THE COMPLETE
TECHNICAL PAPER PROCEEDINGS
FROM:



4K AND HDR, MORE THAN MORE PIXELS
How New Market and Technology Dynamics are Defining Next-Generation TV

Ken Goeller, Vice President, Solutions Engineering
Deluxe OnDemand™

Abstract

As MVPDs shape their approaches to bringing 4K Ultra HD services to market, the emergence of High Dynamic Range (HDR) technology signals that they will soon have an opportunity to deliver a far more compelling viewing experience to their customers.

There's no longer any question that MVPDs must move sooner rather than later to address market pressures that mandate a near-term 4K UHD offering.

Going well beyond the immersive big-screen benefits of 4K, HDR supported TV sets combine greater luminosity, deeper contrast, and a broader color gamut to deliver a stunning picture unlike anything ever seen in mass market television. Just how important HDR is to accelerating 4K adoption can be ascertained simply by viewing a UHD set operating in basic 4K mode next to one equipped to support HDR.

Given market pressures and competitive dynamics, MVPDs can ill afford to wait for HDR to gain traction, even as 4K UHD services are in their nascent phases. Clearly, MVPDs must embrace a migration strategy that embraces the 4K UHD content that exists today, while allowing for the future potential of HDR.

Understanding the distinctions between the most prominent HDR modes and what's entailed in the preparation of content for each, is the first step toward factoring HDR into a UHD service migration strategy. As shall be seen, an appreciation of what's at stake and the attendant challenges, raises the question of what can be done to minimize time-to-market, lower the costs of delivering 4K services that meet consumer expectations, and futureproofing for HDR.

This paper begins with a brief overview of the UHD market, followed by a discussion on the bitrate impact on quality to enable MVPDs to offer 4K UHD content in this early state. We then cover the impact HDR technology is having on the quality of the UHD experience. Next, we turn to a detailed look at HDR technologies, the paths to market adoption and progress toward establishing the new workflow elements that will go into bringing this content to market.

UHD MARKET MOMENTUM

The Big-Screen Factor

Consumer demand for large-screen smart TVs is the greatest force behind the 4K UHD. Early sales are less driven by consumer demand for viewing UHD content, of which there are limited titles (although upscaling capabilities allowing 4K UHD sets to display Blu-ray 1080p HDTV at a higher level of resolution have been heavily promoted as a sales incentive).

A rapid drop in prices, rated by consumers in the CEA study as the biggest consideration in 4K set purchases, is fueling expectations for increased sales.[1] Top-tier brand suppliers are pricing 55-inch 4K sets at under $2,000, which is anywhere from $600 to $1,000 above the cost of a reasonably good quality 55-inch HD smart TV. This is a long way from the

thousands of dollars that separated them three years ago.

The Content Factor

Before long, content will begin to play a larger role in 4K UHD demand, thanks in part to the aggressive steps taken by OTT service providers. UHD services are already available to owners of 2014 or later LG, Samsung, Sony, and Vizio UHD sets. MVPDs are also making moves into the market, starting with Comcast and DirecTV, whose initial offerings are targeted to Samsung TVs.

Adding to the momentum, Dish Network announced it will roll out a new UHD set-top this summer that will work with UHD TV sets equipped with HDMI 2.0 and HDCP 2.2. Major pay TV networks may soon be contributing UHD content as well. A recent survey conducted by researcher FairmileWest found that at least six pay TV broadcasters plan to offer some UHD content online this year.[2]

MOVING AHEAD WITH 4K UHD

Launching a 4K UHD service is a major undertaking that cannot be accomplished merely by adding a new generation of transcoders. Acquiring and ingesting content for UHD is much more complicated than what MVPDs are accustomed to in the HD realm, where they typically receive content that has been pre-qualified with respect to basics like synchronization of video and audio, compilation of metadata, captioning, thumbnail art, etc.

Formatting content from various sources to a uniform level of 4K quality requires expertise in different conversion processes, depending on whether the content was originally shot in 35 mm film, 1080p HDTV or 4K cinematic profile. Unlike the highly automated processes attending format conversions in the HDTV realm, these 4K UHD conversions are resource intensive.

For example, MVPDs must be able to implement quality assurance procedures at every stage of the workflow, from ingest to final staging for distribution. Critical to delivering a superior consumer experience is the ability to take high quality content and make it playable across multiple formats and devices, while also testing that content across the vast range Consumer Electronic companies to ensure the intended playback experience.

Bitrate Impact On Quality

Supporting 4K UHD to both large form factor TVs and lower resolution formats on mobile devices is a balancing act, with the need to minimize bitrates without sacrificing quality in the HEVC transcoding process. This is dependent on the relationship between the source material, size of the master file and the quality of output in the transcoding process.

While there is much fanfare about 4K natively shot content, a significant portion of the libraries will likely be available utilizing the very best 1080p ProRes HQ master source materials available at bitrates of 600-700 Mbps or higher, that can be up-converted to 4K UHD. High bitrate source material, >100 Mbps, is key to achieving high-quality outputs in the HEVC transcoding process. Lower bitrates typically delivered to the MVPD, <15 Mbps, is not a sufficient source for up-converting and transcoding a 4K UHD output.

When using higher quality source material, MVPDs are able to deliver 4K UHD-caliber content that has been up-converted from 1080p masters with output maximum bitrate settings of just 10 Mbps. Furthermore, this content can be delivered as adaptive bitrates (ABR) in lower HEVC transcode outputs of

1080p 7.5 Mbps and 720p 3.8 Mbps to ensure the adaptive bitrate video stream will continue uninterrupted in instances where congestion or other events momentarily cut a user's bandwidth throughput to below 10 Mbps.

Results Using Episodic 1080p Masters

Deluxe OnDemand recently conducted tests with video camera captured, episodic 1080p material with the following source specifications: ProRes HQ 422 10-bit 600 to 700 Mbps. Through advanced "up-res" processing, Deluxe was able to produce content that looks better on a 4K TV than on an HD TV from standard AVC and MPEG2 ready to play 1080p material. Results showed:

- When objects move through the screen, UHD provides more pixels to smooth the motion and reduce image ghosting across multiple frames compared to HD.
- Edges are distinct providing clear separation from backgrounds and other objects in UHD – edges will be slightly fuzzy on HD.
- Static backgrounds do not exhibit blocking or compression artifacts – with less bits HD may mush pixels together to save bits for areas of motion, resulting in macro blocking.
- Straight lines are clean and crisp in UHD – HD may exhibit stair stepping across pixels.
- Skin complexion is more readily apparent in UHD, with greater detail than HD.
- UHD is capable of displaying individual hairs - HD tends to lose thin line details.
- Static backgrounds do not exhibit blocking or compression artifacts – with less bits HD may mush pixels together to save bits for areas of motion, resulting in macro blocking.
- Small details on fabrics are captured in UHD – HD detail will average fine patterns, looking less clear.

Results Using Studio 4K Native Shot Maters

Deluxe OnDemand conducted additional tests on 4K native shot footage that had all of the challenging elements that video compression technology struggles with -- motion, dark footage, water, reflection, skin, hair and CGI.

The content was presented in two bitrates and from two source masters:

Output Bitrate	**Source**
20 Mbps	UHD Master, 8 Gbps, v210 AVI
15 Mbps	UHD Master, 8 Gbps, v210 AVI
20 Mbps	1080p Master, 600-700 Mbps, ProRes HQ
15 Mbps	1080p Master, 600-700 Mbps, ProRes HQ

Clearly, the output based on the UHD master at 20 Mbps yielded the best experience. However, consumers would see discernable improvements from the UHD master source at 15 Mbps and the1080p master source at 20 Mbps.

THE HDR FACTOR

Now there's a new development that could make a big difference in UHD penetration and service expectations – HDR.

As the name implies, High Dynamic Range goes beyond the spatial resolution benefits of 4K by breaking with the long-standing Standard Dynamic Range (SDR or

ITU REC 709) specifications that have defined how TV content is presented since the dawn of the HD era 20 years ago. The various HDR modes now vying for market acceptance all have in common (a) support for a much wider color gamut,(b) much greater contrast dynamics with deeper levels of black in the darker pictorial elements and (c) far greater luminance in the brightest white and color elements.

As depicted in **Figure 1**, the impact of HDR is obvious.



Figure 1: Impact of HDR

The goal is to more closely match the dynamic range of the human visual experience in the natural world. HDR specifications were tied to the limitations of the cathode ray tube (CRT), which can render the darkest scenes at no lower than .117 nits and top out with the brightest white at 100 nits and color brightness ceilings at various levels below 10 nits.

The human eye can detect a luminance range from one millionth (0.000001) of a nit to about 100 million nits. For formulating HDR displays, the goal of course, isn't to match this range but rather, to operate within a dynamic contrast range that is closer to what the eye experiences in the natural world as a person's gaze moves from bright to dark backgrounds – in other words, the instantaneous dynamic range available to human perception as a function of pupil dilation and other opto-physiological processes in real-world situations.

THE TECHNICAL ELEMENTS OF HDR

Contrast Ratio

One way to measure dynamic range in video imaging is to derive a contrast ratio across the entire luminance range based on exposure value or stops, which originated with the f-stops used in camera settings for still photography. One stop is equivalent to a doubling of light value, so that with each stop the luminance increases by a power of 2.

The dynamic range of SDR, at about seven stops, represents a contrast ratio of 128:1. HDR, at the baseline value incorporated in some emerging standards, is about 11 stops, which represents a contrast ratio of 2,048:1, or about 16 times that of SDR. The contrast ratios enabled through enhancements used with templates supported by Dolby, Philips and others go up from there. With everyday experience in the natural world, the maximum contrast ratio registered by the human eye in a given instant with minimal adaptation is about 20 stops, which equates to a 1,048,576:1 contrast ratio.

Dolby, in ascertaining with focus groups the ideal dynamic range that its Dolby Vision enhanced HDR technology should support to create the most realistic human experience possible without producing contrasts that would be jarring to viewers, found that consumers chose as the optimum a range of 0 to 10,000 nits. But while a 10,000-nit dynamic range became the foundation specification for Dolby Vision, the company has settled on 4,000 nits, equating to between 17 and 18 stops or a contrast ratio of 200,000:1, for the initial commercial iteration.

Color Space

Where color gamut is concerned, the goal is to set a benchmark for production that minimizes the amount of color lost from original camera capture, such as occurs when producing content that maps to the REC 709 standard. The ITU has developed REC 2020 as the successor to REC 709, thereby providing a standardized palette for use in video imaging that comes closer to the full color range in nature. Where, with 8-bit encoding, REC 709 encompasses 16.78 million colors, REC 2020 with 10-bit encoding offers 1.07 billion colors. With 12-bit encoding the REC 2020 color count tops out at an incomprehensible 68.7 billion colors.

But a more reasonable target to shoot for is the color gamut devised by the Society of Motion Picture and Television Engineers (SMPTE) with the DCI P3 standard, which is the color range that cinema projectors are pegged to and which, therefore, is the color range used in the filmmaking postproduction process. With 10-bit encoding DCI P3 encompasses a range of 756.6 million colors. See **Figure 2.**

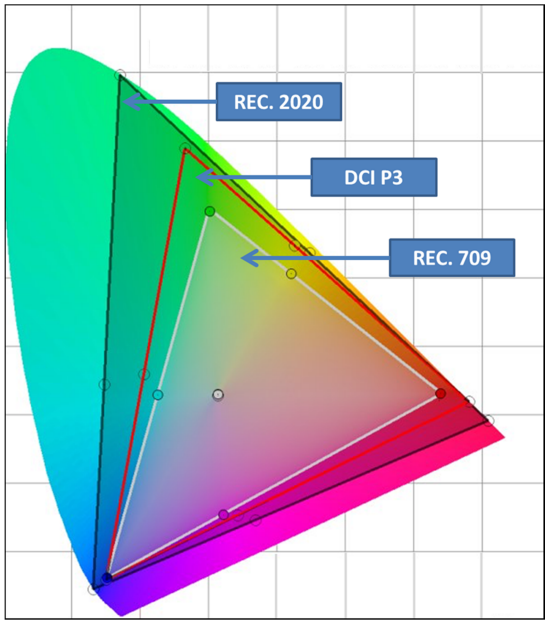


Figure 2: Color Gamut Comparison

Deluxe, which has been working closely with Dolby to set post-production parameters for content to be viewed on Dolby Vision-enabled TV sets, has set the color gamut target to the DCI P3 specifications. This provides a uniform post-production gamut which will also be more than enough to meet the requirements of "baseline" HDR TV sets.

The Transfer Function

A key parameter affecting the ability of manufacturers to repurpose current generation LCD/LED displays to support HDR is what is known as the "gamma function," which is a non-linear transfer function that maps the grey scale signal strength of the display system to the color gamut and dynamic range of the post-production master. Originally set to the display dynamics of CRTs, the gamma function has been enhanced to accommodate the capabilities of LED/LCD displays.

SMPTE 2084 defines an Electro-Optical Transfer Function (EOTF) that is meant to replace the gamma function in order to extend the range of the color gamut and contrast transfer into the realms defined by REC 2020.

SMPTE 2084, using an encoding system known as Perceptual Quantization (PQ), also serves to compress the transferred information in order to reduce the number of bits per color that would otherwise be required to execute the new parameters. For example, without PQ Dolby Vision would probably require 16- rather than 12-bit encoding to signal the high dynamic range and wider color gamut to the display decoder.

Baseline 10-bit HDR systems at luminance levels in the 1,000-nit range and with color a gamut significantly below the full scope of the 12-bit version of REC 2020 or DCI P3, can use the enhanced gamma function. However, Dolby Vision and the enhanced HDR system developed by Philips rely on the more expansive EOTF to enable the display systems to reach the higher color and dynamic ranges achieved with 12-bit encoding. The dividing line between the levels of HDR performance that can rely on the gamma function versus performance levels that require use of SMPTE 2084 may well be a key factor in efforts to define a common baseline HDR standard and a "premium" standard.

Display Systems

Another key factor in setting HDR parameters is the need to minimize costs by leveraging the current generation of SoCs (systems on a chip) used in TV sets. These chipsets have enough processing power to handle the Dolby Vision luminance, color and contrast enhancements. These enhancements, utilizing 12-bit encoding, are conveyed in a metadata overlay, which the SoC decoding processes combine with the regular REC 709 video feed to deliver an integrated signal to Dolby Vision display systems.

This bifurcated overlay approach ensures that content mastered to Dolby Vision specifications will also be viewable on traditional displays, where the SDR decoder does not interact with the metadata. But, from a bandwidth perspective, it adds about 20 percent in bandwidth overhead to the bitstream.

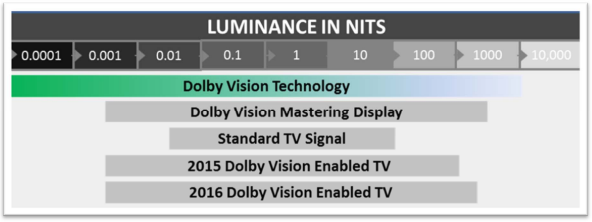


Figure 3: Display Luminance in Nits

As shown in **Figure 3**, the big difference between Dolby Vision, and more basic HDR systems when it comes to manufacturing requirements, is in the displays themselves. Dolby Vision uses individually modulated LED semiconductors that can be switched across multiple steps of luminance from 0 nits to the chosen peak level of brightness, which, as mentioned, is about 4,000 nits for the first-generation Dolby Vision. This requires production of all-new display systems, whereas the current generation of LCD displays, particularly those with better local back-lit contrast controls, can be repurposed with firmware upgrades of current-generation SoCs at the factory to support HDR platforms operating in the 1,000-nit range.

STANDARDIZING HDR

At CES 2015, most of the major brands had prototype HDR sets on display with a bewildering potpourri of labels such as Wide Color LED from LG Electronics, Dynamic Range Remaster from Panasonic, X-tended Dynamic Range from Sony and SUHD from Samsung. Prototypes supporting Dolby Vision were on offer from Toshiba, Hisense, Philips and Vizio.

Efforts to set standards have made significant headway, but there are many initiatives underway that will have to be brought together on the road to true

standardization. These include the ITU's pursuit of a global UHD standard; the U.S. broadcast industry's development of ASTC 3.0; the Blu-ray Disc Association's (BDA's) Ultra-HD Blu-ray platform; the end-to-end UHD agenda at the Ultra HD Forum, and the efforts of the new UHD Alliance.

The UHD Alliance, which was launched at CES, is the latest undertaking, fueled by Samsung's proposal of an open HDR solution. This initiative has brought together some key players from across the consumer electronics, OTT, MVPD, content and studio sectors.

At this early stage, while there may be refinements to Samsung's proposal through contributions from other UHD Alliance members, it appears that open HDR will prevail as a "basic" HDR format, while Dolby Vision could well emerge as a "premium" format. Where the first generation of Dolby Vision has been pegged to a 4,000-nit range in luminance, Samsung's version of open HDR supports a 1,000-nit range.

Meanwhile, the BDA has revealed basic parameters of the Ultra-HD Blu-ray standard scheduled for release later this year. According to press reports, the specs will support an open HDR standard using 10-bit HEVC encoding and SMPTE 2084 signalling while making provisions for two optional solutions, Dolby Vision and the Philip's proposal, that can be layered onto the basic platform. The BDA also includes means by which 10-bit encoding and the wider color gamut can be applied in the presentation of SDR content.

It remains to be seen how all this plays out. While the goal is to deliver products that can be branded with an HDR label guaranteeing a consistent quality of experience on all displays, the industry must reach consensus on basic questions such as what the minimum nit threshold should be, at what point does the nit count exceed viewers' comfort levels, how much contrast is too much and how broad the color palette must be to qualify as HDR compliant.

THE COMMECIALIZATION OF HDR

Whatever comes of the standardization efforts, one thing is certain: HDR when combined with the high resolution offered by 4K UHD, delivers a powerful viewing experience that truly differentiates UHD. While today's LED/LCD TVs deliver luminance at up to 300 to 500 nits, which is sufficient to enable viewing in a brightly lit room, these levels of brightness are achieved with display enhancement techniques applied to content that has been mapped to the contrast and color range of REC 709.

There's no improvement when it comes to fidelity to color or better contrast. For example, there's still no ability to accurately render Coca-Cola Red or the green used on Caltrans highway signs in California. In fact, these brightness enhancements introduce distortions that often take the viewer farther away from the creator's intent with heightened emphasis on "out-of-gamut" colors. And without sufficient contrast, the increased brightness can produce a wash-out effect across color gradations on the limited REC 709 scale.

HDR from OTT Providers

Consequently, buyers of HDR-enabled 4K UHD TV sets will look for 4K content sources that can deliver the far superior experience of HDR. Judging by plans publicized so far by various content providers, HDR set owners won't have long to wait.

One OTT provider, for example, has publicly committed to begin streaming HDR-caliber content this year. While the company has been a vocal proponent of Dolby Vision, executives have said it will be able to support other formats as well. Dolby has announced

commitments from many OTT providers at the present time.

The motion picture studios are beginning to react to the HDR potential as well. Some are already formatting movies for viewing on Dolby Vision displays while others have announced exclusive deal to make some of its movies available for viewing on displays using Samsung's open HDR.

While we don't know how soon HDR-enabled 4K UHD TV sets and the availability of HDR content from OTT sources will begin shaping the expectations of cable subscribers, the fact that early adopters will most likely come from the ranks of MVPDs' highest paying subscribers means their expectations will carry special weight in determining how HDR factors into UHD service strategies. Thus, it's likely MVPDs will need to begin supporting HDR sooner than later, well before the standards issues are fully resolved.

MEETING THE MVPD CHALLENGES

Expediting the Move to HDR

The advent of HDR introduces new challenges for MVPDs. The ability of OEMs to quickly move from 4K UHD to HDR attests to the fact that the systems on chips (SoCs) used in today's television sets are so powerful that new levels of viewing experience can be achieved through advances in software, without having to wait for new generations of hardware.

As a result, MVPDs face a protracted period of transition to mass market adoption, during which they will want to be able to support multiple permutations of HDR without incurring untenable costs. In addition to all the steps associated with ingesting, storing and processing content for higher resolution, MVPDs must be able to establish the workflows essential to ingesting and converting content that can be mapped to the higher color gamut, luminance and contrast ratios of each HDR mode.

Thus, the benefits of the outsource option increase with the addition of support for HDR, which may only be of interest to a small proportion of their subscriber bases. By utilizing an outsource service that is able to support any commercially viable HDR mode, MVPDs will avoid the pain of any format wars preceding industry agreement on standards.

The new procedures start with establishing relationships and a means for acquiring source material that has not been processed in post-production for SDR distribution. In other words, MVPDs want to be able to use source materials that have not lost their original colors, contrast range and brightness in a mastering process. In so doing, they will be able to add value even to old films, which will be converted for TV viewing at a quality level that rivals the original theatrical experience.

MVPDs will need to be able to adopt conversion processes essential to supporting whichever HDR modes gain traction in the marketplace. This will require new expertise, new tools, and in the case of an enhanced HDR service such as Dolby Vision, support for 12-bit encoding and processing steps related to creating the metadata for transmitting enhancements to the baseline REC 709 formats.

Adding to the challenges is the fact that, at this early stage of HDR development, the details of how to execute on all these requirements must be worked out through close working relationships with providers of HDR technology. For example, just as it took time for Deluxe OnDemand to work out all the details that have allowed it to create a fast-to-market launch pad for 4K UHD services, it will take time to develop the optimal HDR platform for baseline and enhanced levels of HDR distribution.

Deluxe, working closely with Dolby and other HDR platform providers, is moving rapidly on this path. As a result, MVPDs can proceed with launching and scaling 4K UHD services in tandem with market demand without waiting until they're ready to support HDR.

When they are ready to add HDR to their UHD options, they will be able to avoid the long delays they would incur by having to climb the learning curve in-house. With the workflows, tools and practices already in place, they will be able to get underway as soon as they determine market conditions have reached the point where they stand to gain by offering their customers the unprecedented viewing experience of HDR.

CONCLUSION

The prospects for consumer adoption of 4K UHD have been increased by virtue of commercial introduction of HDR technology as a major enhancement to the viewing experience. With 4K UHD penetration outpacing earlier industry projections and ever more 4K UHD content entering the market, MVPDs already have ample reason to launch a 4K UHD service, but it won't be long before the same can be said of HDR.

By delivering a quality of experience far beyond that of SDR-formatted content, HDR promises to drive consumer expectations even higher as 4K UHD takes hold. Already, leading OTT suppliers, who were first to market with 4K UHD content, are preparing to introduce HDR content before year's end.

Just as MVPDs cannot afford to cede the 4K UHD viewing advantage to OTT competitors, they cannot ignore the implications of HDR for service development. The question is, with 4K UHD just getting off the ground and so many HDR formats in play, how can these emerging market imperatives be accommodated without spending immense sums on the staffing, workflows, tools and storage capacity these new technologies require, long before the market for these services will grow large enough to generate a reasonable ROI.

The answer can be found in leveraging the shared-cost environment of an outsource supplier for the heavy lifting to launch and scale 4K UHD in the near term, with expectations that similar benefits will accrue when the time comes to introduce HDR services. Companies such as Deluxe OnDemand are already playing the role of the outsource workflow supplier of 4K UHD services to MVPDs, with all the processes essential to delivering a superior 4K UHD service at minimal costs. Furthermore, the outsource strategy has an inherent futureproof component for HDR support. As a result, customers can be assured they will have the opportunity to introduce HDR services whenever they determine the timing is right to enter this next phase in the transition to next-generation TV.

Footnotes:

[1] Consumer Electronics Association, 4K Ultra HD Update: Consumer Adoption & Awareness, January 2015

[2] The Digital TV Group, "At Least Six Pay TV Broadcasters Set to Launch UHD Services this year," January 2015

4K Reasons to Accelerate to All-IP

Dr. Robert L. Howald
Vice President of Network Architecture
Comcast

Abstract

While it may seem like a distant memory, and in retrospect seems hard to fathom, High-Definition (HD) video got off to a very slow and inauspicious start. The adoption cycle was quite long. Obviously display technology, costs at every stage of the content cycle from creation to display, and a very recent massive digital video revolution played a role in the adoption rate. Once the conditions were in place, however, HD took hold and there was no going back.

Fast forward to today, where HD is mature, deployed in scale, and impossible to live without. While an important and profitable component of the operator's business model, there is a resource price to be paid in term of the capacity HD consumes compared to Standard Definition (SD) TV. Operators are regularly adjusting their service mix and introducing technology to increase capacity and make better use of existing capacity to support service evolution. Today's HD falls into this latter category. High Def's maturity, and MPEG-4 encoding's capability present in fielded STBs allows operators to deploy HD content more efficiently.

An HD channel typically has an SD brother in the channel lineup – it's a simulcast. This is not very efficient, but also not terribly penalizing given the relative bandwidth that SD consumes. Now, along comes….4kHD. It's not quite really here, and there will be clear limitations at first. But, it is quite reasonable to expect that 4kHD scales into a mass service within a decade. However, a simulcast approach for current HD channels to be made available in 4kHD would be difficult. The likely option to prepare for 4kHD, in addition to HEVC encoding, is transitioning to an all-IP architecture. Already part of most operator's future plans, 4kHD may be additional inspiration to drive forward with a methodical transition strategy.

This paper will take the reader through an all-IP transition scenario that considers key aspects that an operator must coordinate in phases. Access architecture, bandwidth, encoders, CPE, home architecture, and multicast services represent major balls to be juggled in the transition of services. We will walk through a hypothetical case study based on a typical existing system serving as the Day 0 scenario. We will quantify the CPE situation and migration plans for STB and HSD solutions. We will describe a spectrum management plan with a sensible pace that balances legacy box retirement with capacity demand. And, we will consider how services and capacity tie to network evolution.

An all-IP nirvana has been a future vision for quite some time. With enhanced video and continued HSD speed demand, the future has arrived.

INTRODUCTION

Over the last two decades of service evolution of video, voice, and data, operators have developed in their DNA a keen sense for the planning and implementing growth opportunities at just the right time. MSOs

have been a very successful with a pay-as-you grow approach, adopting new technology at the proper point in their lifecycle to deliver on increasingly demanding consumer services. A major enabling component has been access to new HFC capacity, incrementally exploited by pushing fiber deeper, expanding RF bandwidth in the distribution plant, use of WDM in the Optical Distribution Network (ODN) to fuel continued segmentation and services, and migrating all-digital video, or in some cases deploying switched digital video (SDV).

The challenge of continuing to aggressively deliver on new services, which are being introduced more rapidly than ever, and matching the pace of this with network evolution, has never been greater. It is the nature of infrastructure that changing it does not come simply or inexpensively. Various emerging initiatives in Network Function Virtualization (NFV) and Software Defined Networks (SDN) promise to bring some renewed flexibility to evolving the network quickly, but big iron at the edge, connectivity to the home, and CPE are necessarily part of the equation.

The transition to all-IP has been part of most cable operator's thinking for years, and recent service deployment decisions have been weighted towards the recognition of moving towards an IP end state as opposed to adding new capabilities to the legacy digital video, often referred to as the "QAM" ecosystem (confusingly so since cable's IP pipe – DOCSIS – also uses QAM modulation). As new technology draws out the untapped capacity potential of HFC, it will be used to deliver the next generation of HD services, 4KHD, as well as be called upon the support the unquenchable appetite for IP data capacity and speed. Together, they will fuel the final stages of the all-IP transition.

SEEING MORE BANDWIDTH

The appetite for HD is stronger than ever, and it has become a must-have for most video subscribers. However, the lifecycle of HD, which itself carries a large relative bandwidth penalty compared to the longstanding SD services, has only just begun. Cable systems deliver programming in both 720p and 1080i formats today, while 1080p already exists in consumer electronics such as Blu-Ray and gaming consoles. Flat panel televisions continue to become larger and sizes as large as 84" (7 feet!) have been available on the market for a couple of years.

Of course, such screen sizes are able to take advantage of emerging higher resolution HD formats, such as 4K HD. The 4K HD format represents 4x the number of pixels of 1080-column HD. Figure 1 shows the well-understood relationship between resolution, screen size, and viewing distance that governs the practicality of its use as it relates to improved video quality. Note that 4K HD is more than just more pixels. It also enables "better" pixels and potentially increased frame rates in the future.

By the end of this year, it is projected that nearly all majority of TVs sold new will be 4K HD capable. Obviously, they will not be displaying content that has been delivered in 4K HD format right away. Most cable operators utilize and monetize all or most of their spectrum at all times. They regularly manage that spectrum and adjust services, mostly adjusting the video and data service mix. Most of the spectrum is broadcast video spectrum, and it is not practical that "HD 2,0" in the form of 4K HD is delivered on the cable network as a simulcast alongside the existing SD and HD broadcast services.

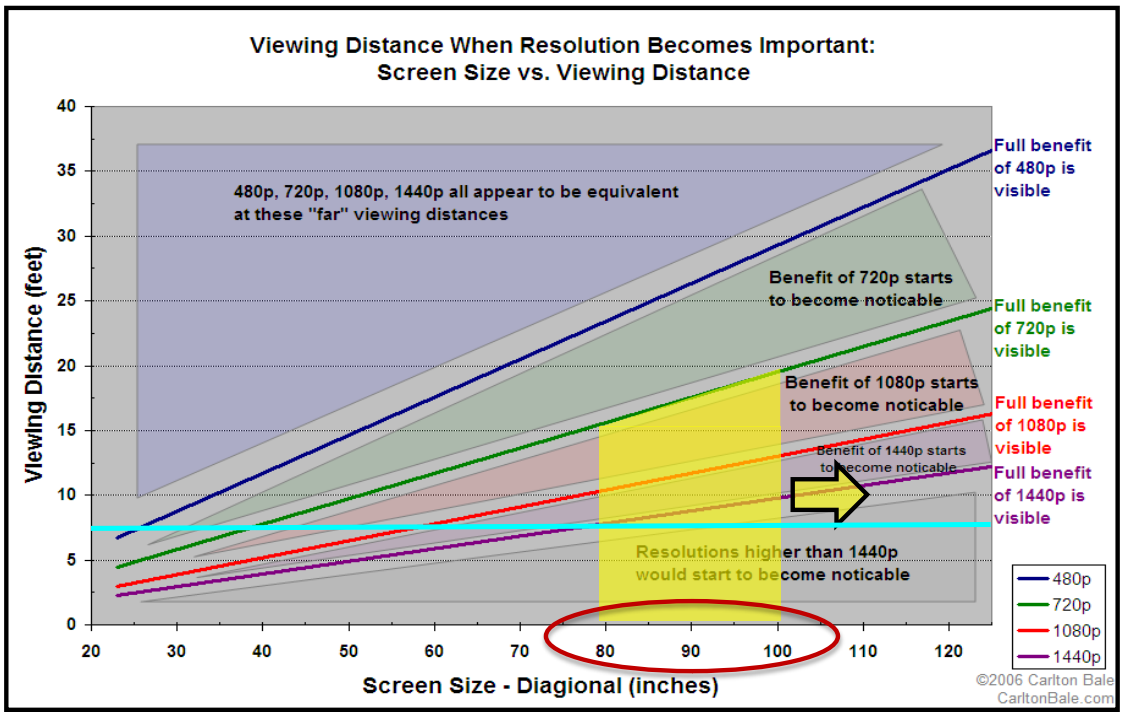


Figure 1 – Screen Size, Viewing Distance, and Spatial Resolution [1]

For example, a 750 MHz system may have 300 SD channels and 150 HD channels. Using MPEG-2 encoding, the spectrum allocated to these channels as broadcast would be about 85 slots – 60 for HD and 25 for SD – for a total of 115. The remaining spectrum can be allocated to On-Demand (VOD) or data services (DOCSIS).

With 300 SD and 150 HD, let's assume half again (75 channels) would be 4K HD broadcasts. Based on Table 1 below [2], those 75 programs would need another 75 spectrum slots encoded as MPEG-4. A 750 MHz system has about 116 spectrum slots in total, so the notion of a simulcast broadcast using MPEG-4 of 4K HD is clearly not in the cards.

Less Filling

MPEG-4 codecs are valuable and mature tools for operators to manage HFC spectrum as HD content and OTT video services continue to grow. However, as was obvious in the 4K HD calculation, MPEG-4 efficiency gains do not offset the resolution increase due to pixel count alone.

MPEG-4 is, of course, not the latest encoding standard to draw on – that would be High Efficiency Video Coding (HEVC), or H.265. HEVC has been successfully shown to achieve its objective of 50% less average bandwidth compared to MPEG-4. It is fair enough to expect that the emergence of HEVC should undoubtedly be turned toward the 4K HD service, and in this case 38 spectrum slots are necessary to support such a broadcast. Not available typically, but we are in a much closer realm of possibility with a combined set of practical set of tools for network evolution, as we shall see.

In addition to linear services, there would be an expected increase in VOD due to 4K HD content being part of the library, roughly doubling the VOD allocation.

Table 1 – Video Streaming Rates [2]

Resolution		Digital Compression Method and Bit Rate		
Resolution Terms	Frame Size / Scanning System / Frame Rate	MPEG2	H.264	HEVC
SDTV	480 / i / 30	3.7 Mbps	2 Mbps	1 Mbps
HDTV	720 / p / 30	6 Mbps	3 Mbps	1.5 Mbps
HDTV	720 / p / 60 (or 1080 / i /60)	12 Mbps	6 Mbps	3 Mbps
HDTV	1080 / p / 60	20 Mbps	10 Mbps	5 Mbps
4K UHDTV	4Kx2K / p / 60	80 Mbps	40 Mbps	20 Mbps
8K UHDTV	8Kx4K/ p / 60	320 Mbps	160 Mbps	80 Mbps

How would 38+ more spectrum slots be possible? Several technology tools similar to those that operators have developed in the past when capacity expansion was required can be exploited towards this goal. In particular in this case, Fiber Deep, which provides new standard spectrum and also expanded DOCSIS 3.1 spectrum, and DOCSIS 3.1 itself, which provides more "effective" spectrum slots, could provide a path to this broadcast 4K HD support.

However, while 4K HD is a step function of new capacity on a relatively full network, video is not the only thing growing in the cable network these days. Wise operators are preparing for all projected service evolutions. And, thinking of video through only the lens of video channels is also no longer sufficient – it now blends with IP data, which has been governed by a persistently aggressive Compound Annual Growth Rate (CAGR).

PERSISTENT AGGRESSIVE CAGR (PACAGR)

Figure 2 captures the breadth of services and applications now common in today's residential service portfolio. The "triple play" has been a valuable bundle, offering an attractive combination of customer satisfaction as well as sound business strategy and opportunity for the operator.



Figure 2 – Residential IP-Based Services Continue to Expand [10]

The IP data component of the triple play has driven many of the new growth applications. Having delivered on the initial game-changer of high-speed web browsing, broadband data service rapidly evolved into the rich content and ubiquity of access we assume today. Operators invested in increasing capacity and speed year after year, as demand grew and applications became more sophisticated.

Many of today's applications take advantage of these ever-increasing network capabilities. Expectations have become gradually raised, and supporting more sophisticated and bandwidth-hungry applications seamlessly is a customer requirement. Key examples include online gaming, Over-the-top (OTT) streaming video (i.e. Netflix), file sharing applications, telecommuting, the exploding use of image and video-centric social media, and access to on-demand programming on all screens, anywhere in the home.

Recently, operators have branched out into whole home services such security, home control and automation, and even energy management. Unlike the prior media/entertainment-centric applications described, these services deliver completely new value to operators and customers. Currently deployed applications are perhaps the tip of the iceberg as the "Internet of

Things" (IoT) possibilities emerge. As envisioned in Figure 3, new service growth opportunities unrelated to typical media-centric and telecommunications services are vast, with promising new categories already underway on or the horizon in telemedicine, health, and wellness.

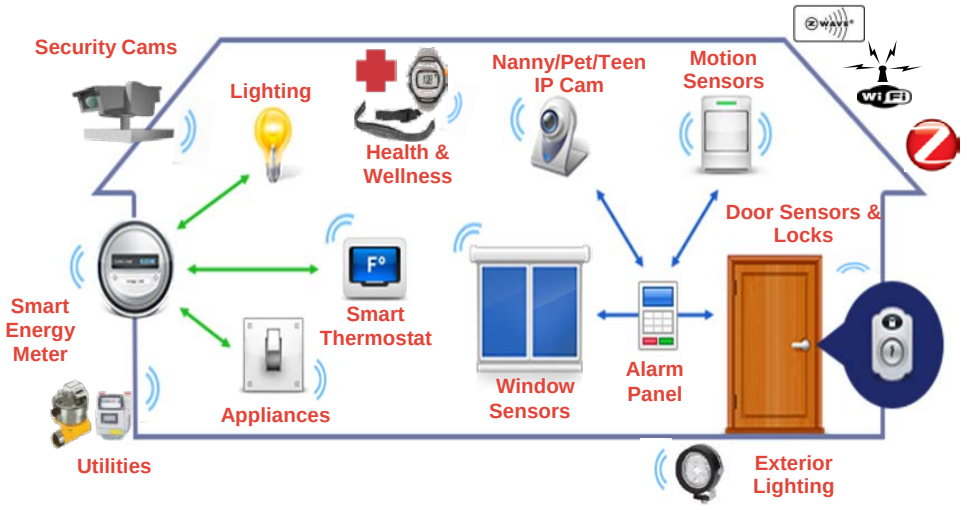


Figure 3 – The Emerging Area of Whole Home Services [10]

All of the above promise to keep PACAGR racing steadily ahead, and perhaps (gulp!) accelerating as we move beyond media delivery, which has finite practical limits of human consumption [3]. And, as noted above, the media consumption component itself continues to grow with the introduction of 4K HD, which in part has to do with the fact that there is room in the human experience for improved video quality – we have not yet achieved the "looking through a pane of glass" experience. Because of this, it remains important to account for these emerging video formats when we analyze long-term capacity management. Herein, that focus is on 4K HD, but higher resolution, 8K HD, formats also exist and could play a future role.

On the bandwidth asset side of the ledger, there is vast, untapped capacity in the HFC architecture. However, it is not infinite. The competing phenomenon of growth and HFC capacity can be observed using the Capacity Management Timeline approach, as shown in Figure 4. This approach represents a simple a tool that allows operators to understand the impacts of PACAGR to HFC resources.

Figure 4 is a basic example using the common 50% CAGR trend on a particular serving group. A 16-channel DOCSIS 3.0 spectrum allocation, operating at half of total capacity (there is runway for growth), follows the red trajectory which compounds at 50% YoY – doubling roughly every 21 months. As the trajectory climbs, node splits occur in 2016 and 2021.

Video and data service evolution occurs in parallel with architecture evolution, and the most daunting of the service evolution steps is the all-IP Transformation. However, this is a an essential transition for the long term, maximizing capacity, simplifying operations, as well as enhancing service flexibility, breadth, and velocity. The path to an IP network nirvana is a complex balance of new services arriving, old services phasing out, and architecture evolutions introduced nimbly.

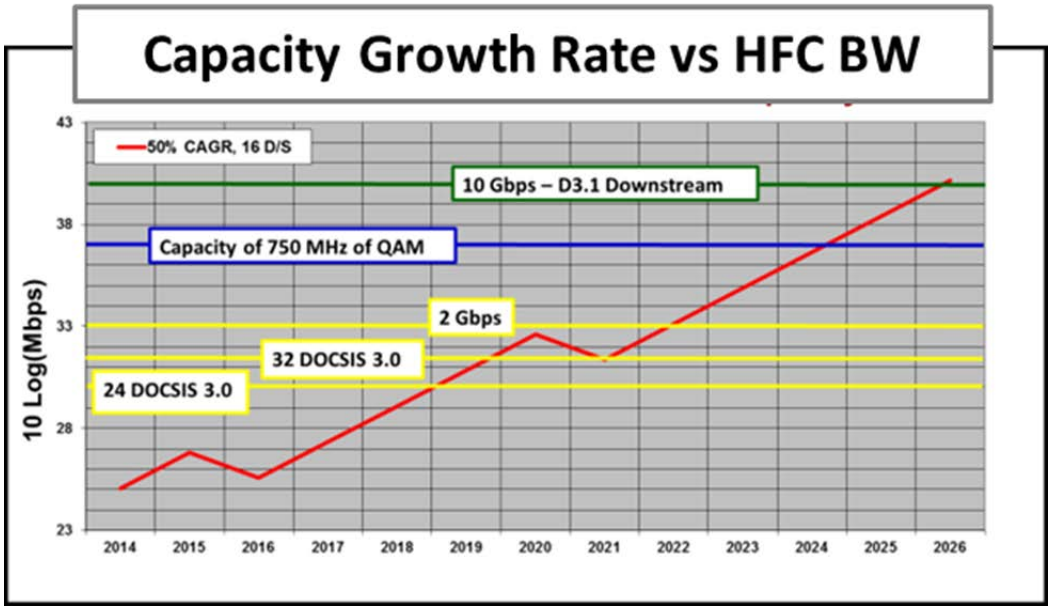


Figure 4 – Capacity Management Timeline Approach: PACAGR vs HFC Capacity [4]

LAST MILES AND MILES

Typical deeper fiber HFC migration has been shown to be a valuable tool for providing continued lifespan runway for video evolution such as HD and OTT, as well as non-video IP data traffic growth [5]. The use of node splitting reaches its final phase when the last active feeding the home becomes a fiber optic node. This architecture goes by various names – Passive Coax, Fiber-to-the-Last-Active (FTLA), or N+0, and the names are not used to refer to the same thing across the industry. Figure 5 illustrates this commonly used multi-phased migration strategy to segmenting a serving area [4].

An "N+0" strategy often involves more than just exchanging amplifiers for nodes, which can be an inefficient way to create a logical N+0. A more strategic implementation optimizes the node placement for reach, balancing the optimal reach with the re-use of existing infrastructure.

Three main benefits of N+0 are:

1) Small serving groups; more average capacity per HHP
2) Opportunity to exploit new coaxial bandwidth; over 60% more downstream and more than doubling of the upstream
3) Higher performance (higher SNR) channel

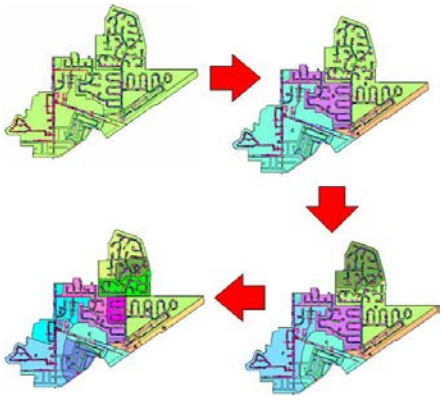


Figure 5 – Service Area Segmentation using Common Node Splitting [4]

The latter two are related to DOCSIS 3.1, which enables 10 Gbps of downstream capacity by enabling use of more spectrum, and by making use of more bandwidth efficient modulation profiles (4096-QAM, possibly higher). DOCSIS 3.1 modernizes the core technology, primarily updating the forward error correction (FEC) and introducing the use of OFDM. N+0 also leaves operators a stone's throw from FTTP.

Figure 6 compares today's 256-QAM modulation with 4096-QAM as enabled by DOCSIS 3.1 [6].

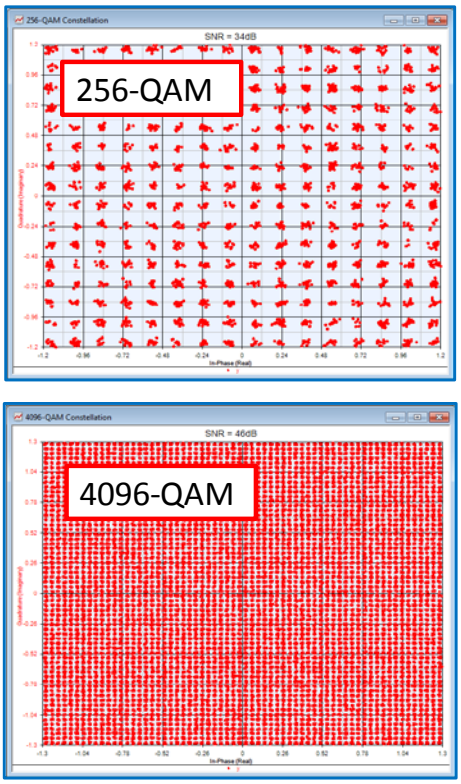


Figure 6 – Comparing Maximum Modulation Profiles of DOCSIS 3.0 and DOCSIS 3.1 [6]

Figure 7 on the following page captures the broad essence of the N+0 or "Fiber Deep" architecture [4].

IP HOME – A HALFWAY HOUSE

IP video access in the customer home and onto customer devices occurs with OTT services as well as MSO-managed services, with operators developing apps for second screens that connect to their networks to authenticate devices and enforce content rights. Traditional STBs will remain in the field for some time. However, as part of the IP transition, as they are removed there is no longer an intent to replace them with a more sophisticated version integrated into the QAM ecosystem.

Hybrid gateways, IP gateways, and IP client STBs, operated from the cloud, is the realm of the majority of new "STB" development. As this transition takes place, both QAM and IP networks (DOCSIS, MoCA™) will be running simultaneously on the coax in the home, as will WiFi™, of course. We are "half-way" to the IP conversion in the home. However, this hybrid state is common during the technology phase of any transition.

Traditional HFC allows for placement of CPE anywhere in the home that a coaxial outlet is located. However, in the case of HFC spectrum re-allocation, spectrum expansion, or even future FTTP extension, a network demarcation point may be established at the point of entry. In this sense, it is a natural "two box" system of network termination device plus in-home CPE.

