped packets at an interface etc., to get an understanding of the normal operation of the network elements. This by itself can reveal patterns in network utilization within a day-to-day time frame and also expose patterns for the longer term. Any

deviations from the norm can be automatically flagged for the operator. In the case where the application is tied into a SDN controller or orchestrator which can effect changes in the network, the application could actually create new paths or bring up new virtual routers etc., to handle the abnormal situation, for example, a temporary overload in data traffic.

Customer Churn Prediction

Given that acquiring new customers is generally three times the cost of keeping existing customers, an analysis of churners and an ability to identify potential churners can be very valuable. The task of churn prediction can be thought of as either a classification problem (classify a customer as a churner or non-churner) or a regression problem (calculate the likelihood that a customer will churn). Usually churn prediction is a classification task, and in particular, a binary classification task (i.e., customers are predicted to either churn or not churn in the next billing cycle). Multi-class classification is another possible approach. In this setting, customers could, for example, be predicted as belonging to either a low-risk, medium-risk, or high-risk class.

In the context of cable, a cable operator might want to identify cord shavers: subscribers with both broadband and video subscriptions who drop their video for over-the-top services such as Netflix, Hulu Plus, Amazon Instant Video, etc. For such a task, data would be required from several different sources. Linear viewing data would be useful as this would identify how much video a customer watches, what channels they generally watch, and what times they tend to watch. These features might be predictive of a cord shaver, as a customer that has a high-degree of interest in live sports may be less likely to drop his or her video service. Deep packet inspection (DPI) could also yield some useful features such as the percentage of bandwidth that streaming services like Netflix consume. Customer billing data would also be useful. For example, a customer who had subscribed to a promotional package for video and broadband might be more likely to drop video service once the promotional period ends. Combining the linear video, DPI, and customer billing record features would likely provide the most value. Some classification techniques that have proven successful in churn prediction include ensemble methods, SVMs, logistic regression, and random forests [19], [22]. There are also some indications that deep learning can yield some promising results in churn prediction [24].

One challenge with churn prediction is that churners tend to make up a small percentage of customers in any given billing period. For example, suppose that for a certain month, 2% of customers drop their video service. A classifier trained on this data would see 98% of the examples as non-churners. Some classifiers, such as Naive Bayes, use the class prior probability in predictions that are made. This would bias the classifier toward the majority class. Some techniques for addressing such class imbalances include undersampling the majority class (using fewer examples), oversampling the minority class (possibly by creating synthetic examples), and weighting the minority class as more important in the classifier being used [20], [22].

SDN Routing

Optimal or quasi-optimal routing has been a well-known challenge in computer networks. In routing, typically the objectives are to configure the routing policies in such a way that fulfill the requirements of the flows or that maximize the minimum link utilization. This area has been strongly limited by the fact that networks are inherently distributed systems where each node has a partial view and control over the network. In addition, the traditional destination-based

routing limits routing granularity and hence, the performance of the solution itself.

With the rise of the Software-Defined Networking (SDN) paradigm, routing becomes a centralized problem where the logically centralized controller receives all the information from routers and switches, computes an optimal solution and provisions the network equipment with the appropriate routing policies. In addition SDN also provides flow-based routing granularity with technologies such as OpenFlow [10] or LISP [11]. A notable example of this is B4, Google's SDN network [12].

In order to compute such algorithm, the SDN controller requires a model of the network, such model can be either an analytical or a computational model. In both cases the model may be hard to obtain or incorporate inaccuracies that usually arise when modeling real systems. In this context Machine Learning techniques can represent a solution for such inaccuracies.

Indeed, Machine Learning (ML) algorithms using a supervised approach can be trained with the monitoring data from the network. Specifically, the ML algorithm (see Figure 2 below) can be trained with the routing configuration, the load of the network (e.g., traffic matrix) and the resulting performance (e.g., delay or link utilization). With this, the ML algorithm is learning the function that relates routing, load with the resulting performance: function(routing configuration, load) = performance. Please note that this function is a model of the network and, if the dataset is large enough, an accurate one that can take into account not just common network behavior such as queuing but also complex ones such as the delay introduced by hardware, etc.

This model can be then explored online by the SDN controller using a traditional optimization algorithm to compute the optimal solution. For instance, the SDN controller may search which is the optimal routing configuration for a particular objective (e.g., delay) taking into account that the network is loaded with a particular set of traffic.

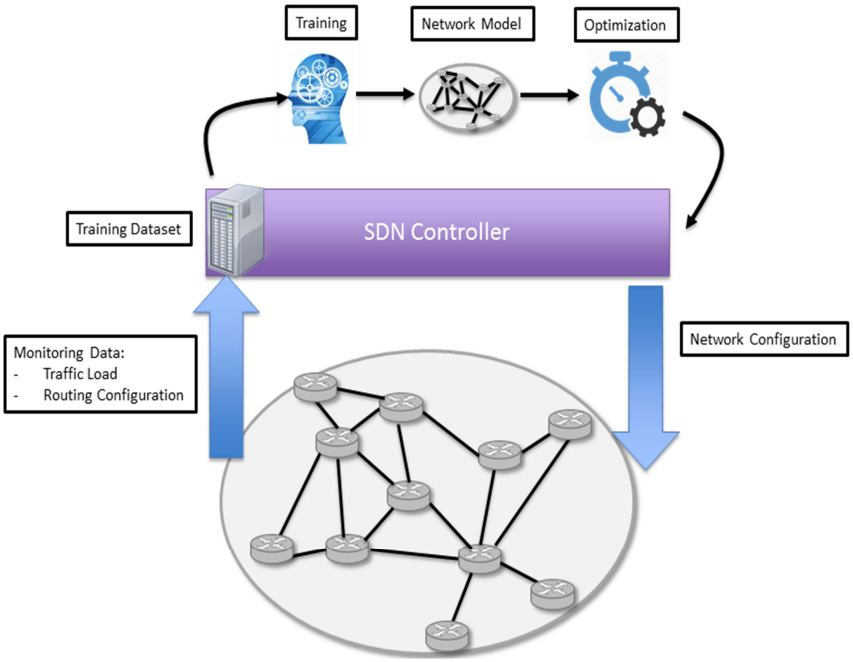


Figure 2. A Machine Learning-enabled SDN routing algorithm

NFV/SFC: Allocating VNFs to Appropriate VMs

Network Function Virtualization (NFV) [13] is a networking paradigm where network functions (e.g., firewalls, load-balancers, etc.) no longer require specific hardware appliances and instead are implemented in the form of Virtual Network Functions (VNFs) that run on top of general purpose hardware. Service function chaining (SFC) defines and instantiates an ordered list of instances of such functions, and the steering of traffic flows through those functions.

The resource management in NFV/SFC scenarios is a complex problem since VNF placement may have an important impact on the overall system performance. The problem of optimal Virtual Machine (VM) placement has been widely studied for Data Centers (DC) scenarios (see [14] and the references therein). However, in DC scenarios the network topology is mostly static, while in NFV scenarios the placement of a VNF modifies the performance of the virtualized network. This increases the complexity of the optimal placement of VNFs in NFV deployments.

In the VNF placement problem all the information is available, e.g., virtual network topology, CPU/memory usage, energy consumption, VNF implementation, traffic characteristics, current configuration, etc. However, in this case the challenge is not the lack of information but rather its complexity. The behavior of VNFs depends on many different factors and thus it is challenging developing accurate models.

In this context, some challenges that NFV resource allocation presents can be addressed by Machine Learning algorithms. Indeed, the SDN controller can characterize, via ML techniques, the behavior of a VNF as a function of the analytics data collected, such as the traffic processed by the VNF or the configuration pushed by the controller. With this model, the resource requirements of a VNF can be modeled by machine learning without having to modify the network. This is helpful to optimize the placement of this VNF, and therefore, to optimize the performance of the overall network.

In this case the scenario works as follows: the SDN controller receives a query by the user/owner to run a particular VNF, which is a black box from the SDN controller point of view. The controller, (e.g., via OpenStack) launches the new VNF and starts monitoring it, specifically it monitors the traffic being consumed by the VNF (e.g., by means of traffic features such as number of flows, distribution of the inter-arrival time, etc.) and a performance parameter such as delay or CPU consumption. With this dataset, the SDN controller trains a ML algorithm with the objective of learning the function(traffic features)=performance. Once the VNF has been characterized, the model can be used by optimization algorithms to optimally place the VNF and/or to provide delay guarantees of the virtual network.

BEST PRACTICES, GUIDELINES, CHALLENGES

This section documents some of the points to be aware of when framing and solving a problem using Machine Learning techniques. For a more detailed review please see references: [1], [2]

Clear Problem Definition

Before starting to solve a problem using any Machine Learning techniques, it is important to define the aim of the Machine Learning. Is the problem one of data distribution, identifying patterns, or making decisions? Is it a regression problem or a classification probem? What are the assumptions and boundaries? How would the

problem be solved manually? These questions help define a problem accurately.

Objective functions

Another important decision is to define how to evaluate the results of the Machine Learning algorithms. What is the metric or objective function which needs to be maximized or minimized. This will help compare learning algorithms and determine how much better it does as compared to simple or random predictions.

Proper Features:

Lots of thought needs to be given to the decision of how to represent the data available. There may be a set of original features which are relevant and other artificial features can be developed as well. Data and feature definition is an essential part of system design.

Suitable Methods and Algorithms

The main guideline to remember is to try out different ML techniques and algorithms. There are many effective methods and algorithms and no algorithms which work perfectly for certain applications/use cases. As the training set increases, it reduces the effect of algorithm selection.

Adequate Representative Data

The learning data needs to represent realistically the space; this is important to make sure the ML algorithm understands the whole range of inputs and outputs. If not, the predictions will be skewed. Random re-sampling of the data is preferred, and data cleaning will be necessary in many cases. Another guideline is that the training samples are expected to be much more in number than the testing samples and samples to be predicted.

Training data Testing and Cross Validation

With any given dataset of training examples, if you use the entire data set to train your model, the final model will normally overfit the training data and the error rate estimate will be overly optimistic. A better approach is to split the training data into disjoint subsets: a training set (e.g., 80%) and a test set (e.g., 20 %). The procedure is to run the algorithm on the training set, minimizing the error of the learning function. Then the idea is to compute the error on the test set to figure out how good the model is. This training vs test split has issues: if the training data is a sparse dataset, one may not afford setting aside a portion of the dataset for testing, and also if the split happens to be not representative of the data set, the model learnt will be off. These limitations can be overcome with resampling methods like cross validation.

Metrics / Evaluation Criteria

When evaluating a classification or regression task, it's important to choose the right metric for the task. Classification is commonly measured by accuracy. Accuracy is defined as the percentage of predicted labels that were correct. However, this metric is not always the best measure of success. For example, in a task with an imbalanced class distribution (such as churn prediction), a very bad classifier can obtain a high accuracy score. Suppose that 98% of the examples in our test set are of one class, *Class A*. A classifier can simply predict *Class A* for every example and obtain an accuracy of 98%. However, in tasks such as churn prediction or fraud detection, identifying the minority class is very important. For this reason, accuracy is not generally a good metric for evaluating the performance of a classifier when dealing with an imbalanced dataset. A better way to view the performance of binary classification would be to examine the number of true positives

(tp), false positives (fp), true negatives (tn), and false negatives (fn). For churn prediction, the class of "churner" would be considered the positive class. So churn prediction success might be measured by Precision, defined as $\frac{tp}{tp+fp}$, or Recall, defined as $\frac{tp}{tp+fn}$ [23]. In general, it's always useful to keep the high-level objective of the task in mind. In tasks such as fraud detection or disease identification, a false negative will likely be considered more costly than a false positive. So in these cases, the classifier would be optimized to minimize false positives.

Evaluating a Hypothesis

A hypothesis may have low error for the training examples but still be inaccurate (because of overfitting). The idea is to troubleshoot errors in predictions by some of the following techniques: increasing the number of training examples, trying a smaller or larger set of features, changing the parameter settings in an algorithm, etc.

Above all are statistically significant results from a given ML technique: It goes without saying that the predictive results of a ML algorithm should be of a much higher possibility than random choices at a very minimum.

CONCLUSIONS

There are very many good use cases within the cable access network for the application of Machine Learning Algorithms. The ultimate goal of applying Machine Learning to networks is to provide automation and remove many of the manual tasks associated to network control and operation. Problems which have large sets of data are tough to solve manually and ML algorithms provide an intelligent way to get knowledge from the data.

This paper describes nine differnet and relevant use cases to which machine learning can be applied today. The application of these will give the operators knowledge about their networks and how to most effectively run the network. ML algorithms will be assimilated into systems which can configure networks, identify issues and take the appropriate corrective action.

There are important open research challenges that need to be addressed before this vision is fully implemented. First, typically ML is applied to scenarios that are resilient to errors, such as image recognition. However, computer networks do not handle errors well. In this context there is a need to understand how such probabilistic Machine Learning techniques can be effectively incorporated to networks. And second, learning requires a representative dataset. But what does representative mean in the context of computer networks? Are observable network loads and configurations (i.e., those that do not break the system) representative enough to provide accurate estimations? Generating a training set that allows the algorithm to learn the boundary between functioning and failing networks can be challenging. At the time of this writing, there is no clear answer to this question and will require further research efforts.

In addition there are also non-technical issues that need to be addressed. Machine Learning builds on top of huge datasets that, in some cases, will contain user information. As a consequence there are privacy issues since there is a potential of leakage of personal information. This will require that the appropriate safety mechanisms are put in place and, if possible, to always learn based on aggregated-information.

As a result, the use-cases that do not directly operate over the network but that rather provide recommendations to the network owner/administrator will be commercialized first. Such systems are safe since they involve human intervention and validation, and will help operators get used to this new technology and better understand its potential.

REFERENCES

[1] Ji Shufan, Case Study: Autonomic Network Configuration Using Machine Learning, https://www.ietf.org/proceedings/94/slides/slides-94-nmlrg-2.pdf

[2] Pedro Domingos, A few useful things to know about Machine Learning, Communications of the ACM, Vol. 55 No. 10, 2012. http://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf

[3] (DOCSIS PHYv3.1), DOCSIS 3.1, Physical Layer Specification, CM-SP-PHYv3.1-I08-151210, Cable Television Laboratories, Inc. http://www.cablelabs.com/wp-content/uploads/specdocs/CM-SP-PHYv3.1-I08-151210.pdf

[4] Stanford Professor Dr. Andrew Ng, Machine Learning Course on Coursera https://www.coursera.org/learn/machine-learning

[5] Clark, David D., et al. "A knowledge plane for the internet." Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications. ACM, 2003. http://groups.csail.mit.edu/ana/Publications/PubPDFs/A%20knowlege%20plane%20for%20the%20internet.pdf

[6] Michalski, Ryszard S., Jaime G. Carbonell, and Tom M. Mitchell, eds. Machine learning: An artificial intelligence approach. Springer Science & Business Media, 2013.

[7] Matlab's Neural Network Toolbox and Machine Learning toolbox : http://www.mathworks.com/products/neural-network/index.html, http://www.mathworks.com/products/statistics/index.html

[8] Theano:, http://deeplearning.net/software/theano/#

[9] Pylearn2, http://deeplearning.net/software/pylearn2/

[10] McKeown, Nick, et al. "OpenFlow: enabling innovation in campus networks." ACM SIGCOMM Computer Communication Review 38.2 (2008): 69-74.

[11] Farinacci, Dino, Darrel Lewis, David Meyer, and Vince Fuller. "The locator/ID separation protocol (LISP)." (2013). (RFC 6830)

[12] Jain, S., Kumar, A., Mandal, S., Ong, J., Poutievski, L., Singh, A., ... & Zolla, J. (2013, August). B4: Experience with a globally-deployed software defined WAN. In ACM SIGCOMM Computer Communication Re-view (Vol. 43, No. 4, pp. 3-14). ACM. Chicago

[13] Matias, J., Garay, J., Toledo, N., Unzilla, J., & Jacob, E. (2015). Toward an SDN-enabled NFV architecture. Communications Magazine, IEEE, 53(4), 187-193.

[14] Berral, J. L., Goiri, Í., Nou, R., Julià, F., Guitart, J., Gavaldà, R., & Torres, J. (2010, April). Towards energy-aware scheduling in data centers using machine learning. In Proceedings of the 1st International Conference on energy-Efficient Computing and Networking (pp. 215-224). ACM.

[15] Bergstra, James, et al. "Theano: a CPU and GPU math expression compiler." Proceedings of the Python for scientific computing conference (SciPy). Vol. 4. 2010.

[16] Collobert, Ronan, Koray Kavukcuoglu, and Clément Farabet. "Torch7: A matlab-like environment for machine learning." BigLearn, NIPS Workshop. No. EPFL-CONF-192376. 2011.

[17] Bastien, Frédéric, et al. "Theano: new features and speed improvements." arXiv preprint arXiv:1211.5590 (2012).

[18] TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, http://download.tensorflow.org/paper/whitepaper2015.pdf

[19] "Handling class imbalance in customer churn prediction", J. Burez and D. Van den Poel, Expert Systems with Applications Volume 36, (pp. 4626-4636), (2009)

[20] "SMOTE: Synthetic Minority Over-sampling Technique", Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, W. Philip Kegelmeyer, Journal of Artificial Intelligence Research Volume 16, (pp. 321-357) 2002

[21] scikit-learn- http://scikit-learn.org/stable/

[22] “Telco Churn Prediction with Big Data”, Yiqing Huang, Fangzhou Zhu, Mingxuan Yuan, Ke Deng, Yanhua Li, Bing Ni, Wenyuan Dai, Qiang Yang, Jia Zeng, SIGMOD '15 Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, (pp. 607-618), 2015

[23] Precision and recall definition: https://en.wikipedia.org/wiki/Precision_and_recall

[24] "A New Neural Network Based Customer Profiling Methodology for Churn Prediction", Ashutosh Tiwari, John Hadden, and Chris Turner, "Computational Science and Its Applications -- ICCSA 2010: International Conference, Fukuoka, Japan, March 23-26, 2010, Proceedings, Part IV", (pp. 358-369) (2010) Springer Berlin Heidelberg

[25] A Machine Learning Approach for Efficient Traffic Classification Wei Li and Andrew W. Moore, http://www.cl.cam.ac.uk/~awm22/publications/li2007machine.pdf

[26] An SVM-based machine learning method for accurate internet traffic classification, Ruixi Yuan & Zhu Li & Xiaohong Guan & Li Xu, https://pdfs.semanticscholar.org/556a/209838cf5c742d2f9a75a3436a0de6611526.pdf

[27] Realtime Traffic Classification Based on Semi-supervised Learning, Chengjie GU1, Shunyi ZHANG, Xudong CHEN, Anyuan DU, http://www.jofcis.com/publishedpapers/2011_7_7_2347_2355.pdf

[28] The BellKor Solution to the Netflix Grand Prize, http://www.netflixprize.com/assets/GrandPrize2009_BPC_BellKor.pdf

[29] Torch | Scientific computing for LuaJIT, http://torch.ch/

[30] Distributed Machine Learning Toolkit, http://www.dmtk.io/

Assessing Network and Equipment Failures in the New SDN/NFV Architectures

Marlon Roa
Chris Liou
Vijoy Choyi
Mark Bieberich
Infinera

Abstract

There is little doubt that SDN (Software Defined Networking) and NFV (Network Function Virtualization) are, jointly, a disruption and the breaking point for the telecommunication industry. The promise of lower cost and enhanced functionality of a service provider network is now palpable as we start seeing both technologies mature.

As these technologies converge toward a standard approach, the next step is to look into the details of actual implementation. Among those details, the issue of failure prevention of the network supporting and implementing the said virtualization stands out since a bad implementation in this new control layer approach could, ultimately, create critical failures in the data layer.

This paper explores this topic aiming to raise awareness of the potential complexity of the said network which cannot be omitted from the overall SDN and NFV business case.

INTRODUCTION

SDN and NFV are being implemented in the transport network attempting to follow the success of the software-defined evolution in data centers along with its focused hardware pieces (i.e. controllers and white-boxes). SDN moves the control layer out of data layer equipment or custom management systems and into a central "controller" platform, which now executes the control layer functions for multiple data-layer devices. This lowers the cost of the data-layer devices and enhances the scalability and flexibility of a particular network. It also allows for the easy addition of new services and includes the potential of interoperability among operators' networks for fast multi-provider service turn-up, for example.

In the case of NFV, service providers are seeking solutions that provide a non-purpose-specific, data-layer hardware platform where diverse data-related functions can run either in a central server or in the data-layer equipment as software applications. Again, the objective is to reduce the cost of equipment and enhance scalability of the providers' service offerings.

As a result, the SDN controller hardware concentrates critical transport decisions for its particular set of data-layer devices. The central NFV Management and Orchestration (MANO) solution controls the interworking of specific functions being run by, or controlling, the function execution in a set of data-layer devices within a particular network segment.

Let's approach the analysis of preventing and resolving network failures in SDN/NFV networks from different perspectives:

- Hardware and site failures of the SDN controller, SDN orchestrator or NFV MANO
- Network failures in the interconnectivity between SDN controllers and orchestrators
- Network failures in the interconnectivity between SDN and NFV devices and the data-layer equipment

- The added complexity and cost as a consequence of a resilient SDN/NFV network

For the sake of simplicity, we assume the SDN controller and NFV MANO coexist in the same server under their own VM (Virtual Machine) space. When this is not the case, and they run in their own hardware, one can extrapolate that the complexity for redundancy must be duplicated for each virtualization scheme increasing the cost of the overall deployment.

SDN/NFV HARDWARE REDUNDANCY CONSIDERATIONS

In a traditional network management system (NMS) reliability is necessary to guarantee constant client-side access and to preserve the databases containing configuration and status information. HA (High Availability) schemes have been used to implement redundant NMS systems including geographical diversity in some cases.

As the number of nodes to be managed increased, it was necessary to change the NMS architecture to allow the implementation of hierarchical or distributed approaches, beyond the initial centralized method. The servers, under which the NMS runs, already include redundancy features such as protected hard drives, fans, power supplies, processor redundancy or full hardware redundancy where software can maintain a load balancing scheme as well as monitor the health of the hardware cluster. In addition, modern NMS implementations can be virtualized over common hardware.

The NMS has usually been a proprietary solution, so the vendor included, or recommended, the server hardware to guarantee proper behavior, especially when in a distributed or hierarchical mode. In addition, NMS redundancy methods also add security measures against malicious attacks from the client access perspective and from the network node intercommunication (security is outside the scope of this paper).

The increased number of managed nodes dictated the implementation of a new server or a new load in a distributed system was needed and it also determined if redundancy was needed for this new added server.

As we turn into SDN/NFV, not much changes in terms of server redundancy as the industry was already starting to virtualize the NMS implementations and loading it in common white-boxes.

If the NMS solution in the network has not been virtualized, then there is the extra cost of switching to such an environment for SDN/NFV which involves not only the hardware and interconnect, but the retraining and/or hiring of staff with the correct skills set to maintain the virtual environment.

Hence and up to this point the level of complexity to fit a redundant hardware scheme to prevent single-point of failures for SDN/NFV deployments does not differ much from the classic NMS approach.

However and as a main departure from the traditional NMS scheme, SDN and NFV are actually part of data processing. In the case of SDN, routing, re-routing and cross-connect decisions are part of its responsibility and in the case of NFV, actual data plane functions can take place in a server and/or a fast and reliable transfer of a data-related function from the NFV server to the node is now critically important. This change in network management paradigm affects the hardware and interconnectivity needs, and the latter is covered in the next sections of this paper.

Also, it is important to point out that *"Service continuity is not only a customer*

expectation, but often a regulatory requirement, as telecommunication networks are considered to be part of critical national infrastructure, and respective legal obligations for service assurance/business continuity are in place" (NFV ETSI [2] specification).

It is then critical that SDN controller functions have a high degree of availability to avoid long periods where the network does not respond to new connections, new labels (Ex: IP addresses, MPLS labels, Ethernet VLANs or MAC addresses) or lack of reconfiguration due to failures or traffic congestion, for example.

To this end, Openflow switch specification 1.2 added the support of multiple controllers, namely slave/master/same controller objects, for redundancy issues, and it leaves it up to the industry the design of the topology to implement the protection.

In the case of NFV, it is even more critical that the server with the NFV MANO function has a high availability as a failure in this server would cause the lack of loading customer requested functions which can violate SLAs or, worse, create a situation where the network assumes an NFV function has been implemented and opens it to customer traffic, which worst case could become a legal matter (Ex: Failed firewall or encryption VNF execution and customer traffic to be protected is enabled to be transported anyway).

In both cases (SDN and NFV) and due to the amount of traffic being processed, latency concerns and/or the number of nodes being managed, the controller scheme is forced to migrate to a hierarchical approach similar to what occurs with the traditional NMS case. In NMS, we could have a slave NMS server that did not need to be backed up since its data was periodically stored in the HA master NMS server cluster and it was seen as not critical for operations since traffic continuity was always maintained by the underlying network elements. Instead and in the case of SDN/NFV, the servers handle portions of data processing which, in most cases, will call for redundancy in these slave locations.

Hence, an increased expense is incurred (vs traditional NMS model) by adding HA schemes to the slave location which can call for cluster structures requiring not only protection servers, but HA monitor solutions where all this hardware may have to be located in non-temperature control environments requiring the extra expense of hardened hardware.

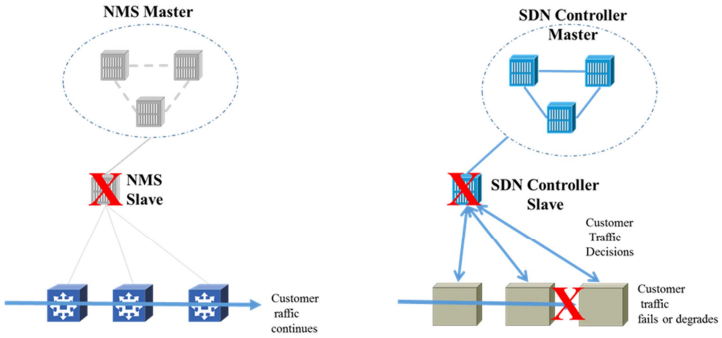


Figure 1 – Management layer failure and traffic effect

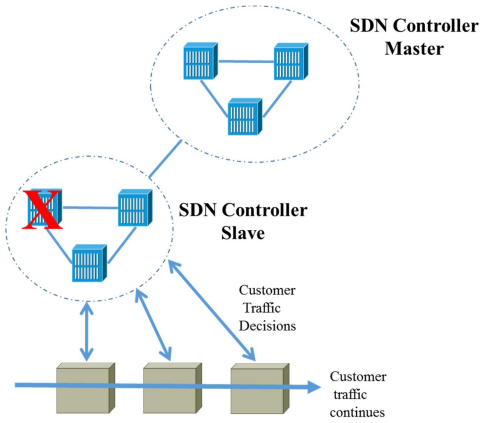


Figure 2 – Required Server redundancy to guarantee traffic continuity

Differing from SDN, NFV requires a level of redundancy on the host of the VNF whether it is a white-box or the network element itself. In the traditional method, the network functions resided in the same hardware processing the data plane without the need to be downloaded, instead they remained idle until required to execute.

This meant there was no need to flag its existence or its correct deployment except for an after-reboot/reset integrity check that protected all logic in the hardware. Under the new NFV scheme, the hardware processing the data plane needs to add a method to check the integrity of the VNF to be run, then exchange a control protocol with the VNF MANO and add cache and logic that can restore the executing VNFs in case of reboots or resets events.

Enhanced complexity is needed to ensure that after a reboot/reset event, the NFV MANO was not in the middle of a VNF transfer along with the resync of the hardware with the MANO to ensure proper operation moving forward and, in the worst case, re-transmit many VNFs from MANO to the network device making the outage even longer for the customer traffic. This prolonged outage could be reduced by adding full hardware protection for the hardware running the VNFs.

Yet, the recommendation would be to establish SLAs for networks with VNFs to include longer outage time allotment after a reboot/reset event than a traditional network. Additionally, extra-testing of the NFV devices reaction to this type of events must be included in qualification testing to avoid possible long-term outages in production due to sub-par implementations.

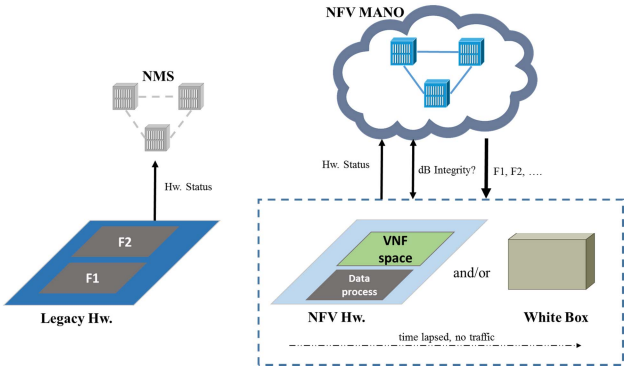
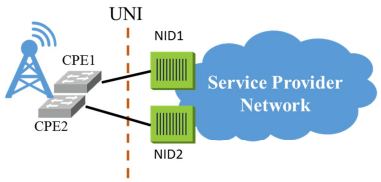


Figure 3 – Reboot/reset scenario – Legacy and NFV network element

In general, NFV ETSI recognizes the need for HA of the new solution and so it requires support for

- *"Service continuity and failure containment*
- *Automated recovery from failures*
- *Eliminating single points of failure in the underlying architecture*
- *Multi-vendor environment*
- *Hybrid Infrastructure*" NFV ETSI

However, it does not enforce a common approach to solving these issues. Hence, the selection among the different vendors of controller software, server hardware, network elements and their interconnectivity to guarantee carrier grade service becomes a new item for operators to evaluate vendors and solutions for.



Legacy Case

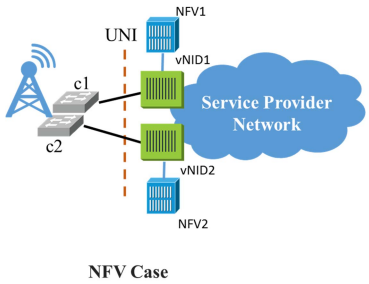


Figure 4 – Customer request for separate hardware at UNI

There is also the very real possibility that customers of a service provider will be unwilling to share the same hardware for their VNFs with other customers VNFs as it is common in demarcation deployments, for example. In this case and if VNFs are running in a common white box near the CPE site, the service provider would be pushed to add a white box and node element for that customer, possibly not being able to charge for the multiple hardware pieces since it is common today to provide the dedicated CPE device at no extra cost.

In all redundancy and load balancing cases, server hardware redundancy is a key part of any protection scheme (including HA, 1+1 or 1:N). SDN and NFV transactions and objects for operation are being standardized, but no effort has been made to normalize the multi-vendor space of server hardware. Hence, an extra operational cost is added to the service provider reflected in extra qualification testing to ensure that the selected vendors servers run the SDN and NFV at a similar processing and connectivity performance level, so if deployed and executing protection switching, the overall data-processing functions will not suffer by a mismatch in performance from one vendor server to another.

Besides the SDN controller, there exists the SDN Orchestrator that is not necessarily part of the traffic processing portion of the network; hence, the approach to redundancy can follow the NMS or datacenter HA concepts without any special provisions.

Then, looking at cost and complexity for the hardware portion of SDN/NFV, the SDN Orchestrator protection is similar to the traditional datacenter and NMS protection schemes, also mixed with OSS. The SDN controller is a new layer of protection where the hardware redundancy requirements are similar to the said NMS and Datacenter schemes, except that these instantiations of carrier-grade redundancy for SDN/NFV controllers can be expected to be far more numerous than in the legacy case and with the probability of requiring hardened server solutions in some locations.

In addition, the operational costs can increase as a network element solution now requires that the SDN management entity software, its host hardware, the NFV MANO, its host hardware and the network element, all of which can be from different vendors, work in fully synchronized and in a manner that their individual execution do not disrupt service continuity or perform below traditional networks. This performance has to be added as new testing by the service provider vendor or solution qualification process.

A way to mitigate this cost increase is to embrace the traditional selection of a handful of vendors where each owns the set of software and hardware of network elements and the management. This basically reduces the service providers cost savings by locking them again to a particular vendor for the network elements, management (SDN controller, NFV MANO) and the necessary software.

Alternatively and as a better option, the standards bodies can define a certification for hardware and software in a multi-vendor space where performance details are included in the test plan as processing performance,

ports speeds, protection topologies and protection switching performance parameters.

HIGH AVAILABILITY CONSIDERATIONS BETWEEN SDN/NFV LAYERS

When considering an SDN-centric solution that involves distributed orchestrators, controllers, and physical network elements, it is necessary that an architecture is put in place that ensures survivability and a proper level of availability under multiple potential failure or communication degradation scenarios. The distributed system as a whole needs to be tolerant to faults and conditions that include loss of communications, loss of data, and loss of synchronization, as well as potential overload conditions where specific application instances may be subject to higher loads than anticipated.

This section describes some of the key issues and considerations regarding the communications network between the orchestration and SDN control layer, as well as between the SDN control layer and the physical network infrastructure.

Resiliency Considerations: Orchestrator & SDN Control Layer

The general commonly-accepted multi-tiered architecture for a scalable & distributable SDN solution involves service orchestration systems interacting with typically multiple SDN controllers. Service definition and orchestration is generally handled by the orchestrator, and the configuration of the service into the often multi-domain, multi-vendor network infrastructure is achieved via programmatic interaction with generally different vendors' SDN controllers, generally through an Application Programming Interface (API). This separation of functionality allows the orchestrator to focus on service abstraction and service lifecycle functionality on an end-to-end basis, while the SDN controllers can focus on network-specific configuration, management, and control for implementing its specific portion of the end-to-end service.

The functional separation of roles and responsibilities helps to facilitate the distribution of software modules running on different compute platforms, interacting via some form of client/server information exchange inherent to a well-defined API. This distributed software building block approach not only helps minimize the impact from changes and independent evolution of each individual software module, but also helps to facilitate software scalability, using techniques described in a later section of this paper.

However, distribution of software functions and the API's between the orchestration and SDN control layers present a potential failure point which requires attention in order to maintain system integrity and resiliency. Examples of failure or degradation modes that an HA solution should address include:

- Loss of data communications between client and server
- Failure of client- or server-side application
- Performance degradation due to insufficient server-side resources or excessive information flooding the client application
- Indeterminate state of server application

These issues are not dissimilar from Data Communications Network (DCN) resiliency issues that exist today between OSS/BSS systems and vendor network management systems. Techniques used in recovery from failures or degradation in these scenarios can be leveraged, with the added benefit of new open-source technologies that may make the implementation more standardized or easier.

Remedies for detection and correction of such issues include:

- Client/server heartbeat monitoring
- Performance & health monitoring of SDN controller application for detection of performance degradation
- Methods for both deep data synchronization as well as rapid re-synchronization to recover from communication lapse
- Active/active or active/standby SDN controller application architecture for redundancy and rapid failover
- Dynamic instantiation and load balancing across multiple server application instances
- Networking resiliency between orchestration and SDN control layer (generally employing mesh networking with multiple communication paths and IP networking techniques)

In scenarios where the orchestration and SDN control layer functions may span a sufficiently large geography, it is necessary to additionally take into consideration the added complexity of increased communications latency, and potentially reevaluation of the use of synchronous or asynchronous client/server communications, as well as whether strong consistency or eventual consistency of state is needed for the specific use-case scenario.

Intra-Orchestrator & Intra-SDN-Controller Resiliency

Within each of the software architectural layers are additional failure points that must be guarded against. Resiliency from internal software or database failures must also be accounted for – watchdog monitoring and general software lifecycle management (including restart, resynchronization, and state recovery) are typical methods employed by the vendor of the solution.

Software functionality within the layers may themselves be distributed across multiple servers, which again presents similar resiliency challenges as mentioned in the previous section, with the exception that the interaction between software modules within the orchestration or SDN control layer is addressable by the vendor's own design options, whereas the resiliency between the orchestration and SDN control layer is generally integral to the formalized API that defines the interaction between layers.

Depending on the implementation of the SDN control layer, additional challenges in maintaining resiliency may exist. SDN control solutions that span large geographies or large numbers of systems, for example, may internally support distributed "root/leaf" architecture to co-locate some functionality closer to the physical infrastructure layer, while still maintaining an aggregated, logically centralized view of resources within the root controller. In addition to survivability of communication lapses between the root/leaf sub-layers of the SDN control layer, database resiliency and state consistency between the root/leaf controllers may present additional challenges, but these would generally be addressed by the vendor of the SDN control layer.

The NFV case

As it was stated in the introduction, this paper assumes SDN and NFV share environment following in the work of ONF and ETSI [8]. Within the NFV MANO, the communication between the different control and monitor parts (orchestrators and managers) follows a similar scheme as in the SDN controller/orchestrator case.

The communication between NFV and SDN will also need to be protected using normal cluster/HA methods. A more delicate interconnectivity situation is present between

the NFV control layer and the data plane device.

Resiliency Considerations: SDN Control Layer & Forwarding Layer

The communication layer between the SDN controller and the physical network presents another important potential failure point that needs to be protected against. SDN configuration and control data, along with associated faults, events, and performance monitoring data, are typically passed between these layers. Here, router-based DCN resiliency to the forwarding elements plays an important role, ideally with some form of low-latency and reliable transport with sufficient QoS to help mitigate communication degradation situations.

In scenarios where full DCN redundancy is not available or cost-effective, such as in some WAN environments, additional communication channels via in-band paths may be required. In these environments, Gateway Network Element (GNE) functionality can be utilized to proxy control plane communications to the target system, ensuring an alternate communication path to the DCN path. This is generally required independent of the device-specific protocol used to communicate forwarding instructions and information.

In scenarios where control plane communications to the end forwarding device is proxied through a pair of redundant system controllers, additional provisions are needed to handle system controller failover and replacement scenarios. Depending on the way system controller redundancy is implemented, this may involve custom recovery schemes.

HIGH AVAILABILIY ARCHITECTURE

The requirements for High Availability can vary depending upon the specific component of the SDN or NFV controller. This is the case when it comes to the connectivity layer at the northbound and at the southbound interfaces of the SDN or NFV controller.

Controller to Data plane connectivity

The HA requirement on the connectivity between the controller and the data plane is quite stringent. Firstly, the network-wide state maintained by the controller will get out-of-sync with the network, and the control plane functions in the controller now makes incorrect or poor decisions based on this data. This would be the case when the packets that were not delivered were routing updates for some link state based protocol. Secondly, the loss of connectivity could result in down time for end users, as would be the case when the network element needed the controller to decide on how to handle some data traffic for which a rule was not yet programmed.

To ensure resiliency in the connectivity between the controller and the data plane, a common technique is to maintain multiple connections between the controller and each network element. Along with this, the network connecting them needs to have multiple diverse paths that the above mentioned connections can traverse.

Figure 5 depicts the redundancy model wherein a connection exists between every node of the controller cluster and a network element. Also, by using techniques such as subnets we can ensure that the path taken by each of the connections does not overlap. In this case, let's say a link gets severed, then we still have 'n-1' connections available between that NE and the controller, assuming there are 'n' nodes in the controller cluster. The connection itself can be established by either side initiating it. It is quite common these days for network elements supporting

Openflow to allow configuring multiple controllers to concurrently communicate with.

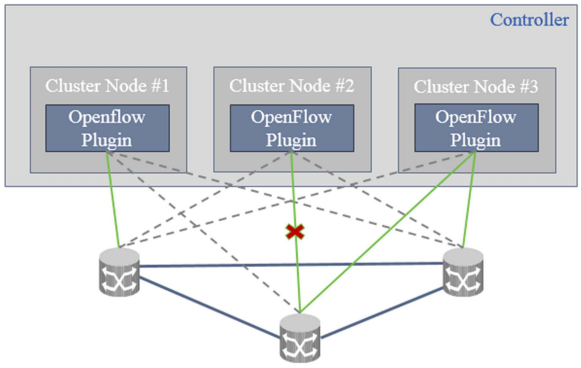


Figure 5 – Connection redundancy between controller nodes and network nodes

Having put in place redundant connections to handle connection failures does take care of making the connectivity resilient, but it does introduce some level of complexity within the controller. When the controller needs to program a set of rules or send down other commands to the NE, it now has the option of distributing the requests across the connections concurrently. Although this may seem like an attractive option w.r.t scalability, it can lead to issues of ordering and dependencies between the requests may break. In order to prevent this, the controller will need to consider just one of the connections active at any given time and the others as standby. This way the write operations from the controller always go on the active connection, while operational status updates coming from the NE can still be accepted over every connection.

The controller cluster can have an election process between the cluster nodes in order to decide which node's connection is to be given the active status. The election can occur on first discovery of the NE and also whenever the active connection fails.

With NEs supporting southbound protocols like Openflow, the default role of each connected controller is equal. However, this can be overridden such that the role can be set to slave. Once the active connection election occurs in the controller cluster, each of the controller cluster nodes can then request the Openflow NE to switch the role accordingly.

In the case of NFV MANO, we must assume that the connectivity is direct between the MANO and the network device. A low latency link is a must as the NFV aims to provide the same performance as legacy carrier-grade data-plane solutions. In addition, the bandwidth must be high enough to handle control layer traffic and any data-layer traffic transported over the same link. A QoS scheme can guarantee coexistence of both low latency data transport and control traffic continuity.

The link can be protected in a scheme that provides 1+1 protection or, in some non-critical cases, an in-band alternate path can be implemented, but the provider must be aware that non-network-element resident VNF execution could degrade its performance to levels that may violate the stipulated SLA.

As it can be concluded, sharing the same access path to the network element by NFV and SDN can hinder scalability of both SDN and NFV or it can impact the performance of VNF execution due to the higher requirement for latency and speed in the VNF case. Then, as VNFs are loaded in a particular network element location, it can be expected that as a worst case, the VNF load and the SDN control tasks can grow to a point where a network element can be required to handle a physical connection for SDN and another physical connection for NFV, both of which can be protected. Another alternative is to upgrade the management link to a high speed, optical-base connection. Yet and in either case, the cost and complexity increase from that of an equivalent legacy implementation.

<u>Controller to Orchestrator connectivity</u>

Although we need High Availability for the connectivity between the controller and the orchestrator, the requirements are not as high as that at the southbound interface. Since the orchestrator is not directly dealing with packet-in messages coming from the dataplane and not in charge of maintaining the global network state, it can afford to sustain occasional glitches in connectivity with the controller. Moreover, with automation and retry logic that orchestrators are typically built to handle, it should be quite trivial for the orchestrator to re-try any provisioning tasks that got impacted while the connectivity was down with the controller.

Given that a typical SDN controller in itself is comprised of a cluster of nodes where each node is active, it is not going to be feasible and is not recommended for the orchestrator to directly be exposed to the cluster nodes. This is because the controller can at any point in time decide to spin up or bring down nodes in its cluster based on the load and this should in no way impact the interaction with external systems. By fronting the controller cluster nodes with a load balancer, we not only achieve this decoupling with external systems but also get the ability to distribute the request load across the nodes of the controller. With this, the orchestrator then only interacts with a virtual ip address that is serviced by the load balancer. The load balancer will appropriately distribute the incoming requests to one of the cluster nodes depending on the load balancing policy configured by the network administrator.

To ensure HA in the connectivity between the controller and the orchestrator, we now need to ensure not only that there are redundant paths deployed in the network to reach between them, but also that there is a redundant load balancer.

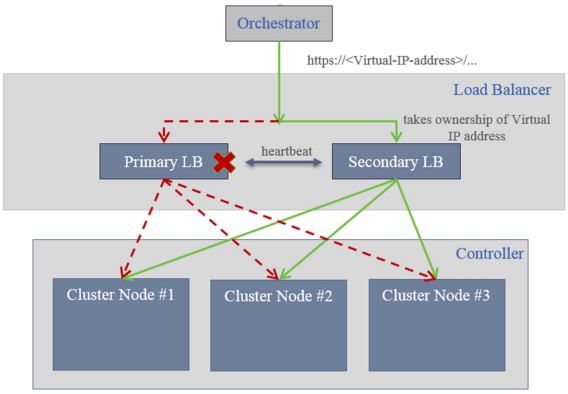


Figure 6 – Redundant load balancing

As can be seen from Figure 6, the secondary load balancer detects that the primary has failed based on loss of heartbeat. The secondary then takes ownership of the virtual IP address and updates the ARP cache. From that point on, any new request from the orchestrator will get routed to the secondary load balancer, which in turn will distribute the requests to one of the cluster nodes of the controller.

This model also handles scenarios where one of the controller cluster node fails. In such cases, the load balancer will detect the controller cluster node failure and then ensure that future requests from the orchestrator are forwarded only to one of the remaining cluster nodes.

ADDED COST AND COMPLEXITY OF A RESILIENT SDN/NFV NETWORK

From the data plane to the controller and orchestration layers, the migration to SDN and NFV requires operators to embrace new approaches to network resiliency. A resilient SDN and NFV network architecture is critical to maintaining and enforcing SLA's; however, there are important cost and complexity implications that must be identified, evaluated and planned for as part of the process to validate the network virtualization business case.

In comparison to more simplistic resiliency measures in traditional NMS and OSS architectures, where data exchanges with the network are administrative in nature, SDN and NFV execute control plane operations and network function instantiation—critical functions that have a direct impact on the ability of the network to respond to end-user demand. As a result, SDN and NFV require more sophistication in network resiliency planning, design and testing to ensure SLA's are met and operations can scale, reliably, over the long term.

Added cost and complexity that result from SDN and NFV resiliency can be grouped into the following categories: qualification testing, hardware and software distribution, and physical interconnectivity.

Qualification Testing

One of the central tenets of network virtualization is openness, the concept of enabling integration and interoperability of disparate applications and network functions—across hierarchical boundaries—using standards-based, open API's. With an open network and operational environment, operators can create and deliver services using solutions from multiple vendors as well as their own and partner-developed applications—an increasing number of which will be based on emerging open source software. As openness increases the number of vendors and systems represented in the network, and as new open source solutions evolve, qualification testing is likely to become more complex.

In addition, in the near term, while vendor hardware and software solutions vary considerably in performance, scale, resiliency schemes supported, and the extent to which they offer open solutions, operators will likely incur higher pre-production qualification testing costs. Nevertheless, as standards-based API's, open source software solutions, and general purpose servers become more mature and commercially validated over the mid to long term, qualification testing should become easier and expenses should decrease relative to total cost of operations.

Hardware and Software Distribution

To achieve scale in SDN and NFV networks and operations, service providers will embrace increasingly distributed architectures of the controller and orchestration layers. As a result, resiliency of key software components and the server infrastructures on which they operate—often spanning multiple, geographically dispersed data centers—will become a more complex challenge than it was in a traditional NMS and OSS environment.

Many of the hardware and software resiliency measures discussed in previous sections are unique to SDN and NFV, and represent added capital and operating expenses to the migration effort. Because SDN controllers play an active, central role in traffic management, operators must consider a wide range of new and proven resiliency measures to employ across network and operational layers. For example, mesh networking between SDN controllers and orchestrators represents a degree of resiliency largely unnecessary in the traditional NMS and operational domains of physical architectures. The use of more containerized, microservices-based software architectures, which allow for more distributed and flexible implementation of SDN control and orchestration functions, is another driver of new resiliency requirements that can increase complexity and cost.

Physical Interconnectivity

As deployments of SDN and NFV scale, the number of servers, clusters and data centers—and the optical paths interconnecting

them across an expanding geographic landscape—will grow steadily. With connectivity among nodes and sites increasing and requiring optimized performance and high availability to assure network services, cost and complexity of implementing resiliency measures is likely to increase.

The previous sections outline physical layer considerations and requirements for connectivity between the network and controller layer and controller to orchestration layer. Connectivity between controller clusters and orchestrators may require additional nodes, such as load balancers, to optimize controller performance. However, complexity between these layers is relatively low; there are more complexity and cost considerations from the controller to the underlying network. Fundamentally, there are more potential failure points, more network performance attributes to control (such as latency and loss), and a larger number of constraints posed to the network planner.

SUMMARY AND CONCLUSIONS

SDN and NFV are changing the way we approach network design. They expand the market for new vendors to compete in this arena while promising reduction of costs for service providers in a similar fashion as in the case of data centers. However, these technologies require the inclusion of known resiliency methods to locations and devices in the network that were not designed to support them. Hence, there is an expense to be included in the business case in order to support new demands like

- The implementation of HA clusters in new non-headend or datacenter locations,
- New low-latency redundant management paths or
- Changes of qualification and testing of network equipment, etc.

In addition, the actual separation of functions among multiple devices forces the protection of the integrity of the information being exchanged. This includes cost and complexity increases due to encryption and firewall schemes. This is an extensive topic, not covered by this paper, but still a key expense and complexity item.

As it is always the case with new technology life cycles, the expectation is that as the hardware and software mature, the cost of implementation will decrease over time. However, and in the near term, protection and redundancy could condition the extent to which some service providers can commit to deploy SDN and NFV in portions or even all of their networks.

REFERENCES

1. ETSI GS NFV 001 v1.1.1. “Network Functions Virtualisation Use Cases”. 2013
2. ETSI GS NFV-REL 001 v1.1.1. “Network Functions Virtualisation Resiliency Requirements”. 2015
3. ONF, “OpenFlow Switch Specification V1.3.0”. 2012
4. Software Defined Networks: A Comprehensive Approach, Paul Goransson, Chuck Black June 5, 2014
5. Kretz, F. Ramos, and P. Verissimo. “Towards Secure and Dependable Software-Defined Networks”.
6. Bernstein. “Availability For Network Function Virtualization”. 2015 Spring Technical Forum
7. A Basta, M Hoffman et al.“Applying NFV and SDN to LTE Mobile Core Gateways; The Functions Placement Problem”. 2014
8. ONF. “Openflow-enabled SDN and NFV solution brief”. 2014

Bringing the power of Analytics to improve end-user Quality of Experience

Sangeeta Ramakrishnan, Xiaoqing Zhu, Frank Chan,
Bhanu Krishnamurthy, Cindy Chan, Zheng Lu, Kashyap Kambhatla
Cisco Systems

Abstract

Broadband speeds continue to increase. As operators cope with growing bandwidth demands, expectations amongst subscribers also continue to grow with respect to how well the broadband service performs. Traditionally a broadband system has been judged solely based on how it manages bandwidth. The assumption was that the higher the bandwidth and the lower the packet drops the happier the end consumer would be. While this assumption is fundamentally true, it creates a system that has no clear bounds and therefore it tends to be over-engineered and expensive. Using analytics the "user happiness" can be measured directly and a cost optimized network can be built. The analytics generated can be used for multiple purposes, including troubleshooting, capacity planning, and optimization.

This paper will present the various ways the analytics can be gathered and how it can be used. A primary usecase for this technology is for IP video delivery and troubleshooting any problems associated with its delivery. Operators spend a large amount of OpEx dollars for troubleshooting their networks and services. Analytics are key to being able to quickly triage and identify sources of problems.

Another way the analytics can be used is for capacity planning. Today much of capacity planning is highly dependent on network interface utilization. We will instead present a novel method to go beyond utilization levels and in fact estimate bandwidth demand on the network. Improved capacity planning techniques can help operators spend their Capex dollars in an efficient manner to target network spend on the areas that need it the most, thereby being cost-effective while yet improving subscribers' experience.

Finally we will present ways in which the analytics can be leveraged to optimize the efficiency of the operator's network. The optimizations can range from video-aware cable modem load balancing, to WiFi optimizations that enhance broadband delivery. Further these analytics can be used to optimize IP video Quality of Experience thereby significantly improving the bandwidth efficiency of operators' networks. Results from such optimizations demonstrating up to 40% improvement in stream packing efficiency will also be presented.

Besides outlining the various ways Analytics can be used in operator networks, we will also present a proposal on how this can be realized with a Software Defined Networking (SDN) approach. Given recent technology advances in the industry with respect to Analytics, Big Data, and the advent of Software Defined Networking, this is the right time for the cable industry to adopt Analytics to improve the service offered to subscribers while yet improving the cost effectiveness of offering such services.

INTRODUCTION

According to Cisco Visual Networking Index (VNI) Global IP traffic has increased more than fivefold in the past 5 years, and will increase nearly threefold over the next 5 years. Video as it is dominates the Internet traffic today, but that trend is only going to accelerate further with 80% of Internet consumer traffic expected to be video by 2019 [1]. The vast majority of video being delivered on the Internet uses HTTP Adaptive Streaming (HAS), also known an Adaptive Bit Rate (ABR) video [3].

In the ABR video delivery systems the video is encoded at various bitrates called profiles. Additionally the video is segmented into short fragments each of which is a few seconds long (typically between 2-10 seconds). At every fragment boundary, the ABR client can dynamically select the rate profile to request. The client typically switches to a lower-rate profile, to avoid buffer underflow during network congestion. ABR is widely deployed with Apple HTTP Live Streaming (HLS) [4], Microsoft Smooth Streaming [5], and Adobe HTTP Dynamic Streaming (HDS) being the most popular ABR streaming protocols. More recently Motion Picture Experts Group (MPEG) has issued the Dynamic Adaptive Streaming over HTTP (DASH) [6] specifications standardizing the ABR delivery. However client rate adaptation logic has been left outside the specification to allow innovation by vendors in this area.

ANALYTICS

The advent of Adaptive Bit Rate (ABR) video has changed the traffic characteristics on Service Provider networks. With ABR video, the bandwidth usage on the network adapts up and down to availability and shortage of bandwidth respectively. This makes it harder for operators to estimate end user Quality of Experience (QoE) simply based on BW utilization of interfaces, like it was done in the past. With ABR video, there can be multiple interfaces with comparable network utilization but very different levels of QoE.

Instead of using interface utilization, by examining the level of video oversubscription, a better understanding of end-user QoE can be derived. In our work we have defined video oversubscription as a ratio of the aggregate bandwidth demand for video on an interface to the aggregate bandwidth utilized by video on the interface, with higher numbers of oversubscription implying a worse QoE on such interfaces. In Figure 1 the blue bars show sample interface utilization for 10 different interfaces. The green bars show video oversubscription for each of those same interfaces. As seen below, while interface utilization looks comparable across several interfaces, the video bandwidth oversubscription is quite different amongst them. The QoE on Interfaces 6 and 8 are significantly higher than others and hence subscribers on those interfaces experience much worse QoE than on other interfaces.

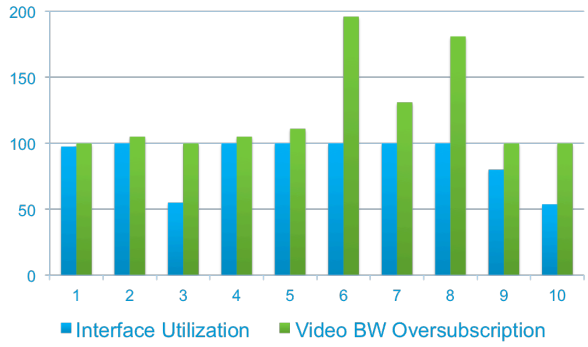


Figure 1 Interface Utilization versus Video Bandwidth Oversubscription

Such an oversubscription metric not only provides an insight into the state of the video delivered on the interface, but also on the overall state of the interface. With vast

majority of traffic on the DOCSIS downstream interfaces being TCP traffic, assuming TCP flows are reasonably sharing bandwidth amongst them, the Video Bandwidth Oversubscription metric not only provides insight into the state of the video flows for which such visibility is available, but also for the remaining TCP flows on those interfaces.

Besides the above-described metric, interfaces to clients can be used to measure buffer stalls, and buffer levels as observed by clients. Aggregating such metrics across clients on a single CMTS or a single CMTS interface can be useful in gauging the health of those interfaces.

ARCHITECTURE

We propose a Software Defined Architecture (SDN) for gathering these analytics as shown in Figure 2. SDN is defined as an architecture that is characterized by the separation of Control and Data planes, and uses an open standard protocol to communicate between them. In an SDN architecture a controller provides network layer abstraction to the applications above it. SDN promises flexibility and rapid innovation by virtue of the fact that Applications can be built almost independent of the underlying dataplane hardware, and need not support network-element-specific protocols, such as COPS, IPDR etc. More information on SDN can be found in [6].

The Video QoE Application collects information from various elements of the end-to-end architecture, from streamers, clients, and various network elements, including the CMTS. The Video QoE application combines information from both video and network domains to provide valuable analytics, that can be used in multiple ways as described in the following sections.

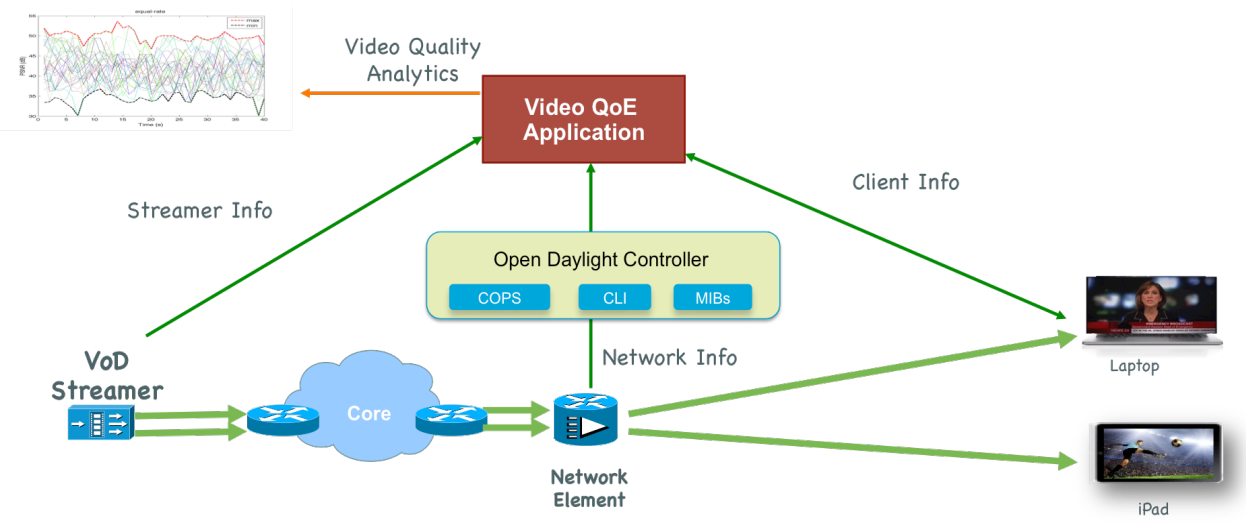


Figure 2 SDN based architecture for Video QoE Analytics

CAPACITY PLANNING

Operators can use the above-described metrics such as oversubscription for capacity planning purposes. It provides better insights into the true state of the network and can thus help them target network upgrades in areas where upgrades are truly needed versus

spreading it out evenly. This enables them to delay capital expenditure and also focus upgrades in a timely manner in the most problematic areas.

With plans afoot to move to virtualization, these kinds of metrics can also be used to spin up additional resources for certain Service Groups over others. Such methods can be used to deploy "virtual capacity" only on a need basis, thereby applying these spare resources selectively to Service Groups where it is needed.

TROUBLESHOOTING

Operators today spend significant amount of OpEx dollars on troubleshooting the network. When customers complain about the video quality they are experiencing, having QoE metrics per user, per network segment etc. can help troubleshoot and root-cause the problem more easily.

For example if only 1% of users are experiencing poor QoE, it may not be clear how big a problem it is and how much resources to expend to troubleshoot the problem. However if the poor QoE users' network path is examined, it may become very obvious that vast majority of poor QoE issues are on a single CMTS or on a single interface etc. This can provide an excellent starting point for operators to dig deeper to troubleshoot and fix the problem.

Although the above-described example limited itself to the CMTS as the only network element, one can see how this can be extended across multiple network elements. In fact in the age of big data, all types of information about flows can be logged, and machine-learning techniques can be applied to identify problem areas. This should help rapidly narrow down problem areas when troubleshooting large scale systems, which may have multiple problem areas, co-existence of which may make it harder to root cause the problem in traditional/manual ways.

NETWORK OPTIMIZATION

Having a better understanding of the QoE of users can help operators improve the QoE of users on their networks. In fact there are two sides to such optimizations – QoE can be improved while keeping network utilization comparable, or QoE can be maintained while reducing network resources required to meet the QoE needs. More realistically as demands on operators' networks grow, they can pack more users and streams on to the same network while maintaining QoE. These types of optimizations can be achieved in a number of ways, and a couple of such usecases are described below.

Video QoE-aware Cable Modem Load Balancing

DOCSIS specifies CMTS to CM interfaces that enable cable modem load balancing across interfaces. It is performed in a couple of ways in today's networks. The first being static cable modem load balancing, where modems are assigned to channels as they come online. In static load balancing typically the counts of cable modems on interfaces are balanced. The second type of load balancing is called dynamic cable modem load balancing and in this approach cable modems are moved across interfaces after they are on-line. Although DOCSIS specifies the messages the CMTS must use to communicate load balancing instructions to cable modems, the standards do not specify how the CMTSs make these load balancing decisions. Typically CMTSs use the interface utilization along with cable modem counts to balance modems across interfaces. With ABR video growing or shrinking to fit the available bandwidth, interface utilization has

diminished in value as a metric to use in load balancing decisions.

Instead a video-aware cable modem balancing application could use the video oversubscription metric to determine which interfaces are more/less oversubscribed to determine in what direction to move the cable modems. Applications can also use the same metrics to evaluate whether the load balancing decisions made were effective in improving QoE or not.

As shown in Figure 3, with the proposed SDN architecture, the Video QoE Application can publish a QoE API, which the Cable Modem Load Balancing Application can use to adjust its decisions to take video QoE into account. This in fact shows the advantage of building applications in a SDN framework and enabling value-added functions across such Applications by leveraging the API exposed by Applications.

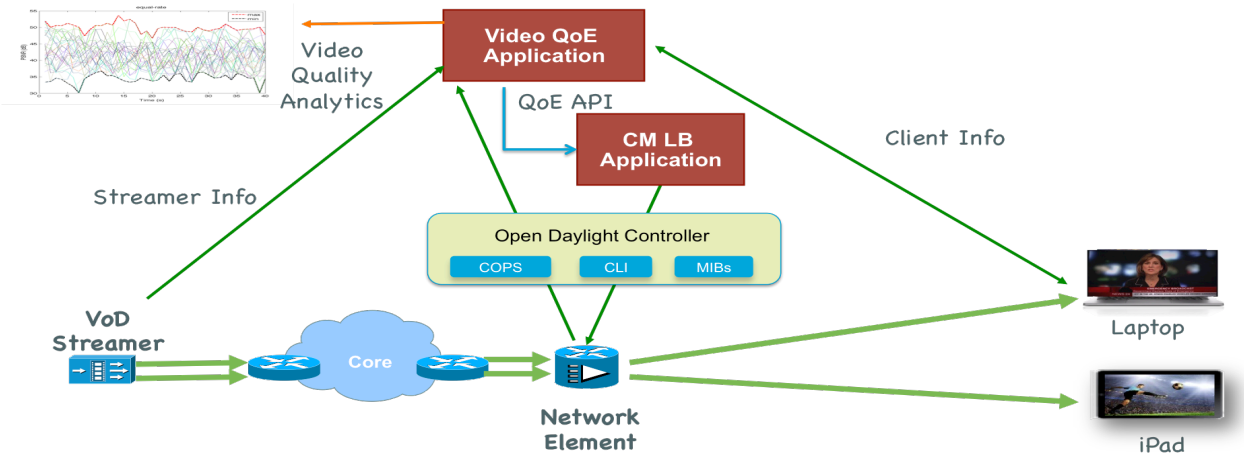


Figure 3 Video QoE Aware CM Load Balancing

Video QoE-aware WiFi Radio Resource Management (RRM)

With increasing number of WiFi devices in the home, and operators' deployment of residential gateways with built-in WiFi, it is becoming more important to manage the WiFi radio resources efficiently to improve subscriber user-experience. As described in [8] WiFi RRM works by collecting various types of information from Wi-Fi Access Points (APs) that it manages and analyses this information and determines whether to apply any changes to the operating parameters of the APs that it manages. Some of the parameters that it may change include Transmit Power, Channel assignment, Channel bandwidth, etc.

The RRM algorithms have generally been envisioned to be fairly static – changing only once a day or few days etc. What we propose in this paper is an enhancement to RRM by which the video QoE information is taken into account in the decisions made by the RRM algorithms. This will likely necessitate more frequent RRM updates, but this can still be done in the order of minutes (not milliseconds). Conceptually this can be viewed as the equivalent of dynamic load balancing as described in an earlier section, except applied to WiFi radio resources. This kind of video-aware RRM can improve video QoE for subscribers.

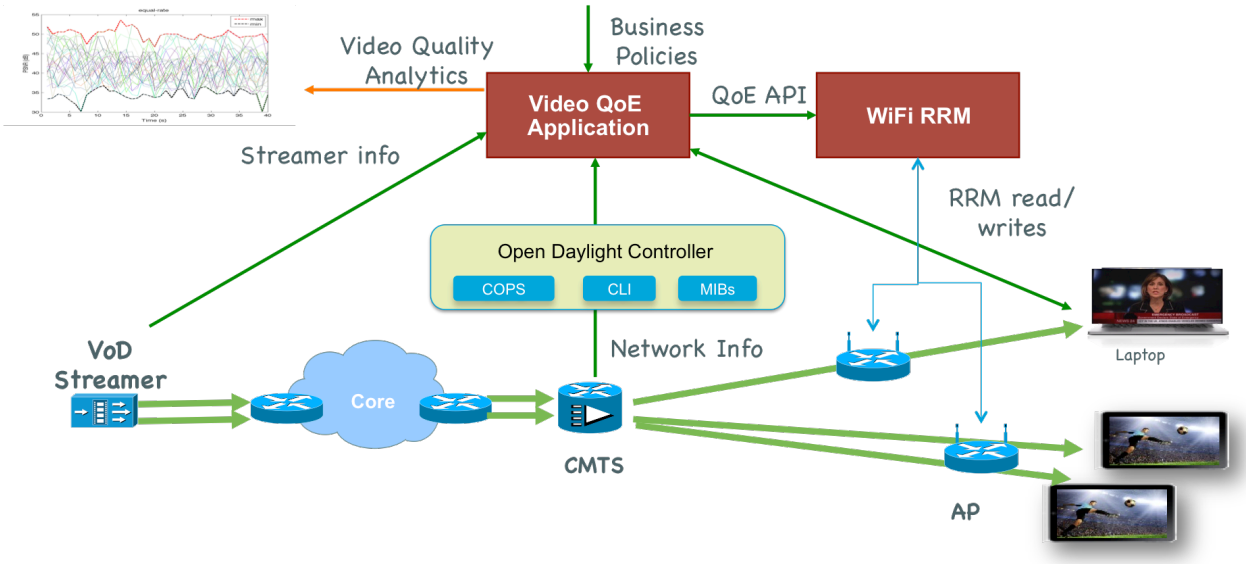


Figure 4 Video QoE Aware WiFi Radio Resource Management

Video QoE optimization over DOCSIS

The proposed SDN architecture can be extended to further optimize the video quality by ensuring fairness in how bandwidth is shared across the DOCSIS interface. In existing HTTP Adaptive Streaming architectures, each client operates completely independently of each other and therefore in a greedy fashion. There is no attempt made to share bandwidth in a fair manner across the competing clients. In the recent past there have been several attempts [9], [10], made in developing client rate adaptation algorithms that share bandwidth fairly across competing clients. In all of the referenced work, fairness across clients is defined as equal bandwidth share for each client. An additional goal of these efforts has been to minimize/eliminate rate profile oscillations of the clients over time.

To compare the behavior of clients with and without our QoE application, we tested 16 clients competing for bandwidth on an 80Mbps link. The clients we used were VLC (an open source client) [11] with a simple rate adaptation logic, and a more sophisticated client with superior rate adaptation logic called Probe-And-Adapt (PANDA) [9]. Figure 5 shows the rate profile oscillations of VLC clients. Figure 6 shows the same setup except with the PANDA clients. Clearly the PANDA rate adaptation algorithm is superior to VLC's as seen by the reduced rate oscillations. Next we repeated the experiments with the Video QoE application programming equal rate allocation on the network element, thereby ensuring that each of the flows gets equal bandwidth share on the congested link. The rate profile selections of the VLC client under this scenario are shown in Figure 7.

The vertical dashed line indicates the time when the network programming was applied. The contrast is stark in that after network programming is applied, the rate oscillations are almost completely gone. From a rate oscillation and bandwidth fairness perspective this approach far outperforms the scenario with even a highly intelligent client such as PANDA as shown in Figure 6.

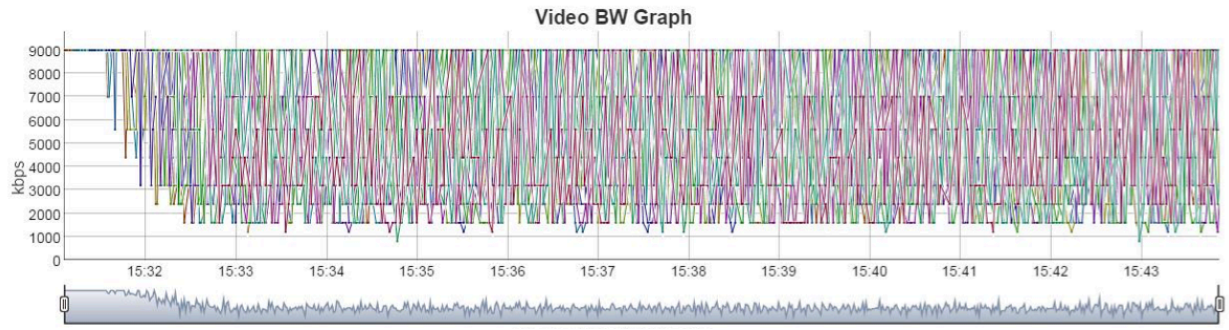


Figure 5 Rate profile selections of VLC clients - baseline

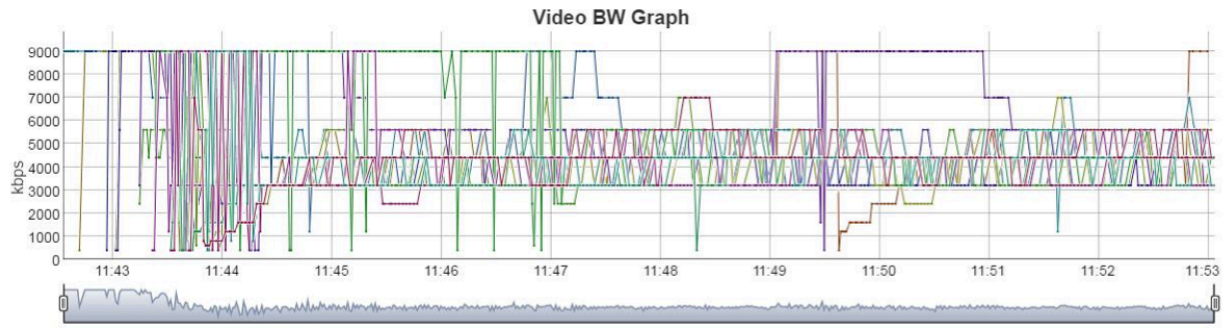


Figure 6 Rate profile selections of PANDA clients – baseline

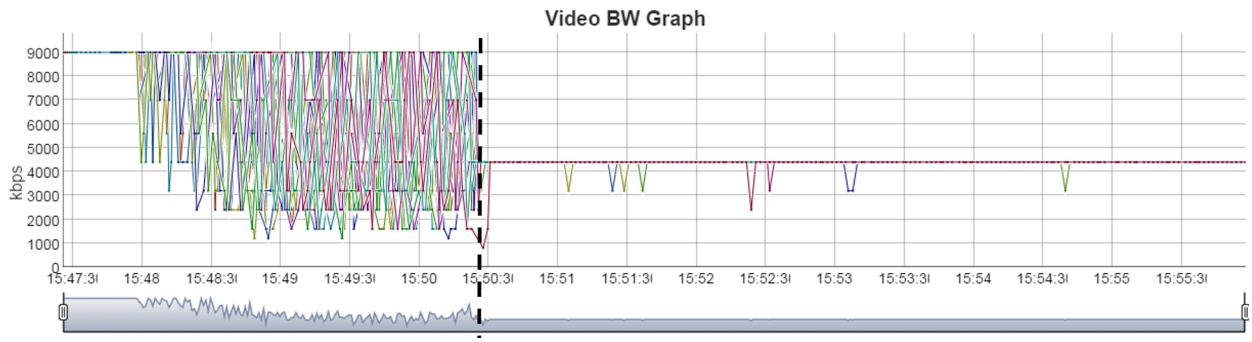


Figure 7 Rate profile selections of VLC clients – with equal rate allocation in network

It is interesting to note that if indeed equal share of bandwidth is an important goal, it is very easily achievable via the proposed SDN architecture. The Video QoE Application can program the CMTS (via PCMM) to allocate equal bandwidth share for each of the streams. This causes the clients to adapt to the perceived network bandwidth. In fact such a straightforward mechanism not only achieves equal bandwidth share, but also helps eliminate rate oscillations in clients' profile selections due to the steady nature of bandwidth available to them. Moreover this architecture can be leveraged to apply business policies that achieve a better fairness model by allocating bandwidth based on content/device type etc.

Beyond equal share of bandwidth, fairness could be defined as comparable perceptual video quality across clients sharing a congested link. Such a goal can again be achieved by our above-defined architecture by re-programming the network at regular

intervals to ensure fairness across clients. That is, instead of bandwidth equally across clients or class of clients, bandwidth can be allocated in a manner to equalize quality across flows. Anyone familiar with encoding technologies will be familiar with the fact that the bits required to encode a video stream at a certain quality will vary depending on the complexity of the video scene being encoded. Or put another way two different videos may achieve very different video qualities even when encoded by the same encoder at the same target bitrate. This property is the basis for the widespread use of variable bit rate encoding in broadcast TV today.

With the migration to Adaptive Bit Rate video, however generally the industry has fallen back to Constant Bit Rate encoding. This loss of efficiency was not a major problem as long as ABR video constituted a small portion of the overall network traffic. Given the popularity of ABR video and its continued growth on DOCSIS networks with the advent of IP video, it is imperative that ABR video operates in as bandwidth efficient a manner as possible.

There have been recent attempts to move in this direction, by large video streaming providers such as Netflix. They describe a per-title video encode optimization method in [9]. This approach is an explicit recognition of the fact that different bitrates are required to achieve the same quality on different titles (content). Hence they propose using different rate profiles for each title. So very simple content (example animated titles) may use lower bitrates compared to more complex titles. While this approach is an improvement over the existing fixed bitrate ladder used in most ABR delivery systems, it suffers from two problems. The first is that the encode may be optimized to that title, but there is nothing in the network that ensures that bandwidth is apportioned fairly per title. Put another way in the face of congestion both flows (complex and simple titles) will settle at the same bitrate, resulting in a much better visual quality for the low complexity title and poorer visual quality for the high complexity title. Our proposed SDN architecture can be leveraged to solve this problem by ensuring the network allocates bandwidth proportional to the title complexity. For example when uncongested, the Video QoE App can allocate the bandwidth for that flow corresponding to the highest bitrate profile available for that title. As the network gets congested (due to say start of busy evening hours) the Video QoE Application can move each of the flows down to the second highest bitrate profile for the title that is being viewed.

The content diversity property is exhibited in the rate-quality tradeoffs depicted in Figure 8 below, where each line depicts the quality of a video encoded at different bitrates. The fact that different lines exhibit different slope means that some videos achieve high quality rapidly at lower rates, while others struggle to reach comparable qualities even at the highest bitrates. Generally speaking the lowest graphs correspond to the most complex videos that are harder to compress and the highest graphs correspond to the simplest videos that are the easiest to compress.

The quality metric we have used is SVQ (Stream Video Quality) a Cisco proprietary non-reference video quality metric, which is lightweight to calculate and has demonstrated via internal investigations high correlations to subjective mean-opinion-scores. The SVQ score ranges from 1 to 10, with 10 being the highest perceived quality. Typically, a score below 6 corresponds to “bad” visual quality whereas a score above “9” is considered “very good”. Irrespective of which quality metric is used, it is well accepted that rate quality trade-off curves do indeed vary from video to video.

The second problem with the per-title encode optimization is that it doesn’t account for the fact that video complexity not only

changes from one title to another, but also from minute to minute within the same title. This is exhibited in the rate-quality trade-off curves in Figure 9, which shows similar patterns as the one above for content diversity, except here the quality depicted is per fragment. So even within 16 consecutive fragments of video (each of 2 seconds) the quality achieved at each bitrate varies significantly.

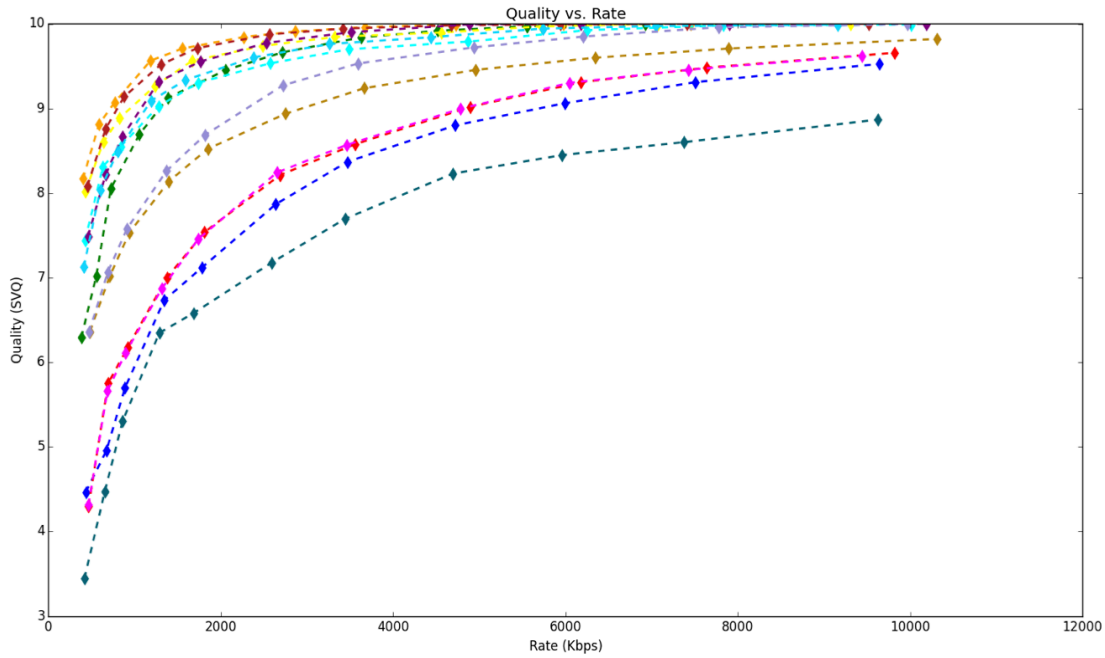


Figure 8 Rate quality trade-off curves for different videos

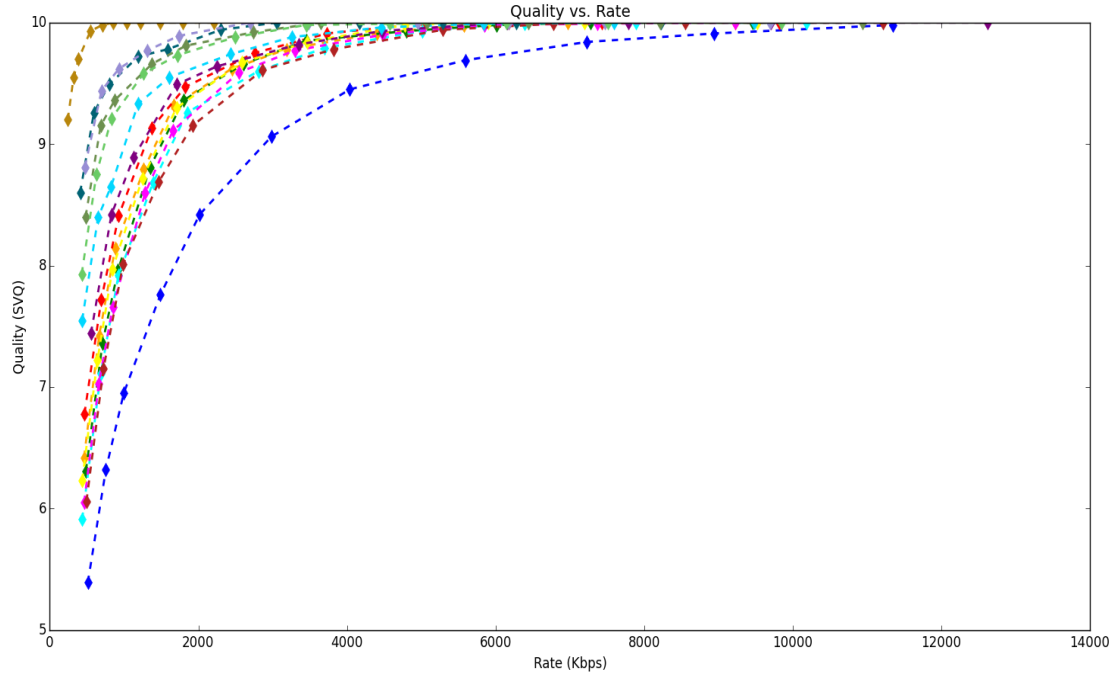


Figure 9 Rate quality trade-off curves for different fragments within the same video

The Video QoE Application can also solve this second problem of content complexity variation over time, within the architecture depicted in Figure 2. By making the Video QoE Application aware of the varying complexity of the video streams via metadata, it can now account for the changing complexity of the content over time, besides the content complexity across the different flows. By taking into account both of these diversities the Application can periodically reprogram the network to account for these differences. ABR clients are already built to adapt to varying network bandwidth, so they will adapt to the network bandwidth made available to it. If it is feasible to modify clients they can in fact be made aware of the decisions made by the Video QoE Application thereby vastly simplifying its bandwidth estimation and associated rate adaptation logic.

We conducted tests with such an optimization method applied in a testbed with a link bandwidth of 100Mbps. We varied the number of clients competing for bandwidth from 15 to 35 clients at any given time. The Video QOE Application was used to collect the minimum and maximum quality across all the clients for each run of the test. We ran 3 flavors of algorithms, where one is baseline, where the Application does not program the network at all, and simply collects and reports analytics. The second case was the "equal-rate" case where the Application programs equal bandwidth for each of the flows on the network element. Finally the third case was where we turned on our optimization algorithm that not only reprogrammed the network but also communicated with the clients to influence their rate adaptation decisions. Results from the above tests are shown in Figure 10. Our optimization method yields the highest minimum video quality compared to other methods at all number of clients. It is not surprising that our method also achieves the lowest maximum quality across all clients. This is in fact fair, because as an operator, the goal is indeed to keep as many subscribers satisfied as possible, which is achieved by ensuring all clients stay as high a quality level as possible without dropping below a threshold. In fact from the graphs it can be observed that, suppose an operator wanted to maintain a minimum quality of 8, they could only pack at most 15 VLC clients on their network, with equal-rate allocation method of Video QoE Application they could pack about ~25 clients on their network, and with the outlined optimization approach they could pack 35 clients on their network. The optimization approach improves stream packing efficiency about 40% over an equal-rate approach. Therefore a significant improvement in stream packing efficiency can be achieved while maintaining video quality. Alternately operators can leverage such technology to improve QoE across their subscriber base.

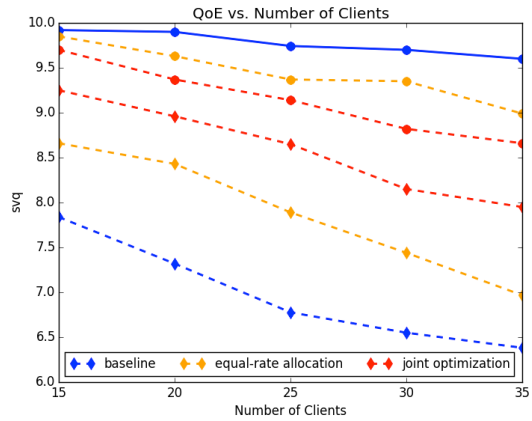


Figure 10 Minimum and maximum quality with varying number of clients

SUMMARY

We have presented methods by which operators can gather video QoE analytics leveraging an SDN architecture. The value of these analytics in capacity planning and troubleshooting has been outlined. Finally optimization methods that leverage the

analytics to significantly improve network efficiency have been presented. All of the above capabilities have been built on top of an extensible SDN architecture.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contributions of various colleagues at Cisco Systems including Joel Schoenblum, Gareth Bowen, Tankut Akgul and Ziv Nuss.

REFERENCES

1. Cisco Visual Networking Index: Forecast and Methodology, 2014-2019 White Paper http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html
2. "HFC capacity planning for IP Video", Sangeeta Ramakrishnan, Cisco Systems, SCTE 2011.
3. History of Move Networks, Available Online: http://www.movenetworks.com/history.html
4. Apple Inc., "HTTP Live Streaming Overview," https://developer.apple.com/library/mac/documentation/networkinginternet/conceptual/streamingmediaguide/Introduction/Introduction.html
5. A. Zambelli, "IIS Smooth Streaming Technical Overview," http://www.microsoft.com.
6. MPEG, "ISO/IEC 23009-1:2012 Information technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats," 2012. https://www.opennetworking.org/sdn-resources/onf-specifications/openflow
7. Software Defined Networking, Open Networking Foundation, https://www.opennetworking.org/sdn-resources/sdn-definition
8. CableLabs Technical Report, "Wi-Fi Radio Resource Management (RRM)/Self Organizing Networks (SON) Technical Report", http://www.cablelabs.com/wp-content/uploads/specdocs/WR-TR-RRM-SON-V01-1409261.pdf
9. Z. Li, X. Zhu, J. Gahm, R. Pan, H. Hu, A. C. Begen, and D. Oran, "Probe and Adapt: Rate Adaptation for HTTP Video Streaming at Scale," IEEE Journal on Selected Areas in Communications, vol. 32, no. 4, pp. 719–733, Aug. 2014.
10. S. Akhshabi, S. Narayanaswamy, A. C. Begen, and C. Dovrolis, "An Experimental Evaluation of Rate-Adaptive Video Players over HTTP," EURASIP Journal on Signal Processing and Image Communications, vol. 27, no. 4, pp. 157–168, May 2011
11. VLC Media Player, http://www.videolan.org/
12. Netflix Blog, " Per-Title Encode Optimization", December 2015, http://techblog.netflix.com/2015/12/per-title-encode-optimization.html

Cable Access Redundancy: Opportunities in Virtual Deployments

Amit Singh, Principal Engineer CTAO Cable Access, amsingh@cisco.com
Cisco Systems Inc.

Abstract

The Remote PHY architecture disaggregates the PHY from the physical CCAP (Converged Cable Access Platform). This enables CCAP to be virtualized. Virtual Cable Access Platform (vCAP) is the virtualized instance of CCAP (Converged Cable Access Platform) core software. vCAP has two disaggregated software components, vDOCSIS and vEQAM. These provide DOCSIS and MPEG video services respectively. Together with a RPD (remote PHY device) vCAP provides CCAP services. vCAP is realized via NFV (Network Function Virtualization) virtual machines (VM). vCAP will be deployed in data centers.

One of the key functions of CCAP is high availability (HA). vCAP presents unique opportunities to design and implement redundancy and high availability that are not possible with custom hardware based CCAP.

This paper and accompanying presentation examine the unique aspects of redundancy in a virtualized environment, discuss their pros and cons and weigh them against the relative cost to deploy and maintain them.

INTRODUCTION

Traditional custom hardware implementations of CCAP provide redundancy and high availability via N+1 redundancy for line cards (1 standby line card backing up several active line cards) and 1:1 for supervisor cards.
The typical subscriber scaling of CCAP platforms is about tens of thousands of per instance and thousands of subscribers per line card.

vCAP breaks this model. There are no redundant line cards or supervisors. The instance of a vCAP needs to be redundant to ensure high availability. The disaggregated software centric data center environment provides unique infrastructure, such as, databases and redundant computing environments. These can be leveraged to offer new models of redundancy. For example, databases can be leveraged to store active modem state and redundant compute can be purposed for running back up instances of vCAP functions.

The allocation of storage and compute resources to redundancy is tunable per deployed instance and can be changed over time or altered via application of redundancy policy. The degree to which data center resources are allocated to redundancy services affects both the capital and operating cost.

With reduced scale of subscribers (in the order of tens or few hundreds) per RPD the new availability models become relevant.

The unique flexibility in resource allocation to redundancy results in multiple availability models which offer varying degrees of availability from almost none to availability levels on par or even exceeding as those offered by physical CCAP devices deployed today.

The remainder of the paper discusses these in detail. It is constructed with the following sections. The first section introduces the components of vCAP. The second section introduces and compares new virtual redundancy models. A possible scheme for disaster protection in case the data center is

lost is discussed followed by the mechanics of switchover and revert back.

vCAP COMPONENTS

A vCAP is composed of several components:

1. DOCSIS and eQAM VMs to realize the vCAP instance
2. Remote PHY Device
3. Network
4. Orchestrator
5. Servers

These components are physically distributed, unlike in a traditional CCAP where they are all contained in a single device. The virtual environment is dynamic as functionality could be moved from server to server, rack to rack or even DC to DC. A highly available solution has unique challenges but could also provide unique value over HA solutions in physical systems.

TIERED vCAP AVAILABILITY

If one views a vCAP VM HA from the perspective of the failure domain it is possible to construct multiple tiers of availability functionality. The deployment of these tiers of high availability depends on the number of modems and the type of service levels and service level agreements (SLAs) provided to the end user. The economics of the solution increases as the system becomes more and more available. Further, it is possible to deploy multiple tiers of HA in the network at various service nodes.

The tiers of high availability could be provisioned in a recovery policy with diminishing levels of service in case the primary recovery policies fail. The level of failure recovery could vary with the time of day or annual calendar. Data center alternatives such as vMotion could also be incorporated in the over all high availability policy. vMotion is not discussed here as information on it is available on the Internet.

vCAP high availability tiers are listed below. It is not necessary that the DOCSIS and eQAM VM comprising the vCAP instance deploy the same identical high availability tier.

1. Stateless Non-Redundant VM

This is the lowest tier of redundancy. When a VM goes down a new one is spawned on the same server. Modem state is not saved or cached. All modems will need to re-register. It is beneficial if the VM hosts a small number of modems and the deployed services can withstand the outage time of a virtual machine coming online.

The advantages are, this is the simplest redundancy policy to orchestrate. It is the most flexible and most economical redundancy scheme.

The time taken for the system to recover is the time to stand up the VM, for it to establish communication with the RPD, synchronize configuration with the RPD, and for modems to register.

Synchronizing configuration could be very quick if the RPD and vCAP maintain a checksum of the configuration file. A possible checksum algorithm could be a SHA512.

The RPD and vCAP just need to quickly compare the SHA512 checksum values to validate that the configuration file is the same on either side.

2. Stateless Redundant VM without Pairing

This is the next higher tier of service, where a pool of standby VMs is spawned a priori. The number of standby VMs is a fraction of the primary functioning VMs. The ratio of standby VMs to primary VMs is set by the operator and could change dynamically

depending on the time of day, annual calendar or the current failure rate.

One of the standby VMs takes over when an active VM fails. It will be a little quicker than the previous case as the VM is already up and running. The VM will still need to establish connections with the RPD, synchronize configuration with the RPD and modems will need to re-register.

3. Stateless Redundant VM with Pairing

In this case, a standby VM is spawned for every active VM and pairing information is maintained. Backup VMs could be over subscribed on backup servers. The degree to which backup servers are over subscribed is tunable via policies based on deployment geography, service tier, time of day and annual calendar.
When the primary VM fails the backup VM takes over. The backup VM will pre-establish redundant tunnels with the RPD.
This scenario is quicker than the previous ones, as the backup VM is already up and running, the tunnels are established and the RPD configuration is also synchronized a priori. The time to recover service is the time it takes for modems to re-register.
When the active VM re-spawns a revert back to the active server will need to be affected to recover over subscribed backup server resources if backup VMs are over subscribed on a server.

4. State-full Non-Redundant VM

In this case, all the state of the VM is stored outside the VM in a database. When the VM goes down, it is re-spawned. The VM reloads its state from the database, re-establishes tunnels with the RPD, does configuration synchronization checks with the RPD and re-establishes service to modems. Some modems might need to re-register depending on the time it takes to detect a failure and spawn a VM.
This solution will require access to high-speed databases for storing VM state. The database could itself be replicated to alternate data centers in case it might be desirable to protect against data center outages. This scenario is described in detail in the following sections.
Here server resources are economized as there are no backup servers deployed, however storage for databases will be needed. The database itself would need to be replicated so it is highly available otherwise the scenario mimics the first scenario where the VM recovers statelessly. The time to recover services will be higher.

5. State-full Redundant VM

This is the highest level of availability. A backup VM is spawned for each active VM. The active VM constantly updates its state in an off VM database. The backup VM synchronizes active VM state from the database. The backup VM establishes tunnels with the RPD a priori. Service outage time is minimal and in the same order as that of purpose built CMTS hardware.
Additionally, the database could be replicated to databases in other data centers to protect against data center outages. Further, it is possible to host multiple backup VMs in the same over subscribed backup server, where only a few of them are expected to become active simultaneously. This helps achieve better high availability economics of the entire solution. The number of backup VMs spawned on the backup server can be controlled via policy such as being tunable by the operator dynamically based on time of day and annual calendar. Figure 1 shows the arrangement of active VMs, backup VMs & the databases.

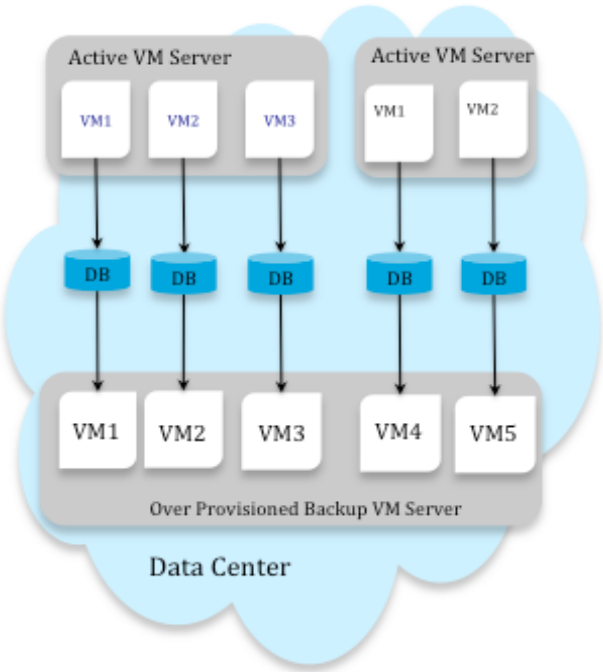


Figure 1: Active & Backup VM Instances

The backup instance does not take much server resources and therefore many virtual standby instances can be hosted on a fraction of the servers used for the active VMs. N+1 can be emulated by creating N instances of vCAP that are on dedicated servers all backed up by a single server. The key statistical assumption is the same as with a traditional CMTS: that N is small enough so that its very unlikely that the backup VM server will have more VMs active than it can support simultaneously.

Further, N is a customer tunable knob, depending on how frequently they experience failures on the vCAP VM. They can also tune N dynamically based on a calendar, for example during important events (such as the Super Bowl) N could be low, to provide higher fault resilience at the cost of more back up servers.

For business services N could be lower than for residential services, thereby providing a higher service level guarantee to business customers.

This flexible and tunable nature of the redundancy mechanism is not offered in the real hardware product. It would require very significant re-architecture of the existing product to offer this level of flexibility and it would cost more in terms of sunk costs of deploying the solution.

If the scale of the VM is small in terms of the number of subscribers it supports a new load of software could be deployed sooner on a limited number of sites with availability cranked up, so if it fails the backup can quickly take over.

The vCAP approach can also simplify the software. Instead of managing N+1 redundancy we mange a collection of 1:1 redundant instances which are over subscribed on the back up server to derive scalable value.

The cost of deploying this flavor of high availability is the highest compared to the rest as it needs backup servers to be deployed and needs databases to manage VM state.

vCAP AVAILABILITY BENEFITS

The other key benefits to arranging HA in this manner are:

1. **Database:** State is stored in an external database. The data center infrastructure inherently provides redundancy of the database. This database redundancy infrastructure can be leveraged as is. This allows for a singe vCAP to have multiple backups. For example, one can keep on backup instance locally to protect from software crashes and at the same time keep another instance in a different part of the data center to protect against hardware failures or at remote data center to cover for the entire data center. Though using a database to store state is not unique, the usage in this case is unique for a couple of reasons, the data base is a transitional

store between VMs that are all running hot and synchronous, with one VM acting as the source of data and the remaining sinks. Also, the database is live and dynamic with real time information such as modem states, service flows, voice sessions, IP video over DOCSIS and subscribers. It can be used to mine and analyze plant and service information.

2. **Backup Tunnels:** Backup tunnels are pre-provisioned so that paths do not have to be calculated on the fly. Intervening switches and networks do not need to be reprogrammed or traffic engineered.

3. **Disaster Recovery:** The approach allows scalable disaster recovery, where the same database is replicated to an alternate data center. Disaster recovery VMs are deployed at even higher over provisioning levels for cost effectiveness. This scenario is depicted in Figure 2

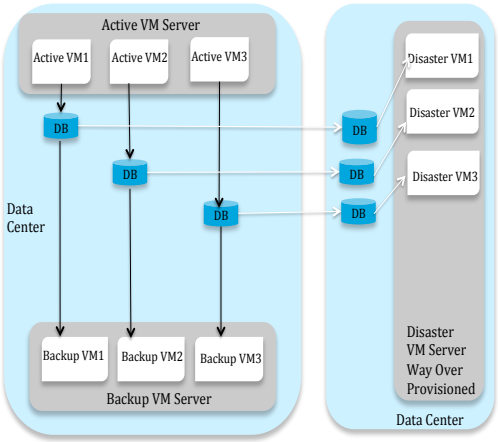


Figure 2: Disaster Recovery via Availability

It is not expected that disaster recovery would switch live modems. RPDs will need to be instructed to connect to VMs in the alternate data centers. All modems will re-register. The network between the disaster recovery data center and the RPDs may not be optimized or traffic engineered to provide an equivalent experience to the primary data center. The service levels of each one of these components are choices for the services provider. They can be tuned via policy, which may include geography or end user service level agreements. Today, there is no cost effective approach to disaster recovery deployed in hubs and head ends with physical equipment.

VM SWITCHOVER MECHANICS

The following sections discuss the mechanics of performing a switchover and revert back in a virtual environment.

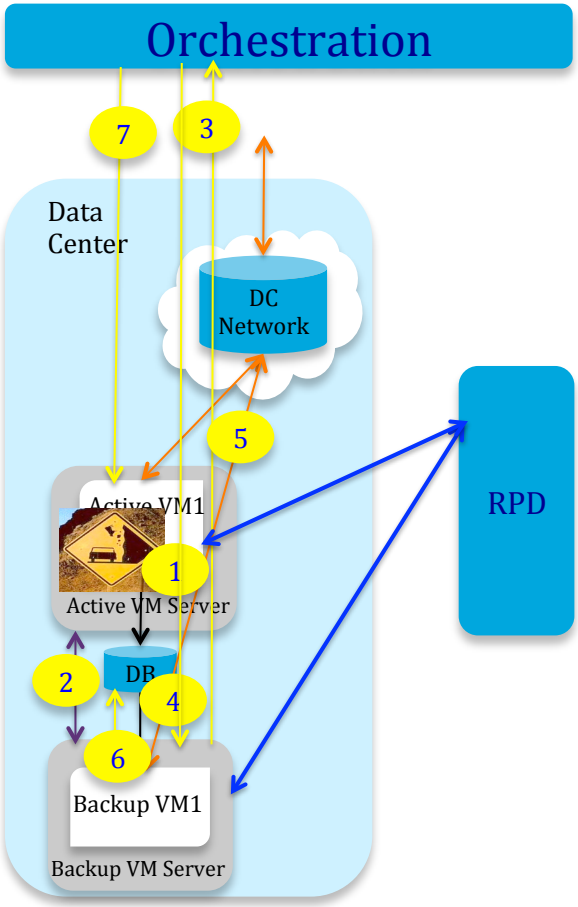


Figure 3: VM Switchover

Figure 3 depicts the typical processing steps for switchover.

1. Active VM1 has an accident.

2. Keep alives to backup VM & orchestration fail; RPD tunnels go down.

3. Orchestration notices keep alive failure OR backup VM notifies orchestration VM1 went down.

4. Orchestration notifies backup VM to take over.

5. Orchestration re-routes backhaul traffic to backup VM.

6. Backup VM becomes DB master & keeps the database updated.

7. Orchestration cleans up active VM1 & spawns new VM1 (waits for 1588 Sync); backup VM1 connects to DB to sync state back.

Figure 4 depicts the typical processing steps for revert back. As the backup server is over provisioned revert back is necessary.

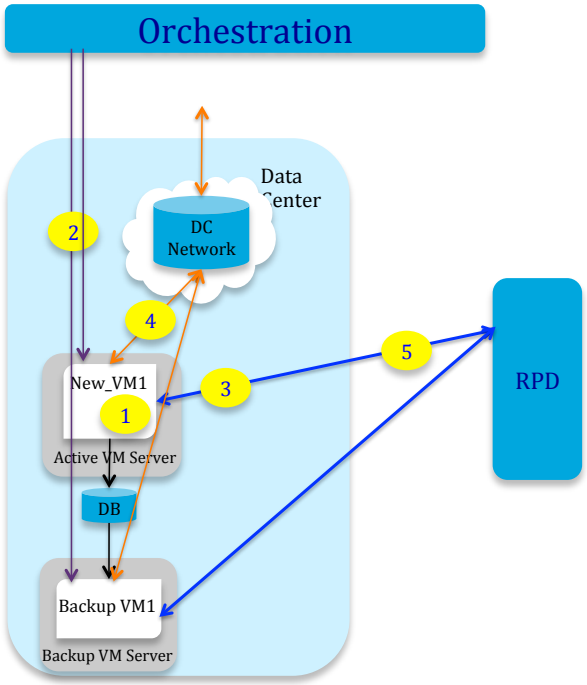


Figure 4: VM Revert back

1. New VM1 initializes & waits for hold over time (~5 minutes). The tunnels/keep alives are up and running. The configuration on New VM1 is applied and synchronized with RPD configuration. NewVM1 connected to the database of the backup VM and is continuously synchronizing real time state from the database.
2. Orchestration notifies NewVM1 & Backup VM NewVM1 is going active.

3. NewVM1 Notifies RPD Its Going Active (RPD tunnel switch is quicker than orchestration re-route). New VM1 becomes the database master and updates state to the database.

4. Orchestration re-routes backhaul traffic to NewVM1. All downstream traffic now goes through NewVM1.

5. RPD starts sending US traffic to NewVM1. The backup VM reverts to backup mode where it synchronizes state from the database.

CONCLUSIONS

The remote PHY architecture enables virtualization of CCAP functions. Virtualization of CCAP functions via a disaggregated vCAP in turn offers new tiers of service, end user SLAs, service availability and tunability via policy. It offers new possibilities in the economics of deployment, service monetization and allows rapid deployment of new services.

REFERENCES

US Docket 03919.0897 (997928): Virtual Cable Modem Termination System Redundancy in a Cable Modem Network Environment, *John Chapman, Jan Medved, Alon Bernstein, Amit Singh, Sep 2015*.

CONSIDERATIONS FOR A SERVICE DELIVERY PLATFORM FOR OPTIMIZING THE DEPLOYMENT OF MSO BUSINESS SERVICES

RAM SRIDHARAN, STEVE DAVIDSON
NOKIA

Abstract

The business environment for MSO's commercial customers has changed; it's a disruptive world where yesterdays technology is hindering the need to accelerate towards cloud-based IT. This is particularly evident in the area of launching new business services. Traditionally, business services are tightly connected to a dedicated network infrastructure and based on a set functionality. This has forced businesses into the complex world of networking and service delivery platforms in order to achieve the network and service capabilities that match their specific business needs. Many enterprise branch locations (as well as Small and Medium Businesses - SMBs) lack the on-site technical support necessary to manage network services at these locations.

This paper examines the industry architecture concepts and delivery platform requirements for the rapid deployment of value-added cable business services that MSOs can provide to new and existing customers. The paper covers NFV (Network Function Virtualization) and SDN (Software Defined Network) -related topics focusing on tools and automation techniques both in a data center and branch office environment that would provide MSO business customers with the power to self manage and deploy their own network services.

INTRODUCTION

Cable multiple system operators (MSOs) have an unprecedented opportunity to accelerate commercial service revenues far beyond current levels by providing businesses the support they need to maximize the benefits of cloud technology.

The worldwide shift by businesses of every size and description towards greater reliance on private and public clouds for access to mission-critical applications has generated demand for lower-cost, more flexible ways to extend utilization of those virtualized resources across multiple locations and to workforces on the go. This in turn has accelerated network service providers adoption of virtualization technology to create software-defined wide area networks (SD-WANs) that greatly expand the flexibility and lower the costs of IP-based connectivity. At the same time, service providers are deploying SDN concepts within business branch offices and datacenters and embracing Network Function Virtualization (NFV) to rapidly and efficiently deploy business services across their customer base.

MSOs can take the lead in satisfying this demand by leveraging a cloud-friendly, multi-layered virtual network architecture to create a connectivity and value-added service model that far outperforms traditional service offerings. Given the growing importance of VPNs in delivering cloud-based IT applications to businesses across all their locations and to employees on the go and a platform to launch value added services, MSOs have a significant opportunity to satisfy market demand through use of SD-WAN technology to deliver software-defined VPN (SD-VPN) services to the SMB and enterprise markets.

By providing customers agile and secured access to cloud-hosted applications, MSOs can make significant inroads into a market

currently dominated by rigid VPN services that are no longer suited to the needs of SMBs or, for that matter, to new cloud-driven needs of larger enterprises. Not only will implementation of SD-VPNs and cloud infrastructure allow cable operators to provide a more flexible and scalable way for businesses to exploit the cloud; it will facilitate operators' expansion into the lucrative managed services market with the bundling of cloud-hosted applications as value-added enhancements to their connectivity services.

COST CONTAINMENT & CLOUD VIRTUALIZATION

Several recent studies underscore the impact that demand for new types of services will have on total spending for network services in the years ahead. Generally speaking, they all reflect the heavy priority businesses are placing on cost containment.

According to an annual survey of IT professionals in the U.S. conducted by Computer World, even as IT budgets are increasing, cost containment ranks as the top business priority, with 53% of respondents citing this as the top priority compared to 35% who cited growing revenue as the top priority. Other top priority items also directly relate to cost containment, including optimizing and automating business processes (47%) and accelerating business process and agility (38%). Looking at the top categories targeted for reduced spending, the survey ranked hardware spending first, followed by legacy system modernization, on-premises data center consolidation and optimization and on-premises software. All of these cost-cutting priorities point to growing reliance on cloud-friendly networks that can maximize companies' ability to leverage private and public cloud technology. With the virtualization of datacenters utilizing SDN/NFV technology, where storage, server and software applications and their backups run on a shared resource pool, efficiencies tied to use of the cloud have reoriented spending across the business landscape, extending down to the smallest SMBs.

Moreover, new developments have added to the need to virtualize the datacenter. Most notably:

- The modularization of software requires scheduling of upgrades far more frequently than in the past, greatly adding to the burden of doing things the old way.
- With the onset of big data analytics, IT teams must continually ensure they are bringing analytics systems into play across multiple workflows, requiring a level of cooperation and coordination that is difficult to sustain in the fragmented traditional operations environment.
- The traditional data center has proven to be a hospitable environment for businesses seeking to exploit opportunities in the fast-paced Internet domain, where the focus is on agility and speed-to-market amid an incessant flow of application innovation.

SOFTWARE DEFINED VPN

At the connectivity level, the worldwide pursuit of cloud-enabled cost containment is driving escalating demand for IP VPNs. Currently this is part of a VPN market dominated by incumbent local exchange carriers delivering secured IP VPNs over MPLS (Multi-Protocol Label Switching) infrastructures.

But traditional carrier VPN service options are inadequate to market needs with regard to both costs and flexibility. From an operational standpoint, manually configured VPNs impose significant constraints on a company's ability to integrate cloud services, turn up new sites and deploy and provision new IT and communications systems.

The need for a new type of VPN connectivity service applies to multiple enterprise scenarios. Companies operating in many locations linked over wide area network connections need to overcome the limitations of the traditional enterprise WAN VPN service in order to more effectively exploit the benefits of IP, the Internet and cloud technology. At the same time, for all companies, including SMBs that have traditionally not been in the market for VPNs, secure VPN connectivity to cloud-hosted apps is now essential.

An emerging need for Enterprises includes VPN connectivity for employees outside the office. Businesses everywhere want to be able to provide secure VPN connectivity to applications, data and other company resources for employees on the go, which requires that network administrators be equipped to quickly set up VPNs for secure access over mobile and Wi-Fi links by any authorized person, wherever they are.

The Advantages of SD-VPN Connectivity

All these business needs require a shift away from inflexible VPN services that do not meet the emerging needs of the new digital enterprise. MSOs, leveraging their high-capacity fiber backbones and broadband access networks, have an opportunity to provide such a cost-effective alternative with superior versatility and ease of use in comparison to legacy connectivity services.

The key to success is utilization of virtualized networking technology to create a software-defined VPN (SD-VPN) architecture, which is a subset of the SD-WAN services that are in growing use among larger enterprises. Rather than using proprietary terminals in the network and at customer premises to nail up dedicated VPN connections, network operators can leverage network functions in the cloud to orchestrate point-and-click provisioning of SD-VPNs in incremental response to customers' needs on a pay-as-you-go cost basis.

With SD-VPNs in place, MSO customers will be able to:

- Leverage the virtualized cloud to lower costs of providing backup and recovery for their storage and network-based operations;
- Expedite connectivity as they add and move branch offices;
- Steer traffic into data centers for advance service chaining use cases;
- Utilize cloud-based applications and third-party services;
- Implement unified communications and network-based collaboration;
- Easily extend VPN connectivity over wireless links to employees;
- Make better use of video for training, in-house messaging and other applications;
- Eliminate the costs of buying and upgrading proprietary CPE.

Moreover, the SD-VPN platform will allow operators to extend the reach of their customers' VPNs beyond their cable service areas to any broadband-connected locations worldwide.

OPEN PLATFORMS AND SDN

Software-Defined Networking (SDN) is a rapid shift to open source technology. Open source technologies drive down costs, reduce vendor lock-in, and accelerate innovation and development cycles. There are many points of integration between the functional layers in a cloud architecture. Adhering to standards around these points of integration enables interoperability between solutions from multiple vendors. In some cases the standard is formalized through a vendor-neutral standards group. In others a de facto standard arises due to the dominant market share of one vendor in one area that drives a large

ecosystem of partners to align with that market.

As a result, cloud technology has quickly evolved into a multi-vendor ecosystem that is commonly called a "cloud stack" (not to be confused with the cloud orchestration software from Citrix® and Apache™ called CloudStack).

A generic cloud stack that is also part of a NFV platform includes all the hardware and server virtualization, including SDN controllers and virtual networking, all the way up to the cloud orchestration software that automates the entire infrastructure of servers, storage and networking.

Looking further into the emerging cloud infrastructure, it becomes apparent how open interfaces and protocols have contributed to real multi-vendor solutions as enterprises and services have built out functioning cloud architectures. For example, SDN controller technology typically supports higher-level cloud orchestration platforms (like OpenStack®) through REST-compliant APIs (generally called northbound APIs in the controller architecture). Standardized interfaces within OpenStack also allow networking vendors to provide a standard Neutron (the networking management framework of OpenStack) plug-in, allowing any vendor to support this open source cloud orchestration system.

By comparison, a combination of long-established and new networking protocols have standardized virtual networking and overlay networks across all network vendor equipment, virtual switches and cloud management systems. Examples of such protocols include VXLAN, OpenFlow™, or control-plane protocols such as BGP.

Open source communities are playing an important role in the evolution of cloud architectures in at least two important areas: the cloud orchestration platform (both OpenStack and Apache CloudStack), as well as the SDN controller software (through OpenDaylight, Floodlight, etc). Another key open source cloud component is the Open vSwitch (OVS), which has a rapidly expanding development community and is becoming a de facto standard for Linux™ deployments.

Open source projects generally ensure interoperability between solutions from multiple vendors and can shorten development cycles. However, they are not strictly required to achieve interoperability if other interfaces and protocols have been sufficiently agreed to by vendors and between products. As a result, there is not a one-size-fits-all approach when it comes to "open" in the emerging cloud ecosystem. The cloud industry has benefited from the open source approach of OpenStack as a multi-vendor cloud orchestration platform.

NFV ORCHESTRATION

Network Functions Virtualization is the key to unlocking the true potential of the cable industry's broadband and data center infrastructure. By supporting the instantiation and delivery of any service—independent of the access and transport technologies—NFV offers MSOs a significant competitive advantage.

Management and Orchestration is a key capability that needs to be implemented on the MSO cloud infrastructure for NFV. The ETSI NFV Industry Specification Group has defined the Management and Orchestration (MANO) functional group to include three components: the virtual infrastructure manager (VIM), the VNF manager (VNFM) and the NFV orchestrator (NFVO). VIMs are responsible for controlling and managing the compute, storage and network resources. Typically there are multiple VIMs responsible

for the resources at different NFV points of presence (PoPs) or data centers.

One or more VNFMs are responsible for managing the life cycle of VNF instances including the deployment, monitoring, scaling, healing, and software upgrade processes. Finally, the NFVO has two quite different responsibilities: the orchestration of resources across multiple VIMs, and the life-cycle management of network services, which consist of one or more VNFs.

While the NFV infrastructure manages sets of compute, storage and network resources, MANO elevates this to the application level; that is, MANO understands which resources belong to specific NFV applications. MANO enables users to monitor and control the state of NFV applications. With the VNFM function, MSOs can deploy applications and execute other life-cycle actions according to pre-established recipes, reducing time-to-market and helping to reduce human error.

Arguably the most important role of MANO is service assurance; that is, it helps providers to deliver services that are at least as highly available as those delivered with traditional physical network functions. Typical production NFV deployments will be distributed across multiple NFV-PoPs to avoid single points of failure due to technical or force majeure events, but even single location events need significant assurance capabilities. MANO is responsible for collecting status information and alarms related to infrastructure resources and also those at the application level related to VNFs and network services. Unlike traditional network functions that are delivered as a bundle of hardware and software, NFV applications and the underlying infrastructure are separate elements that can fail independently. In case of a failure, a critical MANO capability is to help operators understand which infrastructure resources are used to deliver the application services (physical to virtual to application mapping) and determine the real source of the failure (root cause analysis); for example, whether the failure is caused by a problem with a server or network or whether the failure is due to an application issue.

Another aspect of service assurance is disaster recovery. In case of a catastrophic event, MANO can help restore service applications faster through automated deployment, a process that may be less prone to human error than manual restoration.
MANO is also responsible for establishing and monitoring the (virtual) networks need for internal communication among the components of VNFs or between multiple VNFs that may or may not be distributed across multiple locations.

MANO plays an important role in the security of the solution. MANO itself needs to be secured against attacks and MANO should also help to secure NFV applications against outsider and insider attacks. With role-based access control, access rights for different user roles such as cloud admin and application owner can be carefully defined, and permitted user actions should be logged to make sure that a person responsible for all actions can be identified. MANO should be protected against illicit external access using a multi-tier architecture and web application firewalls. The MANO system needs to adhere to national security requirements with the appropriate certification level. MANO should allow the definition of security zones, security groups, and security appliances to isolate applications from each other. The system should be scanned for security vulnerabilities using appropriate tools. Private keys of VNFs need to be protected. All of these functions need to be accessible through an easy-to-use user interface, typically a GUI with appropriate logging capabilities for system events and user actions.

In addition, these functions need to be available through APIs to higher layer management systems such as an operations support system (OSS)/business support system (BSS) or VNF-specific element management systems.

Model Based Orchestration

Due to the challenges of human management and procedural orchestration, a model and policy-based approach to orchestration becomes a requirement. In this approach, VNFs and NFV infrastructure are described through domain-specific languages and different engines automate a variety of operational processes based on these models. A major advantage of the model-based approach is the possibility for different engines to use the same model for their respectivepurposes.

- Deployment/placement
- Capacity management, scaling
- Healing and root-cause analysis
- Maintenance/software upgrades
- Agile development and operations (devops)

On-boarding of Virtual Network Functions (VNFs)

While the ultimate goal of NFV is the full life-cycle automation of VNFs and their management by the NFV platform, a pragmatic step-by-step approach may be required for on-boarding of VNFs. This Multi-step methodology allows for an agile, fast, initial on-boarding of VNFs, while providing a path that will ultimately realize the full benefits of NFV.

The on-boarding of a VNF is a multi-dimensional activity aiming to gradually carry out transformation in various areas such as networking, hardware, applications and their management, and also in the way the service provider is internally organized, in order to progressively take more advantage of the benefits offered by NFV. This means some VNFs are more advanced in one dimension than others.

The iterative on-boarding process allows for a gradual advancement toward the ultimate scenario of fully automated life-cycle management of the VNF. While there is no rigid path toward this progression, we can identify some typical major milestones:

• *Deployment:* At this stage, new instances of the VNF can be automatically created and deployed by the NFV Platform, which uses the information in the VNF Descriptor to fetch the software binaries from the VNF Catalog, and then instantiates and configures them over the selected NFV-Infrastructure.

• *Automated healing:* At this stage, the NFV platform can automatically handle the healing process; for example, the recovery of an application after failure of a hardware or software component. Typically, once an alarm is triggered either by the OSS or by the NFV MANO platform itself, a set of automatic actions are automatically executed to restore the application's functions. These actions can consist of the re-creation of the failed software component on a different server, the modification of the networking configuration, and other actions to connect the newly re-created software with the other components.

• *Automated scaling:* As the load on a network function varies, the network function may require more resources (or fewer resources) to be able to function without being slowed down. Therefore the resources allocated to the VNF need to be increased (scaled up) or be reduced (scaled down). At this stage of integration, the VNF platform is capable of identifying when the resources allocated to a VNF are close to being exhausted (typical example is high CPU usage due to high demand). It will then automatically allocate additional resources, create additional

instances of the software components that need to be scaled out, and most importantly reconfigure the network so that traffic is distributed over a larger pool of resources. A similar set of actions are also performed, when an application no longer requires the extra capacity allocated to it, and resources are automatically freed.

• *Automatic full life-cycle management:* The ultimate objective of NFV is to allow not only for the instantiation and automated management of an existing release of a VNF, but also to automatically update (application of patches) and upgrade (roll-out of new releases) the VNF with minimum or no outages, thus relieving the operational team from the tedious tasks that are often executed outside normal working hours. The automation of life-cycle management also protects the system from potential human errors that can happen during these critical maneuvers.

OPERATIONAL SUPPORT SYSTEM (OSS) SIMPLIFICATION

Today, there is a real opportunity for service providers to simplify their OSS by introducing a highly automated and dynamic approach to service operations. Why is this happening today?

First, the use of virtualization and other cloud software technologies inside network functions (NFV) and their deployment as applications on IT infrastructure provide the opportunity to enrich and align network operations with agile IT operations practices. Dynamic application lifecycle management principles can be leveraged from the IT world and directly applied to network functions, which allows simplification by moving from dedicated to IT standardized processes including DevOps automation processes.

Second, new flexibility is introduced by linking the OSS processes to the features of NFV orchestrators and SDN controllers as programmable networking technologies exposed using open APIs. OSS systems that are able to leverage the capabilities of such dynamic resource managers can then achieve the dynamic operations processes that are required to take full advantage of NFV and SDN.

Assuming that the introduction of new technologies will take some time, an effective evolution strategy must also consider service operations across hybrid networks consisting of both current and new technologies. In particular, the following two aspects of OSS evolution need to be addressed, sequentially or in parallel.

1. NFV readiness: The current OSS needs to be assessed with respect to its capability to move up the value chain from current to dynamic operations. This includes cleaning up current OSS service fulfillment, inventory and assurance stacks (by consolidating and automating where possible), and ensuring OSS readiness for dynamic process support. These steps are required in order to play a part in the target OSS for NFV and SDN and to be ready for dynamic operations.

2. Introduction of new operations solutions linked to NFV and SDN technologies: This includes extending or elaborating new operational processes that work across network/SDN and cloud/NFV platforms, as well as analyzing their impacts on current processes and organizations.

Abstraction is particularly useful for OSS service operation in an NFV & SDN world where highly dynamic resource management may take place in order to avoid service impacts. Abstraction between resource and service layers is key for end-to-end dynamic

operations & OSS simplification. Abstractions of underlying domain capabilities, whether technical, organizational, or geographical domains, are needed to shield the service operations OSS from resource-focused details.

These abstractions are gathered in catalogs, and become the atomic elements for easy service composition. Abstraction and composition can be repeated as often as needed to simplify end-to-end service operations. The availability of a simple information model to represent service instances with minimum information at each abstraction level is essential to this approach. This model allows the required object hierarchy to be implemented, where instances are created dynamically through service orchestration or auto-discovery, and can also be used by automated service assurance processes, thus enabling to close the current gap between fulfillment and assurance.

Hybrid service orchestration is also a key requirement while we are in this transitional state. E.g. using agile orchestration for activation of end-to-end services across Physical Network Function (PNF) and VNF based networks. The solution should allow the MSO to create a simplified and open orchestration architecture that works across networks and IT, and isolates business operations from network management.

OSS simplification should address automated assurance and dynamic operations:

- *Automated assurance*: This capability addresses customer needs by bringing customer care and network operations closer together through intelligent root cause and service impact analysis tools. This typically include analytics-driven automation of service assurance and allows you to build in proactive, automated processes that automatically identify and resolve service and network issues, and possibly avoid them through proactive measures.

- *Dynamic Operations for NFV and SDN*: This capability enables the operationalization of NFV and SDN based services by closing the loop between automated assurance and hybrid service orchestration through a common engine that spans across networks and IT and automates the discovery and reconciliation of network resources.

THE VALUE-ADDED SERVICE (VAS) OPPORTUNITY

Until now, MSOs supplying connectivity services to businesses have largely deferred to third-party CPE vendors, VARs and cloud-based managed service suppliers to provide value added services and applications over those connections that are specific to customers' needs.

But the emergence of cloud-hosted models and with the deployment of flexible networking and delivery platform for virtualized functions, MSOs can create the opportunity to tie into those models to layer robust portfolios of value-added options into their service bundles.

Virtual VPNs can open a path for MSOs into this value-added services market at a moment when access to cloud-based applications is transforming business operations. The wide-spread embrace of the benefits from lower costs, greater agility and better service availability that have softened market resistance to the cloud has made the availability of bundled value-add options a natural next step in market migration to ever greater cost savings and efficiencies.

The significance of MSO's opportunity to bundle VAS with connectivity services is reflected in a global survey of SMBs conducted by researcher AMI-Partners. The

study found SMBs prefer bundled SaaS offerings over single services by a margin of four to one.

Leveraging revenue-sharing partnerships with SaaS providers and developers, cable operators can avoid the painful experience of watching other entities use their high-speed connections to capture all the VAS business. They can up sell their connectivity customers to bundles of cloud-hosted services and applications that can be automatically provisioned through customer service representatives (CSRs) and self-help portals. With dashboard control over those apps, customer administrators can manage usage and other policies on a per-employee basis and set up access for outside collaborators as they're brought in to work on specific projects.

CONCLUSION

The MSO case for the aggressive pursuit of new commercial service opportunities through use of virtualized networking and NFV technology is clear. By deploying some of the capabilities mentioned in this paper, not only will operators be able to take the lead in providing the SD-VPN connectivity that SMBs need to maximize operational efficiency and lower costs, but also will be well positioned to build the in-house capabilities and partnerships with SaaS providers that will allow them to compete aggressively in the lucrative VAS market.

As revealed in research studies of market trends and the analysis supplied by Bell Labs Consulting, there's a vast pool of pent-up demand at all business levels for the benefits MSOs can deliver via SD-VPN connectivity and associated value added services. Costs of implementing these capabilities, unlike many new network technologies, are low enough to ensure payback on MSO investments within a very short time. With the ability to meet customers' VAS needs, MSOs will have an opportunity to drive those returns ever higher over time.

REFERENCES

1. Computer World, 2015 IT Spending Forecast, November 2014

2. AMI-Partners, SMB Cloud Playbook – Opportunities & Actionable GTM Insights, November 2014

Distributed Architectures and Converged Access Network

Jorge D. Salinger
VP, Access Architecture
Comcast Cable
1701 JFK Boulevard
Philadelphia, PA 19103
+1 (215) 439-1721
jorge_salinger@cable.comcast.com

Abstract

Cable operators are now actively converging video and data services into a common CCAP platform. And, as the usage of narrowcast services continues its seamlessly exponential growth, the evolution of the access network progresses relentlessly towards ever-smaller service groups. This results in a continuous growth of headend equipment, which will soon far exceed the capacity that facilities can support.

Therefore, the above trends are now intractably linked to two additional evolutions: distribution of components of the access network, and virtualization of the core network functions.

This paper will begin by outlining why and how the migration to a Distributed Architecture is necessary and beneficial. The paper will then expand into features that can be implemented with Distributed Architectures, and some of the resulting possibilities. Finally, this paper will discuss the topic of Virtualization, and how these additional evolutions relate to each other, as follows:

1. Present why Distributed Architectures are useful, which would cover the following key benefits:
    - HE density increase, which is becoming critical as MSOs are segmenting service groups more and more, including moving towards fiber deep and passive networks
    - HFC network performance improvements by migrating to digital transport, especially as we move to DOCSIS 3.1 and would like to maximize the use of the higher order modulation rates that DOCSIS 3.1 offers
    - Trunk fiber savings as we move to higher capacity digital links that can be muxed much more than analog links, and
    - The ability to eventually virtualize the remaining upper layers of the CCAP
2. Discuss features that could be implemented in Distributed Architectures, which are not possible with the current analog forward and digital return links, such as:
    - Broadcast capacity replication
    - Multiple service groups per HE port
    - Optical path redundancy
3. Present the concept of Access Network Convergence, explaining how a single network could be used for various access technologies, which will cover:
    - The use of Ethernet switches in the node
    - The use of multiple access technology modules in the node (RF, PON, etc.)
4. Finally, outline a network evolution strategy for services, and show how a Converged Access Network would make the evolution easier, including the Virtualization of Access Network Functions.

Typical HFC Networks Today

Most MSO's hybrid fiber-coax (HFC) networks have been designed to either 750 or 860 MHz of spectrum capacity, while a smaller number designed to support 1 GHz and even newer networks designed to support 1.2 GHz. For the more abundant 750 or 860 MHz networks, if not already fully

utilized already, it is expected that use of their capacity will be increased to the point of exhaustion as the use of DOCSIS® increases for the higher high-speed data (HSD) service tiers, additional high-definition (HD) programs for both broadcast (BC) and especially narrowcast (NC) services such as video on demand (VOD) and switched digital video (SDV) are deployed, or new services such as internet protocol (IP) video and cloud-based digital video recorder (cDVR) are added.

In recent years the growth in, and demand for, HD programming has resulted in the need for allocation of large numbers of EIA channels for HD services, both for BC and NC, which has filled every available portion of the spectrum. This is especially true for BC, where large numbers of programs are offered in HD format, while simultaneously the need for distributing the standard definition (SD) version has persisted. This has resulted in the need for use of 3x to 5x the number of EIA channels than previously required. For example, a typical digital multiplex including 10 to 15 programs would require an additional 3 to 5 EIA channels for the HD equivalent streams, even assuming the newer, more sophisticated multiplexing schemes available in the market. Of course not every program is available, or still sought by subscribers, in HD format. But very large numbers of them are, including 100 to 150 BC programs.

The above is also applicable to a great extent in systems utilizing SDV technology for distribution of its content. The difference is that the HD and SD versions of the program are not distributed unless a subscriber is requesting them, which reduces the marginal increase in capacity. Assuming that all programs are distributed in only one format, which is certainly a valid expectation for programs of low viewership, then the increase in capacity for a conversion from SD to HD would just be the increase in capacity required for the transmission of the HD program without requiring the simultaneous use of bandwidth for both formats.

Additionally, considerable spectrum is needed to deploy high-capacity narrowcast legacy video services, especially cDVR, and a full-array of HD video-on-demand services. For the former, initial observations suggest that network requirements for cDVR may be as high as 4x to 5x that of VOD, and that peak utilization overlaps, at least partially, with that of peak use for other narrowcast services.

Finally, the growth in HSD services continues. All network operators have offered increased service tiers and observed an increased use of HSD service capacity for well over a decade now, as shown in Figure 1, which amounts to a year-over-year compounded growth of 40% to 60%. The applications have changed throughout this time, and the demand has continued to increase at the same relentless rate.

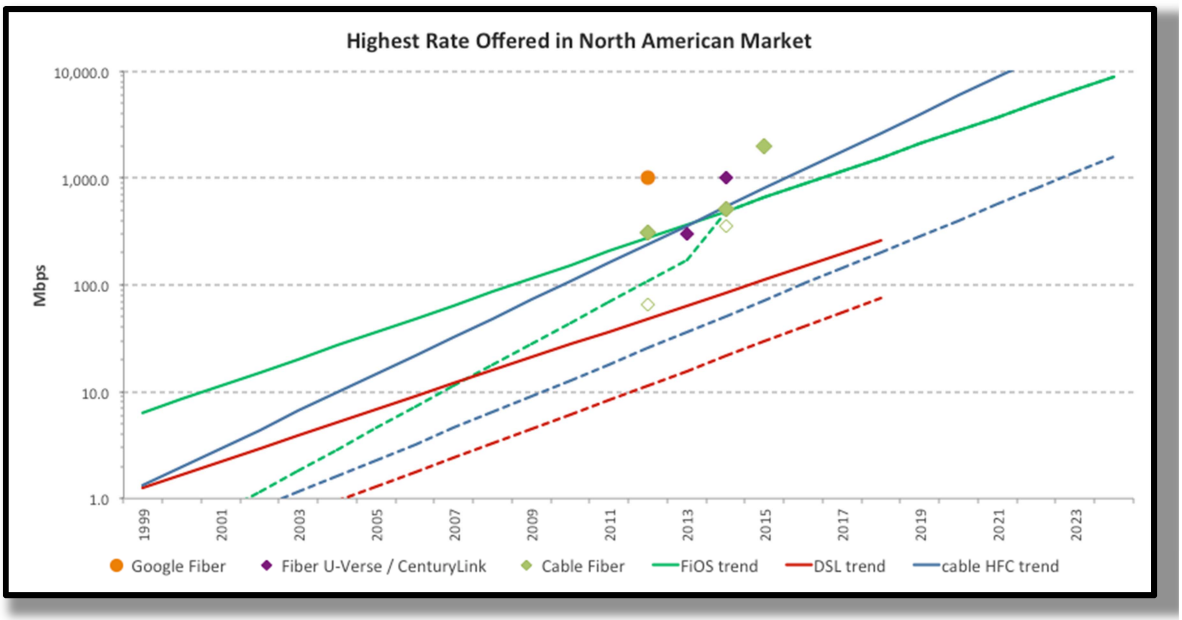


Figure 1: Examples of HSD service tier capacity increase over time

How does this compare to other operator's data services and a longer period? Projecting operator's HSD service growth back in time to when Internet services started, 25 years ago services should have been about 100 bps. This coincides with the history of telephone modems from 110 and 300 baud modems from the mid-80s, to 56 Kbps/V.42, into ISDN services.
This demonstrates that the growth seen in MSO's HSD services is typical over a much longer period of time, rather than an exception observed by operators in recent years.

Growth Projections

From all of the above, it then follows that, should the usage growth pattern continue at the same rate as in the past, networks would be required to provide HSD services in the range exceeding 1 Gbps within the next few years. This growth, coupled with the surge in HD video formats, and more personalized narrowcast services, will result in a significant growth in narrowcast capacity, as shown in Figure 2 below.

To support this growth, MSOs have deployed, or are considering deployment of, bandwidth reclamation tools such as SDV for digital broadcast, digital terminal adapters (DTAs) for analog service reclamation, or a combination of both. These tools have been extremely valuable to MSOs, and their operational complexity and cost well justified.

In the case of SDV, early predictions several years back from industry analysts projected that the efficiency of SDV would reach 40% (e.g., programs requiring 10 EIA channels could be carried in 6). This has proven to be understated, since it was based on the use of SDV for reduction in bandwidth required for existing services. As SDV's role in the network grew, the efficiencies have been even greater, especially as SDV has been used to introduce niche services that have low viewership and would have otherwise been difficult to deploy.

The benefit of DTAs has been just as, or perhaps even more striking. MSOs deploying DTA devices are able to eliminate the need to distribute the analog channels in the network. Even if DTAs are distributed to top analog tier customers, such as only to subscribers of the traditional expanded basic subscribers, such deployment would reduce a channel line up from perhaps 50 EIA channels dedicated to 50 analog programs to perhaps as little as 4 EIA channels dedicated to transport the 50 programs in

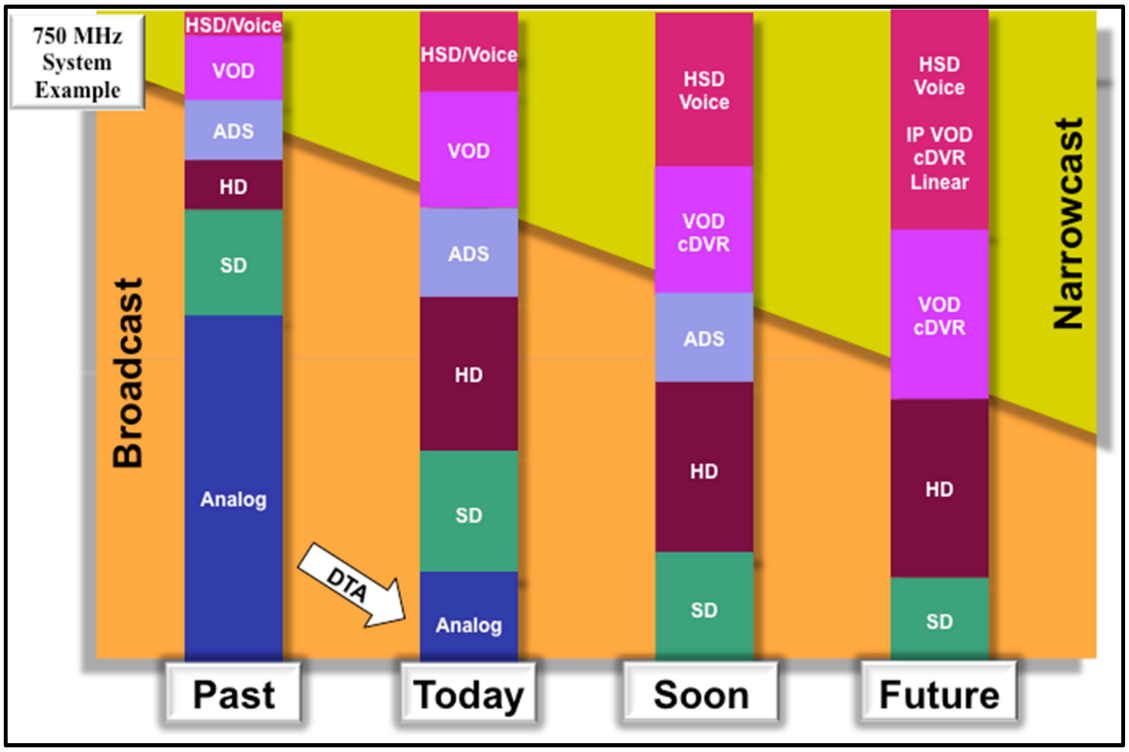


Figure 2: Example of narrowcast service growth over time

their equivalent digital transport. Using the same comparison method as the above SDV case, this is a >90% efficiency. If extended to the entire analog tier the efficiency gains are very significant.

Despite the availability of these tools, they are not universally applicable. With respect to SDV, in general it is not likely that all broadcast programs will be switched since experience shows that many broadcast programs are constantly viewed by someone in the service group during peak hours, which will leave a large portion of the spectrum still used for broadcast. Similarly, not all analog channels can be removed in the short term due to operational and/or cost constraints.

Additionally, while many MSOs will use one or both tools, in general these tools won't be used by every MSO for all applications.

Finally, there are also significant potential gains to be achieved from the use of advanced video CODECs (AVCs) and variable bit-rate (VBR). In the case of AVCs, coding efficiencies of approximately 50%, depending on implementation and content type, can be obtain with H.264[1] and/or MPEG-4 Part 10[2]. Furthermore, with the recent release of the H.265[3] standard in April of 2013, it is possible to achieve a 50% improvement over H.264. And the use of VBR promises to result in a capacity efficiency gain of as much as 70% versus CBR[4]. The combined gains from using the above approaches for multiple services are even more significant.

However, these are difficult tools to take advantage on the network since, proportionally, relatively few legacy set-tops still support all these technical advances, especially H.265. These tools will likely enjoy significant support in newer, IP-video based services equipment moving forward.

But, this approach will require additional capacity on the network. This is especially true when considering that the deployment of these advanced video services will result in an additional simulcast of video programs, at least initially, which is expected since its deployment will not, at least initially, replace the currently deployed service formats.

Furthermore, ubiquitous support for such devices would require considerable spectrum if the legacy services are maintained for an extended period, as it is expected since legacy devices are, and likely will, continue to be deployed. Moreover, this increase in simultaneous use of the more advanced IP video services while maintaining legacy services will be especially impacting over time as its penetration increases given their narrowcast nature.

All of the above, coupled with the success experienced by MSOs in recent years with business services, homes security, etc., will likely require the deployment of IP capacity beyond what can be supported today, requiring the development of tools for increased efficiency in the use of spectrum and/or unlashing of additional spectrum in the HFC network. The following sections of this paper will enumerate ways in which this can be achieved.

The Advent of DOCSIS 3.1

As it has been pretty well advertised in the media, DOCSIS 3.1 is under development and even beginning deployments. Comcast has announced deployments of DOCSIS 3.1 across multiple markets, and several other operators are doing the same. By 2017, DOCSIS 3.1 will have been broadly deployed in MSOs' networks.

The key motivation for the new version of the DOCSIS specification is, in a nutshell, to scale DOCSIS more efficiently, both from the cost and operations perspectives. While for the first 10 years or more it was possible to offer Internet services and support its growth with just 1 DOCSIS channel, services today require many more channels. This is because 1 DOCSIS channel provides almost 40 Mbps, which

---

[1] ITU-T Recommendation H.264: 2005, Advanced Video Coding for generic audio-visual services

[2] ISO/IEC 14496-10: 2005, Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding

[3] ITU-T Recommendation H.265: 2013, High efficiency video coding

[4] Capacity, Admission Control, and Variability of VBR Flows, CableLabs Winter Conference, February, 2009

was well above the data rate of the services offered in the past. However, the year-over-year growth drove service speeds well above the initial levels, to 20, 50 and even higher Mbps tiers today, which can't be supported by the single channel. MSOs then went to multiple DOCSIS channels, now reaching 16, 24, and even 32 channels, and soon requiring well beyond the capacity supported by DOCSIS 3.0.

To that end, the 3 key goals and features of DOCSIS 3.1 are:

1. Much more efficient use of spectrum, with up to 50% improvement in bandwidth efficiency (or bps/Hz, resulting from:

   a. The use of more efficient forward error correction (i.e., replacing the older and less efficient Reed-Solomon approach for the more modern and far more efficient Low Density Parity Check, and

   b. Addition of the higher-order modulations 1024 and 4096 QAM downstream and 256 and 1024 QAM upstream.

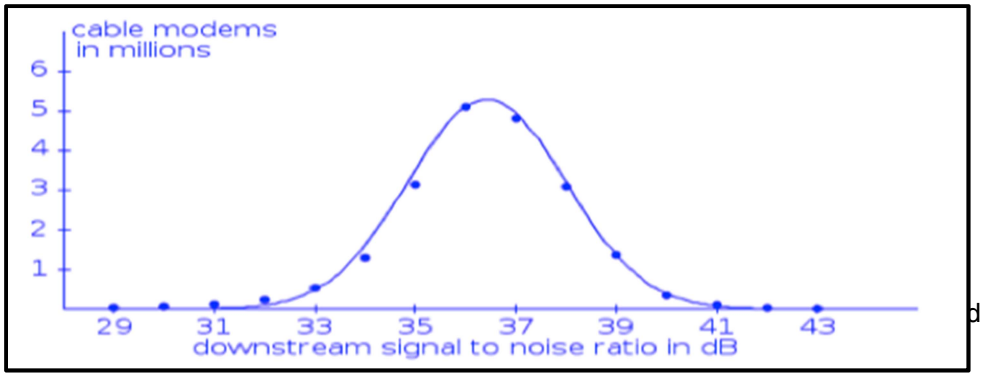
   These new modulation schemes provide 2 and 4 bits/Hertz/second improvement in both upstream and downstream, while the use of the new forward error correction approach provides approximately 5 dB better RF performance. The end result is that MSOs will be able to transport 1 Gbps of DOCSIS capacity in about 120 MHz of spectrum while doing the same with the current DOCSIS approach using single-carrier QAM requires about 180 MHz of spectrum.

2. Cost reduction, mainly by leveraging technologies commonly used in other transmission media, such as the inclusion of Orthogonal Frequency Division Multiplexing, which is used extensively in wireless and wireline transmission media. Specifically, the addition of OFDM for the downstream and OFDMA for the upstream should enable MSOs to reduce costs while "packing" more bits in the HFC network more efficiently since these technologies likely result in a larger supplier ecosystem, increasing innovation and fueling competition.

3. Enable a simple and orderly transition strategy, both with respect to compatibility with previous generation CMTS and CM equipment while supporting an expanded spectrum capacity in the HFC network.

   Specifically, DOCSIS 3.1 cable modems will operate with DOCSIS 2.0 and 3.0 CMTS/CCAP equipment, enabling deployment of DOCSIS 3.1 CPE as soon as available. Similarly, DOCSIS 3.1 CCAPs will support DOCSIS 2.0 and 3.0 CPE allowing MSOs to upgrade headend equipment without having to change any of the existing CPE. And, both DOCSIS 3.1 CM and CMTS equipment will support the currently required upstream and downstream spectrum, plus an expansion of the upstream to 85 MHz and beyond, and of the downstream up to 1.2 GHz.

Figure 3 depicts the downstream signal-to-noise ratio (SNR) as reported by a very large population of cable modems[5]. This data verifies that many cable modems will be able to support the high-order modulation profiles included in DOCSIS 3.1. However, others will not without an increase in SNR.



d to

Figure 3: Example of downstream SNR for a large population of cable modems

Assuming an 8/9 coding ratio, Table 1 shows the required SNR for the modulation rates included in DOCSIS 3.1:

Modulation	Signal-to-Noise Ratio
512 QAM	27 dB
1024 QAM	30 dB
2048 QAM	33 dB
4096 QAM	36 dB
8196 QAM[6]	39 dB
16384 QAM	42 dB

Table 1: SNR required for DOCSIS 3.1

Applying the SNR requirements from Table 1 to the population of modems shown in Figure 3, we can easily see that a large population of cable modems would not achieve sufficient SNR to operate at 4096 QAM. Furthermore, if sufficient headroom is allowed to account for environmental fluctuations, the population of cable modems that would not receive signals with sufficient SNR to operate at 4096 QAM would be significant.

The Analog Modulated Forward Link in HFC Networks

As their name indicates, hybrid fiber-coax networks use a fiber transport between the headend and the coaxial cascade. This fiber link, intended to reduce the size of cascades, mainly driven to improve performance, was originally developed with analog modulated lasers and receivers in both directions, upstream and downstream.

Over time, the performance of the upstream link was improved by replacing the analog modulation with a digital transport. This change improved performance significantly, and allowed for longer distances between the headend and the node. Different vendors implemented their own methods and technical capabilities to implement a digital transport, which resulted in incompatible systems and required the use of the same vendors' components for both the node and the headend.

However, the downstream link remained almost unchanged over time, with the only enhancements focused on improving distance and RF spectrum capacity. Performance has not really been an issue like it was in the upstream.

But more importantly, while the digital capacity of the upstream was limited to a few megabits per second, well under a gigabit of digital capacity which could easily be digitized and carried with Ethernet optics, the downstream digital capacity necessary to transport the downstream spectrum has been considerably higher, reaching and even exceeding 10 gigabits per second.

Because of the above, analog forward links continue to be used to date. And, while headend equipment is currently capable of launching signals with >47 dB MER performance, which would be sufficient to generate and transport 16,384 QAM signals, analog lasers are limited to about 35-38 dB of MER performance, which would limit end-of-line performance to barely enough for 2,048 QAM or 4,096 QAM in short cascades the best of the cases.

Description of Options for Digital Forward Link

As time has gone by, technology evolution and certain developments as described below have enabled options for implementing a digital forward link. These include:

1. Evolution of QAM edge modulators which have gone from single and/or a few modulators to supporting 32, 64 or even more modulators,

2. Development of the CCAP, combining the functions of the video QAM modulator and DOCSIS into a single platform, and

3. Migration to digital video, either partially for now or already completely.

With this technology evolution, it is conceivable to remove the RF combiner network, and instead implement it digitally in the edge device, such as the CCAP.

This evolution of the edge headend devices makes it possible to envision several options for digitizing the forward link.

[6] 8196 QAM and 16384 QAM are included for future consideration in the DOCSIS 3.1 specifications

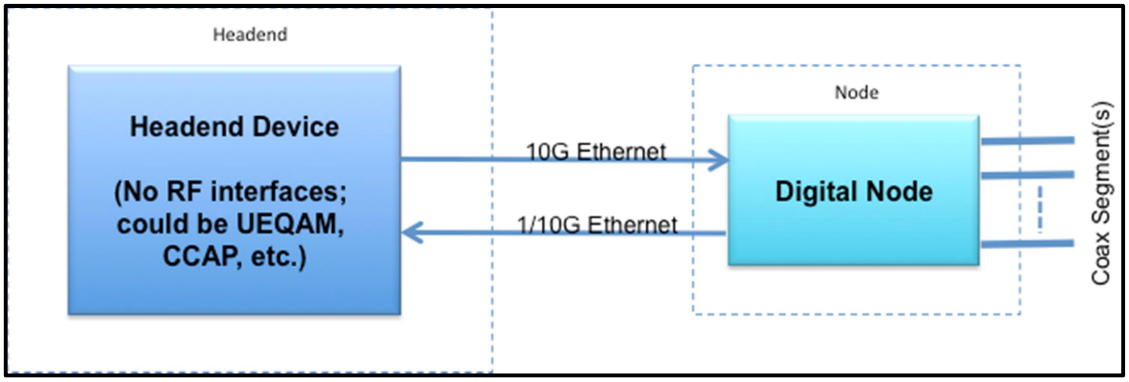


Figure 4: Digital Forward – High Level Architecture

Fundamentally, the migration to a digital forward includes the components included in Figure 4, as follows:

- The headend device, such as a CCAP, which would be a high-density edge QAM comprising QAM modulation for the entire spectrum,
- The node would contain components normally implemented in the edge QAM or CCAP which generate the RF signals,
- The link between the headend device and the node would be comprised of a digital interface, such as an Ethernet link.

There are then various approaches for how a digital forward link can be implemented to replace the currently used analog link. These various approaches for distributing the various components can be categorized into 4 groups, plus 1 option that would still leave an RF generation at the headend device, as outlined in Table 2:

Option	Description and Approach
1. Maintain RF output in the headend	1.a Headend equipment remains unchanged 1.b Headend RF output is digitized, transported digitally, and RF is regenerated in the node
2. Remote the DAC from the PHY	2.a The DAC is removed from the headend 2.b Digital samples are transported digitally to the node where the DAC generates the RF signals
3. Partition the PHY and remote the lower portion of the PHY	3.a The PHY is split between the headend and the node 3.b The digital bit stream between upper and lower PHY is transported from headend to node
4. Remote the entire PHY	4.a The entire PHY modulation is moved to the node 4.b The MAC remains in the headend, and MAC frames are transmitted from the headend to modulator that resides in the node
5. Remote the entire PHY and MAC	5.a The entire PHY and MAC is removed from the headend device and placed in the node 5.b IP frames are transported from the headend to the node.

Table 2: Categories of options for implementing a digital forward link.

Comparison of Options for Digital Forward Link

There are pros and cons for each of the options. The following sections outline these trade-offs.

Option 1: RF remains in the headend

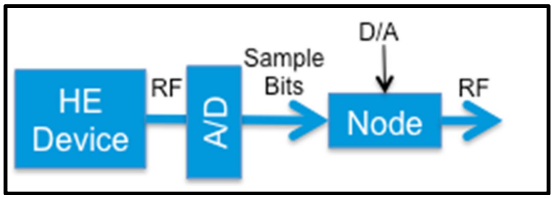


Figure 5: Block diagram for Option 1

- Equivalent to digital return, the RF output from the headend device is digitized, transported digitally, and converted back to RF in the node.
- Maintains HFC transparency
- This option results in the highest bitrate over fiber; the capacity for multiple nodes would not fit into the available capacity of one 10G fiber
- There is a loss of MER in the double conversion, so this option provides the least performance improvement
- Results in the least intelligence placed in the node, but an additional conversion stage is added in the headend

Option 2: Digital-to-analog conversion is moved to the node

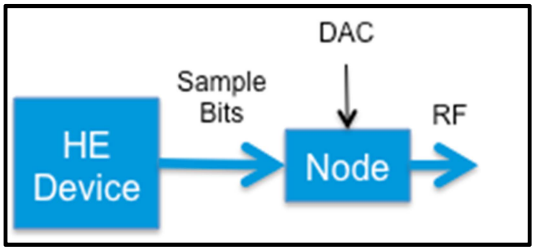


Figure 6: Block diagram for Option 2

- Requires separation of the digital-to-analog conversion from the modulator
- Together with Option 1, results in the least intelligence in node
- Similar high bitrate over fiber as Option 1; capacity for multiple nodes would not fit into the available capacity of one 10G fiber

Option 3: Lower PHY is moved to the node

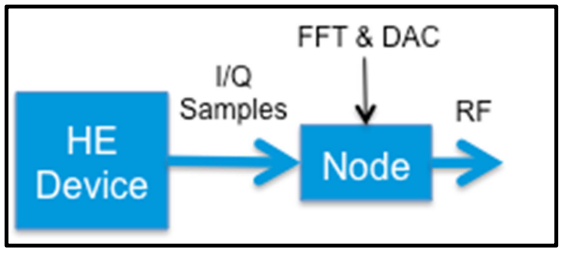


Figure 7: Block diagram for Option 3

- The PHY layer needs to be split into two components: upper and lower PHY
- More intelligence than in either of the previous options is placed in the node
- Although lower than the previous options, this option also results in a very high bitrate over fiber
- This option would require and industry proprietary point-to-point link between the headend port and the node to transport the I and Q samples
- Implementation of this option would require the definition of interfaces which have never been defined in previous versions of the DOCSIS specifications, which in turn would result in modification of the silicon used and/or planned to date, and therefore results in the highest implementation complexity

Option 4: Entire PHY is moved to the node

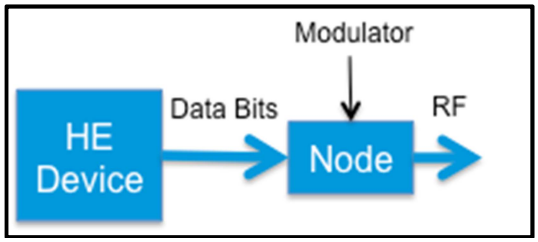


Figure 8: Block diagram for Option 4

- More intelligence is placed in the node than with all previous options
- This option results in the lowest bitrate over fiber; multiple nodes fit into the capacity of a 10G fiber
- Enables a packet-based link between the headend and node, which results in significant benefits outlined later in this paper
- Could use existing/planned silicon devices, and thus may be the easiest and quickest to implement
- Offers the best MER performance improvement over analog

Option 5: Move PHY and MAC to the node

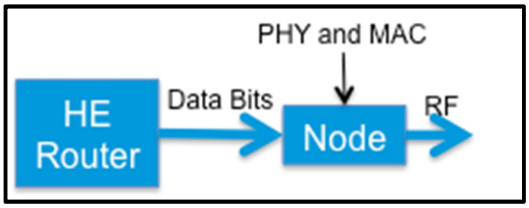


Figure 9: Block diagram of Option 5

- This option puts the most intelligence in the node
- The data rate between the headend and the node is equivalent to the actual data transmitted, except for the addition of ancillary network data
- Same packet-based network benefits as Option 4
- Same highest MER performance as Option 4

Benefits of Digital Forward Link

As described above, one key benefit of Digital Forward Link is the improved performance resulting from the migration from an analog to a digital link. This gain varies depending on the characteristics of the analog link being replaced, but can be generalized as at 5 dB of improved signal-to-noise ratio at the end of the line. This gain will result in higher capacity/Hz as it will be possible to run higher order modulations for more of the cable modems in the network.

In addition, the Digital Forward Link will offer the benefit of enabling longer distances between the headend and the node. This is because digital interfaces, such as an Ethernet link, are designed to operate over much longer distances while carrying the designated capacity. Extending the distance between the CCAP and the Digital Node would enable MSOs to move their CCAP devices back in the network to more centralized facilities, leaving the hub or OTN free of CCAP equipment. The benefit of such change could be very big for some MSOs, especially as segmentation of the network continues towards smaller service groups, for which additional CCAP equipment needs to be deployed.

A third benefit from the Digital Forward Link is improved reliability of the optical link. It is well known that analog links require period maintenance and are subject to the effects of environmental changes. By contrast, Ethernet optical links are far more stable across a wider range of environmental conditions, and require little to no maintenance. The impact of this benefit could be very significant to MSOs.

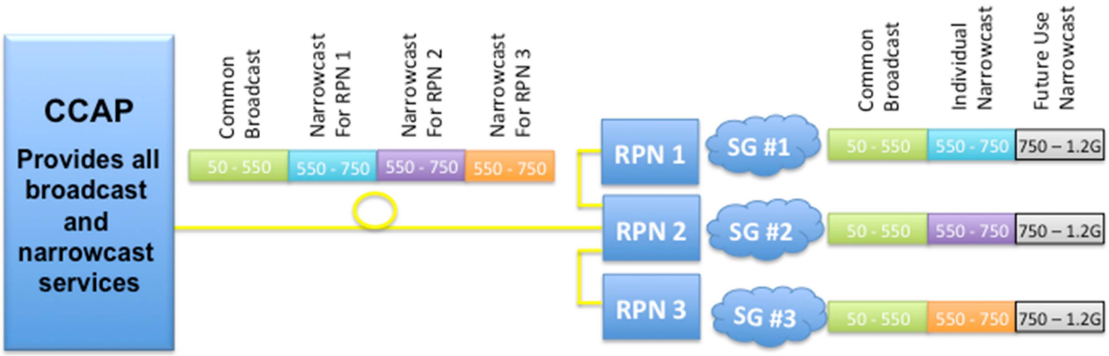


Figure 10: Reuse of broadcast capacity across multiple RPNs

Increased Headend Equipment Density

The implementation of distributed architectures makes it possible to improve the density of CCAP devices in several ways.

First, while CCAP devices are normally implemented via separate upstream and downstream line cards, a distributed architecture line card would be implement both upstream and downstream. This, in effect, doubles the capacity of a CCAP chassis.

In addition, a typical CCAP downstream line card will house 8 or perhaps 12 RF ports because of the printed circuit board space required by the components required for RF modulation plus the sheer connector spacing required. However, Ethernet connectors can be placed considerably closer to one another, allowing a similar line card to easily house 16 to 24 ports. This additional density gain once again doubles the capacity of a CCAP chassis.

Finally, it is possible to consider “daisy chaining” digital nodes (RPNs) off of a single CCAP Ethernet port. This is because, on the one hand the capacity of an 10 Gbps Ethernet link would support the capacity needed for a single RPN, plus in addition it is possible to generate an RPN “channel line-up” by transmitting the broadcast content once to multiple RPNs. As depicted in Figure 10, the data stream transmitted from the CCAP could contain a single “copy” of the broadcast line-up content, plus individual versions of the narrowcast content for each of the RPNs. The RPNs would then reuse the broadcast line-up content to recreate the individual RPN channel line-up. In this way each service group served by the CCAP port would contain the same broadcast line-up but its individually different narrowcast line-up.

Then, as the narrowcast line-up capacity grows over time, CCAP ports would be segmented to support less RPNs, akin to the way service groups are split today to support more narrowcast capacity as it is required.

As summarized in Table **3** below, the combined effect of the 3 factors described above is very significant, ranging from 8x to 18x of headend capacity gain. From a space and power perspective, this is hugely impacting savings.

Density Factor	Density Gain
Combined US/DS line card	2x
Greater number of ports per line card	2x to 3x
Multiple RPNs per CCAP port	2x to 3x
Combined capacity gain	8x to 18x

Table 3: Distributed architecture headend density gain

But, just how meaningful is this headend density gain?

Considering that a migration from an HFC architecture with an average of N+5 (i.e., a node plus 5 amplifiers in average) to N+0 would require about 10x the number of nodes, the headend density benefits resulting from distributed architectures would neutralize the potential increase in CCAP equipment.