A whole-home hybrid architecture that utilizes MoCA™ and WiFi™ to deliver video over IP on the Home LAN while passing through QAM is as shown in Figure 8 [10].

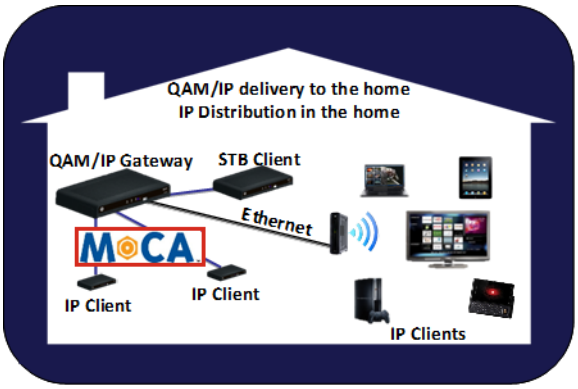


Figure 8 – Cable Home LAN Today [10]

With current HFC, MoCA™ bands are implemented that do not overlap with the spectrum used by current services, because today's CPE terminate the HFC access network inside the home. The Home coaxial LAN, as stated, is part IP – there is an IP Home LAN, but it is a shared resource with conventional QAM signals today.

With the network terminated at the point of entry, the entire coaxial infrastructure in the home can be allocated to the IP Home LAN, and lower MoCA bands, and higher in-home speed and capacity are enabled. Of course, to make this a practical reality, existing QAM digital video services and STBs must be removed from the network. The "pass through" of legacy services shown in Figure 8 must be eliminated. In other words, this architecture is well-suited to the end-state of the all-IP transition.

In many ways, the implementation and turning on of IP-based video services is the easier part – the more difficult and complex part of the IP Transition is the business-driven aspect of retiring legacy QAM video.

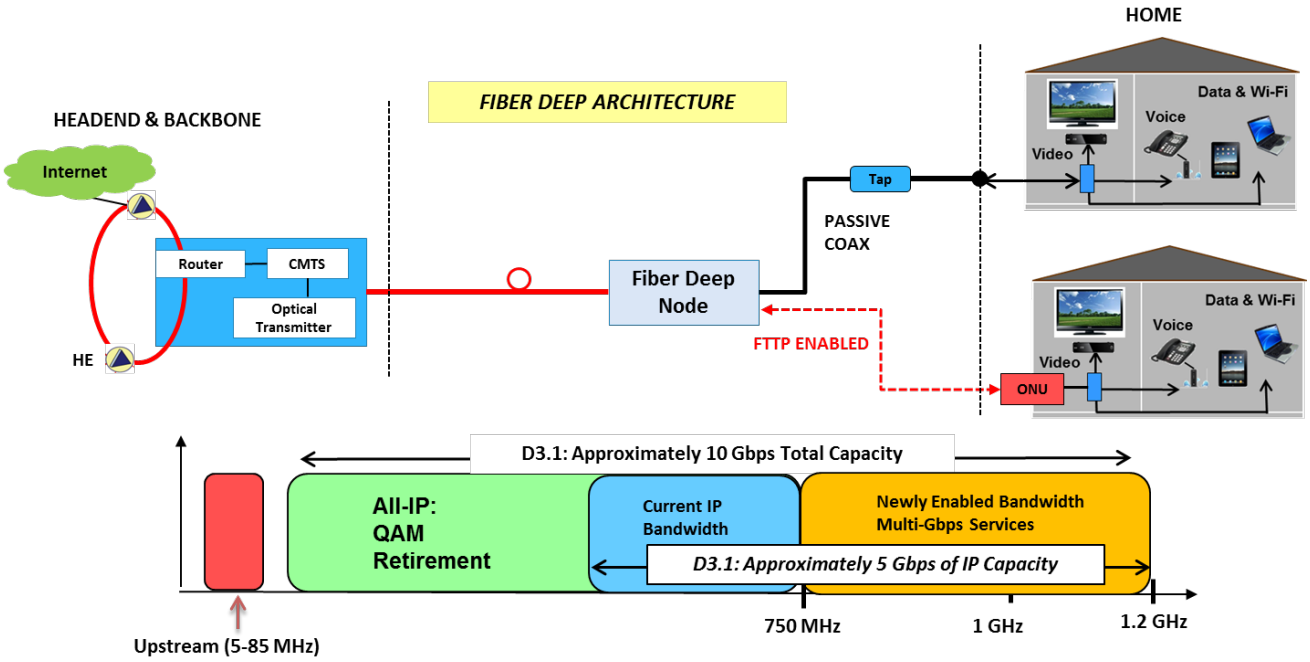


Figure 7 – The Essence of the N+0 or "Fiber Deep" Architecture

IPTV IN THE ACCESS NETWORK: THE LAST FRONTIER

The reasons for cable to migrate to IP for video, or IPTV, are numerous and varied, and have been the subject of many industry papers in the past, and we will not delve into them here. Some key ones are listed below:

- Multi-screen video – IP Devices and Home LANs
- Access to content anywhere
- Blending video with other services (Social media, merchandising, targeted ads)
- Cost efficiencies of single unified network operations
- Leveraging of the global development and equipment ecosystem
- Access agnostic for similar IP-based systems structured on OSI principles
- Software-based digital content rights and management (DRM)
- Flexibility of IP traffic engineering at the edge
- Standard remote management tools

When the network is all-IP, content and transport will have no limiting relationships. Video service is abstracted from access architecture, and the network is consolidated around a single unified delivery system.

The transition to an all-IP network is an end-to-end effort. However, the "transition" largely left to take place is almost exclusively in the access network, as epitomized by Figure 9 [10]. Video gets to distribution hubs and Headends using IP/Ethernet delivery. Behind the access network edge right up to the EdgeQAM or CMTS, IP/Ethernet is the dominant transport and hardware implementation. Broadcast and narrowcast video is delivered to the edge in this format, whereupon it makes its way onto the HFC network as MPEG-TS only.

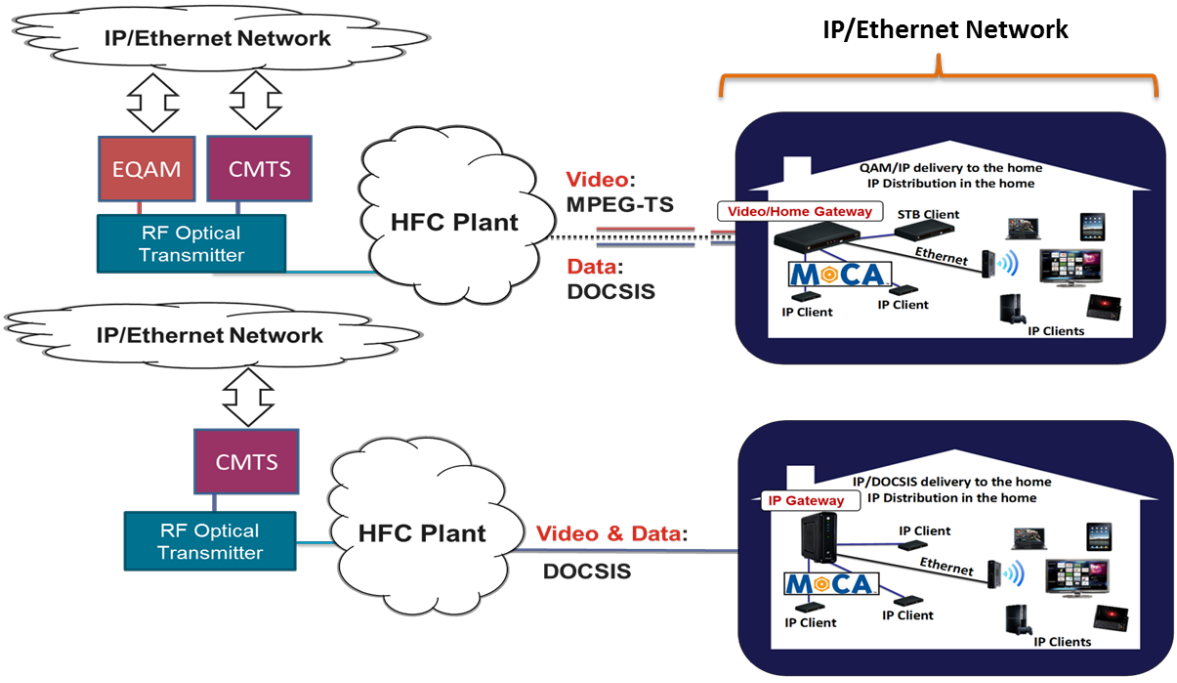


Figure 9 – End-to-End IP – Most Work Remains in the Access Network [10]

In the HFC access network, the IP transition has two major implications. First, the transition of all content to IP eventually enables the potential to eliminate legacy analog and QAM carriers. There will be a period of simulcast of existing video services and IP-based content, but the maximum benefits of the IP transition occur when all legacy spectrum is recovered and freed up for use as IP capacity.

A second implication of the all-IP conversion is the ability to distribute video over any other IP-based transport mechanism – in particular other RF (Wi-Fi™, MoCA), Ethernet, and future FTTP, such as Ethernet PON (EPON).

IP Video – Multicast, Multiplexing & Variable Bit Rates

IP video over DOCSIS 3.0 or DOCSIS 3.1 makes available techniques that provide valuable efficiencies over tradition broadcast QAM – multicast and wideband IP traffic multiplexing gains.

Legacy QAM architectures squeeze a fixed number of video signals in a 6 MHz QAM slot with the help of video statistical multiplexing to make the most of the roughly 40 Mbps per slot. IP Video can be implemented using variable, adaptive bit rates, over bonded DOCSIS 3.0 channels or wider DOCSIS 3.1 spectrum. IPTV over 16 bonded DOCSIS 3.0 channels, for example, is a resource pool of 640 Mbps.

Various analysis have estimated the incremental efficiency associated with wider spectrum and natural IP traffic multiplexing [8]. Bonding, coupled with more streams per QAM of MPEG-4 or HEVC benefits from the law of large numbers, reducing average bandwidth. Many independent streams over *much* more pipe capacity results in effective self-averaging.

Based on simulations and observations of the prior analysis, we use an 80% scaling as the bandwidth required for VBR-based channel bonded DOCSIS 3.0 video in comparison to single carrier QAM transport.

A second valuable benefit is the use of IP Multicast, which means that only when a program is asked for is it put onto the network, in contrast to broadcast channels. Others in the service group access the same multicast stream when they tune to the program, saving access bandwidth.

Trends have been moving towards more unicast for many years and for many reasons:

- Time-Shifting
- 2nd Screen viewing
- Smaller service groups
- Targeted advertising
- Voluminous on-demand offerings

Nonetheless, the most popular programs and channels tend to always benefit from multicast for even shrinking service groups at prime time, although the number of channels this applies to can be surprisingly small – the top 20-30 in the lineup, for example. In IP, of course, we are streaming programs, not channels, which is another layer of granularity and potential for dynamic flexibility.

Figure 10 is an example of the type of analysis showing the dependency of IP multicast to number of concurrent users and content mix. Note that this analysis is based on MPEG-2 video encoding. Since the analysis of Figure 10 in [9], MPEG-4 has become mature and is expected to be the dominant compression format initially for IPTV.

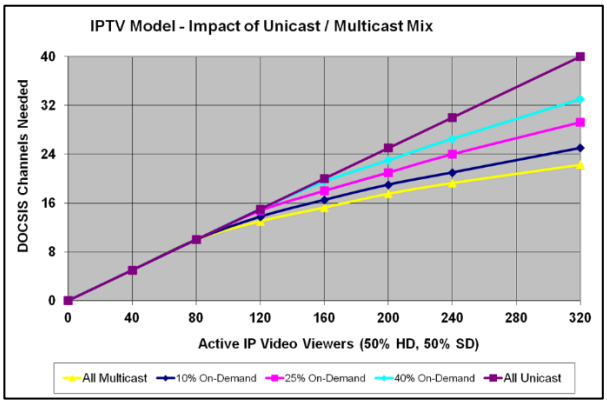


Figure 10 –IP Multicast Efficiency Analysis [9]

An important premise of determining IP Video capacity needs is to recognize that ultimately what matters is the number of eyeballs at primetime. The rest is just mathematical scaling. Customers are agnostic to delivery method and what is underneath the hood.

A WALK IN "IP" PARK

Figure 11 pulls together a possible transitional spectrum map integrating the items we have discussed herein and more. While the approach is to incrementally evolve services and technology in phases such as the eight steps in Figure 11 below, recognizing the impracticality of a straight "cutover" in most cases, the steps shown are still relatively coarse. The primary focusing was to introduce impacts of key technology components while having a reasonable number of permutations, but enough to create a useful timeline of guidance with the chart.

We describe the steps introduced in Figure 11 below.

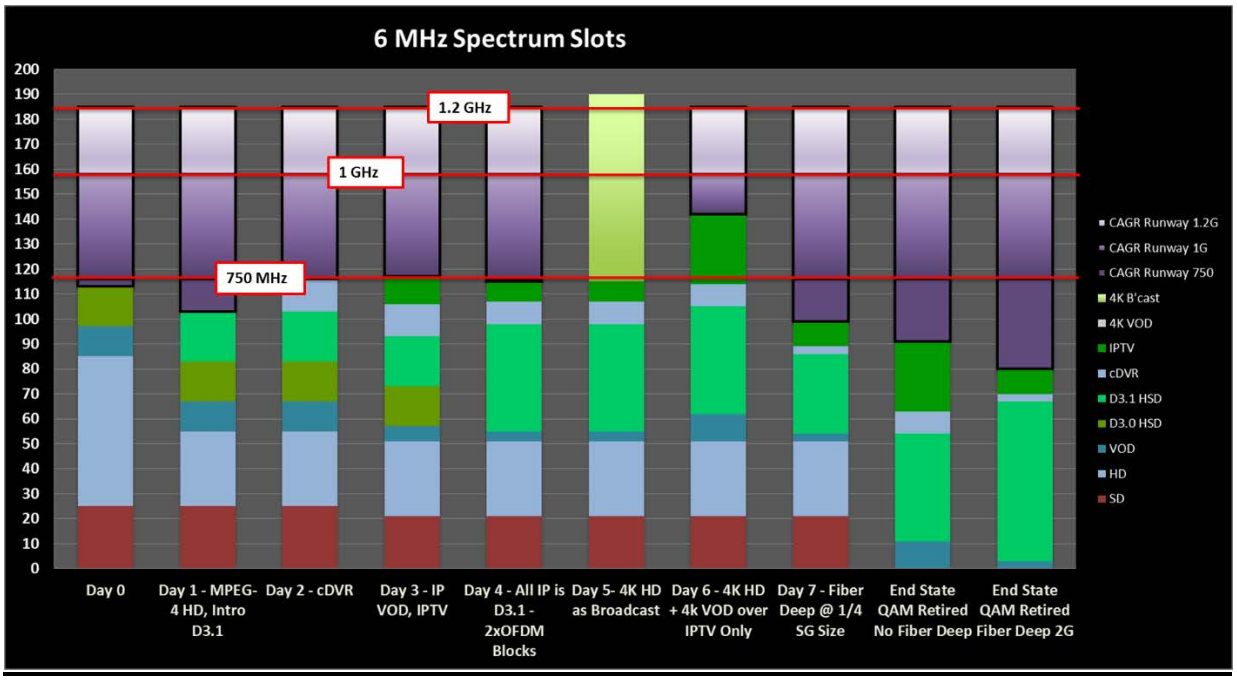


Figure 11 – Spectrum Transition Example Involving 4KHD, IPTV, Fiber Deep, and D3.1

Day Zero:

An HD channel typically has an SD brother in the channel lineup – it's a simulcast. This is not very efficient, but also not terribly penalizing given the modest relative bandwidth that the SD consumes.

As a starting point, we assume 300 SD and 150 HD channels of broadcast, 12 channels dedicated to VOD, and 16 channels for DOCSIS 3.0. It is apparent already (and not surprisingly) that this lineup consumes most of a 750 MHz HFC plant. All of the spectrum is being monetized for services.

Day 1: MPEG-4, DOCSIS 3.1

The introduction of a substantial chunk of spectrum for DOCSIS 3.1 is shown occurring using spectrum mined from transitioning HD to MPEG-4 for HD content.

Some of the spectrum reclaimed from the QAM pool is used to introduce DOCSIS 3.1 and enable Gigabit data service.

Note that the fact that DOCSIS 3.1 uses the 6 MHz spectrum slots more efficiently is embedded in the accounting of Figure 11 throughout all steps that incorporate DOCSIS 3.1

Day 2: Cloud DVR

Cloud DVR is an emerging service that continues the MSO move to IP-based video service. The impacts of cDVR can be projected mathematically based on available data on DVR penetration, video consumption behaviors, and content type.

We will assume an initial recorded viewing percentage at 33% based on [11]. We use 60% video services penetration, 70% peak busy hour (pbh) usage, 1.5 streams per home time-shiftable, and the 33% concurrency of recorded consumption.

We also assume *all* DVR customers have exchanged local storage service for cDVR service. Again, this transition will occur in phases in practice.

Day 3: IP VOD, IPTV

The move of VOD to IP VOD has occurred in this phase. Again, this will happen in phases, but in Figure 11 we show only all-QAM and all-IP cases, with the assumption of equivalent eyeballs meaning an efficiency improvement for IP VOD due to MPEG-4 encoding basis.

In addition, IPTV using IP Multicast is activated as a simulcast service using Figure 10 as a guide for required DOCSIS channels.

To remain contained in a 750 MHz network, the slots allocated to SD services was reduced. The reduction shown is about 16% on average, so more efficient use of existing compression may be sufficient. Of course, a full 50% is obtainable if desired with investment for the SD tier as well.

Day 4: Release 1 of DOCSIS 3.1

Use of DOCSIS 3.1 expands to the two full OFDM blocks per the initial capability requirements. DOCSIS 3.0 equipment has been replaced (again, a phased transition in practice).

Day 5: Intro to 4KHD

We consider the case of introducing 4KHD as a simulcast/broadcast, using HEVC compression.

Day 6: IP 4KHD

We remove broadcast 4KHD, and implement 4K VOD and linear as IP only using HEVC encoding.

Day 7: Fiber Deep

A Fiber Deep architecture (N+0) is considered. The architecture guidelines for Fiber Deep cap the maximum serving group size at 128 hhp.

End States

Both nominal service group size and a Fiber Deep architecture (N+0) are considered as the legacy QAM services are retired.

7 BUSY DAYS – SUMMARY

The *Day 0* state points out the obvious, using the 750 MHz network assumption – operators wisely use nearly all of the spectrum at their disposal for revenue-generating services.

Day 1 makes clear an important and oft-overlooked component of HFC capacity management. That is, while PACAGR garners all of the headlines, the industry has hardly stopped innovating in the video compression space. Since the deployment on a wide scale digital video over cable using MPEG-2 compression, a full two generations of MPEG compression have been completed and become deployable. The vast majority of non-IP based digital cable spectrum is still MPEG-2 based. Obviously there is a STB dependency that makes rapid changeover difficult, but a carefully managed changeover is quite reasonable and underway with many MSOs.

Day 1 indicates the effect of MPEG-4 on the HD broadcast spectrum only. Some of the reclaimed spectrum of the 50% HD savings is allocated to deploy DOCSIS 3.1 with enough capacity that bonded DOCSIS 3.0 and DOCSIS 3.1 channel achieve Gigabit data services.

Note that the net of MPEG-4 conversion and DOCSIS 3.1 introduction is an *increase* in spare spectrum, largely because DOCSIS 3.0 already existed and is providing a good percentage of the total capacity needed to support 1 Gbps.

In *Day 2*, we can see that the spare spectrum set aside is sufficient (not accidentally!) to enable Cloud DVR services (cDVR) at 100% penetration of DVRs. The service is assumed supported over DOCSIS 3.0, as the initial production of these DVRs it is assumed occurred in the DOCSIS 3.0 era (i.e. now)

By *Day 3*, we assume the transition of VOD to all-IP has occurred. Viewing behaviors are assumed held constant, and the IP version benefits from the use of MPEG-4. The introduction of IPTV occurs, based upon analysis as shown in Figure 10 for multicast gain and channel count required (spectrum).

As mentioned, the SD tier was reduced in order to make enough room for linear IPTV service. This can be a sacrifice of SD content, perhaps acceptable this many years down the road, squeezing better efficiency out of existing MPEG-2 compression, or of course MPEG-4 encoding is available for the SD tier if the investment is worthwhile to be made on SD services by this stage.

In *Day 4*, all of the IP services – HSD, cDVR, IPTV, VOD are over DOCSIS 3.1 only. The added spectrum efficiency of DOCSIS 3.1 allows all of these to consume less total spectrum.

However, HSD capacity growth and speed requirements are expected to continue as these phases (years) go by. All of the freed up spectrum is given to DOCSIS 3.1. The total number of DOCSIS 3.1 is increased to allow for the two full OFDM blocks of 192 MHz to be deployed. This now just fits within the 750 MHz envelope (again, not by accident!).

On *Day 5*, we can see the extreme scenario of the 4KHD dilemma. It displays the "what if" a broadcast simulcast of 4KHD (over HEVC) sat alongside SD and HD content. This does not make much sense to consider, and there is likely no one with such a plan. Figure 11 simply makes this abundantly clear.

On *Day 6*, however, we begin to see a pixel of light at the end of the 4KHD tunnel when we deploy 4KHD linear video and VOD over IP. While we have stepped over the 750 MHz boundary, we are comfortably within the 1 GHz boundary of existing HFC equipment and infrastructure with well-understood upgrade cost and logistics.

Alternatively, we have the *Day 7* approach to HFC evolution involving Fiber Deep as shown in Figures 5 and 7. The assumption of Fiber Deep in this case is a service group reduction by a factor of four, effecting the number of channel required for all narrowcast (unicast or multicast) services. This can be calculated. Note that spectrum allocation associated with a peak speed requirement (in this case, a 1G assumption).must still be maintained despite shrinking service groups and less total capacity utilized.

As can be seen in Figure 11, each of the evolutions of Day 6 (1 GHz upgrade) and Day 7 (Fiber Deep) are possible solutions to 4KHD capacity management. It is straightforward to show that the broadcast 4KHD case does not fit within 750 MHz when "Fiber Deep" is implemented. However, Fiber Deep in the form of N+0 and strictly NO amplifiers is inherently a 1 GHz architecture based upon a new Fiber Deep node. In this case the 4KHD broadcast is actually supported. So, HFC could support a 4KHD broadcast if necessary.

However, given that the IP transition is in process, just as 4KHD comes to life alongside HEVC, and given the much higher spectrum efficiency possible deploying over IP and with DOCSIS 3.1, it is understandable how we can look at 4KHD as a logical stimuli to the movement of video on the network to exclusively IP-based

End State makes the correlation between the scaling of 4KHD content and service, and the IP transition, abundantly clear. In this case –Fiber Deep or no Fiber Deep, retiring QAM spectrum and clearing its use for IP only services, delivered over maximally efficient DOCSIS 3.1 is the HFC Network Nirvana.

Note also that in this final state, which is assumed to be some years away, a D3.1 spectrum allocation supportive of 2G speeds is included, still comfortably.

End State – No Fiber Deep – make a pretty compelling case as to why the transition to IPTV includes the technology evolution of full video services over IP, but also the retirement of legacy video services, eventually eliminating broadcast/simulcast spectrum inefficiency. Ultimately, this simply makes the network traffic completely correlated with customer demand, which is significantly skewed today because of broadcast video services.

CONCLUSION

Cable operators are well underway in their transition to all-IP. However, it is along road. Behind the edge, throughout the home, and onto second screens are all mature aspects of IP video for cable operator (non-OTT) programming as well as OTT itself. This paper hopefully described some of the

key puzzle pieces, and shows an example instructive pathway – a mix of technology and service evolutions – that can serve as coarse guidance of the "how-to" of IP transition management.

References

[1] CarltonBale.com

[2] Emmendorfer, Mike and Tom Cloonan, *Nielsen's Law vs. Nielsen TV Viewership for Network Capacity Planning*, Cable Show Spring Technical Forum, Los Angeles, CA, Apr 29-May 1, 2014.

[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, *Growth Architectures: Built to Last, Built to Launch*, Cable Show Spring Technical Forum, Los Angeles, CA, April 28-30, 2014.

[5] Howald, Dr. Robert L, *Looking to the Future: Service Growth, HFC Capacity, and Network Migration*, 2011 Cable-Tec Expo Capacity Management Seminar, sponsored by the Society for Cable Telecommunications Engineers (SCTE), Atlanta, Ga, November 14, 2011.

[6] Howald, Dr. Robert, *The Grown-Up Potential of a Teenage PHY*, Cable Show Spring Technical Forum, Boston, MA, May 21-23, 2012.

[7] Howald, Dr. Robert L and Dr. Sean McCarthy, *Accounting for Techies: Taking it to the Ultra*, The Cable Show, June 10-12, Washington, DC, June 10-12 2013

[8] Howald, Dr. Robert L, Dr. Sebnem Zorlu-Ozer, Dr. Nagesh Nandiraju, and Erik Metz, *Web Surfing to Channel Surfing: Engineering the HSD Edge for Video*, 2009 Cable-Tec Expo, sponsored by the Society for Cable Telecommunications Engineers (SCTE), Denver, CO, Oct 28-30, 2009.

[9] Howald, Dr. Robert L and John Ulm, *Delivering Media Mania: HFC Evolution Planning*, 2012 SCTE Canadien Summit, Toronto, Ontario, March 27-28.

[10] Howald, Dr. Robert L and Robert Thompson, *Tomorrowland®: Customers, Services, Networks*, 2014 Cable-Tec Expo Capacity Management Seminar, sponsored by the Society for Cable Telecommunications Engineers (SCTE), Denver, Co, September 22-25.

[11] www.nscreenmedia.com/tivo-owners-use-time-shifted-viewing-third-time/

A SYSTEMATIC APPROACH TO VIDEO QUALITY ASSESSMENT AND BITRATE PLANNING

Sean T. McCarthy, Ph.D.
ARRIS

Abstract

We present a streamlined method of setting operational video quality and bandwidth using either subjective or objective testing, using individual golden-eyes or focus groups of any size. A key feature of the method we present is it provides a means of validating test content to ensure that it is representative of actual programing. Another key feature is that this method enables bandwidth planning based on both average video quality and the probability that hard-to-compress content will fall below a predetermined just-acceptable video quality threshold. The data and analysis we present are intended to aid in planning video quality and bandwidth resources across a range of service offerings from OTT through Ultra HD.

INTRODUCTION

Cable operators and other MVPDs can significantly reduce the complexities of providing ever more viewing options to subscribers by streamlining their approaches to setting video quality thresholds and the minimum bitrates for achieving those thresholds across all streams.

As pay TV providers continue to pursue ways to improve bandwidth efficiency through better encoding techniques on their legacy SD and HD channels, they must also be able to set new quality and bandwidth benchmarks in response to subscriber demand for access to premium content on every type of IP-connected device. This entails working with multiple adaptive bitrate (ABR) streaming modes to determine minimum bitrates for delivering content over fixed and wireless outlets at acceptable quality levels for each type of device.

At the same time, service providers are exploring how to benefit from the introduction of HEVC (High Efficiency Video Coding) compression technology in the IP and legacy domains. This includes preparations for migration to the next plateau in display resolution, 4k Ultra HD. Quality assessment processes are integral to these efforts as well.

These developments present unprecedented challenges to performing systematic, scientifically valid quality assessment at a time when content licensing policies and new approaches to monetization require consistent cross-platform quality performance. Cable and other network operators must be able to define acceptable viewing parameters for every type of service and set the minimum bitrate requirements for whichever encoding protocols they're using to deliver those services.

As part of this process they must be able to comparatively determine which encoding solutions achieve the best results. And their procedures must include a straight-forward approach to determining whether the technical nuances of the test files they use in focus groups and other testing forums mirror the full scope of nuances in motion, color, zooms, and other dynamics to be found in real programming.

Meeting these challenges requires a practical approach to video quality assessment that is fast, rigorous and cost-effective. This can't be done simply by relying on objective methods like PSNR (Peak Signal-to-Noise Ratio), which were not designed to gauge the actual human viewing experience[1]. Yet there is nothing quick or cost-effective about repeatedly going through all the time-consuming and costly steps associated with a

comprehensive approach to the subjective MOS (Mean Opinion Score) method of assessing real viewing experience, as prescribed in guidelines set by the ITU's BT.500 video assessment recommendations[2].

Instead, operators must be able to take a much more streamlined approach to perceptual video quality assessment which still can be relied on to generate accurate results from the MOS method of measuring real human experience while affording operators the flexibility to employ objective measures wherever appropriate to specific assessment goals. This simplified approach should generate quantifiable results that can be easily communicated in data-driven discussions about video quality on all services. This means the methodology must be sufficiently comprehensive to generate a matrix of average test scores at multiple bitrates across a wide range of video sequences representing the gamut of variables in each service category.

By compiling results from all test sequences in a systematic, quantifiable way operators will be able to use analytic techniques to quickly find answers to different types of questions. For example, operators should be able to use the scores from average viewing experiences at various bitrates across multiple video test sequences to calculate optimal quality thresholds. Looking at test scores generated at different levels of complexity in the video test sequences, they should be able to calculate what the operational bitrate threshold should be for achieving the target quality for a given class of video service.

Operators also want to be able to identify the sequences that have the most negative impact on overall quality performance so that they can focus on finding solutions that will do the most good in raising quality levels for any given bitrate or, alternatively, lowering the bitrate for achieving a given level of quality. They want to be able to identify and quantify the differences in results generated by different encoding systems across multiple test sequences at multiple bitrates. They want to determine whether improvements enabled by new solutions are of sufficient magnitude to merit replacing solutions currently in use.

As explained in the ensuing discussion, extensive commercial experience with a wide range of quality assessment procedures has led to development of a scientific approach to video quality assessment and bitrate planning that will allow operators to achieve highly reliable results in far less time and at far less expense than was previously the case. This is not a one-size-fits-all approach. It provides leeway for using various combinations of subjective and objective testing with reliance on scoring from one or more outside focus groups or from just one individual "golden eye" on the operator's staff.

In all cases, the method involves a means of validating test content to ensure it not only represents the full scope of video characteristics in actual programming but also introduces uncommon extremes that serve to stress test encoders. The method also supports bandwidth planning based on average video quality while accommodating the probability that hard-to-compress content will occasionally fall below a predetermined, just-acceptable video quality threshold.

NEW CHALLENGES IN VIDEO QUALITY ASSESSMENT

From a cable perspective, not very long ago the tasks associated with defining minimum levels of acceptable video quality and the bitrates required to hit those targets focused on just two types of video streams – MPEG-2-encoded SD and HD programming delivered over QAM channels to set-top boxes. Today, operators must also be able to perform quality assessment and bitrate planning on those legacy channel streams

when MPEG-4 H.264 encoding is used. And they must be prepared to apply similar assessment procedures to HEVC H.265 encoding as well.

HEVC and 4k UHD

Encoding methodologies applied with MPEG-2 and MPEG-4 are relatively stable, which means operators need to occasionally look at improved encoding techniques to determine whether such improvements translate into gains in bitrate reduction that justify the costs of replacing existing systems. But, where HEVC is concerned, the protocol is still in a relatively unstable phase of implementation, which means assessments of video quality for purposes of setting bitrates will have to be made more frequently in the early rollout phases and beyond.

This is especially true in the use of HEVC to deliver 4k UHD services, which will most likely be a primary driver to early use of HEVC in cable and other MVPD networks. The emergence of 4k UHD requires the setting of new quality thresholds and bitrates. Given the increase in bitrates required for 4k even with the use of HEVC, incremental improvements in encoders will be even more significant to operators than has been the case with SD and HD encoding. As operators add ever more content to their 4k service portfolios, updated video quality assessment will be vital to ensuring maximum bandwidth efficiency as HEVC matures.

Streamed IP Video

Greatly complicating matters is the need to deliver premium video over ABR streams to smartphones, tablets, computers, game consoles, IP set-tops, and smart TVs over Wi-Fi and mobile 3G and 4G connections. This introduces new form factors requiring different bitrate settings to reach minimum quality thresholds, depending on which encoding protocols are in use.

To determine what the appropriate minimum quality thresholds should be for each category of device requires that operators be able to quantify measured video quality in a way that accurately correlates with the quality of actual human viewing experience in accord with widely accepted parameters. In so doing they can determine what the minimum bandwidth requirements will be for a given range of bandwidth contention levels across multiple types of devices, thereby avoiding excessive spending on network capacity as they seek to accommodate the multiscreen IP service paradigm.

Complicating matters is the fact that configurations required for ABR distribution are different from standard broadcast configurations. Interlaced broadcast video must be converted to progressive scan. Frame rates and horizontal and vertical resolutions must be adjusted to comport with the types of devices targeted by the ABR streams. How these processes are performed in today's automated transcoding systems can impact the quality of the viewing experience and so must be accounted for in analyzing the results from tests used in video quality assessment.

OVERVIEW OF VIDEO QUALITY ASSESSMENT METHODS

Much attention has been devoted to video quality assessment over the decades. There are 2 main kinds of tests: subjective and objective. Subjective methods use systematic and repeatable psycho-visual tests to quantify human judgment of video quality. Objective methods rely on computer-based algorithms to attempt to predict human judgment.

When testing video quality, it is important to understand that "video quality" is an imprecise term. What is really important is being able to answer specific actionable questions such as:

- "What is the minimum operational bitrate that will provide a good video quality of experience to my subscribers?"
- "How often might my subscribers notice problematic video quality?"
- "Is encoder A better that encoder B?"
- "Did the latest encoder upgrade improve video quality or allow a lower operational bitrate?"
- "What is standing in the way of better video quality and lower bitrates?"

For all video quality assessment methods – subjective and objective – the reliability of the results obtained depends crucially on the extent to which the test clips represent the full range of characteristics common to video content. Thus, there need to be clips devoted to emphasizing high motion, random motion, horizontal and vertical pans, zooms and so on. Moreover, the clips chosen from test libraries should go beyond conveying the nuances of these characteristics as they appear in average programming by creating stress conditions suitable to testing the limits of encoder performance.

Subjective Methods

Subjective methods are the golden standard by which all other methods of video quality assessment must be compared. Yet, there is not just one type of subjective test. Which test should be used depends on the question the experimenter hopes to answer.

There are two international standards that give guidance on subjective tests and the methods that should be used to analyze results. ITU-R BT. 500 is a formal recommendation for television. ITU-T P.910[3] is explicitly intended for multimedia applications such as video conferencing, telemedicine, etc.; but may also be applied to television (particularly as the lines between traditional TV and Internet video blur). The aim of all standardized tests is to produce quantitative data such as Mean Opinion Scores (MOS).

Standardized test methods may be classified in various ways. Some display test video in simultaneous pairs either side-by-side or top-to-bottom on one or two displays. Other tests display video test sequences sequentially, thus relying on a viewer's visual memory. An important way in which subjective methods may be classified is based on the manner in which pristine reference video is introduced. In some testing protocols, the reference is "explicit" and the viewer knows which video sequence is the reference. In other protocols, the reference is "hidden," meaning that the viewer does not know which video sequence is the reference. When no reference video sequence is provided, the test may be said to use "implicit" reference, which is usually based on the test viewers' previous experience and expectations with respect to the video quality of commercial television.

The standardized method called Absolute Category Rating (ACR) is worth highlighting as it inspires the method we describe in this paper. The ACR method is a sequential, implicit-reference subjective method. Short test video sequences are displayed and the viewer(s) scores the quality of the video after each test sequence. Each individual test sequence is a particular short clip (10 seconds is a good length) processed or compressed to a different extent. The advantage of ARC is that it is relatively easy to set up. It is also similar to the way in which people watch television. People never see and compare television programming to a pristine reference. A potential disadvantage of ARC is that special care needs to be taken to create a range of test video sequences that include discontinuities such as scene changes, fades, graphic scrawls, etc. typically seen in commercial television.

In addition to the subjective testing methods specified by international standards,

there are non-standardized methods that are routinely used to make business decisions involving multi-million dollar equipment purchases. One example of a non-standard method is the "side-by-side bake-off" in which two versions of test video are compared on side-by-side displays. Often the evaluator will go "eye to the glass" to inspect tiny artifacts at very close distances, much closer than any usual viewer. The advantage of this approach is that it is quicker and easier to implement than any standardized method. It also enables an evaluator to compare different video processing options – often products from competing vendors – to highlight and isolated differences with great precision. The disadvantages are that it does not produce quantitative data nor does it give a good indication of how a subscriber or other viewer might react to the artifacts.

It is important to recognize that subjective testing creates an unnatural television viewing experience that can impact how the results should be interpreted. When someone is watching television for the sake of watching television, that person is not actively looking for artifacts. Rather, the viewer notices poor quality when the artifacts are significant enough to grab the viewer's attention. In subjective testing, a viewer's tasks are to inspect the television display(s), mentally classify the importance of the any distortions that can be seen, and physically record scores. Thus, subjective testing provides information that marks the boundary of what a viewer could see rather than providing direct information about what a viewer would see during normal television viewing.

Objective Methods

Subjective testing can never be as fast and automatable as objective computer-based algorithms. That is the major advantage of objective testing and why it is so attractive and tempting to use instead of subjective testing.

Objective methods and metrics may be classified as either fidelity-centric or perceptual-centric. An objective metric can be said to be useful if it is accurate and precise, is consistent across different kinds of content, and varies monotonically with subjective scores.

Fidelity-centric methods are the quickest and easiest to use. They quantify the numeric difference between processed and unprocessed video sequences. Peak signal-to-noise ratio (PSNR), sum of absolute differences (SAD), and mean-squared error (MSE) are very common fidelity-centric metrics that accumulate pixel-by-pixel differences. The Structural SIMilarity (SSIM) index[4] and the related multiscale-SSIM (MS-SSIM) analyze correlations between groups of pixels in a test and reference. SSIM and MS-SSIM are moderately more computationally demanding than PSNR but tend to be better predictors of subjective scores and have thus been gaining in popularity in recent years.

Perceptual-centric methods are those that model human vision. Research and development into perceptual models has been very active for decades and is ongoing. Among the better known and more commonly used perceptual-centric methods [5-9] are: Video Quality Metric (VQM); Sarnoff JND (JND borrows from the psychophysical term "just noticeable difference"); and Visible Differences Predictor (VPD).

PSNR, SSIM, MS-SSIM, VQM, VPD, and Sarnoff JND are full-reference metrics (FR), which means that test videos are compared to unprocessed reference video. As such, full-reference metrics are analogous to subjective tests that use explicit or hidden references. Consequently, full-reference may be thought of directly addressing the same class of video quality questions: i.e., "How noticeable is the difference between processed and unprocessed video?" rather than "What is the

video quality my subscriber will normally experience?"

Precise spatial and temporal alignment between test and reference videos is a significant issue when using full-reference methods. Being off by a single frame or by a single pixel vertically or horizontally can have a huge impact that could lead to erroneous conclusions. Many objective video quality products have automated alignment methods, but misalignment errors can be so significant that close attention should always be paid to it. Particular caution must be used when the reference and test videos are at different resolutions, different interlacing, different frame rates, or there are skipped, dropped, or repeated frames in a test video.

Because of the recognized issues using full-reference metrics, significant research and development has gone into developing reduced-reference (RR) and no-reference metrics (NR). Reduced-reference means that some metadata concerning the reference video is made available to the video quality analyzer rather than the reference video itself. No-reference means that the video quality analyzer has no access whatsoever to the unprocessed reference video. As such, no-reference metrics are most closely analogous to someone watching television in a normal way.

As tempting as objective methods are, they should be used judiciously and the results interpreted cautiously. No full-reference, reduced-reference, or no-reference metric can yet take the place of people looking at video. Recent tests on FR, RR, and NR methods for HDTV by the Video Quality Experts Group (VQEG)[10,11] give hope and suggest that we are on the right path to developing objective metrics and methods than can serve as proxy for people; but VQEG states in its conclusions that "None of the evaluated models reached the accuracy of the normative subjective testing."

Often, perceptual-centric methods and products that use them will report results as "MOS" or differential MOS ("DMOS"). While understandable, it is worth noting (at the risk of being too pedantic) that this terminology can be confusing and lead to wrong conclusions, particularly by business colleagues who do not have a deep understanding of video quality assessment and the limitations of certain methods. We find habitual use of the terms "predicted MOS" and "predicted DMOS" when referring to the results of objective tests to be a safer more accurate approach.

Objective methods and metrics are very useful even if they are not yet able to take over from subjective testing entirely. Objective methods are great for monitoring systems and for regression testing. The key is to be cognizant of the interpretation of the objective data. Objective results can highlight differences and changes that might require action from an operator or equipment vendor, but objective results should not yet be used synonymously with quality terms such as "better" or "worse." No doubt the day for artificial subjective video quality intelligence will arrive, but that day is not today.

STREAMLINING THE SUBJECTIVE EVALUATION PROCESS

While formal ITU-R BT.500 and ITU-T P.910 testing can provide the results operators need to make mission-critical decisions, the general perception is that standardized subjective testing is too time-consuming and costly to be practically useful in all the situations where video quality assessment is required in today's operations. In truth, however, long experience in video quality assessment shows that a much less time-consuming, less costly iteration of standardized MOS testing can produce results comparable in reliability to those of the comprehensive approach described in ITU-R BT.500 and ITU-T P.910.

As illustrated in Figure 2, there are 4 main components of the streamlined methodology we have found useful: 1) pick an intuitive scoring metric; 2) create a benchmark viewing environment; 3) create a library of test sequences; 4) decide on one or more decision metrics.

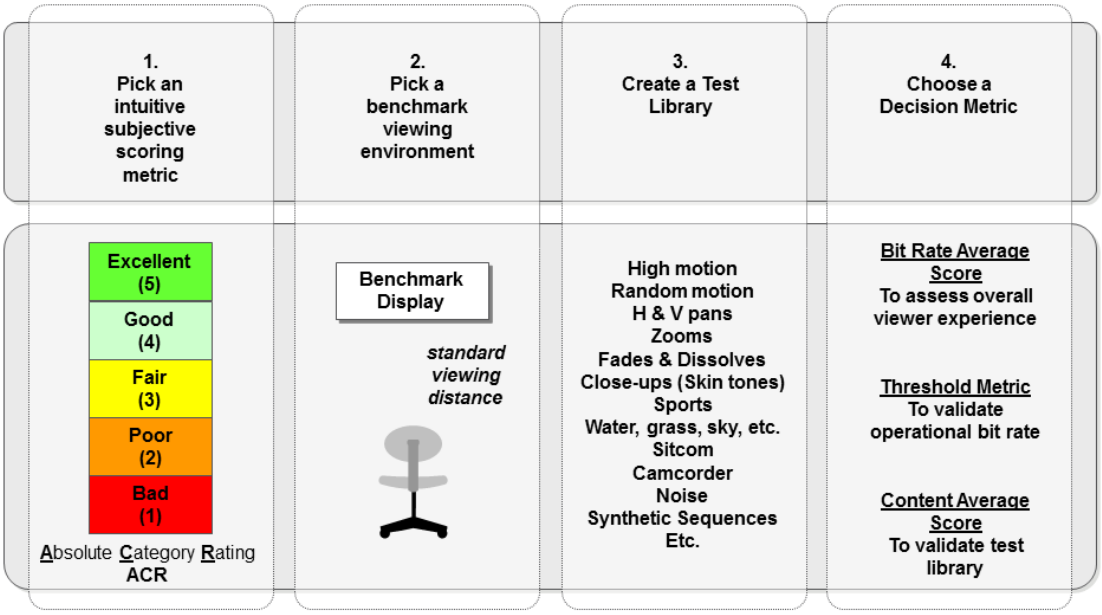


Figure 1. The 4 key components of our streamlined video quality scoring protocol

Simplifying Scoring

Our streamlined method is a simplified form of the Absolute Category Rating (ACR) system specified in ITU-T P.910. We have found that the ARC scoring metric is very easy to explain and most test subjects intuitively understand what is meant by "Bad" (1), "Poor" (2), "Fair" (3), "Good" (4), and "Excellent" (5). The key to choosing a scoring metric is that it not be overly prone to nuance and has stable meaning over time.

Simplifying Viewing Conditions

While the viewing environment must be consistent for all participants in a given testing operation, the conditions that must be factored into creating such an environment are easy to replicate in comparison to the broad set of highly specific conditions such as backlight luminance and minimal ambient noise levels that are described in the ITU recommendations. Test results will be reliable as long as all tests use the same display system with the viewer at a prescribed distance from the screen, room lighting or lack thereof is consistent, and there is no reflected screen glare to interfere with viewing. In our testing, we strive to create viewing conditions that would not be unfamiliar to people watching television at home or viewing content away from home on a mobile device. Although setting up ad hoc testing environments is possible, we have found that having dedicated testing environments, even small ones, helps us be more agile in responding to opportunities to improve our products.

Fewer Test Subjects

The streamlining extends to reducing the number of subjects participating in a test. While there's always an incremental gain in reliability as the number of subjects increases,

those gains become increasingly less significant as the number increases. As long as test viewers are properly vetted to be sure they have good eyesight and no significant perception impairments, good results can be attained from just a handful of viewers. In fact, results obtained from tests involving just a single "golden-eye" viewer on the operations staff whose assessments of video quality have been shown to be consistently reliable can be useful to an operator's decision process.

Building, Validating, and Simplifying a Test Library

Creating a re-usable test library is perhaps the most important aspect of systematic video quality assessment. At a minimum, a test library should represent the variety of content that one would expect to find on commercial television or distributed over one's video service; i.e. "typical" content. However, we have found that also including synthetic and very hard-to-compress "torture" content provides significant advantages.

It is tempting to simply use live programming as input video for video quality testing, particularly for testing transcoders and re-encoding. Using live video has the advantage that, over time, it is representative of one's video distribution service. And it is often as simple to implement as tuning to a live channel. The disadvantage is that it is not repeatable, which is a critical issue for scientifically valid testing. Moreover, live programming rarely provides the "torture" moments for video processing that help define the full range of performance for equipment and services.

We have found that a better approach is to record clips of live programming as either MPEG-2 transport streams or decompressed video for possible inclusion in a master video quality library. Such clips may then be stored and played back in a loop or in concatenation with other test content as many times, and in as many separate tests, as needed.

As mentioned earlier, the test sequences should be chosen to reflect rigorous stress-level representations of all the characteristics that go into assessing video quality. This ensures that the tests will maintain a balance between "easy" and "hard" characteristics, thereby avoiding a skewing of scores too far to the high or low ends of the MOS spectrum. As discussed below, compression-busting "torture" test video can also help to show an operator or equipment vendor what particular technological improvements might be undertaken to improve video quality and/or reduce operational bitrates.

When is a test library complete enough that it can be used to generate actionable data through video quality testing? Validating a test library is very important. It is also important that the test library not become too burdensome. It is often counterproductive to have too many test videos, particularly for subjective testing. It is better to have a smaller set of test videos that strategically spans the range from very simple, through typical, to very hard-to-compress. (Objective methods are a great companion to subjective testing in this regard. They can process larger test libraries to probe for corner cases to be identified for further investigation.)

Extending the Usefulness of Single Stimulus Testing

Most important of all when it comes to simplifying video quality assessment procedures, operators can use the single-stimulus test results to achieve all the goals normally associated with both single- and double-stimulus testing. In other words, they can dispense with the side-by-side and sequential comparative testing between reference and test videos commonly used in encoding vendor "bake-offs" and for

determining video quality and bitrate thresholds.

Our Streamlined Testing Protocol

In our streamlined protocol, the subjective testing is carried out by asking a viewer to watch short test clips from test libraries we have developed over time in our interactions with colleagues and customers in cable, telco IPTV, direct-to-home satellite, industry groups, and standards bodies. Test sequences were 10 to 20 seconds each and were displayed up to 3 times in a continuous loop.

After the presentation of each test video, the viewer scored the video quality of the test clip. Each test video represented a different combination of source content and compressed bitrate. Note that our streamlined protocol may be used with test videos that were compressed in either constant bitrate (CBR) mode or variable bitrate (VBR) mode.

Care should be taken in the preparation of test videos. In most of our testing, we play the source video on a short loop and perform continuous real time encoding so as to avoid artifacts related to start-up effects of the rate control algorithm. An alternative approach would be to create a short uncompressed clip from a longer captured compressed bit stream. The uncompressed clip could then be played on a short continuous loop as the test stimulus.

Figure 2 shows data from a subjective testing session from a number of years ago. Although the scores are now obsolete (encoders keep getting better), they provide a useful way to illustrate the data analysis component of our streamline protocol. The data are also somewhat unusual because they are from a "golden eye" session in which a single expert viewer recorded individual scores. This particular subjective testing session was undertaken to provide information with respect to choosing a particular technology development path that required a quick response on our part. We provide this use case here to highlight that we believe our method has value even in such a case, though more viewers will always provide more reliable results.

Results (after rank sort on content complexity)

	CBR Video Bit rate (Mbps)											
Sequence Name	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5
Test Sequence 1	1	2	3	4	4	4	5	5	5	5	5	5
Test Sequence 2	1	2	2	3	3	4	4	5	5	5	5	5
Test Sequence 3	1	1	2	3	3	4	5	5	5	5	5	5
Test Sequence 4	1	2	2	3	3	3	4	4	5	5	5	5
Test Sequence 5	1	1	1	2	2	3	5	5	5	5	5	5
Test Sequence 6	1	1	2	3	3	3	4	4	4	5	5	5
Test Sequence 7	1	1	2	2	2	4	4	4	4	5	5	5
Test Sequence 8	1	1	2	2	3	3	3	4	4	4	5	5
Test Sequence 9	1	1	1	3	3	3	3	3	4	4	5	5
Test Sequence 10	1	1	1	2	3	3	3	3	4	4	4	5
Test Sequence 11	1	1	1	2	2	3	3	3	4	4	5	5
Test Sequence 12	1	1	1	2	2	3	3	3	4	4	4	5
Test Sequence 13	1	1	1	1	1	1	1	1	2	2	2	2
Total Sequences	13	13	13	13	13	13	13	13	13	13	13	13

Note – These data are old and used for illustration only. They do not represent state-of-the-art.

Figure 2. An example of subjective scores for a particular session. In this session, 13 test sequences were each encoded at 12 CBR bitrates from 0.05 Mbps to 11.5 Mbps. Note that the test sequences/bitrate combinations were presented in random order during the test. The data shown here were sorted during data analysis according to average content complexity (see Figure 4.)

Deriving Multiple Answers from a Single Scoring Matrix

As shown in Figure 3, these results can be analyzed in three different ways to provide answers to key questions associated with video quality assessment. One parameter produced in the analytics process is the degree of complexity of each test clip as reflected in the average score at each bitrate. In this example, the least complex clips produce an average score of 5 ("excellent") at a minimum bitrate of 6.5 Mbps while some of the most complex clips only reach 5 at 11.5 Mbps and one never goes above 2 even at the top bitrate. This analysis suggests that the set of test sequences used in this session represent a range of complexity even in this streamlines testing process.

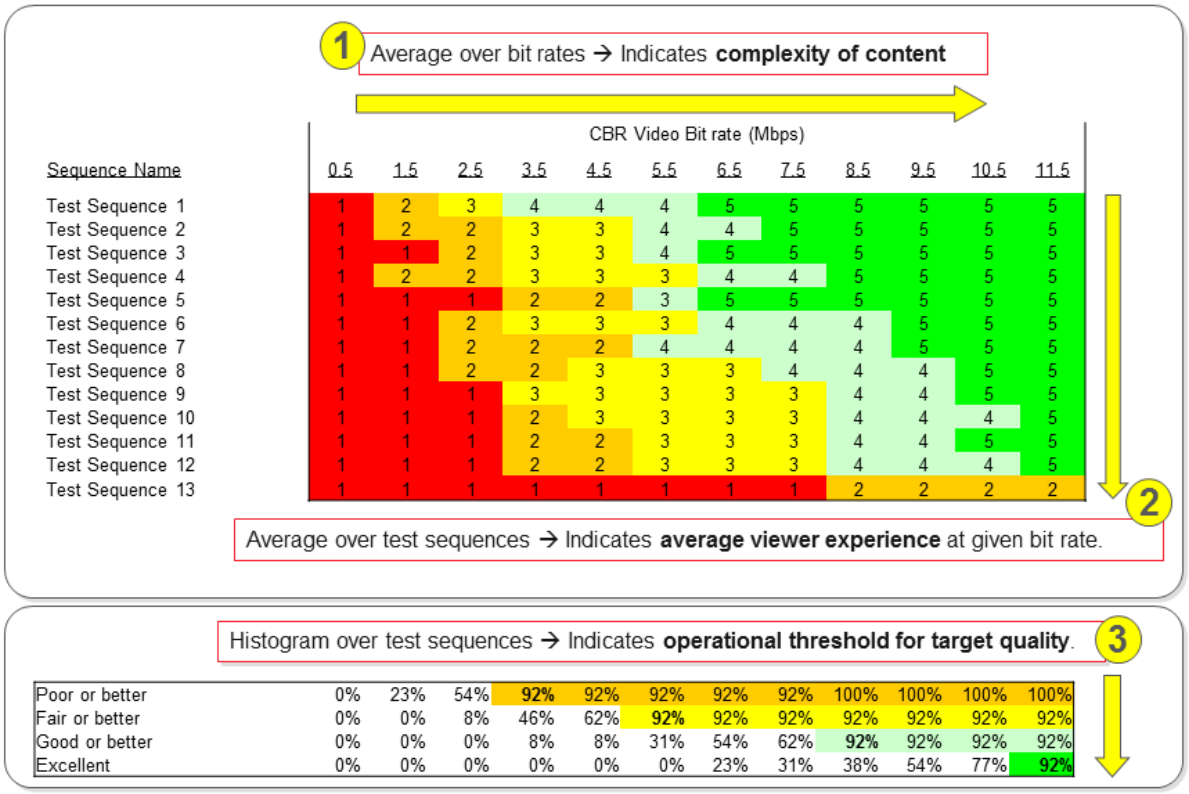


Figure 3. An example of analyzing the subjective scores to help answer 3 important technical questions that affect business: 1) "Is the test library sufficiently varied to allow for solid conclusions from the recorded data?" 2) "What may we expect the average viewer experience to be?" 3) "What bandwidth should I plan for in order to achieve a particular level of video quality?"

A more detailed view of the validity of the set of test sequences used may be obtained by calculating the average score across all bitrates for each test clip, as plotted in Figure 4. It is clear that 12 of the 13 test clips cover the just-below-fair to just-above-good range, but only 1 test sequence probes below poor and none above good. The value of plotting the averages in this manner is that it shows us, in hindsight, our test of several years ago would have benefited from a more diverse set of test sequences. If these were results for a test we were performing today, it would be prudent for us to add more test sequences and record more subjective scores. Over time, analyzing the master test library and updating the library based on the results would help to optimize the library so as to support better decisions about video quality and bitrate plans.

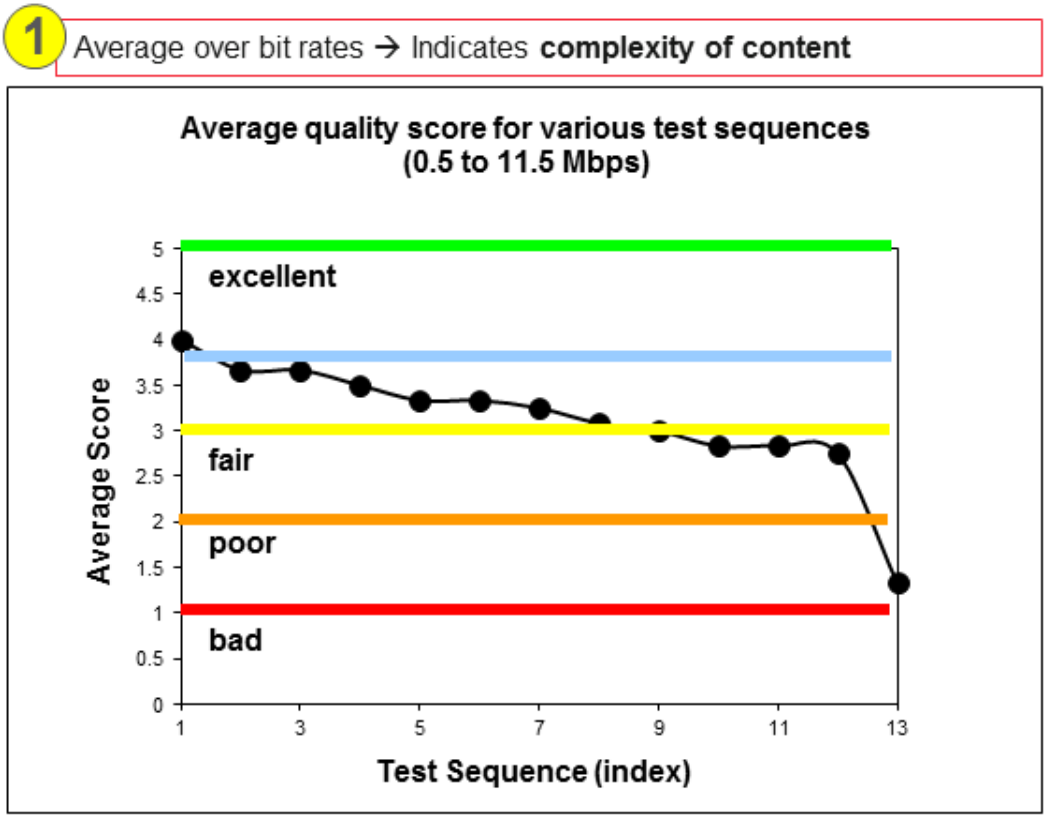


Figure 4. An example of analyzing the subjective scores to determine if the test library is sufficiently varied to allow for solid conclusions from the recorded data

By providing insight into the average viewing experience across all sequences at any given bitrate the methodology allows operators to calculate the overall average score at each bitrate, as reflected in Figure 5. This chart shows that the encoder used in this test produces a good quality of viewing experience, on average, starting at a bitrate of approximately 5.5 Mbps and an excellent video quality starting at around 8.5 Mbps.

These measures can be used by the operator to perform comparative analysis on different encoding systems, as shown in Figure 6. Here the tested performance of the improved encoder shown in Figure 5 is charted with the performance of a previously existing encoder to show key points of disparity between the two, such as the higher bitrates at which the previously existing encoder registers "good" and "excellent" scores.

The third type of analysis provides operators the metrics they need to determine where they want to set the operational bitrate for reaching a targeted level of quality. As shown in Figure 7, the resulting histogram shows the tested encoder achieves fair or better quality on 90 percent of the test sequences at a bitrate of 5.5 Mbps, good or better quality at the 90 percent threshold at a bitrate of 8.5 Mbps and excellent or better at 11.5 Mbps.

Using the same type of histogram results to compare the performance of the two encoders, the operator can assess the degree to which

the bitrate threshold for achieving a given quality target can be lowered with use of the improved encoder.

As shown in Figure 8, the threshold can be lowered by a margin of about 1 Mbps for achieving good-or-better performance on 90 percent of the sequences with the improved encoder.

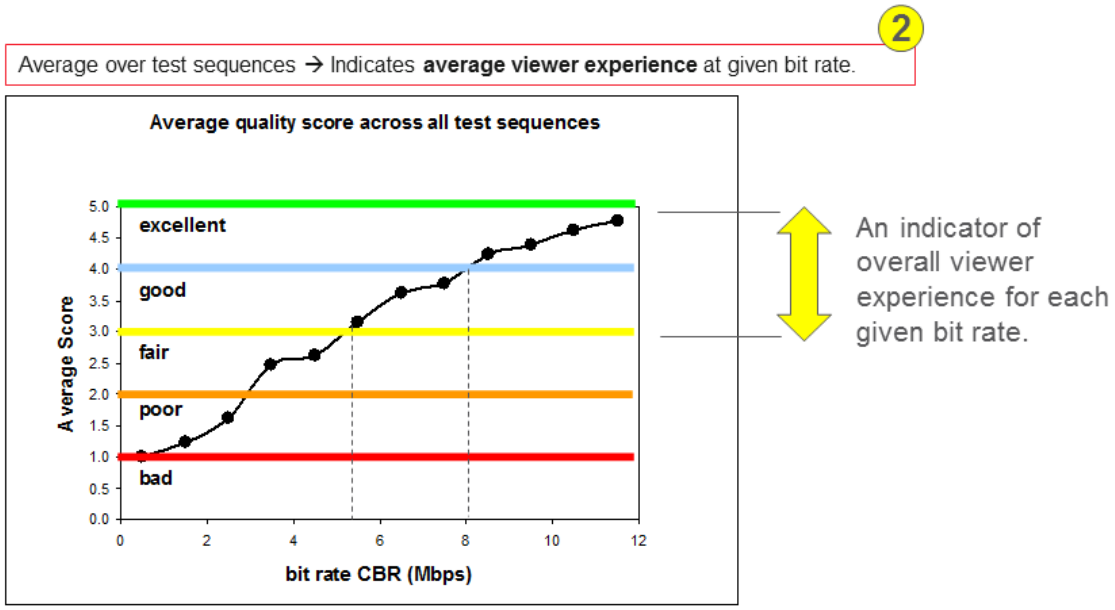


Figure 5. An example of analyzing the subjective scores to provide information that would be predictive of subscriber's overall video quality of experience

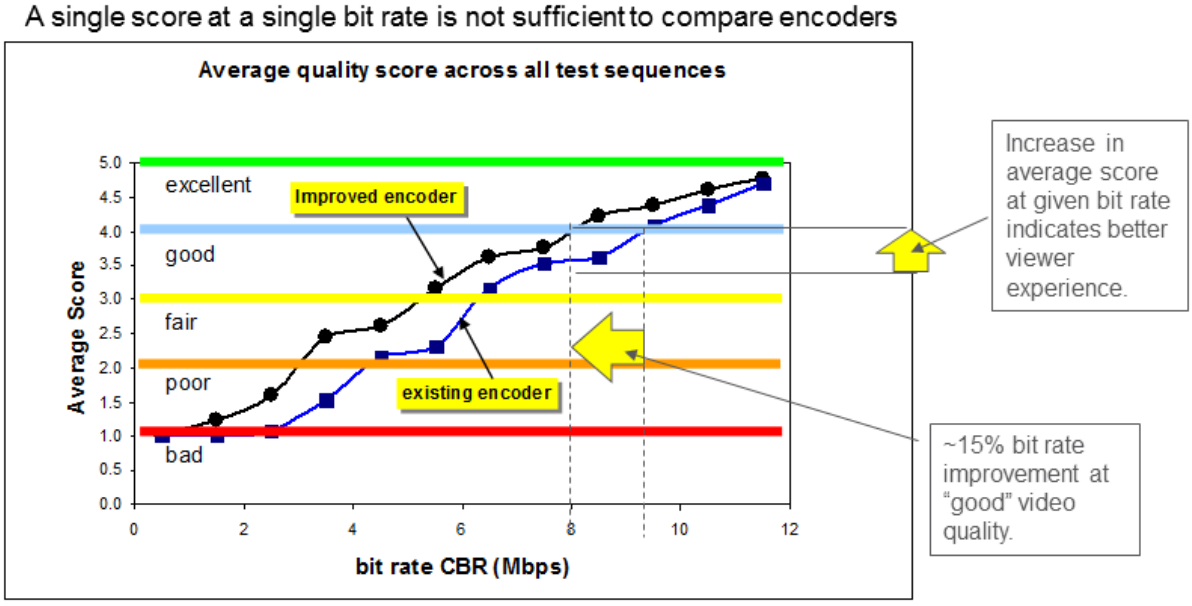


Figure 6. An example quantifying how much one compression option might improve the overall view quality of video experience compared to another compression option

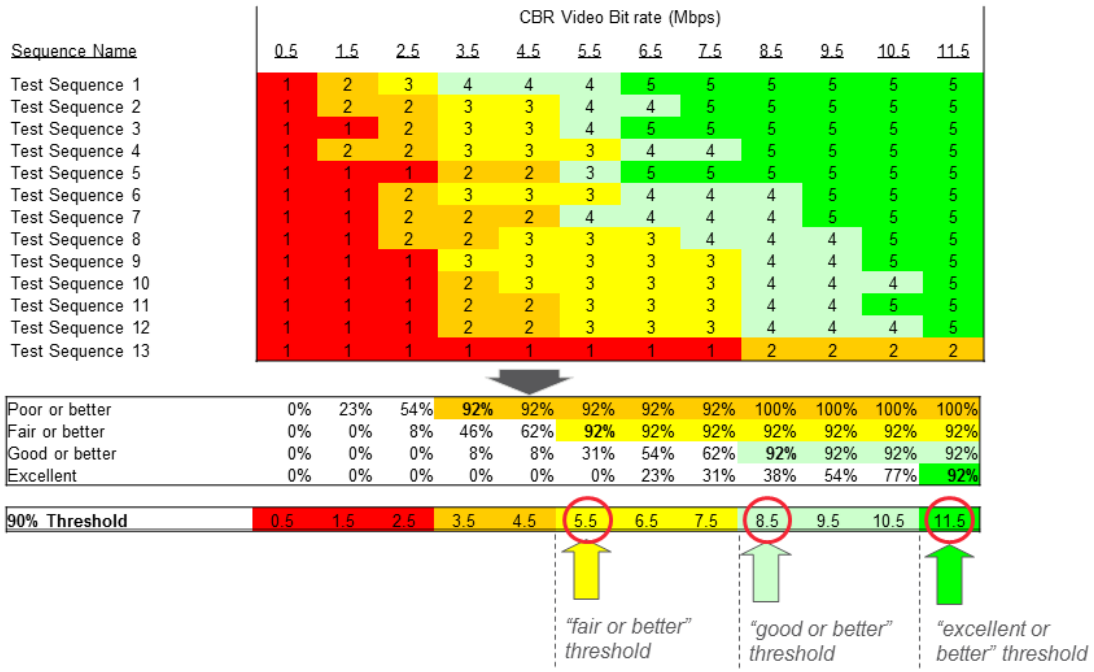


Figure 7. An example of analyzing the subjective scores to help determine how much bandwidth should be planned in order to achieve a particular level of video quality for a certain percentage of programming

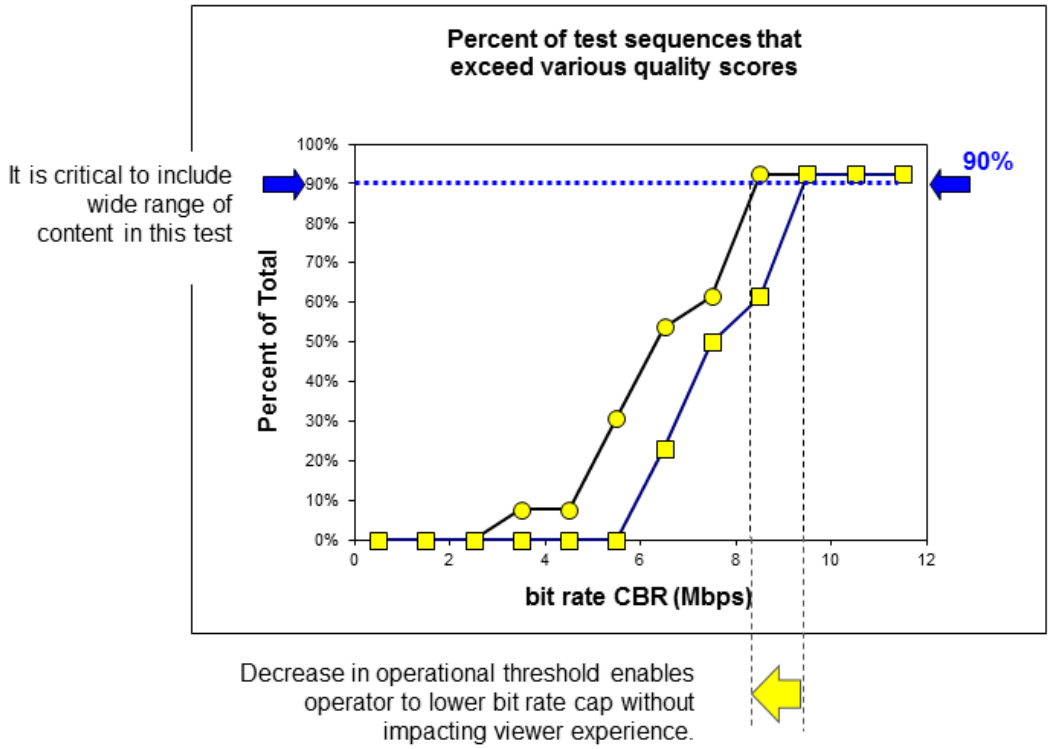


Figure 8. An example quantifying how much one compression option might lower the operational bitrate requirements compared to another compression option

Facilitating Improvements in Performance

The scoring matrix also provides a ready means of identifying the areas of weakest quality performance for a given encoding system, as shown in Figure 9. Whether the matrix is used by encoding designers or operators, they need to be able to quickly see where efforts to improve performance will do the most good.

By focusing on the weakest spots as shown in the test clip scores, developers and operators will be able to obtain the greatest improvement in overall performance ratings. Ongoing efforts to improve performance can be focused on the next most troublesome areas once the performance on the initially targeted set of "troublemakers" is improved as much as technical means allow.

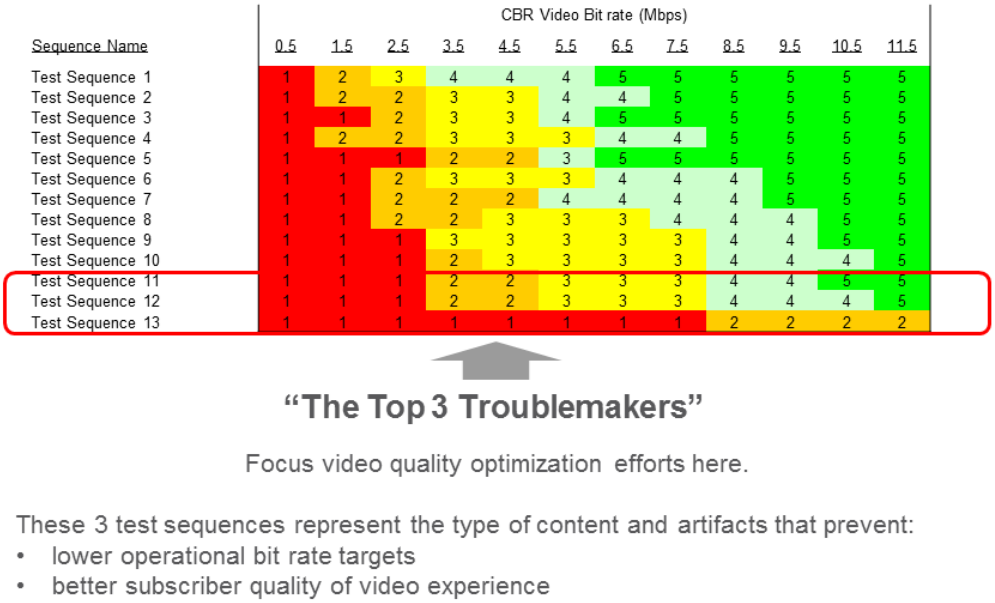


Figure 9. An example of using the subjective scores to help decide where to allocate resources toward improving and optimizing video processing equipment and services

These examples demonstrate how a simplified, systematic approach to single-stimulus MOS testing can enable a data-driven discussion about video quality across all service categories. Such quantifiable results are essential to guiding the selection of quality targets, setting minimum bitrates, choosing encoders, and determining whether to make investments in new encoding solutions.

CONCLUSION

As the need for video quality assessment encompasses an ever-expanding range of cable and other MVPD operations, from legacy SD and HD to TV Everywhere and IPTV to HEVC and 4k UHD, operators cannot afford to base critical decisions on unreliable testing processes. PSNR is not designed to be, nor should it be, used as substitutes for measuring human perception of video quality. Other objective methods leverage knowledge of human vision to a greater or lesser extent, but none of them should yet be relied on as a replacement for people.

Efforts to base key decisions affecting quality thresholds, bitrate settings, and

technology choices on real human viewing experience should not be impeded by unnecessary complications in the testing process. There's not enough time or money for executing all the procedures typically associated with using standardized MOS-based measures of human perception as the foundation for decision making.

Fortunately, there is a much simpler, scientifically valid approach to generating and applying the results of MOS testing in video quality assessment. By reducing the number of participants in tests, eliminating overly precise conditions for creating viable test environments and focusing on using the results of single-stimulus testing to answer questions traditionally addressed through more complicated dual-stimulus procedures, operators can perform reliable video quality assessment essential to answering all key questions much faster and at far lower costs.

REFERENCES

1) Z. Wang, A. C. Bovik. "Mean squared error: Love it or leave it? A new look at signal fidelity measures." IEEE Signal Processing Magazine, Vol. 26, No. 1. 2009.
2) ITU. "Recommendation ITU-R BT.500-13. Methodology for the subjective assessment of the quality of television pictures." 2012
3) ITU. "Recommendation ITU-T P.910. Subjective video quality assessment methods for multimedia applications." 2008
4) Z. Wang, A.C. Bovik,, H.R. Sheikh, and E.P. Simoncelli. "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing 13: 600–612. 2004
5) Mylène C. Q. Farias (2010). Video Quality Metrics, Digital Video, Floriano De Rango (Ed.), ISBN: 978-953-7619-70-1, InTech, DOI: 10.5772/8038. Available from: http://www.intechopen.com/books/digital-video/video-quality-metrics
6) Wang, Y. (2006). "Survey of objective video quality measurements," Technical Report T1A1.5/96- 110, Worcester Polytechnic Institute.
7) S. Winkler. Digital Video Quality: Vision Models and Metrics. John Wiley and Sons. 2005.
8) S. Winkler "Video Quality Measurement Standards – Current Status and Trends" IEEE ICICS 2009
9) S. Winkler and P. Mohandas. "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics" IEEE Trans. Broadcasting Vol 54. No. 3. 2008
10) VQEG. "Final report from the video quality experts group on the validation of objective video quality assessment." March, 2000
11) VQEG. "Report on the Validation of Video Quality Models for High Definition Video Content." June, 2010
12) Winkler, S. "Analysis of Public Image and Video Databases for Quality Assessment? IEEE J. Select Topics Sig. Proc. Vol. 6, No. 6 2012

AN MSO APPROACH TO SECURING THE 'INTERNET OF THINGS'

Jim Poder
IOT Strategy and Innovation

Abstract

As the Cable Industry prepares to support the onslaught of connected devices that make up the 'Internet of Things' we must consider our role in managing and securing the devices and services that appear on our networks and in our customer's homes. Traditionally the Cable Industry has provided these functions in other service areas through heavy specification, certification, and testing regimes, but in this new-world of devices that can not be the case. This paper describes how devices acquired through a variety of channels can successfully be managed and networks and services secured even without the traditional Cable model of specification, certification, and test. Through Device and service "fingerprinting" and taking an analytical approach to traffic monitoring, security can be delivered to the 'Internet of Things' world. This paper also explores the rapidly evolving world of 'Internet of Things' standards and discusses their relative merits where it comes to enabling the MSO to provide management and security while enabling service.

SCOPE OF THE IOT

The Internet of Things (IoT) is rapidly evolving! This is not much of a revelation to anyone who is involved in the tech industry and is not even to most technology users. Sensors and devices have become pervasive in our daily lives. From wearables like Fitbit or the Apple Watch to home automation and security systems like Xfinity Home and Lowes Iris to vertically integrated point solutions like the Nest Thermostat, many people are now living with something that could be described as the Internet of Things. Companies and market researchers will debate the size of the IoT, but most will admit that it is many billions of devices. Despite efforts toward creating and adopting standards like the AllSeen Alliance, Open Interconnect Consortium, Thread Alliance, and others, the "unified theory" of the IoT currently eludes us. There are many reasons why this is the case, but ecosystem control is one of the key drivers. Many device and sensor developers want a clear and easy path to consumers that doesn't leave them beholden to a service or platform provider. For them it can be an easier go-to-market strategy to build their own vertical integration where they build not only the device/sensor, but the communication protocol, data model, server, and web and mobile interfaces as well. The issue is that the power of the IoT is in correlating data from many devices and sensors to provide value to the user. Single data points don't allow for the rich set of services that have been envisioned. Integrated platforms and standards do not provide the panacea for devices either. Today the standards are incomplete, or under-supported, and often there are multiple standards that must be certified to in order to get devices deployed on a platform with a provider. Not only can this be very time consuming, but also very expensive for a smaller company.

Security is one of the largest and most complex issues when it comes to IoT. Systems typically have so many interfaces that it can be difficult to properly test them all to assure that there are no vulnerabilities. Additionally, the software and service layers evolve rapidly making it difficult to assure that new vulnerabilities are not introduced. We have seen many examples in the news over the past couple of years where systems

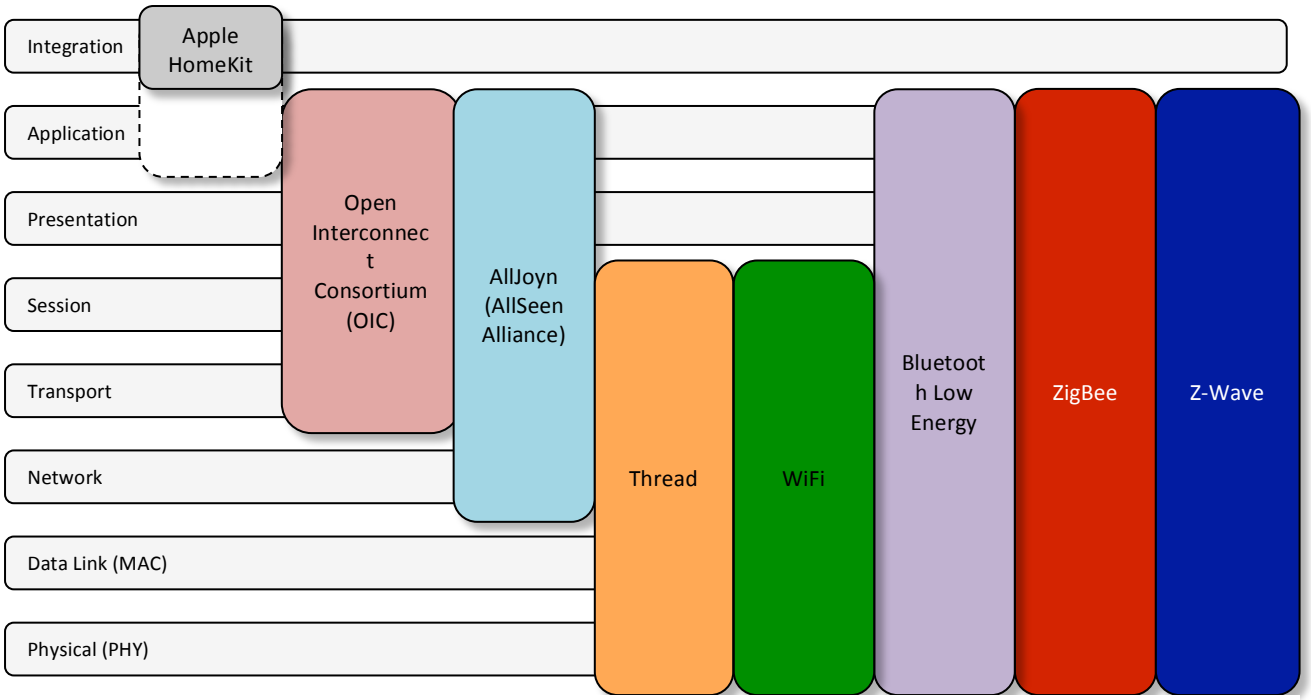


Figure 1: IoT Standards

have been exploited due to vulnerabilities of interfaces. Attacks can come from somewhat unrelated systems that happen to share common server or network elements. Several years ago there was an example of a large retailer in the US that had a significant theft of credit card data. It was discovered that the attack was perpetrated by infiltrating the climate control/HVAC systems for the stores and once the attacker had access to the internal network the Point-of-Sale (POS) system was open. While the POS system had strong security elements associated with it, the HVAC system did not. The moral to the story is that if security is not thought of and addressed holistically, the IoT cannot be made secure.

DEVICES OF THE IOT

As we look at the end devices of the IoT the challenges in securing them are numerous: They will come from a variety of suppliers, utilize a variety of standards and technologies. These devices may not be able to be truly authenticated, might not be trustworthy, and may not be patched or updated on a regular basis.

Role for Standards

There are many Standards being proposed to make up the IoT. Figure 1 shows each of their roles across the standard OSI stack. While these standards take good measure to ensure device authentication and secure communications, they fail to address the issues if the devices themselves have been comprimized.

The concern is that devices already on the home network, but have been comprimized can cause a serious issues such as sending data to non-authorized services or sending commands to other IoT devices.

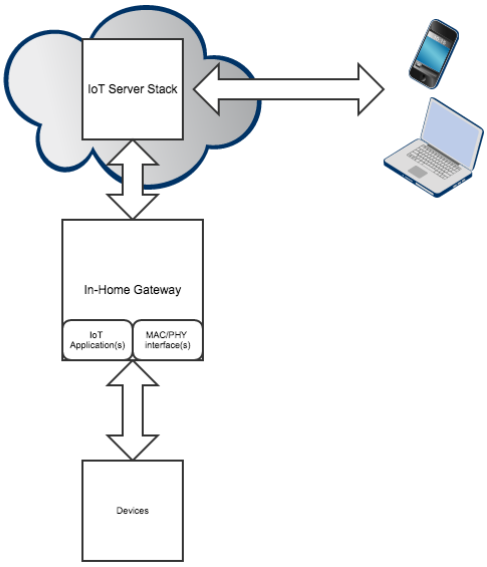


Figure 2 IoT Architecture

Figure 2 shows a high-level IoT architecture with devices communicating with servers/services by traversing the gateway and applications running therein. For devices utilizing TCP/IP protocols and a thin-gateway where the application is expected to reside in the IoT server Stack the device is left open to communicate with other servers. This communication may be part of normal, expected operations such as the device checking in with a manufacturer server to see if new firmware is available to be downloaded, but can also be used maliciously for a device to send data to other locations. Figure 3 describes the scenario where a rogue device (that is authticated to the network and is certified) is listening to messages on the device bus it can then forward these messages to another IoT server. Under most existing IoT standards this would be perfectly acceptable and compliant, but could result in a breach of personal information.

Rogue devices may also send commands to other devices on the network. A compromised device may send a command to a thermostat for instance to change the setpoint to 99 degrees! If the gateway is not blocking these sorts of malicious attacks it could cause real problems.

This really points out the need for device fingerprinting.

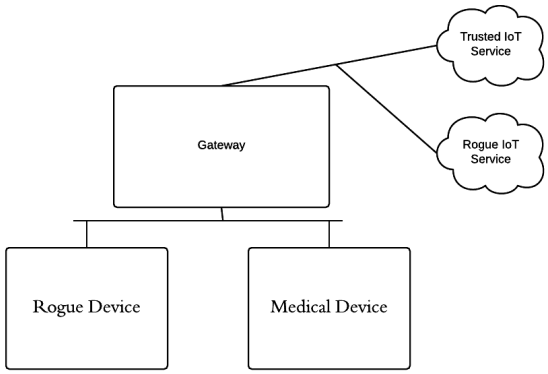


Figure 3: Rogue Device

DEVICE FINGERPRINTING

Device Fingerprinting can be thought of as network-wide pattern matching. When devices are communicating as expected the network runs fine, but should a device begin to operate outside of its expected parameters the network blocks the transactions and alerts are sent to the system and its users.

So how do we determine "expected parameters"?

White List, Black List and Grey List

When a device is trying to join the network and can be authenticated it will declare its capabilities as part of the authentication process (i.e. Thermostat, lighting controller, audio playback device, etc.) along with a manufacturer ID, device ID, hardware revision, and software revision. The gateway must then be responsible for doing a service level authentication for that device. Through a database stored on the gateway or service in the cloud the device will be determined to belong to a white-list of known and trusted devices, a black list of known untrusted devices, or a grey list of unknown or somewhat trusted devices. White list devices are those that have undergone rigorous testing and are certified throughout to be good

players. Black List devices are ones that have been proven to be bad players either through non-compliance with standards or expected operations or through rogue behaviors. These devices should not be allowed to participate in IoT services and the end user as well as the service provider should be warned of there presence so that they can be removed or mitigated. It is expected that the grey list of devices will prove to be the largest of the IoT. Devices that have not been seen before or that have not gone through the stringent certification process of white list devices will remain in this category. They do not have full permissions to access servers and services and any communications sent or received by these devices will be subject to fingerprinting to determine compliance with the security model.

Fingerprinting is the role of the gateway. During authentication and service level authentication the gateway will create a profile for each device containing device manufacturer, model, HW/SW revision. This will also contain expected behaviors of the device. These include:

- Heartbeat frequency, size of packet, and destination
- Normal communications frequency, size of packet, and destination(s)
- Firmware upgrade size and destination

The gateway keeps running counters for each of these attributes for each device on its network and will assure that

If a profile does not exist, the system will start with either a default model for that device type, or will begin the fingerprinting process. All of these are meant to establish what normal communication behavior means for a particular device. This may mean how often the device transmits a message, the destination for that message, or even the content of that message. For example, a connected thermostat may relay its status every 30 seconds, and send a heartbeat every 10 seconds and all of its communications have the climate control app within the gateway as its destination. If at some point the thermostat starts sending messages bound for a server on the internet or for a lighting control device in the home that communication is blocked and the system is notified of a potential issue. Given that there will be many manufacturer of similar devices, the fingerprinting algorithm will compare profiles of similar function devices against one another and should a device be sufficiently out different from its counterparts it will be flagged for possible blacklisting. One challenge to this model is that all intra-device communications inside the home must traverse the gateway where messages will be fingerprinted and evaluated as to their adherence to that device's allowed model before being delivered. This does not mesh with some IoT standards, notably Allseen which relies on peer-to-peer communications over the D-Bus architecture.[1] that rely on device-to-device communication using a publish-subscribe model on the home network.

Source of Rogue Devices

With the huge potential for suppliers for IoT devices it must be assumed that not all will be trustworthy, and some devices may come with software onboard that has alterior motives. This was seen in 2014 on some smart phones sold at retail in the U.S. that contained the malware 'DeathRing' preinstalled and note able to be uninstalled by the user.[2] This malware would transmit personal information from the smart phone to an unknown destination. This type of behavior must be expected, accounted for, and dealt with in the IoT. By fingerprinting devices and comparing device types this type of misbehavior can be rooted out.

Additional sources of Rogue devices involve long-lived devices that may not get maintained.

The Legacy of the IoT

Many of the smart home devices are ones that become part of the infrastructure of the home; Thermostats, outlets, lighting, appliances, motion detectors, door/window sensors, smoke alarms, etc. These devices are long-lived in homes and most consumers will never replace these devices once they are installed. This legacy-building problem is well known to MSOs. Take for instance the claim on an iconic lighting device of a 15 year life. Supporting it and insuring security robustness for that lifespan is going to be daunting. Feature upgrades aside, these devices are unlikely to have any long-term sustaining engineering support to patch bugs and security flaws that are found later in the device's life.

This presents a significant challenge to the smart home. Who is responsible for the security of the overall network and how will they accomplish this when all of the devices are not under their control?

Should a device manufacturer cease to exist or has declared end-of-life for a product there is the chance of a flaw being discovered later in that device's life that can be exploited maliciously. While these devices may be automatically added to the Black List and users notified, that is not a consumer friendly approach. Fingerprinting may help identify devices that have been exploited, but only after an attack has occurred and been identified. Fortunately, these attacks are limited to individual homes and are unlikely to be perpetrated on a wide-spread scale.

Another approach that is being addressed by Arm with their MBED OS is to require devices to run an underlying Operating System (OS) that can be upgraded by the system without harming the application software. [3]

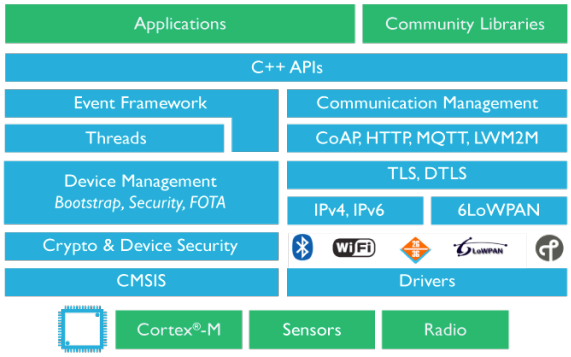


Figure 4: Arm MBED OS software Stack

Using this model, the security features and communications protocols can be upgraded without the need for the manufacturer's intervention. This can be difficult for all devices as many of the single-function, battery powered devices today do not run any sort of OS. As the need for manageability as well as over-the-air software upgrades becomes standard, it is believed that this will change.

CONCLUSIONS

Although it is too early to say which standards and approaches will prevail, it is clear that standardization and open security models are going to be an important piece of the IoT. Security of device and sensor networks must be thought of on a system wide level through device and communication fingerprinting along with dynamic white, black, and grey lists of accepted devices. These must not become an impediment to innovation and adoption, however. Service providers will have to be diligent in selecting architectures and devices, but must realize that they are building a legacy with consumers that will have to be supported for a very long time. We can then begin to realize the promise of services and analytics that can yield tremendous new businesses and services.

[1] Allseen Alliance Security https://allseenalliance.org/developers/learn/core/system-description/alljoyn-security

[2] Security Watch, PC Magazine December 8, 2014 http://securitywatch.pcmag.com/security-software/330164-mobile-threat-monday-deathring-malware-pre-loaded-on-android-smartphones

[3] Arm IOT device Platform http://www.arm.com/products/internet-of-things-solutions/mbed-IoT-device-platform.php

Approaches for IPv4 as a Service

Brian Field Ph.D.
Comcast

Abstract

IPv6 has been gaining traction both as a necessary service provider technology and as an increasing carrier of Internet traffic. Within Comcast, dual-stack has been an important step to help IPv6 deployment mature and grow. As IPv6 traffic grows and as IPv4 traffic diminishes, we would like to avoid the need to make software changes to enable network features, while also avoiding hardware operational costs that accompany the treatment of IPv4 as native transport in our infrastructure. As such, we are looking at ways for our next generation infrastructure to incrementally evolve towards an IPv6 focused infrastructure. We realize we must continue to support IPv4 for the foreseeable future. Our our thinking is to support IPv4 as a Service (IPv4aaS) — meaning we will carry IPv4 traffic over a "lean" IPv6 focused network infrastructure. In this paper, we detail the insights and learnings Comcast acquired in evaluating technologies that enable or support IPv4aaS over a lean core infrastructure. We also describe how these technologies synergize with current Software Defined Network (SDN) methodologies.

INTRODUCTION

The unmistakable rise of IPv6, both as a necessary service provider technology and as an increasing carrier of Internet traffic, prompted Comcast to take a leadership position in IPv6 deployment. Comcast operates a 100% dual-stack network, with a significant portion of our cable modem infrastructure strictly using IPv6. As part of this effort, Comcast has developed and shared with the Internet community enhancements to OpenStack that support IPv6 for virtual machine (VM) and related infrastructure.