It is then quite clear that from a space and power savings, distributed architectures take the benefit of CCAP to a whole new level.

Integration of HFC and Fiber Services

One of the largest areas of growth for MSOs is business services. MSOs have deployed business services via both cable modems and fiber-based infrastructure. The fiber-based services are either point-to-point, using Ethernet and wave-division multiplexing (WDM), or point-to-multipoint, using PON technologies (either EPON or GPON).

This duality results in the existence of two parallel networks. One of them, the HFC infrastructure, uses fiber from the headend to the node via an analog modulated link for the forward direction and either analog or proprietary digital return, followed by coax infrastructure from the node to the home. The other consists of digital fiber from the headend to the subscriber, which is used for commercial services.

Given the use of a digital fiber in both the forward and the return for the RPN, and especially because this digital fiber is based on Ethernet technology, it is possible to

collapse both of these networks into a single infrastructure.

Therefore, the implementation of RPNs with an Ethernet interface between the CCAP and the RPN would make it be possible to implement a PON interface at the RPN.

The benefits from this integration include:

- Reduce the optical link for PON to the distance between the node to the customer premise
    - Since the typical distance from a node to a customer premise in an N+0 architecture would be 1-2 kilometers. This would virtually eliminate any distance limitations for PON, making it possible to implement the largest possible densities as network capacity enable.
    - In addition, this shortened distance would enable the use of lower power optics, which can translate into significant savings, especially for 10 Gbps optics, and especially for the upstream which results in significant savings in the ONU.
- Leverage a single network for multiple services, which will reduce maintenance and increase operational efficiencies.

Migration Strategy

Clearly, one of the more concerning issues to MSOs is the migration strategy.

Any migration that requires synchronized cut-overs, or which requires changes in multiple locations to execute, is problematic, and usually results in a barrier to adoption. Therefore, it is very important that the migration to distributed architectures allow for unsynchronized changes.

Furthermore, ideally the migration to distributed architectures allows for opportunistic changes in the network. For example, one such change would be to migrate a single node, such as would be the case in an MDU to increase capacity.

As it turns our, distributed architectures enable such gradual, unsynchronized and opportunistic changes in the network. What follows is an overview of the steps and components involved in the migration to distributed architectures.

Starting with the components of the network on both sides of the distributed architectures, neither the back-office nor the various components in the customer premise need to be modified in any way. All back-office components are unaffected by the migration to distributed architectures, and any additional MIBs for management and/or commands for configuration as needed can be added well before the first distributed architecture CCAP line card or node are deployed. With respect to customer premise devices, these would not be affected in any way in order to deploy distributed architectures, and any enhancements that are made possible through the introduction of distributed architectures would be implemented in CPE equipment that can be introduced before or after the migration to distributed architectures.

The critical portion of the network were changes need to be made are in the headend and the plant.

To begin with, the changes required in the headend are principally in the CCAP platform. The CCAP architecture was specifically designed to support multiple technologies simultaneously, which makes it possible to install regular RF upstream and downstream line cards and distributed architecture line cards in the same chassis. While some MSOs may chose to deploy a separate CCAP platform for distributed architectures, it is certainly possible to support both types of line cards in the same chassis. Of course, these distributed architecture line cards can be installed at any time prior to beginning the migration in the plant, and any removal of RF upstream or downstream line cards can follow the deployment of any number of distributed architecture line cards or nodes.

Turning our attention to the plant, it is similarly possible to migrate regular nodes to distributed architecture nodes in any sequence. As an example, what follows is a sequence of steps where a single node is

gradually converted from standard HFC to distributed architecture.

Figure **11** depicts a single HFC node connected to a CCAP device.

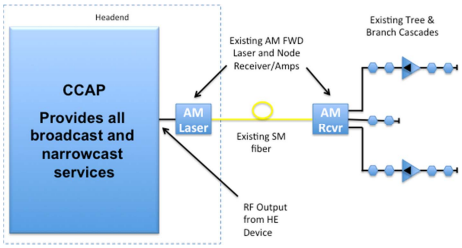


Figure 11: Single traditional HFC node

Figure **12** shows how the HFC node would be converted to RPN while the rest of the HFC network remains unchanged. The distributed architecture line card in the CCAP would have been deployed in the headend a priori, and even the RPN could have been deployed before the day of the cut-over. Then, the day of the change the fiber cable could be swung in the headend from one AM laser to the CCAP distributed architecture card, and in the field from the HFC node to the RPN. Of course it is not necessary to perform the migration in such a fashion, but it would be possible if desired.

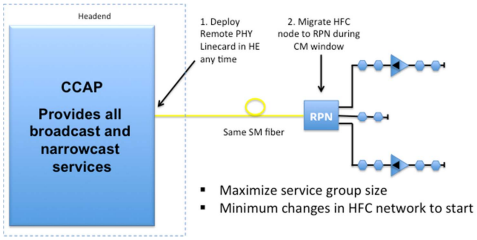


Figure 12: RPN deployment step 1

Figure **13** depicts a possible step 2 in the process, whereby additional RPNs are installed to segment the original service group further. These additional RPNs could be daisy chained from the original RPN node by taking advantage of the broadcast reuse feature, minimizing complexity in the deployment process.

NOTE: The example depicted is one in which fiber is run to every amplifier station. However, a more efficient segmentation scheme would include optimal placement of RPNs in an N+0 HFC architecture with some turn-around of passive components.

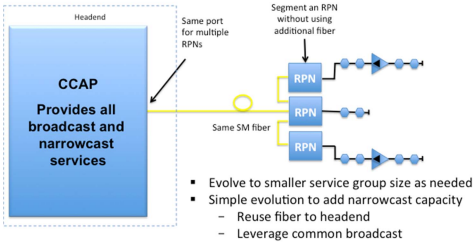


Figure 13: RPN deployment step 2

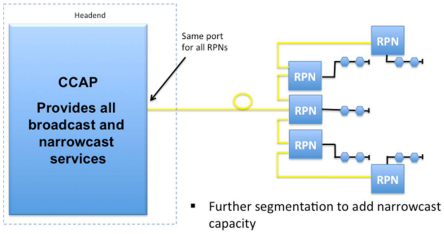


Figure **14** shows how further segmentation could take place by replacing the remaining amplifiers in the network with RPNs.

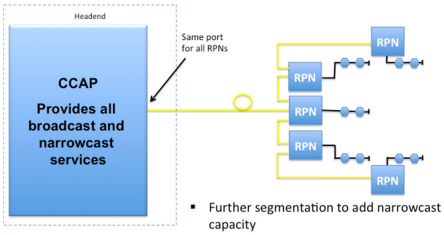


Figure 14: RPN deployment step 3

Figure 17 shows the RPN service group depicted above is segmented as additional narrowcast capacity is required. In this example, 2 of the RPNs from the distributed architecture service group shown in Figure 16 are split into separate service groups using separate CCAP ports.

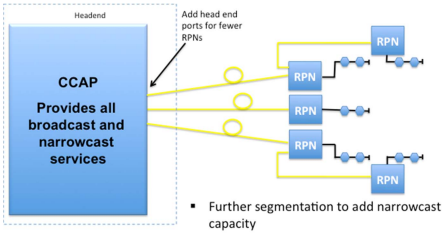


Figure 15: RPN deployment step 4

Eventually each of the RPNs could be connected to an individual CCAP RPN port. This would provide up to 10 Gbps of capacity to each RPN. This could, for example, be desirable to provide both RF and PON services from the RPN.

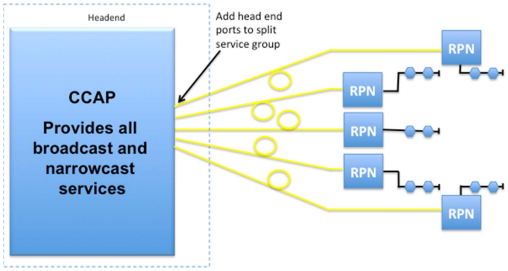


Figure 16: RPN deployment step 5

Similarly, the distributed architecture line card in the CCAP could be upgraded to support even more capacity as such capacity is needed and becomes cost effective. For example, the Ethernet link from the CCAP to the RPN could eventually be upgraded to 40 or 100 Gbps, both of which are already commercially available.

From Today to Virtual CCAP

As the network has to continue in operation through the transition, virtualizing the CCAP requires careful planning and a sensible staged process. As with roads, where cars must be kept moving during any lengthy highway reconstruction, in the network customer traffic must continue flowing day after day. In a sense, while road work is visible to car drivers, in a network the modifications remain invisible to the end user.

One way to do so is to migrate individual functions, one at a time. So, one must develop a list of the functions that would be virtualized, and this list would be prioritized, such as on the basis of complexity of implementation and benefit. Those features with the lowest implementation complexity and the highest benefit would be prioritized higher in the list, and consequently implemented first.

In DOCSIS 3.1, one of the functions that would rise to the top of any such list is Modulation Profile Management (MPM). This is because MPM will take time to be implemented by vendors in a CCAP chassis, but implementing externally via virtualization could be quite simple. In the process, its benefit to operators is quite significant since it would enable bigger benefits from DOCSIS 3.1.

Over time, implementing virtualization of the various functions of the CCAP would lead to a completely virtualized CCAP platform. Such a platform would be more easily scalable than CCAP platforms are today, where segmentation of service groups requires the addition of more chassis in a linear relationship fashion.

In addition, and perhaps more importantly, virtualizing the CCAP will enable the development of functionality, and improvements to such functionality, to occur much more rapidly than it is possible to do today.

Conclusions

Capacity for narrowcast services in HFC networks continues to increase. MSOs continue using well-known techniques for increasing capacity through service group segmentation.

With the advent of DOCSIS 3.1 MSOs will have the opportunity to use much higher modulation orders, which will result in more efficient use of RF spectrum.

One area where improvements can be made is in the optical link from the headend to the node.

There are several approaches to convert the forward link to digital, known as distributed architectures. These distributed architectures offer many benefits, including: improved performance, enable longer distances, and improved reliability.

Distributed architectures make it possible to increase headend equipment density. This results from several factors, such as: combined upstream/downstream line cards, increased port density, and “daisy chaining” RPNs. The combined effect of these density factors results in a density gain of 8x to 18x.

In addition, the use of an Ethernet link between the CCAP and the RPN makes it possible to integrate fiber-based services into a single consolidated network. For example, it is possible to implement a PON OLT interface from the RPN, which being within 1-2 kilometers from the customer premise would enable higher splits and/or the use of lower cost optics.

The migration to distributed architectures should be a very smooth one, requiring no synchronization between network and customer premise changes. Migration could begin with the deployment of CCAP

distributed architecture line cards, followed by migration to digital nodes on a case-by-case basis.

Finally, virtualization of the CCAP will offer significant advantages, such as better scalability and faster evolution of access network functionality.

DOCSIS 3.1 Multicast Profile Management Mechanism

Yanbin Sun, Tao Ouyang, Hesham ElBakoury
Huawei Technologies Co.,Ltd.

Abstract

DOCSIS 3.1 introduced many new features into the access network, including variable bit loading across a channel and the use of multiple modulations profiles for downstream and upstream channels. For multicast, the CMTS should attempt to maximize the link utilization by only sending packets to a multicast group on a single profile when a CM joins or leaves. Currently, there is no definition about the multicast profile management mechanism.

In this paper, the DOCSIS multicast profile management mechanisms based on different implementations were proposed,which can be regarded as the supplement methods of the DOCSIS 3.1 and SDN-Cable TR about the downstream profile management.

INTRODUCTIONS

DOCSIS 3.1 Profile Management

DOCSIS 3.1 [1] introduced many new features into the access network, including variable bit loading across a channel and the use of multiple modulations profiles for downstream and upstream channels. Using multiple profiles can increase throughput per CM, maximize network capacity, and minimize errors. As a new feature, modulation profile management is part of the CMTS but not directly tied to the DOCSIS protocol. This function can be deployed in the CMTS as an internal module or moved out of the CMTS as an external 'Application'. The DOCSIS 3.1 profile management application (PMA), as an external 'APP', was introduced by the SDN Architecture for Cable Access Networks (SDN-Cable) Technical Report [2] firstly and also published in one white paper at SCTE 2015 from Karthik Sundaresan [3], CableLabs. But in these materials, except for the downstream profile management mechanism, there is no definition for the multicast profile management mechanism.

For multicast, based on the description of DOCSIS 3.1 specification [1], the CMTS should attempt to maximize the link utilization by only sending packets to a multicast group on one single profile. The CMTS should use the highest bandwidth profile common to the CMs which are members of the multicast group. When a CM joins a multicast group the CMTS determines if the new CM can support the existing profile in use for the session; if not then the CMTS will have to move the session to a lower common profile which all group members can support or be forced to replicate the multicast on multiple profiles. When a CM leaves a multicast group the CMTS determines if the remaining group members can support a higher bandwidth profile than is currently in use for the session; if yes, then the CMTS MAY move the session to the higher bandwidth profile.

Internal Profile Management Module

The profile management can be deployed in the CMTS as an internal functionality module, which is used to determine the best modulation profiles for each downstream and upstream channel, given the channel characteristics seen by each CM on the network.

External Profile Management Application

In the SDN-Cable Technical Report [2], there are two kinds of application that an MSO can create and deploy:

1. Applications that provide service to the end customer;
2. Applications that provide network services to the MSO.

DOCSIS 3.1 Profile Management belongs to the second class of applications that can be enabled by an SDN infrastructure. Some DOCSIS features that are part of the CMTS but are not directly tied to the DOCSIS protocol interface between the CMTS and CM, include variable bit loading, multiple modulations profiles and upstream probes. Refer to the Figure 1, it is the SDN-based DOCSIS 3.1 Profile Management Application architecture, which is not directly tied to the DOCSIS protocol.

The PMA is responsible for gathering the data it needs to make profile decisions. It interacts with the CMTS through the SDN controller to initiate modulation profile tests, provide new or optimized modulation profiles, and provide suggestions or commands to use these modulation profiles. At high level, the PMA makes requests to the CMTS to get information from the CMTS, which in turn sends out MMMs to the CMs to collect the needed information and send it back to the PMA. The PMA also makes recommendations on the profiles to the CMTS, which the CMTS uses to configure CMs to use at an appropriate time.

For the detailed description about PMA please refer to the [2, 3], which is also designed to help Cable Operators determine the best modulation profiles.

In this paper, the authors only focused on the multicast profile management which deployed internally or externally.

MULTICAST PROFILE (MP) MANAGEMENT

MP Optimization Trigger condition

Based on the DOCSIS3.1 Specification [1], DOCSIS 3.1 is backward-compatible with some equipment built to the previous specifications. DOCSIS 3.1-compliant CMTS seamlessly support DOCSIS 3.0, DOCSIS 2.0 and DOCSIS 1.1 CMs. Also, DOCSIS 3.1 CM can work as DOCSIS 3.0 mode. So, when a CM joins/leaves a multicast group the CMTS needs to determine its working mode, which is DOCSIS 3.0 mode or DOCSIS 3.1

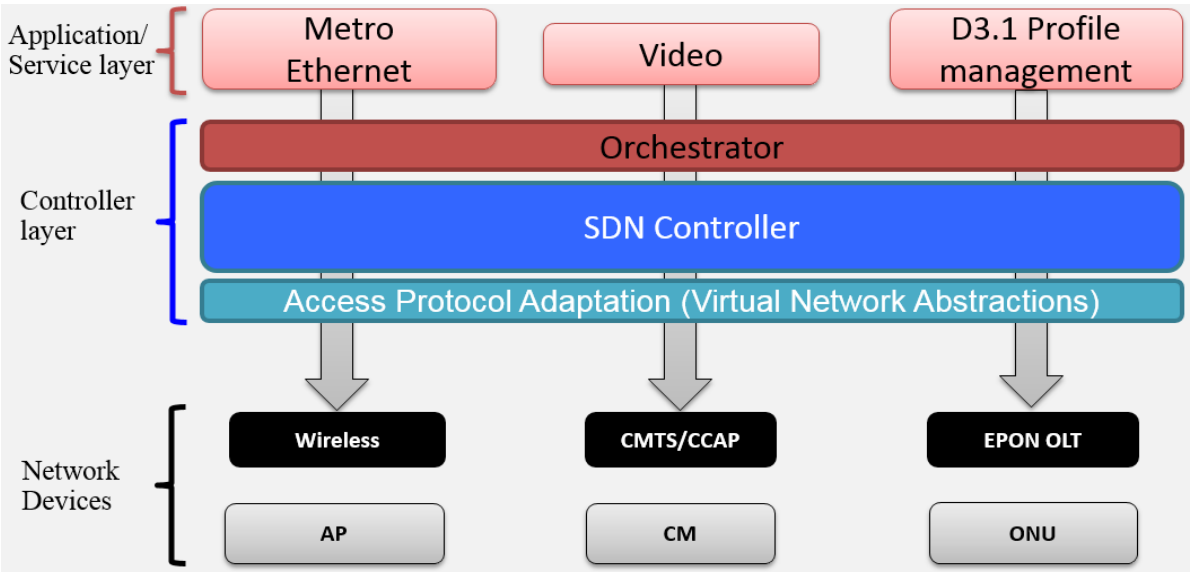


Figure 1. PMA Architecture

mode.

If the CM is working on DOCSIS 3.0 mode, the CMTS will allows it to continue to join/leave the multicast group without doing the profile optimization, otherwise, if the CM is working on the DOCSIS 3.1 mode with OFDM downstream channel, the CMTS needs to determine whether it needs to do multicast profile optimization or not, because each CM can support one or more multicast service which may belong to one or more multicast group, as showed in Figure 2.

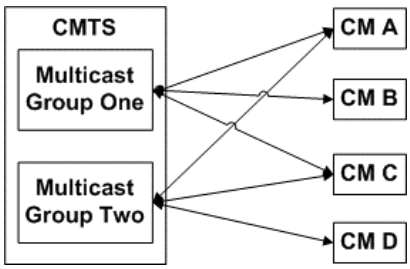


Figure 2. Multicast Groups in CMTS

For the scenario in Figure 2, different multicast groups includes different CMs, CM A, CM B, and CM C belong to multicast group one, and CM A, CM C, and CM D belong to multicast group two, assuming that all the CMs are working on DOCSIS 3.1 mode with the OFDM downstream channel.

Figure 3 shows the case where CM A does not belong to any multicast group.

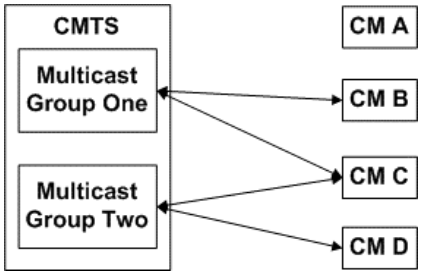


Figure 3. Multicast Groups in CMTS

Now, as shown in Figure 4, client 1 connected to CM A wants to join multicast group one which already includes CM B and CM C using one common multicast modulation profile. At this time, CMTS detects that this is the first client connected to CM A which wants to join multicast group one, then, CMTS will start to do the profile optimization for CM A, CM B, and CM C which belong to the same multicast group.

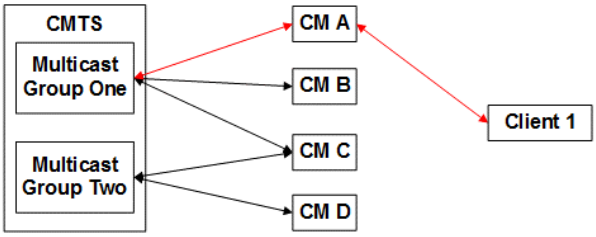


Figure 4. Client Joins Multicast Group

As shown in Figure 5, Client 2 is another client that is connected to CM A and wants to join multicast group one.

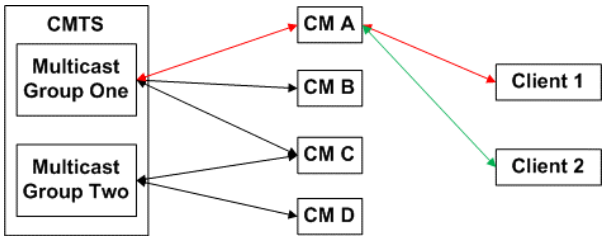


Figure 5. Client Joins Multicast Group

At this time, the CMTS detected that Client 2 connected with CM A wants to join the same multicast group as Client 1, so, the CMTS will not need to do the multicast profile optimization, because the CM A has already joined that multicast group.

In another scenario, if Client 2 wants to join multicast group two, as shown in Figure 6, CMTS detects that this is the first client connected to CM A which wants to join multicast group two, then, CMTS will start to do the multicast profile optimization for CM A, CM C, and CM D which belong to the same multicast group.

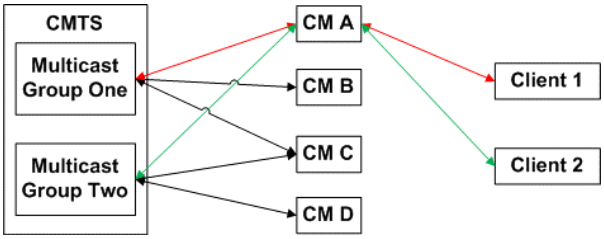


Figure 6. Client Joins Multicast Group

Based on the above analysis, if the CMTS wants to do the multicast profile optimization when a CM joins one multicast group, the following prerequisites need to be satisfied:

1. The CM is working on DOCSIS 3.1 mode with OFDM downstream channel;
2. The CM receives the first join request for the multicast group from one of the clients connected to it.

When a CM leaves a multicast group, the following prerequisites of the multicast profile optimization need to be satisfied:

1. The CM is working on DOCSIS 3.1 mode with OFDM downstream channel;
2. Last Client connected to the CM leaves the multicast group.

Internal Profile Management Mechanism

In this section, the profile management is internally deployed in the CMTS, and assumes that CMTS needs to start to do the multicast profile optimization when a CM joins the multicast group. As showed in the Figure 7, the procedure of the profile management (PM) is as follows:

1. The PM module needs to know the identity of the CM (MAC address, etc.) and the identity of the multicast group. The identity of CM can be obtained from other modules, such as the multicast management module. Using the identity of CM, the PM module can obtain the multicast profiles which the CM supports. And using the identity of the multicast group, the PM module can obtain the multicast profile which is used by current multicast group.

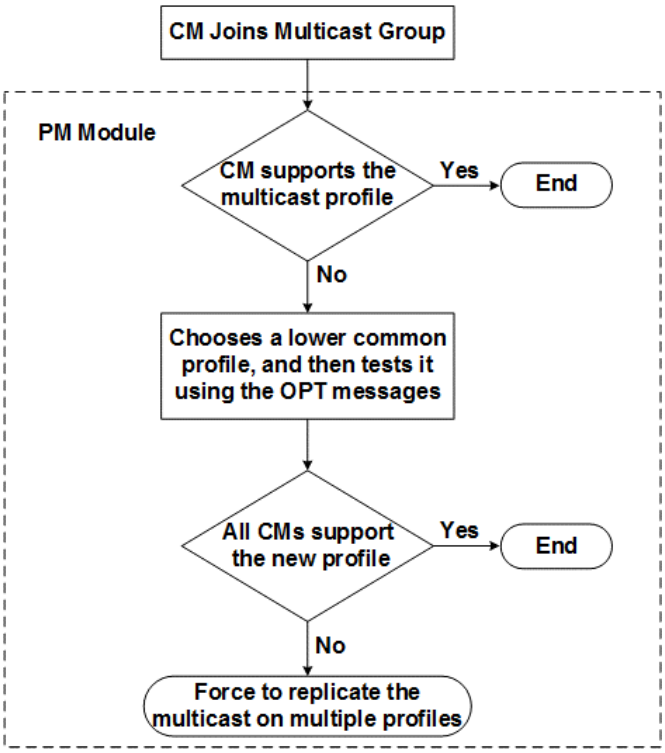


Figure 7. Internal Profile Management (CM Join)

2. Then, the PM module tries to compare the multicast profiles of this CM with the multicast profile which is currently in use by the current multicast group. If this CM can support the current multicast profile, the multicast profile optimization procedure is finished. Otherwise, goes to next step.
3. If this CM cannot support the current multicast profile, the PM module will choose a lower common profile for the multicast group, and then tests it for all CMs in the multicast group using the OPT messages in the DOCSIS 3.1 specification [1]. If all CMs can supports the new profile, the CMTS selects the new profile for the multicast group members, and the multicast profile optimization procedure is finished. Otherwise, goes to next step.
4. If all CMs cannot supports the new profile, the CMTS will force the multicast management module to replicate the multicast on multiple profiles.

When a CM leaves a multicast group, as showed in Figure 8, the PM module determines if the remaining group members can support a higher bandwidth profile than is currently in use; if yes, then the CMTS will want to move the session to the higher bandwidth profile. Based on testing the new profile, if the remaining CMs can support the new higher bandwidth profile, the CMTS will select the new profile for the multicast group members, and the multicast profile optimization procedure is finished. Otherwise, the CMTS will continue to use the profile which is currently in use.

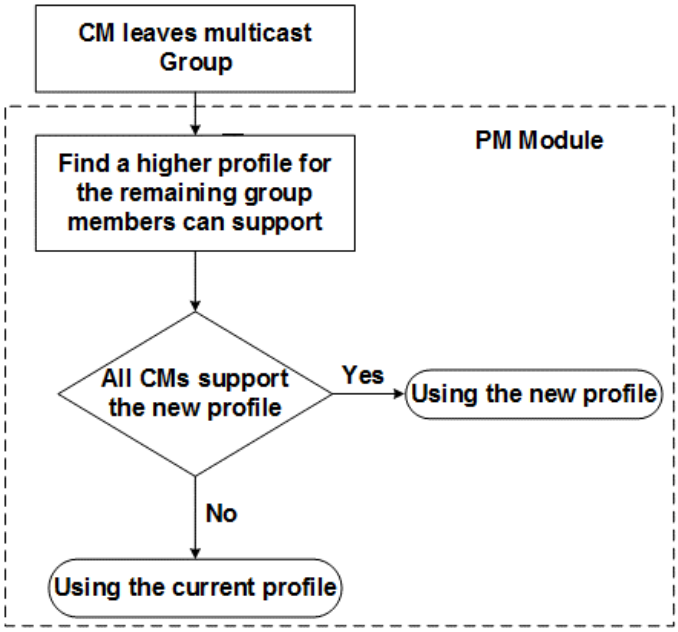


Figure 8. Internal Profile Management (CM Leave)

External Profile Management Application

In [2,3], the profile management functionality is implemented by an external application based on the SDN/NFV cable access architecture. This application can be deployed outside of the CMTS, as shown in Figure 1. But in [2, 3], there is no definition for the multicast profile management.

When one CM wants to joins or leaves one multicast group, if the CMTS wants PMA to do the multicast profile optimization, the CMTS needs to notify the PMA of which CM wants to join or leave the multicast group.

Then, the PMA will work the same way as an internal PM module, to initiate the profile optimization process. After that, PMA will send the new profile to the CMTS to it. If the testing results are acceptable, then PMA will ask the CMTS to assign the profile to all CMs in the multicast group, as showed in Figure 9.

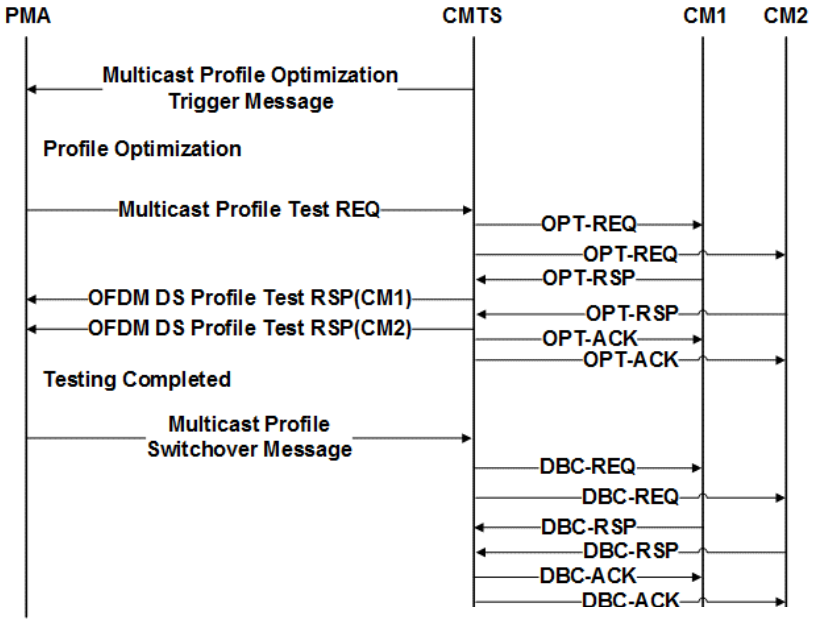


Figure 9. PMA Profile Optimization

In the SDN-Cable TR [2], the downstream modulation profile messages for the PMA-CMTS interface has been defined. As showed in Table 1.

In the Table 1, those messages are not enough for the multicast profile management. Some of them can be reused and some of them need to be modified to accommodate the multicast profile management.

In this section, the high level interactions between a PMA and the CMTS about the multicast profile information were proposed. The interactions messages as follows:

1. CM Join/Leave Descriptor includes four elements:
   1) Downstream Channel ID;
   2) Multicast Group ID;
   3) Indicator for whether the CM joins or leaves the multicast group.
   4) List of CM MAC address;
2. This message is sent from CMTS to PMA to indicate whether one CM or a list of CMs have joined or left the multicast group;
3. OFDM Downstream Multicast Profile Test Request.

   This message is for the PMA to request the CMTS send an OPT-REQ messages to the CMs in the multicast group to test a modulation profile. In order to reuse the messages of OFDM Downstream Profile Test Request in Table 1 for the multicast profile testing, some modification and limitation about the CM MAC address and IfIndex need to be satisfied, as follows:
   1) CM MAC address list;
      A Multicast CM MAC Address indicates a request to send the OPT-REQ to all CMs in that multicast group;
   2) IfIndex:
      Needs to be ignored by all CMs in the multicast group;

Table1. PMA-CMTS Downstream Modulation Profile Messages

Message	Description
Downstream OFDM Channel Descriptor	Conveys the configured channel parameters for a downstream DOCSIS 3.1 channel.
Downstream Profile Request	Either a request from the CCAP for a new profile or a request from PMA for the details of an existing profile.
Downstream Profile Descriptor	Provides the configuration details of a modulation profile.
Downstream Spectrum Request	Request from the PMA for the RxMER values for a channel from a given CM.
Downstream Spectrum Desciptor	Conveys the per subcarrier RxMER measurements for a channel from a given CM.
Downstream Profile Test Request	Request from the PMA for the CCAP to test a specified modulation profile.
Downstream Profile Test Response	Conveys the results of the test of a modulation profile on a specified CM.
CM-to-Profile Assignment Request	Request for a list of CMs that are assigned a specified modulation profile.
CM-to-Profile Assignment Descriptor	Provides a list of CMs that are or are to be associated with a modulation profile.
Profile-to-CM Assignment Request	Request for a list of the profiles that are assigned to a CM.
Profile-to-CM Assignment Descriptor	Provides a list of channels and modulation profiles that are assigned or should be assigned to a CM

3) Profile ID;
4) Op code;
5) OPT-REQ TLV;

4. OFDM Downstream Multicast Profile Test Response message can reuse the OFDM Downstream Profile Test Response messages in Table 1.

5. Multicast Group Information Request, includes two elements:
   1) Downstream Channel ID;
   2) Multicast Group ID;

   When sent from the PMA to the CMTS, this message is used by the PMA to request that CMTS send the multicast groups information for a downstream OFDM channel. A Multicast Group ID of 0xFFFF indicates a request for all multicast groups information on the downstream OFDM Channel. A Downstream Channel ID of 0 and a multicast group ID of 0xFFFF, indicate a request for all multicast groups information of all downstream OFDM channels on a given CMTS.

6. Multicast Group Information Descriptor includes three elements:
   1) Downstream Channel ID;
   2) Multicast Group ID;
   3) List of tuples (Multicast Profile ID, CMs Number, CMs MAC Address);

   When sent from the CMTS to the PMA, this message is used to inform the PMA of that multicast group information on the downstream OFDM channel.

   When sent from the PMA to the CMTS, it describes the CMs to multicast profiles assignment as recommended by the PMA. The message may be sent with or without Multicast Group Information Request message.

   In one multicast group, if all the CMs use the same multicast profile, this message only includes one tuple, in which the downstream channel ID could be 0, and the CM MAC address could be the multicast MAC address; if using the

different multicast profiles, this message will include several tuples.

SUMMARY

In this paper, the DOCSIS multicast profile management mechanisms based on different implementations were proposed, which can be regarded as the supplement methods of the SDN-Cable TR about the DOCSIS 3.1 downstream profile management. For the external case, the authors proposed the multicast profile management interface which is used for the PMA-CMTS profile management interactions.

REFERENCES

[1] CableLabs, DOCSIS 3.1 MAC and Upper Layer Protocols Interface Specification, CM-SP-MULPIv3.1.

[2] SDN Architecture for Cable Access Networks Technical Report, VNE-TR-SDN-ARCH, 2015.

[3] Karthik Sundaresan, Applications Architecture for DOCSIS 3.1 Deployments – DOCSIS 3.1 Profile Management. EXPO'15, Oct. 13-16, New Orleans.

DOCSIS 3.1 OVERDRIVE: DYNAMIC OPTIMIZATION USING A PROGRAMMABLE PHYSICAL LAYER

Saifur Rahman
Senior Engineer
Comcast

Jason Schnitzer
Founder & CEO
Applied Broadband

Joe Solomon
Principal Engineer
Comcast

Dr. David Early
Data Scientist
Applied Broadband

Abstract

DOCSIS 3.1 promises to dramatically increase data rates of the HFC infrastructure through the introduction of downstream Orthogonal Frequency Division Multiplexing (OFDM) and upstream Orthogonal Frequency Division Multiple Access (OFDMA).

However, in order to realize the full potential of these powerful new modulation methods, DOCSIS 3.1 channel parameters must be dynamically adjusted to optimize for the greatest possible throughput given the time varying error performance of the downstream signal path.

This paper describes an experimental software-defined profile optimizer for DOCSIS 3.1 networks. In it, we describe the theory of operation driving automated data collection, optimization analysis, and programmatic control of the DOCSIS 3.1 physical layer.

To illustrate our approach and learnings to date, we'll examine the current state of the art of D3.1 systems to support profile optimization capabilities; which form the basis of an early implementation of SDN architectural principles in cable broadband networks.

BACKGROUND

This section explores the capabilities of DOCSIS 3.1 modulation profiles and builds upon earlier works describing these protocol features [1][2][3].

Modulation Profiles

Traditionally, modulation parameters in the DOCSIS physical layer were statically configured in the CMTS and rarely, if ever, modified. A DOCSIS 3.0 (D3.0) [4] channel provides 6 MHz (North America) of data transport at a fixed modulation order for all cable modems (CMs) that can receive the channel.

In DOCSIS 3.1 (D3.1) [5], significant flexibility in the size and modulation of a channel has been introduced. Channels range in size from 24 MHz to 192 MHz in the downstream and 6.4 MHz to 96 MHz in the upstream. The D3.1 channel is separated into discrete subcarriers; each subcarrier is only 25 or 50 kHz wide, with thousands of subcarriers per channel. Each subcarrier can have a separate modulation order applied, allowing multiple modulation orders to be applied across the spectrum band. As the constellation density increases (from 256-QAM to 4096-QAM or better), the number of bits

transmitted per hertz (bits/Hz) also increases. However, higher modulation orders are more susceptible to channel interference than lower modulation orders. Therefore, higher modulation orders (e.g., 4096-QAM) should be used on subcarriers in areas of spectrum where channel error performance is favorable, and less efficient modulation orders (e.g., 256-QAM) should be used on subcarriers in areas of spectrum where channel error performance is less favorable.

The modulation order applied to a subcarrier is managed through the creation of modulation profiles. Each modulation profile defines, on a per-subcarrier basis, the modulation order for all the subcarriers in that channel. A downstream channel can support up to 16 modulation profiles; an upstream channel can support 8. Why so many? Because the CMTS can assign each CM that uses that channel a different modulation profile from the set of profiles configured (to a maximum of 4 per CM). More robust modulation profiles (e.g. lower modulation order) can be used for important, low bitrate data to ensure it gets through; profiles with more aggressive modulation orders can be used to send data at the highest possible bit rate for that CM. This allows modulation profiles to be designed in a way that accounts for the downstream channel characteristics and performance at each CM. Profiles can be defined to use higher order modulations in parts of the spectrum where the signal-to-noise ratio is high, and use more robust modulation orders in parts of the spectrum where channel conditions are not as optimal. This approach optimizes the amount of data that can be transmitted and received, with a goal of providing the highest data rate possible to each individual CM.

The CMTS decides which modulation profile to apply to a channel for a given CM based on two factors: The quality of the downstream channel signal measured at the CM receiver, and the performance of the forward error correction (FEC). If the quality of the signal being received is high, the CMTS can choose a modulation profile for the CM that uses higher order modulation. If the quality of the signal is less than optimal, a modulation profile that uses more robust modulation orders can be chosen.

The CMTS also monitors the ability of the receiver to correct codeword errors using FEC. A high incidence of FEC errors may indicate an overly ambitious modulation profile; a complete lack of FEC errors suggests a higher performing modulation profile can be used.

One interesting property of D3.1 is the fact that there will be FEC correctable errors even during normal (acceptable) operation. The use of FEC coding gain to push higher modulation orders is made possible by the system's ability to adapt to a lower bit loading profile if the coding gain can no longer provide unerrored frames. This is a key mechanism in D3.1 to provide the maximum capacity of the channel while also running at a lower operating margin.

Modulation Profile Design

In current D3.1 implementations, the CMTS does not participate in the design or creation of modulation profiles – this task is left to the operator. This is where the challenge of D3.1 optimization begins. Determining the modulation profiles that can be used by the best performing CMs and the worst performing CMs is relatively straightforward: The profile for "top tier" CMs, where plant performance is high across spectrum, uses high modulation orders (increased bit rate) across the spectrum. The profile for "bottom tier" CMs, where the plant performance is consistently poor, uses more robust modulation orders (lower bit rate). The CMTS can then assign these profiles to the top and bottom CM performers accordingly.

What about the middle performance tier, where the majority of CMs reside? Given that most service groups show a normal (Gaussian) distribution of downstream MER, setting one modulation order across all

subcarriers is not operationally possible, or is suboptimal in that the lowest common MER value dictates the modulation order in that service group. Therefore, some means of dynamically adjusting modulation profiles is necessary for optimal network throughput. The number of profiles and the bit loading of the profile are based on several factors which are discussed in this paper.

To create effective and efficient modulation profiles across this population, the quality of the signal must be analyzed across the entire downstream spectrum of the channel for all CMs. Patterns in the quality of the signal across the spectrum should be identified and evaluated using criteria including:

- CM grouping: Which CMs have similar enough performance (commonality) across the spectral width of a channel to share the same modulation profile?
- Topological grouping: Are there areas of the plant (in certain geographic regions) where downstream spectral performance is not as strong as in other locations of the plant?
- Diurnality: Are there daily patterns where plant performance degrades or improves?
- Seasonality: Are the plant characteristics consistent between summer and winter, considering the different environmental factors?

Identifying these patterns in signal performance helps to design profiles that work for a large number of CMs on a given channel. Because a given CMTS serves tens of thousands CMs, there is a large amount of data to gather and analyze.

In addition to designing modulation profiles based on CM performance, we must consider the processing load imposed on the CMTS from having a large number of active profiles per channel. CMTS scheduler efficiency can be impacted by the number of profiles that are in use on a channel, especially on wide channels. What is a 'reasonable' number of profiles to have active at any given time? What is the number of profiles that a CMTS can support without impacting overall system performance?

Modulation Profile Optimization

The questions described in the previous section are difficult to answer and, given that conditions in the plant are not static, present challenges to keep the answers fresh and accurate. This necessitates a system that dynamically manages the modulation profiles that are associated with a D3.1 channel and to which CMs each are assigned. The D3.1 profile optimizer will be responsible for the following tasks:

1. Develop a set of modulation profiles for a channel that allows the CMs on that channel to use the most efficient bit loading across the channel.
2. Make recommendations for profile assignment, communicating to the CMTS the best fit profile for each CM.
3. Monitor CM performance using a given profile over time and change profile assignment recommendations, as appropriate.
4. Monitor channel performance across the entire CM population, as measured at each CM, and adjust existing profiles or create new profiles to adapt to those conditions.

This profile optimizer will continuously monitor plant and CM performance. A mature system will consider performance over time, identifying intermittent anomalies and recurring interferers. In addition to designing modulation profiles and managing their assignment in such a way that the best possible bit loading can be achieved, the profile optimizer will also be able to identify parts of the plant where operation is sub-

optimal, flagging issues for operational teams to investigate and fix. The system augments existing Proactive Network Maintenance (PNM) [6] toolsets, optimizing performance where there are issues while also shedding light on those issues.

The remainder of this paper focuses on work completed to date to define an experimental profile optimizer and create the mechanisms and programmatic interfaces necessary to control the management of modulation profiles dynamically. Preliminary work focuses on the downstream, as many operators are focusing on deploying D3.1 downstream channels first.

MODULATION PROFILE OPTIMIZATION USING PROGRAMMABLE NETWORK PRINCIPLES

The Role of SDN

Central to the concept of SDN is the idea of the separation of the *control plane* (how to direct traffic) from the *data plane* (forwarding of traffic according to control plane directives). The advantages of this new architectural approach have been broadly researched and documented [7][8][9], but commonly describe accelerating network innovation while enabling greater resource efficiencies within network deployments (e.g., network and system virtualization).

In a SDN configuration, the control plane asserts all influence over the configuration and state of the all data plane network elements, both intermediate (routers, switches, CMTSs) and endpoints (CM devices). The control plane realizes this via a well-defined Application Programming Interface (API) [Feamster].

Traditionally, DOCSIS CMTS implementations have contained *both* the control plane and data plane tightly coupled and embedded in a single integrated network element. However, specific management tasks such as complex D3.1 modulation profile management, taken the fullest potential of Shannon's limits, are both computationally rigorous and data intensive. [1] Removing these complex profile optimization tasks and the long term data management function from the CMTS reduces the burden on the CMTS focusing resources on its principal role as a termination system within the data forwarding plane.

Figure 1 illustrates this conceptual separation of control and data planes in the context of a D3.1 network supporting an OFDM profile optimization use case.

A second concept central to SDN is the idea of abstracting network layer complexity, like different types of devices and protocols, through a single interface. The control plane interacts with the data plane through a well-defined API. This contains the complexity of vendor-specific implementations and a plurality of diverse protocols within a software layer de-coupled from higher-level application logic.

In developing this experimental D3.1 programmable modulation profile management system, the following SDN principles were successfully implemented:

- Separation of the control plane and the data plane.
- Abstraction of network complexity in terms of both vendor-proprietary CMTS features and low level network management instrumentation through the use of a CMTS Vendor Abstraction Layer (CVAL).
- Interaction between control and data planes through the application of a programmatic RESTCONF [10] API based on a standard YANG [11] data model.

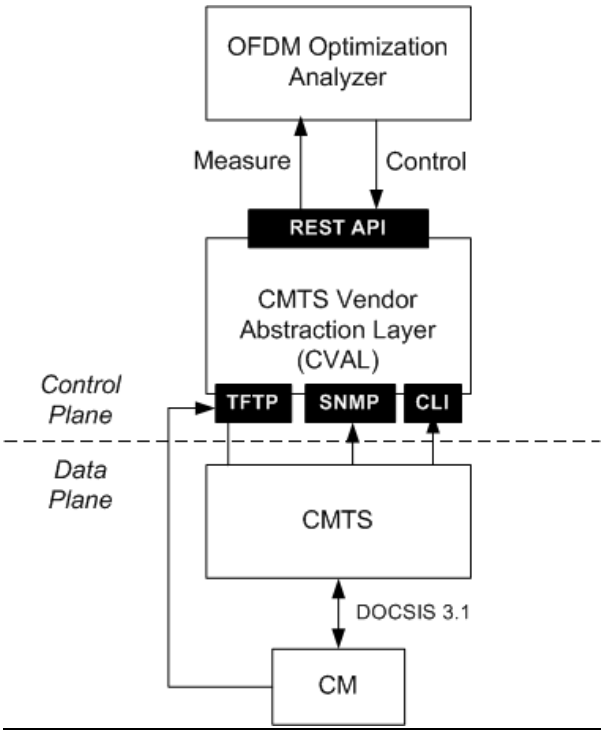


Figure 1: An experimental D3.1 profile optimizer architecture

Theory of Operation

The profile optimizer implements the key functions described in this paper in three iterative steps:

1. **Data Collection:** Key channel performance metrics are derived from network data collected for each channel and CM device.
2. **Optimization Analysis:** Per-CM downstream channel error performance is evaluated using key metrics and network optimization algorithms combined with a profile configuration policy.
3. **Profile Assignment:** Once optimization analysis indicates the need for a CM channel profile to be changed, then assignment occurs through the control plane.

Details underpinning these three conceptual steps are provided below.

Data Collection

Periodically, data is collected from the D3.1 network (both CMTS and CM) to gather key metrics. This is performed using two forms of low-level network layer instrumentation:

SNMP: Simple Network Management Protocol, using the data models described in the D3.1 Management Information Base (MIBs) defined in [5]. Though it is largely held that more contemporary network management protocols exist, SNMP remains widely adopted and universally supported by CMTS vendors. Collection of key metrics described in this paper via SNMP polling is performed on both a periodic and real time basis. SNMP is also used to configure the CM's measurement and uploading of data delivered via TFTP.

TFTP: An important tool to gather channel performance metrics is the Trivial File Transfer Protocol (TFTP), which is used to gather file-based measurement data on downstream channel performance on a per-OFDM-subcarrier basis. Though TFTP was originally defined by the IETF in 1981 [TFTP-RFC], it remains the preferred choice within the DOCSIS specification for the retrieval of CM-sourced data describing PNM file contents as defined in [5].

The data collected to derive key metrics is discussed later in this paper.

Optimization Analysis

Optimization analysis automates the evaluation of CM downstream performance based on key metrics derived from the network performance data collected. Operational thresholds defined for the range of acceptable values are compared to those calculated from measurements. If it is determined that CM's current profile is sub-optimal, the parameters of the desired profile are forwarded to the control plane for CM assignment.

Profile Assignment

CLI: The Command Line Interface (CLI) exposes vendor-specific instrumentation using a remote shell connection. In this system, the CLI is used to instruct the CMTS which profile to assign to a given CM; this interface will be used until such time that the standards-based interface is available. At the time of this paper's writing, the CableLabs PMA working group is evaluating different options for data models and protocols that together define the behavior of a standard CMTS API.

Key Metrics

To effectively design modulation profiles the health of the plant has to be understood. D3.1 defines new MIB objects to provide visibility into plant health. The primary metrics of use for the evaluation of downstream channel error performance are:

1. Receive modulation error ratio (MER)
2. Receive power
3. Forward Error Correction (FEC) statistics

MER: Due to the size of downstream D3.1 channels, an average MER reading for the channel is not an accurate measurement of a channel's health; areas of degraded performance can be hidden when averaged with measurements of good performance, making the overall performance of the channel appear to be better than it actually is. For this reason, a method to measure and record the MER for each downstream channel subcarrier has been defined [5].

To make this measurement, the CM is commanded via SNMP to start measuring the MER of each subcarrier and write that value to a file saved in NVRAM on the CM. Once complete, TFTP can be used to transfer the file to a location accessible by the profile management system. The `docsPnmCmDsOfdmRxMerTable` of the `DOCS-PNM-MIB` provides the mechanism necessary to execute this measurement. The `docsPnmBulkFileControl` object is used to upload the resulting file. The D3.1 Operations Support System Interface Specification details how these objects are implemented and used [5].

Receive Power: Downstream receive power measurements taken by the CM are available on a per-channel basis. Like the MER measurements, an average receive power measurement across the entire width of a downstream channel is limiting because it does not provide a detailed enough view into areas of the channel spectrum where receive power is below operational thresholds, possibly requiring more a robust modulation order. For this reason, D3.1 provides receive power measurements per 6 MHz block within the channel's spectral width. These measurements are available via SNMP from the `docsIf31CmDsOfdmChannelPowerTable` in the `DOCS-IF31-MIB`. The D3.1 channel is divided into 6 MHz bands and a row entry with the measured power is provided for each.

The CM can also perform downstream spectrum analysis, providing the the energy level of the signal at each frequency within a specified frequency range. The CM is signaled to begin the measurement via the `docsIf3CmtsSpectrumAnalysisMeasTable`, which was extended for D3.1. The resulting measurements are written to a file on the CM that can be uploaded using TFTP.

FEC Statistics: Downstream FEC statistics in D3.1 are maintained on a per-CM, per-modulation-profile, basis providing insight into the performance of all active profiles. For each profile, the CM keeps a running count of:

- The total number of codewords received by the CM.

- The number of codewords that were corrected by the FEC.
- The number of codewords that were not correctable.

The FEC counters are sampled on a periodic basis; the amount that each counter increases indicates how the profile is performing. D3.1 uses a new scheme based on Low-Density Parity Check (LDPC) FEC, which is a more robust error correction algorithm compared to that used in previous versions of DOCSIS (Reed-Solomon). As a result, having a high number of corrected codewords indicates that the FEC is working well and that the modulation profile is operating near its performance edge. If the number of uncorrectable code words is high, then the modulation profile is likely not performing well and should be reevaluated.

The FEC counters are provided in the `docsIf31CmtsCmUsOfdmaProfileStatusTable` of the `DOCS-IF31-MIB`. Codeword statistics can also be collected over a period of time (either a 10 minute period or a 24 hour period) on a per-profile basis and written to a file within the CM that can be uploaded via TFTP. When collecting for a 10 minute period, the CM records codeword data every second for a total of 600 measurements. When collecting for a 24 hour period, the CM records codeword data every 60 seconds for a total of 1440 measurements. Each recorded entry is time stamped to indicate when the measurement was performed. Each record contains the total number of codewords, the number of correctable codewords, and the number of uncorrectable codewords received during that interval. The `docsPnmCmDsOfdmFecTable` of the `DOCS-PNM-MIB` is used to conduct these measurements.

The profile management system could collect all of the data discussed here directly from the network layer as required. However, this data is also collected by operators in the context of other network health and maintenance systems, so the data may already be available in an existing data store, queryable by the profile optimizer. The schedule for refreshing this data needs to be considered before this data can be used, since the age of the data could impact its usefulness.

STATE OF THE ART & CURRENT LIMITATIONS

Today's Implementations

At the present time, D3.1 systems are just beginning to become available from equipment vendors with early field deployments underway. Given the richness of the total D3.1 feature set, many operators have chosen a phased approach to deploying the new protocol's full capabilities. The current focus of initial deployments is to increase downstream channel capacity and service speeds by introducing a single OFDM channel to the forward spectrum. OFDM channels deliver data in a more spectrally efficient way (bits/Hz), allowing MSOs to get more bits through the plant by using higher modulation orders.

With compressed time-to-market for CMTS vendors to implement the core functions of the D3.1, the industry has yet to fully focus on developing value-add features such as a full PMA support within the CMTS. Implementations currently support 3 to 4 profiles with statically set bit loading. The modulation order assigned to a subcarrier cannot be changed while the profile is active. This number of profiles should be sufficient to adequately assign groups of CMs to profiles that provide capacity gains in these early deployments. Given the other limitations noted below, an increased number of profiles would not be currently useful.

A significant limiting factor in how many profiles are useful in practice is the bit loading supported by the CMTS. Most implementations today only support a limited number of square constellations. Non-square constellations or mixed modulation orders are

not supported. This limitation yields a step size of ~6bB MER between modulation orders, limiting the number of usable profiles as shown in Figure 2.

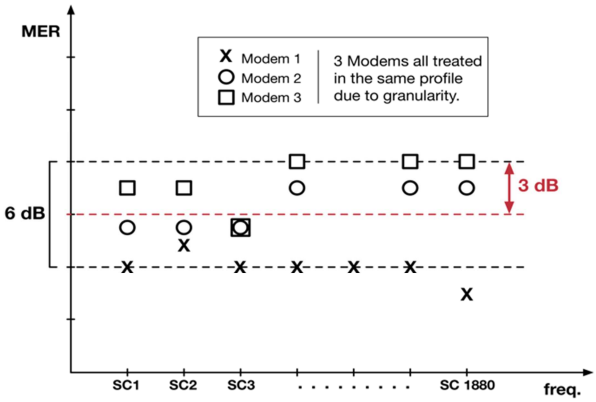


Figure 2: Limited Step Size Challenges

A second limiting factor is that the bit loading for each modulation profile cannot be assigned at the subcarrier level – implementations support 3 or 4 zones (spectral widths) of subcarriers that have the same modulation level. Techniques are under evaluation that would use a "pseudo-mixed modulation" that could produce a step of ~3dB: by using zones with different modulation orders, the average MER needed can be controlled. With this method, 4 profiles could support a total variance of 12dB in MER between CMs within a service group.

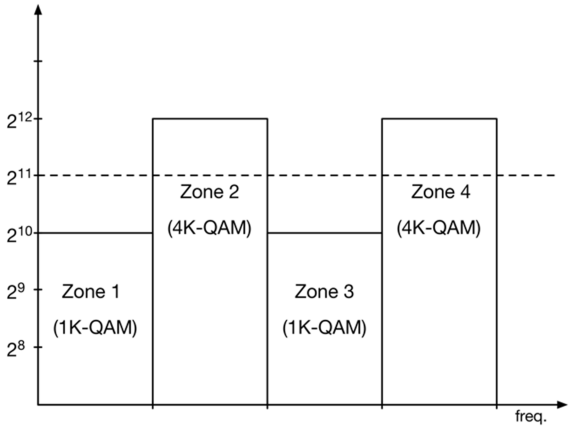


Figure 3: Using Zones to Provide Smaller Step Sizes

Possibly the largest obstacle to realizing full OFDM optimization in practice today is the relatively simple logic a CMTS uses to determine when to dynamically move a CM from one profile to another: the CMTS moves a CM if there are deteriorating RF conditions (move from a higher to a lower bit loaded profile) or if there are improving RF conditions (move from a lower to a higher bit loaded profile). In addition, current CMTSs may only support moving in one direction. This logic uses limited data, such as receiving a CM_STATUS message from the CM with a reason code of high uncorrectables. There is no trend analysis or use of prior events in determining if the status message indicates just a one-time anomaly or is indicative of a more systemic issue that requires a change; the CMTS pessimistically moves the traffic to the lower profile on receipt of the message. Current implementations do support a method to limit cycling of profiles (moving to a worse profile, then moving quickly back to the original) via simple protection mechanisms.

Support for generating a new profile at runtime based on is currently not supported. Today's D3.1 implementations support statically defined profiles that the operator preconfigures on the CMTS. There is no evaluation conducted by the CMTS to determine the suitability of the profile for existing plant conditions.

The final limitation we will discuss is the use of the test profile to optimize modulation profiles. The D3.1 specification [5] recommends that the CMTS supports at least 16 profiles for every channel and that CMs must support simultaneous reception of traffic on 4 profiles, not including the test profile. The specification allows any of the 16 profiles' bit loading configuration to change dynamically in order to better perform in varying channel conditions. The test profile must be supported by the CMTS and CMs; it allows the CMTS to use emulated test data to determine if the CM receiver can function optimally given a candidate profile.

Ideally, these protocol capabilities should be leveraged by CMTS implementations to optimize the generation of profiles, the bit

loading within each profile, as well as to determine the optimal profile for each CM. To determine if a profile is suitable for use, the CMTS should continually use the test profile and/or periodically probe the receiver's MER performance. Current implementations do not support either of these methods to determine the bit loading of a profile.

As mentioned above, the bit loading of the channel in current implementations is not applied at the subcarrier level. Non-uniform bit loading in zones of the frequency band is supported instead. Whether this is approach is sufficient depends on the frequency response of the channel. Currently, operators are deploying D3.1 downstream channels in areas of spectrum that have a relatively flat (ideal) response over hundreds of subcarriers where a broader resolution of subcarrier control is sufficient. The need for finer granularity on a per-subcarrier bitloading basis becomes critical as the operating band is pushed into plant roll-off spectral regions, where there is likely to be more variability in the response due to fewer subcarriers in these regions.

Ongoing HFC plant characterizations have shown that hundreds of megabits of capacity could be gained via proper bit loading of the channel in the roll-off region of the usable downstream spectrum. See Figure 4, which shows a sharp roll-off on the right side of the plot. This will become an important feature as more spectrum is needed to offer higher tiers of service.

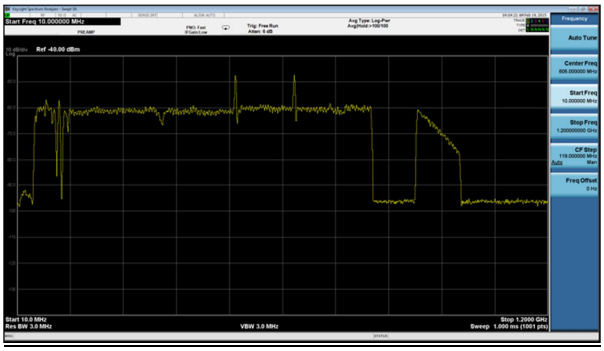


Figure 4: Bit Loading in the Roll-Off Region

Proposed Enhancements

Given these limitations of the current state-of-the-art, what things would be useful to have in the future? There are several functions needed within the CMTS, along with further development of the PMA.

The goal of an external profile optimizer is to remove the burden of optimizing plant performance from the CMTS. Instead, the CMTS could implement just the base mechanisms to react quickly to a problem with a CM and rely on an external entity to make longer-term decisions. For example, the CMTS could move the CM to a very robust profile that is known to work under the worst case MER when the modem reports FEC issues via the CM_STATUS message. Once error free performance is restored, the profile optimizer can determine a more optimal profile and provide this guidance to the CMTS. This requires a new interface between the CMTS and the profile management system. CableLabs is currently working to define new interfaces and APIs to allow such an system to interface with the CMTS.

Regardless of the level of embedded optimization capability, the CMTS will continue to play a part in the process of optimizing and managing profiles. The CMTS needs to provide information via a well-defined interface for things such as the service group to CM mapping, the profile to CM mapping, and the ability to act on suggestions from the system to change a CM's modulation profile.

Test profiles will continue to serve as a powerful method to understand how a candidate profile will perform on a CM. Again, the profile optimizer needs an interface to the CMTS in order to direct it to configure and use the test profiles. The FEC results from the test profile then need to be conveyed back to the system, where the data can be analyzed and the profile optimized. One of the main reasons the test profile is important is that it provides a direct measure of a particular bit loading configuration to support error-free performance. Since the CMTS only supports

one test profile, the profile optimizer will need to communicate different bit loading settings for the test profile to allow direct measurement of the effectiveness of different candidate profiles.

Future CMTS support of mixed modulation formats, as defined in the D3.1 PHY specification [12], would improve the granularity; mixed modulation will support a step size of 1.5 dB in MER, making it possible to differentiate between groups of CMs within a service group. Mixed modulation, coupled with the support of more profiles in the CMTS, can further optimize channel capacity.

FUTURE WORK

What additional steps are needed to develop a full-featured autonomous D3.1 profile optimization system?

Optimizing CM throughput performance requires analysis of several statistics over time in addition to evolving algorithms that evaluate key metrics. Given the early stages of D3.1 deployments, limited operational knowledge exists to fully understand all the behaviors of this new modulation scheme in a modern HFC plant. Having this optimization logic executing in an external application (rather than embedded in the CMTS) allows the algorithms to be adapted more quickly as new trends are understood without having to update the CMTS software.

For example, the following behaviors need to be better understood before the logic of a complete-the-profile optimizer can be fully developed:

- How often should the re-assignment of CM profiles occur?
- What MER value ranges and margins should we use for each profile to limit the cycling of profile changes?
- How frequently should test profiles be used to probe FEC performance?
- How does MER vary from service group to service group?
- What is the frequency response of the plant (MER variation across frequency)?

Although initial estimates for each of these questions are available, more data needs to be collected to characterize the long-term RF dynamics of the HFC plant in which OFDM channels are operating. There are several key D3.1 features that can aid in this data collection and characterization. Key performance from the receiver has to be collected with enough frequency over long durations. As discussed earlier, these are the key metrics that are needed:

- FEC uncorrectable performance for both data and control channels
- Receive power levels of all channels via full band capture
- Receive MER (RxMER) per D3.1 subcarrier

In addition to SNMP-based data collection from the CMTS, data can be captured directly from the CM via periodic TFTP uploads to a data store. This allows every CM in a service group to upload downstream channel error performance data with enough frequency and over a long duration of time without adversely impacting the CMTS performance. Trends in the data can then be analyzed and the characteristic of the plant better understood. In addition to the measurements discussed earlier, the spectrum capture feature of the CM provides the RF power level at the subcarrier resolution for the entire RF band. Also available via file upload is the receiver equalizer coefficient data; this can be used to estimate tilt, ripple, and group delay.

All of this data could be housed in a data store, either external or incorporated within the profile optimizer. This rich set of data will then need to be analyzed to develop the logic of the system. Since these stats are just

beginning to be analyzed and given the breadth of information, an iterative approach will be needed. The exact heuristics and algorithms in the profile optimizer will be in an ongoing state of development as more insight is gained.

CONCLUSION

In this paper we have discussed progress to date on developing a experimental system for the the optimization of modulation profiles for D3.1 networks. We have described the challenges in managing modulation profiles over a large set of CMs with a diversity of downstream channel quality performance levels. We described the use of an SDN-based approach to creating an experimental profile optimizer, and detailed the primary functions of this programmable system for use in D3.1 networks. Finally, we cataloged the obstacles that currently exist to develop and deploy such a system, given the relative maturity of D3.1 CMTS implementations. Given these limitations, we have defined areas of focus and proposed next steps for the evolution of a programmable D3.1 profile optimizer.

We see a great need for a programmable profile optimization software system as a critical component of the full D3.1 service delivery infrastructure. Without a robust system, the full potential of D3.1 will be difficult to achieve.

Work will continue to evolve the experimental implementation described in this paper in order to deploy a robust production system based on both the progression of D3.1 and on practical SDN principles.

ACRONYMNS AND ABBREVIATIONS

API – Application Programming Interface
CLI – Command Line Interface
CM – Cable Modem
CMTS – Cable Modem Termination System
CVAL – CMTS Vendor Abstraction Layer
DOCSIS – Data Over Cable Service Interface Specification
FEC – Forward Error Correction
HFC – Hybrid Fiber Coaxial
LDPC – Low Density Parity Check
MER – Modulation Error Ratio
OFDM – Orthogonal Frequency Division Multiplexing
OFDMA – Orthogonal Frequency Division Multiple Access
PMA – Profile Management Application
SDN – Software Defined Networking
SNMP – Simple Network Management Protocol
TFTP- Trivial File Transfer Protocol
YANG – Yet Another Next Generation

REFERENCES

[1] P. Sowinski. Leveraging external applications for DOCSIS 3.1 HFC plant optimization. NCTA 2014

[2] J. Chapman, The Power of DOCSIS 3.1 Downstream Profiles. NCTA, 2013.

[3] N. Pantelias .Monitoring the Variable Network: Impact of DOCSIS 3.1 on Plant Utilization Metrics. NCTA, 2015.

[4] DOCSIS® 3.0 specifications: Cable Television Laboratories, Inc. "CableLabs Specifications" [Online] Available: http:/cablelabs.com

[5] DOCSIS® 3.1 specifications: Cable Television Laboratories, Inc. "CableLabs Specifications" [Online] Available: http:/cablelabs.com

[6] DOCSIS Best Practices and Guidelines: Proactive Network Maintenance Using Pre-Equalization, CM-GL-PNMP-V02-110623. CableLabs, 2011.

[7] N. Feamster, J. Rexford, E. Zegura. The Road to SDN: An Intellectual History of Programmable Networks. IEEE ACM Queue, December 2013.

[8] Virtualization and Network Evolution Open Networking - SDN Architecture for Cable Access Networks Technical Report VNE-TR-SDN-ARCH-V01-150625, CableLabs.

[9] Bruno Nunes Astuto, Marc Mendon_ca, Xuan Nam Nguyen, Katia Obraczka, Thierry Turletti. A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks. Communications Surveys and Tutorials, IEEE Communications Society, IEEE, 2014, 16 (3), pp.1617 - 1634.

[10] A. Bierman, M. Bjorklund, K.Watsen. RESTCONF Protocol, draft-ietf-netconf-restconf-09. Internet Engineering Task Force, December 15, 2015.

[11] M. Bjorklund, Ed.. YANG – A Data Modeling Language for the Network Configuration Protocol (NETCONF). Internet Engineering Task Force, October 2010, RFC-6020.

[12] DOCSIS 3.1 PHY Physical Layer Specification, CM-SP-PHYv3.1-I02-140320, CableLabs

DOCSIS 3.1 Profile Management Application and Algorithms

Greg White and Karthik Sundaresan
Cable Television Laboratories, Inc. (CableLabs)

Abstract

DOCSIS 3.1 OFDM Profiles provide a wide range of modulation choices that can be used to fine-tune the CMTS's transmissions to get the best performance from the current network conditions. A well-designed, optimized set of modulation profiles allows a downstream channel to operate with a lower SNR margin, allowing a channel to operate at an overall higher throughput.

This paper describes methods for designing OFDM profiles and choosing the appropriate modulation orders for a profile. It answers the questions around which profile is appropriate for a CM and what is the optimal set of profiles to use across the an OFDM channel for a given set of CMs. This paper defines an objective function which can be used to calculate the gain in system capacity resulting from the use of multiple profiles, and then explores approaches to maximizing that objective function for a population of CMs. It proposes one such approach, which is referred to as the K-means Coalescation Algorithm, that appears to provide very good results with low computational complexity.

INTRODUCTION

The DOCSIS 3.1 (D3.1) specification fundamentally changes the nature of information delivery across the cable plant, and along with that, the way HFC networks will be maintained and managed. For the first time in the history of HFC, the downstream channel does not use a one-size-fits-all modulation scheme; rather, the modulation and forward error correction can be optimized based on actual plant conditions at individual devices, so that devices that receive an exceptionally clean signal can utilize very efficient high-order-modulation, while devices that have a degraded signal will use more robust modulation, all on the same channel.

The DOCSIS 3.1 toolbox provides a wide range of modulation choices that can be used to fine-tune the transmissions to get the best performance from the current network conditions. To manage the optimization of these settings across the population of devices, the CMTS uses the concept of downstream profiles. A downstream profile defines the modulation order (i.e., bit loading) on each of the 3840 or 7680 subcarriers across the OFDM channel. D3.1 provides for defining multiple downstream profiles, where each profile can be tuned to account for the specific plant conditions that are experienced by a set of CMs.

A well-designed, optimized set of modulation profiles allows a downstream channel to operate with a lower SNR margin, potentially allowing a channel to operate at an overall higher throughput. In addition, it can account for troubled devices by providing service even in situations where significant plant impairments exist.

The application that implements this optimization logic can be external to a CMTS, enabling the most efficient use of profiles across channels and CMs. For an operator, it also allows uniform operation of such algorithms across different CMTS platforms. This profile optimization and profile creation functionality is implemented as an 'application' running outside the CMTS and is known in this paper as the Profile Management Application (PMA).

This paper briefly describes the data elements needed to enable a Profile Management Application, the key information needed to make decisions (e.g., channel config, profile test results, SNR/MER data for each CM etc.). It then focuses on the PMA itself, and the algorithms it needs to implement. This paper analyzes how a PMA can choose the appropriate modulation orders for a profile and other related questions such as: How does the PMA decide which profile is appropriate for a CM? How does the PMA decide the optimal set of profiles to use across the network?

DOCSIS DOWNSTREAM OFDM CHANNELS

A D3.1 CM supports two or more independently configurable OFDM channels, each occupying a spectrum of up to 192 MHz in the downstream. The maximum channel bandwidth of 192 MHz corresponds to 3841 subcarriers in 4K mode and 7681 subcarriers in 8K mode. The OFDM signal is composed of: Data subcarriers, Scattered pilots, Continuous pilots and PLC subcarriers. A DOCSIS 3.1 downstream OFDM 192 MHz channel's encompassed spectrum (active bandwidth of the channel) does not exceed 190 MHz. So the number of contiguous active subcarriers in a downstream OFDM channel does not exceed 3800 for 4K FFT and 7600 for 8K FFT. The two downstream channel types are summarized in Table 1.

Table 1 – Downstream Channel Options

Subcarrier Spacing	25 kHz	50 KHz
Symbol Period	40 µs	20 µs
FFT Size	8192	4096
Maximum Number of subcarriers	7600	3800

The downstream OFDM channel bandwidth can vary from 24 MHz to 192 MHz. Bandwidths smaller than 192 MHz are achieved by zero-valuing the subcarriers prior to the IDFT, i.e., by adjusting the equivalent number of active subcarriers while maintaining the same subcarrier spacing of 25 kHz or 50 kHz. See the DOCSIS 3.1 PHY Specification [3] for details on the OFDM Channel definition.

Downstream Modulation Profiles

Each of the subcarriers in the downstream OFDM channel can be configured to use a different modulation order. This allows the CMTS to optimize the downstream transmissions across the wide frequency band (192 MHz) of the channel. The specific choice of modulation order selected for each subcarrier is communicated to the CMs in the form of a modulation profile, which allows them to interpret and demodulate the signal.

A modulation profile consists of a vector of bit-loading values, an integer value for each active subcarrier in the downstream channel. Since the modulation orders range from 16-QAM to 16384-QAM, the range of bit-loading values is from 4 to 14 (skipping 5); however, it is expected that very low bit-loading values, 7 or less, will be used very infrequently since most plants support 256 QAM today.

The CMTS generates a "Profile A" that is the lowest common denominator profile, able to be successfully received by all CMs in the Service Group. It can then generate up to 15 additional modulation profiles, which are communicated to the Service Group. Each CM can be assigned up to four modulation profiles, including Profile A (used for broadcast frames), an optimized profile for the CM's unicast traffic, and possibly two additional profiles that could be used for multicast traffic. Since the number of CMs in the Service Group is expected to be larger than 15 in the majority of cases, each profile is expected to be used by a group of CMs that have similar channel characteristics.

This capability, the ability to optimize the downstream transmission for the channel characteristics of the CM population, is a powerful feature that allows for a significant improvement in channel capacity, and at the same time fundamentally changes the nature of RF plant maintenance. Prior to D3.1, all traffic has been sent using the lowest common denominator modulation (64-QAM or 256-QAM), setting the downstream channel capacity to a fixed value, and setting a Modulation Error Rate (MER) target for plant maintenance. Since only a fraction of the downstream traffic in D3.1 will be carried using Profile A, and since the CMTS can automatically determine the modulation to use for that profile, there is no longer a single MER target. Furthermore, since CMs can be assigned to modulation profiles that are optimized for their channel conditions, there is no longer a fixed value for channel capacity. The cleaner the channel is to any CM, the more efficient its traffic becomes, raising the overall average efficiency, and hence capacity, of the channel. Capitalizing on this capability requires that the CMTS can determine the channel conditions to the set of CMs in the Service Group, and that from this information determine the optimal set of (up to 16) modulation profiles.

OBTAINING INFORMATION FROM THE CMTS

There are a couple of methods which a Profile Management Application can become aware of the channel conditions in the DOCSIS network. These are using the OFDM Downstream Profile Test Request/Response mechanism (OPT-REQ/OPT-RSP MAC Management Message exchanges) or using the Proactive Network Maintenance (PNM) toolset.

OFDM Downstream Profile Test

In a D3.1 network, the CMTS can use new MAC Management Messages like the OPT-REQ to ask a CM to test various aspects of an OFDM downstream channel. A OPT-REQ message to a CM can check the CM's ability to receive a specified downstream OFDM profile and/or query the CM's RxMER statistics.

When asked to report the modulation error ratio measurements, the CM includes the RxMER per Subcarrier in the OPT-RSP message back to the CMTS. These are encoded as a packed sequence of 8-bit values for N consecutive subcarriers ($N \leq 7680$) from lowest active subcarrier to the highest active subcarrier, including all the subcarriers in between. Note that while the vector includes values for excluded subcarriers, a PMA can ignore these values based on its knowledge of excluded subcarriers.

In addition, if needed, a CM can compare the RxMER per Subcarrier to a Threshold and report a result, calculate the number of subcarriers whose RxMER is a certain value below a target, or report the SNR Margin of a candidate profile. See the D3.1 MULPIv3.1 spec for details [4]. In essence a CM can help by precomputing some of the data before sending it back in a OPT-RSP.

When requested to test a candidate profile, a CM reports back the number of Codewords received during testing, Corrected Codeword Count (codewords that failed pre-decoding LDPC syndrome check and passed BCH decoding) and Uncorrectable Codeword Count (failed BCH decoding). In addition, the CM can report if the number of codeword failures were greater than a given threshold for the Candidate Profile.

Proactive Network Maintenance (PNM)

The CMTS and CM perform measurements and report network conditions as a part of supporting PNM functionality in the DOCSIS network. The DOCSIS 3.1 Downstream PNM Measurements and Data

include: Symbol Capture, Wideband Spectrum Analysis, Noise Power Ratio Measurement, Channel Estimate Coefficients, Constellation Display, Receive Modulation Error Ratio (RxMER) Per Subcarrier, FEC Statistics, Histogram, and Received Power. See the DOCSIS 3.1 PHY Specification [3] for details.

DOCSIS 3.1 PNM capability assumes a PNM server that initiates PNM tests and receives data output from the CM and/or from the CCAP. Some of the PNM data, like the RxMER numbers, are the same as that computed by a CM for the OPT-RSP. All of this PNM data gets uploaded to a PNM Server When a Profile Management Application (PMA) comes around looking for this data from a PNM server, this data may be dated as compared to obtaining the data using the OPT method.

In this paper we are mainly using the RxMER per sub carrier data from the CMs. The other data as described above are definitely useful and how they can help further refine the profile definition will be a topic for future study.

PROBLEM STATEMENT

Objective of Profile Creation

We assume that a user's demand for bandwidth is independent of the modulation profile to which they are assigned. This assumption is based on the observation that the capability of a CM to use a particular profile is determined by the RF channel characteristics, which are presumably independent of the user themselves. Furthermore, in the absence of a priori information to the contrary, we can assume that, in aggregate, the users assigned to each profile are equivalent in terms of bandwidth usage, and, as a result, each user is expected to place an equal load on the channel on average. (In the case where all profiles contain a significant number of CMs, the law of large numbers would enforce this assumption, but in other cases there may be some achievable gain by taking into account historical usage patterns or weighting users based on their service tier. This is left for future study.) Based on this assumption, we can derive the average channel capacity for a population of CMs divided among a set of modulation profiles as follows.

Each user places data on the channel at a rate of b bits-per-second, and thus each profile of users places data on the channel at a rate of N_x*b bits-per-second, where N_x is the number of users assigned to profile x. This data rate equates to a symbol rate for the profile x, $S_x = N_x b/K_x$ symbols-per-second, where K_x is the total efficiency of profile x (i.e., the sum of the bit-loading values of all of the subcarriers). The total channel symbol rate S is the sum of the symbol rates of all of the profiles (note, this calculation disregards FEC, and the overhead incurred by the use of multiple profiles, i.e., additional NCP blocks, partial codewords, etc.), or:

$$S = \sum_{\forall x} S_x$$

$$S = b \cdot \sum_{\forall x} \frac{N_x}{K_x}$$

Since the symbol rate of the channel is a parameter set by the operator (either 50 Ksym/s or 25 Ksym/s), we can thus derive the per user bit rate for a fully loaded channel as:

$$b = S \Big/ \sum_{\forall x} \frac{N_x}{K_x}$$

And if we calculate the total channel capacity as:

C = N * b, where N is the total number of users.

We can then express this as:

$$C = NS \Big/ \sum_{\forall x} \frac{N_x}{K_x}$$

Or

$$C = S \Big/ \sum_{\forall x} \frac{\Phi_x}{K_x}$$

where $\Phi_x = N_x/N$ is the fraction of users assigned to profile x.

This is the harmonic mean (across all CMs) of each CM's single-user channel capacity.

DOCSIS 3.1 includes the notion of a lowest-common-denominator profile, referred to as "profile A" that can be utilized by all CMs in the service group. It is mandatory that profile A be created (for broadcast data, if nothing else), so a useful metric to assess the utility of a set of candidate profiles P is the ratio of the channel capacity using the set of candidate profiles to the channel capacity only using profile A.

This metric J is thus:

$$J_{P,A} = C_P / C_A$$

$$J_{P,A} = \frac{S / \sum_{\forall x} \frac{\Phi_x}{K_x}}{S / \frac{1}{K_A}}$$

Which reduces to:

$$J_{P,A} = \frac{1}{K_A \cdot \sum_{\forall x \in P} \frac{\Phi_x}{K_x}}$$

And, the objective of profile creation is to select the set of profiles P that maximizes the metric $J_{P,A}$ for the Service Group.

Channel Conditions for the Service Group

Given the population of N CMs in the Service Group, the PMA can generate the N-by-N_sc matrix M describing the bit-loading capability of each CM, where N_sc is the number of active subcarriers in the channel (up to 3800 for the 4K FFT, and up to 7600 for the 8K FFT). The PMA generates this matrix by instructing CMTS/CMs to perform channel measurements and report the MER per subcarrier, and then from this MER, calculating the maximum bit-loading that the CM can reliably receive for each subcarrier. This calculation could include some headroom to ensure reliable communication, or could be aggressive given that FEC will correct for some MER deficiencies. (The DOCSIS 3.1 PHY specification [3] discusses an example where alternating subcarriers are assigned alternating bit-loading values, in order to target a CM that reports an MER that is halfway between the MER required for each of the two bit-loading values, and relying on FEC to enable reliable communication.) The result is a "maximum-bit-loading" vector for each CM.

Calculating Profile A and Φ_x

For a given M matrix, Profile A can be determined by taking the column-wise minimum of M. Furthermore, the values Φ_x are determined by comparing the rows of M to the matrix of profiles P as follows. When considering whether a particular profile is suitable for a CM, the PMA needs to compare the maximum-bit-loading vector for the CM to the bit-loading vector of the profile, and ensure that the CM's vector is greater than the profile vector for all subcarriers. (A more sophisticated algorithm could allow the CM to be somewhat deficient in some number of subcarriers, based on an analysis of the ability for FEC to correct for these deficiencies.) We assume that the PMA will assign each CM to the most efficient profile available, the one with the highest total bit-loading.

This allows us to recast the objective function as:

$J_{P,M} = J_{P,A}$ where A=min(M) and Φ_x is determined as described above.

We can now express the optimization problem more explicitly as: given M, what is the P that maximizes $J_{P,M}$?

M is expected to vary over time, both due to CMs being added to or removed from the Service Group, and due to fluctuations in channel conditions, so it is expected that the PMA would periodically refresh its measurement of M, and update the set of profiles P accordingly.

Profile Cost

As mentioned previously, our derivation for channel capacity, and hence our objective function $J_{P,M}$, doesn't take into account the additional overhead incurred by introducing profiles to the profile set. This "cost" of supporting each additional profile is the result of the additional Next Codeword Pointer Data Message Blocks (needed when the profile in use is changed within a symbol), the possibility of needing to pad a symbol with Zero Bit Loading, and the increased likelihood of using shortened FEC codewords.

Given that $J_{P,M}$ doesn't factor in the "cost" of supporting multiple profiles, it can be shown that the optimal $J_{P,M}$ will increase as the number of profiles in P increases - until the number of profiles equals the number of unique vectors in M (i.e., each CM getting its own personalized profile). So, unless this cost is included in the optimization, the result will be biased toward larger sets of profiles. We leave the analysis of the profile cost for future study, and here focus on the problem of defining/selecting the optimal set of profiles given a target number of profiles, rather than determining the optimal number of profiles.

Multicast Profiles

As mentioned above, some profiles may be used for multicast transmissions that need to be received by a known subset of CMs. The CMTS can very simply generate a multicast profile specifically to target the joiners of each multicast group; however, since the CMTS only supports 16 total profiles, each of these multicast profiles reduces the number of unicast profiles available, and hence reduces the total unicast channel capacity. We leave for future work the problem of optimally defining profiles in the presence of multicast groups, and here focus only on optimizing the unicast profiles.

OPTIMIZATION METHODS

Brute Force Search

One approach to selecting the optimal set of profiles could be a brute force search: try all combinations of all possible profiles, calculating the $J_{P,M}$ value for each combination, and then selecting the best. The difficulty is that the search space is enormous. In a 192 MHz /4K FFT channel, there could be 3800 subcarriers, with 10 choices of bit-loading for each. Thus there are 10^{3800} possible individual profiles, and 10^{3800}-choose-15 possible 16-profile sets (approximately 10^{57000}) to evaluate. In an 8K mode channel (with 7600 subcarriers), the problem is even worse.

The search space can be reduced by observing that there is no need to try profiles that have a lower bit-loading than Profile A. For example, suppose that the geometric mean of the subcarrier bit-loading values of Profile A is 8 (corresponding to 256-QAM), the search space (in the 4K case) is then reduced to 7^{57000} or $\sim 10^{48000}$. This is a significant reduction, but still well beyond being computationally feasible.

Another simplification could be to observe that it is unlikely that subcarrier bit-loading would need to vary by more than one modulation order between adjacent subcarriers. This simplification could bring us down to approximately 10^{27000} profile sets. A further simplification (that may reduce the optimality) would be to assume that bit-loading would only change every 2 MHz (40 subcarriers), and then only by at most one modulation order each time. This reduces the number of profile sets to approx. 10^{700}. This is a dramatic reduction, but alas it is still computationally infeasible.

Assuming 200 CMs, it takes approximately 11 million operations to calculate $J_{P,M}$ for a 16 profile set on a 4K FFT channel. Optimistically, a current top-end Core i7 processor could evaluate 20,000 profile sets per second, or approximately 10^{10} profile sets per day.

Even with significant simplifications, the brute force approach is not realistic.

Grouping CMs

An optimal single profile for a group of CMs can be determined very simply by finding the minimum bit-loading capability for each subcarrier across the group of CMs. In other words, finding the column-wise minimum of the subgroup's bit-loading capability matrix, M_1, where the rows of M_1 represent the bit-loading capability vectors of the CMs in the subgroup.

Thus the problem of finding the optimal set of profiles for a population of CMs can be considered a problem of finding the optimal groupings of CMs.

Profile Coalescation Algorithm

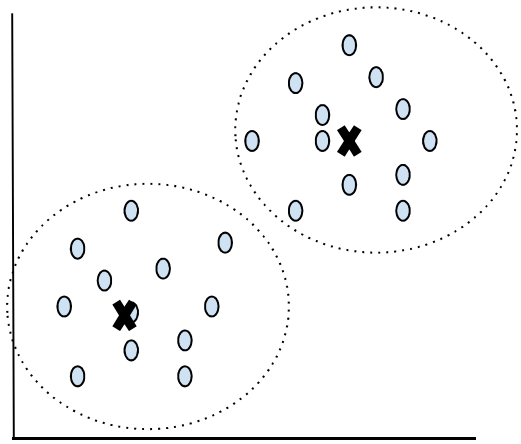
We have experimented with a technique for finding good groupings of CMs that operates as follows. Begin with one profile per CM, where each CM's profile exactly matches its bit-loading capability vector. This set of N profiles is optimal for the population of CMs; however, unless N is less than 16, this is more profiles than the CMTS can support. Next, calculate $J_{P,M}$ for this set of N profiles.

In order to reduce the number of profiles, generate the optimal set of N-1 profiles by combining two of the profiles into a single profile. The optimal value can be found by evaluating the effect of combining each of the N-choose-2 possible pairs of profiles (19,900 pairs for a population of 200 CMs) and then of these potential choices, coalescing the pair that provides the highest value of $J_{P,M}$.

This process can then be repeated to find the best set of N-2 profiles, then N-3, N-4, etc. until achieving the desired number of profiles. With large values of N, this results in a significant amount of processing ($O(N^2)$). This processing load can be reduced by pre-identifying clusters of CMs, and then starting with one profile per cluster instead of one profile per CM.

Clustering and K-Means

Clustering algorithms take an unlabeled data set and identify "clusters" of elements that are close to one another by some distance metric. The K-means [1] algorithm is a popular and widely used clustering algorithm.



K-means Algorithm for K=2 , separates the data points into two clusters, each point being assigned to the cluster whose cluster centroid (x) to which it is closest

Figure 1 - K-means Algorithm

The K-means algorithm takes a set of points and attempts to identify the set of K clusters for which the average distance of each point to the centroid (mean) of its cluster is minimized. The distance metric used in K-means could be simple L1 or L2 norm, or the Mahalanobis distance [2]. Implementations of K-means typically take as input the number of clusters (K) to split the input data set into, and return the set of K centroids as well as a set of assignments of points to clusters. K-means is not guaranteed to find the optimal set of K clusters, and in fact it will always converge to a local optimum. But, it is extremely efficient, and so it is possible to run it multiple times and select the best result.

K-means Algorithm

1. Select K, the number of clusters.
2. Initialize clusters by picking K random starting centroids
   - Typically this is done by selecting K random points from the set
3. Batch Update phase:
   - Assign every point to its nearest cluster (i.e., centroid to which it is nearest)
   - Once all points are assigned, recalculate centroids of the K clusters.
   - When the average distance no longer changes, move to the online phase
4. Online Update Phase:
   - Taking one point at a time, evaluate whether reassigning that point to another cluster (and then recalculating centroids) will reduce the average distance, and if so, reassign it to whichever cluster results in the smallest average distance (and recalculate the centroids).
   - When the assignments no longer change, the algorithm has converged

K-Means Applied to Finding Optimal Set of Profiles

For the profile optimization problem, we have experimented with using the K-means clustering algorithm to find groups of CMs that have similar bit-loading capability vectors, under the assumption that the profiles generated by the K-means clusters would provide an attractive set. This is motivated by the intuition that tightly clustered CMs would be good candidates to assign to a single profile.

The goal of K-means, however, is not directly aligned with our objective (optimizing $J_{P,M}$), and so the set of clusters found by K-means (even if it were to identify the optimal set) is not necessarily the optimal grouping of CMs with respect to our objective function.

In an attempt to sidestep this issue, rather than utilizing K-means to identify profiles, we use it as a pre-processing step for the profile coalescation algorithm to reduce the number of starting profiles, which significantly reduces the computational complexity.

K-MEANS COALESCATION ALGORITHM

Our algorithm, which we refer to as the K-means Coalescation Algorithm (KCA), runs in two stages. The first stage involves running several (e.g., 30) iterations of the K-means algorithm with K=20, and selecting the best result - that is the cluster assignment that results in the best set of profiles, as determined by its $J_{P,M}$ score. The second stage then uses this set of profiles as the initial set for running the profile coalescation algorithm.

RESULTS

CM Data

We created a synthetic data set of N=200 CMs , for a OFDM channel with N_sc=3800 subcarriers, by generating a N-by-N_sc matrix M describing the bit-loading capabilities of each CM. The CMs were divided into five clusters, with the CMs in each cluster having their bit-loading capability values chosen from a normal distribution with standard deviation of 1 and a mean unique to the cluster. The values were quantized (rounded) to integers and limited to the range (4 - 14) of bit-loading values. The mean values for the five clusters were 14, 13, 12, 10 and 9. The distribution of bit-loading values within each cluster are shown in Figure 2.

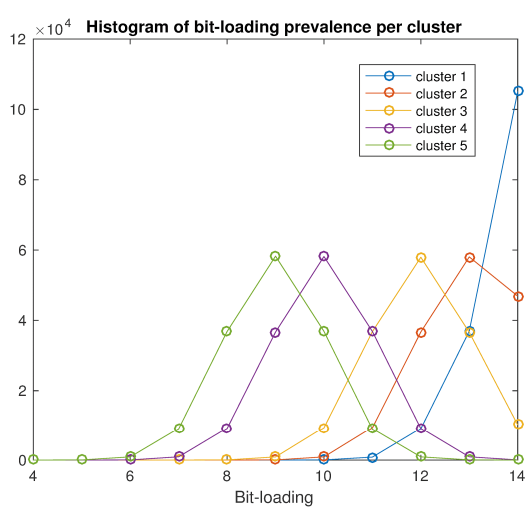


Figure 2 - Synthetic CM Data Set

Profile Coalescation Algorithm & K-means Algorithm Results

The K-means algorithm starts with a randomized set of centroids, and the results vary depending on the starting conditions. To get a sense for how this variation will affect the $J_{P,M}$ metric, we ran K-means 1000 times, and plot the histogram of the results (Figure 3).

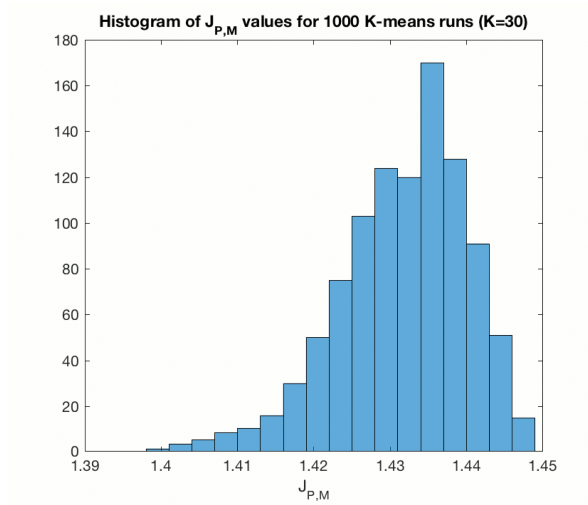


Figure 3 - Variability of the K-means Result

For the remaining profile generation algorithms, we run K-means 30 times and take the best result (as measured by $J_{P,M}$) of the 30 runs. Figure 4 shows the result of running K-means on its own to identify the profiles. The result is compared to the result obtained by the PCA algorithm. Running the full PCA algorithm on this modem population was extremely time consuming, but does provide better results than using K-means on its own.

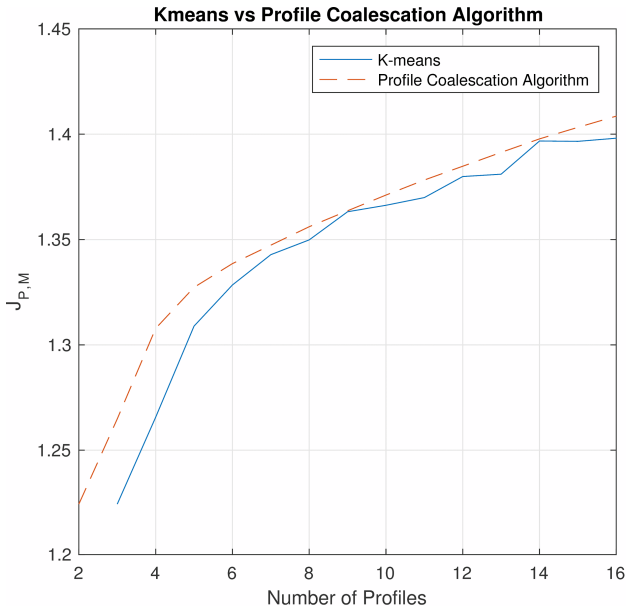


Figure 4 - K-means Compared to PCA

K-means Coalescation Algorithm Results

Our K-means Coalescation Algorithm (KCA) provided results that converge to the PCA results for our CM population. Figure 5 shows the results of the KCA algorithm using values of K from 5 to 125, compared against PCA (dashed line). In this figure, the x's show the $J_{P,M}$ value achieved by the K-means phase, and the solid line shows the progression of the coalescation algorithm.

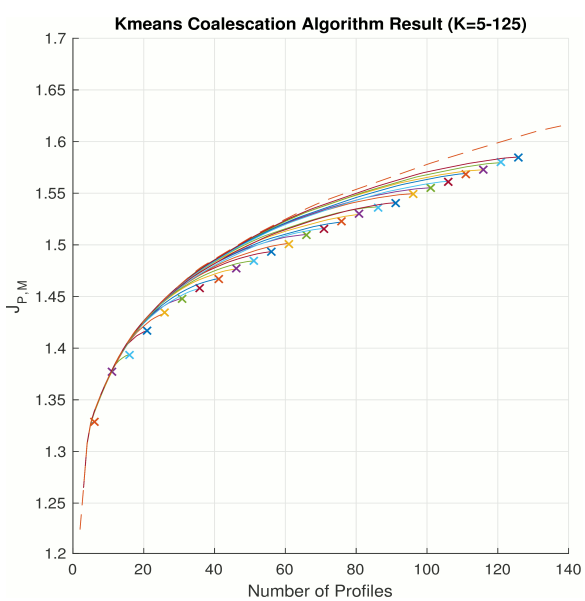


Figure 5 - KCA Results

Zooming in to see the region of most interest (where the number of profiles is ≤ 16), we can see that running KCA with a K value of 20 (i.e. 21 total starting profiles) results in $J_{P,M}$ values equivalent to the PCA for the 2-16 profile range, but with dramatically reduced computational complexity.

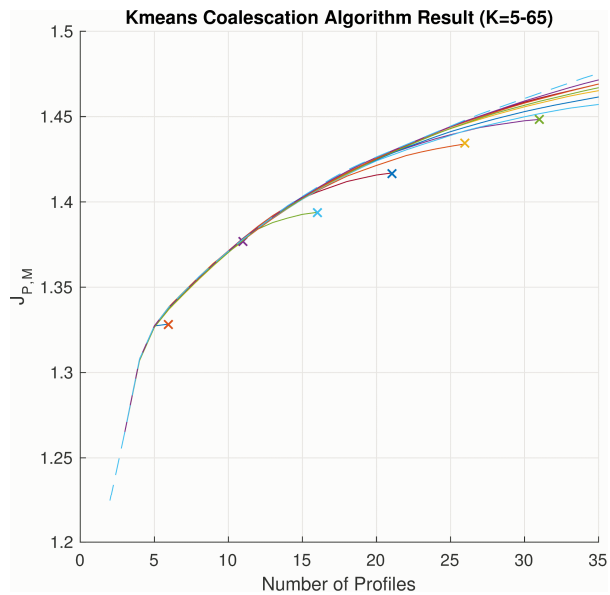


Figure 6 - Zoom of KCA Results

The gain provided by utilizing 16 profiles here is approximately 1.41, or 41% additional capacity. This result is, of course, strongly dependent on the CM population, and so could be much more or much less depending on how disparate the channel conditions are across the CM population.

FUTURE WORK

Our work at present has relied solely on synthetic data sets. In addition to the items mentioned previously, future work includes studying (and possibly refining) the KCA approach using real CM data.

Earlier, we commented that the D3.1 channel capacity is the harmonic mean of the individual CM single-user capacities. It is

important to note that the harmonic mean is dominated by the minimum of the data. So, operators can expect that overall channel capacity in D3.1 is going to be strongly influenced by the worst performing CMs. Furthermore, with a profile management application in place, if the SNR for the worst performing CM degrades, the profiles can automatically adjust to maintain connectivity, eliminating any overt signal that the degradation has occurred. Unless the operator has systems in place to detect this condition, they may be disappointed by the performance of their D3.1 channels. Monitoring the bit-loading of Profile A, and identifying for remediation the CMs assigned to it will likely be an important part of D3.1 plant maintenance.

Other areas for exploration include seeking answers to the following questions. How often does a profile management application need to recalculate the profile definitions? How does a profile management application handle sudden changes in the plant conditions, which make some profiles inappropriate for a set of CMs? When is it appropriate to alert the operator to degraded CMs that are requiring an extremely inefficient profile? This analysis will help an operator develop a solid profile management application and help maintain an efficient and optimized plant.

CONCLUSIONS

In this paper we presented a formulation of the DOCSIS 3.1 downstream modulation profile optimization problem, where we showed that the overall downstream capacity can be calculated as the harmonic mean of the individual CM's single-user channel capacities (i.e. the capacity each CM would achieve if it were the only CM on the channel), and that the profile optimization problem is akin to a clustering problem.

We applied a commonly utilized clustering algorithm, known as K-means, and showed its value and performance in identifying modulation profiles. We also developed a novel approach, which we refer to as the Profile Coalescation Algorithm, which provides better performance than K-means, but comes with significantly greater computational burden. Finally, we described a hybrid of the two approaches, which we refer to as the K-means Coalescation Algorithm, which can provide equivalent results compared to the PCA algorithm, but with much lower computational complexity.

REFERENCES

[1] K-means (Lloyd algorithm): Lloyd, Stuart P. (1982), "Least squares quantization in PCM", IEEE Transactions on Information Theory 28 (2): 129–137, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.131.1338

[2] Mahalanobis, Prasanta Chandra (1936). "On the generalised distance in statistics" (PDF). Proceedings of the National Institute of Sciences of India 2 (1): 49–55.

[3] DOCSIS 3.1, Physical Layer Specification, CM-SP-PHYv3.1-I08-151210, December 10, 2015, Cable Television Laboratories, Inc.

[4] DOCSIS 3.1, MAC and Upper Layer Protocols Interface Specification, CM-SP-MULPIv3.1-I08-151210, December 10, 2015, Cable Television Laboratories, Inc.

[5] MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. pp. 281–297. MR 0214227. Zbl 0214.46201. Retrieved 2009-04-07.

EFFECTIVE UTILIZATION OF M-ABR (MULTICAST ASSISTED ABR) USING BIG DATA AND REAL-TIME ANALYTICS

Sridhar Kunisetty, Jeffrey Tyre, and Robert Myers
ARRIS

Abstract

Adaptive Bit Rate (ABR) streaming protocols support multi-screen devices like tablets, smart phones, gaming devices, and smart TVs for accessing Over-The-Top (OTT) video content. ABR streaming is now making its way to the primary TV screen in the home in the form of next generation IP set-top boxes (STBs), and with this a tremendous growth in ABR generated, HTTP unicast traffic is expected due to higher video resolutions and viewership concurrency as compared to second screen devices. The introduction of Multicast-assisted Adaptive Bit Rate (M-ABR) solutions is intended to help offset the potential impact of this unicast traffic growth by leveraging the bandwidth efficiencies of IP multicast over the operator network, while still enabling all pay TV subscriber devices in the home to support Adaptive Bit Rate streaming.

Actionable insight derived from collecting, processing, and analyzing the huge amount of data from the millions of home gateways provides several levels of valuable information that allows us to better manage the network by understanding the operational issues. Harnessing Big Data technologies and doing this in real-time, provides a feedback loop using which things can be changed dynamically at a granular level -- at a specific time of the day in a specific geographical area -- to better utilize the bandwidth and reduce operational complexity.

This paper provides an overview into how Big Data and Real-time Analytics enables insight into video consumption and network operational aspects to make M-ABR more effective.

M-ABR SOLUTION SUMMARY

Adaptive Bit Rate (ABR) protocols have become the mainstay of multi-screen devices like tablets, smart phones, gaming devices and smart TVs for accessing Over-The-Top (OTT) video content. Because of their explosive popularity, it is highly desirable for an operator to provide existing services to these IP Video, multi-screen devices. The ABR protocols have been optimized to deliver a wide range of bit rates for varying screen sizes from thumbnails to 4K Ultra HD, operating over IP network connections with fluctuating bandwidth and service levels.

Adaptive Bit Rate streaming is already used today by pay TV operators to deliver linear content over IP in the home to second-screen subscriber devices such as PCs, laptops, tablets, and mobile phones. ABR streaming is now making its way to the primary TV screen in the home in the form of next generation IP STBs. The intent of the M-ABR solution is to enable all pay TV subscriber devices in the home to be ABR-based using unicast HTTP, and to leverage the bandwidth efficiencies of IP multicast for linear channel delivery across the operator network. The economic benefits of the M-ABR solution have been previously documented[2].

The following diagram illustrates the major components of a Multicast-assisted ABR solution based on recently released CableLabs specifications. The CableLabs M-ABR specification[1] uses the IEFT RFC-5740's NORM (NACK-Oriented Reliable Multicast) transport protocol to propagate multicast streams over the operator network.

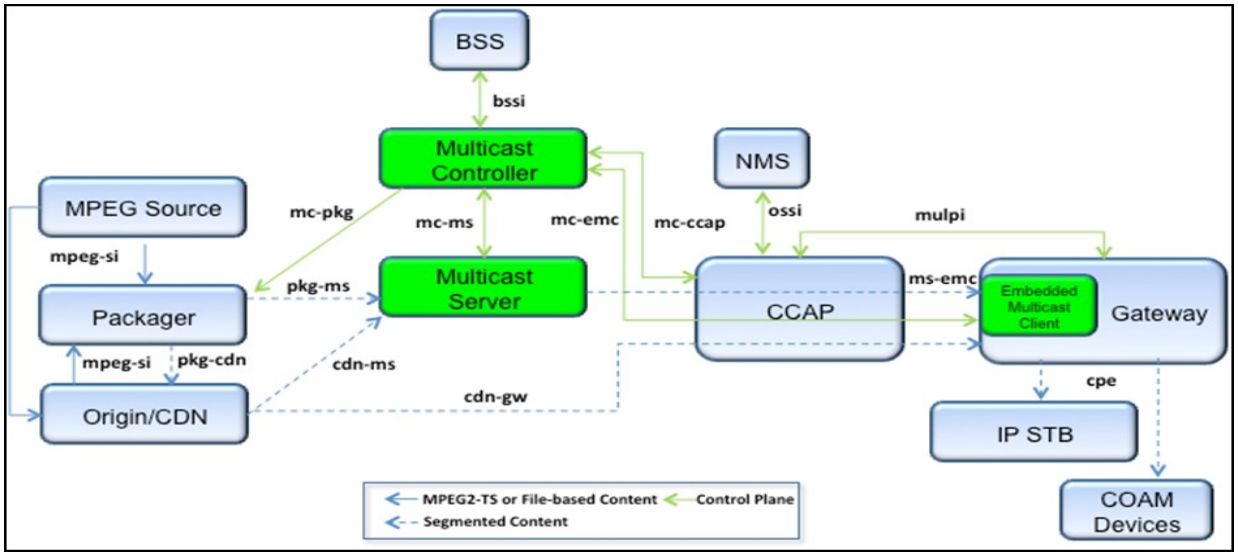


Figure 1: Major components in a M-ABR System

In this pay TV operator solution, an *ABR transcoding-packaging system* publishes ABR content into the operator managed content delivery network (CDN). The *M-ABR Server* pulls this content from the CDN as a normal ABR client would do, and sends the content as a multicast transport over an IP routed network to a *M-ABR Edge Device* with an *Embedded Multicast Client (EMC)*. The M-ABR Edge Device is often a home gateway representing the operator network termination point for the customer premise equipment (CPE). The addition of a *M-ABR Configuration Manager* may be deployed to support the setup and configuration details of the M-ABR Servers and Edge Devices with EMCs hosted on them.

The *M-ABR Controller* is responsible for defining the mapping of specific ABR delivered streams for the M-ABR Servers to multicast delivery of the ABR content and provide mappings to edge devices so they know what services are available and how to connect to them. Management of this mapping can be either operator scheduled or dynamic based on tracking ABR client channel viewership trends across service groups, as well as at regional and national levels.

The M-ABR Controller determines what channels are multicast by instructing M-ABR Servers to setup and subsequently tear down multicast streams as ABR stream viewership changes or based on operator scheduled multicast events. Having a single management platform for both scheduled and dynamic viewership allows optimization of network, M-ABR Server, and edge device cache resources.

The M-ABR Embedded Multicast Client (EMC) – Window to the New Age of IP Video Analytics

The Embedded Multicast Client (EMC) is an application running in the M-ABR Edge Device that manages all aspects of the NORM multicast inputs received from M-ABR Server and their conversion to HTTP unicast ABR streams delivered over a local network (e.g. subscriber home network) to ABR clients for playout. In an effort to standardize its implementation, an EMC application is currently being incorporated into the

Reference Design Kit for Broadband services (RDK-B) library.

EMC communicates with the M-ABR Configuration Manager which provides configuration information to EMC. This includes the M-ABR Controller URL to send heartbeats and receive a multicast channel map and other configuration parameters required for operation.

The following describes the process that occurs when an ABR player client makes a request for content through the M-ABR Edge Device with EMC running as an application. When EMC receives an ABR video segment requests from ABR player clients which have been proxied to it:

1. EMC checks if the ABR client URL request matches a stream that is currently being multicast and in the Edge Device's local cache. If not, EMC requests the video segment via HTTP unicast and returns it to ABR client. In this capacity, EMC acts as a traditional CDN transparent web cache.

2. If the video segment requested by the ABR client is present in the local cache, EMC returns it to the ABR client immediately. A video segment could already be present in the local cache due to a previous ABR client request or based on a scheduled operator request to pre-cache content based on projected program viewership such as for live sports events.

3. If the NORM multicast stream has not been joined, the EMC joins the multicast stream to receive video segments via multicast, then converts these from NORM to an ABR format (e.g. HLS, DASH ISO), and stores these as ABR segments in the Edge Device's local cache.

4. All subsequent ABR segment requests are delivered from the Edge Device cache based on step #1 above.

5. EMC periodically checks for updates to the configuration and multicast stream channel map.

Subscriber devices are HTTP unicast, ABR clients, whether they are IP STBs or other subscriber devices such as iPads, PCs, and smart phones. The fact that some of the ABR content is being carried over the operator network as multicast traffic is completely transparent to them. End ABR clients could themselves support the M-ABR processing. However, it is generally expected that this functionality to be the most prevalently implemented in home gateways to minimize CPE costs by deploying the M-ABR functionality in a single home device.

EMC periodically sends statistics to the Analytics System about the ABR stream sessions and other details, including what ABR streams are requested, the cache hit/miss ratio for multicast streams, multicast data reliability, etc. This data will be the backbone of the operational and subscriber viewership analytics driving which streams should be available via multicast.

With the introduction of the M-ABR and EMC-enabled home devices, a new window to pay TV operator IP Video-supported services will emerge. Similar to current legacy MPEG video delivery technologies, new tools will need to be developed in order to effectively manage the associated new service deployments. This paper explores the analytics potentially derived from collecting and processing the huge amount of data generated from millions of EMC-enabled home devices that could be instrumental in ensuring the successful deployment of Multicast-assisted ABR solutions.

OVERVIEW OF BIG DATA & REAL-TIME ANALYTICS

The advent of the Internet and the resulting data that needs to be managed, stored, and analyzed, over cheap commodity hardware that can easily scale horizontally, led to the emergence of modern day Big Data technologies. Terms typically used to characterize Big Data are volume, velocity, and variety[3]. Volume refers to the sheer amount of data that is being created continuously. Velocity is the speed with which this data can be processed and analyzed for a meaningful use. Variety is the different types and formats of data that is collected.

The core tenets of any Big Data system are its ability to collect a variety of large amounts of data, use parallel processing to analyze and process that information to derive meaningful insight, and to provide access to the derived information through dashboards, reports, and APIs[4]. All this can be achieved at near linear scalability - as the data size increases, one can throw more hardware at it and see processing complete in the same amount of time – and at the same time being tolerant to failures on any of the computer nodes processing that data.

Over the years, many companies, universities, and open source communities contributed to Big Data related technologies such as Hadoop [5], MapReduce[6], YARN[7], SpringXD[8], Spark[9], etc. While Hadoop is good at handling Big Data, it primarily deals with batch-oriented data processing and storage. Traditional Hadoop technologies are not well suited for real-time data processing needs. In recent years, the open source Spark framework has become the defacto choice for processing and analyzing big data in real-time.

While Spark's streamlined API allows representing complex pipelines in a few lines of code making distributed programming easier, its in-memory processing makes it an order of times faster than traditional Hadoop technologies[9][10]. Spark also makes the life of a data scientist easier by providing APIs for statistics, machine-learning, graph processing, etc. – all in a single framework while providing linear scalability.

Spark provides strong integration with Hadoop ecosystem and in many cases both of them are used together – Spark for real-time data processing & Hadoop for batch processing and to store the data.

In this paper, our primary focus is on real-time analytics and not batch processing of Big Data.

Applicability to M-ABR

A well-instrumented M-ABR eco-system mirrors the web world in terms of volume, velocity, and variety of data that needs the processing power of a Big Data system. A typical operator has a few million subscriber homes, each of them having a home gateway (containing EMC), along with one or more STBs and second-screen devices connected to the home gateway. Each of these millions of home-based subscriber devices are constantly producing events i.e., we have a system that is generating millions of data events every minute. In addition to the devices in the subscriber's home, the components in the content delivery side, such as Multicast Server, also produce huge amount of useful data.

Analyzing this data provides insights into the M-ABR ecosystem for understanding operational issues and generating a feedback loop to help in the automation for fixing the operational issues that were detected. In order for this to be effective, the collected data need to be analyzed in real-time. For example, if we detect that the network latency is beyond a particular threshold, then this information can be used for corrective actions immediately. Similarly, if the cache hit ratio is below the

99% watermark, then it tells you, in real-time, that there are some problems in the network that need immediate attention.

While, addressing operational issues is the main focus of this paper, the M-ABR Analytics system can also gain insight into quantifying user behavior to facilitate various types of analysis and recommendation systems. Client-facing applications resident on a EMC device or a web-application can be instrumented to collect user interaction and content consumption information to derive insight into user behavior with respect to video consumption. And when combined into external data such as EPG metadata & demographics, we can get an even deeper insight into user behavior. This paper does not elaborate on the user behavior analytics aspect.

ANALYTICS - DERIVING INSIGHT

Analytics System Overview

The Analytics System is broadly classified into three different layers:

- *Data Ingest Layer*: This layer facilitates collection of multi-formatted information from a variety of data sources in real-time, doing parsing and pre-processing as needed and passing it to the Analysis & Processing Layer. Optionally, a copy of the raw data is stored in Hadoop for long-term storage needs.

- *Analysis & Processing Layer*: This is where all the real-time processing and analytics happen. It consists of different processes all running asynchronously in parallel. One set of processes operate on the raw data, process them and stores in a No-SQL database. The other set of processes perform the next level of processing by doing summarizations, correlations, deeper analysis, etc.

- *Data Output Layer*: The resultant data is exposed through multiple methods – dashboards and charts that provide a visual view into the different operational and usage aspects of M-ABR and REST APIs that can be programmatically consumed by various services to act on the analyzed data.

A Spark cluster, running on multiple nodes, provides the Analysis & Processing Layer the ability to horizontally scale as the data traffic increases.

Goals of Real-time Analysis

The objective of the Analytics System is to benefit the operator by providing meaningful information to understand and address operational issues in the M-ABR eco-system.

High-level goals include:

- Understanding usage of M-ABR components, services, and network

- Understanding and classifying network and system limitations by analyzing trends of system usage with respect to scalability, capacity planning, and bandwidth management

- Creating dashboards to provide business intelligence, and APIs for downstream usage

- Answering key questions about the performance of multicast over DOCSIS, including the telemetry of any packet loss detected by the EMC

- Evaluating the workload within the EMC to perform the functions of the “transparent proxy cache”

- Examining the performance of the CMTS/CCAP for multicast delivery and

implications related to its support for IGMP.

At a more detailed level, the Analytics Systems helps in analyzing and answering questions such as:

- How often does a file segment experience any packet loss?
- Is the packet loss characterized by factors such as time of day, downstream capacity, CMTS type/configuration, aspects of the HFC access network, IGMP traffic management, etc.?
- How does channel surfing affect the IGMP behavior of joins and leaves and when does it become more of a burden and less of a benefit?
- How quickly can initial tuning to a new channel result in moving from unicast video segments to multicast (cached) video segments?
- How long should the EMC remain on a multicast stream before it determines there is no benefit to caching stream segments?
- What is the timing of a cache segment to client request?
- What is the behavior of client change in quality using alternate, non-multicast variant playlists?

EXAMPLE DATA EVENTS AND INSIGHTS

Data Events

The Analytics System captures a variety of information. Some of them are:

- Segment data
- Multicast Session data
- Client Session data

Session Data

For each multicast video segment received on the EMC, this data event is produced containing attributes such as multicast service name, receive timestamp, missing packet count, client request timestamp, etc.

This data helps in analyzing the following:

- Segment packet loss details and whether the packet loss was sequential or multiple packet-loss across the segment
- Segment timing details such as, when it is received with and without error and how soon after the client requested the video segment
- Segment details about how long it is maintained in the EMC cache

Multicast Session Data

For each IGMP session on the EMC this data event is produced to track the lifecycle of the multicast stream. This data event contains attributes such as multicast service name, multicast join & leave timestamps, first segment received timestamp, total segment count, segment errors count, etc.

This data helps in analyzing the following:

- IGMP behavior on the EMC and the success of each multicast video segment received
- Estimate of the IGMP load on the CMTS as a viewer transitions from service to service
- Details on the multicast session, such as when it is initiated, when it receives data, when the first segment starts, how many segments had an error, etc.

Client Session Data

For each client HLS session associated to a multicast service, this data event is generated containing attributes such as multicast service name, start manifest timestamp, total segment count, initial unicast segment count, user agent, etc.

This data helps in analyzing the following:

- Whether the video segment was provided from the cache or required the EMC to pass the request through unicast on the DOCSIS network
- Estimate the initial unicast load for the client buffer fill as well as how often the client request a segment that results in a cache miss and goes unicast
- The experience of a client session using specific multicast service

Insights

Information that is collected and processed must provide actionable value to the operator. Most of the insights can be categorized into the following keys areas:

- Multicast Channel Usage / Channel Popularity
- Initial Unicast Traffic & Multicast Cache efficiency
- Network Behavior & Packet Errors
- EMC Client Interaction Behavior

This section shows a few example charts providing meaningful insights.

Channel Utilization

Figure 2 shows the number of client sessions that exist on a per multicast channel (stream) basis over a specified time period, termed as the Popularity of a Channel.

This information is helpful in environments where the number of channels that can be simultaneously multicast is limited due to bandwidth capacity limitations. It can help to determine which channels are more popular and when.

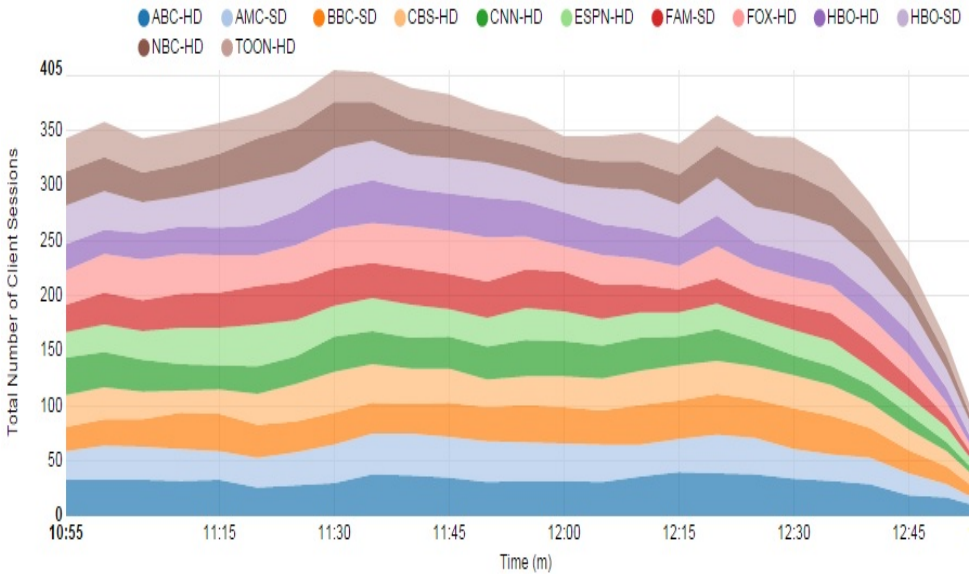


Figure 2: Channel Utilization

Missed Segments

Figure 3 shows the peak and average number of segments that could not be delivered from the EMC cache to a client device after the EMC has started receiving the multicast stream for that client session. In an optimal network this should be zero. Missed segments but can be impacted by issues in such as network errors or multicast server failures.

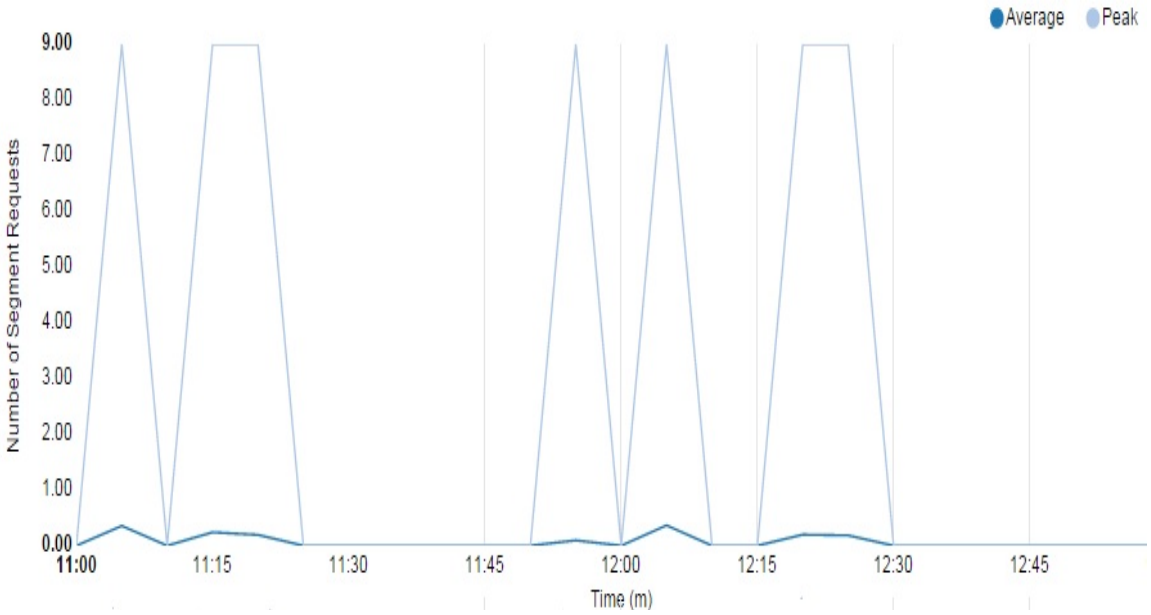


Figure 3: Missed Segments

Network Latency

Figure 4 shows the peak and the average time that it took for the EMCs to receive a segment from the multicast stream. This value is dependent on other factors such as duration of the segment, the available bandwidth and the multicast rate percentage. Optimally, the average value should be less than the segment duration to ensure that the EMC will always have available segments in its cache. If this value becomes greater than the segment duration it could indicate an issue somewhere in the network or in the multicast server.

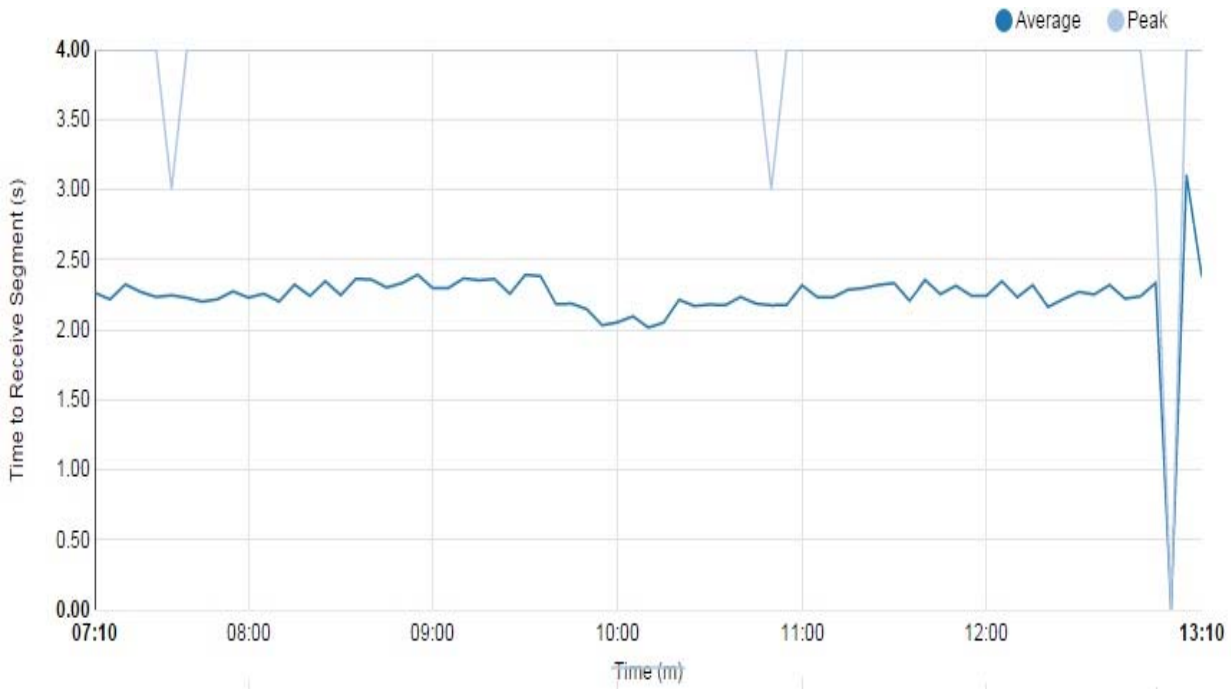


Figure 4: Network Latency

CONCLUSION

Multicast-assisted Adaptive Bit Rate (M-ABR) is a promising solution for operators as they roll out IP Video to Multiscreen devices since it helps in leveraging the bandwidth efficiencies of IP multicast for linear channel delivery over the operator network, while still enabling all TV subscriber home devices to support ABR streaming. Understanding the operational issues becomes essential for this new technology so that we can address those issues to improve efficiency, reduce cost and provide better user experience.

Actionable insight derived from collecting, processing, and analyzing data events provides several levels of valuable information. The volume, variety and speed of data generated from the M-ABR eco-system needs a Big Data and Real-time Analytics system.

In this paper, we provided an overview of the Big Data and Real-time Analytics technologies, discussed the sources of events in the M-ABR system, and provided examples of meaningful conclusions that can be ascertained from analyzed events useful in understanding and driving operational efficiencies.

ACKNOWLEDGEMENTS

This paper represents work that is being done in ARRIS Network & Cloud business unit.

REFERENCES

1. "IP Multicast Adaptive Bit Rate Architecture Technical Report", Cable Labs document OC-TR-IP-MULTI-ARCH-V01-141112, Nov 2014

2. Ulm, J., “Challenges delivering Multiscreen Linear TV Services: Multicast assisted ABR to the Rescue”, SCTE Cable---Tec Expo, Fall 2014

3. Couts, A., “Breaking down 'big data' and Internet in the age of variety, volume, and velocity”, Retrieved from

http://www.digitaltrends.com/web/state-of-the-web-big-data/, Sept 2012

4. Gandhi, B., Martinez-Smith, A., Kuhlman, D., “TV INSIGHTS - Application of Big Data to Television”, IBC 2015.

5. “Apache Hadoop”, Retrieved from http://hadoop.apache.org, Feb 2016.

6. "MapReduce Tutorial", Retrieved from http://hadoop.apache.org/docs/r2.7.2/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html, Jan 2016.

7. "MapReduce NextGen aka YARN aka MRv2", Retrieved from http://hadoop.apache.org/docs/r2.7.2/hadoop-yarn/hadoop-yarn-site/index.html, Jan 2016.

8. "Spring XD Guide", Retrieved from http://docs.spring.io/spring-xd/docs/1.3.1.RELEASE/reference/html/, Feb 2016.

9. “Spark, Lightning-fast cluster computing” Retrieved from https://spark.apache.org, Mar 2016.

10. “Spark the fastest open source engine for sorting a petabyte”, Retrieved from https://databricks.com/blog/2014/10/10/spark-petabyte-sort.html, Oct 2014.

Evaluation of virtualizing DOCSIS MAC by software in data center

Li Zhang, Xin Yang, Lifan Xu
Huawei Technologies Co.,Ltd.

Abstract

Virtualized CCAP was thought as the best cable access architecture in the future, but now the industry hesitates in what parts of the CCAP functionality should be moved to the cloud and what parts should be kept in the physical hardware. It is not only related with technical feasibility of what can be done by software, but also economic consideration.

We all agree on moving the CCAP configuration and subscribe management modules to the cloud, and remaining the DOCSIS PHY in the physical hardware. But as to the DOCSIS MAC, it is not easy to make that kind of decision because the MAC is part of hardware like Encryption and Framer, and part of software like scheduling.

This paper will look deep into each functionality module of the DOCSIS MAC including data plane, control plane and management plane, then give an analysis of virtualized DOCSIS MAC base on the evaluation of generic computing resource spent by each MAC module.

1. DOCSIS MAC and NFV Platform

DOCSIS MAC consists of data plane, control plane, and management plane. Now the control and management plane are all software in CCAP implementation, but the data plane is almost hardware to guarantee high throughput and low latency. Virtualized CCAP is considering about moving the DOCSIS MAC to cloud, before that, we need to think about the cost and performance of virtualized MAC in first.

In order to estimate how many generic CPU cores are needed for DOCSIS MAC virtualization and what performance it can achieve, we break down the DOCSIS MAC to several functionality modules, then build a NFV simulation model driven by packets for each module. We can get the CPU cycles consumption of the whole upstream and downstream MAC data path by simulation, based on that data, we can know the throughput performance in various packet size with a given X86 server.

1.1 DOCSIS MAC Break Down

DOCSIS MAC is a quite complex protocol compared to other access MAC technologies like PON, Figure 1 show the whole DOCSIS MAC modules which includes Upstream and Downstream Data Plane, DOCSIS MAC Control and Management Plane.

Now the DOCSIS downstream and upstream data plane are all hardware implemented by FPGA or ASIC chipset. According to DOCSIS MULPI specification, the downstream frame processing includes the following steps.

1) ***Ethernet RX MAC:*** it receives L2 Ethernet packet from IP router.
2) ***Service Flow Classifier:*** it searches the MAC address table and get destination CM ID firstly, then determine the SFID of the packet base on CM classifier profile, meanwhile, get the QoS attributes.
3) ***Downstream Scheduler:*** it puts packets to different downstream queue according to the QoS attribute, and schedule them out base on specific algorithms and rules.
4) ***Bonding Channel Distributor:*** it distributes packets to different downstream bonding channel base on load balance strategy, the DOCSIS MAC management packets usually have higher priority than data packets.
5) ***AES/DES Encryption:*** it encrypts packet according to the CM individual keyword.

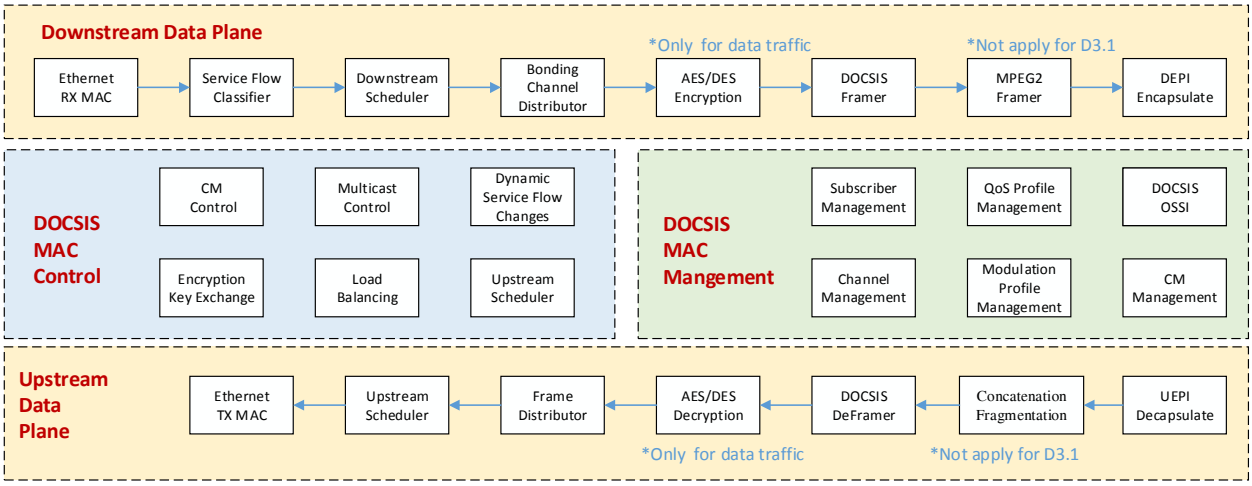


Figure 1. DOCSIS MAC Functionality Break Down

6) ***DOCSIS Framer:*** it adds DOCSIS Header for each packets.
7) ***MPEG2 Framer:*** it slices the DOCSIS stream to 184bytes and adds 4 bytes MPEG Header. It is only for D3.0 and below version because D3.1 has removed the MPEG framer.
8) ***DEPI Encapsule:*** it encapsule the downstream DOCSIS or MPEG frames into L2TP tunnel connection between MAC and PHY.

The upstream data plane is almost reverse path of downstream except some small differences. The processing steps are explained in the below.

1) ***UEPI Decapsule:*** it stripes L2TP header and get DOCSIS or MPEG frames.
2) ***Concatenation and Fragmentation:*** it concatenates the DOCSIS upstream burst segments.
3) ***DOCSIS DeFramer:*** it checks the CRC, then stripes the DOCSIS header.
4) ***AES/DES Decryption:*** it decryptes upstream frames for each service flow.
5) ***Frame distributor:*** it pickes up the DOCSIS management messages from data traffic and forward them to control and management software.
6) ***Upstream Scheduler:*** it schedules the upstream packets into different priority queue buffer.
7) ***Ethernet TX MAC:*** it forwards the upstream L2 packets to IP router.

DOCSIS MAC control plane modules are real-time dynamic operation, unlike static service configuration, they are triggered by service application or MAC internal messages, for example VOIP initialize a call, CM request a upstream bandwidth, or multicast video apps request a program. Those modules are software threads and running on CPU with real-time OS or NP, they may not consume too much CPU cycles but always be required to response immediately. Actually the DOCSIS control functionality is fully defined in MULPI specification, a brief explaination of each control modules is provided in the below.

1) ***Upstream Scheduler:*** it collects Bandwidth Request Message from CMs and calculates upstream grants for each CM in every scheduling cycle.
2) ***Dynamic Service Flow Changes:*** it creates, changes or deletes service flow dynamically, usually the policy server will instruct CCAP to change service flow.

3) ***Load Balancing:*** it is a feature of CCAP that controls the dynamic changes to the set of downstream and upstream channels used by a CM.
4) ***Multicast Control:*** it responses multicast group join, leave request from CPE.
5) ***Encryption Key Exchange:*** it updates CM encryption key words periodically complying to BPI+ specification.
6) ***CM Control:*** it authorize CM when power on and monitor the connection between CM and CCAP.

DOCSIS MAC management plane is about configurating different kinds of profiles like channels, MAC domain service groups, QoS, D3.1 modulation orders and so on. Usually almost configurations are few possibility to be changed unless operators want to extend the upstream or downstream channel number, add QoS profile for a new service etc., so the management modules are all regular software threads which don't have strict real-time requirement.

1.2 NFV DPDK Framework

NFV(Network Function Virtualization) is trying to replace network equipment with generic x86 CPU and software, the hardware can be physical server or virtual compute resource in the cloud. Operators like NFV because it decouples the network functions from proprietary hardware appliances and offer a new way to design, deploy and manage networking services.

DPDK(Data Plane Development Kit) is a widely used NFV programming framework for Intel x86 which enables faster development of high speed data packet networking applications. It provides a set of data plane libraries and Ethernet drivers that run in linux user space for fast packet processing. In order to evaluate virtualized DOCSIS MAC, we change the orginal FPGA MAC data plane modules to software unit base on DPDK framework, and try to get the CPU instruction cycle consumption of each MAC packet processing unit.

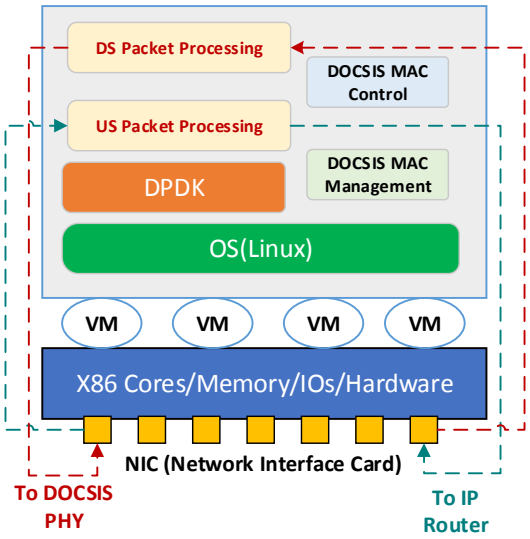


Figure 2. Virtualized DOCSIS MAC

2. Evaluate DOCSIS MAC virtualization

Building a real and complete DOCSIS MAC software and making a end to end system level testing is a tough work. In this paper, we only do real testing in module unit level and sum them together to evaluate the system performance. The hardware we refer to is Huawei FusionServer RH2288 which can provide Intel Xeon 12 cores 2.6GHz-64bits processor, 8*10GE NIC and 128G RAM.

2.1 MAC data plane evaluation

The data plane evaluation is the most difficult, you should design the software processing flows and tasks very well to achieve high throughput and low latency. In

order to make the evaluation as exact as possible, we use the optimized API in DPDK library to implement queue scheduling, classification and packet slicing modules which can guarantee high processing performance, and as to AES encryption and decryption, we choose intel AES-NI library which is a special instruction set designed for security applications.

The MAC data plane modules process and forward traffic by packets, the performance of forwarding software relates with packet size, we choose 64, 128, 256, 512, 1024, 1518 and 2000 bytes as input packet size, then count the processing cycles. Obviously longer packets can get higher bits/s throughput, Table 1 and Table 2 provide the downstream and upstream processing cycles for 2000 bytes frame which is the maximum PDU size in DOCSIS 3.1 MULPI specification. We can see the downstream and upstream computing cost are very close, and AES encryption and decryption consume more than half computing resource.

DS Data Plane Modules	CPU Cycles
Eth RX and Classifier	1948
Downstream Scheduler	1318
Bonding Channel Distributor	218
AES Encryption	16384
DOCSIS Framer	353
MPEG2 Framer slicing	7631
DEPI tunnel	651
Total	28503

Table 1. DS data plane CPU cycles

US Data Plane Modules	CPU Cycles
Eth TX	1552
Upstream Scheduler	818
Frame Distributor	318
AES Decryption	16384
DOCSIS Deframer	2553
Concatenation Fragmentation	5931
UEPI tunnel	502
Total	28058

Table 2. US data plane CPU cycles

Table 3 gives the total downstream and upstream CPU cycles in different packet length.

Packet Size	DS Cycles	US Cycles
64 Bytes	21065	20672
128 Bytes	21103	20768
256 Bytes	21814	21410
512 Bytes	22602	22244
1024 Bytes	24811	24362
1518 Bytes	27009	26453
2000 Bytes	28503	28058

Table 3. CPU cycles for typical packet length

2.2 MAC control and management evaluation

Unlike data plane, MAC control and management threads are driven by events not packets. The events usually occur randomly and can't be measured like packet throughput, fortunely those modules only require real-time operation but don't cost too much computing resource except upstream scheduler module. Thus we decide to port the upstream scheduler code to X86 processor and evaluate the performance by real testing, but as to other modules, it is hard to port them all to X86 processor, so we estimate computing cost by the CPU load usage experience data in current device.

Upstream scheduler operation is cycle base, it collects all the request from cable modem and then allocates grants in every cycle. The scheduling cycle length will impact data traffic upstream latency, usually the scheduling cycle is set to be 1.5-2ms. The computing resource consumption of upstream scheduler relates with how many cable

modem it supports. Figure 3 gives a X86 CPU computing usage curve in different number of Cable Modem, the maximum Cable Modem number one CPU core can support is about 1800, so we can estimate the CPU consumption of upstream scheduler is:

$$\frac{X86\ Cores}{1K\ CMs} = 1/1.8 \approx 0.556$$

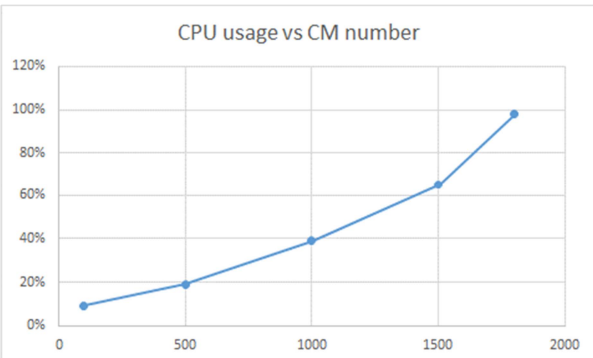


Figure 3. Upstream Scheduler CPU usage

Now we run the other MAC control and management modules except upstream scheduler on a ARM processor with 1.2GHz speed, when we load 1000 CMs and each CM has 4 upstream service flows and 4 downstream service flows, the CPU usage is about 70% in peak time and 20% in idle time. If we think the X86 CPU in data center can be designed to 100% usage, and know the 2.6GHz X86 CPU computing ability is 3.2 times of 1.2GHz ARM through running same testing code, thus we can estimate the X86 CPU cores per 1K CMs for other MAC control and management modules as:

$$\frac{X86\ Cores}{1K\ CMs} = 70\% \div 3.2 \approx 0.218$$

3. Data analysis

Data Plane CPU consumption relates with the packet length, for real data traffic, the Ethernet packet length varies from 64 to 1518 bytes, typically we choose iMIX profile shown in Table 4 as access network packet length distribution model, the CPU consumption can be calculated as:

$$\frac{\text{Cycles}}{bps} = \sum_{l=64-1518} \frac{bits\ percent(l) * Cycles(l)}{packet\ length(l) * 8}$$

Length	64	128	256	512	1518
Packets Percent	50%	10%	15%	10%	15%
Bits Percent	8.9%	3.5%	10.6%	14.1%	62.9%

Table 4. Variable Packet Length Distribution

If we input Table 3 and Table 4 data to the above formula, we can know the upstream data plane is 7.55 cycles/bps and the downstream data plane is 7.7 cycles/bps. Assuming the downstream and upstream bandwidth ratio is 5:1 and the X86 processor speed is 2.6GHz, we can calculates the CPU cores needed to achieve 1Gbps as below.

$$\frac{Cores}{Gbps} = \frac{1G * 7.7 + 1G * 0.2 * 7.55}{2.6G} = 3.54$$

The CPU cores consumption of the whole DOCSIS MAC software can be summed as:

$$\text{Cores} = 3.54 * \text{T} + 0.556 * \text{C} + 0.218 * \text{C}$$

Where:

- T is throughput in Gbps;
- C is the number of cable modem in one thousand.

For the data shown in Figure 4 it gives 3 CPU consumption curve for small HUB with 10K HHPs, typical HUB with 30K HHPs and big HUB with 50K HHP. Given a HUB with 30K subscribers and 80Gbps bandwidth, the virtualized DOCSIS MAC needs about 300 X86 cores which means it needs 25x2U Fusion Server.

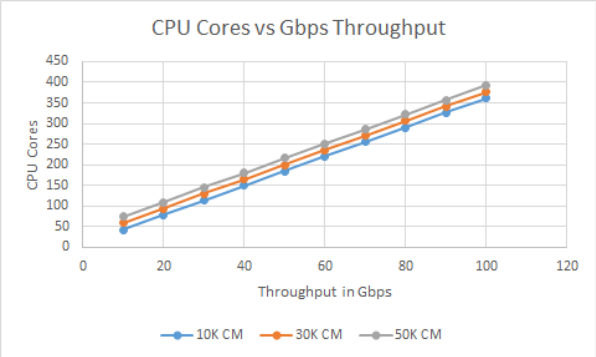


Figure 4. Virtualized MAC CPU consumption

Figure 5 gives the power consumption comparison between the DOCSIS MAC X86 CPU implementation, FPGA implementation and ASIC implementation. We can see the X86 power consumption is 38.35 watts per Gbps bandwidth, it is really a huge number compared to FPGA and ASIC.

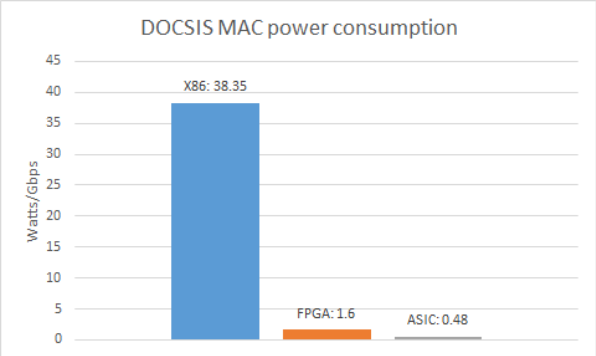


Figure 5. MAC power consumption comparison

4. Summary

NFV changes various network devices to generic hardware and specific software, virtualizing some network devices can get a huge benefit with less pay but some devices may be not.

This paper provides a detailed evaluation of virtualizing DOCSIS MAC by generic CPU, base on the data in this paper, we can see there is still a big gap between DOCSIS CPU implemented MAC and FPGA MAC. The power consumption of X86 CPU MAC is 24 times of FPGA MAC to achieve the same bandwidth, and the hardware cost is about 77 times. We believe the generic hardware will get faster and faster in the future, but 24 times is not the number can be eliminated in short term. Now the generic X86 servers in the cloud are designed for applications which need huge memory and storage, actually the access forwarding device don't require that, that's why the X86 implemented DOCSIS MAC is lack of economic competition. If some new generic hardware platform designed for forwarding device comes out in the future, the DOCSIS MAC NFV dream may come true.

Full Duplex DOCSIS

John T. Chapman, Cisco Fellow & CTO Cable Access, jchapman@cisco.com
Hang Jin, Distinguished Engineer, hangjin@cisco.com
Cisco

Abstract

A method is proposed to use the same spectrum for simultaneous downstream and upstream transmission. This technique could allow a DOCSIS 3.1 plant to run at 10 Gbps downstream and 10 Gbps upstream in the same spectrum.

This technology would boost the upstream throughput by 50x what it is today without having to extend the maximum frequency of the plant. It will provide the ability to deploy true symmetric services.

This white paper will discuss the hardware and software technology of FDX. It will also discuss different deployment plans and options.

Cisco started working on a full duplex concept for DOCSIS in 2014. In early 2015, the FDX DOCSIS project received internal research funding from the Cisco Technology Fund. This white paper is a result of that work and is checkpoint half way though our project.

TABLE OF CONTENTS

TABLE OF CONTENTS 1
INTRODUCTION 2
SPECTRUM PLANS & USE CASES 4
- WHAT IS POSSIBLE? 4
- FREQUENCY TEMPLATES 4
- STANDARD FDD SPECTRUM PLANS 5
- VIDEO RESTRICTIONS 6
- FDX HIGH BAND 7
- FDX MID BAND 7
- FDX LOW BAND 8
- FDX SPECTRUM SUMMARY 9

DEPLOYMENT CONSIDERATIONS 10
- SIMPLIFYING ASSUMPTIONS 10
- PASSIVE PLANT MODEL 10
- ADJACENT DEVICE INTERFERENCE ANALYSIS 11
- FDX OPERATION MODE 13
- ECHO CANCELLATION AT RPD 13
- INTERFERENCE MITIGATION AT CM 14
- INTERFERENCE SUPPRESSION AT CM 14

OPERATIONAL DETAILS – MAC 19
- DOWNSTREAM CONSIDERATIONS 19
- DOWNSTREAM DATA PATH TRAFFIC 21
- UPSTREAM CONSIDERATIONS 22

FDX SCHEDULING 23
- INTERFERENCE GROUPS 23
- FDX – FDD 25
- FDX – TDD 26
- FDX – FDD/TDD 28

CONCLUSIONS 28
REFERENCES 29

INTRODUCTION

Looking Back

For the first 30 years, say the mid-60's to the mid-90's, the cable plant was just a cable plant. There was no fiber and generally there was no upstream return path. When VOD services where introduced in the 80s and 90s, there was a need for upstream signaling for STB and an upstream OOB carrier was established (there was already a broadcast downstream OOB carrier).

This large coax system would not scale for data so in the 1990's optical nodes where deployed to create an HFC plant that had a limited number of households passed in a coax segment so that the noise funneling from the homes was tolerable enough that data could be passed over the upstream path.

Data started off with a shared 2.5 Mbps upstream channel. That later grew to 10 Mbps with DOCSIS 1.0 and to 30 Mbps with DOCSIS 2.0. The number of channels that the silicon could support also grew from one to four. Today with bonding and DOCSIS 3.0, the return path data capacity is around 90 Mbps in a 42 MHz or 65 MHz return path.

In practice, the spectrum below 42 MHz is full of legacy traffic and will likely stay as ATDMA traffic. Although the OFDMA traffic from DOCSIS 3.1 could theoretically be mixed with the ATDMA traffic, the bandwidth gain provided by OFDMA and LDPC would be quickly offset by the time and frequency guard time requirements that must exist between DOCSIS 3.0 and DOCSIS 3.1 spectrum.

With DOCSIS 3.0 and DOCSIS 3.1, it is possible to run a higher frequency upstream return path. The two favorite options are 85 MHz and 204 MHz. Both these return paths push the start of the downstream spectrum to a higher frequency. In todays FDD (Frequency Division Duplex) system, new upstream spectrum is taken from existing downstream spectrum. As an added penalty, more bandwidth is lost to the cross-over band as the split increases in frequency.

Enter Full Duplex.

Today's plant is run as FDD (frequency division duplex). The proposal is to run the plant in FDX (full duplex). This comparison is shown in Figure 1.

The physical path of a N+0 FDX DOCSIS system is shown in Figure 2. The

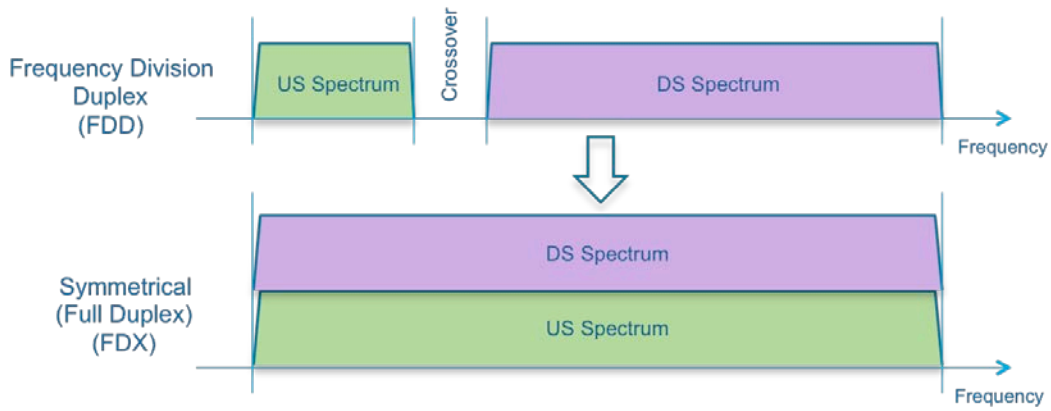


Figure 1 - FDX Spectrum

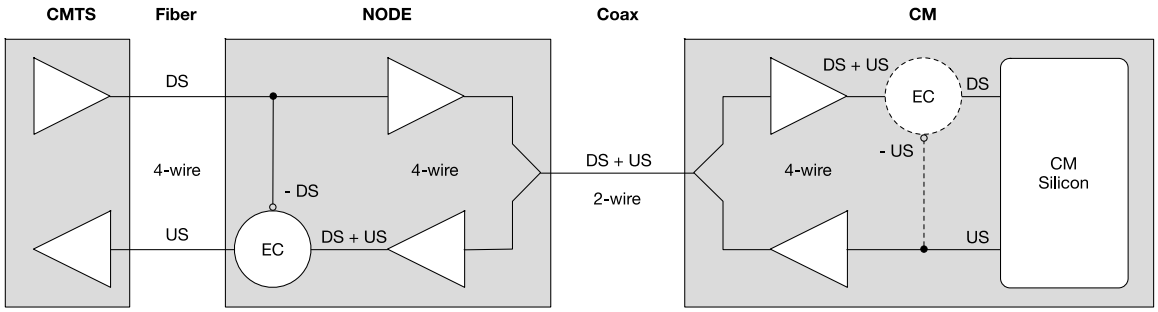


Figure 2 - FDX Path

CMTS is a 4-wire path. A 4-wire path is a path with separate transmit/ground pair and a receive/ground pair. The CMTS signal travels to the optical node where is converted to a 2-wire path. A 2-wire path is where transmit and receive signals are on the same wire and there is a common ground.

Note that energy can travel simultaneously in opposite directions on a passive coax. It is only when amplification is needed that the transmit and receive signals need to be separated into different paths. For the node and the CM, this can be done with a power splitter/combiner.

The combined DS and US signals are combined and arrive that the CM receiver. As we will see, this is not a problem that is solved by an echo canceller, as there can also be interference from adjacent CMs.

In the reverse direction, the upstream path from the CM leaves the 4-wire domain of the CM and becomes 2-wire. The combined DS + US energy from the coax enters the node upstream. Here there is an echo canceller. The echo canceller may be a combination of an analog echo canceller followed by a DSP based digital echo canceller. If the echo canceller can calculate the proper level of the DS signal to cancel out from the DS + US signal, only the US signal will be left. In practice, the remaining attenuated DS signal becomes the new noise floor.

Before diving into the technical details of how the system works, lets look at how the system might be deployed. Lets define the problem first, and then look at the solution.

SPECTRUM PLANS & USE CASES

What is Possible?

Full Duplex DOCSIS has two fundamental impacts on the DOCSIS implementation. The first impact is the full duplex operation itself. This includes the echo cancellers instead of diplexers, interference mitigation, TDD or FDD CM operation, and a series of RF challenges. The second impact is just more upstream bandwidth. Maybe 10 to 50 times more bandwidth. That's a lot.

This section looks at the impact of that available bandwidth and how it would be used. To define a product that is deployable, we need the following things to be true:

1. Working CM and CMTS products with the right silicon and software.
2. Working Optical Nodes at the frequencies of interest. This implies a higher frequency and higher power return path.
3. A spectrum plan that makes sense.

Frequency Templates

To illustrate the FDX spectrum plan, it is convenient to use a log 2 frequency base. This allows observation of the many frequencies of interest at lower frequencies. This is shown in Figure 3. Note that as you look towards the right of the figure, the amount of spectrum per inch is doubling. So on the left, an inch might be 25 MHz of bandwidth but on the right it is 400 MHz of bandwidth.

For the sake of calculating throughput and providing comparison between all approaches, the DOCSIS 3.1 OFDM spectrum will be evaluated at 4096 QAM in the downstream and 1024 QAM in the upstream. In practice, some deployments may require lower modulations at the higher frequencies due to lower CNR.

The spectrum graphs will also include basic services such as OOB (Out-of-Band), FM, and Video.

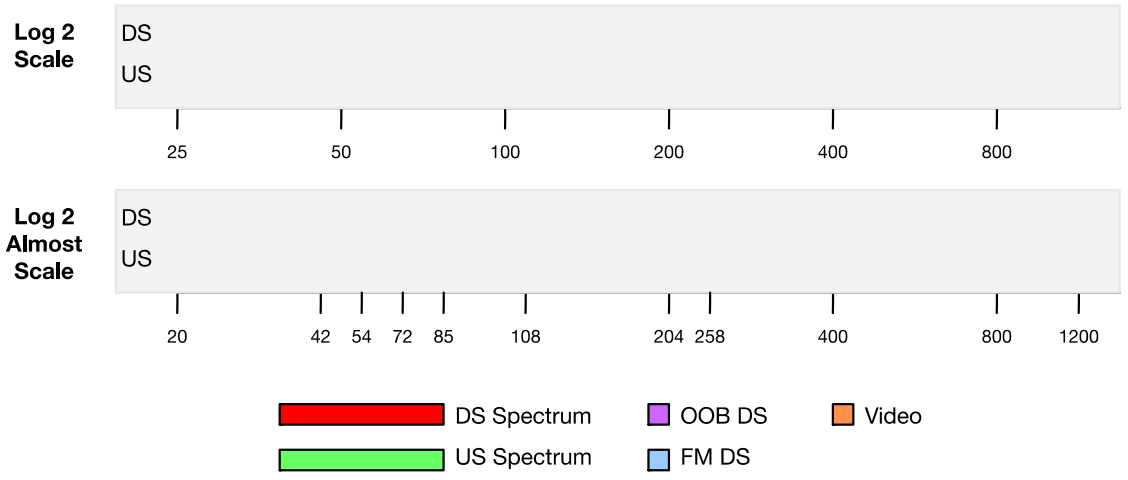


Figure 3 - Frequency Template

OOB Channel

The OOB is very significant as almost all North American upstream expansion plans are limited by the location of the OOB channel. The OOB channel is well over 20 years old and was designed prior to DOCSIS as a two-way channel for managing set-top boxes. The OOB downstream was specified as being located between 72 MHz and 130 MHz. The typical deployment for OOB was the 4 MHz gap at 72 to 76 MHz that is between channels 4 and channel 5.

As we will see, the existence of the OOB channel interferes with the growth of the return path in North America.

FM Band

The FM band in North America is more about interference. Caution is taken that the over-the-air FM band does not interfere with video signals on the cable plant, and that cable signals on the plant do not interfere with over the air FM signals. FM used to be service over cable in North America, but has long been removed.

In Europe, the FM band is an actual service over the HFC plant. Subscribers can plug in their FM radios to the cable plant and receive FM radio. Since it is a one-way service, the cable operators cannot tell how many people actually use the service and thus how many subscribers would be disrupted by the removal of that service.

As we will see, the existence of the FM band interferes with the growth of the upstream band in Europe.

LTE Band

Another band worth noting is the LTE band located between 698 MHz and 806 MHz. LTE shares the same concern that the FM band does. Good HFC plant design with low leakage must be done so that LTE does not interfere with cable signals and cable signals do not interfere with LTE.

Standard FDD Spectrum Plans

In Figure 4, three common spectrum plans that were defined with DOCSSI 3.0 and DOCSIS 3.1 are shown.

Figure 4 - Standard FDD Spectrums

42 MHz Return

The first plan is the 42 MHz return path. This is the default return path deployed in North America. Europe uses a 65 MHz return path. In the 42 MHz return path, frequencies below 20 MHz are generally not usable. Frequencies close to 42 MHz, say above 40, are often not usable due to group delay. The remaining spectrum is generally good enough to support three 6.4 MHz and one 3.2 MHz D3.0 upstream channels with a throughput of about 90 Mbps.

The maximum downstream frequency depends upon the age of the HFC plant. The example provided in Figure 4 is 750 MHz. Almost all plant is at least 750 MHz today. Newly upgraded plant is 870 MHz and 1002 MHz.

85 MHz Return

This is a new return path option that is just now being deployed. The return path is extended up to 85 MHz and the forward path begins at 108 MHz. Since it is a new install, it can also take advantage of the new optical nodes that can support 1218 MHz.

To accommodate 85 MHz, the OOB channel is moved up in frequency toward the upper end of the FM band, just below the new 108 MHz start of the downstream band. The analog or digital video channels from 54 MHz to 108 MHz must also be removed.

For the North American plant, the additional bandwidth is about 40 to 45 MHz, depending upon group delay limitations. Compared to the previous 20 MHz, an 85 MHz return path has about 3.5x the bandwidth as before. In data capacity, this is about 300 Mbps.

204 MHz return

This alternative return path is 5 to 204 MHz with the downstream starting at 254 MHz. This upstream has 10x the effective spectrum of a 42 MHz return path system with an effective shared throughput on the order of 1.4 Gbps (this is based upon 1024 QAM, 40 MHz to 204 MHz, adding ~1.3 Gbps to the existing 90 Mbps from D3.0 located at 20 MHz to 40 MHz).

1.4 Gbps in the upstream is *5 times more bandwidth that an 85 MHz return path* (1.4 Gbps / 300 Mbps) and *16 times the bandwidth of a 42 MHz return path* (1.4 Gbps / 90 Mbps).

The thing to notice about the 204 MHz return path solution is that there is no support for OOB or for FM. For FDD, both OOB and FM are downstream services and must be removed for the upstream return path to extend to 204 MHz.

Conversely, it is the existence of OOB and/or FM that is preventing many operators from using a 204 MHz return path. This is one of the features that FDX could potentially address.

Video Restrictions

In the beginning, analog video occupied the spectrum from 54 MHz to 552 MHz. Digital video was then added from 552 MHz to 750 MHz. DOCSIS was added either into the 550 to 750 spectrum or above 750 MHz.

Today most, if not all, analog video has been removed from today's HFC spectrum. The digital video has been moved into the lower part of the spectrum. There are several concerns that limit the location of video.