Within Comcast, dual-stack represents an important step to the deployment, maturation and growth of IPv6. None the less, we are looking at ways for our next generation infrastructure to natively support *only* IPv6.

We believe this direction will result in a number of benefits. One such infrastructure benefit would be a reduced Forward Information Base (FIB) size, because the network will only need to carry IPv6 routes. This translates into much less high-speed memory required on routing line cards, which reduces the cost to build them, and the energy they consume. Less memory on router line cards also translates into more compact form factors, which makes for better port densities. Further, less memory means less heat generation, which lowers overall environmental costs.

If we are able to totally eliminate IPv4 as a native citizen in the network, we can also eliminate IPv4-specific features in the router configs— which, in turn, reduces both our configuration and operational overhead. Further, we could operate router code revisions that only contain IPv6 capabilities— which yields more efficient code, fewer bugs, and fewer unintended operational interactions with our equipment vendors. It follows that a simpler router environment lowers the bar for new vendors to enter the eco-system and makes it more plausible to seamlessly mix and match differing vendor router platforms.

While we certainly see value in making the infrastructure IPv6 only, we can't abandon IPv4 — even as its traffic share decreases. By supporting IPv4 as a Service (IPv4aaS), we can carry IPv4 traffic over our "lean" network infrastructure.

Lean, IPv6 Focused, & Lean IPv6 Networks

The technologies we present in the subsequent sections represent different capabilities and solve slightly different problems. We define a "lean" network as one with a relatively small FIB table, with a goal of keeping the FIB size small, to reduce capital and facility expenses. A lean network can be dual-stack.

An IPv6-only network is one in which the network infrastructure only supports the IPv6 protocol natively. The goal of this network architecture is to drive towards a IPv6-only feature set; a small FIB is not a specific optimization.

A "lean IPv6 core" is one where the underlaying network has the characteristics of reduced FIB size, and is a single stack network, namely based on IPv6.

Before considering the technology options that enable a lean or IPv6-focused network, we need to consider how much IPv4 traffic currently exists on our network. For instance, if we attempt to move to an IPv6-only network today, how much IPv4 traffic might need to be carried, via an overlay technology?

To determine the traffic load, we used Netflow information processed through Deepfield[1] to show how much traffic was traversing the Comcast Backbone. The amount of traffic observed over a five-day period is show in Figure 1. In the following sections, we have opted to use qualitative terms to characterize the load on our network rather than specific values.

Figure 1 shows that the Comcast backbone carries, at peak, multiple Tb/s of traffic. Even during non-peak hours, the backbone carries traffic a significant IPv4 traffic load. This is a sizeable amount of traffic to carry via an alternative technology were we to attempt to migrate to an IPv6 only network.

Thus, in the near term, it might not be cost effective to transition all of this IPv4 traffic to an overlay technology in order to obtain an IPv6 only infrastructure. However, there is value in understanding the specific traffic load, by IPv4 prefix, as a means to determine if we can incrementally evolve to a lean network core.

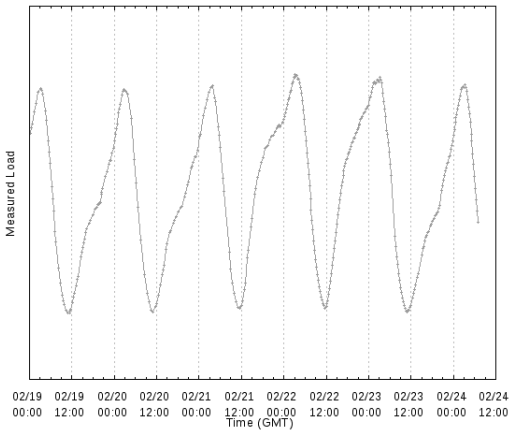


Figure 1: Measured Load

IPv4 FIB and Per Prefix Usage

Our IPv4 Internet routing table contains a mix of Comcast local IPv4 routes (for internal connectivity) and Internet routes (routes received from peers that provide the Internet-wide connectivity). As of Q1 2015, the size of the IPv4 routing table is 575,000 routes. Each route has a "prefix" and a "mask" and consumes a FIB entry.

Operators of ISPs around the world tend to differ in business and operational directions, and those directions tend to influence how an ISPs announces its IPv4 address space to the Internet. ISPs (us amongst them) often have limited means to manage these received announcements, while still maintaining full Internet connectivity.

While full Internet connectivity is a requirement for the backbone, it wasn't clear how much traffic was being delivered to each IPv4 route in the FIB.

Using Deepfield, we performed an analysis on the IPv4 traffic observed on the backbone. Specifically, we configured Deepfield to monitor and report on the amount of traffic observed per IPv4 prefix contained in the routing table. We configured Deepfield to report data on five-minute intervals, meaning the information returned as average over a five-minute window.

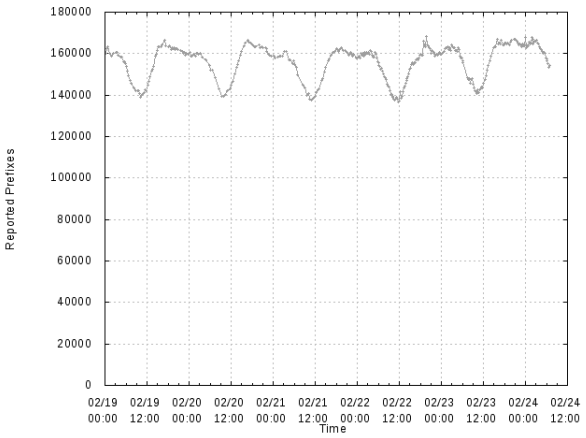


Figure 2: Observed Prefix count per Sample

Figure 2 shows the number of prefixes Deepfield reported in each five-minute window, observed during a five-day stretch in February 2015. As it turns out, the number of prefixes varies between 140k (in the very early morning hours in the U.S.) to just over 160k prefixes during the day and into the early peak hours.

Recall that the full Internet routing table is about 575k routes, and at peak, Deepfield sees traffic of only 160k prefixes. This means that Deepfield *doesn't* see traffic for 415k prefixes (or more) in each 5-minute sample.

Part of the reason we might not see traffic to these prefixes is related to our Netflow sampling configuration (1 packet per 8000). However, not seeing traffic to many prefixes (415k, in our case) likely indicates that there is very little traffic to those prefixes in the first place.

The first observation from this data: Nearly three quarters (72%) of the IPv4 prefixes in our FIB have no measureable traffic.

Details on the Prefixes Observed

We continued our analyses by looking at the load reported for the prefixes within one five-minute snapshot. We picked a snapshot near peak load -- specifically, the 3:15 GMT snapshot on 2/20/2015. Figure 3 displays analysis from this five-minute snapshot.

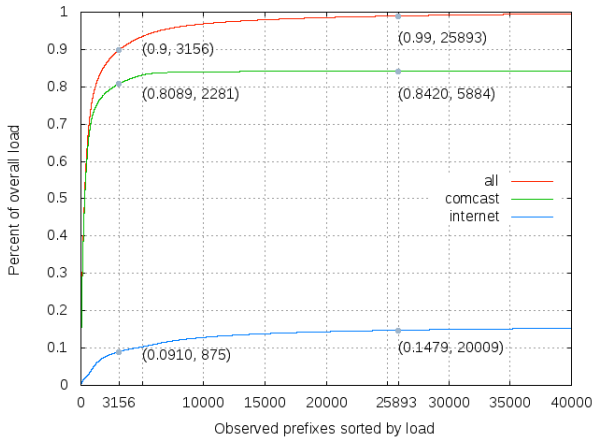


Figure 3: Cumulative Load Percent

Specifically, the graph in Figure 3 shows the cumulative load observed from the most loaded prefixes, as a percent of the total overall load.

The first item of note: While the snapshot analysis contained 160,102 prefixes, the graph reports information for the 40k most heavily loaded prefixes (x-axis). The graph doesn't show the remaining 120,102 other prefixes because while there is some load to them, it's insignificant compared to the total load.

A number of additional key points are highlighted on the graph. First, the 3,156 most heavily loaded prefixes account for 90% of the total traffic on our backbone (red line). Recall that the backbone FIB consisted of 575k routes. What this means is that 0.0054% of the IPv4 prefixes carry 90% of the load in

this snapshot. The 3,156 most heavily loaded prefixes are derived from 2,281 Comcast prefixes (green line) and 875 Internet prefixes (blue line).

A second highlight involves the point at which 99% of the traffic is carried. The top 25,893 prefixes carry 99% of the overall total load. These prefixes are made up of 5,884 Comcast prefixes and 20,009 Internet prefixes. What this means is that while our full Internet routes consist of 575k routes, 99% of the IPv4 load is carried via 25,893 prefixes—or 4.5% of the FIB. The inverse of this is that 95.5% of the prefixes in our full route FIB carry just 1% of the load, roughly a few hundreds of Gb/s.

What this analysis suggests is that a very lean FIB architecture is possible today—meaning we could significantly reduce the size of the IPv4 FIB with very little traffic (relative to the overall traffic load) being carried in an overlay IPv4aaS infrastructure.

More details on the relationship between various possible lean FIB sizes and the corresponding load that would be carried in an overlay network are provided in the Figure 4.

Figure 4 shows the amount of traffic that would need to be carried in an overlay IPv4aaS network, if the underlay (physical network) IPv4 FIB was limited to a number of lean sizes. While not specifically shown on the graph, the y-axis maximum value is on the order of a few 100s of Gb/s—so all lines represent loads well below this. Take, for instance, the red line in Figure 4, which corresponds to the IPv4 FIB of 25k prefixes. The graph shows the amount of traffic that would be in the overlay if the IPv4 FIB was limited to 25k prefixes. As we increase the size of the underlay FIB, we can carry more prefixes and hence the load to the overlay network decreases.

In this analysis, we choose which routes to put into the underlay FIB -- and thus what traffic would be carried in the overlay network. To simplify the operation of the overlay network, we opted to pack the underlay FIB with routes in a particular order. The order in which we selected routes to put into the underlay FIB was 1) our routes (highest loaded prefixes, 2) Internet aggregate routes (highest loaded prefixes first), and 3) Internet most specific routes (highest loaded prefixes first).

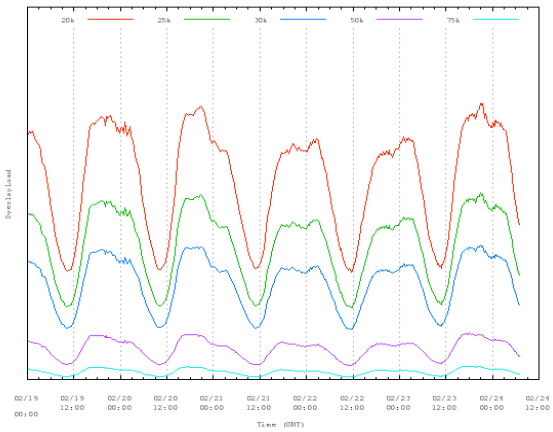


Figure 4: FIB sizes and Overlay Load

By putting all Comcast prefixes into the underlay FIB, the overlay network would only contain Internet prefixes and thus would only need to operate in one direction. This simplifies the routing architecture needed to pull traffic into the overlay network. This approach does, however, increase the load on the overlay network, because we service several thousand Comcast prefixes with very little traffic (e.g. customers with static address prefixes.) Internet aggregates were added before more-specifics, as they tend to be catchall announcements that subsume the load of the more-specifics, if the more-specifics disappeared from the upstream peer. This approach results in a higher amount of traffic in the overlay network, but we think it a reasonable tradeoff, because it makes the overlay network architecture simpler.

Traffic Patterns Enable Thin FIB

The data collected and corresponding analysis strongly suggests that today's IPv4 traffic patterns easily lend themselves to a Thin FIB underlay solution, without requiring excessive resources to build an overlay network. As shown in Figure 4, one could deploy routers with IPv4 FIBs holding well under 100k entries, and the overlay network would only need to carry upwards of a few 100s Gb/s at peak. Note that this overlay load would be sourced from our regional networks, such that any individual regional network would source only a portion of this load.

This means that a collection of cloud infrastructure resources, located within each regional network, should be sufficient to support the overlay infrastructure.

Next, let us examine the different technologies which could be used to create this overlay network.

IPV4AAS TECHNOLOGY OPTIONS

There are multiple technologies that could be used as the overlay mechanism. In this paper we focus our analysis and work on two technologies, namely MAP and LISP.

MAP

MAP (Map Address and Port) is an approach that enables a customer to have access to a global IPv4 address, but removes the need for this IPv4 address be natively carried by the underlying infrastructure. (The customer home network is assumed to be dual stack.) The customer home gateway (HG) is assigned an IPv6 address and a provisioned with a set of MAP rules. The combination of the assigned IPv6 address and the MAP rules are used by the HG device to algorithmically derive two things: Its globally (possibly shared) IPv4 address, and the range of ports it can use for its NATP (Network Address and Port) function. Because MAP is used to connect IPv4 islands over an IPv6 infrastructure, the MAP rules also define how to map the IPv4 addresses into IPv6 addresses. Two approaches are defined, one using encapsulation (MAP-E) and a second using translation (MAP-T).

MAP-E

The key aspect to MAP-E[2] is that traffic between IPv4 islands is fuly encapsulated in an IPv6 header. For traffic to an end-point outside of the MAP domain, packets must be encapsulated and sent to a MAP Border Router (BR), which then de-encapsulates the packet and forwards the IPv4 packet natively. When both IPv4 end-points are within the MAP domain, it is possible for the sending end-point to derive the destination end-point's IPv6 address, based on the mapping rules provided. In this case, traffic need not transit through a BR. It can be sent to the remote end-point directly, where it is de-encapsulated and IPv4 processing occurs. In the current MAP-E Internet draft, the model is that the HG device would have a default route to an identified BR for off-domain traffic.

MAP-T

In MAP-E, the IPv4 packet (header and payload) is encapsulated within a newly created IPv6 header. This extra level of overhead is avoided in MAP-T[3], because in MAP-T, the IPv6 header is synthesized from the original IPv4 header. The IPv4 payload (UDP, TCP) is then carried directly within the synthesized IPv6 header. In this approach, the IPv4 addresses are embedded into the synthesized IPv6 header. Much like the MAP-E system, for end-points within a domain, traffic can be directly transmitted. For an off-net end-point, the traffic must be sent to a Border Router (BR) device, which performs the de-translation process—meaning it takes the IPv6 header info and derives from it the original IPv4 header. The IPv6 payload

is then attached to the IPv4 header, and the reconstituted IPv4 packet is then forwarded by the BR.

Analysis of MAP

In both MAP approaches, the goal is to eliminate the carriage of IPv4 packets over the underlying infrastructure. A consequence is that it effectively enables a HG based stateless version of a Carrier Grade NAT—meaning an IPv4 address can be shared by many individual customers. The MAP solution is IPv4-specific, meaning its applicability is to carry IPv4 within an IPv6 infrastructure. It can't be extended to non-IPv4 payloads. MAP lends itself to an IPv6-focused network -- but doesn't eliminate the need for a portion of the network infrastructure to be fully IPv4-aware.

In order for a MAP solution to carry the same traffic engineering capabilities as our fully routed IPv4 network today, the BR devices likely need to be interconnected via a IPv4-based infrastructure, with corresponding IPv4 routing mechanisms in place. MAP can eliminate the need for IPv4 routing in the domain covered by MAP --it can enable a lean core in that portion of the network. However, the practicality of lifting the current scale of our IPv4 traffic into a MAP solution would be a serious undertaking. Further, the MAP solution is geared towards a solution where the domain edge is the customer home gateway (HG)—which means the HG device must support the MAP functionality.

LISP

An IPv4 address is both an identity (what an application binds to) and a location (used to announce how to reach an end-point.) The semantics of an address being both an identity and locator contributed to the scaling issues we see with IPv4 today. LISP[4] (Location Identity Separation Protocol) has its roots back in the late 1990s, when an active discussion thread about the "Next-Generation" Internet Protocol led to the development of IPv6. During the discussions leading up to the creation of IPv6, it was suggested that this "Next-Generation" IP address should have new semantics. Specifically, it was suggested that the identity of a device itself did not imply the location of the device.

In the end, IPv6 carried forward the same semantics location-identify semantics of IPv4. However, researchers continued to see value in technologies which decoupled a device's location from its identify. This thinking evolved into a protocol called LISP.

LISP really consists of two functions: an encapsulation protocol and a control plane. The encapsulation mechanism allows the carriage of one address family, inside another address family. LISP could, for instance, encapsulate IPv4 inside IPv4—if one wanted to create a lean IPv4 infrastructure. In this paper, we will examine LISP encapsulating of IPv4 packets inside of IPv6 packets.

The second LISP component is a control plane—a mechanism by which one determines the current, best remote-end point for the outer encapsulation layer address. The default LISP control plane is called DDT (Delegated Database Tree). It's a DNS-like query mechanism, where the LISP encapsulation engine receives a packet, looks at its destination, and then queries the DDT tree to determine the remote end point to send the packet to. The structure of the DDT tree is similar to DNS, in that there's a root node, and then portions of the "name space" are delegated to sub-nodes, with further sub-delegation possible. The LISP encapsulation engine performs the iterative queries through the DDT that determine the best current remote end-point and its IP address. Once this end-point IP address is determined, the packet is encapsulated and sent to the remote end-point. Much like DNS, scaling is

achieved by caching information learned from the DDT tree during the lookups. Like DNS, the cached information has a TTL (time-to-live) value to prevent the encapsulator from having stale information. Besides a TTL value associated with cached entries, the LISP encapsulation mechanism can also perform "probes" to remote end-points, to make sure they're still reachable and functioning correctly.

LISP Analysis

LISP is well suited to enable a lean core, carrying one address family as an overlay to another address family, operating as the underlay. For instance, one could build a IPv6-focused underlay and use the LISP encapsulation and control protocol for an IPv4 overlay. Overlays could also be created for L2 VPNs or L3 VPNs.

One of the downsides to LISP is the query-based control plane, in that the time to detect a remote end-point failure is a function of a cached TTL value, or how often probes are sent. One powerful aspect to the route announcement paradigm (not present with the DDT control plane) is that it is push-based. Changes to the routing system get pushed asynchronously, which makes it well suited in a dynamic environment.

Given LISP's ability to enable a lean IPv6 focused core for both IPv4 and other address families, the remainder of the paper details a LISP-based network overlay solution.

To address the issues with LISP's DDT control plane, we instead considered a routing approach --except this routing model leverages aspects often associated with the SDN. Details of an SDN-powered LISP solution are presented in the following sections.

IPV4AAS

Given the insights from our data analysis, we can incrementally move to a lean IPv6 focused architecture by carrying the IPv4 prefixes with no or little traffic on them via an overlay network. We propose that as a starting point, only the unused or low bandwidth internet prefixes be carried in the overlay network. As such, we detail an overlay network architecture that initially operates only for a subset our Internet traffic.

The overlay architecture can be built on our elastic OpenStack cloud infrastructure, and will require the cloud infrastructure to migrate from national data centers, down into the regional networks, and into the larger hub site locations. This evolution of the cloud eco-system closer to the edge (both hub site edge and peering edge) is being driven by multiple service and business requirements—IPv4aaS is but one of the service drivers for an edge-based cloud infrastructure

Architecture

In order to lift traffic onto this overlay network, we will deploy two or more "Ingress Tunnel Routers" (iTR) per regional network. An iTR is a LISP term for a device that takes native traffic (in our case, IPv4 traffic), performs a look-up on the destination address to determine the remote end-point (in our case an IPv6 address), and then encapsulates and forwards this encapsulated packet. The iTR can operate as VM within the regional network edge cloud infrastructure.

The encapsulated traffic is addressed to an Egress Tunnel Router (eTR). Its responsibility is to de-encapsulate the packet, then forward it, based on the inner IPv4 packet contents. eTRs can be co-located with our peering routers, and can also be VMs operating out of our cloud infrastructure. This high layer architecture is depicted in Figure 5.

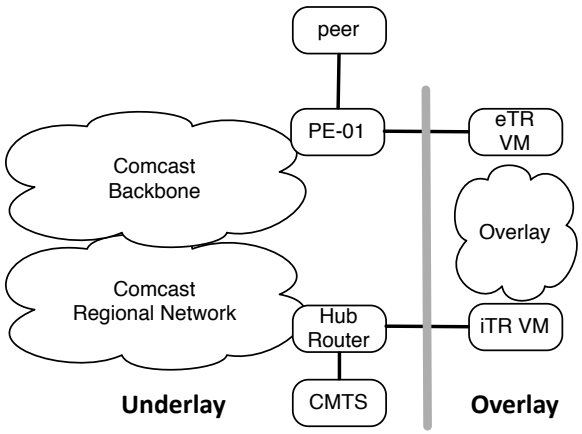


Figure 5: IPv4aaS Logical Architecture

Routing Architecture

In order to pull traffic into the overlay network, the iTR will need to inject routes into the underlay network. At the time of this writing (winter 2015), we are evaluating a number of routing designs, and present one of the approaches below.

Consider the overlay network as depicted in Figure 6. A peer will send IPv4 prefixes to the Comcast underlay network's peering router PE-01. PE-01 then sends all received routes to the overlay network eTR router. PE-01 also sends a subset of these routes towards the core underlay network's backbone routers. The selected underlay routes are the combination of Comcast prefixes, and the highest loaded Internet prefixes. Let's call these the Thin Fib Routes (TFRs). The TFRs are propagated from the backbone routers into each regional network and down into each hub site router, effectively populating the underlay network's IPv4 FIB.

Netflow data from the underlay network allows us to determine loads on the prefixes in the underlay network. For the overlay network, we are looking at how Openstack's Ceilometer[5] could be used as an alternative paradigm for Netflow-like data collection and back-end data analysis. Further, we are looking at how Openstack's Heat[6] might allow us to adjust iTR or eTR capacity as needed.

These collective systems would continue to monitor load in the underlay and overlay and over time adjust the prefixes in the underlay accordingly.

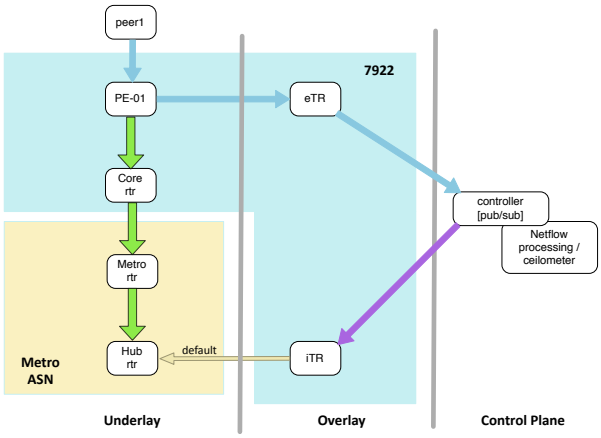


Figure 6: IPv4aaS Routing Architecture

BGP, JSON and HTTP

The eTR router takes the received BGP[7] (Border Gateway Patrol) announcements and translates them -- from classic BGP protocol format to BGP semantics encoded within a JSON structure. The eTR wraps the BGP JSON announcement information into an encompassing JSON structure that contains information for LISP. Specifically, the encompassing LISP JSON structure provides the RLOC (Routing Locator) which is the IPv6 address of the eTR.

What we are effectively doing is associating a BGP IPv4 prefix announcement with an IPv6 Next Hop. Note that this is not possible with BGP today, but, these semantics are easily encoded via the JSON logic performed on the eTR. The eTR then performs an HTTP PUT of this JSON object to a Pub/Sub entity. One can think of this Pub/Sub logic as being part of an SDN controller. The controller can process the incoming announcements, and adjust the set of announcements sent to an iTR. The iTR then receives the set of JSON

based BGP IPv4 prefixes with an IPv6 Next Hop, and performs classic BGP path selection. For each chosen IPv4 prefix, the iTR LISP forwarding engine is programmed with the target eTR IPv6 RLOC address. The iTR announces a default route, via classic BGP, to the hub router in the regional routing system.

The hub router will see the set of more specific TFRs and a default from the iTR. This way, traffic only gets pulled into the overlay iTR if the packet doesn't match against any routes in the underlay network. We envision the overlay network, and the SDN controller, to be part of the backbone ASN (Autonomous System Number). This enables the backbone operational team to adjust the amount of load which traverses the overlay network based on the routes present in the underlay FIB, and based on business rules applied by the SDN controller on the routes and associated BGP attributes sent to each iTR.

Prototype Development Status

To date, we have developed a working prototype of a LISP overlay network, including iTR and eTRs coupled with an implementation of the routing control plane architecture described above. This includes the BGP JSON encoding with the associated LISP extensions to enable a IPv4 prefix to have a IPv6 next-hop value.

One issue that we have run into is related to limited Network Function Virtualization (NfV) capabilities in Openstack. Specifically, a key feature needed for this overlay implementation is expected in the Kilo[8] Openstack release slated to be released in April 2015.

Alternative Encapsulations

While this work has leveraged LISP as the encapsulation technology, other encapsulation technologies could be used including Segment Routing[9] (SR). One could envision that an eTR could associated not just an RLOC with a IPv4 prefix, but also a stack of SR labels (MPLS) or Segment IDs (IPv6) that would be applied by the iTR for a prefix. Note that changes to the routing announcements and associated encapsulation details could also be performed by the SDN controller before sending the update announcement to an iTR.

Future Work

Future work includes considering how we might use SR and the above routing based push control plane to signal service chaining semantics for specific prefixes and or possibly more specific (prefix and port) announcements.

Summary

In this paper we have measured the loading of IPv4 prefixes carried in the Comcast full routing table, and determined that a significant number of FIB entries have little or no traffic. We used that information to suggest that a lean, IPv6-focused core network is possible today. We then then described a technology and routing architecture to support an IPv4aaS overlay network. We show that this overlay architecture can be instantiated on an OpenStack elastic cloud system, and can leverage SDN approaches to build both the routing control plane and a real-time measurement system. The measurement system ties classic network performance information (Netflow) to OpenStack SDN systems, including Ceilometer and Heat.

Acknowledgements

We would like to thank John Leddy, John Brzozowski and Paul Mabey for their thoughts in the many discussions we have had in this subject area. Further, we would like to thank Dino Farinacci for the use of his LISP

implementation and associated software changes to support the described IPv4aaS routing architecture.

---

[1] www.deepfield.net
[2] https://tools.ietf.org/html/draft-ietf-softwire-map-13
[3] https://tools.ietf.org/html/draft-ietf-softwire-map-t-08
[4] https://tools.ietf.org/html/rfc6830
[5] https://wiki.openstack.org/wiki/Ceilometer
[6] https://wiki.openstack.org/wiki/Heat
[7] http://www.ietf.org/rfc/rfc4271.txt
[8] https://wiki.openstack.org/wiki/Kilo_Release_Schedule
[9] http://www.segment-routing.net/home/ietf

Availability For Network Function Virtualization

Alon Bernstein
Cisco Systems

ABSTRACT

High availability (HA) is one of the most critical requirements of a service provider network and HA solutions for storage and compute have been deployed for some time. HA is also a major concern when considering network function virtualization (NFV) for service provider networks. The cloud culture of "just spin a new Virtual Machine (VM)" in case of a software or hardware fault seems at odds with the %99.999 availability requirements for service providera. This paper explores ways to meet the HA requirements of a service provider network with the relatively unreliable cloud infrastructure.

DATA CENTER AVAILABILITY

The following table translate the common benchmark for availability to time duration in year/month/week:

Availability	Downtime per year	Downtime per month	Downtime per week
99.99% "four nines"	52 minutes	4.3 minutes	1 minute
99.995%	26.2 minutes	2.2 minutes	32 seconds
99.999% "five nines"	5.2 minutes	26 seconds	6 seconds

Unfortunately some of the most visible data centers (DC) have suffered failures of several hours in recent years. In 2013 the Amazon cloud was down for 49 minutes exceeding it 99.995% claim. Google had a 4-hour outage in 2015. However the types of issues that caused those global outages, such as overload conditions or connectivity failures are out of scope for this paper and the assumption is that as technology matures this type of catastrophic events would became more and more rare. We will focus on basic single tenant availability where a single VM fails and another has to cover for it. Furthermore the focus is on packet forwarding VNF (virtual network function) as opposed to storage or compute availability.

NFV AVAILABILITY

The data center world likes to categorize VMs into "cows" and "pets". Cows are VMs that don't need much care, if one fails another quickly takes over. There are many cows and an individual failure is not critical. Pets are unique, relatively few and require care. If a pet fails it's a big deal.

A typical "cow" would be a VM that load shares a task with other VMs, for example serving web pages. Recovery of a cow is fairly straightforward. In case of a software failure the VM can be restarted and re-run. In case of a hardware failure a replacement VM can be started on another server. The process for creating a new VM is pretty much the same process as recovering from a failure and there is no state shared between the old and new VM. This greatly simplifies the handling of failures.

NFVs on the other hand are certainly under the "pet" category and have a few requirements for HA:

- Minimal packet loss: the role of an NFV VM is to process packets. It is obvious that in the case of failover packet loss must be minimal. Since the total number of drops depends on the time it took to detect the failure as

well as the recovery time it's hard to guarantee a zero packet loss.

- Stateful recovery: many NFVs require state. For example a virtual CMTS needs to keep track of E.911 calls. In case of a failure these states need to be carried on to the new VM.

A typical solution for achieving high availability that meets the requirements above is by means of redundancy; one VM is "active" and the other VM is "standby" – essentially the same solution used in the physical world but with one big difference. In the physical world this redundancy is typically intra-box and the implementation details are completely hidden from the user. In contrast the components of a virtualized redundancy solution are visible. This is not to say that it will be more complex from an end user point of view. The process of provisioning redundancy and handling failures can be fully automated. The components of the solution however are visible and can even be selected by the operator from either open source or vendor specific modules.

REDUNDANCY MODELS

There are several possible redundancy models. Most cable providers are familiar with the modes described below. All of which can be emulated with Virtual Machines (VM):

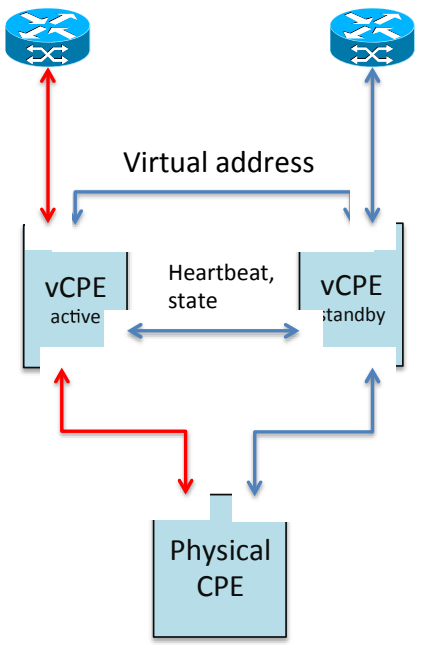
- Active/active: two separate instances share the traffic. If one fails then capacity is halved but selected services can still be prioritized and handled over the reduced capacity. In a physical CCAP device these could be two supervisor cards.
- Active/standby (1+1): as opposed to active/active the standby component carries no traffic and in case of a failure it assumes the identity of the failed card.
- N+1: in CCAP devices the linecards are relatively expensive and so to reduce the cost of high availability several linecards are backed up by a single linecard.

Other redundancy modes include:

- N+M: a group of N active instances are backed up by M standby.
- N-to-1: Similar to N+1 but the back up entity switched back to the primary instance once its recovered. As we will see later this mode may work well for NFV.
- N-to-N: its possible to combine the concepts of N+M and active/active so that each element has some extra capacity to cover for a failure but no specific element is designated as standby or active.

In this paper we focus only on the 1+1 and 1-To-1 redundant configurations and refer to "active" and "standby" VMs.

The data center uses the term "cluster" to describe a group of servers that appear as a single entity to the outside world (for example an e-mail server) and "HA cluster" for a solution that includes availability. In NFV virtual devices tend to be implemented as a single VM. We can still relate to the DC term of cluster with the understanding that it's a very simple cluster: an active VM and a standby VM.



.

Figure 1 Redundant Virtual CPE use case

Figure 1 depicts a particular use case of a physical CPE (pCPE) with an active and standby virtual CPE (vCPE) providing extra services. It builds upon existing use cases in the data center. For example the case of two e-mail servers using the same database would look very similar (where the e-mail server corresponds to the vCPE and the database is the shared resource and equivalent to the pCPE). Both active and standby vCPE have the same address to the outside world and are connected to redundant routers. This can be done in active/standby cases since only one path actively carries data. The obvious benefit is that a switchover is transparent to the rest of the network.

The red path in Figure 1 represents the packet flow. Since it's an active/standby packets will flow only along this single path.

The two vCPEs exchange heartbeat to detect failure on the active VM. Note that the heartbeat is only assisting in making a switchover decision and a failure of the heartbeat does not automatically cause a switchover. The criteria for switching over will be discussed in more detail in the "split brain" section.

CONNECTIVITY

The "cluster" that both the active and standby VM belong to is a single L2 domain from a networking point of view (see Figure 2). Even if we use geo-redundancy, where the active and standby VMs are in physically distributed data centers we would still connect the standby and active VM with an overlay (typically a tunnel of some sort) that would make them appear as belonging to the same network.

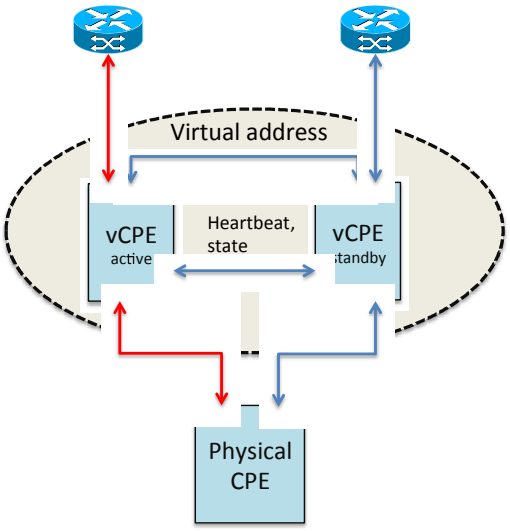


Figure 2 Connectivity overlay

It's the role of an orchestration system to create this overlay as part of the complete setup and life-cycle management of the active and redundant VM pair.

Note that in some cases, such as geo-redundancy, the redundant path may not be as optimal as the active path. For example, the redundant VM may be intentionally placed in a data center further away from the pCPE. For that reason the N-to-1 model may work better because as soon as the fault is resolved control will be returned to the optimal

instance as opposed to leaving the recovered VM as a "standby".

SPLIT BRAINS

When two VM try to decide which one is active there might be situations where both become active. The classical condition that can cause that condition is when the link they used for heartbeat check fails: the originally active VM stays up and the standby VM assumes the active is down (since the heartbeat stopped) and becomes active as well. When redundancy is managed within a box (either 1:1 or N+1) then the box management handles this ambiguity internally. Furthermore, in a traditional appliance the heartbeat is carried over a highly reliable internal path, which is not likely to fail in the first place. In the DC the "split-brain" is a well-understood issue and the solution used for redundancy of storage and compute can apply to NFV as well. As depicted in Figure 3 it requires the introduction of an object called "Quorum":

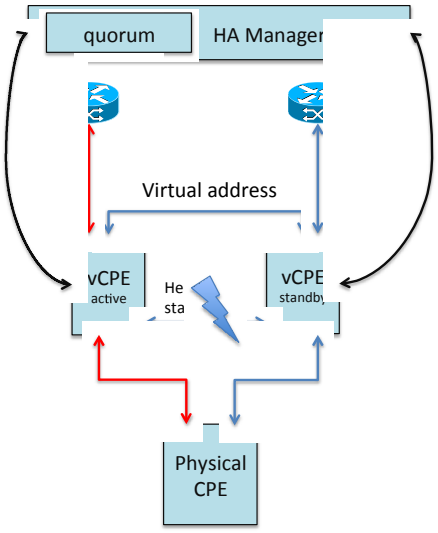


Figure 3 Redundancy Quorums

In order to become active a VM has to "own" the quorum. In the example outlined above the standby VM would try to acquire the quorum from the HA manager but would not succeed (because the active still owns it) and therefor it will stay in standby mode.
It is statistically possible that under certain conditions a couple of system are down, e.g. the redundancy manager itself and both vCPEs (and one would consider that as "double /triple faults" that are not covered by HA). Even in the above case it is impossible for two systems to be active at the same time.

AVAIABILITY AND ETSI-NFV MODEL

ETSI-NFV is an industry consortium that defined a unified framework for NFV deployments. The first phase of ETSI-NFV did not address the issue of availability in detail, but it is addressed in phase 2 (see ref 1). While this paper is not meant as an ETSI-NFV tutorial, it will outline the main functional components that are directly involved in creating a high availability solution.

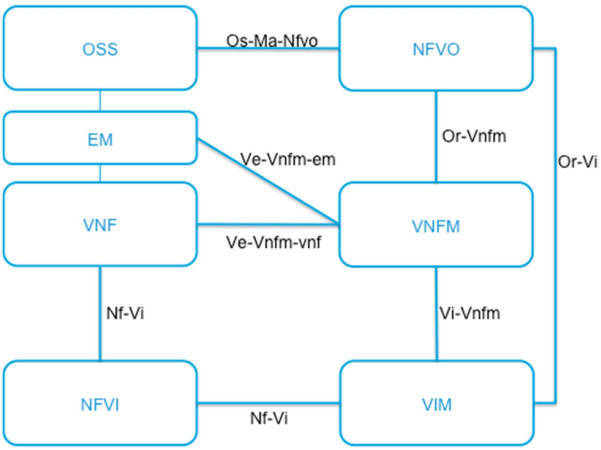


Figure 4 ESTI-NFV architecture

- The NFVI (Network Function Virtualization Infrastructure) should be unaware of a redundancy solution. A VM is a VM regardless of whether its active or standby
- The VIM (virtual infrastructure manager) should also be relativily unaware of redundancy. From the VIM perspective all it has to do is

create/destroy a VM regardless of its state.

- The VNF (Virtual Network Function) itself should know if its active or standby but should be the same SW package for both states.
- The VNFM (Virtual Network Function Manager) can act as a health monitor point for both the active and standby VM and will report the health monitor results to the NFVO layer that would make the final decision on making a switchover. In that sense even the NFVM is not that aware of the states of the VM (active/standby) and from its point of view its just two VMs it had to manage.
- The NFVO (NFV orchestration) is the element that triggers the creation of a redundant instance, the ownership of the quorum and the setup of the network overlay between the active and standby VM. Furthermore, the orchestration system has the final say on weather or not to perform a switchover. While it may seem risky to have the orchestration system be so involved in redundancy (as opposed to lower layers or pure peer-to-peer between the active/standby instances) it is the entity that can make the best decisions because of its system wide view.

STATEFUL RECOVERY

In some cases active and standby VM need to share state, e.g. routing tables or session information.
The active instance can update the standby either in a peer-to-peer fashion by sending state directly from the active VM to the standby or through a shared database where the active is a writer and the standby is a reader. RESTCONF/YANG can be a good transport of state between the VMs.
Some virtualization environments offer "VM mirroring" where the whole memory context of an active VM is synchronized to a redundant one. This greatly simplifies the SW since no state synching code has to be written and the applications are completely unaware of the fact they are synchronized, however, it takes a lot of BW to pass memory blocks around. It may make sense to do that in compute or storage use-cases where network bandwidth it not a bottleneck, but mirroring clearly does not make sense for NFV packet processing.

PLANNED FAILURES

Once a redundancy framework is in place it can be used as a platform for "planned failures" as well.
A prominent example is software upgrade. Instead of bringing the service down during an upgrade procedure the following workflow can achieve the upgrade with a minimal service disruption:

- Update standby VM with new SW.
- Switchover and verify operation. If failed switch back and downgrade
- If upgrade operational then update the other VM
- Switch back. Now both instances run the new software.

CONCLUSIONS

Typical cloud deployments are not %99.999 available. Redundancy for NFV can help in reaching this availability goal.
In traditional appliances redundancy is an internal implementation issue. In the cloud the details of managing redundancy are visible and the modules that comprise them interchangeable. Thanks to automation this extra visibility and flexibility need not make operation more complex.

REFERANCES

1. ETSI-NFV phase 2 : http://www.etsi.org/news-

events/news/850-2014-12-news-etsi-network-function-virtualization-enters-phase-2

BEYOND FTTH - FROM FTTH TO FIBER IN THE HOME

Dr. Yuxin (Eugene) Dai, Network Architecture group, Cox Communications

Wei Dai, Computer Engineering Department, University of California Santa Barbara

Abstract

FTTH has enjoyed worldwide deployments since the mid 2000s. The main technology for FTTH today is TDM PON, but in the future this may migrate to other technologies such as WDM PON, OFDM PON, and advanced modulation. In this paper, we first review the FTTH standards development and then discuss the migration of the last-mile fiber networks beyond FTTH – the extension of fiber network infrastructure from access networks to home networking, related architectures, technologies, as well as applications.

Keywords: FTTH, Fiber in the Home, TDM PON, WDM PON, POF, M-PAM

I. INTRODUCTION

FTTH (Fiber to the Home) is a last mile optical access network architecture. It has Point-to-Point (P2P) and Point-to-Multi-Point (P2MP) architectures. These architectures are distinct from the fiber topology and the protocol layer. The class of P2P architecture can be further split into P2P fiber and active Ethernet. P2P fiber architecture, as its name indicates, has a fiber termination between a headend or Central Office (CO) and a customer premise. Due to the high cost of outside plant (OSP), P2P fiber architecture is relatively unattractive in access networks. Active optical Ethernet, as shown in Fig. 1, has a P2P architecture from the protocol point of view. However, it actually has a P2MP fiber topology.

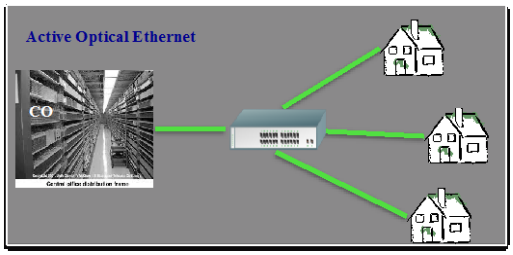


Fig. 1. The active optical Ethernet architecture

In Fig. 1, if we replace the active Ethernet switch, which is located in the field, with a passive optical power splitter, we get a passive Optical Distribution Network (ODN) that Time Division Multiplexed Passive Optical Networks (TDM PONs) are built upon today, as shown in Fig. 2.

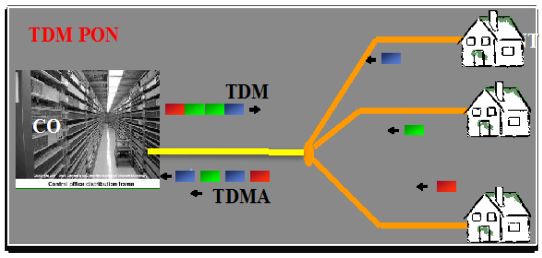


Fig. 2. TDM PON with P2MP passive ODN

Wavelength Division Multiplex (WDM) PON also has a P2MP fiber topology, though an Arrayed Waveguide Grating (AWG) could replace the passive optical splitter in the ODN. Alternatively, a WDM PON can have a hybrid AWG and power splitter ODN architecture.

TDM PON has reached 10Gbps today. For example, IEEE 10G EPON supports up to 10Gbps in both downstream and upstream

directions [1]. ITU-T XGPON1 [2] supports 10Gbps in the downstream direction and 2.5Gbps in the upstream direction. The FSAN/ITU-T is currently developing a NG-PON2 standard that has aggregated 40Gbps system capacities with 4 pairs of wavelengths on a hybrid AWG/power splitter ODN. Meanwhile, IEEE has begun Next Generation EPON studies.

WDM PON is still in the research stage; there are no industry standards for WDM PON as of today. However, NG-PON2 can be viewed as a hybrid TDM and WDM PON.

With FTTH system capacities reaching 10Gbps and beyond, service providers now offer Gigabit services to residential and business customers. When bandwidth reached Gigabit/s and beyond to the home gateway, the old copper-based home networking infrastructures became the bottleneck. New home networking physical media and standards are needed for Gigabit service FTTH. Plastic/Polymer Optical Fiber (POF) is a new promising medium for Fiber in the Home networking (FITH).

In this paper, we will first review the next generation PON standard developments at IEEE and ITU-T, followed by a discussion of current home networking standards, and then discuss the Fiber in the Home and Fiber to the Device (FTTD) concepts with POF for home networking and related architectures, technologies as well as applications.

II. NEXT GENERATION PON STANDARDS

TDM PON is the main technology behind FTTH. The first TDM PON standard was APON (ATM PON); it has symmetric 155Mbps upstream and downstream bandwidth and was developed at FSAN/ITU-T in 1998 as G.983.1 [3]. APON has very limited deployment; it was shortly replaced by asymmetric Broadband PON (BPON), which has 622 Mbps downstream and 155Mbps upstream bandwidth. The BPON standard was published in ITU-T G.983.3 [4] in 2001. BPON had mild deployment, and was quickly replaced by GPON, which was published in G.984.1 to G.984.6 [5] in the 2004 to 2008 time frame. GPON has been deployed worldwide. ITU-T since then completed G.897 asymmetric 10G/2.5G XGPON1 [2] in 2010. There is little to no commercial deployment of XGPON1 as of today. FSAN and ITU-T are currently working on NG-PON2 that is based on hybrid TDM and WDM technologies.