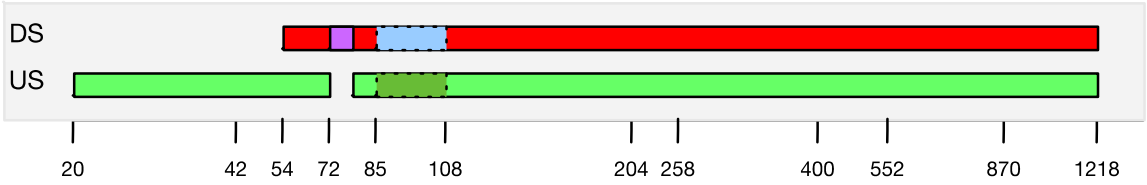


Figure 5 - FDX 1218 MHz with no Video

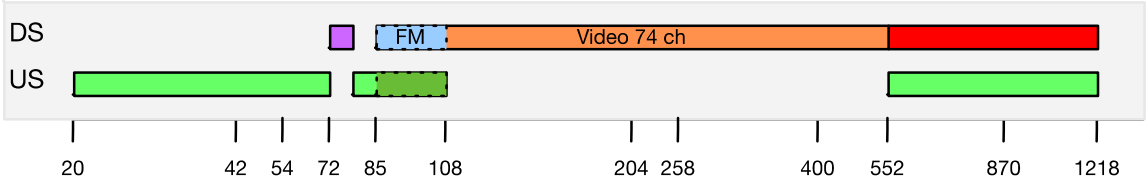


Figure 6 - FDX 1218 MHz with Video

The first is legacy STBs and DTAs (Digital Terminal Adaptors). The upper end of these devices are either 750 MHz or 870 MHz. Some even got to 1002 MHz. However, 750 MHz is the comfortable upper limit.

The next impact is the LTE band. The LTE band starts at 698 MHz and extends to 806 MHz with even more bands beyond that. DOCSIS is more tolerant to a lower CNR (Carrier to Noise ratio) than video, so the general preference is to use the spectrum above 696 (6 MHz band edge) for DOCSIS only.

This means that video – broadcast, SDV, and VOD – generally is between 54/108/258 MHz and 696 MHz.

FDX High Band

Lets start with the maximum FDX use case shown in Figure 5 where the entire upstream would support FDX. One of the rules that will be discussed is that the FDX upstream is not allowed under any broadcast service such as MPEG-TS video. Thus, this solution would not support any MPEG-TS video. All video would have to be video over IP over DOCSIS.

The DOCSIS throughput of this solution could be 11 Gbps x 9.6 Gbps.

This solution would require high power return path amps in the CM that would be able to transmit at 1218 MHz and have enough power to cover the increased attenuation at these higher frequencies.

In Figure 6, a MPEG video band from 108 to 552 MHz (74 channels) has been added. This pushed the FDX action to a high band above 552 MHz. The DOCSIS throughput for this scenario is 6.4 Gbps x 5.8 Gbps. As the MPEG video is decreased, this scenario will approach the previous scenario.

If a video guard band is required (see Section “Video SC-QAM”), there will either be less video channels or less DOCSIS FDX upstream throughput.

FDX Mid Band

Maybe a 10 Gbps upstream is not needed in the near term, or the cost of the amplification and silicon is too much or the power of the return amp is too much. There

may be an interim step partially way through the band. As a somewhat arbitrary mid-point, the spectrum in Figure 7 uses 552 MHz. Years ago, 552 MHz was a top frequency in the downstream. 552 MHz is also a good number that matches current analog-to-digital converter performance.

In this scenario, the video is pushed above 552 MHz. To stay below 696, there is only room for 24 QAM channels (at 6 MHz each). This may actually be enough for a plant with small node sizes – say 80 HHP – and where the main video services are over IP and 24 channels are for legacy.

If a video guard band is required, the upstream may be limited to 435 MHz (552 / 1.27).

FDX Low Band

The FDX low spectrum scenario is targeted at existing silicon and existing return path amplifiers, both of which can operate at 204 MHz.

In Figure 8, the upstream spans from 20 MHz to 204 MHz. 2 MHz is skipped to allow the OOB channel at 72 MHz. Below 20 MHz is omitted just to allow proper comparison of spectrum. The FM band may also be skipped for European operation.

In this scenario, it is also illustrated how the 20 to 42 MHz bandwidth may be used by legacy DOCSSI 3.0 and earlier ATDMA services. 42 MHz to 72 MHz are left for non-FDX DOCSIS 3.1 services. OOB is 72 MHz to 74 MHz. Above 74 MHz is FDX. This example also leaves a video guard band between 204 and 258 MHz. There are many variations of these themes.

In this particular scenario, there is room for up to 73 video channels. If all those video channels are used, the data capacity of the system is 7 Gbps x 1.5 Gbps. This is sufficient bandwidth to support a 1 Gbps

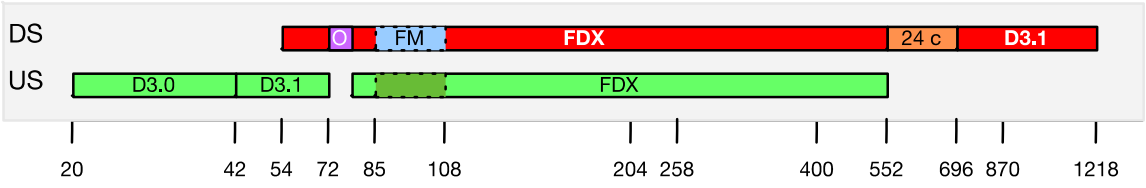


Figure 7 - FDX 552 MHz

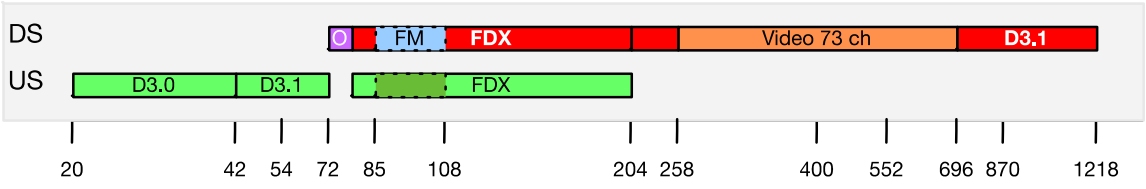


Figure 8 - FDX 204 MHz with Overlap

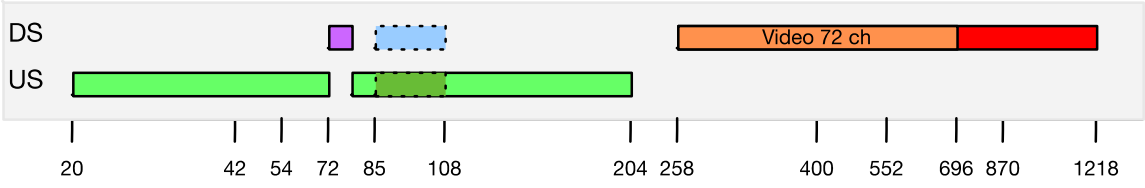


Figure 9 - FDX 204 MHz without Overlap

upstream service tier.

A variant of this channel plan is shown in Figure 9 where these is actually no overlap in spectrum. However, it is still a FDX system as there is no diplexer in the system. As a result, the upstream spectrum can be multiplexed with the OOB and/or the FM services in the downstream.

The throughput of this system, assuming up to 72 channels of video, is 5.2 Gbps x 1.4 Gbps.

This example illustrates that with sufficient downstream bandwidth, it may not be necessary to add complexity to recover downstream bandwidth from 54 MHz to 258 MHz.

In this simplest of deployment scenarios, with no over lapping spectrum, FDX allows an operator to retain OOB and/or FM and also deploy a 204 MHz upstream path.

FDX Spectrum Summary

Table 1 - FDX Spectrum Choices

Option	**Low Band**	**Mid Band**	**High Band**
US Max Frequency	204 MHz	552 MHz	1218 MHz
DS Max Frequency	1218 MHz		
Max Data Capacity	11 x 1.5 Gbps	11 x 4.2 Gbps	11 x 10 Gbps
Cost	$	$$	$$$
TTM	Sooner	Later	Maybe Even Later

In this section, we have discussed three fundamental deployment scenarios. These are a low band, mid band, and high band. These solutions are summarized in Table 1.

There are many variations of these themes. The common consideration fro design, though, it the maximum upper frequency. So a system that extends to 1218 MHz in the upstream is still a high band system, regardless if it uses a 100 MHz or 1000 MHz of spectrum for the upstream FDX path.

In the grand scheme of things, the choice will be driven by the intersection of:

1. CM silicon availability and price
2. Return path bandwidth/power availability and price, for CM, amps, and node.
3. CMTS silicon and throughput capability
4. The needs of the market for higher bandwidth services.

The low band solution is very interesting if existing silicon works, but that depends if higher bandwidth silicon become available before the FDX market is established.

The mid band solution is very interesting in that it delivers more than enough bandwidth for the next few years without stressing out the design elements and without having to deal with the extreme attenuation at higher frequencies.

The high band solution is interesting if technology permits it to come into existence at the right price point. Then we are done.

DEPLOYMENT CONSIDERATIONS

Simplifying Assumptions

N+0 (Deep Fiber)

FDX requires careful use of echo cancellers. Every time there is a 4-wire to 2-wire conversion, the signal from one direction become noise in the other. The easiest scenario to design for would be a passive coax plant where the only 4-wire to 2-wire conversion occurs on the optical node and the CM.

This is not to say that an FDX amp could not be inserted, but success would be declared if the system worked with N+0. (Optical Node plus 0 amps)

By definition, since there is no amplification, the length of the HFC plant is on the order of 500 to 1000 feet.

R-PHY Only

Since the preferred architecture for deep fiber is Remote PHY, a good starting assumption is that the FDX node will be an R-PHY Node. That allows the RPD (Remote PHY Device) silicon to potentially host FDX circuits and allows the FDX to re-use the ADC and DAC components in the RPD silicon.

Same Home Works

This assumption recognizes that when an FDX CM is installed in a home, whatever changes are required to make FDX for that home to work can be assumed. This might extend to a swap out of all the video gear in that home.

This reduces over specifying the FDX solution and focuses the FDX technology on being compatible with equipment in neighbor's homes that are separated by the tap.

Silicon

It is recognized that the CM and CMTS silicon will have to be upgraded to include any protocols changes but also to accommodate the increased upstream speeds.

Passive Plant Model

A passive plant is referred to as N+0 for optical node followed by 0 amplifiers. A node typically has 4 inputs/outputs. Each I/O feeds an independent segment of plant.

Each plant segment is composed of a series of 5 to 7 taps. Each tap has 1 to 8 (typically 4) drop cables to the home.

So, on a maximum dense node, there would be 7 taps x 4 HHP/tap = 28 HHP per I/O . This would be approximately 120 HHP for a node with 4 I/O. Given large neighborhoods, a minimum might be 25 HHP per node with an average being 60 HHP per node.

For FDX, each node I/O is isolated, so and CM to CM interference is isolated to a single segment.

Since the first four to five taps have a higher downstream signal and a lower upstream signal, the interference groups discussed later will tend to be per tap. The last few taps may get combined into a single interference group.

Adjacent Device Interference Analysis

Table 2 - Adjacent Device Interference Analysis

Scenario	Different Frequency	Same Frequency
Same Home	• Impact • Solve at install time within the home with new equipment • Allow for QAM video in the spectrum	
Adjacent Home	• Should be no impact • Evaluate each scenario	• Impact likely • CMTS Interference Mitigation Required
Remote Home	• No impact	

When the FDX CM is installed, it may interfere with legacy equipment. That equipment can include other DOCSIS CMs and other STBs. Any interference has be either acceptable or mitigated somehow.

Based upon location, there are three scenarios as show in Table 2. The type of interference can be within the same piece of spectrum (co-channel) or an adjacent piece of spectrum (adjacent channel).

ADI – Same Home, Same Frequency

This only applies to FDX and will be managed using interference mitigation (see section "Interference Mitigation").

ADI – Same Home, Different Frequency

Devices within the same home as the FDX CM would be the most subject to interference. The default plan is to replace those components with components that work with an FDX CM. Some allowance may still have to be provided for video.

Regardless, there may be a need for a video guard band for MPEG QAM video if the video receiver is relying on the attenuation of a diplexer for it receiver isolation. This would only impact spectrum allocation. This item is under study.

ADI – Adjacent Home, Same Frequency

This is only going to happen with other FDX devices. FDX scheduling will be used to prevent CMs that can interfere with each other with actually interfering with each other. See Section FDX Scheduling" for more information.

If the downstream service is broadcast in nature, the upstream spectrum will be notched. See the section "Downstream Considerations" for more details.

ADI – Adjacent Home, Different Frequency

The testing results in the SCTE 2012 White Paper: "Do You Have Legacy Issues in Your New Upstream Plant?" by John T. Chapman [1] showed that there should be enough attenuation between homes combined with enough out-of-channel rejection and that STB in adjacent homes where not impacted by high power upstream carrier. The reference diagram used in [1] is shown in Figure 10.

This should also be true for CMs, although a guard band may have to be observed. More study is needed.

ADI – Remote Home

For devices located far from the home, either outside of the interference group or even on the other side of the optical node, there is not interference. This case is worth mentioning, but no work needs to be done.

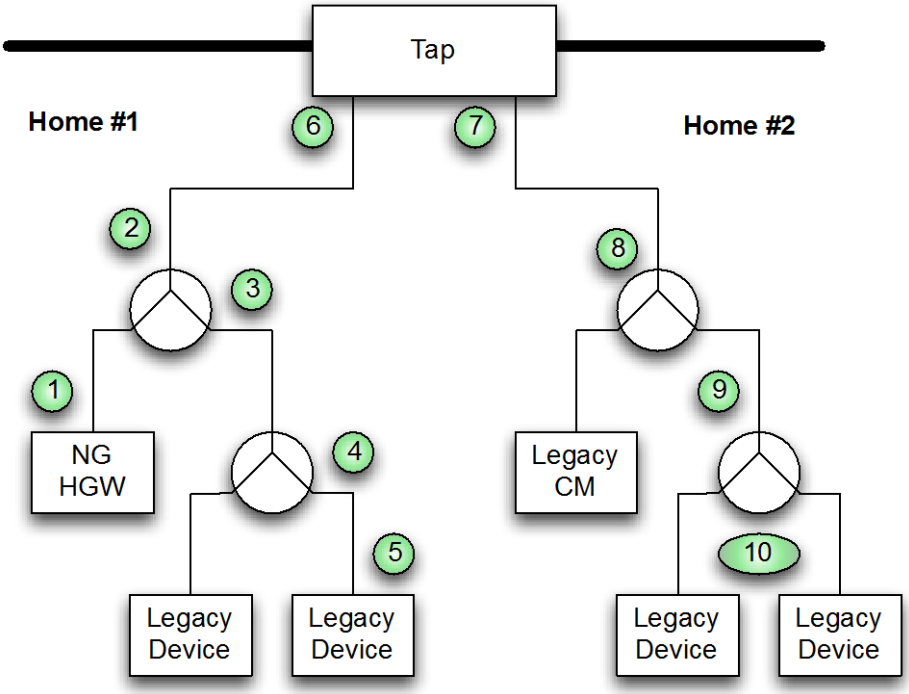


Figure 10 - Adjacent Home Splitter Model

OPERATIONAL DETAILS – PHY

Here is where we get into the nitty gritty of how the PHY layer works. If you are the RF type, this section is for you. If you are not the RF type, feel free to jump to the next section and tell your buddy to read this section.

FDX Operation Mode

A hybrid operation mode is adopted for HFC full duplex network. Same spectrum will be used twice at the same time, one for DS, and one for US. From the RPD point of view, it is a true full duplex operation, it transmits and receives signal on the same spectrum at the same time, but the same spectrum is not assigned to the same CM. So, CM still runs in the simplex mode. This hybrid mode requires minimum changes to CM.

Since HFC runs in a hybrid mode (i.e. full duplex on RPD, simplex on CM), the interference resulting from full duplex operation will be different in the RPD and in the CM. The RPD receiver sees the co-channel interference and adjacent channel interference coupled from its own transmitter.

The CM receiver sees the adjacent channel interferences from its own transmitter and sees both co-channel and adjacent channel interferences from neighboring CM transmitters.

Echo Cancellation at RPD

To suppress the co-channel and adjacent channel interference at RPD, echo cancellation must be implemented at the RPD for supporting full duplex operation. The echo cancellation will be implemented in two stages: the RF stage and digital stage.

Echo cancellation at RF stage

Echo cancellation at RF stage will cancel out the echo before it hits receiver ADC. RF echo cancellation is readily done through port-to-port isolation. Typically one could have a 2:1 combiner that offers 30dB ~ 35dB port-to-port isolation for the whole bandwidth (10 MHz to 1218 MHz).

Higher port-to-port isolation can be achieved. Nevertheless, one needs to consider all the possible paths through which the transmitter signal is coupled/ or leaked to receiver.

Direct coupling

Transmitter signal can be coupled to receiver locally at RPD due to limited port-to-port isolation. Improving port-to-port isolation will help reduce the echo that is coupled directly through the 2:1 combiner.

Reflections

The transmitter signal can go down the coaxial cable and get reflected back at any HFC dis-continuities. In full duplex operation, any reflections of transmitter signal will appear as co-channel interferences on receiver. For example, say there is a tap 100 feet away from the RPD and its input port has 20dB return loss.

A transmitter signal will be then be reflected back with 20dB attenuation. Consider the loss of 100 foot cable (both ways) and a 4dB loss of the 2:1 combiner, when it hits the receiver, the signal level of the reflection will be -25 dB to -30 dB down compared to the transmitted signal level. So,

effectively, you will have 25 dB to 30 dB port-to-port isolation, even there is no direct coupling.

One could build a RF echo cancellation to cancel out not only the echo through the direct coupling, but also the reflection. To achieve this, you need an adjustable attenuator with a resolution of a fractional dB, and a tunable phasor with a resolution of <1ns, and they all need to work from 100 MHz to 1218 MHz.

You probably need to cascade a few pairs of them as there may be multiple reflections resulting from multiple discontinuities. As we speak, a tunable phasor with <1 ns resolution and 100 MHz to 1218 MHz bandwidth does not exit.

In summary, echo cancellation at RF stage has its limitation: it can cancel out the direct coupling, but will hardly cancel out all the reflections. In reality, most of echoes seen by the receiver result from reflections, which is in range of 25dB to 30dB down from the transmitter signal level.

A 2:1 combiner with 35dB port-to-port isolation is adequate, as it has suppressed the direct coupling echo under the reflections.

Echo cancellation at digital stage (DSP)

Any echo residue that does not get cancelled out at the RF stage will pass through the ADC and be further reduced through digital echo cancellation. Digital echo cancellation utilizes modern DSP algorithm and can achieve up to 55 dB echo cancellation.

The actual echo cancellation required depends on the echo and input signal levels that in turn depends on transmit and receive signal levels and HFC network condition.

Interference Mitigation at CM

In a full duplex HFC network, CM still runs in simplex mode, which means CM will not transmit and receive on the same frequency/channel at the same time. The FDX is realized through pairing CMs in the network: one CM transmit at certain frequency/channel, and other CM may receive at the same frequency/channel if there are enough isolation among them.

Which CMs can be paired for FDX operation depends on the isolation among them: sufficient RF isolation is required to ensure the receiving CM will not be impaired by the interference of the transmitting CM if they are on the same frequency/channel. The isolations among CMs depend on the HFC topology and will vary case by case.

To sort out specifically the isolation among CMs, a sounding scheme for supporting FDX is proposed. In the sounding scheme, a CM transmits with a known signal level, and all other CMs listen and report the received signal levels.

Based on the reported received signal levels, CMTS scheduler will know exactly the path losses (isolation) among CMs, and which CMs may or may not be paired for FDX operation.

The DS and US scheduler will schedule the DS and US transmissions accordingly based on these pairing information.

Interference suppression at CM

To further suppress the interferences among CM, one could leverage the subcarrier orthogonality of OFDM. OFDM subcarriers remain orthogonal if their frequencies are perfectly synchronized and times are synchronized within cycle-prefix (CP) window. As a result, there are no adjacent

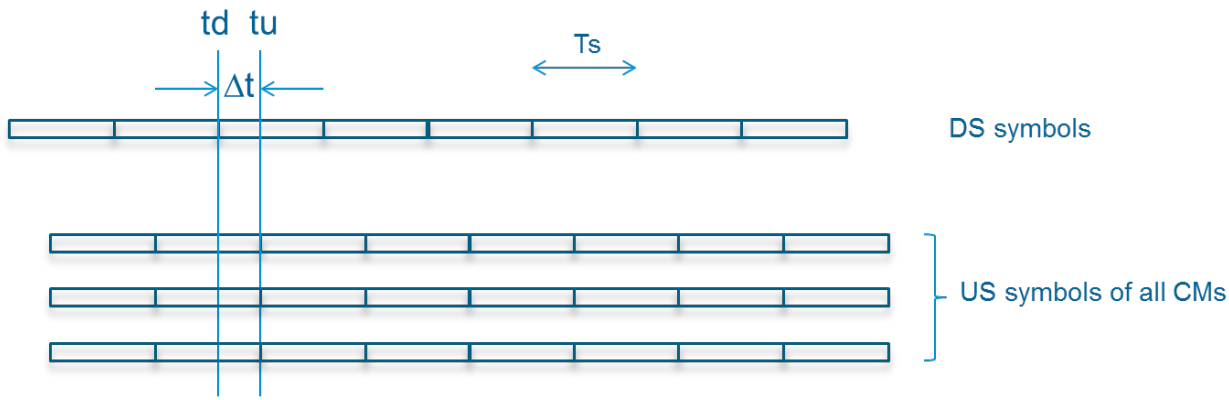


Figure 11 - DS and US Symbol Alignment

carrier leakages among neighboring subcarriers.

Therefore, if there are active subcarriers on certain spectrum and null subcarriers (excluded) on its neighboring spectrum, there will be no energy leakage from the active subcarriers onto the null subcarriers. DOCSIS 3.1 requires subcarrier orthogonality among all the DS subcarriers, and subcarrier orthogonality among all the US subcarriers.

Please note: DOCSIS3.1 requires DS and US are synchronized on frequency but does not require them to synchronize on timing. In reality, DOCSIS 3.1 DS and US timing may have arbitrary offset. As a result, DOCSIS 3.1 DS and US subcarriers are not orthogonal, and DS subcarriers may interfere with US subcarriers if they locate on their neighboring subcarrier, and vise versus.

In the case of FDX operation, a CM receives composited signals that are a superposition of its desired DS traffic and the US interferences from its neighboring CMs.

The subcarriers of the US interference need be aligned with the subcarriers of the DS in frequency and time to make them orthogonal if one likes to utilize the subcarrier orthogonality to suppress the US interferences to DS.

To align DS and US timing, DS and US must use the same cyclic prefix (CP) and symbol lengths and have timing alignment on symbol boundaries. This is shown in Figure 11.

Where:

- Δt is the timing offset that can be set:
- Δt (ns) = 0 to align DS and US symbols at RPD
- Δt (ns) = 2*L(ft)/0.87 will align DS and US symbols at a distance L from RPD

One could set L equal to the max coverage of the RPD, which effectively to make the DS and US subcarriers synchronized at the node edge.

With a single Δt, there will be some offset between US and DS timing at CM, depending on the CM distance to RPD. Given the N+0 deployment case, the offset will be within or close to CP length, thus maintaining ~100% orthogonal (quasi-orthogonal) among DS and US subcarriers throughout the whole node coverage .

This will ensure there are no or little US interference to DS on the adjacent subcarriers.

FDX US to DS pilots and PLC

The subcarrier orthogonality between the DS and the US through the DS and the US timing alignment is required to prevent the DS pilots and the PLC from any US interference. CMs need to lock to the DS pilots and the PLC to remain frequency and time synced with the RPD.

There are three types of DS pilots:

1. Preamble in PLC – 8 symbols x 8 or 16 subcarriers.

2. Continuous pilots – pilots on every symbols and on some fixed subcarriers

3. Scattered pilots – pilots scattered over time and frequency

To avoid the DS pilots from being interfered from any CM US transmissions, the CMTS must designate all the subcarriers that are occupied by the PLC and all continuous pilots as excluded subcarriers for US transmissions. As PLC and continuous pilots take <1% spectrum, leave them un-used for US wouldn't cause much overhead.

As explained earlier, due to the subcarrier orthogonality through DS and US timing alignment, as long as those subcarriers that are occupied by the pilots and PLC are excluded from US transmissions, there will be no or little interferences from US adjacent subcarriers on the continuous pilots and PLC.

CM uses scattered pilots for channel estimation, thus it needs scattered pilots only on the channels where it has DS traffics.

Full duplex DCOSIS enables full duplex operation for the network (same frequency used for both DS and US at the same time), but each individual CM is still on simplex, i.e., its DS and US frequencies do not overlap. Thus CM US does not interfere with its own DS, including the scattered pilots in its DS traffics.

Moreover, CMTS has the knowledge on whose US will interfere with whose DS through the sounding process and the scheduler will avoid assigning any US transmissions that may impair the DS of any CM in its coverage.

The scheduler will assign the US resources in the 2-dimensional time and frequency space and ensure no US/DS overlaps among interfering CMs.

Pilots are just one type of DS signal. In a big picture, full duplex DOCSIS needs to solve the interferences of:

a) RPD transmission (DS) to its reception (US) at RPD

b) CM US to its own DS at CM

c) Adjacent CM's US to adjacent CM's DS at CM

d) CM US to DS broadcast messages.

a) is resolved through echo cancellation, b) is resolved by limiting CM to simplex, c) is resolved through sounding and intelligent scheduling, and d) is resolved through null (excluded) US transmissions on the spectrum where DS broadcasts locate on.

FDX US to OOB

The OOB occupies 2MHz spectrum and centers on the spectrum 72MHz to 76MHz. With FDX operation, one needs to protect the OOB from CM US transmissions.

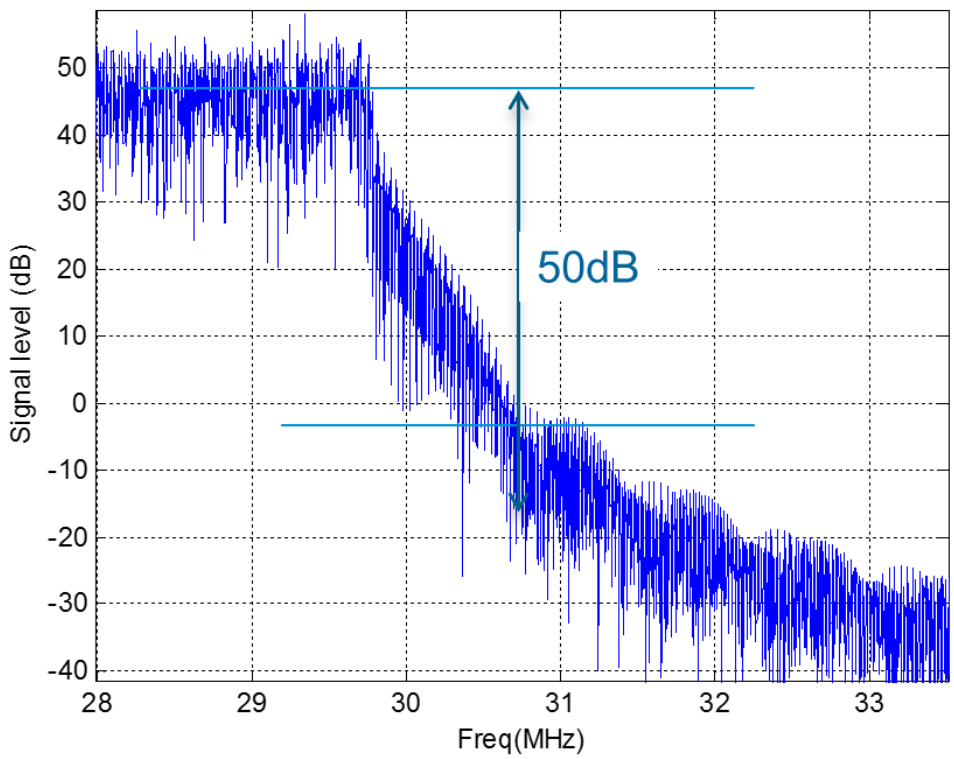


Figure 12 - OFDM Adjacent Channel Leakage

To protect the OOB, the CMTS needs to null the US transmissions on 72MHz to 76MHz, this effectively leaves 1 MHz guard band on both sides. With 1 MHz guard band, OFDM signal leakage into OOB will attenuate by ~50dB.

This is shown in Figure 12. For a CM, the max signal level of US interferences from the neighboring CMs will be 17dBmV/ch (6MHz) (max 57dBmV/ch – 40dB tap-to-tap isolation).

With 50dB attenuation on adjacent channel, OFDM leakage into OOB bandwidth will be -33dBmV/ch. OOB employs low QAM orders with input signal level close to ~0dBmV, it is not expected there will be any issues on OOB reception if the interference level is as low as -33dBmv/ch.

FDX US to FM

The same scheme that is used to protect OOB from FDX US interferences can be used here to protect FM from FDX US interferences.

FM spectrum occupies total 20.5MHz in 87.5MHz to 108MHz. CMTS will exclude US transmissions on 86.5MHz to 109MHz, leaving 1MHz guard band on each side to protect DS FM radio.

FDX US to Video

Similar to OOB and FM, CMTS needs to carve the DS spectrum that is used by video and do not run FDX on that spectrum. As a general rule, leave 1MHz guard band on each side to ensure no or little leakage from OFDM channel into video spectrum.

FDX US to FDX DS – Same CM

In a FDX network, the CM still runs in simplex mode. For each individual CM, the US and the DS will be on different frequency or channel. As a result, CM does not have co-channel interference from its own transmitter.

Nevertheless, this doesn't mean that CM receiver is 100% immune to its own transmissions. For supporting FDX operation, CM needs to have full frequency agility, that is, it's US and DS channels can be assigned on any FDX spectrum, as long as they do not overlap.

For that purpose, the diplexer currently used in CM to separate DS and US spectrums needs be replaced with 2:1 combiner. The common port connect to the HFC coaxial cable, one individual port connects to TX and the other individual port connects to RX.

The US signal will be sent to the HFC through the common port of the 2:1 combiner and DS will be received through the same common port of this 2:1 combiner.

Due to limited port-to-port isolation, the US signal will leak over to the RX. There is a concern that even if the US and the DS are on different channels, and there are no co-channel interferences, the level of the US leakage into the RX may be so high that RX front end may get saturated.

With nominal received signal level 0 dBmV/ch for DOCSIS and 6dB power boosting for video, the total received DS power at CM is 26 dBmV.

The max interferences seen by CM from its own US will be 64 – 35 = 29dBmV, 3dB higher than the nominal DS signal at the CM; where 64 dBmV/ch is the max CM output power, 35dB (average) is the port to port isolation at the CM.

The CM may require some tweaks to its front end to ensure its receiver will not be saturated and can maintain proper operation with the presence of intermittent interference which are 3dB higher in power than its nominal DS received power.

FDX US to FDX DS – Neighboring CM

In an FDX network, the intelligent scheduler will co-schedule DS and US to ensure there are no co-channel interferences to DS. The intelligent scheduler utilizes the sounding scheme and interference groups to achieve this purpose.

Nevertheless, CM can still see US interference on adjacent channels. This could occur with CMs on the same tap or on different taps. A brief calculation shows this adjacent channel interference would not be an issue.

The max interferences seen by CM from its neighboring CM's US will be 64 – 4 – 39 = 21 dBmV, 5 dB lower than the nominal DS signal at the CM; where 4 dB is the 2:1 combiner loss and 39 dB is the minimum isolation among CMs (tap-to-tap isolation).

CM should be able to maintain its normal operation, given the interference level is a few dB below its nominal DS receiver level. In the case where new CM (FDX CM) and legacy CM (non-FDX CM) co-exist in the same network, one could keep the legacy CM on FDD and only operate FDX on new CM.

A new CM may be assigned to transmit on the spectrum that is a part of the DS spectrum for legacy CM as long as the legacy CM does not receive any DS traffic on that spectrum.

Downstream Considerations

When the same spectrum is used for both upstream and downstream transmission, coordination is required. The general scheme used in this white paper is that any one CM does not transmit at the same time on the same frequency.

This poses some challenges for any downstream broadcast/multicast or common upstream spectrum usages. Lets go through each use case and determine the best course of action. Most of these cases will result in upstream exclusion bands and are illustrated in Figure 13.

OOB

The out-of-band channel, as defined by [2] and [3], typically is located at 72 to 74 MHz, which is a gap between channels 4 and 5 [4]. The OOB channel is used to control legacy MPEG set-top boxes. There is also an upstream OOB channel typically located at 8 to 12 MHz.

In a conventional FDD system such as DOCSIS 3.1, the existence of the OOB downstream channel often limits the upstream spectrum to an 85 MHz return path.

The plan would be to notch out the upstream spectrum under the downstream OOB channel. There should not be a need to have guard band.

Since the upstream OOB is at a sufficiently low frequency, it is not impacted by the FDX operation other than having to be included in any upstream digitization of the spectrum.

FM

FM is a service on the downstream plant. Years ago, FM was a service in North America. However, that is generally not the case now. Europe still has FM as a service. Since FM is a one-way service, there is not easy way to determine how many users exist. So it is hard for European operators to judge the impact of removing FM as a service.

In a conventional FDD system such as DOCSIS 3.1, the existence of the FM band often limits the upstream spectrum to an 85 MHz return path.

The plan would be to notch out the upstream spectrum under the FM band. There should not be a need to have guard band.

DOCSIS 3.0 SC-QAM

DOCSIS 3.0 deployments today have

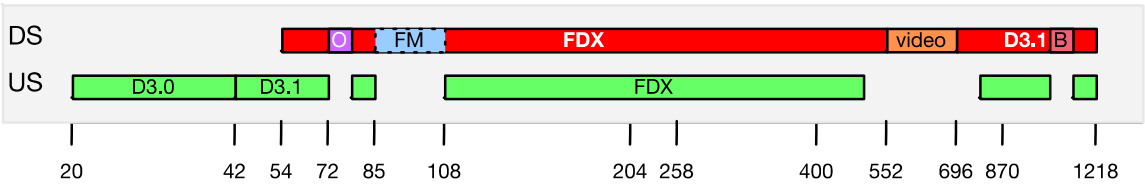


Figure 13 - Downstream Exclusion Bands

between 8 and 32 QAM carriers on the plant. If the FDX protocols, specifically the sounding and the scheduling cannot be retrofitted back into DOCSIS 3.0 and earlier CMs, then the upstream spectrum below DOCSIS 3.0 cannot be used.

If the FDX protocols can be extended to DOCSIS 3.0, then the DOCSIS 3.0 traffic would be broken up into unicast QAMs and multicast/broadcast QAMs.

The unicast QAMs would participate in the FDX protocols, while the broadcast and multicast QAMs, such as primary QAM channels that contain MAPs and other MMM would have spectrum notched out in the upstream.

2 MHz guard band is required between DOCSIS 3.0 and DOCSIS 3.1 per DOCSIS specifications. This 2MHz guard band rule shall be applied here as well to separate DOCSIS 3.0 spectrum and FDX spectrum.

DOCSIS 3.1 Non-FDX

DOCSIS 3.1 in a non-FDX mode is just being deployed. If the FDX protocols – namely sounding and scheduling – can be applied to non-FDX DOCSIS 3.1 CMs and there are no RF issues, then it should be possible to use the upstream bandwidth below the unicast DOCSIS 3.1 spectrum. Otherwise, non-FDX DOCSIS 3.1 spectrum would have to be excluded.

DOCSIS 3.1 PLC

The PHY Link Channel (PLC) is a broadcast channel that is used for CMs to register. It has a preamble for syncing, a timestamp, and some basic signaling. The upstream spectrum must be notched out below the PLC channel.

For supporting FDX operation, DS and US subcarriers need to maintain orthogonality by synchronizing DS and US timing. With DS and US subcarriers orthogonality, no additional guard band is required around PLC subcarriers.

DOCSIS 3.1 NCP

The Next Code Word Pointer (NCP) is a challenge. The NCP exists initially in every symbol and points to the beginning of the next code word. The NCP is then scrambled across time and frequency due to the interleaver process. This makes it hard to omit upstream bandwidth for the NCP field.

The easiest solution is to allow the CM to ignore the NCP when the upstream transmission trashes the downstream transmission. This requires some research to see if this is acceptable. There is one aspect of the NCP where there is a countdown when the NCP modulation is going to change. Missing that count down could be problematic.

Both these problems could be addressed

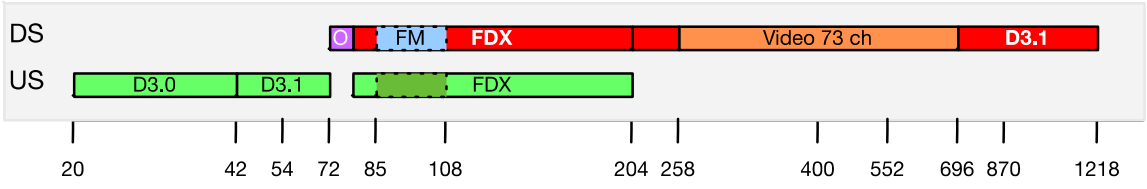


Figure 14 - Video Guard Band

by introducing additional signaling in the PLC channel. This is an issue that requires further study.

Video SC-QAM

Video has three forms – linear which is broadcast in nature, Switched Digital Video (SDV) which is multicast in nature, and Video on Demand (VOD) which is unicast in nature. All these three forms are uncorrelated with DOCSIS and hence with potential FDX protocols. As such, all upstream bandwidth at the same frequency of video must be excluded.

There may be a necessity to have a video guard band as explained in section "Video SC-QAM". If this is the case, here is how to calculate the guard band. The amount of guard band is dependent upon the order of the filter used in the diplexer. This means that it takes a certain amount of frequency band to create enough isolation.

The classic diplexer ration is 1.27. For example:

54 / 42 = 1,285
108 / 85 = 1.27
258 / 204 = 1.265

These ratios vary slightly as the frequency cut-offs are rounded to the nearest 6 MHz channel boundary. So, if the upstream band extended to 204 MHz, video could not start until 258 MHz (204 * 1.27). If video stopped at 696 MHz, and an upper guard band is required, the upstream spectrum could not restart until 882 MHz. (696 * 1.27 rounded down to a channel boundary).

It is not clear at this time that an upper video guard band or even a lower video guard band is required. This is under study.

Another factor that comes into play is the migration of video services to all IP. If a home that upgrades to FDX DOCSIS can also upgrade to all video over IP, then no video guard bands may be required.

Downstream Data Path Traffic

In the earlier section, we covered layer 1 and lower layer 2 messaging such as PLC and NCP. In this section, we will look at the layer 2 and layer 3 traffic that are in the data path. The focus will be on DOCSIS 3.1 traffic.

Downstream Broadcast

This would include the MAP message that is sent to all CMs. It would also include certain MMM (MAC Management Messages). All these are contained in profile A. However, we cannot cut out the spectrum below Profile A as Profile A is multiplexed with the other profiles, and the result is then interleaved across time and frequency.

The only reasonable solution seems to be to isolate the traffic from profile A onto its own channel with its own time and frequency assignment, with its own interleaver. Another solution would be to put the MAP and MMM on the PLC channel, although this might increase latency and exceed that bandwidth of that channel.

Both of these approaches require a change to the usage of the DOCSIS 3.1 protocol. For backwards compatibility, it would probably make sense to make this part of the standard DOCSIS 3.1 operation.

Downstream Multicast

Downstream multicast traffic would refer to any IP multicast traffic over DOCSIS. This typically might be a video stream. It also might be some sort of IP signaling such as an IPv6 signaling multicast. There are at least two solutions.

The first solution is to put all multicast traffic in with the broadcast traffic in a dedicated DOCSIS 3.1 channel. This is simple. It is also useful if you believe that there is not much multicast traffic to begin with, so optimization is not needed.

CMs are grouped into interference groups (IGs). Those IGs could be further sorted to line up with multicast groups. This is a bit more complex, but not impossible.

Upstream Considerations

Upstream Contention Requests

During a contention interval, any CM may send a request. To avoid interference from contention requests, CMTS needs to assign sub-slots for contention requests and exclude the corresponding DS subcarriers from DS usage to avoid interferences from contention requests to DS.

To avoid overhead, the sub-slots will be assigned on fixed mini-slot locations over time.

This would be a *downstream* exclusion band.

Upstream Initial Ranging

During initial ranging, any CM may send a request. To avoid interference from initial ranging, CMTS needs to designate initial ranging zone and exclude the corresponding DS subcarriers from DS usage to avoid interferences from initial ranging signals to DS.

Initial ranging zone does not have a frequency permutation with the rest of US mini-slots. This helps to locate the corresponding DS subcarriers that need be excluded to avoid interferences

To avoid overhead, the initial ranging zone should take a narrow width on subcarriers (take one mini-slot)

This would be a *downstream* exclusion band.

Interference Groups

Sometimes there are just noisy neighbors and you are stuck with them. When they party, you can't hear yourself think. The same is true in the tiny world of FDX.

If one CM is too close to another CM, its upstream transmission may interfere with the other CMs downstream reception. And since the interference is coming from another CM, a classic echo canceller cannot cancel the signal out.

FDX deals with situation using discovery followed by mitigation. Basically, FDX finds the noisy neighbors and forces them to get along.

Interface Group (IG) Discovery

The CMTS is generally not aware of the physical topology of the HFC plant. It may know the name of a particular node group and can attach CMs to that node group, but it does not have observably on which tap a CM is on or how long the cables are.

There is only one real way to determine if one CM will interfere with another CM and that is to test for interference. To do this, one particular CM will send out an interference signal and the other CMs will listen for it. If the transmit power is known and the receive power is measured, an actual CM to CM path attenuation can be calculated. If signals are sent at different frequencies, tilt can also be calculated. Multiple tests will probably be needed that correspond to each channel boundary.

The interference group testing will either re-use existing signaling within the DOCSIS 3.1 complex – such a PNM – or new signaling can be created.. It will be important for the CM to be able to accurately measure the power of the received signal. One concept would be to reserve specific

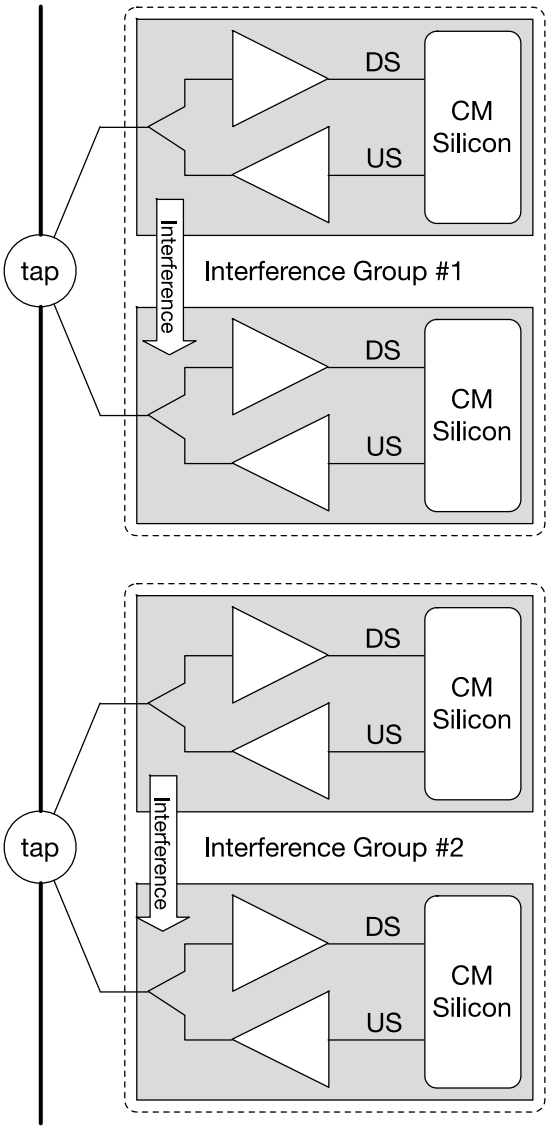


Figure 15 - Interference Groups

frequencies that are only used for IG Discovery. That way there will not be any artificial background noise.

When interference is discovered between two CMs, those CMs are placed into an interference group. Typically, it is expected that each tap group will also be an interference group. The last few tap groups at the end of a line may also group together into an interference group.

Transmission Groups (TG)

IGs are based on actual measured interference between neighboring CMs. If there is good isolation between taps, there may be quite a few independent IG. For scheduling convenience, it may be convenient to have a minimum number of groups.

This goal is achieved by assigning interference groups (IG) into transmission groups (TG). So, whereas IG are groups based upon a physical property, TGs are based upon a logical need.

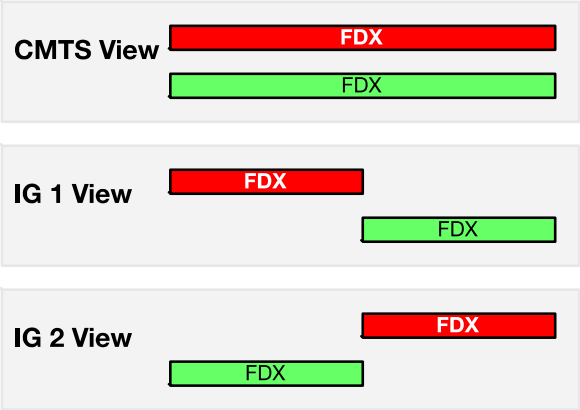


Figure 16 - Duplex - Simplex Scheme

For example, a 80 HHP node may have 20 taps. If each tap were a separate IG, then there would be 20 IGs. For scheduling purposes, a simple scheduler may only need two TGs. IG are assigned to TGs based on traffic density, multicast groups, or any number of criteria determined by the CMTS.

For nomenclature, it is proposed that IGs are numbered where as TG are assigned capital letters. So IG-1 and IG-2 may belong to TG-A.

Interference Mitigation

The rule is that *within an interference group one CM is not allowed to transmit at the same time and frequency that another CM is receiving*.

This rule is enforced by the CMTS through scheduling. To schedule properly, there needs to be a frame work to schedule within and rules that can be enforced. This is what we will explain next.

Duplex-Simplex Scheme

What this really implies is that *a CM IG operates in a simplex mode while the overall plant operates in a duplex mode.* This is illustrated in a simple way in Figure 16. Figure 16 shows frequency allocation within a spectrum at a specific point in time.

So, within an IG (or a TG), the CMs cannot transmit and receive at the same time on the same frequency. They may transmit and receive on different frequencies (FDD) or at different times (TDD). Meanwhile, while one IG is receiving only, another IG can transit only. This is represented in Figure 16 with IG View #1 and IG View #2. This is what is implied with simplex within an IG.

When scheduled right, the entire downstream spectrum, at each point in time, is used for CM reception and the entire upstream spectrum can be used for CM transmission. This is duplex. This is represented in Figure 16 with the CMTS view.

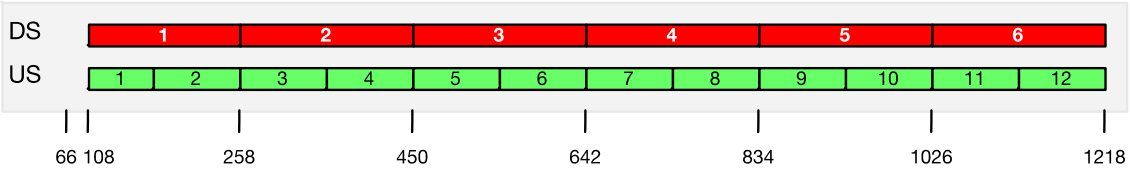


Figure 17 - High Band Channel Plan

There are three fundamental ways to create these simplex partitions: FDD, TDD, and a mix of FDD/TDD. Note that Figure 16 is equally applicable to FDD where the horizontal access is frequency, or TDD, where the horizontal axis is time.

FDX – FDD

This would be spoken as "Full Duplex with FDD". FDD is Frequency Division Duplex and would the technique for enforcing transmission rules for TGs and hence IGs. FDD is how DOCSIS and the cable plant works today. For the sake of illustration in the following scenarios, a full spectrum system is assumed.

FDX Channel Plan

So, how many channels are required for an FDX full spectrum system?

In the downstream, the answer is six. Six 192 MHz OFDM channels will span from 1218 MHz down to 66 MHz. If there is video to be skipped over, the number might be less.

If a dedicated broadcast channel of say

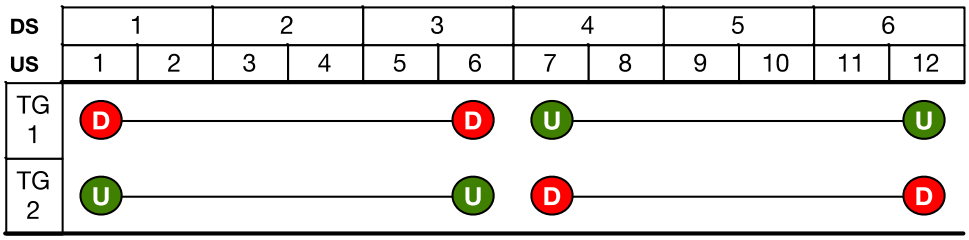


Figure 18 - FDX-FDD With 2 TG

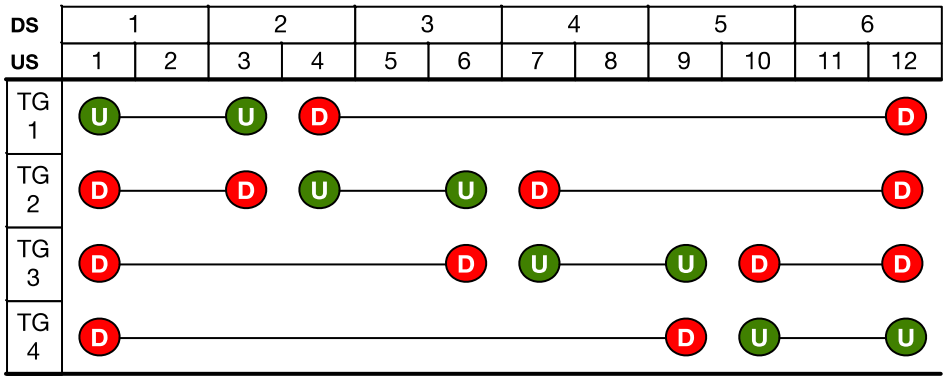


Figure 19 - FDX-FDD with 4 TG

24 MHz is used, then one more channel would be needed (so seven). Note that 66 – 24 = 42 MHz, which conveniently is the upper part of the existing upstream spectrum.

In practice, FDX would not go below 42 MHz so as to not disturb legacy DOCSIS 3.0/2.0 traffic.

In the upstream, the answer is twelve since each OFDMA channel is 96 MHz, which is half the width of a downstream OFDM channel.

Other combinations or partial channel maps are possible. However, Figure 17 will be used for illustrative purposes.

Scheduling a FDX-FDD System

In Figure 18, the top row shows 6 downstream channels and 12 upstream channels. On the left are transmission groups (TG) one and two. The barbell style diagrams are bonding groups.

In this example, each TG is given 50% of the bandwidth for downstream and 50% of the bandwidth for upstream.

Within a TG, different frequencies are used for isolating the downstream and upstream transmission. Between TGs, it does not matter. Thus, upstream transmissions will never interfere with downstream reception.

From the CMTS viewpoint, 100% of the spectrum is used for both downstream and upstream transmission.

What if the SLA that you are trying to create is not symmetrical? Figure 19 shows a system with four TGs where the upstream is 25% of the bandwidth and the downstream is 75% of the bandwidth within each TG. Overall, though, 100% of the spectrum is used for downstream and upstream.

FDX – TDD

TDD – Time Domain Duplex – is another scheme all together. It is not compatible with the current DOCSIS architecture. Let analyze what it would look like and then see if it is needed.

TDD at the CM would mean that a CM, or rather the IG or TG associated with that CM, is either transmitting across the full OFDM channel or receiving, but not both. That decision to transmit or receive would come from scheduling at the CMTS.

To do that, the CMTS has to know that there is downstream traffic on the wire for the TG that cannot be interrupted. That is not possible today. Today, the CMTS does not really know what is on the downstream wire and when. Take a look at the DOCSIS 3.1 system today in Figure 20 (source: [5]).

When the downstream packets hit the DOCSIS QoS engine, they are queued and potentially delayed. This is the first delay. The profile buffers and the code word builder are the second delay. Then there is the time and frequency interleaver which really slices and dices things. Even if you could predict where a packet was, its bits are now scrambled all over the place.

The DOCSIS upstream is better. You actually know exactly what packet from which CM will be sent at what exact time. You just have to calculate it 2 ms or so in advance.

The only way to have a TDD system is to construct one.

First, the downstream and upstream symbols have to be aligned. This means that the downstream and upstream channels that are occupying the same spectrum will use the same cyclic prefix.

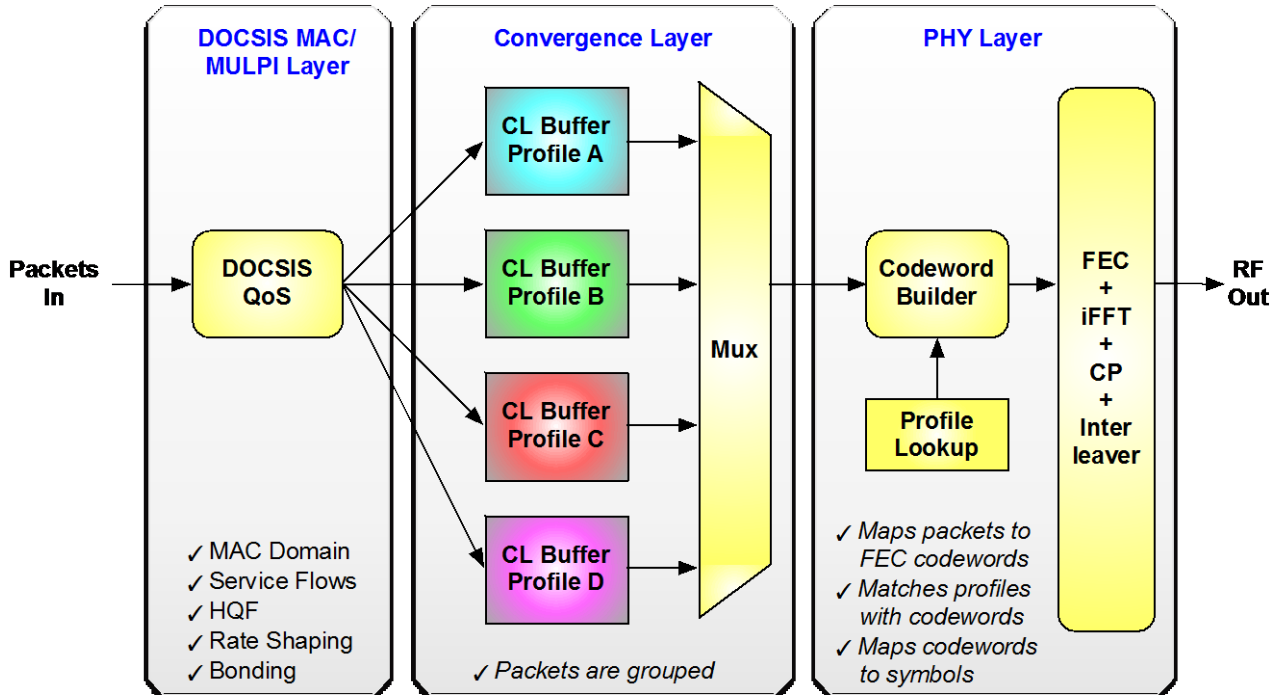


Figure 20 - DOCSSI 3.1 MAC-PHY

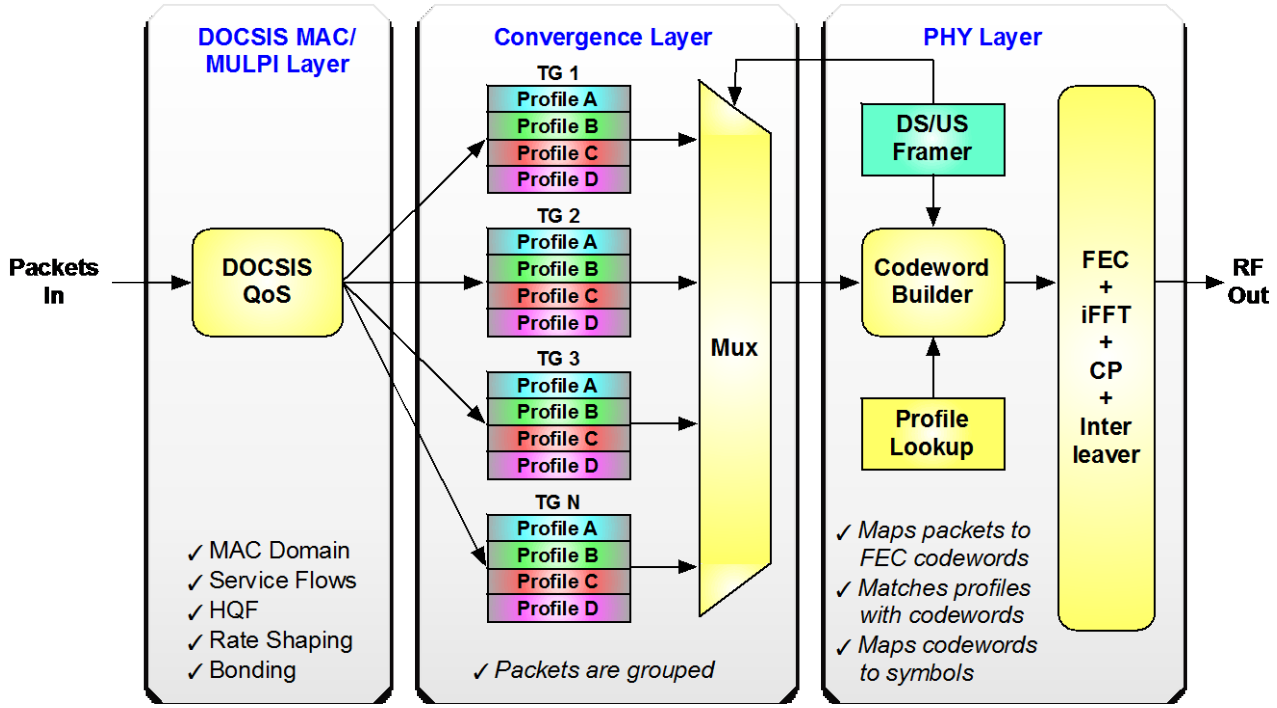


Figure 21 - DOCSIS 3.1 MAC-PHY with TDD

A frame structure has to be introduced. Since the downstream already has inter-leaver boundaries and the upstream already has minislot boundaries, a frame structure

that recognizes the superset of both of these is required.

Next, the queuing and profile hardware needs to support these frames boundaries. A proposed system is shown in Figure 21.

When the packets are released by the DOCSIS QoS stage, they are directed towards a downstream TG. Each TG has its own profile buffers. There is than a framer that pulls an entire frame contents from one set of profile buffers.

The assignment of DS and US frames to TGs is dynamic and can be done based upon the relative bandwidth that each TG requires.

FDD vs TDD

So, is a TDD system for TGs needed if there is already a FDD system available? It probably boils down to granularity. An FDD system has the granularity of channels and might break the available bandwidth into two or for frequency chunks.

A TDD system would break the bandwidth into multiples of frame boundaries. It is quite possible that a TDD system would have better granularity than an FDD system. However, in a large system, the granularity of a FDD system may be enough.

Ironically, a TDD system must support more channels than an FDD system. For a given bandwidth resource block, an FDD system can use less channels but is continuous in time.

A TDD system uses less time and therefore requires more channels. For example, for the same throughput, an FDD system may use one channel 100% of the time while an equivalent TDD system might use two channels 50% of the time.

FDX – FDD/TDD

It is worth noting that FDD and TDD systems could be combined to provide a three dimensional matrix of time, frequency and DS/US space, all under full scheduler control. The theoretical ability to do this would have to be compared to the practicality of doing this. But that discussion is for another time and place.

CONCLUSIONS

10 Gbps x 10 Gbps theoretically possible for DOCSIS. That put DOCSSI on equal footing with 10 Gbps Ethernet and 10 Gbps PON.

The optical node silicon (which is probably the CCAP silicon for an RPD) have to support an echo canceller. However, the biggest impact of FDX is more upstream throughput. This means more OFDMA channels and more CPU cores for scheduling.

CMTS and HDC plant are FDX. The entire spectrum is used for both downstream and upstream transmission. Each CM, however, is simplex using FDD or TDD. (partial spectrum or partial time is possibly a more accurate way of describing the concept.)

The design is being optimized for R-PHY Node and N+0 deep fiber. This is a realistic deployment plan going forwards and represents the minimum viable product.

Getting some proof of concepts, writing the specifications, creating the silicon, building the product, and establishing the market will determine availability of FDX. The next step then is to create a plan to match silicon and product availability with market requirements.

REFERENCES

Related Papers by the Same Authors

[1] John T. Chapman, Do You Have Legacy Issues in Your Upstream Plant? Dealing with ADI and OOB on a High-Split Plant, SCTE Cable-Tec Expo 2012, Oct 17, 2012, http://www.johntchapman.com/Chapman-LMA-(2012-08-30).pdf

Related Industry Specifications

[2] SCTE 55-1 2009, Digital Broadband Delivery System: Out of Band Transport Part 1: Mode A, SCTE, http://www.scte.org/documents/pdf/standards/SCTE%2055-1%202009.pdf

[3] SCTE 55-2 2008, Digital Broadband Delivery System: Out of Band Transport Part 2: Mode B, SCTE, http://www.scte.org/documents/pdf/standards/ANSI_SCTE%2055-2%202008.pdf

[4] North American Television Frequencies, Wikipedia, https://en.wikipedia.org/wiki/North_American_television_frequencies

[5] DOCSIS 3.1 Physical Layer Specification, CM-SP-PHYv3.1-I08-151210, CableLabs, Dec 12, 2015, http://www.cablelabs.com/wp-content/uploads/specdocs/CM-SP-PHYv3.1-I08-151210.pdf

Full Duplex DOCSIS Technology over HFC Networks

Belal Hamzeh
CableLabs, Inc.

Abstract

DOCSIS 3.1 technology provides a significant increase in network capacity supporting 10 Gbps downstream capacity and 1 Gbps upstream capacity. To support Gbps symmetric services over the HFC plant, increasing the upstream capacity is required. Full Duplex Communication using DOCSIS 3.1 technology provides the ability to increase the upstream capacity without sacrificing downstream capacity. In this paper, we provide an overview of how full duplex communications using DOCSIS technology can be implemented over the HFC network and the potential capacity gains it can deliver.

INTRODUCTION

Wired and wireless technologies use either frequency division duplexing (FDD) or time division duplexing (TDD) to enable bidirectional communications.

In FDD, the available spectrum is divided between the upstream (US) and downstream (DS) channels (uplink and downlink channels in wireless) without overlap between the channels to prevent interference as shown in Figure 1.

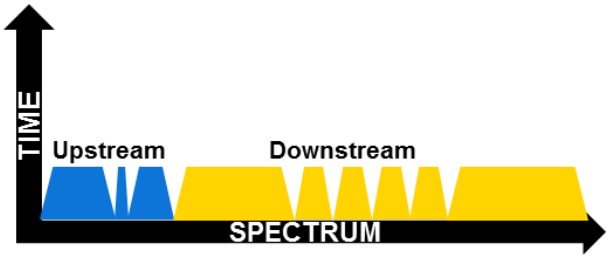


Figure 1: Frequency Division Duplexing

The spectrum split between US and DS channels is the main factor deciding the ratio between US and DS capacities. Thus increasing the US capacity will decrease the DS capacity, and increasing the DS capacity will decrease the US capacity.

In TDD, the same spectrum is used for both channels, but shared in time to avoid interference as shown in Figure 2. The amount of time allocated to each channel is the main deciding factor between US and DS capacities. Similar to FDD increasing a channel's capacity will decrease the other channel's capacity.

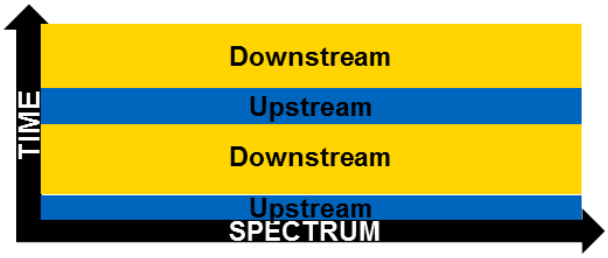


Figure 2: Time Division Duplexing

DOCSIS technologies operate using FDD; with the US channels typically residing in the 5-42 MHz, 5-65 MHz or 5-85 MHz bands while the DS channels share the upper frequency bands (up to 1 GHz) with the video channels [1].

Increasing a DOCSIS network capacity can be achieved by either increasing spectral efficiency; for example by upgrading from DOCSIS 3.0 technology to DOCSIS 3.1 technology, or by increasing allocated spectrum. In a spectrum limited network (upper band edge cannot be extended), increasing the US capacity will reduce the DS capacity

FULL DUPLEX COMMUNICATION

As discussed in the previous section, the goal from using TDD or FDD is to prevent interference between US and DS channels by

preventing overlap. Recent developments wireless R&D have shown that advanced interference cancellation techniques make full duplex (FDX) communication possible [2].

In full duplex (FDX) communications, the system supports concurrent transmissions on the US and DS on the same frequencies. Interference between the US and DS transmissions is cancelled out at the receiver using interference cancellation techniques.

Interference cancellation at the receiver is required to cancel out self interference. The transmit signal which is at a high power relative to the received signal interfers with the low power received signal due to the limited isolation levels between the transmit and receive ports as shown in Figure 3. Ideally, the combining circuit would comprise of a circulator with infinite port to port isolation which would prevent any cross coupling between the transmit port and the receive port.

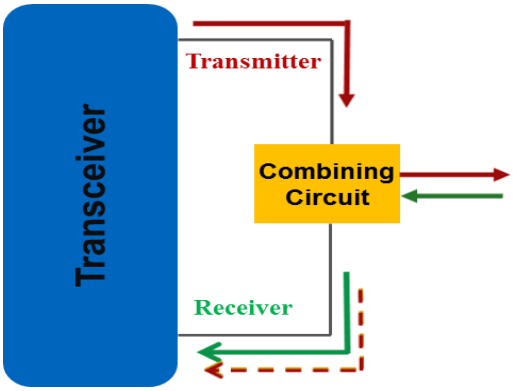


Figure 3: Self Interference in Full Duplex

Self interference cancellation at the receiver can be performed in the RF domain, digital domain or a combination of both [3].

FULL DUPLEX DOCSIS NETWORKS

Current DOCSIS networks have a limited amount of bandwidth available for US channels. FDX communication over the HFC network can provide additional bandwidth for US channels, while maintaining the available bandwidth for DS channels.

To implement FDX communications over the DOCSIS network, several things need to be considered; such as: Backward compatibility with previous DOCSIS generations, performance, scalability, and device complexity.

Backward compatibility: For a FDX device to be backward compatible with other DOCSIS devices, devices need to be able to coexist on the same network while achieving the target KPIs. Operating spectrum can be either shared spectrum or dedicated spectrum; for efficiency, shared spectrum is preferred but device complexity is another factor to be considered when looking at spectrum sharing. For example, in DOCSIS 3.1, the US spectrum can be shared between SC-QAM and OFDMA channels using Time and Frequency Division Multiplexing (TaFDM), while on the DS, SC-QAMs and OFDM channels operate in dedicated spectrum.

Performance: Symmetric services is a growing trend in broadband networks, especially in FTTH deployments. With 1 Gbps DS tiers being offered on current DOCSIS networks 1 Gbps US tiers is a reasonable initial performance target to be able to deliver 1 Gbps symmetrisc services, keeping in mind that the technology needs to be able to scale to higher tiers as needed.

Scalability: Technology transition within a network is closely managed to minimize any impact to current customers. Solutions that can be deployed in a gradual manner based on network capacity demands help facilitate technology transitions and lower the associated cost of such transitions. Solutions need to be able to scale efficiently based on the number of users and service requirements while at the same time maintaining backward compatibility. This enables an operator to introduce new technologies gradually without

the need to do a complete network upgrade ,replacing devices at all customer premises in addition to service interruption.

Device Complexity: Managing device complexity is critical to managing technology deployment costs. With current HFC network architecture, where the CMTS is a common resource shared across many users, minimizing CM complexity in exchange for driving additional (reasonable) functionality into the CMTS might be a viable approach.

Based on the above, CableLabs has proposed a DOCSIS 3.1 technology based FDX solution as a means to improve US network capacity.

Proposed FDX Architecture:

Based on the previously discussed requirements, CableLabs has proposed a FDX DOCSIS technology based solution. In the proposed solution, the CMTS operates in FDX mode on the DOCSIS channels while at any point in time, a CM is either transmitting or receiving on a certain channel with the option that it might be able to concurrently transmit and receive on different channels as shown in fig.

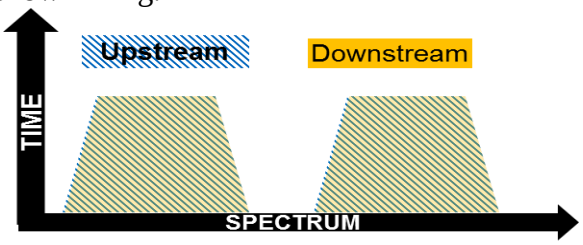


Figure 4a: FDX Channels as seen by CMTS

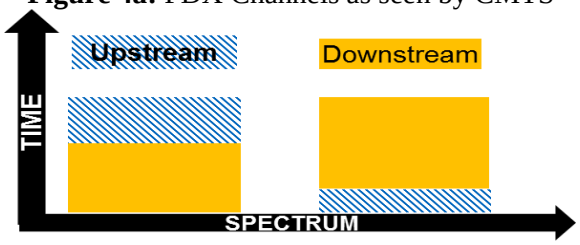


Figure 4b: FDX Channels as seen by CM

This approach allows the interference cancellation to be implemented only at the CMTS and thus minimizing CM complexity. The performance of the interference cancellation at the CMTS impacts directly the achievable US capacity, while the DS transmit power and interference patterns impact the achievable DS capacity.

Coexistence with legacy DOCSIS devices also needs to be addressed especially if sharing the DS spectrum is desired. Legacy devices continuously listen to the DS channels from the CMTS. US transmissions from a nearby device operating in FDX mode can potentially interfere with the DS signal being received by a nearby CM.

Interference from a FDX device's US transmission to the DS signal being received by another device is dependent on the level of RF isolation between both devices. US transmissions from FDX devices will impact nearby devices with low inter-device RF isolation. Based on experimental measurements conducted by CableLabs, devices operating on the same tap as a transmitting device will observe high levels of interference and thus severely impairing their ability to receive the DS signal, while devices connected to different taps will either not be impacted by the US transmissions, or observe reduced levels of DS SNR that will require them to temporarily operate using a lower modulation order.

To minimize the probability of impairing a device's DS reception; the CMTS would need to manage device to device interference. One possible approach to achieving this is to identiy the interference group (CMs that will observe interference due to transmissions from a FDX device) associated with the FDX devices, and based on the interference group, the CMTS could align scheduling such that a devices in the interference group are not scheduled to receive data (tor more optimally operate at a lower modulation order if

possible) while FDX device is scheduled for an US transmission.

Measuring the device to device isolation to identify interference groups can be achieved by channel sounding; where CMs measure the received power due to FDX device's transmission, and then report the recived power to the CMTS.

PERFORMANCE

The potential capacity achieved by using FDX DOCSIS communications on the HFC network is dependent on the HFC network RF characteristics, CM and CMTS capabilities and the performance of the self interference cancellation at the CMTS.

We use the DOCSIS 3.1 specifications as a baseline for the CM and CMTS capabilities, and the measured RF characteristics of a passive HFC network shown in Figure 5 with the assumptions listed in Table 1.

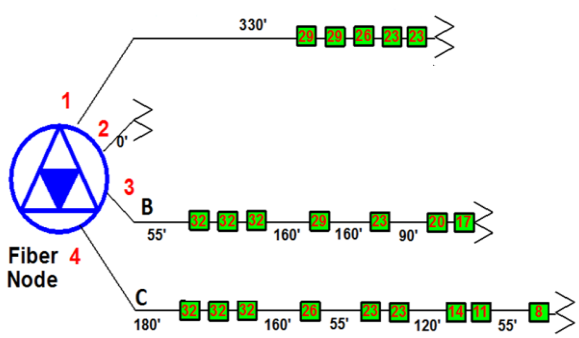


Figure 5: Passive HFC Network

Parameter	Value
CMTS Launch Power	41 dBmV / 6 MHz
Target US Signal Receive Power	0 dBmV / 6 MHz
Maximum CM Tx Power	65 dBmV
DOCSIS Channel Width	192 MHz
Drop Cable + Splitter Loss	4.5 dB
System Efficiency	80%

Table 1: HFC Network Assumptions

Assuming the parameters given in Table 1, we analyze the achievable US and DS capacity as a function of interference cancellation capabilities at the CMTS as shown in Figure 6. It is assumed here that the CMTS scheduler allows CMs within the interference group to receive at a lower modulation order in the presence of a nearby US transmission.

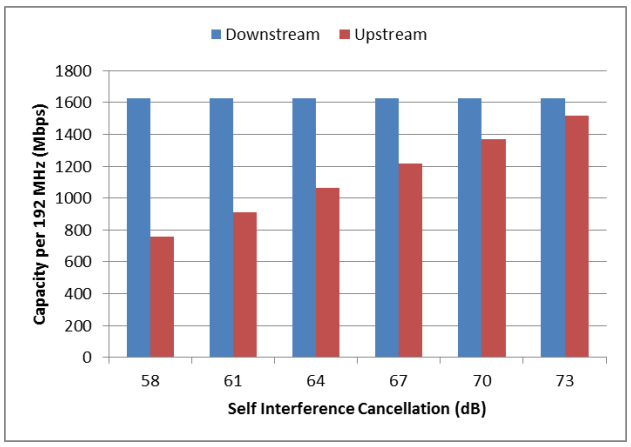


Figure 6: US and DS Capacities for FDX DOCSIS System

Due to the interference of US transmissions with the DS channel, there is a 11% reduction in DS capacity in comparison to a DOCSIS 3.1 system operating under the same assumptions but with no interference.

Further enhancements in US and DS capacities are possible by jointly optimizing the US transmit power, DS transmit power with respect to the self interference cancellation capabilities. Additionally, the DS capacity can be further optimized by scheduling DS traffic to devices part of an interference group when no other device in the group is transmitting, thus reducing the need to use lower modulation orders and thus improve DS capacity.

SUMMARY

FDX DOCSIS communications over the HFC network holds potential to significantly increasing upsteam capacities and enabling

Gbps summetric services over the HFC plant. The proposed FDX DOCSIS solution leverages DOCSIS 3.1 technology, the RF characteristics of the HFC plant, self interference cancellation at the CMTS and intelligent scheduling.

REFERENCES

1. Data-Over-Cable Service Interface Specification, Physical Layer Specification DOCSIS 3.1, Cable Television Laboratories, Inc. http://www.cablelabs.com/specification/physical-layer--specification

2. J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti. Achieving single channel, full duplex wireless communication. In Proceedings of the sixteenth annual international conference on Mobile computing and networking

3. D Bharadia, E McMilin, S Katti. Full duplex radios. ACM SIGCOMM Computer Communication Review 43 (4), 375-386

HEAD IN THE CLOUD, FEET ON THE GROUND: VIRTUALIZATION OF THE RESIDENTIAL GATEWAY

Authors: Barak Hermesh, Shaul Shulman, Guy Ray, Paul Mannion, Eoin Walsh
Intel Corporation

1 Abstract

The benefits of virtualizing the functionality of the broadband home gateway into the cloud have been discussed for some time in the telecommunication industry. These include OpEx reduction through ease of home gateway SW deployment, faster time-to-market of new features, and ease of service-chaining with little dependency on capabilities of the residential HW resources.

As we enter the DOCSIS 3.1 era, the transition to an end-to-end virtual gateway architecture not only is needed but also technologically attractive.

This paper describes the Virtual Home Gateway using Software Defined Network principles. We will discuss a reference architecture and present analysis of benefits in terms of easier software maintenance and reduction in time to market of new services. We will also discuss some of the implementation challenges in the path of making the Virtual Home Gateway mature for mass deployment: open networking while preserving subscriber privacy and intelligent hair-pinning over the wide-area network under constrained uplink resources.

2 INTRODUCTION

Transformation of the networking architecture into a Software Defined Network (SDN) seems no longer to be a question of "if," but a question of "how" and "when." Successful implementation of SDN in the Data Center space has triggered interest to apply the same principals to the telecommunications network due to the many similarities in both the challenges that SDN can overcome and in the methods that can be reused or replicated in telecommunications and broadband residential access networks. The ecosystem of SDN based solutions is growing rapidly and is available for MSOs. These solutions include Commercial Off-The-Shelf (COTS) hardware platforms and software tools. Since these are in general applicable for any networking application, a strong business incentive drives a very high speed of innovation that any particular industry can benefit from.

In the broadband residential access industry, two major growth vectors have been present since the early days of the introduction of Internet connectivity to the home. These are the connection speed race and the growth in the diversity of the services provided over the data connection. Both have expanded rapidly over the years, presenting a challenge of scalability of the infrastructure to continue supporting this growth. In the case of the speed of access, a new DOCSIS standard (DOCSIS 3.1) was recently introduced. It allows a much more efficient use of spectrum, enabling scalability of the residential connection data rates into the Gigabits per second (Gbps) range over the existing Hybrid Fiber Coax (HFC) infrastructure. In parallel, introduction of Fiber to the Home (FTTH) technologies easily allows multi Gbps speeds. These two advances create a future proof, scalable physical layer infrastructure, ready to accommodate anticipated as well as unanticipated demands in communication bandwidth.

On the side of the services however, nothing has significantly changed in terms of infrastructure. Conceptually, the network architecture for delivery of Gateway services looks the same as it looked 15 years ago. Functionality of the Gateway is embedded into the GW's hardware, firmware and software on a multitude of proprietary platforms from different vendors. Adding a new service requires a field upgrade of software and firmware on the embedded CPE device (different update per platform type). This severely limits the TTM of deployment and raises cost, putting a burden on operations and risking future scalability. An anecdotal indication of complexity growth of the embedded gateway is the fact that its SW stack image has grown about 100 times in size since the early days of a Cable Modem.

As a parallel process, over the top (OTT) providers are pushing forward to enable new services which are not dependent on the home user platform. The MSO's control of the whole MSO network and the CPE is not utilized to its full potential under existing networking architecture.

SDN and Network Function Virtualization (NFV) technologies present a wonderful opportunity for MSOs to upgrade the L2-L3 networking infrastructure into a more agile, scalable and future proof architecture. Virtualization of the network functions is also symbiotic with existing trends, including "fiber deeper," digitization of the last mile access from analog fiber to digital fiber, a move to distributed architectures of the Cable Modem Termination System (CMTS), and transformation of linear video into IP video. With virtualization, Service Providers and Operators also can offer services never before possible or not practical to implement, such as those based on "hair-pinning" and those that require intensive data path computation.

Application of SDN and NFV into Cable MSO networks has been talked about in the past several years ([1],[2]). As the ecosystem matures it seems we are near the time of moving from discussion into implementation stage. The purpose of this paper is to expand on the architecture, review and compare different architecture alternatives and offer our view of what the virtual CPE will look like and how commercial service providers and the surrounding ecosystem can benefit from the new architecture.

The paper is structured as follows. Part 3 sets a common terminology. Part 4-7 describes a proposed virtualized architecture for the Gateway and discusses the advantages of the architecture. Part 8 further articulates the advantages of the virtual CPE, Part 9 discusses the challenges and a summary follows in Part 10.

3 TERMS AND DEFINITIONS

As the industry has not converged on terminology of the network elements of the future architecture and in order to avoid ambiguity, the following terms are proposed and will be used as defined here in the rest of this paper.

pCPE – A physical CPE. The term describes the actual hardware residing in the subscriber's premise, which could be residential or business subscriber.

vCPE – A virtual CPE. This is a logical entity providing services that were traditionally provided by the pCPE. A vCPE is a software entity residing in the provider edge (PE), Place of Presence (PoP), data center or a mixture of the locations.

Home Gateway – The collection of functions that provides WAN (broadband) connectivity, routing, and bridging into the home and across the wireless and wireline networks in the customer premise.

Embedded Home Gateway – A home gateway whose execution is done exclusively on a pCPE

Virtual Home Gateway – A distributed network solution implementing the home gateway functions across a pCPE and a vCPE, in which the pCPE is a slave to networking decisions made by service logic executed on the on the vCPE.

vCPE Home – the vCPE component of a **Virtual Home Gateway** solution

PoP –A regional Place of Presence on the network infrastructure.

Customer Edge (CE) – The topological location of the customer facing edge of the MSO network. For a cable network, this would typically be where the pCPE is.

Provider Edge (PE) - The topological location of the provider's network edge, facing the access network. For a cable network, this would typically be the edge of the MSO MPLS network facing one or more CMTS/CCAPs. Some or all vCPE functions are hosted by a server hub located at the PE.

4 NETWORK ARCHITECTURE

In this section we analyze and compare several possible network architectures. We compare the architectures using the following parameters:

Efficiency – how efficient is the architecture in terms of cost of equipment.

Agility – how complicated it is to introduce new services or to upgrade services.

Operational expenses – how are the operational expenses affected by an architecture compared to alternative architectures.

4.1 MSO Network Topology

A cable MSO network consists of many elements. This section sets the terms of the elements relevant for the rest of the paper, illustrated in Figure 4-1.

The CE or Customer Edge is the edge of the MSO network residing in the customer premises. The CE equipment consists of a Cable Modem with potential Embedded Service Functional Entity (ESAFE) such as an embedded router.

The CMTS or CCAP terminates the HFC network and provides access to the MSO network.

The edge of the provider core network which is usually MPLS/Metro Ethernet is the location where the access network meets the MSO core network. The Provider Edge (PE) is usually located within the vicinity of the access termination equipment and usually covers a few CMTS/CCAP nodes. Data path processing related functions of the virtual home GW usually reside at the PE hub. Other functions may reside further away. The PE is sometimes known in the industry as the Central Office (CO) but in this paper we use the term PE.

The Point Of Presence (PoP) is a regional center aggregating traffic from multiple CE entities. Typically, a country or state has one or a few PoP centers. A PoP usually employs cloud technologies.

A service provider data center is usually a single location in one geographic region providing functions for the whole service provider network. The data center usually employs large-scale cloud technologies.

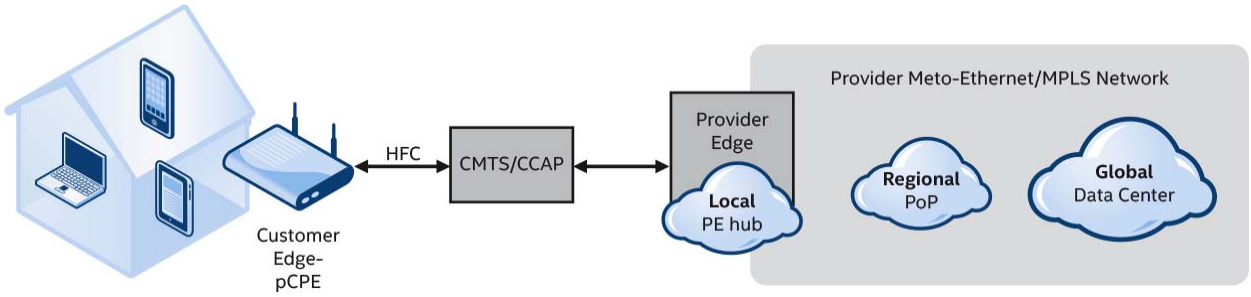


Figure 4-1 MSO Network Topology

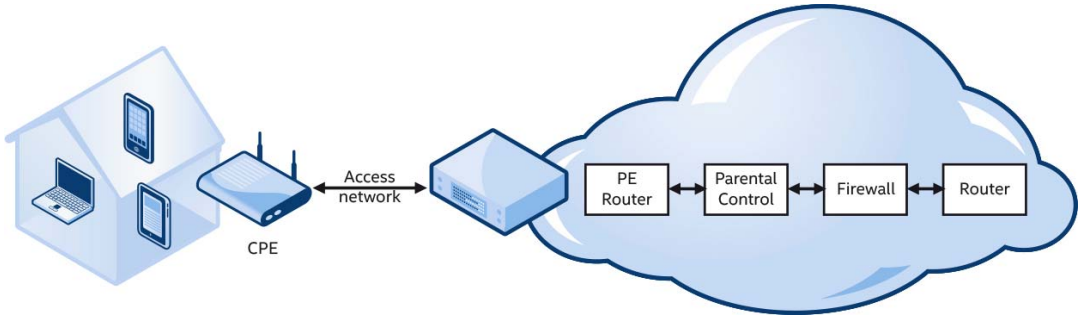


Figure 4-2 Legacy Service Provider Network

4.2 Existing Architecture

In today's existing architecture, the Home Gateway is an Embedded Home Gateway executing all gateway functionality locally. Traffic between devices in the home and the Internet flows through and is processed in the Home Gateway. It is then forwarded through the access network to the remote termination point (CMTS, DSLAM, OLT or other) and goes to the service provider's network. The service provider deploys carrier grade services like firewall and parental control at the provider's network. In a non-virtualized network, these services are implemented in proprietary blue-box equipment. Traffic flows through the network equipment where it is being processed.

Figure 4-2 illustrates the existing non-virtualized service provider network with parental control and firewall functions as examples. A typical home with connected devices contains a single embedded Home Gateway connected to the access network termination point. The service provider network contains blue-box network equipment implementing services (also known as network functions).

The existing architecture is efficient in the short term as proprietary equipment implementing a single network function is usually cost optimized. However, upgrade, enhancement and replacement of network functions has high cost overhead.

Introduction of new services becomes very complicated with the existing architecture, mainly due to the large variety of proprietary network equipment. A cable operator has on average about six different pCPE hardware platforms deployed, coming from different OEMs. Adding an embedded network function requires at least a software upgrade to all pCPE devices, assuming the pCPE hardware can even support the new function.

One lesson from the past illustrating the complexity is DOCSIS multicast forwarding:

DOCSIS 3.0 introduced a technique called Multicast DSID Forwarding (MDF) [3]. Its purpose is to allow for the CMTS to dictate the multicast forwarding at the CM. This was done, in part, because of the huge complexity involved with upgrading all CM (pCPE) devices to support IGMPv3. In a sense, MDF follows SDN concepts, where the forwarding device (CM) is a slave and the actual logic is determined by the software residing on a remote device (the CMTS).

A report by Analysis Mason [4] shows that moving customer households from an embedded Home Gateway to a fully virtual one reduces operational expenses by 64%. The two most significant contributors to operational expenses are customer visits and service calls. Both are much higher with the existing architecture, compared to a fully virtualized environment.

4.3 The Fully Virtualized Network

The Home Gateway in a virtualized environment is split between the pCPE and the vCPE. The pCPE owns the physical LAN ports and the actual packet forwarding in the home domain, both in-home and through the service provider.

In a virtualized environment, the physical elements are reduced to provide physical layer (Phy), Media Access Control (MAC) connectivity and data forwarding. The actual control and decision making is done by external software running on a virtual platform. The pCPE in this case would serve as a SDN switch exposing LAN and WAN ports to the SDN controller. The Home Gateway logic is virtualized by the vCPE-Home software running as a collection of virtual network functions in the service provider network.

The vCPE performs traditional Home Gateway functions like configuration of the LAN IP network by DHCP services, NAT, IP routing and more. For each pCPE, an instance of a vCPE is generated by the orchestration upon pCPE initialization. Each CPE obtains an individual configuration and state maintained mostly by the vCPE.

The vCPE is the management endpoint for the Home Gateway. As such, it exposes management interfaces like TR069 and web UI towards the home and the service provider. As the owner of the home GW management and configuration, the vCPE manages the configuration of the pCPE. This includes querying for counters and statistics but also controlling the port configuration (like WiFi SSID configuration).

A network orchestrator is software running on a virtual platform. It controls the instantiations and monitors the lifecycle of the Virtual Network Functions (VNFs) at the operator network including the vCPE. The orchestrator also communicates with the SDN controller, feeding it with forwarding policies based on the VNFs and on the service chaining policies.

In a fully virtualized network, the network access equipment (CMTS, DSLAM, OLT) is also virtualized. A CMTS, for instance, would have its forwarding logic reduced to a SDN switch and the control would come from the SDN controller. It is however expected that at first stage the access network would remain as today and the traffic from the pCPEs to the operator network would be tunneled over the underlay access network as explained in 4-6.

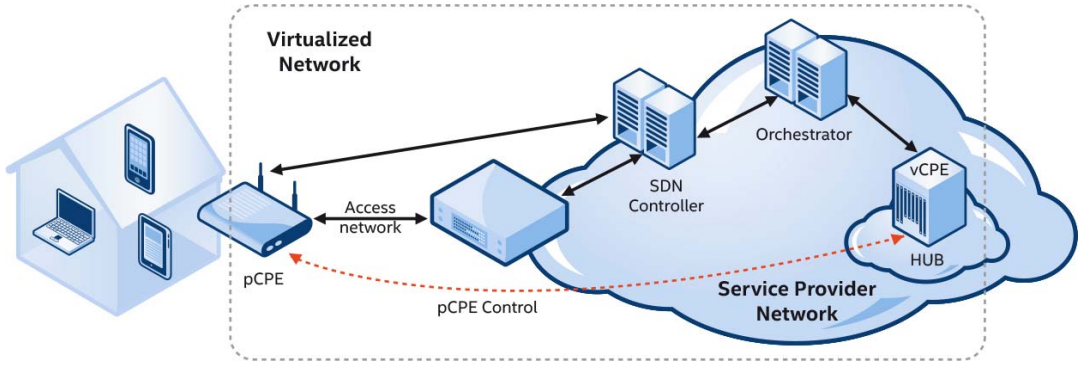


Figure 4-3 The Virtualized Network

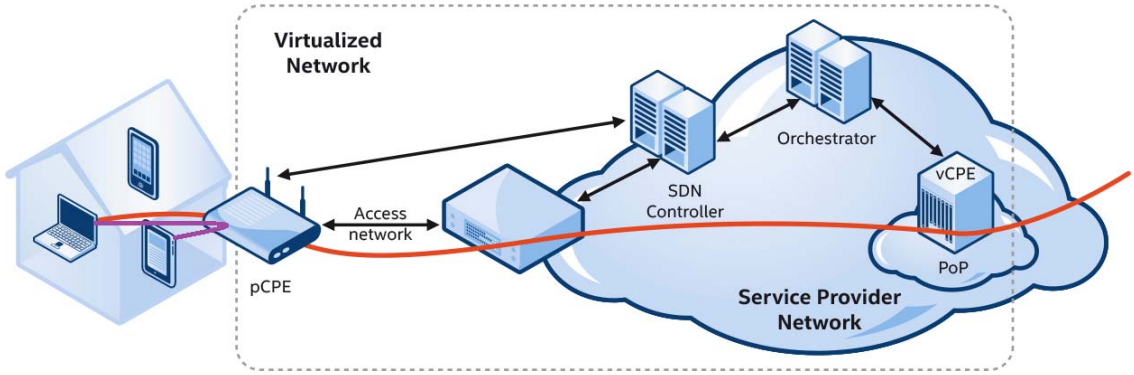


Figure 4-4 Straightforward pCPE Forwarding

Figure 4-3 illustrates the major elements of the virtualized network.

4.4 Software Defined Forwarding at the pCPE

The physical CPE performs packet forwarding. The rules by which the pCPE forwards the packets are set by the SDN controller. It is expected that in conventional cases traffic on the same LAN segment will be simply forwarded from the source LAN port to the destination and that traffic to an external destination will be forwarded to the access network termination point. Due to the power of SDN, exceptions could be made.

Figure 4-4 shows two forwarding paths. One LAN-WAN marked in orange and one LAN-LAN marked in violet.

The flow would start with a pCPE having pre-configured forwarding rules set by the SDN controller (this is called proactive mode). The rules would say roughly:

1. For any unknown packet, consult with the SDN controller.

2. For all packets whose destination is not in the home network, forward the packet to the WAN port (potentially with tunneling)

3. Specific rules based on destination MAC address for L2 forwarding in the home, set by dynamic consultation with the SDN controller

Since the forwarding logic at the pCPE is dictated by the SDN controller, topology changes and exceptions could be easily made and deployed. One novel example is hair-pinning.

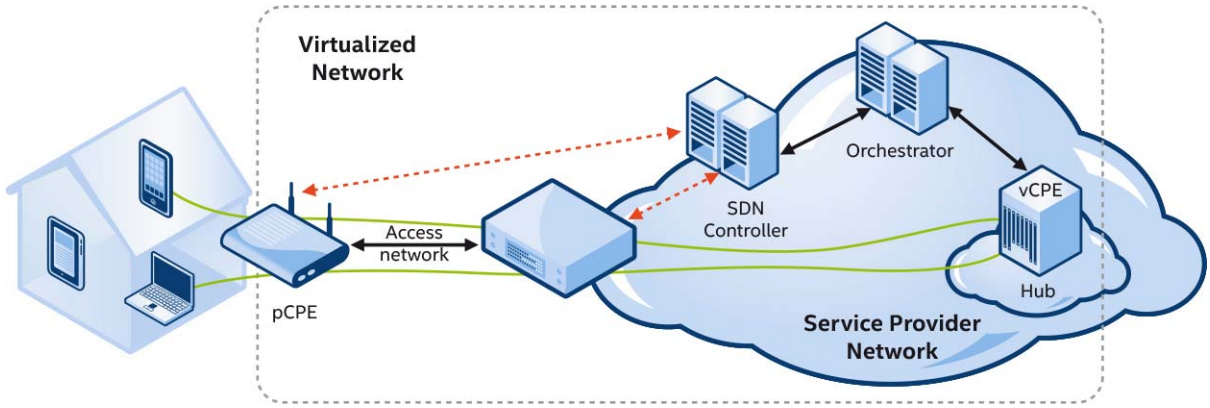


Figure 4-5 Hair-Pinning

Hair-pinning, in this context, is the concept of sending a stream of packets outside of the home network for processing at the vCPE and then looping it back in to the target port. Figure 4-5 shows a case of hair-pinning between a laptop computer and a smart phone. Although both are on the same network, an administrative decision had been made to hair-pin the traffic through the vCPE. Thanks to the flexibility of SDN, hair-pinning could be done for selected flows or even selected packets.

There are multiple usages for hair-pinning, starting with enhanced in-home network security and going through remote diagnostics and troubleshooting. At this point in time, since the broadband access network is asymmetrical, hair-pinning is limited by upstream capacity. Thus, it should be applied with care.

4.4.1 Network Security via Hair-Pinning

In the modern home, the assumption that in-home traffic is safe is no longer valid. Malware can be injected into the home via a guest joining the home network, an attacker connecting to the home wireless network and in other ways. As the home network becomes more essential, it becomes mandatory that traffic originated in the home and terminated in the home is subject to monitoring by Intrusion Prevention Systems (IPS).

Running a full blown IPS on the pCPE is not feasible today and will not become feasible in the future as the computation power available for applications grows at a rate not higher than home network throughputs. In other words, the gap between the computation the pCPE can perform at a reasonable cost and the computation required by IPS does not close. The only feasible approach is therefore to take advantage of the elasticity and the power of the cloud and run IPS as a VNF. The SDN policy becomes such that when required (and no more), selective sessions get hair-pinned through the IPS for inspection.

Imagine a case where the alarm system starts communicating with the smart TV. This might be considered an irregular pattern and will then be hair-pinned for inspection. After a few packets, the IPS, being a logical part of the vCPE, may conclude there is no threat and instruct the SDN controller to terminate the hair-pin. The SDN controller will then reconfigure the pCPE SDN forwarding logic to switch the traffic locally.

4.4.2 Real Time Troubleshooting via Hair-Pinning

From time to time, subscribers experience in-home network connectivity issues. Since MSOs are responsible for home networking, such service calls frequently end up with technician visits generating high expenses to the MSO.

With SDN it is possible to troubleshoot such issues remotely. The technical support team sets up hair-pinning to all of the subscriber's in-home traffic, or to selective flows if identified as faulty. Using standard monitoring and analysis tools, the technician may detect the issue and apply corrective measures.

Other advantages of supporting SDN at the pCPE include flexible residential topology setup, simple creation and teardown of tunnels and future proofing.

4.5 The Virtualized Hub

An alternative to a fully virtualized network is a hybrid where the existing gateway services are embedded in the pCPE and additional services are located in a hub at the MSO core network. The MSO core network is SDN operated and network functions are VNFs.

A hub is given here as a generic term for the cloud compute cluster where VNFs run. A virtualized hub operating with a pCPE that is SDN unaware hosts VNFs at the hub without extending the software defined network into the customer premises. The pCPE in this case contains an Embedded Home Gateway with multiple WAN-facing tunneling interfaces. Classifiers provisioned to the pCPE direct traffic into tunnels based on the VNF service chain required to operate on the traffic. Figure 4-6 shows an example of three tunnels representing three classes of traffic. The pCPE performs some GW functions locally while other services are executed at the hub.

Compared to the existing architecture, this architecture's operational expenses for the VNFs are reduced, while the same level of complexity and operational expenses at the pCPE is maintained.

Service agility is achieved by deploying and upgrading VNFs. Orchestration includes setting up tunnels, tunnel terminators are service chains. Since the pCPE is not a part of the virtualized network, agility of traffic forwarding is limited to the service provider network. Flows like hair-pinning are not possible as LAN to LAN traffic is always terminated at the pCPE. pCPE related OpEx and complexity remain the same as in the existing architecture.

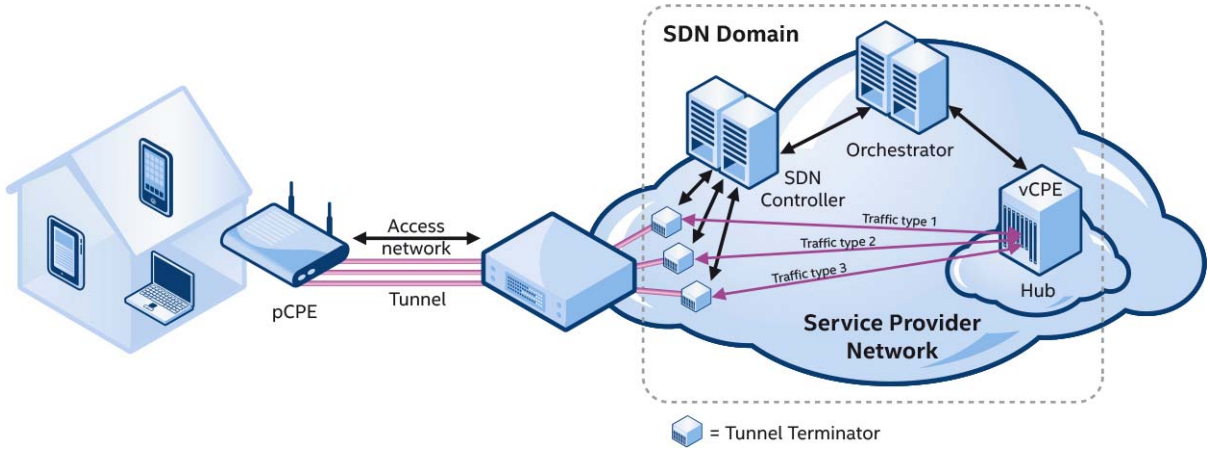


Figure 4-6 The Virtualized PoP

4.6 The Hybrid Network: Overlaying Non-Compliant Network Equipment

It is expected that not all network equipment will become SDN compliant at once, there will be gradual transition from the existing architecture to a fully SDN compliant architecture. In order to traverse through non-SDN network equipment, tunnels are created as bridges between SDN domains. Tunnels are logical point to point connections on top of an underlying network. Popular tunneling techniques include Layer 2 General Routing Encapsulation (L2GRE also called EoGRE) over IP and Virtual Extensible LAN (VXLAN) over UDP.

The most probable scenario is that while the home and the service provider core network become SDN compliant, the access network is still not SDN compliant. In this case, a tunnel is used to connect the residential SDN domain and the service provider SDN domain. The pCPE is located at one end of the tunnel and a dedicated tunnel termination network function is located in the service provider network and serves as other end of the tunnel.

It is allowed that multiple tunnels exist for the purpose of load balancing, redundancy and service segregation. The SDN controller instructs the pCPE and the tunnel terminator to direct traffic into a selected tunnel using standard SDN rules.

Figure 4-7 shows a case of a single tunnel from one pCPE to a tunnel terminator. Note that the SDN controller controls both the pCPE and the tunnel terminator. The tunnel bridges between to SDN islands comprising a single SDN domain together.

4.7 Tunneling Overhead

One of the drawbacks of tunneling is that it adds network overhead. L2GRE [5]over IPv4 overhead is 14+20+16 bytes for plain Ethernet header+IPv4+GRE. 50 bytes appended before each L2 frame, assuming an IMIX [6] average packet of 700 bytes, GRE tunneling adds an overhead of 7% resulting in a drop of 7% of effective data throughput.

5 THE pCPE ARCHITECTURE

The pCPE consists of four layers. The physical ports, the SDN forwarder, the forwarder control and the provisioning/OSS layer.

The bottom layer is that of the physical ports. All network ports are exposed to the layer above.

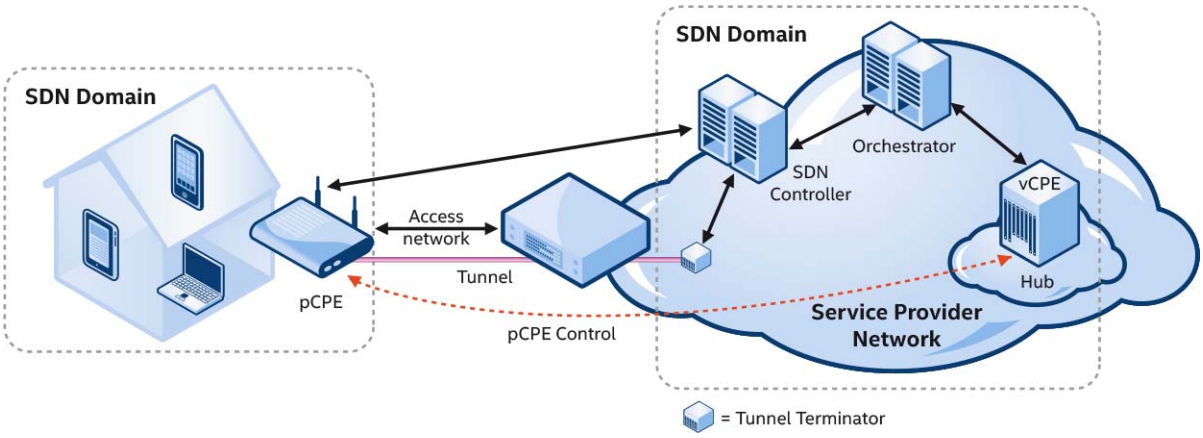


Figure 4-7 Transition Phase via Overlay

The layer above the ports is the SDN rule based forwarding layer. The forwarding layer receives packets from the ports and forwards them to their destination ports based on the SDN rules dictated by the SDN controller. One possible implementation of this layer is the Open VSwitch forwarder [7].

The layer above the SDN forwarder is the forwarding control layer which consists of two mutually exclusive logics. The SDN protocol termination point comes to action when the link to the vCPE is established. In this case, all forwarding rules are set by the SDN controller by sending them to the protocol termination point. In the case of Open Flow, this would be an Open Flow protocol stack endpoint.

The offline mode logic comes to action when the link with the vCPE is down. This logic makes sure that LAN to LAN forwarding is made possible by setting L2 forwarding rules to the forwarder and by providing L3 configuration to the hosts on the LAN if required.

The top layer consists of the device management, provisioning and state monitoring.

Provisioning starts when WAN connectivity is established and includes notifying the orchestration logic that a pCPE is up, pairing with the vCPE and obtaining runtime configuration.

The device management module is the local endpoint of the vCPE-pCPE device management. The vCPE configures the pCPE runtime parameters by utilizing this module. The vCPE also queries the pCPE device status and statistics by communicating with this module.

The link state monitor is a daemon monitoring the connectivity between the pCPE and the vCPE periodically. If the link goes down, the monitor alerts the provisioning module which may change the operational state from SDN forwarding to offline forwarding until connectivity is re-established.

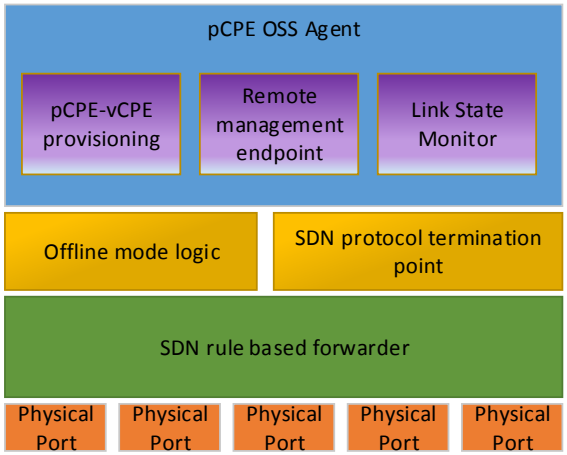


Figure 5-1 pCPE Architecture

6 THE vCPE ARCHITECTURE

The vCPE is a set of software services executed on virtualized platforms located at the PE and sometimes also in other locations. Unlike the pCPE, the vCPE can be implemented in various ways starting with each vCPE being a completely independent self-contained micro service, containing all vCPE functionality for a single home GW and going to a highly distributed architecture where single tenant network functions support multiple pCPEs and service chaining is applied in order to link all network functions together.

The vCPE provides the following services –

- Routing with address translation (NAT)
- Port forwarding, UPnP, port triggering
- Firewall
- IP configuration for hosts on the LAN
- DNS
- Advanced services like parental control, centralized storage (NAS), UPnP media server, DLNA media server and more

In this section we discuss and analyze three possible vCPE architectures. The key difference between the architectures is in the grouping of functionalities into network functions. The tradeoffs are simplicity vs. performance and scale.

First, we provide an observation as to the relation between the MSO oversubscription ratio and the computational power of the vCPE server.

All architectures analyzed here comply with ETSI ***Network Functions Virtualization (NFV); Architectural Framework*** [8]

6.1 vCPE Data Path Computation

All MSOs over-subscribe their subscribers. If the physical network infrastructure is capable of carrying a given bandwidth, the sum of the bandwidth provisioned to all subscribers would be much higher. The ratio between overall subscribed bandwidth and available bandwidth is defined as the oversubscription ratio. The oversubscription ratio is carefully set by the MSOs to meet the statistics of bandwidth consumption of the subscribers, relying on the spread of bandwidth demand (across time) among a large number of subscribers.

Since each unit of data going through the access network is also processed by a vCPE, the overall processing power required for all vCPEs is linear with the physical network infrastructure bandwidth. The compute load of the vCPE server platform benefits from oversubscription.

As an example, a node of 100 subscribers, each provisioned with 100Mbit/Sec is connected to the MSO network over a 1Gbit/Sec link. The oversubscription ratio here is then [100 x 100Mbit] / 1Gbit = 10. A server capable of processing 10Gbit/Sec of CPE network traffic at the Provider Edge can support 10,000 subscribers. Each is provisioned with 100Mbit/Sec given the oversubscription ratio of 10.

6.2 The vCPE as a Micro-Service

Figure Figure 6-1 shows multiple instances of vCPEs, each instantiated as a micro service on a virtual machine. The number of services is dictated by the number of pCPEs currently provisioned, but the number of virtual machines and the allocation of a vCPE service to a virtual machine is dictated by the orchestration layers and is based on optimization and load balancing. The micro services are expected to run within a virtual isolated environment such as a ***Linux Container*** [9]. Note that a real world vCPE is expected to have more services than illustrated in the figures below. The figures below show a few services just for illustration.

Each micro service is a fully self-contained vCPE. Each micro service is very similar to an embedded home GW module. This architecture is simple to orchestrate and offers the simplest development model. However, from the performance aspect this solution is the least efficient. Efficiency is low since packet forwarding, which is the most intensive computational task of the vCPE, is done in many different contexts resulting in context switching and cache thrashing.

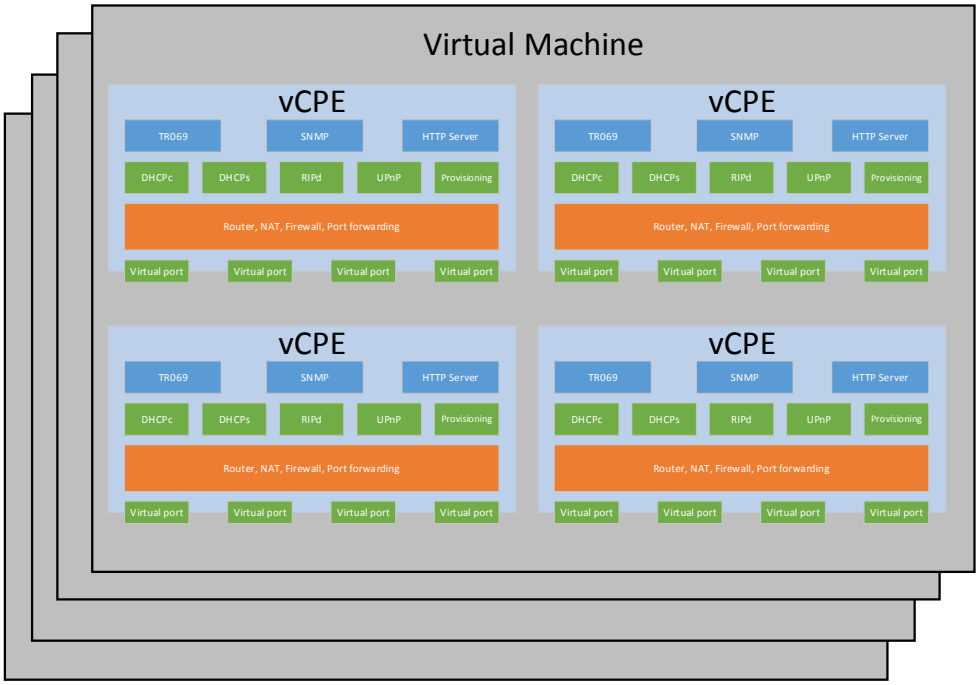


Figure 6-1 vCPEs as micro services

6.3 Split Control and Data Path

Another alternative that is expected to be more performance-efficient is to separate the data processing from the vCPE control. The vCPE control is instantiated per pCPE as in the previous case. However, the data path processing including routing, address translation, port forwarding, firewall and other network functions is done on optimized dedicated single tenant virtual network functions. The advantage of this approach is that it is more efficient both in resource consumption and in compute resources to run all data through a single VNF than to run it through many smaller micro services.

The VNFs are CPE aware so there is a logical context per CPE. The vCPE control installs configuration rules for the centralized VNFs (router, firewall etc).

As an example, if the vCPE-Control TR069 agent is instructed to set up a new firewall rule, it configures the firewall VNF to add the rule for the CPE it represents. The SDN rules are set such that control traffic from the pCPE is forwarded to the vCPE-Control entity where other traffic is forwarded to the service chain containing the NAT, router, firewall and other VNFs.

Figure 6-2 shows multiple vCPE-Control micro services and two centralized VNFs – one for routing and address translation and one for firewall.

This approach is more performance optimal but it comes with the extra complexity induced by the need to orchestrate multiple functions per pCPE and by having a control and communication channel between the vCPE-control and the data path VNFs.

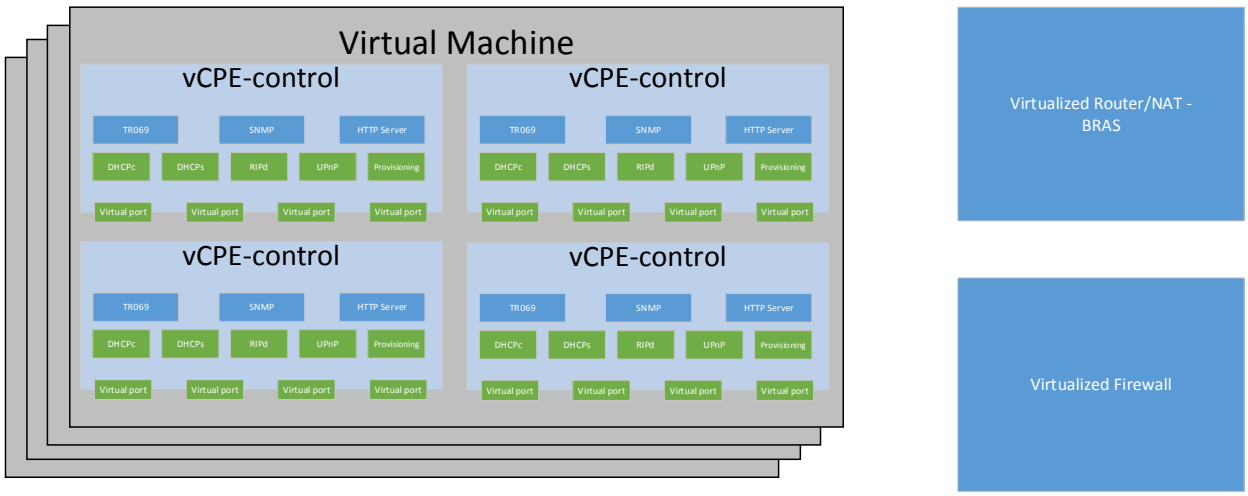


Figure 6-2 vCPEs with data path functions are centralized entities

6.4 Centralized Control and Data Path

An even more centralized alternative runs some of the control entities as centralized VNFs. Two candidates are the vCPE DHCP server and DHCP client. Instead of instantiating one per pCPE, two centralized, multi context VNFs are instantiated. The DHCP VNFs run all DHCP sessions for all vCPEs services and communicate with the vCPE-control services. Communication includes the vCPE-control setting the configuration to the DHCP VNFs (like the subnet and IP pool for the DHCP server) and the DHCP VNFs reporting to the vCPE-control on the DHCP session results. The SDN controller sets rules at the pCPE such that DHCP sessions are forwarded to the DHCPs VNF.

The advantage of this approach is in higher efficiency – centralized DHCP services are more resource efficient than distributed services. Yet, as shown above, orchestration and control are more complex.

6.5 CPE-vCPE Architecture Dependencies

Other architectures also may be considered, some being more centralized than other. The pCPE architecture does not dictate a specific vCPE architecture, as long as some conditions are met:

1. The pCPE-vCPE interface is not affected by the vCPE architecture
2. The SDN protocol is generic enough to set rules for the pCPE supporting all architecture

The vCPE orchestration is flexible enough to instantiate vCPEs elastically as pCPEs come and go. At this time, it appears that the optimal vCPE architecture is defined in 6-3 (Split Control and Data Path). Early feedback from vCPE manufacturers shows that having a vCPE data path with dedicated network functions has great efficiency benefits, yet is reasonably simple to manage and orchestrate.

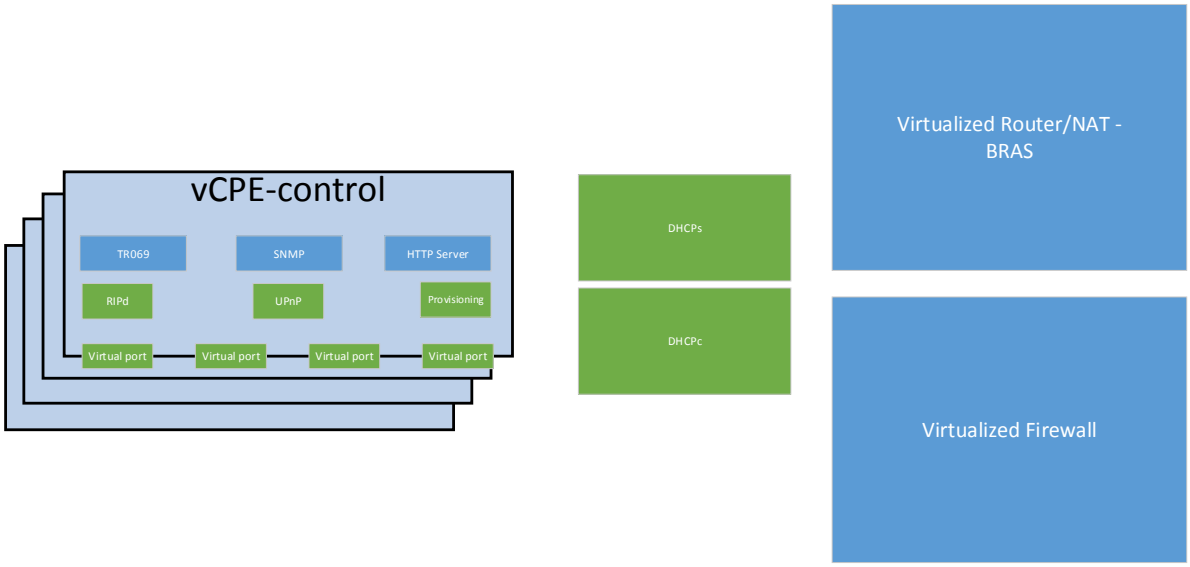


Figure 6-3 vCPEs with control as micro service

7 COMPARISON OF OTT AND SERVICE PROVIDER MANAGED NETWORK FUNCTIONS

With the evolution of broadband service, it is often argued that commercial service provider offerings will transition to being a "dumb pipe." Network and service virtualization provides advanced capabilities compared to the existing architecture, thus allowing for service providers to bring added value to their subscribers that would help compete with Over-The-Top (OTT) service providers. This section discusses those capabilities and highlights inherent advantages of MSO hosted SDN/NFV network compared to OTT VNFs.

7.1 Personalized Services

Making the pCPE a part of the SDN domain opens the door for user and physical port aware virtualized services. As an example, device aware parental control policy can be performed as a part of a parental control VNF, as the device MAC address and optionally the port information can be carried with the packet and traverse SDN switches.

If the pCPE is not a part of the SDN domain and acts as an Embedded Home Gateway a lot of valuable information is lost, including the user device identification and even IP address (in case of NAT) and the physical port. Over The Top (OTT) providers relying on embedded gateways do not have access to this valuable information and can therefore offer limited services. To follow the parental control example, an OTT parental control VNF can only associate traffic with a CPE device, not with the actual end device or the physical port.

7.2 Responsive Services

OTT also has the major disadvantage of remote data centers. While the MSOs hubs are located both physically and topologically within the vicinity of the subscriber premises, public clouds are located further away and limited in ability to provide services which require fast response time.

7.3 End-to-End Service Assurance

Many times it is desirable to assure the availability of a network service. Best effort services do exist but are harder to monetize and often provide poor user experience.

SDN network flows are set by the SDN controller. Apart from setting the path, the SDN controller can also set up end-to-end QoS on a per-flow basis. The MSO is a fully managed network so service assurance can be provided while in the MSO network using SDN techniques.

OTT services reside on the public Internet and are therefore not subject to controlled QoS and are at the mercy of the Internet bottlenecks and peak loads.

8 TTM AND FEATURES DEPLOYMENT BENEFITS – QUANTITATIVE ANALYSIS

The major benefit called out for vCPE is service agility. The new approach enables a new multi service x86 platform which enables seamless introduction of new services, service upgrades and simplified end user support. By moving to vCPE-enabled services, some analysts [4] believe CSPs can achieve up to ~75-80% of CPE-related cost savings per household per year after the migration of a majority of the customer base. Others [10]claim a cost saving of up to 24%.

The move can be broadly categorized under 3 headings.

1. Flexible Multi Technology CPE

The CPE can be used in both modes, today where it runs all the legacy functions or as described above where some of the CPE functions are moved to the core network. This Home Gateway can now act as L2 or L3+ device allowing the network edge to orchestrate the appropriate level of intelligence and the Home Gateway can provide connectivity functionalities including an access point (WLAN), switching (Ethernet) and modem (xDSL, Cable), which from the OpEx perspective makes it cheaper to resource, manage, and repair.

2. Cloud vCPE solution

The ability to move advanced IP functions (L4-7) from the Home Gateway to to the Provider Edge or Point of presence (PoP), deployed as virtualised network functions (VNFs) running on COTS servers. Example functions include: CGNAT, Firewall, routing, DHCP, Web management, UPnP.

3. OSS evolution and Customer Support

A fully SDN orchestrated CPE enables automation of the vCPE service deployment, troubleshooting and service upgrades. L2 visibility to the home which allows full visibility to the orchestration layer enables CPE configuration, testing and troubleshooting and operations through a customer self-service portal. This reduces the number and duration of customer calls and truck rolls.

There are varying estimates from different analysts and reports on TCO savings based on the headings above, but they assume varying consumer scenarios such as double-play routers and triple-play set-top boxes. One thing that is consistent is that ability to co-locate services and service deployment, reduce CPE boxes that need to be managed, deployed and maintained, and the ability to up sell services (e.g. try before buy and profit share with OTT's) offers ~75-80% total cost savings per household

New Value Added Services

Operators do not necessarily know which new services may arise, but they do know they can no longer operate in the traditional mode of 18-24 month HW development cycles to enable such service introduction. They must provide flexible multi-locational platforms which enable faster Cloud/OTT like service deployment capabilities which enable new service introduction or indeed the flexible platform onto which external platforms can deploy revenue sharing service models.

Roll back

CSP's nominally do 2-3 firmware updates per year. The logistical, technical and financial impact of an update going wrong is such that every upgrade is a heavily controlled and rehearsed event. The cost of a failed upgrade carries a huge opex and capex penalty to the operator resulting from a rise in support calls, truck rolls for repair/replace and the potential for customer loss. This is not the case for cloud services as these can be spun up and down without the end user even noticing. Cloud service providers can update several times a week without the user seeing any noticeable impact to service.

The VNF eco system

We are already seeing a large ecosystem develop in this space. Traditional telecom vendors such as —Alcatel-Lucent, Cisco, Ericsson, Fujitsu, Huawei, Juniper, NEC, and ZTE—remain the favored vendors from which operators will buy SDN and NFV software to satisfy and fulfill the vCPE enablement and roll out.

Ecosystem vendors are likely to deploy multiple functions (such as firewall, IP/MPLS VPNs, load balancing, QoS support, VPN termination, CG-NAT, DPI, IDS/IPS, WAN optimization controller, or BRAS/BNG) from physical edge routers to software vRouters or VNFs running on commercial servers

Taking Nokia/ALU as an example, they are demonstrating six x86 servers behaving as a single router that can offer up to 2Tbit/s throughput, with the control plane and EMS [element management system] regarding it as a single instance. Alcatel-Lucent has architected its routing software to get the best performance, resilience and reliability on general purpose (x86) compute platforms. The company is demonstrating [11] 320G half-duplex, or greater than 2x better than competitor offers, for a virtualized Provider Edge routing application in a single x86 server.

Brocade in conjunction with Telefonica have shown their Vyatta 5600 vRouter on a commercial off-the-shelf based x86 server within a Red Hat KVM environment. Tests have shown 8x the performance in the control plane over competitor offers [12]. Deployed as a single virtual machine, the Vyatta 5600 was able to support all of the server's available ports at line rate by hitting the 80Gbit/s mark, Brocade actually exceeded its own original goal, which was to prove that a software router can support the 10Gbit/s performance that is mainstream in carrier environments

Table 1 – Scope of Visibility per CPE Technology

	L3 NAT decisions	L2 switching decisions	Level of visibility from outside the home
Embedded Home Gateway	By the pCPE	By the pCPE	Traffic between the home and the internet
Virtual Home Gateway	By the vCPE	By the pCPE	Traffic between the a client within the home and the internet
Virtual Home Gateway with selective hair-pinning	By the vCPE	By the vCPE	Traffic between clients and the internet or between themselves

9 CHALLENGES

9.1 Security and Privacy, DOS

Different degrees of virtualizing Home Gateway functions imply different levels of visibility from outside the home. We expect that the Virtual Home Gateway will be tightly linked to new services for subscribers to allow it. These services may include community WIFI and home automation with advanced network protection.

An architecture where the pCPE is a complete slave of the vCPE (as proposed in 4-3) exposes some security and privacy vulnerabilities that have to be addressed. Impersonation of the vCPE and the SDN controller can result in having subscriber traffic, including LAN to LAN traffic, diverted to an arbitrary path chosen by the attacker. It is expected therefore that virtual Home Gateway operation would imply two-way authentication between vCPE and pCPE. New generations of threats could include *man in the middle* attacks between vCPE and pCPE. Also, DDOS attacks by pCPE on the shared L3 NAT resources of the vCPE will drive motivation for authenticating "traffic patterns" generated by pCPE – even when not violating the bandwidth service level.

Another effect of the proposed architecture is the potential exposure of LAN to LAN traffic to the service provider via hair-pinning. Although it does allow the exposure of more content to the service provider, the convention today is that all traffic being sent in the clear is exposed. Specifically, all Internet traffic is exposed to the service provider today so the additional exposure induced by hair-pinning is not likely to be perceived as violating the level of privacy provided today.

9.2 Intelligent Hair-Pinning Under Limited Uplink Resources

The hair-pinning mode of operation between vCPE and pCPE makes it possible for Service Providers to associate advanced processing functions to specific traffic flows between two clients connected to the same residential local area network – wireless and wireline.

Advanced processing of LAN to LAN traffic could be used by holistic smart home network security solutions designed for protection in an IOT environment with no assumption on the source or the capabilities of the "Things." This includes the scenario of friendly visitor clients being legitimately authorized to onboard one's home network. Such solutions will rely on consuming just as much traffic as processing and bandwidth allow in search for suspicious traffic patterns across the residential network.

As broadband technologies are asymmetrical and as home traffic volume and traffic peak are increased, managing upstream resources under hair-pinning

operational mode between a vCPE and a pCPE becomes a new challenge. We expect that the vCPE would have to be real-time aware to actual utilization of the upstream and to the type of clients involved in each session. As an example, traffic between the Refrigerator and the TV is likely to be more suspicions and might require advanced inspection. The vCPE will therefore be smart enough to make real-time intelligent decision with respect to which LAN to LAN traffic should be processed on the vCPE platform and which should be offloaded for switching by the pCPE.

Designers of next generation broadband technologies should be aware to possible increased demand for improved upstream/downstream ratio.

10 SUMMARY

Virtualization of the home gateway as a broadband service provider has many advantages including significant reduction in OpEx, service agility, enablement of compute intensive services and enablement of services at the subscriber LAN scope. Several network architecture alternatives are made possible, each with its own advantages and complexities. Rather than pointing to an optimal solution, this paper compares the alternatives and analyzes them.

We believe that the transition of the home gateway architecture from an embedded one to a virtualized one will bring many benefits to the broadband service provider. We also believe that the technological infrastructure for the transition is for the most part ready and that the transition should happen in the near future. We encourage Service Providers and ecosystem vendors to contribute to making this transition as soon as possible.

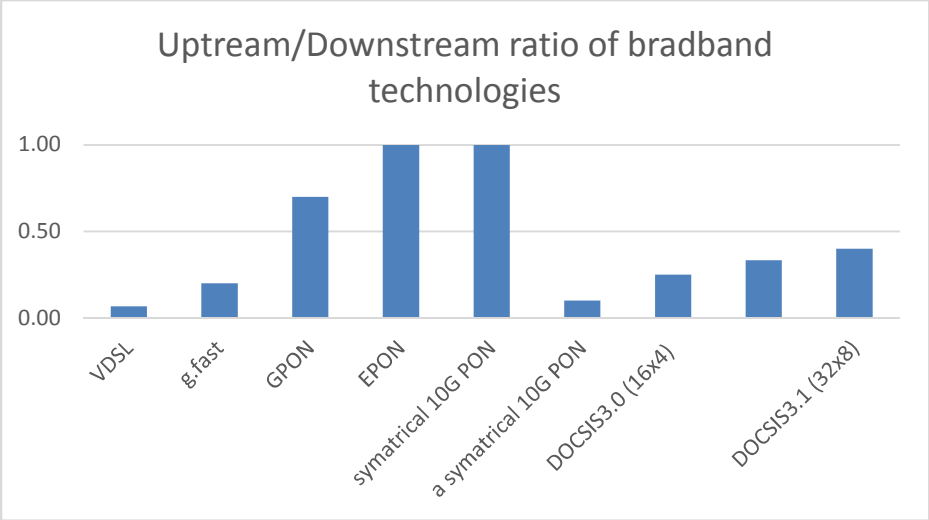


Figure 9-1 - Up/Down Bandwidth Ratios per Technology

11 REFERENCES

[1] Nagesh Nandiraju , Ph.D. and Sebnem Ozer, Ph.D.; Arris Inc., "Applying the Software Defined Networking Paradigm to MSO Commercial networks," in *NCTA Spring Technical Forum*, 2013.

[2] Nagesh Nandiraju Ph.D., Yiu Lee and Jorge Salinger; Comcast Cable, "Delivering Seamless Subscriber Aware Services over Heterogeneous Access Networks using a SDN and NFV framework," in *INTX Spring Technical Forum*, 2015.

[3] Cable Labs, *Data-Over-Cable Service Interface Specifications MAC and Upper Layer Protocols Interface,* Cable Labs, 2015.

[4] G. Ragoonanan and Y. Gorkem, "vCPE services business case: potentially billions of dollars payback for fixed CSPs," Analysys Mason, 2015.

[5] D. Farinacci, T. Li, S. Hanks, D. Meyer and P. Traina, "IETF.org," IETF, March 2000. [Online]. Available: tools.ietf.org/html/rfc2784.

[6] W. community, "Wikipedia," Wikipedia, [Online]. Available: en.wikipedia.org/wiki/Internet_Mix.

[7] "openvswitch.org," [Online]. Available: openvswitch.org.

[8] European Telecommunications Standard Institute, *Network Function Virtualization (NFV); Architectural Framework,* ETSI, 2014.

[9] LinuxContainers.org, "LinuxContainers.org," [Online]. Available: linuxcontainers.org.

[10] "The reality of cost reduction," HP, 2015.

[11] Light Reading, 11 12 2014. [Online]. Available: http://www.lightreading.com/nfv/nfv-elements/alcatel-lucent-joins-virtual-router-race/d/d-id/712004.

[12] Light Reading, 19 8 2014. [Online]. Available: http://www.lightreading.com/nfv/nfv-elements/telefonica-proves-brocade-router-performs-for-nfv/d/d-id/710397.

Hi Ho, Hi Ho, to a Gigabit We Go
Positioning the HFC Network for the New Gigabit Era

Phil Miguelez
Comcast

Abstract

The broadband landscape is rapidly changing driven by market competition from non-traditional providers, a resurgence in new home construction, and a vast array of new services to support the connected home environment. Traditional node splitting is not enough to keep up with bandwidth demands and high speed data rate comparisons to predatory fiber to the premise over-builders. The decisions that cable operators must make to meet these challenges include options that affect every part of the infrastructure network: facilities, OSP, fiber augmentation, and plant powering. Capacity planning which barely existed just a few years ago is now the major focus of every MSO.

This paper will compare and contrast the different network architecture options that are available today and developing network technology that will be ready to deploy in the very near future – DOCSIS 3.1 with RF bandwidth expansion, Fiber Deep cascade reduction, FTTH, and R-Phy / R-MacPhy distributed architectures. The paper will also explore the relative cost implications of each architecture design and the capability to migrate to the next generation as data capacity demands continue to grow and accelerate.

INTRODUCTION

Background

MSO HFC access network architectures have changed very little for over a decade. The arrival of DOCSIS 1.0 in 1997 and the subsequent enhancements leading to the release of DOCSIS 3.0 in 2006 helped to spur the last major network upgrades. During that time QAM video channel expansion dominated the growth of the downstream bandwidth while the still emerging HSD, IP telephony, and VOD services drove expectations of further bandwidth expansion requirements in the future.

GaAs technology allowed the development of 1 GHz nodes and amplifiers that fulfilled the potential needs of D3.0 access architectures. GaN technology followed providing the increased output levels necessary to convert 750 MHz or even older 550 MHz systems into 1 GHz networks with little or no expensive respacing of existing amplifier locations. HFC appeared to have all the tools necessary to compete with Direct TV satellite delivery of video content and the newly minted Regional Bell Operating Companies (RBOC's) that began construction of massive and expensive fiber systems to replace the twisted pair copper wireline networks that had been in place for nearly 100 years.

A negative consequence of the last rebuild cycle was that Wall Street increasingly took a dim view of further network upgrades without a clear perspective on the competitive or cost benefit. The economic recession that started in 2008 also dampened any incentives to expand the capabilities of the existing network.

Narrowcast and HSD Growth

Downstream broadcast channel growth has continued to expand fueled by broad adoption of High Definition (HD) video along with an array of sports, special interest, and ethnic diversity programming. At the same time the

number of narrowcast channels has also steadily grown although at a slower pace. DOCSIS 3.0 provides for up to 32 bonded DS channels representing just over 1 Gb of digital capacity. Until recently almost all systems operated with only 8 to 12 NC channels. These channels carried voice, VOD, and high speed internet data sufficient to meet the tier rates guaranteed by the MSO based on the number of subscribers in the node service area.

High speed data usage has also been steadily increasing since the early days of audio modems. In 1998 Jakob Nielsen created a predictive law in which he stated that high end home users' bandwidth would grow by 50% per year. Subsequent data has validated what is now widely known as Nielsen's Law.[1] Figure 1 shows a plot of the projected data rates extended out for the next 15 years. The compound annual growth rate (CAGR) predicted by Nielsen has driven the steady BW expansion of the HFC network along with node segmentation and the gradual reductions in HFC amplifier cascades, resulting in lower numbers of homes passed per node. Every MSO has created their own set of CAGR charts reflecting factors specific to their network, customer base, or personal beliefs. These growth rate predictions will continue to drive investment for many years to come.

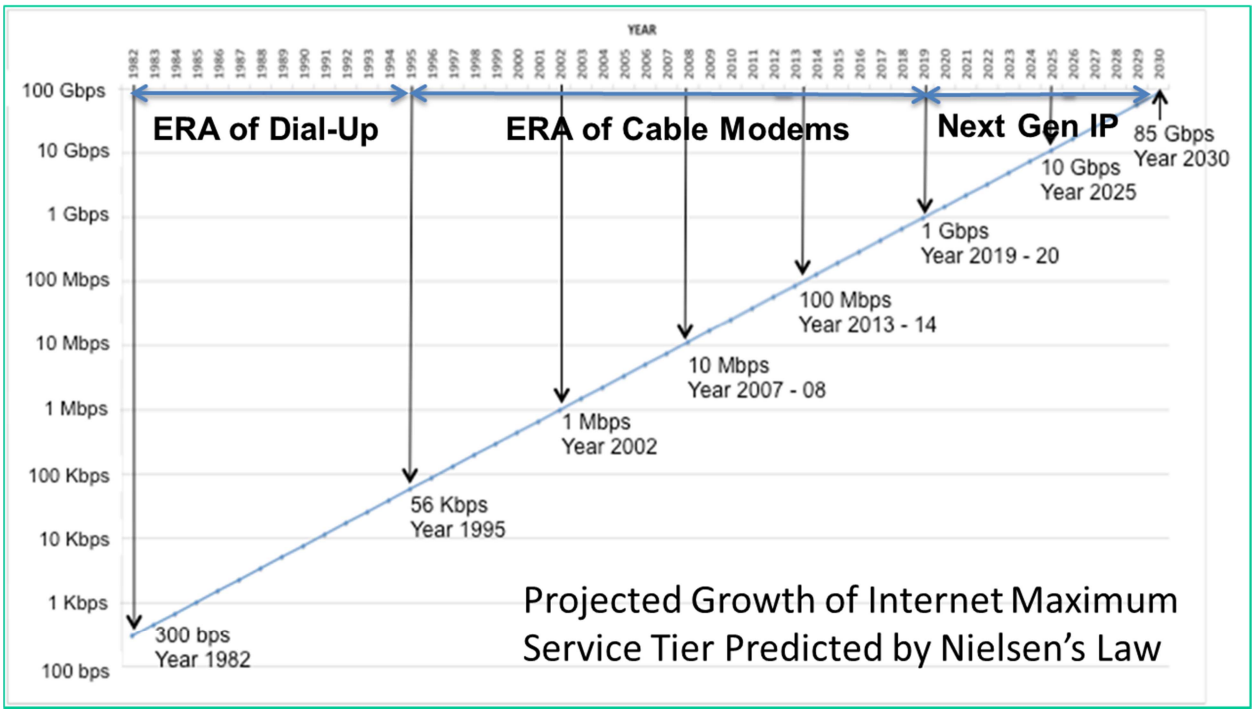


Figure 1 – Nielson's Law Predicts 50% Year Over Year Growth Rate

Downstream channel growth has historically been the largest driver of bandwidth in MSO networks. In order to create additional DS bandwidth without making changes to the outside plant operators have reclaimed analog channels, incorporated higher digital compression rates in QAM video channels, and installed Switched Digital Video (SDV) systems. SDV takes advantage of the fact that a number of channels have limited peak viewing time so these channels can be dynamically reclaimed when no one is actively watching.

Increasing narrowcast bandwidth per subscriber requires an additional set of network tools. Data consumption has continued to rise in line with predicted rates.

The primary driver of this consumption is the increasing popularity of social media and streaming video. Netflix, You Tube, and other popular video streaming services have had a major impact on the amount of data consumed by individual subscribers. Netflix alone accounts for roughly 35% of all internet downstream bandwidth.[2] Figure 2 shows the most recent data available on wireline data consumption.

	Upstream		Downstream		Aggregate	
Rank	Application	Share	Application	Share	Application	Share
1	BitTorrent	28.56%	Netflix	37.05%	Netflix	34.70%
2	Netflix	6.78%	YouTube	17.85%	YouTube	16.88%
3	HTTP	5.93%	HTTP	6.06%	HTTP	6.05%
4	Google Cloud	5.30%	Amazon Video	3.11%	BitTorrent	4.35%
5	YouTube	5.21%	iTunes	2.79%	Amazon Video	2.94%
6	SSL - OTHER	5.10%	BitTorrent	2.67%	iTunes	2.62%
7	iCloud	3.08%	Hulu	2.58%	Facebook	2.51%
8	FaceTime	2.55%	Facebook	2.53%	Hulu	2.48%
9	Facebook	2.25%	MPEG - OTHER	2.30%	MPEG	2.16%
10	Dropbox	1.18%	SSL - OTHER	1.73%	SSL - OTHER	1.99%
		65.95%		78.69%		76.68%

sandvine

Figure 2 – Q4 2015 Top Ten Peak Period Applications – North American, Fixed Access

The number of internet connected devices per home has also been a driving force in the amount of narrowcast bandwidth capacity needed per home.

THE NEW MARKET DYNAMIC

The Impact of Google Fiber

In 2010 Google announced a new business initiative – Google Fiber. Google's announced plan was to create a lottery of sorts to identify cities and towns that would be willing to provide access for new fiber deployment and a sufficient number of potential customers that would pre-subscribe for Google Fiber installation. The winning locations would be built out with FTTH PON and have HSD tier rates up to 1 Gb which far exceeded anything available from cable or other wireline broadband service providers at that time.

Google Fiber's actual competitive impact on the ground has been muted. After 5 years, the total number of video connected homes was recently reported to be only 53,000.[3] The number of internet data only homes is assumed to be higher and has been estimated at around 100,000 which is still quite low considering the number of years since the initial deployment. However, Google has had a major impact on the broadband market by raising the bar on subscriber expectations for digital delivery of high speed data rates and an even greater impact on the potential providers of broadband service delivery.

Google's size and name recognition allowed them to negotiate concessions with numerous municipalities in exchange for the investment that Google Fiber would provide to their community. The concessions included right of way easements and access to utility pole attachments which have always been costly barriers to new entrant providers. A

consequence of the negotiations by Google was that they created a template that smaller FTTH broadband service provider start-ups could use to get equivalent deals when they entered a new market to compete with an incumbent provider. Within a few years FTTH organizations were springing up touting plans to "Gigafy America" and teaching startup companies how to sell and construct FTTH systems.

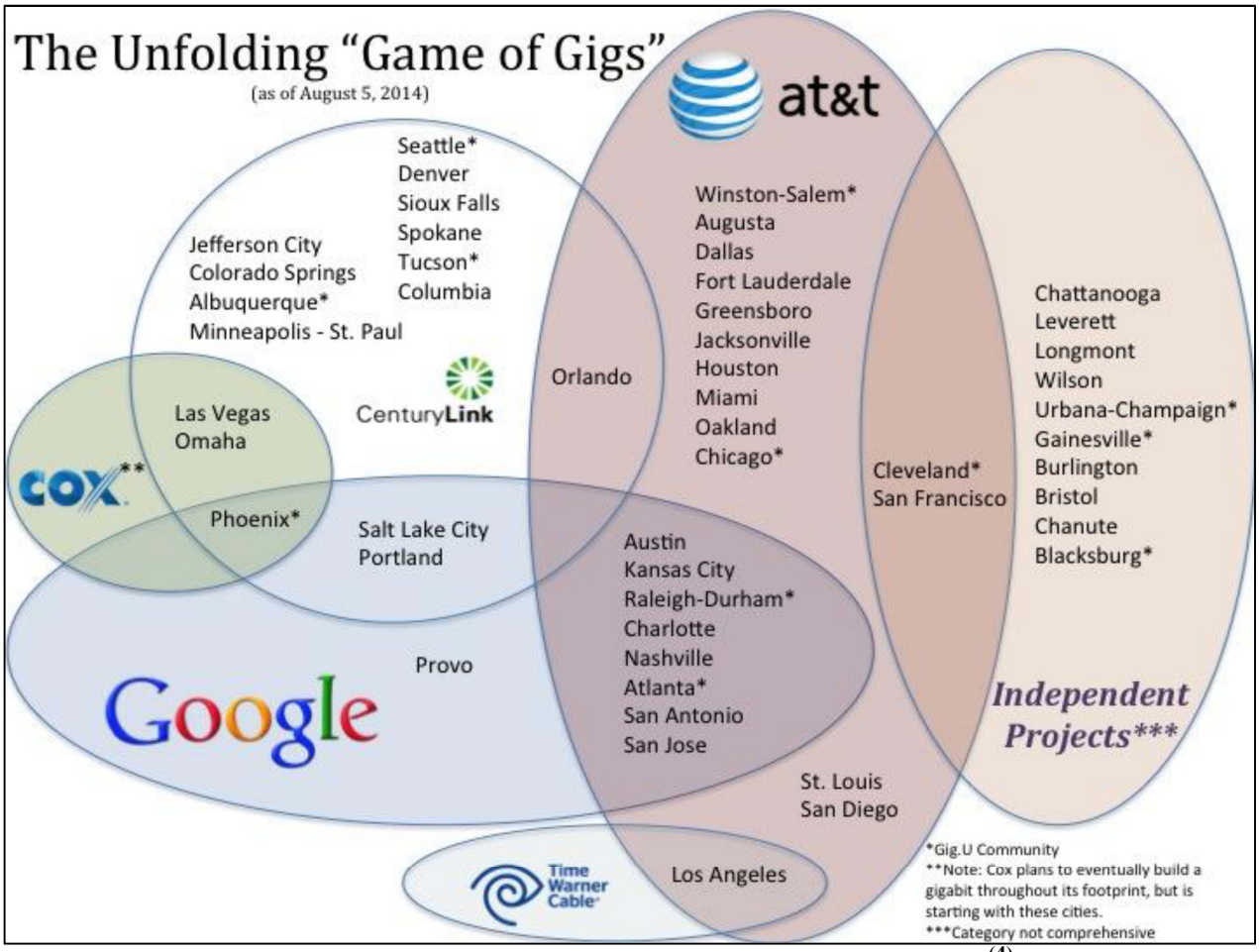


Figure 3 – Target City Deployments Announced by Major Over-Builders[4]

Figure 3 shows a graphical representation of the cities targeted by over-builders for Gigabit FTTH deployments as of August 2014. Many of the listed cities are Comcast, Time Warner Cable, or Cox Communications franchise locations. Additional cities have been announced in 2015.

More recently, AT&T has modified their direction announcing "Gigapower"[5] a fiber delivered 1 Gb internet data service. AT&T is now expanding their reach into targeted cities and residential areas wherever they have fiber assets. Verizon FIOS has taken a slightly different tact with clever ads such as the "Half Fast" campaign which is intended to point out that the cable upstream data capacity is much lower than the downstream data capacity. Verizon has followed up this marketing campaign by introducing symmetric (DS/US) data plans putting pressure on cable operators serving the same markets to expand their HSD tiers and add US capacity. Century Link, Hotwire, and countless smaller regional over builders are cherry picking MDU and gated community properties that provide 100%

penetration opportunities with long term contract guarantees. The market landscape is now filled with large and small competitors offering gigabit speed service and threatening to overtake the HFC cable space.

While it's true that the only immediate application for gigabit data rates today is the speed test, history has shown that new applications seem to always appear once the network BW is available to support them. One example is the number of smart devices in the typical home. A few years ago the network assumption was 3 to 5 devices on average and 5 to 7 devices in a high end user home. Today those numbers have jumped to an average of 5 to 7 connected devices per home and 10 to 12 for a high end subscriber.

Another pending driver is the Internet of Things (IoT). The number of internet connected devices is exploding. Smart home applications and monitoring devices especially video cameras are driving both US and DS bandwidth usage. Figure 4 shows a graphical representation of the various smart home devices available today.

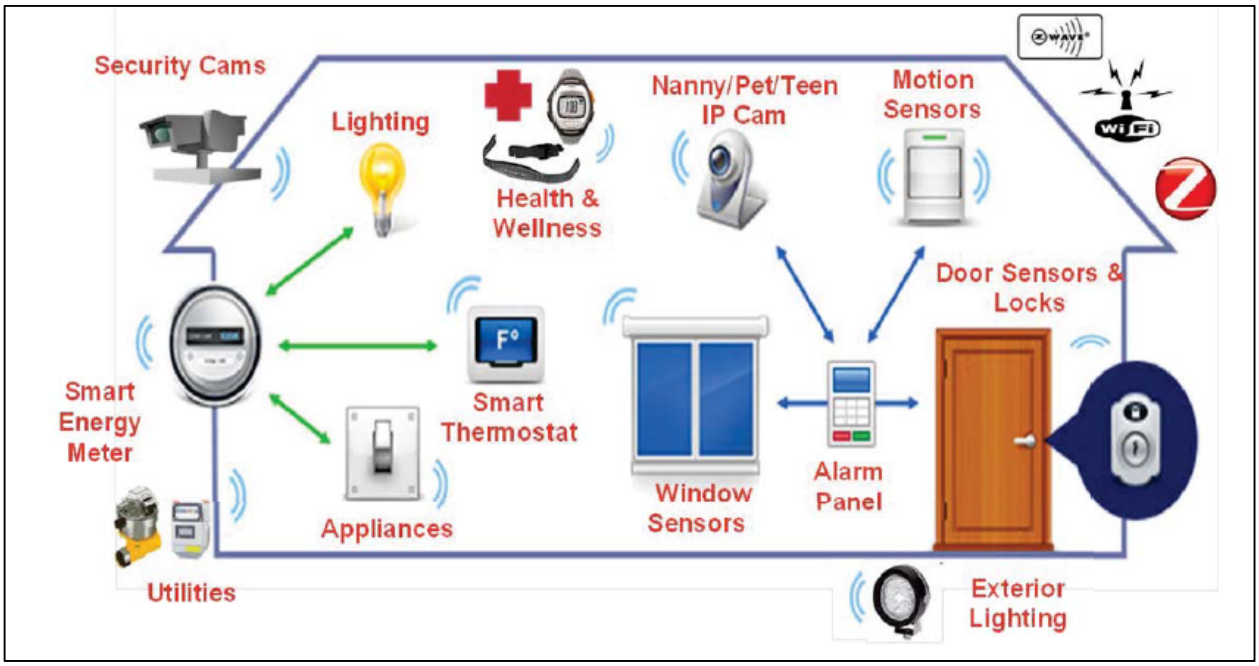


Figure 4 – Examples of the New Connected Devices That Make Up the Internet of Things

MEETING THE GIGABIT CHALLENGE

Realities of Existing HFC Networks

The DOCSIS 3.0 specification bounds the DS modulation format to 256 QAM in 6 MHz channels. With 32 bonded Single Carrier (SC) channels the usable data throughput is 1.216 Gb. Finding open channels or reclaiming existing channels in a typical 750 MHz system is a daunting task. Transitioning from analog video to all QAM is almost a mandatory first step. Replacing MP2 with MP4 modulation can also provide additional BW relief. Even these steps may not be enough considering that any increase in DS bandwidth will precipitate the need for added US capacity enhancement. This means added US channels must be allocated plus an appropriate guard band to adequately isolate the forward and return signal paths.

In higher density HP areas and especially urban serving areas, the number of targeted channels needed to satisfy multiple ethnic and special interest communities makes finding open channel space virtually impossible without an added RF bandwidth expansion. The same situation exists in low density and rural HP serving areas where many networks have been stretched to reduce the number of actives that must be maintained. This usually results in lower performance margins and reduced total bandwidth due to increased roll off at the upper band edge.

The digital capacity in an HFC system represents shared bits for all subscribers connected to a particular node. The typical homes passed node size has been steadily decreasing over the years. Physical node splits and node segmentation have reduced a node's serving area from >1500 HHP to the current average of 500 HHP. Even with this significant reduction, a few simultaneous shared users could potentially drag down the usable delivered data rates during peak time periods. Using historical CMTS concurrency rate assumptions the maximum guaranteed downstream data rate tier that could be supported with 32 bonded D3.0 channels would be 250 to 300 Mb at the current average node HP size.

The traditional method to increase narrowcast digital BW available per subscriber is to reduce the number of HHP per node. Node segmentation provides operators with a minimally disruptive method to significantly increase the delivered bandwidth per subscriber. The cost of the initial primary node segmentation is typically estimated at $20,000 since the majority of the expenses are usually limited to material costs rather than new fiber deployments.(6) By comparison, subsequent node split costs can increase almost exponentially due to fiber construction expenses, Headend expansion costs to support the additional service groups, and when calculated based on the fixed number of subscribers served by a particular node.

The high costs of node splits when evaluated as a means to compete with accelerating high speed data requirements led Comcast and other cable operators to examine alternative architectures.

D3.1 Availability in the HFC Network

In June 2012 CableLabs initiated a new specification effort to establish requirements for a major revision of DOCSIS. The DOCSIS 3.1 specification which was released in October 2013 defines a new modulation standard for HFC networks with a raw data capacity of 5 Gbps downstream (DS) and up to 1 Gbps upstream while maintaining the current 1 GHz RF bandwidth capabilities of existing D3.0 cable plant. The D3.1 specification provides an intermediate downstream RF bandwidth target of 1218 MHz which aligns with the deployment of near term achievable hardware that is already coming to market. The potential for 10 Gbps downstream capacity is achievable with an RF spectrum expansion to the full spec 1.794 GHz D3.1 DS BW allocation or eliminating the QAM broadcast channels. Figure 5 shows the subscriber data rates available for the various DOCSIS releases over the years.

Upstream capacity growth is provided by DOCSIS 3.1 with the implementation of RF bandwidth expansion to either a mid split of 85 MHz or a high split to 204 MHz. The high split option is required to achieve the 1 Gb raw capacity capability. More details on the D3.1 specification can be found on the CableLabs website.(7)

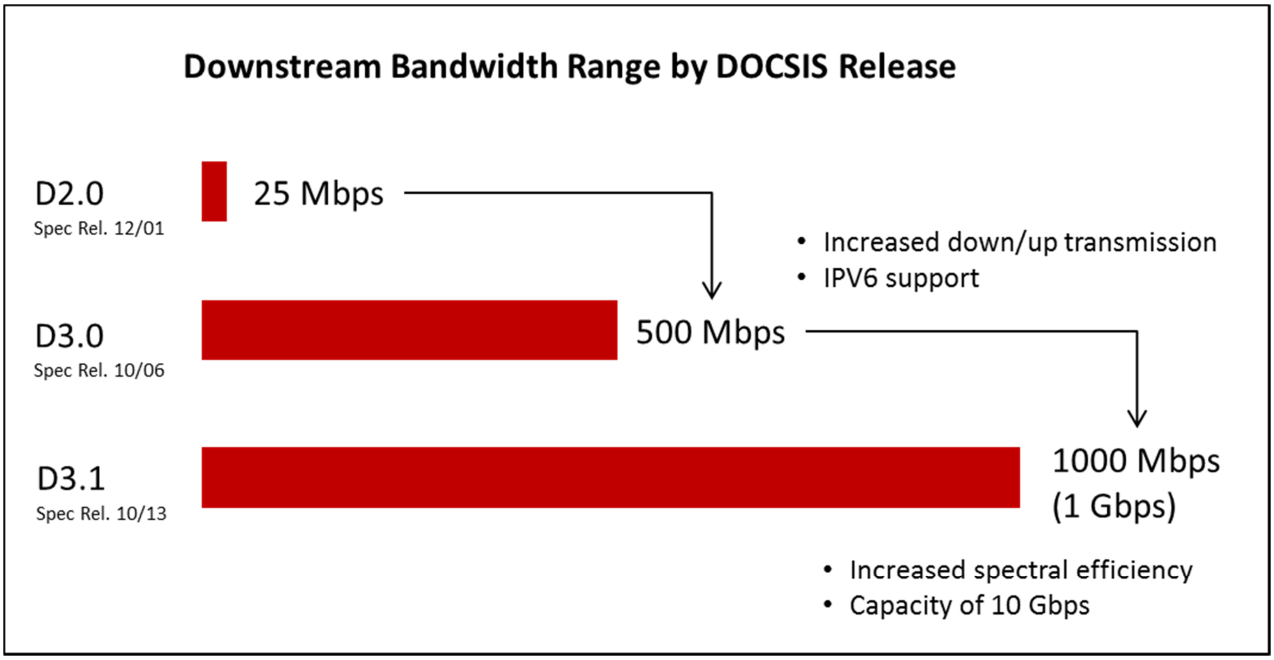


Figure 5 – DOCSIS Path to Gigabit Speeds

DOCSIS 3.1 initial silicon has been released and compatible cable modem samples have been in lab evaluation for more than 6 months. Preliminary field trials have been deployed beginning in Q3 of 2015. System level deployments have been announced and are expected to start in late 2016 / early 2017[8].

OFDM modulation and higher order subcarrier QAM formats allow DOCSIS 3.1 to increase the number of bits in a given channel block thus improving the available capacity. But this is still shared spectrum depending on the number of homes served by the local node. Subsequent sections of this paper will describe the architecture choices that determine the ultimate data capacity available per individual subscriber.

THE GIGABIT ERA
WHICH PATH TO CHOOSE?

Competition and HSD growth curves will continue to push MSOs to deploy gigabit and higher HSD capable networks or risk losing subscribers to a new entrant service provider with claims of a faster network. There are multiple solutions that a cable operator could employ to migrate the existing HFC network to gigabit per subscriber capability. The tradeoffs between different solutions usually are centered on the network migration costs and the expected useful lifetime of the network change.

The next sections describe the features of the possible migration options followed by details on the cost analysis of each migration path.

Fiber to the Home

The housing boom that occurred in the early 2000's saw many developers insisting on a fiber ready or fiber to the home solution for their properties. Fiber was considered a future proof technology and home builders were able to charge up to $10,000 more for a fiber enabled home. When the housing market began to recover after the 2008-2009 recession a shift to MDU greenfield properties had occurred but the demand for fiber only entry into the building continued. In most cases contracts for these properties can extend

for 5 to 10 years and many times the selected service provider has an exclusive contract making these MDU's the prime target for over builders and too important to ignore for MSOs.

Fiber to the basement (FTTB) has been a long standing MDU design solution for many years. Fiber is brought into the building to feed a standard HFC node. The output of the node is distributed over coax to the individual subscriber housing units. In cases where the developer demands fiber to the home service, RFoG is the deployed solution since it is transparent to the existing HFC back office and installed CPE equipment.

Unfortunately, FTTB and RFoG are limited to current HFC data rates and vulnerable to gigabit competitors. RFoG is not an ideal solution for other reasons. It's biggest weakness is optical beat interference (OBI). MDU's tend to be a worst case environment due to potential ingress sources and the fact that many of these new communities are premium subscribers with a higher than average number of connected devices that tend to increase the chances of OBI.

To counter the gigabit competitive threat a PON overlay must be deployed. The most commonly deployed PON formats today are GPON which provides 2.5 Gb DS / 1.25 Gb US and EPON which provides 1 Gb DS / US. These are PHY layer rates including packet overhead that will reduce the actual delivered data rate. The primary attraction of GPON and EPON is the use of lower cost uncooled optics. This made PON CPE pricing more affordable.

GPON is capable of supporting peak rates of 1 Gb symmetrical service while EPON's lower 1 Gb Phy rate means that packet overhead limits the peak rates to the low 900 Mb range. Speed test measurements can also be highly variable since the Ethernet switching chip sets used in commercially available modems, laptops, and WiFi equipment are limited to a maximum of 1 Gb. Measurements made in a home environment can typically range from 800 Mb to 940 Mb.

A challenge unique to cable operators is that DOCSIS HSD tiers are commonly guaranteed delivered rates rather than best effort. A 100 Mb tier subscription results in a verifiable 100 Mb delivered data rate. Offering 1 Gb FTTH service using GPON/EPON could put the operator in a difficult position since the delivered rate could not be guaranteed similar to DOCSIS tier rates. Also, MSOs deploying GPON/EPON have no competitive advantage compared to rival PON over-builder solutions.