IEEE developed the symmetric 1Gbps EPON standard 802.3ah [6] in 2004, and large-scale deployment started shortly afterwards (mostly in China and Asia-Pacific countries), and continues today. The asymmetric 10G/1G and symmetric 10G/10G EPON standard 802.3av [1] was completed by 2009. The large-scale deployment of 10G EPON started in China in late 2013. IEEE started NG EPON Industry Connection (IC) activities in early 2014. The NG EPON IC may eventually lead to the NG EPON Study Group at IEEE.

A. ITU-T NG PON2

FSAN/ITU-T is currently working on NG-PON2 standardization. When completed, it will be published in the G.989.x series. The main parts of the NG-PON2 standard are expected to be complete at the end of 2015.

NG-PON2 has a hybrid WDM PON and TDM PON architecture, as shown in Fig. 3. NG-PON2 can be viewed as a TDM PON operating over WDM PON channels. As shown in Fig. 3, the first TDM PON OLT is connected to the first WDM wavelength channel, and the second TDM PON OLT is connected to the second WDM wavelength

channel, and so forth. The output port of the Arrayed Wavelength Grating (AWG), a wavelength multiplexer, is fed into a power splitter ODN. In NG-PON2, an ONU can be assigned to any of the TDM PON OLTs during the ranging of the WDM PON. The TDM PON wavelengths are assigned during the WDM PON ranging. In order to do so, tunable lasers and tunable optical filters are needed at the OLT and ONU as well.

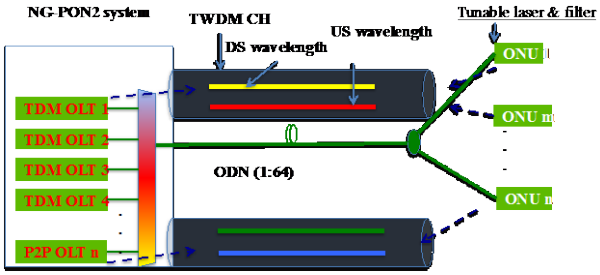


Fig. 3. NG-PON2 architecture

A NG-PON2 system has aggregated capacities of 40Gbps with 4 pairs of wavelength channels, assuming a 10 Gigabit TDM PON OLT is in each of the NG-PON2 channel.

In summary, the NG PON2 is a hybrid WDM PON and TDM PON, sometimes referred to as TWPON (TDM WDM PON). It supports four 10 Gigabit TDM PONs with an aggregated system capacity of 40Gbps. It is optional to support P2P WDM channels for a mobile front haul, or more TDM PON channels.

B. IEEE 802.3 NG EPON

IEEE started Next Generation EPON activities recently. The NG EPON IC (Industry Connection) activity was formed in Q4 2013. Industry Connection is a pre-standardization activity at IEEE. The goal of NG EPON IC is to raise awareness of the need for EPON beyond 10Gbps. Hopefully, the NG EPON IC will lead to the formation of a NG EPON study group at IEEE 802.3, and therefore lead to NG EPON standardization at IEEE.

There are many ways to achieve higher capacities for PON. As shown in Fig. 4 there are several architectural choices for NG EPON [7]. For example, WDM can be used at 10G EPON to achieve 40G aggregated capacities with 4 lambdas like that in NG-PON2. Or, we can go to a higher TDM rate first, for example, 25Gbps EPON, and then employ WDM to have 100G system capacities with 4 lambdas.

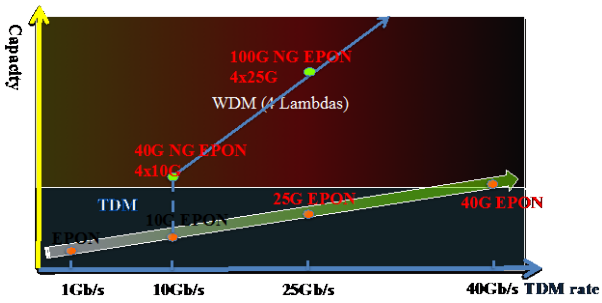


Fig. 4. NG EPON architecture choices

C. WDM PON

WDM PON has a P2MP ODN with P2P terminated wavelength channels between the subscribers and the headend or CO. As a result, it has the advantage of low CAPEX with P2MP passive ODN, high bandwidth, and high QOS with P2P wavelength channels. WDM PON is ideal in providing Gigabit services today and has the ability to provide scalable bandwidth to even higher data rates in the future.

WDM PON has been a research topic for many years; there is no single type of WDM PON architecture. The choice of different

"colorless" technologies could affect the passive ODN architecture. Also, the choice of OLT light sources such as tunable lasers, laser arrays, single wavelength lasers, etc., may also affect WDM PON ODN architectures and protocols.

WDM PON is considered to be a future-proof architecture. However, major problems for WDM PON include immaturity of the key enabling technologies, the high cost associated with these technologies, too many architecture variations, and a lack of standards. However, the recent Gigabit services over FTTH may provide market motivation, and NG-PON2 may accelerate the cost reduction of tunable optics. Therefore, the time for WDM PON may come sooner than expected.

III. A BRIEF REVIEW OF HOME NETWORKING STANDARDS

Advancement of FTTH technology enables gigabit services to the home. IEEE 10G EPON provides 10Gbps symmetric shared bandwidth to 32 or more customers, and ITU-T NG-PON2 can provide 40Gbps aggregated capacities. However, the physical media for home networking are age old, and some can date back to 1880s. The most common media for home networking today are twisted pair phone lines, power lines and coax cables—or, air with Wi-Fi. Although these are different media, they share the same limitations:

- Limited spectrum
- Limited bandwidth
- Vulnerable to RF interference
- Vulnerable to EMI
- Low PHY and MAC efficiencies
- Complexity at PHY & MAC layers

A. Twisted pair phone line

The application of twisted pair phone lines can be dated back as early as the 1880s when the telephone was invented. Although twisted pair phone lines are still used in copper access networks for POTS phone services and data services with XDSL technologies, they were rarely used for home networking until G.hn [8] was recently published. G.hn is a new home networking standard published by ITU-T in 2010. The main application of G.hn is for home networking with twisted pair phone lines, although it has variations for coax cables and power lines. The G.hn twisted pair version supports PHY rates up to 900Mbps with 500Mbps MAC rates.

B. Power line

Power Line Communication (PLC) has been used in the power industry for many years for internal communications. Only in recent years with HomePlug AV [9] has PLC been used for home networking. The HomePlug AV standard published in 2005 supports 200Mbps PHY rates with 80Mbps MAC rates. The newer HomePlug AV2 [10] specification published in 2012 supports 750Mbps PHY rates in Single Input Single Output (SISO) mode with MAC rates up to 300Mbps. HomePlug AV2 also supports Multiple Input Multiple Output (MIMO) mode with PHY rates up to 1.5Gbps.

C. Coax cable

Coax cable has been the main physical medium in home networking for video and data services since the mid 1990s when the Internet became popular. Multimedia over Coax Alliance (MoCA) published MoCA 1.0 [11] in 2006 and MoCA 2.0 [12] in 2010. MoCA 2.0 base mode supports 700Mbps PHY rates with 400 Mbps MAC rates. MoCA 2.0 extended mode supports up to 1.4Gbps PHY rates with 800Mbps MAC rates.

Table 1 shows a comparison of wire line based home networking standards. It is interesting to note that the PHY rates of home

networking standards in Table 1 are 1.8 times to 2.5 times higher than the corresponding MAC rates.

Table 1. Comparison of home networking standards

Media	Standard	PHY rate	MAC rate
Twisted pair	G.HN	900 Mbps	500 Mbps
Coax	MOCA 2.0 base	700 Mbps	400 Mbps
	MOCA 2.0 Extended	1.4 Gbps	800 Mbps
Power line	HomePlugAV	200 Mbps	80 Mbps
	HomePlug AV2 (SISO)	750 Mbps	300 Mbps

We can see from Table 1 that a commonality between these home-networking standards is that they have rather low PHY efficiency. The excessive PHY overheads come from several sources. All these copper-based PHYs have to operate with various EMI and RF noise sources, and therefore strong FECs are needed. This adds relatively large overhead. Since copper-based media have limited bandwidth, advanced modulation, such as OFDM, is used that add additional overhead for channel estimation, pilot tones, channel bonding, etc.

Coaxial cables are shielded media that are less vulnerable to RF noise. Needless to say, MoCA still has rather low PHY efficiency. There are many contributing factors, one is the complexity associated with the Multi-Point-to-Multi-Point topology. MoCA has a MP2MP architecture using coax cables. For a meshed MP2MP coax network using RF power splitters, some connections only experience splitting loss while other connections experience isolation loss that is normally >25dB. Take a 1X2 RF splitter for example, as shown in Fig. 5, connection A-B has a 3 dB splitting loss. In contrast, connection B-C has >25dB isolation loss since it has to cross port B and port C. Therefore, the transmitter for link BC has to have higher burst power than that of link BA. This high power transmission will cause strong reflections that behave as noise.

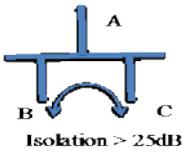


Fig. 5. Links across 1X2 RF splitter in MoCA architecture

Low PHY layer overhead is essential for media that have limited bandwidth, such as twisted pair phone line, PLC and coax cable. However, the home networking standards with all these media have rather high PHY overhead. The excessive PHY overhead further limits copper-based media for Gigabit Smart Home networking applications.

D. Air

For Wi-Fi based home networking, air or vacuum (better be air) can be considered the physical medium. The capacity of Wi-Fi for home networking is essentially limited by the available RF spectrum.

IV. FIBER IN THE HOME WITH POF

The FTTH technologies continuously evolve to higher bandwidths; Gigabit services to the home gateway are available today. The Smart Home requires high bandwidth connectivity in the home, driven by services such as cloud, 4K video, multi-room DVR, high speed Internet, Internet of Things, etc. The old home network physical media have become the bottleneck. The copper-based home networking media and standards have limitations on bandwidth and low PHY efficiency. The capacity of Wi-Fi home networking is limited by the available RF spectrum and by regulations. FITH is the natural and inevitable path to extend FTTH bandwidth within homes.

POF is the most promising solution for FITH. Compared with Glass Optical Fibers (GOF) such as Single Mode Fiber (SMF) and Multi-mode Fiber (MMF), POF has a large core diameter. For example, Step Index POF (SI-POF) has a core diameter up to 1000um that translates into low cost connectors as well as easy installations and terminations. Fig. 6 illustrates the relative diameters of POF with MMF and their capacities.

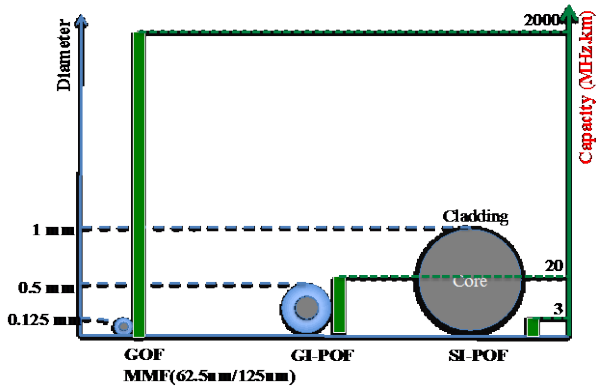


Fig. 5. Comparison of POF with MMF

The low cost POF solution for home networking is also reflected in the low cost light sources for POF transmissions. Unlike GOF, incoherent light sources such as LEDs can be used for POF transmission, which greatly reduces the cost of POF transmitters. The coupling of light sources to POF is much easier than to GOF. Vertical-cavity Surface-emitting laser (VCSEL) arrays can be used with POF for high density P2P links.

The capacity of SI-POF is around 3MHz-km, see Fig. 5. In order to achieve Gigabit rates with 50-100m distances, advanced modulation has to be used. The choice of modulation formats includes QPSK, QAM, OFDM, M-PAM, etc. Among these modulation formats, M-PAM is suitable as a low cost solution with relatively high bit/s/Hz efficiency. M-PAM is a one-dimensional code; it is much less complex than two-dimensional codes such as QAM and OFDM. Another advantage of M-PAM is that it uses low cost Intensity Modulation – Direct Detection (IM-DD) methods for transmission and detection.

The bit error rate (BER) of M-PAM can be expressed as

$$BER = \frac{M-1}{Mlog_2 M} erfc\left(\sqrt{\frac{3}{2(M^2-1)}\frac{S}{N}}\right) \quad (1)$$

where M is modulation order and S/N is signal to noise ratio. Assuming BER=10^{-6}, the required S/N for 8-PAM is 26.56 dB.

The theoretical spectral efficiency of M-PAM is $2\log_2(M)$. With the consideration of FEC, the true spectral efficiency be expressed as:

$$C_{PAM} = 2\log_2(M)*R \quad (2)$$

where R is the FEC code rate. Take 8-PAM with a FEC code rate at 0.90 for example, the spectral efficiency is:

$$C_{PAM} = 2\log_2(8)*0.90 = 5.40 \text{ bit/s/Hz} \quad (3)$$

The Shannon limit is expressed as:

$$C_{Shannon} = \log_2(1+ S/N) \text{ bit/s/Hz.} \quad (4)$$

The low boundary of S/N according to equation (4) can be expressed as

$$10\log_{10}(2^C -1) \quad (5)$$

For spectral efficiency at 5.40 bit/s/Hz, the Shannon limit gives 16.2 dB as the low S/N boundary.

The low cost M-PAM modulation with POF gives sufficient spectral efficiency for Gigabit/s home networking. When transmitting Gigabit/s Ethernet over POF, the PHY efficiency can be increased by changing

the 8B/10B Gigabit/s Ethernet line code to more efficient 64B/66B line code. This alone can reduce about 25 percent PHY overhead. Optimizing FEC is another area to improve the efficiency of Gigabit/s Ethernet over POF. LDPC can be used for higher coding gain and lower overhead. In general, Gigabit Ethernet POF (GEPOF) is expected to have lower PHY overhead than that of copper-based solutions for home networking.

IEEE started Gigabit Ethernet over POF standardization in May 2014 under the 802.3 Working Group.

V. FTTH WITH FITH

Fiber in the Home with Gigabit/s Ethernet POF can be used with FTTH to extend Gigabit/s services to home networking devices.

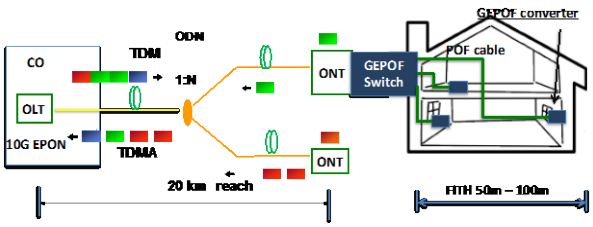


Fig. 6. FTTH + FITH solution

Fig. 6 shows the FTTH + FITH solution [13]. In this scenario a 10G EPON provides up to 10 Gigabit/s connectivity to the ONT at customer premises with Gigabit/s service to the home. A Gigabit POF switch/converter converts the Ethernet frames onto the M-PAM modulated GEPOF links to provide P2P Gigabit/s Ethernet links to the GEPOF converters in the home. At the GEPOF converters 1000BASE T interfaces are provided for Ethernet devices in the home. The GEPOF converters can be powered by DC power, or, they can be powered by Power over Ethernet (POE) from consumer devices.

FITH is a solution where P2P POF cables are terminated at the wall jacks with a 1000BASE T interface to consumer devices via a short CAT5 cable.

VI. CONCLUSION

FITH is a natural and inevitable next step to extend FTTH in order to serve the Smart Home with Gigabit/s and above bandwidths. POF is a promising physical medium for next generation home networking technologies.

REFERENCES

[1] 802.3av Part 3 Amendment 1: Physical Layer Specifications and Management Parameters for 10 Gb/s Passive Optical Networks

[2] G.987, “10-Gigabit-capable passive optical network (XG-PON) systems: Definitions, abbreviations and acronyms”, ITU-T 2010

[3] G.983.1 “Broadband optical access systems based on Passive Optical Networks (PON)” ITU-T 1998

[4] G.983.3, “A broadband optical access system with increased service capability by wavelength allocation” ITU-T 2001

[5] G.984.1, “Gigabit-capable passive optical networks (GPON): General characteristics, ITU-T 2008

[6] 802.3ah Part 3 Amendment: Media Access Control Parameters, Physical Layers, and Management Parameters for Subscriber Access Networks

[7] Yuxin (Eugene) Dai “NG EPON requirement and architecture considerations II” IEEE 802.3 Plenary Meeting Next Generation EPON Industry Connections, March 16 - 18, 2014, Beijing, CHINA

[8]G.9960 “Unified high-speed wireline-based home networking transceivers - System architecture and physical layer specification” ITU-T 2011

[9] Homeplug AV, HomePlug Powerline Alliance, 2005

[10] Homeplug AV2, HomePlug Powerline Alliance, 2012
[11] MoCA 1.0, Multimedia over Coax Alliance,2006
[12] MoCA 2.0, Multimedia over Coax Alliance,2010
[13] Yuxin (Eugene) Dai, “GEPOF for home networking” IEEE 802.3 Interim Meeting, Gigabit Ethernet over POF Study Group, September 8 -12, 2014, Ottawa, Canada

CPE CACHING – USING NETWORK INEFFICIENCY TO DELIVER BETTER INTERNET STREAMING

Scot Loach
Aterlo Networks

Abstract

We are in the midst of a shift in TV viewing behavior from a traditional broadcast-based experience to an on-demand experience delivered over the Internet. At the same time, video bit-rates are increasing as 4K resolution becomes mainstream. These trends are putting pressure on the economics of providing Internet access, since networks are built to peak usage and the majority of video streaming still occurs during "prime time". Caching technologies are being used to respond to this problem, but traditional caching technologies only address the transit edge of the network. Caching does not help the access network, which is often difficult and expensive to upgrade, since it is constrained by physical properties such as limited radio spectrum.

This paper presents the concept of caching within the customer premises. We discuss the opportunities and challenges associated with this approach and illustrate the architecture of such a system. We present a study that models customer premises caching on a consumer broadband network, and show the potential cache efficiency and resulting savings predicted.

INTRODUCTION

Video on Demand (VoD) over the Internet is replacing traditional linear broadcast video, as subscribers increasingly use services such as Netflix and Hulu to watch TV episodes and movies. These services have made great strides in recent years in the amount of content they offer, their ease of use, and their brand awareness with the average consumer. Many TVs natively support these services, so a consumer can simply connect a TV to their home network and use these services without any additional hardware. This is causing an upward trend in the number of subscribers using these services.

Although the technology used to watch TV is changing, the behavior of TV viewers still follows a "prime time" usage pattern, where the majority of users watch TV during a few hours in the evening. At the same time, streaming bitrates are on the rise, as more consumers adopt High Definition (HD) and Ultra High Definition (UHD, 4K) screens. HD video at 1080 lines of resolution is typically encoded at about 5Mbps, and UHD video at 2160 lines of resolution is typically encoded at 16-20Mbps. A standard definition video is encoded at about 1.5 Mbps.

This combination of increased adoption and higher bitrates is causing an exponential increase in the volume of traffic on the Internet during prime time hours. Figure 1 shows a projection by Cisco showing the increase in prime time relative to the average hour of the day. One consequence of this is an increasingly inefficient network, since the cost of building a network is proportional to its peak required capacity. Figure 2 shows a typical average day on a consumer broadband network; in coming years, the peak of this chart will move higher relative to the rest of the day. This represents a high cost of network upgrades and an increasing waste of resources, as much of the network capacity sits unused for most of the day. Consumer broadband networks are designed with oversubscription, assuming that usage is spread out over the day. The trend towards the majority of subscribers using the Internet to watch TV puts pressure on the economics of providing Internet access to consumers.

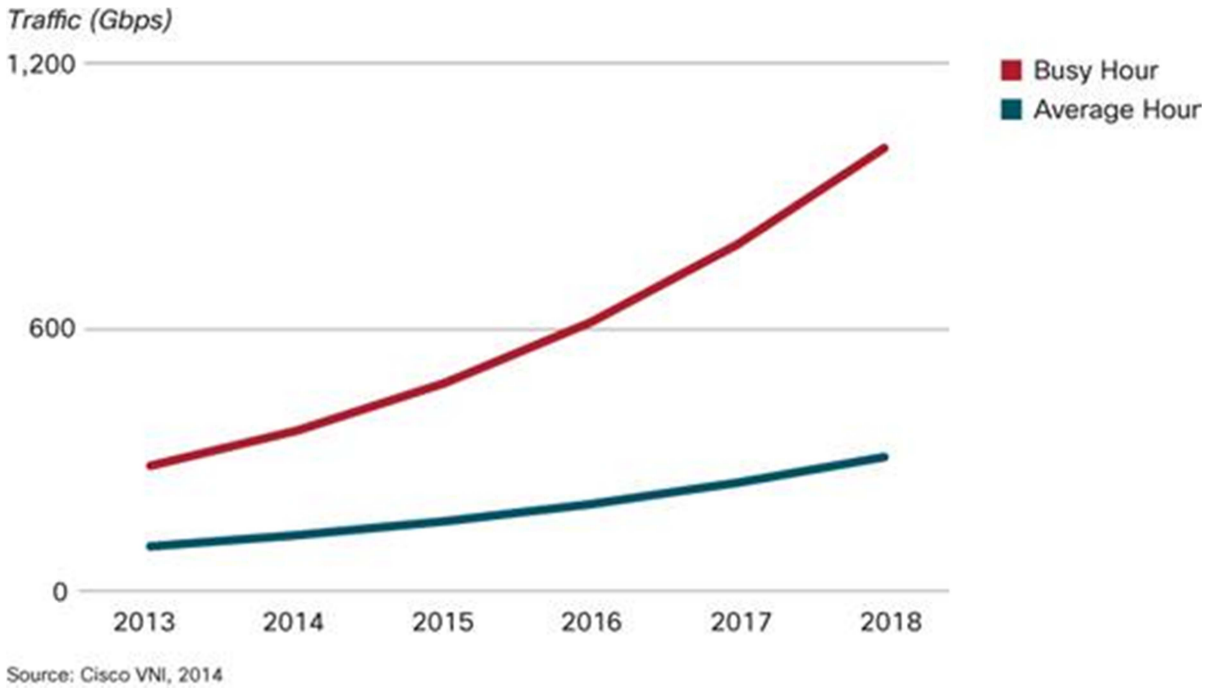


Figure 1. Busy hour traffic is growing faster than average hour traffic.

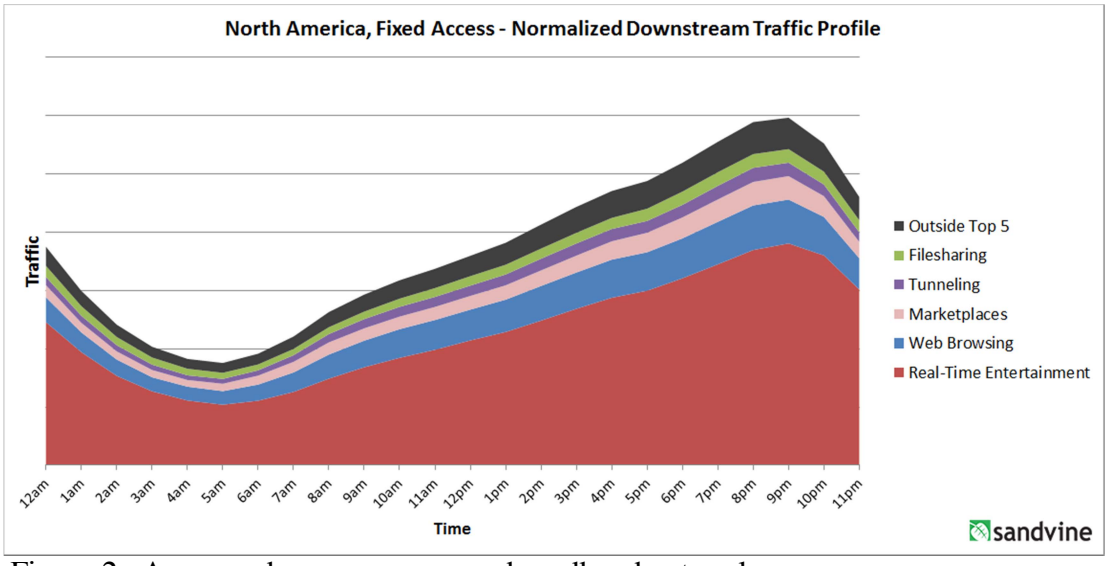


Figure 2. Average day on a consumer broadband network.

Communications Services Providers (CSPs) react to this situation with a variety of strategies. The most common strategies are increasing capacity, reducing demand, and increasing efficiency.

Increasing capacity means upgrading the network. This is a continuous business process for CSPs, and can be extremely expensive, especially on the last mile access network. The access network is typically limited by physical constraints such as

wireless spectrum or physical cables in the ground; increasing capacity is a costly endeavor that takes a long lead time to plan. The planning cycle for increasing capacity is usually an annual process that is tied to the budget cycle of a CSP.

Reducing demand is usually done by putting policies in place to incent subscribers to stream less video, or to stream it during off-peak times. Subscribers may be convinced to stream less video or at a lower resolution by enforcing usage quotas on the network, and subscribers may be convinced to stream video at off-peak times by putting an un-capped time period into place, for example during the night. These policies are unpopular with subscribers and may cause churn as subscribers move to competitors with fewer restrictions.

Increasing efficiency can be done by putting optimization systems in place such as network caches. Service-specific network caches are offered by some of the major content providers such as Google and Netflix, and service-agnostic network caches are offered by several network equipment vendors such as PeerApp, as well as open-source implementations such as Squid. Network caches are appliances that sit in the network, load popular content into large storage arrays (passively or during off-peak hours), and serve the content from their storage instead of the Internet. Caches are an effective way to reduce the amount of data volume between the CSP's network and the content provider's servers; however, a cache can only improve efficiency on the part of the network that is upstream from it. Therefore traditional network caches that sit in the core network cannot improve the last mile access network, which is often the most expensive part of the network to upgrade. A cache in the customer premises could improve the efficiency of the entire network, including the access network. This would make it possible to address the bandwidth crunch caused by the increasing usage of VoD over the Internet, while keeping the capacity upgrade budget under control, all without alienating subscribers.

This paper examines the concept of a "CPE Cache" which is a network cache implemented within the customer premises. An overview of caching technology is provided, with a discussion of how this relates to CPE caching. Architectural considerations and tradeoffs for building a CPE cache are discussed. Finally, the results of a case study applying a CPE caching model to real network traffic are presented.

CONCEPT

Network Cache Overview

A network cache can be modeled as a system that transforms one traffic pattern on a network link to a different traffic pattern. The purpose of a network cache is to reduce the peak bitrate of its upstream link, and therefore the entire network upstream of it. Figure 3 shows this model of a network cache. In this paper, we assume a cache does not change the traffic pattern of its downstream link, although in reality it is possible for this to occur if a link is congested, for example by

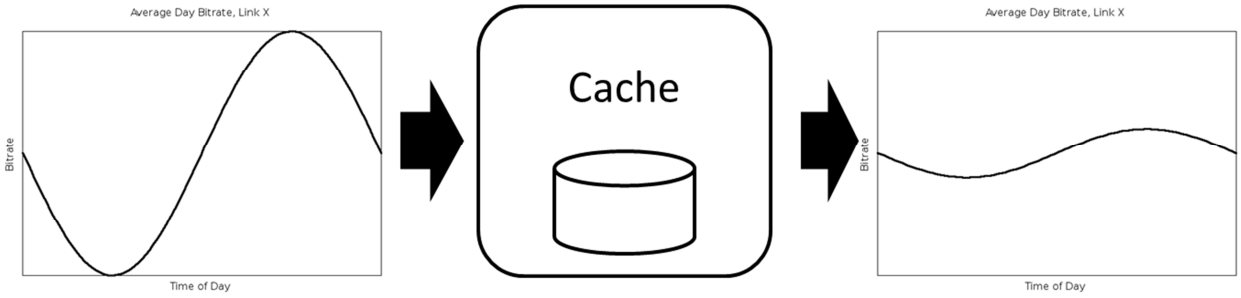


Figure 3. High-level model of a network cache

adaptive video streams shifting to a higher bitrate when congestion is relieved by caching.

The peak bitrate of a link is defined in terms of how it is measured, and the interval it is measured over. This may depend on physical and financial considerations. For example, a transit link that is billed monthly based on 95th percentile downstream bitrate may use the 95th percentile downstream bitrate as the peak, and measure this at 5 minute intervals over a calendar month. For a last-mile network segment such as a DOCSIS channel that is a capital expense, the link efficiency may be monitored quarterly as part of a quarterly reporting package, and again the 95th percentile may be used as a way to eliminate outliers.

The ultimate measure of a cache's effectiveness is the reduction in peak downstream bitrate on its upstream link (the network link that packets take between the cache and the Internet). A cache reduces peak bitrate by serving bytes from its storage; the number of bytes served from cache is noted B_C and the total bytes transmitted downstream from the cache (from the Internet and from cache) is noted B_T. Traditionally, a network cache is evaluated based on cache efficiency or cache hit rate, which is defined as $R_{CH}=B_C / B_T$. However, such a measure only correlates to an improvement in peak bitrate if it is measured exclusively during the peak time of network traffic. During off-peak times, there is idle network capacity that can be used to improve the effectiveness of the cache by speculatively downloading content to the cache. This activity lowers the cache hit rate during those times, since B_T is increased with no increase in B_C; but there is no added cost to doing this since the cost of the network is defined by its peak utilization. For example, the Netflix OpenConnect appliance works by downloading new content to its storage during off-peak hours. This means it can service even the first request for any stored content during peak hours. This strategy increases link efficiency and overall traffic volume in order to reduce peak bitrate.

The effectiveness of a network cache depends on its ability to have requested content in its storage. Assuming infinite storage, and technical feasibility of filling it, a cache could remove all static content from its upstream link. However, there are both physical and financial limits to storage, so caches have methods for filling the storage to optimize their effectiveness. This paper defines storage efficiency as the ratio of bytes served from the storage on the cache to the total amount of bytes stored on the cache. $R_S=B_C/B_S$. The storage capacity B_S is the number of bytes available for storing cached content.

Another consideration related to network caching is the concept of Quality of Experience (QoE). Subscribers judge quality of experience by factors such as the quality of the video encoding, the resolution of the video, and any visual impairments in the video due to network congestion or hardware capability such as dropped frames, buffering "wait" messages, and transitions between different resolutions when adaptive algorithms are operating in changing network conditions. This paper does not undertake a detailed discussion of Quality of Experience, but we assume that the quality of experience for video delivered from a cache is equal to or better than the quality of experience delivered from a remote CDN, since there is less probability of experiencing network congestion the closer the content source is to the user.

CPE Caching Technology

A traditional network cache covers some sample of the subscriber population, and is popularity-based, meaning it fills its storage with content that is most likely to be consumed by the largest number of

subscribers. This works well, since most Internet content follows a power-law distribution where a small subset of the total content will be accessed by many users within a time period. For example, although a video streaming service may have 50,000 videos, in general 80% of the videos streamed in a day may be within the top 20% of the total content. Therefore a network cache could reduce the upstream bitrate of that service by 80% if it is able to store the most popular 10,000 videos.

In contrast, a customer premises equipment cache (CPE Cache) only serves a small number of subscribers - in many cases only a single subscriber. A naive approach to achieving a result equal to a traditional network cache would be to store the same 20% of popular content on every CPE Cache. Although this would achieve the same overall savings as a traditional cache, it would reduce the storage efficiency by a factor of the number of CPE Caches deployed. It would be prohibitively expensive to do this; for example, a typical traditional cache may have several terabytes of storage capacity, so replicating this into each CPE would have a high cost both in hardware and also in power and cooling. So an approach is needed that can achieve a high storage efficiency without replicating all popular content.

This paper considers two techniques for maximizing both link efficiency and storage efficiency within the customer premises: predictive and participative caching. Participative caching is when a user explicitly flags some content to be watched later, and a user interface tells the user what is available in the cache to be consumed. This approach is similar in concept to a PVR, and it can achieve a very high link efficiency and storage efficiency. The downside of this approach is that the user has to take extra steps and wait to consume content. Predictive caching uses historical data to pre-position content into the cache that is likely to be consumed. The advantages of this approach are that there is no user interaction required, and the user does not have to wait for content to be loaded into the cache before they can consume it. The disadvantages of predictive caching are that the content needs to be predictable, and since the user is not in control, the link efficiency and storage efficiency are lower than with participative caching. A hybrid approach of these techniques can be a good tradeoff between user experience and link efficiency; for example, predictive caching could be used to keep the cache filled with content that is likely to be consumed, and a user could choose to pre-select content they know they will want to consume later.

The considerations discussed above lead to a relationship between storage efficiency, storage capacity, and user participation. Increasing any of these increases cache effectiveness, and decreasing any of these reduces cache effectiveness.

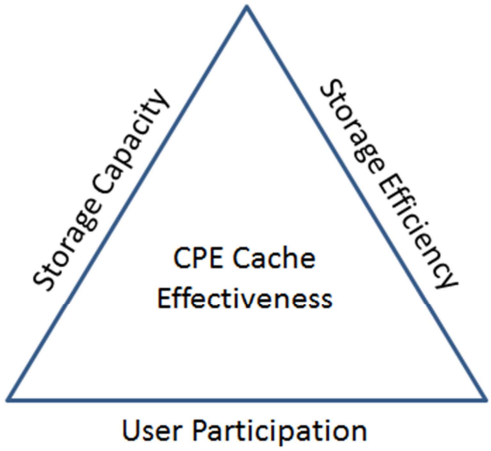


Figure 4. Storage capacity, storage efficiency and user participation.

A traditional network cache improves QoE since the video is delivered from the core network instead of going over transit or peering links to the Internet. A CPE Cache can further improve QoE, since delivery of cached content is not affected by access network congestion.

ARCHITECTURE

The CPE cache system has two major components: the CPE Cache Controller ("Controller") and the CPE Cache. Each CPE Cache is deployed in a customer premises, and is responsible for storing content and serving it to client devices on the LAN. Each CPE Cache is managed by a Controller. The Controller is responsible for providing the CPE Cache with operating instructions such as what content to store and when to download it. The Controller may also provide an interface (API) for other systems such as content providers.

Figure 5 shows how the CPE Cache system fits into the network, and how the parts of the system are interconnected. The client device is a device running on the LAN in the customer premises, such as a laptop, phone, tablet, or TV. The client device communicates with the content provider to authenticate and interact with the content provider's application, and it downloads content from a Content Provider CDN if no CPE Cache is present or if the content is not cached. If the content is cached, the client device downloads content from the CPE Cache. The CPE Cache downloads content from the Content Provider CDN to fill its storage when there is idle network capacity. Each CPE Cache also talks to its Controller. Finally, content providers may communicate with the Controller to pre-position specific content to a CPE Cache.

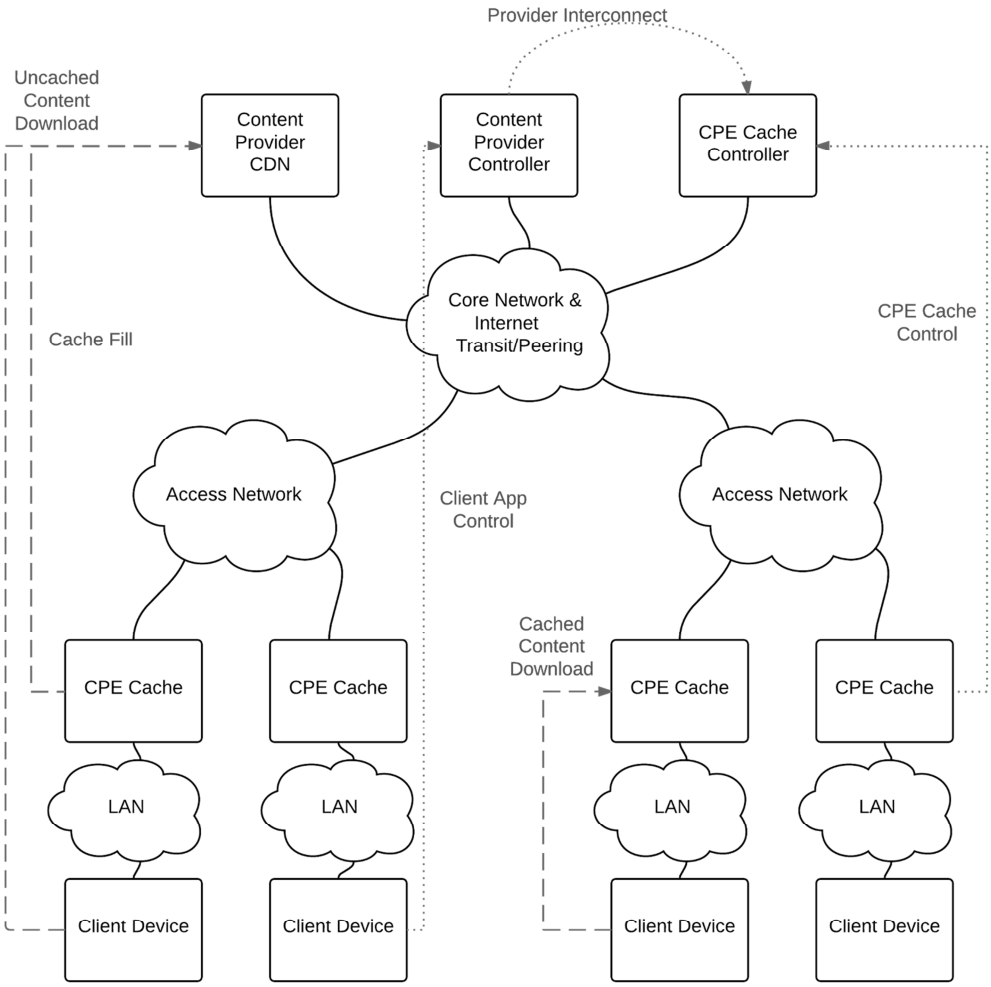


Figure 5. Network Architecture

The following sections discuss several considerations and constraints for the design of the system, where the overall goal of the system is to optimize the peak bitrate of network links on the access network.

Transparent Caching vs Provider Interconnect

Transparent caching works by inserting a transparent proxy into the path of traffic between the client device and the content provider's CDN that serves the content. The transparent proxy analyzes network protocols used to download content, such as HTTP, and may also understand higher-layer protocols such as adaptive streaming protocols that are used to tie several downloads into a single media stream. For example, a transparent cache that works with HTTP and the HTTP Live Streaming (HLS) protocol can intercept the HLS manifest exchange and read the metadata. When chunks of video are subsequently downloaded, the cache can act as the server, and serve the chunks from its storage instead of the Internet.

Provider Interconnect works by the cache system interoperating with the content provider. The Controller exposes an API to the content provider. This API can be queried to expose information about which users have a CPE Cache running in their premises. It exposes a control API for the content provider to push pre-positioning requests that will be delegated to the CPE Cache.

A CPE Cache may use a combination of transparent caching and provider interconnect to achieve the goal of reducing peak bitrate. The key difference between these is that with provider interconnect, the content provider directly participates in the system. Transparent caching can work without any participation from the content provider. Predictive caching, participative caching, or both may be used in either mode. The following table summarizes how content is selected and pre-positioned in each combination.

	Transparent Caching	Provider Interconnect
Predictive Caching	The Controller selects the content to pre-position.	The content provider selects the content to pre-position and programs the Controller.
Partitipative Caching	The Controller is programmed with what files to load by a separate web portal or app.	The user selects the content to pre-position using the content provider's app or web portal. The content provider programs the Controller.

To implement transparent caching, the Controller must be able to select what content to pre-position. This is different from a traditional transparent cache, such as Squid, which saves content as it is consumed and serves the same content on repeated requests. The Controller builds a schedule of content to pre-position and sends this to the CPE Cache. The CPE Cache must be in the path of traffic; for example, it may be implemented as software running on the CPE router, or as a separate device that bridges traffic between the CPE router and an Internet gateway such as a modem. The transparent proxy accepts connections (such as HTTP connections) from a client device, and analyzes the request to see if it is for content that is stored locally. If the content is in cache, it is transparently served from local storage; the client device does not know that it is being served from a cache. If the content is not in cache, the transparent proxy forwards the request to its intended destination and forwards any resulting content back to the client device.

To implement provider interconnect, an API is exposed on the cache controller. The IETF has defined such an API under its Content Delivery Networks Interconnection (Cdni) working group. In the context of this standard, the CPE Cache Controller is the downstream CDN, and the content provider is the upstream CDN. A CPE Cache doing only provider interconnect can be much simpler than the transparent caching case; it needs to have access to storage and it needs to be always on and addressable on the LAN, but there is no requirement for it to be in the direct path of traffic, since the content provider will instruct the client device to connect directly to CPE Cache.

Content Selection

The CPE Cache works by pre-positioning content that has a high probability of being consumed by a downstream client device in the near future. This can be done using participative content selection, predictive content selection, or a combination of the two. The goal of content selection is to maximize the storage efficiency of the CPE Cache, with the minimum amount of user interaction. A perfect content selection algorithm would be able to automatically predict with perfect accuracy what content any downstream client device will consume during the peak hours of the following day, so that the content could be loaded into the cache overnight when the network is idle. This section discusses some possible approaches and heuristics for automatically predicting what content is likely to be consumed in the future.

Storage efficiency can be defined by two factors: content accuracy and format accuracy. *Content accuracy* is a measure of how well the system predicts what content will be consumed in the future. For example, if the system predicts that a user will watch a particular movie from a streaming service, and the user watches that movie, that prediction was successful from the point of view of content accuracy. A specific piece of content may be encoded many times for different devices, screen sizes, and network conditions. *Format accuracy* is a measure of how well the system predicts how content will be consumed. For example, if the system pre-positions every available resolution for a predicted movie into the cache storage, it may have a poor storage efficiency even though the content accuracy was good. A user may only have a device capable of watching standard definition video, so any high-definition video loaded into the cache would have wasted significant storage that could have been used to improve content accuracy by loading a larger selection of content.

The following sections discuss some methods for predicting future videos and other forms of content.

Historical Content Analysis. One way to predict future content is by analyzing historical content. A CPE Cache that is protocol-aware is able to analyze content requests and responses from downstream client devices, and send relevant metadata to the Controller. The Controller can then do an analysis of the total sum of data from the CPE Caches to build a prediction model of content and format for each user.

For example, a popular video streaming service works by sending MP4 files over HTTP. This service is popular for “binge viewing” of TV shows. The service uses a URL format that includes a unique hash for a show as part of the URL. It encodes videos at several different bitrates to adapt to different device capabilities and different network conditions, and each of these encodings is a separate MP4 file. A protocol-aware CPE Cache sends information about the client device, the identifier of the show, and the bitrates played to the Controller every time a client device plays a show. The Controller analyzes this data and for every show, it finds the show that is most frequently played next.

This way, it can find the series order of shows without knowing what the show actually is. It can also analyze the device and bitrate data to predict what video bitrates are likely to be played. With this information, a prediction can be made for what content and formats are likely to be played in the future.

Content Awareness. If the system is content-aware, it can predict future content by comparing historical content to a categorized index of available content. For example, the historical content analysis described in the previous section may be improved on by partnering with the content provider and using their API to obtain information about each TV series, the order of episodes in the series, and the identifier of each episode. Then the Controller can predict the next show without doing an analysis of all historical data.

Premises Awareness. If the system is premises-aware, it is able to collect data about the environment within the customer premises that is relevant to predicting future content consumption. For example, a protocol-aware CPE Cache can analyze the user agent sent in requests it intercepts and determine that there is a device running Apple iOS within the premises. A CPE Cache that optimizes OS updates could use this information by pre-positioning new iOS upgrades to homes that have devices that will consume the upgrade.

Trending Content. The system can use broad trending information across a large population to predict content to pre-position. For example, if the CPE Cache network is deployed across several time zones, and a high percentage of users are consuming a new viral video during peak hours in the leading time zone, the trailing time zones can predict that this will be watched during peak time and pre-position the content in advance. As another example, if new content is being released that is known to be popular with some demographic of the population, and the Controller has this demographic knowledge about its CPE Cache users, it can pre-position this content to selected users in advance.

Deep Learning. The previous paragraphs discuss methods and heuristics for predicting content. A deep learning algorithm may be able to do similar predictions if it is trained with attributes of the historical data seen by the CPE Caches.

Storage

Storage is an important consideration for a CPE Cache. Traditional caches contain expensive, high-capacity storage that is centralized in the core network. A CPE Cache is highly distributed and usually placed in customers' homes, so a variety of different factors must be considered.

Availability and performance are important considerations. Storage must be "always on" and available at any time, and it must be able to serve data at high speeds, since a cache may be serving several concurrent sessions at high bitrates. Storage that is directly connected to the device where the CPE Cache runs will have a better and more consistent performance than network-attached storage.

Another important factor is cost. This includes both the cost of the storage device, and the cost of operating it, such as power and cooling. The CPE Cache system needs to have a good storage efficiency to optimize the storage size required, but there is a base storage capacity needed to serve a typical household. At this time, a 64GB USB drive is a good tradeoff between cost and capacity; these cost about $30 for a high-end USB drive and are powered directly by the USB port, so they draw up to 2 watts of power when fully active, and much less when idle. Since no separate power source is required, this makes them a simple option to install and operate, and in the case of failure they can be replaced without the expense of replacing a device with internal storage.

Reliability is another important factor, since many storage devices will be distributed and a high failure rate would be expensive and inconvenient. The quality of USB drives is extremely variable; the more expensive devices based on SLC or MLC memory are much more reliable and long-lasting than low-end drives based on TLC memory.