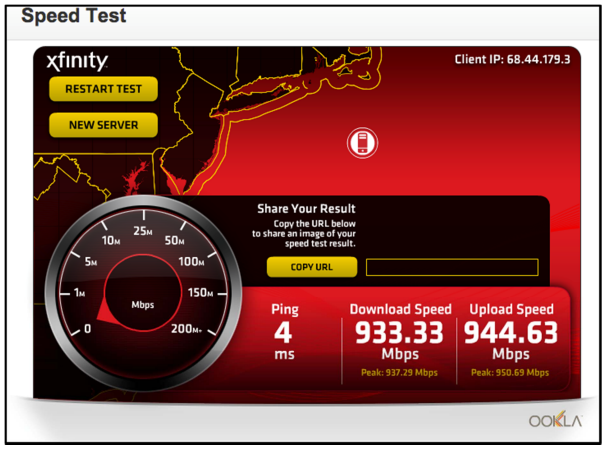


Figure 6 – Symmetric Gigabit Speed Test

10G EPON is an available solution that can solve the issues described above as well as provide expandable capacity for >1 Gb tier offerings. The cost of 10 Gb optics while higher than today's 2.5 Gb un-cooled optics are quickly decreasing based on growing volumes in North America and China. The Telco market is beginning to feel the pressure from emerging 10G EPON competitors. FSAN, the ITU standards organization that defined the GPON protocols has recently initiated a new XGS PON spec that is compatible with 10G EPON optics. Before long current GPON competitors will be raising the bar with 10 Gb service offerings.

FTTH construction costs are significantly higher than any other HFC alternative. While fiber cable and coax material costs are nearly equal the added material and labor costs associated with optical passives, fiber splicing, and connectors are very different. In building fiber to the home wiring is the major cost driver for this architecture.

Business as Usual (BAU) Node Splits

This is basically a strategy of delay allowing usage based node splits to keep up with increased data requirement hot spots until a clear view of the future migration path becomes evident. This migration tactic does not provide a direct path to gigabit per subscriber HSD capacity.

Historically, the number of node splits as a percentage of the total network has been consistent at roughly 2.5% to 3% for several years and remains very manageable. Since this level of node splits is simply a sustaining operation it does not protect against the accelerated growth rates predicted by Nielsen and similar CAGR estimates. It also does not provide any protection from competitive threats.

Traditional node splitting results in the highest network migration costs. Based on the 50% yr/yr capacity growth projections the number of node splits needed would go expediential within 4 to 5 years. HE and Hub capital equipment, rack space, power, and HVAC additions needed to support the increase in node splits will be a major driver to the costs of this migration approach.

1 GHz Drop-in RF BW Expansion

Another often discussed migration approach is to continue the traditional RF BW expansion strategy that has been used many times in the past. Changing out the RF electronics in the existing plant node and amplifier actives would provide additional RF spectrum that could be used for D3.0 channel bonding or to create room for a future channel block of D3.1 OFDM. Improvements in GaN RF hybrid and MMIC gain blocks allow the expansion to 1 GHz without the need to respace the location of the existing plant actives. All of the OSP actives would need to be touched in order to effect the DS BW enhancement. Therefore the opportunity exists to modify the upstream split at the same time thereby increasing the return path digital capacity. While the initial cost of this network upgrade method is very favorable compared to every other alternative, there are still a number of very real downsides.

1 GHz lasers, nodes, and amplifiers have been available for over a decade. 1 GHz taps and passives have been the standard for even longer. Every OEM manufacturer has obsoleted 750 MHz and 870 MHz versions of their actives portfolio and today only offers 1 GHz qualified products. Plug in accessories allow the gain and tilt to be adjusted such that the drop-in RF electronics module can be backward compatible with legacy equipment bandwidth designs or provide the necessary drive to obtain 1 GHz extended performance. In actual operation many cable operators have discovered that cascaded roll off factors at the high end of the RF spectrum limit the usable bandwidth to roughly 940 MHz. The frequency roll off is a cumulative effect caused by the RF response signature of the distribution taps and passives as well as some older HE lasers and OSP actives. The end result is a reduction in signal to noise performance that makes operating channels above 940 MHz extremely marginal. As a result, the cost of the BW expansion may not be as attractive with only a 70 MHz (for an 870 MHz legacy system) or 190 MHz (for a 750 MHz legacy system) frequency spectrum gain. This is especially true if the upstream split is modified at the same time further reducing the total DS frequency increase.

Another downside factor to consider is that while a 1 GHz drop-in frequency expansion increases the potential digital capacity of the system this is still shared bandwidth in a relatively large node serving group. Also, roughly 75% of existing multi-port nodes are unbalanced meaning that one leg of the RF distribution access link leaving the node has considerably more subscribers than the other RF legs. As a result, unless there is a significant reduction in the amplifier cascades on the unbalanced node leg, the effective data rate increase on a per subscriber basis will be relatively low. The drop-in approach will therefore still require future node splits to meet the competitive threat. These added node splits increase the capital expenditures required to support the new serving group additions.

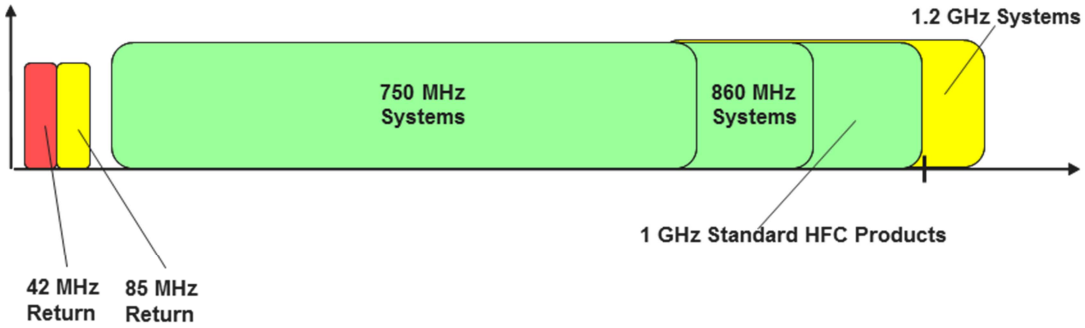


Figure 7 – HFC RF Spectrum

<u>1.2 GHz Drop-in RF BW Expansion</u>

A 1.2 GHz drop-in BW expansion will be several orders of magnitude more difficult and costly than 1 GHz. Increasing the usable bandwidth to 1.2 GHz means that every part of the network must be touched. CMTS, headend RF distribution, OHE lasers, nodes, amplifiers, taps, and CPE. The sheer scale of this level of changes without reducing the serving area of the nodes makes this architecture choice very questionable.

The higher coax attenuation at 1.2 GHz requires higher output levels and higher tilt to compensate for the increased loss. Node designs for 1.2 GHz have been available for a few years but these early products have all been designed using hybrid gain blocks that are essentially 1 GHz devices with the RF response re-tuned to meet the 1.2 GHz bandwidth requirement. The output level specified for these nodes is typically lower than a 1 GHz comparable model due to the increased channel loading. More recent development work geared to N+0 Fiber Deep requirements has generated higher power hybrid gain blocks including versions with more appropriate levels for an N+X cascade design.

The added channel loading for a fully loaded 1.2 GHz laser results in a 0.6 dB drop in laser OMI per channel which is equivalent to a 0.6 dB reduction in carrier to noise at the node receiver. The OMI reduction also requires an increase in node receiver sensitivity and gain to compensate for the lower modulation level.

RF Bridger and Line Extender amplifiers for 1.2 GHz are not yet generally available. Vendors supplying the European markets are several months ahead of their north American counter parts since they are seeing a higher level of interest from western European cable operators. Concerns about the increased levels needed to avoid re-spacing and the actual

value proposition for a 1.2 GHz extended BW cascade have delayed the introduction of these products in the north American market.

FIBER DEEP N+0 ARCHITECTURE

Overview

Fiber Deep architectures have been discussed for many years and have different definitions depending on the author. The variations are largely based on the number of amplifier actives allowed after the node (N+small or N+0). N+small network designs provide a lower initial cost but eventually will require additional segmentation to fully optimize the digital capacity per subscriber. The end state of HFC Fiber Deep is N+0 which is the definition used for this paper.

An N+0 architecture comes closest to the ideal of a last touch HFC network design. A brief list of goals for this topology is the following:

- Centrally locating the node to provide balanced outputs
- Optimally reducing the node serving area to achieve the desired data rate tier capacity per HHP
- Eliminate all actives following the node to improve OpEx performance and provide a passive coax access link to the home
- Maximize the effectiveness of D3.1 and Distributed Architectures
- Provide the same DS data capacity as 10 Gb PON with the migration to D3.1 and 1.2 GHz bandwidth expansion
- Establish the FD node location as the Edge Network demarcation launch point for future system migrations such as Full Duplex DOCSIS or FTTH
- Extend the competitive life of the HFC network for many years to come

A Fiber Deep architecture is not without challenges. The previous section of this paper described the high touch impact of a 1.2 GHz network migration. Node + 0 increases the complexity due to the changes that need to be made to the actual HFC plant configuration.

Fiber Deep relies on new technology developments in order to be a successful and practical architecture. Depending on the homes passed density of the target system the number of new nodes required can range from 10:1 to as high as 16:1 compared to the existing N+X cascade design. The impact on the available rack space within the local Hub or HE can be dramatic. CCAP and higher density OHE platforms help to conserve rack space but the transition to smaller serving group's means shifting to 1:1 port count associations with lasers and nodes. There are very few Hubs with the space, power, and HVAC capability to accommodate the growth in equipment associated with this significant increase in nodes. Without a relatively near term solution new Hubs would need to be built to support the increase in FD nodes. The cost and time needed to acquire property, route connections and build out the infrastructure is prohibitive. This is the inevitable problem with continuing BAU node splits but has also been the main show stopper for alternative cascade reduction architectures, including fiber deep. The solution is to disaggregate and distribute the Layer 1 elements and some of the management functions from the CCAP platform out to the node. Distributed Access Architecture (DAA) solutions are being developed and evaluated today that will significantly reduce Hub density. These solutions could be available in the market starting the second half of 2017.

Other development efforts such as all IP transport and SDN/NFV will further reduce the current equipment density in the Hub and could allow the eventual consolidation of secondary Hubs into a master headend.

N+0 Node Size and Effective Reach

Up until the release of DOCSIS 3.1, fiber optic PON networks have had the potential advantage of providing the highest delivered data rates to an individual subscriber. In actual systems this was rarely realized since FTTH networks typically over subscribe the available data capacity to keep capital equipment costs as low as possible. Google Fiber, AT&T, and the multitude of new over-builders that are now challenging the established MSO networks have changed the paradigm by offering 1 GB data at a low monthly cost. DOCSIS 3.1 provides the digital capacity to compete with PON. Reducing the node serving area size allows an HFC network to optimize the data rate capacity available per subscriber. A target node size of 128 HHP delivers the same data rates as a 10G EPON port and positions the network for a future transition to PON when needed.

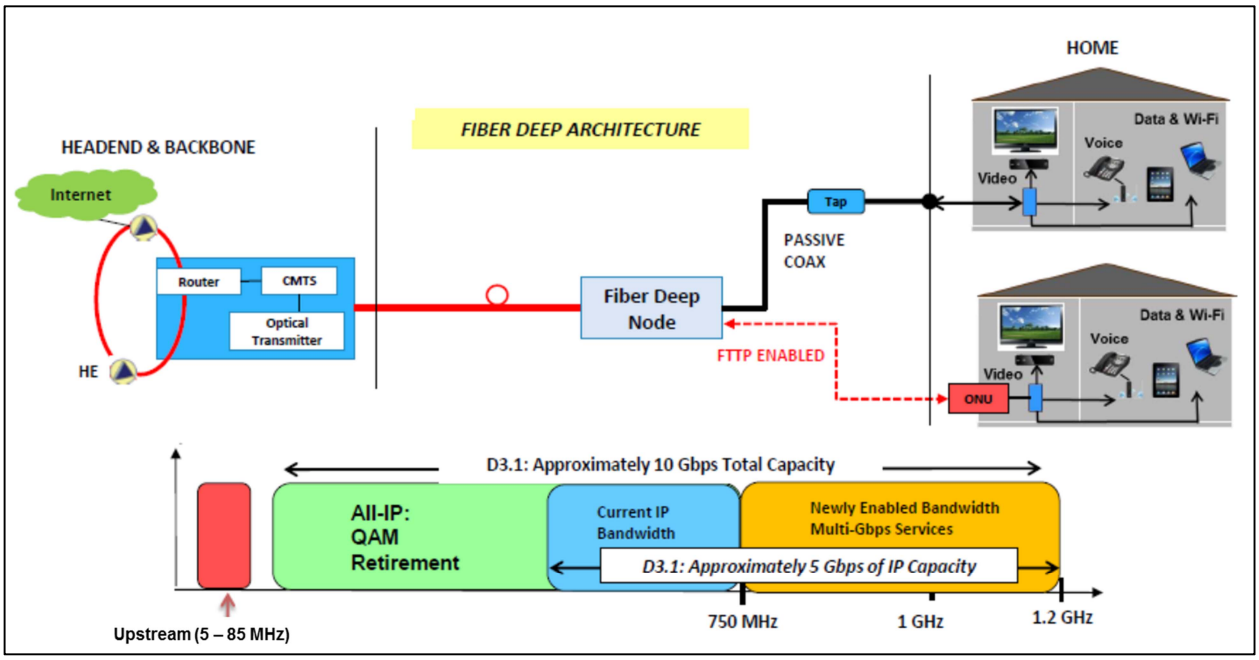


Figure 8 – Fiber Deep Architecture and Bandwidth Diagram

Figure 8 shows a pictorial of the Fiber Deep architecture design along with a bandwidth diagram depicting the current and future loading. The diagram also shows the potential for multiple services such as PON or P2P Ethernet connections emanating from the HFC FD N+0 node location.

As discussed previously, providing sufficient BW to take advantage of the full benefits that D3.1 offers requires extending the BW to 1.2 GHz. Coax plus passive losses at 1.2 GHz are considerably higher than a 1 GHz HFC network. By comparison, a DOCSIS 3.0 N+0 node design requires 60 dBmV output (analog equivalent power) and a tilt of 18 dB measured from 54 MHz to 1002 MHz. To achieve the same equivalent reach and performance at 1218 MHz the tilt line must be linearly extended 4 dB with an equivalent increase in the node output level. This seems like a daunting task but the limits of GaN semiconductor technology have not yet been reached and after a year and a half of development effort two major device suppliers succeeded in creating qualified

hybrid gain blocks that are now commercially available. This accomplishment was only possible as a result of working in close partnership with multiple node design teams to solve technical hurdles related to power consumption and thermal capacity limitations of the node housing and cable plant powering.

The physical characteristics of an HFC node are constrained by OSHA regulations for the safety of the installers. Other standards are also closely watched such as the National Electric Code rules, Utility service provider requirements, and intra-service agreements which all define the physical spacing between different services hanging on the outside aerial plant or coexisting in underground conduits and access tunnels. The mechanical dimensions of the housing play an important role in determining the maximum power dissipation that the node can handle. Careful design including optimizing switching power supply efficiency has resulted in the high output Fiber Deep node consuming the same or lower power than a legacy 4X4 node today. Further improvements are still possible including efforts underway today to introduce predistortion techniques to the RF power amplifier stages in order to lower power consumption while maintaining performance.

At a target node size of 128 HP only very high density urban serving areas would possibly require a segmented node. In almost every other case a 1X1 centrally located four output node will accommodate the current legacy network densities. Unfortunately, even D3.1 mid split band plans will not resolve the asymmetric capacity of the available US BW so provision for 1X2 segmentation is incorporated in the node requirements.

Field trial designs have shown that at 64 dBmV node output the 128 HP maximum cap is an appropriate target. In reality, home densities vary from system to system and neighborhood to neighborhood so there is a wide distribution that occurs in most FD designs. The average node size based on field trial deployments is 64 HP. This value is used to estimate BOM equipment requirements in the HE and outside plant for all new fiber deep builds. At an average 64 HP node size the typical FD node ratio is a 12:1 increase over the existing legacy network.

Fiber Deep RF Distribution Components

The configuration change to 1.2 GHz N+0 requires modifications to every aspect of the network. Distribution Taps and Passives have been designed that maintain the current 1 GHz performance plus extend the bandwidth to 1218 MHz maintaining the same monotonic insertion loss response characteristic. A major cost saving feature of these devices is that the Tap faceplate is mechanically identical to the legacy housing allowing the tap housing body to stay in place and only the faceplate to change. In FD field trial designs 60% of the tap values must be changed. The ability to change the faceplate instead of cutting out and replacing the entire tap housing amounts to a considerable construction cost savings.

While creating the new 1.2 GHz faceplate designs the tap vendors also made improvements to the surge resistance and shielding effectiveness of these devices. These changes ruggedize the product against electrical storms and power line issues as well as improve ingress resistance to LTE and other interferers. Every major tap manufacturer has followed this same guideline and plans to obsolete and replace the current 1 GHz devices with 1.2 GHz standard product later this year.

The Mid Split Dilemma

The most difficult part of implementing any new architecture is making any change to the upstream that affects customer premise equipment. Data usage between downstream and upstream has always been asymmetrical. In the late 1990's and early 2000's as internet

traffic was growing by leaps and bounds the ratio of DS:US traffic had narrowed to 5:1. In 2007 Netflix initiated video streaming and since then the popularity of over the top video caused the ratio to widen again to 10:1. Today, home monitoring and the increasing number of smart connected devices in the home is beginning to pressure the ratio back downward once again. DOCSIS 3.1 and Fiber Deep do not remedy the asymmetric upstream situation. They do add incremental bandwidth that can keep the DS:US ratio in check.

The dilemma with moving the upstream frequency split is the nature of the wide range of legacy set top boxes that are still in service today. Modern STBs have the capability to be frequency agile meaning that the low data rate communications pilot that the STB uses to validate the channel map of video signals that can be received can be moved across a wide range of frequencies and the STB will search, find, and lock in the new pilot channel. Different generations of STB's had different levels of frequency search capability. Almost all of these newer boxes can be reset via uploaded software.

One of the most widely distributed digital set top boxes is also one of the oldest in service, the General Instrument DCT2000 developed in the mid 1990's. This box has very limited memory size and therefore cannot be reprogrammed to increase its agility. The only solution in this case is to remove and replace the DCT2000's in service. This entails identifying and informing the affected subscribers, arranging to swap the box with a new model, and verifying that the changes are all in place before transitioning to the new mid split. Any boxes that have not been replaced will lose connectivity when the mid split is activated. The replacement process can take 3 to 4 months if everything goes well. The customer care aspects of a process like this must be handled very carefully and professionally to be successful.

Addressing MoCA

MoCA is the primary home networking solution for HFC CPE equipment. MoCA runs over the existing in-home coaxial cabling, enabling whole-home distribution of high definition video and content. MoCA 1.1 and MoCA 2.0 are the currently available versions. MoCA operates in 50 MHz wide channels starting at 1125 MHz (Ch D1) through 1650 MHz.

MoCA 1.1 is the most widely deployed version and provides 175 Mbit/s net throughputs (275 Mbit/s PHY rate). MoCA 2.0 supports two performance modes, Basic and Enhanced, with 400 Mbit/s and 800 Mbit/s net throughputs (MAC), using 700 Mbit/s and 1.4 Gbit/s PHY rates, respectively. MoCA 2.0 is backward compatible with MoCA 1.1.

MoCA allows frequency agile channels and most equipment performs a search for the optimum channel when initially configured. Due to coaxial losses and in-home frequency response characteristics, the lower channels tend to be the most utilized. Many MoCA 1.1 consumer off the shelf products only provide channel D1 support. The end result is that MoCA operates well with 750 through 1 GHz D3.0 in-home networks but the lower D1 and D2 channels will directly interfere with 1218 MHz D3.1 when deployed.

MoCA 1.1 and 2.0 MAC throughput rates will not support D3.1 targeted gigabit data rates. Next generation MoCA 2.5 and 3.0 specifications are currently in development with plans for a phased release schedule. Both will significantly increase the usable data rates providing true gigabit per subscriber rates. MoCA 3.0 includes a shift in the RF spectrum requirements that removes channel D1 and D2 allowing the full 1218 MHz D3.1 bandwidth to be used in the future.

Power Plant Considerations

Cable plant power supplies use a ferro-resonant transformer design that provides 60 or 90 VAC that powers the node, RF amplifiers and other actives connected to the coax access network. The legacy cable plant was originally designed with nodes located near trunk fiber access routes and local utility power locations which typically ran along major highways. This provided easy accessibility to the node in case of failures or normal maintenance. The cable power supply which powered the node and associated amplifier cascade was almost always co-located with the node. Additional power supplies were placed in the system if needed to overcome I^2R losses in long access cable spans or in serving areas with a large number of active elements. Cable plant power supplies are available in a wide range of amperage. The most common power supply size is 15 amps. The primary input power source for the cable PS is local utility AC power. MSOs have traditionally designed back up power systems in case of outages in the local utility AC mains.

There are many degrees of freedom when designing and implementing access network architecture changes. Changing the power grid is not one of them. Adding, relocating, or significantly modifying a supply requires high level interaction and permits from the local municipality and utility power provider. A change in the power grid can take a year to negotiate and implement. This is practically a show stopper for most system deployments. Therefore, for everything except new greenfield construction, any new equipment deployments must safely fit within the margins of the existing AC power capacity design limits. It is also critical to consider future application module deployments and their impact on total network power consumption.

As previously described, an N+0 architecture requires a significant increase in the number of nodes needed to cover the existing serving area. Additionally, Fiber Deep D3.1 capable nodes include higher output RF amplifier stages in order to reach the maximum number of homes possible at 1.2 GHz. Another tenet of N+0 design is moving the node to a central location within the serving area of homes passed so that the node has balanced output ports or as close as can be reasonably achieved. These changes are made without relocating the power supplies in the area or adding new supplies. In order to power the relocated nodes, coax feeder lines are added between the PS and closest node(s). AC power is also fed from node to node by bridging the access coax tap strings between adjacent nodes.

Construction Realities – Lessons Learned

The very first in-house N+0 design of a target node area yielded 20 HP per node. After passing around the smelling salts and consulting with some of the top tier industry designers the subsequent designs raised the node service area to 70 HP per node. Since then design guidelines and training have helped to keep the node designs within the established target window to meet internally established cost targets.

Balancing HP per node RF performance design targets with the power grid design constraints has been another challenge. The ultra-conservative criteria created to accommodate higher output power nodes while maintaining total amperage capacity margins caused design time to more than double. As FD Node product development progressed it was possible to confidently reduce the margin added for each node and bring design time back to normal levels.

There has not been a major construction cycle in the HFC network space for more than 15 years. The number of skilled construction

crews has decreased dramatically as these craft people moved on to other work in different industries. Ramping up construction of a completely new architecture such as Fiber Deep N+0 has highlighted the issues of finding and training new crews. Retaining these crews is another challenge since AT&T and other over-builders are also ramping up their own projects to compete with cable.

An example of the initial skill level of the current crews is the number of coax splices performed per hour. During the peak of the 1990's upgrade cycle the average number of splices completed was 4.5 /hr. In 2015 during the build out of the first N+0 trials the average number of splices completed was only 1.2 /hr. Even bucket trucks have become a rare commodity these days with backlogs extending up to a year in some cases. The rollout of Fiber Deep will accelerate as construction crew training improves and new practices become standards. FD is a multi-year deployment effort.

The success of commercial services product offerings has created a business segment that is a growth leader in our industry. Business services customers were practically non-existent at the time of the last network upgrade cycle so establishing procedures to inform and work with these customers during construction is critical. Node cut in schedules must be determined based on the minimum impact to the customer. This takes a lot of extra planning and over communicating to the businesses that will be affected by the network interruption.

NETWORK MIGRATION COST COMPARISONS

Cost comparisons of the various network migration options were made based on a 10 year outlook. The comparisons assumed that in the case of BAU node splitting more FTTH deployments would be needed to counter PON over-builders. Another assumption was that D3.1 would provide a defense against FTTH competition. Therefore 1 GHz RF Active Drop-In (ADI) would provide some relief against having to build FTTH in order to retain subscribers. Even less FTTH builds would be needed in the case of Fiber Deep since the data capacity of FD was equal to PON competition. Figure 9 shows the relative cost comparisons between the four most likely deployment scenarios.

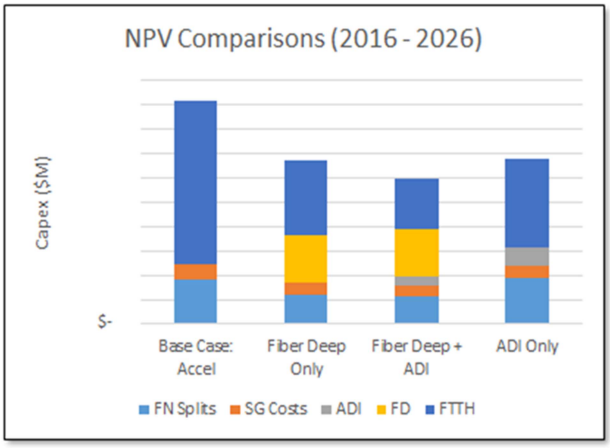


Figure 9 – 10 Year NPV Cost Comparison

Note: ADI includes the cost of subsequent node splits over the 10 year period.

The cost of Fiber Deep is almost the same as 1 GHz Active Drop-In although the expected lifetime of the plant built with ADI is significantly shorter. The best case scenario as seen in the graph is a combination of Fiber Deep plus some amount of Active Drop-In. The assumption in this case was that the ramp up of FD would take time and ADI could fill in the gap to enhance that would be under threat before FD could be built out in these locations. The worst case scenario was the base case of accelerated node splits.

WHAT COMES AFTER FIBER DEEP?

The first 12 to 18 months of Fiber Deep deployments will be built taking advantage of the higher density improvements coming from CCAP and new RF distribution modules in the HE and Hub. Older equipment will be cleared out to make room for optimally laid out racks supporting the new FD serving areas. Available rack space in the existing Hubs is not capable of meeting the 12:1 increase in nodes that complete conversion to FD will create.

To be successful, Fiber Deep must as quickly as possible change to a Distributed Access Architecture by changing out the analog optics with digital optics and adding an R-Phy or R-MacPhy module to the nodes. DAA will bring significant improvements in Hub density and provide the pathway for virtualization allowing even higher density results. Both R-Phy and R-MacPhy designs are in the lab today with early trials of both technologies being planned for later in the year.

A goal of the Fiber Deep N+0 design has been to strive for a last touch architecture. The introduction of DAA has always been known to be a necessary second phase of the Fiber Deep evolution and roll out plan. Nodes deployed prior to the changeover to DAA will be converted by removing the DS analog optical receiver and US digital return modules and inserting a DAA module which includes 10Gb digital optics. All of the Fiber Deep nodes qualified for deployment are designed to accommodate and provide for a simple change over to Distributed Architecture.

As stated earlier, DOCSIS 3.1 provides over 5 Gb of downstream data capacity but upstream usable data rates will remain asymmetrical for 1 Gb and higher tier offers. A new concept that has been demonstrated in the lab provides the capability to transmit DS and US data over the same frequency spectrum simultaneously. The benefit is an immediate expansion of the upstream bandwidth and the ability to carry multi-gigabit symmetrical US traffic.

This technology christened “Full Duplex DOCSIS” is an extension of principals used in DSL system designs. Full Duplex DOCSIS requires a passive coax network to operate effectively so N+0 access networks are ideal. Full Duplex DOCSIS was recently announced at the 2016 CableLabs Winter Conference.

CONCLUSIONS

Competitive threats from FTTH over-builders and the continuing growth of digital high speed data is driving the adoption and eventual deployment of DOCSIS 3.1 to provide expanded gigabit capacity and extend the usable life of the HFC DOCSIS network.

1.2 GHz Fiber Deep N+0 network design provides the DS and US bandwidth to maximize the benefits of DOCSIS 3.1. This implementation allows 10 Gb data capacity equivalent to 10G EPON FTTH downstream. The combination of N+0 FD plus D3.1 delivers the capacity to support gigabit per subscriber HSD tiers.

The incremental cost per subscriber of N+0 Fiber Deep is lower than most of the available alternatives such as traditional node splits, or FTTH. An N+0 network design provides a demarcation point at the node location that can be migrated to a distributed architecture or FTTH architecture or whatever path future technology innovation holds.

N+0 Fiber Deep deployments will accelerate and become standard practice as designers and construction crews gain experience and confidence in this new HFC topology.

The cable plant power grid is a fixed resource. For network operators the goal is to stay within the existing plant power capacity.

REFERENCES:

1. https://www.nngroup.com/articles/law-of-bandwidth/

2. https://www.sandvine.com/

3. Fierce Cable March 4, 2016 http://www.fiercecable.com/story/report-google-fiber-accounts-only-5100ths-1-us-pay-tv-market/2016-03-04?utm_medium=nl&utm_source=internal&mkt_tok=3RkMMJWWfF9wsRokuanIde%252FhmjTEU5z17%252BklW6Gyh4kz2EFye%252BLIHETpodcMScFiNLHYDBceEJhqyQJxPr3HJdQN18R7RhHnDg%253D%253D

4. Chart from Third Annual Report of Gig.U (August 2014)

5. http://www.bizjournals.com/atlanta/blog/atlantech/2015/04/source-at-t-ready-to-rollout-gigabit.html

6. Miguelez, Phil, Fred Slowik, Michael Aviles, "Lowering the Cost of Network Segmentation", SCTE Expo, Orlando, FL, October 17 – 19, 2012

7. www.cablelabs.org

8. Fierce Cable March 15, 2016 http://www.fiercecable.com/story/comcast-rolls-out-docsis-31-powered-advanced-trials-atlanta/2016-03-15?utm_medium=nl&utm_source=internal&mrkid=%257B%257Blead.Id%257D%257D&mkt_tok=3RkMMJWWfF9wsRokuK3Kce%252FhmjTEU5z17%252BklW6Gyh4kz2EFye%252BLIHETpodcMSsBiMLrYDBceEJhqyQJxPr3HJdQN18R7RhHnDg%253D%253D

Acronyms and Abbreviations:

ADI	Active Drop-In
BOM	Bill of Material
CAGR	Compound Annual Growth Rate
CPE	Consumer Premise Equipment
DAA	Distributed Access Architecture
DOCSIS	Data over Cable System Interface Specification
DS	Downstream
FD	Fiber Deep
FTTB	Fiber to the Basement
FTTH	Fiber to the Home
GaN	Gallium Nitride
GaAs	Gallium Arsenide
Gbps	Gigabits per second
GHz	Gigahertz
HD	High Definition
HFC	Hybrid Fiber Coax
HHP	Households Passed
HP	Homes Passed
HSD	High Speed Data
HVAC	Heating, Venilation, and Air Conditioning
IoT	Internet of Things
MoCA	Multimedia over Coax Alliance
MP2	MPEG2
NFV	Network Function Virtualization
OBI	Optical Beat Interference
OEM	Original Equipment Manufacturer
OHE	Optical Headend
OMI	Optical Modulation Index
OpEx	Operating Expense
OSP	Outside Plant
PS	Power Supply
QAM	Quadrature Amplitude Modulation
R-Phy	Remote Phy
RFoG	RF over Glass
SDN	Software Defined Network
SDV	Switched Digital Video
STB	Set top Box
US	Upstream
VOD	Video on Demand

HOW AUTOMATIC CONTENT ANALYSIS ENABLES THE ENTERNAIMENT EXPERIENCES OF THE FUTURE

Jan Neumann[1]
Comcast

Abstract

Metadata has always been one of the essential ingredients to help customers find something to watch. It varies from the TV Guide with high-level descriptions to the extended descriptions of recommended items.

These days we have machine learning driven content discovery experiences and clip-based navigational capabilities which rely on the availability of descriptive and semantically meaningful metadata.

Unfortunately, the availability of the metadata is limited due to the high cost of creating such metadata which require significant amount of human supervision. We show how automatic content analysis combines video, audio and text processing with machine learning algorithms to identify relevant moments or temporal segments and their descriptions without or with very limited human interaction.

INTRODUCTION

Nowadays entertainment is consumed in many different ways. Besides the traditional linear consumption of live TV, we regularly enjoy entertainment on our own schedule via DVR recordings or on-demand content services. Vast amount of content is provided to consumers, thus we want the applications we use to consume videos to guide and assist us in a personalized way while we are looking for something to watch that we like and that fits our current mood.

In addition, due to increased consumption of content on mobile devices, we also often do not want to consume whole shows anymore, but only the "best" parts, i.e. the segments that are (or our social networks deems to be) most relevant and interesting to us.

To enable such personalized entertainment experiences, we need to know more about the content shown than just the metadata at the asset level such as titles, credits, keywords. More information includes also semantic segments and moments of a program, who is on screen and when (Figure 1), the theme or topic of the segment, and how relevant a segment is. In addition, if the information is aligned with the timeline of the video, it allows for advanced navigation within and between assets and can be indexed to enable relevant search and recommendations within videos.

Figure 1 - Face Recognition in TV Shows

Moment and segment based metadata has many applications, such as chaptering of shows for improved navigation, enhanced search within captions and speech from audio so that search interfaces like the Xfinity Voice Remote allow users to search for specific phrases within movies or shows (e.g. "Life is like a box of chocolates").

We can also extract the current game state from the video at regular intervals, and use that information to align the video to an external

[1] The assistance of Hongcheng Wang and Jonghyun Choi during the preparation of this document is greatly appreciated!

meta-data sources such as sports play-by-play information as seen in Figure 2.

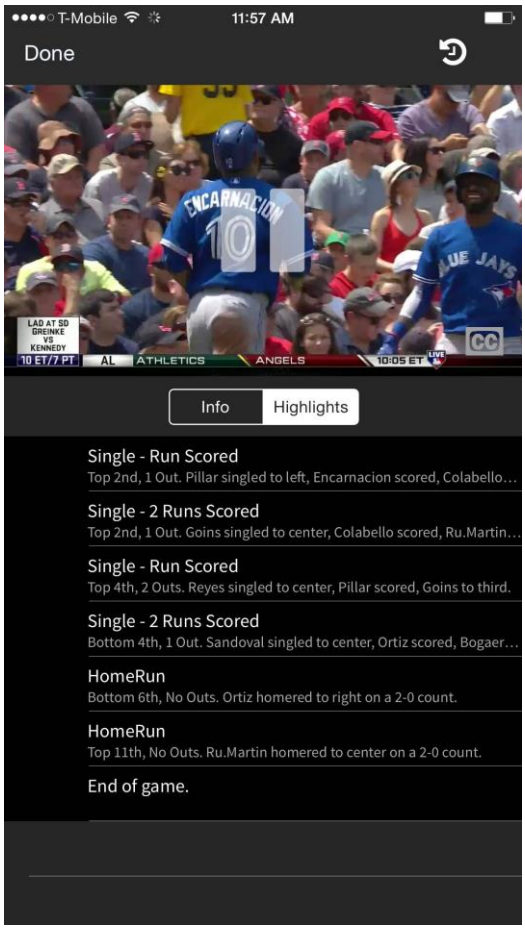


Figure 2 - Mobile App Allowing User to Select Playback of Individual Highlight Plays

Finally, automatic content analysis can be used to classify the activity and theme happening within segments of a show, it can be used to automatically classify content as suitable for various audiences by using adult content and language recognition [Ries], and object and content-based video compression can also impact the coding efficiency of the video pipeline positively.

The availability of such metadata, however, is limited due to the high cost of creating such metadata with human input or supervision. Metadata can be crowd-sourced as on sites such as YouTube or Facebook by analyzing the tags and comments from consumers themselves. This user generated data is usually unconstrained and unstructured though and it requires sophisticated algorithms to make good use of it.

We show how automatic content analysis combines video, audio and text processing with machine learning algorithms to identify relevant moments, segments and their descriptions without or with very limited human interaction. The generated data can then be used to provide consumers with a more interactive and personalized experience. Examples are more accurate program (segment) recommendations, in and between program navigation, and enhanced search capabilities including free-form queries using voice interfaces.

We also describe the details of how we use automatic content analysis techniques to create metadata for navigating from highlight to highlight within DVR recordings of sports broadcasts (Figure 2), organizing news segments according to their topics (Figure 5), and recognizing the persons currently visible in a scene similar to Amazon's X-ray feature, see Figure 1.

WHY DO WE NEED AUTOMATIC CONTENT ANALYSIS?

The availability of metadata is a fundamental building block to help customers interact with content and engage with it at a deeper level.

Unfortunately, the content metadata that is provided by the programmers and third-party metadata providers is often limited, i.e. it only contains high-level tags and descriptions that apply to assets as a whole. This is due to the amount of time it takes to create accurate descriptions and segmentations of content manually.

For live shows where the actual content is not known with sufficient details before the actual broadcast, any metadata would need to be generated on the fly if we are interested in more granular data than is available via the guest list or the genre information.

Companies like Google (YouTube), Amazon, Netflix, Apple, Pandora and Facebook rely heavily on accurate content descriptions to recommend related content so that their customers always have something else to consume while staying in the same eco-system.

Amazon and Netflix have dedicated groups of "movie watchers" that annotate their respective video on-demand catalogs, so that the content discovery algorithms can utilize this information and deliver relevant suggestions, accurate content descriptions and fine grained browse and search capabilities.

Pandora is famous for their collection of 1.3M annotated music tracks which they label as the "music genome" and is the backbone of their recommendation system.

All these companies use automatic content analysis techniques to speed up the generation of their metadata. They use automatic face detection and tracking technology to help their annotators identify the actors within scenes, and are developing techniques to automatically label the scene content in user-generated videos and images [Venugopalan]. Even Pandora who prides themselves on the superior quality of their metadata due to highly trained annotators, uses "machine listening" approaches to scale the annotation process.

As mentioned, another important application of content analysis is temporal metadata that labels contents at the moment or segment level, e.g. ad break detection, when someone appears on screen, when a touchdown is scored, etc.

Navigating through a video to manually annotate it in a frame accurate manner without algorithmic assistance is extremely laborious and time intensive since we can only navigate linearly through a video with standard video editing tools.

Automatic content analysis is very beneficial in this context because it can automatically identify potential annotation locations with associated confidences and present them to the annotator. This transforms the annotation task into a verification and selection task which can be performed much more efficiently by a human.

Finally, automatic tools can also be coupled with crowd-sourcing approaches like Amazon Mechanical Turk to scale up annotation resources on demand.

HOW DOES THE TECHNOLOGY WORK?

In this report we define automatic content analysis as a combination of computer vision, audio analysis, natural language processing and machine learning algorithms that allow the generation of semantic information for media assets. These analytics capabilities can be applied as a batch process to existing content catalogs (e.g. video on-demand content) and in real-time to linear broadcast streams. Many of the algorithms rely on machine learning techniques such as classification and clustering, and lately deep learning approaches.

Visual Analysis Techniques

Visual analysis is critical to detect and recognize text, objects and logos visible in the scene, segment programs, and classify segments. The standard approach to detect objects in an image and determine the concept category they belong to is usually done in stages.

During the training stage we apply image processing operators such as line, corner or histogram of gradients to create representations or features of examples for each category. Machine learning algorithms then use these representations to train a model that maps from features to categories using either classification or clustering approaches or a combination of the two.

During the detection phase we then process selected frames of the incoming video, apply the trained model to the resulting feature

representations and return the object categories with the highest score according to the model.

Implementations differ in the type of representations chosen, how the data is being stored (raw features or hashed representations), and how the representations are classified, e.g. different classification algorithms such as K-means clustering, support vector machines, neural networks, or random forests. More information about these approaches can be found in classic text books such as [Bishop].

One important aspect of object detection and recognition is to recognize the persons on screen in shows as shown in Figure 1. The first real-time computer vision face recognition system was developed by [Viola & Jones] and a variant of their approach is the basis for the face detection algorithms used in nearly every consumer camera these days. They combined the use of simple features that are easy to use with an efficient learning and classification algorithm that minimizes the number of computational steps necessary for the classification.

Face recognition for content analysis is different from this canonical approach because in contrast to the frontal views that we see in passport portraits we want to identify actors or people on screen who might only be partially visible, are looking away from the camera or are partially in the shadows. This scenario is often described as the "face recognition in the wild" problem [Huang]. At the same time we also have the advantage that for the celebrities that we want to detect in the videos, many images are already available to create accurate models.

Furthermore, since the same actors appear frequently in the same movie or across different episodes of a series, we can improve our results by labeling complete tracks of detected faces and use global constraints such as that the same actor cannot appear in the same scene simultaneously in two different image locations to find a globally optimal labeling solution.

A special variant of object recognition is optical character recognition, OCR in short, where the goal is to read the text that is present on the screen at a given time. This is helpful to establish the context of a show as in news shows, or it can be used to establish timing information in sports broadcasts by analyzing the game clock. Since most text in broadcast videos is artificially inserted as part of the post-processing, we can take advantage of such constraints as increased contrast, static horizontal text on a moving background and canonical size of the fonts. Furthermore, we can also utilize language models and dictionaries to constrain what sequences of characters are likely to appear next to each other (e.g. words, names).

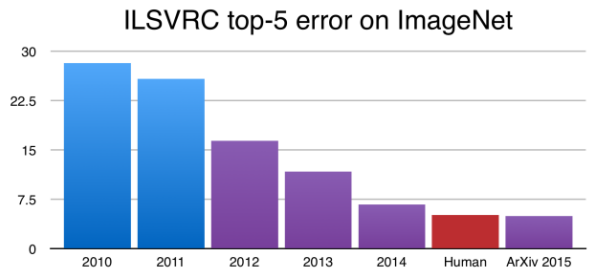


Figure 3 - Improvements in the Accuracy of Automatic Large-scale Object Recognition Algorithms over the Last Years [He]

Advances in deep learning [Bengio] have recently achieved close to human-level accuracy which allows for the automatic creation of metadata with very high accuracy [Krizhevsky].

Figure 3 shows the improvement of large scale object recognition performance for the ImageNet Large Scale Vision Recognition Competition which evaluates how accurately object class recognition algorithms can correctly identify 1 out of 1,000 categories for more than 1.2M images.

The large improvements starting in 2012 are due to the use of deep learning algorithms where the combination of large amounts of data with extensive computing capabilities have moved the field forward at a rapid pace. Recent deep learning based approaches perform face recognition beyond human capabilities [Huang].

Deep learning has not just applications for the classification of static images but can also

be applied to video sequences and is the gold standard for speech recognition. Recently, researchers demonstrated how deep learning models can even learn to "translate" a video to text for automatic content descriptions which can further accelerate the progress in creating metadata for programming [Venugopalan].

The main advantage of deep learning approaches is that they allow us to learn feature representations automatically based on the underlying structure inherent in the data instead of having to create hand-crafted feature extractors based on heuristics which are often suboptimal as shown for both visual and audio features in Figure 4.

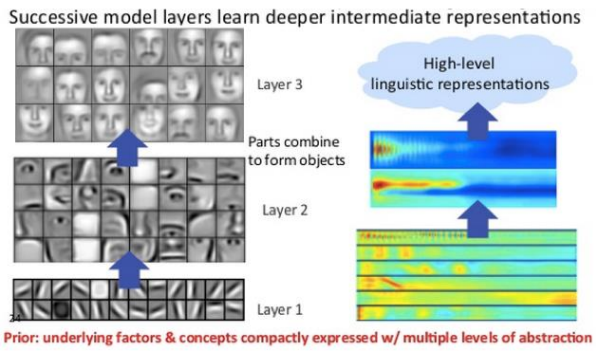


Figure 4 - Multi-layer Representation of Information in Deep Neural Networks [Lee]

Finger printing of content is another application of visual content analysis, for example to identify recurring video segments such as ads or news segments. For efficiency, video finger printing can use a hash representation for each frame, find likely matching candidates using an inverted search index, and then refine the matching using temporal constraints [Kanth].

Another important component of content analysis is the temporal segmentation of videos resulting in chaptering information, as well as the alignment of external meta-data such as sport highlight feeds or movie scripts with the video. We can detect all scene boundaries in a video and group them into larger segments to create chapters that are semantically distinctive.

An essential component of segmentation algorithms is shot-boundary detection which detects the presence of camera view discontinuities, e.g. the boundaries between sequences of contiguously recorded video frames..

We can efficiently find hard cuts between shots by thresholding the difference between the frame signatures of consecutive frames, such as quantized color histograms. We then choose a statistically motivated similarity metric function such as chi-square distance to compute the distances, and choose our thresholds appropriately depending on the content.

More complex transitions such as fades and wipes can also be detected by analyzing the spatial and temporal coherence of the mean and standard deviation of the image brightness values for each frame [Lienhart].

Once we have found shot boundaries, we can use various techniques to group shots into scenes, and scenes into program segments. Helpful cues here are the rate of shot changes, the presence of channel logos, the presence of black frame sequences, etc. These segments can be further classified into programs, commercials, credit roll, etc. using sequential classification approaches such as CRFs [Lafferty] or Recurrent Neural Networks (RNN) [Ng].

Audio, Speech and Closed Caption Analysis

Audio features are also very useful for program segmentation. Typical audio features include the volume, zero-crossing rate, short-time energy ratio, and spectrum flux. These features tend to be complementary to the visual features.

Mining and understanding what is being said in a program is especially helpful to describe what is happening on screen, specifically for talk shows where it is hard to make sense of what is happening on screen based on visual information alone [Schonfeld].

We can detect the topics that are being discussed at the moment on TV by reading all the text visible on the screen and combine it with the information mention in closed caption and speech transcripts. For news shows this can provide us with natural and accurate segment descriptors.

To analyze the text and closed caption we can first apply standard natural language processing techniques [Manning] to detect the parts of speech such as subject, verbs, and objects, before modeling the statistical properties such as co-occurrences between the words and sentences. This can then be used to find representative words and segment the video into coherent clips (Figure 5).

Figure 5 - Chaptering of a News Show

Finally, we could achieve a better and more robust result by a fusing visual, audio, speech and closed caption analysis results due to the complementary nature of different features [Tzoukermann].

EXAMPLE APPLICATIONS

Video distributors are selling a premium video product to its customers. To make this product appealing it is important that they offer an enhanced and more personalized entertainment experience for customers than competing products.

Having access to detailed metadata at the moment and segment level allows them to identify people, locations, activities, music, concepts in scenes, leading to stronger engagement of customers within their ecosystem instead of losing them to second screen devices.

Enhanced and real-time metadata will lead to more interactive and engaging experiences, for example a Trending Topic Guide could show which channel is currently showing a discussion of a topic that a customer is interested in. The viewer could then directly access the relevant programs without having to scan through the channel guide

The guide could also indicate breaking news by noticing that similar topics are being discussed at the same time on multiple channels and further incorporate information from Twitter and other social media sources to enhance the news experience.

The automatic creation of relevant short content from long programming would also be an appealing application, for example one could automatically create collections of the "best" SNL or Jimmy Fallon's late night show sketches from last night, the highlights for my favorite sports team or the news segments that match my interest. Chapter annotations allow for easy in-asset navigation of sports highlights and news segments that is more similar to how customers navigate clip based video platforms as seen in Figure 5.

To find these chapter markers, we first use visual analytics to temporally segment the videos into shots and scenes, then we extract topics and summary sentences from the closed caption co-occurrence statistics to group the segments into clusters, and finally use an exemplar based clustering approach to relate clips across broadcasts.

Sports highlights can be automatically aligned with a video by using OCR to automatically detect read the current game clock in the video. Once the game clock is analyzed, we can align the detected clock times with an external game highlight feed to create a temporal annotation of each highlight moment.

More detailed descriptions of the media assets lead to improved search and personalized recommendations that goes beyond simple keywords but instead allow one

to search via semantic concepts which is important as voice interaction with the TV becomes the norm rather than the exception, e.g. Xfinity X1 Voice Remote, Apple TV, Amazon Fire TV.

Finally, automatic content analysis could also create visual descriptions as an assistive technology and thereby noticeably increase the coverage of these solutions.

SUMMARY

In this article we described how the availability of detailed content metadata enables a range of new and exciting user experiences. Due to the time and labor intensive nature of creating such metadata manually it is essential that human annotation is augmented with automatic content analysis techniques such as described in this paper. Only then can the metadata creation process scale to handle large number of assets and near real-time constraints.

Finally, we gave an overview about common approaches to analyze the visual, audio and language content of an asset and how these approaches are combined in the industry today to enable the entertainment experiences of the future.

REFERENCES

[Bengio] Y. Bengio. Learning deep architectures for AI, Foundations and Trends in Machine Learning, 2(1):1-127, 2009.

[Bishop] Christopher Bishop. Pattern Recognition and Machine Learning, Springer, 2007

[He] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, ICCV 2015

[Huang] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts, Amherst, Technical Report 07-49, 2007.

[Kanth] K. V. Ravi Kanth, Divyakant Agrawal, Amr El Abbadi and Ambuj Singh. Dimensionality reduction for similarity searching in dynamic databases. SIGMOD 1998

[Krizhevsky] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012

[Lafferty] John Lafferty, Andrew McCallum and Fernando Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, ICML 2001

[Lee] Honglak Lee, Roger Grosse, Rajesh Ranganath and Andrew Y. Ng. Convolutional Deep Belief Networks for scalable Unsupervised Learning of Hierarchical Representations, ICML 2009

[Lienhart] Rainer Lienhart. Reliable Transition Detection in Videos: A Survey and Practitioner's Guide. International Journal of Image and Graphics (IJIG), Vol. 1, No. 3, pp. 469-486, 2001.

[Manning] Christopher Manning and Hinrich Schuetze, Foundations of Statistical Natural Language Processing, May 1999, MIT Press

[Ng] Joe Yue-Hei Ng, Matthew Hausknecht, Sudheendra Vijayanarasimhan, Oriol Vinyals, Rajat Monga, George Toderici, Beyond Short Snippets: Deep Networks for Video Classification, CVPR 2015

[Ries] Christian X. Ries, Rainer Lienhart. A Survey on Visual Adult Image Recognition. Multimedia Tools and Applications, Springer, 2012

[Schonfeld] Dan Schonfeld, Caifeng Shan, Dacheng Tao, Liang Wang. Video Search and Mining, Springer, 2010

[Tzoukermann] Evelyne Tzoukermann, Geetu Ambwani, Amit Bagga, Leslie Chipman, Tony Davis, Ryan Farrell, David Houghton, Oliver Jojic, Jan Neumann, Robert Rubinoff, Bageshree Shevade, and Hongzhong Zhou. Semantic Multimedia Extraction using Audio and Video. Multimedia Information Extraction. Wiley-IEEE Computer Society Press, 2012.

[Venugopalan] Subhashini Venugopalan, M, Rohrbach, Jeff Donahue, Raymond J. Mooney, Trevor Darrell, and Kate Saenko, Sequence to Sequence – Video to Text, Proceedings of the 2015 International Conference on Computer Vision (ICCV), 2015.

[Viola] Paul Viola and Michael Jones. Robust Real-time Object Detection, IJCV 2001

IMPROVING CUSTOMER EXPERIENCE THROUGH COOPERATIVE IN-HOME CACHING AND PRE-POSITIONING (CIHCP)

John Jason Brzozowski
Jan van Doorn
Comcast Cable

Abstract

While discussions of content delivery often center on backbone architecture, a substantial portion of the cost of delivering content from Internet peering points to subscribers is spent on Access Networks.. Content Delivery Networks (CDNs) have done a good job optimizing backbone traffic, but to date, there has been less innovation around optimizing traffic flows at level of Metro and Access Networks, which represents a major opportunity.

ISP networks, including Metro and Access Networks, are scaled for peak usage, and are underutilized at off-peak times. In this way, user demand and usage is to some degree, predictable.

In this paper we explore a potential method of cooperative in-home caching and pre-positioning. This approach seeks to leverage the predictable nature of user demand to reduce the highest peaks, by pre-caching content in the home during the traffic valleys. In this paper we explore only the potential technical function of a cooperative in-home caching and prepositioning approach. Related issues such as cache management, telemetry, and application interfaces would be the subject of future work on this topic.

INTRODUCTION

Traffic patterns in ISP IP networks are very predictable, with high peaks at primetime and low valleys late at night / early morning. See Figure 1 for a typical example of traffic in an ISP's network segment. *In most ISP networks it is not uncommon for traffic peaks to represent three times their corresponding valleys.* Obviously, the network has to be scaled to at least the highest peak in these normal traffic patterns. Specific events, like a software or a game release, can spike these peaks even higher: 1.5x the normal daily peak is not unheard of, so for the ISP to maintain quality of service in those cases it would need to scale the network much higher than even these daily peaks. It follows that there is unused capacity in the ISP network for a very large percentage of the day.

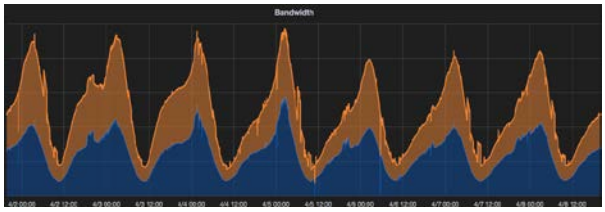


Figure 1: Typical traffic pattern

CDNs do important work in reducing backbone or transit traffic but that is typically the smallest portion of the network cost and they are often not in the optimal position in the network. One of the goals and ongoing challenges has been to move content delivery to be as close to the end user as possible. In many cases the deepest (meaning closest to the customer) caches are outside of the ISP network. Recently, many ISPs have started deploying CDN caches deeper in their networks. Even when deployed in the ISP network, the caches require extensive Metro resources and are still "above" the CMTS driving most of the costs of delivering the bytes to the customer (i.e. below the CMTS [1].)

Individual bandwidth usage is, to a certain degree, predictable as well. For IP video, a

viewer who scheduled a DVR recording for a certain show is most likely going to watch that show. After seeing episode 1 through 4 of a series, it is likely that episode 5 is next. Some OTT video providers have a "My List" or "Watch List" feature that the viewer can use to queue things to watch later – the videos queued here are likely to be watched by this person. For software releases, a person who downloaded MacOS 10.11.3 is likely going to download version 10.11.4 of that same OS shortly after it is released. Similarly, a household that has several devices of the same make and model, i.e. iPhone6, is likely to download the same update to each and every device. Other possible use cases of cooperative in-home caching and pre-positioning could be cloud-based backups that can be scheduled for upload overnight instead of when children are actively gaming online and pre-positioning of popular game release, or updates, gradually in advance of the official release date.

Individuals and families also often have unused storage and computing resources in their home. The household PC is often sitting idle, as is the game console's storage. Those, along with retail devices, like Network Attached Storage (NAS) appliances, represent opportunities to leverage large amounts of idle in-home storage.

In this paper we try to make use of the above observations by exploring on-demand and pre-caching of content at low traffic times in an in-home cache; for the purposes of this paper it is assumed that the cache is customer-owned and maintained equipment, which cooperates with either the CDN or the client. However, the technological concepts detailed here apply equally to operator-provided in-home cache appliances.

In all cases, these models are premised on a cooperative, opt-in approach that would involve the participation of customers, content providers and ISPs.

A comprehensive in-home caching strategy including content pre-positioning is mutually beneficial to multiple stakeholders across the Internet ecosystem. The primary beneficiary is the consumer, who will realize the benefits of massively improved performance, efficiency, and a heightened customer experience. Content owners who elect to participate can benefit by way of increased delivery efficiency by transmitting fewer copies of the same content to a single household, and off-peak content delivery infrastructure, which manifests lower costs and improved customer satisfaction. Broadband network operators with the ability to deliver content off-peak can especially benefit, by way of overall network performance and capacity management. Tests have illustrated that a basic in-home caching solution minimally yields a 300% increase in download performance when in- home caching is utilized. The evolution of in-home caching to include off-peak content pre-positioning introduces the additional benefit of delivering content when network utilization is lower, which helps to utilize valuable network resources when demand is at its lowest.

The rest of this paper is organized as follows: First we examine the different types of HTTP proxies in use today, and how CDNs utilize them. Then we explore the options available to insert the in-home proxy into the client's request path. Next, we describe a prototype implementation that was built to help think through the problems and options. We conclude with final thoughts and suggested future work.

HTTP/1.1 AND PROXY/CACHES

This section is an abbreviated version of a section of the Traffic Control documentation written by one of the authors of this paper [2]. The main function of a CDN is to proxy requests from clients to origin servers and

cache the results. To proxy, in the CDN context is to obtain content using HTTP from an origin server on behalf of a client. To cache is to store the results so they can be reused when other clients are requesting the same content. There are three types of proxies in use on the Internet today: The "reverse proxy," the "forward proxy," and the "transparent proxy." They are described below.

Reverse Proxy

A reverse proxy acts on behalf of the origin server. The client is mostly unaware it is communicating with a proxy and not the actual origin. The typical use case for the reverse proxy is a CDN - to the end user, a CDN appears as a reverse proxy because it retrieves content from the origin server, acting on behalf of that origin server. The client requests a URL that has a hostname which resolves to the reverse proxy's IP address and, in compliance with the HTTP 1.1 [3] specification, the client sends a `Host:` header to the reverse proxy that matches the hostname in the URL. The proxy looks up this hostname in a list of mappings to find the origin hostname; if the hostname of the Host header is not found in the list, the proxy will send an error (`404 Not Found`) to the client. If the supplied hostname is found in this list of mappings, the proxy checks the cache, and when the content is not already present, it connects to the origin the requested `Host:` header maps to and requests the path of the original URL, providing the origin hostname in the `Host:` header. The proxy then stores the content for this URL in cache and serves the contents to the client. When there are subsequent requests for the same URL, a caching proxy serves the content out of cache thereby reducing latency and network traffic.

For a reverse proxy to be injected into the path of delivery the content owner changes the delivery URL to not point to the origin server, but to the CDN, and the CDN is configured to map this delivery URL back to the origin.

Forward Proxy

A forward proxy acts on behalf of the client. The origin server is mostly unaware of the proxy; the client requests the proxy to retrieve content from a particular origin server. The typical use case for the forward proxy is an organization trying to save on Internet resources, either bandwidth or address space. In a forward proxy scenario, the client is explicitly configured to use the proxy's IP address and port as a forward proxy. The client always connects to the forward proxy for content. The content provider does not have to change the URL the client obtains, and is unaware of the proxy in the middle.

For a forward proxy to be injected into the delivery path, the client is configured with the forward proxy information, and often, for what domains to use it.

Transparent Proxy

In the Transparent proxy case, neither the content owner nor the clients are aware that the proxy is in the path. The traffic is intercepted by putting the cache directly in the path as a router, or by a router with a special configuration to intercept traffic and redirect it to the proxy. The typical use case of the transparent proxy is a school or business that wants to force all Internet traffic through the proxy to be able to block some sites (parental controls) and to share Internet resources.

For a transparent proxy to be injected into the path, a network administrator in the layer 3 network path needs to configure the interception.

Even though caching and proxying are distinctly different functions, in the rest of this paper we will refer to a *cache* as the

combination of these two functions in one device or piece of software running on a multi purpose computer.

CDN INTEGRATION

An in-home cache needs to be inserted into the CDN delivery path (with the cooperation of the CDN of course). In this section, we describe the options that were evaluated. Before we do that, we first look at how a CDN normally gets inserted into the request path. The client will do a DNS address lookup of the hostname of the content URL, and in most cases the CDN is authoritative for the domain-name used in the content URL. The CDN can do one of 2 things. For short-lived sessions, the CDN will return an IP address of an edge cache that is closest to the local DNS (LDNS) server that handles the DNS resolution for it. This is often referred to as *DNS content routing*. DNS content routing is low-latency, but has some disadvantages: the CDN doesn't actually know the client's IP address, only the IP of the LDNS server, and the client could have a remote LDNS configured, such as Google's public DNS [4], moreover, the CDN only knows the hostname of the content URL, not the path. To address this, some CDNs will send the IP address of a content router in response to the DNS query of the hostname in the content URL. The client will now connect to this content router and this router will send a HTTP 302 redirect to the client based on the actual client IP address *and* the requested content – a much more informed decision that enables better localization and content affinity. This is often referred to as *HTTP content routing.* DNS content routing is usually preferred for objects and images in web pages, since there are usually many of these in a web page, and the added latency of the 302 redirect is too costly. For video and game or software downloads, the added latency is not that much overhead compared to the expected download time, and the added precision is worth the extra round trip to the content router.

Next, we look at the options of inserting an in-home cache into the client request flow.

1) Domain Name System

In this case, the CDN authoritative DNS server will respond with a CNAME to a `.local` address that the in-home cache advertises using mDNS, and has a reverse proxy rule for. The biggest advantage of this approach is the simplicity, but that is also it's drawback – there is no way the CDN can determine what client is getting the DNS response (only the DNS caching server or LDNS server knows), so the CDN will have to send this response to everyone asking for the address of this name, regardless of presences of the local proxy. HTTP only allows for A/AAAA records in resolving hostnames, and while you can send multiple IP addresses in the answer to an A/AAAA query, these answer records don't allow for any prioritization. This means we can't send a DNS response to instructs the client to first try the in-house cache, and if that fails, then try the CDN cache higher in the network.

2) HTTP 302 Redirect

With HTTP 302 redirect, we can target individual clients, and we can send the 302 to a URL that "points to" the in-house cache. The advantage here is that we now can only redirect to the local cache if it is running and "registered" with the CDN, and only for content that is known to be cached, since the registration could tell the CDN content router what content is pre-cached. The disadvantage is that the CDN now needs to keep track of all the clients that have an active caching proxy in their home, as we as the content that is pre-cached. At scale, this could be a challenge.

3) Transparent Proxy

Transparent caching could be an interesting option, but we are looking to do this in cooperation with the content owner, and, with the advent of HTTPS in standards [5, 6] and practice [7] it seems that transparent proxies

are becoming more and more difficult (if not impossible) to implement without major security compromises.

4) DNS Intercept

In this scenario, the DNS request for the CDN domain is intercepted by software on the in-house cache using ARP (Address Resolution Protocol) spoofing, special software on the router, or even on the LDNS server itself. For domains that have content pre-cached, a CNAME response is sent to the client that points to a .local name the cache is advertising using mDNS. Another variant of this method is the Smart DNS proxy method, where the client is configured with LDNS servers that send the proxy's IP address in response to a DNS query for certain origin / CDN domains. Finally, interaction between the local in-home cache and the local DNS server can also be leveraged to update the local DNS with the IPv6 and/or IPv4 address of the in-home cache. Often the home gateway or wireless gateway in the house is configured as the LDNS server for all downstream clients, and it simply forwards all DNS requests to the ISP's LDNS servers. This local LDNS server in the gateway could be modified to be authoritative for a domain that is not `.local` (and thus doesn't have the mDNS quirks and drawbacks), for instance `cache.lan.` When the in-home cache pre-caches content for a domain, it updates the gateway and the CDN serving this domain with that information, and this update would ensure that the LDNS server in the gateway is able to resolve the CNAME response that was provided by the authoritative CDN DNS server locally.

With the growing and pervasive deployment of Domain Name System Security Extensions (DNSSEC) it is important to note that some DNS approaches outlined here may no longer be viable or the use of the same may become increasingly complex. Cooperative in-home caching and pre-positioning that properly integrates every player in the content ecosystem could conceivably incorporate the use and support of DNSSEC including any local DNS infrastructure in the home. However, it is important to note that the use of DNSSEC may not be strictly required by local DNS. The detailed analysis of DNSSEC and content caching and pre-positioning are out of scope for this document.

5) HTTP Intercept / 302

A cache can be inserted by "snooping" the connection that carries the HTTP GET request, then returning a 302 to the cache on the connection without the knowledge of the destination server of the HTTP GET request. This is what Laguna [8] does. This has the same restrictions as method 3.

6) None – Client configuration and forward proxy

In this and the next option, we don't cooperate with the CDN, but instead with the browser. The browser is configured to use the in-home proxy/cache as a forward proxy for certain domains, or for all domains. The drawback of this system is the client configuration, and the fact that now the proxy is in the client path for all requests in the configured domain.

7) None - Proxy Auto Configuration (PAC)

This option is very much like 6, but here, we use the in-home proxy/cache as the PAC server, and it can insert domains to dynamically use the in-home cache, as the content is pre-positioned. Since most PAC systems do not have an auto update, and the proxy configuration only gets loaded on client start, this option does not seem to add much over option 6.

<u>Notes about mDNS</u>

The multicast Domain Name System, (commonly referred to as mDNS) [9] is a system for advertising and resolving host names to IP addresses (as well as services locally) in small networks that don't have a local name server. It is most often used with

zero configuration services like DNS service discovery (DNS-SD) [10] or Apple's Bonjour. In this case we use the simplest form of mDNS. We only use it to do hostname advertisement, and to advertise A or AAAA records, with no associated services.

Using mDNS carries some drawbacks. During testing, we found that some applications assume it is used in conjunction with DNS-SD, and won't work when you just use it for hostname resolution. Specifically, we were not able to have a CNAME in the unicast DNS system point to a `.local` mDNS announced record, and have browsers "follow" that CAME. This severely hampers some of the DNS based CDN integration options.

In an IPv6 capable system the mDNS announcements in the CDN integration options could be replaced with real DNS AAAA records that point to the IPv6 address of the in-home cache. This has the advantage of not requiring mDNS, but adds some scaling and other requirements to the DNS system overall, specifically for authoritative DNS server operators which also tend to be content owners and/or CDN providers.

THE PROTOTYPE SYSTEM

The three options explored in the prototype system outlined in this paper are option 2, cooperation with the CDN, a combination of option 2 and 5, where we 302 redirect to a unicast announced DNS with gateway modification, and option 6, cooperation with the content consumer.

The sequence of events when using the system in CDN cooperation mode is depicted in figure 2. Steps 1 through 5 are only done in low traffic times, while the rest of the steps can be done at peak or low traffic times.

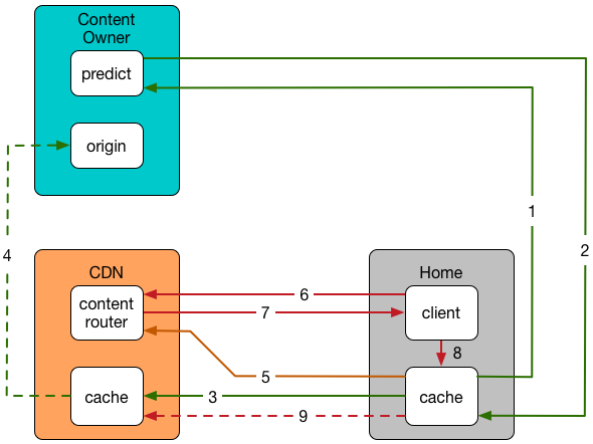


Figure 2: in home pre-caching system with CDN/content owner cooperation

1) The in-home cache registers with the content owner(s).
2) The prediction engine at the content owner instructs the in-home cache to pre-cache an (or set of) object(s).
3) The in-home cache fetches the content from the CDN and caches it.
4) On miss, the cache fetches the content from the origin server.
5) The in-home cache registers the pre-cached content with the CDN's content router. In an IPv4 system, the registration will likely arrive at the CDN's content router using the same IPv4 NAT address that the client will use. In an IPv6 system, it will need to tell the CDN content router the delegated prefix. At this time, the in-home cache also adds an mDNS announcement of the content URLs hostname with a .local appended (<original FQDN>.local).
6) The client requests the content from the CDN
7) The CDN content router, knowing that this content is pre-cached at the in-home cache, redirects the client to <original FQDN>.local
8) The client fetches the content from the in-home cache
9) On miss, the in-home cache fetches the content from the CDN cache.

Another similar option is the use of unicast DNS. This approach shares many similarities with that of the mDNS-based solution. The authoritative DNS server for the origin content must still be able to detect and reply accordingly to end user and in-home cache requests separately. An example sequence of events in this option is show in figure 3:

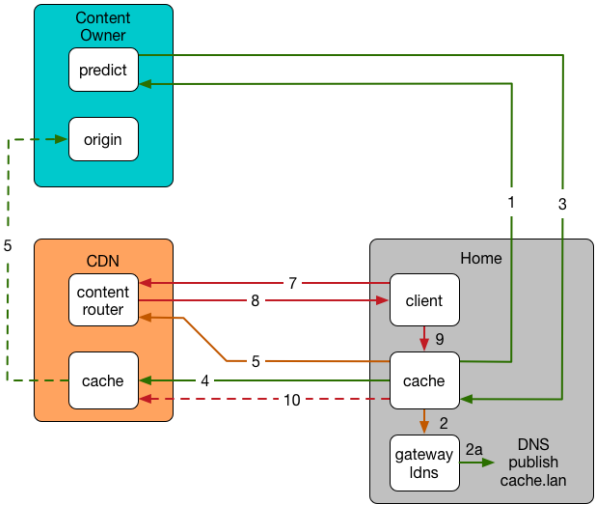


Figure 3: option 2 and 5 combined

1. The in-home cache registers with the content owner(s).
2. The in-home cache must publish itself to the local, unicast DNS server in the customer's premises. This could very well be an API or dynamic update to the DNS server running within the Wi-Fi router or a standalone DNS server. The in-home cache would update DNS with an FQDN that is resolvable locally (i.e. `cache.lan.`) and that would be used by the authoritative DNS server as a CNAME to redirect content requests to.
3. The prediction engine at the content owner instructs the in-home cache to pre-cache (a set of) objects. This is the same whether unicast DNS or mDNS is used.
4. The in-home cache fetches the content from the CDN and caches it.
5. On miss, the cache fetches the content from the origin server.
6. As part of the registration the in-home cache can for example be added to an authoritative view that will allow it to properly resolve DNS for the content origins, instead of the resolution that will redirect content requests to the local, in-home cache by default. In an IPv4 system, the registration will likely arrive at the CDN's content router using the same NAT address that the client will use. In an IPv6 system, it will need to tell the CDN content router the delegated prefix. At this time, the in-home cache is expected to be resolvable using a local FQDN, i.e. `cache.lan.`
7. The client requests the content from the CDN
8. The CDN content router, knowing that this content is pre-cached at the in-home cache, 302 redirects the client to `cache.lan.` with the same path. Conversely, the CDN's authoritative DNS servers may use CNAME resource records to re-direct the client content request to the local, in-home cache.
9. The client fetches the content from the in-home cache
10. On miss, the in-home cache fetches the content from the CDN cache.

Separately, option 6 has also been implemented and assessed. This is the option where the content consumer is cooperating with the in-home caching system.

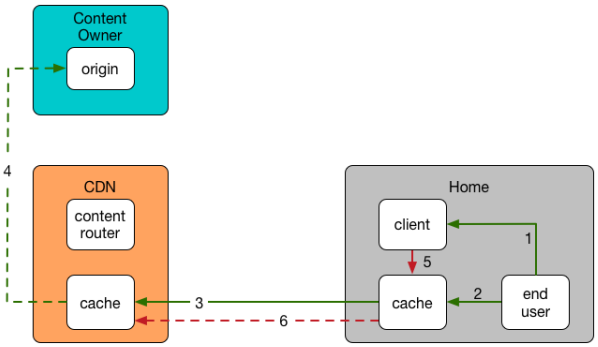


Figure 4: in home pre-caching with content consumer cooperation

1) The end user configures the client to use the in-home cache as a proxy for certain domains.
2) The end-user tells the in-home cache what to pre-cache, or the in-home caches uses the "watch list" for the content to pre-cache.
3) The in-home cache pre-caches the content.
4) On miss the CDN cache fetches the content from the origin.
5) The client requests the content with the original CDN hostname, but the client will use the configured forward proxy, and thus go through the in-home cache.
6) On miss, the in-home cache fetches the content from the CDN cache.

Candidate requirements for the end-to-end system are listed below:

Secure
Security and privacy are of paramount importance for a system like this. This applies equally to content consumers and creators.

Open
The system must be based on open standards, and must completely comply with the same. Ideally key elements of the solution itself may be open sourced and/or include open APIs (application programming interfaces) to enable wide spread adoption and seamless integration. Some of the CDN integration options that intercept traffic are disqualified by this requirement, since they create traffic patterns that violate the specification.

Non-intrusive
The system must be seamless to the end-user once installed, and even more so if failures were to occur; there must be no impact to the customer experience. The last thing we want is for a failure in this system to prevent viewers from accessing content they would've otherwise been able to access.

Scalable
A CDN often serves many millions of users simultaneously. A cooperative in-home caching and pre-positioning system needs to scale to the same numbers.

Content-Aware
Given the fact that a large proportion of the Internet traffic these days is video, the system should understand popular ABR video formats. It should understand, for instance, the most-used manifest file formats, so that the in-home cache can be instructed to cache a video asset just by giving it the manifest URL.

<u>The in-home cache requirements:</u>

Requirements for the in-house cache include:

Portability
The software should run on as many operating systems and environments as possible, with as few dependencies as possible. While a hardware appliance may be utilized, the in-home caching system should be able to run on consumer-provided hardware, such as a personal computer or Network Attached Storage (NAS) with sufficient capacity.

Lightweight
The software should be able to run on small systems as a background daemon with little or

no noticeable impact to performance of other tasks.

User-Friendly
The software should be easy to install and configure. Any end-user should be able to install and configure it in minutes, and no additional management or maintenance should be necessary.

Given these requirements we decided to implement the in-home cache using the Go language. The Go language is supported on many platforms, and has the big advantage that the binaries produced are self-contained, and not dependent on libraries or other external files, and don't require installation of any packages. See github.com [12] for the source code for the prototype in-home cache and CDN integration implementation cited in this paper.

Predicting what to pre-cache

Predicting what to cache is easy when we cooperate with the consumer and/or content owners – the user will simply tell the system what to cache either by directly interfacing with the in-home cache, or by keeping a watch list up to date. Consumers, in cooperation, with the content owners and broadband operators can further enhance the in-home caching experience. People consuming content may optionally share additional information with the prediction engine that can be used to more accurately predict future video selections.

In the case that the content owner is providing the pre-cache instructions, the content owner can use its own usage data to create the pre-cache rules. How to do this in the most optimal way is out of the scope of this paper.

CONCLUSION

Over the years, the way people use the Internet has rapidly evolved -- from web pages and email, to rich applications, video, and interactive content. This transformation has dramatically changed how players across the Internet ecosystem engineer their infrastructure, products, and services. Modern applications and rich content has pushed the limits of modern software and networking technologies resulting in new, innovative approaches to support the delivery of the same. Innovation and engineering in these areas has been largely focused on increasing efficiency, lowering costs, and ultimately improving the customer experience.

The approach documented in this paper is intended to take these innovations to new levels. A cornerstone principle of this work has been to assess, technically, how close content can be positioned to the consumer, and what benefits can be expected from such close proximity. A material and intended byproduct of this work has been to develop a proof-of-concept, intended to derive data that can be used to quantify the performance and efficiency of in-home caching and pre-positioning. As cited in this document, the performance improvements are non-trivial, as are the benefits for the broadband network -- and ultimately the customer experience.

In summary, the principles outlined in this document, when done cooperatively, represent distinct, measurable benefits to every stakeholder across the content ecosystem. Cooperative in-home caching and pre-positioning effectively represents a win-win-win scenario for content consumers, owners, and for those who facilitate delivery of the same including ISPs.

REFERENCES

[1] http://traffic-control-cdn.net/
[2] [RFC2616] Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L.,

Leach, P., and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1", RFC 2616, DOI 10.17487/RFC2616, June 1999, http://www.rfc-editor.org/info/rfc2616

[3] https://developers.google.com/speed/public-dns/

[4] Dorwin D., Smith J., Watson M., Bateman A. “Encrypted Media Extensions”, W3C Working Draft, March 2016, https://www.w3.org/TR/encrypted-media/

[5] [RFC7540] Belshe, M., Peon, R., and M. Thomson, Ed., "Hypertext Transfer Protocol Version 2 (HTTP/2)", RFC 7540, DOI 10.17487/RFC7540, May 2015, http://www.rfc-editor.org/info/rfc7540

[6] Dan Goodin, “It wasn’t easy, but Netflix will soon use HTTPS to secure video streams”, April 2016, http://arstechnica.com/security/2015/04/it-wasnt-easy-but-netflix-will-soon-use-https-to-secure-video-streams/

[7] https://github.com/concurrentlabs/laguna

[8] [RFC6762] Cheshire, S. and M. Krochmal, "Multicast DNS", RFC 6762, DOI 10.17487/RFC6762, February 2013, http://www.rfc-editor.org/info/rfc6762

[9] [RFC6763] Cheshire, S. and M. Krochmal, "DNS-Based Service Discovery", RFC 6763, DOI 10.17487/RFC6763, February 2013, http://www.rfc-editor.org/info/rfc6763

[10] https://golang.org/

Leakage Detection Using Test Signal Phase with GPS

Authors: Thomas H Williams and Colin Justis
Cable Television Laboratories, Inc.

Abstract

This paper describes a leakage detection system that measures received phase and magnitude of a stable continuous wave (CW) test signal versus distance. The received complex signal samples are used to mathematically construct a synthetic phased array (SPA), which points to (vectors) the leakage signal source. The latitude and longitude of a leak source are located where the vectors intersect. The use of a CW test signal with synthetic phased array provides superior receiver sensitivity. Having both signal strength and a detection distance determines if the leak is over Federal Communications Commission (FCC) limits. Practical test considerations and samples of field tests are provided.

BACKGROUND

The FCC has mandated for many years that cable operators routinely test their plant to verify that signal leakage field strengths are below prescribed thresholds. Historically, testing has been done in or near the 108 MHz to 137 MHz very high frequency (VHF) aviation band. In 2012 Hranac and Tresness presented a paper (1) that showed that leakage in the ultra-high frequency (UHF) band is not tightly correlated to leakage in the VHF band. Testing in only the VHF band does not guarantee that UHF leakage is not taking place. The conclusion that can be drawn is that to avoid interference with services like long term evolution (LTE) tower sites and mobile devices, testing should be done in both bands.

THE INVENTION

CableLabs engineers came up with an idea to assist technicians with their routine testing for leakage by pinpointing the source's location. Most test gear in the past used signal strength, or magnitude, and the signal's phase angle was ignored. As a test antenna was moved closer to the leak, the field strength was expected to get stronger.

Our initial experiment was to simulate a leak by radiating a stable CW signal and receiving it with a complex In-phase and Quadrature (I-Q) demodulator using a matching stable CW signal for demodulation. This is illustrated in Figure 1. Inserting a single CW signal is relatively easy for an operator, because only a small narrow vacant bandwidth is required. If a CW signal is being used for AGC (automatic gain control), that signal can also be used for leakage detection. To provide the necessary local oscillator stability, 10 MHz rubidium frequency sources were used on both transmit and receive sides. A global positioning system (GPS) disciplined clock could have alternately been used, but the rubidium sources were chosen to allow indoor testing and to simplify headend installs. As a test antenna was moved closer to a leakage source, a dot (single constellation point) on the test equipment's

I-Q polar plot could be observed to move in a counter-clockwise circle, and in a clockwise circle as the antenna got further from the leak. If the received signal got stronger, a radius associated with the dot increased. The rotation of the dot on the I-Q plot is essentially a manifestation of Doppler shift caused by a decrease of phase angle between antennas.

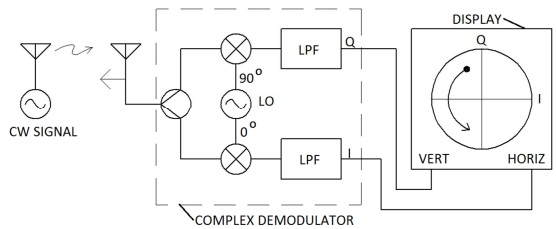


Figure 1. Constellation point rotates counter-clockwise when the test antenna approaches the leakage source, and clockwise when the test antenna is moved away from the leak.

It is mathematically possible to determine a bearing angle between the leak and the test antenna's drive path by measuring the velocity of the test antenna, and comparing it with an expected Doppler shift. For example, if the test signal was 900 MHz, its wavelength would be 0.3 meters ($\lambda=f/c$). If the test antenna was traveling directly towards the leak at 10 meters per second (m/sec), a Doppler shift of +33.33 Hz would be expected. If a lower frequency was measured, such as 20 Hz, one could compute that the test antenna was not traveling directly towards the antenna, but at an angle of ACOS(20/33.33) = 53.13 degrees. The leak source lies somewhere on a 3-dimensional cone, with the test antenna at the cone's apex, and the angle is half of the cone's apex angle. See Figure 2. Since a cable operator generally knows what side of the street the plant is on, where the cone intersects the cable plant is the leak's source (2).

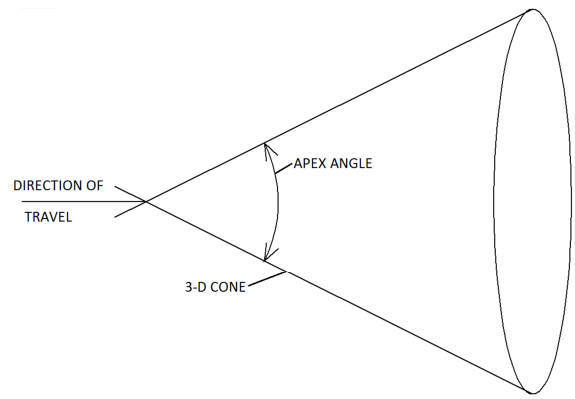


Figure 2. The source of a leakage signal lies somewhere along a three dimensional cone, where the test antenna is at the cone's apex.

This test worked well with a single leakage source, but the dot was observed to move erratically when there were multiple leaks that added in and out of phase as the test antenna moved. A solution was to capture the complex I-Q samples as the test antenna was moved and insert the values into an inverse fast Fourier transform (IFFT) after windowing. The net result was the creation of a synthetic phased array, where each I-Q voltage sample formed another virtual element. Typical element separation was 10-15 cm. Because the Fourier transform is a linear transform, it became possible to observe multiple leaks simultaneously. A magnitude plot of the transformed coefficients makes a Doppler plot. An example Doppler plot is shown on Figure 3. Each leak's bearing angle can be read, as well as the strength of each leakage signal. As one drives by a leak, it appears on

the right (positive frequency), increases in strength as the test antenna approaches, passes through zero hertz (Hz) (center) as the test antenna passes, and then moves to the left side (negative frequency) and then disappears.

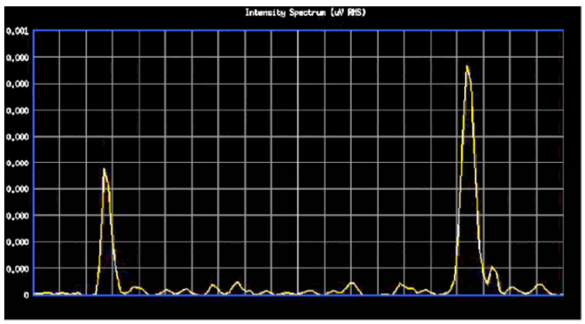


Figure 3. A Doppler plot formed by inserting I-Q values into an inverse discrete Fourier transform. 0 Hz is in the middle. Note that there is a leak (on right) ahead and a leak (on left) further behind.

By plotting bearing angles (vectors) from different test antenna positions, the intersections of the vectors locate the leak's source location. In practice, objects can obstruct the radiated leakage test signal. For example, a leak may originate behind a house and cannot be observed while the test antenna is in front of the house, but can be measured when approaching and again when driving away from the house.

Figure 4 shows a Doppler frequency vs. distance plot illustrating a rapid frequency shift for a leak near the drive path, and a more gradual frequency shift for a leak that is distant from the drive path. Figure 5 is an integral of Figure 4 and shows phase angle vs. distance for a travel path 'A' that comes close to the leak and phase angle vs. distance for a travel path 'B' that does not come as close to the leak.

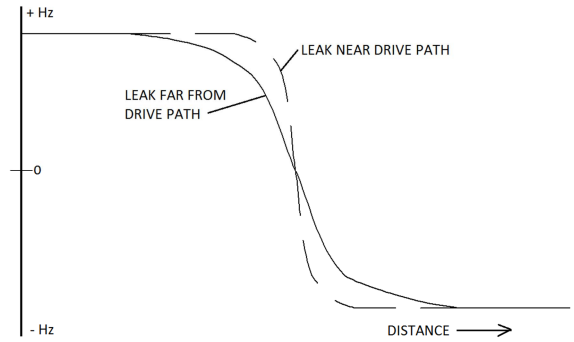


Figure 4. A plot of Doppler shift for two drive-by cases, leak source far from road and leak source near road.

The cone of Figure 2 intersects level ground on two lines, and the vector intersections create a false leak source location and a true leak source location. If one wishes to get rid of the second (false) intersection point, it can be accomplished with a second antenna that rides alongside of the first antenna. As the test vehicle passes the leakage source, the phases of the two antennas' received signals will diverge, as illustrated in Figure 5, and then re-converge.(3) Simple phase subtraction reveals what side of the antennas' drive path the leak is on.

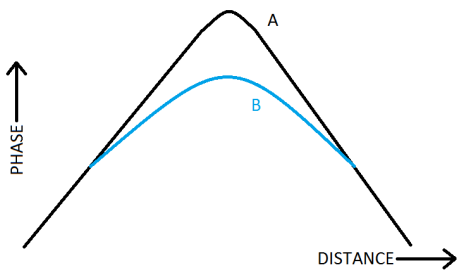


Figure 5. A plot of phase vs. distance for an antenna 'A' that comes closer to a leakage signal vs. an antenna 'B' that does not come as close to the leakage signal. The antenna that achieved the highest phase value came closest to the leakage source.

Figure 6 is a plot of phase difference for a drive-by of a leak. The upper line is when the leak is on the right side of the vehicle and the lower line is for when the leak is on the left side of the vehicle.

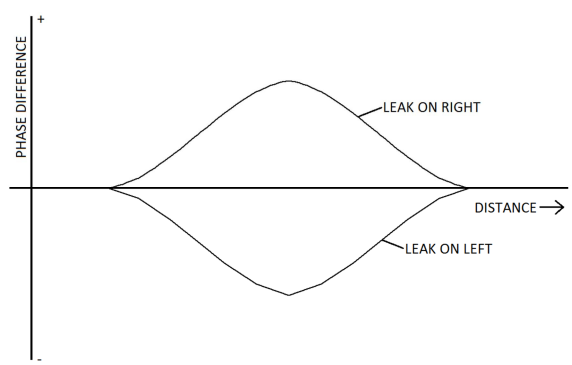


Figure 6. Phase difference between 2 antennas vs. distance. Upper plot is for leak on right and lower plot is for leak on left. Peak phase angle is a function of CW test signal wavelength and separation between the two antennas.

Figure 7 is a visual aid showing a static wave pattern for a source in the center. Note that antenna 'A' comes near the source while antenna 'B' does not get as close. For illustration, the waves are all shown equal in amplitude, but in practice the amplitude of the waves drops 6 dB every time distance from the center doubles. Only the In-phase (I) component is illustrated. The quadrature (Q) component will be shifted by 90 degrees.

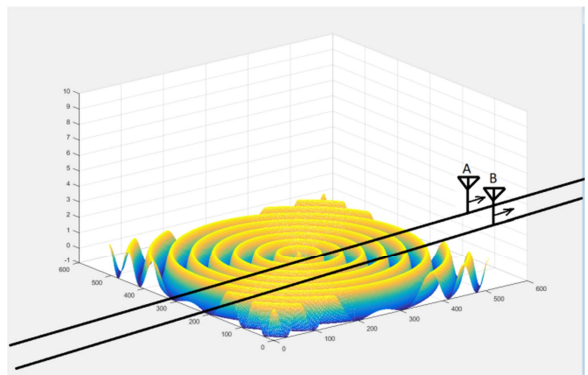


Figure 7. A leakage source static wave visual model. Leak source is in the middle and phase radially increases from the center. Antenna A, which passes close to the source, experiences more phase shift (wave cycles) than antenna B, which passes further from the source. Each full wave cycle corresponds to 360 degrees rotation on the DISPLAY of Figure 1.

The testing was done using an Ettus B210 software defined radio (SDR) connected to a MAC laptop computer. Monopole omnidirectional antennas were used. Incorporating a GPS with the SDR allows the latitude and longitude of the leaks to be recorded. Knowing the velocity of the test antenna along with the exact time when measurements are taken also allows spacing between synthetic phased array elements to be calculated, as well as providing a compass function.

FIELD TESTS ON THE SYNTHETIC PHASED ARRAY

A CW test signal was inserted on the downstream spectrum of live plant at 800 MHz and a hunt for leakage signals began. Testing was not done at VHF frequencies, but this technique should work in this band as well. A GPS module provided latitude and longitude coordinates to locate the test antenna's position. Driving experience revealed what was expected: leakage signals can be dispersed by passing through an object, like a tree, a structure, or by reflections. One helpful way to visualize diffusion from inside a house at UHF frequencies is like a transparent structure illuminated from inside with visible light. The house is made of frosted glass with multiple curved mirrors inside. Walls and roofing diffuse RF signals while metal, such as appliances and cars in a garage, reflect them.

Fortunately, dispersed signals are generally weaker, and they can frequently be ignored when computing intersecting vectors. Generally outdoor environments are more benign for dispersion relative to indoor environments. Figure 8 is a 'fan' display showing how a leak from outdoor overhead plant appeared as it was passed. There is yet another leak ahead. Our evaluation software assumed far-field conditions, which were not necessarily always valid. Our synthetic phased array typically used 128 virtual elements, but for near-field detection the length of the synthetic phased array may be reduced. The fan plot also illustrates the antenna's speed (upper center), a compass function (upper right), and a Doppler plot in the upper left graph.

With calibration, the system method knows the field strength of a leakage signal. The proximity of the leakage source to the drive path is revealed by the grid on the circular fan diagram of Figure 8. Factoring test antenna distance to the leak source with the field strength can reveal whether FCC limits were exceeded.

Figures 9, 10, and 11 are examples of causes of UHF leaks detected with this test method. Figure 9 illustrates a hole worn through the hard line and Figure 10 illustrates a radial crack found underneath heat shrink tubing. Figure 11 shows a radial crack created on an expansion loop. The smaller geometries of defects on UHF leaks explain one reason why the strength of a VHF leakage signal was reduced: the antenna aperture is smaller.

The defects illustrated in Figures 9, 10, and 11 have an effect on cable signals, particularly downstream signals. When a transmission line has an impedance discontinuity, some of the energy will be reflected backwards towards the source. If the RF source (frequently an amplifier) has a good output return loss, the reflected energy will be absorbed. Transmitted energy will then have much lower amplitude and be reduced at end points, such as cable modem (CM) terminals, tap ports, or other downstream test points. As mentioned previously, Figure 11 shows a radial crack. After repair, the operator's operations support system (OSS) revealed 43 homes in the node's service area fed by this cable went from "yellow" status to "green."

One application for this measurement technique is automatic data logging as a test vehicle, perhaps even a pizza delivery truck, is driven around town. Leakage data and GPS coordinates can be uploaded and analyzed overnight when the vehicle is parked. Another application is hand-held test devices and yet another application is airborne testing, both using fly-overs and drones. With a typical UHF amplifier launch signal level being 48 dBmV per 6 MHz channel, and a tap level of 15 dBmV, a leak on a hard line has a potential for 2000 times more power (33 dB). Because of the sensitivity of this technique, individual houses with leakage problems are detectable, even when their leakage signals are well below FCC limits. These houses will be vulnerable to ingress from LTE and other signals.

This test method has exceptional sensitivity for a number of reasons. One is the use of a CW signal and a resulting narrow receiver bandwidth, typically only a few hundred Hz. The CW test signal's narrow bandwidth also makes it easier to find and use a smaller part of the spectrum on a cable system. Another reason is averaging of multiple points to make a single I-Q sample. And finally, use of a synthetic phased array further enhances sensitivity due to directionality.

One weakness of this test system is standard GPS accuracy. The system knows accurately where the leaks are relative to the test antenna, but a low cost GPS receiver system only determines the absolute location of the test antenna to within 4-5 meters. A premium GPS system could provide better absolute accuracy. A photograph of a leak source, with a source in the center of the picture, could also provide improved reporting.

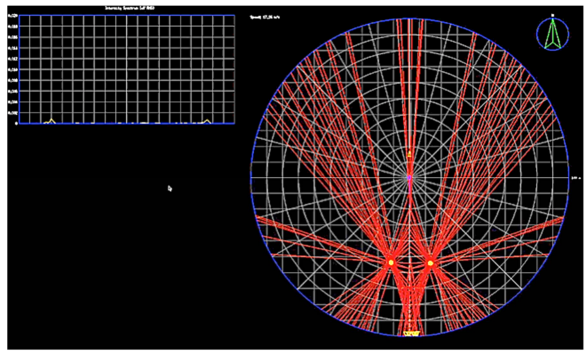


Figure 8. A GPS driving plot featuring drive direction (compass arrow in upper right), drive velocity (upper center), Doppler shift plot (upper left) and fan plot (right side). The polar grid is 10 meters and test antenna's current position is in the center of the polar plot. A purple line in the center of the polar grid displays the length of the synthetic phased array. Using a second antenna reveals which side of the road the leakage source is on. Leak is at the intersection of vectors.

Figure 9. A hole was rubbed through a hardline coax shield causing UHF leakage.

Figure 10. A radial crack causing UHF leakage.

Figure 11. This is a radial crack that was not producing a VHF leakage alarm. After repair, signal quality improved for homes fed by this cable.

BENEFITS

This technique has operational benefits above the required FCC testing and responding to LTE interference reports. The traditional techniques require human support to direct the antenna toward the leaks in order to discriminate which side of the road the leak is on, and to discrimination of multiple leaks. This system allows you to unequivocally pinpoint location of leaks. The unmanned nature of this technique enables these leak finding systems to be placed in any and all vehicles from an operator's fleet to collect leak location data and feed that information into a common database. All technicians may have that information available, and the operator will be able to plan ahead of time how many leaks need to be fixed by which technician. Leakage information can also be used to collaborate other problems, such as UHF or LTE ingress, weak receive levels, high transmit levels, intermittent service, etc.

CONCLUSION

Using GPS data with a leakage signal's phase data as well as magnitude can improve leakage detection by providing latitude and longitude coordinates for the leakage source. A leakage signal's source is identified at the intersection of bearing angles. Signal leakage elimination is more than just a FCC testing requirement; it is an opportunity to improve operations. A repaired leak means less possibility for leakage-related interference to over-the-air services and ingress interference from those same over-the-air services, and less degradation to the cable network's signals.

REFERENCES

(1) Another Look at Signal Leakage – The Need to Monitor at Low and High Frequencies by Ron Hranac and Greg Tresness, 2012 SCTE Cable-Tec Expo

(2) If the leakage testing is being done by an aircraft, piloted or a drone, an intersection of the ground with the cone will form a parabolic shape, assuming flight is straight and level.

(3) Also, if the test antenna's drive path varies, by changing lanes or turning, the false image diverges and the true image converges.

Many thanks to reviewers of this article who provided valuable feedback, and to cable operators who hosted our testing.

LEVERAGING OPEN SOURCE BROWSERS TO OPTIMIZE APPS AND UI PERFORMANCE ON SET-TOP BOXES

Albert Dahan, Co-Founder and CTO, Metrological
Wouter van Boesschoten, VP of Technology and Innovation, Metrological

Abstract

Embedded browsers present a host of performance challenges for operators. Today, the user experience of enhanced browser-based services is largely dependent on the performance of the browser itself. Embedded browsers face set-top box (STB) CPU power and memory limitations that affect performance and fidelity. As consumer demand for personalized OTT offerings increase (Netflix, Hulu, etc) operators need a browser experience that can support a greater number of apps and the mechanisms in place to search and discover the right apps for each person.

HTML5 creates a standard, but it doesn't solve the issue of having to create a uniform experience on every different device. Operators need an in-house browser supported by a framework that can run cloud-based services consistently across all devices. Using an open-source approach enables operators to leverage the rapid increase in innovations coming from contributors to open-source browsing.

The solution is to adopt a browser approach that leverages open-source "WebKit for Wayland" (WPE) components. This new software approach delivers high performance rendering of HTML5 apps and next-generation user interfaces, increasing browser performance with a smaller software footprint, and requires significantly less memory usage. WPE components enable robust rendering of cloud-based applications and next-generation user interfaces and provide better window management to control multiple applications. Operators can enable cloud-based apps to run on STBs with the speed and consistency of native or local apps while avoiding the costs normally associated with proprietary based approaches.

This paper will outline how to implement an open sourced based approach that enables high performance rendering of HTML5 apps and user interfaces. It will discuss the capabilities of the browser in detail along with its pros and cons when compared to other approaches and its ability to integrate into the STB ecosystem.

BROWSER ORIGINS

In the early 90's as the first building blocks of the Internet became available there was a need for a program to retrieve, traverse and present information from internet resources. This ultimately became what we now refer to as an Internet browser. The first widely adopted browser was the NCSA Mosiac browser, which was later renamed to Netscape in 1994. Microsoft responded with Internet Explorer in 1995 and Opera Software ASA responded with Opera version 2.0 in 1996. Apple's Safari was first released in 2003 however, the origins of WebKit go all the way back to 1998, as part of the KDE HTML Layout engine (KHTML) and KDE JavaScript Engine (KJS). The WebKit project was started within Apple in 2001 as a fork of KHTML and KJS.

EVOLUTION OF WEBPAGES

In 1992 CERN launched the first webpage under the HTML1.0 specification. At the time, a webpage was nothing more than text, static images and hyperlinks to other resources. During the first wave of browser launches the HTML specification evolved at a high pace. Within a few years HTML 2.0 (1995), HTML 3.2 with CSS1 (1996), and HTML 4.0 with CSS2 and ECMAScript 1 (1997) were released by the World Wide Web (W3) consortium.

With the general adoption of the Internet, webpages became less static with server side scripting executed through technologies such as Common Gateway Interface (CGI) and HyperText Preprocessor (PHP).

The Asychronous JavaScript and XML (AJAX) technology paved the road for webpages to start changing from simply offering static information to dynamically providing the latest (server side generated) information such as news and web portals. This transition from static webpages to web-applications meant that webpages could now load dynamic data by making client originated queries towards (other) Internet resources; more data could now be loaded from different resources without traversing towards a new webpage. With the rise of social media in the mid 2000's websites dynamically retrieved data and provided (through plugins such as Flash) the ability to playback various media. This in turn, allowed users to share their videos, pictures and other updates with one another through the browser. Webpages were no longer static single load and click to the next resource applications, but interactive applications that retrieved data and updated information as the user stayed within that Internet resource.

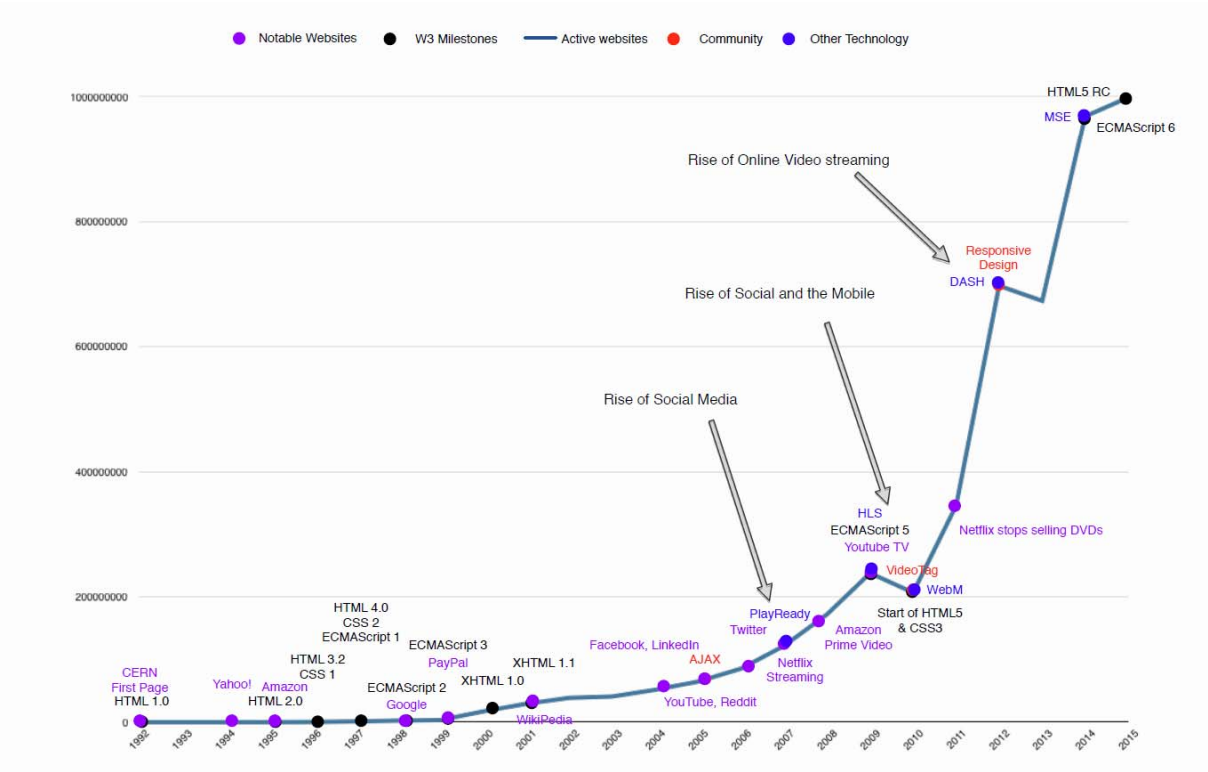


Figure 1: Evolution of website development over time

As mobile devices became popular in the late 2000's (ex: iOS, Android) the concepts of responsive design and single page Model-View-Controller web applications became popular. Various JavaScript frameworks provided the developer with tooling to build full-blown stateful graphical applications on the browser platform.

As mobile devices began to struggle with browser plugins such as Flash, Video Tag was proposed and early drafts of HTML5 and CSS3 started to emerge. Interestingly, the drafts of HTML5 came almost along 10 years after its predecessor (HTML4). While browsers started to adopt portions of the HTML5 specifications, social media platforms were dominating the Internet. Following the rise of social media, along came the rise of online video streaming, which provided further focus on the browser's capability to retrieve and render video. With websites such as YouTube serving millions of videos every day, companies like Netflix, Amazon Prime Video, Hulu started to get immensely popular. This in turn drove the need for adaptive streaming and common content encryption capabilities in the browser.

Today's websites, which are loaded by browsers, are no longer web pages. Even the term 'web page' seems to indicate something static. These words depict something that resembles a page in book or text in a newspaper. Today's web pages do not fit that description at all. These web pages provide user interaction, dynamically loaded information (sometimes even real-time), advertisements, two-way communication, audio/video streaming and 3D graphics on a platform that is uniform across multiple devices. Interestingly, this application environment isn't controlled by a single technology company. It is truly an open platform where the community drives the need for changes. Open source is leading the way and the browser is by far the most widely used application platform on the Internet. The browser is no longer 'a program to retrieve, traverse and present information from Internet resources'. It is a full application environment.

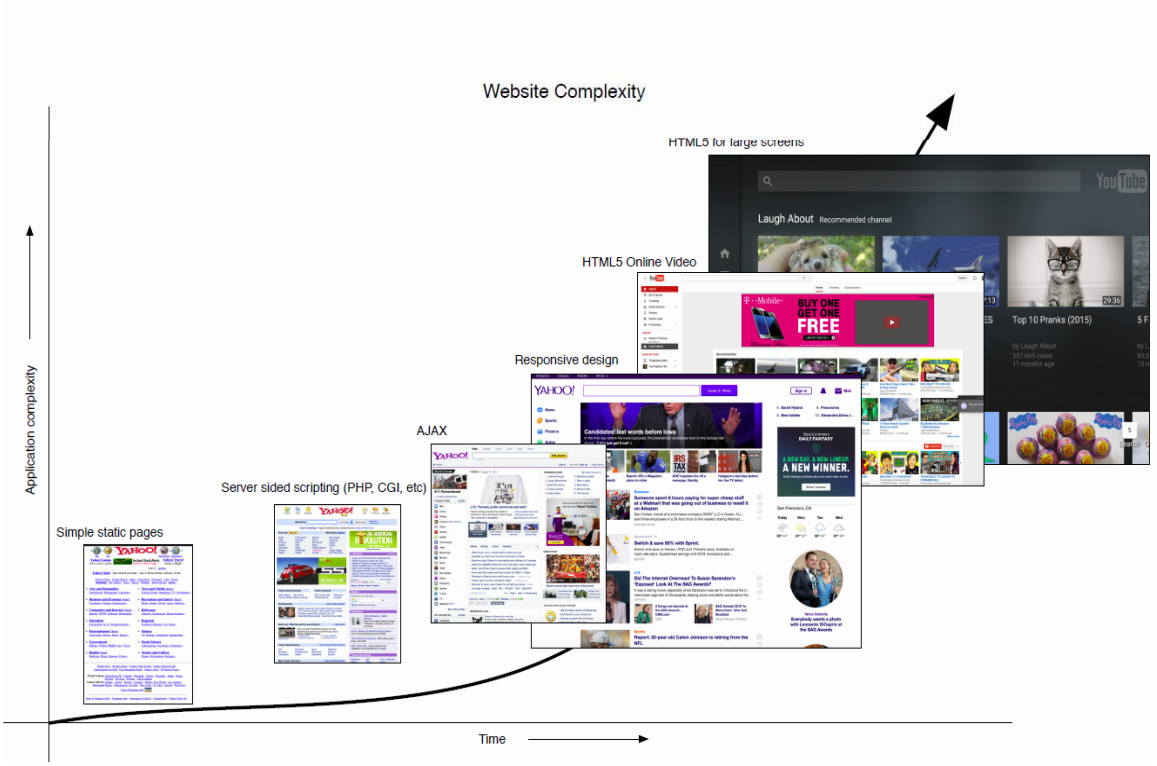


Figure 2: Website complexity over time

EMBEDDED BROWSER IMPLEMENTATIONS

Embedded devices have always been a bit of a niche market for browsers. With the rise of mobile devices there was no unification on the browser and no massive adoption of one or two browsers. Unlike the desktop market, mobile users do not have a 'preferred browser'. Historically, PC users install their own browser on their desktop. This is because at least initially operating systems did not come with a latest and greatest browser (i.e. Windows and Internet explorer). Mobile devices, on the other hand, are different. They are already bundled with a browser from its respective company (e.g. iOS comes with Safari, Android with Chrome, etc.) and use the app store mechanism to keep it up to date (an approach later adopted by desktop environments).

For embedded devices that are not running iOS, Android or Windows Mobile there is not a lot of choice. There are a few proprietary solutions available which require a license, but often these proprietary solutions struggle to keep up with the high pace of new HTML5 specifications.

Through the Chromium project, the Blink-based source code was available for porting towards embedded devices. This provides a Chrome-based browser for mobile devices. However, Blink is built for a desktop and requires desktop resources in terms of available memory, Central Processing Unit (CPU) and graphics power. It runs great on expensive embedded hardware, but not all embedded projects can afford 600 dollars' worth of hardware.

For quite some time the only license free solution was a WebKit port on top of the QT application framework. With the introduction of QT 5.4 this became increasingly harder due to licensing changes by QT. Since QT port isn't maintained upstream, WebKit is clearly the best choice due to its lightweight nature and BSD v2/LGPL v2 licensing. Being maintained by Apple, Adobe, KDE (graphical desktop environment for UNIX workstations) and others, the WebKit project is sure to quickly adopt the latest W3 specifications for HTML5.1 and beyond. All that was missing was a free and lightweight graphics framework.

FUTURE TRENDS OF W3C AND THE IMPACT ON EMBEDDED DEVICES

The shift from static web pages to an application environment isn't accidental. The W3 consortium started its catch up with other application environments with the start of the HTML5 specification. The major highlights in past iterations included video tag, drag and drop, offline apps and Canvas. The new HTML 5.1 specification enables media source extensions, encrypted media extensions, full screen Application Program Interfaces (APIs), geolocation, Indexed Database API (IndexedDB) and Web Audio support.

However W3 isn't stopping there. A closer look at the task forces that are present within the W3 reveals that they are planning to add functionality for automotive, mobile devices and TV and broadcasting. There are already W3 drafts/proposals for functionality such as screen orientation, lock screen for mobile, reading metrics for cars and the TVAPI for tuning and recording functionality on broadcasting devices.

W3 is clearly going after these other application environments and with the focus on mobile, cars and TV, the need for an

embedded browser is much more important than ever.

PERFORMANCE CHALLENGES OF EMBEDDED BROWSERS

Anyone who has developed on embedded devices will acknowledge that performance is the Achilles' heel of embedded solutions. Embedded solutions continuously balance hardware costs with performance. Like all computer hardware, the more expensive the hardware components the more resources a software program will likely have. The more resources a software program can utilize the better the performance, especially for graphical applications.

Those who expect high-end smartphone and gaming console performance out of a 60 dollar device should think again. The actual cost of a smartphone or gaming console that can support cutting edge graphics and high performance capabilities would run almost tenfold of that in terms of costs. The right balance can be hard to find. If the manufacturing costs are too high, the end consumer product will likely be too expensive. A product with functionality that is too limited, even with a small price tag, will have too many hardware compromises. This results in an end product that underperforms, yielding a negative end user experience.

Software matters when considering the capability of hardware. It won't make a 60 dollar device outperform a 600 dollar one. However, software should be at a level where it meets the hardware capabilities. In order to achieve the right performance there are several factors, such as complexity and optimizations that are crucial. However, complexity and optimizations are almost orthogonal, as optimizations can quickly get complex. The key is to combine both at the right place. Use simple solutions where possible and leverage complex optimizations that are already out there and vetted by the community.

Need for a new approach

WebKit for Wayland (WPE) does exactly that. Leveraging state-of-the-art browser optimizations capabilities provided by WebKit, such as the JavaScript Core Fourth-tier optimizing (FTL) JIT compiler, and combining that with Wayland. Wayland provides simple, elegant, graphics compositing integration between different layers using EGL (interface between Khronos rendering APIs). Wayland started as a replacement for the X Windowing System (X11). Wayland is not an implementation but a protocol specification between a display server and its clients. Its implementations are lightweight with a small footprint. Wayland is primarily focused on performance, code maintainability and security.

Integration effort, time to market

Creating proprietary solutions for a single device with limited requirements will always be viable. Performance can be safeguarded as the complexity of the solution is within controllable limits. However these kind of proprietary solutions do not scale very well over multiple devices. Even using open source does not solve the scaling issue. Using community driven components does not necessarily mean it will lower integration efforts for every device. In other open source browser solutions substantial time is lost on the integration of low-level primitives for the browser such as graphics and input. This means a full stack integration and effort on both the Software Development Kit (SDK)/driver level as well as how the browser uses those primitives.

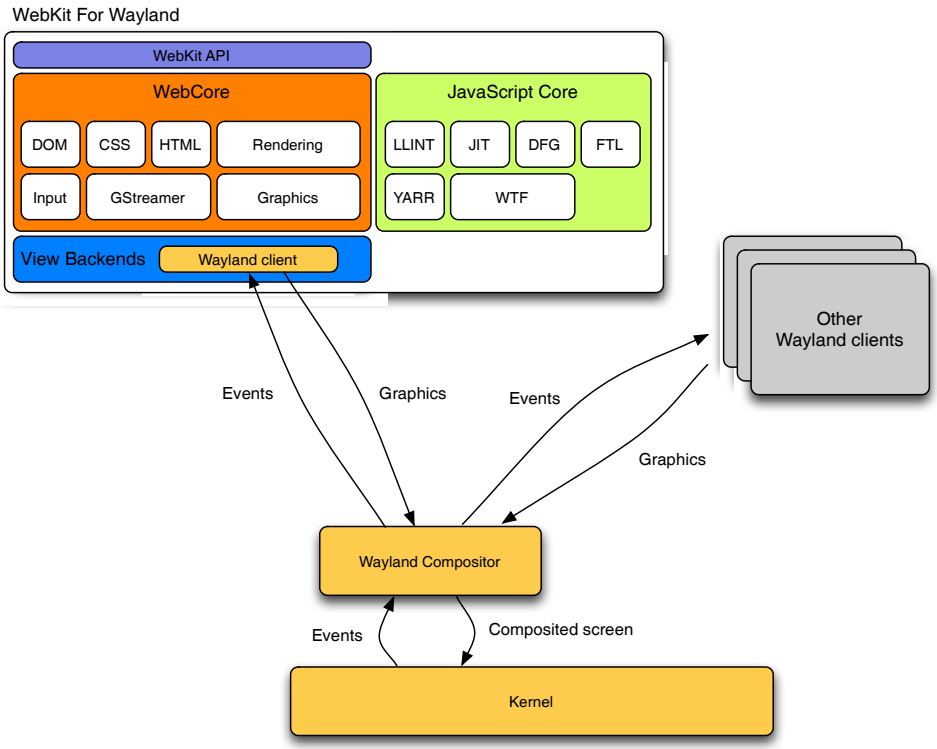


Figure 3: WebKit for Wayland architecture

To solve the issue, Wayland separates these responsibilities. It provides a unified protocol specification for a multi-program graphical compositing environment and user input. Simply put, the hardware/SDK supplier can ensure that their drivers work with Wayland by using a reference implementation (e.g. Weston) without even touching a browser. These hardware/SDK suppliers can ensure the drivers are operational with a minimal stack. And, the browser can also be validated and tested without doing full-stack integration on the target device. The browser will still need to be validated when the two are put together, but the SDK/drivers and the browser components can be independently validated. This independent validation lowers the risk of integration issues and integration effort, thus speeding up time to market.

Code maintainability

The bigger the footprint, the harder it is to maintain something. Simply put, more lines of code equal more maintenance. Open source is great, but it's important to use the right solution for the problem. The QT application framework served a broader, bigger scope than just rendering webpages. This broader scope meant a maintaining a lot of layers, abstractions and lines of code for a browser.

A browser consists of a piece that processes, parses and runs the application (DOM, JS and CSS) and a piece that renders graphics. WebKit handles the first series of functions. Wayland handles the rendering with a much leaner fit. Not only does Wayland more concisely solve the rendering problem, but it also provides an easy path forward for multiple instances of WebKit and

sharing graphical/input resources between other applications through the Wayland protocol. This functionality is often sought after in any browser solution and is more important than ever in the embedded market where multiple video streaming applications are emerging fast.

When selecting an open source project another requirement is to stay open. Quite often in the past, proprietary solutions were based off of something that was open source, but then diverged from open source with proprietary or closed modifications. Breaking too far away from the open source architecture means that the two are so far apart they can't be merged anymore. Staying code compatible and even bug-to-bug compatible with the upstream project is paramount. This way, if someone finds an issue they can solve and share it upstream and upstream fixes can be easily pulled in. Keeping the code upstream compatible is essential in order to leverage the open source community. By contributing solutions back into the community everybody benefits.

WPE contributes back to the community and is on average seven days behind the main WebKit trunk. The WebKit for Wayland browser is upstream compatible and will continue to follow the tip of the trunk from WebKit.

Native or local implementations

Quite often skeptics say, "A native user interface (UI) implementation is better then a browser". Let's focus on what this implies.

First, what is a native UI? Technically, native means something that exists or belongs to one by nature, but in this context that meaning doesn't apply. Software is (cross) compiled on a machine that uses human readable source code to turn it into something a machine can understand. That doesn't fit the meaning of native here. So let's paraphrase native and assume skeptics mean something that is closer to the hardware primitives. Often with embedded development, as mentioned above, fewer layers mean less complexity and could actually benefit performance. And that's correct. Writing code in assembly (if you know what you are doing) will always outperform code that is written in a high level programming language. Writing code in assembly would take years, if not decades. Which is why brilliant engineers in the early 1970's came up with generic purpose programming languages, such as C/C++ and beyond.

In essence it's not about writing the entire application in assembly or low-level hardware primitives. Let's focus one level up from those primitives like C and OpenGL and look at the second part of the sentence: "A native UI implementation is **better then a browser**". One can't compare a browser to a specific UI application written in C. The browser by itself doesn't do anything. It's necessary to write an application that runs within the browser. Would an application written straight on top of OpenGL outperform an HTML5 application that runs in the browser? That really depends on the implementation details of the HTML5 application. Would it outperform a set of animations in CSS? Not likely, as a browser is very efficient in determining its strategy to render CSS and it has years of experience to back that up. Technologies such as the threaded compositing provide huge benefits in terms of performance that will need to be replicated in the C application. This in turn, takes a lot of time (and money) and adds to the complexity.

Next, let's look at complex 3D graphics. Will an application that renders complex 3D graphics written in C directly on top of OpenGL outperform a HTML5 application using CSS animations? Yes, without a doubt. CSS is not meant to compete against complex 3D graphics. Cascading style sheets are created to style the text and blocks within the Document Object Model (DOM) tree. In lieu of that, the browser supports doing Canvas and Web Graphics Library (WebGL) straight from the browser for complex graphics. The browser exposes methods to access those OpenGL primitives straight from the HTML5 application. You get access to the same low-level interface. In that sense it is as 'native', to use the words of the skeptics, as writing your application straight on top of OpenGL for 3D graphics in C. So if both components are using the same APIs there isn't much difference in the one or the other for graphics. This makes the argument of 'native' no longer about access to graphical primitives, as both environments have the same accessibility.

That leaves the difference of writing an application in C versus writing an application in the browser in terms of CPU and memory performance. What is the added value of a browser over doing an application directly in C? An application in C, if written well, will be faster than anything one can create in a browser. However the key thing to note here is "if written well". C and C++ are general-purpose low-level programming languages, meaning a software developer can write code in a generic syntax and access low-level APIs. This has benefits over writing the code in assembly. C and C++ still provide access to those low-level primitives such as access to memory and other hardware APIs. However, writing against low-levels takes a great deal of time and it's important to note that C and C++ can be very unforgiving. Since it has access to low-level primitives, making a mistake can have disastrous effects to the runtime of the application. Due to its low-level nature it takes a lot of time to do basic things that are not exposed the same way as in high-level languages such as JavaScript. This includes type casting, threading (service workers/web workers), non-blocking code and all the utility the browser gives (network stack, player interfaces, etc.).

If someone makes a mistake in JavaScript the compiler will deal with it. Mistakes are unavoidable. The JavaScript Core in the browser combined with four tiers of JIT compilers will kick in as functions are used more frequently within an application. The JavaScript Core JIT compilers optimize the code to a level that is hard to accomplish with one's own C/C++ implementation. A lot of research and development went into these different JIT tiers. Because of this, it's hard to beat with your own application. At least not without spending millions of hours on optimization alone. Hours that someone has already spent and made available within the community as part of the browser.

On top of that an HTML5 application is a lot easier to maintain with equal amount of functionality in a C/C++ application. Since the code does not need to deal with a lot of primitives the code becomes smaller, simpler and less complex. This makes future maintenance easier. Finding developers is easy and often cheaper as it does not require very specific expertise that would otherwise be required in a C/C++ application. It provides equal or better performance due to the level of optimizations the browser can apply without involving the application developer.

Back to the question: "What added value does a browser have over doing your application directly in C?" Well, quite a few: ease of large-scale development, code maintainability, time to market, threaded compositing and built-in performance through its JITs to name just a few.

WEBKIT FOR WAYLAND INTEGRATION

With the removal of the QT application framework and providing a lean and mean integration towards Wayland the WebKit for Wayland browser can be easily integrated in low cost devices. If the device supports a Wayland compositor the integration is extremely straightforward. However in cases where the System on a Chip (SoC) vendor does not provide the required support for the Wayland integration the WPE ViewBackend can be extended to support direct graphics integration with the required SoC drivers. This can be used as a fallback mechanism in cases where the SoC is no longer actively developed on and where it is unlikely the hardware drivers will be extended or modified to meet the requirements of Wayland.

WPE comes with reference build environment recipes for Buildroot and OpenEmbedded for easy adoption within (existing) build systems. At the writing of this article WPE is mostly supported on all major hardware platforms, including various Broadcom chipsets, Intel CE chipsets, Nvidia gaming platforms and the Raspberry Pi family. The latter is mainly used for validation of WPE and because the Raspberry Pi is widely adopted within the open source community. This enables the open source community to quickly develop and validate using a widely available and cheap device.

Much of the development done for WPE is fed back to the community. Lots of changes originally developed for WPE have already been provided back to Apple WebKit and G-Streamer. WPE will continue to pull in changes from upstream WebKit. WPE is dedicated to continued support and contributions to the open source community.

Because of its open source nature anyone can access the source code. The source code is available on a GitHub repository: https://github.com/Metrological/WebKitForWayland.

Obtaining the source code is just a matter of cloning the repository and building for the right target machine (don't forget to checkout the Buildroot and OpenEmbedded repositories). Contributing to WebKit for Wayland is as simple as forking the repository, make your changes, run the tests and create a pull request. The changes will be reviewed by the open source community, and if approved, merged into the trunk. That's it. No license fees, no hours of meetings, no masses of documents and no other overhead. Just code.

CONCLUSION

Embedded browsers have always been the underdog of browsers. Historically, they were available solely as proprietary or license restrictive solutions, which didn't offer a lot of choice. With HTML5 features growing at a rapid pace and entering new territories such as the mobile, TV and automotive industries, the need for a fast, open, bleeding edge embedded browser is greater then ever.

Embedded browsers should be widely available to anyone. Similar to most desktop browsers the embedded browser should be

free and open source. By combining the latest WebKit and Wayland integration is simple and straightforward.

The browser should follow the bleeding edge of the W3 standards. By following the WebKit main trunk closely, all new features can be pulled in with ease.

WPE provides a simple, high performance, low footprint, well maintained open source browser. By truly being part of the open source community, everybody is invited to contribute to the better future of embedded browsers.

Machine Learning: The Past, Present and the Future

Narayan Srinivasa
Intel Corporation

Abstract

Machine learning algorithms of the past were designed to capture the learning capabilities of the brain but with a high level of abstraction of its learning mechanisms. These abstractions resulted in shallow learning models that relied on hand crafted feature extraction on a problem specific basis with limited practical applications. The deep learning models of today represent the next generation of machine learning algorithms that can be trained from raw data using multiple processing layers due to novel modifications to the learning architecture compared to shallow learners. This development combined with the availability of raw data to train these models and the availability of fast affordable computers have enabled a great surge in its utility for many applications including video and audio pattern recognition. By exploiting the fundamentally different computing architecture and mechanisms prevalent in the brain, we believe that the next generation of machine learning called neuromorphic computing will advance the state-of-the-art in this field. In particular, it has the potential to realize energy efficient learning machines that could support a wide range of applications including internet of things, sensor processing, cybersecurity, robotics, mobile devices, diagnostics and prognostics and exoscale computing systems.

INTRODUCTION

Can we build machines that can exhibit intelligent behavior? The answer to this question has drawn our attention to understand and mimic how the human brain functions. So far, most attempts to understand brain function has focused on how the brain can compute intelligent behavioral responses from internal representation of stimuli and stored representations of information of past experience. The roots of this approach can be traced back to two key ideas. Alan Turing's pioneering work [1] in machine theory defined computation as formally equivalent to the manipulation of symbols in a temporary buffer. Similarly, the pioneering work on telephone communication by Shannon and Weaver [2] resulted in a formal definition of information where informational content of a signal is inversely related to the probability of that signal arising from randomness.

As these developments launched computer science into prominence, and as computers grew in functional complexity, so did the analogy between computers and brain. The basic premise for this analogy was that computers and the brain received information from the external environment and both acted upon this information in complex ways. This analogy (also known as the computer metaphor) provided a candidate mechanism to explain intelligent behavior as akin to a digital computer program that can manipulate internal representation according to a set of rules. Furthermore, mental entities in the brain were akin to software, whereas physical mechanisms were akin to hardware.

The extensive use of the computer metaphor resulted in the brain models using network of neural cell like computational elements [3], or artificial neural networks (ANN), as a proxy for the computation in the brain. This was the origin of the field of machine learning. In this paper, we will briefly outline the key ideas that shaped this field in the past and its status today. We will then contrast it with the emergence of neuromorphic computing as the next generation of machine learning, highlight its salient features and related challenges, discuss potential applications of this technology and how it could influence our future.

ARTIFICIAL NEURAL NETWORKS

In an artificial neural network, simple artificial "processing elements" or "neurons", are connected together in layers to form a network which mimics a biological neural network. The connections between the artificial neurons represent tunable adaptive weights (Fig. 1(a)) that serve as a proxy for the synapse in biological neural networks. These neural networks (Fig. 1(b)) are abstracted versions of biological neural networks but similar in performing of functions collectively and in parallel, rather than there being a clear delineation of subtasks to which individual neurons are assigned.

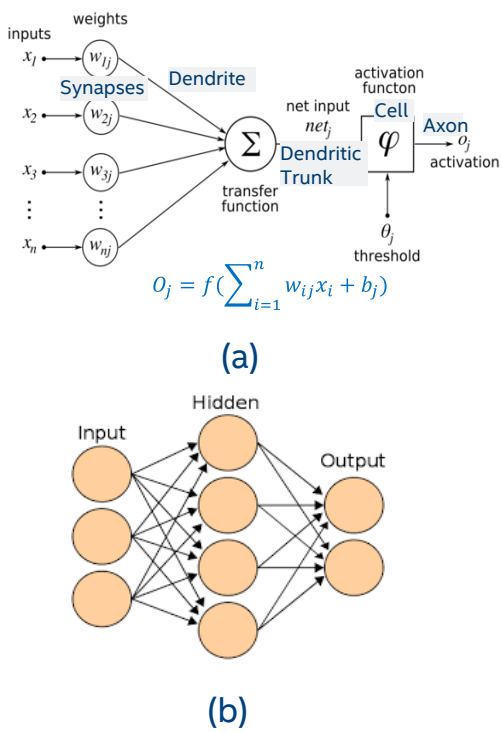


Fig. 1 (a) the artificial neuron model with the biological analogs appropriately marked. The output was a nonlinear function *f* of the weighted inputs. (b) A three layered network with each circle representing one of the artificial neurons connected.

The first interesting network called the *perceptron* was a two-layer network designed for pattern recognition [4]. After going through road bumps with the basic perceptron in its inability to solve the exclusive-or problem [5], the creation of the back-propagation (BP) algorithm [6] not only solved the exclusive-or problem but also enabled the training of these networks resulting in the first generation of practical machine learning models.

Shallow Machine Learning Models

The first generation of machine learning models was commonly designed to be supervised where the goal of the system is to classify input patterns into desired output class labels. During training, the machine is shown a sample from the input data and the corresponding output score for all the categories is provided as ground truth. The ideal outcome after the training process is for the machine to able to infer the correct class for inputs both within and outside the training data. To accomplish this, the training process leverages the BP algorithm. The first step in the process is to compute an objective function that measures the distance (L_2 norm) between the desired and actual scores for the class labels. The BP algorithm then modifies its internal adjustable parameters such as the adaptive weights to minimize this distance across all training data. The BP algorithm computes a gradient vector for each weight that evaluates how the error changes (either up or down) as a function of weight changes (by a small amount up or down). The learning step then adjusts this weight in the opposite direction of the gradient vector. This process is repeated for all the weights in all the layers in the network [7].

After training, the performance of the system is measured on a new data set called a test set that is used to compute the accuracy of the classification on new inputs that the machine has never seen before. While this algorithm worked very well for hand crafted input features in the training data, it did not work well for raw data. This is because with small initial weights, error gradients using BP in early layers are very small and this gets worse with depth (number of layers) of network. Thus only *shallow networks* (depth <= 3) could be trained. Furthermore, BP algorithm was also found to be susceptible to becoming stuck in local minima with large initial weights. This made shallow learners sensitive to irrelevant

details such as illumination changes or variations in the pitch or accent of speech [8]. Thus, there was a need for considerable skills to create hand crafted features that could ensure high performance. This was a key limitation in the adoption of shallow learners for real-world applications.

Deep Machine Learning Models

The most common machine learning models used today are deep learning models that are composed of multiple stacks of layers or modules with millions of parameters for tuning (Fig. 2). Each layer computes a nonlinear input-output mapping that increases both the selectivity and invariance of the representation. This feature of deep learning models enables the network to compute intricate functions of inputs from raw data that are sensitive to object details while being insensitive to irrelevant variations –illumination etc. There were two flagship models that achieve these capabilities. The first deep learning model are the deep belief nets based on the Restricted Boltzmann machine (RBM) and the second is the convolutional neural network.

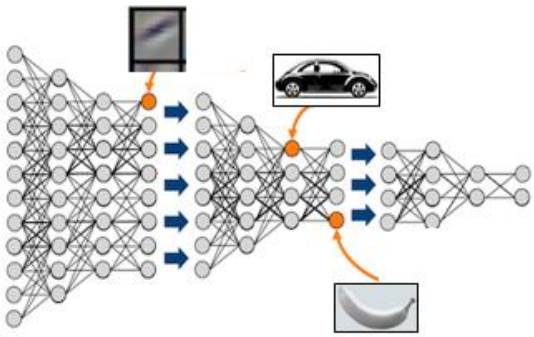


Fig. 2, A deep learning model with multiple layers where the neurons in the earlier layers represent low level features such as oriented edges while the neurons in the higher level represent high level concepts such as cars and banana.

The RBMs were invented by Geoffrey Hinton and his group [9] for the purpose of unsupervised of features from unlabeled data [10]. The objective of each layer was to reconstruct or model activities of features (or raw inputs) in the layer below. As a first step, the raw data in the first layer was used to construct the feature detectors in the second layer and the second layer in turn was used to reconstruct the raw data (Fig. 3). This process was repeated pairwise for all pairs of layers to create a deep auto encoder also known as the RBM. This unsupervised process forms a crucial first step in training RBM based deep learning networks and is commonly referred to as *pretraining*. Using pre-training enabled the deep learning system to be seeded with sensible initial weights in an *unsupervised* fashion and the process was scalable to arbitrarily large depths of network layers.

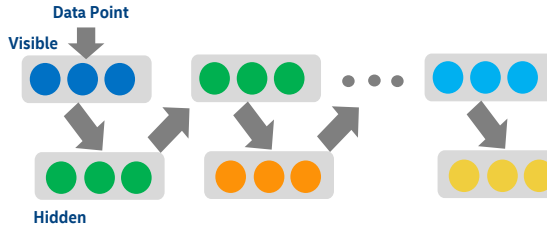


Fig.3. the pre-training process for the RBM is illustrated here. The input data point forms the visible layer that the second layer of neurons are trained to represent during learning. This hidden layer then forms the visible layer for the second pair of layers and the process is repeated until all sequential pairs of layers in the forward direction of the deep learning network are trained.

As a second step, all the layers are then combined and the full network is trained using BP with these initialized weights. This step is referred to as *fine tuning*. The interesting aspect of fine tuning is that the network converges readily compared to shallow learners while producing very low errors. Another huge advantage is that both pre-training and fine tuning scale linearly in space and time with the number of training cases. This enables RBM based deep learning models to be trained with very large data sets but without needing much training time.

Inspired by the hierarchical visual cortex model by Fukushima [11], another deep learning model was developed by Lecun [12] called convolutional neural networks. This model was designed to process data that come in the form of multiple arrays – 2D/3D images, signal sequences including language etc. This deep learning model was composed of a series of stages (Fig. 4). Each stage composed of two layers: convolutional layers and pooling layers [12]. The neurons of the convolutional layer is organized into feature maps where each neuron in this map is connected to local patches in map of previous layer via a set of weights called a filter bank. Since this filtering operation is a discrete convolution, the deep learning network was called convolutional networks (ConvNets).

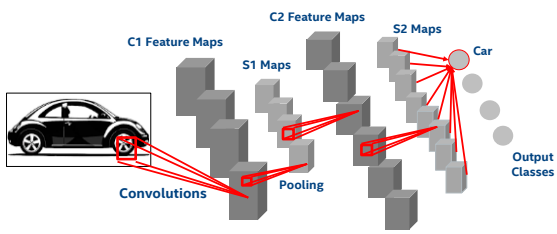


Fig. 4. A two layered convolutional network showing the convolution (C#) and pooling (S#) operations where # represents the layer number. The last layer represents the classification layer or the outputs of the network.

The pooling layer merges semantically similar features into one. A typical pooling neuron (also known as a rectified linear unit (ReLU)) computes the *maximum* of a local patch of neurons in the feature map (Fig. 5). Neighboring pooling neurons take input from patches that are shifted by more than one row/column thus providing shift and distortion invariance in its processing. Multiple stages of convolution and pooling with a final coding layer forms the ConvNet.

The training for the ConvNets was based on the BP algorithm where the weights in all the layers of the deep learning model was adapted based on the errors in classification. Unlike shallow learners, these networks also converged readily since the ReLU units offers more structure to the network and the weights are well-conditioned as a result for learning.

There are three factors that enabled the rapid adoption of deep learning models as a flagship approach in machine learning today. The first factor is the ability of these networks to be trained even with many layers of depth thanks to innovations such as the RBM and ConvNets. The second factor is the availability of fast enough computers with affordable large memories that enabled these deep networks with many millions of parameters to be trained. The third factor was the availability of large data sets for training and benchmarking due to emergence of the Internet and fast computers that provided the raw data (such as images and audio data) with ground truth. These approaches are now being explored from a software perspective by major companies like Google, Facebook, Microsoft, IBM, Yahoo!, Twitter, and Adobe. They are rapidly developing image understanding products and services using deep learning models. In parallel, these models have been implemented in hardware by companies such as Intel, NVIDIA, Mobileye, Qualcomm and Samsung to enable real-time vision applications for smartphones, cameras, robots and self-driving cars [8].

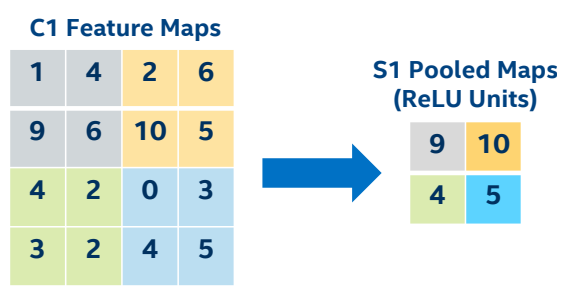


Fig. 5. A ReLU based computational process where the max of each local patch of features is extracted by the pooling process.

NEUROMORPHIC COMPUTING

The original term in this field was the term "neuromorphic electronics" coined by Carver Mead [13] at CalTech in the late 80's to describe electronic *analog* circuits that *mimic*

neurobiological circuits and architectures in the nervous system. Lately the term "neuromorphic engineering/computing" was introduced to expand the scope to include analog, digital, mixed-mode analog/digital VLSI and software systems and algorithms. This is a multidisciplinary field with skills ranging from computer science and engineering, physics, mathematics, neurobiology, psychology and computational modeling. It is also a very dynamic field where our understanding of the nervous system is changing all the time due in part to better measurement tools such as optogenetics, viral tracing, better understanding of cellular features, better models that capture this understanding and new insights and theories that makes this field both exciting but also challenging.

The foundation of neuromorphic computing unlike the machine learning algorithms lies in understanding and exploiting how biological computation is very different from digital computers of today. Brain is composed of very noisy analog computing elements including neurons and synapses. Neurons operate as relaxation oscillators. Synapses are implicated in memory formation in the brain and can only resolve between 3-4 bits of information at each synapse [14]. The dynamics of these elements are asynchronous and thus *clock-free* [15, 16]. Since these synapses are fully distributed and they are implicated in memory, this implies that, in general, there are no single synapse or single neural firing activity that corresponds to a particular item or concept and so are *symbol free* [16]. The effective integration of heterogeneous and non-local sources of information for multiple goals and is a hallmark of human-like cognition [16, 17]. The brain thus operates in a *grid-free* fashion. Finally, the scale of neuronal interactions can span from a few neurons all the way to the entire network of neurons in the brain during various time instances that depends upon on the context of its interaction with its environment. This implies that the brain is a complex physical system whose dynamics is *scale-free* [18].

While these features of neuromorphic computing are foundational, there are four clear focus areas that are being actively explored today to design and architect the next generation of machine learning algorithms and hardware systems. We will now highlight these areas and highlight the progress and challenges in all these four focus areas.

Spike Based Representations

Brains operate using spike-based codes that often appear sparse in time and across populations of neurons. A neuron generates these spikes in response to spikes from other neurons by integrating the current that leaks

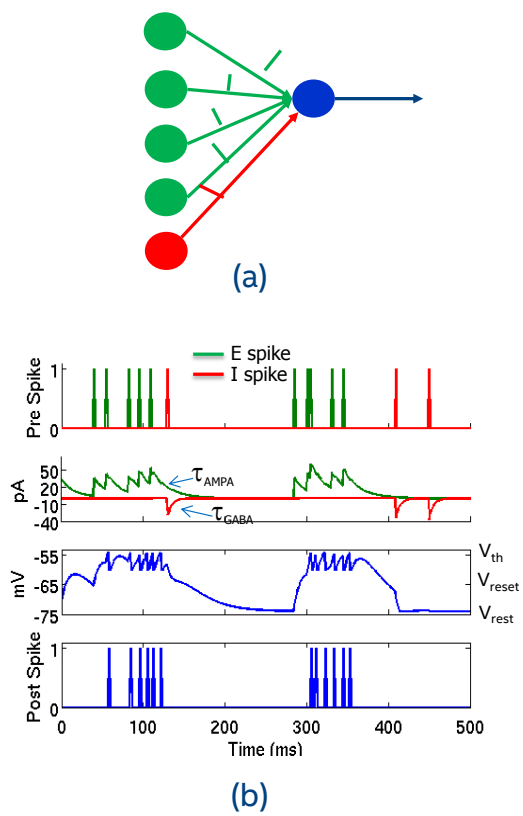


Fig. 6. (a) A simple 5 input neuron network with four excitatory neurons and one inhibitory neuron to a single post-synaptic neuron. (b) The input spikes shown in the top row gets converted to input currents (the green and red traces – the second row). The membrane voltage and the corresponding spikes of the post-synaptic neuron are shown in the last two rows.

into the soma and generating a spike of its when the net current exceeds a threshold (Fig. 6). Spike-based representations offer the advantage that sensory information is encoded by relatively small populations of spikes and their precise relative timing.

This capability can be very useful because it offers an efficient solution by generating sparse codes [19]. In particular, it provides an approach to extract and represent useful information from high-dimensional data such as video images by extracting the most relevant information from just a few spikes [20]. There seems to be some understanding on how such spike-based representations can be realized for small scaled feedforward and recurrent spiking neural models [21, 22].

Spiking neural networks (SNN) based architectures have recently been implemented [23, 24] to mimic deep ConvNets. The communication between the various stages of ConvNets are through spikes but the synaptic weights are trained off-line and programmed into the model.

Benchmarking of these models using CIFAR-10 datasets composed of real world images show that the performance of these models are comparable with state-of-the-art ConvNets algorithms while realizing energy efficiencies of ~10-20x. A recent spiking model [25] inspired by auditory processing in mammals was found to exhibit feature sensitivity by being robust to a range of variations in the stimulus such as found in naturalistic stimuli and human speech samples. In robotics, applications of spiking neural models including the control of a simulated robot arm [26] and application in Robocup competition [27] have recently appeared.

There are several recent hardware systems that have utilized spike based representations. In the case of sensor processing, a silicon retina was developed [36]. This sensor is a spike-event generating image sensor consisting of 128 x 128 pixels which asynchronously outputs streams of address events in response to relative light-intensity changes (Fig. 7). The events are tagged with the address of the

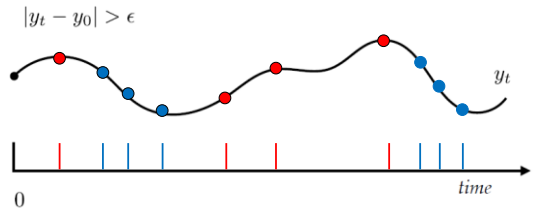


Fig. 7. An event or asynchronous coding of intensity y_t via spikes where the red spikes represent a nominal upward change in intensity compared to a baseline y_0 and blue spikes represent a nominal downward change. The parameter € provides the threshold for change.

creating pixel, a time stamp, and an ON or OFF polarity tag, which indicates whether the event was created in response to an increase or decrease of light intensity over that pixel. The sensor is able to respond to these changes with extremely low latencies of 15 µs. An address event based silicon cochlea [28] was developed to mimic the biological cochlea. This device transforms input sounds into streams of spikes in 64 channels responsive to different frequency ranges (350 – 1200 Hz) where the inter-spike-intervals was a more precise indicator of the frequency of pure input tones than the distributions of channels from which the spikes originated. These sensors were also recently fused to perform real-time classification that is comparable to conventional CovNets based algorithm performance on the MNIST dataset [29]. There have also been several recent applications that have exploited these types of sensors for real world applications including motion sensing [30] and visual shape tracking [31] and navigation [32] and also for visual information processing [33].

IBM developed a TrueNorth chip [34] that integrates 1 million programmable spiking neurons and 256 million configurable synapses. Chips can be tiled in two dimensions via an interchip communication interface. The architecture applied for example, multiobject detection and classification but the synaptic weights were obtained in an offline fashion. For a 400-pixel-by-240-pixel video input at 30 frames per second, the chip consumes 63 mW. Inspired by the brain's inherent dynamics and

plasticity, HRL has developed an efficient, scalable, and flexible non–von Neumann architecture that leverages contemporary silicon technology to learn on-chip [35]. They recently demonstrated the application of this chip on a nano air vehicle [36] to learn on-chip to recognize three rooms in under 8 minutes with ~50x improvement in energy and ~10x improvement in throughput compared to a conventional processor.

The Neurogrid project [37] at Stanford provides an option for brain based simulations by using analog computation to simulate ion-channel activity and uses digital communication to simulate synaptic connections. The Neurogrid can simulate a million neurons. There also other efforts in Europe including the work on FACETS and BrainScaleS projects [38-40] and in the UK SpiNNaker Project [41]. The current version of the Neuromorphic Physical Model (NM-PM) from the BrainScales project [42] incorporates $50*10^6$ plastic synapses and 200,000 biologically realistic neuron models on a single 8-inch silicon wafer in 180 nm process technology.

Challenges with Spike Based Representations: There are gaps in our understanding of how large scale spiking networks could help compute and stably represent information using recurrent architectures that seem to be ubiquitous in the brain. There are interesting hypothesis [44, 45] emerging on how this may be performed but still lack strong experimental support due to lack of experimental tools that can monitor large populations of neurons at the spike level. Another gap in our understanding comes from the fact that spike-based representations are inherently noisy (e.g., latency, jitter etc.). The fact that the brain appears to embrace this as a feature rather than avoid it remains to be fully understood. There are some interesting recent ideas (e.g., Bayesian computation [46-48]) on how this could happen in the brain based on spikes as its representation. Finally, the spike based representations in the brain are invoked both by sensory stimuli and spontaneously when there are no stimuli. Another key gap to address is in our understanding of the role of these two modes of operation for computing (unlike Von Neumann machines that do nothing if there is no input) and interacting with a changing environment.

Asynchronous Computing

It is well known that the brain operates using a plethora of brain rhythms (see [16]) but without any global clock. This asynchronous mode of operation seems to rely on surprisingly precise timing information that cannot be recovered by counting spikes over long temporal windows [53-57]. A recent study [58] suggests that acquiring temporally precise spikes from a visual sensor [28] provides up to 70% more information than conventional spikes generated from a frame-based acquisition as used in standard artificial vision thus enabling more accurate classification. Furthermore, the response times of these sensors are much faster thus enabling the detection of features that cannot even be sensed with regular cameras. This advantage combined with lower energy consumption in producing these precisely timed spikes will result in dramatically lower energy and high throughput products making next generation computing systems endowed with sensors that are more sensitive, efficient and fast in its processing capability.

Recent neuromorphic research chips have demonstrated the same mixed-mode model of computation as found in biological neural systems, namely local analog computation with global digital asynchronous communication [34-43]. Unlike traditional computing applications, these systems typically forego the determinism and reliability of synchronized digital computation. The systems loosely synchronize through the dynamics of time-based (spike) interactions (see [55]).

These neuromorphic architectures optimize for energy efficiency above all else, relaxing the performance, reliability, and deterministic

computational constraints of conventional ASIC (Application Specific Integrated Circuit) and CPU architectures. Through its 2011 acquisition of Fulcrum Microsystems [56], Intel now has the most advanced and commercially proven asynchronous design technology and tools in the industry [57] demonstrating best-in-class latency, bandwidth, and power metrics. Today this technology is being directed to more basic research including neuromorphic computing.

Challenges with Asynchronous Commputing: The gap in our understanding stems from how the brain resolves this temporal precision in the cortex and deeper brain areas (i.e., far away from the sensory inputs and motor outputs) given the amount of noise due to the stochastic nature of neural firing and also the interaction between millions of neural cells that are all affected by this muddling of precisely timed spikes with imprecise noise. One school of thought suggests that the cortical areas of the brain operate in a Bayesian fashion [48, 58] to help resolve this ambiguity. Another view point is that while the brain is not deterministic, neuronal variability may often be overestimated due to use of inappropriate reference times and also due to uncontrolled internal variables [59].

The argument is that the problem of reference time can be largely avoided by recording multiple neurons at the same time and looking at the statistical structures in relative latencies with the added advantage that they are insensitive to the variability that is shared across neurons.

While much of the work on spike latency based models emphasize synchrony of firing to enable robust learning in a sea of asynchronous spikes, an alternate viewpoint called polychronization [60] was recently developed. This alternate viewpoint relies on the observation that axonal conduction delays in the mammalian neocortex has a large range depending upon the type and location of neurons [61]. Using these delays, a model developed in [60] showed that when spikes originating from these set of neurons with an appropriate pattern of delays become coincident at a recipient neuron that neuron can generate a strong postsynaptic potential. This in turn can cause synapses to adapt thus forming a polychronous group between these neurons. The number of such polychronous groups far exceeds the number of neurons in the network resulting in an unprecedented memory capacity. New experimental work with the ability to measure large populations of neurons in the future will be needed to help resolve some of these differences in viewpoints in how the brain operates under asynchronous conditions.

<u>Plasticity and Learning</u>

The shallow and deep learning machine learning algorithms relies on the BP algorithm which is not a biologically plausible algorithm. In biology, the synaptic weights are adapted based more local influences. In particular, each synaptic weight between a pair of neurons modifies its strength based on the temporal correlations in spiking activity between the

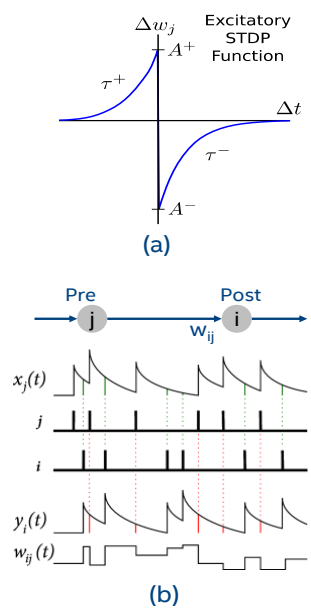


Fig. 8. (a) The temporally asymmetric spike timing dependent plasticity (STDP) rule. (b) The additive STDP changes (red for negative and green for positive) using this rule for synaptic weight w_{ij} as a function of the pre spikes j and post spikes i is illustrated here.

pair. This idea was originally postulated by Donald Hebb [62] and subsequently many forms of Hebbian plasticity have been discovered in the brain. The key observation is that brain exhibits plasticity of various kinds that operates at many spatial and temporal time scales and is continuously on. Thus, the brain is learning using these plasticity rules online.

Online learning enables a computer to directly and continuously interact with its environment thereby enabling a self-organized approach to understanding its world. If the interaction is such that there is feedback provided on the computers performance, then it allows the system to also continuously improve its ability to perform as per the expectations of its user/environment.

The learning mechanisms at short time scales are becoming clear and seems to primarily consist of Hebbian homosynaptic spike-timing dependent plasticity [63-65] (Fig. 8) and other heterosynaptic plasticity [66, 67]. Mechanisms for long-term plasticity appears to involve synaptic consolidation and maintenance [68]. There are other non-standard forms of plasticity including homeostatic plasticity [69], structural plasticity [70, 71] and short-term plasticity [72]. In addition there seems to be neuromodulatory influences [73, 74] on learning as well.

There have been many applications of online learning but using models that are not biologically plausible such as the Adaptive Resonance Theory [75] based models. These models exhibit online learning but address the question of stability by preventing any overwriting of already formed memories using a reset mechanism. As a result, the network does not scale gracefully with increasing number of classes to be stored. The other common online learning approaches are based on the deep learning and its variants [8-10, 76].

An example of an electronic chip exhibiting plasticity was the HRL spiking neuromorphic system. This system has short-term plasticity, spike-timing dependent plasticity and homeostatic plasticity all integrated into a single chip. The demonstration of online learning was performed using this hardware for the application described in [37]. However, in that application, only the short time scale based spike-timing dependent plasticity was used. As mentioned above, this chip experiment resulted in a ~50x improvement in energy and ~10x improvement in throughput compared to a conventional processor and learning happened in self-organized fashion without any human intervention.

Challenges with Plasticity and Learning:
One key gap in our understanding is how these various mechanisms interact during learning to form memories (but see [77] for a recent attempt). The other gap is in our understanding of how memories formed during online learning can be stabilized despite a constant barrage of new information being encountered by the brain and in the presence of noise. A related problem of an extreme kind is that of one-shot learning [78] where the system is only exposed to stimuli once and yet is capable of reliably responding to future instances of that one time event. Finally, there is no clear consensus on how the brain can learn complex episodes (spatiotemporal patterns) and then recall information from its learned memory in a robust and yet rapid fashion.

<u>Co-located memory and computation</u>
The average operating speed of neural circuits is in the order of a few Hz compared to current generation computers that operate in the tens to hundreds of GHz range. Evolution appears to have designed these low speed brain circuits to optimize on energy by avoiding having to transmit information long distance thus facing communication delays and interference. But unlike “weak co-location” in current generation computers where memory units are physically co-located with the processing units but serve the same decoupled functions, neural circuits in the brain are “strongly co-located” (Fig. 9) where in addition to physical proximity of memory and computation, there are learning and plasticity mechanisms co-located with memory and computing. In other words, in this

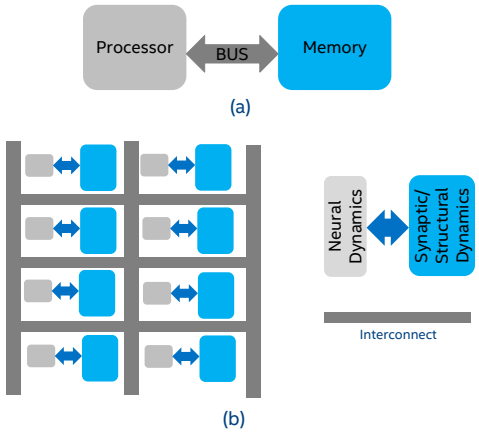


Fig. 9. (a) Memory and processing are fully decoupled in both physical location and functional role in a Von Neumann architecture (b) In a neuromorphic architecture, they are strongly co-located both physically and functionally where neuronal dynamics influences synaptic dynamics and vice versa. Furthermore the connectivity (1 neuron connects to 10,000 synapses) is very large for each neuronal processing unit.

system all memory resources, both for routing digital events and for storing synaptic and neural circuit parameters are tightly integrated with the synapse and neuron computing circuits. The changes in the timing of spike streams passing through a synapse effects changes in the synaptic strength which in turn affects the timing of the spikes and communication speeds to other neurons in the future. Furthermore, since this learning happens in a fully distributed fashion among many synapses, memory of events being processed is stored in the dynamics of the circuits rather than at a single location making it more robust to faults and tolerant to damage unlike current computational systems.

There have been several recent neuromorphic chips that have incorporated co-located memory and computation elements. NeuroGrid [38] designed at Stanford and the TrueNorth [35] designed at IBM are two examples of chips with a large number of processing synapses and neurons where different memory structures are distributed across the network (e.g., in the form of routing tables, parameters, and state variables). The ability of the shared synapses to integrate incoming spikes reproducing biologically plausible dynamics provide the system with computational primitives that can hold and represent the system state for tens to hundreds of milliseconds. However, the design choice to use linear synapses in the system excluded the possibility to implement synaptic plasticity mechanisms at each synapse, and therefore the ability to model on-line learning or adaptive algorithms without the aid of additional external computing resources.

The HRL neuromorphic chip [36] overcomes this bottleneck with a variety of plasticity mechanisms that support complex synaptic dynamics. However, that chip is limited in the number of neurons and synapses supported to address large real world tasks. The BrainScales [39] effort in Europe implements a wafer scale neural simulation platform with co-located neuron and synapse elements. However, in order to maximize the number of processing elements in the wafer, they chose to implement relatively small capacitors for modeling the synapse and neuron capacitances. As a consequence, given the large currents produced cannot achieve the long time-constants required for interacting with the environment in real-time. Rather, their dynamics are "accelerated" with respect to typical biological times by a factor of 10^3 or 10^4. This has the advantage of allowing very fast simulation times which can be useful e.g., to investigate the evolution of network dynamics over long periods of time, once all the simulation and network configuration parameters have been uploaded to the system. But it has the disadvantage of requiring very large bandwidths and fast digital, high-power, circuits for transmitting and routing the spikes across the network.

Applications of Neuromorphic Computing

By enabling autonomous computing for a low energy cost, neuromorphic computing hardware could play a big role in the future of large scale data centers and exoscale computing systems. The main contribution here will come from a mixed mode of

computing where computing sub-tasks that can tolerate approximate solutions will be enabled by energy efficient neuromorphic hardware that is integrated within a Von Neumann machine to provide large gains in energy efficiency.

The energy efficient neuromorphic hardware will also create a whole new set of applications from intelligent mobile device to intelligent and efficient IoT devices and systems. This is because neuromorphic chips offer size, weight and area efficiencies as well. Furthermore, since these systems can learn online, they can enable these devices to operate in a self-organized fashion without any manual intervention.

The complex spatiotemporal learning capability will also enable autonomous self-driving cars and drones as well as provide novel cyber security solutions. This online learning capability will also usher in a whole new class of sensors such as cameras that can learn about the pictures it may capture and label the pictures when they similar objects in the future. They will also enable novel home appliances and health monitoring devices that could learn to customize the user experience to their tastes and state of health respectively.

Since these system are capable of learning in a supervised fashion as well, they will usher in novel computing frameworks that learn the physics of complex systems by observation and then replace these systems with matching performance while being faster and much more energy efficient in comparison.

CONCLUSIONS

In this paper, we discussed the field of machine learning by highlighting the salient aspects of the past and present. We believe that the future will be based on neuromorphic computing. With architectural features including asynchronous processing, distributed memory and computation and spike based representations combined with online learning capability, this next generation of machine learing will usher in a whole new computing paradigm and potentially revolutionize our way of living.

REFEENCES

1. A. M. Turing, “On computable numbers with an application to the Entscheidungs problem,” *Proc. London Math. Soc.*, vol. 2, no. 42, pp. 230–265, 1936.
2. C. Shannon and W. Weaver, *The Mathematical Theory of Information.* Urbana, IL: Univ. Illinois Press, 1949.
3. W. McCulloch and W. Pitts, “A logical calculus of ideas immanent in nervous activity,” *Bulletin of Mathematical Biophysics*, 5(4), pp. 115-133, 1943.
4. F. Rosenblatt, "The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain". *Psychological Review* 65 (6): 386–408, 1958.
5. M. Minsky, S. Papert. *An Introduction to Computational Geometry*. MIT Press, 1969. ISBN 0-262-63022-2.
6. P. J. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Ph. D. Thesis, Harvard University, 1975.
7. D. E. Rumelhart, G. E. Hinton and R J. Willliams, “Learning Representations by back-propagating errors”, *Nature* 323, pp. 533-536, 1986.
8. Y. Lecun, Y. Bengio and G. Hinton, “Deep Learning”, *Nature*, vol. 521, pp. 436-444, 2015.
9. G. E. Hinton, S. Osindero, Y. W. A. Teh, “A fast learning algorithm for deep belief nets”, *Neural Computation*, 18, 1527-1554, 2006.
10. G. E. Hinton and R. Salakhutdinov, “reducing the dimensionality of data with neural networks,” *Science*, 213, 504-507, 2006.
11. K. Fukushima and S. Miyake, “Neocognitron: a new algorithm for pattern recognition tolerant to deformations and shifts in position,” *Pattern Recognition*, 15, 455-469, 1982.

12. Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, Gradient based learning applied to document recognition. *Proc. of IEEE* 86, 2278-2324, 1998.
13. C. A. Mead, Neurmorphic Electronics Systems, *Proc. of IEEE*, vol. 78, no. 10, pp. 1629-1636, 1990.
14. A. B. Barrett, M. C. W. van Rossum, "Optimal Learning Rules for Discrete Synapses," *PLoS Computational Biology* 4(11): e1000230. doi:10.1371/journal.pcbi.1000230, 2008.
15. A. Renart, J. De la Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, K. D. Harris, "The asynchronous state in cortical circuits," *Science*, 327, pp. 587–590, 2010.
16. G. Buzsaki, Rhythms of the Brain, Oxford University Press, USA, 2009.
17. Edelman, G. M., *The Remembered Present: A Biological Theory of Consciousness*, Basic Books, NY, 1989.
18. W. J. Freeman, "A field-theoretic approach to understanding scale-free neocortical dynamics," *Biological Cybernetics*, 92(6):350-359, 2005.
19. D. J. Graham, D. J. Field, "Sparse coding in the neocortex". In: Kass JH, editor. *Evolution of the nervous system*, Vol III. Oxford: Academic Press. pp. 181–187, 2007.
20. R. Van Rullen, and R. Guyonneau, and S. J. Thorpe, "Spike times make sense", *Trends in Neuroscience*, 28:1-4, 2005.
21. N. Srinivasa, and Q. Jiang, "Stable learning of functional maps in self-organizing spiking neural networks with continuous synaptic plasticity, *Front. In Comp. Neuroscience*, doi: 10.3389/fncom.2013.0001, 2011.
22. N. Srinivasa, and Y. K. Cho, "Unsupervised discrimination of patterns in spiking neural networks with excitatory and inhibitory synaptic plasticity," *Front. In Comp. Neuroscience*, doi:10.3389/fncom.2014.00159, 2014.
23. Y. Cao, Y. Chen, and D. Khosla, "Spiking Deep Convolutional Neural Networks for Energy-Efficient Object Recognition," *International Journal of Computer Vision*, vol. 113, no. 1, pp. 54–66, 2014.
24. E. Hunsberger, and C. Eliasmith. "Spiking Deep Networks with LIF Neurons." arXiv:1510.08829v1, 2015.
25. M. Coath, S. Sheik, E. G. Chicca, G. Indiveri, and S. L. T. Wennekers, "A robust sound perception model suitable for neuromorphic implementation," doi:10.3389/fncom.2013.00278, 2014.
26. P. Joshi, and W. Maass, "Movement generation and control with generic neural microcircuits, in Proc. of BIO-AUDIT, pp. 16-31, 2004.
27. M. Oubatti, P. Levi, M. Schanz, T. Buchheim, "Velocity control of an omnidirectional robocup player with recurrent neural networks," In Proc. of Robocup Symposium, pp. 691-701, 2005.
28. P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128×128 120 db 15 μs latency asynchronous temporal contrast vision sensor," *IEEE J. Solid-State Circ*. 43, 566–576, 2008.
29. S. Liu, A. Van Schaik, B. Minch, and T. Delbruck, "Event-based 64-channel binaural silicon cochlea with Q enhancement mechanisms," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, 2027–2030, 2010.
30. P. O'Connor, D. Neil, S. C. Liu, T. Delbruck, and M. Pfeiffer, "Real-time classification and sensor fusion with a spiking deep belief network," *Front. in Neuromorphic Engineering*, http://dx.doi.orgns/10.3389/fnins.2013.00178, 2013.
31. G. Orchard, and R. E. Cummings, "Bioinspired Visual Motion Estimation," *Proc of the IEEE*, vol. 102, 10, 1520-1536, 2014.
32. Z. Ni, B. Bolopion, J. Agnus, R. Benosman and S. Regnier, "Asynchronous Event-Based Visual Shape Tracking for Stable Haptic feedback in Microrobotics. IEEE Trans. On Robotics, Vol. 28, no. 5, pp. 1081-1089, 2012

33. T. K. Horiuchi, "A spike-latency model for sonar-based navigation in obstacle fields," *IEEE Trans on Circuits and Systems – I Regular Papers*, vol. 56, no. 11, pp. 2293-2401, 2009.
34. R. J. Vogelstein, U. Malik, E. Culurceillo, G. Cauwenberghs, R. Etienne-Cummings, "A multichip neuromorphic system for spike-based visual information processing," *Neural Computation* 19:2281-2300, 2007.
35. P. A. Merolla, J. A. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, et al, "A million spiking neuron integrated circuit with a scalable communication network and interface," *Science*, Vol. 345, Issue 6197, pp. 668-673, 2014. DOI: 10.1126/science.1254642.
36. J. Cruz-Albrecht, T. Derosier, and N. Srinivasa, "Scalable neural chip with synaptic electronics using CMOS integrated memristors, *Nanotechnology, Special Issue on Synaptic Electronics*, vol. 24, 384011 (11pp), doi:10.1088/0957-4484/24/38/384011, 2013.
37. http://www.technologyreview.com/news/532176/a-brain-inspired-chip-takes-to-the-sky/
38. B. V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran , J. M. Bussat, R. Alvarez-Icaza, J. V. Arthur, P. A. Merolla, and K. Boahen, "Neurogrid: A Mixed-Analog-Digital Multichip System for Large-Scale Neural Simulations," *Proceedings of the IEEE*, vol 102, no 5, pp 699-716, 2014.
39. *Brain-inspired multiscale computation in neuromorphic hybrid systems (BrainScaleS)* FP7 269921 EU Grant, 2011–2015.
40. G. Indiveri, E. Chicca and R. Douglas, A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity, *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 211-221, 2006.
41. R. Vogelstein, U. Mallik, J. Vogelstein and G. Cauwenberghs, Dynamically Reconfigurable Silicon Array of Spiking Neurons With Conductance-Based Synapses, *IEEE Transactions on Neural Networks*, vol. 18, no. 1, pp. 253-265, 2007.
42. N. Qiao, H. Mostafa, F. Corradi, M. Osswald, F. Stefanini, D. Sumislawska and G. Indiveri, A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses, *Frontiers in Neuroscience*, vol. 9, no. 141, 2015.
43. S. B. Furber, F. Galluppi, S. Temple, L. A. Plana, "The SpiNNaker Project. *Proceedings of the IEEE*: 1. doi:10.1109/JPROC.2014.2304638, 2014.
44. A. M. Bastos, W. M Usrey, R. A. Adams, G. R. Mangun, P. Fries, K. J. Friston, "Canonical microcircuits for predictive coding," *Neuron* 76(4):695-711, 2012.
45. R. P. Rao, and D. H. Ballard, "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects," *Nature Neuroscience*, 2(1):79-87, 1999.
46. W. Maass, "Noise as a resource for computation and learning in networks of spiking neurons," *Special Issue of the Proc. of the IEEE on Engineering Intelligent Electronic Systems based on Computational Neuroscience*, 102(5):860-880, 2015.
47. D. Kappel, S. Habenschuss, R. Legenstein, R., and W. Maass, "Network plasticity as Bayesian inference," *PLOS Computational Biology*, 11(11):e1004485, 2015.
48. B. Nessler, M. Pfeiffer, L. Buesing, and W. Maass, "Bayesian computation emerges in generic cortical microcircuits through spike-timing-dependent plasticity," *PLoS Computational Biology*, vol. 9, no. 4, p. e1003037, 2013.
49. D. S. Reich, F. Mechler, and J. D. Victor, "Independent and redundant information in nearby cortical neurons," *Science*, 294(5551), 2566-2568, 2001.
50. P. Kara, P. Reinagel, and R. C. Reid, "Low response variability in simultaneously

recorded retinal, thalamic and cortical neurons," *Neuron*, 27, 635-646, 2000.
51. B. Haider, M. R. Krause, A. Duque, Y. J. Yu, J. Touryan, J. A. Mazer, D. A. McCormick, "Efficient discrimination of temporal patterns by motion sensitive neurons in primate visual cortex," *Neuro*n, 65, 107-121, 2010.
52. G. T. Buracas, A. M. Zador, M. R. DeWeese, and T. D. Albright, "Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex," *Neuron*, 20, 959-969, 1998.
53. S. Panzeri, N. Brunel, N. Logothetis, and C. Kayser, "Sensory neural codes sing multiplexed temporal scales," *Trends in Neuroscience*, 33, 111-120, 2010.
54. H. Akolkar, C. Meyer, Z. Clady, O. Marre, C. Bartolozzi, and S. Panzeri. "What can Neuromorphic Event-Driven Precise Timing Add to Spike-Based Pattern Recognition?," *Neural Computation*, 27, 561-593, 2015
55. J. Binas, G. Indiveri and M. Pfeiffer, Spiking Analog VLSI Neuron Assemblies as Constraint Satisfaction Problems. arXiv, 2015.
56. The Register, "Intel snaps up network chipper Fulcrum," [Online]. Available: http://www.theregister.co.uk/2011/07/19/intel_acquires_fulcrum_microsystems/.
57. R. Wilson, "Fulcrum," EE Times, 6 9 2011. [Online].Available: http://www.eetimes.com/document.asp?doc_id=1279057.
58. L. Büsing, J. Bill, B. Nessler, and W. Maass, "Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons," PLoS Comput. Biol. 7:e1002211. doi: 10.1371/journal.pcbi.1002211, 2011.
59. P. Tiesinga, J. M. Fellous, and T. J. Sejnowski, "Regulation of spike timing in visual cortical circuits," *Nature Review Neuroscience*, 9, 97–107, 2008.
60. E. M. Izhikevich, "Polychronization: Computation with Spikes," *Neural Computation* 18, 245-282, 2006.
61. H. A. Swadlow, and S. G. Waxman, Observations on impulse conduction along central axons. *PNAS* 72, 5156-5159, 1975.
62. D. Hebb, *The Organization of Behavior*. New York: Wiley, 1949.
63. H. Markram, J. Lubke, M. Frotsche, and B. Sakmann, "Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs," *Science* 275, 213-215, 1997.
64. G. Q. Bi and M. M. Poo, "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength and postsynaptic cell type," *Journal of Neuroscience*, 18, 10464-10472, 1998.
65. P. J. Sjostrom, G. G. Turrigiano, and S. B. Nelson, "Rate, timing and cooperativity jointly determine cortical synaptic plasticity," *Neuron* 32, 1149-1164, 2001.
66. M. Chistiakova, N. M. Bannon, M. Bazhenov and M. Volgushev, "Heterosynaptic plasticity: multiple mechanisms and multiple roles," *Neuroscientist* 20, 483-498, 2014.
67. G. S. Lynch, T. Dunwiddie, and V. Gribkoff, "Heterosynaptic depression: a postsynaptic correlate of long-term potentiation," *Nature* 266, 737-739, 1977.
68. U. Frey and R. G. M. Morris, "Synaptic tagging and long term potentiation," *Nature* 385, 533-536, 1997.
69. G. G. Turrigiano, "Homeostatic synaptic plasticity: local and global mechanisms for stabilizing neuronal function. *Cold Spring Harbor Perspective in Biology*, 4, a005736, 2012.
70. A. Stepanyants, P. R. Hof, D. B. Chklovskii, "Geometry and structural plasticity of synaptic connectivity," *Neuron* 34, 275-288, 2002.
71. J. T. Trachtenberg JT et al, "Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex," *Nature* 420, 788-794, 2002.
72. H. Markram, Y. Wang, and M. Tsodyks, "Differential signaling via the same axon

of neocortical pyramidal neurons," *Proc. Natl Acad. Sci USA*, 95 5323-5328, 1998.
73. V. Pawlak, J. R. Wickens, A. Kirkwood, and J. N. D. Kerr, "Timing is not everything: neuromodulation opens the STDP gate," *Front. Synaptic Neuroscience* 2, 146, 2010.
74. W. Schultz, P. Dayan, and P. R. Montague, "A neural substrate of prediction and reward," *Science* 275, 1593-1599, 1997.
75. G. A. Carpenter, and S. Grossberg, "Adaptive resonance theory," *Encyclopedia of Machine Learning and Data Mining.* C. Sammut and G. Webb, Eds. Berlin: Springer-Verlag, 2015.
76. J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," *Neural Networks* 61: 85–117, 2015.
77. F. Zenke, E. J. Agnes, and W. Gerstner, "Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks," *Nature Comm*. Doi: 10.1038/ncomms7922, 2015.
78. P. D. Balsam, M. R. Drew, et al, "Time and associative learning," *Comparative Cognition and Behavior Review*, 5, 1-22, 2010.

MERGING CONSUMPTION DATA FROM MULTIPLE SOURCES TO QUANTIFY USER LIKING AND WATCHING BEHAVIOURS

Sashikumar Venkataraman, Craig Carmichael

Rovi Corporation

Abstract

We describe an integrated platform that aggregates consumption data from multiple sources towards building a unified model for collaborative filtering for more accurate user and content representations. The goal is to provide a framework that combines various signals spanning explicit ratings, implicit information of watching behaviors and meta-content information in a single model that potentially goes beyond the usual goal of maximizing consumption and incorporates metrics that capture "likeness" and "discovery". We also feed the usage data back into meta-content to determine more accurate content representations that aid in targeting content-based recommendations more effectively.

INTRODUCTION

Recommendation systems are becoming an increasingly key component in many media and entertainment content retrieval systems by providing a powerful, efficient method for users to easily sift through large catalogs of media content and finding valuable programming [1-3]. Within the last couple of decades, several algorithms have been developed that leverage metadata, such as information on programs, casts and genres, and user consumption data to provide users with more targeted content recommendations through these recommendation systems [4-5]. However, in most systems, much of the viewership data is typically implicit in nature, and models that are based on over-simplified estimations of viewing behavior are often not wholly comprehensive and intuitive. A user is further predisposed to watch certain shows and channels often based on pure habit, content they are already aware of, or the sheer popularity of a show, and these aspects are usually unaccounted in the recall score of a recommender system.

In this paper we discuss a unified framework that aggregates consumption data from multiple sources and fuses them with meta-content to create an efficient recommendations framework capable of both significant discovery and high precision. First, we begin with the cold start problem where no consumption data is available, and create a baseline recommendation model that applies word-to-vector [6] factors solely from metadata content. An aggregation process is used to accumulate these word-level vectors to content level factors for each show and channel potentially consumed. Next, we derive the usage factors for each media asset considering both explicit and implicit ratings from various sources. While the explicit ratings provide a more direct notion of "likeness", the implicit signals only provide a measure of watching behavior. We discuss methods to correlate these notions of likeness and watching attributes in a more formal way.

A central part of the framework for merging consumption data from multiple sources is the Rovi Knowledge Graph that incorporates factual information of all 'known' or 'named' things. This includes all movies and TV shows, music albums and songs, as well as all known people such as actors, musicians, celebrities, music bands, known companies and businesses, places, sports teams, tournaments and players, etc. All the facts pertaining to these entities are synthesized from multiple publicly available

sources such as Wikipedia, Freebase, and many others and correlated so as to create a unique smart tag (with a unique identifier) to represent each entity in the Rovi Knowledge Graph. The main utility of the knowledge graph for our problem is in aiding to merge information across different data sources. We have a merge system that can take any data source and does a best-effort to merge the entities in that data source with the underlying knowledge graph. Such a system makes the aggregation easier since different data sources have variations in referring to entities and carry different level of meta-information.

One often encounters rich meta-content associated with media assets, such as genre, keywords, and description. However, the relevance or weight of each individual piece of meta-content (for finding similar movies or recommendations) is often lacking, missing or wrong due to multiple sources, algorithms, and/or manual entry. For example, a show is a comedy but exactly how funny and how it impacts other comedies is more of a viewing sentiment. Usage data, on the other hand, provides a different kind of information in conveying what programs co-occur in watching behavior across users. Analysis of usage data involves very different techniques and algorithms from those for analyzing meta-content. There has been some effort in the context of recommendations wherein one uses meta-content to filter in the post-process of the collaborative filtering algorithm or mixes results coming from collaborative filtering and meta-content algorithms. However, no prior efforts make use of usage data to better understand metadata relevance and enrich meta-content directly. It would be desirable if usage data along with the implicit/explicit ratings of users could be leveraged to determine the relevant weights of different pieces of meta-content.

OVERALL ARCHITECTURE

We begin by describing the key steps involved in merging deep and dynamic metadata with usage data across various data sources. These steps are described in brief below and will be discussed in further detail in the following sections.

Step 1: We first merge the assets from the data source into the central knowledge graph, to enable infusing of usage and meta-information to and from the knowledge graph. This step serves a dual purpose. Firstly, it helps in augmenting the meta-content of the assets in terms of keywords, genres and deep-descriptors from the knowledge graph and hence aids in getting more accurate factors derived from meta-content representations. Secondly, it aids in merging the usage factors from this data source with usage factors from other data-sources that are also merged with the central knowledge graph.

Step 2: Using word-representations, we next determine the meta-content factors corresponding to the media-assets for cold-start baseline with no or minimal usage data. The word2vec model developed in [1] is used to determine the word-representations corresponding to each of the meta-content information and these are aggregated to form a vector for each asset in K-dimensional vector space.

Step 3: Next, we build a model that merges implicit and explicit information from multiple data sources and fuse them into the meta-content factors to create more accurate asset representation. This naturally results in a model to estimate "likeness" from user-watching behavior even in absence of explicit information.

Step 4: Finally we feed the usage data back into meta-content in the form of coefficient weights of each individual meta-content factor involved in each asset. This enables the

creation of more accurate representation of the individual meta-content factors that further increase the precision in content-based recommendations.

MERGING INTO KNOWLEDGE GRAPH

The Rovi Knowledge Graph is a dynamic system that has been created by synthesizing multiple metadata sources and evolving and refreshed continuously over time. Each smart tag in the Knowledge Graph has rich meta-content built by a combination of manual tagging and automatic aggregation from multiple sources along with several machine-learning algorithms.

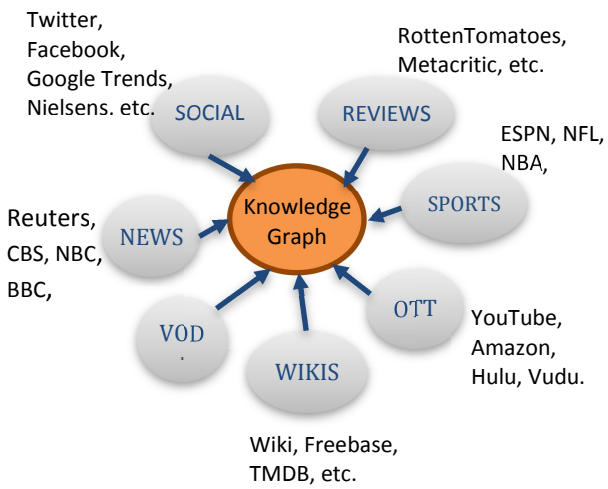


Fig 1: Merging of assets from multiple data sources into a central knowledge graph

Fig 1 below shows the merging of information from various sources into the central knowledge graph. While certain meta-content fields, such as cast-members, roles, and release year are unambiguous, while other fields such as genres and keywords are more subjective. These fields are assigned using a combination of manual tagging and automatic aggregation from multiple sites like Wikipedia and Freebase. Genres can also be assigned by tagging directly from keywords gleaned from descriptions or plots. This is done using machine-learning algorithms by first determining the set of keywords associated with a genre, using movies known to have that genre, and then predicting them on other movies that have a strong overlap in keywords with the genre keywords.

The core part of building the Knowledge Graph is a "merge" function that takes in any source and tries to map the entities in that source with the entities in the existing graph. Whenever a new entity, such as movie or personality, is merged with an existing smart tag, the metadata corresponding to the smart tag gets augmented from the entity. If the new entity does not get merged with any smart tag, then a new smart tag is created for the entity in the KG. An important aspect of the merging is the allowance of slight variations in the meta-content fields between the two assets. For example, the titles can differ slightly due to several reasons. One reason for the difference in titles is the fact that some sources put the season number and episode numbers in the title, while other sources only put the episode title in the title. Sometimes they may differ due to some lexical error or absence of common words like articles. Release year could also vary by one or two units, and sometimes cast members could be be missing. All such variations are considered during the merge process by considering all the fields simultaneously and coming with a combined match score considering all fields. Following are 2 examples of approximate matching with inexact titles which got matched due to matching of other fields:

BUILDING WORD-REPRESENTATIONS FROM META-DATA

We now describe the step of determining the baseline factors for the media-assets from meta-content. These factors are especially useful in cold-start situations with no or minimal usage information. Over time the usage information from multiple data sources play a key role in how the asset factors evolve; nonetheless the meta-content factors remain valuable to bias recommendations

towards more meta-content oriented discovery without losing high precision. In our collaborative filtering algorithm, we represent meta-content information of each unique media-asset as a weighted combination of individual meta-content pieces, such as genre, category, and keywords. Each individual piece of meta-content can be represented as a vector in a K-dimensional vector space (K is usually 100-300). Each asset vector is then a weighted sum of individual vectors and hence also a vector in this space. The individual vectors could have been determined independently by other known algorithms based on co-occurrences of terms in large corpus (such as word2vec [6]).

While a naïve algorithm in determining the asset vector could be to just average over all the vectors of keywords and genres, there are several problems in this approach. Firstly, not all keywords are of same importance and some keywords may be noisy and unrelated to the theme of the movie. Secondly, the genre vectors are not readily available in the word2vec model and needs to be explicitly computed from the movies. To address the first problem, we process the movie keywords to form keyword clusters based on the closeness of the keyword-vectors and then filter those keywords that don't fall in a significant weighted cluster. For the second problem, we treat a genre as a collection of movies and hence cluster all the keywords across all the movies belonging to that category. The top keyword clusters found by this clustering are then used to find the final genre vectors. After we determine the keyword and genre vectors, we can then find the final asset vectors as a linear combination of the corresponding keyword and genre vectors. While these asset vectors are a good starting point in our collaborative filtering model, it may not be the most accurate representation due to the lack of knowledge of the genre and keyword cluster weighting coefficients. We will discuss techniques to optimally determine those weights in a later section.

MULTI DATA-SOURCE COLLABORATIVE FILTERING MODEL

As mentioned earlier, one of the main challenges in doing collaborative filtering is the lack of explicit ratings in dealing with usage data from many data sources. Most of this usage data results from linear TV watching behavior where the users are limited in content choice and no explicit feedback is provided that indicates how much the user liked a particular show or movie. We also get (to a lesser extent) usage data that involves VOD (video on-demand). Though the signals from VOD (and DVR) do provide a greater correlation to user liking behavior, they too are not as accurate as explicit ratings due to the limitation in the number of media assets in typical VOD catalogs. At best, they capture the kinds of genres and categories that the user typically watches, but fails to capture more fine-grained likeness or dis-likeness within a category. So the most valuable information that we get is the usage data from a couple of sources that involve explicit ratings. However, the amount of this kind of usage data is a lot less compared to the usage data relating to linear TV and VOD. The challenge then is to build a model that can unify all these forms of usage data and provide a model to carry over information from data source to augment the usage factors in another source.

There are several models existing that convert the ratings (explicit or implicit) to similarities between media assets. These include Pearson correlation coefficient, cosine similarity, log-likelihood and jaccard coefficient. Yet another interesting coefficient is a notion of probabilistic similarity as discussed in [7] referred as ProbSim in sections below. During merging ratings from multiple sources, the data sources explicit ratings are given a higher weight than data

sources with implicit ratings. We also treat VOD/DVR watching as a more explicit intent and give higher weight to those data points. The final result of merging all these forms of usage data is a NxN similarity matrix where each element *(i, j)* gives the similarity between the items *i* and *j*. The similarity matrix can either be created with only explicit ratings (referred as *XSim(i, j)*) or implicit ratings (referred as *ISim(i, j)*) or combining both (referred just as *Sim(i, j)*).

Our next step is to determine the factors for the media-assets that most closely match these similarities derived from the usage data. The starting point for the asset vectors are the factors determined in the previous section from the word-representations corresponding to the meta-content. The asset vectors are then let to vary so as to match the usage similarities as much as possible in an iterative fashion. For every item-item pair, the corresponding item factors are made to come closer or away from each other based on how much the cosine distance between the asset vectors match the computed item-item similarity. These final asset vectors represent a more accurate representation that reflects usage based similarity and simultaneously remaining as close to the meta-content representation as possible. Hence, these form an ideal representation for hybrid recommendations.

We next use these similarities to create a model to determine explicit user likeness from implicit watching behavior. While explicit ratings were unambiguously absorbed in the similarity computation, implicit watching were accompanied with several signals that needed to be translated to some form of implicit rating in a systematic fashion. Some of the these signals include duration of watching, number of times/episodes watched, number of similar items user has watched around that asset, the price the user paid for watching, the average rating users have rated that particular item in other data sources, popularity of the asset, etc. The goal is then to build a model that uses these signals and uses a model to create an optimal expression for "implied rating" or "likeness". For this purpose, we create a model that translates the implicit ratings to similarities and then match them with the similarities obtained from explicit ratings. Figure 2 shows the model that is used for this purpose.

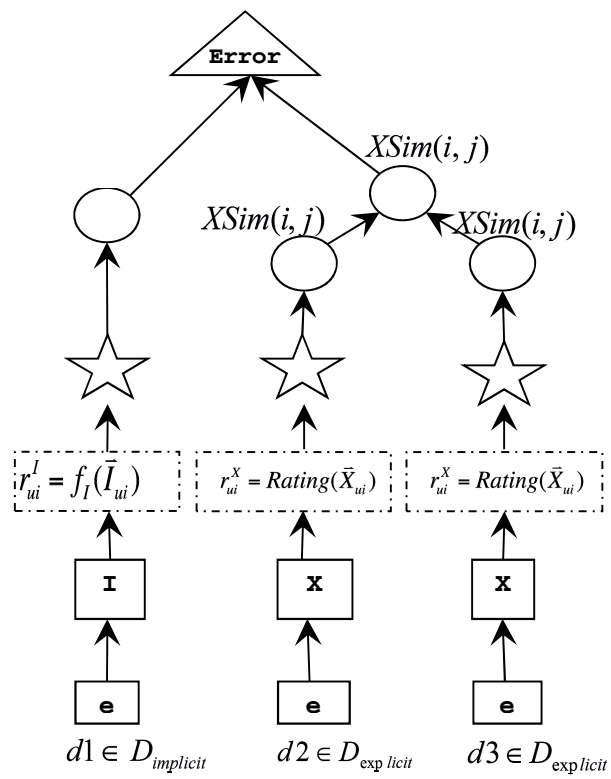


Fig 2: Merging usage from multiple data sources with explicit and implicit ratings

In the above figure, we aggregate the raw user event data in the first or bottommost layer, *L1*. Examples of explicit indicators include specific data that contain explicit information such as ratings on the scale of 1-5 in increments of a half-star, or a like/not (0/1) binary indicator. This would comprise an explicit (input) vector $X_{ui}(d)$. Example of implicit indicators for user-item interaction would include implicit watching signals such as the ones described earlier. This would comprise an implicit vector *I*. The next layer

L2 contains a mapping from consumption details provided by *L1* (both $\vec{I}_{ui}(d)$ and $\vec{X}_{ui}(d)$) to the output to a form of preference or likeness of a user *u* to item *i*, $r_{ui}(d)$ in the data source *d*. If the data set d involved is an explicit one, the mapping is often straightforward. However, for implicit data sets, the mapping is more challenging and the mapping would be in the form of a (preference) model taking various forms $f_I(\vec{I}_{ui}(d^I),\vec{W})$ that ultimately depends on a set of trainable weights $\vec{W}$. Different functions can be used for various embodiments of f_I which represent trainable $\vec{W}$ in different ways. For example, we can use a linear estimator with a sigmoid function at the output node, a general regression neural net, a random forest, or various others.

The predicted ratings are then passed to a similarity layer *L3*, that takes the preference detail estimates for each (u,i,d) across all common media assets and users, and produces a similarity estimate between media assets i and j. While different kinds of similarities (discussed earlier) can be used, we use similarities based on a weighted Pearson coefficient as mentioned below:

$$Sim(i,j,d)=\frac{\sum_{u\in(i,j)}^{U(d)}\left(r_{ui}^{X}(d)-\bar{r}_i^{X}(d)\right)\left(r_{uj}^{X}(d)-\bar{r}_j^{X}(d)\right)}{\sqrt{\sum_{u\in(i,j)}^{U(d)}\left(r_{ui}^{X}(d)-\bar{r}_i^{X}(d)\right)^2}\sqrt{\sum_{u\in(i,j)}^{U(d)}\left(r_{uj}^{X}(d)-\bar{r}_j^{X}(d)\right)^2}}$$

It is assumed that whether similarities are determined using implicit or explicit data, they should roughly equate to the same values per (i,j) pair across multiple data sets. So first we obtain an estimate *XSim(i,j)* based on the usage data among data sources with explicit ratings. This is then compared with the data sources with implicit ratings and a difference is then used in the final layer *L4* to compute the error between the two observed similarities as shown below:

$$Error=\sum_{d=1}^{Dimplicit}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(XSim(i,j)-ISim(i,j,d)\right)^2$$

Then we can take the error, and propagate it backwards, layer by layer until the derivatives are estimated across the trainable weights $\vec{W}^I$. The iterations are performed until the error is minimized to below a certain threshold. We then use those weights as the optimal coefficients to combine the implicit signals and create an implicit rating.

FUSING USAGE INFORMATION INTO META-CONTENT

The collaborative filtering model used here is referred as Weighted Vector Collaborative Filtering (WVCF) where the meta-content information of each media-asset is represented as a weighted combination of individual meta-content pieces, such as genre, category, and keywords and each individual meta-content piece is treated as a vector in a K-dimensional vector space. A meta-content similarity of the two assets is then modeled as a function of these individual meta-content pieces of information (such as a dot product). However, the weight coefficients for each individual piece of information are usually not known apriori and our goal is to use usage information to best predict these weights. Once these weights are determined, we can create more accurate item-item similarities and thereby more accurate recommendations.

Importantly, the WVCF baseline model consisting of a single trained vector per asset can be broken up into two finely tuned predictors in the modeling pipeline, one that targets aspects of watching (WVCF-watch), and one that targets liking characteristics of the user (WVCF-like). Later in the results section is an example of recall performance using Gradient Boosted Trees as a local corrector to target both sides of precision (liking) and recall (watching).

As described in the previous section, the usage information is separately modeled to produce item-item similarity wherein items watched together and similarly evaluated/rated (common sentiment) across multiple users have better usage-similarity. This similarity also takes into account the user's sentiment along with viewing information (i.e. both watched and similarly liked/reviewed as opposed to only watched). All similarities are in some sense based in part on co-occurrence; usage-based similarity is the co-occurrence of users watching the same program, while meta-content similarity is the co-occurrence of some meta-content (crew or genre or keyword). The strength of co-occurrence in the meta-content case is based on the weights of the individual meta-content information within that asset, while the strength of co-occurrence in usage-based similarity is governed by the user sentiment/rating.

The system then tries to align the media asset vectors as close as possible to the usage-based similarities. An error function is then constructed that compares the modeled meta-content similarity to the usage-based similarity (based on co-occurrence combined with sentiment factors).

$$E = \sum_{ij} \left(s_{ij} - m_{ij} \right)^2$$

where s_{ij} denotes the observed "sentimental" similarity between items i and j (as determined from usage data discussed in previous section) and m_{ij} denotes the modeled similarity based on the asset meta-content vectors. The objective in this error minimization is to explain the observed sentimental similarities between many asset pairs with a model that also has deeply embedded metadata. The modeled similarity m_{ij} defined by the dot product of asset vectors $\vec{a}_i$ and $\vec{a}_j$ over all the latent factors:

$$m_{ij} = \sum_f a_{if} a_{jf}$$

We next break down the asset vectors into individual metadata components consisting of keywords and genres:

$$a_{if} = v_i^{genres} w_{if}^{genres} + v_i^{key} w_{if}^{key}$$

$$w_{if}^{genres} = \sum_{g \in i}^{genres\ in\ i} v_g^{genre} w_{gf}^{genre}$$

$$w_{if}^{keywords} = \sum_{k \in i}^{keywords\ in\ i} v_k^{keyword} w_{kf}^{keyword}$$

where w_{if} denotes the total genre and keyword factors, and v_i denotes the corresponding weights of those fields. The model continues to extend so far as metadata is available. Unknown components can also be added to address when metadata is lacking or unavailable. We then break down the genre and keyword vectors further into factors corresponding to individual genre and keywords. Next we minimize the error function using stochastic gradient descent that changes the weights of the individual meta-content components so that the net error between the meta-content similarities and usage-based similarities is minimized. After iterating over all the usage data, the individual meta-content weights are stored as the best predictors for the corresponding meta-data relevance for the media-asset.

RESULTS

Table 1 below shows 10 examples of the most similar shows based on the cosine similarity between sentiment vectors prior to training the model using only a small subset of the terms in each a_{if}. This is using only raw source combined with weighted word vectors without the use of metadata filtering based on advanced tags, moods or deep descriptors. So the starting point a_{if} used for training each media asset vector is already capable of overcoming cold-start issues.

Table 2 below and Figure 3 demonstrate the trained accuracy of our model for various number of meta-content factors. Recall@K for various values of K was computed over a VOD data set. Note the objective here was to show that precision and recall can be contained within a vector rather than a similarity matrix without significant loss, while holding metadata seamlessly across usage spaces as vectored sub-components. Further improvements are seen using WVCF-watch and WVCF-like (beyond the vector). With the vector, other training was based on 15K assets and 1 million users over the span of a year. It appears little is lost in performance when comparing the current method to a purely usage based approach such as ProbSim, which has proven to be a top method in recommender systems space in terms of precision/recall.

Embedded into our model is the potential to also normalize to a space that can be consistent across multiple usage data sets. So when recommending an asset that is either completely new or not yet popular, this model will tend to significantly outperform purely usage-based approaches.

Title 1	Title 2
Top Chef	Top Chef: Texas
Scooby-Doo! Pirates Ahoy!	Scooby-Doo and the Alien Invaders
28 Weeks Later	Resident Evil: Apocalypse
Pirates of the Caribbean: On Stranger Tides	Pirates of the Caribbean: The Curse of the Black Pearl
Kansas City SWAT	Detroit SWAT
Stuart Little 2	Stuart Little
Modern Vampires	Vampire in Brooklyn
Mad Max	Death Race 2000
Godzilla vs. Mechagodzilla II	Terror of Mechagodzilla
Attack of the Killer Tomatoes	Return of the Killer Tomatoes

Table 1: Example most similar shows using meta-content tags before training usage data

Model	Recall@K				
	K=5	K=10	K=25	K=50	K=100
Probsim	0.192	0.286	0.445	0.551	0.681
WVCF, F=20	0.163	0.249	0.365	0.452	0.575
WVCF, F=50	0.163	0.266	0.392	0.495	0.626
WVCF, F=100	0.186	0.289	0.425	0.528	0.664
WVCF, F=300	0.206	0.296	0.432	0.551	0.678

Table 2: Recall performance of WVCF vs Probsim

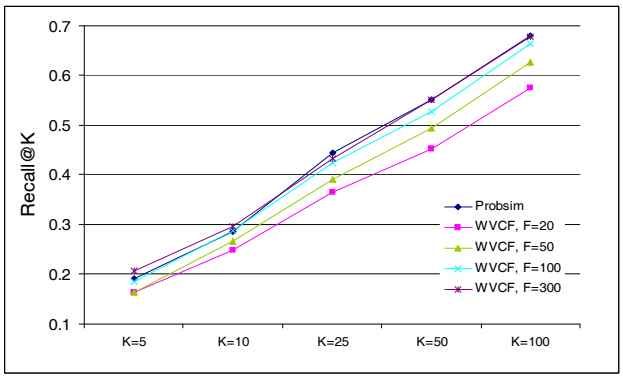


Fig 3: Recall performance of WVCF baseline vs Probsim for various factors, prior to local affects

The difference between WVCF and ProbSim becomes more prominent when we consider the "discovery" factor along with recall precision. This factor is based on how many overall programs of the corpus get recommended in recall to a set of users. It has been noted that models like ProbSim have an extremely low diversity since most of the recommendations are based from the top 1% popular programs.

Table 3 shows the top 15 programs from the bottom quartile of the MovieLens (20M, shows having less than 100 ratings discarded) dataset using the total number of positive and negative ratings and sorting by the ratio of positive ratings to the total number of ratings. While these "hidden gems" are virtually never recommended by the ProbSim model, they do have a high positive rating (though overall number of ratings is small), and are therefore

appropriate for more targeted user-recommendations. This is versus at least 4% hidden gem recommended for a modified approach, and thus capable to have a better spread in recommendations w.r.t. the discovery of "long-tail" programs.

Figure 4 shows the variation among users in MovieLens [9] for their preference in watching different grades of popular shows. Recommender systems targeted to high recall and high precision tend to overshoot for the most popular shows (Q1) to score well on recall (and precision automatically since there are not many "2" or lower ratings for most popular shows), while not discovering quality shows at lower popularity.

Title(i)	N(i)	Rating Avg(i)	G(i) +ve rating	B(i) -ve rating	P(i) = G(i)/N(i)
Intimate Strangers	126	3.655	117	2	0.9832
Follow the Fleet	118	3.712	111	2	0.9823
Ax, The (couperet, Le)	114	3.807	107	2	0.9817
One Man Band	105	3.895	98	2	0.9800
Angels' Share	105	3.710	96	2	0.9796
Queen of Versailles, The	144	3.628	137	3	0.9786
Yes Men Fix the World, The	194	3.843	179	4	0.9781
Big Clock, The	162	3.756	153	4	0.9745
Hundred-Foot Journey, The	156	3.750	143	4	0.9728
Short Film About Love, A	224	4.025	214	6	0.9727
Our Man in Havana	153	3.748	142	4	0.9726
Pixar Story, The	186	3.747	177	5	0.9725
Criss Cross	115	3.639	103	3	0.9717
Advise and Consent	183	3.811	170	5	0.9714
Something the Lord Made	214	3.895	199	6	0.9707

Table 3: 15 Hidden Gems in the fourth quartile of popularity in MovieLens

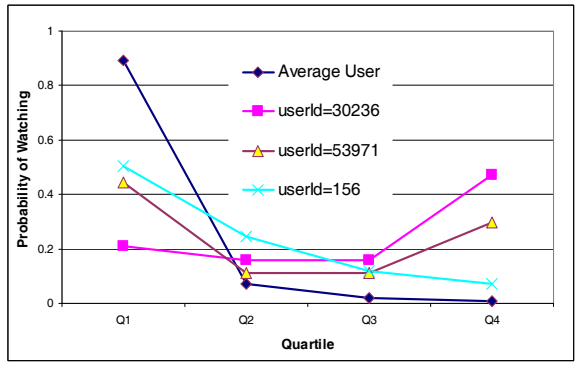


Fig 4: Popularity effect by quartile in MovieLens

To address the above issue, we introduce a penalty for not matching the user's overall popularity preferences over multiple bins using a loss such as:

$$LOSS(u) = 1 - \alpha \sum_{b=1}^{Nbins} \left| P(u,b) - p(u,b) \right|$$

Averaged over all u users, with $\alpha = 1$ and $N_{bins} = 10$ popularity bins, for example, the new recall score is normalized to better match user distribution $\vec{P}(u)$ with the recommended distribution $\vec{p}(u)$. This penalty reduces Probsim's recall score over MovieLens from roughly 0.3 to 0.1, with Q4 hidden gems still at nearly 0%. Furthermore, different shows and movies have quite different popularity characteristics across data sets, both explicit and implicit. To this extent, recall scores should be penalized even further to remove arbitrary popularity artifacts associated with each data set. Although the choice of α was arbitrary in this case, without having a "budget" to place recommendations into popularity bins which fits known user preferences, the model otherwise appears lacking in diversity and drastically under-recommends hidden gems, for the short-sighted purpose of scoring high recall.

If the model were tweaked for purely high recall, where watching is predicted rather than liking (to find hidden gems), Figure 5 below shows example performance of WVCF-watch along with liking and the combination of

like/watch. Figure 6 shows the precision equivalent. Here Gradient Boosted trees take the baseline WVCF (vector only), one movie at a time, and apply the baseline predictor as an attribute for that movie. It then expands the input deck to also include many attributes in the surrounding neighborhood of ratings. This produces a very strong estimator targeted to watching. In this case, Figure 5 shows SVD (F=100) versus Probsim versus WVCF-watch for the (budgeted) popularity bin of the top 25 most popular shows. Recall/Precision points were tallied based on each user/item sample in the cross validation being somewhere in that bin (with the predictor answering which movie was watched of the 25). The important thing here is that with both WVCF-watch and WVCF-like, the underlying recommender engine can control effects such as the dial for liking and watching, compared to other models with high precision/recall that only provides a single answer (both like and watch simultaneously).

Table 4 below shows the results of applying the 208 word emotions vocabulary from [10] using pre-trained word vectors to MovieLens usage within a modified SVD framework. As these moods serve as additional metadata, the experiment here shows the power of Word Vectors fused into the model. The training process tuned the relevance of each vocabulary word vector to each movie, within SVD latent movie factors, to best model explicit likes. For top word ranking, interesting patterns emerge between overall, most popular, and gems. For example "contrary" and "bored" is universally important, "obnoxious" matters for most popular, and "powerful" really applies to gems.

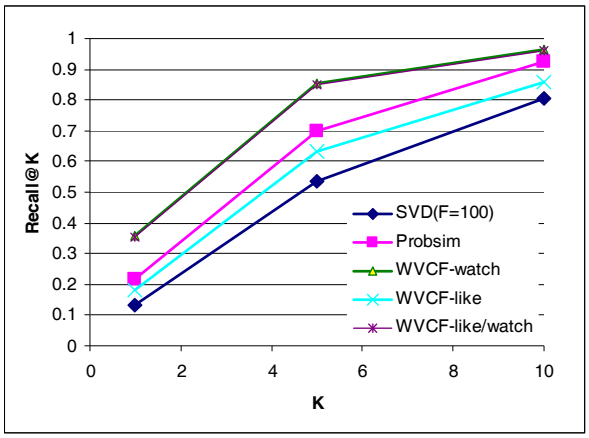


Fig 5: MovieLens Recall@K for top 25 Most Popular bin during "budgeted" recall

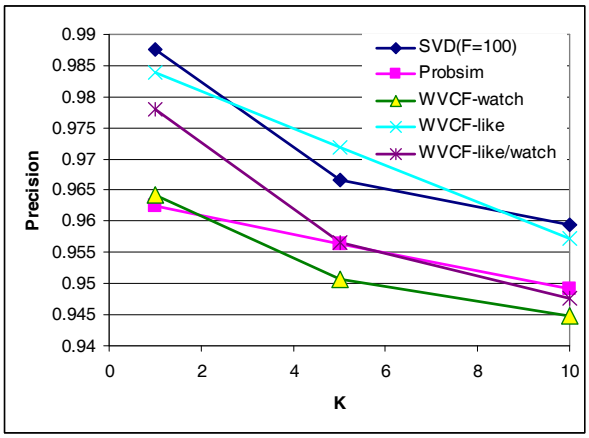


Fig 6: MovieLens Precision for top 25 Most Popular bin during "budgeted" recall

Rank	Overall	Popular	Gems
1	contrary	lazy	honest
2	angry	inconsiderate	contrary
3	brilliant	pity	cordial
4	wonder	obnoxious	powerful
5	bored	annoyed	lucky
6	insightful	woeful	wise
7	rough	haughty	aggressive
8	affectionate	uncertain	sad
9	weird	alive	nasty
10	folksy	silly	folksy

Table 4: Relevance of emotions/feeling/mood word vectors trained over MovieLens usage

CONCLUSION

In this paper, we experiment with the idea of providing powerful recommendations ranging from discovery to high precision across an arbitrary space of multimedia assets. We discuss a unified framework that aggregates consumption data from multiple sources and fuses them with meta-content to obtain more accurate content and user representations. The explicit ratings are used to convert the implicit watching user behavior to a notion of "likeness" based on ground-truth. The usage information is then used to feedback into the meta-content to determine more accurate weights of the individual meta-content factors thereby enabling richer content-content recommendations.

Methods within the unified framework were applied to MovieLens data to frame some common issues involved in the tradeoff between discovery (of hidden gems) versus recall/precision. Both watching WVCF-watch and liking WVCF-like models provide further flexibility to personalize recommendations. By budgeting recall to within the same popularity bins, hidden gems are not as likely to be ignored due to a prediction of not watching caused by a lack of awareness of the (unpopular) show's existence. Modifications to scoring functions were discussed to better personalize recommendations to best fit each user's watching preferences for maximal viewing enjoyment. An example fusion of word vectors to explicit ratings usage was provided based on a specified vocabulary, with patterns provided overall, for most popular shows, and gems. This is complementary to tag fusion, as in this case each vocabulary word was universally applied to each show.

There are several interesting directions in which our research can continue in the future. Care was taken to seamlessly convert all available data spaces, including metadata, into factored components, but integrated local effects such as [8] as well temporal effects can be better added. Even though the choice of gradient boosted trees works well for WVCF-watch and WVCF-like, other local correctors and stacking may further be beneficial. Another key issue is extensive data of explicit ratings are usually not available in linear content. Additional ways may exist, such as latent factor and nearest neighbor combinations, to close the gap between an arbitrary notion of "liking" in implicit space and explicit and/or ground truth. Another area of research is to further formalize the notion of "discovery" and relating it to likeness and precision. Further modifications of recall scores based on popularity biases between data spaces may help also. It will be interesting to see how this can then be used to create a parameter knob that can be controlled to either increase discovery vs. precision in recommendations.

REFERENCES

[1] Gediminas Adomavicius, I. and I. Alexander Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. Knowledge and Data Engg, IEEE Transactions, 2005. 17(6): p. 734-749.

[2] Melville, P. and V. Sindhwani, Recommender Systems. Encyclopedia of Machine Learning, 2010.

[3] Miller, B.N., et al., MovieLens unplugged: experiences with an occasionally connected recommender system, in Proceedings of the 8th international conference on Intelligent user interfaces 2003: New York, NY.

[4] Su, X. and T.M. Khoshgoftaar, A survey of collaborative filtering techniques. Advances in Artificial Intelligence, 2009.

[5] Musto, C., Enhanced vector space models for content-based recommender systems, in Proceedings of the fourth ACM conference on Recommender systems. 2010: Bari, Italy.

[6] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In Advances in Neural Information Processing Systems 26, pages 3111–3119.

[7] O. Jojic, M. Shukla, N. Bhosarekar, *A Probabilistic Definition of Item Similarity*, Proceedings of the 2011 ACM Conference on Recommender Systems, RecSys 2011, Chicago, IL, USA, October 23-27, 2011.

[8] Y. Koren, "Factorization Meets the Neighborhood: a Multifaceted Collaborative Filtering Model", Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008.

[9] *GroupLens Research*. Available from: http://grouplens.org/datasets/movielens

[10] *Vocabulary word list*. Available from: https://myvocabulary.com/word-list/emotions-feelings-mood-vocabulary

Modernizing In-Home Wi-Fi for Video: Multiple Access Points, Intelligence, and Mesh

Bulent Celebi, Executive Chairman and Co-Founder
AirTies

Abstract

Service providers are investing heavily to bring faster Internet speeds to the home, but conventional Wi-Fi equipment generally prevents subscribers from actually experiencing those speeds consistently throughout the home itself. More devices streaming video are straining the home network, and most subscribers do not realize the worst performing device (usually a smartphone or tablet) drags down the speed of all other devices in the home. Progressive operators need new solutions to ensure quality of service (QoS) for all video-capable devices and to provide a seamless TV viewing experience throughout every room in the house.

INTRODUCTION

Factors including home construction materials, the types of devices being used, and the distance to the nearest in-home access point (AP) are just some of the issues preventing subscribers from experiencing broadband speeds consistently throughout the whole home. Unlike traditional Wi-Fi, which relies on a single AP from a router/gateway, intelligent in-home Wi-Fi can be used not only to improve coverage, but to serve as the central video distribution system in the home – transforming all devices (including set-top boxes) into coordinated Wi-Fi hotspots. These APs are networked together using software to create an intelligent in-home network that prioritizes video, and dynamically routes data to ensure optimal performance across all screens in the home. In addition, these APs can be connected by a combination of wired connections – such as Ethernet, MoCA, and powerlines – and wirelessly by Wi-Fi Mesh.

THE PERFECT WI-FI STORM

Taken together, the rise in IP-connected devices connecting to domestic networks and mobile streaming creates the perfect Wi-Fi storm. For some operators, this means that

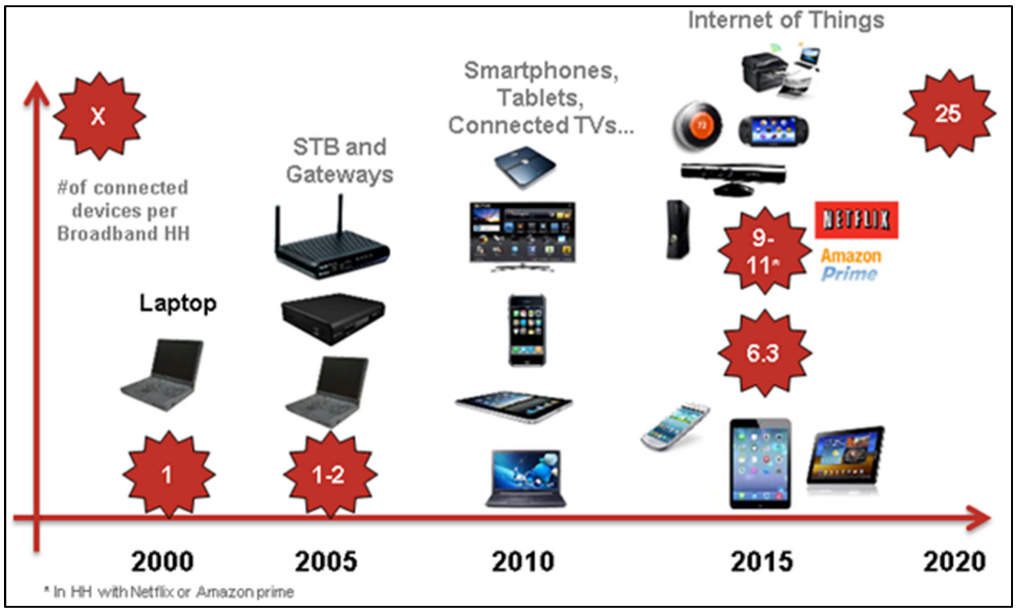


Figure 1: The Perfect Wi-Fi Storm (MRG Research)

more than half of all call-center traffic is Wi-Fi related. As shown in Figure 1, according to MRG Research, the average household in 2015 had more than six Wi-Fi devices capable of steaming video. MRG expects this figure to reach 25 devices by the year 2020. Notably, in 2015, those households that also subscribed to Netflix or Amazon Prime had between nine to 11 Wi-Fi devices capable of steaming video.

Likewise, according to the 2015 Cisco Video Index (VNI) Global IP Traffic Forecast, shown in Figure 2, Internet video represents 43.4% of exabytes of IP traffic per month.

Research is clear: the amount of video streaming is rising dramatically (and will continue to do so in the years ahead), which will place increasing strain on the home network, as the slowest performing device in the home drags down the speed and performance of all other devices.

DEVICE AND HOME LAYOUT IMPACT ON OVERALL IN-HOME PERFORMANCE

Based on results from AirTies' study of Wi-Fi performance in homes with just one AP, distance, barriers such as walls and ceilings, smartphones and tablets equipped with low end chips, and contention between multiple users degrades not just individual device performance but total Wi-Fi capacity as well. The problem here is not Wi-Fi itself. The latest IEEE 802.11ac provides plenty of bandwidth on paper.

As shown in Figure 3, simply moving a device away from the home's main router can dramatically impact performance for every

Figure 2: Global IP Traffic by Application Category (Cisco 2015 VNI)

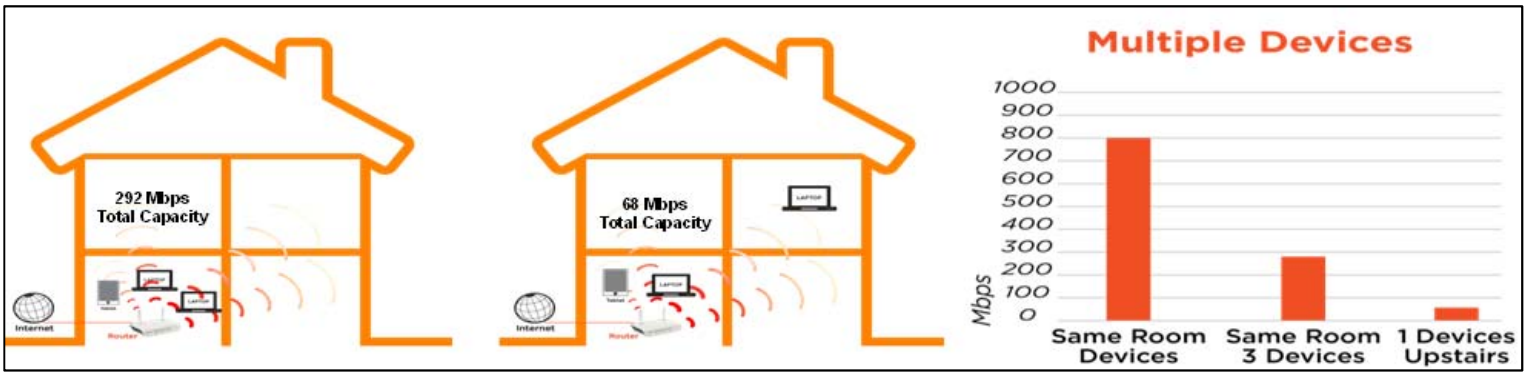


Figure 3: Device Impact on Overall In-Home Performance

device in the house. Our data shows that this is true even with the most advanced Wi-Fi devices. For example, in a room receiving 802 Mbps, sharing data amongst three devices – between an iPad 4 (2X2 11n), and two MacBooks (3X3 11ac) – total capacity drops to 292 Mbps in the same room because of the iPad (11n device). Moving one MacBook (3X3 11ac) upstairs drops the total capacity to 68 Mbps in the room previously receiving 802 Mbps, shared between the three devices. If the iPad 4 was moved upstairs, much worse results would be obtained.

Our research shows that streaming device-to-device (as opposed to each device connected to the Internet individually) reduces total capacity by 96%. For example, in a room with an iPad 4 (2x2 11n), DVR (3x3 11ac), and set-top box (3x3 11ac) total capacity drops from 292 Mbps to 175 Mbps, as displayed in Figure 4. Moving the DVR (3x3 11ac) upstairs drops total capacity even further, to 38 Mbps.

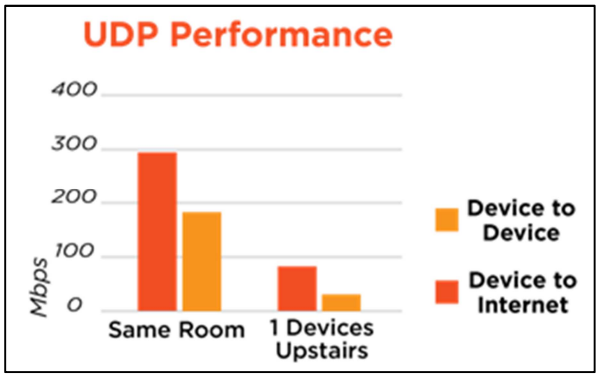


Figure 4: Device Streaming Effect on In-Home Performance

Patchy coverage is also a common issue in larger homes or in homes where walls impede wireless signals. Things as trivial as bathroom tiling, stucco walls, concrete support beams, and more all play a factor.

For example, AirTies ran tests in multiple sites around the globe. At an apartment in Europe with 20 Mbit/s of broadband access, streaming video from an iPhone in the same room as the Wi-Fi router provided the expected or better performance than what the subscriber paid for. However, moving the iPhone to the next room over dropped the performance from 20 Mbit/s down to 1Mbit/s. (Anything below 5 Mbit/s on Wi-Fi will be perceived as poor performance by the user.)

In an Asian apartment (MDU) with a 1 GBit/s fiber connection, using the best available routers and laptops, where the receiving device is located in the same room as the router, the laptop downloaded at 177 Mbit/s. However, moving the laptop down the hall dropped the performance from 1 GBit/s to 108 Mbit/s. Moving the laptop even further into the bedroom dropped the performance from 1 GBit/s to 0.8 Mbit/s.

Another AirTies test-case showed the worst-case scenario in a three-story European home built with concrete walls and floors, whereby coverage was only available on the same floor as where the residential gateway came into the house.

This leads some users to deploy repeaters to boost the signal, which may partially solve coverage issues but also creates new problems by consuming extra bandwidth that results in service degradation. This can happen when users are within range of the repeater and the original AP at the same time, consuming twice as much airtime capacity as necessary, and negatively effecting overall network performance – especially in the case of video streaming. In essence, repeaters can half the performance of in-home Wi-Fi rather than helping.

Traditional PLC/Wi-Fi or MoCA/Wi-Fi extenders treat the Wireline as a backbone. But, actual Wireline throughput varies widely. Frayed wiring, distance or noise are typical causes that limit the total network performance to the weakest link.

TABLETS AND SMARTPHONES MAKE POOR CONNECTION DECISIONS

As seen in Figure 5, mobile phones and tablets often make poor decisions as to where to connect to an AP or band – typically based on signal strength, with no knowledge of the network or optimal path(s). The RF characteristics of 2.4G give higher signal strength and therefore most mobile clients tend to connect to it, although the 5G 11ac connection typically performs 10X better.

Furthermore, mobile/tablet clients are often "sticky", in that once they make a connection to a particular AP they stick with it, even if the signal strength has dropped significantly or there is a better device that is closer. Due to the poor connection, they start consuming more of the total available airtime capacity, in particular if they are streaming video. This is often referred to as the "Bad Apple" Wi-Fi performance problem, where the slowest performing device degrades overall performance for all other devices in the home.

Mobile devices are equipped with low end 1X1 chips due to size, power and costs and thereby have inherently low performance to begin with. As the device moves around the home, away from the AP it was initially connected to and behind walls, the original AP devotes more and more capacity to serving that particular device's needs. In reality, from a network management point of view, it should have re-connected to a different AP, if available, or switched to a different channel. The combination of low end chips and distance from the connection consumes large amounts of airtime and dramatically reduces overall network performance, which becomes more acute when streaming video.

HYBRID MESH SYSTEMS TO CONNECT MULTIPLE ACCESS POINTS

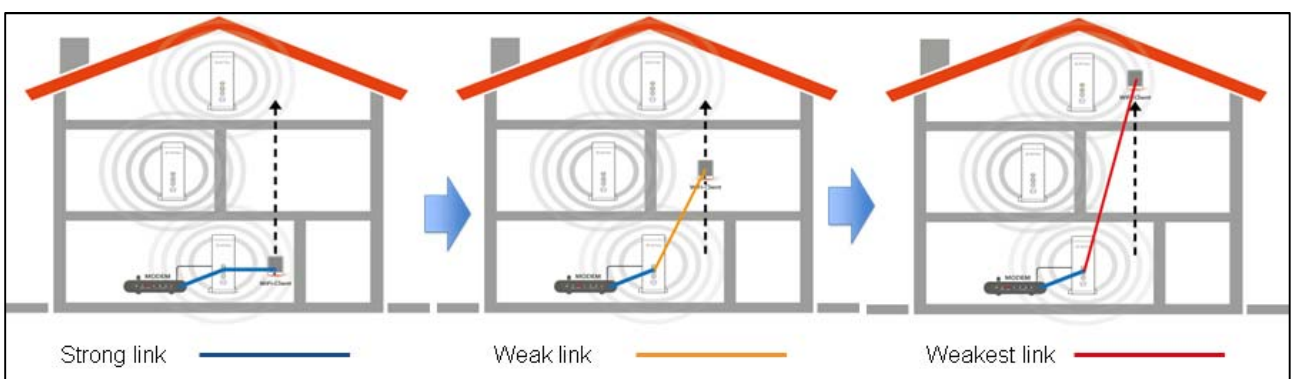


Figure 5: Mobile/Tablet Clients Make Poor Connection Decisions

Many of the issues discussed can be resolved by deploying intelligent Wi-Fi Mesh. Mesh creates a managed Wi-Fi domain that exploits multiple APs to provide multiple paths to overcome obstacles, and allows mobile clients to connect at their maximum capability, while creating resilience, extra capacity and eliminating dead zones. The Mesh operates logically as a single network with just one SSID (Service Set Identifier), dynamically configuring links between them, as the APs are relocated or as clients move around the house – and also switching Wi-Fi communications between paths as signal conditions change. Intelligent networks also use both 2.4 GHz and 5 GHz Wi-Fi signals to expand capacity, with the latter effectively providing backhaul between the APs for added performance.

For example, in the aforementioned three story home, with AirTies' solution installed, the user was able to simultaneously stream three HD streams around the house (one to the basement, one to the first floor, and one to the second floor), while simultaneously downloading a movie to a laptop from the in-home network-attached storage (NAS) device. There were up to 90 Mbit/s worth of traffic going throughout the concrete home with only 50% consumption of airtime capacity. In this scenario, the user was able to get 180-200 Mbit/s on a sustained basis.

AirTies' hybrid Mesh treats Ethernet, PLC, and MoCA as individual point-to-point Mesh connections, in addition to the wireless Mesh links available. This not only overcomes corner cases like steel walls, but can also be used to increase the total network capacity. AirTies' Mesh nodes select a combination of wired and Wi-Fi hops to best route packets, and run links in parallel, which dramatically increases total network capacity. A best path is determined by taking factors such as source and destination location, number of hops required, and best point-to-point link speed into account, rather than just signal strength.

Hybrid Mesh supports failover, so that for example, a PLC link could cut in when a Wi-Fi signal fades to ensure service continuity

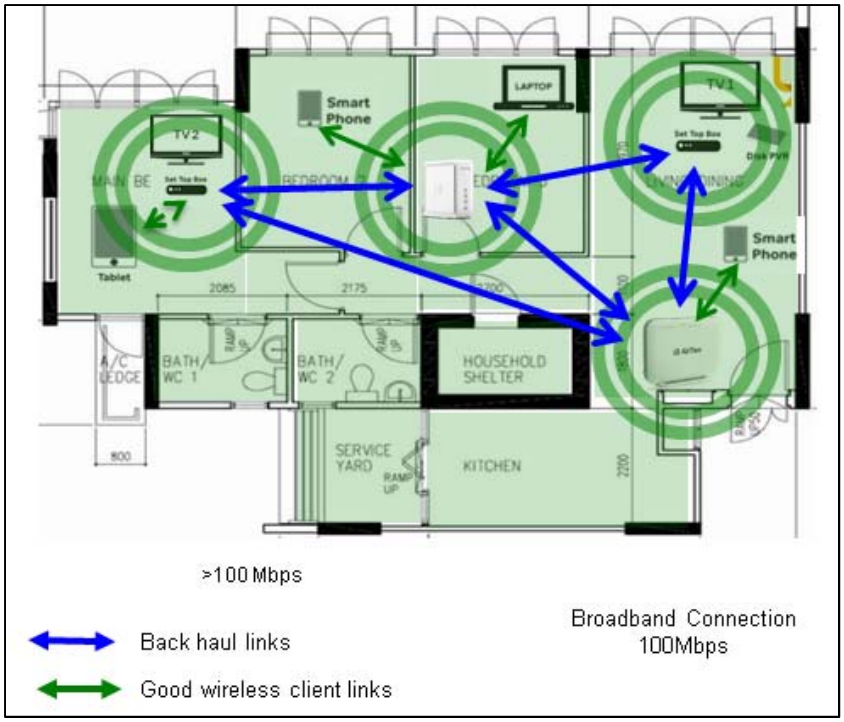


Figure 6: Hybrid Mesh Connects Multiple APs

and graceful degradation.

AirTies' modular approach to video and Internet distribution in the home gives operators a solution for guaranteed QoS to the end device, whether wired or wireless. AirTies' software lets operators meet the needs of individual homes with a complimentary and dynamic wired/wireless approach. Previous dark spots can be eliminated by the simple addition of another device using the network resource best equipped to reach it.

A major attraction for operators is that Mesh achieves effective control over the Wi-Fi domain that they have been lacking, while also giving consumers freedom to deploy the APs themselves and thereby avoiding the costs of onsite visits. Operators can be confident that Wi-Fi will extend their broadband service to the end device and enable the same performance guarantees. Operators can simultaneously keep their customers and support staff happy by reducing the number of call-center issues related to Wi-Fi.

NETWORK DECIDES WHERE DEVICES SHOULD CONNECT

An intelligent network uses a technique called "client steering", which effectively decides where and how devices connect within the home – whereby an intelligent home network, not the subscriber's devices, routes data in the most efficient way.

Traffic is routed around the network in a manner analogous to the way in which IP packets are routed around the Internet: across a best path instead of a single one, which is dynamically configured and re-configured depending on the number of active devices, the traffic profile, and the QoS required by the subscriber's agreement with the service provider. As seen in Figure 7, the APs communicate with each other and decide what is the least network capacity cost path, and the network steers the mobile device to that path. Based on AirTies' deployment data, client steering can improve average performance eight-fold by keeping mobile devices connected to the best AP location and 5G/11ac connection speeds.

THE NEED FOR MULTIPLE ACCESS POINTS TO SUPPORT VIDEO DISTRIBUTION

Instead of relying on one Wi-Fi hotspot in the home, every managed CPE can become a smart Wi-Fi AP as well.

These APs can include not only routers but also IP-enabled set-top boxes and DVRs.

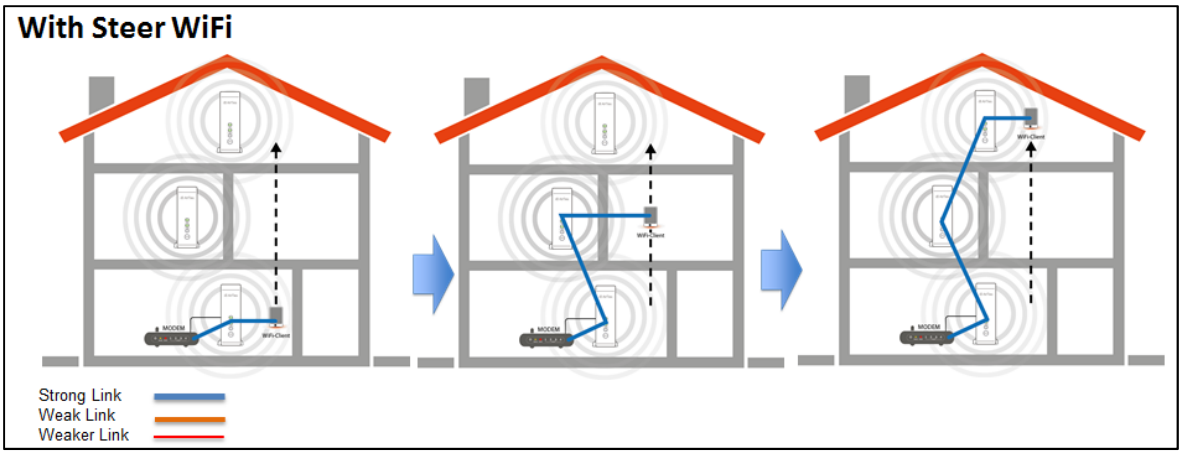


Figure 7: Network Steers Device to Best Connection

For example, Sky Europe's Sky Q is using AirTies' software in all of its devices (set-top boxes included) to transform them into coordinated Wi-Fi hotspots, which creates a robust network of multiple APs and results in better coverage and consistency throughout the entire home.

REAL-WORLD DEPLOMENT OF VIDEO-CENTRIC HYBRID MESH WI-FI

Sky Europe's Sky Q is a family of advanced products that connect wirelessly to create a new ecosystem that makes it easier for customers to access their favorite TV content. Sky Q's "Fluid Viewing" enables subscribers to watch content on up to five screens simultaneously, while recording four other channels, pause on one device and resume viewing on another, watch live, recorded or on-demand content from any device around the home, and take recordings on-the-go with a mobile device.

The AirTies Mesh solution is part of Sky Q's next generation of Internet Routers, TV set-top boxes and accessory devices, and is a key contributor to Sky's Fluid Viewing by enabling a superior Wi-Fi and video experience throughout the home. All of Sky Q's CPE is also part of the smart in-home Wi-Fi network to support streaming on both managed and unmanaged devices.

With a single, open Wi-Fi network delivering video content from broadcast, OTT and broadband, as well as locally stored content, customers can access Internet and video anywhere around the home. As mobile devices move throughout the house, AirTies' Mesh software maps the home network, monitoring the shifting demands for broadband and TV streams from a multitude of devices. It decides in near real-time which path to send data packets through to optimize performance. AirTies' Mesh improves connection speeds for Sky customers, while also enabling a simplified and improved user experience.

According to Andrew Olson, Director of New Products at Sky: "As Europe's leading entertainment provider, ensuring that our customers enjoy the best TV experience possible is our top priority. Working with AirTies has helped us create an amazing new customer experience across screens with Fluid Viewing, as well as take the Sky experience to the next level by enabling Sky Broadband customers to turn all their Sky Q boxes into smart hotspots."

Utilizing Wi-Fi Mesh enables a convergence of video IP systems throughout the home, as demonstrated by Sky Q's next generation, open Wi-Fi network that delivers video across every room. This type of deployment validates the fact that having multiple APs smartly linked together is the way forward for optimum home network performance. It is also likely to be a bellwether for other operators around the globe that are looking to provide a ubiquitous viewing experience throughout the whole home, on any screen.

CONCLUSION

Compared with typical Wi-Fi configurations based on a single AP, AirTies' modular approach to video and Internet distribution is an optimal choice for progressive operators that are looking to provide new solutions that ensure QoS for all video-capable devices and provide a seamless TV viewing experience across the whole home. Hard to reach areas see performance improvement by the simple addition of another device. The AirTies system can utilize MoCA, PLC, Ethernet, or Wi-Fi Mesh

to create smart links between the APs to ensure video quality in every room, even in thick, stone-walled homes. And, the Bad Apple Wi-Fi performance problem is resolved via an intelligent network of APs spread throughout the home.

NETWORK MIGRATION DEMYSTIFIED IN THE DOCSIS 3.1 ERA AND BEYOND

Ayham Al-Banna (ARRIS), Tom Cloonan (ARRIS), Frank O'Keeffe (ARRIS),
Dennis Steiger (nbn)

Abstract

The spectral efficiency of DOCSIS 3.1 networks is analyzed based on real-world measurements from the nbn network. It is shown how DOCSIS 3.1 can reduce the required number of node split operations as well as postponing them. The article proposes a decision tree to select the appropriate technology enablers such that traffic demand is met with minimum cost. DS and US network migration to DOCSIS 3.1 is described. Finally, long term network evolution scenarios are provided.

I. INTRODUCTION

The cable industry has achieved tremendous progress over the past decade in offering high speeds that met customers' demand and defended against competitive threats of speed wars. This journey is nowhere close to an end! As MSOs continue their network evolution, they are currently faced with an interesting time since many options are available to augment their existing HFC networks. For example, Fig. 1 shows multiple potential evolutionary paths that the MSOs can select from. Herein, the network architecture (e.g., I-CCAP/DAA/PON) is plotted against the topology which is presented here as the depth of the fiber in the network (e.g., HFC, FTTLA/FTTC, FTTT, FTTH).

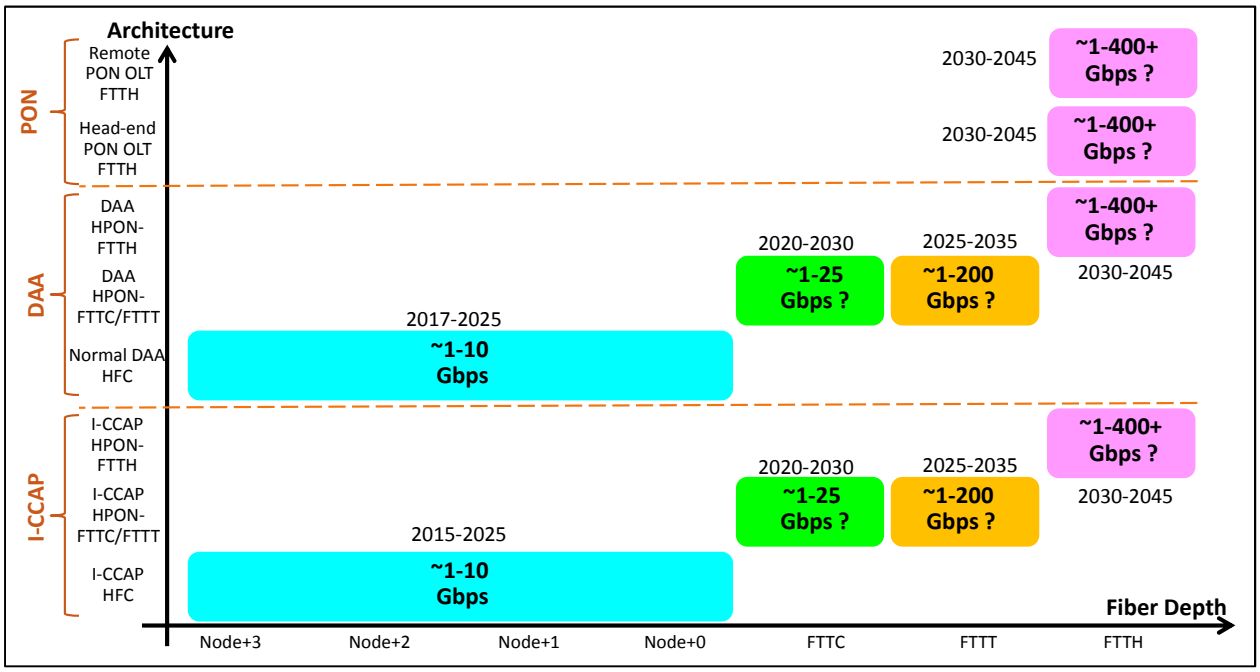


Figure 1. Evolution of cable networks in the next 2-3 decades.

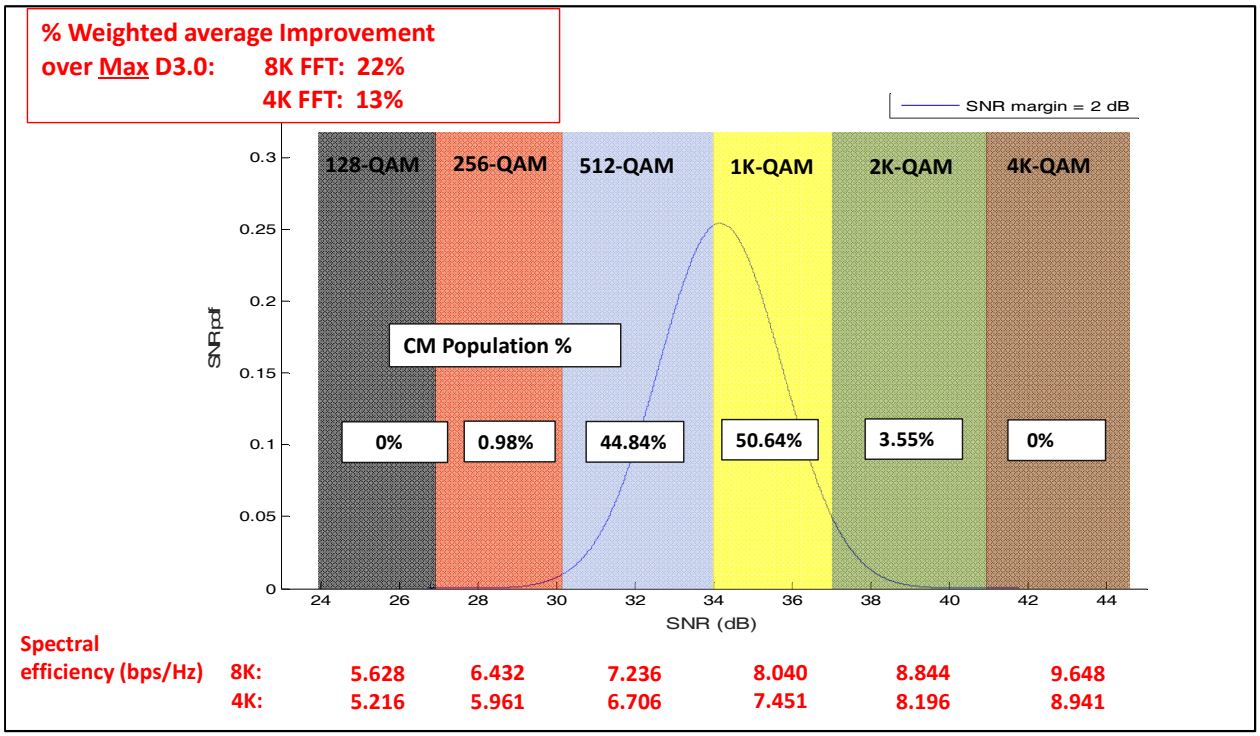


Figure 2. Spectral efficiency of DOCISIS 3.1 signals (pdf is plotted for SNR margin of 2 dB) based on Comcast data [1]

Note that the potential capacities shown in Fig. 1 increase as the fiber depth increases. This is because the analysis assumes that more fiber and less coax will yield less signal attenuation. Moreover, smaller service group sizes leads to less noise especially in the US direction. Both of these factors lead to higher SNR values. For instance, the I-CCAP capacities with HFC networks (N+x to N+0) can offer anywhere between 1 Gbps and 10 Gbps of total capacity depending on the upper edge of the DS spectrum (e.g., 550, 750, 1218, 1794 MHz) assuming that the spectrum is occupied by 402 MHz of broadcast digital video, 16 SC-QAM channels (with 6.33 bps/Hz spectral efficiency [1] [2]), and the rest is filled with DOCSIS 3.1 signals. The spectral efficiency of DOCSIS 3.1 signals is assumed to be ~7.7 bps/Hz based on previous analyses of current MSO plants and assuming an MSO operating margin of 2 dB as shown in Fig. 2 [1] [2]. Observe that the weighted average spectral efficiency of 7.7 bps/Hz is equivalent to an average mix of 512QAM/1KQAM.

For the I-CCAP FTTC capacities of 24 Gbps, the spectrum is assumed to be extended to 3 GHz and DOCSIS 3.1 signals are assumed to fill the whole spectrum with an average of 2K QAM and 0 dB SNR operating margin (8.8 bps/Hz as shown in Fig. 2). For the FTTT option that can offer a total capacity of ~200 Gbps, the spectrum is assumed to be extended to 20 GHz and 4K QAM DOCSIS 3.1 signals are assumed to fill the whole spectrum. The spectral efficiency of those DOCSIS 3.1 signals is assumed to be 9.6 bps/Hz with 0 dB SNR operating margin (see Fig. 2). Observe that the capacities of the DAA architectures in Fig. 1 are assumed to be similar to those of the I-CCAP architecture in this analysis in order to be conservative since the main benefits of moving to distributed access architectures are the saving of power/space in headend and supporting more

lambdas with digital optics. While digital optics can lead to higher SNR values and therefore higher capacities [3] [4], those potential capacities cannot be guaranteed due to potential bottle necks in the coax portion which is not affected by the architecture (I-CCAP vs. DAA). The coax portion consists of the following plants pieces: Tap, Drop cable, Home network in subscriber's home, and the Modem (TDHM) [2].

Figure 1 above also showed potential time frames for transitions between different phases of the same architecture or moving from one architecture to another. Those time frames overlap in time because different MSOs may select to transition to a particular alternative at different times. For example, it is shown that the current HFC networks with the normal practice of node splits going to N+0 may live until year 2025. At the same time, the graph shows that some MSOs may choose to move to a FTTC architecture as early as 2020 (until 2030?). Similarly, MSOs may choose to move to FTTT architecture as soon as 2025 (until 2035?). Finally, it is assumed that some MSOs may choose to migrate their networks completely to FTTH as early as 2030. Note that the capacities of all architectures (I-CCAP/DAA/PON) in an FTTH environment in the 2030 time frame are assumed to be similar (~400 Gbps+) because it is assumed that those architectures will leverage similar technologies then.

Different factors can cause MSOs to move from one architecture to another. For instance, moving from centralized (I-CCAP) architecture to distributed architecture (DAA) can provide multiple benefits such as the reduction of the headend space/power requirements and increased number of simultaneous lambdas that can be supported on a single fiber via wavelength division multiplexing. Moving from I-CCAP/DAA to PON architecture permits MSOs to move to a technology (PON) that may be their last-mile technology of the future. Moving from N+x/N+0 to FTTC and FTTC to FTTT permit the MSOs to increase the BW capacity to a service group by making the service group smaller and by reducing the length of lossy coaxial cables as well as extending the spectrum. Alternatively, some MSOs may view OBI-free RFoG) as their long-term last mile technology of the future which can be supported using FTTH architecture.

Given the large combinations of the various network architectures shown in Fig. 1 (I-CCAP/DAA/PON) and different fiber depth topologies, selecting the appropriate architecture/topology transition path is not a trivial task. The challenge at hand is to develop a methodology that utilizes the available technology enablers in selecting the appropriate transition path. These technology enablers include node segmentation, node splitting, DOCSIS 3.1, plant upgrades, spectrum management and reclamation, Selective Subscriber Migration (SSM), extended spectrum DOCSIS, DAA, and others.

Most of the technology enablers listed above are known and understood. However, some enablers like extended-spectrum DOCSIS and selective subscriber migration may need to be explained before being proposed for use in migration strategies. The Extended-spectrum DOCSIS refers to extending the spectrum used in cable networks above and beyond of what DOCSIS 3.1 can support [5] [6]. This can be effective in different network topologies like FTTC and FTTT where no amplifiers or diplexers are present. The coaxial cables can support very high frequencies like 3-6 GHz for rigid cables and 25 GHz for RG-6 drop cables. With extended spectrum DOCSIS, the proposed network topology is illustrated in Fig. 3 for the FTTT case.

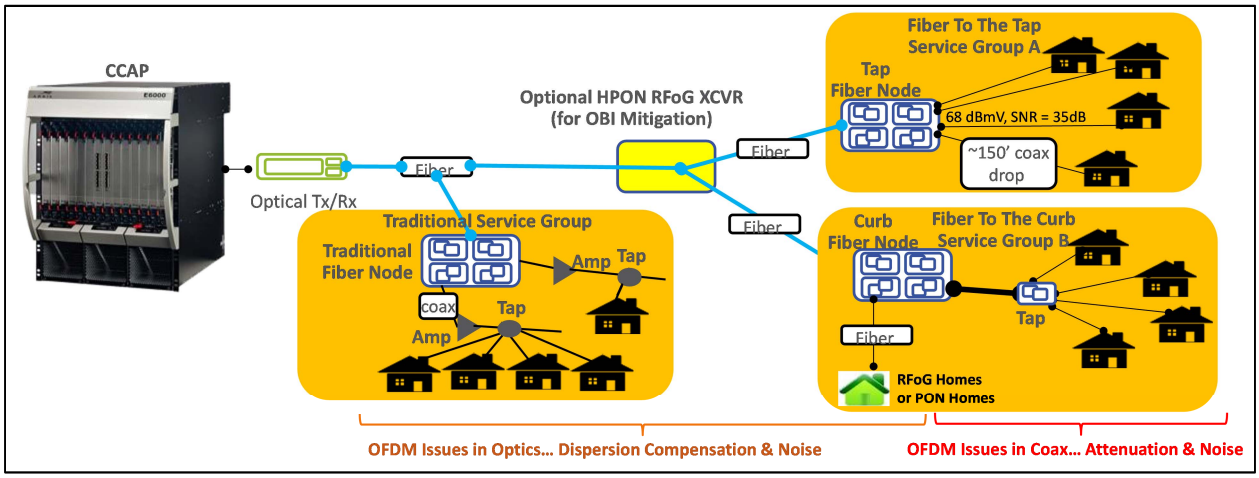


Figure 3. Extended-Spectrum DOCSIS (FTTT & FTTLA examples)

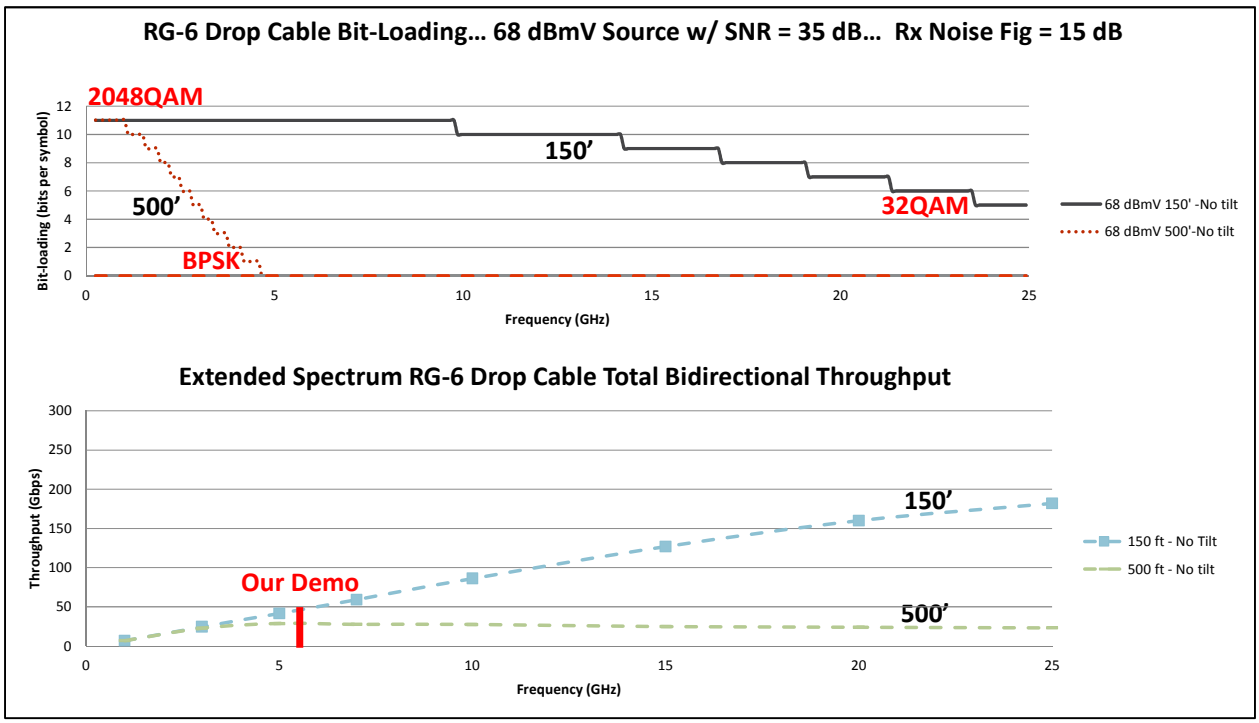


Figure 4. Extended-spectrum DOCSIS for the FTTT case: Modulation order decreases as frequency increases but total system capacity monotonically increases

Extending the spectrum to high frequencies will cause significant signal attenuation which causes a drop in the signal SNR. However, this can be accommodated by using lower order modulations at higher frequencies as shown in Fig. 4. Observe that while the modulation order drops as frequency increases, the total system capacity monotonically increases due to the expansion of the spectrum.

As for the Selective Subscriber Migration technology enabler, it refers to moving subscribers that require high peak rate service (and potentially heavy users) to a different topology like FTTH as shown in Fig. 5 [6]. This process significantly reduces the need to upgrade the network since the majority of the subscribers after the SSM process have moderate peak rates and reasonable consumption that can be accommodated by the existing network for a long period of time.

An example illustrating this concept is shown in Fig. 6, where the average throughput (Tavg) per subscriber in 2015 is assumed to be 600 kbps and the peak rate (Tmax) is 200 Mbps. According to the QoE formula [7] [8], the required capacity to support 500 subscribers is 520 Mbps (assuming k = 1.1). Ten years later (2025), Tavg and Tmax are assumed to be 35 Mbps and 11 Gbps, respectively, assuming 50% ACGR. The number of users is assumed to be 128 in 2025, which has dropped after multiple node splits. In this case, the total required capacity will be 16.6 Gbps. However, if only 1-2% of subscribers require high peak rates and they get moved to a different technology, then Tmax will likely be significantly less (e.g., 5 Gbps instead of 11 Gbps), which in turn drops the total required capacity to 10 Gbps (instead of 16.6 Gbps).

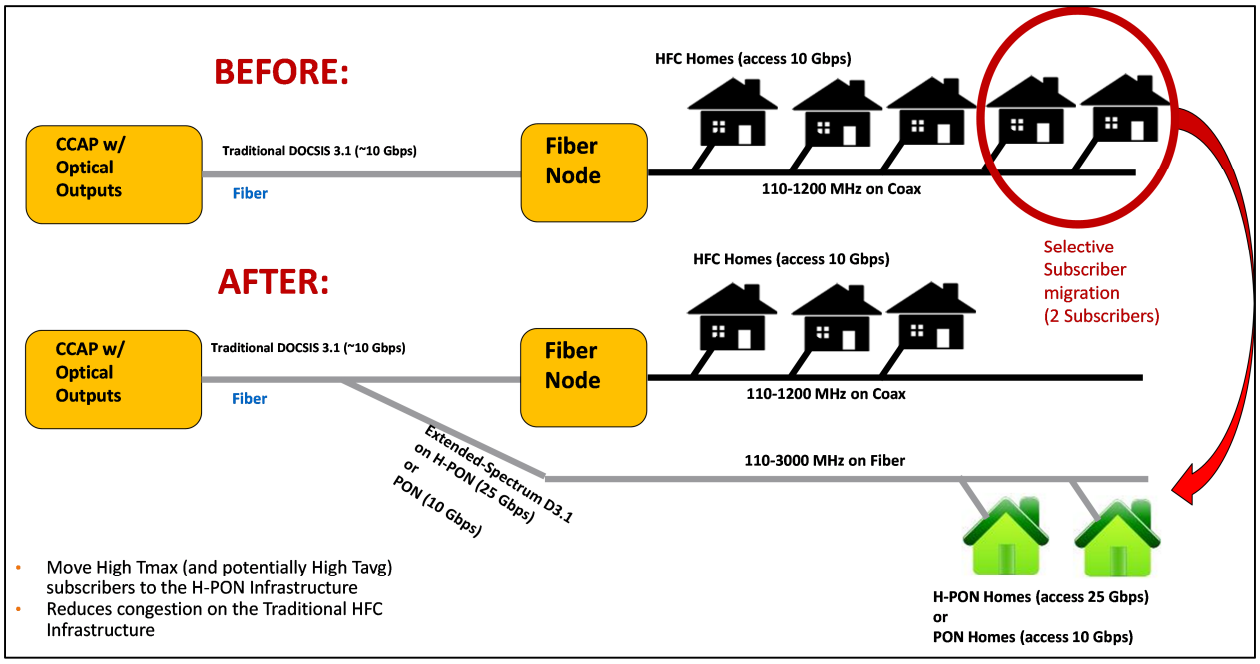


Figure 5. Selective Subscriber Migration (moving high peak rate users to a different topology like FTTH)

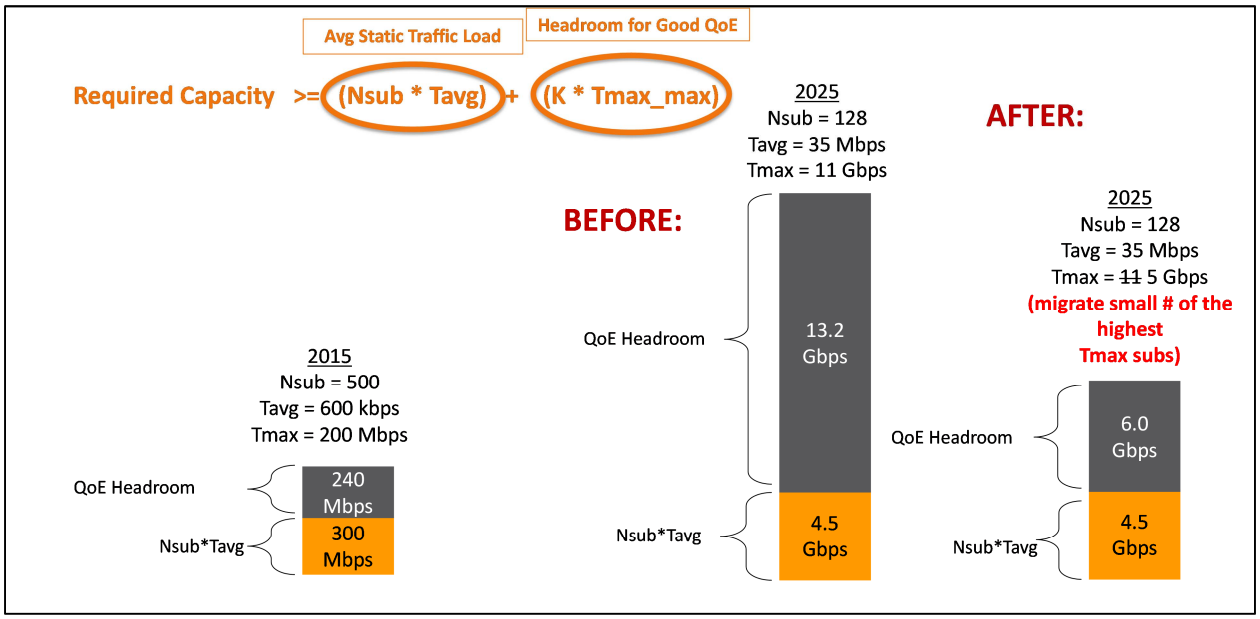


Figure 6. Selective subscriber migration extends the life of the existing network

The purpose of this paper is to develop a gradual network migration strategy that accommodates the traffic demand in a cost-effective manner using the various technology enablers listed above. It is believed that the best way to illustrate this methodology is via relevant and practical examples, which are based on real-world plant measurements and MSOs' particular plans for service offering. The analysis in this paper is based on collaboration between ARRIS and nbn that provided plant measurements as well as example service offering plans to be used in the analysis. While the results could be specific to nbn network, similar methodology can be applied by other MSOs networks through incorporating their own plant measurements and plans for service offering.

The paper is organized as follows: Section II analyses the potential DOCSIS 3.1 spectral efficiency gain based on measurements collected from the nbn network. In section III, traffic engineering and service offering plans are studied to understand how DOCSIS 3.1 can help in extending the life of the network. A methodology to use the different technology enablers for migration strategy is proposed in Section IV. Deployment scenarios specific to DOCSIS 3.1 are found in Section V. Section VI describes different example paths for the long term network migration, and finally, the paper is concluded in Section VII.

II. DOCSIS 3.1 SPECTRAL EFFICIENCY & CAPACITY ANALYSIS

MER measurements were collected and provided by nbn in order to perform the DOCSIS 3.1 analysis provided in this section. The histograms of the collected DS and US MER values are shown in Fig. 7 and Fig. 8, respectively.

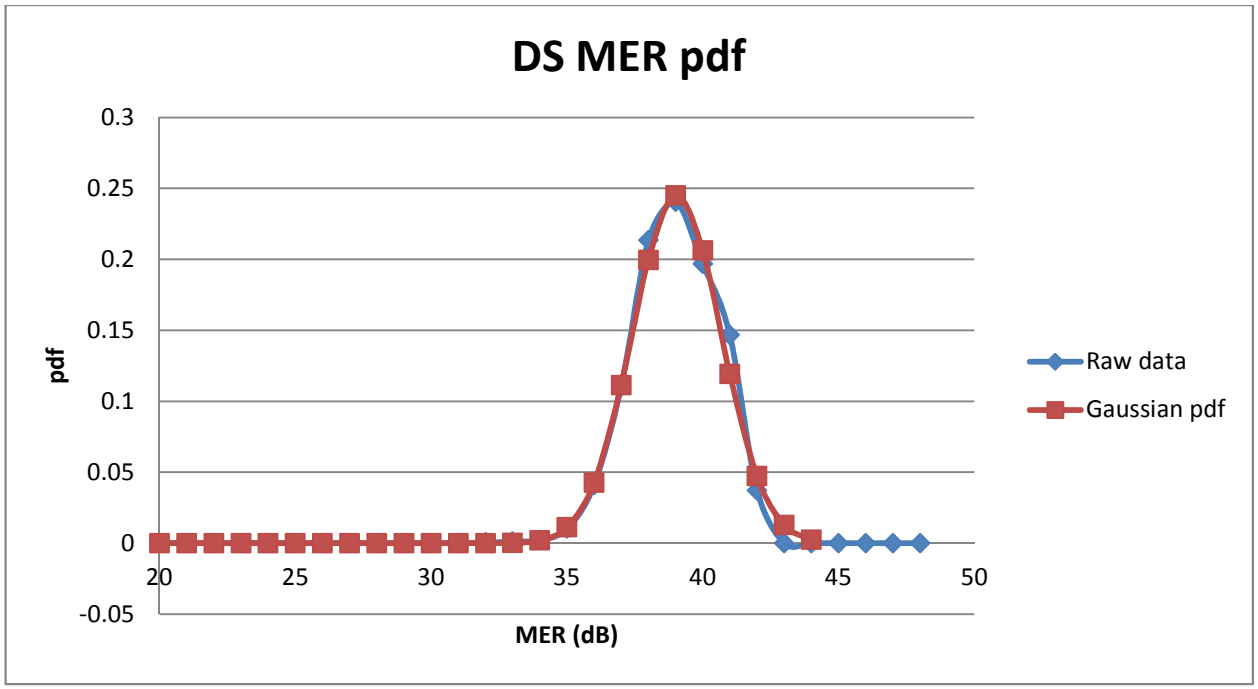


Figure 7. DS MER measurements collected on nbn network

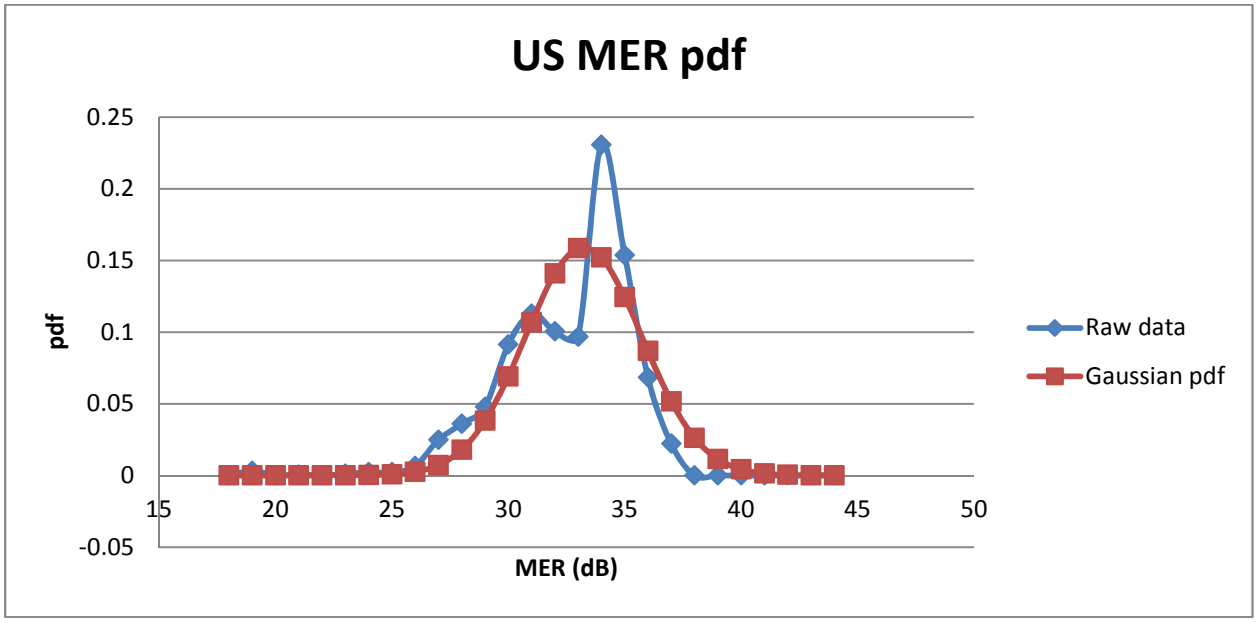


Figure 8. US MER measurements collected on nbn network

The spectral efficiency was analyzed following the same approach that was described in [1] [2] and using the same DOCSIS 3.0 and DOCSIS 3.1 parameters listed there. The DS spectral efficiency analysis for 0 dB and 2 dB MSO SNR operating margins are shown in Fig. 9 and Fig. 10, respectively. The DS spectral efficiency analysis results are summarized in Table 1 and Table 2. Similarly, the US spectral efficiency analysis is shown in Fig. 11 and Fig. 12, and is summarized in Table 3 andºTableº4.

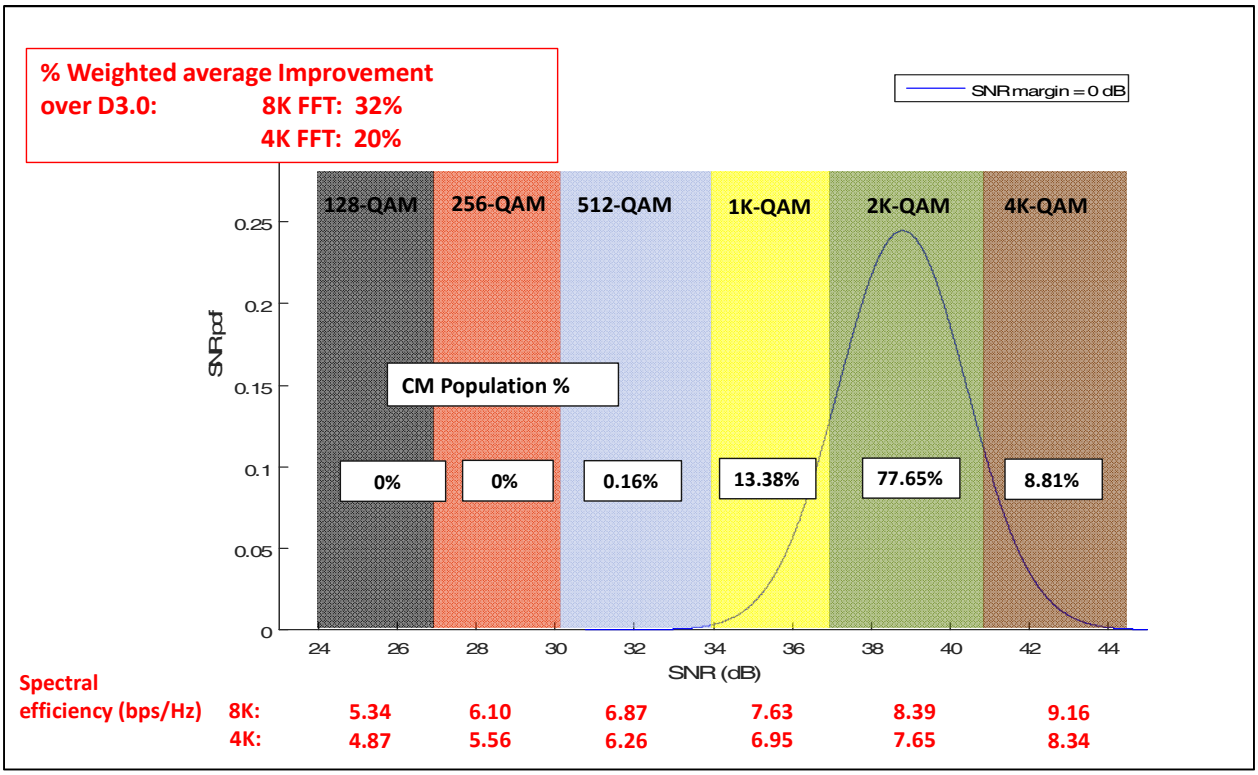


Figure 9. DS DOCSIS 3.1 Spectral efficiency analysis for 0 dB SNR operating margin

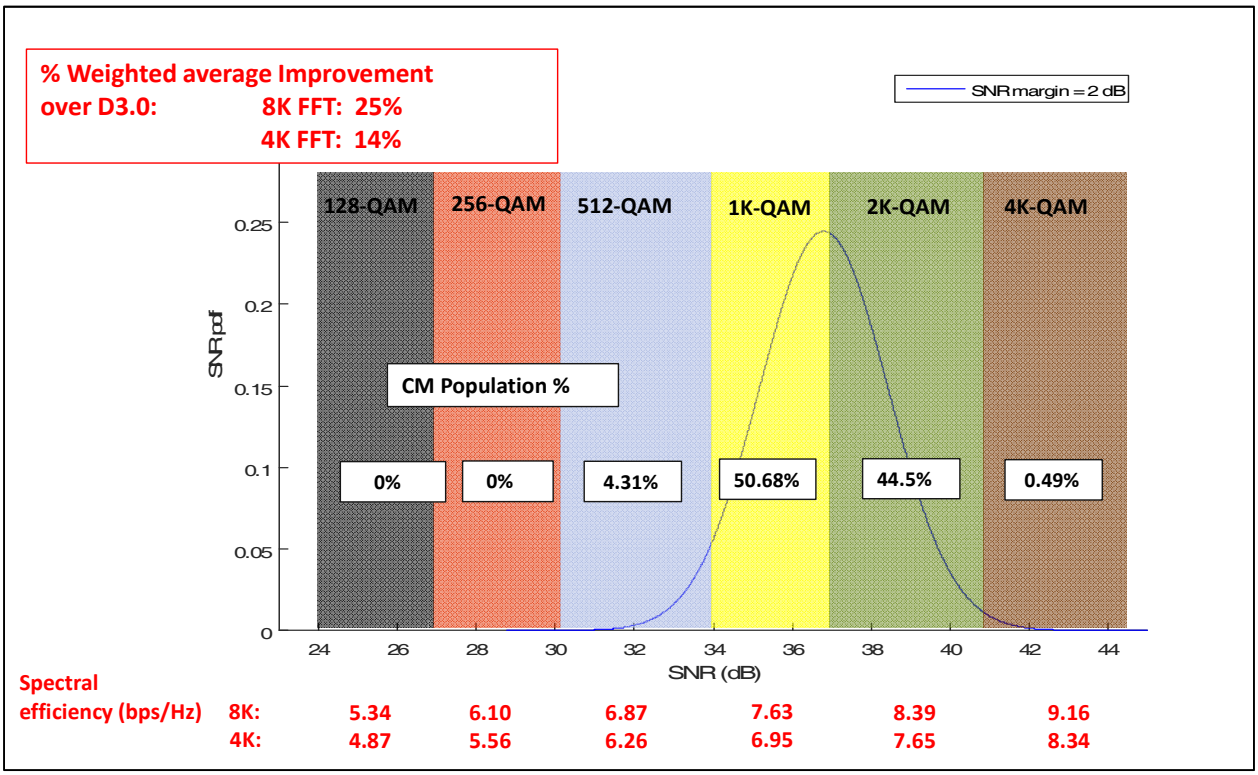


Figure 10. DS DOCSIS 3.1 Spectral efficiency analysis for 2 dB SNR operating margin

Table 1. DS DOCSIS 3.1 Spectral efficiency (bps/Hz) analysis summary

MSO SNR Operating Margin (dB)	4K FFT	8K FFT
0	7.61	8.36
1	7.44	8.16
2	7.24	7.94
3	7.02	7.70

Table 2. DS DOCSIS 3.1 Spectral efficiency gain summary

MSO SNR Operating Margin (dB)	4K FFT	8K FFT
0	20%	32%
1	17%	29%
2	14%	25%
3	11%	22%

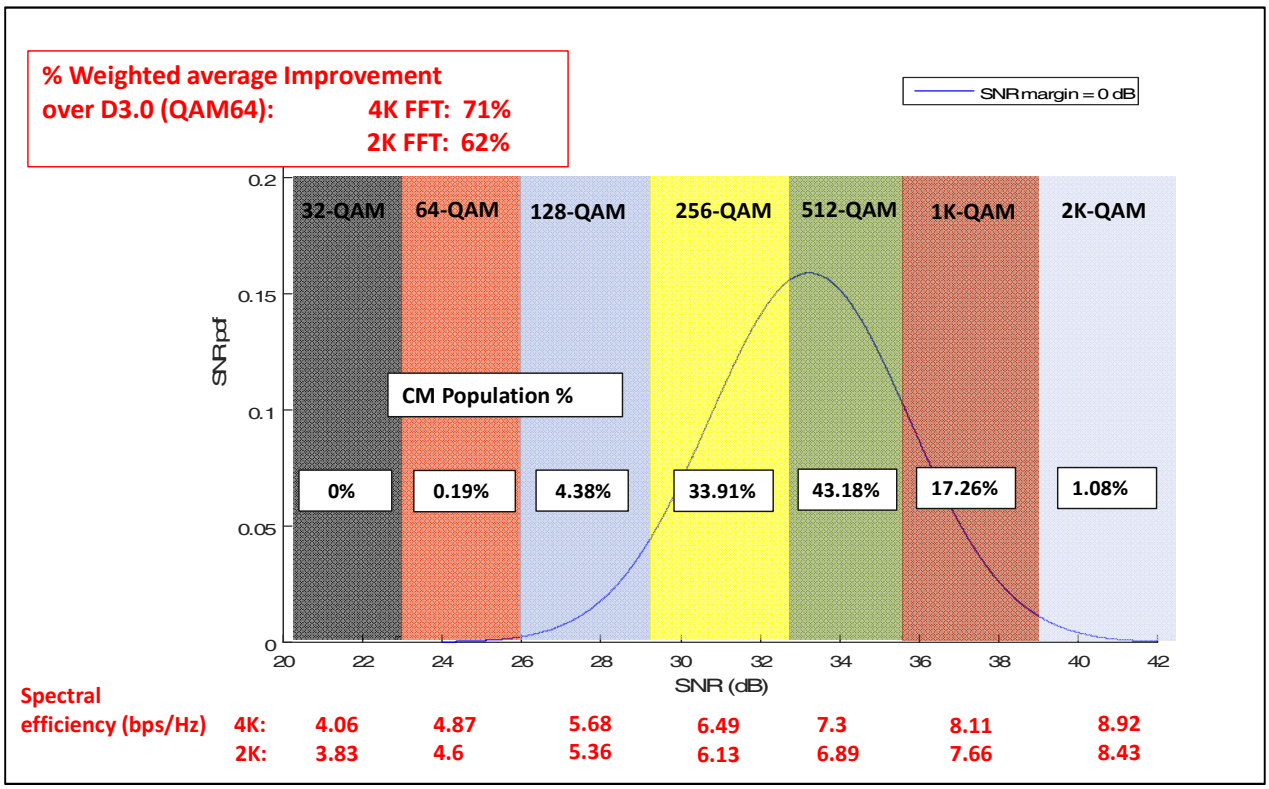


Figure 11. US DOCSIS 3.1 Spectral efficiency analysis for 0 dB SNR operating margin

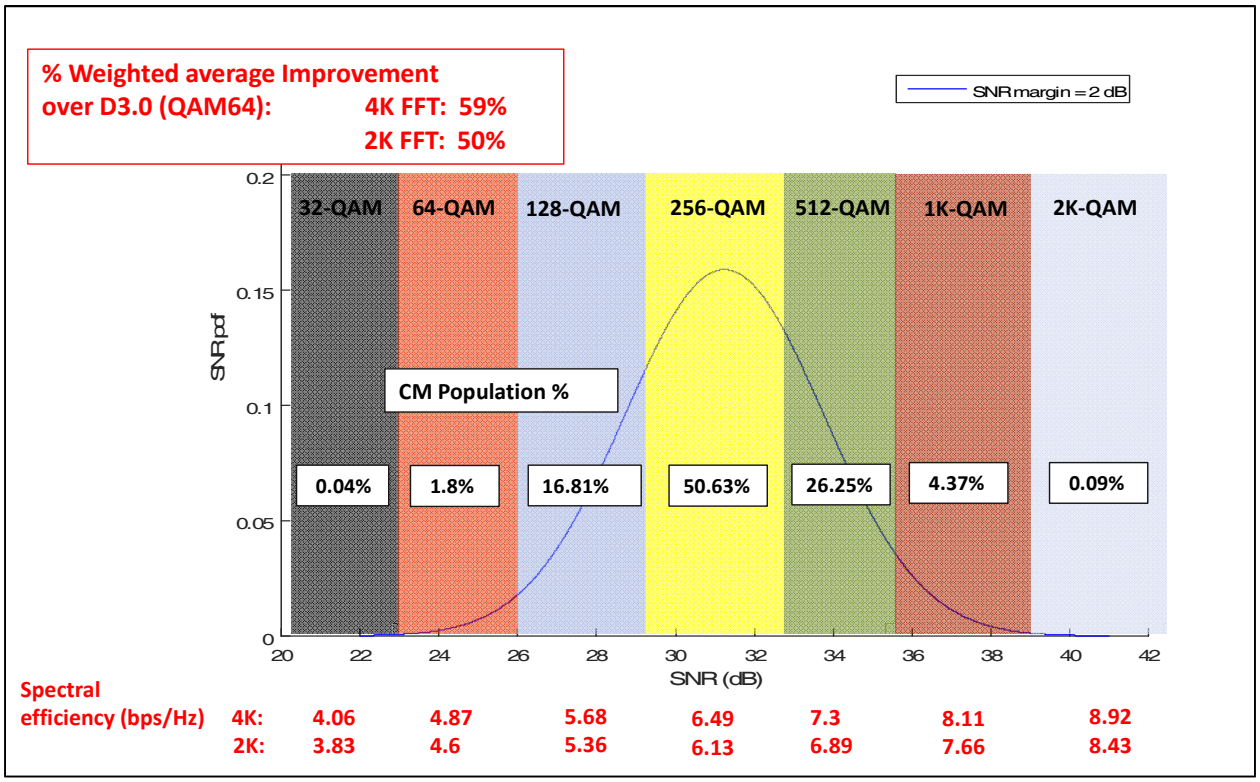


Figure 12. US DOCSIS 3.1 Spectral efficiency analysis for 2 dB SNR operating margin

Table 3. US DOCSIS 3.1 Spectral efficiency (bps/Hz) analysis summary

MSO SNR Operating Margin (dB)	2K FFT	4K FFT
0	6.71	7.11
1	6.48	6.86
2	6.24	6.61
3	6.01	6.36

Table 4. US DOCSIS 3.1 Spectral efficiency gain summary

MSO SNR Operating Margin (dB)	2K FFT	4K FFT
0	62%	71%
1	56%	65%
2	50%	59%
3	45%	53%

DS Capacity analysis was also performed to estimate the rough potential gain that DOCSIS 3.1 can offer. The nbn network is acquiring their infrastructure from two different operators (Optus and Telstra). Once nbn gets a specific network portion, they will offer their DOCSIS services in a coexistence phase with the old operator. This coexistence phase lasts for 18 months were both companies have to share the network infrastructure. Therefore, nbn will not have full access to the spectrum during the coexistence phase. The last coexistence phase will end in 2020. Note that coexistence phase implies that the available spectrum for nbn is limited. While this spectral restriction is particular to nbn network, lack of spectrum is an issue for MSOs in general due to many reasons including analog video channels, legacy DOCSIS channels, regulatory constraints, etc. Therefore, the analysis is applicable to other MSOs. The DS capacity analysis was done for both Optus and Telstra cases and is shown in Table 5. The DS capacities in Table 5 represent the estimated DOCSIS 3.0 and DOCSIS 3.1 capacities that can fit within the available spectrum for nbn assuming a DS DOCSIS 3.0 spectral efficiency of 6.33 bps/Hz and 7.94 bps/Hz spectral efficiency for DOCSIS 3.1 (assuming 2 dB MSO SNR operating margin and 8K FFT. Refer to Table 1 for details).

As for US capacities, similar analysis was performed with 4.15 bps/Hz DOCIS 3.0 spectral efficiency and 6.61 bps/Hz DOCSIS 3.1 spectral efficiency assuming 2 dB MSO SNR operating margin and 4K FFT size (refer to Table 3 for details). For the Optus case, during the coexistence phase, the available spectrum for nbn is 28.7 MHz where a large portion of it is in the noisy part of the spectrum leaving clean spectrum only enough for 3 6.4 MHz DOCSIS 3.0 channels. In the DOCSIS 3.1 case, it is assumed that the OFDMA signal is able to operate and utilize the whole 28.7 MHz spectrum. After coexistence, it is assumed that nbn can run up to 7 6.4 MHz DOCSIS channels in the 5-65 MHz for DOCSIS 3.0 but OFDMA is able to fully utilize the 60 MHz with DOCSIS 3.1. Similar assumptions are used for the Telstra case except for the coexistence phase where the available spectrum is 34.4 MHz where nbn can use 3 6.4 MHz and 1 3.2 MHz DOCSIS channels (instead of 3 6.4 MHz channels in the Optus case) due to access to larger portion of the spectrum during coexistence. The results are summarized in Table 6.

Table 5. DS Capacity gains offered by DOCSIS 3.1

DS Spectrum	D3.0 Gbps	D3.1 Gbps	Gain %
Optus Case			
Co-existence 134 – 198 MHz [8 8 MHz channels]	0.41	0.51	25.4
Post Co-existence [32 8 MHz channels]	1.62	2.03	
Telstra Case			
Co-existence (291-355) + (498-562) MHz [16 8 MHz channels]	0.81	1.02	25.4
Post Co-existence (291-419) + (434-562) MHz [32 8 MHz channels]	1.62	2.03	

Table 6. US Capacity gains offered by DOCSIS 3.1

US Spectrum	D3.0 Gbps	D3.1 Gbps	Gain %
Optus Case			
Co-existence 43 – 63 MHz	0.08	0.19	138
Post Co-existence 5 – 65 MHz	0.19	0.40	113
Telstra Case			
Co-existence 5 - 39.4 MHz	0.09	0.23	145
Post Co-existence 5 – 65 MHz	0.19	0.40	113

III. HOW CAN DOCSIS 3.1 HELP?

In order to show how DOCSIS 3.1 can help in the transition for MSOs, we consider the nbn application. While the numerical results in this section could only be applicable to nbn, the methodology is applicable to any MSO network in general. nbn is faced with several challenges as they acquire new infrastructure from the other operators. These challenges include large number of HHP per node, which could be as high as 1580 in the Optus network. Other challenges are also attributed to the lack of spectrum due to the limited access to the spectrum during the coexistence period, when both nbn and other operator have to share the available spectrum. Finally, the subscriber demand continues to grow!

In order to satisfy the traffic demand generated by the large number of HHP per node via the limited available spectrum, node splits is considered as a potential solution. However, performing large number of node split operations can be challenging for multiple reasons including cost which is affected by the number of nodes that increases exponentially after each node split operation. Another challenge is the availability of large

number of trained workers that can perform the required number of node splits operations within the required time frame such that the traffic demand is met. It should be noted that the node split operation can only help in cases where the capacity of the network is not enough to support the average throughput of the users during busy hour. This is because the node split operation reduces the number of subscribers per service group.

On the other hand, node splits cannot help when the network capacity is not enough to support very high peak rates (Tmax), which is significantly higher than the average subscriber rate during busy hour (Tavg). To illustrate this point, consider a hypothetical scenario where a network has two subscribers with DS Tmax of 200 Mbps and DS network capacity of only 150 Mbps; performing a node split that yields one user per service group will not help because the network capacity of 150 Mbps is still not adequate to support a single user with Tmax of 200 Mbps. Note that deploying DOCSIS 3.1 can use the limited spectrum more efficiently and therefore increase the network capacity in a given spectrum portion, which can be helpful in meeting the demand during busy hour as well as offering high peak rates.

This section presents some traffic engineering analyses and shows how DOCSIS 3.1 can be used to delay and reduce the required number of node split operations as well as avoiding them altogether in some cases. This can be very beneficial especially during the coexistence phase that was referenced above when the available spectrum is very limited.

The traffic engineering analyses are based on input from nbn which was obtained via collecting subscriber statistics and providing nbn's vision regarding potential tier speeds that need to be supported in the next few years. The traffic engineering assumptions are listed in Table 7 for both Optus and Telstra networks. Note that the parameters and plans are mostly the same for both networks except for the number of HHP per node which will affect the results significantly. Table 8 provides the take rate assumptions as well as the calculated number of subscribers which were used in the analysis.

Table 7. Traffic engineering assumptions

Details	Parameters	Optus	Telstra
Downstream Maximum Throughput in 2015	DS Tmax (Mbps)	100	100
Downstream Average Throughput in 2015	DS Tavg (Mbps)	0.3	0.3
Upstream Maximum Throughput in 2015	US Tmax (Mbps)	40	40
Upstream average Throughput in 2015	US Tavg (Mbps)	0.06	0.06
Downstream Compound Annual Growth Rate for Tmax	DS CAGR for Tmax	0%	0%
Upstream Compound Annual Growth Rate for Tmax	US CAGR for Tmax	0%	0%
Downstream Compound Annual Growth Rate for Tavg	DS CAGR for Tavg	23%	23%
Upstream Compound Annual Growth Rate for Tavg	US CAGR for Tavg	23%	23%
Household passed per node	HHP	1580	850

Table 8. Projected take rates and corresponding numbers of subscribers

YEAR	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
Take rate	0.1	0.15	0.35	0.54	0.66	0.74	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
Number of users Telstra	85	127.5	297.5	459	561	629	637.5	637.5	637.5	637.5	637.5	637.5	637.5	637.5	637.5	637.5
Number of Users Optus	158	237	553	853.2	1043	1169	1185	1185	1185	1185	1185	1185	1185	1185	1185	1185

The total system capacity required to support the number of subscribers with the specified Tmax and measured Tavg values is calculated via the QoE formula that was described earlier and provided here again for convenience [7] [8]

$$C = (Nsub \times Tavg) + (K \times Tmax),$$

where

- C -Required bandwidth capacity for a particular service group
- Nsub -Number of subscriber in the service group
- Tmax -Highest peak rate offered to subscribers
- Tavg -Average busy hour per subscriber bandwidth consumption
- K -QoE factor (assumed to be 1.1 for the analysis)

Observe that the first part of the formula compensates for the average traffic load whereas the second part relates to the headroom needed to support the peak rate such that good QoE service is provided to all subscribers. The required DS and US capacities for the Optus network are shown in Fig. 13 and Fig. 14, respectively. The available network capacity corresponding to different phases (refer to Table 5) are also illustrated on figures. For convenience, the available DS capacity numbers from Table 5 are provided here: the DS D3.0 capacity during coexistence phase is 0.41 Gbps, the DS D3.1 capacity during coexistence phase is 0.51 Gbps, and the DS D3.1 capacity after coexistence phase is 2.03 Gbps. Similarly, the US network capacity numbers are obtained from Table 6, where the US D3.0 capacity is 0.08 Gbps during coexistence phase, the US D3.1 capacity during coexistence phase is 0.19 Gbps, and the US D3.1 capacity after coexistence phase is 0.4 Gbps.

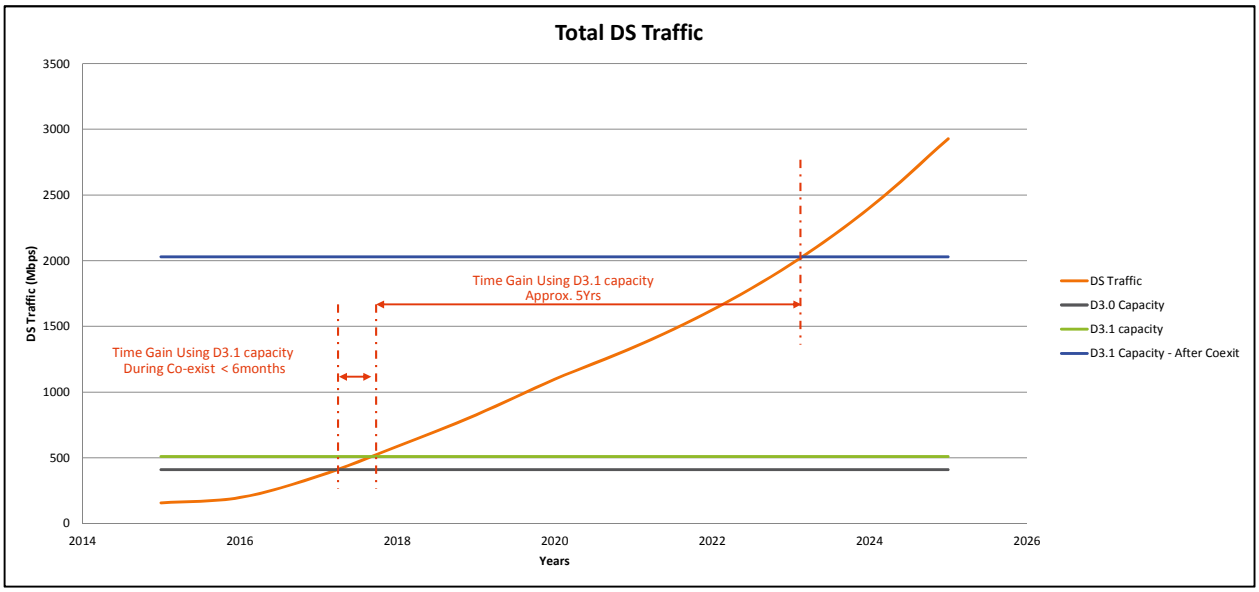


Figure 13. DS traffic demand and network capacity alternatives for Optus networks

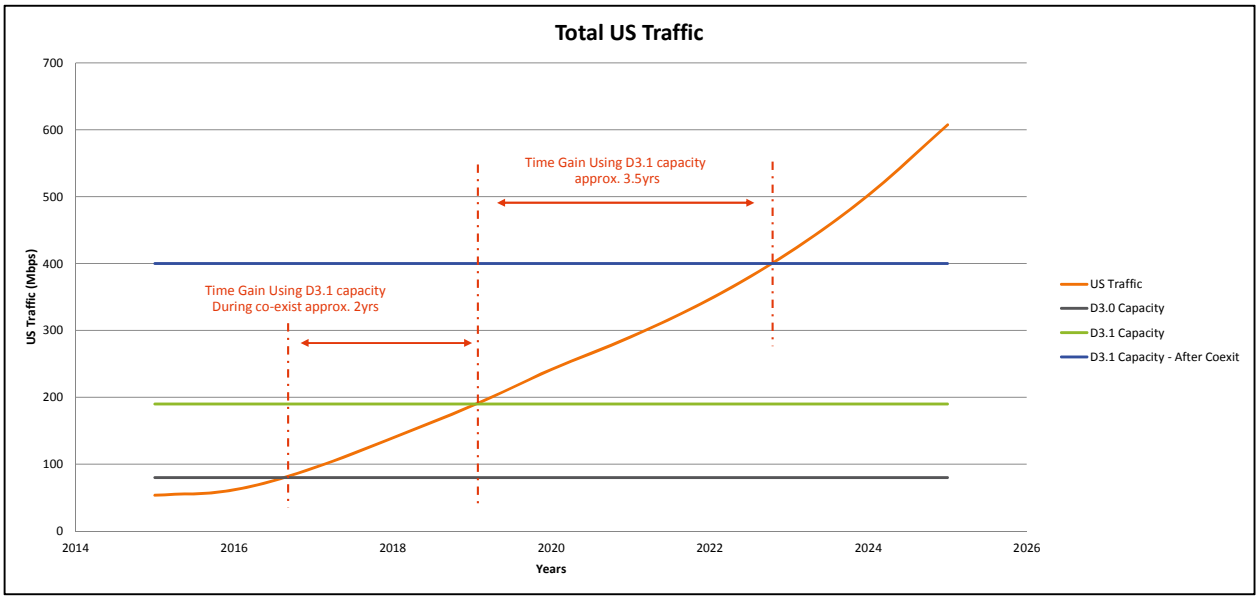


Figure 14. US traffic demand and network capacity alternatives for Optus networks

The results in Fig. 13 show the network capacity when DOCSIS 3.0 is deployed will be exhausted in 2017. Moreover, during the coexistence phase in 2017 to 2018 time frame, deploying DOCSIS 3.1 delays the need to perform node splits by 6 months. Although this may not look significant, it is definitely beneficial given that the coexistence phase lasts only for 18 months. The analysis in Fig. 13 suggests that it is important to start the coexistence phase early for as many nodes as possible to capitalize on the low amounts of DS traffic and which will lead to an early finish of the coexistence phase leading to more available spectrum which translates to significant DS capacities using DOCSIS 3.1. In this case, the amount of additional capacity that DOCSIS 3.1 can offer when full access to

the spectrum is available is significantly higher than the capacity that DOCSIS 3.1 can offer with the limited spectrum during the coexistence phase. The difference is worth approximately 5 years!

As for the US, the findings in Fig. 14 suggest that the US traffic demand may trigger a node split in 2016 (i.e., earlier than the DS) if DOCSIS 3.0 is used. However, if DOCSIS 3.1 is used, the US traffic would present less pressure to perform node splits than the DS did. In particular, deploying DOCSIS 3.1 in the US can delay the need for node splits by about 2 years which is beyond the coexistence phase. After the coexistence phase, more spectrum is available and therefore the available US capacity using DOCSIS 3.1 is significant (i.e., worth an additional 3.5 years after the end of coexistence phase).

Similar analysis was performed for the Telstra network and the required DS and US capacity results are shown in in Fig. 15 and Fig. 16, respectively. Observe that the network capacity numbers for the DS and US were obtained from Table 5 and Table 6 as was done earlier with the Optus analysis. Note that the pressure to perform node splits in the Telstra network is significantly less than that in the Optus case because of smaller number of HHP per node.

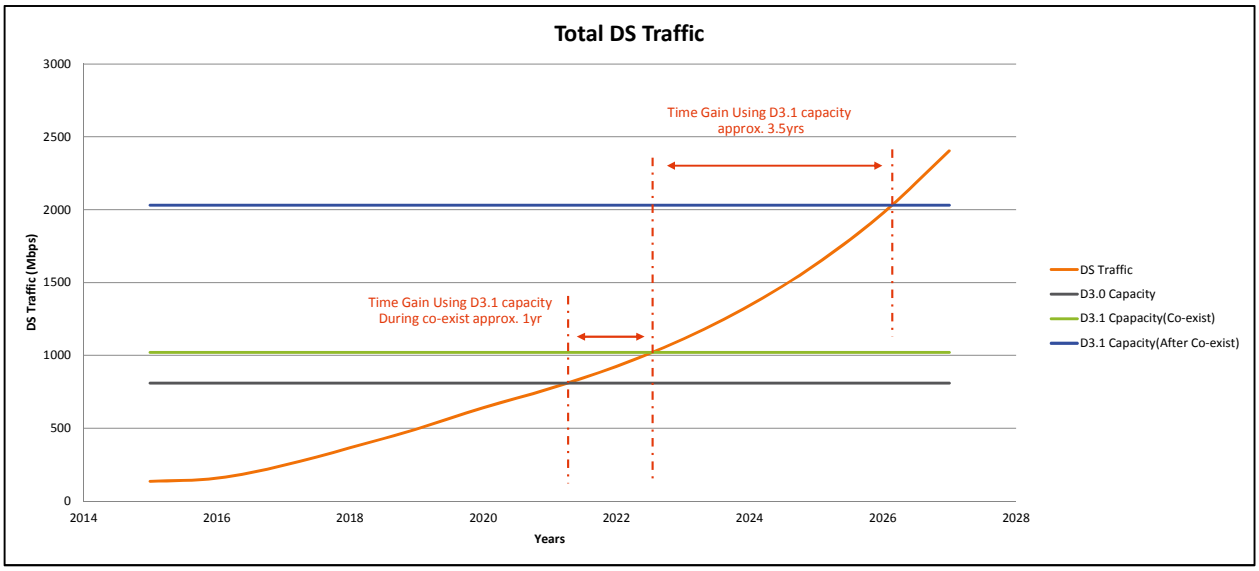


Figure 15. DS traffic demand and network capacity alternatives for Telstra networks

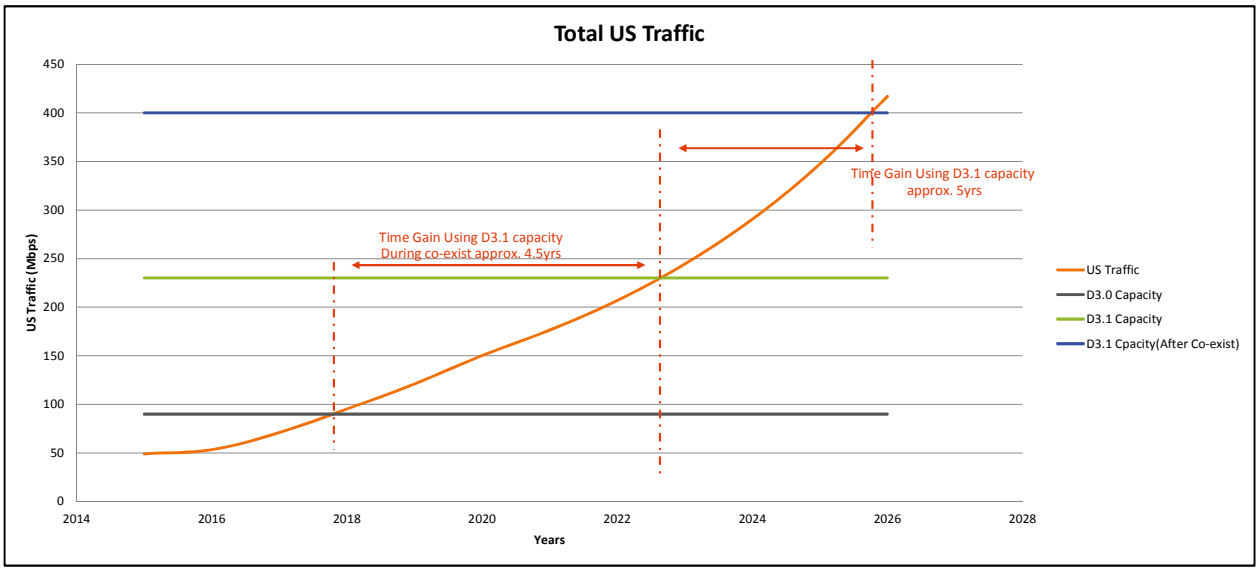


Figure 16. US traffic demand and network capacity alternatives for Telstra networks

HHP Sensitivity Analysis

While Optus and Telstra networks had only one difference in the analysis assumptions which was the HHP per node parameter, they had very different results in terms of the value of DOCSIS 3.1 in regard to the number and timing of node splits operations. Therefore, more analyses were performed corresponding to different HHP per node scenarios, where same assumptions as before were used for the other parameters. The DS and US results are illustrated in Fig. 17 and Fig. 18, respectively. Observe that these figures also contain the throughput curves corresponding to the Optus and Telstra networks.

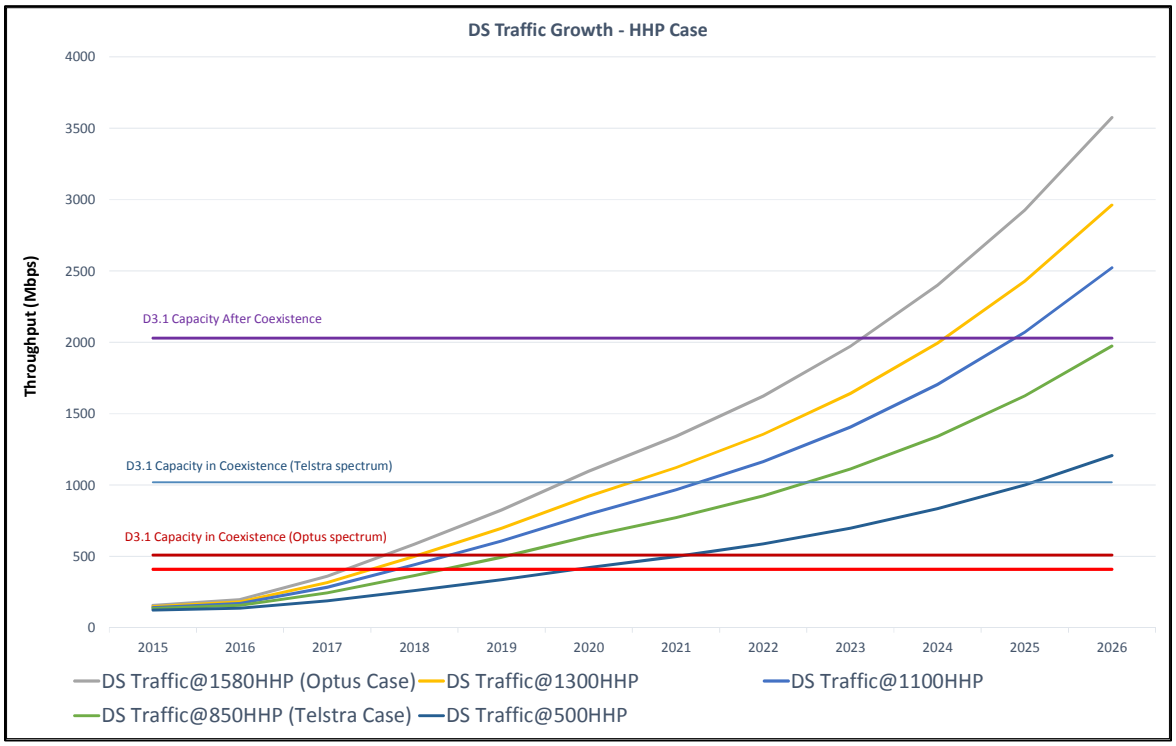


Figure 17. HHP per node sensitivity analysis for DS capacity

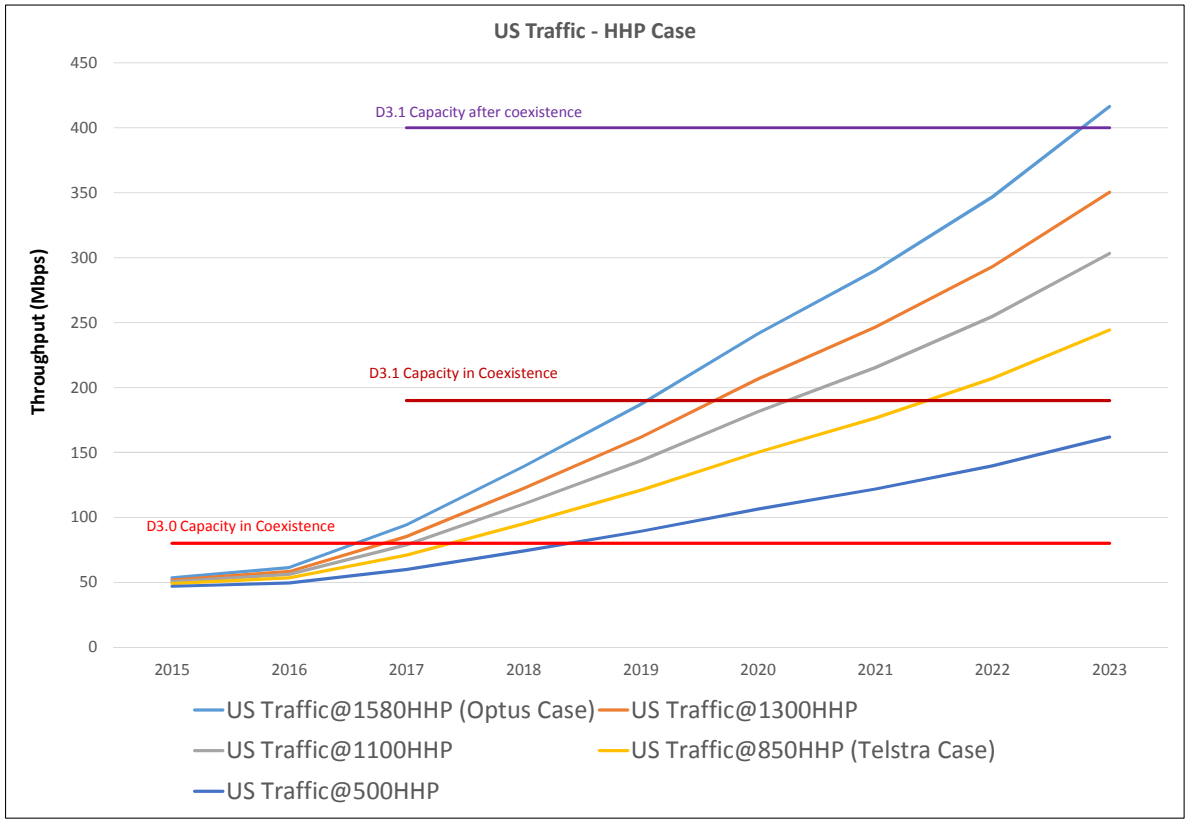


Figure 18. HHP per node sensitivity analysis for US capacity

The analysis that was presented thus far in this section assumed a CAGR of 0% for the DS and US Tmax. In other words, it is assumed that the highest DS and US peak rates will stay at 100 Mbps and 40 Mbps, respectively, for a few years. This may appear to be an unrealistic assumption. However, if the selective subscriber migration strategy is used to accommodate the highest tier subscribers, most other subscribers will likely stay satisfied with the peak rates specified above. Also, the analysis was mostly concerning the near future where studying the effect of DOCSIS 3.1 in reducing the required number of node splits is critical and therefore those rates appeared adequate to nbn for the next few years. Having said this, the analyses presented above were regenerated using 25% CAGR for both US and DS and the results are shown in Fig. 19 and Fig. 20, respectively.

It should be noted that the analysis approach described here can be used for any assumption made by the MSOs. It is actually very possible that each of the MSOs will have their own set of assumptions that fits their networks, subscribers, service offering plans, etc. Those assumptions can be applied to the methodology presented here in order to estimate the value that DOCSIS 3.1 can bring to their network.

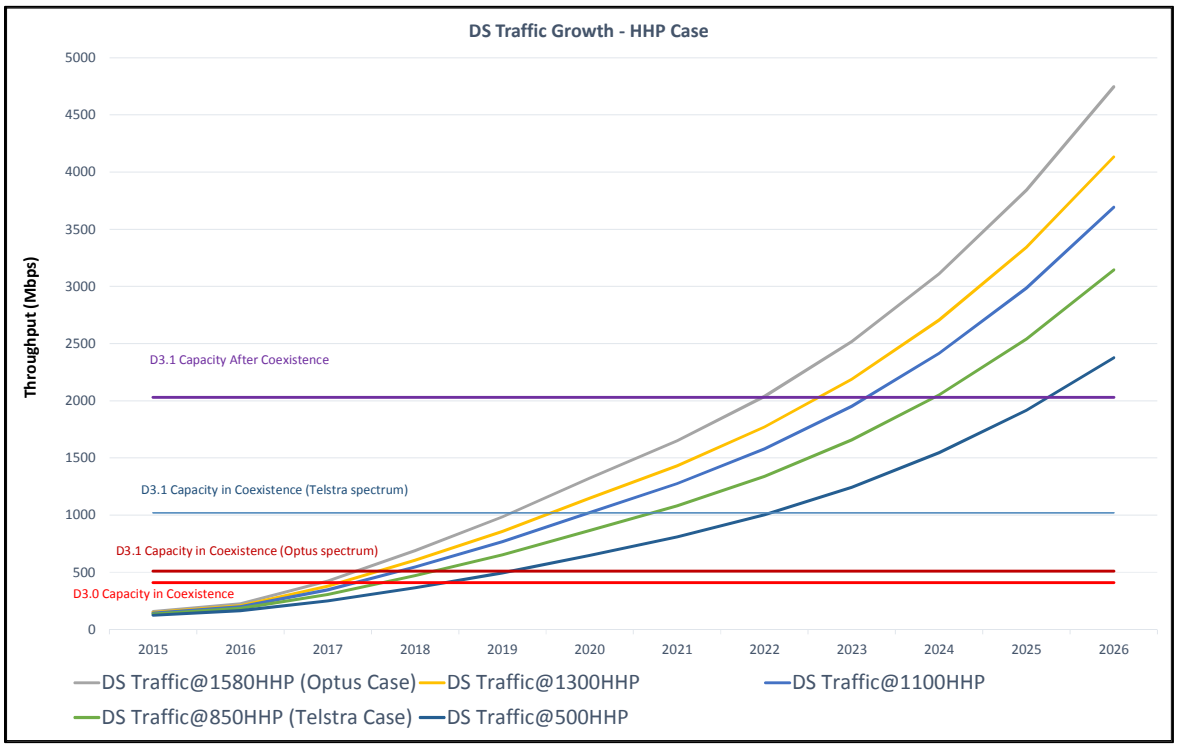


Figure 19. HHP per node sensitivity analysis for DS capacity (DS Tmax CAGR = 25%)

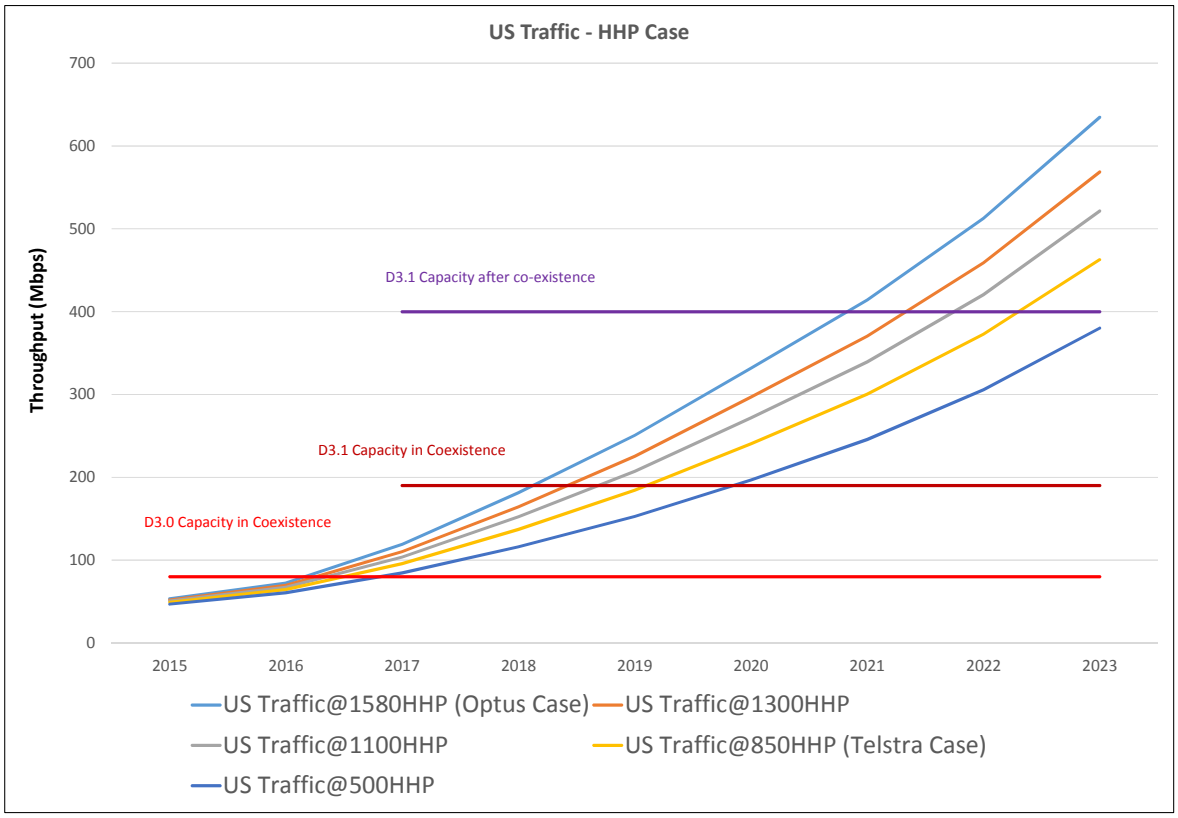


Figure 20. HHP per node sensitivity analysis for US capacity (US Tmax CAGR = 25%)

IV. HOW TO PROCEED WITH NETWORK MIGRATION?

The analysis in the previous section showed that DOCSIS 3.1 is an excellent tool that is available for the MSOs to extend their network capacities, reduce number of required node splits, and meet the offered traffic demand using existing infrastructure. As described in Section I, DOCSIS 3.1 is actually one of many tools that are available for the MSOs to choose from. These tools include node splitting & segmentation, selective subscriber migration, HFC vs. FTTH, RFoG vs. PON for FTTH, DAA vs centralized, etc. The optimal choice depends on the network parameters, offered demand and statistical distribution of subscribers among services, and MSO's restrictions (e.g., logistics/operational/resources constraints, infrastructure, budget, etc.). Therefore, a solution that perfectly works for one MSO may not be optimal for another MSO!

This section proposes a methodology that can be beneficial in selecting the appropriate near term transition path. Later in Section VI of this paper, multiple long term transition examples are described. For the near term transition paths, Fig. 21 illustrates a decision tree that can be considered when making a decision in response to a bandwidth problem. The decision state flow diagram starts when the network available capacity is not enough to meet the desired demand and/or address the competition. When that situation occurs, the decision state flow diagram proposes that the insufficient bandwidth issue is thoroughly understood before taking any actions because there is no one particular solution that fixes all problems. There are different solutions to different problems!

For instance, at high level, the network capacity can be deemed inadequate for two different reasons: The first scenario is when the network capacity is not enough to enable offering the high peak rate (Tmax) service which is required to address competition or to satisfy business customers and super users; The second scenario is when the network capacity not enough to accommodate the traffic demand offered during busy hours (Nsub*Tavg).

For the first scenario, when the network capacity is not adequate to offer high Tmax values, there are multiple solutions. For instance, DOCSIS 3.1 can be deployed, if not already, to get an instant boost in the network capacity without changing the outside plant. Observe that deploying DOCSIS 3.1 in the DS assumes that spectrum is available for DOCSIS 3.1 and CMTS/CMs that support DOCSIS 3.1 are available. If DOCSIS 3.1 is already deployed and the capacity is still not enough, there are two potential solutions:

1. Selective subscriber migration (via FTTH).

2. Plant upgrades (change split, deeper fiber, new equipment with support for larger spectrum, etc.).

It is important to study the cost as well as the logistic/operational complexity of performing selective subscriber migration vs. plant upgrades. For instance, if the number of business customers/super users is small and the cost/logistics of upgrading their paths to FTTH is manageable, then upgrading those portions of the network selectively and leaving the larger portion of the network untouched could be the best option. This not only allows satisfying business customers/super users by moving them to their own network infrastructure, it also alleviates the traffic and high peak rate service demand on the rest of the network; which in turn elongates the life of the network at minimum cost.

On the other hand, if the MSO decides the number of business customers/super users is large or the likelihood of the penetration percentage for that service is high, then it may makes sense to upgrade the whole network all at once. This case study needs to be performed for every portion of the network where this decision needs to be made because one decision for one portion of the network may not be appropriate for another portion.

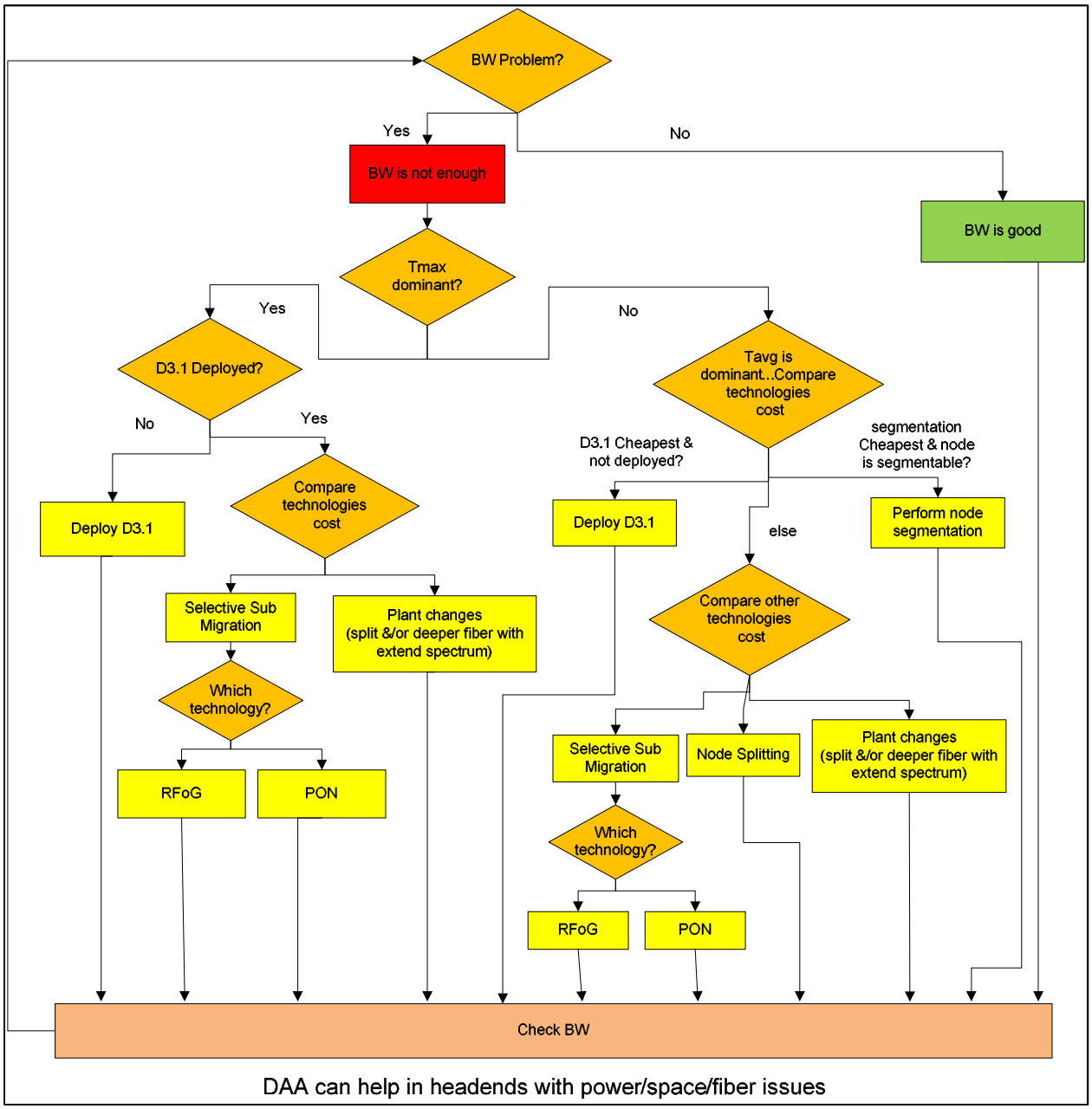


Figure 21. Proposed state flow diagram to make a decision in a response to a bandwidth problem

If the MSO decides that selective subscriber migration is more adequate than full plant upgrade, then there will be two alternatives for signal transmission and resource allocation for the FTTH infrastructure: RFoG and PON. The RFoG approach has multiple benefits such as:

1. There currently exist OBI-free solutions.
2. OBI-free solutions allow for multiple US transmissions and therefore uses the system capacity more efficiently.
3. OBI-free solutions can work with any CMTS and CM without any modification.
4. Works with any DOCSIS version including DOCSIS 3.1.

5. Can support future DOCSIS features like extended spectrum DOCSIS [5] [6] and full duplex DOCSIS [5] [9].

6. Reuses same back-office systems, CMTSs, CMs, and therefore protects the MSO's investment.

While RFoG has many advantages as described above, some MSOs may still select PON as their transition step if they believe it better fits their long term plans. Both approaches have their own pluses and minuses as discussed in other papers [10] [11]. The topic of comparing these technologies against each other is outside the scope of this paper, which mainly focuses on proposing the decision tree.

Regarding the second scenario where the network capacity is not adequate to accommodate the traffic offered during busy hours, different solutions are required. In particular, when that situation occurs, the MSO may have to select from the two options below after studying the cost of deployment and operational implications:

1. Deploying DOCSIS 3.1

2. Node Segmentations

Observe that the two options listed above are relatively cheap because they do not require immediate plant upgrades. DOCSIS 3.1 can offer an instant boost to the network capacity over the existing infrastructure and node segmentation can reduce the number of subscribers per service group and therefore increase the available capacity per subscriber. Note that node segmentation does not require changes to the physical plant if the node design is segmentable.

If the MSO has already deployed DOCSIS 3.1 and segmented the node (if segmentable), and the available capacity is still not adequate, then the MSO can further advance in the decision process and select from one of the following solutions:

1. Node splitting

2. Selective subscriber migration (via FTTH)

3. Plant upgrades (change split, deeper fiber, new equipment with support for larger spectrum, etc.)

All of the above approaches require physical changes to the plant. Therefore, a detailed business case study needs to be performed in order to select the most appropriate option which can offer the longest plant life at minimum cost. Again, part of this process is to decide whether to use RFoG or PON if selective subscriber migration or FTTH infrastructure is chosen as the appropriate solution.

Finally, the decision tree assumed that DAA network architectures can be beneficial in addressing headend power/space issues as well as the limited number of wavelengths that are supported by the AM optics. Since the capacity gain from transitioning to DAA networks can be limited depending on the network parameters and/or the TDHM bottle neck, increasing the network capacity was not assumed to be a key driver for the MSO to move to DAA network architecture.

V. DEPLOYMENT OF DOCSIS 3.1

The previous section provided a proposed decision tree that can be considered when decisions are needed to perform near term transitions to the existing network when the available capacity is not enough. One of the available tools in the decision tree was the deployment of DOCSIS 3.1. This section provides more insight regarding DOCSIS 3.1 deployment due to its importance as multiple MSOs are currently thinking of or have already started migrating to DOCSIS 3.1.

As implied in the previous section, deployment of DOCSIS 3.1 does not have to be done across the whole network at once. It can be done as needed and where needed. Once DOCSIS 3.1 has been selected as the appropriate next step for the node or service area, then it can be deployed gradually as the functionality is verified via lab trials, field trials with friendly customers, and finally using production offering.

The first question that comes to mind when thinking of DOCSIS 3.1 deployment is how to come up with spectrum for DOCSIS 3.1 channels. The answer could be different for the DS and US directions. For the DS, the MSO can reclaim spectrum via:

1. Phasing out analog channels

2. Use IPTV for video delivery

3. Use more efficient encoding schemes for video

4. Use of switched digital video

5. Transfer spectrum between services (e.g., MPEG VoD to DOCSIS)

Beyond spectrum reclamation, other alternatives can be utilized. These include:

1. Turning off some of the legacy SC-QAM channels and assigning the spectrum to DOCSIS 3.1 (especially if heavy and super users are moved to DOCSIS 3.1, i.e., selective subscriber migration to DOCSIS 3.1 is performed).

2. Us of the top part of the spectrum that could be noisy and interference-prone for DOCSIS 3.0 operation but could be usable by DOCSIS 3.1.

3. Use of roll off spectrum which can be utilized by using DOCSIS 3.1 features such as Multiple Modulation Profiles (MMP) and Variable BitLoading (VBL).

The DOCSIS 3.1 DS channel does not have to be 192 MHz wide. It could be smaller [12]. A small portion of the spectrum assigned to DOCSIS 3.1 can still be beneficial because DOCSIS 3.1 CMs can receive bonded traffic over OFDM and legacy SC-QAM channels. As spectrum is secured for DOCSIS 3.1 channels, the MSO can simply add the OFDM channel to the DOCSIS 3.1 CM's Receive Channel Set (RCS) and use the OFDM channel.

If a small number of CMs experience issues in locking to the OFDM channel, those will still be connected to legacy channels and trouble-shooting can be performed without affecting the subscribers' service. If a large number of DOCSIS 3.1 modems experience issues with the OFDM signal, the MSO may choose to turn off that signal all together and get to the bottom of the issue before putting the OFDM channel back in service.

As for the US, several techniques can be utilized to secure spectrum for OFDMA channels. These include:

1. Use of DOCSIS 3.1 Time and Frequency Division (TaFD) multiplexing to share the US spectrum between legacy and OFDMA channels. This feature allows dynamic allocation of US spectrum to SC-QAM and OFDMA as needed. That is, the same part of the spectrum can sometimes be used for SC-QAM and some other times used for OFDMA [13]. This feature can be beneficial when no or little spectrum is available for OFDMA channels.

2. Use of the lower and/or top part of the spectrum that could be noisy and interference-prone for DOCSIS 3.0 operation but could be usable by DOCSIS 3.1 OFDMA channels.

3. Use of roll off spectrum which can be utilized by utilizing DOCSIS 3.1 features such as Multiple Modulation Profiles (MMP) and Variable BitLoading (VBL).

4. Turning off some of the existing legacy SC-QAM channels and assigning the spectrum to DOCSIS 3.1 (especially if heavy and super users are moved to DOCSIS 3.1 OFDMA channel, i.e., selective subscriber migration to DOCSIS 3.1 is performed).

If, for any reason, the CM is not able to range and register on an OFDMA channel, it will likely try SC-QAM channels. Therefore, the CM will probably not be completely offline due to issues with an OFDMA channel. After registration, there are two scenarios:

1. If SC-QAM and OFDMA channels do not overlap and/or the TaFD feature is not available (i.e., they overlap but exclusions within the OFDMA channels are used to accommodate SC-QAM channels), then the traffic from the SC-QAM and OFDMA channels can be bonded together. That is, the Transmit Channel Set (TCS) of the CM will contain OFDMA and SC-QAM channels that are bonded together. In this case, if the OFDMA channel gets impaired, the traffic gets scheduled on the SC-QAM channels (i.e., partial service).

2. If the TaFD feature is used, then a slightly different approach can occur. There are two cases:

    a. The TCS of the CM contains only an OFDMA channel that services the SF of the CM. If the CM experiences issues with the OFDMA channel, it will likely reset and tries to range on the available SC-QAM channels.

    b. The TCS of the CM contains an OFDMA channel plus one or more SC-QAM channels that act as backup to the OFDMA channel in case of a failure. For normal operation, the traffic is scheduled on the OFDMA channel only. If the OFDMA channel gets impaired, then the CMTS schedules grants on the SC-QAM channels. This operation does not require a modem reset. Observe that the CMTS can add more SC-QAM channels via the DBC operation if needed.

VI. LONG TERM NETWORK MIGRATION

As MSOs plan their long term migration, they can take different approaches in reaching their end goal architecture. In particular, they can take one of the following evolution approaches:

1. Gradual evolution

2. Large-step evolution

3. Very large-step evolution

Different scenarios illustrating examples of the various approaches listed above are provided in Fig. 22 to Fig. 25. While only a few examples are shown, many alternatives can be envisioned. This paper encourages a gradual approach for network migration to capitalize on the investment protection and perform upgrades only where and when needed. This methodology was presented via the analyses and decision state flow diagram in Sections III and IV, respectively. Observe that the decision state flow diagram presented earlier can be considered as the MSO takes gradual transition steps leading to the end goal network architecture.

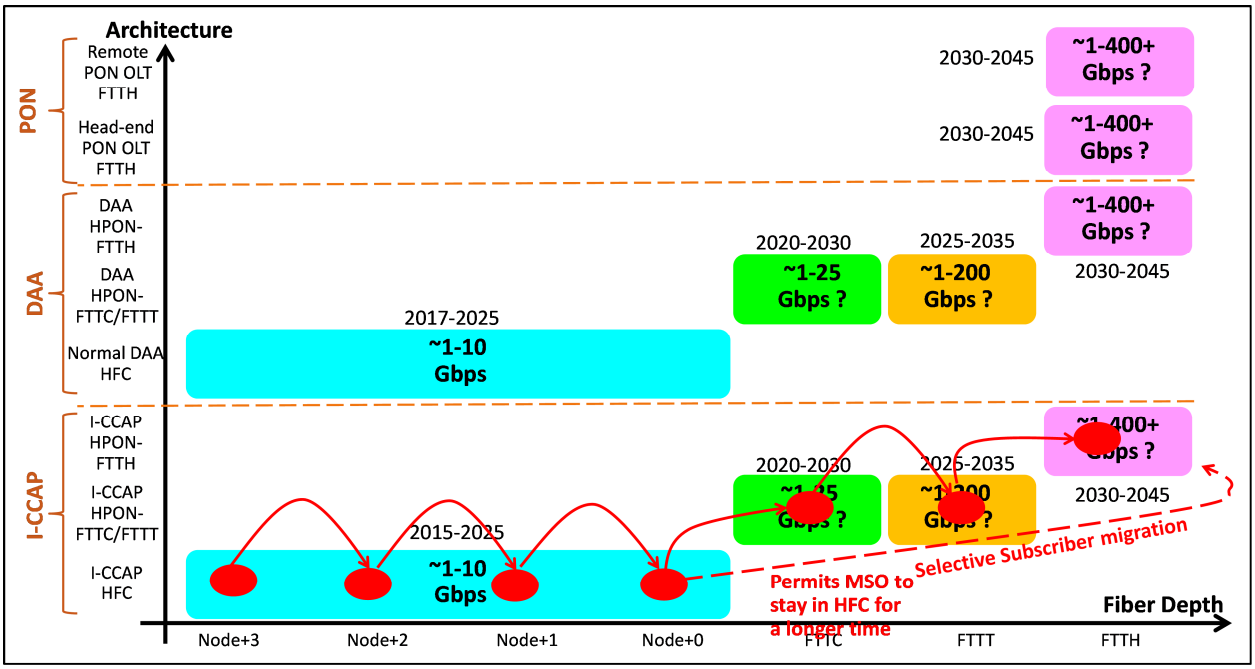


Figure 22. Example of gradual migration strategy – Scenario 1.

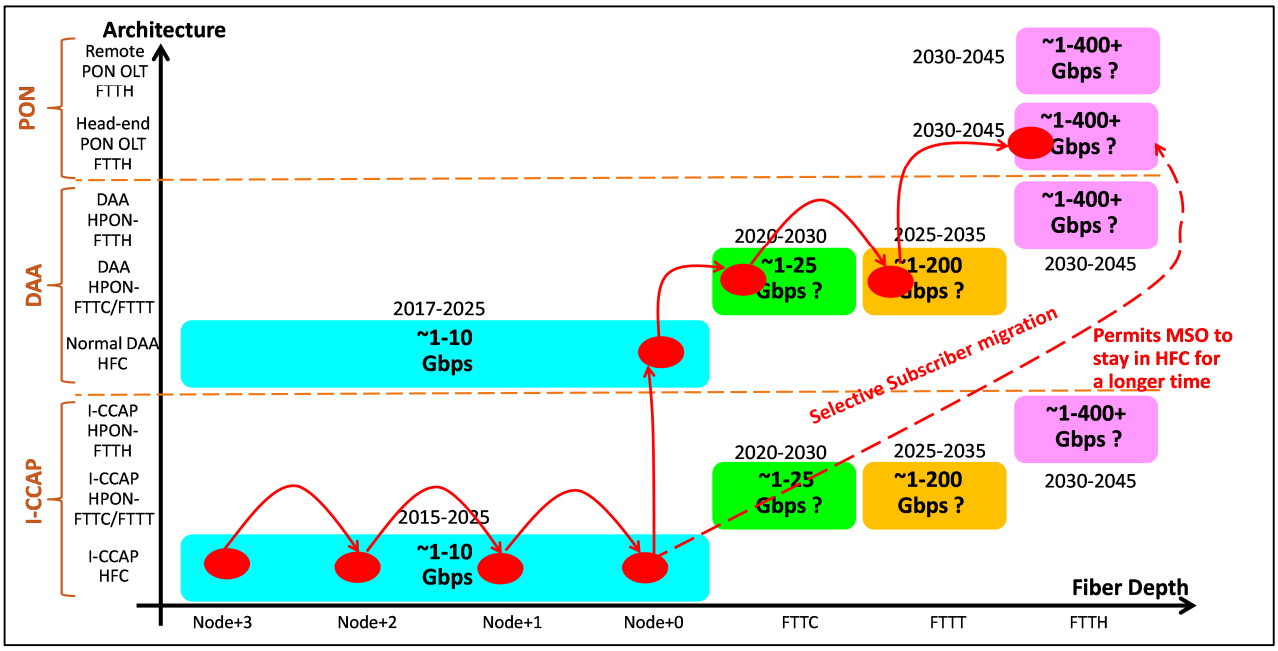


Figure 23. Example of gradual migration strategy – Scenario 2.

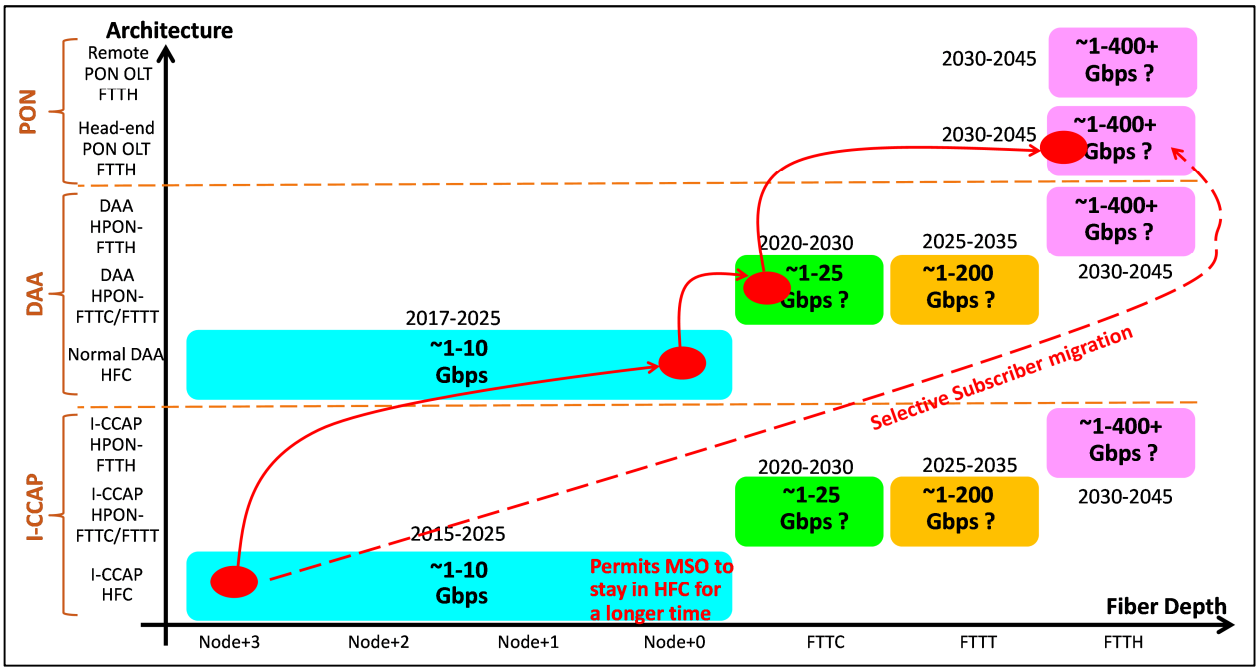


Figure 24. Example of large-step migration strategy.

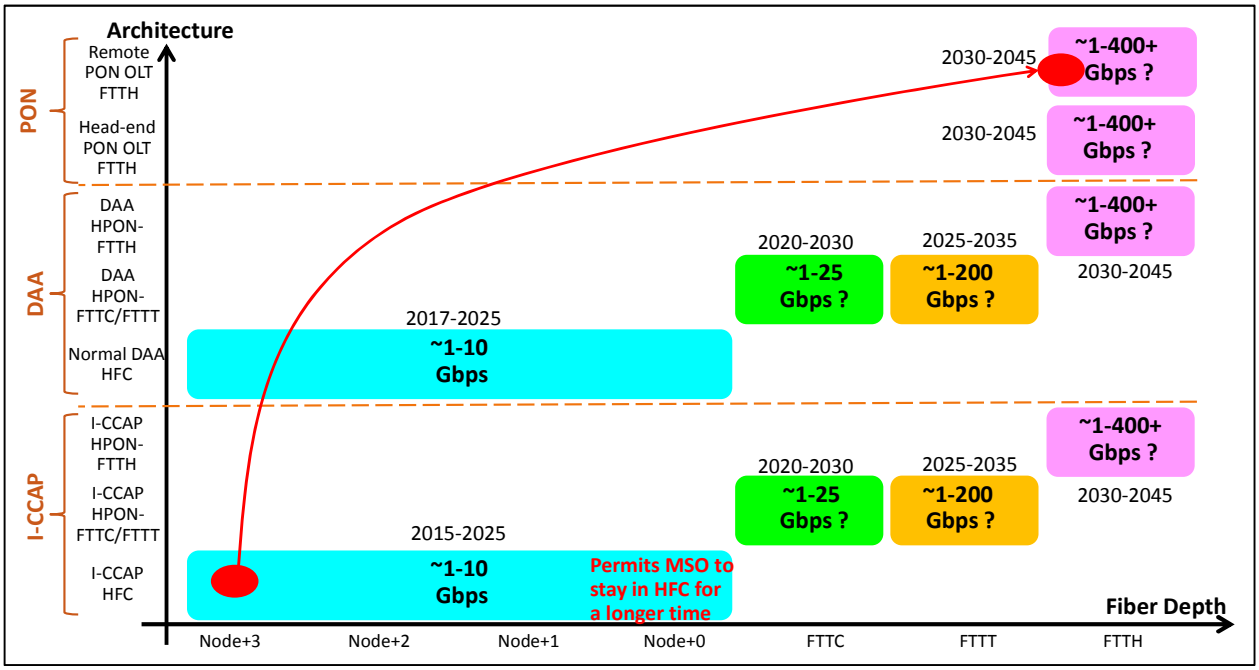


Figure 25. Example of very large-step migration strategy.

VII. CONCLUSIONS

The article analyzed the value of DOCSIS 3.1 in increasing the network capacity and reducing the OPEX costs. Based on real-world scenarios, it was shown that deploying DOCSIS 3.1 can delay node-split operations needed to address inadequate DS capacity by 6 months to 1 year. As for the US, DOCSIS 3.1 deployment can delay node split operations to address inadequate US capacity by 2-4.5 years. The article described different

technology enablers including DOCSIS 3.1, node splits/segmentation/selective subscriber migration, plant upgrades, RFoG vs. PON, DAA vs. I-CCAP. A decision tree state flow diagram was proposed to select the appropriate combination of the above technology enablers to form an optimal near term transition path. Finally, the paper described the network migration process to DOCSIS 3.1 in the DS & US directions and also showed different long term network migration scenarios.

ACKNOWLEDGMENT

The authors would like to thank their colleagues for their great contributions to this paper including Geoff Lee, Jason Cheung, Saad Ahmed, and Greg Gohman.

REFERENCES

[1] Ayham Al-Banna et. al., "The Spectral Efficiency of DOCSIS® 3.1 Systems," SCTE Cable-Tec Expo 2014.

[2] Ayham Al-Banna et. al., "Measurement-Based EoL Stochastic Analysis and DOCSIS 3.1 Spectral Gain," Spring Technical Forum conference INTX 2015.

[3] Venk Mutalik et. al., "Value of Transparency in the Evolution of HFC Networks: How Optical Transparency Enhances Capacity, Promotes Innovation and Preserves Infrastructure in HFC Networks," SCTE Cable-Tec Expo 2013.

[4] Mike Emmendorfer et. al, "Estimating Downstream Performance and DOCSIS 3.1 Capacity in CAA and DAA Systems," Spring Technical Forum conference INTX 2015.

[5] Tom Cloonan et. al., "Using DOCSIS to Meet The Larger BW Demand of the 2020 Decade and Beyond," Spring Technical Forum conference INTX 2016.

[6] Tom Cloonan et. al., "Lessons From Telco & Wireless Providers: Extending the Life of The HFC Plant with New Technologies," Spring Technical Forum conference INTX 2015.

[7] Tom Cloonan et. al., "Simulating the Impact of QoE on Per-Service Group HSD Bandwidth Capacity Requirements," SCTE Cable-Tec Expo 2014.

[8] Mike Emmendorfer et. al., "Nielsen's Law vs. Nielsen TV Viewership for Network Capacity Planning," Spring Technical Forum of the Cable Show 2014.

[9] http://www.cablelabs.com/full-duplex-docsis-3-1-technology-raising-the-ante-with-symmetric-gigabit-service/

[10] Venk Mutalik et. al., "The Yin and the Yang of a Move to All Fiber: Transforming HFC to an All Fiber Network while Leveraging the Deployed HFC Assets," Spring Technical Forum INTX 2015.

[11] John Ulm et. al., "HFC Transformation to FTTP: The Role of RFoG, PON and Hybrid solutions," SCTE Cable-Tec Expo 2015.

[12] CableLabs, "Data Over Cable Data Interface Specifications DOCSIS 3.1 Physical Layer Specification," CM-SP-PHY-v3.1-I08-151210.

[13] Ayham Al-Banna et. al., "US Scheduling In The DOCSIS 3.1 Era: Potential & Challenges," SCTE Cable-Tec Expo 2015.

ABBREVIATIONS & ACRONYMS

CAGR	Cumulative Annual Growth Rate
CM	Cable Modem
CMTS	Cable Modem Termination System
DAA	Distributed Access Architecture
DOCSIS	Data Over Cable Service Interface Specifications
DS	Downstream
FTTLA	Fiber To The Last Active
FTTC	Fiber To The Curb
FTTH	Fiber To The Home
FTTT	Fiber To The Tap
HFC	Hybrid Fiber Coax
H-PON	Hybrid PON
I-CCAP	Integrated-Converged Cable Access Platform
IPTV	Internet Protocol Television
MER	Modulation Error Ratio
MMP	Multiple Modulation Profiles
MPEG	Moving Picture Experts Group
MSO	Multiple System Operators
OBI	Optical Beat Interference
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OPEX	Operational Expenses
pdf	Probability density function
PON	Passive Optical Network
QoE	Quality of Experience
RCS	Receive Channel Set
RFoG	Radio Frequency over Glass
SC-QAM	Single Carrier-Quadrature Amplitude Modulation
SNR	Signal to Noise Ratio
SSM	Selective Subscriber Migration
TaFD	Time and Frequency Division Multiplexing
Tavg	Average Throughput per subscriber during busy hour
TCS	Transmit Channel Set
TDHM	Tap, Drop, Home network, Modem
Tmax	Highest peak rate
US	Upstream
VBL	Variable Bit Loading
VoD	Video on Demand

Reexamining The Design of Next Generation Video Streams

Yasser Syed, PhD
Comcast Distinguished Engineer
T&P/VIDEO Architecture Group
Comcast Cable

Abstract

The popularity of on-demand viewing means that content is being stored and retrieved more than ever before. Oftentimes, it is video playout on handheld devices, and resultant battery drain, that impacts performance more so than any lack of bandwidth. Similarly Content Delivery Networks/CDNs can impair viewing because too many copies of video packets are being stored. Yet for certain live events, the audience is larger than ever before.

Video streams traditionally have been designed for compression efficiency and with a specific focus on linear systems. This has affected codec selection, transmitted resolution, GOP length, random access, conditional access and hardware. Such linear-related restrictions have done a good job at balancing good linear video experience with good content delivery. But while traditional video streams may be suitable for linear systems, they may actually cause scaling issues for newer approaches that deliver IP video content over CDNs, mobile networks, mesh systems, and other unmanaged infrastructures.

This paper examines how video codec design could be optimized for the rising gamut of IP delivery mechanisms, and how a re-examined codec could impact the design of next generation video streams using IP delivery.

INTRODUCTION

The IP transformation of video systems is already happening with a transition between Linear-based QAM systems to CDN-based IP systems. Along with this transition is the shift to non-linear viewing models that can be viewed not just on television displays, but also on multiple, smaller display devices owned by the consumer. This results in increased demands for storage capacity -- which will, in the future, rival the bandwidth delivery constraints already occurring in IP unicast delivery.

Figure 1- Delivery Over Hybrid Networks

Along with this IP transformation is the expansion of new features for video services. This includes features based on new display technologies, such as Ultra High Definition/ UHD and High Dynamic Range/ HDR. These additional features can increase the amount of content versions that are stored as well as adding pressure to bandwidth capacity.

In addition, video services are also transversing across hybrid delivery networks (cable, WiFi, mobile—both over managed and unmanaged bandwidth) and are expanding into different devices (television, PCs and portable devices such as laptops, tablets, and smart phones). The over-arching benefit of switching to an IP infrastructure is robustness in delivery design, despite variations in networks and device types. This allows for networks and devices to improve without significantly changing the delivery approach.

DESIGNING TRANSITIONS

To aid in this transition to an IP infrastructure, a separated transcoder/packager structure is being developed. The transcoder prepares the content into multiple Elementary Streams (ES) wrapped by a delivery transport structure (in this case, an MPEG2-Transport Stream/TS structure). The transcoder also marks up the stream to create virtual segments (or Encoder Boundary Points /EBPs) that can carry timing, boundary, and labeling points. The stream output from the transcoder is known as an Adaptive Transport Stream

(ATS). The ATS is then ingested by a packager, which can be co-located with the transcoder, or located further down in the network. The packager uses the marked-up points to create a manifest. It also segments the content into pieces in a single media format, similar to a DASH Manifest with MPEG-TS segments. This format is then called by the client or CDN and is transmuted in a lightweight process into the particular client-requested manifest type and media segment format (HLS, DASH, HDS, HSS, ISOBMFF).

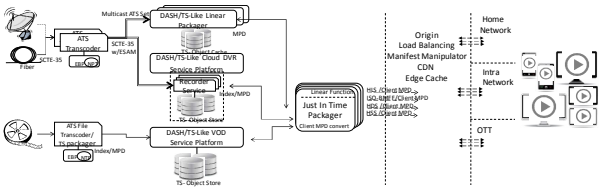


Figure 2- Transcoder & Packagers in the IP delivery Network

This transcoder/packager structure assists in the IP transformation in three ways: 1) by allowing access to the content in a stream format for point-to-point, intra-network distribution (push model); 2), by creating a segmented and indexed format for long term storage (pull model w/ manifest); and 3) by modification to whatever client-specific adaptive streaming format needed for loading into CDN (pull model w/ manifest). The combination of these formats allows us to use existing equipment to encode and check for video quality, store content in a single format, and only output in several formats when content is pulled out to the client delivery network. This also allows us a pathway to start utilizing a pull-model-based IP delivery architecture.

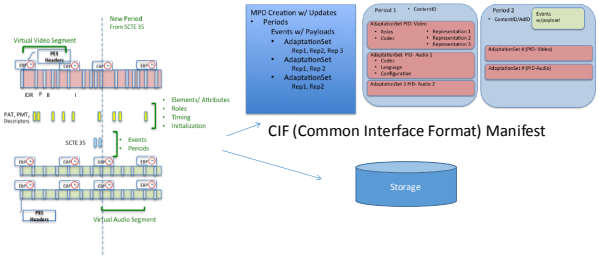


Figure 3- Conversion from Push to Pull Model Using Packagers

MANAGING OPTIMIZATIONS

It is usually the case that operational factors in switching to new technologies or different delivery systems can impede adoption of new delivery designs. While moving through the IP transformation, both for non-linear viewing and for new service expansion there will be a need to focus on optimizations in the following areas:

- Reducing the amount of identical copies and perceptually similar variations of content stored in the network (storage efficiency)
- Localizing content traffic demands through the use of well-designed CDN and edge servers (bandwidth efficiency)
- Reducing the volume of serially-dependent transactional traffic in the network (throughput efficiency)
- Extending battery lifetime of portable devices through the optimization of decoding complexity (display life optimization)
- Increasing the quality bitrate ratio of content streams to allow for video service expansion (future feature optimization)

In the new delivery system, the tradeoff between how to store content and how to transmit content will need to be considered simultaneously, because being efficient in one medium may cause waste in the other. Additionally, with IP delivery, the requests

caused by content selection will increase, because each fragmented piece of content is requested separately -- and if missed, then re-requested -- which subsequently will cause more content to be sent across the network. With respect to devices, extending battery lifetime is not only important to consumers, but operationally important as well, to avoid disruption of services simply because the device powers down frequently. Lastly, with the additional features of 4K/8K and HDR that are being added to IP video services, better video compression will be needed. Better compression will help in avoiding any lack of perceived functionality of the service (i.e. UHD with Stereo Audio) because of its inability to scale up with increased traffic. Any of these factors can easily affect the deployment of video services over IP delivery systems, if the system does not accommodate for them, and as traffic volume increases.

USING SEGMENTED CONTENT

To address these transformations, Adaptive Bitrate (ABR) Streaming technologies are being used in combination with CDN delivery structures. The benefit of this is the ability to deliver bitsize chunks of video in a transactional delivery protocol for data IP packets (e.g. HTTP). This allows for the payload to be switched through a manifest indicating the fragment options that are available to the player. It converts video into packetized data IP delivery that can easily be routed through data handling devices using port 80 designations.

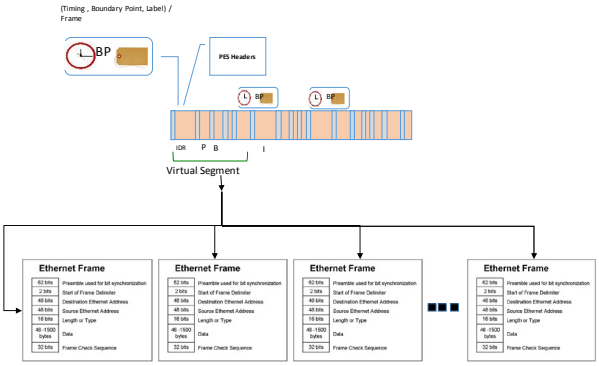


Figure 4- Fitting a Segment into IP Encapsulation

The video stream itself is encoded into a specific codec format, which in itself is a NAL (Network Adaptation Layer) unit. The packetized video stream is buffered and synchronized for both compression and constant framerate playback.

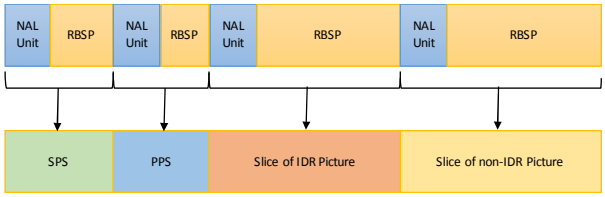


Figure 5- Video Stream Packetization

When an individual client makes a serial request for a video stream, in a best effort network, the trick to ensuring smooth playback is to deliver that stream --which has been encapsulated into multiple IP frames -- while being fast enough to keep the video CPB buffer filled. The rate of the buffer being filled is dependent on the fragment rate, since the whole fragment needs to be received in order to load it into the buffer. If the network is not fast enough to keep the buffer filled, then a lower bitrate fragment is requested, if the manifest knows that it is available to the player. The benefit of ABR is that it allows a consistent way to do video streaming over different network conditions and devices.

In first-generation ABR technologies, the design of the video streams was not drastically changed from the GOP design used in linear QAM streams. Rather, it was adapted and restricted to fit ABR technologies. At the beginning, in the mid-to-late 2000s, some of the earliest implementations, video Streams were adjusted around the following guidelines:

- GOP duration and the fragment duration were tied together to make

the linear stream or VoD file easily divided into fragments by a packager

- All content segments were of the same duration
- All segments were independently decodeable
- All ABR streams had to have the same GOP structure (or a decimated version of it) to allow easy bitstream switching at every fragment
- All streams were constant bitrate streams to allow the fragment to be switched to a known bitrate
- All segments were meant to be delivered to a client player (so a server/ client relationship)

Improvements came later to allow for one or more GOPs to be contained within a segment duration. This was accomplished by creating a signaling construct that identified specific IDR frames in the stream, so as to indicate the beginning of the segment and allow the codec design flexibility. The result was increased quality. Notably, not all segments had to be the same duration (most of them, but not all of them), which allowed for splice-accurate ad insertion. Lastly, the delivery of segments did not have to be just from a server to a client player. This introduced the idea of storing content in one ABR format and allowing a packager to rewrap it into other ABR formats. Further work is occurring in second generation ABR technologies and corresponding video stream conditioning, particularly in standards development organizations and forums such as MPEG-DASH, MPEG Systems (virtual segmentation), JCT (HEVC), the DASH Industry Forum (DASH-IF), and SCTE WG-7.

REEXAMING VIDEO STREAM DESIGN

As this second generation ABR work is occurring, it is timely to consider how the video stream could be adapted to fragmented IP delivery, rather than vice versa. It comes down to how we handle video streams of the future. Granted, there are transitory issues to consider, but rethinking how we will use video streams in an IP delivery infrastructure could bring about a fresh approach in the design of the video stream. The rest of this paper asks questions that encourage the reader to relook at the design of the video stream, including what activities are already occurring in these areas, and how some modifications may help in handling new demands in next generation systems.

What if we accessed content differently?

With fragmented content, manifests, and the existence of several MBR (multi-bitrate) streams for content, there are many possibilities for new approaches to content access. Content could be indexed to handle program starts/end, chapter starts/end, and ad starts/ends. These can be indicated in the manifests through the period mechanism or url mechanisms, and triggered in the encoded stream through the SCTE 35 constructs and labeling af_descriptors (ATS/DVS 1196). The player could access content through a set of content indices, as described in the manifest. This allows the viewer to intelligently access the content by skipping to meaningful points (seek vs scan) -- but it is also dependent on how much of the content is indexed before and during the transcoding step.

At a more myopic level, random access of stream fragments takes on a new meaning with the existence of MBR streams. With several streams present, random access could now be defined differently in each separate MBR stream. If one needed quick access into the content, a representation could be selected where random access would happen at each fragment boundary. In another representation,

the quality of the stream could be improved by reducing the number of random access points in the stream. Bitstream switching can still happen between representations but not at every fragmen. The effect of this could be twofold: 1) Either longer duration fragments, or fragments that may not be independently decodeable because they are dependent on the preceding fragment in sequence (playthrough on higher bitrate representations) or 2) shorter duration fragments that are independently decodeable (random access on lower bitrate represenations). This MBR approach could re-introduce useful "new/old" concepts, such as open GOP, which was once needed to increase the quality of the stream by allowing for larger fragments to be broken up, which subsequently allows some flexibility in fragment delay at the player buffer. The new requirement would determine some level of alignment of different streams in the MBR set.

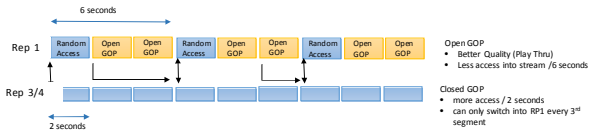


Figure 6- Setting different Random Access Points for Each Representation

Another consideration in video access is the level of handling of the stream. Because video services need to extend beyond just a server/player distribution, and into hybrid networks and backbone infrastructures, there are several intermediary devices that need to handle, store, or route the fragments or stream. These devices are not necessarily involved in the playout of the content. Partial encryption of the stream allows for this handling, where header or marker information is in the clear, but actual video frame content is encrypted. This "partial reveal" allows for stream handling to happen, without the access for playout. Common encryption takes this a step further, in that it doesn't need two different sample variations of the remaining parts of the still-encrypted content for storage and distribution. This allows for devices to decrypt and re-encrypt for each mode, while still allowing DRM policy to be applied in both client-based and transactional models. This reduces the number of different versions of content stored, while also reducing the amount of transactional traffic in the network to decrypt and re-encrypt content.

What if we stored differently than we packaged?

Traditional storage systems were mostly file-based designs, which kept the number of items in the storage hierarchy small -- even though each item was rather large and non-uniform in size. Access to files at faster speeds required increasingly complex read/write/ or caching strategies. Additionally, file integrity required redundant approaches, like RAID-striped storage, to preserve content in the event of storage equipment failure.

New object-based storage systems have started increasing in popularity, given recent technology improvements that allow handling of lots more items, while flattening the storage hierarchy, and increasing the speed of read/write access to content. File integrity in object-based storage technologies moved from RAID-based striping approaches to the storage of redundant copies, in different pieces of hardware, across federated-based storage systems (which is very cloud-friendly). Also, moving to an object-based system allows for faster read/write access, which can accommodate easier parallel retrieval of content.

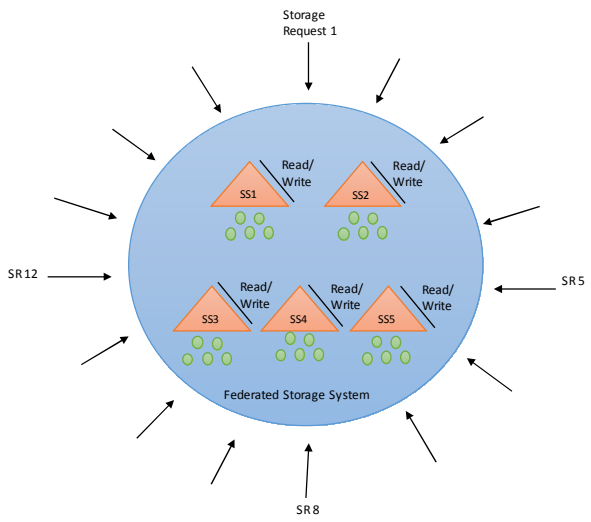


Figure 7- Access into Federated Storage Systems with multiple parallel requests

Object-based storage systems can be very accommodating when retrieving small content fragments in parallel request situations. In file-based solutions, a byte offset is needed to retrieve the content fragment if it is part of a bigger file. In object-based solutions, such retrieval could be simplified, made parseable, or altogether eliminated in certain designs. With storage design, and especially in object store designs, there is just as much awareness of segment size as segment duration. If duration-varying segments were used more, this could help in storage designs and segment retrieval. Methods to do this could include changing the placement of virtual segmentation boundaries, to accommodate more frames in a segment, or to assign different partitions to the streams. A drawback in using time-varying segments is that it would increase the size of the manifest.

<u>What if we "switched" on something other than bitrate?</u>

In traditional adaptive streaming, the client player monitors the CPB buffer and sends out a message to the server to switch either the representation or the profile to a higher or lower bit rate, based on current network conditions. The player tends to behave in a greedy fashion, always asking for more, unless it risks stopping the display of the video -- but at the same time, this response is not immediate. The issue here happens when thousands of players are all acting greedily, which can cause pulsating waves of bitrate demands. This is usually mitigated through the use of policy enforcement on the server side. Policy enforcement can make adaptive streaming systems more complex and very state-aware, so simplifying the strategy and making it less invasive could help to avoid what is a burdensome policy enforcement strategy.

How could players assist in this issue? Simply by acting less greedily. This could be done by making players more informed -- not just switching representation levels solely based on bitrate. For instance, if the quality of the representation level is known to the player, it could decide not to ask for the highest-level representation, especially if the quality is not discernably better. With this approach, the system could be saving bits that may be needed by other players. Other dimensions to this would involve a decoding complexity factor as well. Other approaches could address the design of representation levels by defining them in terms of constant quality factors. This, in turn, would bring in video coding concepts like Variable Bit Rate /VBR (alright capped VBR), which is proven to save on bits. Using a quality level to define stream representation could still allow the player to switch based on bitrate (lower quality usually means lower bit rate, or even lower complexity).

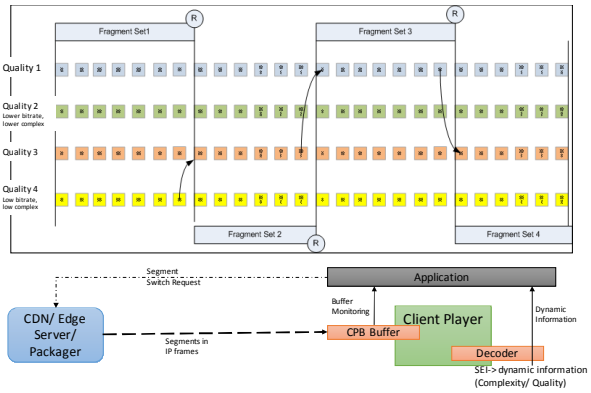


Figure 8- Switching Between Different Representation using more than Buffer Monitoring

Some of these approaches work under the assumption that information about switching choices resides only in the manifest -- but this doesn't have to be the only place. The stream itself could carry information that could be useful to the decoder. This information could be more dynamic than information carried in the manifest. If the player can access information that is available at the time of decode, the player could then make better decisions about switching to another representation. For instance, if a complexity indicator was sent as a Supplemental Enhancement Information / SEI message, the SEI message could be used by a player with limited battery power to switch to a lower quality representation, which would, in turn, extend the device's battery life. Another example could embed a dynamic quality indicator in the SEI that could be used by the player to rachet up or down a representation level, and manage its bit rate demands. Yet another example could indicate the brightness characteristics (nits) of a scene, which could again be used by the player to alter the display, so as to save on battery power. (Actually, the application of SEI messaging is a popular concept for both HDR display adaptiveness and "green" metadata for battery conservation.) These type of approaches can work well in reducing the "greediness" of players, while not demanding an intrusive policy enforcement strategy.

CONCLUSIONS

With next generation systems, there is an opportunity to re-examine stream design in light of the different demands being placed on the network for IP delivery. Specifically, we need to consider not just bandwidth and storage demands, but now display device battery lifetimes, and related delivery across a hybrid of networks. Optimization, and especially optimization of content handling, is key in reducing some of these demands, especially as the network scales up.

This paper provides some different perspectives in terms of processing, storage, and handling of content at a video stream level -- especially in a system with ATS streams, packagers, and adaptive streaming technologies. Other efforts exist to reduce the complexity of storage, including the standardization of media segment formats though MPEG's CMAF (Common Media Application Format) efforts, but these improvements can be orthogonal without issue. In addition, we need to consider traffic optimization and latency, which can be worked through strategies relating to CDN architectures, multicasting, edge servers, and data caching. Ultimately, such an approach could yield a transition plan and delivery system for on-demand content over IP that is both robust and stable enough to sustain across many types of networks -- from cable to WiFi to mobile, mesh, and others, as they materialize.

REFERENCES

[1] ISO/IEC 13818-1:2013, Information Technology - Generic coding of moving pictures and associated audio information - Part 1: Systems.

[2] ANSI/DVS 1196, Adaptive Transport Stream, www.scte.org

[3] ANSI/SCTE 214-1, MPEG DASH for IP-Based Cable Services, Part 1: MPD Constraints and Extensions, www.scte.org

[4] ANSI/SCTE 214-2, MPEG DASH for IP-Based Cable Services, Part 2: DASH/TS Profile, www.scte.org

[5] ANSI/SCTE 214-3, MPEG DASH for IP-Based Cable Services, Part 3: DASH/FF Profile, www.scte.org

[6] ITU-T Recommendation H.264 | ISO/IEC 14496-10:2012, Information Technology - Coding of audio visual objects - Part 10: Advanced Video Coding.

[7] ITU-T Rec. H.265 | ISO/IEC 23008-2:2014 – MPEG-H Part 2: High Efficiency Video Coding

[8] Y.F. Syed and A. Giladi, Hybrid Content Generation Workflows: from MPEG-2 TS to Adaptive Streaming and ISO-BMFF, proc of INTX 2015

[9] ISO/IEC 23009-1:2014 Information technology -- Dynamic adaptive streaming over HTTP (DASH) -- Part 1: Media presentation description and segment formats.

[10] DASH-IF Interoperability Points, Version 3.0, http://dashif.org/wp-content/uploads/2015/04/DASH-IF-IOP-v3.0.pdf

[11] F. Fernandes, X. Ducloux, and S. Cheng, "Proposed White Paper on Green Metadata" M37090 MPEG 113 Meeting, Geneva Switzerland, Oct. 2015.

[12] ISO/IEC 23001-7: 2015, Information technology- MPEG systems technologies—Part 7: Common encryption in ISO base media file format.

[13] ISO/IEC 23001-9: 2014 Information technology- MPEG systems technologies—Part 9: Common Encruption of MPEG-2 transport streams.

[14] "Proposed Requirements for Common Media Format for Segmented Media" M37756 MPEG 114 Meeting, San Diego USA. February 2016.

Remote PHY Deployment Options

Jeff Finkelstein, Cox Communications
Alon Bernstein, Cisco Systems

Abstract

All Internet Service Providers are challenged by the ever increasing demand of our customers for higher speeds and more consumable bandwidth. In order to meet that need we must come up with new ways to optimize our spend to keep capex and opex under control, while we also try to find new opportunities to increase our revenue with exciting new services.

This paper describes how to use a new technology known as Remote PHY Device (RPD) to develop a well thought out strategy to increase capacity by using node splits, fiber deep deployments, DOCSIS 3.1, IP video and mid-split. Our use of the term RPD is not limited to a PHY-only device, it would also be applicable to a remote MAC-PHY or remote CCAP device, we use it generically to mean placing the PHY into the outside plant and do not differentiate where the L2 and L3 functions would exist.

1.0 INTRODUCTION

"The report of my death was an exaggeration." – Mark Twain, 2 June 1887

"Mine too…" – DOCSIS, 9 March 2016

The demise of DOCSIS has been predicted since it first appeared in the mid 1990's. Yet, to this day it continues to be deployed and to drive new technologies into its ecosystem enabling a bright future for high-speed internet over cable.

It has been said that cable, in particular hybrid fiber coax based networks can not compete with other technologies. As an industry we have proven that false by intelligent use of node splits, adding new DOCSIS carriers, channel bonding, higher order modulations, new parity coding, OFDM, and on…

However, even with the innovations brought forth by a number of creative individuals and companies, we recognize that we need to continue developing new technologies and uses of existing ones to provide the long useful life that many of us believe that DOCSIS has left.

One specific technology we believe will have a significant impact on the future potential of DOCSIS is distributed access architectures (DAA). DAA places pieces of the CCAP architecture into the HFC plant closer to the end users. In our example we use the generic term RPD to mean remote PHY, remote MAC-PHY and remote CCAP. The end result is the same in that we get fiber and PHY layer technologies closer to the customer.

2.0 BACKGROUND

To start, we need some common terminology which will be used to describe the service provider (SP) network. For the purpose of this discussion, we will break the SP network into five layers.

1. Backbone
2. Metro
3. Edge
4. Outside Plant
5. CustomerPremise

For the purpose of this document we will be focusing on the edge, outside plant and customer premise. We can draw the network that we historically have deployed as depicted in Figure 1.

In Figure 1 we show the traditional placement of fiber nodes and a migration path from node plus 5 actives, to node plus 1 active, to a node plus no active layout (passive coax).

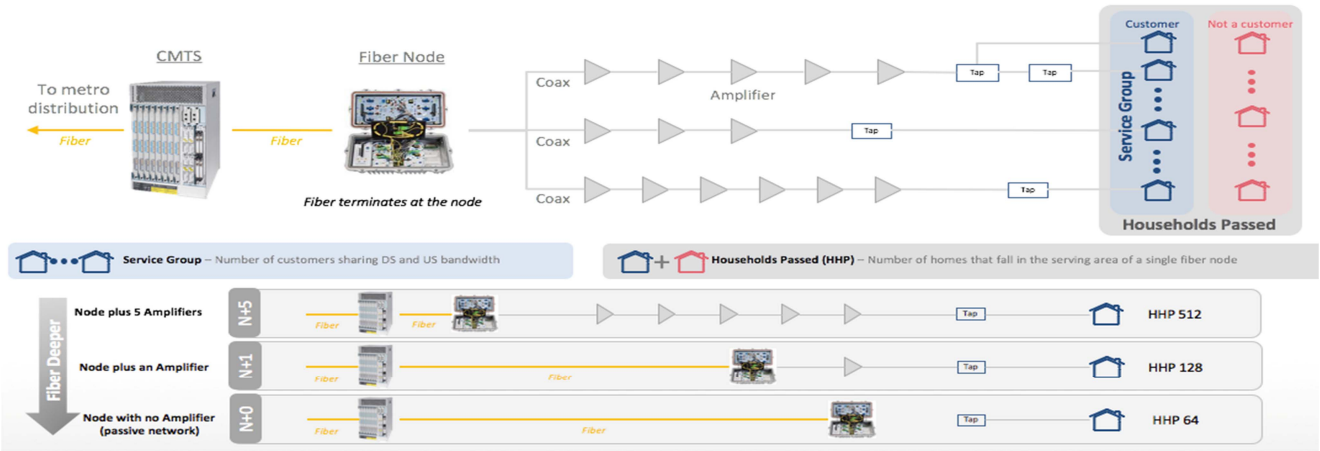


Figure 1 - CCAP Access Network

If we show the same drawing, but this time replacing the CMTS with a CCAP and the fiber node with a remote PHY device (RPD), the drawing would look as shown in Figure 2.

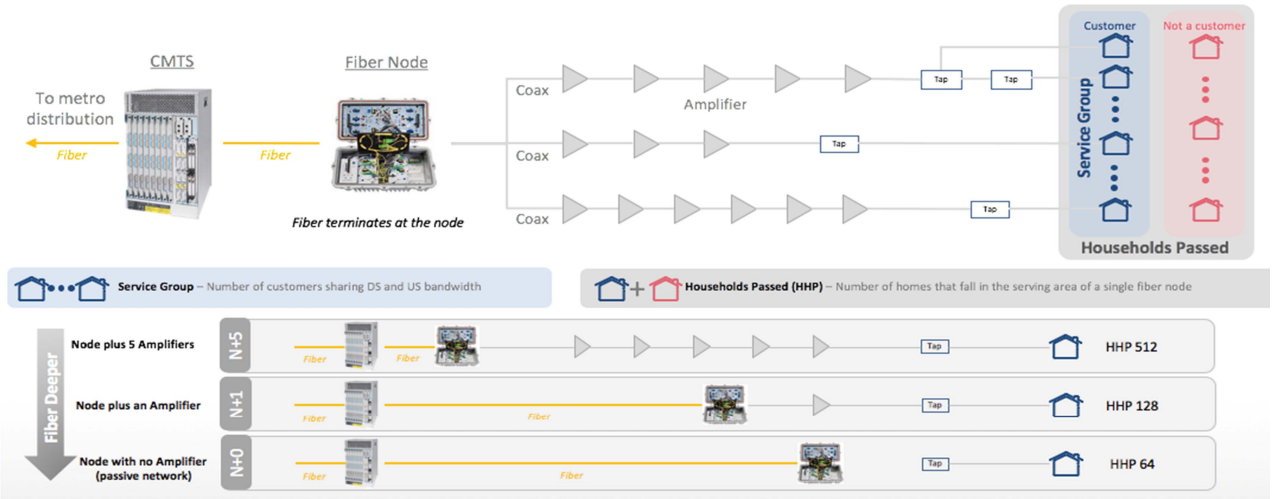


Figure 2 - Remote PHY Access Network

The biggest change in this scenario is that the CCAP core no longer contains a PHY chip (or at least does not use, even if placed in its integrated PHY chips). The MAC uses the Downstream External PHY Interface (DEPI) and Upstream External PHY Interface (UEPI), to communication with the RPD over an ethernet link.

MHAv1 was the separation of downstream only, with keeping the upstream on the CMTS. This became known as the "Modular CMTS" architecture and facilitated use of edge QAM devices for the physical layer.

3.0 REMOTE PHY ARCHITECTURE

The Remote PHY architecture is described in the "DCA-MHAv2" CableLabs specification (ref [1]). DCA stands for "Distributed CCAP Architecture" and MHAv2 is "Modular Head-end Architecture Version 2".

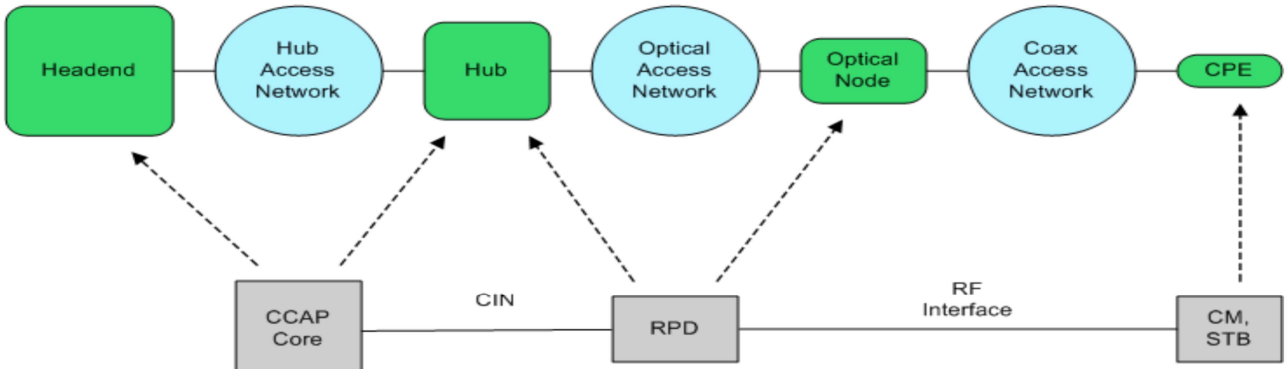


Figure 3 - Remote PHY Architecture

At a basic level, the Remote PHY Device solves a simple problem: the number of coax connectors that can be attached to a linecard in a CCAP box is limited. As a result the physical connectivity of a CCAP might be lower then the actual traffic processing capacity.

For example, consider a case where there is room to attach only 64 coax connectors, each carrying 5 Gbps of data with a forwarding engine capable of forwarding 500 Gbps. In that case the total amount of traffic that can be connected is 5x64 = 320 Gbps, which means all the forwarding capacity of the box is not utilized. With a Remote PHY this basic problem is resolved because we can now use regular Ethernet switches to

aggregate a large number of Remote PHY devices.

However, the ability of the RPD to serve as a "port extender" for CCAP is only the beginning. With the remote core we can both:

1. Place the RPD at a great distance from the core, only limited by the contraints of the timing method used to synchronize the downstream and upstream. This can work over 100 miles or more.
2. Allow a flexible allocation of RPD to core. In essence, we have a pool of Remote PHY Devices and a pool of core resources that can be matched in many configurations. This allows for the creation of "virtual serving groups" that share a common MAC
3. Attached to different cores, e.g. CMTS core, broadcast EQAM core, VOD/SDV core and OOB .

In the following section we will see how these capabilities allow for a flexible and scalable deployment model.

4.0 REMOTE PHY DEPLOYMENT OPTIONS

The following section outlines the remote PHY deployment models, but it also presents them as a transition strategy. The CCAP architecture can be migrated from the current "hub in a box" monolithic model of a fully converged access solution into a virtual cloud application in several steps, all enabled by the Remote PHY technology:

1. The first phase is to deploy the RPD in a "port extender" mode. There are only two components required. The physical core and a set of Remote PHY Devices.

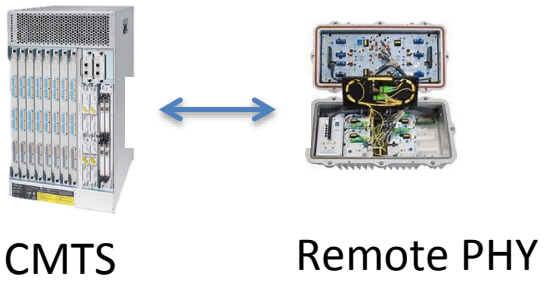


Figure 4 - Physical CMTS With RPD As A Port Extender

2. The second phase would be to attach the RPD to a remote CCAP core, still in a port extender mode, but over a greater distance and over multiple network hops.

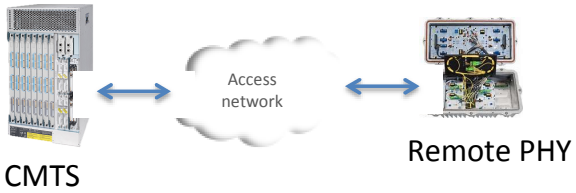


Figure 5 - Physical CMTS with RPD as a port extender across a network

3. For the next phase an operator can use a 3rd tool : a orchestrator and a set of controllers that will mix and match remote phy and cores as needed without fixing a particular core to a particular remote phy. The role of the orchestrator is to implement a "workflow" of instructions needed to select the RPD to core association as well as create and validate the connection between them. Under the orchestrator there can be a couple of controllers, each specializing in a particular domain, e.g. access, network etc.

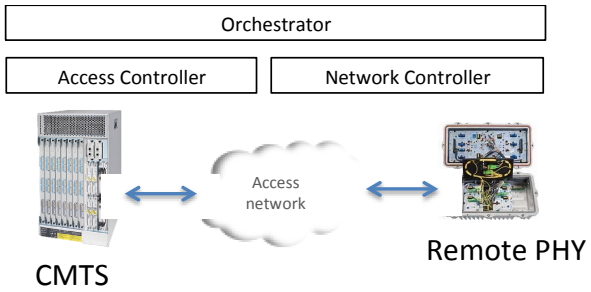


Figure 6 - Orchestrated Virtual CMTS

4. The controller can assign some of the remote phy devices,, to a virtual instance. Initially the virtual instances can be standalone – in that mode they are a CMTS implemented over a generic server platform, but not yet part of a "real cloud"

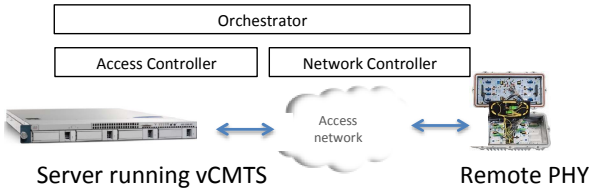


Figure 7 - vCMTS In An Appliance Mode

5. A fully orchestrated virtualized solution where virtual instances are created on demand and linked to other Network Function Virtualization (NFV) service chains. Note that one controller that's added in this configuration is the "data center controller" as depicted in the figure below:

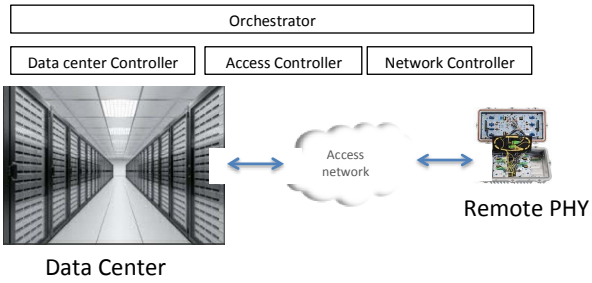


Figure 8 - Remote PHY Data Center deployment

An operator can create systems that have a bit of any of the architectures mentioned about as "transition phases", for example, its possible to have the active instances of a vCMTS deployed as an appliance and only the redundancy instances deployed in a data center. The flexibility is because the remote PHY has to point a tunnel is a core, without caring about how far the core is or whether it's physical or virtual.

5.0 SERVICE GROUP SIZING

Because the remote PHY processes information at layer 1 only its completly imprevious to the service group size or the number of subscribers in the service group. It is simply a "Coax-in-fiber-out" device. Therefor when planning for serving group sizing the only planning criteria are those related to cost, bandwidth and RF span.

6.0 THE UNCONVERGED CCAP

A lot of the "Convergence" in CCAP (Converged Cable Access Platform) was achieved by frequency stacking on the physical port. A single port that can support a full spectrum motivated a system design that generated the full spectrum, including MPEG transport and data, in a single box. However, in the remote PHY architecture the frequency stacking occurs at the remote PHY itself, and that enables a "de-convergence" of the MAC where separate appliances or virtual machines drive different parts of the spectrum, each being able to scale, test and upgrade independently.

7.0 CONCLUSION

There is no simple answer of how to best deploy distributed access architecture. As service providers we desire to deploy new technologies in ways that are least impactful to our existing customer base. This many times requires that we slow down our deployments and plan our strategy carefully to minimize outages.

Remote PHY Devices, are particularly well suited for smaller service group sizes, that in the future will reduce our downtime as many plant related conditions today require truck rolls and hands-on tuning of lasers and amplifiers. In the RPD world, lasers are digital which will lead to lower maintenance costs over analog lasers and simplified configuration. The pairing of RPD with Proactive Network Maintenance means that output power levels may be fine-tuned remotely, which means a better user experience from those customers closer to the RPD to those at the end of the coaxial network.

Finally, with RPD comes the ability to position ourselves for a lower cost passive coaxial plant, which leads us to the possibility of a future version of DOCSIS that may even be full duplex.

These benefits are not limited to a pure remote PHY, but are equally a consideration for other distributed access architectures including remote MAC-PHY and remote PON. In all scenarios the one undeniable fact is we end up with fiber closer to the customer, which is something that prepares us for future needs.

SDN Ground Truth:
Implementing a Massive Scale Programmable DOCSIS Network

Sameer Patel
Manager
Network Policy Engineering – T&P
Comcast

Jason Schnitzer
Founder & CEO
Applied Broadband

Mohammad Kabir Chowdhury
Product Engineer
Comcast

Dr. David Early
Data Scientist
Applied Broadband

Abstract

SDN carries the promise of enhanced and expanded services with a more manageable operational environment. Though many standards groups and open source projects exist, deployments remain in their infancy.

This paper provides an overview of Comcast's SDN-driven implementation of programmable Service Flows through the integration of the OpenDaylight PacketCable Multimedia plugin. This paper also provides practical considerations for successful analysis of SDN architectures and technologies when applied to DOCSIS Broadband network deployments.

MOTIVATION

Making The Case for SDN and DOCSIS

The application of a software-defined network (SDN) architecture to DOCSIS [1] broadband access allows resources to be virtualized and allocated to whatever application or service requires them.

Though advantages to SDN-based architectures and applications have been proposed [2], the specific benefits when applied to a programmable DOCSIS policy environment are as follows:

1. **System and network resource elasticity**. Traditionally, deployments of network provisioning and management software for DOCSIS networks has been burdened by rigid environmental constraints and restrictive resource dimensioning limitations. The virtualization of DOCSIS low-level services using contemporary SDN architectures enables greater design flexibility in networks of scale, while affording discrete capacity upgrades at time of need.

2. **A common platform across network applications.** Capital, operational, and organizational economies of scale result when using multiple applications that access the same network services and related resources in a common way.

3. **Service delivery agility**. As a consequence of both (1) and (2) above, new subscriber services that benefit from

QoS treatment in the access network can quickly be supported.

Business agility is essential to the very survival of broadband organizations, therefore the ability to implement new services quickly is vital.

Traditional DOCSIS network architectures and services lack the flexibility required to support an agile business, and over-provisioning the network for peak utilization inflates costs. A forward-looking SDN approach is required for any organization seeking to become agile, because this approach brings dynamic features to the network and gets maximum benefit from network virtualization functions for a minimum overall cost. Virtualization has largely transformed the data center with flexible and automated server provisioning, but networking and storage infrastructure have not kept pace.

Current Use Cases

Within the cable access network, SDN has the potential to reduce service deployment and maintenance times, reduce errors associated with those setups, offer new management options, and facilitate the introduction of more fine-grained service offerings. While SDN can apply to a wide array of applications in the cable access network [3], our implementation addresses three specific use cases:

- **Telephony** – The application of QoS to enhance managed VoIP traffic.
- **Congestion management** – Dynamic, rule-based implementation policy controls to assure fair access to resources during times of network congestion.
- **Acceptable Use Policy Management (AUPM)** – The application of network-based controls to address service abuse.

APPLYING SDN PRINCIPLES TO BROADBAND NETWORKS

SDN (at least in today's leading initiatives) finds its genesis in switch networks largely aimed at enterprise network environments. OpenFlow [4] is the leading standard, driving many SDN implementations today. OpenFlow is largely concerned with the programming of packet paths (flows) across a network of IP switches.

When comparing OpenFlow to Cable access network technologies such as DOCSIS, fundamental differences exist:

- **Topology:** The underpinning HFC network is a physical tree/branch topology. The CMTS (master) and CM (slave) topology, and resulting protocol design, are fundamentally different than port-to-port topology of switched networks such as Ethernet.
- **Scale**: The number of managed devices (entities) in a single managed DOCSIS network is well into the tens of millions of and expected to grow further.
- **Complexity**: The information model representing a complete DOCSIS network describes objects and attributes numbering in the thousands.

The following section provides a brief overview of basic SDN principles. Readers familiar with these concepts are encouraged to skip to the next section.

The SDN architecture abstraction presented by the ITU-T in [5] describes three functional layers and two interfaces between the layers (illustrated in Figure 1). The layers are:

1. **Application Layer:** SDN and/or business applications that specify network behavior.

2. **SDN Control Layer:** "a means to dynamically and deterministically control the behavior of network resources…as instructed by the application layer" [5].

3. **Resource Layer:** Network devices and element management systems.

The interfaces – the Application-Control Interface and the Resource-Control Interface – represent bi-directional communication between the components of each logical layer.

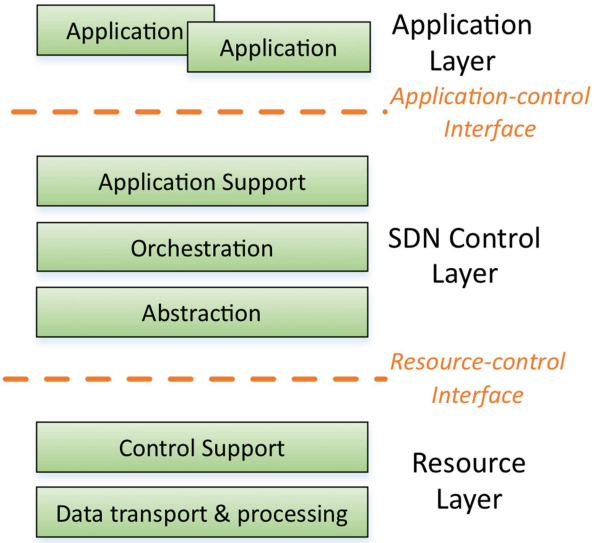


Figure 1: SDN High Level Abstraction (ITU-T Y.3300 [5])

For the purposes of our implementation, the architecture is consistent with the general description above with the following additional details:

1. The Application Layer includes a total of three (3) software applications, each representing a current use case.

2. Each of the three Applications communicate with the SDN Control Layer by way of an API based on a common data model abstracting the Resource Layer (network).

3. The SDN Control Layer forwards Application requests with minimal modifications to the incoming messages. The Control Layer is also responsible for the storage and management of both the desired Resource Layer configuration and the actual operational state.

4. The Resource Layer is implemented by the DOCSIS network infrastructure which features control support as well as data transport and processing.

ARCHITECTURE

This new programmable architecture proposes enhancements to current DOCSIS access network Policy Management infrastructure in order to support both legacy and emerging use cases by leveraging Software Defined Networking (SDN) architectural principles.

The system architecture described in this section presents an overall vision for how the OpenDaylight (ODL) [6] SDN platform is used to implement specific PCMM functions within the DOCSIS policy infrastructure. The proposed architecture is aligned with next-generation access systems and programmable network goals. The architecture leverages new development and operations models to create a more durable, open, and extensible policy infrastructure for DOCSIS deployments.

OpenDaylight

The OpenDaylight (ODL) [6] initiative is a collaborative open source project that aims to accelerate adoption of SDN and create a solid foundation for Network Functions Virtualization (NFV) enabling a more transparent approach that fosters new innovation and reduces risk. Founded by industry leaders and open to all, the

OpenDaylight community is developing a common, open SDN framework consisting of code and blueprints.

Porting PCMM to ODL

Though a rigorous review of the CableLabs PacketCable Multimedia (PCMM) [7] protocol will not be provided in this paper, it is useful to understand a few key concepts in the context of integrating DOCSIS and PCMM capabilities into an SDN platform architecture. Originally released in 2003, the intent of the PCMM specification was to support the deployment of general Multimedia services by providing a technical definition of several IP-based signaling interfaces that leverage core QoS and policy management capabilities native to DOCSIS Versions 1.1 and greater.

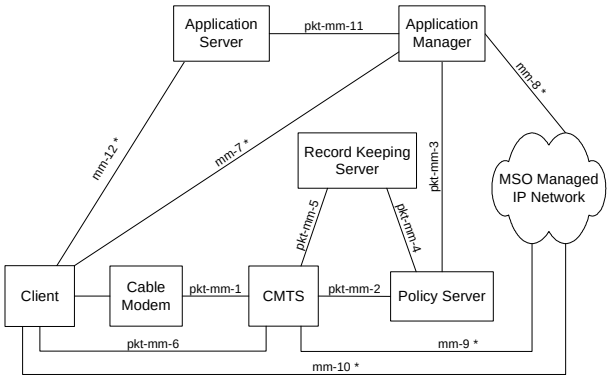


Figure 2: Complete PCMM Architectural Framework

Figure 2 illustrates the complete PCMM architectural framework, defining a collection of elements and interfaces which together support the service objectives of Dynamic QoS (DQoS)-enabled capabilities in DOCSIS access networks.

A limited subset of the overall capability described in the most current PCMM release [7] is needed to support the use cases and requirements necessary for SDN implementation. For the purposes of our application, only three (3) of the seven (7) elements formally described in the PCMM architectural framework are addressed:

1. Cable Modem (CM),
2. Cable Modem Termination System (CMTS), and
3. Policy Server (PS).

It follows that a small subset of the interfaces (two of the twelve total) are relevant to future SDN applications:

1. **pkt-mm-1:** Using CMTS to CM DOCSIS messaging, the CMTS instructs the CM to setup, teardown or change a DOCSIS service flow in order to satisfy a QoS request via DSx signaling.

2. **pkt-mm-2:** This interface (PS to CMTS) controls policy decisions, which may be: (a) pushed by the Policy Server (PS) onto the CMTS, or (b) pulled from the PS by the CMTS. In some scenarios, this interface may also be used to inform the PS when QoS resources have become inactive. Common Open Policy Service (COPS) [8][9][10] is used to transfer policy information between the PS and the CMTS. The PS initiates communication for a Multimedia session by sending a DQoS Gate-Set message (which is an unsolicited COPS message) to the CMTS.

It is worth noting that, along with the unnecessary interfaces and elements, other historically relevant concepts used to describe PCMM capabilities were not carried forward into the SDN-based architecture. These include more rigid COPS/PCMM-centric concepts such as Service Control Domain (SCD), Resource Control Domain (RCD), Policy Decision Point (PDP) and Policy Enforcement Point (PEP). In some cases, these protocol semantics were preserved but

are now expressed using different structures and terminologies within a new SDN context.

OpenDaylight PCMM Plugin

Figure 3 represents a high-level view of selected PCMM capabilities within the context of the ODL's SDN Controller architecture. In it, the PCMM capabilities to be integrated into the ODL Controller take the form of a PCMM device communication plugin as a SouthBound (SB) device communication plugin – that is, a low level protocol implementation above the network resource layer that implements the desired PCMM PS capabilities and select protocol features of the PCMM pkt-mm-2 interface. PCMM-driven applications (representing the use cases described) interact with the PCMM-plugin using the northbound (NB) abstraction layer/API as described by the PCMM YANG data model (packetcable-policy-model).

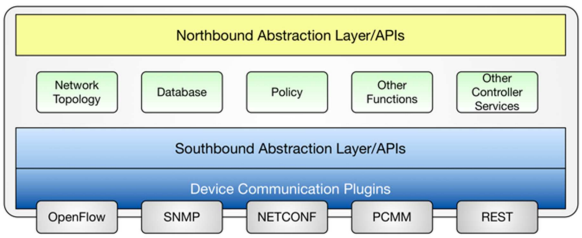


Figure 3: PCMM plug-in integrated into the ODL SDN Controller

The PCMM plugin for ODL sponsored by CableLabs was originally conceived to use the ODL platform as a proof-of-concept for the Application Manager (AM) and parts of the Policy Server, both components of the PCMM architecture. In 2014, the source code representing this reference implementation capability moved from CableLabs and became a formal PacketCable PCMM project under ODL stewardship [11].

Module	**Description**	**Associated Feature(s) or purpose**
features-packetcable-policy	Defines the bundles that are part of the Policy Servicer.	The Feature module
packetcable-driver	Contains the logic required for communicating or creating a PCMM CCAP/CMTS.	features-packetcable & features-packetcable-policy
packetcable-policy-model	Contains the YANG data model.	features-packetcable-policy
packetcable-policy-server	Responsible for processing RESTCONF calls and maintaining persistent socket connection to the CCAP/CMTS.	features-packetcable-policy
packetcable-policy-karaf	Creates the OSGi distribution for feature including ODL and other dependencies.	features-packetcable-policy
packetcable-emulator	Standard Java library that acts as an emulator for a PCMM CCAP/CMTS.	test tool

Figure 4: OpenDaylight PacketCable Multimedia (PCMM) Plugin Components

Figure 4 provides an inventory of the ODL PCMM plugin components, which together provide the following functions [12]:

- RESTCONF APIs for provisioning CMTS network elements.
- RESTCONF APIs for dynamic creation of service flows for registered CMs.
- RESTCONF APIs for dynamically controlling service flow QoS parameters.

- A complete PCMM/COP-PR messaging library and protocol stack implementation

IMPLEMENTATION

In this section we will provide an example of a complete PCMM QoS session created and terminated using the ODL system described. Figure 5 illustrates the example configuration which includes a DOCSIS CMTS, CM, ODL Controller, and a PCMM ODL Application.

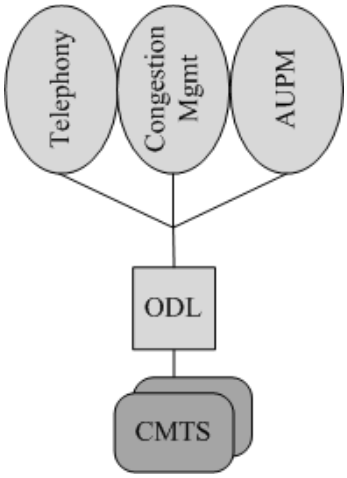


Figure 5: Connecting PCMM Applications to ODL Controller

In this example, the ODL Application is a simple HTTP client capable of performing CRUD (Create, Read, Update, Delete) operations on the ODL's internal MD-SAL datastore using data defined by the PCMM YANG module. The ODL Controller is based on the Beryllium release and includes the PCMM plugin [11][12].

The PCMM plugin extends the capabilities supported in Beryllium to implement additional PCMM protocol features not available in the open source release. Specifically, the PCMM plugin fully supports the expression of traffic profiles via a FlowSpec. A FlowSpec object defines the traffic profile associated with a Gate through an RSVP-like parameterization scheme [12].

PCMM Gate-Set

Referring to Figure 6, an Application generates a RESTCONF PUT operation to the ODL Controller's RESTCONF API exposing the YANG data model representation of the PCMM Gate-Set request (Figure 7). The content of the PUT is a JSON formatted list of parameters that together describe the desired Gate configuration (Figure 8).

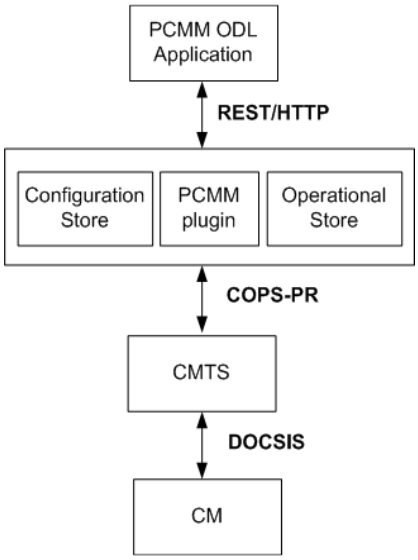


Figure 6: Example ODL PCMM System Configuration

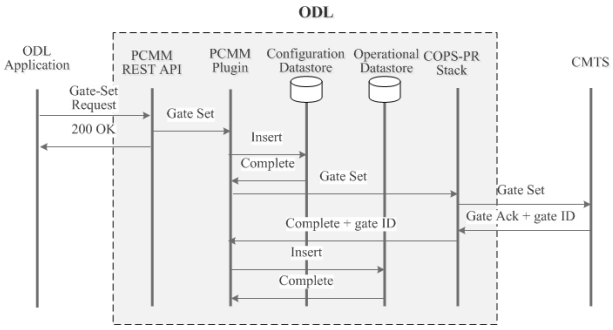


Figure 7: PCMM Gate-Set Transaction Sequence Diagram

```
http:/restconf/config/packetcable:qos/apps/app/<appId>/subscribers/subscriber/<subscriberId>/gates/gate/<gateId>/
```

Figure 8: Format of RESTCONF Gate-Set request in a HTTP PUT operation

The HTTP PUT format for an Upstream (US) FlowSpec Gate-Set request from the southbound PCMM Plugin (within ODL) to the CMTS is as follows:

- <appId> - A string used to uniquely identify the requesting application.
- <subscriberId> - The subscriber identity, usually a CM IP address.
- <gateId> - ODL Gate identifier. An internal-only reference to a unique node containing configuration parameters within the ODL datastore. The ODL gateId value is determined by the ODL Application generating the request. Note that the ODL gateId and PCMM GateId (described later in this section) are two logically isolated concepts.

Figure 9 shows example JSON content for the CM Gate-Set request. As described earlier, our use cases required a FlowSpec traffic profile, which necessitated an extension of the Beryllium release of the PCMM plugin, which only supports a Service Class Name (SCN) defined Gate profile.

The PCMM plugin NB REST API receives the Gate-Set request message from the ODL Application, and performs the following steps:

1. Basic verification of the format of the request content is made against the PCMM YANG data model. If one or more mandatory elements are missing or malformed, the request is rejected and an error is returned to the calling ODL Application.
2. Upon acceptance of the Gate-Set request message by the NB REST API, an acknowledgement message (“200 OK”) is returned to the calling ODL Application. It is conceptually significant to recognize that the receipt of this acknowledgment only indicates that the REST API successfully received the message. It is not an acknowledgement that a Gate was successfully created on the CMTS.

```
{
  "gate": {
    "gateId": "<gateId>",
    "classifiers": {
      "classifier-container": [{
            "classifier-id": "1",
            "classifier": {
              "srcIp": "10.20.0.3",
              "dstIp": "10.20.0.5",
              "protocol": "0",
              "srcPort": "54322",
              "dstPort": "4322",
              "tos-byte": "0x01",
              "tos-mask": "0x00"}
          }]
    },
    "gate-spec": {
      "dscp-tos-overwrite": "0x00",
      "dscp-tos-mask": "0x00",
      "direction": "us"
    },
    "traffic-profile": {
      "flow-spec-profile": {
        "token-bucket-rate": "400",
        "token-bucket-size": "40",
        "peak-data-rate": "400000",
        "minimum-policed-unit": "400",
        "maximum-packet-size": "400",
        "rate": "30000",
        "slack-term": "0"
      }
    }
  }
}
```

Figure 9: Example JSON representation of a PCMM Gate-Set Request message

ODL Internal Message Processing

There are several steps that a request goes through within ODL, including creation and management in the internal configuration and operational datastores as defined in NETCONF [13]:

1. The NB REST API unmarshalls (deserializes) the inbound Gate-Set request.

2. ODL then semantically evaluates the contents of each parameter in the JSON content of the Gate-Set request. (Figure 9) against the constraints of the PCMM YANG model.

3. Upon successful validation, the MD-SAL creates a node within the configuration datastore and populates it with the desired Gate-Set parameters referenced by a unique ODL GateId.

4. The SB PCMM API (containing the PCMM messaging library), acting as a COPS PS or Policy Decision Point (PDP), then forms a PCMM Gate-Set message for transmission to the CMTS.

5. The CMTS receives the Gate-Set message from the PCMM Plugin, and processes accordingly.

6. The CMTS either accepts and establishes the Gate or rejects it. Upon acceptance, an acknowledgement containing a unique PCMM Gate ID is generated by the CMTS.

7. If a PCMM Gate ID is returned by the CMTS, the established Gate information, including the Gate ID, is stored in ODL's MD-SAL Operational Datastore. If no Gate ID is returned, then relevant error information is stored in the Operational Datastore.

CMTS Message Processing

As described in PCMM, the CMTS assumes the role of Policy Enforcement Point (PEP) and is responsible for processing the Gate-Set request from the ODL's SB PCMM plugin (pkt-mm-02). The CMTS executes the relevant procedure as described in [PCMM]. Once a DSA-RSP is received by the CMTS from the CM confirming a successful transaction, the CMTS will send a Gate-Set-Ack to the Policy Server. The CMTS issues a Gate-Set-Ack to the ODL PCMM plugin (PS) in response to the Gate-Set message.

If successful, the CMTS and CM collectively execute a DOCSIS DQoS Dynamic Service Add (DSA) operation to admit the Upstream Service Flow. If CMTS admission control succeeds, the CMTS will initiate the process of reserving and committing the access network resources by issuing a DSA to the Cable Modem. At this point, the requested US PCMM Gate is now installed and the traffic for the PCMM session is forwarded from the CM to the CMTS according to the DOCSIS US QoS parameters provided.

PCMM Gate-Delete

In this implementation, the ODL Application (representing one of the use cases described) is responsible for signaling in order to explicitly terminate the PCMM session.

Again assuming the role of a PCMM PS, the ODL Application terminates the session by sending a Gate-Delete message to the CMTS.

Referring to Figure 10 and Figure 11, The ODL Application generates a RESTCONF DELETE operation to the ODL Controller's RESTCONF API .

If successful, the CMTS will tear down the access network resources by sending a DSD-

REQ to the CM. The CMTS will complete the Gate-Control transaction with a Gate-Delete-Ack.

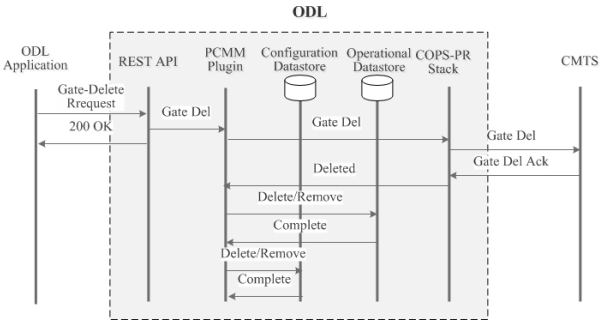


Figure 10: PCMM Gate-Delete Transaction Sequence Diagram

```
http:/restconf/config/packetcable:qos/apps/app/<appId>/subscribers/subscriber/<subscriberId>/gates/gate/<gateId>/
```

Figure 11: Format of restconf Gate-Delete request in HTTP DELETE operation

STATE OF THE ART & KNOWN LIMITATIONS

Based on the scope of the initial CableLabs PCMM plugin development, the implementation has the following current known limitations:

- Though the PCMM plugin messaging library includes full support for COPS-PR, not all PCMM Traffic Profiles and DOCSIS MAC layer scheduling types are supported in the plugin itself.

- Event messaging as described in PCMM [7] is not supported.

- Testing has been limited to small scale proof-of-concept testing. Large scale and broad functional testing is required.

- Security has not been formally addressed.

- The ODL REST API does not return any explicit (ACK/NACK) operational information to the requesting ODL Application. The only message is a “200 OK” message indicating correct syntax and acceptance of the request. Any additional information about the Gate, including whether it was successfully set, must be explicitly requested by the calling application in a separate transaction.

FUTURE WORK

Beyond addressing known limitations described in the prior section, the following areas must be addressed in order to deploy the PCMM ODL platform within a network of massive scale:

- Scalability: The role and persistence of state within the ODL Controller vs. the ODL Application layer.
- Survivability: ODL persistence and HA/DR implications.
- Manageability: Visibility and control of the system from an operations perspective in terms of fault and performance management.

Emerging Use Cases

It is anticipated that new applications providing QoS-driven enhancements to services delivered within the DOCSIS infrastructure, such as video delivery, will evolve. Each new service will be added to a common SDN infrastructure, represented by a new PCMM-based ODL Application.

Emerging Networks

DOCSIS 3.1 [13] describes a new framework for Hierarchical QoS (HQoS), which enables operators to define QoS policies on an aggregation of Service Flows. If adopted by CMTS vendors, this new mechanism for the organization of access network resources will enable new service configuration and bandwidth management scenarios not available today.

Emerging SDN Platforms

Though considerable progress has been made to realize SDN-based principles in deployment, implementations such as ODL remain in their infancy. As the realities of deploying SDN architectures into massive scale networks are learned, it is expected that new general SDN controller implementations will emerge that may address the known limitations of ODL.

CONCLUSION

In DOCSIS broadband networks, the practical application of SDN differs from those defined for other networks. The nature of the DOCSIS protocol, deployment topology, and management capabilities form the basis for unique requirements and modifications to existing SDN approaches.

To date, the integration of PCMM and DOCSIS capabilities into an SDN platform has demonstrated encouraging progress. The implementation described in this paper

highlights the functional goals of ODL acting as a control plane for the creation of DOCSIS/PCMM service flow management.

By leveraging an emerging SDN platform (ODL) and taking a modular crawl-walk-run approach to development and deployment, a practical, scalable implementation is becoming reality. While further development and field experience are required, there is a clear path forward for SDN in the new programmable DOCSIS network.

ACRONYMNS AND ABBREVIATIONS

ACK – ACKnowledgement
AM – Application Manager
API – Application Programming Interface
AUPM – Acceptable Use Policy Management
CRUD – Create, Read, Update, Delete
CM – Cable Modem
CMTS – Cable Modem Termination System
CCAP – Converged Cable Access Platform
COPS – Common Object Policy Service
COPS-PR – COPs
DOCSIS – Data Over Cable Interface Specification
DQoS – Dynamic Quality of Service
DSA – Dynamic Service Add
DSA-RSP – Dynamic Service Add Response
DSD – Dynamic Service Delete
HA – High Availability
HFC - Hybrid Fiber/Coax
HQoS – Hierarchical QoS
IETF – Internet Engineering Task Force
ITU-T – International Telecommunication Union
JSON – JavaScript Object Notation
MD-SAL – Model Driven SAL
NACK – Negative ACKnowledgement
NB – Northbound
NFV – Network Function Virtualization
ODL – OpenDaylight
PDP – Policy Decision Point
PEP – Policy Enforcement Point
PS - Policy Server
QoS – Quality of Service
RCD – Resource Control Domain
REQ – REQuest
RFC – Request for Comments
RSVP – Resource Reservation Protocol
SAL – Service Abstraction Layer
SB – Southbound
SCD – Service Control Domain
SDN – Software Defined Networking
SD – Software Defined
SF – Service Flow
VoIP – Voice Over IP

REFERENCES

[1] DOCSIS® 3.0 specifications: Cable Television Laboratories, Inc. “CableLabs Specifications” [Online] Available: http://cablelabs.com

[2] N. Feamster, J. Rexford, E. Zegura. The Road to SDN: An Intellectual History of Programmable Networks. IEEE ACM Queue, December 2013.

[3] Virtualization and Network Evolution Open Networking - SDN Architecture for Cable Access Networks Technical Report VNE-TR-SDN-ARCH-V01-150625, CableLabs.

[4] OpenFlow® [Online] Available from: http://openflow.com.

[5] Recommendation ITU-T Y.3300 (06/2014) Framework of software-defined networking, ITU, 2014.

[6] Open Daylight Project. http://www.opendaylight.org/.

[7] PacketCable Specification Multimedia Specification PKT-SP-MM-I06-110629, CableLabs, June 2011.

[8] IETF RFC 3084, COPS Usage for Policy Provisioning (COPS-PR), March 2001.

[9] IETF RFC 2748, The COPS (Common Open Policy Service) Protocol, January 2000.

[10] IETF RFC 2753, A Framework for Policy-based Admission Control, January 2000.

[11] ODL PacketCable Plugin Documentation, https://wiki.opendaylight.org/view/PacketCablePCMM:Main

[12] OpenDaylight Users Guide, https://www.opendaylight.org/sites/opendaylight/files/bk-user-guide.pdf

[13] IETF RFC6241, Network Configuration Protocol (NETCONF), June 2011.

[14] DOCSIS 3.1 MAC and Upper Layer Protocols Interface Specification, CM-SP-MULPIv3.1-I02-140320, CableLabs, 2014.

SERVICE AND MANAGEMENT ORCHESTRATION IN DISTRIBUTED HIGH SPEED DATA NETWORKS

Dan Torbet, Tom Cloonan, and Andy Aftelak
ARRIS
Steve Krapp
MaxLinear

Abstract

High speed data networks to the customer's premise are evolving rapidly, driven by the need to innovate quickly and to reduce the cost of both operation and capital spending. Distributed DOCSIS architectures, whether remote PHY or remote MAC/PHY, will create more intelligent nodes in the network that require management and create valuable data on network performance and effectiveness. Mixed distributed and centralized network Architectures will create management domains with different requirements, straining traditional management infrastructures and organizations. Management paradigms from the management of data centers, such as software defined networking (SDN) and network virtualization (NV or NFV) are being applied to telecom networks promising agility and innovation in the creation and monitoring of network services. Managing and innovating on these evolving network platforms will be key to both maximizing the return on past and future investments, enabling service velocity, and containing the operational costs of the network.

We describe an approach of using a service orchestrator that evolves with the network architecture. Using SDN principles, we show how such a cloud based orchestrator in conjunction with purpose built applications can begin to unburden the existing network elements in the collection and management of network data, and to create value added services for network configuration, network optimization, and ultimately network planning. The purpose built applications in conjunction with the Service Orchestrator can be well placed to orchestrate network management between traditional and distributed equipment domains, and play an essential part rapid configuration and control of the entire network.

INTRODUCTION

The cable industry is currently experiencing a wide range of rapid changes in its services as it moves into the future. These changes encompass not only the access networks that have been leveraging the Hybrid-Fiber Coaxial (HFC) plant, but how those networks are architected, provisioned, and managed. The HFC plant continues to carry a wide variety of services that depend on multiple transmission methods including traditional broadcast, IP HSD, video, and telephony over DOCSIS and PON, and cell backhaul over fiber based Ethernet. Increasingly, operators are migrating into a highly service oriented methodology of enabling, managing, monitoring, and provisioning services over this HFC architecture As a result, the HFC plant as a last-mile delivery system will undoubtedly change immensely in the next several years to support these changes in services.

The changes in the access network and HFC network will be driven by many advances, including:

Video Upgrades

- A transition from Standard Definition/High Definition video to Ultra-High Definition (4K, 8K) video
- A transition from MPEG-2 and H.264 encoding to HEVC encoding

- A transition from QAM-based video distribution to IP-based video distribution

High-Speed Data Upgrades

- A transition to 1+ Gbps downstream service level agreements (SLAs)
- A transition to symmetrical data services
- A transition from SC-QAM-based DOCSIS 3.0 channels to OFDM-based DOCSIS 3.1 channels
- A transition from predominantly 750 MHz downstream systems to 1.0/1.2 GHz downstream systems
- A transition from 42/55 MHz upstream systems to 85/204 MHz upstream systems
- Transitions from integrated and purpose built CMTS equipment to distributed devices where specific functions can be unbundled

Alternative Wireless Services

- The addition of cell tower back-haul support
- The augmentation of existing commercial Wi-Fi support
- Wi-Fi services such as voice
- IoT

While some of the above changes are definitely aimed at reducing the bandwidth requirements on the HFC network (ex: HEVC encoding, OFDM), most of these changes will push the network in the other direction and will likely force much higher bandwidth capacity requirements on the network equipment of the future. Each of these changes also requires that for the operator to be successful, the overall introduction and velocity of services must evolve.

Multiple System Operators (MSOs) are planning several actions and investigating many new technologies and architectures as they prepare to provide for this increased bandwidth capacity. As an example, High-Speed Data traffic engineering studies are currently underway to predict the future bandwidth capacity requirements as MSOs look towards the 2020 decade and beyond. These studies indicate that by the year 2020, the average downstream bandwidth per subscriber (in the busy-hour) may grow from its current level of 1 Mbps to ~5 Mbps, and the average upstream bandwidth per subscriber (in the busy-hour) may grow from its current level of 100 kbps to ~330 kbps. Maximum downstream SLA levels that are (in many cases) at 200 Mbps today will grow to be greater than 1 Gbps by 2020, and maximum upstream SLA levels that are at ~20 Mbps today may grow to be 75 Mbps by 2020. Some MSOs are even beginning to explore the likelihood of providing symmetrical 1 Gbps services by the 2020 time-frame.

All of these bandwidth trends will place a heavy strain on the existing HFC infrastructure and will lead to capitalizing on the higher spectral efficiencies and larger spectral widths supported by newly-arriving DOCSIS 3.1 equipment. While DOCSIS 3.1 improvements will undoubtedly help MSOs with increased bandwidth capacities, other techniques for supporting the bandwidth requirements will also be required. In particular, MSOs are looking at many new architectural approaches in an effort to provide even more bandwidth improvements.

One approach is to continue to perform node-splits to reduce the number of subscribers that share the bandwidth within a Service Group. As Service Group sizes decrease, the number of Service Groups will obviously need to increase. Future I-CCAP systems are being developed with much higher Service Group densities than exist today. However, another approach that permits MSOs to increase the density of Service Group support in the headend is to move toward Distributed CCAP Architectures.

DISTRIBUTED CCAP ARCHITECTURES

Distributed CCAP Architectures (DCAs) are

a new class of architecture for providing broadband digital services (such as high-speed data, video, and voice) over HFC plants. These DCA architectures move some (or most) of the CCAP functionality (including PHY and optionally MAC processing sub-systems) into the fiber nodes that exist within the outside plant. They therefore require the use of Ethernet-based or Passive Optical Network (PON)-based digital optics (instead of analog optics) for transmissions across the fiber within the HFC plant. DCAs offer many potential benefits to MSOs, including:

a) Reducing the required space and power requirements in the headend or hub

b) Increasing the number of simultaneous lambdas that can be wavelength-division-multiplexed on a single fiber (for support of future node-splits) based on the switch from amplitude modulation (AM) optics to digital optics

c) Circumventing nonlinear optical noise effects and increasing the End-Of-Line Signal-to-Noise Ratios (SNRs) and increasing the corresponding spectral efficiencies of DOCSIS 3.1 transport by moving the transmission of the signals closer to the device that a given node serves

Two types of Distributed CCAP Architectures are currently being explored for DOCSIS based access networks. They are:

1) Remote PHY- This architecture moves the DOCSIS PHY (Downstream and Upstream) into the fiber node or a remote PHY shelf and keeps the DOCSIS MAC in a CCAP core in the headend

2) Remote MACPHY- This architecture moves the DOCSIS MAC and PHY into the fiber node

Each of these architectures has its own pros and cons, but both suffer from some common challenges. One of these common challenges is the increase in the number of intelligent devices (managed elements) within the network. Whereas an integrated CMTS or CCAP would be managed as single managed element, the same system built using Distributed CCAP Architectural approaches might have hundreds (or even thousands) of managed elements (including CCAP cores and Remote PHY devices or Remote MACPHY devices). This is because fiber nodes will take on these traditionally integrated functions in a DCA. This increase in the quantity of managed devices will create new challenges that existing management paradigms will struggle to address. A new approach to management will be required. The remaining sections will discuss some techniques that could simplify the management of the large number of DCA elements.

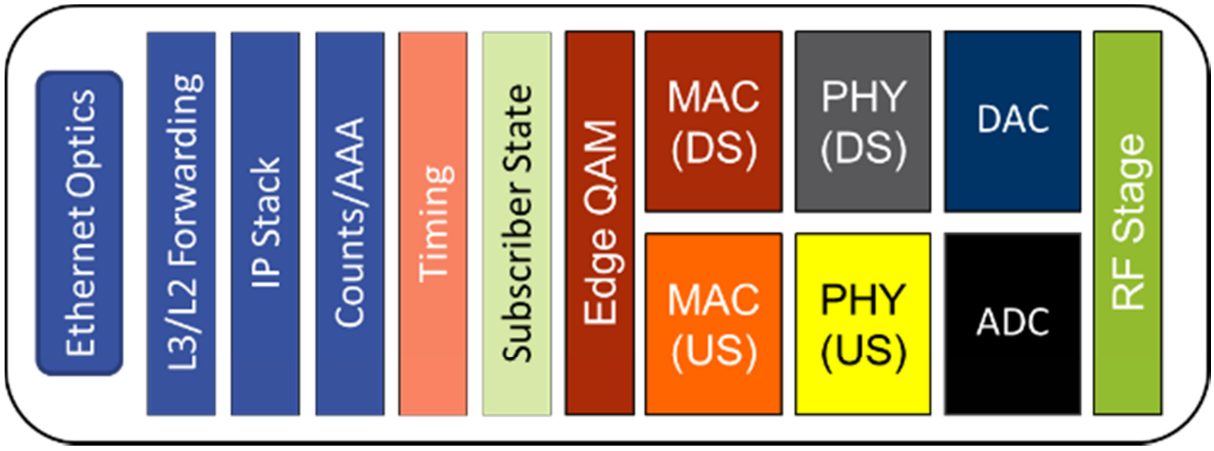


Figure 1: Basic CCAP Data Plane Architecture

Figure 1 shows a very basic data plane structure of a typical Converged Cable Access Platform (CCAP). A cable modem termination system (CMTS) would be similar minus edge QAM functionality. All of the components that are needed for a CCAP or CMTS to forward a given packet to and from the customer premises equipment (CPE) devices are integrated into a single enclosure and located in a headend. Figure 2 depicts the same basic breakdown when a Remote PHY architecture is implemented.

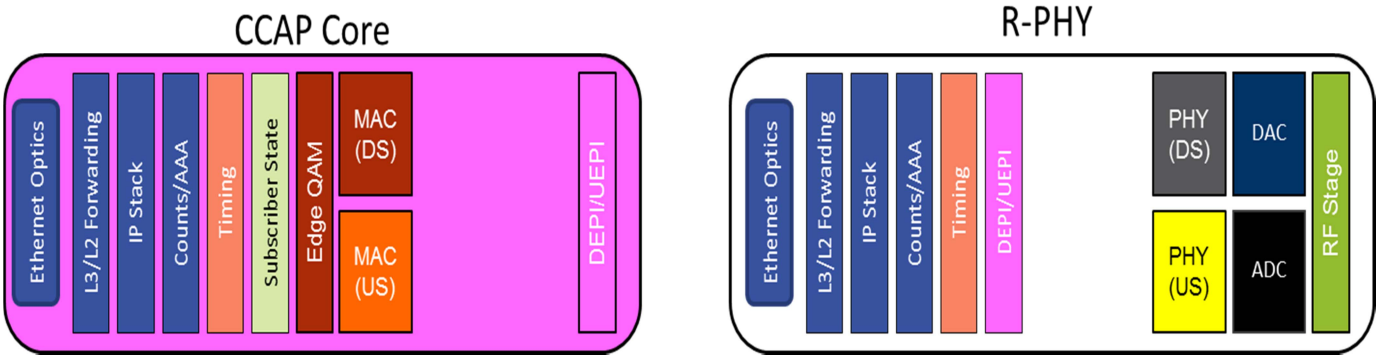


Figure 2: Remote PHY Architecture

In this scenario, the packet handling and DOCSIS MAC related functions are in a CCAP core element (traditionally located in a headend or hub site) and the downstream DOCSIS and video channels as well DOCSIS upstream channels are moved into the optical nodes in the outside plant that service customer homes. Where a single I-CCAP chassis may have serviced 40 fiber nodes, a CCAP core that supports the remote PHY architecture might support many times that number of fiber nodes.

In order to manage this explosion of devices, each of which needing to be provisioned and managed over their lifespans, operators are increasingly looking to evolving technologies like Software Defined Networking (SDN) and Network Function Virtualization (NFV) to help them cost effectively roll out these networks and

the services that run over them. In the next section we'll discuss SDN and NFV in more detail.

SDN, NFV AND ITS APPLICATION TO ACCESS INFRASTRUCTURE

Software Defined Networking (SDN) as a networking paradigm has been evolving since its inception the early 2000's along with the OpenFlow Specifications which were developed in 2008 and 2009. This paradigm is based on the notion that traditional networking equipment has been based on an integrated control and data plane built on proprietary hardware and software modules that are expensive and require long cycles for new features. SDN seeks to address two main goals. The first is the separation of the control plane from the data plane. As a result of this modification, several benefits can be realized, including:

- Networking equipment can become more operationally cost effective
- Commercial off-the-shelf (COTS) servers can be utilized, resulting in the ability to capitalize on the elasticity of COTS servers

The second goal of SDN is the enablement of a higher degree of network programmability, which results in a reduced time to market for features and capabilities. Where a new routing protocol or filtering method may have taken many months for a given vendor to implement on their hardware and software, an SDN approach may deliver that same functionality in a fraction of the time and, potentially, at a fraction of the cost. These benefits are driving the desire to migrate as quickly as possible to architectures that can adopt the new SDN paradigm.

A second technology trend has emerged alongside the SDN efforts. This effort is called Network Function Virtualization (NFV). It has taken advantage the notion that with the separation of the control and data planes, portions of the functions that were previously performed by proprietary equipment and servers could be moved to servers within the network and scaled separately from the hardware and control planes as virtual network functions (VNFs).

Where SDN provides the basic framework for increased programmability of the network and its equipment, Network Function Virtualization (NFV) takes this concept another next step forward by providing capabilities to house these data plane functions on hardware/servers that can be independently scaled to match the needs of the services they support. While SDN and NFV can be deployed independently of each other, the combination of the two technologies makes for a powerful combination.

Increasingly, MSOs and other broadband network providers have been working diligently to determine how and where these technologies can map into their business plans and how to apply these concepts to the access network and data center architectures. Multiple international standards organizations like the BroadBand Forum (BBF), European Telecommunications Standards Institute (ETSI), Internet Engineering Task Force (IETF), and the Open Network Foundation (ONF) have taken up the call to develop standards for their respective members.

Indeed, many telco providers like AT&T and Verizon have already made significant headway in the deployment of SDN and NFV in their networks and already realizing significant gains from it. In the cable space, several MSOs have been very active in defining how they would like to see these concepts deployed in their network. Some MSOs have chosen to focus SDN and NFV operations on the data center before moving out to the access network equipment while others have been equally

focused on the access network and the home CPE.

As mentioned previously, the DCA brings with it several challenges. One challenge is the management and configuration of all of the Remote PHY Devices (RPDs) that are serviced by a given CCAP core. Each of the RPDs is connected to a given CCAP core using an interconnected network of some size. RPDs can either be directly connected to the CCAP core, or there may be a network consisting of one or more network switches / routers between the CCAP core and the RPDs it services.

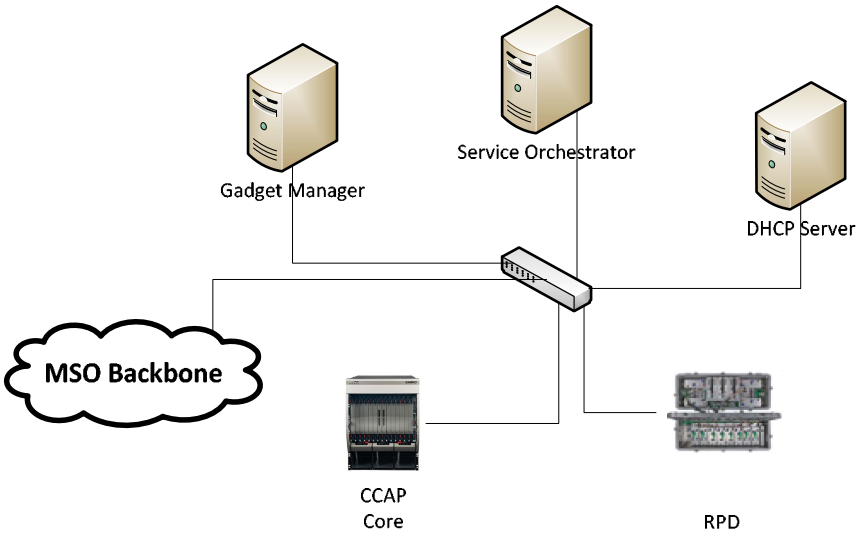


Figure 2: Generic Physical Network

As shown in Figure 4 below, when an MSO migrates from an integrated CCAP chassis to a DCA based architecture, the number of provisioned and managed devices increases very quickly. Where a DCA node in the diagram may have represented a MAC domain and a port on an upstream or downstream line card, the DCA node is an entity that has a MAC domain in the headend / Hub (perhaps virtualized across several servers in the headend /Hub) and then an RPD and the intervening network switches between the node and the hub site.

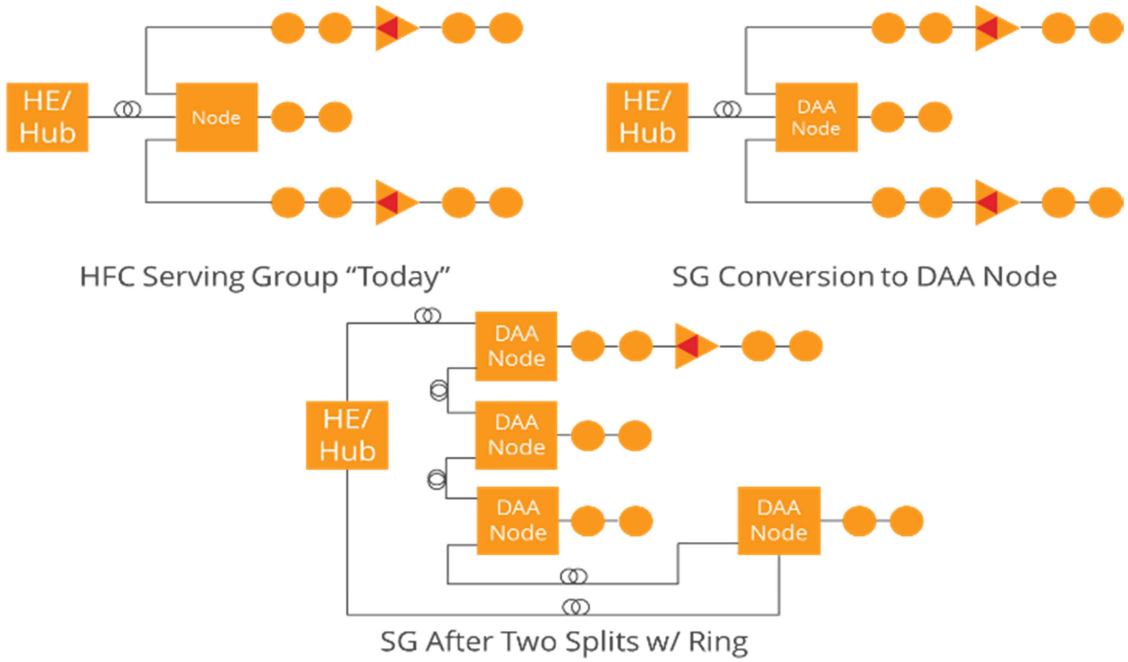


Figure 3: The Managed Device Explosion Brought on by DCA

The increased number of devices will make a CCAP Manager a necessary reality for MSOs. One of the benefits of migration to an SDN configuration model is that centralized control can provide programmatic deployment of systems and services. The ability to manage, configure, and update/expand network capacity utilizing an API based configuration model, rather than SNMP and command line interface (CLI) based model, provides significant benefits to the operator. Eventually the multiple functions integrated in the CCAP today can be pushed to VNFs and these can be located in the network where they make the most sense.

SERVICE ORCHESTRATION - AN EVOLUTION PATH TO A VIRTUALIZED NETWORK

What is Service Orchestration?

Service Orchestration is a concept that has evolved as part of the SDN and NFV efforts. It is the name for the function that facilitates the creation of a given end-to-end service among different Virtual Network Functions. It includes not only the virtual services, but also touches the physical network elements – like CCAP or CCAP Core, Remote PHY devices, and cable modems. The Service Orchestrator along with a NFV Orchestrator (NFVO) can be utilized to implement both the end-to-end connectivity, as well as the virtualized network functions and physical network equipment needed to fully realize a given service that a service provider wishes to make available. The NFVO in this context manages the virtual machine infrastructure that houses the virtualized functions – like RADIUS and related authentication servers, the DHCP services for the RPDs, and related functions.

To better understand the role of the Service Orchestration function in the context of the evolving Distributed CCAP Architecture, we need to look at several issues that are present today. One of these issues is the configuration, or on-boarding, of the RPD with its set of CCAP and Auxiliary Cores. In the current DOCSIS Remote PHY specifications, this is done using

the DHCP server and DHCP Options. In a network where a Service Orchestrator is present, the Service Orchestrator would be able to assign existing or new CCAP MAC and Auxiliary Cores dynamically and provide that information to the DHCP server function which can then populate the values in the DHCP reply messages. Should the Service Orchestrator need to create additional CCAP MAC or Auxiliary Cores, it can contact a NFVO to create additional core functionality and have it ready for the RPD to connect to it. This process shows the power of this dynamic provisioning or on-boarding model – the ability to add in capacity when new devices are discovered. Likewise, functions like the authentication of the RPD and the CCAP cores can be virtualized and those resources managed by the NFVO.

The Service Orchestrator can also be teamed with an SDN Controller in a given location that will help to manage the custom forwarding and message handling that is employed during the RPD initialization and alert the Service Orchestrator that a new or recognized RPD has been discovered. This helps then further distribute the load generated by the on-boarding of the RPD. In the past this was managed by the wiring of nodes to RF cards in the Integrated CCAP and EQAM network in a hub.

The Service Orchestrator can be further augmented by a Gadget Manager application as described in the next section.

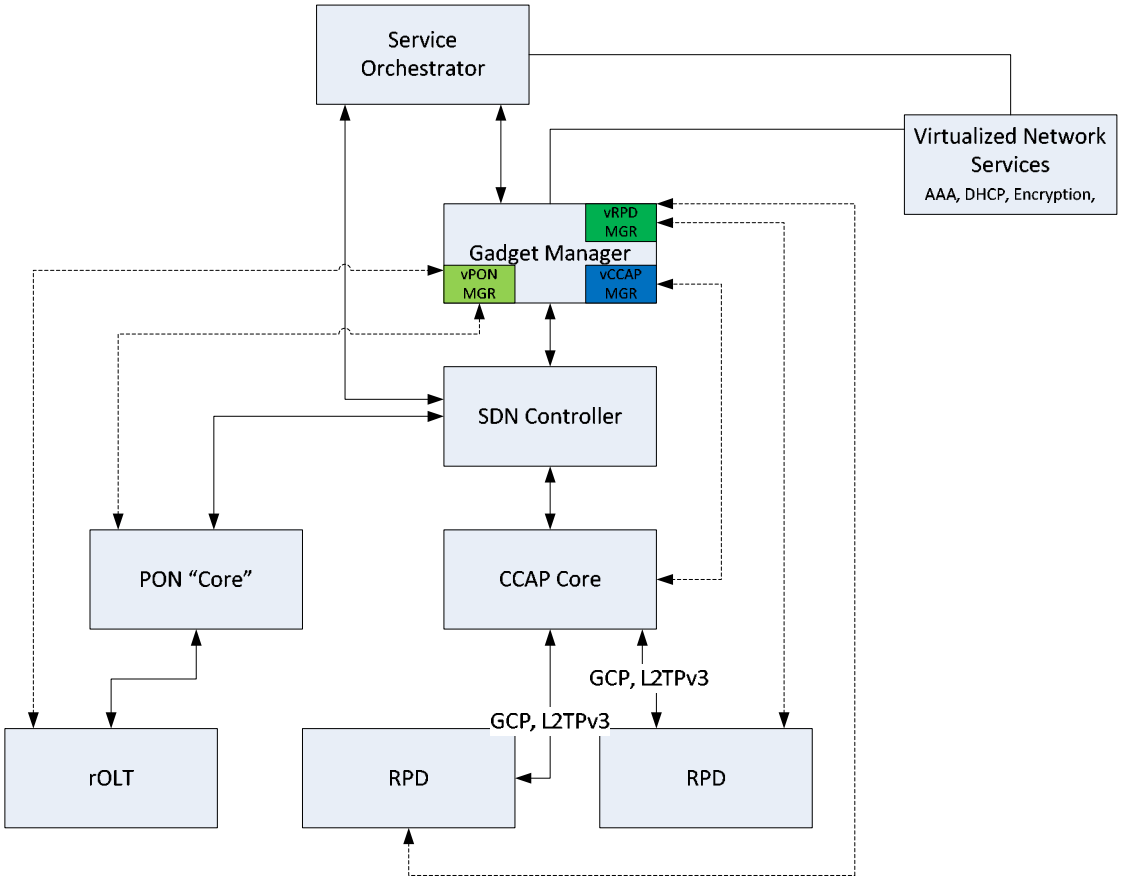


Figure 4: A Logical View of the Network

The Gadget Manager

As discussed above the advent of DCAs it has become apparent that a more centralized and programmatic method is needed to manage the large number of distributed access devices that would be deployed throughout MSO networks in the coming years. This presents both a challenge and an opportunity. The challenge is to build an Element Management System (EMS) that can manage thousands of devices and give the operators the control and monitoring required to support and maintain such a network. The opportunity is the realization that this system can be more than a very elegant EMS with a nice graphical interface. Rather, can be both the User Interface (UI) and the machine-to-machine interface for the system. Next, since this system doesn't live on the access device itself, rather it lives in a virtual compute architecture, it can be used to abstract the access layer from both management and control, thus realizing one of the main goals of SDN – the separation of the control plane logic from the data plane. Ideally the "Gadget Management" function can be used for management and presentation for any access technology including: I-CCAP, DCA, PON, Managed Ethernet, or even G.fast.

This new system needs a name that will communicate that the system is more than simply an EMS and that it can be used to manage any flavor of access device. At the time of the gadget manager's conception the industry had not yet settled on whether the DCA world would be an R-PHY world or an R-MACPHY world. "Distributed CCAP Access Device" could have been used but this seemed too formal, and would potentially create a name that was cumbersome. Also it would not communicate the idea's access agnostic nature. At the same time a small group was referring to DCA devices as "gadgets." A gadget can be just about any device or function, and the term was found to be endearing as it was used by the Manhattan Project as a handle for the design of the first atomic bomb. The atomic bomb's existence radically changed the world, and there is a belief that DCA, access agnostic architectures, and this new EMS that is more than an EMS, will radically change the cable and access worlds. Thus this new EMS has been dubbed the "Gadget Manager" (GM). The GM would be access agnostic striving to eventually manage any access gadget.

The GM is the Presentation Layer

"Presentation Layer" is a fancy name for an API, but in the case of the GM this idea can be extended to be any interface to CCAP system including both well-defined APIs of the DOCSIS OSS system and loosely or un-defined or even proprietary interfaces such as the Command Line Interface (CLI) with which users interact when managing CCAP devices. Thus the GM will "present" as a CCAP to the outside world weather that outside world was a back office management system, or was a human being banging on a CLI.

This idea tracks well with the virtualization efforts that are being discussed in the industry today. For example, consider a collection of R-MACPHY nodes. In the non-virtual world, each would have its own IP address, each would be managed as a single device, each would have its own set of DOCSIS MIBs to be managed and monitored, and each would have its own CLI. In the GM world each still has its own IP address, however users in this world don't use this address to interact with these devices, rather it's the IP address (or one of the IP addresses) associated with the GM with which users and devices will interact.

This is a virtual system in few ways. First, the world is interacting with an intermediate application, and not the devices themselves. The devices are hidden behind a veil of interfaces that make it look and feel as if directly

connected to the device. Second, the logical partitioning of gadgets to "virtual" CCAP is flexible. Meaning the operator can group, un-group, and re-group the DCA devices based on resource constraints that are not related to where the devices physically live. Lastly, the GM itself is virtualized and runs on virtual machine in a compute farm.

The example above can be further filled in. Consider an R-PHY deployment that consists of 100 such gadgets and the operator decides to partition these devices into four virtual CCAPs. The question becomes how these gadgets should be grouped. One method would be to simply split them into four groups of 25 nodes per CCAP core. In this case, one group of 25 nodes could come from then north end of town, 25 from the south, 25 from the east and 25 from the west. This is a simple partitioning that maps the physical world almost directly to the virtual world.

However there may be other methods for partitioning that are more advantageous. Perhaps the number of subscribers attached to each of the 100 gadgets is not uniform. Maybe 50% of the subscriber population lives on 20 of the gadgets. In this case, it may be desirable to balance the subscriber population between the various virtual CCAP instances. Here, we may end up with 30 gadgets associated with one virtual CCAP, another with 10, and two with 20. They may or may not be geographically partitioned. It could be that the 30 gadget V-CCAP consists of the least used gadgets for all parts of town, East, South, North, and West and the two 20 gadget V-CCAP's live only in the South and West parts of town where the bulk of the customer traffic is located.

It should be noted that this system is not virtual in some respects. It may be argued that a virtual CCAP as a system that has had its data plane processing virtualized. Here, with the GM, only the control plane has been virtualized. It may be that the deployment being managed by the GM is an existing physical I-CCAP or OLT, a physical CCAP-Core, or even a virtualized CCAP-Core. The GM is agnostic to the implementation of the data plane. It will be shown later that the benefits of the GM can apply to both physical and virtual data planes.

Purpose Built Gadgets are Compute Bound

Gadgets, whether they are in I-CCAP or a Remote PHY or Remote MACPHY node are all built on special purpose hardware and this purpose built hardware has an upper bound to the amount of processing power, or compute, of which it is capable. This in itself is not a bad thing as these systems are designed with excess compute power. The system designers target is to have at least enough compute – both in terms of control plane and data plane for the last day the system will be deployed. Operators would be reticent to purchase without the belief that these systems will last for a reasonable length of time.

It is difficult to know ahead of time exactly what an operator is going to demand of a system on the last day of deployment during the design phase of a product. It should be noted that the design phase typically occurs 10 years prior to the last day of deployment. Even if a designer did know exactly how an operator would use a system 10 years after it is first deployed, it may not be cost effective with the current generation of hardware to implement.

Thus designers and operators typically live in a world where they must live with the amount of compute that is available, and many times the amount of control plane compute is strained over the life cycle of a product.

The GM is Elastic (Not Compute Bound)

This is crux of the opportunity of the GM. Moving the presentation layer off of the gadget

and onto a virtual machine makes it possible to refactor functions that were once integrated into a more distributed environment. Separating data plane and control functions allows for them to be independently scaled. In a compute bound world an MSO may have wanted to poll all MIBs every five seconds on an integrated CCAP, but was unable to do so due to the compute limits of the system. With the GM an operator will have the ability to do this type of rapid polling; however, it now becomes a cost to the operator to add enough compute to the GM to do so. The operator can thus weigh this cost against any perceived or real benefit.

A more real world scenario is that an operator's needs grow over time. With elastic compute the operator can add compute "just-in-time." This means that at the start of a project, less compute can be deployed initially than would have to be deployed with compute bound solution that cannot be changed over the life cycle of the deployment. In the elastic compute world as more compute is needed it can be added. Also of benefit is that late added compute can take advantage of the performance gains that are occurring throughout the time of the deployment. This has two affects. The first is that the costs are spread out through the duration of a deployment, and second the total cost of the compute is reduced assuming the price of compute goes down during the deployment.

Challenges to Elasticity

While the GM is elastic it is still dependent upon the compute capabilities of the gadget. This is because the GM presents a view of the gadget to the world that is built upon data collected from the gadget. Thus the poll all MIBs at a one second interval example above is only meaningful if the GM can get all MIB data from the gadget every one second. There is also the cost in terms of bandwidth to deliver this data at this rate from the gadget to the GM. Therefore the problems of limited gadget compute and limited network bandwidth available for gadget data collection need to be addressed in order for the GM idea to be successful. A potential solution is presented next section.

Connectionless Push and Gadget Specific Protocol Design

As the gadget Presentation Manager, the GM has to be able to access each gadgets configuration and operational data store. Gadget implementations that support traditional IPDR protocols are possible. However, given a strong MSO desire for each gadget to be low in cost or specific in functionality, there may be limited compute capacity within each of the gadgets. As a result, it may become critical that the communication path between the GM and the individual gadgets be as lightweight as possible. This will help us reduce the work load required of the gadget to processes data and minimizing the amount of data that must traverse the network. Thus a connection-less protocol is preferred over a connection oriented protocol. The reason for this is that with a connectionless protocol, no state is needed for the protocol itself.

By going connectionless the sending of IP data grams can be simplified and even implemented with simple logic in an FPGA or an ASIC. This has a secondary benefit in that the sending of data can be distributed within a gadget to different functional blocks where the data being collected actually lives and since the protocol is connectionless these functional blocks do not have to coordinate with each other, nor does their respective data need to be curated by a central control function within the gadget. All of this simplifies the design and reduces overhead. It would also allow for individual blades such as a multi-bladed I-CCAP device to send data grams. Lastly, previously sent values would not need to be

maintained either in shadow memory or on either volatile or non-volatile storage.

The only downside to a connectionless protocol is the potential for out-of-order or lost packets. One way to mitigate this is to have a sequence number associated with each datum or set of data that is produced by the same packet source within a gadget. A time stamp could be used but could incur undue overhead to distribute and synchronize it. All data sent must be sent as absolute values. Thus packets coming out-of-order to the GM can be re-ordered by the GM which has elastic compute, and any lost packets would be detected, and would only represent a gap for the specific datum or data lost, and only for one collection period.

If that temporal data collection gap due to lost packets is deemed to be undesirable, then alternative protocols (such as those based re-transmission requests or those based on multicasting of the collected data to redundant, check-pointed GM receivers) can also be considered.

Another design consideration that one should take into account when looking at the design of this protocol – should the data transfer be a push between the gadget and the GM or a pull from the gadget? So called pull protocols, like SNMP, poll the gadget at regular and random intervals and pull data from the data store. Push based protocols, like IPDR, are based on a set of predefined service definitions that include elements from the operational data store that are pushed to the GM at regular intervals. IPDR also allows for ad-hoc requests to be made of the operational data store. REST based interfaces like RESTCONF are more API driven and thus suitable for both push (post) or pull (get) and can be implemented using the north and southbound APIs that can be accessed using and SDN controller. In the end, the GM may utilize both types of protocols when gathering the needed data from a given gadget. Likewise, the GM must be able to provide data to operator management stations using both push and pull protocols that SNMP and IPDR so that these systems can continue to function and gather data for applications that rely on that data.

Following the above principles will allow for a fast and robust data collection between gadgets and the GM. While gadgets are typically thought of as new devices that are yet to enter the network, the GM can be deployed with existing I-CCAPs and with the same principles of connection-less, non-prescriptive, and single collector, to improve data collection rates in today's systems.

THE GADGET MANAGER

We have discussed the challenge facing network evolution and the applicability of SDN and NFV technology and the principle of service orchestration- all of these technologies enable network evolution and provide productivity and OPEX gains in managing the network. In applying these methods, we propose the idea of a Gadget Manager. The genesis of the Gadget Manager is that it can immediately address the issues of managing the explosion of devices in a distributed access architecture by managing the deployment and configuration of these new devices. The Gadget Manager can centrally manage devices in operation. It will also become an aggregator of data from remote devices, legacy CCAP and CCAP core, reducing the burden on the infrastructure to collate key performance parameters and establishing a central repository for operations critical data and key performance indicators. The Gadget Manager also plays a vital part in configuring the network as DOCSIS 3.1 is rolled out, and will be the basis of a service orchestrator and SDN controller.

Figure 6 shows the conceptual locations of the Gadget Manager.

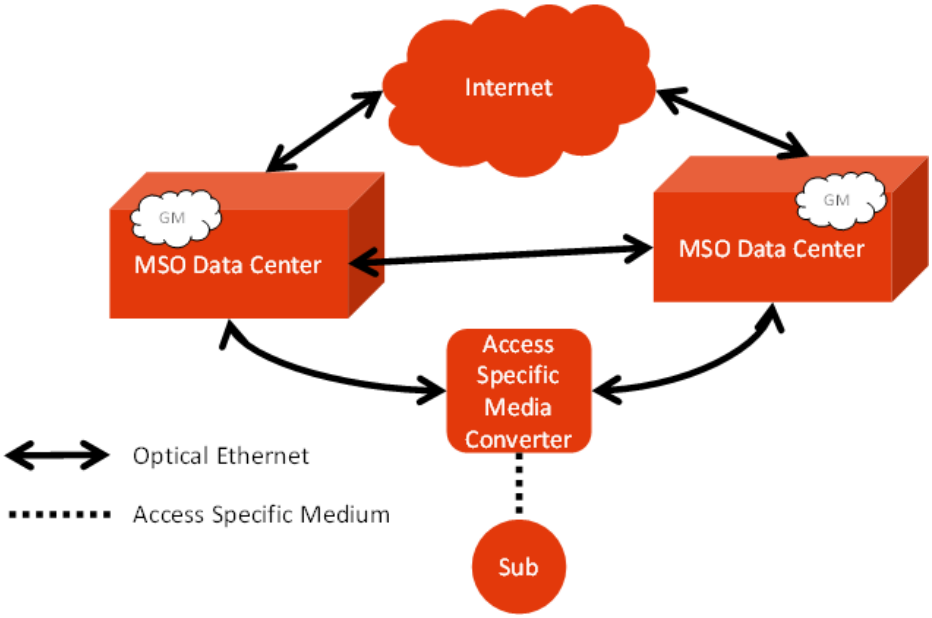


Figure 5: Gadget Manager Architecture

Alternatively the Gadget Manager can be configured as a hosted application for Operators that may not have the resources or choose to operate in a more centralized fashion. This architecture is depicted figure 7.

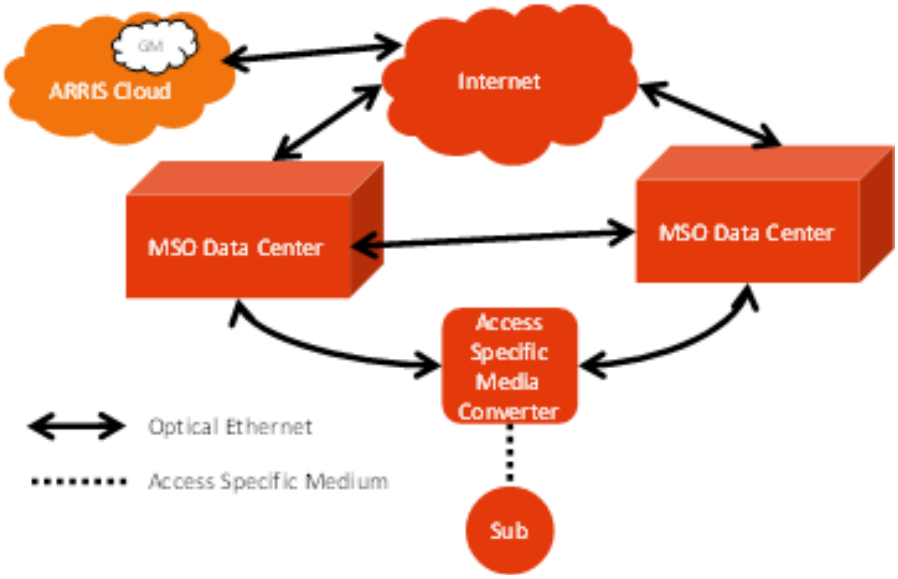


Figure 6: Gadget Manager Hosted Architecture

The Gadget Manager is developed as a cloud-based software that provides:

- An aggregate Interface between the MSO and multiple remote gadget devices
- Manages remote gadget versions and configuration
- Optimizes gadget monitoring
- An aggregate Interface between the MSO and legacy CCAP and new CCAP cores
- Fits into the existing MSO network and allows the MSO to provision and implement the same services that they provision and implement on CCAPs today
- Value added functions such as data aggregation, CCAP core and Remote device license management and reporting, back office integration, device provisioning and traffic management tools
- Eventually, a controller that uses SDN and SON to manage forwarding rules and plumb in NFV functions

Evolution of a Gadget Manager

The evolution of a Gadget Manager could occur in a few phases.

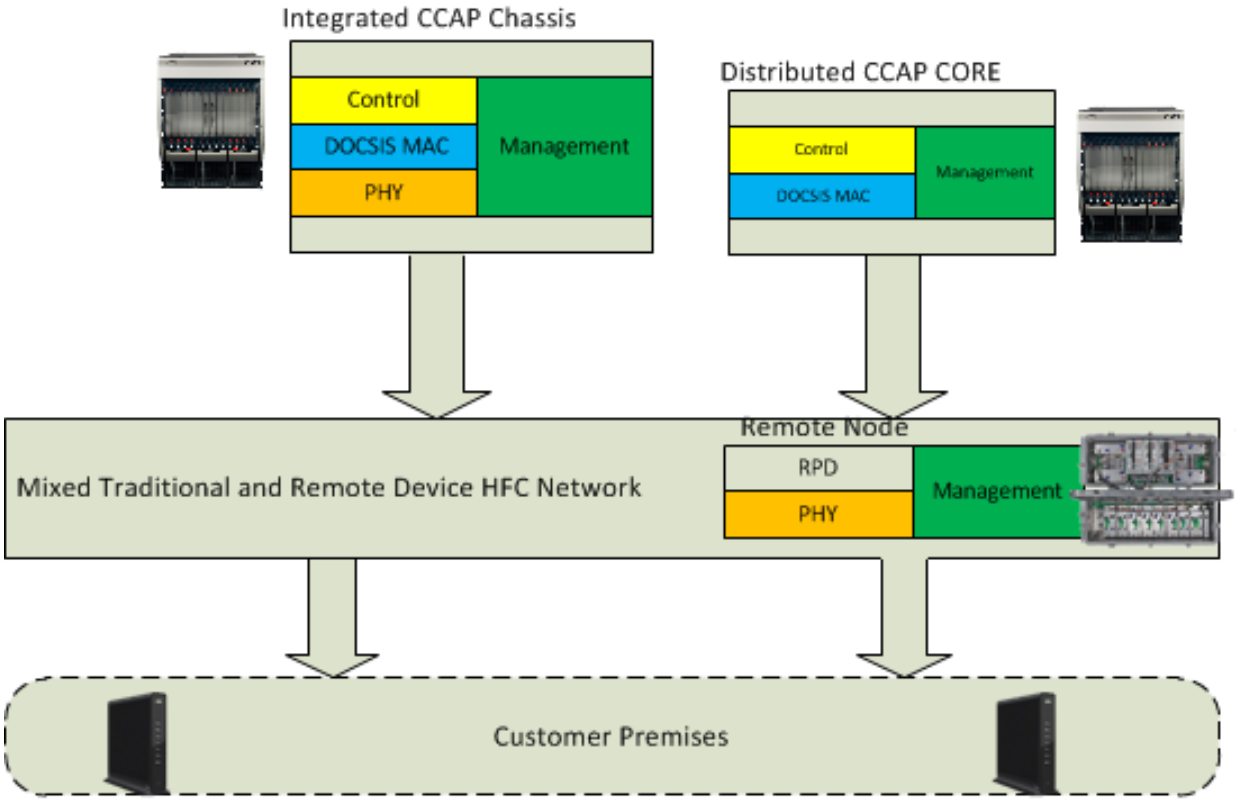


Figure 7: Current Day Network View

Phase 1: the Gadget Manager hosted in the data center provides value added network services in the deployment of DOCSIS 3.1 and remote devices. In this phase, the Gadget Manager can start to provide value added functions like DOCSIS 3.1 profile management and the ability to manage the on-boarding of RDPs to CCAP Cores and discussed previously. This is depicted in figure 9.

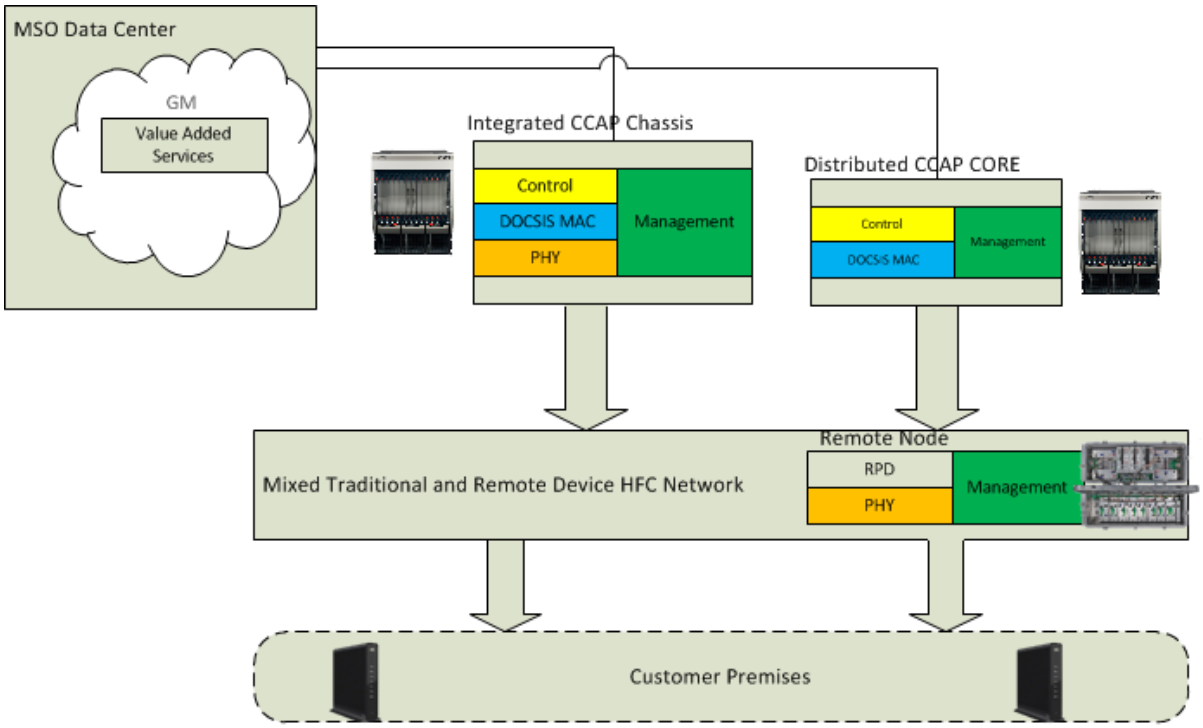


Figure 8: Introduction of the Gadget Manager

Phase 2: In this phase, the Gadget Manager begins to take on more and more of the management functions for the managed devices. This includes the beginnings of the Gadget Manager serving as the management entity for the CCAP and RPD resources by gathering and managing all the performance and configuration data for the core and RPD devices. The GM also provides the configuration interface for the mapping of individual RPDs and CCAP and Auxiliary cores. License management and additional performance applications are also integrated under the Gadget Manager umbrella or interface with the GM to perform their work. This is phase is depicted in Figure 10 below.

The Gadget Manager takes the role of SDN controller for CCAP cores in a distributed access architecture. Its ability to provide control and management of networks with full CCAP, CCAP core/DCA deployments and virtualized deployments protects and extends the life of the investment made in CCAP, providing the benefits of separated control and data planes, simplified management and rapid service creation without having to undertake the risk laden move to NFV before data centers can provide the performance required to effectively run a full CCAP core.

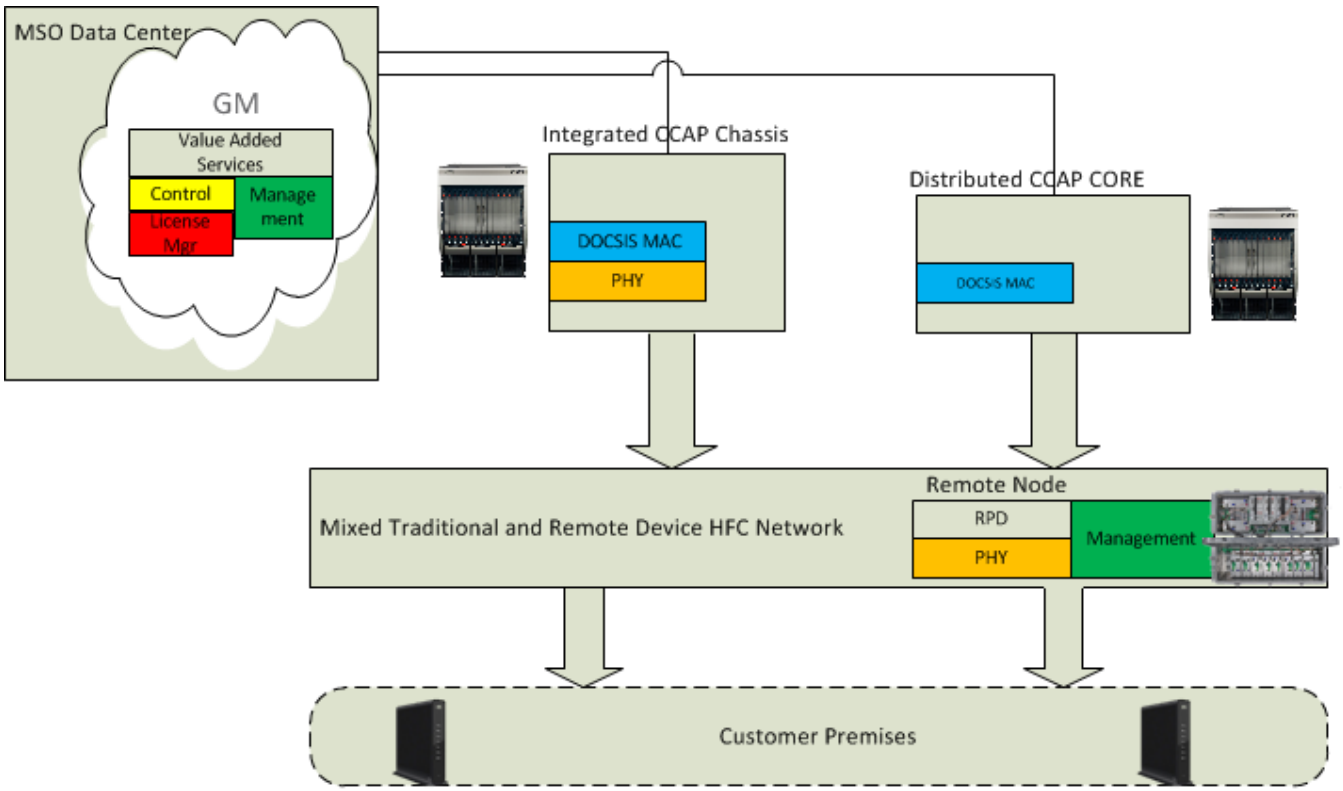


Figure 9: Migration of Functionality to GM

As more and more MAC Core functionality is virtualized, the Gadget Manager becomes more and more integral in managing and configuration of the mapping of the mac core virtualized functions and related traffic to the RPDs and customer homes. Eventually, the Gadget Manager and the Service and NFV orchestrators are enabled to dynamically manage and on-board resources as those resources are needed. This end state is depicted in Figure 11.

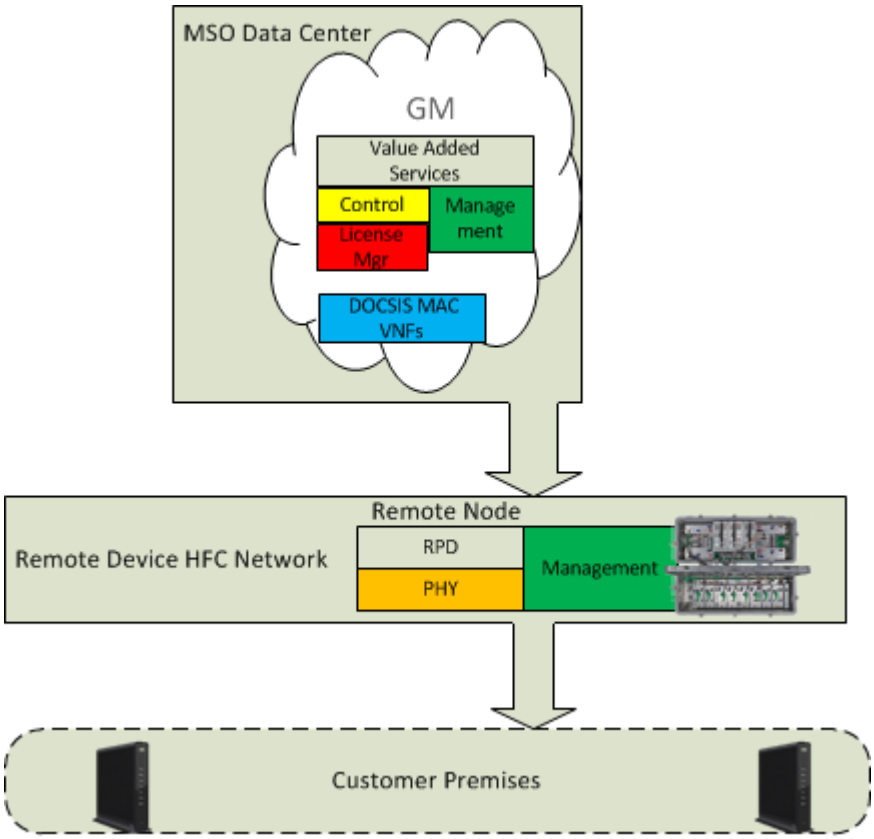


Figure 10: Gadget Manager Fully Realized

SUMMARY

High speed data networks to the customer's premise are evolving rapidly, driven by the need to innovate rapidly and to reduce the cost of both operation and capital spending. These changes encompass not only the Access networks that have been leveraging the Hybrid-Fiber Coaxial plant, but impact how those networks are architected, provisioned and managed. Distributed DOCSIS architectures, whether remote PHY or remote MAC/PHY, will create more intelligent nodes in the network that require new levels of management and will also create valuable data on network performance and effectiveness that must be collected and processed. Mixed distributed and centralized network Architectures will create management domains with different sets of requirements, straining traditional management infrastructures and organizations. Operators will increasingly look to the emerging paradigms like SDN and NFV to help provide a more services-based approach to the provisioning and management of the overall service delivery including the HFC based access network. Using technologies like the Service Orchestrator and a Gadget Manager to configure and manage the increased numbers of distributed HFC gadgets that MSOs will deploy over the next several years, the MSO can realize the benefits of migrating their current service provisioning environment to a more elastic and dynamic model. This paper has shown one possible migration strategy that can be used to migrate from the current generation of provisioning and service enablement to a much more dynamic, cloud based model. This model will permit MSOs to easily change the form and scale of their management system to match the ever-changing demands of the evolving network infrastructure being managed.

ACKNOWLEDGEMENTS

The authors would like to acknowledge and thank some subject matter experts for their invaluable inputs and insights. In particular, Jeff DeMent, Bill Hanks, and Erich Arnold (of ARRIS) were instrumental in giving guidance on several sections within this paper.

ABBREVIATIONS & ACRONYMS

API	Application Programming Interface
BBF	BroadBand Forum
CCAP	Converged Cable Access Platform
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
DCA	Distributed CCAP Architectures
DHCP	Dynamic Host Configuration Protocol
DOCSIS	Data Over Cable Service Interface Specifications
EMS	Element Management System
EQAM	Edge Quadrature Amplitude Modulation
ETSI	European Telecommunications Standards Institute
HFC	Hybrid Fiber Coax
HSD	High Speed Data
I-CCAP	Integrated-Converged Cable Access Platform
IETF	Internet Engineering Task Force
MPEG	Moving Picture Experts Group
MSO	Multiple System Operators
NFV	Network Function Virtualization
NFVO	Network Function Virtualization Orchestrator
NV	Network Virtualization
OFDM	Orthogonal Frequency Division Multiplexing
ONF	Open Network Foundation
PON	Passive Optical Network
RPD	Remote PHY Devices
SDN	Software Defined Networks
SNMP	Simple Network Management Protocol
SNR	Signal-to-Noise Ratios
SO	Service Orchestrator
UI	User Interface
V-CCAP	Virtual CCAP
VNF	Virtual network functions

SMALL CELL BACKHAUL SERVICES WITH A DISTRIBUTED ACCESS ARCHITECTURE AND PACKET BASED CABLE METRO NETWORKS

Andrew Smith
Juniper Networks

Abstract

The growth in mobile voice and data usage is placing tremendous pressure on operators to build scalable backhaul networks to support emerging LTE and LTE-A architectures. At the same time, wireless spectrum is at a premium and must be used in an efficient manner, and the competitive market for cellular customers requires a consistent user experience. For these reasons, operators are exploring architectures to enable a small cell radio to be placed in or near residential or business environments. A modernized cable last mile and metro network could be one path forward. This paper will explore a possible architecture for large scale, small cell backhaul over a Distributed Access Architecture (DAA) and a contemporary all-packet metropolitan transport network.

SMALL CELL OVERVIEW

In mobile telephony and data, the goal is to get data off the air and onto a wire as quickly as possible. A "small cell" is a cellular radio that services a limited area and number of devices. For many years, the notion of a small cell was seen as a device to fill geographic service gaps, such as in rural areas, or as a method to offload traffic from larger macro cell facilities.

In LTE and LTE-Advanced (LTE-A) networks, the small cell is an integral part of the network architecture. The demand for mobile data services simply cannot scale with the existing macro cell architecture, and the constraint of spectrum availability means what capacity exists over the air must be managed with utmost efficiency. These constraints force the operators to deploy small cell architectures to deliver the expected service to end-users. In the LTE and LTE-A architecture, all cells will be self-organizing and are based on the architectural principles specified in Home eNodeB.

In order to make the small cell useful, traffic to and from the small cell must be backhauled to the mobile operator's core network. This can be quite challenging as the small cell will be located in much harder to reach locations than macro cells, and there will be many more of them. The connectivity to the small cell must be of a service provider grade, and the cost to deliver such connectivity must not impede the ultimate goal of blanketed service. Indeed the problem of backhaul of small cell services is cited by most operators as one of the biggest challenges to small cell rollout.

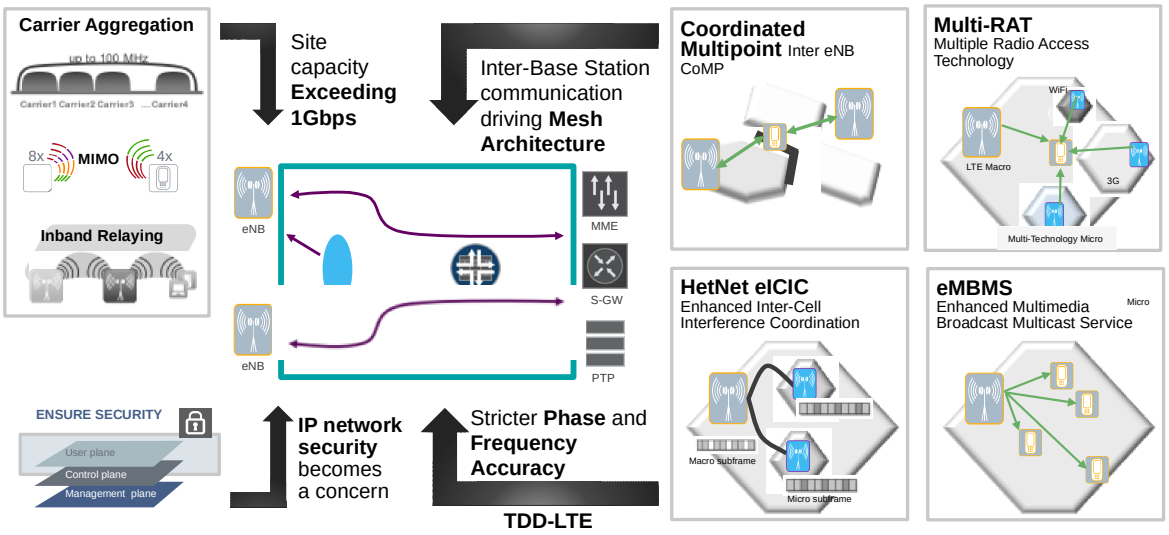


Figure 1: LTE Drivers for Modern Backhaul

DISTRIBUTED ACCESS ARCHITECTURE OVERVIEW

An emerging architectural shift in the cable industry could be tapped to provide effective small cell backhaul services to mobile operators, and solve the challenge of geography, capacity, reliability and cost. This new architecture is known as a Distributed Access Architecture (DAA).

The cable industry is undergoing a dramatic change in the type of product consumers are demanding, and the network must adjust to these new demands. The rise of unidirectional over-the-top video services, such as Netflix, Amazon and others, has changed the economics and expected value of Internet services to the home. As the demand for increased bandwidth capacity grows, the incumbent network architecture centered around a chassis-based CMTS or CCAP, analog optical transport and analog last mile systems will not economically scale. As a result the cable operators are actively exploring inverting the architecture, such that the centralized CMTS device is distributed into many hundreds of elements, deeper (closer to the home) in the network, and connected back to the cable facility with digital Ethernet. In such an architecture the coaxial last mile is preserved. It is envisioned that such an architecture can defer the large capital investment of alternative last mile modernizations, such as PON, for a very long time if not eliminate such investment all together.

A benefit of DAA is the RF characteristics of the analog signal carried by the coaxial last mile dramatically improve. In a typical scenario, this enhancement would be leveraged to provide very high capacity, on the order of many hundreds or even thousand of megabits per second broadband Internet, to the home. In this paper, we explore leveraging part of that bandwidth for small cell backhaul. A second benefit is the transition from analog optical transport to Ethernet enables a large capacity increase and a huge increase in system reliability and dependability. All are key requirements if we are going to leverage DAA for cellular services.

There are, as of this paper being written, two "flavors" of DAA being proposed to the cable industry. One is called "Remote PHY" where the DOCSIS MAC and PHY are "split", with the MAC remaining in the cable facility on a CMTS device and the distributed device (the "node") containing only the PHY circuitry. The other architecture is a Remote MAC-PHY approach, where the centralized CMTS device is eliminated and/or replaced with a permutation of carrier grade Ethernet switching and a service provider edge router, or merged into an existing service provider IP/MPLS infrastructure. In the Remote MAC-PHY case, all cable specific functionality is distributed to the node, freeing the rest of the network from the burden of implementing and managing cable specific technology. In both the Remote PHY case and the Remote MAC-PHY case, the analog transport is replaced with Ethernet.

Because of the applicability of carrier and service provider grade IP/MPLS infrastructure to the Remote MAC-PHY case, this paper will focus exclusively on that architecture. There are several shortcomings of Remote PHY that would make implementation of small cell services quite challenging in that architecture. A specific critique of Remote PHY is beyond the scope of this paper.

CORE AND METRO CABLE NETWORK

The third major element of this architecture is the cable IP/MPLS core or regional network itself. Ultimately all traffic to and from any device in a cable network must transit the regional or core network, and in this case small cell backhaul services are no different.

In a typical regional network there exists a rich deployment of IP/MPLS transport, augmented with MPLS Traffic Engineering such that cellular backhaul services, typically for macro cell, are transported along with other types of VPN traffic and standard Internet access traffic. This document does not explore the specifics of this part of the network, except to indicate that a tight integration between the MPLS VPN infrastructure, the DOCSIS service flows provisioned for small cell backhaul, and timing & synchronization services all must work in harmony for this system to be deployable.

PROPOSED ARCHITECTURE

Cable modem and eNodeB

With the collection of previous elements, we can now begin to explore how to assemble an architecture capable of small cell backhaul over DOCSIS.

The first step is the actual provisioning of the last mile DOCSIS service flows to a DOCSIS 3.1 cable modem or Home NodeB with an integrated DOCSIS 3.1 cable modem. In either case, the Ethernet side of the cable modem will hand data and synchronization services to the small cell device itself. These sync services will be established via a combination of 1588v2 and SyncE from the cable modem. The cable modem will receive it's synchronization from the CMTS via the DOCSIS 3.1 specified DOCSIS Timing Protocol (DTP). This will ensure precise synchronization between the eNodeB, the cable modem, and the upstream CMTS.

A typical small cell will require on the order of 150-200 Mb/s of symmetric bandwidth. To achieve this level of bandwidth on a sustained basis with DOCSIS 3.1, an upstream/downstream split of 200 MHz must be deployed, which typically can produce a usable upstream bandwidth capacity on the order of 1.2 Gb/s.

The distributed nature of the Remote MAC-PHY architecture plays a critical role in upstream bandwidth considerations. First, the notion of deploying a 200 MHz split broadly in a traditional cable network is a daunting task, requiring a large degree of re-engineering and physical plant modifications to accommodate. Because the Remote MAC-PHY node can be placed in a targeted fashion, deployment of the 200 MHz upstream split can be made strategically and tactically. Second, the deployment of the Remote MAC-PHY node lends itself well to the creation of much smaller DOCSIS Service Groups. A Service Group is a collection of homes or end users that share a common pool of bandwidth. In the small cell backhaul architecture, bandwidth in the upstream will be reserved using UGS (Unsolicited Grant Service) DOCSIS flows or RTPS (Real Time Polling Service) type of provisioning based on the specific cable environment.

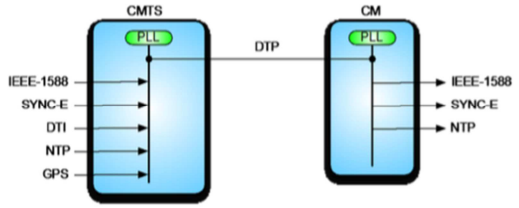


Figure 2: DOCSIS Timing Protocol

TIMING AND SYNCRONIZATION

LTE and LTE-A have very strict timing requirements that must be provided by the backhaul network. All elements in the network must participate as 1588v2 clocks (boundary or transparent) and have SyncE enabled.

LTE-A CoMP

The LTE-A architecture makes use of a scheme called Coordinated MultiPoint Radio. In this scheme, the handset will associate itself with multiple, different base stations to diversify the cellular service delivered to the end user. The benefits are much better utilization of the network and enhanced reception performance of the cellular device.

This means your cell phone will attach to your small cell, and simultaneously your neighbor's small cell and your neighbor's neighbor small cell. In order to make this work, the round trip time (latency) between the small cell eNodeB and the LTE Security Gateway must be very strict, on the order of 3-5 ms.

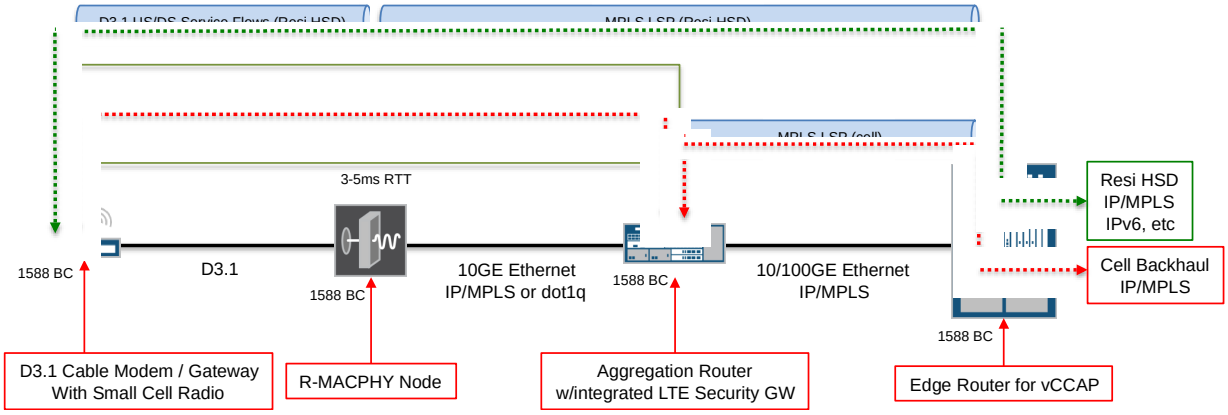


Figure 3: Overall system architecture

In this architecture, we propose locating the Security Gateway close enough to the Remote MAC-PHY node such that the latency to the small cells can be accounted for. In this way, the traffic for small cell backhaul with CoMP can be treated differently and with more accuracy than standard high speed data Internet services. The Security Gateway will then label switch cell backhaul traffic to the Edge Router.

vCCAP EDGE ROUTER

The Edge Router in this architecture is the central administrative and policy enforcement point for all services transiting the vCCAP architecture.

In this architecture, the Edge Router will handle small cell backhaul and traditional residential internet access (HSD) independently. That is, small cell can be treated with a degree of service that is different than regular HSD. Leveraging the featureset inherent in MPLS, the edge router can hand cell backhaul traffic directly to the cell operator via any standardized mechanism – optical wavelength, dedicated interface, or an MPLS LSP. At the same time, residential internet can be simply IP routed or can also be encapsulated into an MPLS LSP, depending on the cable operator's existing infrastructure.

In this way the addition of small cell backhaul services is a net additive operation on top of the existing edge router, and not a wholesale architectural change for this portion of the network.

SUMMARY AND CONCLUSION

This paper has explored, at a high level, how an ensemble of emerging technology could be applied to the challenge of small cell backhaul services. We have attempted to build off of shipping, standards based technology and to create a unique integration of the cable access environment with the specific demands of cellular services.

Admittedly there exist gaps in this architecture. No vendor currently is building a combination small cell eNodeB with DOCSIS 3.1 cable modem. The R-MACPHY architecture has yet to be deployed in mass at any operator (to be fair, no DAA has been deployed as of this writing). And there are some operational considerations with MSO's when extending MPLS VPN technology to the cable last mile that must be accounted for.

But overall, there exists a tremendous opportunity for the cable industry to leverage a very lucrative asset in the coaxial last mile, and apply it to services other than traditional high speed Internet. DOCSIS and coax cable can be used to carry just about any sort of data service, and with efforts like that explored in this paper we predict a very strong case for continued investment in cable infrastructure for the foreseeable future.

Ian Goetz of Juniper Networks contributed to this paper.

Streaming OTT TV Apps as Virtual Channels

Moiz Lokhandwala
Time Warner Cable

Abstract

Over the last few years we have seen the transition of TV from channels to content. There is no limit to the media content found on the internet, this content is either free or fee based depending on the content provider. All the Media OTT services are accessed via 3rd party platforms, such as Apple TV, Roku, Fire TV and the list goes on. These Media OTT apps also provide services to watch movies or TV shows or even Live TV. This App model is good when you have 1 or 2 apps to watch, but if the content which the subscriber enjoys is spread across multiple Apps than it becomes frustrating for him to switch between the Apps and is not at all user friendly. The content switching between the Apps needs to be as easy as switching Live TV channels. A new approach on how the OTT apps can be as user friendly as Live TV is needed across all 3rd party platforms.

The approach to resolve the issues is to get single sign in to all OTT Apps and bring the content across different Apps to the same screen. The idea would be to develop a common framework which interacts with different Apps and pass this information as live content and can be appended as Virtual channels to the existing framework of the Set Top Box (STB).

As more Apps are being used by customers, different login credentials across different Apps become tedious and time consuming. This issue also needs to be resolved by means of saving the customer profile and sharing it across different Apps.

This paper highlights some of the challenges for OTT TV Apps and also some ideas on the possible solutions. The OTT Apps have the content and now they need a common frame work to make the App more user friendly across different platforms. Also the paper outlines on how the common framework approach can be utilized across different devices to get a more generic user interface and user friendly Live TV experience.

BACKGROUND

The TV industry is changing and the cable TV is no more limited to one screen. We have seen that the cable companies have their own TV Apps for live TV and On Demand content. This approach gives their customers Live TV experience on multiple screens. Also the MSO's are trying IP only Video services using 3rd party platforms which could work in the favor of developing a common framework to integrate the OTT Apps with Live TV MSO App.

A common framework strategy by an MSO can help them deliver new set of Apps branded exclusively for them. MSOs can deploy their own branded app store that complements their Live TV experience, On Demand TV, and other catalog of apps and widgets.

CHALLENGES

With so many media OTT Apps available, the consumer is challenged on which Media App should he watch and why should he invest in that App.

Figure 1: Sample Customer challenge

The following are the challenges the consumers face.

- Different login credentials for different apps.

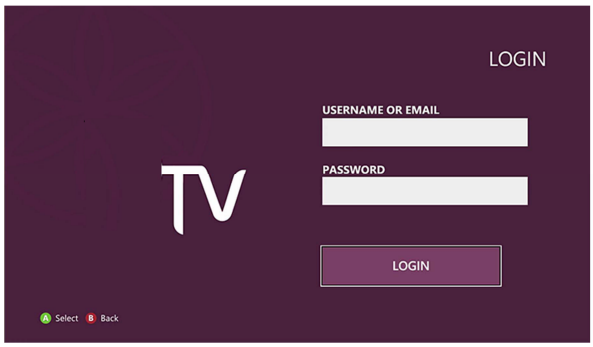


Figure 2: Sample TV App login page

- Not flexible as changing TV channels.

Figure 3: Sample TV App channels available on a platform

- Incompatible ecosystems on different platforms

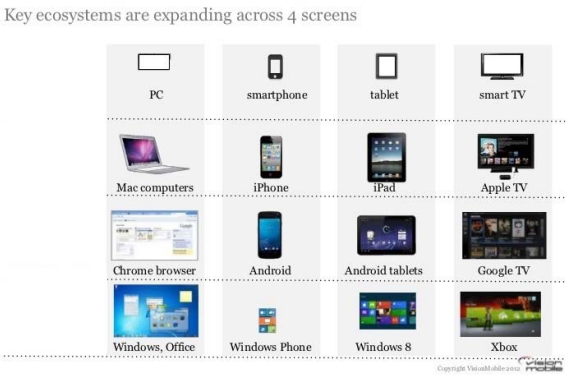


Figure 4: Sample Ecosystems

SOLUTION IDEAS

- MSO App can be an App aggregator. Content from various channels made available regardless of the origin. MSO App can be Master App. This App provides a common framework for all the other OTT TV Apps and add those TV Apps as virtual channels to the MSO Master App. The Master App can integrate content from Live TV, On Demand, App content and also Live free channels available online without copyright infringement.