Storage Management

Storage on a CPE Cache is a limited and valuable resource. Storage management may need to trade off between the technical goals of maximizing storage efficiency, and commercial goals of different content providers wanting to maximize their content. Therefore one of the roles of the CPE Cache system is to arbitrate the storage space between different content providers, while maximizing storage efficiency.

Privacy and Encryption

One consideration for a CPE Cache is the privacy of the content individuals are watching. In the absence of any cache, a user's viewing history is stored in the systems of the content provider, and not anywhere else without the explicit permission of the user. Media files transferred over the network usually do not contain any information that can be used to map them to specific content metadata such as the title of a show. A CPE Cache system should ensure that this privacy is maintained. Any APIs to content providers should be based on obfuscated data, such as opaque IDs for shows that cannot be mapped back to a title. If a cache is running in transparent mode without a provider interconnect, predictions should be based on analysis of content without knowledge of the content metadata such as titles. If a transparent cache is interacting with a user, for example to do participative caching, then there may be a requirement to present video title information to the user in the form of a web site or app. In this case, the system must ensure that user privacy is secured by keeping user viewing history as isolated and secure as possible, by purging historical data after some period of time, and by not sharing personal data with anyone except for the user.

Much of the content that is consumed over the Internet is copyrighted content owned by some third party. This content is frequently protected by digital rights management (DRM) technology. When a client device wants to play some content, it may perform a negotiation with a DRM server to obtain a key to decrypt the protected content, use this key to play the content. This negotiation may be done over an encrypted channel. A CPE Cache that stores copyrighted content must keep the content stored in its DRM protected form, so that a user cannot play locally stored content without being authorized to play it by the content provider. This is the natural implementation of a CPE Cache, since the files served from the Internet are already DRM protected. An implication of storing DRM protected files is that a CPE Cache does not allow offline viewing of content.

Most content providers encrypt user sessions which carry authentication information, usage information such as viewing history and recommended content, and user control such as the ability to initiate streaming sessions. Most content providers deliver the DRM protected content over plain HTTP, rather than encrypting it inside a HTTPS session. This is because the content being transferred is already encrypted, there is usually a high volume of content which would use significant server CPU to encrypt, and some client devices may not be able to easily handle the added load of decrypting content in addition to playing it. Another reason not to encrypt the content sessions is to allow transparent caching, which has historically been used to improve network efficiency and reduce transit cost for CSPs.

A minority of content providers, most notably YouTube, encrypt their streaming sessions. With continued improvement in CPUs and dedicated encryption hardware, it is possible that more content will be delivered over encrypted sessions in the future.

For a CPE Cache that uses content provider interconnect, session encryption is not an obstacle, since the content provider instructs the client device to connect directly to its local CPE Cache. However, for a transparent cache, there is no standard way to operate on sessions that are encrypted end-to-end. Organizations such as the Streaming Video Alliance are working on developing standards for open caching and content delivery; these may facilitate transparent caching of such sessions in the future.

Transfer of Content

CPE Caching technology works on the basis of downloading content from a content provider to a device in the customer premises for temporary storage. The CPE Cache downloads content during off-peak hours. The Controller provides instructions to the Cache; these instructions include what content to download, when to download it, and how to download it.

One consideration for pre-positioning content is whether it should be pre-positioned using unicast or multicast technology. Regular video on demand always uses unicast technology, because subscribers are rarely watching the same content at the same time. However, by pre-positioning content, the CPE Cache system has the opportunity to take advantage of multicast and broadcast technology to push popular content to multiple CPE Caches on the same access network segment. This approach can be used to make more efficient use of idle network capacity, but it is more complex than a unicast approach, since issues such as retransmission of dropped packets must be handled by the system.

For unicast pre-positioning, the HTTP protocol is normally used to download content to the Cache. The Controller distributes URLs to the cache that can be used to download the content from a CDN.

The CDN hosting content is often geographically distributed, and for best performance the CPE Cache should download content from a nearby server. If Provider Interconnect is used, the content provider will provision a URL using the API; the content provider knows the location of the subscriber and can target the server. In the case of transparent caching, the Controller is responsible for choosing a server; this means it must have knowledge of the location of the CPE Cache and the locations of potential servers hosting the content.

Another approach is to use a combination of CPE Caching and traditional network caching in the network core, organized into a hierarchy of caches. In this case, each caching layer acts as an “upstream CDN” to the layer below it, so the upstream CDN of each CPE Cache is a network cache. This architecture has several benefits. The CPE Cache system does not need to take the location of a server into account; it obtains content directly from the network cache, and the network cache is responsible for obtaining the content from an upstream CDN. This approach makes more efficient use of transit capacity. The network cache will have standard interfaces to the CPE Cache that are agnostic of the content provider; without it, the CPE Cache may need special logic for some content providers. Furthermore, in a targeted deployment where a CPE Cache is only distributed to some sample of users, the overall population can still take advantage of increased efficiency and quality from the network cache.

STUDY

A simulation was done to evaluate the effectiveness of content selection for a transparent CPE Cache. A monitoring probe was installed for a sample of subscribers on a fixed wireless CSP in the US. The probe logged anonymous information about Netflix usage, which was post-processed by a simulated CPE Cache Controller. Netflix was chosen for this study because it represents a large percentage of traffic on most residential broadband networks, usually between 35-50% of peak traffic, and because it represents a good sample of video on demand traffic with a mix of TV shows and movies, so the results from it may be applicable to the more general case of all Internet streaming video.

Figure 6 shows the traffic distribution of Netflix over an average day. The amount of traffic varies significantly over the day, with its peak between 6-10pm. This 6-10pm period is used for peak time data for the remainder of the study.

Figure 7 shows the relationship between the time of day and the average bitrate of an average Netflix stream. This chart shows that average stream bitrates are lowest during the peak time of day; this suggests that the network may be congested during this time. Also, this network experiences lower speeds than the average US network; according to the Netflix ISP Speed Index, the average Netflix stream in the US is 3Mbps but in the study it is under 1.5Mbps. This may be a result of usage policies and/or technology limitations at this particular CSP, but this indicates that the content provider's end users on this network are getting a less than desired quality of experience. A CPE caching system would alleviate the congestion, which would provide a better experience to subscribers during peak hours. It would also improve the overall average bitrate, since speed limitations in the network can be overcome by pre-positioning files before they are played. This would help both the CSP and the content provider ensure subscriber satisfaction, and both of these could use CPE Caching to realize a competitive advantage.

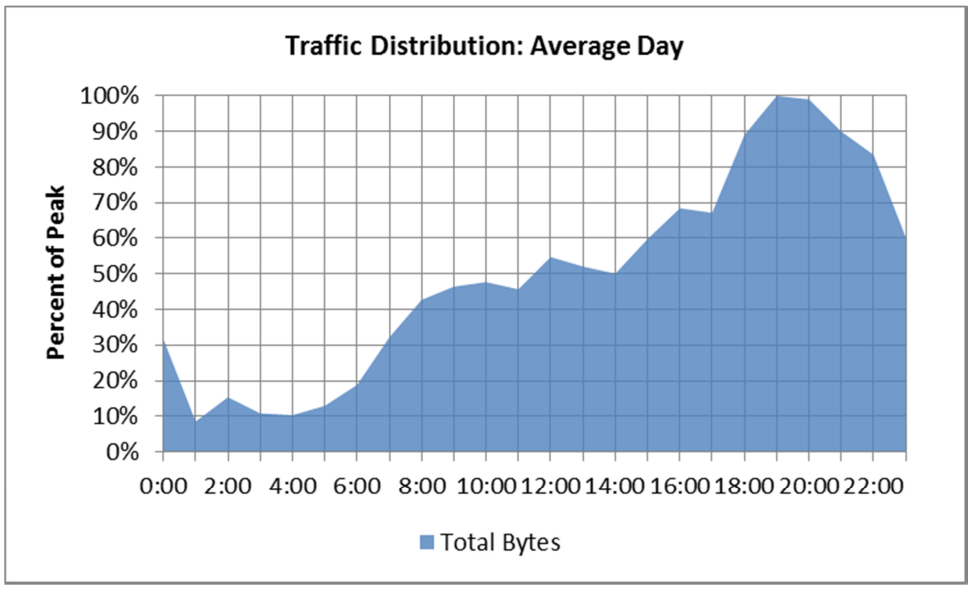


Figure 6. Netflix traffic distribution over an average day of the study

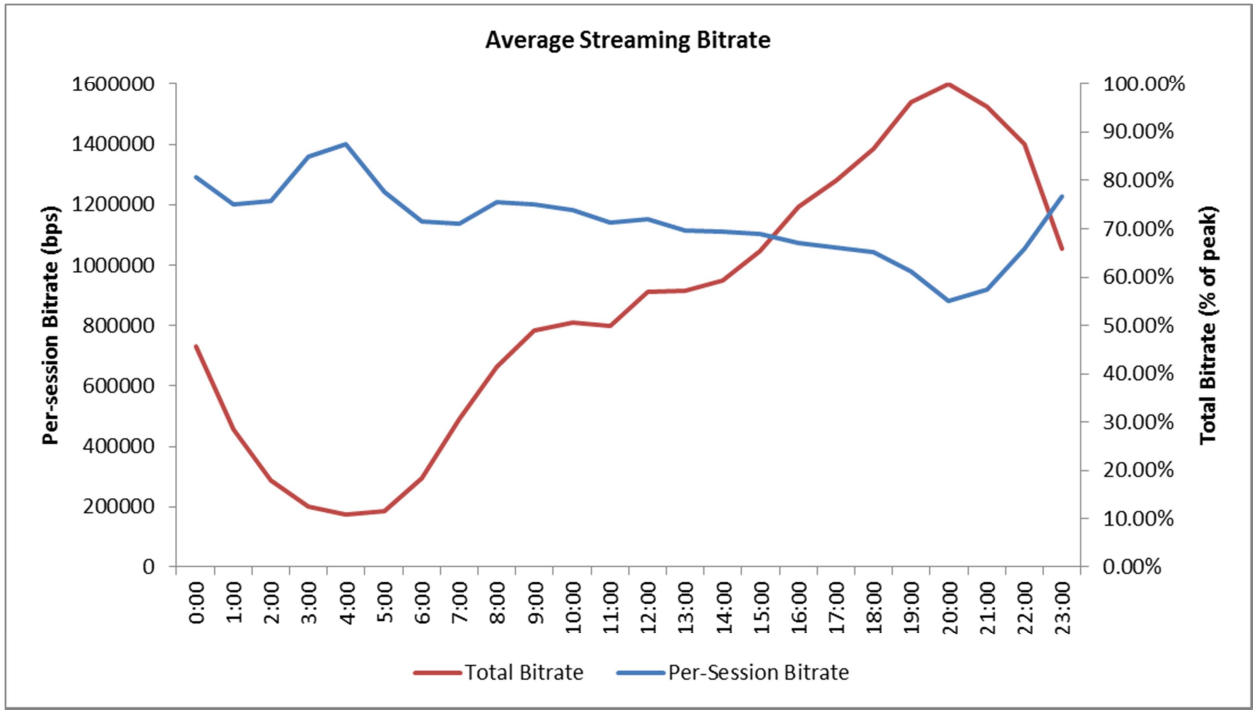


Figure 7. Streaming bitrates over an average day of the study

Figure 8 shows the resulting hit rates during peak time from the simulation, run with four different storage capacity settings. With the content selection algorithms being used, the maximum hit rate achieved is about 47%. The storage capacity has a significant impact on the hit rate until 64GB of capacity; after that, additional storage capacity does not significantly change the cache efficiency. The content selection algorithm used in this study is based on historical content analysis, and uses premises awareness to optimize storage based on the capability of devices in the customer premises. Since this network has a lower bitrate than the US average, this result may not translate well to other networks; a higher storage capacity may be needed to achieve the same savings on a network where higher-bitrate video is the norm.

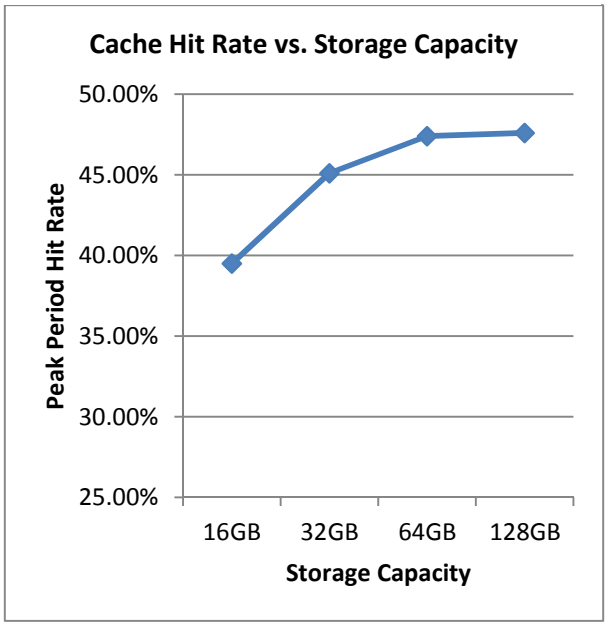


Figure 8. Hit Rate vs. Storage Capacity

Figures 9 and 10 show the results of the simulation for an average day, using a 64GB storage capacity. The peak network utilization by Netflix with CPE caching is 47% lower than the peak network utilization without CPE caching. This study did not examine the effect of downloading content into storage; that would add traffic during the lowest-usage hours of the day, and would make the traffic graph much flatter over the average day.

Since Netflix is about 35% of peak traffic on the average consumer broadband network, implementing CPE Caching for Netflix as shown in this study would save about 16% of peak traffic. According to Sandvine, overall streaming video traffic is 67.53% of peak downstream traffic in North America, so extrapolating the results of this study to all video content implies a potential savings of about 1/3rd of all peak downstream traffic.

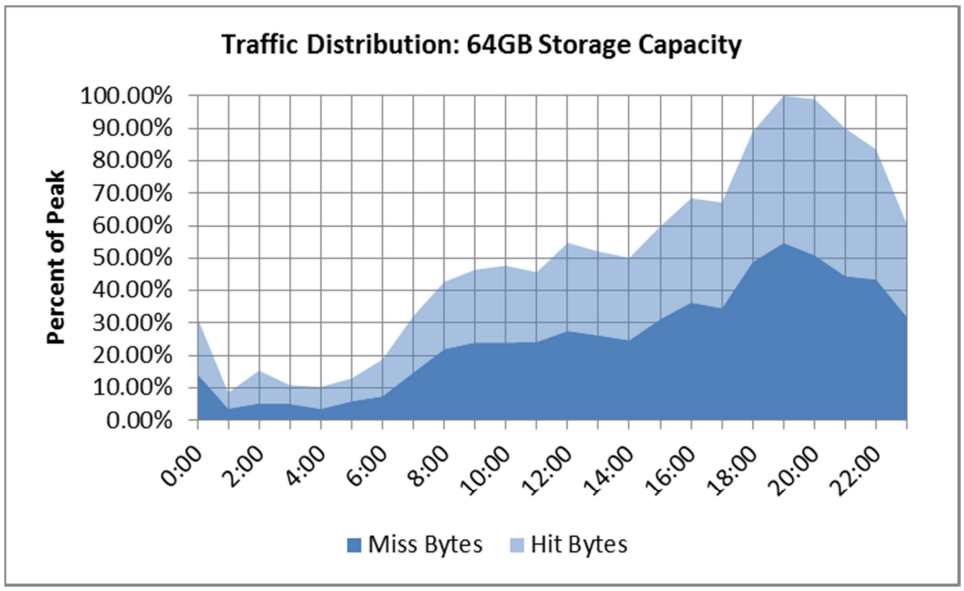


Figure 9. Average Traffic Distribution with 64GB Storage Capacity

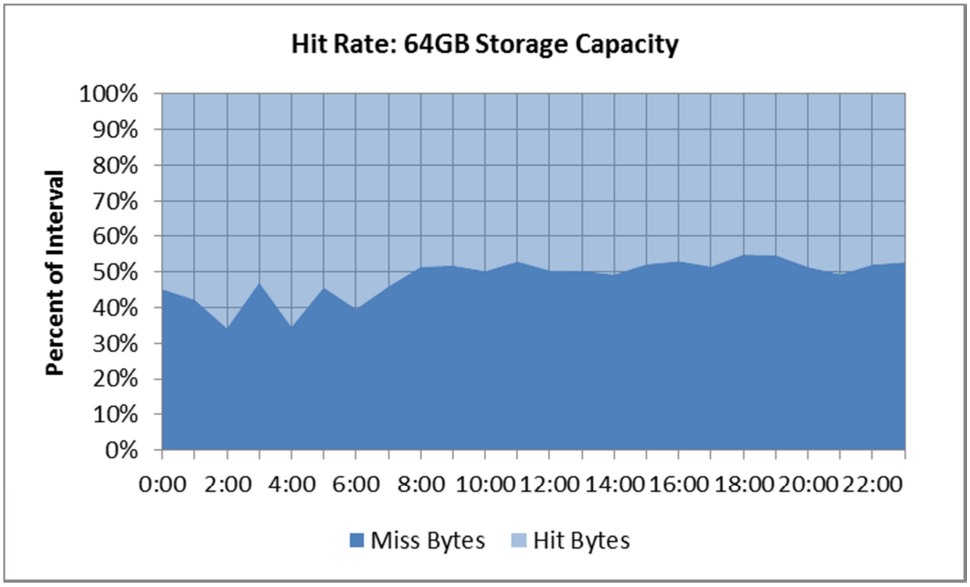


Figure 10. Average Hit Rate with 64GB Storage Capacity

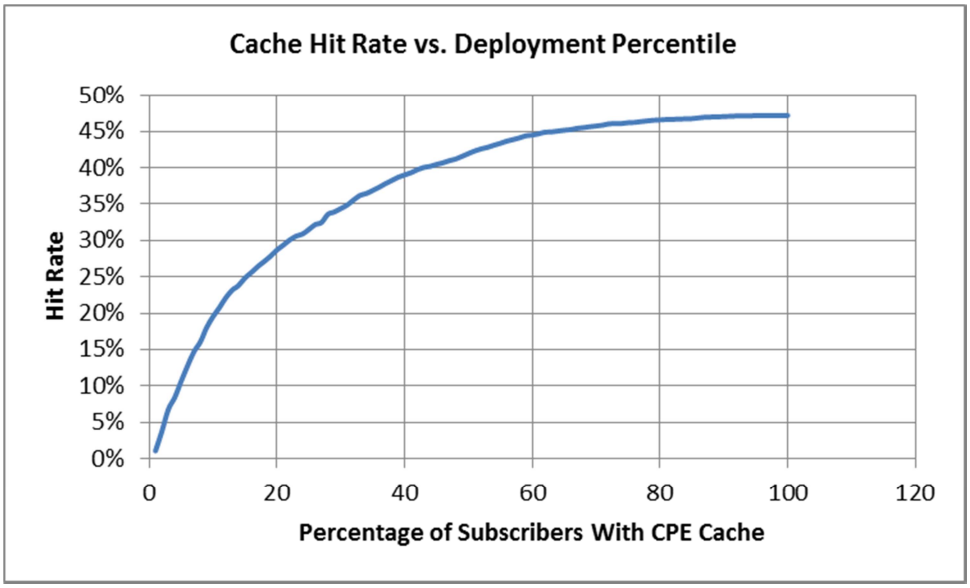


Figure 11. Cache Hit Rate vs. Deployment Percentile

Finally, Figure 11 shows a comparison of the percentage of Netflix users that have a CPE Cache to the benefit overall traffic reduction during peak hours. This shows that about 90% of the benefit can be achieved by deploying a CPE Cache to the top 50% of peak Netflix consuming subscribers. So a targeted deployment can be an economical way to improve network efficiency while minimizing the overall storage capacity across the network.

CONCLUSIONS

The Internet is replacing traditional TV viewing and video bitrates are increasing. Both of these trends are contributing to an exponential increase in peak time Internet traffic, which challenges the economics of providing consumer broadband Internet access. However, as the peak hour usage out-grows the average hour usage, the resulting inefficiency creates an opportunity for optimization using CPE Caching technology.

This paper introduced the key metrics for evaluating a CPE Cache system, and discussed several architectural considerations. A successful CPE Caching system achieves a reduction in peak network bitrate by maximizing storage efficiency while mimimizing required storage capacity and user participation. A study was done evaluating one content selection algorithm that showed a savings of nearly 50% of peak bitrate, for included network traffic. Most of this benefit can be realized by deploying CPE Caches to 50% of the subscriber population.

Future study should focus on what impact the CPE Cache has on the network during off-peak hours; this paper did not examine that aspect of the system.

CPE Caching is a technology that is beneficial to all stakeholders in the Internet video delivery ecosystem. It helps subscribers by providing better quality video even if the network is congested. It helps CSPs by using the network more efficiently, which makes providing consumer broadband access more economical. Finally, it helps content providers by reducing their video delivery costs and improving the quality of their service to their subscribers.

REFERENCES

1. Cisco Visual Networking Index (VNI)

2. Sandvine Global Internet Phenomena Report 2H 2014

3. Application of Policy Based Indexes and Unified Caching for Content Delivery. 2014. Andrey Kisel, Alcatel-Lucent.

4. Netflix ISP Speed Index, February 2015.

CREATING THE VALUE PROPOSITION FOR MSO AND CONSUMER ALIKE IN THE MORE CONNECTED HOME – A GUIDEBOOK TO SUCCESS

Mark Francisco, Comcast Cable
Charles Cheevers, ARRIS

Abstract

The even more connected home is quickly becoming a reality. Much has been discussed about the rise in the number of connected devices in the home to increase from 10 today (on average); to 40 in five years with more than 40 billion connected devices on the planet.

Yet, we have seen the slow progress to mass adoption of home automation, Home Control, home security, and other home connectivity solutions. Why is this stuck and how can it be unstuck to gain a more saturated deployment level of 80% of US homes with one or more new connected home devices?

This paper takes a look at the following key elements in this growing and important area:

- *The current status of technology enablers and how technology can move the needle now to help the overall value proposition*
- *The key technical building blocks required to create an MSO led Internet of Things (IoT) architecture and unify much of the elements of the connected home*
- *The importance of simplicity and curation of the service in this area – and the pros and cons of Do it Yourself (DIY) vs. Managed Connected Home/IoT services*
- *Standards and proprietary solutions – how they intermix and can there be only one or does the MSO have to look at embracing and adding value to the Do it Yourself devices?*
- *The Value and Cost equation with a close look at Effort and Cost vs. Return – why some of the current approaches to getting consumers to spend more of their money on these lifestyle, comfort, or convenience solutions don't work and how to focus on the value returns that do work*

THE INEVITABLE PROGRESSION TO A MORE CONNECTED LIFESTYLE

We are entering the "peak of inflated expectations" for the Internet of Things where there is a huge expectation that we are moving towards a world of more automation of tasks and lifestyle through connected smart devices that can be united together to provide a relevant and important new task that improves costs, efficiency, people's lives or in some cases makes the task completely transparent to the beneficiary.

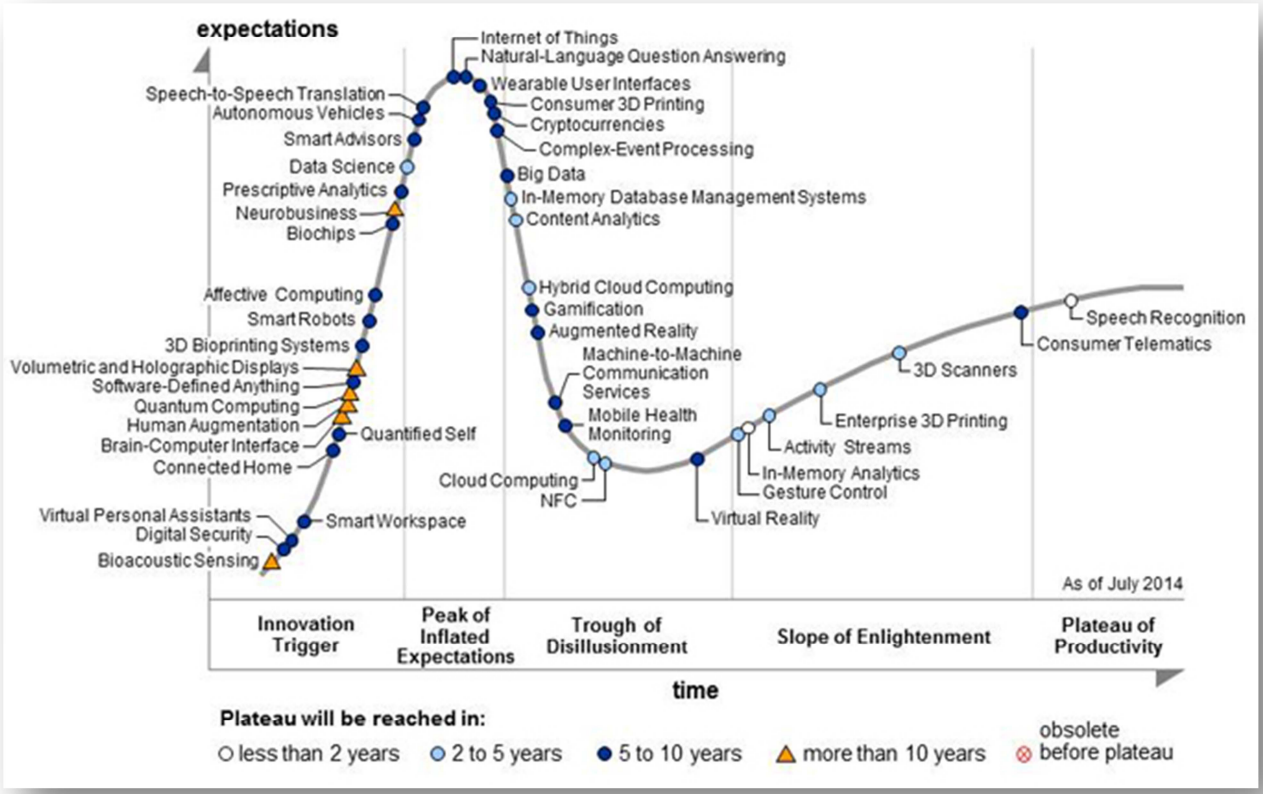


Figure 1 – Gartner's 2014 Hype Curve

We see new connected devices emerge daily ranging from the connected 'fork' to the myriad of wearable devices that record all facets of someone's health and exercise habits.

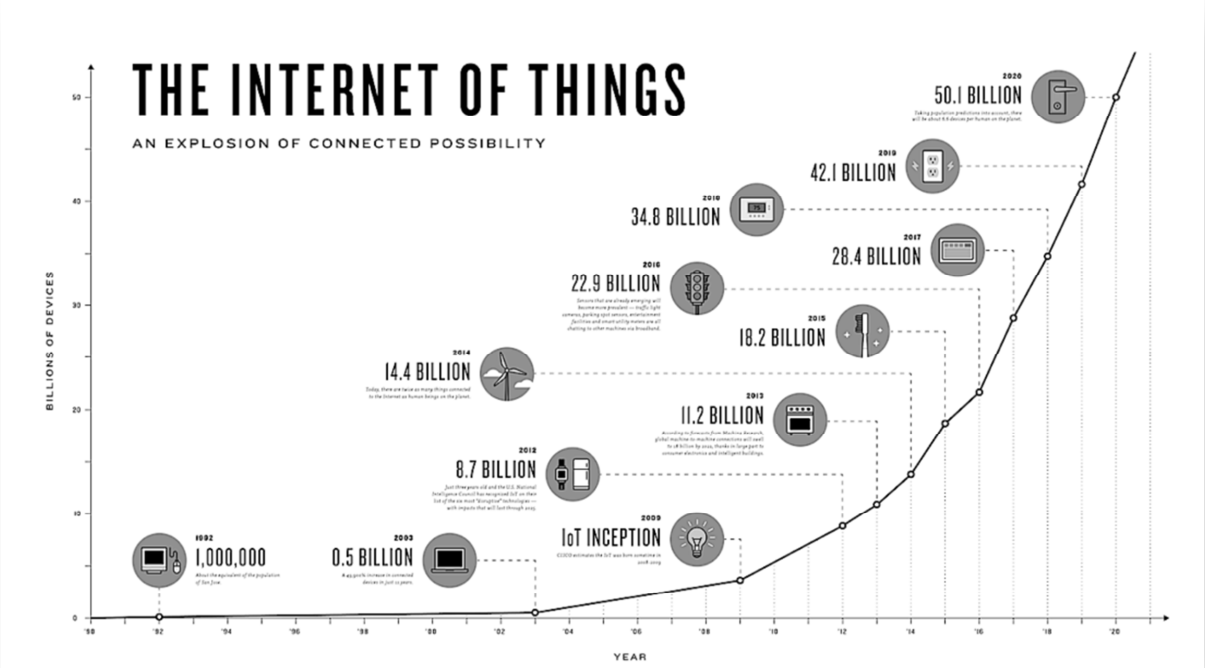


Figure 2 – The projected connected devices curve

There is also critical momentum in a number of areas that shows that the technology is readying to enable the next phase of IoT driven services. These technology pillars include:

- Effective price points at scale for wireless technologies and silicon development with attractive processing power and energy consumption
- Sensor technology and harvesting energy solutions allowing wireless sensors to become more ubiquitous in the home and city environment
- Protocol standards emerging to unite the application space for connected devices
- Cloud-to-Home architectures that can also provide low latency communication for home applications
- Smartphone applications that can drive IoT interaction and provide simple dashboard to monitor or trigger events
- Data analytics and intuitive rules engines for value to consumer, retailer and service provider

There is also a desire in the services sector to leverage the increase in connectivity and connected device to increase their visibility to the consumer or provide Business-to-Consumer (B2C) or Business-to-Business-to-Consumer (B2B2C) services through direct connectivity or Cloud-to-Cloud exchanges.

Typical areas include:

- Energy Management
- Water, Gas and Utility Management
- Medical and Health Management
- Insurance and Wealth Management
- Security and Automation
- Education
- And others

These areas funnel existing spend into the IoT derived services or in some cases generate new sources of income.

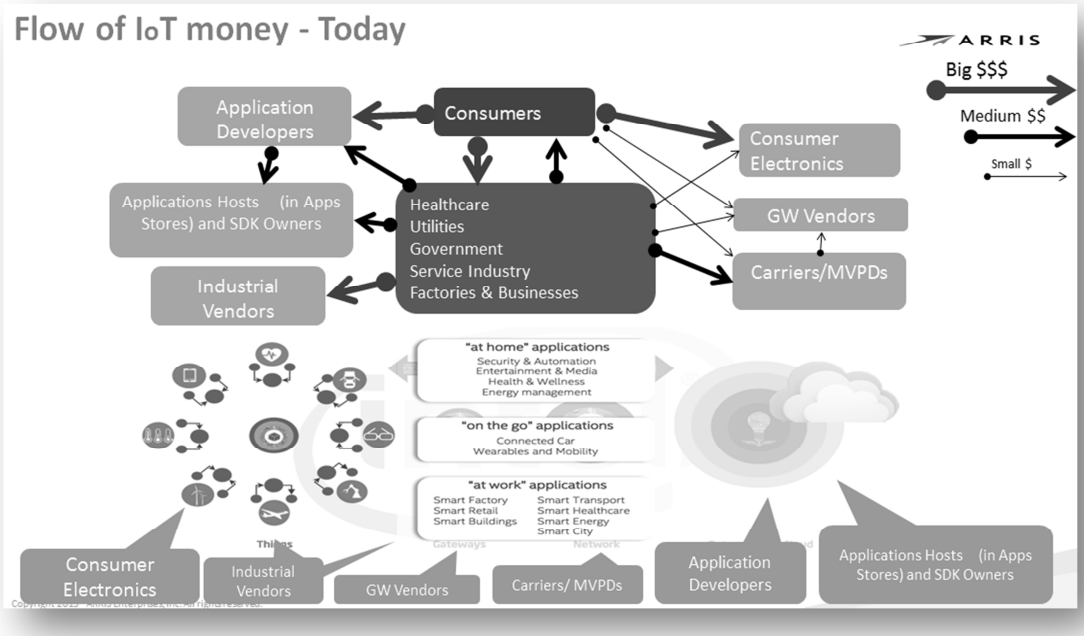


Figure 3 – The flow of money in the IoT ecosystem

There are several macro segments of IoT derived revenue:

- Industrial and Machine-to-Machine
- Smart Cities
- Wearables
- Medical
- And many, many more…

The above segments account for over 66% of the estimated five-year revenue of $5 billion globally. However, the connected home segment is the sweet spot for the MSO and one where the projected revenue in 5 years is about $2 billion of which Smart Energy and Home Security services account for about $1.8 billion of the total.

It's clear that the MSO cannot ignore this trend of connecting more and more devices in the home. Rather than let it just be a burden on the connectivity infrastructure that the MSO has invested in – it seems that there is real opportunity to drive new revenue opportunities from onboarding these new devices and creating value-add elements in the connectivity and service ecosystem.

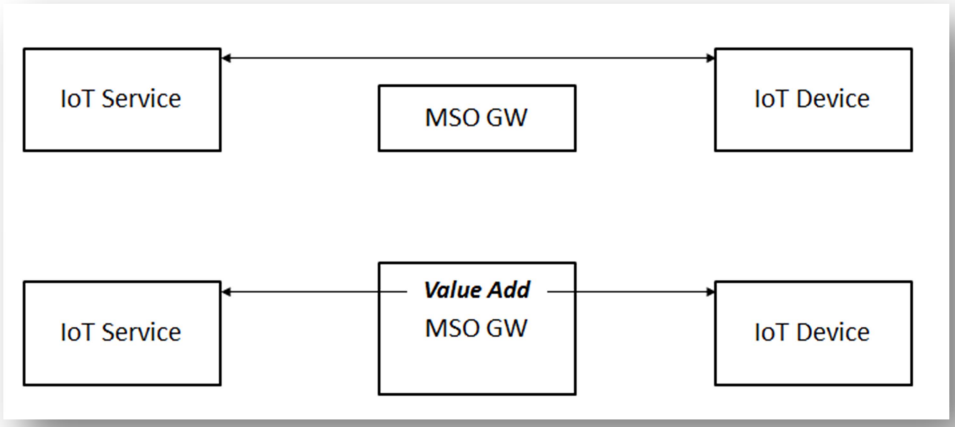


Figure 4 – Creating MSO value-add demarcation to IOT services

This value-add insertion in the IoT service chain is something that the MSO can achieve for many services that could otherwise run Over-the-Top (OTT) from IoT service providers to the MSO's own customer. The remainder of this paper will suggest the path forward for an MSO to evolve a revenue generating position in this growth area of connected things.

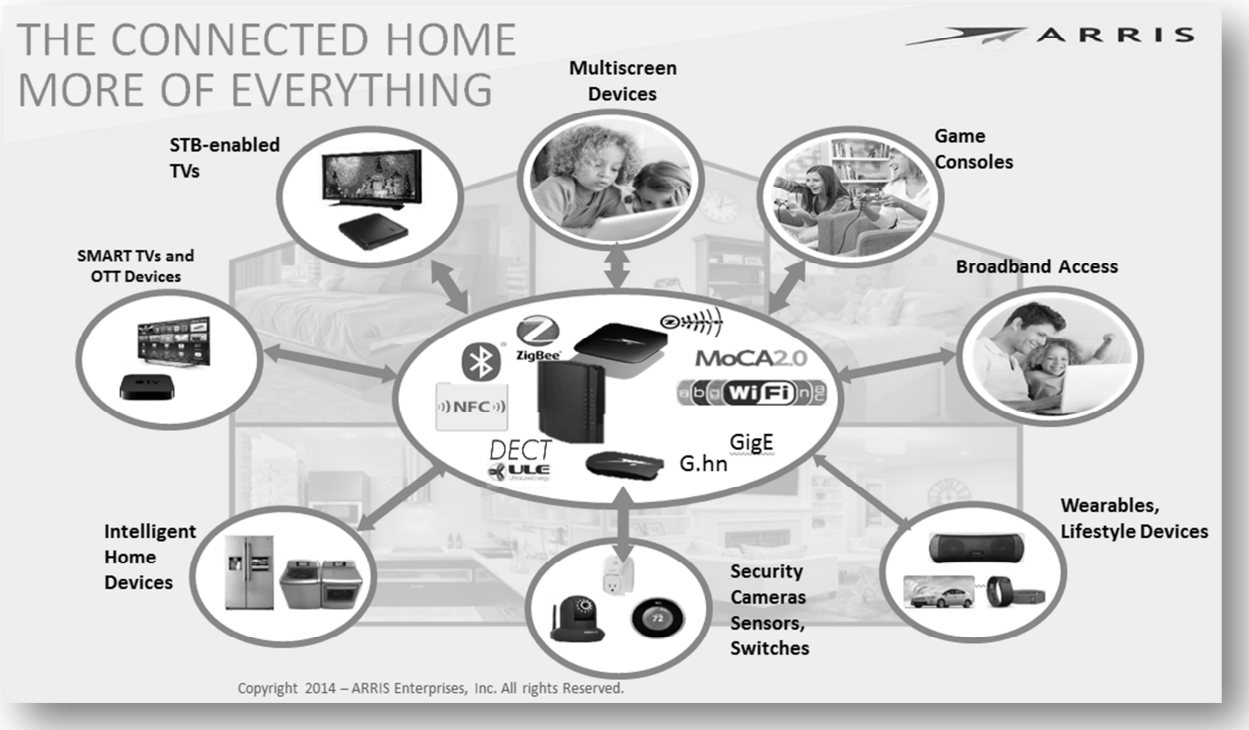


Figure 5 – The Connected Home: Service provider opportunity from their existing devices

The Cable Ecosystem's Role

A service provider may create the Internet to the home connectivity upon which all IoT solutions need to ride. The following set of capabilities that can be leveraged for IoT based services:

- Connectivity infrastructure to Internet
- Technician resources
- Existing devices in the home with wireless connectivity
- Prominent definition of TV user experience
- Existing base and products in market offering home security and automation solutions

The operator is positioned well to extend more into services that touch the end subscriber. In particular, the above set of capabilities really does allow the MSO to enable a faster adoption of connected services.

To baseline potential new solutions, let's take a quick look at the current home security and automation service offered by many service providers.

We typically see the following three services:

1) Home security solution – For example, a comprehensive package of home security solutions that require a typical CAPEX investment of estimated $500-$800 and at a typical monthly payment of up to $39.95 can have an ROI on CAPEX of ~21 months. This type of service requires longer service contracts or requires the customer to purchase some or all of the devices initially.

2) Home automation solution – a smaller package of sensors and switches and cameras that provide home automation services. Typically being offered for a lower subscription cost and a shorter ROI than that of a full security solution.

3) Lower end introductory to (1) or (2) such as a camera solution having a lower subscription cost offering a typical 9 month ROI on CAPEX investment.

Home security and automation services often require professional installation so additional costs of truck roll and technician time and cost increase the ROI by another 2-6 months based on labor and time. Added to this is the network operations center (NOC) and customer support costs causing an even higher time to ROI on CAPEX investment.

The penetration rates of these services have been typically below 5% of the homes served by an MSO. What might be the issue with these service offerings?

It is primarily the customers' view of the cost/value equation of the service. These services, particularly security – have inertia factors from cost-per-month to the hassle factor of installation in the home. These services are typically also not integrated into the MSO's own gateway and TV experience thus they do not leverage any real CAPEX synergies to improve the ROI or capability to lower the overall cost per month.

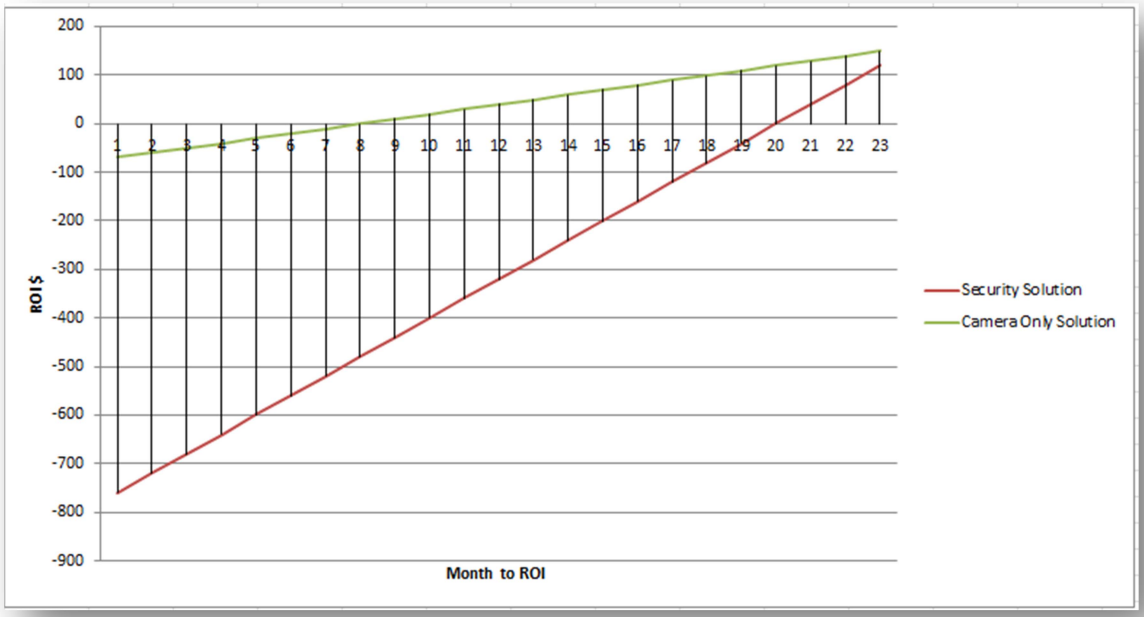


Figure 6 – Home automation/security logjam

How can an MSO improve the potential for revenue for the more traditional connected home services and also begin to also pull in non-traditional revenues in supporting health, energy, utilities, wealth and education services – each of these with large new potential vertical revenue opportunities?

Here are a couple of fundamental stepping stones:

- The inclusion of more IoT radio solutions supporting 802.15.4 and Bluetooth low energy (BLE) as well as Wi-Fi
- Integration of radio solutions in MSO devices such as the gateway (GW), Wi-Fi extender, set-top box (STB) and remote control
- Development of onboarding software solutions to create MSO gateway hub(s)
- Development of the new IoT services that are derivatives of the home automation directions but made simpler and coupled with the voice, video, and data solutions of an MSO



Figure 7 – Service providers have IoT hubs already

Additional to the more traditional connected home revenue opportunities, there is the potential for the MSO to move their connectivity and IoT home presence solution to work with the large vertical opportunities of

- Medicine and Home Health
- Energy
- Utilities – including Water, Gas, and Garbage Disposal etc.
- Health
- Wealth
- Education
- Smart Cities
- Industrial
- Machine-to-Machine

Table 1 and Table 2 below outlines a series of incremental revenue creating capabilities and opportunities that in some cases are unique to each service provider. They leverage absorption of cost of deployment across voice, video, and data and now home automation and IoT services.

<u>The service provider call to action for IoT opportunities</u>

The cost/value tradeoff for the consumer is something that the cable operator can help with by drive providing a value proposition for the consumer.

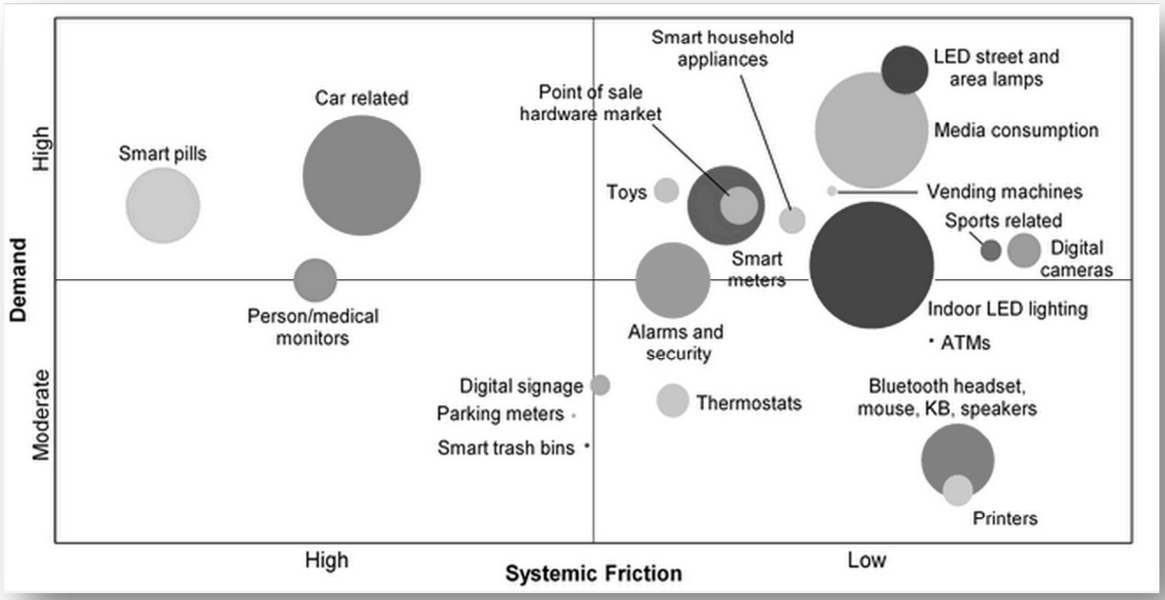


Figure 8 – IoT Service demand and systemic friction

As we can see above, services like alarms and security are relatively low on the friction axis and moderate to high on the demand axis. However, services like Medical Monitors and Smart Pills are highly desirable to automate but are high on the system friction axis – as they require a large change in philosophy, business models, trust, technology, and other factors.

Another component of the cost/value proposition is the absorption of the acquisition costs for any IoT service. These costs can be leveraged by using the MSO installer resources to also support IoT services and can be blended across the MSO's other business of voice, video, and data to leverage the touchpoints that are factored into the operating expenses of its business.

A key element of the MSO capabilities to drive a successful value chain for IoT services is its technician resource base to both support IoT devices and also upsell services when on callout for video, data, and voice installations. This may require potential retooling and reskilling.

Table 1 – MSO key areas of IoT engineering

Potential Action	Description
Reduce the CAPEX investment for home security and automation devices Drive down the ROI time OR decrease the monthly charge to the consumer	Integrate more functionality into gateway, extender, and set-top box for IoT Leverage consumers own devices like tablets, smartphones and TVs more Offer virtualized environments in gateway devices to providers of OTT services
Leverage the technician installs of broadband voice and video services to upsell and install IoT services	Reduce the truck roll costs for connected home services Retrain the technician staff to sell connected home services
Integrate 802.15.4 and BLE into other devices like gateway, extenders, set-top boxes and even remote controls	The service provider can own the radio/PHY level onboarding of the majority of connected devices
Add LTE backup solution to homes requiring security or high availability medical applications	Leverage low usage data rates for LTE call out when required
Offer one set of service provider branded connected home devices – supporting many applications	Provide a service provider branded connected home experience with range of own devices
Support onboarding of BYO devices into the connected home environment	Through supporting the main IoT protocols allow integration of the users own BYOD with service provider solution and the Entertainment system
Develop collaborative value-add applications for integration of service providers and BYOD devices	Create partnerships with different device and solutions to onboard into the service provider curated home experience. Integrating smartphone applications to a single console. Creating IFTTT rules that work across different solutions.
Develop a range of solutions and services that fit into all budgets • CAPEX only investment entry • <$5 per month services • <$20 per month services • <$50 per month • High Value services like Aging-in-Place > $50 per month	Allow subscribers to enter the connected home environment with no monthly fee and CAPEX only services Design a set of applications and services that run without the addition of any specific dedicated device or users own device that range from $1-$5 per month Upsell the consumer to solutions in the $20 range including targeting something like a Security service as a $20 p/m service with reduced CAPEX investment to ROI High Tier multiple services in the $50 range Specific subsidized services like Aging-in-Place commanding $50-$300 per month depending on the level of service

The big adjacent market opportunities from the service providers current baseline

Additional to the traditional connected home derived services above, the following are potentially larger revenue services and the ones that need a deeper look to drive the big step function changes in how a service provider adds new overlay services.

Table 2 – Potential overlay services

Potential Service	Description
Medicare and telemedicine services	Significant dollars are available from insurance companies and government agencies to provide monitoring and future drug administration services Integration of the collector gateway into the service provider gateway as well as using STB/TV instead of tablet devices reduces CAPEX. Adding reminder applications through the TV UX also improve monitoring frequency
Virtual Machine access and APIs to service providers to add application to service provider gateways, extenders and entertainment services	Ability to sell to IoT service providers either client access in home devices or Cloud interfaces
Brokerage for services interfaces for utility and digital services.	Creation of a M2M or B2B2C interface where consumer can broker their utilities to different providers and get lowest and off peak service rates Offer the pool pump 4 hour run to the brokering service and receive the best rate/time from the utility
Analytics and telemetry from the home	Selling telemetry data from the home including supporting Big Data monetizing opportunities
Location and presence triggers and services	For better and more accurate targeted advertising. Can be a source of additional revenue to service provider
Accessibility and voice controlled services	Another specific market to create a more accessible home environment for the blind with voice controlled services
Education services	Leveraging the entertainment, connectivity, and multiscreen control the service provider has to drive a better home education solution including creative task reward interaction with kids the TV and access to the Internet

Integrate 802.15.4 and BLE into other devices like broadband gateway, home network extenders, set-top boxes, and even remote controls

Getting presence of a device in the home is a difficult thing to achieve. Google is moving from its presence in PCs, tablets, and smartphones to other home devices through its acquisitions of NEST, Dropcam, and the recent addition of the Revolv HUB solution.

Apple is advancing its position from its presence in smartphones, desktop, laptop, and tablets. They also supply airport access points, Wi-Fi enabled storage devices and Apple TV. With the addition of HomeKit to IoS, Apple has indicated its strategy to add HomeKit functionality to wireless attached devices and create opportunities for new home control and automation applications which interact with media experiences. One can also see in the Apple Store the rise of partnered IoT and connected devices and solutions.

Consider the general connected home use case for security, in particular where the initial investment of capital items includes a ZigBee Hub Protocol converter, a touch panel, and LTE radio. The potential exists to absorb these radios and functions into products like the:

- Home gateway – already Wi-Fi based yet now add 802.15.4 and BLE support
- Wi-Fi extender – an important emerging additional device in the connection to services. There is an opportunity to add the additional 802.15.4 and BLE radios to the extender device
- Set-top box – perhaps the most powerful addition to the IoT ecosystem because of its presence near consumers as well as having the TV screen to drive integrated IoT based-applications

In particular, the strategy to add 802.15.4 and BLE radios to the gateway, extenders and set-tops – at a small additional cost to those devices to reduce any additional cost of protocol/PHY conversion is key. This allows for the easy onboarding of any service provider IoT devices or consumers own purchased wireless smart devices.

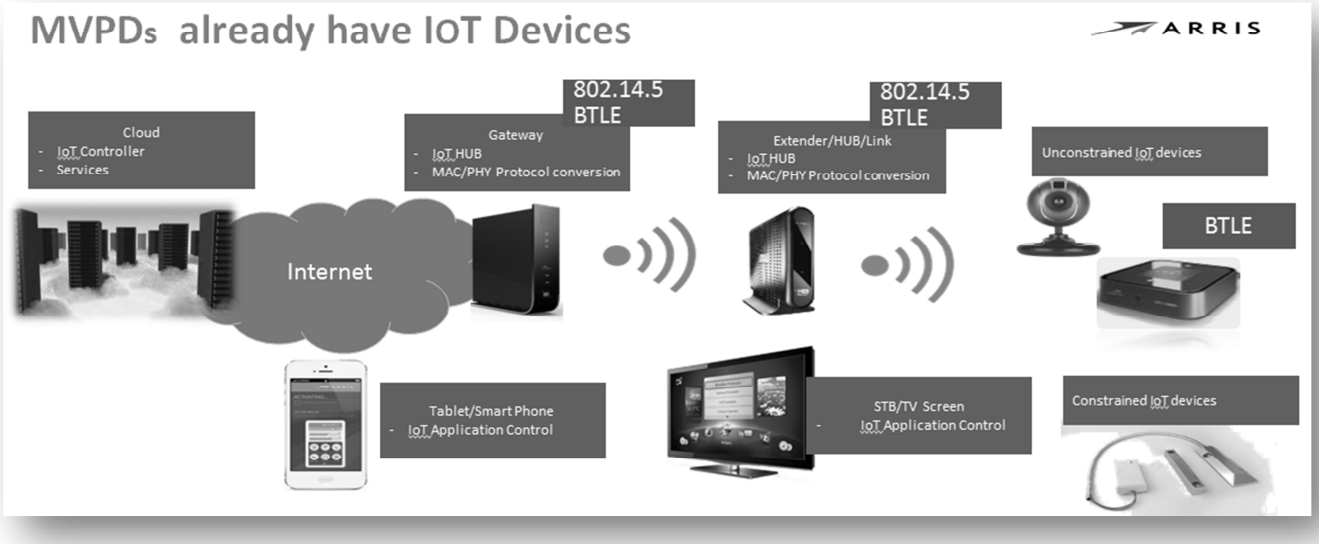


Figure 9 – The service provider's IoT devices

In current proprietary IoT solutions there are those devices that require low power wireless solutions:

- ZigBee and ZWave, for example, are commonly used for low power sensors and inexpensive solutions that run from coin cell battery solutions. They often harvest energy from leakage or kinetic energy
- Bluetooth Low Energy is a popular physical layer of many medical and wearable devices

These solutions typically require the addition of a Hub/Link or protocol convertor from 802.15.4/BLE to Wi-Fi and from ZigBee to IP. This hub is both an additional cost item and an additional device in the home. It takes the data path typically in an over-the-top trajectory. The integration of this device into existing MSO devices is a key part of the MSO strength in the IoT ecosystem and home.

Why 802.15.4

802.15.4 supports the physical layer of ZigBee based devices and aligns with the PHY layer selection of the Thread group of companies in the IoT space. It is likely to be used for low power and more constrained IoT devices, but also emerging with Thread promoting 6LoWPAN direct IP connectivity.

The ZigBee group is also advancing solutions for ZigBee IP and 920IP as well as a 6LoWPAN so another option to consider.

Why BLE

Almost every wearable and hold-able device supports a Bluetooth interface. BLE and BT4.2 are the right interception points to allow the onboarding and presence detection of devices. It provides a good low power solution and offers an opportunity to the service provider to also combine the use of Bluetooth with remote control and set-tops. This allows the cost of BLE addition to the set-top to be amortized over both remote functionality and IoT connectivity to wearables, and also to afford location and presence detection to applications for advertising and other security solutions.

In particular, adding the BLE support into a set-top makes it a more likely solution to connect to BLE devices and wearables when consumers are in the main rooms of their homes. Those rooms typically have a set-top box present. Additionally, applications that allow interaction between the consumer's BLE IoT devices and the TV screen can be developed.

One further big BLE/BT4.2 opportunity is in the Medicare and Wellness area. Typically, most of the monitoring devices connect via Bluetooth to send readings and there is a large opportunity to support Medicare solutions just by adding an IP set-top with BLE to the home.

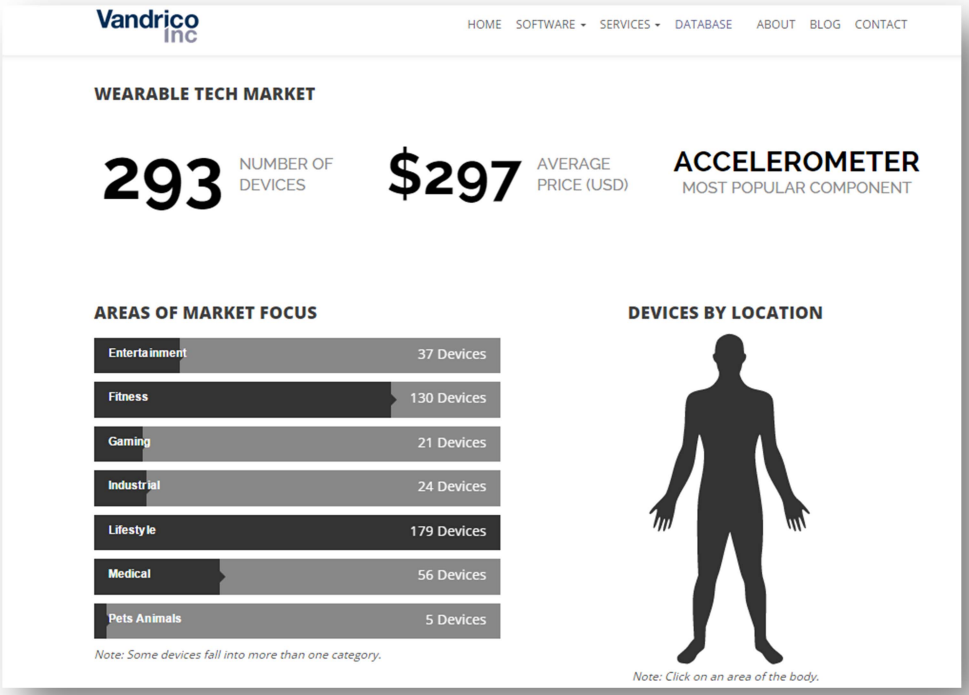


Figure 10 – Almost 300 wearable devices already exist

Add LTE backup solutions to homes requiring security or high available medical applications

While the reliability of the cable network is high, a backup path is required to provide high availability solutions like security and some medical applications.

A 3G/LTE backup has typically been a separate device integrated into either the control panel for security or in medical applications a device like the Qualcomm 2Net device that provides dedicated connection to network side database.

The pricing models for SIM based devices are changing to allow very cost effective 'rarely used' data connections and in many applications the LTE radio can be powered off until needed not requiring additional power overhead in steady state operation.

Leverage the technician installs of broadband voice and video services to upsell and install IoT services

This is a key aspect of a service provider's ability to accelerate adoption of connected home and other IoT services.

One truck roll and $150 investment in technician time could be offset by the following re-engineering of the technician install process.

For example, if a technician is installing a new gateway or set-top, the truck roll cost may also be amortized over any additional connected home installs or upsells.

As another example, if the service provider also partners with water, gas or electric companies, either company can also install metering devices as part of the one visit install.

Even if an MSO doesn't initially partner with any utility companies – the MSO can make an investment in installing devices such as gas leak detection, water flow meters, HVAC systems, and even sprinkler or pool pump systems as part of a larger brokering of service or selling of analytics or telemetry data.

The consumer would have to approve the installation of these devices and opt-in on the sending of information to utilities or insurance company – something that can be managed to high opt-in rates from consumers.

The amortization of multiple activities in the home by the field technician would potentially:

- Decrease the present cost of IoT services install and OPEX considerably
- Incur initial capital outlay for devices like water flow monitors that could be offset with substantial potential monthly fees from insurance companies for leak damage mitigation
- Incur initial capital outlay for items like sprinklers, pumps, HVAC controllers and energy management devices, however brokerage fees and utility pass through of revenue sharing to the service provider would be possible and potentially profitable

The process and amortization changes could enable the viability of a cable operator based IoT solution and its associated services.

<u>Offer one set of service provider branded connected home devices – supporting many applications</u>

It is important for each service provider to create and provide a set of appropriate devices that identify that service with the service provider offering it. In the case of current home automation and security solutions, partnering with home security and automation companies – and leveraging their software solutions – allows for an MSO to enter the market and deploy the solution on its own devices. However, service provider branding may ensure that the consumer identifies more with the service provider than the security company. The other devices that link into the controlling device or hub are many and varied. Typically the current focus devices include

- Thermostats
- Cameras
- Contact and Motion Sensors
- Lights and Light Switches

Future potential additional devices and services include additional devices on the list below.

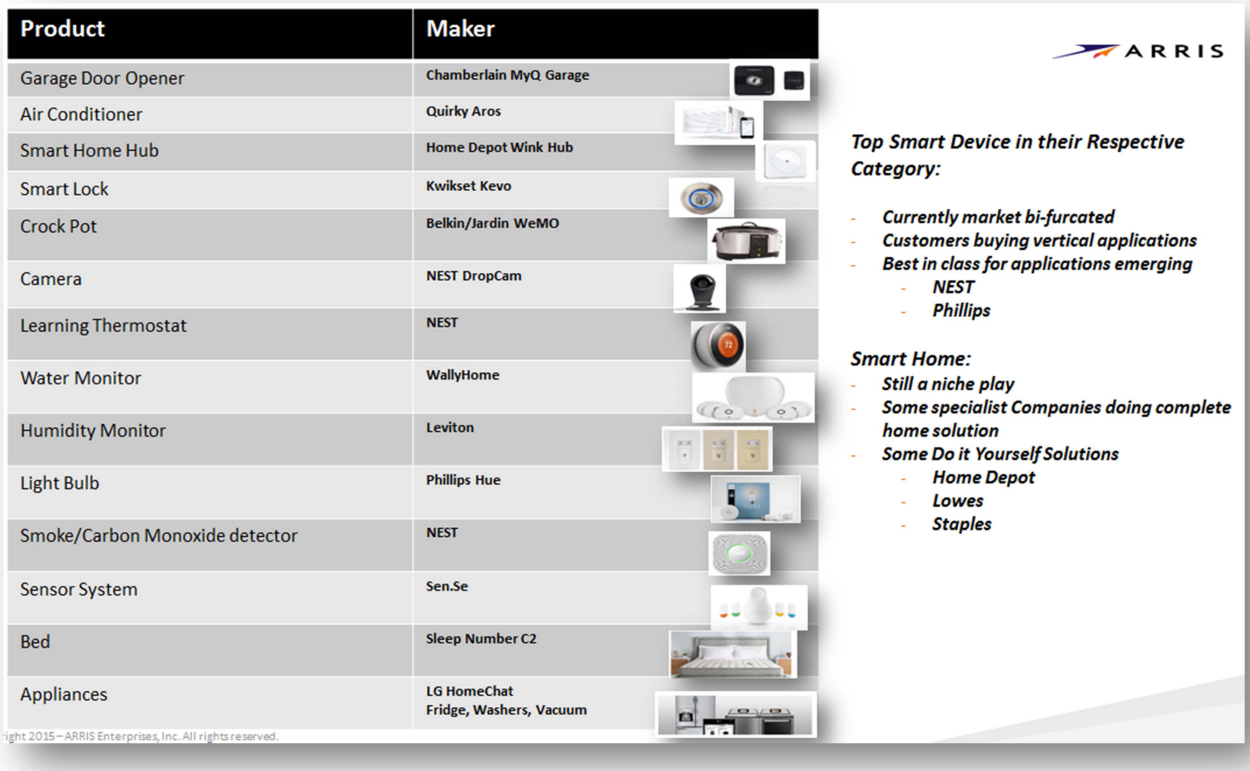


Figure 11 – The top IoT device and service opportunity

Support onboarding of third party IoT devices

There are literally a 100 different solutions for IoT applications – and all come with proprietary protocols. In some cases, they converge on ZigBee or Zwave for non-IP devices but generally, companies like Belkin/WeMO, Insteon, PeQ, Phillips HUE, Withings, Google NEST, Quickset, Chamberlain, Kwikset and Apple/HomeKit all have their own solutions built around many different protocols. This has caused fragmentation in the application space for IoT and has created user experience issues where the consumer needs to use multiple applications and interfaces to access their home automation, control or IoT solutions.

There is an opportunity to make a "Home" experience rather than a WeMo or Google experience and aggregate the different solutions to one interface – an MSO application on smartphones, tablets, PCs, and the set-top box for a user experience which is part of the complete MSO TV experience.

To do this requires the support of more than one framework and/or set of protocols. The logical choice is to:

- Adopt one protocol for each service providers own solution
- Support the protocols moving the herd (market) and closest to standardization or major adoption
- Support protocols which are open standards or on a path to OpenSource
- On a solution by solution basis, if the business analysis dictates, add support for other solutions
- For hardware enabled solutions like Apple's HomeKit, make decisions on the inclusion of the security module in some or all of the onboarding solutions
- Develop strong relationships with the Top 10 BYOD (Bring your own device) and BYOS (Bring your own service) providers

- Develop a value-added proposition for third party solution providers to integrate their solution into a singular home experience. Examples include:
  - Smart Oven – notifications on TV that something is burning
  - Smart Garage Door open – notifications on TV that the garage door is open
  - Integrate audio and voice control into IoT devices
  - Integrate into the MSO Entertainment Platform for integration with video, voice, and data services

The primary IoT protocols that fit the criteria above are:

- Open Internet Connect
- Alljoyn/Allseen Alliance
- Thread
- ZigBee
- UPnP

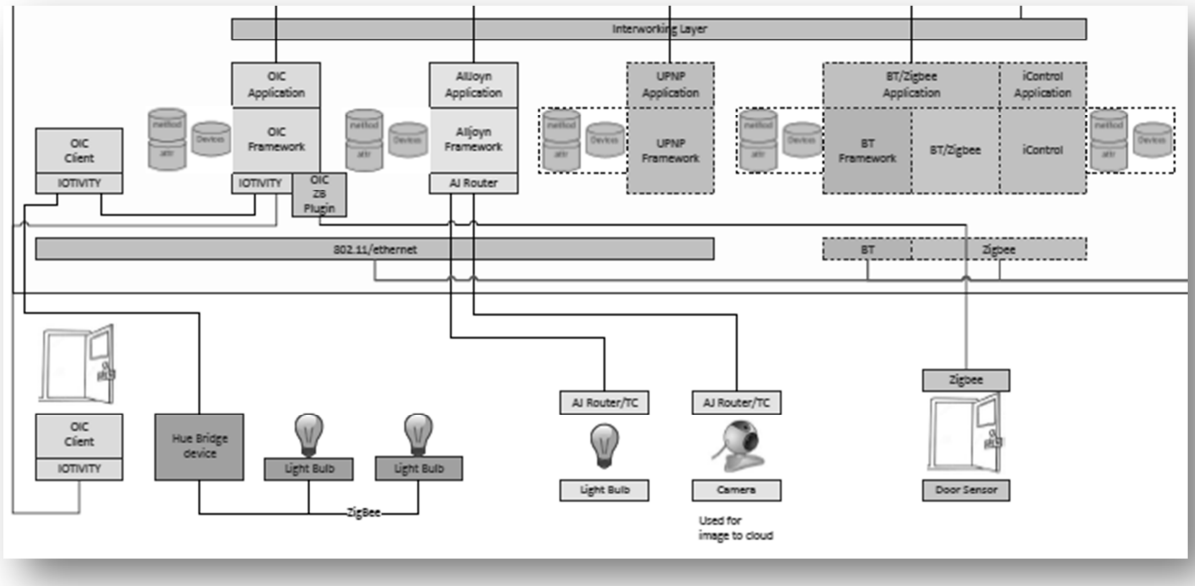


Figure 12 – Multi IoT protocol stack for an MSO gateway

The Business Decision inputs for this multi-protocol support include:

- Can the service provider afford not to provide this feature? If the value-add to the service provider solution is to aggregate fragmented IoT solutions then targeting the high runners for inclusion in a single MSO controlled solution is important
- What is the cost of supporting many different protocols?
  - Each solution has to be weighed on its popularity, opportunity and ROI
  - Additional code overhead has to be reviewed for RAM and Flash implications on devices
  - Supporting the main OpenSource standards solution seems a prudent decision
  - Partnering with the high runner proprietary IoT devices and service providers on a mutually beneficial basis
  - Additional software and testing resources for multiple combinations of protocol and device may be prohibitive

- How are these protocols maintained and integrated into the gateway, extender, and set-top box?
  - One proposal could be to develop a solution that works with the RDK-B architecture. This could be an abstracted and containerized architecture that messages to the RDK-B framework and keeps the IoT modules – modular and protected from the connectivity operations of the gateway
  - Integrate OIC, AllJoyn and Thread and others into the runtime as needed
  - Develop a common data model, since each has very different protocols in their implementation and coverage,
  - Leverage templates from UPnP and the machine-to-machine IoT standardization work
  - Dynamically load as the devices are sensed in the home

<u>Develop collaborative value-added applications for the integration of MSO and BYOD devices</u>

As defined in the preceding section, one of the key strategies for a service provider is to ensure that from a physical layer all major IoT devices and applications can make their way into the ecosystem of the service provider.

To add value to the third party solution, the protocol of the solution must be:

- Part of the supported protocols in the stack – examples OIC, Alljoyn, and Thread
- Partnered disclosure with the solution provider

The opportunities for increased revenue and application opportunity are to:

- Aggregate a single dashboard experience on the single service provider's application. In the case of MSO, the TV UX application becomes the single interface point to both MSO and Third party devices. For the MSO provided set-top boxes, the UX based dashboard and interface to the home experience is offered
- Charging subscribers for an aggregated experience of their third party and MSO-based solutions can be done through the purchase of an application. For example, a nominal monthly fee could be charged for using each MSO provided IoT portal application

Examples of such applications are could include:

- Smart Garage Door opener – The added value is having the event from garage door also sent into the MSO Home Framework and being able to display notifications on the standard TV or TV UX. In addition, if the MSO has a voice services on their TV platform, those services could also be leveraged to close the garage door from the smartphone or a remote control application
- Adding all the IoT service and devices into a singular dashboard where the devices can be viewed for connectivity health without having to fire up multiple smartphone applications.
- Being able to monitor and control the use of security cameras from third parties to help address some of the privacy concerns consumers have about the use of cameras and sensors in the home

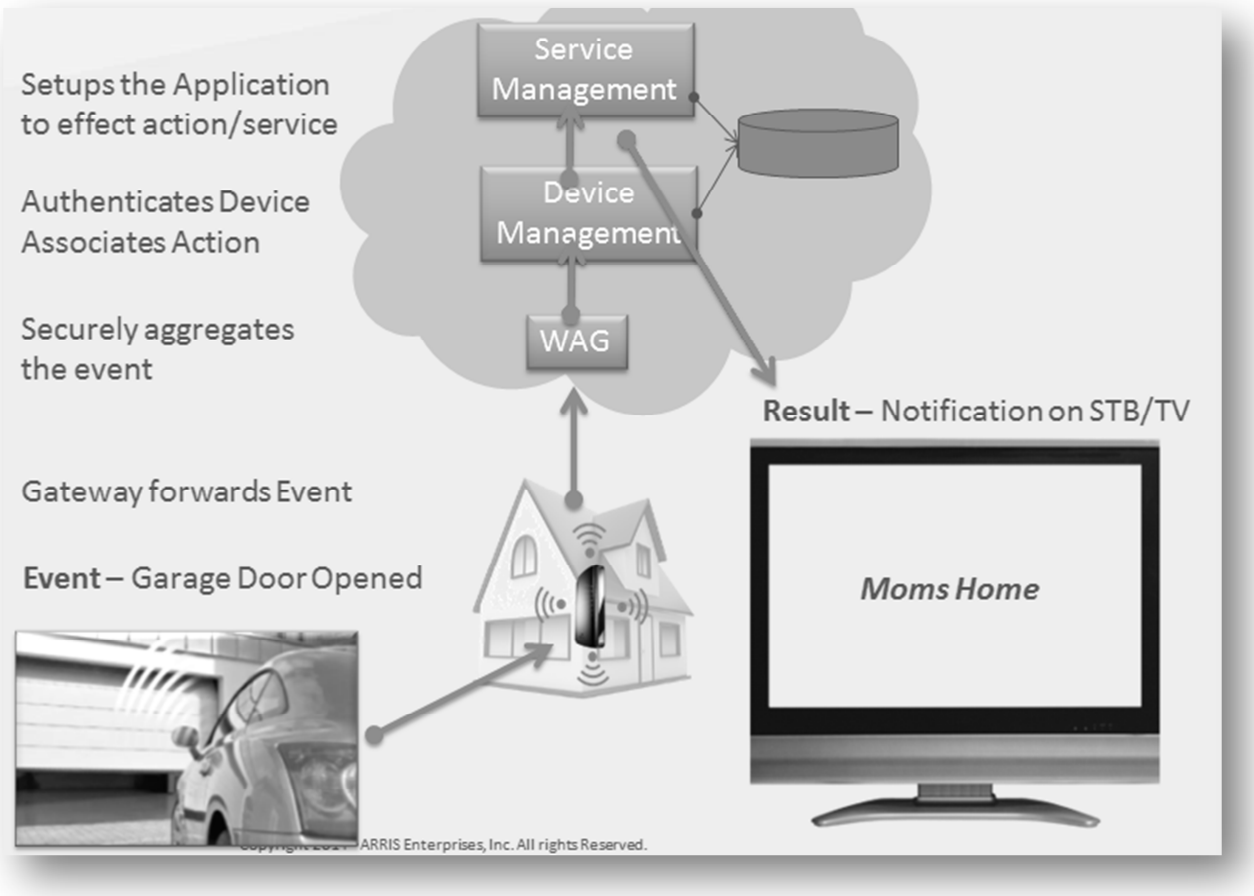


Figure 13 – Onboarding a notification service from third party application using Cloud Services Architecture

Develop a range of solutions and services that fit into all budgets

A key change in the current home automation and home security solutions of the service provider is to create something for everyone's budget in the connected home and IoT space.

Current MSO home security and automation solutions approximate range per month:

- $9.95 per month for camera IoT solutions
- Home automation and control solutions at $19.95 per month with a limited number of IoT devices and sensors
- Entry level security solution at $19.95; a home security solution with provided devices
- Deluxe security solution for $39.95; a home security solution with additional devices

As we have mentioned earlier – these solutions currently have:

- Low penetration with customer base
- ROI ranging from 8 to 20 months
- Requirements for 2+ year minimum contracts
- Have technician installation costs

There is opportunity to improve either ROI or base monthly cost by leveraging the integration of functionality into other devices like gateways, extenders, and set-top boxes.

Additionally, there are some new models that could allow everyone to dip their toes

into the IoT, connected home and security services. These include:

- Zero $ per monthly fee onboarding solutions
  - CAPEX only purchase of IoT device
  - Free giveaways to customers of IoT Switches Sensors to tease them into additional IoT services
- $1-$5 applications that are pay once with no monthly fee
  - Modesty protection for camera; enabling this feature on the MSO home security/automation application
- $1-$5 applications – per month
  - Onboarding your BYOD into the MSO home security/automation application
  - Enabling voice control of IOT devices
  - Elderly panic button based on RF4CE or BLE remote control
  - Adding security cameras including BYO cameras to the TV dashboard
  - $1-$5 per month for each additional IoT devices to the initial bundled packages
- $5-$30 per month applications
  - $30 per month for the MSO security package and "all you can eat" application services from lower tiers
  - $15 per month for a home automation package
  - Second home/vacation home packages
  - Direct to water utility sprinkler solutions leveraging IFTT based weather and sensor information to optimize to a fixed monthly fee for water use and lawn perfection
- Higher end per month fee bundles for MDU and SMB environments

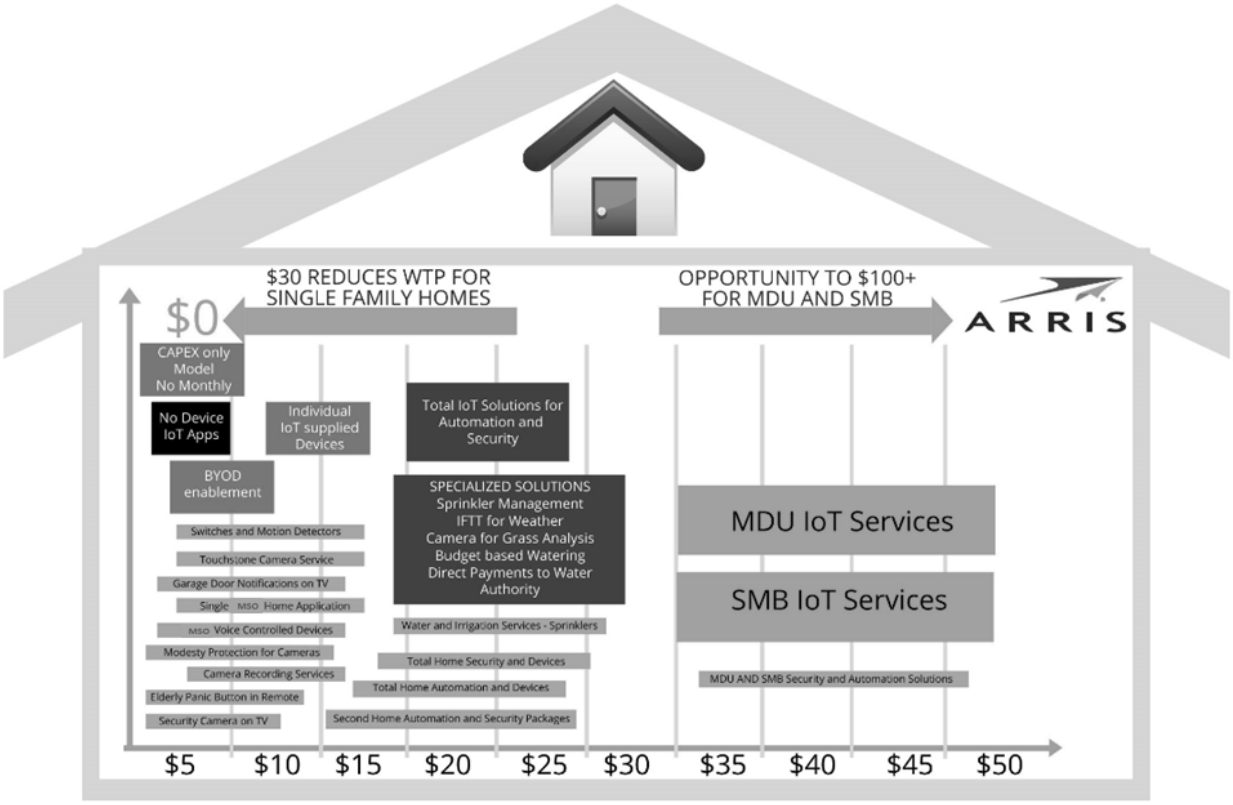


Figure 14 – Potential range of IoT services for everyone



Figure 15 – Simple monthly panic solution using RF4CE and BLE remote

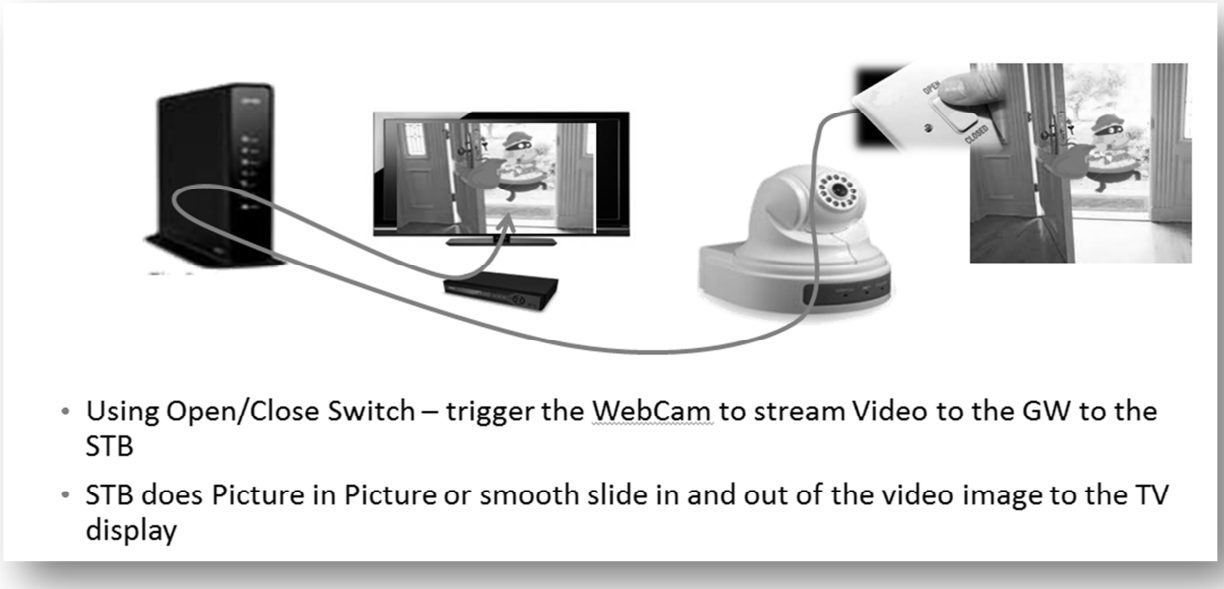


Figure 16 – The better integration of a third party camera

MEDICARE AND TELEMEDICINE SOLUTIONS

One of the most exciting opportunities for the service provider is the Telemedicine, Medicare, and Aging-in-Place business opportunity.

Aging-in-Place services focus on monitoring the elderly and ensuring that clear communications and connectivity are provided to caregivers.

Typically most applications require monitoring and use specific hub devices to connect to Bluetooth-based health monitoring devices from weighing scales to blood pulse-oxygen monitors. There is now even progress in monitoring the drug administration itself from sensors that detect medicine packaging access activity to devices that are worn by the patient and are connected to the Pharmaceutical or Caregiver drug administration system.

There are service provider opportunities to reduce the capital expenditure on the Medicare equipment by using gateways, extenders, and set-top boxes with BLE to connect to the wireless Medicare device. This could effectively create a solution where the patient is checked to see if they are a MSO subscriber with a 'Medicare Ready' home package. They are then sent home with just the monitoring devices – and not the hub device.



Figure 17 – Service provider Medicare solution

A service provider could create the best solution to gather the daily readings by incorporating the service providers TV platform into the solution. This would also allow for training about the medication and devices to be given through the TV platform.

An MSO TV platform typically has many accessibility features to aid any disabled patients. Features might include syncing the patient's monitoring schedule to the TV system; the TV could prompt the patient to take their readings. Strong stick (and even some carrot opportunities) could be created by pausing TV content to force the patient to take a reading. Additionally, blood sugar checks and insulin reminders could be given on the TV, again with a forced pause until the patient completes the requested action item (reminder).

Aging-in-Place communities could be serviced by a single MSO deployment, because of the integration of some of the Medicare solutions. The revenue opportunity for the MSO would be in addition to the already existing voice, video, and data services they might be providing that community, or end up providing that community by winning an Aging-in-Place service bid.

Virtual Machine (VIM) access and APIs add applications to service provider gateways, extenders, and entertainment services

The opportunity is to offer third party companies the ability to have access to the customer's device(s) as part of an overlay service that the third party offers. Services could be potentially hosted in the Cloud with only one service provider interface in the home gateway since VM architecture in the gateway potentially offers the scope for a more hands-off pass-through approach to some specific services.

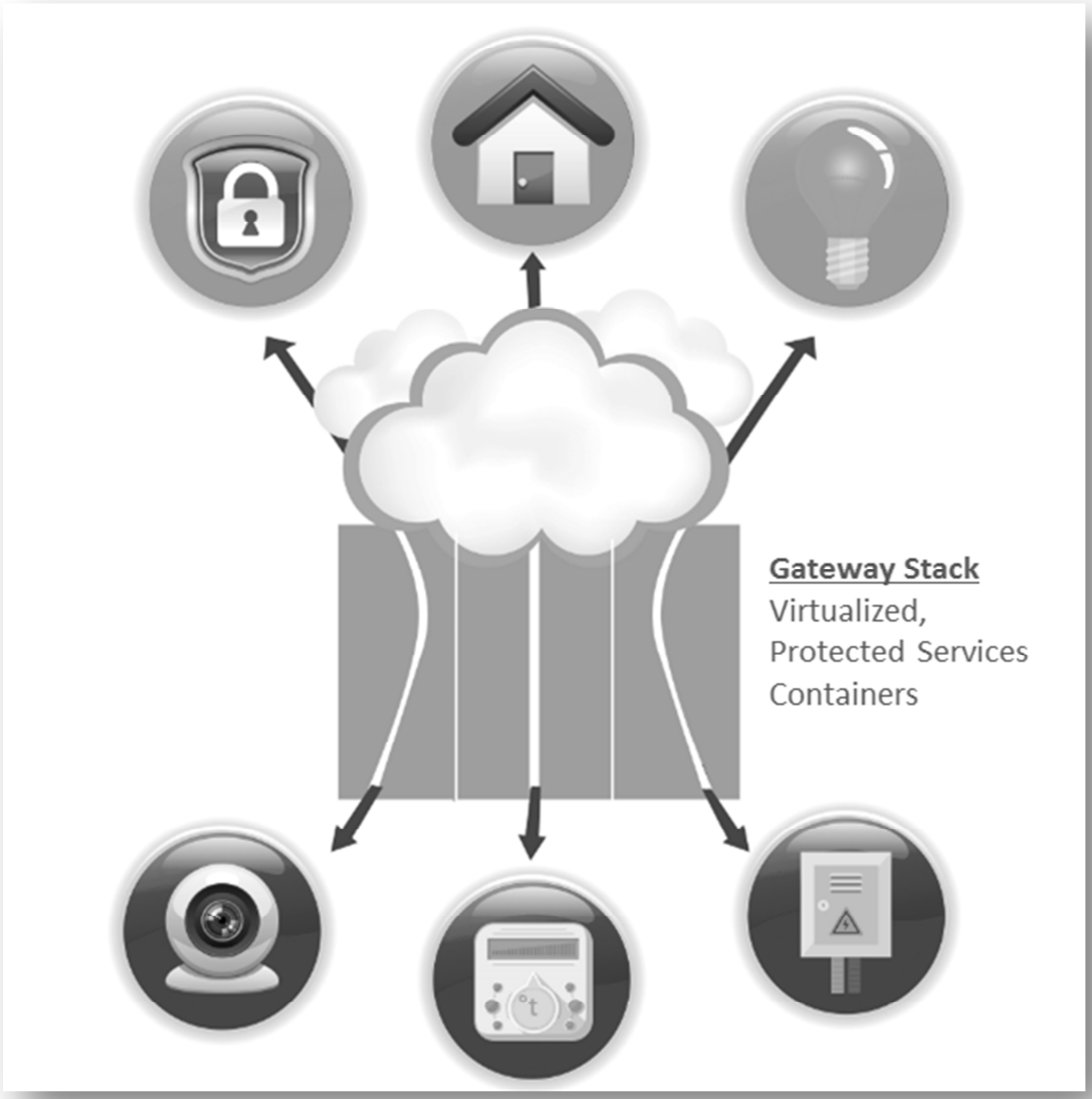


Figure 18 – Hosting third party applications in the IoT gateway

The revenue opportunity is to charge for access to the VM in the gateway and other devices. Fees could be application or service based. The opportunity costs include the development of the VM hosting model, along with the associated APIs and Web services that would allow a third party service provider to access the home information.

For example, offering any security company the ability to leverage the gateway for access to Wi-Fi, 802.15.4, and BLE services, as well as LTE radios in the future, by overlaying their own devices into the home, could be a service that mutually benefits both companies. The service provider doesn't have to get into the security business, yet could make revenue from the devices they have in the home.

The Virtual Machine makes this type of application possible – allowing the service provider and third party to protect each other's services and not impact the consumer experience.

Brokerage for services interfaces for utility and digital services

The commercial model for service provider is to create this brokering service ecosystem. By leveraging its position as the connection to the consumer's IoT devices, the offer/bid solution is to create an automated switching of utilities to the consumer on a monthly or lower granularity. Whether this can be done on a per device or per service basis should also be reviewed. The MSO would then take a percentage of the transaction for providing the brokering of the services.

The MSO also provides onboarding of the utilities as well as details of the power or water or gas requirements of the home through intelligent monitoring.

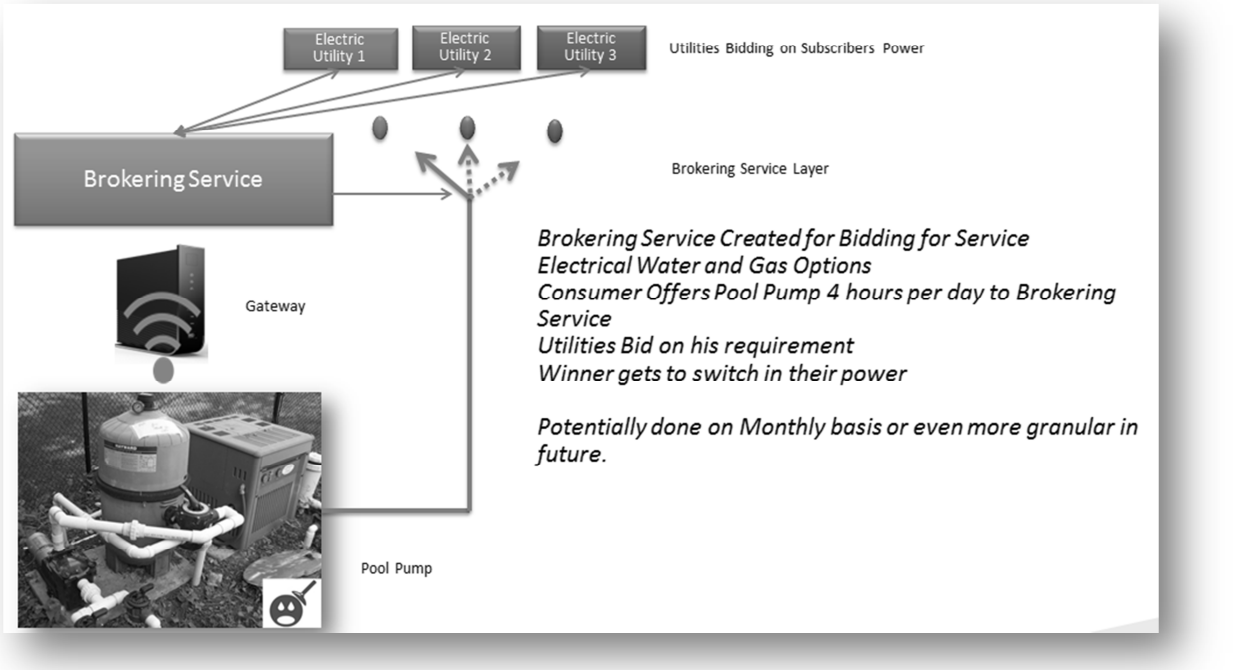


Figure 19 – Example simple brokering architecture

Show me the (IoT) money!

If we take a hypothetical look at a YoY projection – for an IoT revenue stream from the above services example and the resale and use of analytics from the home – the following graph could potentially be the ramp projection of IoT services for an MSO with over 1,000,000 subscribers. The may generate potentially upwards of $50 of revenue, on average, across the entire customer base.