Figure 5: Sample Integration of Channels

- Treat an App as Multiple channels. Populate the Media OTT App as a Virtual channel on the Set top Box

(STB) or on the MSO TV everywhere app. This would require change in the STB guide with the focus on moving to web-based technologies. From a technology point of view, HTML5 provides the best abstraction level. HTML5-based app platforms are attractive from a development perspective, but pose a host of complexities which needs to be overcome by the App developers.

- Create a common frame work for authentication
  Create a profile on the device from where these multiple Media App's are accessed, so that the user does not have to authenticate always. Create a profile with MAC address of the device, so that the credentials are not shared and are not violated. This could be easily achieved by utilizing the same authentication done as in case of Wi-Fi services.
  Also in case of a cable operator if a user is already a subscriber for a TV service then he should be able to use his profile to login to an App service of that same TV service. This would create a lot of simplicity for the user and also less user credential data for the App service.

- Provide better Remote experience
  Integrating all the channels in multiple app environments to one common App will provide better remote experience. One of the major drawback in the OTT TV App is that the remote are different and the interface is different for different platforms. Remote interface could be the simplified if the Apps are integrated in one common framework. Also the search across different Video Apps using the Voice commands would be easier, flexible and more efficient. We know Voice integration is Remote is now available in almost all the platforms and the key to it is how it gets extended across multiple platforms.

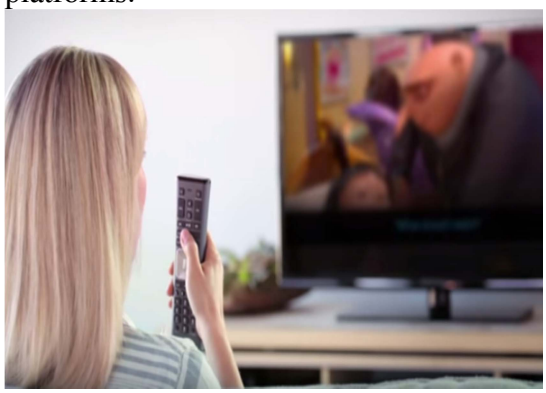


Figure 6: Sample Voice Remote

COMMON FRAMEWORK

Common framework can be achieved by creating a controlled environment, using a specific operating system and a platform where Video Apps can be developed in one unified way.

Developers also need to ensure that this common framework can handle integration of old style STBs UI and the new Video Apps UI. It should be also be scalable to support advanced features which would enhance the user experience.

Advantages of Common Framework

The common framework offers a lot of advantages, some of which are listed below

- Common framework can be integrated in the cloud which could be used across different STB and browser based platforms.
- Apps need to be developed once across different platforms.
- The common framework will support Antenna channels, Live TV, On Demand and free web based channels all on the same TV screen.

NEW FEATURES

The common framework would support content integration agnostic of the source of the content. Also it would allow the operators to support new features and enhance customer experience to both legacy STBs and new browser based platforms. Some of the examples of new features are listed below.

- One of the new feature is Voice search on Remote which could be enhanced to work across all Apps on common framework.
- Common framework will also allow the cable operators to integrate social media and live video experience with family and friends.
- Pairing of Mobile and other devices to common framework could simplify and content sharing would be easy.
- Digital tracking of content could be leveraged across devices.

CONCLUSION

Common framework will provide the Cable Operators a mechanism to integrate live TV, app content and VOD seamlessly. This framework will also work with legacy STBs and will allow MSOs to deploy branded App store and Apps. It will also allow to integrate new features and improve user experience.

The Architecture of the Open-Source Remote Phy Device

Alon Bernstein
Anlu Yan
Cisco Systems

Abstract

Cablelabs has recently announced the open-sourcing of the remote phy device (RPD). This paper outlines the vision and main components of the open software that will drive the RPD. For those who seek an in-depth review it's recommended to sign up to cablelabs openRPD project, read the documentation and of course read the source code !

MOTIVATION AND VISION

Open source has evolved over the years, from a lets-all-be-friends culture of code sharing to an economically driven joint development framework. The original 1985 GNU manifesto stated that "The fundamental act of friendship among programmers is the sharing of programs," and that the commercialization of software is bad. For years to follow the mental image of an open source contributor was that of a poorly groomed collage kid hacking away code late at night. While there are elements of this visual that are true, it turns out that open source is a big business that employs many salaried adults. "Open" should not be confused with "free" and as we explain in this paper at the base of it open source is a joint development framework that can help reduce costs, accelerate feature development and increase software quality for the following reasons:

1. Cablelabs members have jointly developed specifications ever since cablelabs was created. This has proved a very successful model for both operators and vendors. Open source represents a natural next step for this existing collaboration model since published code can be viewed as a super-detailed definition of how a feature should work.
2. As the industry has experienced in the past, publishing an interface specification does not guarantee smooth sailing for inter-operability tests. Open source has great potential in shortening the interop stage because all vendors share the same interface code base.
3. Open source does not mean that there is no vendor differentiation. On the contrary: the open source components should focus on the functions that provide no vendor differentiation, for example parsing messages, reading various sensors, writing alarms to logs etc, thereby allowing a vendor to focus on those features and components that provide added value.
4. The parts of the code that are open source get tested and deployed in a large number of systems, not only a single vendor system. Because of the licensing agreement any fix to the open source part of the code gets committed back to the project resulting in higher software quality overall.
5. With open source the operator community and vendor community have an opportunity to closely collaborate, this is especially relevant as many of the larger operators are large software development houses as well.
6. "open source" is not only about source code, it's a complete SW development environment including source control, bug tracking and simulation code. All of the above is supported by Cablelabs.

7. And last but not least, open source is good for the engineering staff. Their open source contribution are the ultimate referral tool, and the communities that they build is a safety net and support system.

LICESNING AND PRODUCTIZATION

Open source projects are not products. There is no commercial support structure around them. Open source is a base on which products can be built. For example, the open source code from linux.org is not used directly. There are several commercial distributions (e.g. redhat, android) that use the linux kernel and provides support and enhancements around it.
The open source licensing is put in place to make sure there is reciprocation in the community that the software can be used but bug fixes and feature enhancements are provided back to the code base. There are a couple of licenses that are typically used in open source projects, most notably GPL and Apache. The Apache license is considered a very permissive type of license.
From the licencing point of view, the OpenRPD project consists of two parts. One is the OpenRPD packages and the other is enhancements and modifications to OpenWrt. The OpenRPD packages consist of Remote PHY specific contents, to run as user level processes and communicates with the OpenWrt part via IPCs. This part is licensed under Apache 2. The OpenWrt part consists of Linux kernel and other user level processes, some with OpenRPD specific modifications and enhancements. This part retains the licenses of the original packages, many of which are GPL.

KEY ARCHITECTURAL CHOICES

The following sections outline the main tenants of the open RPD software and the reasons for eastablishing them.

This document does not discuss the hardware design of the RPD, however as a reference it's useful to present an overview of the hardware. Figure 1 depicts a high level diagram of the main hardware components in the RPD design.

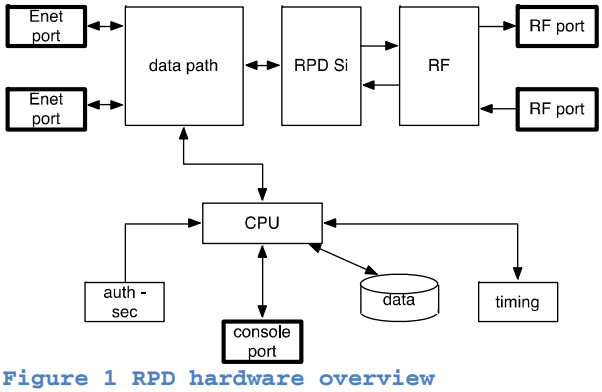


Figure 1 RPD hardware overview

For the most part the vendor specific portions of the software are to be the drivers for the data path, the RPD Silicon and the RF interface. As noted later there is a common HAL (hardware abstraction layer) that provides a uniform, yet extensible, interface to program and read stats from the hardware. Many of the software modules above the HAL are generic and provide little differentiation as they focus on the standard registration/authorization/authentication/etc. processes that have been already standardized by Cablelabs.

OPERATING SYSTEM

The base OS for the openRPD is openWRT, an embedded operating system based on Linux kernel. While openWRT was originally designed for home gateways it is a good base for the openRPD because its designed for embedded systems and has and active community around it and a rich library (more then 3500 packages) that guarantees its longevity as a platform and the ability to add capabilities into it.

MODULARITY, MODULARITY, MODULARITY

The key to a successful project, and an open source project specifically, is a modular design. This is critical because of several reasons:

1. For openRPD vendors developing a product in parallel to the open source system. Certain modules can be implemented in a proprietary way because of time to market pressure or other reasons. If the base infrastructure is modular it's easy to swap these components if and when an open source equivalent exists.
2. The same argument applies in the reverse as well. A vendor may start with an open source module and later replace it with a proprietary one.

The main modules for the open RPD are depicted in `Figure 2` and for the most part correspond to the Remote phy architecture as defined by cablelabs in the remote phy specitication specification (http://www.cablelabs.com/specification/remote-phy-specification/):

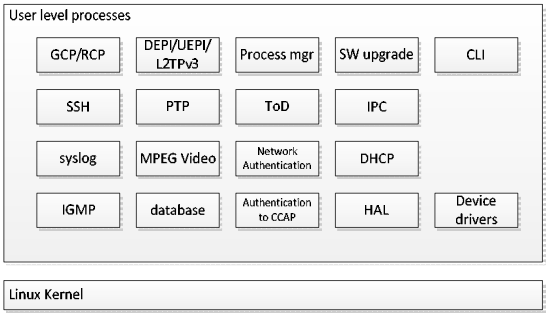


Figure 2 Main software modules for open RPD

To assure modularity the system components communicate over a bus as depicted in `Figure 3`:

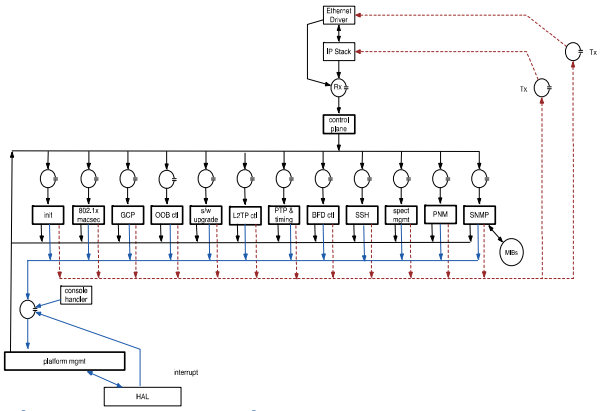


Figure 3 Bus architecture for open source RPD

The IPC technology used for open RPD is ZeroMQ. It's a high-speed asynchronous message library designed for distributed applications and for modularity reasons we treat our components as part of a distributed system even if physically located on the same hardware. ZeroMQ also provides sub/sub and push/pull capabilities for message exchange. With ZeroMQ different modules can be developed in different programming languages and easily swapped into and out of an implementation.

A modular design requires a small overhead because modules communicate over a message infrastructure instead of sending messages directly. However this is relativity insignificant and as Donald Knuth has declared "**premature optimization is the root of all evil**". Clean separation of modules is key to the success for both the RPD open source and RPD products based on it.

GOOGLE PROTOCOL BUFFERS

For internal communications over ZeroMQ the RPD software uses Google protocol buffers (GBP). GBP is a method for serializing data and looks a little bit like JSON definition, mostly used for defining messages in a distributed system.
This choice of GBP further supports the modularity as well since its endian-independent, programming language independent and processor independent

DATABASE

OpenRPD includes a small database (DB) as a generic service which different processes can use in specific ways. One common use is to store persistent data. Persistent data can be used for fast error recovery in cases of process crash. As we can see in the following section the DB plays a critical role in the hardware abstraction layer (HAL) process.

HARDWARE ABSTRACTION LAYER

There is a lot of innovation that can occur at the RPD layer with new technologies coming along such as DOCSIS 3.1 and Full-Duplex DOCSIS. To help accelerate this innovation the open RPD project defines a HAL (hardware abstraction layer) that would make it easy to register multiple drivers from multiple vendors onto the same platform framework. These drivers can be different implementations of existing physical layers or new ones.

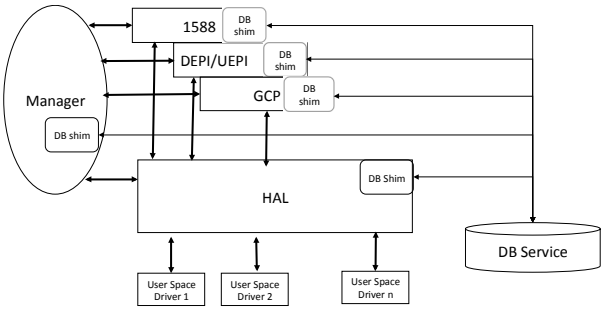


Figure 4 Hardware Abstraction Layer

A key part of the HAL is the database (DB). The HAL uses the DB in the following way:
- GCP may use DB to store shadow data.
- The HAL may use DB to manage data transactions, i.e. GCP stores data in the DB and HAL can pick it from the DB in the following way:

- Each HAL app client (GCP, DEPI, etc.) registers it's supported notification message type to HAL
- Each driver registers its supported message to HAL
- Using a set of notifications the client app knows when the HAL has new data and vice versa.

.

BUILD ENVIROMENT

As open source project is more than a collection of files. It includes a build environment, a bug tracking environment and a test harness. All the above are considered part of the open RPD software.

CONCLUSION

Open source is the next step in vendor and operator collaboration, bringing more consistency and interoperability then the traditional specification writing process and at overall lower cost of implemetation. The open RPD project is hosted and supported by Cablelabs. Readers are encouraged to contact CableLabs and become active contributors to the openRPD project !

REFERNCES

1. openWRT : https://openwrt.org
2. google protocol buffers : https://developers.google.com/protocol-buffers/
3. ZeroMQ : http://zeromq.org
4. Cablelabs remote phy specifications : http://www.cablelabs.com/specification/remote-phy-specification/

The Capacity of Analog Optics in DOCSIS 3.1 HFC Networks

Zian He, John Skrobko, Qi Zhang, Wen Zhang
Cisco Systems

Abstract

The DOCSIS® 3.1 (D3.1) HFC network, supporting OFDM, requires potentially higher system performance (CNR or MER) for both downstream (DS) and upstream (US) signals compared with the previous DOCSIS® 3.0 (D3.0) counterpart. This pushes the limitations of current analog optics technology. How much optical link budget can current technology support? It mainly depends on the per channel optical modulation index of the laser, and optical detection limitations which are dominated by receiver (Rx) equivalent input noise (EIN). This paper will mainly focus on evaluating the noise limitation of the downstream and upstream analog optical links. We explore whether D3.1 PHY layer defined OFDM/OFDMA QAM signals and the extended channel loading bandwidth can be satisfied using an analog optical link. The results show that DS 4K-QAM OFDM signals with full loading (5x192 MHz) can be supported when the optical Rx input power is no less than -2 dBm. Similarly, US 1K-QAM OFDMA signals with full loading (2x96 MHz) can be supported when the US optical Rx input power is -13 dBm. In other words, analog optical links can still support the D3.1 PHY layer defined 4K-QAM/1K-QAM for DS/US, but there are limitations of fiber link loss. Digitizing the optical link of HFC using EDR or Remote PHY can help break the power budget limitations of the analog optical link.

INTRODUCTION

The evolution of the DOCSIS specification has occasionally impacted the familiar HFC network. The D1.1 release had no impact because the addition of QOS using service flows improved utilization of existing bandwidth. The D2.0 release, on the other hand, added a higher order modulation format and wider channels. Although some effort was expended to increase robustness through enhanced FEC and interleaving, the inevitable increase in channel loading caused a review of the upstream plant. Upstream Fabry-Perot lasers, for example, fell out of favor. The significant milestone in throughput brought about by D3.0 and channel bonding had little impact on HFC because it was also an improved utilization of existing bandwidth. Our most recent evolutionary step to D3.1 has forced us to revisit HFC because of both new modulation formats and broadened bandwidth [1]. Higher modulation formats require review of CNR requirements and bandwidth changes require review of per channel power.

The D3.1 Physical Layer Specification defined the DS cable modem minimum required CNR performance of OFDM signals with LDPC error correction in AWGN (Additive White Gaussian Noise) channel as shown in Table 1 [2]. Typical 1K-QAM OFDM required signal performance is 34 dB CNR/MER (assuming CNR and equalized MER has a similar value in dB units), or ~41-41.5 dB CNR/MER for the 4K-QAM OFDM DS option. Note that these values of CNR are only meant to achieve post-FEC packet error ratios of less than or equal to 10^{-6} with 1500 byte Ethernet packets.

Constellation (QAM)	CNR (dB) Up to 1 GHz	CNR (dB) from 1.0-1.218 GHz
4096	41.0	41.5

2048	37.0	37.5
1024	34.0	34.0
512	30.5	30.5
256	27.0	27.0
128	24.0	24.0
64	21.0	21.0
16	15.0	15.0

Table 1 – CM Minimum CNR Performance in AWGN Channel

A similar situation has occurred for the D3.1 US path. The US CMTS minimum received CNR of OFDMA signals is shown in Table 2.

Constellation (QAM)	CNR (dB)
4096	43.0
2048	39.0
1024	35.5
512	32.5
256	29.0
128	26.0
64	23.0
32	20.0
16	17.0
8	14.0
QPSK	11.0

Table 2 – CMTS Minimum CNR Performance in AWGN Channel (non-boosted pilots)

Fundamentals of CNR of a Linear Optical Link in Fiber Deep HFC

Generally, the quality of the RF signal channel can be described by Composite Carrier to Noise (CCN). The CCN of an HFC fiber link is the combination of noise components coming from the carrier to noise ratio (CNR), carrier to intermodulation noise (CIN), and the carrier to crosstalk noise (CXN). CCN can be written as [3]:

$$CCN^{-1} = CNR^{-1} + CIN^{-1} + CXN^{-1}$$

where:

- *CNR* is the total noise contribution of the Rx (thermal noise, shot noise) and the Tx (laser noise).
- *CIN* is the distortion noise that comes from second, third, and higher order components.
- *CXN* is noise coming from crosstalk noise from non-linear fiber interactions, or other fiber transmission induced noise.

For a well-designed linear optical link, the RF non-linear distortion can be controlled to making a very low contribution (CSO, CTB <-65dBc), if the gain block is kept in a linear operation range and the laser is not pushed to clipping. For mitigating the optical non-linear interaction, it is a more complicated topic, which should also consider Four-Wave-Mixing (FWM), Cross-phase-modulation (XPM), Raman Crosstalk, and WDM filter slope induced distortion. But by careful configuration, such optical crosstalk can be reduced to an acceptable level [4, 5]. With

these pre-conditions, CCN = CNR for linear optical downstream and upstream path design.

DS Analog Optical Link

Refer to the CNR equation of a linear optical link [6]. The calculation parameters of a typical downstream path are shown in Table 3.

Parameter	Value
Rx Input power (dBm)	-25 to 0
Rx Responsivity (A/W)	0.95
OMI per 6 MHz	0.021
RIN (dB/Hz)	-155
EIN (pA/√Hz)	3.5
Noise Bandwidth (MHz)	5.36

Table 3 – CNR calculation parameters

The calculated CNR result of the downstream optical link is shown in Figure 1. The EIN of the state of the art is 3.0-3.5 pA/√Hz for a 1.2GHz DS Rx so here we use 3.5 pA/√Hz in the calculation. The per channel (6 MHz) OMI used in the calculation is 2.1% for typical HFC DS with SC-QAM full loading (190 QAM channels, ~20% composite OMI). The noise bandwidth is 5.36 MHz for a DS 6 MHz SC-256QAM. It can be seen in Figure 1 that for the optical input power (Pin) above -8 dBm, the total CNR is dominated by shot noise. The thermal noise of the Rx does not contribute much to the total CNR. The total CNR curve deviates from the linear relation with the shot noise line when Pin goes below -8 dBm. Pin = -14 dBm is a turning point, where below it the total CNR is, instead, overtaken by the thermal noise of the optical Rx.

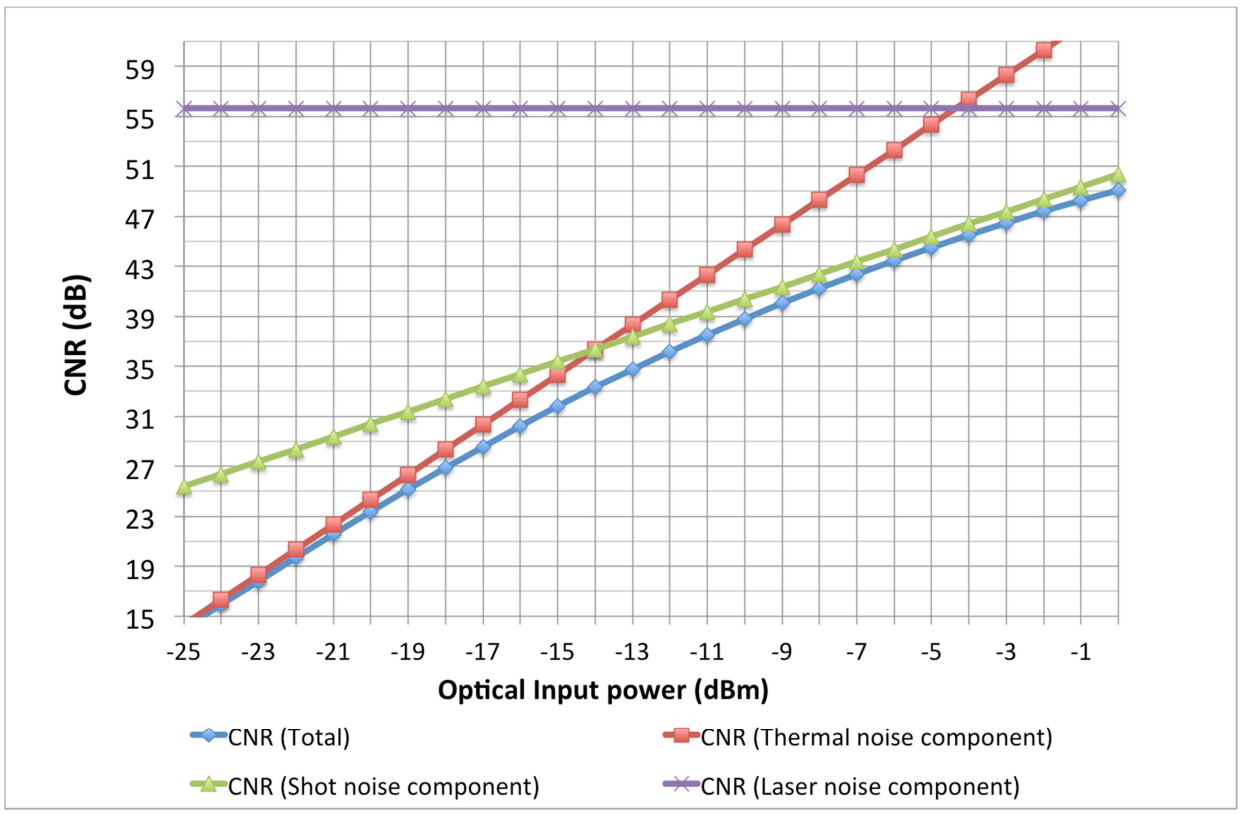


Figure 1 – CNR vs Optical Input Power (Downstream path)

Modulation error ratio (MER) is a "CNR-like" parameter for QAM signals compared with analog signals. To be simplified, in this paper we assume equalized MER and CNR

have the same value in dB units for a QAM signal [7]. In addition, the total CNR/MER should include the signal source baseline. In the past, we have been satisfied with SC-QAM baseline MER values of 44 dB. Figure 2 shows the calculated CNR/MER with this limitation. With the introduction of higher order modulation and OFDM, we are using source baselines at and above 50 dB. Figure 3 demonstrates the impact of this higher expectation. The optical input power in simulation ranges from -12 to 0 dBm. It can be seen in Figures 2 and 3 that as EIN increases, it induces a CNR/MER degradation, especially at lower optical input power levels. So, for higher loss links, such as for fiber deep applications using WDM, an optical Rx with low EIN is even more important than previous HFC scenarios.

Parameter	Value
Rx Input power (dBm)	-12 to 0
Rx Responsivity (A/W)	0.95
OMI per 6 MHz	0.021
RIN (dB/Hz)	-155
EIN (pA/√Hz)	1.0 to 8.0
Station base MER (dB)	44 (SC-QAM)/50 (OFDM)
Noise Bandwidth (MHz)	5.36 (SC-QAM)/6.0 (OFDM)

Table 4 – QAM CNR/MER calculation parameters

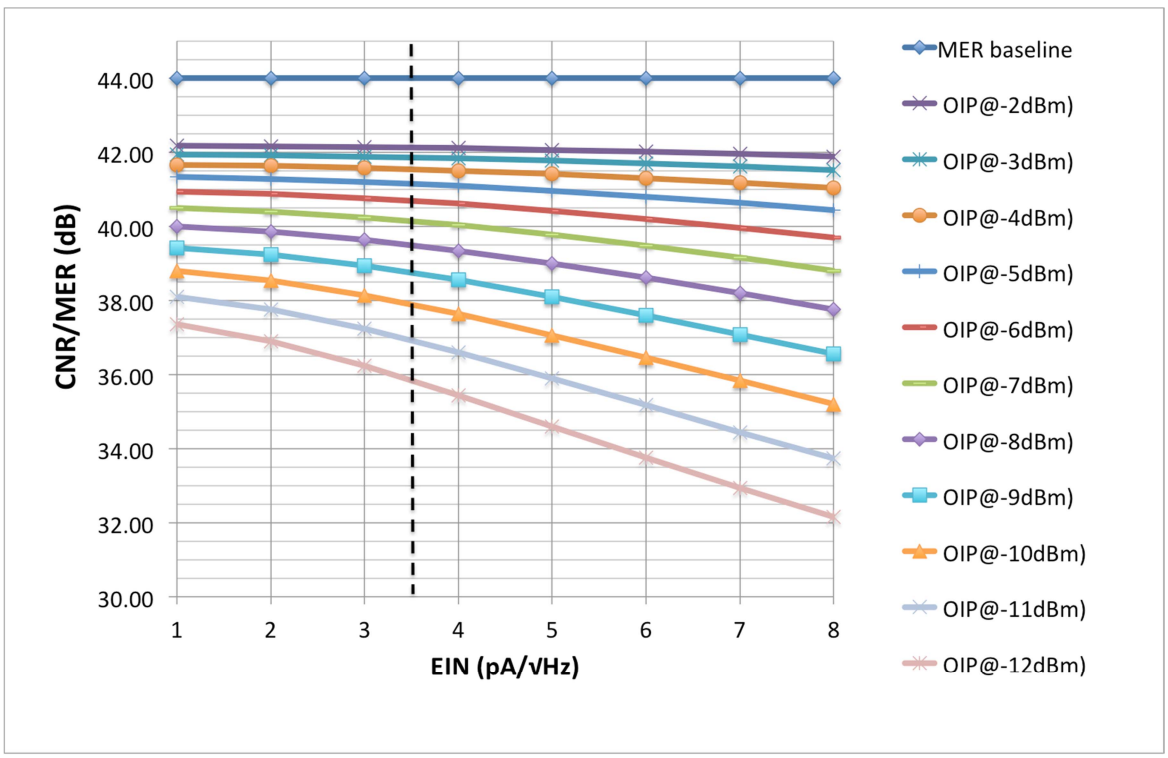


Figure 2 – CNR vs EIN of SC-QAM (calculated)

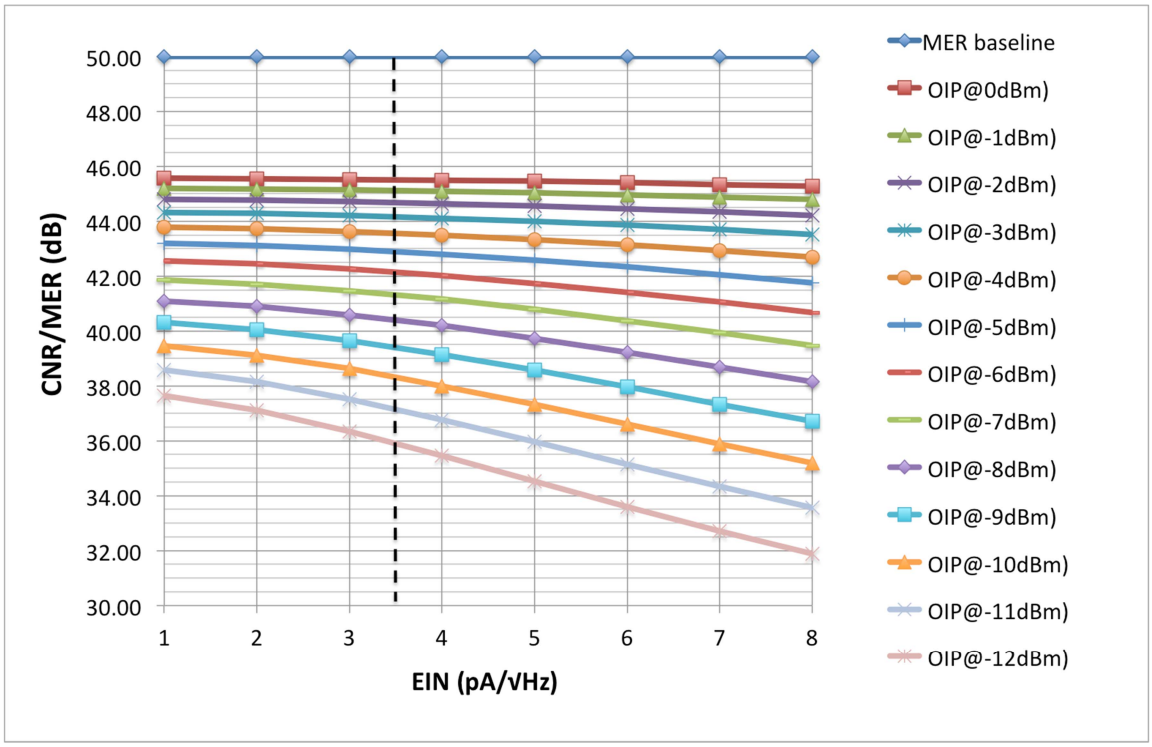


Figure 3 – CNR vs EIN with improved baseline (calculated)

Practically, we need to keep a certain margin (usually 3 ~ 6 dB) above the minimum required CNR for both SC-QAM and OFDM signals [1]. Table 5 shows the marginal CNR (with 3 dB margin) and the corresponding required minimum optical input power (OIP), while assuming the current achievable EIN is ~3.5 pA/√Hz (the vertical dashed line in Figures 2 and 3). The result demonstrates that the optical Rx with 3.5 pA/√Hz EIN can support SC-256 QAM and OFDM 1024QAM if the OIP is above -11 dBm, which is enough for most of the existing homes passed (HP) coverage. However, for D3.1 4K-OFDM DS signal, the OIP should be no less than -2 dBm.

Constellation (QAM)	SC-QAM		OFDM	
	Marginal CNR (dB)	Min OIP (dBm)	Marginal CNR (dB)	Min OIP (dBm)
4096	NA		44.5	-2
2048	NA		40.5	-8
1024	NA		37	-11
512	NA		33.5	< -12
256	37	-11	30	

Table 5 – Minimum Optical Input Power for Downstream D3.1 HFC

US Analog Optical Link

For upstream, the Noise-Power-Ratio (NPR) dynamic range is often used to specify the operational range of the upstream RF or optical link [8]. The NPR value can be loosely interpreted as CCN of the QAM signal. The left side of the NPR curve is mostly established by the system noise, while the middle and right side are dominated by

distortion and laser clipping. In practice, a minimum NPR dynamic range is required for an US analog link. Typically it should be >12 dB [1]. To provide more data throughput upstream, D3.1 extends the upstream bandwidth from 5-85 MHz to 5-204 MHz, more than twice the bandwidth compared with D3.0. At the same per channel power, the heavier channel loading of 5-204 MHz will inevitably overdrive the optical Tx "earlier" than light loading.

Figure 4 shows the measured NPR curves of different channel loading. The upstream Tx used in this test is a 1550 nm DFB type. The upstream Rx input power is -13 dBm after 25 km of fiber and optical attenuation. The EIN of the Rx is ~1.3 pA/√Hz. It can be seen that the right side of the NPR curve shifts to the left with the increase of the noise loading bandwidth. The dynamic range difference between 5-85 MHz and 5-204 MHz loading is about 4-4.5 dB, which also matches well with the theoretical expectation. Notice for this example of -13 dBm optical input power, the 5-204 MHz loading NPR curve has 12 dB of dynamic range @ 35.5 dB NPR. This is what is required for transmission of a 1K-QAM OFDMA signal. In addition, the same dynamic range is available for a 2K-QAM OFDMA signal transmission with 5-85 MHz loading at this optical input level.

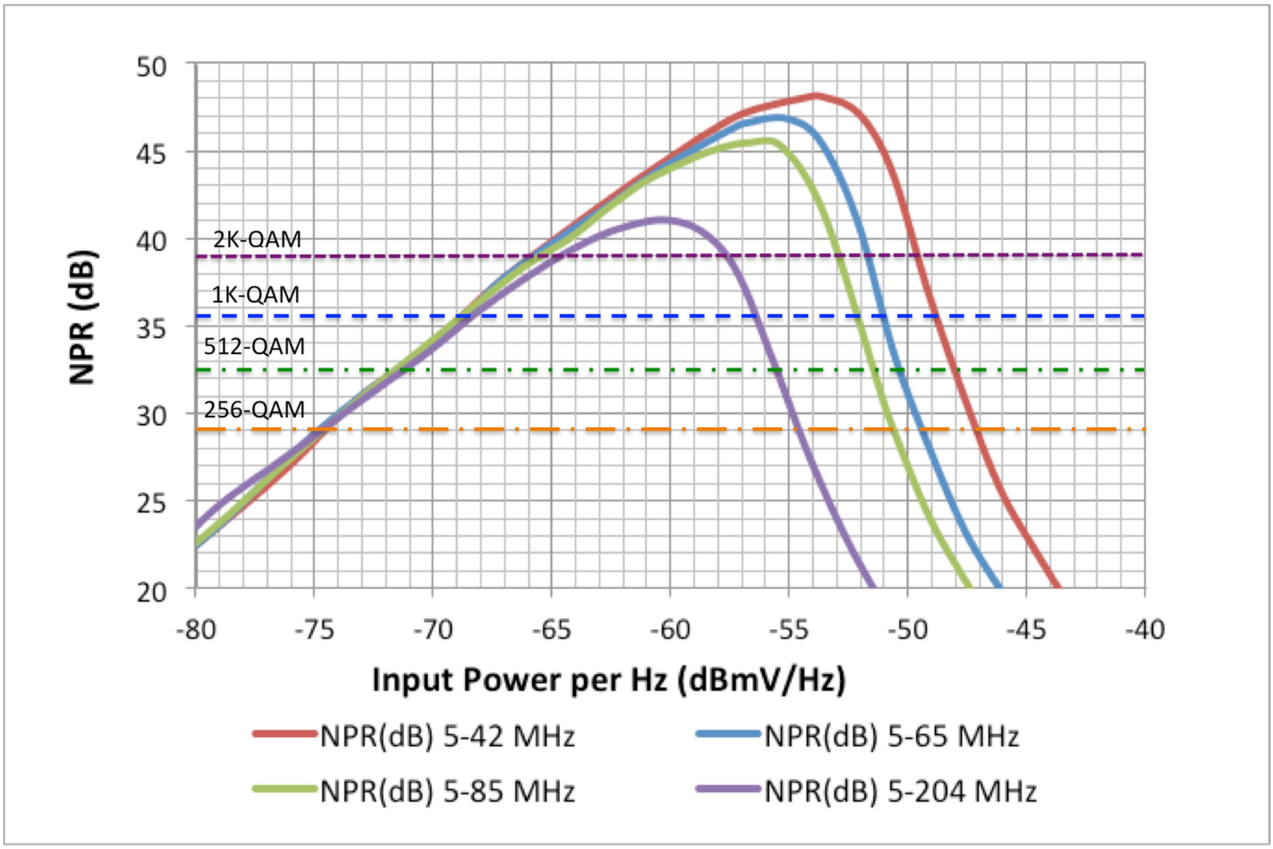


Figure 4 – Measured NPR of different channel loading

However, -13 dBm may not be a low enough optical input power for an analog fiber deep (especially for cascaded WDM scenario) or RFoG application. Assuming the EIN of an US optical Rx to be 1.3 pA/√Hz, the CNR result in Figure 5 shows that the total CNR deviates from the linear relation with optical input power when the optical input power is lower than -17 dBm. If looking at -22.5 dBm optical input power, the total CNR will be 2.5

dB lower than the linear extrapolation, which means the NPR dynamic range at -22.5 dBm will have -13-(-22.5)+2.5 = 12 dB degradation compared to -13 dBm. Consider that at -13 dBm OIP, 24 dB dynamic range of 29 dB NPR is achievable for 5-85 MHz loading, then at -22.5 dBm OIP, 5-85 MHz signal loading will have 12 dB dynamic range at 29 dB NPR, which can only be used for an OFDMA-256QAM signal.

Parameter	Value
Rx Input power (dBm)	-25 to -10
Rx Responsivity (A/W)	0.95
OMI per 6.4 MHz	0.05
RIN (dB/Hz)	-155
EIN (pA/√Hz)	1.3
Noise bandwidth (MHz)	6.4 (OFDMA)

Table 6 – CNR calculation parameter

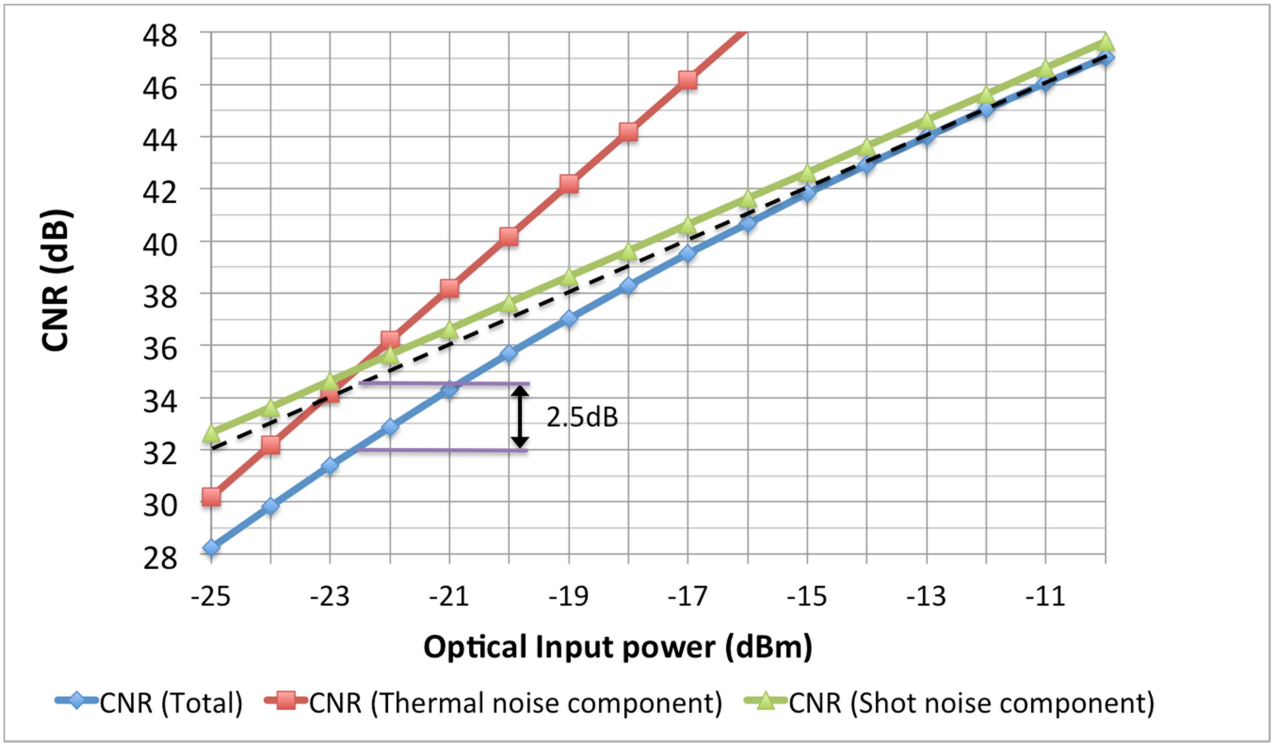


Figure 5 – CNR vs Optical input power (Upstream path)

As a summary for the upstream path, we list the required minimum optical input powers for different OFDMA QAM levels and different loading cases in Table 7. The result shows that 1K-QAM OFDMA can be supported with OIP higher than -17 dBm for 5-85 MHz loading, and -13 dBm for 5-204 MHz loading.

Constellation (QAM)	Required CNR (dB)	5-85 MHz loading (1x24 MHz+1x48 MHz)	5-204 MHz loading (2x96 MHz)
		Required Min OIP (dBm)	
1024	35.5	-17	-13
512	32.5	-20	-16
256	29	-22.5	-19

Table 7 – Minimum Optical Input Power for Upstream D3.1 HFC

Advances of Digital Optical Link

Based on the analysis above, we know that for D3.1 HFC, the analog optical link is still workable for 4K OFDM DS and 1K OFDMA US, but the optical link budget has certain limitations. For the DS, the typical Tx output power is +10 dBm, while for the US, it could be +3 dBm. So the DS & US link budget is 12 dB and 16 dB, respectively. From the analysis above we can see analog optical links will degrade the CNR with the reduction of the Rx input power. Alternatively, digital optical links are less affected by link loss providing the optical input power is above the digital Rx receiving sensitivity. EDR (Enhanced Digital Return) and R-PHY are two viable digital methods for HFC fiber links. Consider typical 10GE optical transmission link budgets (excluding the dispersion penalty), with the EML Tx minimum output of 0 dBm, and APD Rx receiving sensitivity of -23 dBm. The link budget is 23 dB, much better than the analog optical link (see Table 8).

OFDM/ OFDMA	Constellation (QAM)	OFDMA signal bandwidth (MHz)	Optical link budget (dB)		
			Analog Link	EDR	R-PHY
DS	4096	5x192	-12	NA	-23
US	1024	2x96	-16	-23	-23

Table 8 – Optical Link Budget Comparison

CONCLUSION

We have analyzed the current analog optical Tx/Rx performance for its data transmission capability in DOCSIS 3.1 defined OFDM (OFDMA) HFC networks. The results show that analog optical links can still support most of the scenarios which DOCSIS 3.1 requires with a proper optical link budget.

REFERENCES

1. J. T. Chapman, M. Emmendorfer, and R. Howald, "Mission is Possible: An Evolutionary Approach to Gigabit-Class DOCSIS", NCTA, 2012.
2. "Data-Over-Cable Service Interface Specifications DOCSIS®3.1 Physical Layer Specification", CM-SP-PHYv3.1-I08-151210
3. S. L. Woodward, and M. R. Phillips, "Optimizing Subcarrier-Multiplexed WDM Transmission Links", Journal of Lightwave Technology, Vol. 22, NO. 3, March 2004.
4. T. Brophy, F. Villarruel, and Kuang-Yi Wu, "The Delicate Balance - Choosing Robust Alternatives for Multi-Wavelength Systems", Cisco white paper, G1667A
5. F. Villarruel, V. Mutalik, Kuang-Yi Wu, and M. Schemmann, "When Wavelengths Collide Chaos Ensues: Engineering Stable and Robust Full Spectrum Multi-Wavelength HFC Networks", SCTE Cable-TEC Expo paper, 2011
6. "Using a Philips Optical Receiver in CATV Applications", Philips Semiconductors application note of AN98060, 2001
7. R. Hranac, and B. Currivan, "Digital Transmission: Carrier-to-Noise Ratio, Signal-to- Noise Ratio, and Modulation Error Ratio", Cisco white paper, 2006
8. ANSI_SCTE_119_2011 "Measurement Procedure for Noise Power Ratio", SCTE, 2011

ABBREVIATIONS

APD	avalanche photodiode
AWGN	additive white Gaussian noise
bps	bits per second
BW	bandwidth
CCN	composite carrier to noise ratio
CNR	carrier to noise ratio
CIN	carrier to interference noise
CXN	carrier to crosstalk noise
CSO	composite second order
CTB	composite triple beat
D3.1	DOCSIS®3.1
DFB	distributed feed back
EDR	enhanced digital return
EIN	equivalent input noise
EML	externally modulated laser
FEC	forward error correction
FWM	Four-Wave-Mixing
GHz (MHz)	giga Hertz (mega Hertz)
HFC	hybrid fiber-coax
LDPC	low-density parity-check
MER	modulation error ratio
MSO	Multiple system operator
NPR	noise power ratio
OFDM(A)	orthogonal frequency-division multiplexing (assess)
OIP	optical input power
OMI	optical modulation index
QAM	quadrature amplitude modulation
RIN	Relative intensity noise
R-PHY	Remote PHYsical interface
Rx	receiver
SNR	Signal to noise ratio
Tx	transmitter
WDM	wavelength division multiplexing
WL	wavelength
XPM	Cross-phase-modulation

THE FIBER FRONTIER

Dr. Robert L. Howald
Vice President, Network Architecture
Comcast

Abstract

There is a natural conflict between the pace of technology change and the practical pace of change that can occur across the network in a large scale operation, such as that of an MSO. Because of this, it is always the case that there will be significant uncertainties to weigh when making network investment decisions.

A commonly encountered uncertainty is assessing the merits and timing of an investment path associated with development and mass deployment of a new technology. In particular, the question that must rightfully be asked at each technology turn is whether one investment path that appears sensible is instead more wisely passed over in favor of a succeeding emerging technology path. This reflection and assessment must continually be made even should the former path be "plan of record," and even if it is already underway. With the pace of technology change accelerating and in many directions, the timing of new technology introduction is never perfect. It is rarely even good.

There is no simple solution to this puzzle, only a logical "best effort" path that requires operators to increasingly understand that business plans are living documents, and sometimes short-lived at that. Investment decisions should place a premium on flexibility and open standards, and be based on a vision of a long term end state to serve as a guidepost.

There are many example of the above dynamic for MSOs. One of the more timely examples this paper will focus on is fiber evolution for Cable Operators. Operators have been putting more and more fiber into their networks for years. This will surely continue. The question that the industry is grappling with more so than ever is the nature, scale and appropriate path forward for fiber usage beyond its long-standing HFC and business services roles, and the implications this has for operators given the aforementioned imbalance of rate of change between technology and network evolution.

This paper will take a look at how the various stakeholders of one major US operator defined the network evolution problem statement with a long term quantified perspective, determined a vision forward, and developed a plan to deliver on this vision. We will discuss and enumerate some of the assumptions, levers, compromises, platforms, and customer objectives driving the recommendations. We will consider alternative scenarios and reflect on the deviations this may cause in a long-term plan. Finally, we will postulate long term states of services and architecture, and discuss heavy-hitting variables and their potential impact. In short, we will provide an example template for the Fiber Frontier, with an eye towards investment criteria, practical change, and customer experiences.

PROBLEM STATEMENT: PERSISTENTLY AGGRESSIVE CAGR

Part of being a network operator in the age of the "lifeline" service known as broadband Internet access is having an obsessive attention to capacity management. As is commonly referenced in nearly every network

architecture paper, the growth of consumer bandwidth has been persistent and aggressive since the introduction of the Internet, and particularly so in the downstream-to-the-home direction. This all makes logical sense as the technology quickly advanced to support all of our historical communication and media types – text characters to web pages to pictures to video and HD video, massive storage and file transfer, and now towards emerging Internet of Things (IoT) applications. While predicting the next "big" thing to drive Compound Annual Growth Rate (CAGR), something continually emerges to keep the CAGR train moving forward. A simple and sound rule of thumb for planning purposes is to assume this persistently aggressive CAGR charges forth at 50% per year. This means a doubling of traffic roughly every 21 months.

Figure 1 is a cable-specific look at how a 50% growth rate and a starting point of an existing capacity deployment reflect on cable system thresholds of capacity over time.

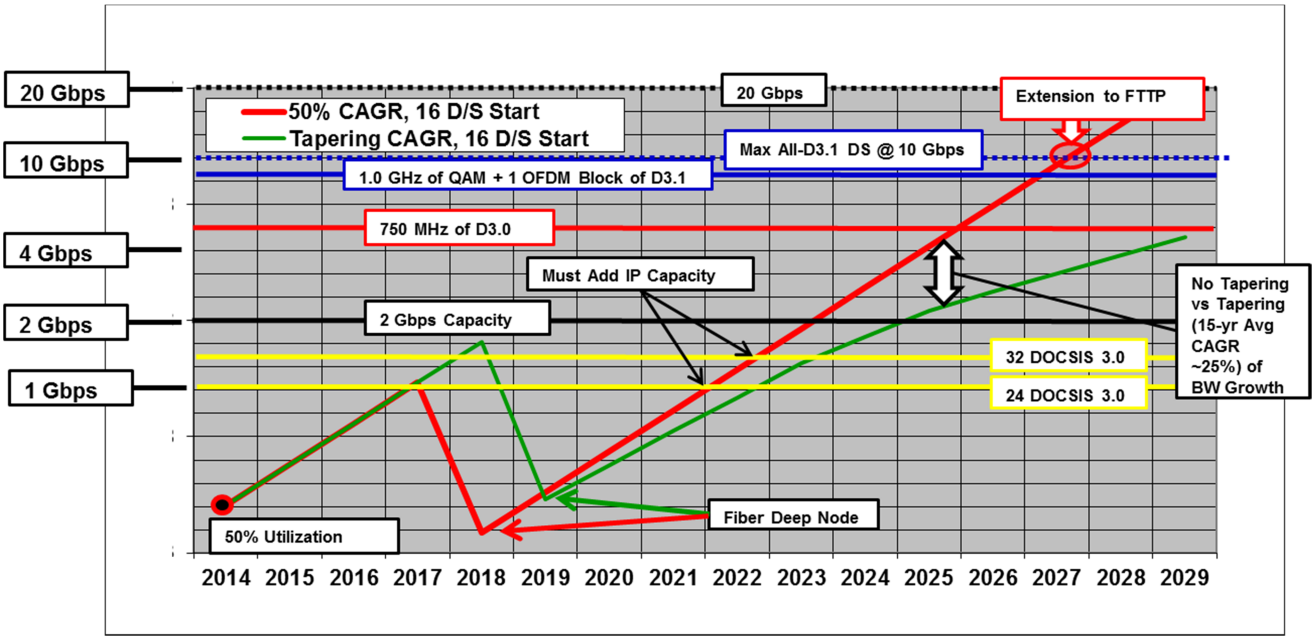


Figure 1 – Capacity Management Timeline view of HFC Long-Term Capacity [6]

Cable operators utilize this form of Capacity Management Timeline to understand the implications of growth on their network and the effects that various knobs and levers at their disposal have in keeping a healthy network capacity margin. This ensure a high quality of experience for customers.

Figure 1 is complex-looking but is in fact relatively self-explanatory when broken down into its descriptive parts. It can be considered in five understandable components that taken together tell an illustrative story:

1) The first year depicted on the chart – 2014 – is when Comcast was widely deployed with 16 downstream DOCSIS 3.0, which is approximately 600 Mbps of capacity for the IP network across a downstream service group. The utilization assumed of this allocation is 50%. This point is the black dot circled in red at the bottom left corner of the curve, and is expressed in expressed in decibels (approximately 25 dB).

2) Capacity is expressed logarithmically (dB) because traffic growth is exponential. The result is that the 50% CAGR then displays as a linear slope upwards. This makes long-term analysis very straightforward to visualize and threshold comparisons – the most important component of this type of tool – much clearer.

3) It was also in 2014 that Comcast began migrating the HFC network purposefully to a "Fiber Deep" architecture [6]. This will be detailed later in this paper, but, by definition, this means that no amplifiers exist after the HFC node. This was foreseen and proposed as the optimal MSO path for managing long term growth, for the reasons to be identified in 2009 [5].

A Fiber Deep migration will take place over time across the national footprint. Fundamentally, it is a more aggressive variant of the natural node splitting cycle that takes place today, and is based on priorities associated with relative network congestion, nature of the competition, status of existing infrastructure, nature of the regulatory environment, and several other criteria.

In Figure 1, this Fiber Deep (FD) phase happens in 2018 (red) or 2019 (green). This is the explanation for the step downward – the service group size is reduced considerably by Fiber Deep, increasing the average bandwidth available per subscriber and "resetting" the growth trajectory to this new reality. In practice, not all of this serving group reduction may take place at once (there may be HE aggregation at first), but a fundamental principle of FD is a last touch of the HFC network for as long as it is reasonable to consider "last," with any additional disaggregation – including all the way to 1:1 service group-to-node – easily applied at the HE when needed.

4) After this step, the traffic growth continues at the same slope. In the case of the green curve, some tapering of the CAGR is shown to understand the very powerful impacts if this were shown to be a trend in the future. The working assumption at Comcast is not to bet on the Internet slowing down, especially given the historical trend which has seen the aforementioned growth for roughly two decades.

5) Lastly, there are horizontal lines on Figure 2 that represent various thresholds of capacity the network can support depending on the total spectrum available, the percentage of it allocated to IP data, and the technology utilized (such as DOCSIS 3.1). When a red or green curve crosses a threshold, that architecture is no longer sufficient to support the capacity demand and new steps must be taken.

Note that capacity and data speed are not synonymous. Operators must overprovision their network capacity relative to the highest speed tier offered because the network resources are shared. After years of providing HSD service, operators understand the relationships between capacity and speed very well, at least under the existing deployment scenarios. Speed growth is predictable in the

sense of expectation, but not necessarily with respect to timing, as the phenomenon is based on more competitive market dynamics rather than customer demand.

Growth Management and Today's Toolkit

Through nearly 20 years of HSD service and the associated bandwidth growth, cable operators have managed growth effectively and efficiently. There is a simple explanation why. When HSD service was launched, capacity and speeds to homes were set by the capability of the telephone wires, which was, and still is, very limited. The cable medium to the home provided orders of magnitude more IP capacity that MSOs were able to take advantage of, which has led to cable broadband dominating the broadband access business to this day.

The tools most frequently deployed to manage the growth were then and still are the addition of new DOCSIS channels to the CATV spectrum, and the adding of HSD ports at the Headend to reduce the sharing of resources into smaller pockets of users. Ports being added in the context of the HFC plant correlates with node splitting – a fiber node that once served, say, 1000 homes, would be segmented, or the network augmented with a new node, such that roughly 500 homes were served from each, and two 500-home groups each terminated at a CMTS port. This "BAU" process has been very effective, is mature and well-understood.

Fast forward 20 years, and as MSOs do their obsessive and long-term capacity accounting, it has increasingly become clear through analysis such as Figure 1 that the once-tiny network burden that the HSD service entailed, because of this persistent, aggressive CAGR, has become a roaring beast not as routinely accommodated. The relatively comfortable 20-yr phase is evolving to a need to respond more aggressively in the network itself as the heyday of excess resources on the cable are over.

Another mathematically descriptive way to state the problem is to recognize that, while traffic has continued to grow exponentially, operators cannot realistically upgrade the network at an exponential pace. Furthermore, they would not want to do so even if it were possible because of the disruptiveness this would mean for customers. Simply put, the pace of technology change at the root of traffic growth exceeds the pace of network infrastructure change – at least physical infrastructure – every time.

Instead of incremental repetitive network adaption that is difficult to sustain, new and more impactful bets must be placed that reset the game with a new set of thresholds of capability and technology. This is precisely where strategy, vision, timing, and uncertainty collide. Network infrastructure investment can be capital intensive. Because of this, investment that has directional alignment to a broader vision is essential. And, as new technology is considered and with an objective of longer term vision alignment, the visionary promise also comes with increased uncertainty.

We elaborate on the known and deployed tools, discuss and then evaluate their role in strategy development, and identify potential next-generation disruptors.

Node Splitting

All MSOs know how to exercise this muscle. When the growth of narrowcast (VOD, HSD) traffic of a node exceeds some threshold of available capacity, beyond which experience says customer experience could be effected, the node is split. Theoretically, the node service group size is halved. In practice the more nodes are split, the more unbalanced

they are likely to become. In some cases, additional steps must be taken to balance them or risk not achieving enough capacity benefit for customers.

In any case, node splitting divides up a serving area into smaller pockets as shown in Figure 2, and each pocket afforded its own narrowcast bandwidth. In so doing, the average capacity of each user is increased – the same Mbps of capacity divided over fewer subscribers.

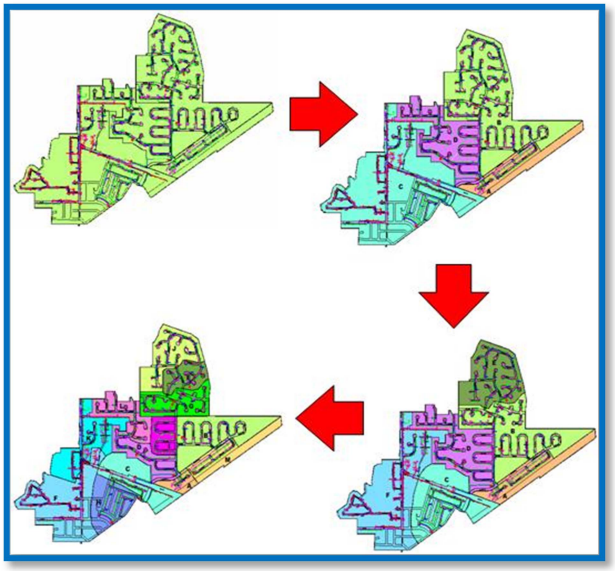


Figure 2 – Node Splitting is a Common Tool for Managing HFC Capacity [8]

Node splitting is efficient in the sense that it puts new capacity exactly where it is needed, and also because MSOs are very well-versed in this practice and can execute it quickly and with minimal customer interruption.

Fiber Deep

Fiber Deep is often referred to as "Node Splits on Steroids." The term "Fiber Deep" has different meanings for different operators. For Comcast, this is the term used for "Node + 0," whereby there are, by definition, zero RF amplifiers after the fiber optic node. Fiber Deep is largely similar to network activity MSOs have been engaged with for years – namely RF upgrades and node splits. The primary difference is that Fiber Deep is a more methodically implemented holistic architecture with specific objectives to maximize coverage and network capability.

Specifically, the many compelling advantages of Fiber Deep are:

1) Familiar HFC tools
2) Much smaller serving group size
3) More freedom of spectral allocation
4) Better end-of-line SNR and MER
5) Lower operational costs
6) Close access to direct FTTH connection for homes
7) Long-term HFC capacity runway under aggressive CAGR assumptions

Figure 3 describes the key principles associated with Fiber Deep.

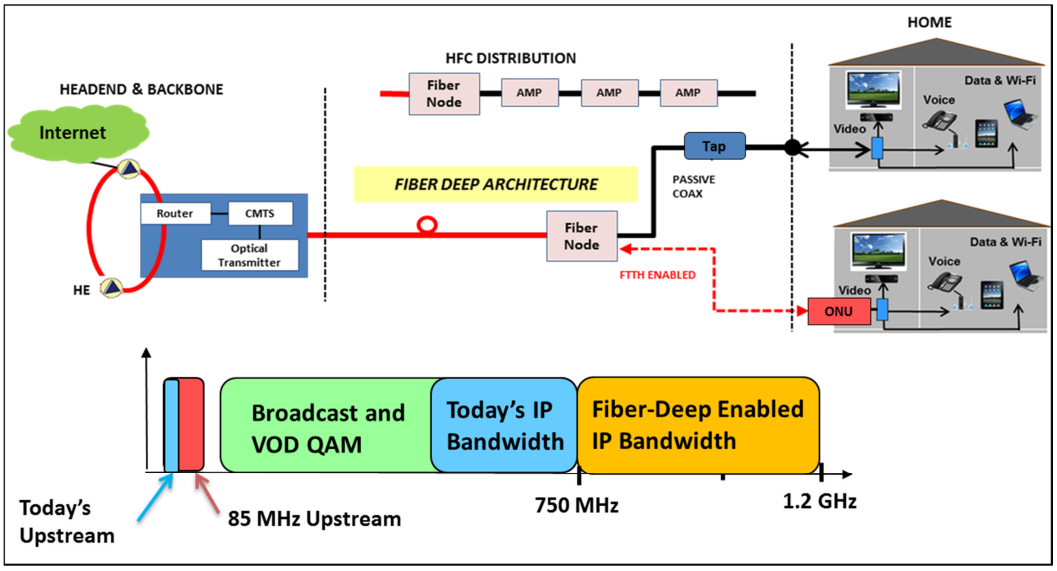


Figure 3 – Fiber Deep Architecture and Key Principles [6]

Spectrum Allocation

Operators have historically added downstream spectrum as the appetite for more video options increased. For Comcast, the transition to completely digital video (QAM) delivered 10:1 efficiencies of spectrum for SD carriage, and 2:1 or 3:1 for HD, essentially staving off the need for new spectrum since that transition took place. Those efficiencies have improved over time with performance advances and optimizations in video encoding. However, the margin in spectrum resources put in place by the use of digital only video (QAM) has nonetheless now been mostly consumed by the explosion of demand for IP data services, so the objective to add new spectrum re-emerges – both downstream and upstream.

Spectrum is a cable operator's most precious resource, so a reallocation must be very carefully thought through and justified. Tools for long term capacity must consider the ramifications to spectrum and the benefits of having it more flexibly allocated in the face of the uncertainty of new demand trends.

Also, while the allocation of spectrum must be done with support for projected CAGR for the long term in mind, it also recognize the market reality of the data speed wars. While CAGR is tied to average bandwidth per user sharing the resources, speed requires MSOs to allocate a dedicated amount of spectrum available over and above the maximum speed offering to ensure that customers will consistently experience the service speed they have signed up for.

For Fiber Deep, a proper balance of the aforementioned objectives is achieved with an 85 MHz upstream and 1.2 GHz downstream.

DOCSIS 3.1

The latest DOCSIS standard offers both the opportunity to expand spectrum beyond bandwidths that had previously been unavailable, both downstream and upstream, and to use the spectrum itself more efficiently. The latter is accomplished primarily by updating the physical layer (PHY) tools.

The combination of technologies in DOCSIS 3.1 provide the means to make the HFC network capable of approximately 10 Gbps downstream and 2 Gbps of upstream

capacity, thresholds which promise runway for long term bandwidth growth when used synergistically with others tools, in particular Fiber Deep. Besides the extension of spectrum, the key new technologies of DOCSIS 3.1 are

1) Full spectrum of multiple simultaneous blocks of up to 192 MHz wide Orthogonal Frequency Division Multiplexing (OFDM)-based QAM. DOCSIS 3.0 is single carrier QAM (SC-QAM) and has a maximum of 32 bonded 6 MHz wide channels (192 MHz), which makes 1 Gbps service speeds impractical in scale. OFDM is the go-to signal structure for virtually all modern communications systems, leveraging the adaptability and flexibility of many narrowband carriers, as shown in Figure 4, to important system advantages and robustness to impairments that can be very deleterious to SC-QAM.

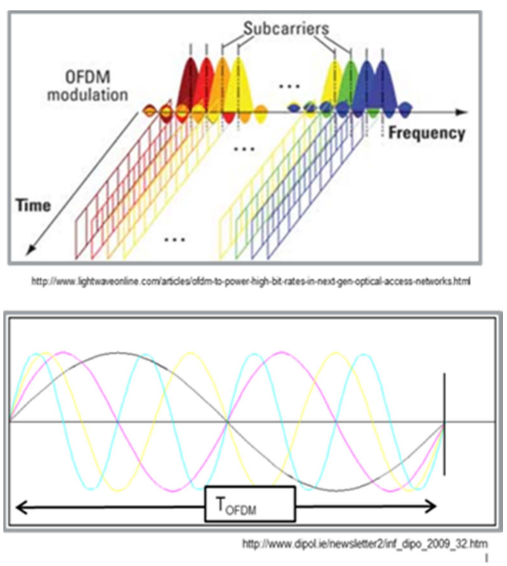


Figure 4 – OFDM as the New Signal Structure in DOCSIS 3.1 (note the underlying modulation is still QAM)

2) Updated Forward Error Correction (FEC) from Reed-Solomon based to Low Density Parity Check codes (LDPC). LDPC enables existing QAM at lower SNR than DOCSIS 3.0 by about 4 dB, or higher order QAM at the same SNR.

3) Availability of higher order QAM formats up to 4096-QAM downstream (from 256-QAM, or 50% more efficient) and 1024-QAM Upstream, (from 64-QAM, or 67% more efficient). Usability of these advanced formats takes advantage of both the FEC upgrade as well as the gradually improving network performance over time that has occurred as fiber has been pulled deeper and HFC optical technology has improved. Figure 5 compares the QAM constellation of DOCSIS 3.0 based 256-QAM and DOCSIS 3.1 based 4096-QAM.

Fiber-to-the-Premises (FTTP)

Many cable operators have been deploying FTTP for years, largely for commercial customers that rely on symmetrical data traffic needs. Residential data traffic is, and has always been, significantly asymmetric in nature in favor of the downstream. The Downstream-to-Upstream ratio has increased over the past few years due primarily to over-the-top (OTT) IP Video (i.e. Netflix, Hulu, etc). Nonetheless, as shown in Figure 6, the data speed wars continue, and FTTP options offer operators immediate paths to delivering these speeds if demanded by customers. The vast majority of customers will benefit from the simplicity of mass deployment of DOCSIS 3.1 with Gigabit speeds based on the coaxial connection to the home that already exists.

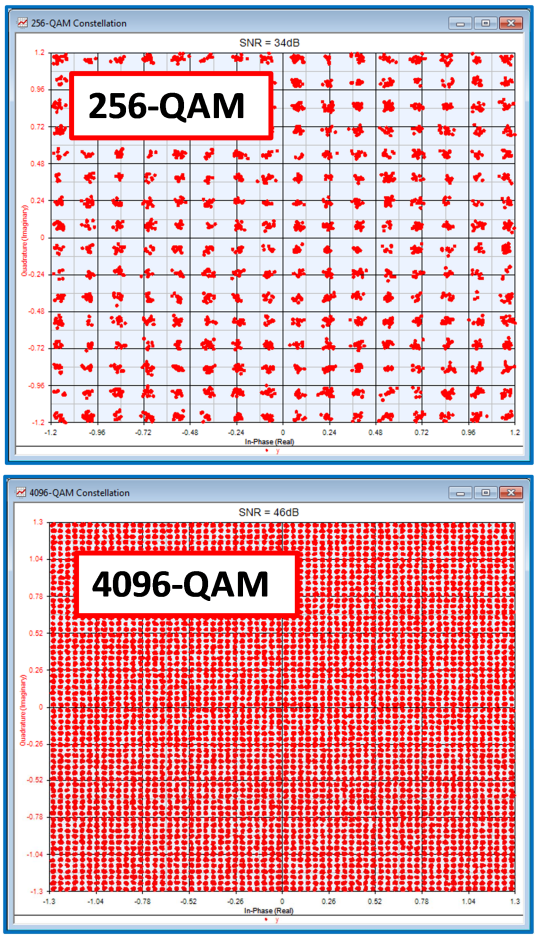


Figure 5 – DOCSIS 3.1 Bandwidth Efficiency Increase due to M-QAM Format

Figure 6 – Market Speeds Continue to Increase

Many operators have already deployed FTTP in residential footprint, and this is becoming more commonly the case for new construction as the costs for coax and fiber builds are nearly at parity. The evolution of HFC that takes advantage of the coaxial network without disturbing neighborhoods and customers will leverage many of the aforementioned tools of the previous section. For fiber systems, there are multiple options to choose from also – SCTE 174 [10], EPON, Ethernet in particular. The move to all-IP will further simplify and unify service delivery across different physical architectures. Indeed, this has always been one of the key value propositions of all-IP – any IP pipe can deliver the services, regardless of how the bits are framed, how the waveforms are generated, or what medium is used.

Lastly, as was shown in Figure 3, it is a fundamental premise of the Fiber Deep architecture that it is deployed with a "2nd putt" strategy to be FTTP capable in the future if needed. Fiber Deep (FD) provides physical access to fiber within approximately 1000 feet of all subscribers. By proactively designing a Master Architecture with a possible FTTP end-state, the optical distribution network (ODN) is deployed in a way that prepares the network for FTTP during the FD cycle. Since the deployment of new fiber is the most intensive and costly phase of new construction, it is therefore utilized to maximize long-term evolution efficiency via consideration of this end-state architecture expectation.

In general, it is a powerful strength of cable operators that they can deliver their full service suite of broadband services over coaxial or fiber last miles.

Solutions Options to the Problem Statement

Figure 1 demonstrates the fundamental problem statement of cable operators. Under an assumption of a persistently aggressive CAGR, such as the verifiable 50% per year, the point at which capacity is exhausted on the network as implemented today is moving into the long-range planning window – in the calculation of Figure 1, this is 2027.

The key specifics of this projection are a transition to a fully deployed, all-IP DOCSIS 3.1 network achieving 10 Gbps, a Fiber Deep (N+0) serving group size. Under these steps, the growth curve crosses the threshold of available capacity in 2027. While this is a long runway, the significance of this for operators and potential shifts in network investment type requires beginning the planning steps for "what's next."

Note that Figure 1 captures only the Downstream growth, as it is this aggressive CAGR dynamic that determines the ability of the network to support future capacity. There is a parallel analysis for upstream that is not shown because the driving dynamic for HFC evolution turns out to be the persistently aggressive DS growth, once the upstream relief of the 85 MHz mid-split is applied. The upstream does indeed grow – less consistently but at a measureable average CAGR – it simply turns out that without a major change in the traffic dynamics – to be wary of – it is secondary to the downstream impacts.

Now consider Figure 7. Figure 7 shows one scenario of many that were run to evaluate investment levels (in Net Present Value terms) for various migration approaches to the HFC Brownfield, including the BAU option. The scenario shown is a culmination of the optimization of a 6-month engagement process with key business, financial, and technical stakeholders, among others, at the table with their various perspectives:

1) Capacity Engineering – How much will we need and when?
2) HSD and Video Product Management – What product targets must be met for our Customers?
3) Architecture and Technology – How do we meet capacity and product needs, when, and at what cost?
4) Construction Engineering – How and how fast do we build it and at what cost?
5) Finance – What does the investment profile vs time look like for the various options?
6) Human Resources – How to achieve the necessary balance of skill sets?

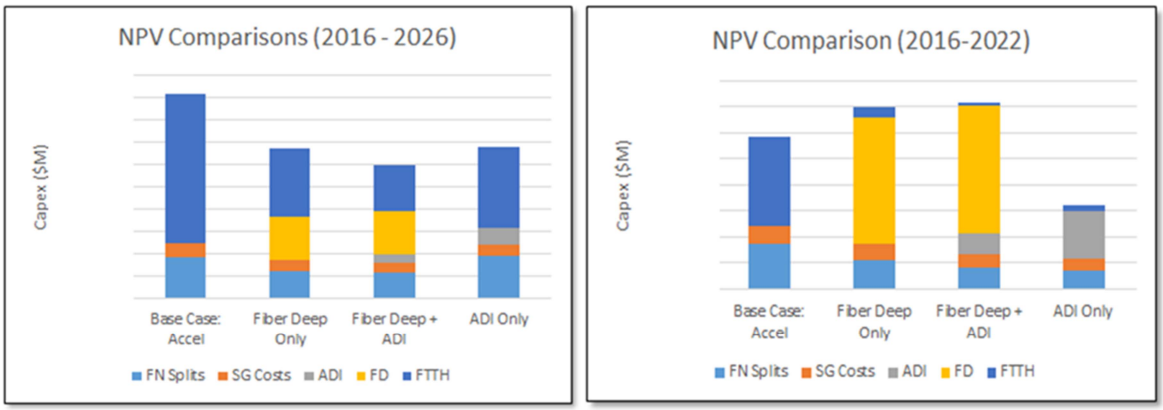


Figure 7 – Network Evolution NPV Investment Comparisons

The evolution scenarios shown in Figure 7 are:

Base Case

This represents the existing baseline level of network investment projected over 6 and 10 years. The "BAU" or Baseline foundation is node splitting on-demand at a standard threshold of capacity utilization and adding DOCSIS spectrum. Both of these are methodically planned, budgeted, and implemented as standard practice today.

Included in the DOCSIS spectrum are existing assumptions around the introduction and growth in allocation for DOCSIS 3.1 "channels," as a path to an all-IP network. All remaining channel slots (we use 6 MHz equivalent slots even though this channel basis is eliminated in D3.1) are for 6 MHz QAM channels for broadcast or narrowcast video, and a small number of slots carved out for overhead functions, OOB, etc. These QAM channel slots are gradually reduced over time with technology, shrinking serving groups, and transition to IP Video delivery.

Notably for "Base Case," while the service group size is reduced because of the node split, the basic HFC architecture is unchanged. Fiber is not necessarily any deeper, the same cascade of amplifiers generally exists, and the network spectrum allocation has not changed either downstream or upstream. Using 750 MHz as the available downstream bandwidth, approximately 108 total channels would be available in an all-IP end state (2025). This is less than would be available using a more aggressive drop-in upgrade (Active Drop-in or "ADI", to be described) or as compared to FD which extends the downstream to 1.2 GHz.

Finally, "base case" assumes that there is some threshold of HHP (several scenarios run to optimize) below which it does not make sense to split further because of the inefficiency. Instead, a proposed next step is to offload a portion of the "top talker" traffic to FTTP over nearby existing deployed fiber and/or via WDM and reduce the load on the HFC network for the majority of customers.

Fiber Deep Only

This category is as previously detailed – the network is taken to N+0, also known as passive coax. Fiber Deep by definition also includes an 85 MHz Upstream band edge and 1.2 GHz Downstream band edge.

An example scenario for DOCSIS channel growth transition (and thus QAM reclamation) is shown in Table 1 for FD, which has the most available spectrum of the HFC-centric options under consideration.

A key dynamic in the transition to DOCSIS spectrum is the practical pace at which QAM can be reclaimed as it relates to existing deployed QAM CPE. This requires extensive logistical planning and coordination, very similar to what was executed in the transition from analog video to all-QAM using the Digital Terminal Adaptors (DTAs).

ADI Only

The acronym "ADI" stands for Active Drop-In, which is also known as "plant upgrade." Nodes (if necessary) and amplifiers are upgraded to 1 GHz, thereby adding significant new bandwidth, in particular for 750 MHz systems. ADI also includes 85 MHz upstream, as expansion of the return is an important capacity strategy to achieve the long-term health of the network. Note that the upstream is not changed in the BAU base case – a significant disadvantage to that approach. The effect of this is to drive more splitting to handle congested upstream even at the moderated CAGR compared to Downstream.

	DOCSIS 3.0	DOCSIS 3.1
2016	24	16
2017	24	16
2018	24	40
2019	24	40
2020	16	64
2021	16	80
2022	8	104
2023	8	182
2024	8	182
2025	0	182

Table 1 – Example DOCSIS Growth vs Time for All-IP (Fiber Deep case only)

Because of the expansion of the downstream spectrum, about 140 channels would be available by installing actives that are 1 GHz – about 30% more spectrum. Note that some of the spectrum added is converted to upstream, where spectrum increases by over 200%. Also, it is anticipated that the network may not quite reach 1 GHz everywhere, at least not with a flat frequency response, because spacings of amplifiers may not support it without some roll-off attenuation.

ADI is expected to provide a fast, cost effective capacity relief valve to complement a Fiber Deep transition strategy. The nature of the Fiber Deep migration is that it takes time to complete in large scale as it system-wide (vs on-demand by node), and involves design optimization, the addition of fiber, and facilities work. There are also practical resource and logistical limitations to simultaneous construction projects

The strategy limitations of ADI are:

1) Node splits are still required, but are deferred due to the spectrum added both Downstream and Upstream
2) Loss of spectrum as compared to Fiber Deep
3) Lower RF performance than Fiber Deep (but the same as today), leading to lower D3.1 capacity
4) Not positioning the fiber any deeper to enable more potential FTTP customers now or in possible future transition
5) As we shall see – and perhaps falling into the category we have alluded to of a plan put in place that becomes superseded by the "next" plan as technology advances – not consistent

with the fundamentals of a full-duplex DOCSIS architecture

The category "*Fiber Deep + ADI*" is self-explanatory, meaning deploy some of both. The scenario depicted in Figure 7 assumes roughly a mix of both of the above, in approximately equal amounts.

<u>As of May 2016, a Winning Architecture Evolution Strategy is……</u>

What is immediately interesting in the scenario comparison depicted in Figure 7 is that in the very long-term view – in this case the 10-year view – the most costly path is the BAU path of attempting to node split your way through the growth. This is due the nature of the traffic growth (exponential), and that this approach does not address the spectral limitations of BAU, nor simple access to nearby fiber for FTTP offload. Note that there is in fact a large portion of expenditure allocated to FTTP. In the analysis, this takes place when s node sizes get below a certain (adjustable in the model) homes-passed threshold after which it becomes impractical to efficiently split further. Note also that the size of the bar for "FTTP" in Figure 7 is not a proxy for households passed (HHP), since FTTP in the HFC brownfield is costlier on a per-HHP basis compared to other tools shown and different depending on whether it is a Fiber Deep architecture or not, such as this case.

Just as interestingly, in the 6-yr view, this is not the case even with an FTTP transition underway. This 6-yr and 10-yr difference is insightful in what it says about the natural life of the as-is network when a "Base Case" approach is taken. Capacity constraints are within the planning window, and this is what drives the need to consider alternative strategies and assess their implications.

Note that what appears as most cost effective approach in the 6-yr view is "ADI" – suggesting a plant upgrade cycle is preferred. ADI addresses the priority capacity needs of Downstream and Upstream spectrum such that node splits and offload to FTTH are deferred. However, as the 10-yr view shows, node split re-emergence takes place and the cycle begins again. Node splits again become prominent in trying and keep up with exponential traffic growth, and FTTH offload begins as the deferment period ends and node size thresholds come back into view.

It is clear from the long-term view that the "winner" under the assumptions of the scenario in Figure 7 is a mix of Fiber Deep and ADI. Fiber Deep offers the order-of-magnitude sized capacity upgrade necessary to paint a long-term lifespan view of the HFC network in one methodical and well-understood step. In addition, it positions the fiber with close proximity to affect FTTP offload if necessary, so this "2nd putt" is then most efficient and cost effective. Perhaps most importantly, it defers FTTP offload as evidenced by the smallest of FTTP investment bar colors in the 10-yr view shown in Figure 7.

A long runway of lifespan with which to assess trends in applications, services, and technology is a comforting situation. It is not possible to predict where these trends lead ten years hence, nor would it be fruitful to try and predict these. But, we can use historical experience as a meaningful guide to what's next. So, while adjustments to the plan can be assured, the probability of an urgent, major shift in plan is lowered with such runway ahead.

<u>How Long Does a Long Term Plan Last?</u>

A dedicated team worked many scenarios across a range of assumptions, projections of

demand, and product offerings to arrive at a go-forward near term and long-term strategy. There is no illusion that 10-yr plan remains unchanged for the duration of the migration. Acceleration of the pace of technology change is nothing new. New technologies will be developed that demand consideration because of the performance, savings, simplicity, or all-of-the-above that they bring.

With this in mind, unsurprisingly, just as the next generation access evolution journey begins, a wave of new technology possibilities are emerging. This is already happening a mere 18 months into the plan. These advances are certain to have an impact on the journey, suggesting a veering off of the defined path for the promise of some of those very compelling benefit mentioned above.

We discuss several of the most appealing opportunities, their effects, and implications below.

Distributed Access Architectures

A Distributed Access Architecture (DAA) breaks the CCAP function into two pieces and distributes a portion to the node. The functional portion distributed can vary. While there are a range of possibilities, the debate has mostly narrowed to two architectures for cable "DAA" – Remote PHY and Remote MACPHY. Regardless of which architecture wins the marketplace, DAA delivers some of the most powerful advantages *ever* deployed into the HFC network, primarily through the use of digital Ethernet optics. This relatively simple change creates a necessary reconsideration for the plan. Every node deployed today is implemented with AM optics. The Fiber Deep architecture defines a node with specific requirements for spectrum plan, output level, segmentability, all based on HFC optics – downstream AM optical receivers and upstream digital returns. However, just as the Fiber Deep plan is in the midst of implementation, along comes the opportunity to drastically improve upon it with DAA.

The major effects of a DAA implementation are as follows

1) Wavelength multiplier of DWDM vs the limitations of WDM for AM optics

With a migration to fiber deep, a significant number of new nodes are installed, as amplifiers are eliminated. Note that a "Fiber Deep" design is not the same as a "Fiber-to-the-Last-Active" (FTLA) approach, whereby existing amplifier stations are converted to nodes. Fiber Deep optimizes the location of the new nodes for HHP to optimize economics via reach. The addition of many new nodes demands the use of WDM to maximize the use of the existing fiber infrastructure. The use of digital optics means that there is a path to 80 DWDM wavelengths/nodes, for example, aggregated on a single fiber (or more). Standard HFC AM optics are restricted typically to 16 wavelengths to manage the nonlinear effects on MER for common HFC optical links lengths. More are possible, but there are performance and distance dependencies to manage that do not exist with digital optics. This means a more cost-effective deployment and a much lower probability of installing new trunk fiber to support the nodes.

2) Reach of digital optics vs AM optics

The longer the optical link from Headend transmitter to node, the fewer wavelengths that can be used at a fixed end-of-line (EOL) MER. Digital optics completely eliminates this dependency in practice for common HFC optical links. In fact, the CCAP core with digital optical outputs could be moved further back into the core if Headend consolidation is an objective.

The above aspects of reach and wavelength multiplication taken together massively simplify the HFC network and add never-before-seen flexibility that can pay large economic dividends as more fiber is pulled deeper into the network, which is exactly what fiber deep does.

3) SNR and MER performance improvement of eliminating the DS and US AM HFC optics

DOCSIS 3.1 enables higher order modulation profiles, increasing the bandwidth efficiency up to 50% over DOCSIS 3.0 n the Downstream and 67% in the Upstream. It also allows the QAM profiles to be varied and optimized to the network performance. In standard HFC and especially in Fiber Deep using HFC-style nodes based on AM optics, the EOL MER is dominated by the performance of the AM optics. The CCAP device transmits a very high MER (set by the Downstream RF Interface Specifications, or DRFI) that is standardized by CableLabs specifications. The DRFI requires higher fidelity transmission for DOCSIS 3.1 ports then for DOCSIS 3.0 ports.

In standard HFC, the DRFI spec sets a high level of fidelity to the optical transmitter – 43 dB minimum for DOCSIS 3.0 and 43 dB up to 48 dB for DOCSIS 3.1, depending on frequency band. This higher fidelity is set to support the higher order DOCSIS 3.1 QAM profiles.

A Headend optical transmitter then acts on the signal and degrades this, and a common requirement is to exit the node at a minimum MER of 38 dB. There is therefore significant fidelity loss in the AM optics of 5-10 dB. The SNR/MER that exists at the node degrades slightly more as it is passed through amplifiers. In Fiber Deep, this degradation does not take place because there are no amplifiers.

With DAA, the link to the node eliminates the AM optics, and eliminates the SNR/MER degradation caused by this link. Instead, the DRFI requirement is met at the node equivalent CCAP port, gaining back the lost fidelity of the AM optics nearly completely (the node has RF actives between CCAP port and output). This maximizes the capacity possible for D3.1 output signals.

The most capable HFC network that can be deployed from a network capacity, operational robustness, and consistency of performance perspective is an N+0 Fiber Deep architecture implemented with a DAA approach, and a point-of-entry home gateway architecture.

4) Space, power, cooling efficiencies in the Hubs and Headends

It seems obvious that moving some of the CCAP functionality into the plant (i.e. the node), leaves less behind in the Hub or Headend to power, cool, and consume space. This is, in fact, the case. However, the benefits are perhaps more outsized than intuitively obvious. The density of RF connectors and supporting electronics and isolation requirements on today's CCAPs tend to set the density of these chassis. With the RF port of the CCAP distributed into the plant, the density of the CCAP core can now be redefined since the density of optical connectors is much higher.

Secondly, the use of DAA drives an accelerated migration to CCAP convergence for video, eliminating Edge QAM boxes and RF combining networks.

The benefits to facilities of DAA cannot be understated in a migration to Fiber Deep, which installs many new nodes for every

single existing node. A degree of aggregation occurs initially in Fiber Deep to avoid excess investment in capacity before it is needed in practice. Nonetheless there is a significant equipment burden placed on facilities to activate a Fiber Deep network.

5) Alignment to strategic NFV and SDN evolution across multiple last-mile coaxial and fiber access technologies

This will be discussed in the next section.

In summary, the significance of DOCSIS 3.1 and Fiber Deep to a long term evolution strategy are magnified by the value of bringing DAA systems to market given the wide ranging, incredibly powerful benefits identified above.

NFV / SDN

Cable operator's networks are a collection of purpose-built video, voice, and data platforms of integrated hardware and software. This has been a successful formula for over 30 years of service introduction including digital video, HD video, voice, data, and VOD.

As detailed throughout this paper, developing any long range plan must deal with the tsunami of persistently aggressive traffic growth that threatens to exhaust network capacity without some of the tools described herein. Another way to think about this is that the "linear" scale of new hardware development does not keep pace with exponential traffic increases now that traditional HFCs resources are fully deployed or nearly so. This extends beyond just network capacity, but also into service velocity enabled by the transition to all-IP, cloud, and software-based systems.

Network Function Virtualization (NFV) and Software Defined Networking (SDN) are enablers of cost effective, efficient, exponential network and service change velocity. The deployment of Fiber Deep and DOCSIS 3.1 to support the continued growth in traffic traditionally means the addition of more and more RF ports and CCAP line cards.

Bringing this lofty concept down to operational practice, in a DAA implementation, these RF ports are distributed into the field serving smaller groups of subscribers. However, supporting line cards of the CCAP core would still necessary with the disaggregation of an existing Integrated CCAP (I-CCAP). Figure 8 conceptualizes this logical architecture [9].

In a virtualized implementation, however, this purpose-built hardware core, designed to be tightly coupled to the output RF interfaces, is instead implemented in Commercial Off-the-Shelf (COTS) server hardware. This is made possible simply through the continuance of Moore's Law. The compute power and resources needed is available in standard processors today, allowing CCAP functions to be executed in such platforms and be software-based.

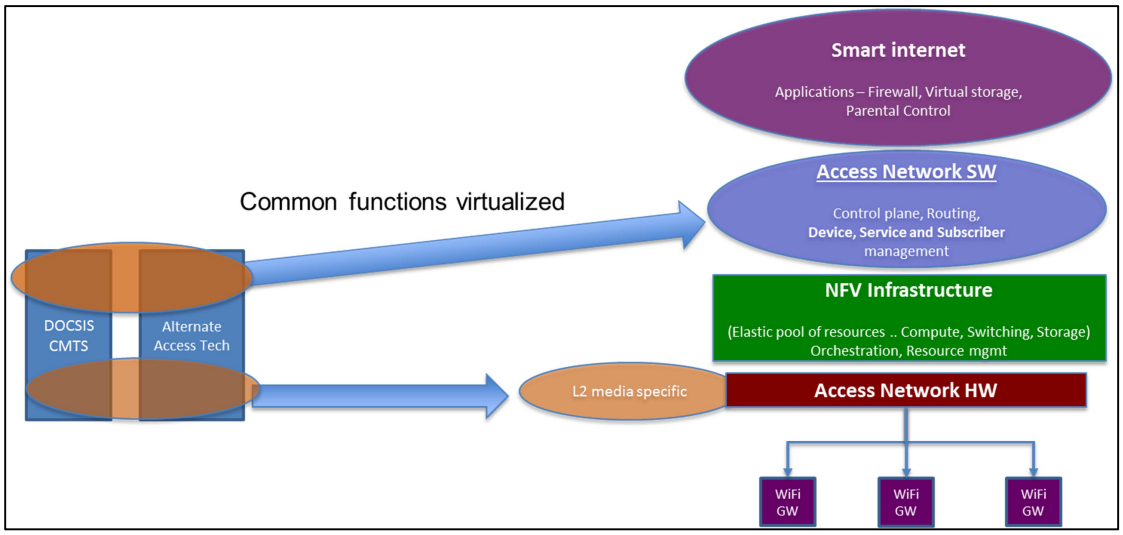


Figure 8 – Virtualization of Access Edge Platforms [9]

The potential to create SW-based CMTS on standard server platforms has enormous implications for cost, facilities, and service velocity – cheaper, faster, smaller, and more cost effective. This virtualized path applies in the case of I-CCAP and to DAA. A virtualized I-CCAP system is actually a form of DAA, since a COTS server-based CCAP core is never going to have RF outputs. And, similar to the DAA topic previously, and again representing the natural conflict of the acceleration of technology and limitations of practical infrastructure cycles, the NFV/SDN development activity will lead to a significantly superior architecture within a few years of deploying DAA systems. Initial DAA systems being invested in and deployed into the field will likely include some that are based on purpose-built hardware cores, such as today's CCAPs.

Also just as in DAA, the advantages of a migration to virtualized edge platforms are so compelling that figuring out this transition from integrated architectures to distributed architectures to virtualized *and* distributed is essential.

Full Duplex (FDX) DOCSIS

Whereas DOCSIS 3.1 pushed the network to 10 Gbps / 2 Gbps, a "Full Duplex" version of DOCSIS, currently in feasibility analysis phase, aims even higher. Taking advantage of another key tool – echo cancellation – borrowed from the telco world, FDX DOCSIS targets up to 10/10 Gbps of symmetric or near-symmetric capacity. Figure 9 demonstrates the fundamental concept behind full duplex DOCSIS.

FDX DOCSIS comes with some network expectations of various levels of dependency. Among them, in decreasing level of essentialness, are:

1) Fiber Deep, N+0
2) Distributed Access Architecture
3) Transition to all-IP

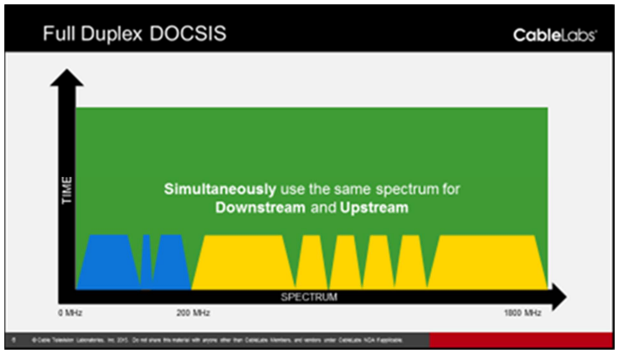


Figure 9 – Full-Duplex DOCSIS Fundamentals [2]

The concept of full duplex for DOCSIS borrows heavily from technology developed in other wireline and wireless industries, but comes with its own complications unique to the cable architecture that are still being evaluated. However, the potential benefits are so valuable – symmetrical capacity and speed regardless of whether customers are connected to a coaxial or fiber drop – this technology could also cause an adjustment to the Fiber Deep plan to not just migrate to DAA, but DAA upgradeable to full duplex.

Wireless

The vast majority of consumers receive their data services wirelessly in the home, and by far prefer they this be the case. Thus, the capabilities of WiFi and where this technology is headed are critical to customer Quality of Experience (QoE). Recognizing the value of WiFi, Comcast has deployed millions of WiFi hotspots that provide customers access to their HSD services when they are outside the home.

While WiFi is an extremely powerful and critical component of a service provider's end-to-end architecture, it is not ubiquitous and is arguably the 2nd most valuable wireless link to the end consumer – their mobile connectivity being even more essential to daily life.

Cable operators have flirted with the "quad-play" for many years. However, with the transformation from mobile phone to smart phone, the role of these devices for media consumption and broadband access has never been higher and continues to increase. As such, renewed inspiration to engage between cable and mobile operators exists.

Today, as 5G comes into view, it comes with promises of tremendous speeds and capacities, but also some critical RF, Millimeter Wave, and home architecture dependencies to achieve its loftiest objectives. One of these dependencies is accessibility to many accessible fiber connections and a widely distributed powering grid. In this way, it is quite synergistic to the infrastructure being built as markets deploy the Fiber Deep architecture. As such, we again come across a long-term plan in motion being confronted with some exciting new possibilities to consider.

The same logic applies to various "Internet of Things" (IoT) municipal and enterprise applications. The "Smart Cities" initiative is based on ubiquitous coverage of low power sensors of various types with RF connectivity to wireline (or wireless) backhaul points.

Thus, as a Fiber Deep architecture is being designed and with a consideration for possible wireless integration, it makes sense to ask: What module types should a modular Fiber Deep node support? What other access point types (in addition to WiFi) might need to be powered and enabled on the strand?

More complex still, in particular in the case of 5G applications, is to go beyond the adaption to a wireless edge on a Fiber Deep architecture, but also to think about the processing core. A value of virtualized is COTS platforms with application-specific software and common CPU resource pool. The platform itself can thus be simultaneously supporting DOCSIS platform, PON, and wireless cores as well. Each access technology having their own farm of server silos seems to be at odds with basic efficiency and flexibility arguments for NFV and SDN. This is where open interfaces and standards offer tremendous value by allowing differences in service provider architectures

delivering common services across them to be integrated efficiently.

Somehow, this long-term vision must be coordinated with the already underway plan, including the nearer term path adjustments described above.

Summary

Service providers of all kinds are grappling with the same problem statement – a consumer bandwidth explosion that shows no signs of settling down. Capacity management using common tools of the trade that have been so successful for so long are being rapidly consumed, and the next round of network evolution is now at hand with many new options on the table.

Traffic growth acceleration is easily quantifiable, but other aspects of service needs and new technology that effect decisions on investment are not as predictable and are moving faster and faster, with shorter and shorter life cycles. It is the architect's job to develop a plan to support the fact that service and technology change outpace the ability to adapt the network at each turn. Bigger, more impactful, longer term bets must be placed using the most information that can be gathered or projected over the expected lifecycle of capital investments such as network infrastructure.

The goal of this paper was to describe some of the processes, analysis, and personnel that one operator used to develop a model and determine a possible path forward in this challenging environment.

References

[1] Chapman, John, Mike Emmendorfer, Dr. Robert Howald, Shaul Shulman, *Mission Is Possible: An Evolutionary Approach to Gigabit-Class DOCSIS*, Cable Show Spring Technical Forum, Boston, MA, May 21-23, 2012.

[2] Hamzeh, Belal, CableLabs Winter Conference Innovation Seminar, Feb 8-12, 2016, Orlando, FL

[3] Howald, Dr. Robert L, *Boundaries of Consumption for the Infinite Content World*, SCTE Cable-Tec Expo, New Orleans, LA, October 20-22, 2010.

[4] Howald, Dr. Robert L, *Breathing New Lifespan into HFC: Tools, Techniques, and Optimizations*, Cable Show Spring Technical Forum, Washington, DC, June 10-12, 2013.

[5] Howald, Dr. Robert, "Fueling the Coaxial Last Mile," SCTE Conference on Emerging Technologies, Washington, DC, April 3, 2009.

[6] Howald, Dr. Robert, "Growth Architectures: Built to Last, Built to Launch," Cable Show Spring Technical Forum, Los Angeles, CA, Apr 29-May 1, 2014.

[7] Howald, Dr. Robert L, *Looking to the Future: Service Growth, HFC Capacity, and Network Migration*, SCTE Cable-Tec Expo Capacity Management Seminar, sponsored by the Society for Cable Telecommunications Engineers (SCTE), Atlanta, Ga, November 14, 2011.

[8] Howald, Dr. Robert L and Robert Thompson, Tomorrowland: Customers, Servcies, Networks, SCTE Cable-Tec Expo, Denver, CO, Sepember 22-25, 2014.

[9] Nandiraju, Nagesh, NFV and SDN – A Comcast Perspective, Open Networking Summit, March 14-17, Santa Clara, CA.

[10] SCTE 174 2010, "Radio Frequency Over Glass Fiber-to-the-Home Specification."

The impact of Remote PHY on Cable Service Convergence.

Pawel Sowinski
Cisco Systems Inc.

Abstract

A Converged Cable Access Platform (CCAP), defined only a few years ago, was intended to enable efficiencies such as higher channel densities and deliver operational savings such as a simplified RF splitting-combining network and unified network management. Convergence in delivery of MPEG video and DOCSIS services was the main driving idea behind CCAP.

Today, CCAP convergence has become a mainstream reality as multiple vendors provide fully featured integrated CCAP systems supporting DOCSIS and MPEG video. Next, the industry will be re-focusing its development efforts on a new challenge – the Remote PHY (R-PHY) – an architecture where the PHY components can be separated from service delivery engines.

The R-PHY architecture was conceived with considerable flexibility. It enables numerous deployment scenarios including pairing of Remote PHY Devices (RPDs) with one or multiple, physical or virtualized core service delivery systems. In this environment, the traditional approach to CCAP convergence must be a reexamined.

The paper examines the technical and operational aspects of CCAP convergence starting with the analysis of the rationale for I-CCAP convergence. The authors explain how the original rationale can be optimally fulfilled in R-PHY architecture.

The paper reviews typical I-CCAP deployment scheme, new deployment scenarios enabled by R-PHY, R-PHY operational impact on service scaling, high availability and OSS systems. Finally, the paper provides recommendations for CCAP and post-CCAP planning and discusses which deployment scenarios are best suited to converged or diverged service delivery.

In summary, the paper debates the impact of R-PHY architecture, how the R-PHY technology establishes new trends and redefines the traditional meaning of CCAP convergence, and suggests the tactical approaches that best serve the needs of cable operators and their customers.

Introduction

According to the recent study, Cisco VNI [4] reports that Global IP traffic has increased more than fivefold in the past 5 years, and predicts a nearly threefold increase over the next 5 years. Overall, IP traffic will grow at a compound annual growth rate (CAGR) of 23 percent from 2014 to 2019.

Over the last several years, the architecture of cable access networks has undergone a remarkable transformation in order to keep up with ever increasing bandwidth growth and to better serve the needs of Cable Operators and their customers. The industry capitalized on innovation in silicon integration and signal processing, thus translating Moore's Law into concrete technology milestones: CCAP, DOCSIS 3.1 and most recently Deep Fiber Architecture with Remote PHY.

Before CCAP

Before we dive into the topic of cable service convergence let us briefly review the traditional cable access network before the introduction of CCAP. A representative architecture of access network before CCAP is shown on Figure 1.

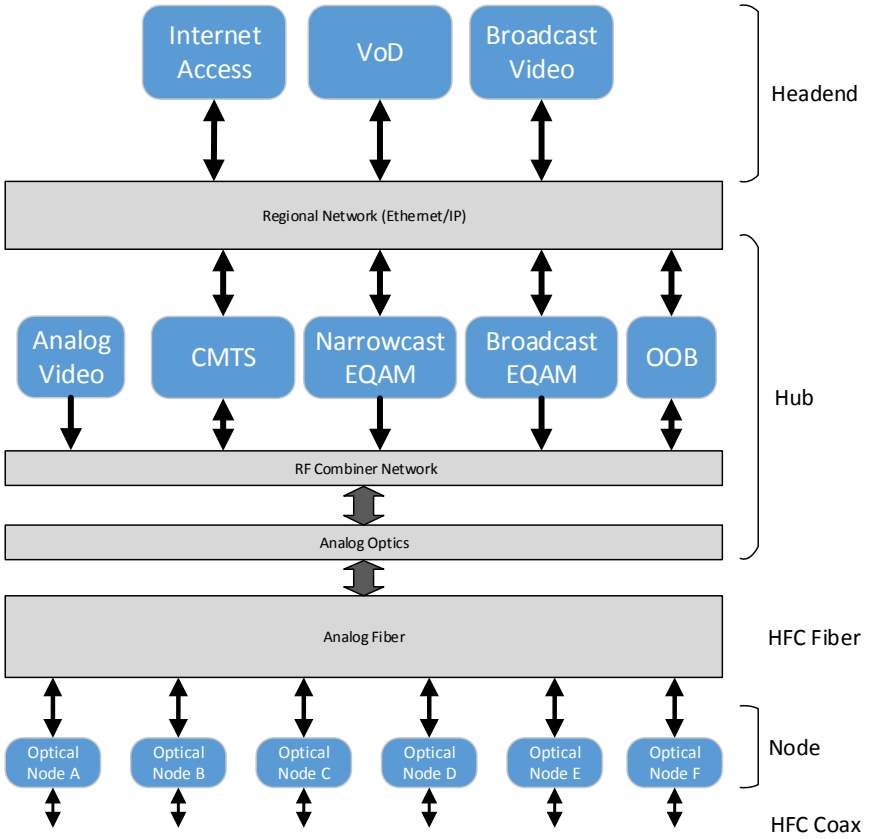


Figure 1 - Traditional Cable Access Network.

In traditional cable access networks, the service delivery platforms, namely CMTSs and EQAMs, are installed in the distribution hub at the edge of the HFC plant. These platforms directly interface to the HFC network because, at the physical layer, they convert digital signals into analog RF spectrum. In the downstream direction, the physical layer along with the lower part of the MAC layer (sometimes referred to as the Physical Media Dependent (PMD) convergence sublayer) are largely identical between MPEG video and DOCSIS 3.0 channels. The functional overlap ends there; EQAMs and CMTSs incorporate a multiplicity of higher level functions which are generally dissimilar between these platforms.

CCAP

At the start of the decade, Cable Operators came to realization that the bandwidth demand progresses along the exponential curve with seemingly no end in sight. A new approach was needed to affectively address accelerated bandwidth needs in the foreseeable future. A Converged Cable Access Platform (CCAP) emerged from industry initiatives as the result of combining separate efforts led by Comcast, Time Warner and other Cable Operators. The adjective "Converged" headlined each of these efforts (CMAP - Converged Multiple Access Platform, CESAR - Converged Edge Services Access Router) before they were merged under a single nameplate "CCAP". The notion of cable convergence associated with CCAP was generally understood to signify the integration of DOCSIS and MPEG video services into a single delivery platform with combined RF output at very large scale. As its predecessors, CCAP was designed for operation in the distribution hubs.

In this document, we'll start referring to the original CCAP as the Integrated CCAP or I-CCAP in order to distinguish from the CCAP-Core which is a CCAP derivative concept introduced with Remote PHY architecture and described further.

The I-CCAP requirements extended far beyond simply repackaging CMTS and EQAM technology into a single device. In addition to combining video and data, I-CCAP was envisioned to deliver multiple benefits over traditional cable access equipment. Let us reiterate a number of these benefits which may be relevant to our discussion:

- Efficient implementation of QAM channel generation allowing flexible configuration of MPEG Video and DOCSIS service groups. Each QAM resource in I-CCAP can be configured to operate as a DOCSIS channel or a video channel.
- Reduced power and space requirements.
- Higher port and channel densities comparing to legacy EQAMs and CMTSs.
- Eliminated or simplified RF splitting-combining network due to common RF output for DOCSIS, MPEG video and OOB channels. This benefit was materialized only partially. I-CCAP implementations have not yet completely integrated OOB functionality.
- Standardized modularity and serviceability requirements.
- N+1 redundancy and non-service-impacting failure repair requirements.
- Operational savings via standardized and unified network management interfaces.

Since I-CCAP leveraged existing cable technologies, it provided a very good fit into existing cable access network architecture. Thus, the transition to I-CCAP did not require fundamental changes to the access network around it. This is illustrated on Figure 2.

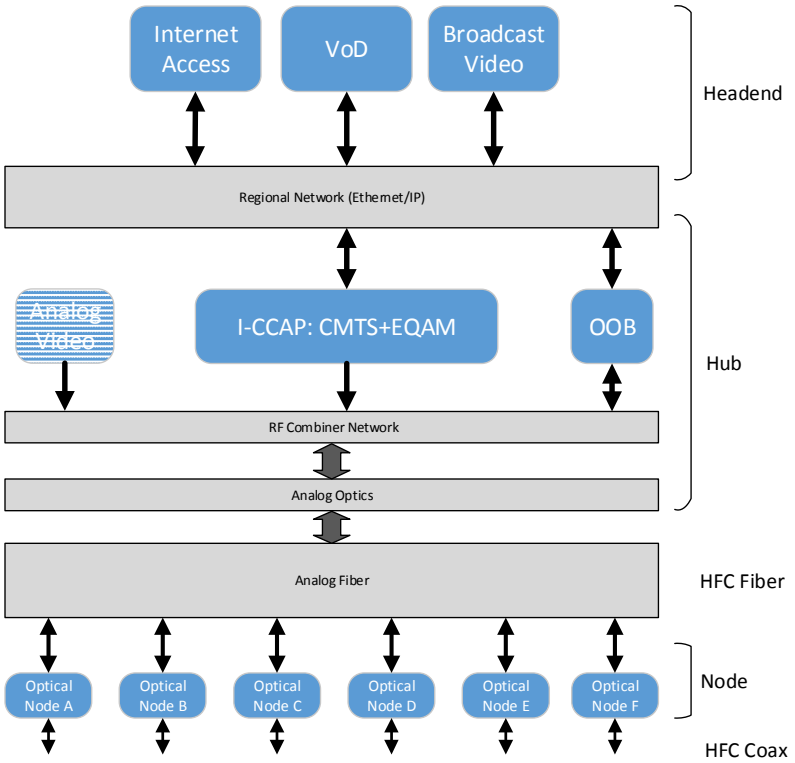


Figure 2 - Typical Access Network with CCAP

As mentioned earlier, I-CCAP was defined as a rack mounted, modular chassis which accommodated a number of line cards, including:

- **SRE Line Card**, which incorporates the management module, the forwarding packet processor and backhaul network interfaces.
- **RF Line Card**, which incorporates upstream or downstream RF ports or a combination of both.
- **EPON Line Card**, which is out of scope of this paper.

The key technology enabling convergence of data and video was the invention of cost effective full-spectrum downstream modulators. A number of technical constraints dictate that the PHY components (DS PHY modulators and upstream demodulators) are implemented on the I-CCAP RF Line cards along with MAC processing resources. This is illustrated by an example of data plane components of the I-CCAP RF line card, shown on Figure 3. For clarity, the diagram includes only one DS and one US port. I-CCAP RF line cards typically incorporate 8-32 RF ports.

The diagram presents logically separated MAC subsystems for DOCSIS and MPEG video because their functional requirements are different. The actual implementation of MAC subsystems can be based of common hardware components such as FPGAs or ASICs. While there are many design options, the author believes that the presented example reflects a typical I-CCAP RF Line card implementation with US and DS functions.

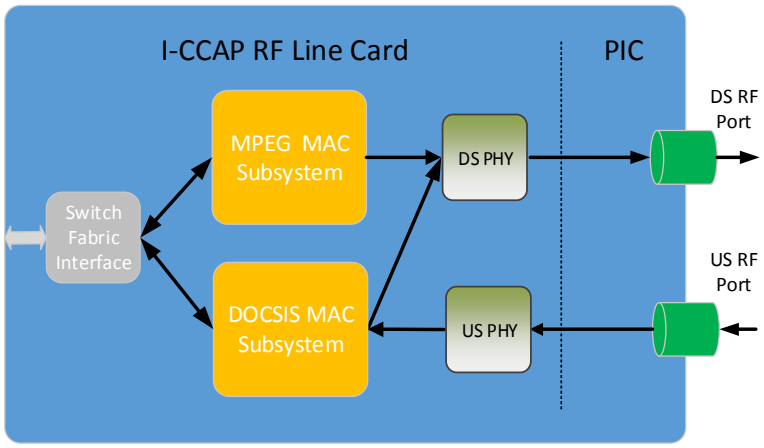


Figure 3 - Example of I-CCAP RF Line Card Components

The collocation of MAC and PHY processing resources on common H/W components results in two of issues.

- Availability. CCAP improves the overall service reliability as non-DOCSIS services get an upgraded N+1 HA architecture, and that there is an overall decrease in chassis, cables, and power supplies. The implementation of multiple services on shared hardware modules increases the system complexity and may result in a decrease of the overall reliability of the CCAP system. Operators have been wary of a situation where a failure such as a hardware or software fault within a subsystem dedicated to one service can interfere with a subsystem dedicated to another service. Unless the cable operator deploys I-CCAP with N+1 redundancy, software or firmware failures or upgrades may affect both DOCSIS and Video services simultaneously.

- Fixed scaling for MAC and PHY. I-CCAP RF line card is the main element of modularity for I-CCAP. The service capacity of I-CCAP can be increased or decreased by the addition or by the removal of RF line cards. Collocation of PHY and MAC results imposes the need to scale them in the same increments. This is undesired, especially considering that broadcast video lineup occupies a significant portion of the over bandwidth. I-CCAP implementations partially mitigate this issue by supporting an additional level of modularity - the removable PHY modules.

Remote PHY

With DOCSIS 3.1 development well under progress, the industry has shifted its focus to a new challenge, Remote PHY (R-PHY). R-PHY is the next step in the evolution of DOCSIS and video service delivery. It is a progression of the current CCAP architecture in which the PHY components are moved from the CCAP platform into a separate Remote PHY Device (RPD). The RPD is connected to the CCAP-Core (CCAP minus the RF PHY functions) by an IP network.

The combination of the CCAP-Core and the RPD provides the functional equivalent of the integrated CCAP. Essentially, R-PHY takes the digital interface to the PHY component from the circuit board in the CCAP and extends it over the IP network to the RPD using pseudowire technology as shown on Figure 4. From a protocol perspective, the MAC layer and the layers above MAC are implemented within the CCAP-Core while the physical layer and the PMD Convergence Layer are implemented in the RPD. The upstream and downstream channel resources in the MAC (MAC-level channel resources) are associated with channel resources in the PHY (PHY-level channel resources).

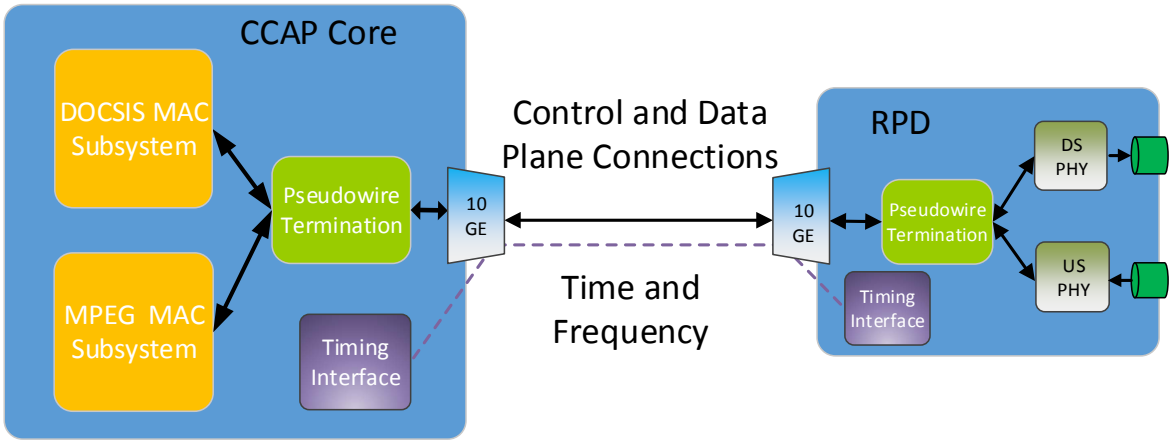


Figure 4 - Remote PHY System.

The connections between the CCAP-Core and the RPD can be established over any L2/L3 network. Both CCAP-Core and the RPD typically facilitate network connections over Ethernet, 10 Gbps or higher. An example of an access network architecture is presented on Figure 5.

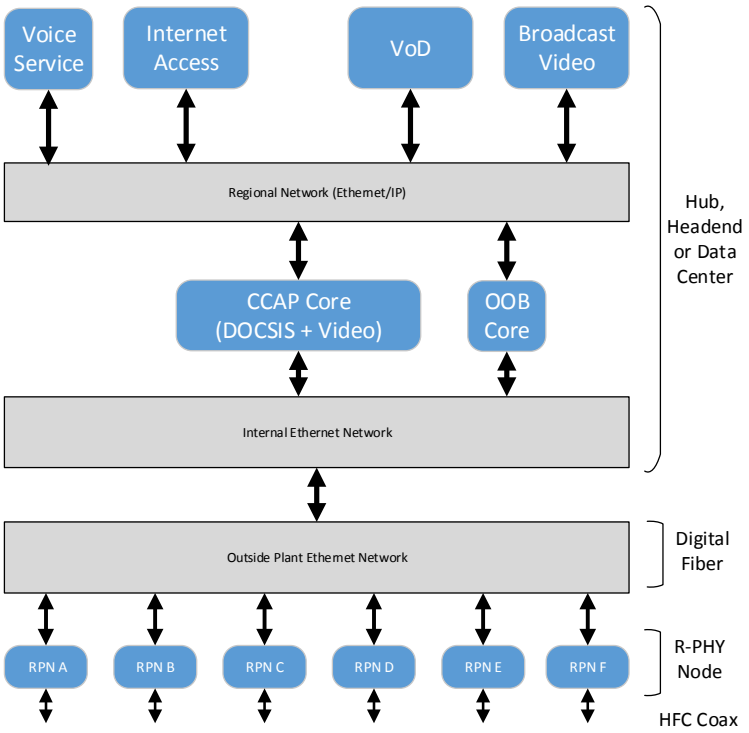


Figure 5 - Access Network Architecture with R-PHY

The CCAP-Core and the RPD communicate over a set of IP connections which includes a Generic Control Plane (GCP) connection for exchange of control information and a set of L2TPv3 pseudowires for transfer of service data between the systems. Additionally the CCAP-Core operates an L2TPv3 control connection to manage the L2TPv3 pseudowires and their network attributes.

The L2TPv3 pseudowires carry data for single DOCSIS or video or OOB channel between the MAC and the PHY. DEPI (Downstream External PHY Interface) specification defined downstream and overall L2TPv3 control plane operation. UEPI (Upstream External PHY Interface) specifies upstream data plane operation. To reduce the number of required pseudowires, DOCSIS or video channels can be grouped and their data carried in a single pseudowire. The R-PHY specifications define certain implementation simplifying rules on what types and how many channels can be grouped in a single pseudowire.

The basic connectivity between a MAC channel and the PHY channel is defined at pseudowire level and expressed as a combination of a pair of IP addresses and a negotiated L2TPv3 session ID and under the control of CCAP-Core software and its configuration logic. Per channel granularity in connectivity permits the Remote PHY architecture to support considerable flexibility in how the PHY-level channels in the RPDs can be connected to MAC-level channels in the CCAP-Core(s). This flexibility can be illustrated as a combination of three methods built in the R-PHY protocol.

<u>Independent Grouping of MAC Resources.</u>

As explained earlier, in Remote PHY protocols the data plane exchange is accomplished via L2TPv3 pseudowires. The protocol provides the means for creating the data plane connections between an RPD and one or more CCAP-Cores. In other words, MAC-level channels corresponding to the

PHY-level channels in a single PRD can be located in one or in multiple CCAP-Cores.

Two examples of MAC resource grouping are illustrated on Figure 6 and Figure 7.

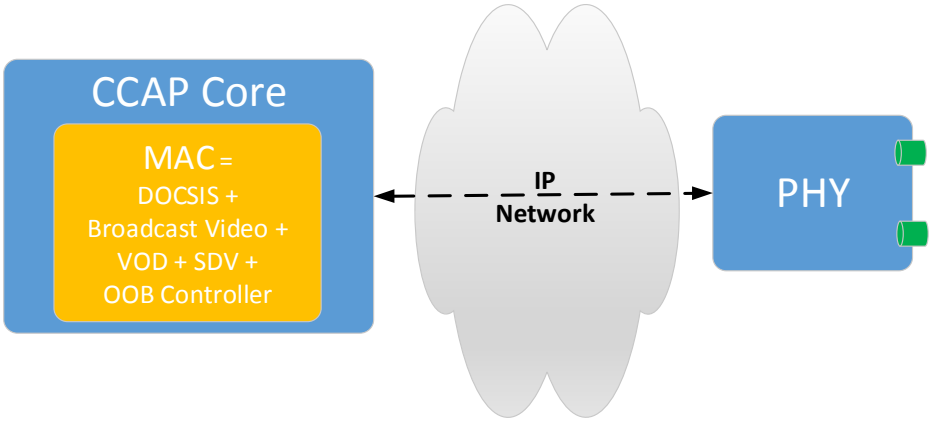


Figure 6 - All MAC resources located in a single CCAP-Core

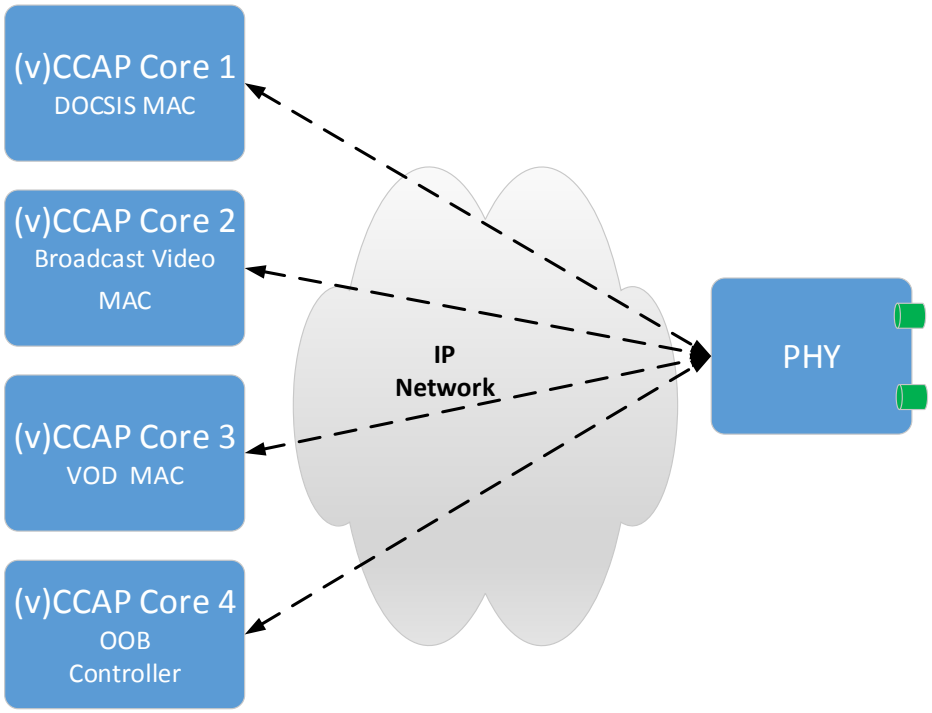


Figure 7 - Logical separation of MAC Resources to multiple CCAP-Cores

Figure 6 depicts a scenario with a single CCAP-Core which provides all MAC resources for all types of channels terminated by the RPD. This diagram exemplifies a scenario most closely resembling the pairing of MAC and PHY resources as implemented on I-CCAP. Such scheme represents the most straightforward step in evolution of the I-CCAP architecture. The virtual RF ports on the CCAP-Core are one-to-one matched to the physical RF ports in the interconnected RPDs. Each RPD is attached to exactly one CCAP-Core. This approach, perhaps combined with certain optimizations may be best suited for those vendors who are planning to evolve their existing I-CCAP systems which already implement DOCSIS and MPEG video services into CCAP-Cores.

Figure 7 illustrates the opposite approach to the example shown on Figure 6. Each CCAP-Core accommodates the MAC resources for only one service type. The diagram shows a separate DOCSIS CCAP-Core, Broadcast Video CCAP-Core, VOD Core and OOB Controller CCAP-Core. Each of the CCAP-Cores is dedicated to only one service category.

There are several benefits of separation of the CCAP service functions. Such an approach not only allows the operators to independently scale MPEG video, DOCSIS and OOB services but also removes many interdependencies in the systems providing these services resulting in improved availability and manageability. Unlike in integrated CCAP, the system maintenance functions, such as software upgrades or module replacement can be performed independently for each service. With proper accommodation in R-PHY protocols, failures within the DOCSIS domain will not affect the MPEG video domain and vice versa.

This approach may be best suited for those vendors who are planning new development of CCAP-Core systems, for example systems leveraging NFV technology.

When CCAP functions are separated to specialized systems providing a single service function, the RPD becomes the central point where the cable services actually convergence. Consequently, the acronym CCAP becomes obsolete. Is this a paradox or a paradigm shift? Does R-PHY architecture reinterprets "Cable Convergence" by ultimately forcing convergence deeper into the access network? For the first time that all services have to be converged. Before it was an option.

Efficient Use of Multicast Replication

R-PHY protocols define downstream data plane operation with IP multicast. The operators can provision the CCAP-Core to create a multicast DEPI pseudowires, each representing a single MAC channel or a set of MAC channels. A multicast group (S, G) sourced by the CCAP-Core can include one or more multicast pseudowires. The CCAP-Core advertises the multicast groups it originates via a multicast routing protocol, such as PIM.

During the RPD configuration process, the CCAP-Core instructs a number of RPDs to join the multicast group via IGMP/MLD in order to receive the data at the end of distribution tree. The RPDs subscribed to a multicast group can pick which pseudowires to receive by filtering based on L2TPv3 Session ID. DEPI multicast can be configured with granularity of a single DS channel.

An example DEPI multicast operation is shown on Figure 8.

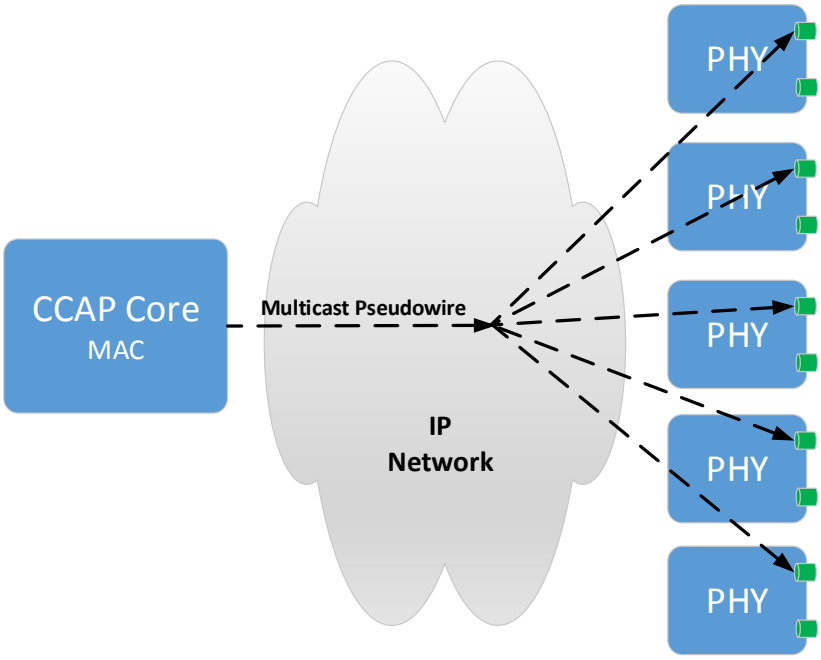


Figure 8 - DEPI multicast

Multicast operation can provide tangible savings to the operator. Multicast reduces the number of required MAC channel resources in the CCAP-Core and the bandwidth of the network connecting the CCAP-Core to the RPDs. One MAC resource in the CCAP-Core may serve tens of even hundreds of RPDs.

The primary use case for DEPI multicast is to distribute of video broadcast service where the same content, typically consisting of tens of MPEG video channels is intended to reach a larger serving group including possibly hundreds of RPDs.

In another use case, the operator may create downstream service groups consisting of several RPDs' downstream ports by configuring the CCAP-Core to replicate the complete lineup of downstream channels, or a set of downstream channels belonging to a particular service to RPDs in the selected group. This technique is referred to as virtual splitting because is the protocol equivalent of electrically or optically splitting the RF signal to multiple fiber nodes.

In such scenario each RPD DS RF Port individually converts identical digital signals to RF signals and all downstream CPEs, i.e. settop devices and DOCSIS Cable Modems receive identical content. The virtual splitting scheme may be useful to construct individually sized service groups for each service type or to support independent scaling of MAC and PHY resources.

For example, let us consider a transitional scenario where an operator builds up a complete deep fiber plant, splitting the nodes and installing as many RPDs as necessary to fulfill long term bandwidth needs. With virtual splitting, such a plant can operate with reduced number of downstream MAC resources in the CCAP-Core until the actual bandwidth demand rises to the level exceeding CCAP-Core capacity. Only then the downstream MAC resources in the CCAP-Core need to be scaled up.

Certain downstream PNM (Proactive Network Maintenance) functions require customization due to virtual splitting because these functions are implemented in the downstream modulator. To illustrate the

problem let us consider a virtually split OFDM channel. The downstream symbol capture function requires that the DS OFDM modulator in the RPD captures the set of digital samples representing a complete OFDM symbol. In the case of virtually split OFDM channel, there are multiple RPDs and multiple modulators involved. Therefore the symbol capture function needs to take place in multiple DS OFDM modulators. Current PNM specification do not cover such possibility. An adequate solution is currently under study.

Virtual Combining of Upstream Channels.

Remote PHY offers another technique, virtual combining of upstream channels which serves a similar purpose in the upstream direction as virtual splitting serves in the downstream. As the name suggests, virtual combining provides functionality equivalent to combining input signals from two fiber nodes. Conceptually a single MAC-level upstream channel resource in the CCAP-Core can be associated with more than one upstream PHY-level channel resources in the RPDs. All PHY-level US channels in a combined group have the same PHY-level parameters and share combined spectrum.

Unfortunately, since in this topology the data flows from many sources to one destination, virtual combining cannot rely on multicast transport. The number of upstream pseudowires scales accordingly with the number of involved upstream PHY-level channel resources in the RPDs. For example, if the US PHY-level channels are combined from four RPDs to feed into a single MAC-level US channel, the CCAP-Core needs to create four separate data pseudowires.

Virtual combining can be best described a method of multiplexing of data at PHY layer. There is impact of to CCAP-Core operation at the MAC layer and to higher layers. The upstream bandwidth scheduler in the CCAP-Core generates a single MAP stream, which is distributed to all burst receivers in a combined group. The CCAP-Core software manages a single set of resources and protocol identifiers such as SIDs.

The CCAP-Core needs to account for minor side effects of virtual combing. Certain statistical counters, for example the “burst collision” counts and “not energy burst” counts will be reported by burst receivers independently and need to be corrected as if they were a result of operation with a single burst receiver. Also, those upstream PNM functions which are implemented in the US burst receiver (Spectrum Analysis, Upstream Histogram) require re-work because the current specifications assume one to one mapping between the MAC layer and the PHY layer.

Virtual combining is completely transparent to the RPDs. RPDs are not at all aware that they participate in virtually combined group; no changes are needed to Remote PHY protocols. All knowledge of virtual combining, necessary configuration and packet mapping functions are implemented solely on the CCAP-Core.

In theory, virtual combining can be deployed with the granularity of a single DOCSIS US physical channel. In practice, more likely virtual combining will be applied on a service group level; each DOCSIS channel in the upstream lineup will be combined together with corresponding channels in a combined group of RPDs.

By deploying a combination of the techniques described above the operators can construct R-PHY access networks which exhibit a number of advantages over I-CCAP based design:

1. The size of the service delivery platforms can be scaled independently and tuned to the desired service footprint.

2. Broadcast services can be cost effectively scaled through network based multicast replication.
3. Virtual splitting and combining enables operators to independently craft per service serving group sizes and to flexibly match the ratio between the MAC-level resources to the PHY-level resources.

An example of a completed HFC access network is shown on Figure 9.

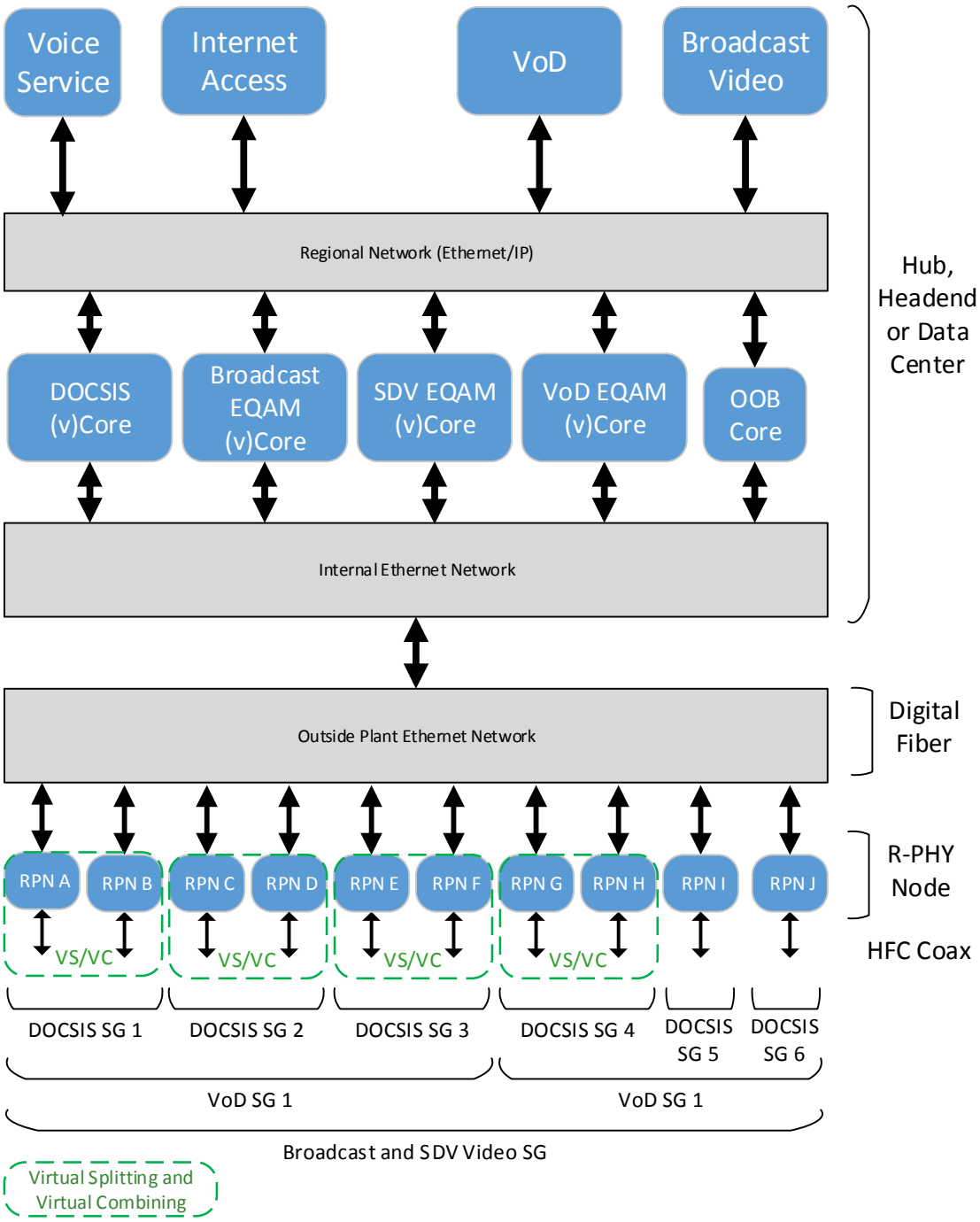


Figure 9 - Service Group in HFC access network

Conclusion

The paper demonstrates that Cable Operators can leverage Remote PHY technology to construct cable access networks that fulfill or even exceed the convergence goals for CCAP. Remote PHY replaces analog HFC connections with Ethernet and IP network, as result the operators are given new options to effectively configure service groups and match MAC and PHY resources with unparalled flexibility.

References

1	http://www.cablelabs.com/wp-content/uploads/specdocs/CM-SP-MULPIv3.1-I08-151210.pdf
2	http://www.cablelabs.com/wp-content/uploads/specdocs/CM-TR-CCAP-V03-120511.pdf
3	http://www.cablelabs.com/specs/specification-search/?cat=docsis&scat=dca-mhav2
4	http://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html
5	"A Next Generation Architecture for Business Services Delivery Based on SDN and NFV", Gerry White, SCTE Expo 2014.
6	"SDN As A Matchmaker For Remote Phy Architecture" Alon Bernstein, Sangeeta Ramakrishnan, INTX 2015.

Abbreviations & Acronyms

CCAP	Converged Cable Access Platform
CMTS	Cable Modem Termination System
CPE	Customer Premise Equipment
CPU	Central Processing Unit
DEPI	Downstream External PHY Interface
DOCSIS	Data Over Cable System Interface Specification
EQAM	Edge Quadrature Amplitude Modulator
HFC	Hybrid Fiber Coax
I-CCAP	Integrated CCAP
IP	Internet Protocol
MHAv2	Modular Headend Architecture version 2
MPEG	Moving Picture Experts Group

MSO	Multiple System Operator
NFV	Network Function Virtualization
OAM&P	Operations Administration Maintenance and Provisioning
OOB	Out of Band
OSS	Operational Support System
PIM	Protocol Independent Multicast
PMD	Physical Media Dependent
RF	Radio Frequency
RPD	Remote PHY Device
RPN	Remote PHY Node
SDN	Software Defined Networking
SG	Service Group
UEPI	Upstream External PHY Interface

The World Is Flat
Capacity Optimization in a Coaxial Network, Constrained by Total RF Power

Karl Moerder PhD, Futurewei Technologies Inc.
Fred Harris PhD, San Diego State University

Abstract

As data traffic continues to grow, the cable industry is reaching the point where downstream capacity in a coaxial network is completely limited by the total integrated RF output power available from a node or amplifier. MSOs today are faced with the problem of increasing the peak network capacity to meet marketing and customer demand. Pushing fiber deeper into the network is one solution, but it is expensive. Working with the existing coaxial network and efficiently utilizing the RF power for maximum capacity can be a cost effective alternative.

Increasing the system bandwidth to 1.2 or 1.8 GHz with a constant quadrature amplitude modulation constellation size and pre-emphasis, while maintaining amplifier spacing based on a 750 or 860 MHz plant design requires a positive slope in the power spectral density (PSD) of as much as 22 dB across the 1.2 GHz band and is far from optimal. Transmitting a signal centered at 1.2 GHz costs over 100 times as much RF power as transmitting the identical signal centered at 100 MHz. It is hard to imagine that this could be sustained as the system bandwidth is increased.

We show that when constrained by the total RF power available, the power spectral density at the output of the amplifier should be nearly flat. The small amount of ripple in the optimal power spectral density is due to the discrete set of QAM constellation sizes.

THE PROBLEM

As data traffic continues to grow, the cable industry is reaching the point where downstream capacity in a coaxial network is completely limited by the total integrated RF output power available from a node or amplifier. MSOs today are faced with the problem of increasing the peak network capacity to meet marketing and customer demand. Pushing fiber deeper into the network is one solution, but it is expensive. Working with the existing coaxial network and efficiently utilizing the RF power for maximum capacity can be a cost effective alternative.

Increasing the system bandwidth to 1.2 or 1.8 GHz with a constant quadrature amplitude modulation (QAM) constellation size and pre-emphasis, while maintaining amplifier spacing based on a 750 or 860 MHz plant design requires a positive slope in the power spectral density (PSD) of as much as 22 dB across the 1.2 GHz band. Coaxial loss increases with frequency, and the pre-emphasis compensates for the increased loss at higher frequencies. Using pre-emphasis with a constant constellation size to maintain a constant signal-to-noise ratio (SNR) over the entire system bandwidth at the receiver is far from optimal.

As an example, transmitting a 6 MHz 256 QAM signal centered at 1.2 GHz costs over 100 times as much RF power as transmitting the identical signal centered at 100 MHz. It is hard to imagine that this could be sustained as the system bandwidth is increased.

Any combination of coaxial loss, amplifier spacing and system bandwidth, results in a negative gain slope across the system bandwidth. Figure A illustrates the typical cable loss and PSD found in a traditional cascade of coaxial amplifiers with both analog and digital television. We express this slope in dB/100 MHz and as a simplification, approximate the slope as constant over the system bandwidth. For any particular slope and system bandwidth, this paper addresses the problem of selecting the QAM constellation size at each frequency to (1) minimize the RF power required to achieve the desired capacity, or similarly to (2) maximize the capacity achievable with a limited amount of RF power.

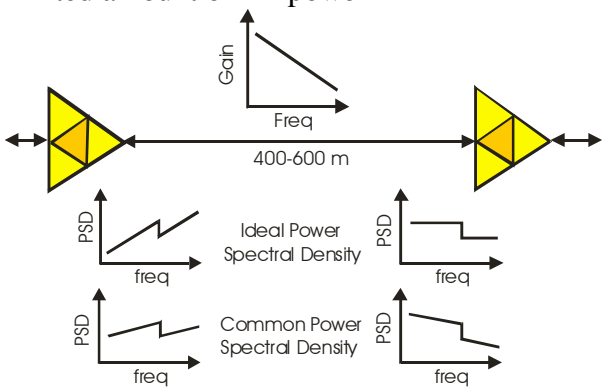


Figure A Traditional Amplifier Cascade

Figure A shows a typical span between amplifiers of 400 to 600 Meters. The total RF power at the output of each amplifier ranges from 73 to 76 dBmV. The frequency dependent loss of the cable is shown directly above the cable. The ideal PSD for a constant QAM constellation size, from an SNR point of view, at the input to each amplifier is flat across the band. This requires pre-emphasis at the amplifier output to offset the slope of the coaxial cable loss. The step drop of about 6 dB in power spectral density between analog television channels at lower frequencies and digital transmission at higher frequencies protects the analog television channels from interference by the digital transmission.

Many older amplifiers were incapable of providing sufficient linearity with a large slope in the PSD across the band. Figure A also shows a second example where the pre-emphasis is split equally between the input and the output of the amplifier. That is, each amplifier provides only half of the pre-emphasis at its output required to offset the gain slope of the coaxial cable. The remaining pre-emphasis is provided at the input of the next amplifier in the cascade. While the amplifier linearity is improved, a SNR penalty is paid.

In many hybrid fiber coaxial (HFC) networks around the world, the elimination of analog television signals is either underway or completed. This eliminates the need for protection of the analog television channels and simplifies the design of the HFC network.

In this paper, we use system capacity to mean the total system throughput in bits per second. We show that the PSD out of the amplifier and into the coaxial cable should be nearly flat over the system bandwidth for maximum system capacity when the total RF power out of each amplifier is limited. So, the PSD at the input to each amplifier should be decreasing with frequency. The total system capacity is maximized with higher-order constellations at lower frequencies and lower-order constellations at higher frequencies. This can be accomplished by filling the downstream bandwidth with DOCSIS 3.1 OFDM channels or by using a larger range of constellation sizes than DOCSIS 3.0 provides. Constellations ranging from 16384 QAM at low frequencies down to 64 QAM at high frequencies, or 4096 QAM at low frequencies down to 16 QAM at high frequencies, maximize capacity for most coaxial networks.

Figures B and C illustrate the PSD into the amplifier, and out of the amplifier and into the coaxial cable for two cases. Figure B shows the PSD for a constant 1024 QAM

constellation size from 100 MHz to 1.2 GHz. Figure C shows the PSD for QAM constellation size that decreases from 16384 QAM at 100 MHz to 64 QAM at 1.2 GHz. The QAM constellations are scaled to maintain the same Euclidian Distance before pre-emphasis. Maintaining the same Euclidian distance, maintains nearly the same symbol error-rate performance with respect to SNR or modulation error ratio (MER). The small difference in symbol error-rate performance is due to the changes in the fraction of edge and corner points for different QAM constellation sizes. The capacity in both figures is 8.8 Gb/s, but the total RF power consumption in Figure C is 56% less than that in Figure B.

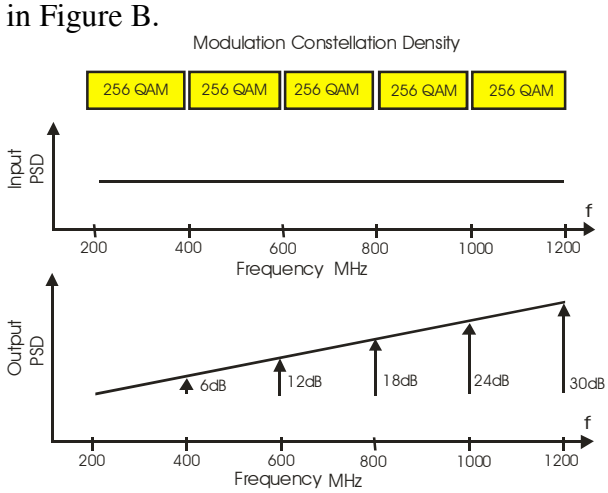


Figure B 1024 QAM System Capacity

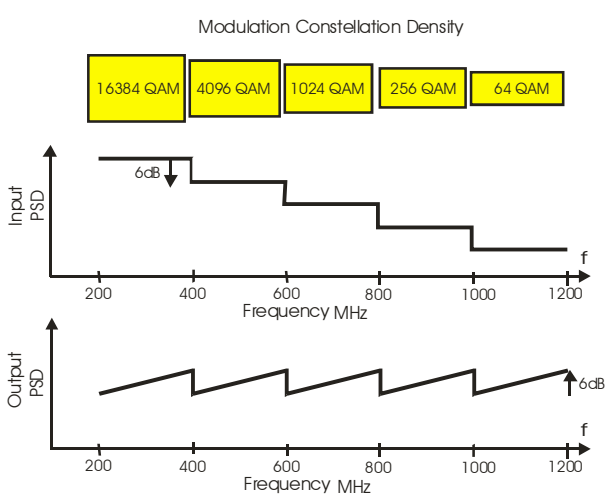


Figure C Optimal System Capacity

While new amplifier designs using state of the art second generation GaN technology can produce more total RF output power than earlier designs, total RF power and the resulting DC power consumption for each amplifier remains an issue for HFC network planning. The distortion in the amplifier depends on more than just the total RF output power. A large positive slope in the PSD increases the amplifier distortion. While distortion is more severe at higher frequencies, the magnitude of the difference in PSD between lower and higher frequencies further stresses the amplifier dynamic range. Distortion components from the higher frequencies can fall back on the lower frequencies that are disadvantaged by 22 dB or more.

A VIEW OF RF POWER AND CAPACITY

We view RF power in relative terms and capacity discretely in bits per second. That is, the cost of each increment of capacity is the RF power required at the output of the amplifier relative to that required for transmitting 4 QAM at the lowest frequency. This simplifies the discussion by eliminating specific link budgets and absolute power levels.

We divide the system bandwidth into equal sized blocks of frequency and consider square QAM constellations, so one unit of capacity is two bits per symbol for a fixed size frequency block. Figure D shows a log plot of the relative power, starting at $\log_{10}(1) = 0$, required to transmit a QAM constellation from 4 QAM to 16384 QAM at any frequency across the system bandwidth. We assume that the code rate, coding gain and any other performance parameters are independent of constellation size. These simplifications can be relaxed as discussed in the conclusion. In Figure E, there are three curves representing the linear RF power required for 4 QAM, 16 QAM and 64 QAM, as a function of

frequency. The frequency range and QAM constellation sizes in the figure are limited to make the size of each cell visible. Each cell in the graph with a width of 50 MHz, below the first curve, or from one curve to the next represents one unit of capacity. So, one unit of capacity can be obtained by transmitting 4 QAM in any frequency block. A second unit of capacity can be obtained either by transmitting 4 QAM in a second frequency block, or by changing from 4 QAM to 16 QAM in the first frequency block. A third unit of capacity can be obtained either by transmitting 4 QAM in a third frequency block, or by changing from 16 QAM to 64 QAM in the first frequency block. Allocating capacity can be viewed selecting cells in the figure. Each cell represents the cost (RF power = power spectral density * frequency) to transmit the corresponding unit of capacity. An optimal solution to (1) is a minimum cost selection of cells with the required number of cells. An optimal solution to (2) is the maximum number of cells selected without exceeding the allowed total cost.

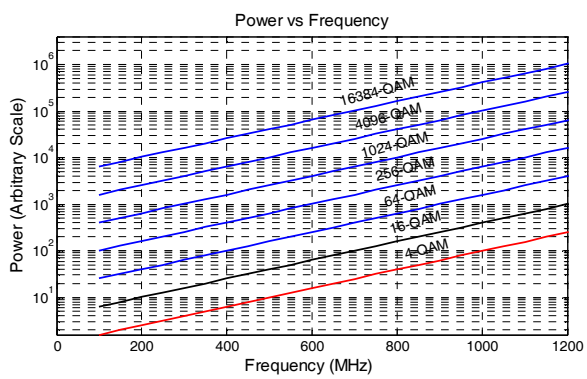


Figure D Power Cost

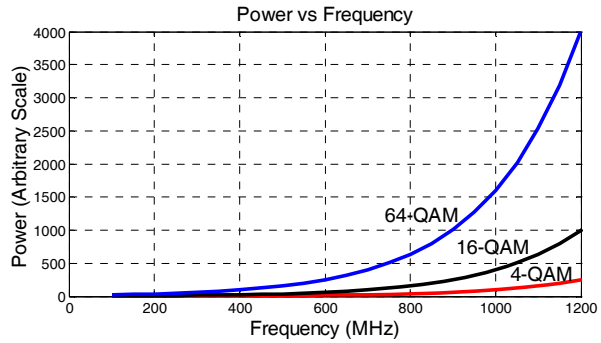


Figure E Power Cost

OPTIMAL CAPACITY AND RF POWER

We now address constructing a solution to (1) and (2) by adding capacity until obtaining the desired outcome. Each unit of capacity has equal value because it provides the same increase in b/s throughput. The cost of each unit of capacity is different. At each step in the process, we add the least expensive increment of capacity.

An optimal solution, S, to (1) is obtained by starting with zero capacity and repeatedly allocating the globally least cost unit of capacity until the desired capacity is obtained.

Suppose that the above solution, S1, were not optimal. That means that there is another solution, S2, with lower total cost than S1. By construction, the cost of every unit of capacity in S1 is less than or equal to the cost of every unit of capacity not in S1. Both S1 and S2 contain the same number of units of capacity, so any other solution to the problem, such as S2, may be obtained by removing some number of units, k, of capacity from S1 and replacing them with k units of capacity from not S1. But the cost of each unit of capacity removed from S1 is less than or equal to the cost of each unit of capacity selected from not S1. So the sum of the costs of the k units of capacity selected from not S1 is greater than or equal to the sum of the costs of the k units of capacity removed from S1. Therefore, the cost of S2 is greater than or equal to the cost of S1. This contradicts the assumption that S2

had lower total cost than S1. Therefore S is an optimal solution to (1).

An optimal solution, S, to (2) is obtained by starting with zero capacity and repeatedly allocating the globally least cost unit of capacity until all available RF power is used.

The proof of this is similar to that of (1).

In order for a solution to (1) or (2) to be feasible, it must satisfy the following constraint, as shown in Figure E: If any cell is allocated, then all of the cells below it must also be allocated.

As an example, it makes no sense to allocate 16 QAM for a particular frequency block without previously allocating 4 QAM for that same frequency block. The cost of 16 QAM is 4 times the cost of 4 QAM in the same frequency block. However our approach would only add the incremental cost, three times the cost of 4 QAM, under the assumption that 4 QAM has previously been allocated.

Fortunately, the incremental cost of 16 QAM is three times the cost of 4 QAM in the same frequency block. So selecting the globally least unit of capacity will always select 4 QAM in any frequency block before 16 QAM in the same frequency block. Therefore the above condition will be satisfied and the solution obtained will always be feasible. Although 4 and 16 QAM are used in this example, the same argument holds true for each increasing QAM constellation size.

THE ALGORITHM

With the above discussion, a simple algorithm for an optimal solution to (1) can now be described.

For any particular frequency block, the cell under the lowest curve is the RF power cost for 4 QAM, the cell between the first and second curves is RF power cost for 16 QAM, and the cell between the second and third curves is the RF power cost for 64 QAM.

Consider the following algorithm.

1. For each frequency block and for the curves for each size QAM constellation, create a list, L, of the height under the first curve, or between consecutive curves for that frequency block.
2. Sort the list, L, from least to most cost. If two or more cells have the same cost, the order is arbitrary.
3. Select the first n cells, where
   n = (desired capacity) / (capacity of a cell).
4. For each frequency block, we interpret the constellation size as the largest constellation size selected for that block.

This is, in fact, an optimal solution to (1). After sorting, the first cell in L is the globally least cost unit of capacity. Removing the first cell, leaves the second cell as the globally least cost unit of capacity in L. And so on. Therefore this algorithm satisfies the condition stated above for an optimal solution to (1).

CONCLUSIONS

We have shown that when constrained by the total RF power available, the power spectral density at the output of the amplifier should be nearly flat. The small amount of ripple in the optimal power spectral density is due to the discrete set of QAM constellation sizes. Extending the allowed set of constellations to include non-square QAM constellations will reduce the magnitude of the ripple. Care must be taken that we do not violate the feasibility constraint.

Using a fixed size frequency block allows for the simplest algorithm for both optimizing system capacity and demonstrating the algorithm's correctness. However, the fixed size frequency block is only a simplification; arbitrary frequency block sizes are possible.

A flat power spectral density at the output of the amplifier minimizes the total RF power required and thus the DC power required for the amplifier. It also simplifies the design of the amplifier. We mentioned that a large slope in the power spectral density makes the amplifier design more difficult. A flat power spectral density can eliminate the compromise of splitting the pre-emphasis between the input and output of the amplifier.

The effects we have shown in this paper are much less dramatic for a narrower total system bandwidth. It is only as we push the bandwidth to 1.2, 1.8 or even 3 GHz that a new look at total RF power, system bandwidth and system capacity becomes necessary.

RELATIONSHIP TO PREVIOUS WORK

Claude Shannon [2] developed Information Theory during World War II. From an information-theoretic standpoint, he solved the problem of selecting the optimal PSD for communication over an additive, colored Gaussian noise channel in his original paper. Robert Gallager [1] provides a description of the Water-Pouring algorithm for this selection process. Our algorithm provides a discrete calculation for the capacity of a coaxial channel limited by total availableRF power.

In a CableLabs internal white paper, *Shannon's Limits Applied to Cable Networks above 1 GHz*, Tom Williams [3], Gregg White and Alberto Campos suggest using "wide bandwidth transmissions above 1 GHz for best performance when transmit power is limited." Using a larger number of smaller QAM constellation sizes to achieve the same capacity is one way to use wider bandwidth transmissions at higher frequencies.

REFERENCES

[1] R. C. Gallager, Information Theory and Reliable Communications, Wiley, 1968.

[2] C. E. Shannon, "A Mathematical Theory of Communication," Bell System Technical Journal, Volume 27, 1948.

[3] T. Williams, G. White and A. Campos, "Shannon's Limits Applied to Cable Networks above 1 GHz," CableLabs, 2014.

TOWARDS SDN FOR OPTICAL ACCESS NETWORKS

Dr. Eugene (Yuxin) Dai*, Wei Dai**

Cox Communications, Network Engineering, 6305 Peachtree Dunwoody, Atlanta GA 30328
***Appfolio, Software Engineering, 50 Castilian, Santa Barbara CA 93117*
**Eugene.dai@cox.com, **Weidai@appfolio.com*

Abstract

The SDN (Software Defined Network) concept has made impressive progress in both enterprise and transport networks. The core idea behind SDN is the separation of control and data forwarding planes. However, the SDN concept is still in an early stage of penetrating optical access networks, especially passive optical networks. A major difficulty of SDN for PON (Passive Optical Network) is how to separate control and forwarding planes in optical access networks. In this paper we discuss the first step towards SDN for optical access networks – determining the control plane of optical access networks, and separating it from the data forwarding plane. We will first discuss SDN for Active Optical Network (AON) and then discuss SDN for PONs.

1. INTRODUCTION

The SDN concept has become a new paradigm for modern communication networks. It has made impressive progress in both enterprise and transport networks in recent years. The core idea behind the SDN is the separation of the control plane from the data forwarding plane. The SDN controller has an updated view of the entire network topology. In principle, through virtualization the SDN controller using OpenFlow or other protocols, can make routing & forwarding decisions based on dynamic multiple layers of network resource availability information. There are two major benefits of SDN. The first benefit is that SDN can more efficiently use network resources. Since the centralized SDN controller has a dynamic global view of the network topology, it can more effectively route traffic through the network. Secondly, SDN/NFV (Network Function Virtualization) can lower network CAPEX and OPEX costs by separating the control plane from the forwarding plane. That in turn makes the network forwarding elements simpler and less costly.

At conception, the SDN was initially developed for packet based enterprise networks where the control plane is separated from the data plane. A vendor agnostic interface is defined between the control and the data planes of the network forwarding elements (switch and routers), and a logically centralized SDN controller is also defined.

Transport SDN has been introduced recently as an effort to more effectively use network resources in optical transport networks. It can be viewed as the natural extension of ASON (Automatically Switched Optical Network) and ASTN (Automatically Switched Transport Optical Network), which were standardized at ITU-T in mid 2000. ASON and ASTN introduced a control plane with distributed path computation on top of the optical transport network that perfectly fits the idea of separation of control plane and data plane in SDN, and enables a holistic view of the network. Centralization of path computation is a novel idea in transport SDN. Through virtualization, the SDN controller can make routing & forwarding decisions based on multi-layer network availability information.

The passive optical networks (PON), including IEEE EPON family (EPON, 10G EPON, 100G EPON) and ITU-T GPON family (GPON, XG-PON, XGS-PON, NG-PON2), are becoming the most promising technologies for FTTH (Fiber to the Home). However the SDN concept is still in an early stage of penetrating access networks, especially passive optical networks. A major challenge of SDN for PON is how to separate the control plane and data forwarding plane. That raises the question of where the SDN boundary should extend to, OLT or ONUs. The PON control and forwarding functions, regardless of EPON family or GPON family, are vertically integrated in a mostly closed system between the OLT (Optical Line Termination) and ONUs (Optical Network Unit). In this paper we discuss the first step towards SDN for PON – the separation of control planes of optical access networks from their data-forwarding planes. We propose a novel method to separate the control plane of the PON from its data plane. We will first discuss SDN for Active Optical Network (AON), and then discuss SDN for PON.

2. CURRENT OPTICAL NETWORK ARCHITECTURES

Active Ethernet is shown in Figure 1 (a). Although Ethernet in the optical access network has point-to-multipoint fiber topology viewed from the Central Office (CO) or Head-end (HE), it is a point-to-point network from the MAC (Multiple access control) layer. Active Ethernet consists of a P2P trunk fiber from a Central-office to an Ethernet aggregation switch in the field location, and multiple P2P (Point-to-Point) fiber terminations from the Ethernet switch to the subscribers. The active Ethernet OSP infrastructure can be easily converged to PON ODN (Optical Distribution Network) by replacing the Ethernet switch in the field with a passive optical power splitter as shown in Figure 1 (b). TDM (time division multi-plexing) PON is the most deployed technology for FTTH.

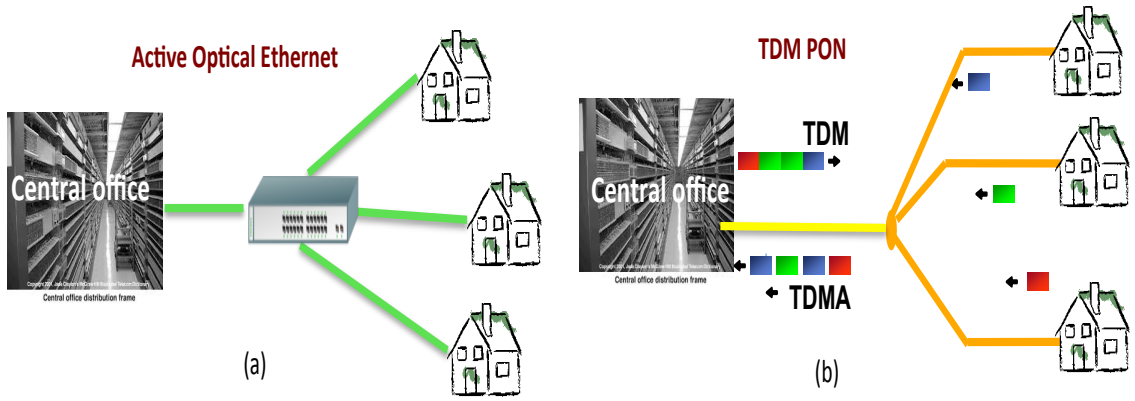


Fig. 1 Active and passive optical Networks, (a) Active optical access network and (b) Passive optical network

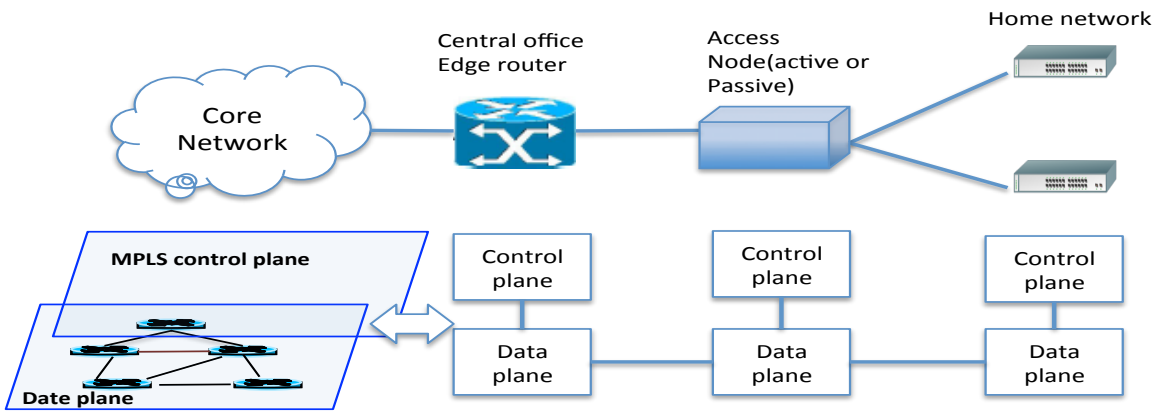


Fig. 2 Current optical access network architecture

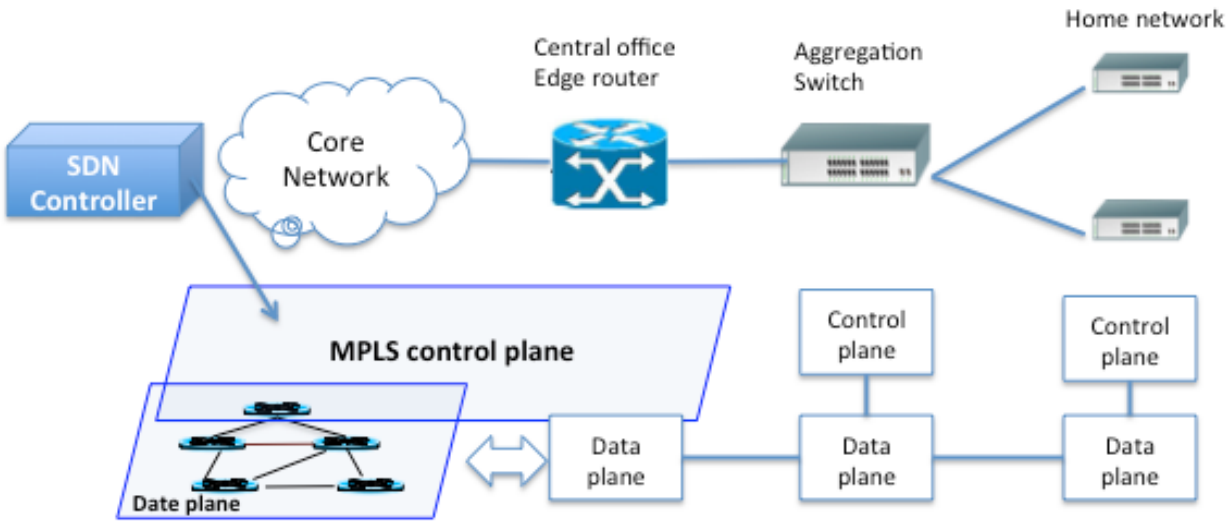


Fig. 3 Extends MPLS/GMPLS control plane to edge router

The current optical network architecture from core optical networks to optical access networks is shown in Figure 2. Most of the core optical networks already have a separated control and data plane. The control plane of the core network is built on MPLS and/or GMPLS, and the data plane runs distributed routing protocols.

The edge router at the service providers' Control-office or Head-end can be viewed as the gateway, for example, the the BNG (broadband network gateway) of the access network to core network. The control plane of BNG can be separated from its data plane using SDN concepts. Therefore, the MPLS/GMPLS control plane of the core optical network can be extended to the edge router in the central-office, as shown in Figure 3.

A SDN control plane then extends its control to the edge router, or BNG. Then the problem remaining is the separation of control and data planes in optical access networks. We will first discuss the Active Optical Network for access.

3. SEPARATE CONTROL AND DATA PLANES OF AON

The access network attached to the BNG shown in Figure 3 is active Ethernet. The IEEE 802.1D Ethernet physical layer topologies can be point-to-point or mess. Using the spending tree protocol, the Ethernet topology can be viewed as tree and branch, as shown in Figure 4. The 802.1Q introduced VLAN and the 802.1ad introduced customer VLAN (C-VID) and service provider VLAN (S-VID) concepts.

With 802.1Q and 802.1ad, Ethernet can be virtualized into VLANs. The virtualization of Ethernet provides a way to separate the control and data planes for Ethernet. Instead of discussing the control plane for forwarding based on Ethernet source and destination addresses, we can now introduce a control plane for VLANs. The multiple VLAN control protocol (802.1ak Multiple VLAN Registration Protocol-MVRP) can be used as one of the multiple VLAN control protocols for the virtualized the Ethernet data plane, as shown in Figure 5.

MVRP is a layer 2 protocol that creates and manages the VLANs. Other Multiple VLAN control protocols include Rapid Spanning Tree protocol, Shortest Path Bridge, Multiple Spanning Tree protocol, etc. With multiple VLAN control protocols as the Ethernet control plane protocols, the SDN controller can then extend its control to both MPLS control plane for core optical networks, in addition to Ethernet control plane in optical access networks.

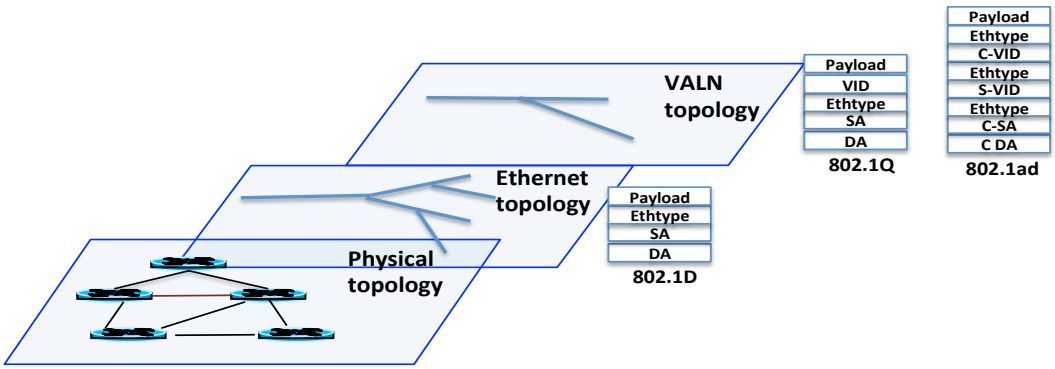


Fig. 4 Ethernet and topology and virtualization

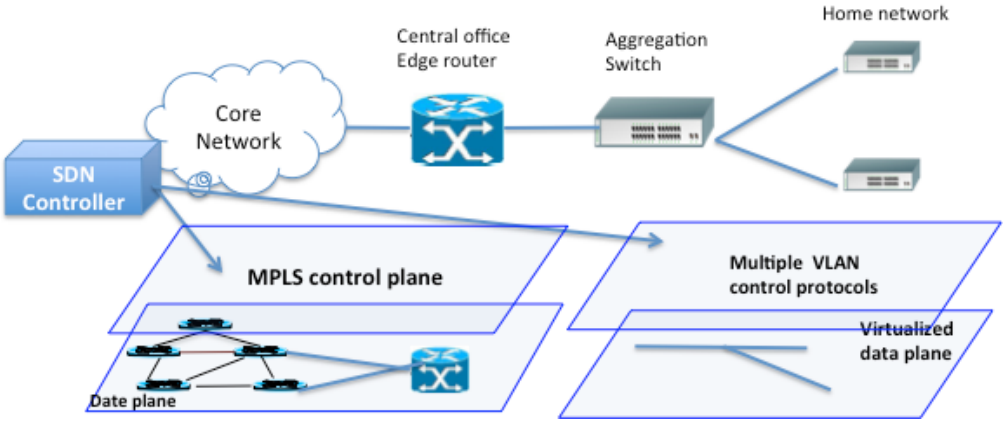


Fig. 5 Ethernet control plane and data plane

4. SDN VIEW OF CORE NETWORK AND AON

When the Ethernet control plane and data plane are separated based on VLANs as proposed, a unified SDN based control plane can be created to control both core networks and access networks. The SDN view of the entire network is shown in Figure 6. The SDN control plane interfaces with network switching and forwarding elements via southbound switch APIs to achieve the first layer of abstraction. It abstracts away the vendor specified hardware interfaces [1]. The SDN controller also has the global view of the network topology. In this global common map abstraction, the Ethernet access network portion is represented by VLAN topology.

The SDN control plane interfaces with the application layer via northbound APIs. Together, the unified SDN control plane extends automated provisioning to the Ethernet access networks and optimizes the use of the network resources across the entire network. Using OpenFlow, the flow forwarding decisions can be based on L2 to L4 information, for example any combination of MAC addresses, VLAN tags, IP addresses, TCP ports, UDP ports, etc.

5. SEPARATE CONTROL PLANE AND DATA PLANE FOR TDM PON

The access network attached to the BNG shown in Figure 3 in this case is PON. There are several literatures on SDN for TDM PON. It is proposed in literature [2] for an OpenFlow controller to replace Dynamic Bandwidth Allocation (DBA) in GPON. There are many problems to link SDN with DBA. First, DBA is a mechanism to enable more efficiently use upstream resources for the EPON family or GPON family. Although DBA controls upstream bandwidth allocation, it is hard to qualify it as a control plane protocol. It is not related to services creation. Second, DBA is a closed protocol between OLT and ONUs in a TDM PON. Take GPON for example, in state report-DBA, the ONUs report their transmission buffer depths to OLT and OLT grants upstream transmission time slots accordingly. It is not feasible for moving the DBA function to an external OpenFlow controller because of delays. Thirdly, the service creations are controlled by LLID (Logical Link identifier) in EPON family and GEM/XGEM Port-ID in GPON family. The LLID and GEM/XGEM Port-ID are also closed parameters between OLT and ONUs in EPON family and GPON family respectively. It is hard to move them to an external OpenFlow controller.

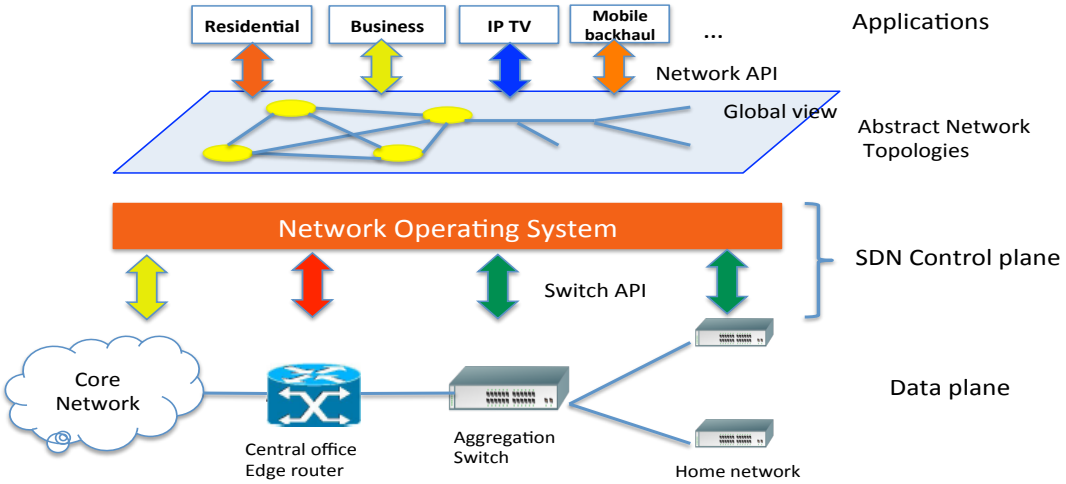


Fig. 6. SDN view of core network and active Ethernet access network

Since DBA, LLID, GEN/XGEM Port-ID are closed between OLT and ONUs in EPON family and GPON family, they should not be visible to the SDN controller. The TDM PON can be viewed as an N+1 ports distributed Ethernet switch: one port for OLT, N ports for ONUs, as shown in Figure 7. This distributed Ethernet switch can be virtualized using VLAN. For this distributed VLAN switch, the control plane and data plane can be separated as discussed in section 3 using MVRP. In this way, a common SDN controller can control both core networks and PON based access networks. Therefore the discussions in section 4 apply to PON based access networks as well.

6. CONCLUSIONS

VLANs can be used to virtualize Ethernet based optical access networks, and the control plane and data plane of active Ethernet based optical access networks can be separated using MVRP. The TDM PON based optical access network can be viewed as a distributed VLAN switch, and therefore the control plane and data plane for TDM PON can be separated using MVRP as well. As a result, a unified SDN control plane can control the optical access network and the core optical network.

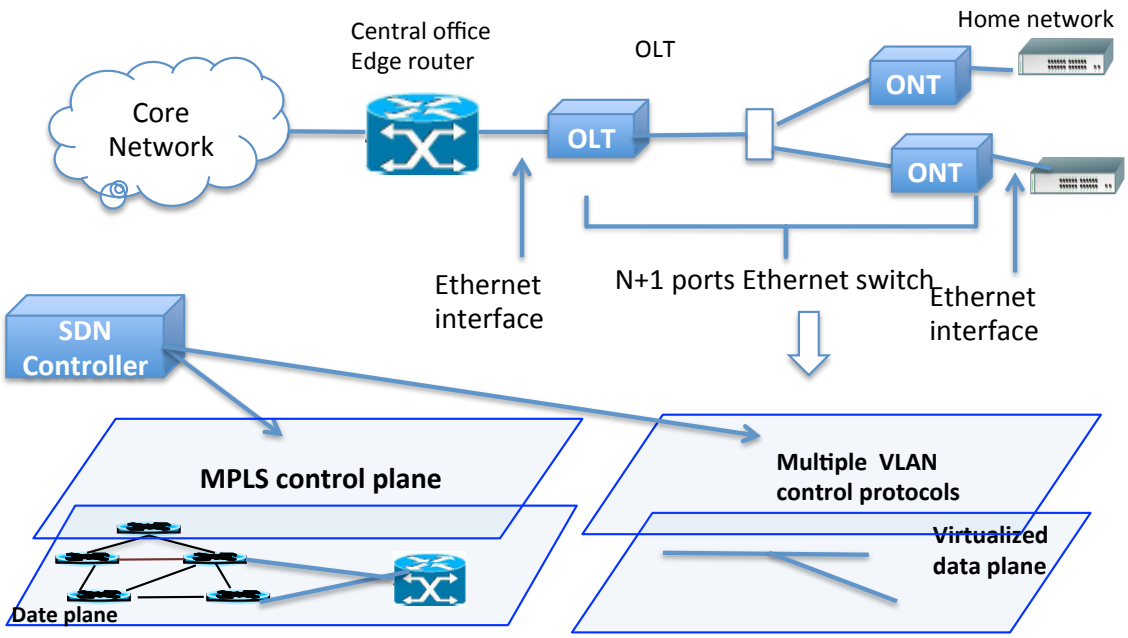


Fig. 7. Separate control plane and data for TDM PON

7. REFERENCES

[1] Saurav Das, Guru Parulkar and Nick McKeown, "SDN Based Unified Control Architecture", page 778-779, IEEE Photonics Conference, 2012

[2] Hagen Woesner, Daniel Fritzsche, "SDN and OpenFlow for Converged Access/Aggregation Networks", OTu3E.4, OFC/NFOEC Technical Digest 2013

USING DOCSIS TO MEET THE LARGER BANDWIDTH DEMAND OF THE 2020 DECADE AND BEYOND

Tom Cloonan, Ayham Al-Banna, Frank O'Keeffe
ARRIS

Abstract

Forward-looking predictions on future bandwidth demand indicate that the 10-15 Gbps BW capacities of DOCSIS 3.1 systems may become limiting by the middle of the 2020 decade. This paper will explore novel ideas that propose to extend the DOCSIS 3.1 specifications to permit DOCSIS to support subscriber Bandwidth (BW) needs into the 2030 decade and beyond. The paper will study BW trends, discussing the details of different options for augmenting plant BW capacity. The paper will then describe extended-spectrum DOCSIS, showing that it offers many advantages to Multiple System Operators (MSOs), including backward compatibility, investment protection, the avoidance of large capital costs, elimination of unnecessary operational challenges, and the utilization of the full capacity of their existing Hybrid Fiber-Coaxial (HFC) networks.

The paper will analyze the performance of HFC networks at higher frequencies and propose a gradual migration plan that increases the BW from 10 Gbps to (perhaps) 200+ Gbps. The paper will also analyze the effect of multiple parameters such as attenuation, signal source tilt, multiple modulation profiles, variable bit loading, and type and length of the hard / drop coaxial cables. The paper will conclude by describing a potential evolutionary path that the MSOs may follow as they migrate from DOCSIS 3.1 systems to extended-spectrum DOCSIS systems.

INTRODUCTION

Traffic Engineering And Motivation

The Cable Industry is now a recognized leader in the delivery of Last-Mile Services to residential subscribers, and it is continually expanding its access to business subscribers. The offered services include Video, Voice, and Broadband Data services. While the Cable network started out as a simple coaxial infrastructure optimized for the delivery of Analog Video, it has evolved into a powerful Hybrid Fiber-Coaxial (HFC) system over the years to support these many varied service types.

The HFC network has proven time and time again that it is an extremely flexible and robust infrastructure that can be morphed to accommodate the unique and continually-changing needs of the subscribers and the different services. The addition of amplifiers for longer reach was the beginning of this evolution. That was followed by spectral expansions to deliver more TV channels. That was then followed by the addition of fiber to further increase reach and reduce the effects of electrical interference. The addition of DOCSIS for Voice and Broadband service was another powerful modification. Fiber Deep solutions (based on continual node-splits) have carried the network to its current state of existence. In the near future, DOCSIS 3.1 augmentations will expand the spectrum and the spectral efficiency of the HFC network to support (perhaps) ~10 Gbps Downstream bandwidth capacity and ~2 Gbps Upstream bandwidth capacity on a 1.2 GHz plant with a 204 MHz high-split. Discussions are now underway on the topics of Full-Duplex DOCSIS that may permit even higher Upstream bandwidth capacities in the near

future. Future DOCSIS 3.1 expansions to 1.7 GHz may even permit the HFC network to even support ~15+ Gbps of Downstream bandwidth capacity.

Is the 15 Gbps bandwidth capacity of DOCSIS 3.1 the last change that we will make to the HFC network? Is the bandwidth capacity offered by DOCSIS 3.1 (in its current form) adequate for the long-haul into the future? To answer these questions, we must explore the likely bandwidth requirements that may be placed on the HFC network over the next fifteen years. Voice and Video will always consume a reasonable percentage of the HFC spectrum. Voice bandwidth requirements for small service groups of the future will be essentially negligible, but video bandwidth for those service groups will become quite large as Ultra-High Definition (UHD) feeds become more prevalent. For example, if an MSO plans to deliver 500 programs of UHD Video (with (say) 12 Mbps per stream) using switched techniques (SDV or Multicast IP Video) to a futuristic Service Group with ~100 subscribers (50% active, 3 viewers per active home), then these programs might require ~1.4 Gbps of bandwidth capacity. In addition, nDVR/VoD viewing from those ~100 subscribers might require an additional 300 Mbps of bandwidth capacity (assuming a 30% usage rate). Thus, video feeds to a small, futuristic 100-subscriber Service Group might consume ~1.7 Gbps of bandwidth capacity.

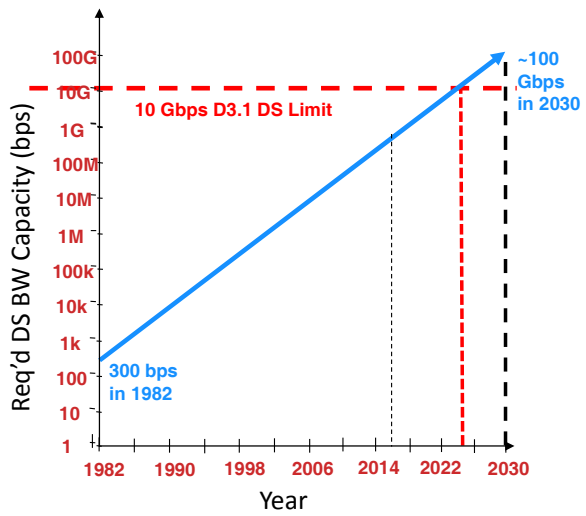


Figure 1- DS BW vs. Year

Broadband Services (with expected 50% annual growth rates) will also add an ever-growing demand of additional bandwidth capacity needs on top of that Video bandwidth capacity. If the Broadband growth rates are extrapolated into the future, the curve in Figure 1 is obtained (for a typical 100-subscriber Service Group of the future). The red, dashed horizontal line shows the 10-15 Gbps capacity level that may become a limiting factor in the future. This figure illustrates that the DOCSIS 3.1 bandwidth capacity limits may permit Broadband data to be supported on DOCSIS 3.1 networks until the mid-2020's.

But what follows that? Will the bandwidth capacity that subscribers require in the distant future (the late 2020's and beyond) be satisfied by the 10-15 Gbps offered by DOCSIS (as currently defined)? Apparently not.

Thus, is it time to look at replacement technologies for DOCSIS? Are there any other solutions to the problem of limited bandwidth capacity? These are all important questions that will be addressed in this paper.

POSSIBLE PATHS SUPPORTING THE FUTURE CAPACITY NEEDS OF BROADBAND

Overview

It is clear that the Cable Industry must plan for some changes if it is going to continue to support the growing Broadband capacity needs of subscribers into the deep future. There are two possible paths that could be followed.

The first path requires the MSO to invest heavily in a new Last-Mile Delivery infrastructure (and technology) that will ultimately replace the HFC network. The second path requires the MSO to continue incremental investments in the existing Last-Mile Delivery infrastructure and to find new ways to augment the capacity of the legacy HFC network.

The First Path (Revolutionary Change)

MSOs who choose the first path are likely to build a parallel network over the top of their existing network, giving them the opportunity to operate both networks in parallel during the transition period. In fact, many MSOs that follow this path may actually keep their legacy HFC network operational for decades to help provided extra bandwidth and to help support legacy in-home CPE equipment.

MSOs have many technology options to select from as they look at potential new delivery infrastructures. The list includes (but is not limited to) RF over Glass, Hybrid PON, EPON, GPON, and point-to-point Ethernet [EMME]. Interestingly, all of these technologies utilize a Fiber-To-The-Home (FTTH) architecture to eliminate the high-frequency attenuation that occurs within coaxial systems. While fiber does greatly reduce the attenuation issues, it does introduce other issues (such as dispersion) that must also be accounted for.

FTTH systems are oftentimes assumed to be the desired end-state technology for Last-Mile Delivery systems, and that is probably a correct assumption. However, FTTH systems do tend to have one undesirable trait (for many MSOs). In particular, FTTH systems can be expensive to initially install and deploy. Estimates vary and depend on many conditions, but the cost of installing new fiber over the last 150' to a typical home and equipping that home with a new ONU equipment can fall in the $500-$1000 range [EMME]. For many MSOs, the additional expense associated with adding a new FTTH connection to a home is not desirable and therefore other approaches must also be considered. A very promising approach is outlined in the following section.

The Second Path (Evolutionary Change)

The second path that MSOs can choose to follow as they expand their per-subscriber bandwidth capacity is a path that attempts to augment the capacity of the existing Last-Mile Delivery infrastructure. In a sense, this approach is not new. This is the "business-as-usual" approach that has been applied to the HFC network continuously for the past sixty years. It uses new technologies as they become available to continually push more and more bandwidth through the HFC system.

One of the key points that became apparent to the authors within another study [CLO1] is that MSOs are barely beginning to scratch the surface of the full potential bandwidth capacity of the HFC infrastructure. Both the fibers and the coaxial legs within the existing HFC network can carry much more bandwidth capacity than they are transporting today. So it begs the question- should MSOs build a costly FTTH system to replace their existing HFC network when smaller investments and augmentations within the

existing HFC network can be made to permit them to deliver much more bandwidth capacity to their subscribers in the future. For some MSOs, this evolutionary approach may be more desirable than the revolutionary approach required by FTTH deployments.

It is interesting to note that Telcos were facing a similar situation in the mid-1990's. Their existing twisted-pair connections to homes were viewed by many as having reached the end of their realistic lifespan, and many actually proposed FTTH as a technology to replace the twisted-pair connections. While some Telcos did deploy FTTH solution in some areas, many more Telcos actually found that it was more cost-effective to continue to use the existing 100-year-old twisted-pair technology. They found innovative ways to boost its performance level with DSL and (ultimately) G.Fast technologies. These approaches have permitted the Telcos to extend the life-span of their twisted-pair copper by at least 20 years, and twisted-pair is still used as a Last-Mile Broadband infrastructure by the Telco industry today [CLO1].

MSOs have an opportunity to follow a similar path that could greatly extend the lifespan of their existing HFC plant. In particular, they can decide to extend the spectrum of their DOCSIS network well beyond the Downstream RF upper band edge of 1.794 GHz that is specified in the current DOCSIS 3.1 specification [DPHY].

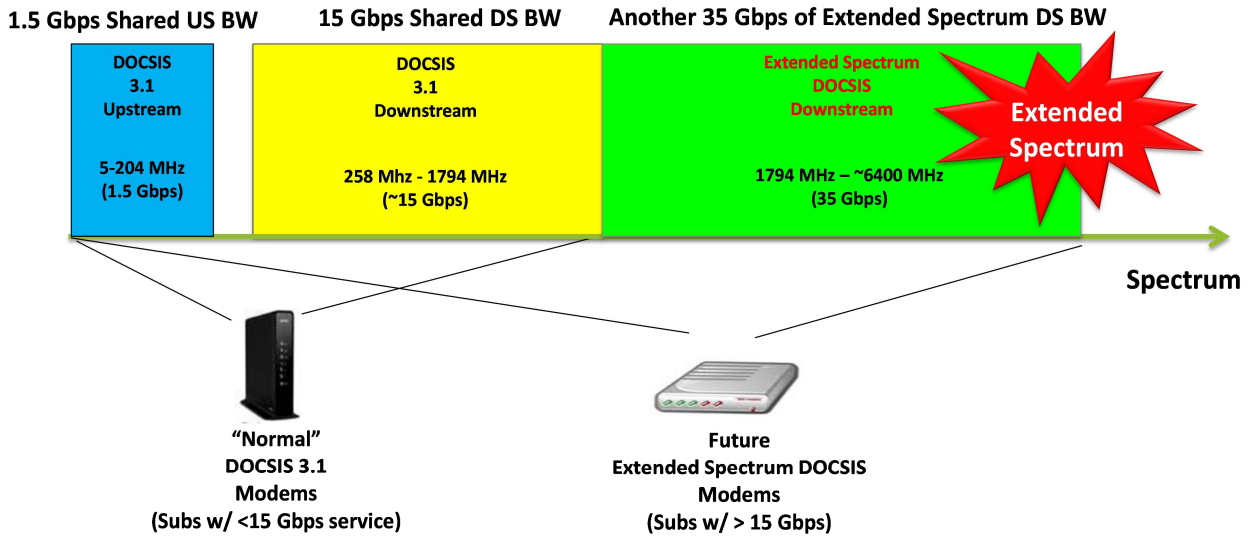


Figure 2- Extended Spectrum DOCSIS in Spectrum Offering ~50 Gbps in ~6 GHz

EXTENDED SPECTRUM DOCSIS

Basic Idea Behind Extended Spectrum DOCSIS

One approach focusing on the second path defined above is an approach known as "Extended Spectrum DOCSIS." The idea is quite simple. Rather than change out the entire HFC plant for a new technology to increase their bandwidth capacity levels, MSOs can instead choose to continue to use the HFC plant by extending the spectrum that supports DOCSIS 3.1 OFDM blocks beyond the 1.794 GHz limit that is specified in the DOCSIS 3.1 specification today (see Figure 2). The top frequency in the Extended Spectrum might be 3 GHz or 6 GHz or 12 GHz or higher. The actual top-end frequency and bandwidth capacity that can be utilized

will undoubtedly be function of the manner in which the signals are delivered to the home. (Several different techniques will be outlined below).

The idea is based on a simple application of the well-known Shannon-Hartley Theorem from the field of Information Theory [SHAN]. In particular, the Shannon-Hartley formula states that:

$$C = B*\log_2(1+S/N) \qquad (1)$$

Where C is the maximum channel bandwidth capacity (in bits per second), B is the channel frequency bandwidth (in Hertz), S is the average signal power level across the bandwidth (in Watts), and N is the average noise power level across the bandwidth (in Watts). In essence, the formula states that the channel capacity is the product of two terms. The first term is the bandwidth (B) of the channel- a wider bandwidth produces a higher channel capacity. The second term is $\log_2(1+S/N)$, which is closely related to the signal-to-noise ratio on the channel. One illustrative interpretation of the formula is that 1GHz of frequency bandwidth, with ~31dB SNR, has theoretical channel capacity of 10 Gbps!

If we assume that the signal power (S) launched into a channel remains fixed (due to transmitter power limitations or due to FCC power level limitations), then the expansion of channel bandwidth capacity (C) depends only on the B value and the N value within the above formula. For simplicity, we will assume that the noise power spectral density is a constant value across the entire channel bandwidth, given by N0 (in Watts per Hz). Thus, the noise power (N) is given by the simple relationship:

$$N = N0*B. \qquad (2)$$

If we assume that we have a relatively large signal-to-noise ratio (with S/(N0*B) >> 1), then (1) can re-written as:

$$C = B*\log_2(S/(N0*B))$$
$$C = B*[\log_2(S/N0) - \log_2(B)]$$
$$C = B*[K - \log_2(B)] \qquad (3)$$

Where K=log2(S/N0).

A typical doubling of the channel bandwidth from B to 2B will therefore result in a New C value (C2) given by:

$$C2 = 2B*[K - \log_2(2B)]$$
$$C2 = 2B*[K - \log_2(B) - \log_2(2)]$$
$$C2 = 2B*[K - \log_2(B) - 1]$$
$$C2 = 2B*[\log_2(S/N0) - \log_2(B) - 1]$$
$$C2 = 2B*[\log_2(S/(N0*B)) - 1]$$
$$C2 = 2B*[\log_2(S/N) - 1] \qquad (4)$$

In most real-world systems, the initial value of S/N is quite large (ex: SNR_{dB} = 40 dB implies S/N = 10,000), so $\log_2(S/N) = \log_2(10{,}000) = \sim 13.3$ is relatively large (compared to 1). As a result, even though the SNR drops by 3 dB, it is still high (i.e. $\log_2(S/N) >> 1$), and therefore the new channel capacity value (C2) is close to double the channel capacity that existed before the channel bandwidth B was doubled. Thus, repeating this B-doubling process over and over can lead to great increases in the overall channel capacity (given by successive calculations of the C2 value). In most cases, the doubling yields beneficial results until the resulting modulation profiles begin to approach BPSK levels. This is clearly illustrated in Figure 3.

B (GHz)	S/N	SNR(dB)	C = B*log2(1+S/N) (Gbps)
0.75	10000	40.00	9.97
1.5	5000	36.99	18.43
3	2500	33.98	33.86
6	1250	30.97	61.73
12	625	27.96	111.48
24	312.5	24.95	199.02

Figure 3- C vs. B

HFC Plant Topics For Extended Spectrum DOCSIS Systems

Extended Spectrum DOCSIS would require that several changes be made within the typical HFC network. Obviously, the HFC plant must have the ability to transmit and pass the higher-frequency signals from the head-end and across the fiber portion of the plant, through the fiber node, across the coaxial distribution leg of the plant, through amplifiers and taps, across the coaxial drop portion of the plant, through the coaxial in-home network, and to the modem in the home.

There are many ways to pass these Extended Spectrum DOCSIS signals through these elements. While it is theoretically possible for high-frequency amplifiers to be designed for passing these signals, the authors believe that most MSOs will likely prefer to wait until their HFC plants have been converted to Node+0 Fiber Deep architectures before Extended Spectrum DOCSIS will be considered as a desirable and feasible technology. (Note: This eliminates the cost of upgrading many amplifiers to the higher frequencies required for the Extended Spectrum network).

It seems fortuitous that many MSOs predict that they may be performing node splits down to Node+0 architectures before or around the same time that the 10-15 Gbps DOCSIS 3.1 systems will be "running out of gas." (Note: It is also possible for MSOs to get to Node+0 architectures even if they keep their service groups large). As seen in Figure 1, it is expected that the required bandwidth capacity of a typical Service Group will exceed the available DOCSIS 3.1 bandwidth capacity in the early- to mid-2020's. That is likely to be the time-frame when Extended Spectrum DOCSIS may prove to be valuable. It is also the time-frame when Moore's Law silicon improvements will likely permit Extended Spectrum DOCSIS systems to be deployed (assuming designs begin in the latter portion of the 2010 decade). This "perfect storm" of events implies that the early 2020's may be a perfect time to consider for initial deployments of Extended Spectrum DOCSIS systems.

Once an MSO begins to deploy Node+0 Fiber Deep architecture designs, many variants of Extended Spectrum DOCSIS systems can be envisioned.

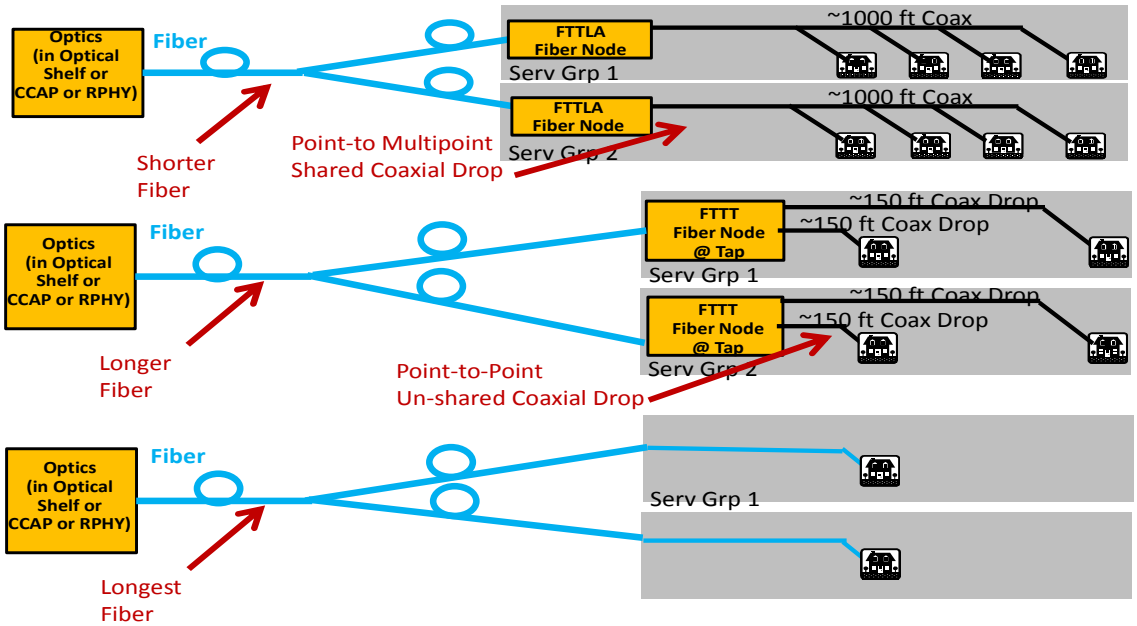


Figure 4- Variants of HFC Plant Configurations for Extended Spectrum DOCSIS Deilvery

Some MSOs may choose to have shorter fiber runs and longer coaxial runs, keeping their coaxial run lengths the same distances with Node+0 network architecture. This would result in 600' to 1250' coaxial runs (and the lowest bandwidth capacities), and we will designate these types of systems as Fiber-To-The-Last-Active (FTTLA) systems (since the fiber is essentially run to the same location where the last amplifiers used to be). These FTTLA system would have hard-line coax carrying the Extended Spectrum DOCSIS signal across up to six taps, with each tap having up to four subscribers connected to it.

Other MSOs may choose to have longer fiber runs and shorter coaxial runs, permitting their fiber to run all the way to the tap and requiring the coaxial run to only hop across the drop cable. This would result in 100' to 250' coaxial runs (and higher bandwidth capacities), and we will designate these types of systems as Fiber-To-The-Tap (FTTT) systems (since the fiber is essentially run to the tap). It is possible to design the Fiber Node within the FTTT system to have a single point-to-point coaxial connection from the Fiber Node's Extended Spectrum DOCSIS transceiver directly to the tap, and that design approach may greatly reduce signal attenuation and noise funneling effects. Alternatively, the Fiber Node can be at the tap location, which reduces the coaxial attenuation even more.

Still other MSOs may choose to have even longer fiber runs that go all the way to the home. This would result in zero-length coaxial runs (and the highest bandwidth capacities), and we will designate these types of systems as Fiber-To-The-Home (FTTH) or Fiber-To-The-Premise (FTTP) systems (since the fiber is run directly to the home). In essence, these types of systems are actually RF over Glass (RFoG) solutions. (Note: All of these variants are illustrated in Figure 4). A key potential benefit of these FTTH/FTTP systems is that they can use different lambdas for the Downstream path and the Upstream path, permitting much higher bandwidth capacities to be offered in both directions.

MSOs could even choose to have fiber lengths inbetween those lengths described in

the previous three paragraphs, so it should be apparent that many different design approaches to Extended Spectrum DOCSIS are permissible.

In general, longer coaxial runs tend to lead to lower bandwidth capacities due to the higher attenuations (which increase as a function of both distance and frequency). However, longer fiber runs can also reduce bandwidth capacity due to dispersive and nonlinear noise effects (coupling energy between lambdas). A compromise will need to be found to maximize overall throughput.

Tap Topics For Extended Spectrum DOCSIS Systems

Some of the architectures defined in the previous section require that the Extended Spectrum DOCSIS signals pass through taps. For example, the top-most architecture in Figure 4 would require the taps to be updated.

As a result, it is a given that new taps will need to be designed and deployed to carry the higher-bandwidth signals. This is still an area of research that needs to be performed before Extended Spectrum DOCSIS systems of that nature become practical.

Optical Beat Interference Topics For Extended Spectrum DOCSIS Systems

One of the obvious side-effects of Fiber Deep systems (which are pre-requisites to Extended Spectrum DOCSIS deployments) is the fact that there will be many Fiber Nodes to manage and feed.

For MSOs, the delivery of signals to and from these many Fiber Nodes may require the use of wavelength-division multiplexing on the fiber or other fiber sharing techniques to minimize the number of unique physical fibers that must be pulled out to the Fiber Nodes. RF over Glass is one of the other popular fiber sharing techniques that permits multiple Fiber Nodes to share a lambda on the fiber. It is therefore quite possible that MSOs may choose to use RF over Glass (RFoG) techniques to deliver the signals from multiple Fiber Nodes up to the head-end. [MUTA]

One of the obstacles to any Upstream Multi-point-to-point optical system (including Upstream RFoG) has been the existence of Optical Beat Interference (OBI). OBI is a dramatic form of signal degradation that occurs whenever two or more optical transmitters simultaneously transmit with closely-spaced optical wavelengths.

OBI problems are exacerbated whenever the number of simultaneous upstream transmitters is increased. This can occur with increases in the number of channels (as with DOCSIS 3.0) or simultaneous transmitted codewords (as with DOCSIS 3.1). Thus, Extended Spectrum DOCSIS will only serve to exacerbate this problem by creating more bandwidth that is shared by the modems. As a result, good solutions to the OBI problem will be required if MSOs attempt to utilize RFoG techniques to combine signals from multiple Fiber Nodes in an Extended Spectrum DOCSIS system.

Two powerful approaches to OBI mitigation already exist. These include planned wavelength separation techniques and Hybrid PON techniques [MUTA]. With solutions like these in hand, MSOs should be able to roll out their Extended Spectrum DOCSIS systems and utilize RFoG techniques to connect to the many Fiber Nodes.

Distributed Access Architecture Topics For Extended Spectrum DOCSIS Systems

Extended Spectrum DOCSIS offers a very flexible design. Due to this flexibility, it should be able to work very well with traditional Amplitude Modulated Optical signals being carried over the fiber portion of the HFC network. (Note: This is the type of

solution that would be used within the RFoG solutions described in the previous section).

However, nonlinear optical effects resulting from interactions between lambdas on any lengthy wavelength-division multiplexed fiber may reduce the SNR values and reduce the throughput of the Extended Spectrum DOCSIS system.

As a result, MSOs may alternatively choose to use Distributed Access Architectures (DAAs) to deliver the signals over the fiber portion of the HFC network. With DAAs, the fiber carries digital optics (Ethernet or xPON signals) from the head-end to the Fiber Node, and the Fiber Node produces the Amplitude Modulated signal that is ultimately transmitted over the coaxial portion of the HFC network. [CLO2]

Two different variants of DAA architectures are being considered by MSOs- Remote PHY architectures and Remote MACPHY architectures.

Extended Spectrum DOCSIS systems could be built using either of these DAA variants. In both cases, there are several benefits that would result. First, the use of a DAA approach would ensure that the SNR of the signals would not be significantly reduced by the nonlinearities within the fiber portion of the HFC network. Second, the use of DAA systems would help to reduce the power and space requirements in MSO head-ends when fiber deep solutions have created the need to support many small Service Groups. Third, the use of DAA systems (and digital optics) would also permit MSOs to place more lambdas on their wavelength-division multiplexed fibers. Fourth, the use of DAA systems (and digital optics) eliminates the presence of OBI within the digital fiber (since OBI only occurs in Amplitude Modulated optical systems).

Because of all of these reasons, MSOs who move to Extended Spectrum DOCSIS will have the option to use either centralized or distributed access architectures.

Spectrum Topics For Extended Spectrum DOCSIS Systems

With Node+0 systems in place, MSO networks will be ready to carry Extended Spectrum DOCSIS signals. The signals themselves may take the form of a stack of 192 MHz OFDM Downstream blocks (or a stack of 96 MHz OFDMA Upstream blocks) that inhabit regions of the spectrum beyond 1794 MHz.

The actual amount of spectrum that might be useable for this stack of OFDM/OFDMA blocks is a function of many parameters, including the signal launch power, the noise power injected on the fiber, the length of the fiber, the number of lambdas on the fiber, the wavelengths of the multiplexed lambdas, the noise power injected on the coax, the length of the coax, the attenuation of the coax, the amount of loss in taps, the amount of loss in splitters, modem receiver noise figure, etc.

Thus, depending on the design of the Node+0 system, different spectral widths will be allowed to carry Extended Spectrum DOCSIS signals. Thus, it is possible that Extended Spectrum DOCSIS systems of various flavors could be deployed differently by different MSOs. Some may choose to limit Extended Spectrum DOCSIS operation to ~2.5 GHz. Others may choose to limit Extended Spectrum DOCSIS operation to ~7 GHz. Still others may choose to push Extended Spectrum DOCSIS operation all the way to ~10-25 GHz. In each case, the HFC plant must be appropriately conditioned (creating deeper fiber runs and shorter coaxial runs) to guarantee successful OFDM and OFDMA operation.

As an example, consider a system with coaxial lengths that are short enough to permit ~6.4 GHz of spectrum to be passed. If an Extended Spectrum DOCSIS system using ~6.4 GHz of spectrum were deployed, then it is possible that the system might place DOCSIS Downstream spectrum from 258 MHz to 6402 MHz. This spectrum is large enough to hold thirty-two 192 MHz OFDM blocks. It is expected that channel bonding can be used across all of the 32 OFDM blocks. This is illustrated in Figure 5. (Note: It is also possible that larger OFDM blocks could be defined in a future Extended Spectrum DOCSIS specification. This would lead to the need for less OFDM blocks in the spectrum and less use of DOCSIS channel bonding).

The spectrum in Figure 5 is quite asymmetrical, with much more Downstream bandwidth capacity than Upstream bandwidth capacity. The spectrum could also be divided in a different way.

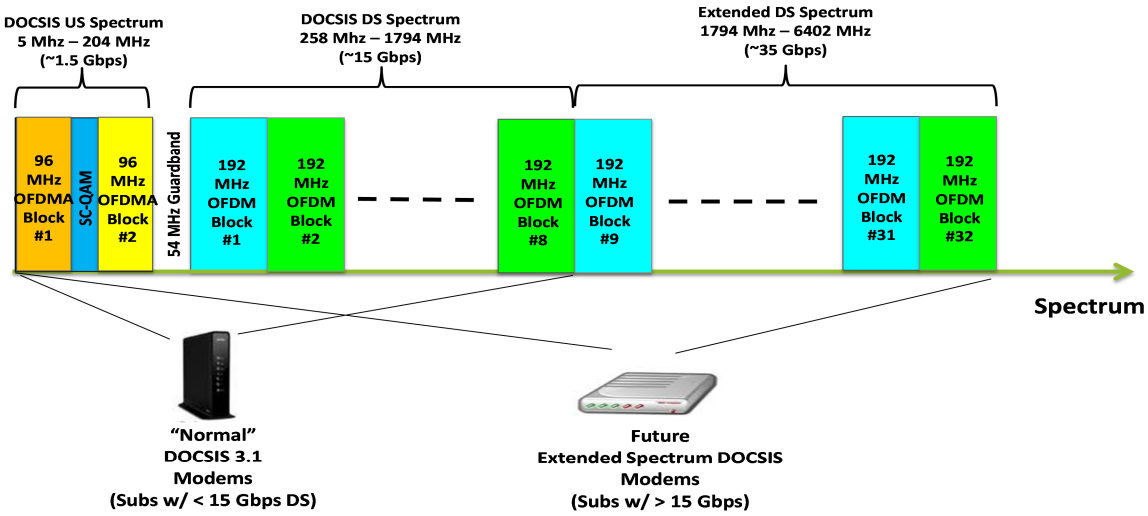


Figure 5- Asymmetrical Extended Spectrum DOCSIS in 2-band FDD Spectrum Offering ~ 50 Gbps DS x ~1.5 Gbps in ~6.4 GHz

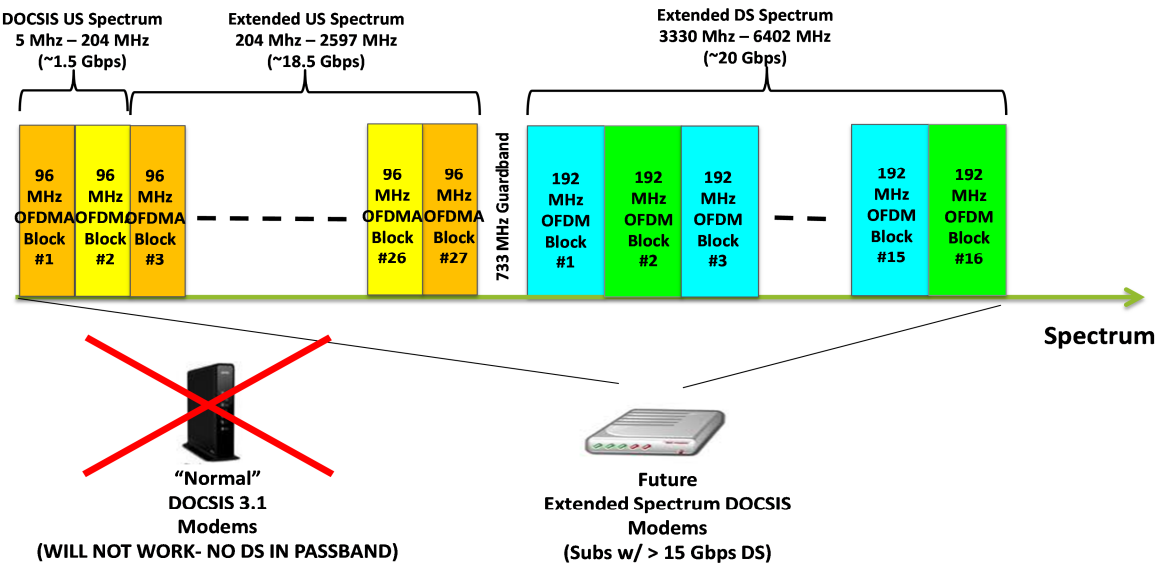


Figure 6- symmetrical Extended Spectrum DOCSIS in 2-band FDD Spectrum Offering ~20 Gbps x ~ 20 Gbps US in ~6.4 GHz

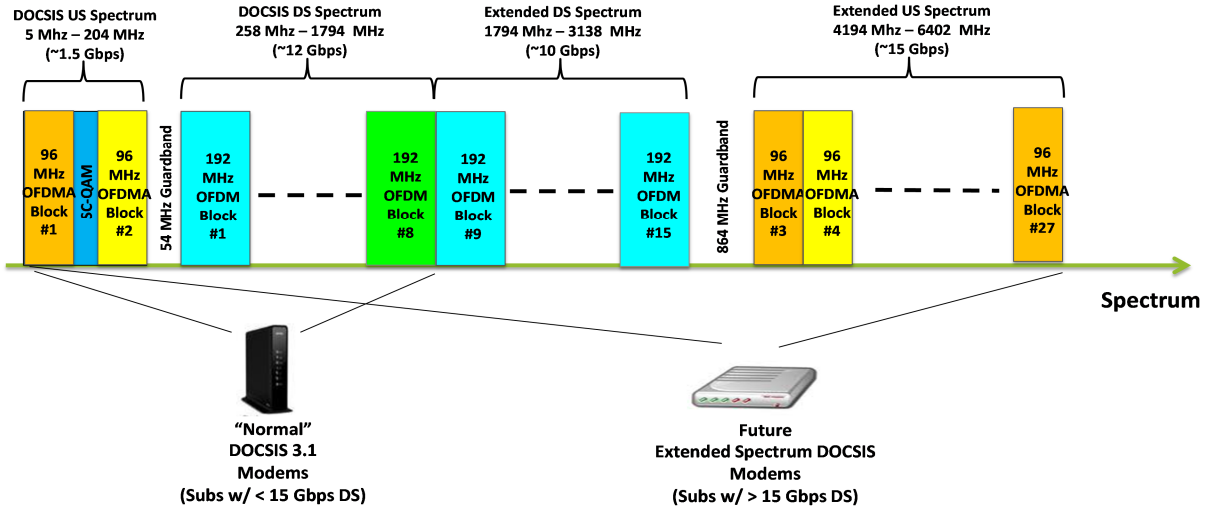


Figure 7- Symmetrical Extended Spectrum DOCSIS in 3-band FDD Spectrum Offering ~22Gbps DS x ~16.5 Gbps US in 6.4 GHz

For example, a more symmetrical version of Extended Spectrum DOCSIS could also be configured as shown in Figure 6, where the Upstream spectrum ranges from 5 to 2597 MHz and the Downstream spectrum ranges from 3330 to 6402 MHz. This requires a larger 733 MHz guardband between the two directional spectra, and that leads to slightly lower overall bandwidth capacities. (Note: It may be possible to eliminate the guardbands in these designs, which is a topic that is still under study within the Full-Duplex DOCSIS effort [FDX] among MSOs, vendors and CableLabs). Note that this particular configuration, shown in Figure 6, will not permit DOCSIS 3.1 modems to operate, because there is no Downstream spectrum in

the "normal" Downstream portion of the spectrum where DOCSIS 3.1 operates.

Another symmetrical version of Extended Spectrum DOCSIS is shown in Figure 7. This particular solution has the benefit of supporting symmetrical bandwidth capacities (with similar bandwidth capacities in the Upstream and Downstream directions) while also permitting DOCSIS 3.1 modems to operate in the system. The disadvantage is that three bands (two Upstream bands and one Downstream band) must be utilized, and this leads to even more guardband being added to the system. Furthermore, the limited upstream transmit power along with the high attenuation at the upper frequencies will cause the overall bandwidth capacity to be further reduced.

Full Duplex And Half Duplex Topics For Extended Spectrum DOCSIS Systems

MSOs are continually being pushed by subscribers and competitors to increase their bandwidth capacities. In recent years, xPON service providers have begun challenging MSOs and are beginning to push MSOs to increase their DOCSIS 3.1 Upstream bandwidth capacities. In response to these challenges, MSOs and their vendors are exploring many ways to provide more bandwidth capacity. As an example, there are currently very active studies under way to identify powerful, new techniques for operating the DOCSIS 3.1 Upstream channels on the same frequencies as the DOCSIS 3.1 Downstream channels. At least two different approaches for frequency spectrum sharing are being considered.

One of the frequency spectrum sharing approaches is based on Full Duplex operation, whereby the Upstream and Downstream signals occupy the same coax and the same portion of the spectrum at the same time. The signals essentially pass right through one another, and the receivers must detect and demodulate the arriving signal in the presence of the interference or "noise" from the signal propagating in the opposite direction. This requires the use of noise cancellation techniques in the CCAP Upstream Receiver. There must also be ways to circumvent noise at the modem Downstream Receiver, because non-ideal isolation in taps can permit energy from Upstream transmissions from a modem to couple into the Downstream spectrum on nearby, neighbor modems (see Figure 8). While solutions to these problems have some challenges and are still being studied, it is quite possible that some form of Full Duplex DOCSIS [FDX] may be operating on DOCSIS 3.1 systems in the next few years.

It should be noted that an extended spectrum DOCSIS system based on FTTT will likely result in a point-to-point coaxial connection between the fiber node, which could be at the tap location, and the modem. Therefore, interference between neighboring modems in a Full Duplex DOCSIS system is essentially eliminated. This may lead to many simplifications as well as performance enhancements within Full Duplex DOCSIS systems because echo cancellation can be added at both ends.

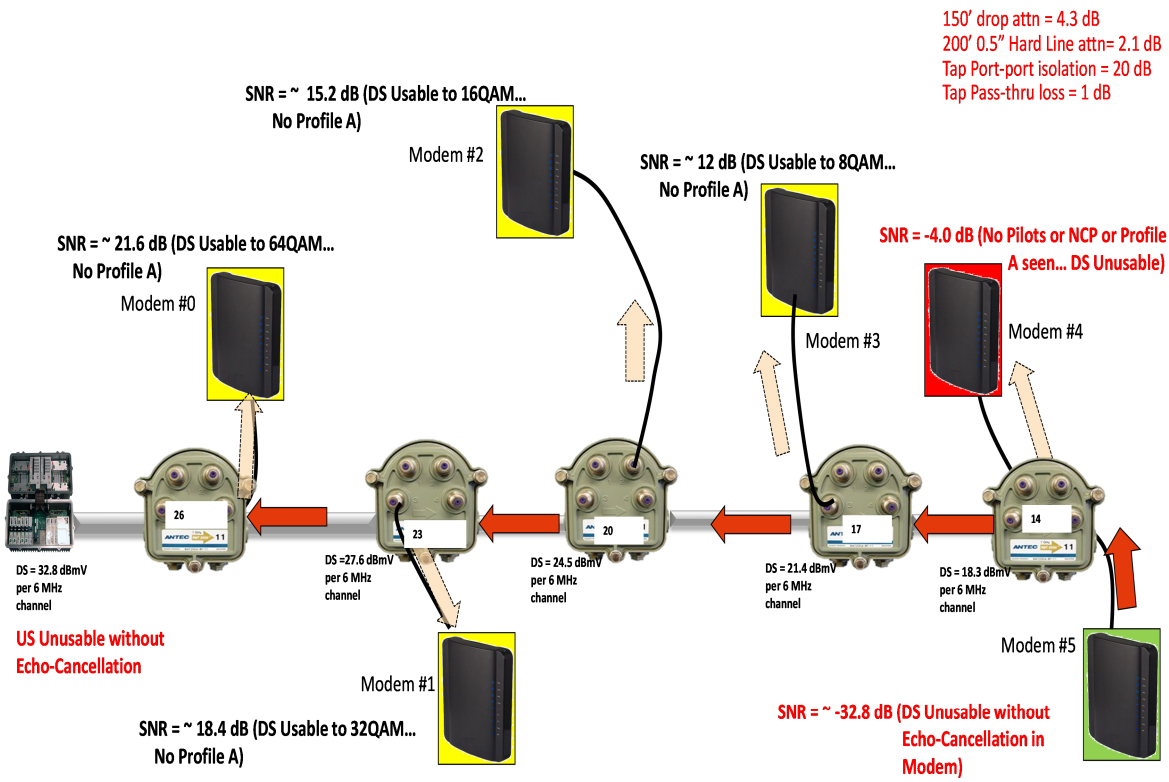


Figure 8- SNRs in Modems due to Upstream Transmission from Neighbor (Modem #5)

The second frequency spectrum sharing approach is based on Half Duplex TDD operation, whereby the Upstream and Downstream signals occupy the same coax and the same portion of the spectrum, but they share the resources in time by allowing only Upstream transmissions at one moment in time and by allowing only Downstream transmissions at a different moment in time. This requires well-synchronized MAP generation for Upstream transmissions and scheduling for Downstream transmissions. Solutions to these scheduling problems have some challenges and are still being studied, but it is also quite possible that some form of Half Duplex DOCSIS may be operating on DOCSIS 3.1 systems in the next few years.

If either Full Duplex or Half Duplex DOCSIS operation becomes a reality for DOCSIS 3.1, it is likely that it would also be used in an Extended Spectrum DOCSIS environment as well. Fortunately, the similarities between DOCSIS 3.1 and Extended Spectrum DOCSIS should permit any successful Full Duplex or Half Duplex approach to be easily applied to Extended Spectrum DOCSIS systems.

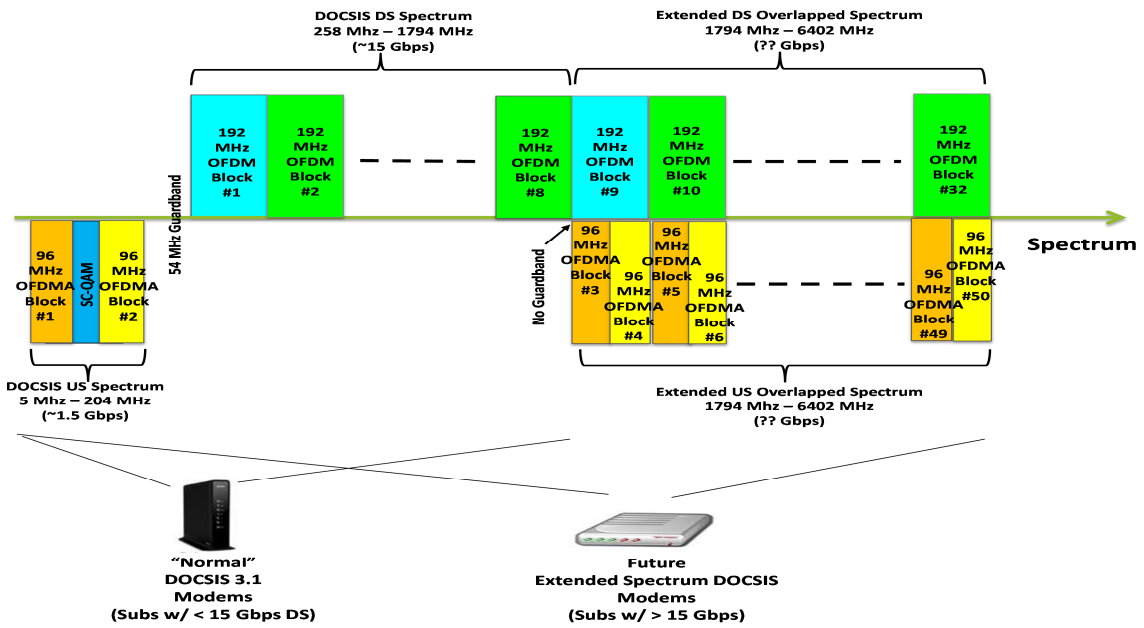


Figure 9- Asymmetrical Extended Spectrum DOCSIS w/ Shared Spectrum in ~6.4 Ghz

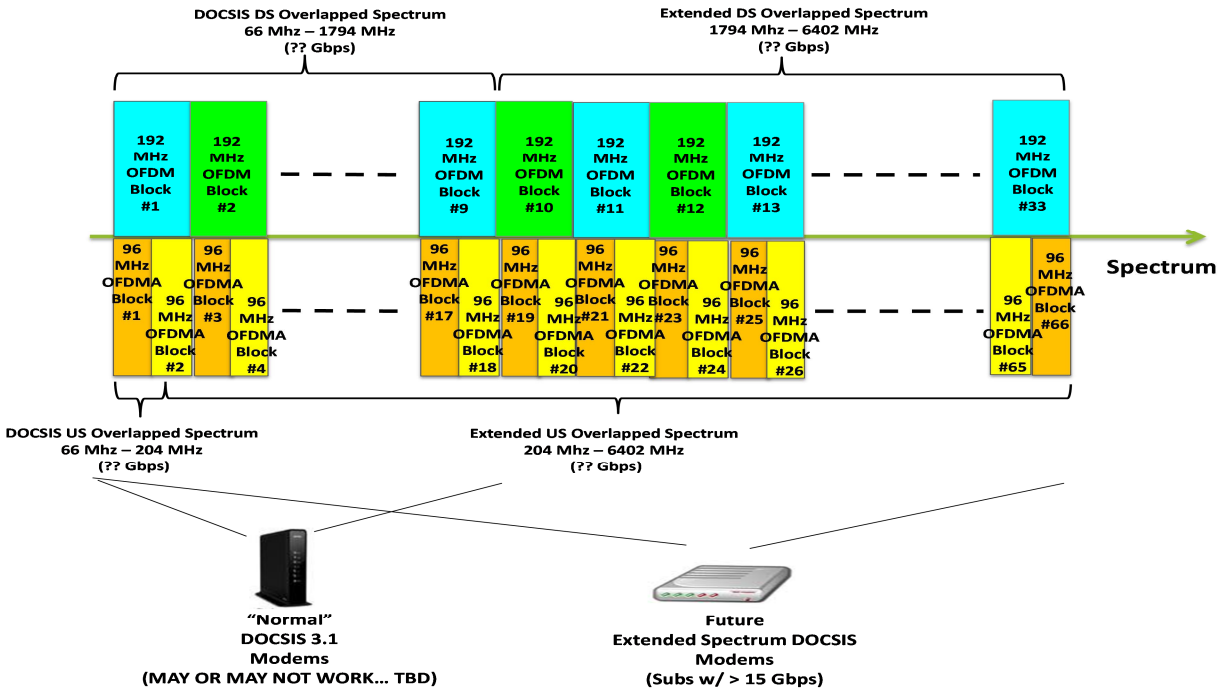


Figure 10- Symmetrical Extended Spectrum DOCSIS w/ Shared Spectrum in ~6.4 Ghz

As an example, Figure 9 and Figure 10 show examples of spectral maps with overlapping Upstream OFDMA blocks and Downstream OFDM blocks. Figure 9 shows a map that may permit DOCSIS 3.1 modems to interoperate on the spectrum with Extended Spectrum DOCSIS modems, whereas Figure 10 shows a map that attempts to maximally

fill the spectrum with Extended Spectrum DOCSIS operation (which likely precludes the use of DOCSIS 3.1 modems on the spectrum). These spectral maps can be used for either Full Duplex or Half Duplex operation.

OFDM/OFDMA Topics For Extended Spectrum DOCSIS Systems

It should be clear that Extended Spectrum DOCSIS is leaning heavily on the good work already done within the DOCSIS 3.1 specification committee. At a fundamental level, Extended Spectrum DOCSIS basically proposes to extend the use of Downstream OFDM and Upstream OFDMA techniques to other portions of the RF spectrum.

There are several reasons why the authors believe that this is a good idea. First, it permits Extended Spectrum DOCSIS to be created without a lot of new specification work. But more importantly, the authors believe that there may be important benefits that result from the use of DOCSIS 3.1's OFDM and OFDMA techniques. In particular, it may greatly aid in the passing of the high-frequency signals across both the fiber portion of the HFC network and the coaxial portion of the HFC network.

For transmissions across the fiber portion of the HFC network, one of the biggest challenges for Extended Spectrum DOCSIS will be to find ways to overcome the signal-distorting dispersion effects (resulting from varying propagation speeds for signals with different frequencies on the fiber). The impact of this dispersion on the OFDM signals is still being studied, but there may be some interesting results that may be found.

For transmissions across the coaxial portion of the HFC network, one of the biggest challenges for Extended Spectrum DOCSIS will be to find ways to overcome the large attenuation that will occur at the high frequencies (due to skin effect issues within the coax). Fortunately, OFDM can again be used to help circumvent this problem, because it permits the use of the variable bitloading feature. This feature permits the system to use lower modulation orders for high-frequency regions of the RF spectrum where attenuation leads to reduced signal powers and therefore reduced SNRs.

As a result, it is quite possible that the use of OFDM for Extended Spectrum DOCISS may greatly help with the transmission of the high-frequency signals in the resulting systems.

EXTENDED SPECTRUM DOCSIS COAXIAL SIMULATION RESULTS

As mentioned in the previous section, carefully-optimized variable bitloaded profiles for the OFDM signals will undoubtedly help with the transmission of the high-frequency Extended Spectrum DOCSIS signals across the coax.

The authors have begun exploring the performance levels of bitloading the Extended Spectrum DOCSIS signals. In particular, simulations of two possible configurations of the future were created: an FTTLA (Fiber to the Last Active) system and an FTTT (Fiber to the Tap) system. In each case, an attempt was made to identify the performance limits that impact the optical link SNR. Several simplifying assumptions were made. These include:

1. The only sources of noise are the noise at the output of the last amplifier (expressed through the source SNR), the Johnson-Nyquist thermal noise floor at the CPE and the CPE input circuitry's own noise contribution (expressed as the CPE noise figure). We explicitly excluded external noise components due to damaged cables or poor in-home wiring because the task at

hand is to explore the limiting factors in the capacity of the HFC plant itself.

2. The total RF power at the output of the Fiber Node is maintained at a fairly reasonable level of (say) 68 dBmV, although a range of values are examined.
3. The distribution cable between the last active component and the taps is 0.625" hard line. Its attenuation is calculated over an extended frequency range (beyond the normal recommended frequency range of 0-3 GHz) using standard formulae for resistive loss in the conductors and tan-δ loss in the dielectric.
4. The coaxial drop cable is assumed to be RG-6. Its attenuation is similarly calculated over an extended frequency range using the same approach as was used for the distribution cable.
5. The CPE's demodulator has an implementation margin of 3 dB at high SNR values. Practical demodulators will generally not achieve theoretical performance for a variety of reasons (e.g. quantization noise, limited number of iterations in LDPC decoders, etc.); this margin ensures that the demodulator's MER exceeds the theoretical minimum MER required to decode a signal with particular modulation depth.
6. The total amount of PHY and MAC overhead is assumed to be 20%. PHY overhead includes the loss of capacity associated with guard bands, pilots, signaling channels, LDPC FEC, and cyclic prefixes. The MAC overhead is associated with framing overheads and MAC messaging.
7. For half-duplex systems, the calculated capacity is the total bitrate in either direction; it may be split in any ratio between upstream and downstream traffic.
8. For full-duplex systems, the calculated capacity is the bitrate in one direction; the total capacity may be double this value for true full-duplex operation.
9. All possible QAM modulation orders (and associated bit-loading values) are used, down to BPSK; this differs from the current DOCSIS 3.1 standard, which uses a minimum of 16-QAM modulation for data. However, the specifications can be augmented with lower QAM orders to support the extended spectrum feature.

Given the above assumptions, the simulation explored the impact of various system parameters:

1. The effect of total transmitted RF power on the system capacity is examined.
2. The transmitted RF power may be distributed over the occupied spectrum either in a uniform manner (i.e. a flat Power Spectral Density) or with a tilt being applied. Both of these cases are examined, and the optimum PSD that maximizes the capacity is determined.
3. The effect of CPE input stage quality, i.e. the Noise Figure (NF) of the CPE's input circuitry in front of the demodulator, is explored.
4. The effect of source quality, i.e. the available SNR at the output of the last amplifier as signals transit over the optical link and previous amplifiers, is explored. In the remainder of this article, we use the term "source SNR" or "amplifier SNR" to refer to this parameter.
5. Different coax drop lengths are used in the simulation to examine the impact of this parameter on throughput.

The available RF bandwidth is assumed to be a little less than the cutoff frequency of the hard-line coax (for FTTLA) or RG-6 cable (for FTTT), with a somewhat arbitrary absolute maximum of 25 GHz. The relevant cutoff frequencies are 8.05 GHz for 0.625" hardline and 28 GHz for RG-6. For FTTLA, we consider signals up to 7 GHz; for FTTT, we consider signals up to 25 GHz. Multiple simulation runs, corresponding to different spectrum sizes, were performed to determine the optimal spectrum size that maximizes the throughput. For each of these runs, the total

power was divided across the spectrum supported in that case.

Note that, while we use the downstream terminology in the remainder of this paper (i.e., the "source" is the last amplifier, the demodulator is at the CPE) the analysis applies equally to the upstream direction, where the source would be the CPE and the demodulator would be either at the node (DAA) or the headend (CAA).

Extended-Spectrum System Capacity for FTTLA Systems

Using the assumptions listed above and the architecture shown in Figure 8, the resulting received signal SNRs, supported modulation levels, and bandwidth capacities can all be calculated for different spectrum sizes and different distances of typical RG-6 drop cables. The resulting bandwidth capacity is shown in Figure 11 and the corresponding bit loading values (in bits per symbol) are also shown in Figure 12, where the QAM order can be calculated from the bit loading values using the following formula:

$$\text{QAM Order} = 2^{(\text{bit loading})}. \qquad (5)$$

For this first-stage analysis, the following parameters have been used:

1. Total transmitted power of 68 dBmV (equivalent to 100 mW).
2. Amplifier SNR of 40 dB.
3. CPE NF of 10 dB.
4. No tilt is applied.

Variations in these parameters will be explored later in this paper.

From these figures, it can be seen that there is a definite bandwidth capacity benefit associated with the extension of the spectrum. This benefit is realized until the bit loading drops to a value of 1 (which corresponds to BPSK modulation).

It is also evident that the peak capacity is achieved at a particular spectrum size which is less than the maximum spectrum supported by the cable, but more than the current DOCSIS 3.1 limit of 1.794 GHz. This seems to contradict the capacity trend that was described in Figure 3, where increasing the spectrum always leads to an increase in capacity. However, this is only true if the SNR is high. For example, at the typical drop length of 150 ft, the peak capacity of ~27 Gbps is seen with a spectrum size of ~4.5 GHz. This behavior can be explained as follows: as the spectrum increases, more transmitted power is diverted to higher frequencies. Initially, as the spectrum size is small, the SNR stays high as we increase the spectrum, leading to increased overall capacity. However, if the spectrum increases further beyond a certain point, power is distributed to frequencies where the attenuation is so great that no modulation can be supported. This lost power detracts from the throughput at lower frequencies, reducing the overall capacity.

As one would expect, the capacity decreases with increasing drop length. Ironically, the optimal spectrum size for the longest drop lengths of 400 ft and 500 ft, is equal to, or slightly higher than, the maximum supported DS frequency in the DOCSIS 3.1 specifications (i.e. 1.794 GHz).

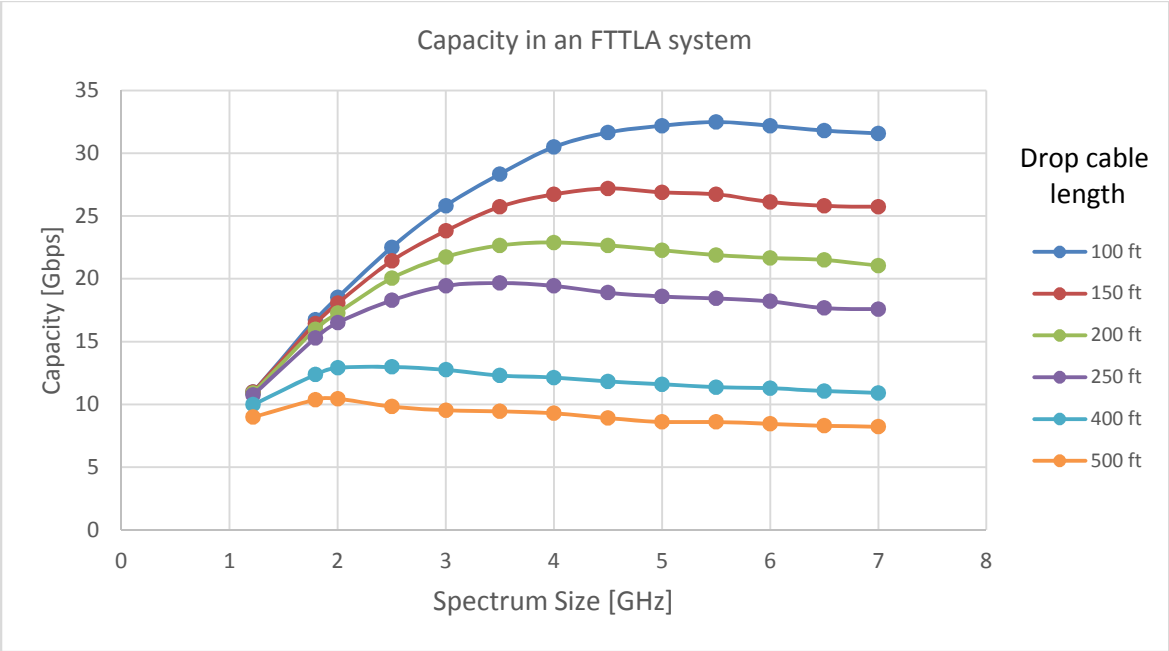


Figure 11. Capacity vs. Spectrum Size in an FTTLA system

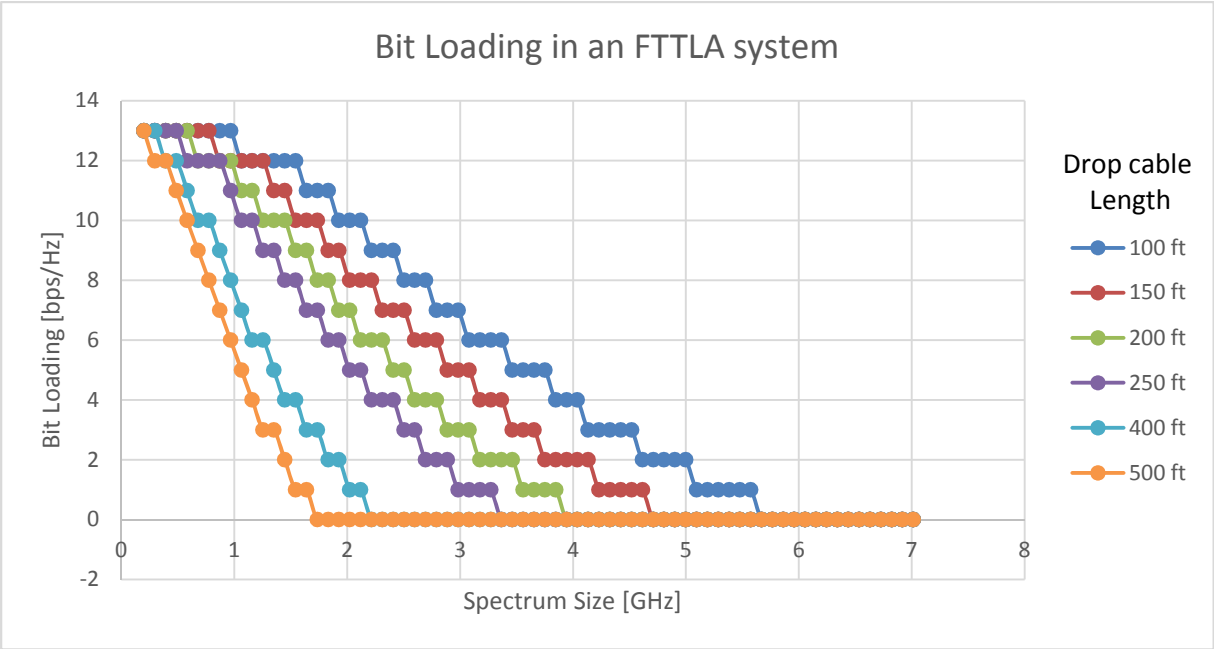


Figure 12. Bit loading vs. Spectrum Size in an FTTLA system

Extended-Spectrum System Capacity for FTTT Systems

The simulations of the FTTLA system above indicated a potential capacity increase of 2x to 3x relative to the capacities promised by the current DOCSIS specifications. As MSOs go deeper in fiber, more capacity can be offered. For example, in an FTTT system, hardline and tap losses are eliminated; only the RG-6 drop line attenuation remains. Additionally, the drop cable, due to its smaller geometry, has a much higher cutoff frequency. Admittedly, its attenuation per unit length is higher than that of the hard-line, but this is compensated by the shorter run lengths.

Figure 13 below shows the throughput vs. spectrum width for FTTT systems with various drop lengths. Moreover, Figure 14 shows the corresponding bit loading levels. It is compelling to note the potential high throughputs that are achievable across a point-to-point RG-6 coax carrying DOCSIS 3.1 OFDM signals. As an example, a 500 ft. length of RG-6 carrying a 5 GHz-wide OFDM spectrum would be capable of transmitting ~30 Gbps of throughput from the Fiber Node at the tap location to the home. On the other hand, a 150 ft drop cable can offer about 200 Gbps with 25 GHz spectrum. While second-order effects would likely reduce these bandwidth capacities to some extent, the message is clear; support for relatively large throughputs is likely to be possible on reasonable lengths of drop cables into the homes.

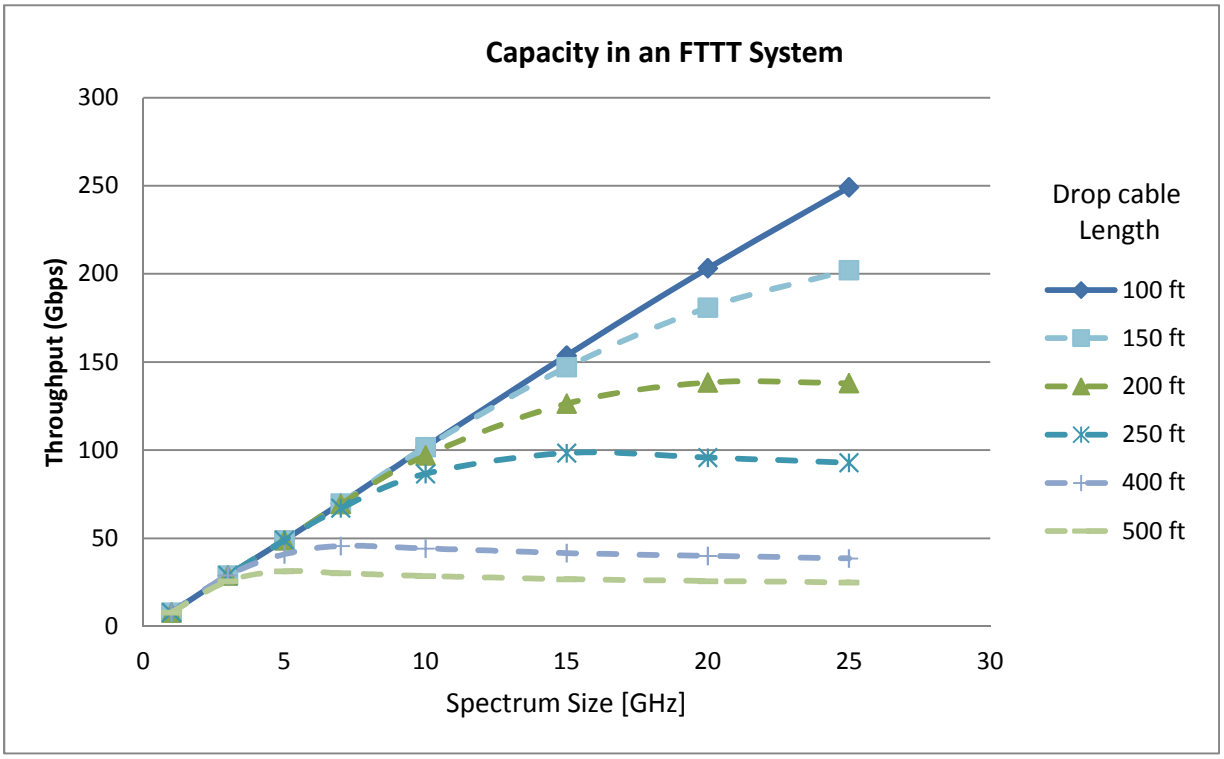


Figure 13. Capacity vs. Spectrum Size in an FTTT system

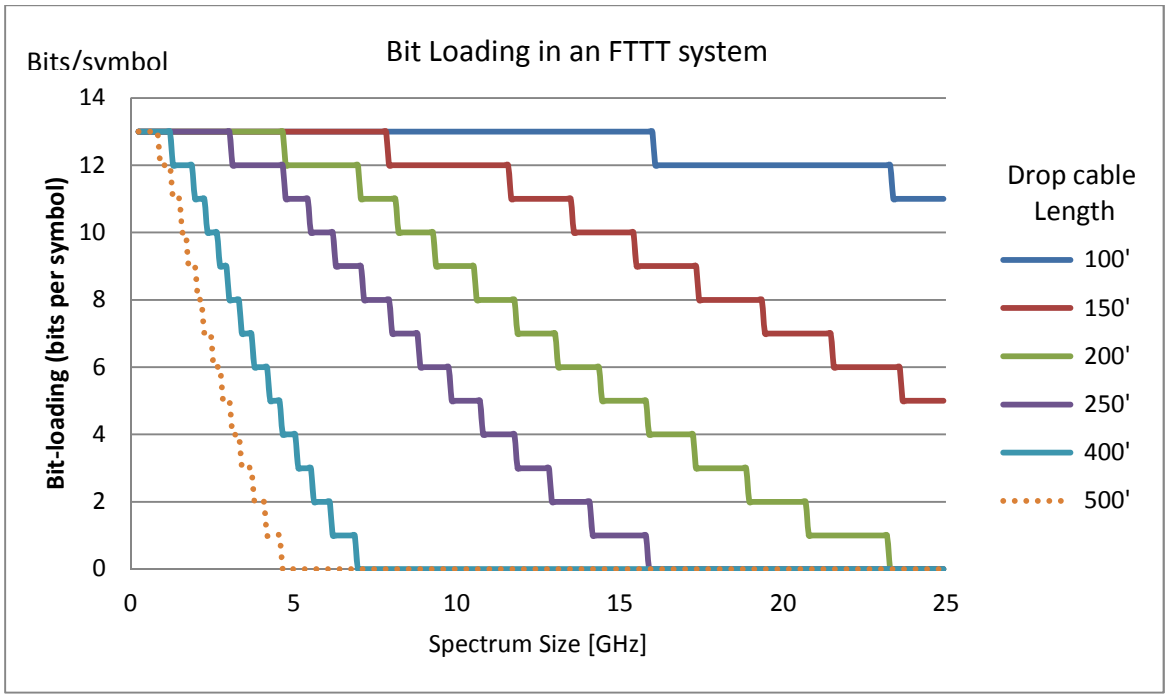


Figure 14. Bit Loading vs. Spectrum Size in an FTTT system

<u>The Effect of Total Transmitted Power on Extended Spectrum Capacity</u>

The simulations above were performed with a total power setting of 68 dBmV (100 mW), which is a typical value for existing nodes. However, higher power levels (e.g. 74 dBmV or 400 mW) have been used in some circumstances, at the cost of higher power consumption in the RF amplifier. On the other hand, lower power amplifiers might well be useful in order to facilitate a low-cost, low power design; this is particularly true of the upstream, where CPE power consumption is likely to be constrained for environmental reasons. Therefore it is useful to examine the effects of different RF power levels on system capacity.

Figure 15 shows the effect of changing the total RF power in an FTTLA system. Three scenarios are studied: one for a "high fidelity" scenario, with an amplifier SNR of 45 dB and a CPE NF of 5 dB; the second is a "mid-fidelity" scenario, with an amplifier SNR of 40 dB and a CPE NF of 10 dB; the third is a "low-fidelity" scenario, with an amplifier SNR of 35 dB and a CPE NF of 15 dB. All other parameters were maintained the same as in the previous simulations. In each case, the peak capacity of the system was calculated at a particular transmitted RF power. Figure 16 shows the equivalent result in an FTTT system.

It can be observed from these plots that the effect of transmitted RF power (which is broadly proportional to amplifier power consumption) on capacity is relatively small. A 3 dB reduction in RF power only produces ~10% reduction in capacity. This is one of the tradeoffs that would need to be considered in designing extended-spectrum equipment.

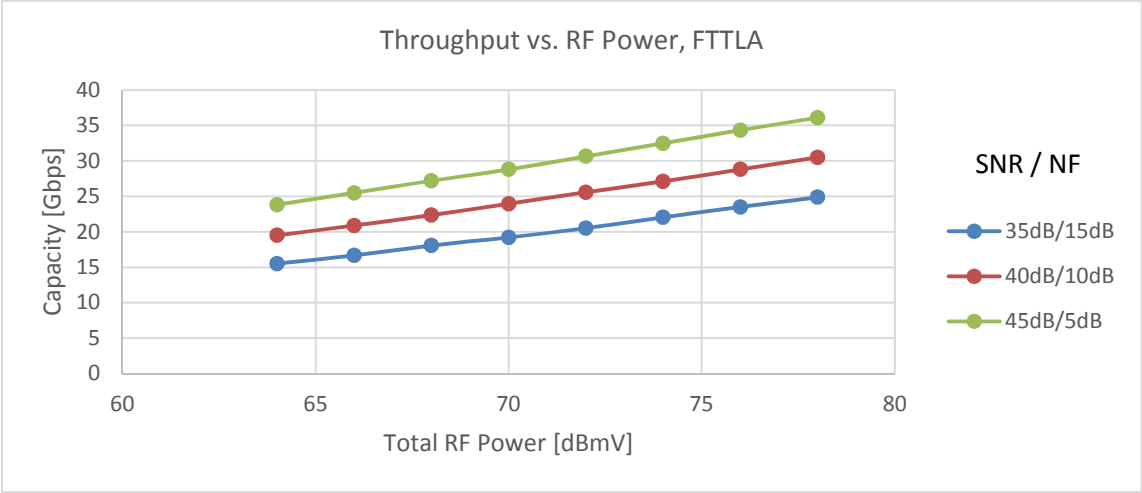


Figure 15. Capacity vs. Amplifier Output Power in an FTTLA system

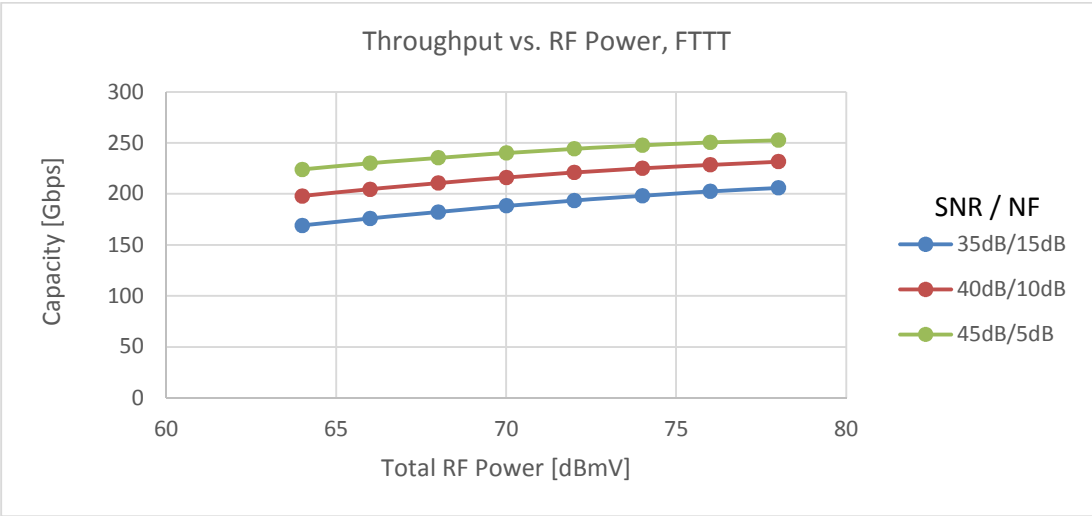


Figure 16. Capacity vs. Amplifier Output Power in an FTTT system

<u>The Effect of Source SNR and CPE Noise Figure on Extended Spectrum Capacity</u>

Regarding source SNR, it is clear that the current generation of CMTS and CCAP downstream modulators can comfortably generate full-spectrum (1.0 or 1.2 GHz) downstream signals with SNR and MER values exceeding 50 dB; DAA devices (Remote PHY, Remote MAC-PHY) are likely to have similar signal quality. The question is whether extended-spectrum DOCSIS equipment could be expected to have similar performance over the full frequency range.

While DAC speeds have increased dramatically over the last decade, and will probably continue to increase, it may be too optimistic to assume that direct-RF DACs generating signals at these high frequencies

(i.e. > 20 GHz) with DRFI-type fidelity and reasonably low cost and power consumption to be available in the 2020-2030 decade. While such devices may become available if the recent trends continue, our analysis will not make this assumption. If such equipment is not available in the time frame when extended spectrum devices are needed, then various techniques can be utilized to achieve the required performance over a full-spectrum of 7 GHz or 25 GHz, needed for FTTLA or FTTT, respectively. For example, multiple DACs in a single package could feed overlapping upconverters with portions of the required full-bandwidth signal; the DOCSIS 3.1 standard lends itself very well to this kind of technique, where multiple synchronous OFDM signals can be supported without guard bands between them. However, one would expect some reduction in signal quality in these conversion stages. Other techniques can be used but they would presumably carry similar penalties. Therefore, the effect of lower SNR values was included in our analysis, where the range of 35 dB to 45 dB is considered.

In Fig. 17, the CPE noise figure, which is largely determined by cost considerations, is investigated because it is difficult to predict the scale of improvements in the performance of electronic devices over the next 10-15 years. In particular, a NF of 5 dB might be expensive to achieve in a CPE, whereas a NF of 15 dB should be achievable relatively cheaply. The analysis here was conducted with NF values that range from 5-15 dB.

Figure 17 shows the effect of varying source SNR on system capacity in a FTTLA system for several values of CPE NF. Figure 18 shows equivalent plots for an FTTT system. It can be seen that a 3 dB change in NF changes capacity by 10-15%, but this is modest compared to the effects of other parameters (especially spectrum size and coax drop length). Similarly, a change in transmit SNR of 3 dB changes the capacity by ~2%, which is relatively insignificant.

In order to simplify the analysis, we will assume a source SNR of 40 dB and CPE NF of 10 dB. These values are both in the middle of the somewhat conservative ranges. Moreover, the effect of changing these values is relatively modest, as explained earlier.

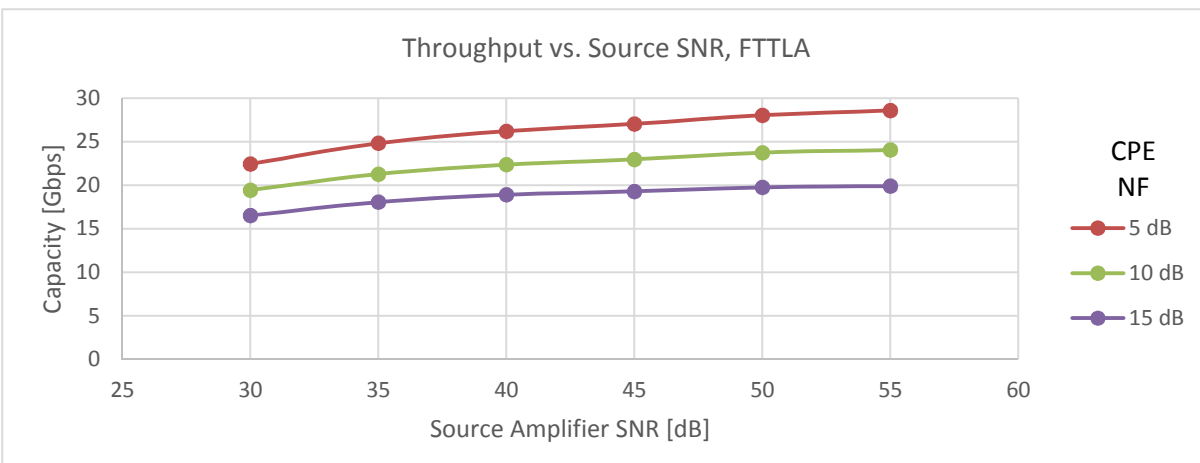


Figure 17. Capacity vs. Source SNR in an FTTLA system

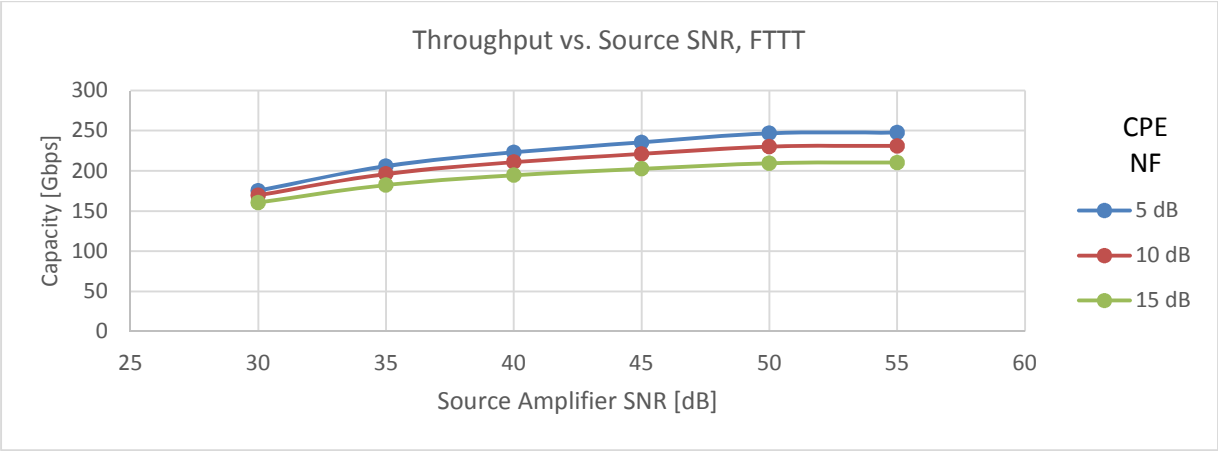


Figure 18. Capacity vs. Source SNR in an FTTT system

<u>The Effect of RF Tilt on Extended Spectrum Capacity</u>

In HFC systems, it is common practice to apply an RF "up-tilt" to the transmitted RF signal at the amplifier. The goal is to compensate for the attenuation of the coax cables (hard-line distribution cables and RG-6 drop lines) which causes a downward tilt across the spectrum. If it were possible to do this perfectly (which it isn't, as different modems will suffer from different tilts due to differing lengths of coax run), the result would be that every 6 MHz channel would be received at the same RF power. This has obvious advantages in a traditional digital (or mixed analog/digital) TV system, where a receiver would demodulate a single channel, and would only accept a relatively limited range of channel powers (typically $\pm$ 15 dB) in order to successfully demodulate a QAM signal (usually 256-QAM) with an SNR target of ~35 dB.

While this target of flat received power simplifies the design of single channel demodulators, it is not clear that it optimizes the capacity of a multichannel system, in terms of total throughput per unit transmitted power, especially when multiple QAM orders are available. In this section, we look at the effect of applying tilt to the extended-spectrum DOCSIS signal.

The previous capacity curves in Figures 11-18 were all calculated with no tilt; the RF power was evenly distributed across the spectrum, giving a constant PSD *at the RF amplifier output*. By way of comparison, one representative case (FTTLA with a 150-ft drop cable and 68 dBmV Tx power) was re-simulated with a tilt applied. The tilt was calculated to exactly cancel the attenuation of the 2 types of coax, while maintaining the same total amplifier output power. The result was a signal with a flat PSD *at the receiver*. The results are shown in Figure 19.

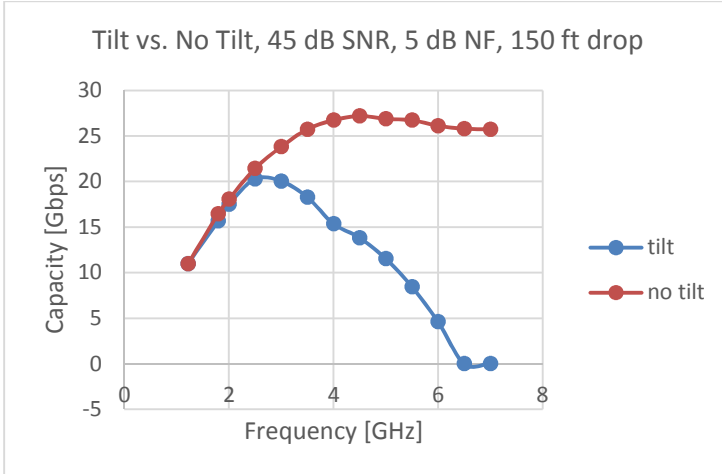


Figure 19. Effect of tilt on capacity at different spectrum widths in an FTTLA system

It is clear that the tilted transmission produces a lower capacity than the flat transmission. This can be explained by the fact that the tilt process, by its very nature, pumps more power into the highly attenuated upper frequencies than it does into the lower frequencies; much of this power is ultimately lost in the noise.

While it is clear from Figure 19 that a no-tilt system has a higher capacity than a fully tilted one, it is not obvious that this is the optimum setting. To explore this possibility, simulations were run with different degrees of tilt. Herein we introduce a parameter, the "tilt factor", to indicate the amount of applied tilt. Full tilt (a tilt factor of 1) is the case where the applied tilt completely cancels the coax tilt; any other tilt (a tilt factor of *f*) is where the applied tilt (in the dB scale) is *f* times the coax tilt (but in the opposite direction). In this way, even negative tilts (which reinforce the coax tilt instead of cancelling it) may be considered. The results of these simulations are plotted in Figure 20. It is clear from this figure that the optimum tilt is zero (or very close to it).

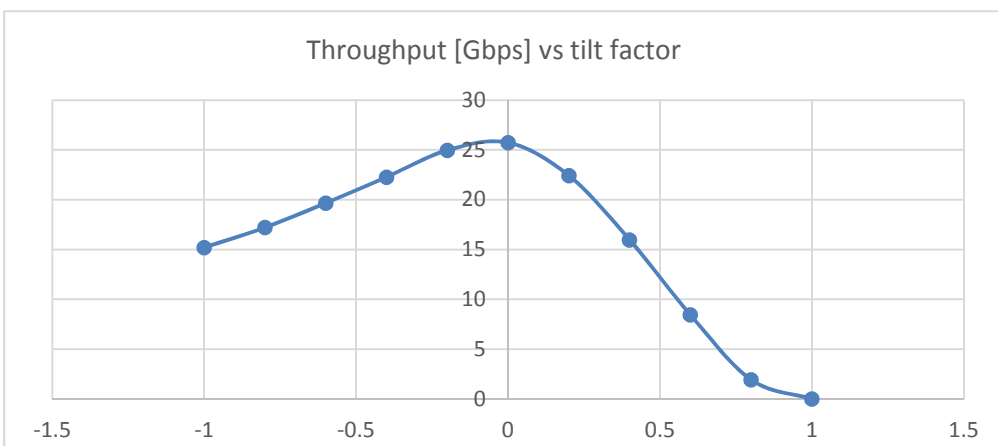


Figure 20. Capacity vs. Tilt Factor

EXTENDED SPECTRUM DOCSIS EXPERIMENTAL RESULTS

To begin showing that Extended Spectrum DOCSIS operation is a feasible solution for the future, an experiment was conducted to transmit a single 96 MHz OFDM block across a typical 150' coaxial drop cable.

In this experiment, a 96 MHz 4K QAM DOCSIS 3.1 OFDM signal was generated using the E6000 CER at 600 MHz center frequency. The samples of this signal were recorded using Keysight vector signal analyzer (PXA N9030A). These samples were then stored on a keysight function generator (MXG N5182B) which was used to upconvert the OFDM signal to 6 GHz center frequency. The ~50 dBmV OFDM signal was then transmitted over the 150' RG-6 coaxial drop cable and was received and demodulated by a Keysight vector signal analyzer (N9030A). The results of the experiment are shown in Figure 21, where the 4K QAM signal was successfully received with no codeword errors by the vector signal analyzer. A live 8-minute demonstration of the experiment described here can be seen via the video posted by Cablelabs at [Demo].

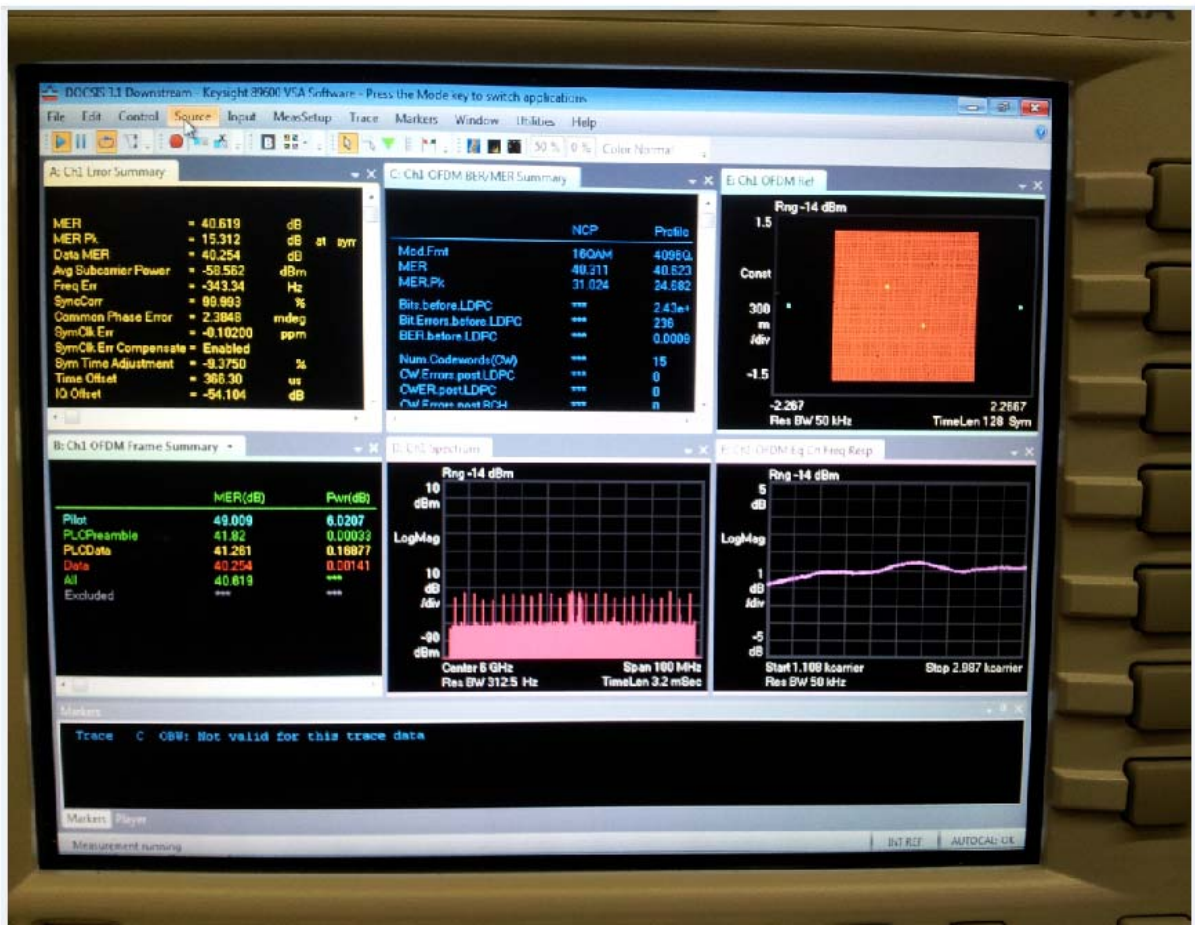


Figure 21. Experimental Results

CONCLUSIONS

This paper proposed a new flavor of DOCSIS called Extended Spectrum DOCSIS, which is aimed at increasing the overall bandwidth capacity that can be provided to subscribers within an HFC network or within a FTTH RFoG network.

The basic idea is to extend the use of OFDM Downstream blocks and OFDMA Upstream blocks into regions of the spectrum that are not currently permitted by the DOCSIS 3.1 specification. Thus, OFDM Downstream blocks would be placed above 1794 MHz, and OFDMA Upstream blocks would be placed above 204 MHz. The blocks could be positioned in a way that yields FDD operation or could be positioned in a way that yields Full Duplex or Half Duplex TDD operation.

Analyses within the paper show that it may be possible to get 50 Gbps or even 100-200 Gbps of bandwidth capacity to be passed across an HFC network (if the plant is appropriately adjusted). Typically, higher bandwidth capacities result from deeper fiber runs and shorter coaxial runs. However, more research is still required before the details of Extended Spectrum DOCSIS system operation can be outlined.

Extended Spectrum DOCSIS has many potential benefits that can help MSOs satisfy the exponentially-increasing bandwidth demands of the future.

First, it is a technology that is based on the existing DOCSIS 3.1 technology. As a result, complicated, time-consuming specification work is likely not going to be required. It could be a relatively simple addendum to the existing DOCSIS 3.1 specification.

As a result of its re-use of DOCSIS 3.1 modulation approaches, it should be clear that Extended Spectrum DOCSIS will be backwards-compatible with existing DOCSIS equipment. It is likely that DOCSIS 3.1 equipment and pre-DOCSIS 3.1 equipment will be able to coexist and operate in lower portions of the spectrum, and Extended Spectrum DOCSIS modems will be able to operate in both the lower portions of the frequency spectrum (along with earlier DOCSIS equipment) and in the higher portions of the frequency spectrum.

Since Extended Spectrum DOCSIS uses the same provisioning, configuration, and management systems as DOCSIS, another key benefit is its ability to eliminate the unnecessary operational challenges that oftentimes accompany the introduction of a new technology.

Finally, Extended Spectrum DOCSIS offers MSOs the opportunity to protect their existing investment and capitalize on their already-deployed HFC infrastructure. By squeezing the most bandwidth possible out of their existing HFC plant equipment, MSOs can reduce their future operational costs and can realize the full potential of their network investment.

Because of all of these benefits, MSOs may want to consider Extended Spectrum DOCSIS as an important tool within their toolkit of solutions that may be used in the 2020 decade and beyond when DOCSIS 3.1 bandwidth capacities become inadequate for the growing bandwidth demands of the future.

REFERENCES

[CLO1] Tom Cloonan et. al., “Lessons from Telco and Wireless Providers: Extending the Life of the HFC Plant with New Technologies,” in Proceedings, INTX 2015, May, 2015.

[CLO2] Tom Cloonan et. al., “Looming Challenges and Potential Solutions for Future Distributed CCAP Architecture Systems,” in

Proceedings, SCTE Cable-Tec Expo 2015, October, 2015.

[Demo] https://www.youtube.com/playlist?list=PLBuD5Alv3NuMUDJjxYjV9HS-JlV94yph7

[DPHY] CableLabs, "Data Over Cable Data Interface Specifications DOCSIS 3.1 Physical Layer Specification," CM-SP-PHY-v3.1-I08-151210.

[EMME] Mike Emmendorfer, "Comparing the DOCSIS 3.1 and HFC Evolution to the FTTH Revolution," in Proceedings, SCTE Cable-Tec Expo 2015, October, 2015.

[FDX] http://www.cablelabs.com/full-duplex-docsis-3-1-technology-raising-the-ante-with-symmetric-gigabit-service/

[MUTA] Venk Mutalik et. al., "The Yin and the Yang of a Move to All Fiber: Transforming HFC to an All Fiber Network while Leveraging the Deployed HFC Assets, in Proceedings, INTX 2015, May, 2015.

[SHAN] Claude Shannon and Ralph Hartley, "The Shannon-Hartley Theorem," https://en.wikipedia.org/wiki/Shannon%E2%80%93Hartley_theorem

VIRTUAL MACHINE PLACEMENT STRATEGIES FOR VIRTUAL NETWORK FUNCTIONS

Adam Grochowski
Juniper Networks

Abstract

Current Virtual Machine (VM) scheduling services, such as Openstack's Nova only have awareness of the CPU, RAM, and storage utilization from a hypervisor perspective. While this has proven to be sufficient for traditional cloud applications and their associated workloads, as they are rarely limited by network bandwidth, Virtual Networks Functions (VNFs) by comparison require fairly static and known quantities of CPU, memory and storage. However, properly placing their workloads depends upon a knowledge of hypervisor network resources and external network topology.

This paper will compare and contrast traditional cloud workloads and their placement with VNF workloads and their respective placement. We will discuss why theoretical extensions to cloud management software are needed to improve placement, quality of service and utilization of commodity hardware.

VNF/NFV BACKGROUND

Network Functions Virtualization (NFV) is an effort originally proposed by the European Telecommunications Standard Institute (ETSI) at the SDN and OpenFlow World Congress in 2012. [1]. The objective or focus of the NFV initiative, is to use standard virtualization technology to consolidate various network equipment types onto a common platform that can be easily distributed throughout the network. The NFV platform comprises high capacity servers, switches and storage, which will utilize software applications running on virtual machines to perform functions like routing, switching and security. These platforms can then be located in the datacenter, head end, or on the end users' premises.

Service Providers will benefit from NFV in the following ways. First and foremost, NFV provides for the ability to increase agility and the velocity of new service deployment, improving time to market. Virtualizing network equipment improves the ability to automate, which allows for faster service instantiation. NFV can help operators by reducing the required initial investment in equipment cost and power consumption as many network functions can be combined onto standard servers. As in traditional cloud application environments, services can be more quickly and even automatically scaled to meet increased demand.

With NFV service providers now have the ability to create virtual devices providing network services on demand that can be dynamically inserted into the end-user's path.

CURRENT WORKLOAD SCHEDULING

The industry has been moving toward Openstack as the standard for managing cloud computing platforms for private and many public cloud installations. Therefore, because Openstack is becoming the de-facto standard and it shares many things in common with other cloud management packages, this paper will use Openstack and its components as a representative example of cloud management software.

Before we dive into how cloud workload scheduling works, it is important to understand how cloud management platforms are structured. Openstack, as an example, comprises several sub projects, which represent the required functions or systems that manage the different aspects of a cloud installation [2]. These projects and their codenames are listed below:

Network – Neutron
Compute – Nova
Authentication – Keystone
UI – Horizon
Storage – Swift, Cinder
Image Service – Glance
Telemetry – Ceilometer, Gnocchi

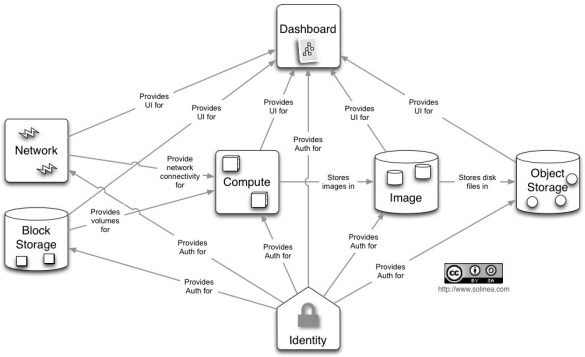


Figure 1: Openstack components. Source: openstack.org

All of these subprojects work in concert to orchestrate the placement and instantiation of a virtual machine as well as its associated network, storage, and other required resources. This placement is broadly based on the following: Availability zone, available RAM, storage, compute availability, custom filters (NB: additional items exist but for the purposes of this paper the most common metrics are referenced). When a user issues a request for a new virtual machine be created, the scheduler will check for available resources, select the appropriate hypervisor to service the request, and trigger the VM creation/boot on the selected host from a virtual image.

TRADITIONAL CLOUD WORKLOADS

Servicing VM requests based on the above metrics has proven adequate for traditional cloud workloads, since memory has in general been the primary point of contention for web-based apps and services, followed by the availability of CPU and storage resources.

In the case of traditional cloud workloads, per-node bandwidth utilization is an infrequent issue, and on the occasion that it is, 'noisy neighbors' are usually prevented from affecting other workloads with virtual switch based QoS, or by over-provisioning hypervisor resources.

NFV Requirements

With the entrance of NFV into the equation, requirements for scheduling need to be re-considered. Providers are starting to move network functions into the virtual world. These functions include, but are not limited to:

Firewalls
Intrusion detection and prevention systems
Route reflector
WAN acceleration appliances

In general, purpose built VNFs have fixed/known requirements for CPU, storage, and memory that can be scheduled by VM orchestrators such as Openstack Nova. It is important to note that in the case of NFV oversubscription does need to be considered. In the world of networking, sharing CPU cores or RAM will create sub-optimal network performance, so this does need to be taken into account when making scaling decisions.

Additionally, the network requirements for each subscriber can vary greatly. As an example, the requirements for a small business, branch office, or residential customer might require a virtual firewall capable of passing an average of 25 Mbps (IMIX), while a business providing free Wi-Fi to customers would potentially need far more bandwidth. At the time of this writing, Nova does not have the ability to schedule based on network utilization or for VNF workloads, this is a gap that needs to be addressed.

SCHEDULING OPTIONS

Since we have established different scheduling needs for traditional cloud workloads vs. VNF, we can consider some options to make the cloud orchestration software take available network bandwidth into consideration.

Turning again to Openstack as our example, we can look at the workflow of the filter scheduler, which makes the decision on workload placement. When a new instance is created a series of filters are applied in order to choose the appropriate hypervisor upon which to place that instance. Filters are binary, either a host passes filtering or it's rejected from consideration.
Current filters include [5]:

AggregateCoreFilter
AggregateDiskFilter
AggregateImagePropertiesIsolation
AggregateInstanceExtraSpecsFilter
AggregateIoOpsFilter
AggregateMultiTenancyIsolation
AggregateNumInstancesFilter
AggregateRamFilter
AggregateTypeAffinityFilter
AllHostsFilter
AvailabilityZoneFilter
ComputeCapabilitiesFilter
ComputeFilter
CoreFilter
NUMATopologyFilter
DifferentHostFilter
DiskFilter
GroupAffinityFilter
GroupAntiAffinityFilter
ImagePropertiesFilter
IsolatedHostsFilter
IoOpsFilter
JsonFilter
MetricsFilter
NumInstancesFilter
PciPassthroughFilter
RamFilter
RetryFilter
SameHostFilter
ServerGroupAffinityFilter
ServerGroupAntiAffinityFilter
SimpleCIDRAffinityFilter
TrustedFilter
TypeAffinityFilter

During the scheduling process, these filters are applied based on resources requested. Additional weights are applied, and finally workloads are created on the appropriate compute nodes.

The following diagram shows the filtering workflow. [5]:

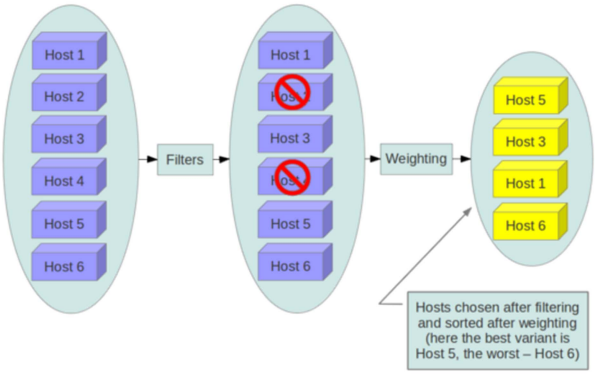


Figure 2: Filtering Workflow. Source: openstack.org

NetworkBWFilter

The contention of this paper is Network awareness is a requirement of NFV, and is absent from the metrics currently used to schedule virtual workloads.

One proposed solution is to create a a new filter that would allow the scheduler to take network requirements into account during into the decision making process. Taking the operative example from the functionality of the IoOpsFilter, a filter such as NetworkBwFilter would be created. As part of the NetworkBwFilter, a max_bw_per_host would be set to specify a high water mark allowable on a particular compute host.

When the request for a new VM instantiation is made, a required amount of bandwidth for that workload would be specified. As with existing filters, each time a host is selected for placement of VNFs, those resources are consumed virtually, and subsequent host selections can be adjusted accordingly. This would guarantee that the amount of requested bandwidth would never exceed that which is available.

TelemetryFilter

Another possibility is to create a filter that would take into account usage data from a telemetry package (e.g. ceilometer) and use actual usage data to make host filtering decisions. This would allow for even more efficient use of network bandwidth.

This filter could even make decisions based on the results of API calls to external systems. An example of this type of operation would be that during the workload creation process, the scheduler makes API calls to a network analytics server.
Utilization information gathered from the network analytics system could be applied as either a weight or a filter during the scheduling process.

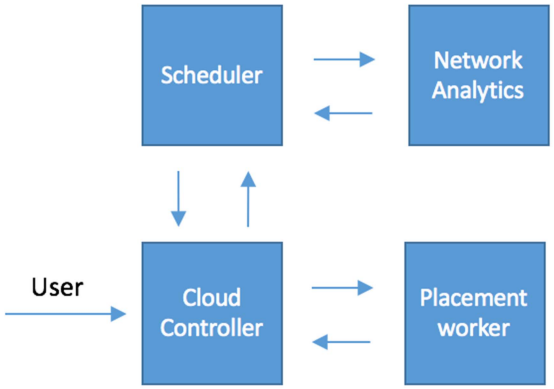


Figure 3: External Interaction with Network Analytics

Since reducing contention between workloads is a primary concern, continual monitoring would be required, with the potential need to re-spawn workloads elsewhere and redirect traffic to them when utilization changes.

Note: There exists a blueprint within Openstack for 'Utilization aware scheduling'. Which is meant to take into account transient resources, allowing decisions to be made based upon gathered usage statistics. [3] Work on the network monitor portion of Utilization aware scheduling appears to have been abandoned[4]. Were this work to continue, it would be a positive step in the goals laid out by this paper, at least where the Openstack project is concerned.

SUMMARY AND CONCLUSION

This paper has compared and contrasted traditional VM workload scheduling with the requirements of NFV applications and included a high level discussion of potential solutions. A gap exists in current cloud

management software which needs to be addressed in order to make it more suitable for NFV applications. Taking network utilization into account during workload scheduling decisions can improve ROI for standard hardware by providing the ability to more densely pack virtual workloads and avoid stranded resources.

REFERENCES

[1] Chiosi, Margaret, et. al, (2012). *Network Functions Virtualisation, An Introduction, Benefits, Enablers, Challenges & Call for Action*, SDN and OpenFlow World Congress, Darmstadt-Germany
[2] http://docs.openstack.org/icehouse/training-guides/content/operator-getting-started.html
[3] https://wiki.openstack.org/wiki/UtilizationAwareScheduling
[4] https://review.openstack.org/#/c/44007/
[5] http://docs.openstack.org/liberty/config-reference/content/section_compute-scheduler.html

WiFi Optimizations to Improve SSID Priority to Enhance Overall Quality of Experience

Ivan Ong
Comcast

Abstract

Mobile devices (laptops, tablets, and phones) may not associate with the best Wi-Fi access point when returning to a home network. This paper will explore how to tune parameters on the PHY and MAC layer of 802.11 to improve association in the absence of a controller or without the need of a third party connection manager app to ensure prioritization from the client itself. It will also include an empirical analysis of the behavior of certain clients as they attempt to associate to their residential wireless gateway routers and conclude with some guidelines for potential tuning parameters.

Introduction

The preference from a service provider's perspective is for a user's mobile device to associate to their home network to ensure proper policies are applied and downloadable content (eg. Podcast, personal video) is able to synchronize but devices often associate to less desirable networks. In part, this is due to the vast array of mobile operating systems and how they decide to associate to a Wi-Fi network. For example, an iOS device merely relies on the last known service set identifier (SSID) when performing a Wi-Fi network selection; this can be problematic because not all networks are equally desirable but iOS does not provide any facility to specify priorities or preferences among Wi-Fi SSID. To add to the complexity of Wi-Fi SSID selection, Apple's iOS Deployment Reference documentation states that the current basic service set identifier (BSSID) connection will be maintained until the RSSI level crosses a -70dBm threshold [4]. This is a complex issue in a multi-dwelling environment where a user's mobile device has associated to a public or neighboring SSID and, upon return to home, is unable to intelligently associate to the home network.

WiFi Scanning

Before exploring the potential of MAC and PHY layer parameter tuning, it is critical to understand the fundamentals of an 802.11 client joining a network. This three stage process consists of [5]:

- Probing
- Authentication
- Association

In the probing stage, a mobile client supports both active and passive scanning; in active scanning (See Figure 1), probe requests are sent at defined intervals to look for either any SSID or specific SSID defined within the frame. Access Points (AP) can then respond to these probe requests to begin the second part of the three stage process of joining a network. In passive scanning (See Figure 2), the client radio cycles through each [operating] channel listening for beacons that may be advertising their SSID. If it receives multiple beacons from various neighboring APs for the same SSID, it will then attempt to associate to the best received signal strength indicator (RSSI) level.

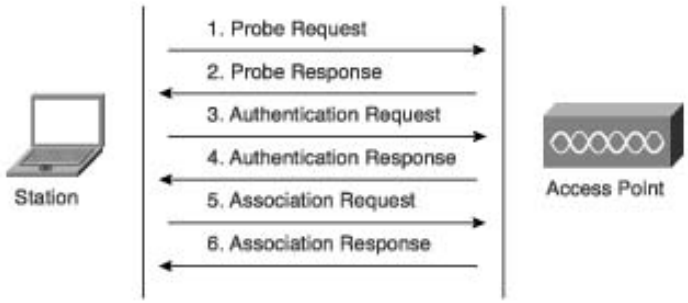


Fig 1: Active Scan

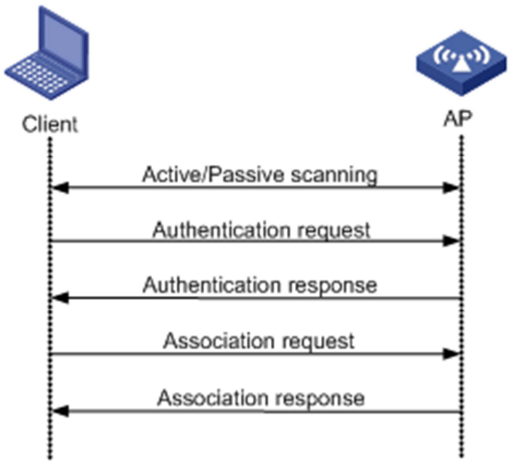


Fig 2: Passive Scan

Once the scanning process has been completed and the client has made a decision to join a network, it will need to authenticate against the AP. Upon successful completion of the authentication process, the association stage will then begin (refer to Figure 3).

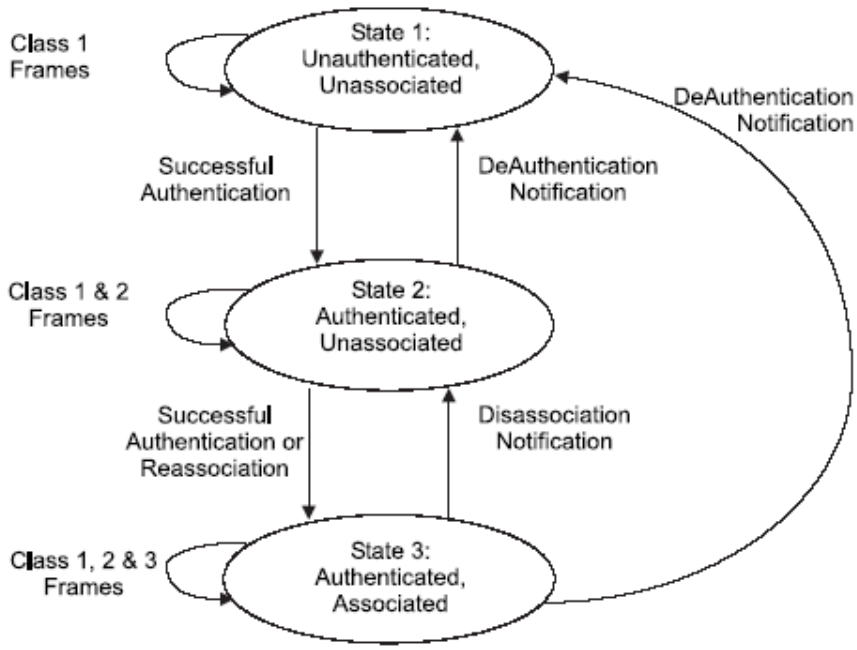


Fig 3. State Machine for Client Association

802.11 Frame Format

The 802.11 Frame format is represented in Figure 4, it is the underlying foundation for Wi-Fi communication in terms of information between client and AP. The structure consists of a MAC header, frame body, and frame check sequence (FCS). The first two bytes of the frame structure specifies the Frame Control block (see Figure 5) which further defines 'control' information to assist the type of information necessary for the receiver on how to process the MAC frame.

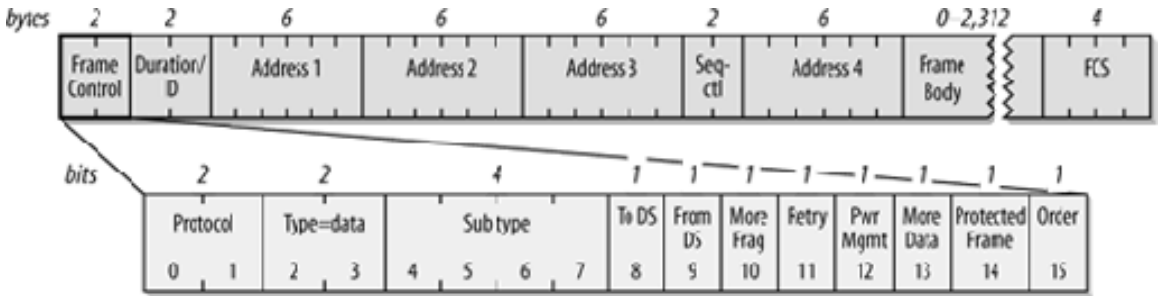


Fig 4: 802.11 MAC Frame Format

The third block of the frame control subfield 'subtype' is a byte block to indicate the type of control frame that is being transmitted. The following are a subset of subtype values and the control frame identified:

Subtype Value	Subtype Name
0000	Association Request
0001	Association Response
0100	Probe Request
0101	Probe Response
1000	Beacon

Table 1: Frame Control Field

By performing some basic parameter tuning on the MAC layer of a home wireless gateway router, the home SSID can be optimized as the preferred network of choice. In particular, Beacon Interval is a MAC layer attribute existing on most Wi-Fi chipsets that may potentially be exploited to optimize SSID prioritization.

Beacon Interval

Beacon frames are used in communicating the SSID information from an access point (AP) to the station (STA) or client. The beacon interval block is defined within the MAC header's frame body as a 2 byte mandatory value. 16 bits are used to define one time unit which typically correlates to 1 millisecond. Figure 5 provides the layout of the beacon interval block located within the frame body of the MAC header.

Within the 802.11 WLAN frame, the subtype block defining a beacon in binary is 1000 (or subtype 8), this value is conveniently visible when decoding beacon type frames in a packet sniffing tool such as Wireshark. The interval of when a beacon is transmitted is typically set at 100ms as a default value. Each SSID constitutes its own beacon frame, so as an example, if there are 10 SSID defined on an AP, it would mean a beacon frame is transmitted 10 times per second, each unique SSID will be advertised every 100ms and cycle each second. In the case of a residential wireless router, there are typically 3 SSIDs defined with potential SSID to be added in the very near future. The 3 SSIDs will be broadcasted every 100ms and cycles every 300ms. There are no control mechanisms to manage beacon transmission; they are subjected to the same 802.11 CSMA/CA algorithm for when to transmit.

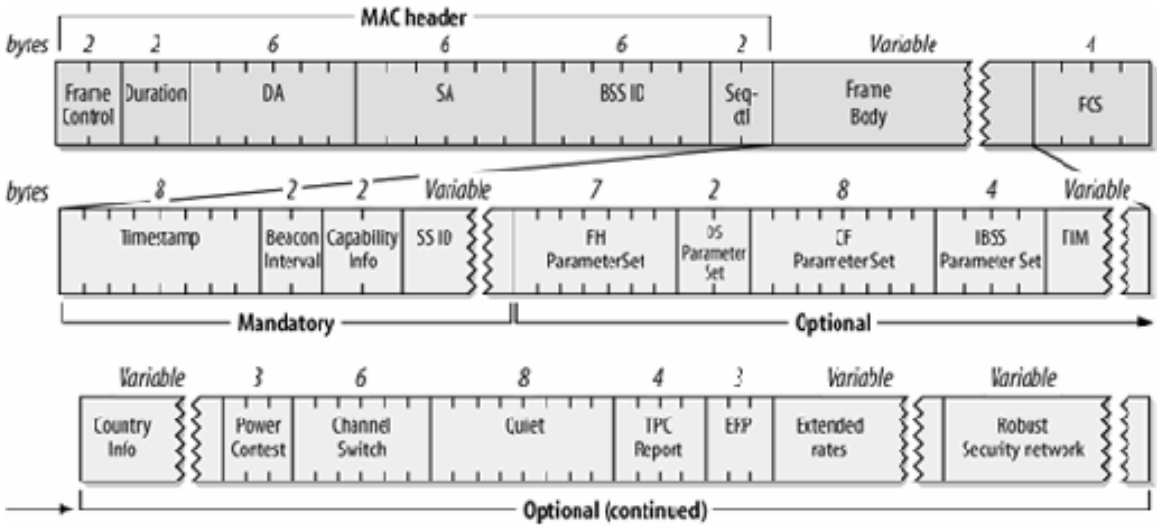


Figure 5: Beacon Frame structure

Proposed Logic and Implementation

Building on our understanding of the 802.11 frame structure, we now describe three approaches to ensuring prioritization of home SSID over that of a public one. The first approach incorporates scheduling based on fingerprinting the behavior of the client. The second introduces a simple algorithm to drop all initial probe requests to a public SSID, hence, increasing the probability that the innate client connection manager will select the home SSID. The third requires account validation of the device to ensure the client is at home and issuing a disassociation frame and forcing a re-association to the home SSID.

Prior to describing and understanding the three SSID prioritization approaches, it is important to understand the behavior of the proposed tweaks to the MAC layer frame such as beacon interval values. The results of some empirical and lab experiments involving beacon interval manipulation will provide a more informed perspective on the first proposed methodology and how that relates to second and third methodology. Once the analysis of the experiments is understood, the logic and algorithms make more sense.

Analysis of Client Behavior

I conducted an empirical analysis of client behavior given the use case of returning to the user's residence after a day of associating to hotspots or non-home SSID.
The results will be discussed in the following sections, the behavior differs slightly between the test clients involved in the experiments. Four scenarios were evaluated:

- Non-Chambered with network preference defined
- Non-Chambered with no network preference defined
- Chambered with network preference defined
- Chambered with no network preference defined

Experiment

An attempt to simulate MAC layer tuning lead to the following basic experiment:

- Increasing beacon interval for a public SSID to be longer than that for a private SSID on a residential wireless gateway router. The private SSID beacon interval will remain at default value of 100ms to avoid interfering with a pre-existing client response to anything other than the default value

The setup of the experiment occurred in two environments: chambered and non-chambered in a residential setting. The equipment consists of two Ericsson AP6120 Access Points (AP) with DOCSIS backhaul, each with the ability to configure beacon interval rates per radio. The clients consist of an Ipad Mini, Samsung Galaxy S4, and Windows 7 laptop. The test case separated each AP into two distinct SSIDs on different channels. Only 5 GHz radios were utilized to ensure a cleaner spectrum to test in. Beacons ranging from 125ms, 150ms, 200ms, 300ms were configured on the secondary AP to simulate the delayed frequency in beacon broadcast. The default 100ms value was held constant, 10 sampling runs per radio were performed, then configuration would be switched on each AP to ensure that there was no BSSID dependency or hardware affinity by the client.

Test Devices	Model	Description
Windows 7 Laptop	Dell Precision M4500	Intel Centrino Ultimate N 6300A/G/N wifi card, Windows 7
Apple Ipad Mini 2	ME276LL/A	802.11A/B/G/N dual band , iOS 8.3
Samsung Galaxy S4	SGH-M919	Android 4.4.4 (kitkat), 802.11A/B/G/N/AC

Table 2: List of clients used in test

I. Non-Chambered Experiment with Network Preference Defined

A real world approach to this experiment was conducted in a residential setting with both APs connected to cable modems so they can be remotely accessed. The same SSID was broadcast on different channel to avoid co-channel interference, the client used were an Apple Ipad Mini 2 and Samsung Galaxy S4 mobile phone. The average RSSI was logged around -40 to -50dB, or more. The SNR was measured from the same location as the clients using a Fluke Wi-Fi Air analyzer provided values around 40-50dBm. See Figure 6 for the test setup

Both iOS and Android test clients were associated to public hotspot SSID and then introduced to the non-chambered test environment. The Android test client associated to the 100ms AP in all sampling runs and to the private SSID as the beacon interval increased for the public SSID. The iOS test client tended to associate to the last known SSID which was the public SSID in all sampling runs. The explanation for this could be attributed to the fact that iOS innate connection behavior favors the last known private SSID which was the public SSID as indicated in the iOS Deployment Reference Guide [4]. Even as the beacon interval increased on the public SSID, it did not steer the iOS test client to the private SSID. This behavior was tested only on one specific client with one version (iOS 8.3).

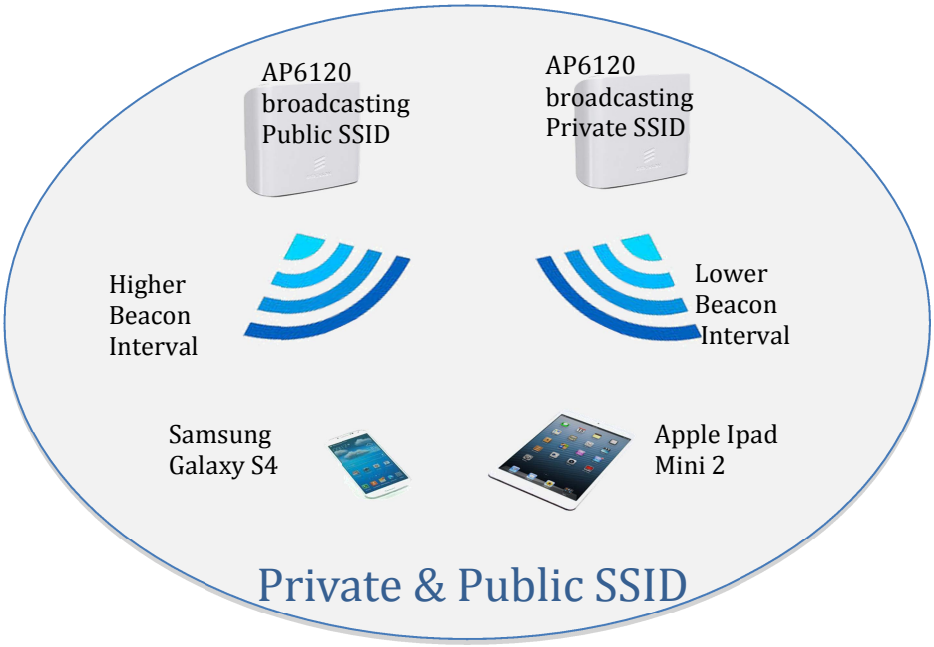


Figure 6: Non Chambered Experiment Setup

An important caveat to note here is the simulated analysis is performed on two separate radios/AP on different channels. Access to a reference Wi-Fi MAC layer to tweak beacon interval per SSID would have been preferred, but due to slow or lack of response from various vendors, we used a simulated approach instead.

II. Non-Chambered Experiment with No Network Preference Defined

The environment remained the same as in the prior experiment; the only difference was no network preference was defined on the test clients. After each sampling run, the networks were 'Forgotten' and Wi-Fi interface was disabled and re-enabled.
A sample of 10 runs was performed per client before alternating the values on each AP to ensure no hardware affinity. The results are summarized in Table 2

When Beacon Rate Increases to: /Devices	Apple Ipad Mini (% tendency to associate to 100ms radio)	Samsung Galaxy S4 (% tendency to associate to 100ms radio)
125ms	80%	55%
150ms	65%	45%
200ms	65%	50%
300ms	50%	60%

Table 3: Non-Chambered with no Network Preference Experiment Results

Based on the empirical analysis of the non-chambered experiment, there are tendencies for the Ipad Mini client to associate to an SSID with lower beacon interval (100ms). As the Beacon interval increases, there is a lesser tendency to associate to the AP with 100ms setting. There is still a minimum of 50% improvement in associating to the lower beacon interval radio.

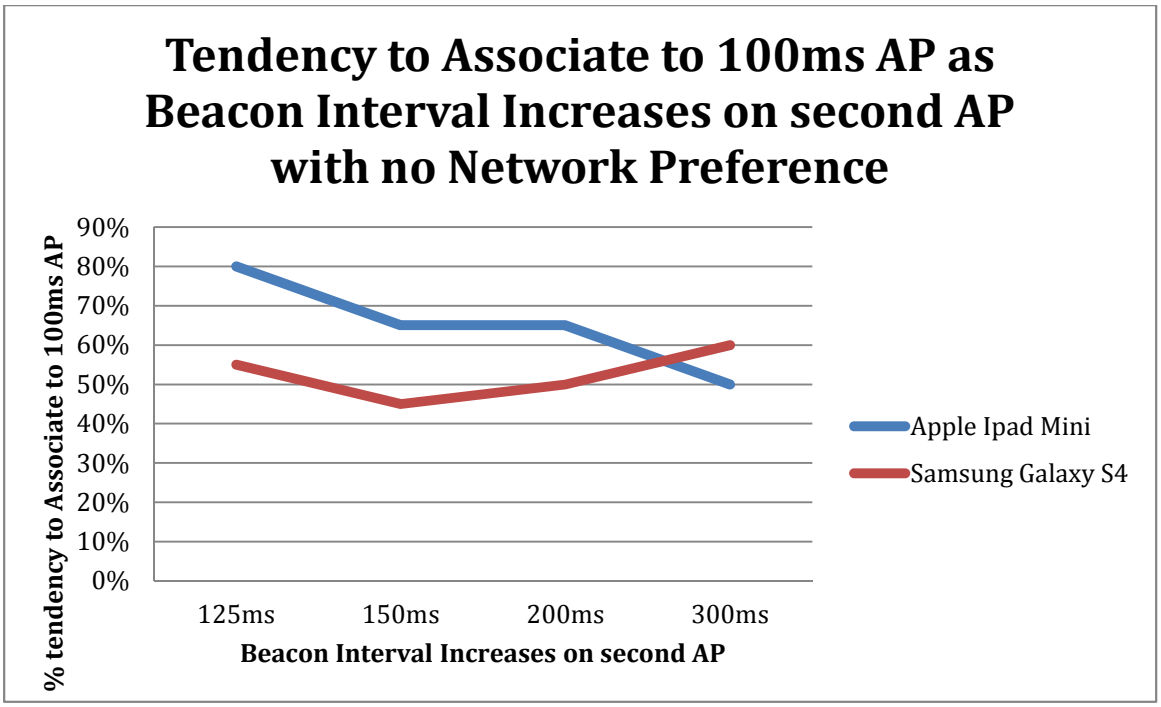


Figure 7: Chart inferred from Table 2

For the Samsung Galaxy S4, the reverse behavior is observed. As the beacon interval increases, there is a slight tendency to associate to the lower beacon interval AP (100ms). The overall results yielded a positive observation of an innate tendency by both iOS and Android connection manager to lean towards lower beacon interval SSID in the presence of differing beacon rates. Referencing Figure 7, the iOS test client, an Ipad Mini 2, has an 80% probability of associating to the default 100ms radio as the other radio's beacon interval increases. Even when beacon interval increases to 300ms on the secondary radio, there is still a 50% likelihood it favors the default 100ms beacon interval radio.

III. Chambered Experiment with No Network Preference Defined

In this environment, each AP was place in a separate chamber with each radio cabled to another chamber containing a laptop running windows 7, (see Figure 8). Each radio was broadcasting on different channels to avoid co-channel interference. Separation between adjacent channels was accounted for to ensure adjacent interference wouldn't skew the results. The average RSSI was logged around -47dBm as reported by both AP. The Wi-Fi radio on the laptop was then enabled and disabled for each sampling run to determine to which AP it would associate. No priorities were set in the laptop's connection manager. The initial test was a difference between 100ms vs. 125ms for the beacon interval. The beacon interval value was incremented from 125ms, 150ms, 200ms, and 300ms. Unfortunately, the results were inconclusive as the AP kept crashing during a majority of the sampling runs.

Figure 8: Chambered Experiment Setup

IV. Chambered Experiment with Network Preference Defined

The chambered analysis was further conducted on the Windows 7 laptop with the order of wireless networks prioritized; the results were unanimously favored the lower beacon interval AP for all runs, (Table 4 and Figure 9).

When Beacon Rate Increases to: /Devices	Windows 7 Laptop (% tendency to associate to 100ms radio)
125ms	100%
150ms	100%
200ms	100%
300ms	100%

Table 4: Chambered with Network Preference Experiment Results

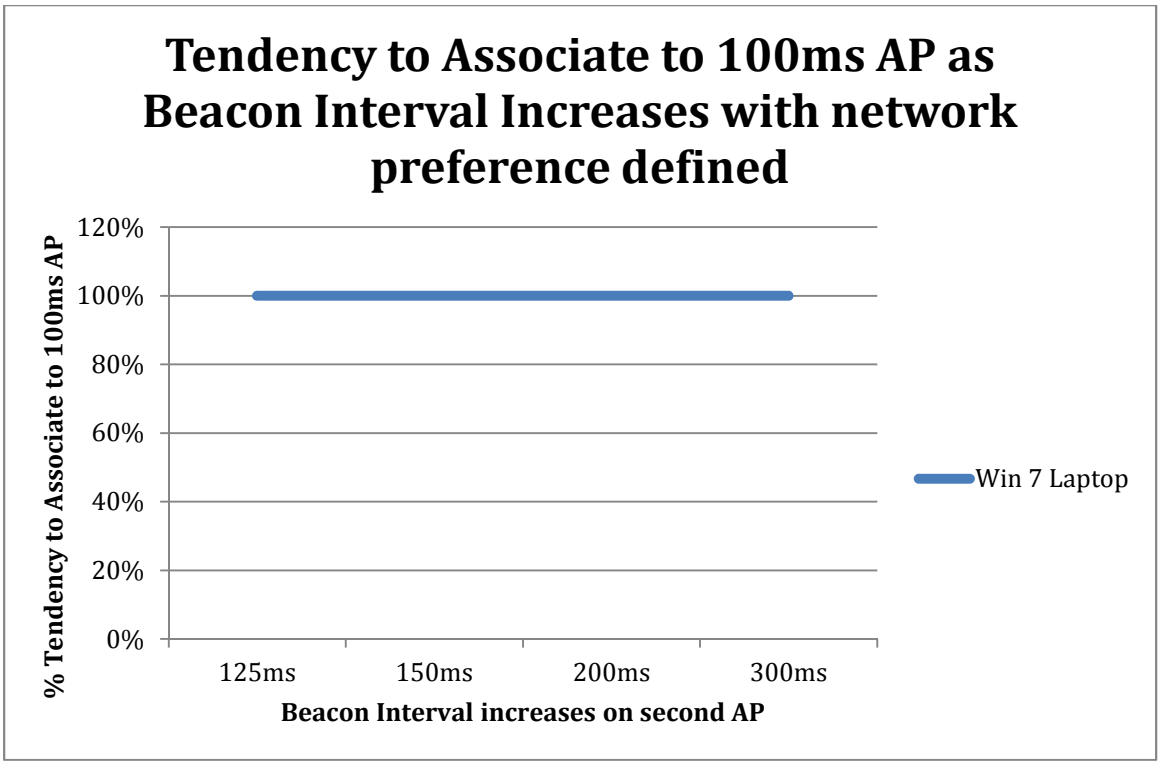


Figure 9: Chart inferred from Table 3

The pertinent observation in this experiment was the network SSID preference was defined on the Windows 7 laptop, the private SSID was set as the 1st SSID to seek then the public SSID as 2nd. As the beacon interval increased on the simulated public SSID, the test client still retained the tendency to associate to the 100ms private SSID. The experiment switched configuration to ensure there was no BSSID or hardware affinity between AP. To expand on the likelihood of the test client associating due to network preference, the same simulated private SSID AP was then configured with a higher beacon interval than the private SSID AP. The results in the next few sampling runs were a little more random as the test client attempted to seek the private SSID AP but due to the higher beacon interval, couldn't do so, and associated occasionally to the public SSID instead.

Proposed MAC Layer Frame Manipulation to Enhance SSID Prioritization of Private over Public SSID

Based on the observations of the experiments, the three proposed resolutions were inferred as a means of ensuring SSID prioritization.

I. Approach 1: Scheduling Implementation

Given that each beacon frame sent from the AP to STA will contain a beacon interval block defined within the frame body, the proposed implementation employs a change in beacon intervals between SSID or even suppression of all other SSID other than the home SSID through cognitive scheduling. A client device will associate to the home AP and the client MAC will typically be cached in the list of connected devices even after disassociation. The daily schedule of client association can be learned and a schedule developed based on defined sampling rates. This fingerprint will then enable application of a cognitive algorithm that either increases beacon interval or suppress all other SSIDs other than the home one in the Beacon

Interval block of the beacon frame during specified times of the day.

A use case to better describe this implementation is a client that returns home anytime between the hours of 5-7pm after work or between certain time blocks on weekends and 'sees' more of the home SSID than other SSID broadcasted by the AP.

Between the predefined hours of return, the AP will broadcast beacon frames with home SSID interval set to 100ms (default value) and increase the value for other SSIDs, thus increasing the frequency that the home beacon may be seen over others. Figure 10 provides the data flow diagram of the proposed implementation:

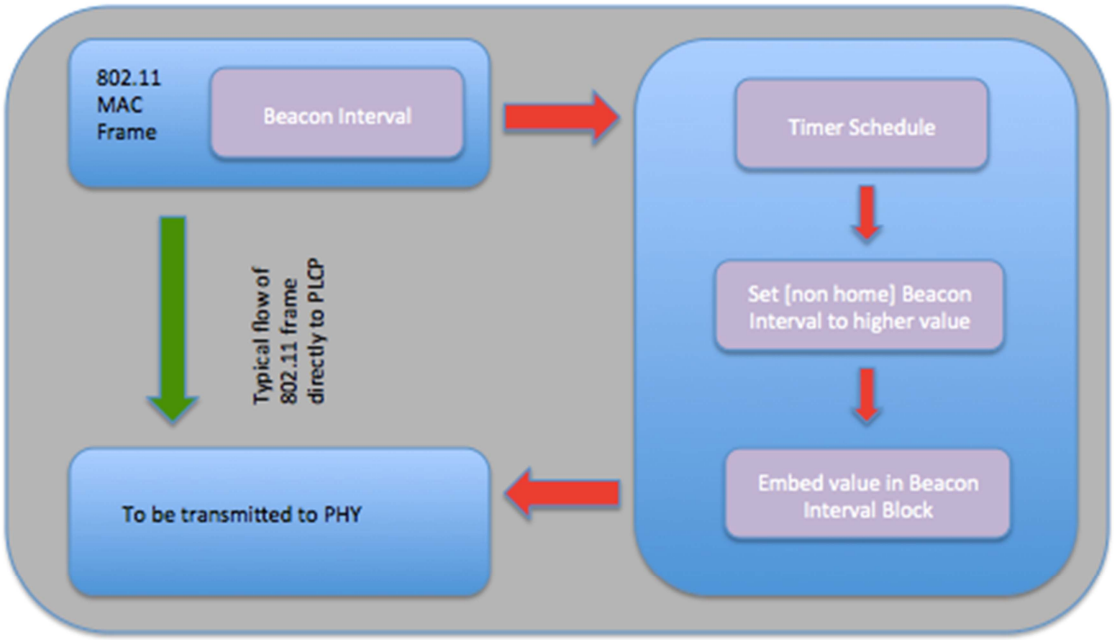


Figure 10: Approach 1: Cognitive Scheduling for SSID Prioritization within MAC layer

The green arrow line in Figure 10 represents the normal flow where an 802.11 frame embeds the necessary beacon values and sends it to the PLCP sub layer for transmission via the PHY layer. The implementation proposes an alternate step in terms of embedding a beacon value to insert higher frequency values for non-home SSIDs to increase the probability of the home SSID being detected by the client's passive scan or probe request when the client returns home within the specified timeframe. The following pseudo code describes the logic within the MAC layer:

```
While timestamp (i) is between 5pm-7pm,
implement scheduling algorithm
{
        i=0
        Write a higher frequency beacon interval
(>100ms) to non-home SSID;
        Replace the more significant bits with higher
defined B (Beacon Interval) values
                i= i+B
}
```

II. Approach 2: Reject Initial Probe Requests per Unique Client MAC

A client's innate connection manager priority is simple for iOS, the order is as listed:

1. The private SSID is most recently associated

2. A private SSID
3. A Hotspot SSID

In contemporary deployments, the hotspot and home SSID have no differentiating trait at the MAC layer level other than the SSID name. Passpoint 2.0 has the ability to address some of this issue but has not been deployed nor supported widely yet. A mobile client will associate to the public SSID throughout the day and upon return to its residence will attempt to re-associate to the same public SSID. This behavior was observed in the 'non-chambered with network preference' defined scenario. The variations in beacon interval did not deter the iOS device from associating to its last known network. Even though more analysis is warranted, this methodology was conceived as a result.

The proposed implementation is to reject all initial probe request to the public SSID [within the residence], this will force the innate connection manager to select the next best private SSID which will be the home SSID.

The sequence diagram in Figure 11 describes the rejection of the initial probe request of the client upon entering the proximity of the residential Wireless Gateway. The probe response will contain [optionally] values forcing the client to proceed to the next valid SSID. These values may include a null SSID along with a reason code indicated within the Frame block of the 802.11 MAC management

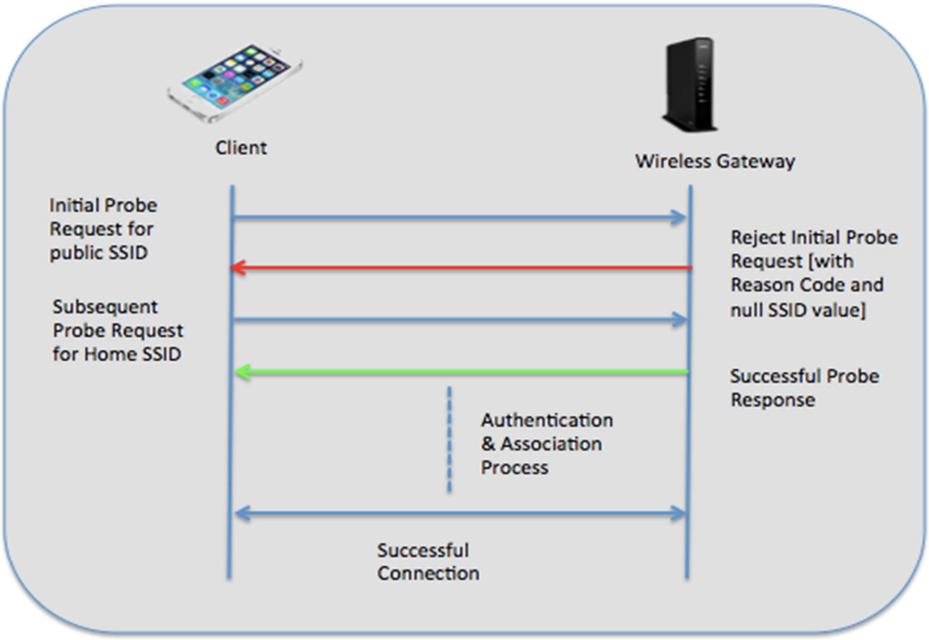


Figure 11: Approach 2: Rejection of initial Probe Request

Note that this algorithm is not a blacklist but merely a lack of response to the initial probe request of each unique MAC and if user desires to associate to the public SSID, the process can be manually triggered to force an association. This process is no different than what is available today by entering the connection manager User Interface and selecting the public SSID.

III. Approach 3: Smart Association to Home SSID

There are mechanisms within a service provider's core network today to indicate a user is sending and receiving traffic behind their home network. When this use case applies, and the client attempts to join a non-home SSID, the initial

association request will be validated the client is behind the home DOCSIS network based on billing account information. The Association response from the AP to the STA will in fact be a disassociation and deauthentication frame used to end the initial association to the non-home SSID. A slight delay timer will begin after this frame is sent to allow time for the client to receive and process the disassociation. A Beacon frame will then be issue after the delay timer expires containing the Receiver Address (client) and the home SSID to connect to within the Probe Request frame.

Subsequently, the client's next action is to perform an [active/passive] scan and potentially may attempt to re-associate with the public or home SSID. The behavior observed all experiments lead to this methodology as evident by how both iOS and Android client behaves. In the event that the client does associate to a non-home SSID, there is reliance on the AP to intelligently switch the client back to the home SSID.

This algorithm will ensure the client in fact will respond to the probe request frame directed from the Source AP (home) with the proper home SSID. Figure 12 describes the process flow of the proposed approach.

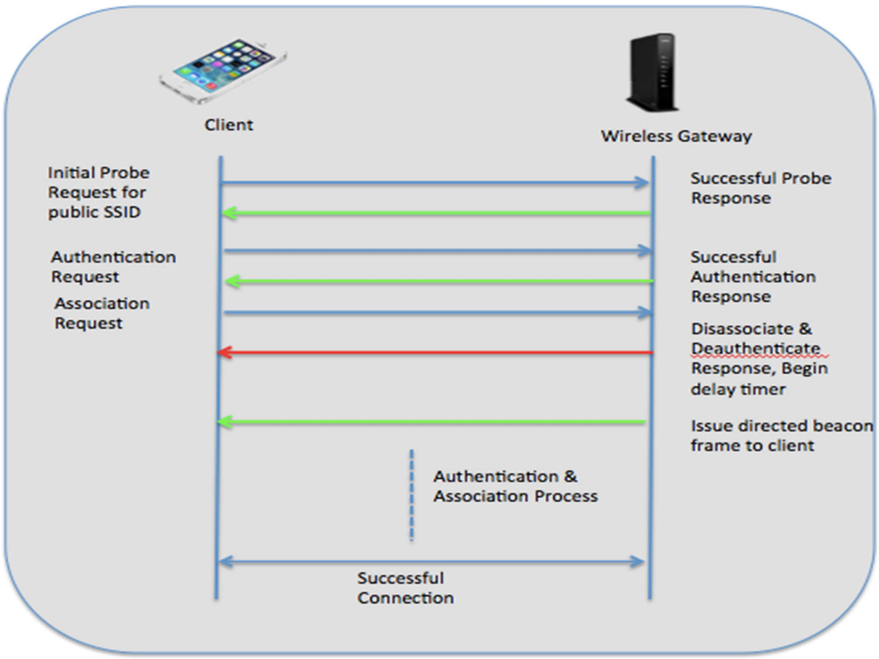


Figure 12: Approach 3: Smart Association to Home SSID

The following pseudo code describes the logic within the MAC layer:

```
When client associates to public SSID behind home cable modem
{
        Validate client traffic is sourced from home cable modem;
        Respond with Disassociation and Deauthentication frame;
        Begin delay timer i=0;
        When i=3ms
        {
                Send Beacon Frame directed to client with Home SSID value embedded;
        }
}
```

Further Analysis

The proposed methods described in this paper has been submitted and accepted for patent filing as potential implementation on wireless gateway routers used in the service provider industry. There are additional analyses required to optimize the design and adapt to the evolving changes in 802.11. For example, manipulating the beacon interval may affect the client behavior; increasing beacon interval to a larger value may cause delays for clients attempting to connect to the AP as the clients scans through every channel seeking for the beacon frame, potentially leading to a longer association time. Conversely, if the beacon interval were to decrease, more bandwidth is consumed but the client will rapidly associate to the AP; as a fallout, power consumption of the client may be affected due to more frequent scanning that occurs.

References

[1] http://www.huffingtonpost.com/vala-afshar/50-incredible-wifi-tech-s_b_4775837.html

[2] http://marketingland.com/nielsen-time-accessing-internet-smartphones-pcs-73683

[3] http://www.ideaconnection.com/blog/2014/02/the-internet-of-things-predictions-for-home-automation-in-2014/

[4] https://manuals.info.apple.com/MANUALS/1000/MA1685/en_US/ios_deployment_reference.pdf

[5] http://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Borderless_Networks/Unified_Access/CMX/CMX_802Fund.pdf

[6] Figure 5 Beacon Interval frame : http://flylib.com/books/en/2.519.1.35/1/

[7] Figure 1 Active Scan: http://www.ciscopress.com/articles/article.asp?p=360065&seqNum=3

[8] Figure 2 Passive Scan: http://www.h3c.com/portal/Technical_Support___Documents/Technical_Documents/Security_Products/H3C_SecPath_U200-A_U200-M_U200-S/Configuration/User_Manual/01-Firewall_Web_Configuration_Manual-5PW100/201205/746164_1285_0.htm

[9] Figure 3 802.11 Client State Machine : http://cecs.wright.edu/~pmateti/InternetSecurity/Lectures/WirelessHacks/Mateti-WirelessHacks.htm

[10] iOS Deployment Reference, 2015, Apple.

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.