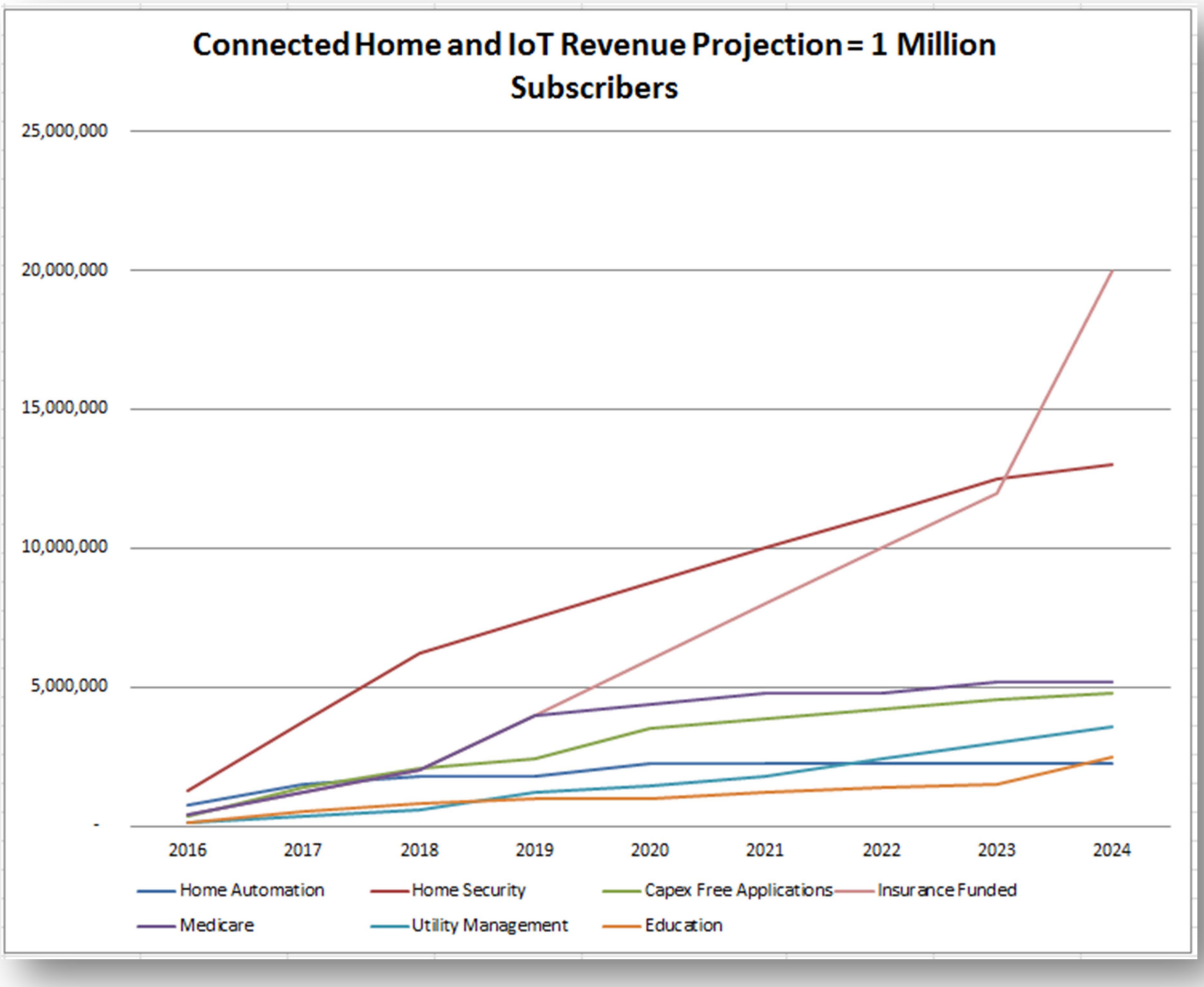


Figure 20 – Hypothetical model of IoT revenue growth by service

	Per Year adoption								
	2016	2017	2018	2019	2020	2021	2022	2023	2024
Home Automation									
<$20 p/m Service	5%	10%	12%	12%	15%	15%	15%	15%	15%
Home Security									
<$30 p/m Service	5%	15%	25%	30%	35%	40%	45%	50%	52%
$5pm Applications	5%	20%	30%	35%	50%	55%	60%	65%	68%
Insurance Funded	1%	3%	5%	10%	15%	20%	25%	30%	50%
Medicare	1%	3%	5%	10%	11%	12%	12%	13%	13%
Utility Management	1%	3%	5%	10%	12%	15%	20%	25%	30%
Education	1%	5%	8%	10%	10%	12%	14%	15%	25%
Analytics/Big Data	25%	50%	75%	85%	86%	86%	86%	87%	87%

Figure 21 – YoY % of IoT service growth

This model illustrates a hypothetical revenue stream from the selling of new IoT based services and could potentially have some of the following market penetration and revenue characteristics to:

- Drive security and home automation to ~ 67% take rates with decreased per month cost and CAPEX investment
- Drive up to ~70% of subscribers to spend about $7 per month on IoT CAPEX free applications based on presence and existing hardware
- Get insurance funding for ~50% of homes for preventative measures
- For Assisted Living / Aging-in-Place and sent home to recuperate applications we may get ~13% take rates with an average of $40 per month service to insurance companies and care providers
- Providing energy and other utility connectivity and monitoring, and analytics may yield another $12 per month and take rates to 30%
- Other Big Data analytics sales of connected home data for about $2 per month (conservative) and ~85% take rate (opt-in typical % number)

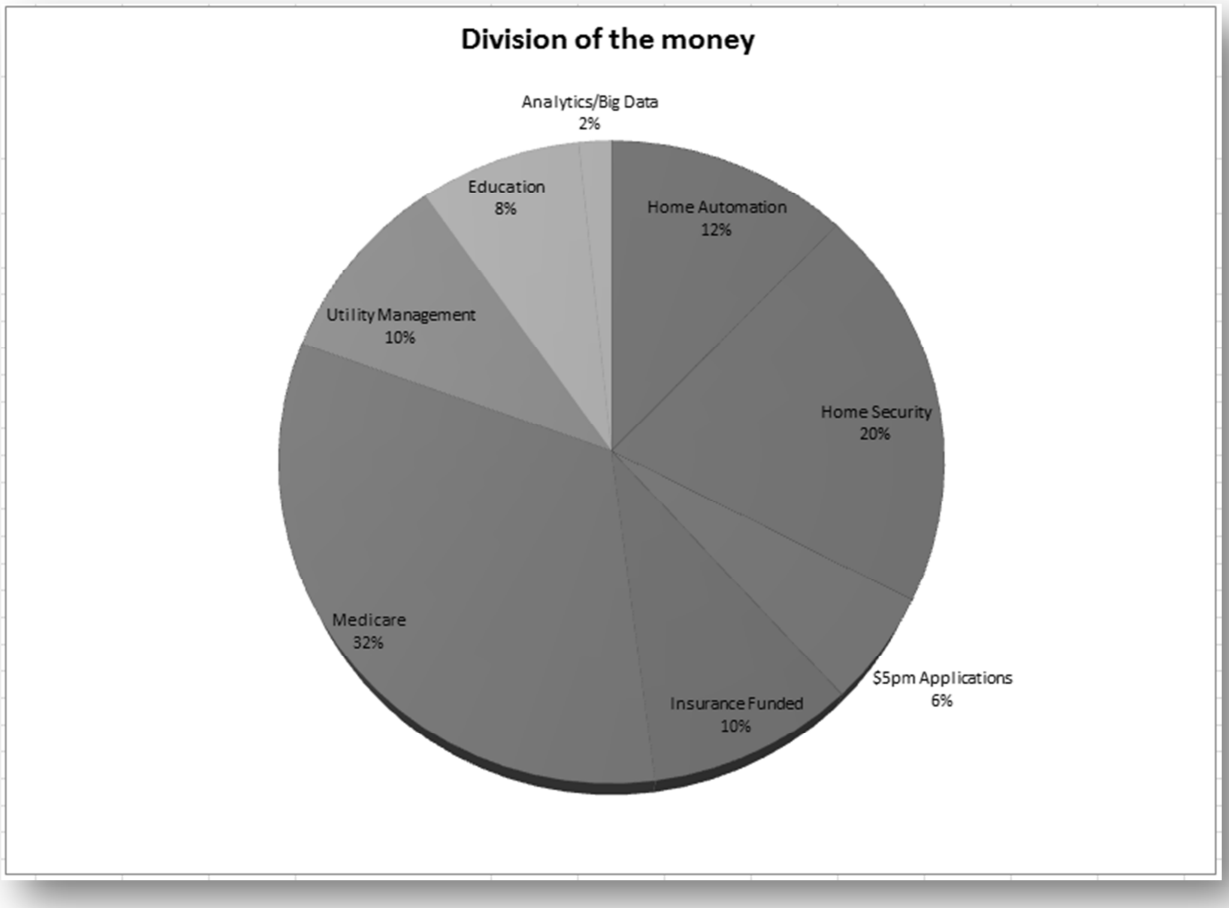


Figure 22 – Breakdown per IoT service in $

Location and presence triggers and services

As we discussed, the connected home needs the connected person to affect some or many of the services that may be desirable to have as a consumer. To be able to create that simple and best positioned and directed user experience – device and people location and presence is a big part of a successful home solution.

The best IoT services will be the services that don't require any specific intervention using smartphone or applications and are driven by ones presence and location. Many home 'scenes' or regular routine actions may be automatically learned or triggered to execute.

For example, the home security lockdown action could simply be done by sensing the TV being turned off, no presence in the bottom floor of the house, or the time of day, etc. Alarm setting and security lockdown could be triggered from the TV being turned off or a number of additional events. In the case of the last TV off at night this could also present the user with a quick camera scan, sensor/contact status, garage door open notification and even police or other social media information relating to the neighborhood or district

As another example, sprinkler systems would not have to use rain detectors. They could instead be driven from cloud IFTTT

weather applications and even soil/moisture analysis sensors.

Additionally, an application could automatically open the garage door when presence is detected via a designated device or set of devices.

There are even more interesting presence and location services that can also be achieved. Consider that as you move from room to room – your home application interface on your tablet or smartphone also changes. When you are in a room with set-top box and TV – the remote for the TV may be presented as the first and landing zone application. When you are in a room with audio speakers, you may be presented with the Playlist and speaker control app as first on your control application landing page.

Who is in the room, vicinity of each other and in front of the TV? There are many ways to try and define who – but one of the easiest is to use iBeacon technology with Bluetooth to declare a person's presence with a smartphone or potentially a wearable fitness device. This allows interesting applications particularly with the service providers TV user experience. For example:

- Recommendations can be done without cameras being required
- Personal preferences can be loaded in the TV UX

Interesting applications can be built as the consumer leaves one room and goes to another room. For example, a TV notification could offer to resume where a consumer stopped viewing a program when they left the other room, even if they forgot to pause the DVR.

ACCESSIBILITY AND VOICE CONTROLLED SOLUTIONS

One of the strongest assets in the MSO IoT play is the integration of an IoT service with the TV experience. This integration provides access for the IoT devices to a visual feedback portal (STB generated TV UX) and input methods to the IoT services. These inputs might be through a TV remote control, an MSO application on TV and smartphone/tablet, and with the increase in voice control of the TV experience the integration of this voice control into the whole home ecosystem.

As was mentioned in the last section, the ability for the MSO to detect presence and movement in the home can also be leveraged into the overall smoothness and integration of the complete home experience.

If we envisage a home with both audio and motion sensors, we can create a simpler interaction with the IoT services and devices through detection of presence and the use of voice control. There is a growing use of applications for "always on" audio sensors – where the user is hands-free and speaks naturally; in this case typically using a code word to trigger the action sequence.

This audio sensor can be positioned in MSO devices like set-tops, gateways, and extenders, and in particular devices like remote controls to pick up the user commands. Of course, the same application can be built on PC, tablet and smartphone using the audio inputs on those devices and leveraging an MSO 'soft' presence through applications on those devices. Voice recognition is particularly useful for control, authentication, and personalization and thus has an important role to play in the future MSO and IoT home.

The growth of IoT is critically dependent on both simplicity of user interaction with any specific device, as also that the complexities of user interaction grow very slowly with the number of devices. In that context, voice is a good modality as it allows for users to interact with IoT devices while doing other things

(e.g. watching TV), thus enabling the user to multiplex IoT interactions with other (media) manipulation. Because IoT interaction in a home setting is likely to be repetitive and with a bounded set of devices, the voice interaction problem in the home has some distinct characteristics from general purpose voice UXes (e.g. the 'Siri' problem on Apple devices).

Companies like Wit.AI are attacking the 'Voice for IoT' problem by making bounded vocabulary command and control interfaces easy to create. Bounded vocabulary command and control recognition is also likely to be more resilient in noisy settings than unbounded voice recognition. Another opportunity for an industry value-added service is the creation and distribution of curated command libraries for commonly used commands (and associated diverse and colloquial utterances).

The use of voice as a modality also solves the input friction in authentication. Given that there are a bounded set of voices in a household, the act of interacting with an IoT device also enables implicit user authentication (and personalization) without explicit effort from the end user.

B2B2C services

It's clear that there will be a number of macro business models in the IoT value-add and revenue chain. These include:

- IoT service direct to Consumer as B2C but OTT to the MSO
- IoT service direct to Consumer as B2B2C with the MSO adding value with relationships with IoT service provider
- MSO Direct to Consumer – as B2C
- B2B2C with Cloud-to-Cloud exchanges between MSO Cloud services and IoT services providers own Cloud infrastructure

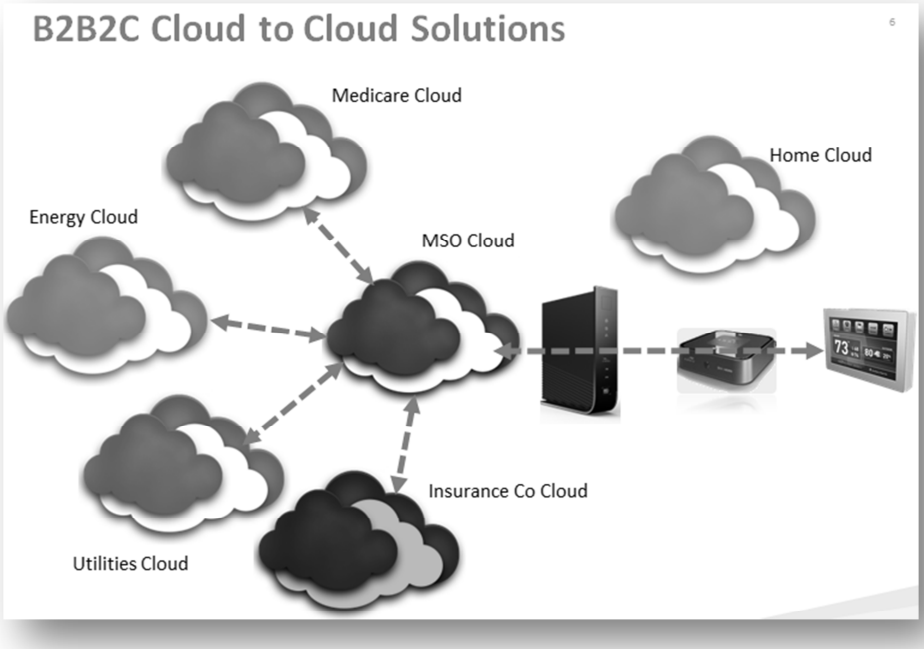


Figure 23 – B2B2C possible IoT services network

The development of these Cloud-to-Cloud interfaces and interchanges will take time. However, the potential exists for an MSO to partner with many of the current services vying to provide connectivity based IoT services to the subscriber. The MSO may optimize the cost and investment chain to provide the service. In particular, as has been mentioned throughout this document, the MSO may also leverage the existing CAPEX investment in device hardware as the IoT gateway and its fleet of technicians.

Data mining and the revenue opportunity for analytics

IoT technology needs to be strongly data driven for two reasons: the superior and personalized user experience, and monetizability.

On the former, IoT operations need to auto configure to a user's 'typical' expectations and context. Take the simple example of a smart light bulb. User expectation of the speed of light bulb operations (e.g. on/off) can range from 50 to 500 msec. Further, the likelihood that the automated turning on of a light may be positive or disruptive (waking up a sleeping spouse already in the room) is a function of social and personal context. Data collection and mining enables a self-learning capability (with potentially periodic or seasonal adjustments) for IoT devices that adjusts to the evolving behavior of users.

On the monetization front, data collection and mining enables the MSO to monetize (and subsidize) IoT services via the creation of third party relationships, both advertising and commerce. As articulated in http://wfoa.wharton.upenn.edu/perspective/venuvasudevan/, proactively understanding a consumer's use of a device like a washing machine enables an MSO to create just-in time advertising relationships with the detergent advertising companies, or Angie's list style intermediaries for periodic repair.

Supporting services

Assuming IOT is valuable enough that a subscriber finds benefits in a service tier to support it, the tier may include associated services such as data archival, security, rules engines, remote and mobile access, expedited customer service, etc. This is the place that consumers pick their own (hopefully standards-based) devices and they are discovered and integrated into the MSO management and presentation

The network and network management

How does the MSO scale for the millions of devices creating billions of transactions on the network?

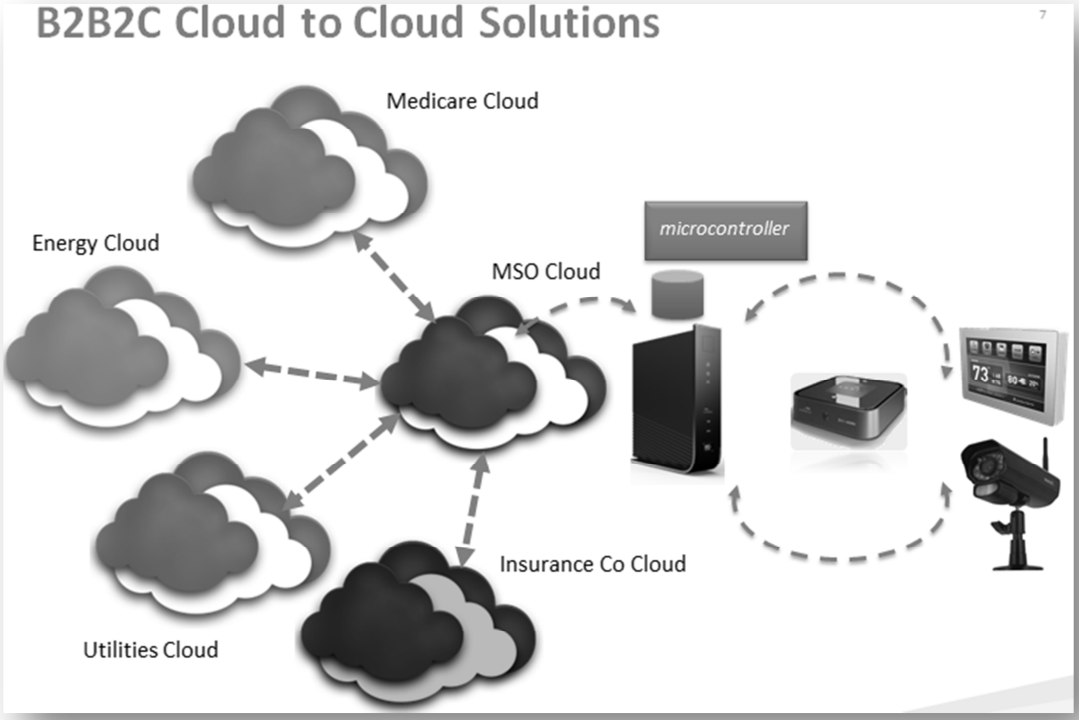


Figure 24 – Leveraging the gateway to scale transactions

There are many performance requirements for IoT services and devices. Some require very low latency, for example turning off and on a light switch (<500ms for action/response desirable). Some require lots of bandwidth, like HD security cameras, most of the devices are 'chirpy' and require maintenance level information to make sure they are still alive and ready to trigger an IoT action or service.

Some of the services require high reliability and availability, such as security and high end Medicare monitoring services, where they need reliable home and network connections and in some cases require an additional LTE backhaul capability in case both the primary Wi-Fi and MSO Internet connection is down. The creation of Wireless Personal Area Networks (WPANs) will generate a significant amount of new traffic for the MSO. With IPv6 addressability like 6LoWPAN architectures, there will also be an increase in addressing scope.

Protocols such as ZigBee have been designed to minimize the power used by a protocol like IP (Internet Protocol) and create their own connection over a home network that sleeps/wakes to preserve power in the end device. Other protocols such as AllJoyn, OIC, and Thread are TCP-IP/UDP-IP based and drive the power requirements of the end devices higher and are more suited to unconstrained devices with higher power and higher MIPS capabilities.

One of the key elements for consideration in IoT architecture is the MSO Gateway. As we have seen when the 802.15.4 and BLE radios are added to the Wi-Fi radios – it becomes the complete IoT gateway with capabilities to:

- Onboard the IoT devices for connectivity
- Provide protocol conversion at both the PHY and the software protocol level with the addition of the main IoT stacks

- Provide the scope to become a microcontroller for the IoT devices and to provide WPAN functions locally. These functions:
  - Minimize latency for in house trigger/actions
  - Provide a local rules capability to improve IoT experience and minimize the traffic to the cloud
  - Cache local IFTTT actions for improved latency and robustness when the Internet is down
  - Work more efficiently with Cloud IoT controller and rules based systems

While there will undoubtedly be more upstream and downstream traffic to/from the home because of IoT, the MSO can provide a local WPAN microcontroller in the gateway to minimize this traffic while simultaneously adding value with lower latency and robustness in the absence of the Internet. Additional to latency improvements, there is also an added value to security by having the gateway play a proactive role in securing IoT; a topic covered later in this paper.

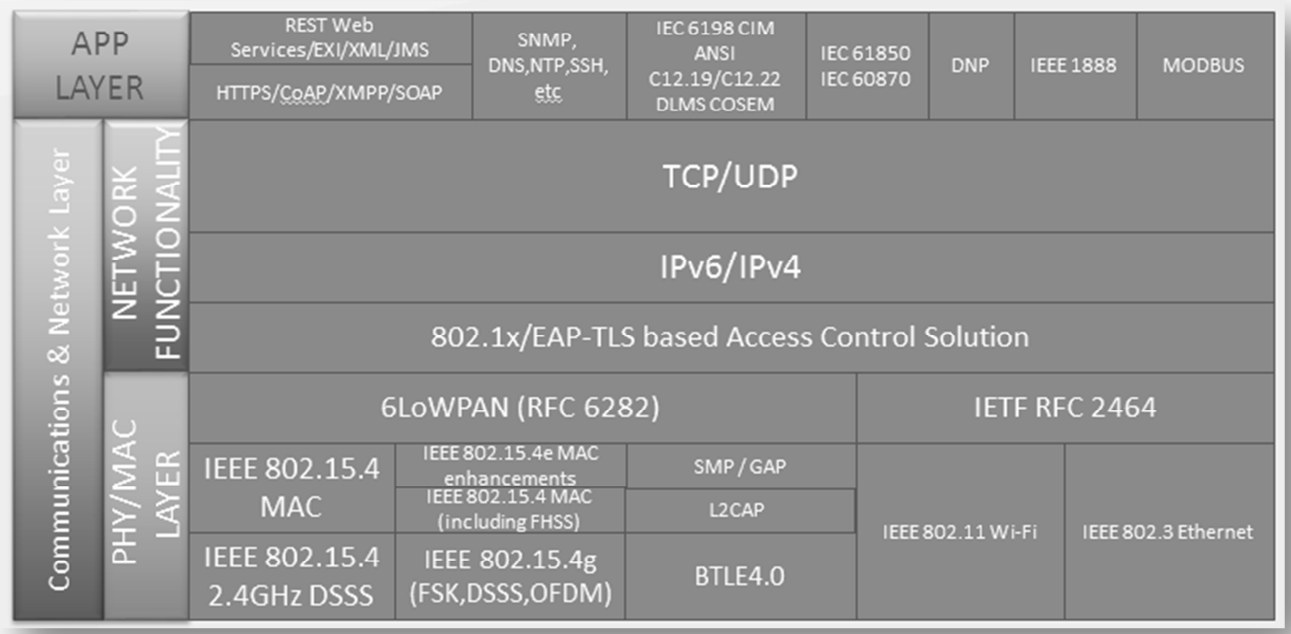


Figure 25 – Example of the IoT protocol stack

IOT will involve a number of different physical layers, primarily focused on low power per transaction. Running an IPv6 protocol over 802.15.4, Bluetooth low energy, and potentially 802.11ah low power Wi-Fi, may all need to be supported at the Gateway in this MSO IoT architecture. While Cloud control of these services is desirable and required in some applications, it will be key to ensure that the gateway supports the ability to apply IoT rules locally to the WPAN.

Data aggregation

The Big Data repository will certainly reside in the Cloud. This will be a key learning and analytics point for macro level views of large data sets. Performance will be

critical as the cloud IoT aggregation service will accept literally billions of transactions per day, some with very low latency demands.

If the actual number of IoT devices even remotely approaches the 26 billion estimate, then the resulting data deluge presents a formidable problem. But is this a 'vanilla' big data (i.e. lots of data) problem, or are there unique characteristics to IoT Big Data? The local gateway microcontroller will be the arbiter for this and ensure that only the true unique data is sent to the cloud aggregator.

Taking patterns, usability, and societal impact perspectives suggests the following distinctive components:

- IoT data is likely to be bursty, both because people and place activity is, and also that anomalous situations (e.g. a storm) are likely to cause large step function changes in data traffic
- Unlike some other data sets (e.g. e-commerce), IoT data can contain long stretches of boring 'life as usual' data, where collection and distribution costs might overshoot the user and business benefit
- Almost all IoT data is personal, so the privacy implications of any breach or misuse are disproportionately high. Even at this early stage in the industry evolution, this issue has caught the attention of the Federal Trade Commission (FTC) who are concerned with these issues
- Given the fragmentation of IoT platforms and devices, any single entity is likely to only have a partial IoT view of the connected home

These characteristics of home IoT data suggest the following attributes to the data aggregation architecture:

- Decentralized data aggregation architecture with CPE level intelligence (and 'Fog Networks'; a contrasting architecture to the cloud in the home) to filter and aggregate 'life as usual' data in latency and cost minimized ways. This augmented with edge intelligence to deal with locale specific responses in low latency ways.
- Dealing with security and privacy concerns via 'limited data window' aggregation technologies in the cloud (and subsequent deletion of the raw data) and edge node summarization techniques that limit the collection and transmission of raw data to the cloud
- An ability to rapidly move data processing functions between cloud, edge and customer premise in response to exogenous network or environmental conditions

IoT and Security and Privacy

The recent press has numerous articles concerning the impacts of poor security practices, as well as on attacks on systems thought to be secure. Introducing new IoT devices that are connected to the outside world magnifies the problem of security.

These devices have access to new data about the home, and about the people living there, report that data, and exert controls on the home environment. Security for the Internet of Things is absolutely an essential aspect to provide, or the IoT will be a source of new attacks compromising consumer privacy and device functionality. For devices that relate to medical data, or significant environmental control, security weaknesses could even create serious consequences for the consumer. Additionally, we see that 20% of people are not comfortable with adding devices because of security and privacy concerns. Privacy is front of mind of many consumers and probably one of the biggest inertia factors for use of sensor devices with video (cameras) and audio.

What should the attributes of a secure IoT implementation be? At a high level, a secure IoT system must protect the privacy of the individual, and by extension, all sensor data collected that pertains in any way to that individual. When an infrastructure server element must process such data, only the required data should be provided for the functions expected, and any Big Data examination of such data must be anonymized. For IoT devices that exert some control over the environment, control commands must be protected against hacking. Entry points to the home for any specific control must be minimized to that system exerting that specific control. The following sections provide specifics.

Privacy and IoT Devices

Even with device robustness, communication security, and public key identities, IoT devices require an additional type of protection. Quite often, the patterns of sensor reports themselves can convey information that needs protection. For example, if cameras are set to report movement only, then the absence of reports may indicate the home is not occupied , and a convenient target for criminals. In such a case, occasional (encrypted) reports need to be issued that in fact report no movement, to break these patterns. Another type of attack might involve moving devices between homes to spoof information or behavior. Location needs protection by including environmental information to assist in proving identity. As an example, a router might report the networks it discovers near the home, including SSIDs and RF channels.

This sort of information is only "approximate" in the sense that it is not repeatable. However, algorithms exist to determine enough similarity to identify location in this fuzzy logic sense.

Key elements in the privacy model that need additional work may include some or all of the following:

1) Leveraging the security models built into solutions like AllJoyn, OIC, and Thread, but not relying exclusively on them.

2) Adding additional security layers over the standard solutions may be a way of adding additional differentiation to the service provider's IoT solution. A particular device or protocol may report being hacked in some OTT environments. However, it may not be hacked in the MSO additional security layer.

3) Making it more difficult to hack an MSO IoT solution using additional information about the home relationship to the device particularly for onboarding could be a key value-add differentiator. The service provider already has trusted elements like a gateway and set-top to use to validate the correct presence of the new device and aid in the security of both onboarding and persistent communications.

4) Developing anti-spoofing solutions for IoT devices to make sure they are authenticated and the right connection chain authenticates their authorization in the home environment.

5) Developing solutions to mitigate against repetitive 'tell all' patterns that may determine who is home alone, if anyone is home could be an important security and privacy solution.

6) In particular, for video and audio devices, ensure that this traffic is transparent for the user with leverage of TV UX for camera and audio sensor transmissions. Also, include gateway and IP level traffic blockers when the home owner is not comfortable with camera and audio sensor devices from third party sources.

7) Make this easy to do and potentially incorporate a privacy button on a remote control, offer modes that when there is presence of one or more of the inhabitants no external transmission of camera information is made.

Communication

All communications between IoT devices and any related servers must be encrypted and authenticated. Typically, for IP end points, web security standards such as SSL and IP-Sec should be employed. All end points should have X.509 public key certificates and keys, signed by accepted and validated trust authorities. Unfortunately, many simpler constrained IoT sensors and devices are not IP and security capable and may use other protocols like ZigBee to a hub that then provides security solution. This IP gateway node in the home serves as an intermediary, providing the IP security desired for the node, and the translation to simpler forms of networks and security.

Consumer Data Access Management

All sensor or media related data collected within the home, and intended for processing in an infrastructure server, belongs to the consumer. The consumer must be able to define its permitted usage in all respects. Some collected data remains within the home for access by the consumer; other data must be passed out to a server node for processing and response. Typically, different device classes are associated with different such servers. Thus, the consumer must be able to set which data can go to which server, and permit specific controls to return from specific servers. This data access management function is akin to the permissions requested for each Android and iOS application installation. However, too often those requests are simply "rubber stamped" by the consumer. In the IoT system, data management has to be available in a user-friendly manner, clearly setting forth the impacts and purposes of the devices in question, and the related data and control. In many cases, the consumer would not understand the subtleties of data management.

Fortunately, recent federal laws are helping IoT component manufacturers steer clear of deceptive practices. IoT devices will be held to the higher standard of identifying clearly the data required as mandatory for the device to work, and optional for additional functions. Further, this minimum access set must be the default set-up. Once the data is processed at the server, any use of that data in a Big Data aggregate is optional, and must be agreed in advance by the consumer. If so agreed, the data still can never be identified as to the consumer it is associated with, without additional agreements in place.

Remote Access

In many use cases for IoT devices, the consumer can remote access infrastructure servers or in-home gateways to effect control or check status. Such communications falls within the "always encrypted and authenticated" regimen described above, but often the mobile devices used are generic non-IoT devices, and access of this type is quite often logon/password in nature. This should not be the preferred mechanism for remote access, as it is subject to hacking and attack. Password attacks are well described in the literature. Fortuitously, the Wi-Fi standards community has created a new approach based upon installed certificates and keys, called Passpoint. IoT remote access should employ a similar public key identity based approach, or perhaps Passpoint itself.

Device Robustness

Often security systems focus on the communications between devices, and not on the devices themselves. However, in commercial video delivery, very robust in-

home platforms have been built cost effectively, where all devices boot securely, have only signed authorized code on them, and where physical access to keys and cryptographic functions is prevented. This should be the high ground for all in-home IP connected nodes for IoT systems. For devices that conform to simpler standards, robust designs should be encouraged as an evolutionary path.

SUMMARY

The Internet of Things (IoT) is at the peak of inflated expectations. With even a small percentage of the predicted numbers of devices involved and potential revenue realized, the benefits are substantial to consumers in terms of lifestyle and businesses in terms of productivity and profitability. Additionally, the impacts to the interconnecting networks will be considerable as the trends are already pointing to tens, if not hundreds of devices, connected per person. The opportunity to excel at or surrender to the impact of IoT will be based on designs with consumer value in mind, scalable network architecture, meaningful data analytics, and attention to matters of data integrity. A sound model for security and privacy of devices through the life of the device and the end user has not been developed and should be point of focus both in technical and business policy synthesis.

The MSO has several differentiated capabilities to leverage their success while participating in the IoT. Cable networks are well positioned to handle the new influx of telemetry and control points. Big Data capabilities already in place for triple-play services can be adapted for the new commodity of sensor data. Operations, Administration, Management and Provisioning (OAM&P) facilities provide the foundation for onboarding and service assurance of IoT devices, although currently are not capable of accepting the wide variety of device types. MSOs have a prominent position on the screen most used in the home, the television. Coupled with voice control and eventually voice response, the television is very convenient and useful, particularly with the graying of our population.

A business plan for the IoT related business should involve facilitating a wide range of consumer-selected devices to offer data and control to and through the cable network. This step allows both the collection of data sets for analytics and the ability to integrate into MSO provided user interfaces on television, mobile, web and telephone. The business models discussed in this paper suggest the need to reduce the equipment costs associated with today's smart home solutions to allow ease the ability to achieve large-scale deployment and allow profitability with one or more services that generate revenues in the order of a few dollars / month. Large-scale deployment is needed to collect a sufficient data set to attract advertisers and other target markets for B2B2C opportunities.

The technical solutions outlined in the paper involve several recommendations for standard technologies involved with communications and device ecosystems. These areas are relatively immature and likely to see entrants, exits, and consolidation therefore a pluggable architecture is recommended. A set of recommendations for fundamental inclusion of robust security capabilities is key to ensuring the emergent IoT remains in positive light during its formative state.

Contributing Authors

Paul Moroney
Venu Vasuvedan
Ian Wheelock
Wade Carter
Mark Bugajski
Eli Baruch

Delivering Seamless Subscriber Aware Services over Heterogeneous Access Networks using a SDN and NFV framework

Nagesh Nandiraju Ph.D., Yiu Lee and Jorge Salinger
Comcast Cable

Abstract

The cable industry is seeing two major shifts: adoption of new access technologies that are capable of delivering higher data rates and a growing interest towards new advanced services that go beyond traditional service flows. Access network technology is undergoing an evolution. Next generation DOCSIS (DOCSIS 3.1) and Passive Optical Networks (e.g. 10G EPON) enable MSOs to deliver higher data rates to subscribers. On the other hand, Network Function Virtualization (NFV) and Software Defined Networking (SDN) have certainly excited the industry and many innovative network architectures and applications are being proposed using these concepts. MSOs are researching to develop and offer more advanced and seamless subscriber-aware services to users that go beyond the DOCSIS service flows. These services will be abstracted from access network and end-to-end network technology has to be agile and capable to deliver these services.

In this paper, we aim to establish a SDN and NFV framework that enables delivering a uniform set of new services across multiple access networks. We will start by elaborating on what are the new services and key drivers for these services, how operators would offer and manage them. Then we will dive into the changes required in the network that will enable MSOs to quickly deploy new services to subscribers. As a part of this effort, we will present our thoughts on what functions can be virtualized in the access network that will improve the agility of the network.

INTRODUCTON

MSOs are currently seeing several transitions that can help drive their next generation network architectures. First, over the past decade or so, Data has been growing significantly at 40-50% compounded annual growth [1]. The advent of smartphones and cloud-based services has only acted as a catalyst to increase the data consumption. Consumers are increasingly relying on new services and applications that present data to end-user. This translates to more bits delivered per customer but not necessarily increased revenue for service providers. Advances in network equipment and technology have been unable to push down cost per bit at the same rate of bandwidth demand. Many MSOs and telecom operators are increasingly seeing this trend and are looking at ways to enrich the customer experience by adding new value added services to their portfolio in addition to just delivering data.

Second, in the access network, MSOs are continuously improving existing protocols (DOCSIS 3.1) and continuously increase capacity to meet the customer demands for higher speeds. DOCSIS has proven to be one of the most successful broadband access technologies. DOCSIS 3.1 can deliver 1Gbps and will be capable of up to 10 Gbps in an all IP network.

Next, the infrastructure to deliver data in the last mile can use several media – HFC, Fiber, Coax, CAT6. There are considerations of using fiber to the building or units in new green field deployments. Several protocols such as GPON, EPON, RFoG, Ethernet can run over these fiber networks and are capable of delivering 10Gbps or even higher speeds.

On the technology front, Network function virtualization (NFV) and Software Defined Networking (SDN) have clearly been the top buzzwords in the networking industry over the past couple of years. The "talk" has transformed into many interesting developments from vendors. Several vendors are demonstrating general-purpose hardware capable of performing network functions previously only possible over purpose built hardware. These activities are primarily accelerated by the increasing use cases of networking by cloud service providers.

MSOs are actively researching to establish their next generation network architecture to better support these transitions, and how they can offer a wide variety of services with an enriching experience to end customers. As the hype phase of NFV and SDN has passed over, time is ripe for investigating how and where these technology advancements can be applied into the network to monetize and improve operational efficiency. Particularly, MSOs are researching to develop and offer more advanced and subscriber-aware services to users that are seamless and go beyond the DOCSIS service flows. These services will be abstracted from access network and end-to-end network technology has to be agile and capable to deliver these services. The questions that often get asked on NFV are: "Will MSOs embrace NFV and SDN? Can NFV offer performance similar to purpose built hardware and is it cost effective?"

Will MSOs Embrace NFV and SDN?

In the not-so-distant past, many of us were accustomed to carrying a variety of accessories and specialized gadgets such as MP3 players, Cameras, GPS, cell phones, USB sticks, and this list goes on. This trend has suddenly changed with the advent of smartphones. Now, people's traveller kits are much lighter as they only carry smartphones, and all the above-specialized applications are virtualized and underlying hardware components are either natively supported by mobile chipsets or packed into the smartphone. The services provided by these accessories are transformed into applications (apps) running on a generic piece of hardware and an operating system sitting on top of that hardware and interworking with the server functions running in the "cloud" to provide necessary functionality and services to the software applications (API). Applications are almost agnostic of the hardware and somewhat tied up with the operating system. The number of apps in both Apple and Android app stores is well over a million. According to Apple, it paid $13 billion to app developers [1].

NFV and SDN are two concepts to rationalize conventional network design. In traditional model, every day MSOs face two challenges in their network. First, MSOs deployed many purposely-built network equipments such as routers, switches, load-balancers and firewalls. Those equipments are designed to serve specific set of functions. They usually work anomously and require different operational procedures to interoperate them. Network integration and operation is well-known time-consuming and complex. Second, local applications such as home security and VoIP require to implement client applications in the Customer Premise Equipment (CPE). MSOs must develop the applications or contract the CPE vendors to develop the applications. Then, the MSOs must upgrade thousands of CPEs when they are ready to introduce the new service. NFV and SDN attempts to apply software development paradigm into networking world to improve the level of automation and deployment efficiency. At high-level, NFV aims to virtualize conventional network functions into softwares running on Commercial Off-the-Shelf (COTS) and Virtual Machine (VM). SDN aims to orchestrate virtual network functions (VNF) that are running on COTS and VM. The potential benefits of NFV and SDN are enormous. Similar to the

transformation brought by smartphones, NFV and SDN aim at bringing similar transformation for service providers by moving most of the services from running on purpose-built hardware platforms to COTS compute, storage and networking infrastructure. Some benefits are obvious: flexibility to scale for service demand, ability to launch new services with ease, agility in service provisioning, automation in operation, and dynamic scalability. But a fundamental question that is asked by almost everyone in the service provider environment is, how can performance be effective with COTS?

Indeed, similar questions were asked when voice/camera/GPS and other critical apps have been proposed to run over generic smartphone software instead to their purpose-built hardware predecessors. The questions are now history as the industry and smartphone users have overcome the doubts and almost all these apps are running on everyone's smartphone fairly regularly.

Similarly, with NFV, a large variety of service provider applications (e.g., firewall, Intrusion Detection System (IDS)/Intrusion Prevention System (IPS), malware detection) can run over a common infrastructure with generic hardware and virtualization. There are primarily two classes of applications in a service provider environment: data plane and control plane. Control plane applications are relatively easy to run over the COTS platform while data plane applications require hardware accelerated paths and orchestration features that prioritize packets, minimize latency and provide line rate throughput. For many applications not only raw bandwidth, but packet processing in terms of million packets per second (pps) and ultra low latency are vitally important.

CURRENT TRANSITIONS IN MSO ACCESS NETWORKS

MSOs are always in the quest to bring the best and reliable services to their customers. They are constantly augmenting their existing infrastructure or building new infrastructure to support new and upcoming customer demands. Below are some of the transitions that help address the changing demand for current services, diverse geographical serving areas and landscape of new services.

- Multi access architecture
  - DOCSIS has certainly been the most successful access technology for MSOs. With technology advancements and maturity of various fiber based PON solutions, some MSOs are building fiber all the way to new buildings and making use of different PON technologies. Alternate last mile access tehnologies such as Ethernet and WiFi are also seeing some adoption to serve certain end customers. This implies that supporting uniform services across multiple access technologies will be quintessential.
- Remote Architecture for DOCSIS
  - There are several drivers behind the remote architecture in DOCSIS. As a natural evolution, service group sizes in DOCSIS networks are becoming smaller in an attempt to provide higher capacity to subscribers. DOCSIS 3.1 with the help of OFDM enables an efficient use of spectrum by supporting higher order QAM modulations. However existing analog optics and density of RF can be limiting factors to fully reap the benefits. One way to address and benefit

from these changes is by adopting a remote architecture and move from analog to digital optics.

- Remote architecture for PON
  - In an all fiber access network deployments, PON technologies (EPON or GPON) have reach and density limitations. PON technologies have been designed for a 32/64/128 split ratios and typically operate within 20 kilometers. Since MSO networks have traditionally longer reach with their HFC networks and higher densities, the need for a remote architecture in PON deployments is desired.
- Vitualization of Home Network (vHN)
  - MSOs use their robust access network to deliver various services such as voice, video and data to their customers. As we enter into the arena of operating multiple access technologies such as DOCSIS, PON and WiFi, these services should be agnostic to the access technologies and work seamlessly on any type of access networks. However it often isn't the case. Each technology has its network characteristics and its deployment scenarios. The current model couples the service characteristics to the network characteristics. For example: to deliver linear content on DOCSIS network, the current practice requires MSO to statically or dynamically provision a dedicated DOCSIS service flow for the video delivery platform. This process will not seamlessly work in PON or WiFi. The goal of vHN is defining an architecture to decouple the services from the access network technology. Application or service logics are virtualized and could be run inside the network or in the CPE.
- IP CDN
  - With the growing web content, network DVR and IP video, the edge caches will naturally move deeper into the networks, perhaps next to the access network gateways (closer to subscribers).

With these transitions in mind, MSOs are interested in building an architecture that leverages the synergies between these changes. Specifically some of the key value propositions and aspects we are interested in are: How can NFV

a) improve our ability to launch new features and services at a much faster pace reducing the cycle time

b) increase the utilization of common compute, storage and networking resources to save space, power and increase operational efficiencies

c) reduce risk associated with rolling our new services (e.g. lower upfront cost per service per subscriber, enable the elasticity of services).

APPLICATIONS - SERVICE PROVIDER'S APP STORE

Modern MSOs data service is more than just a conduit to deliver IP packets. MSOs are actively working to enhance existing application and introduce new applications to their customers. Some examples are parental control, remote storage, home network health dashboard and home automation. Parental control is one of the key applications that is being virtualized to provide the necessary control.

The number of personal devices per home consuming data is steadily growing and currently ranges from 4-15. This number is only increasing every day. As a result, home networking is getting more complicated.

Customers are expecting more information from their networks with deep data analytics of their data usage and easy access to device policy management. This requires applications at the higher layer to provide accounting with some hooks from the access technology layer.

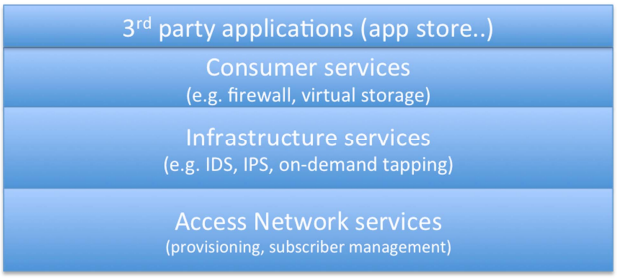


Figure 1: Service Provider Application Stack

Network attack is a real threat not only to big enterprises but also to home users and small businesses. It is not uncommon to hear hackers exploit security vuluerabilies and steal personal data and take control of victim machines to launch attack. Hacking techniques are evolving rapidly, becoming much more sophisticated and harder to detect. Recently, as many high profile data breaches (e.g., Target, Home Depot, Turbo Tax) have been posted in the media, users are more aware and concerned of network security. The traditional client based protections such as anti-virus and personal firewall software require software installation in hosts. Users must constantly upgrade the thread signatures to protect from newly discovered malwares. Besides, client based protections require the software to run on the host, Internet of Things (IoT) such as webcam and smart thermostat may not have the processing power and memory to run the software. The next logical place to protect a user premise is in the ISP network. Customers are expecting their ISPs to offer more protection.

Figure 1 shows a simplistic view of the different layers of services that are likely to run in an MSO network. The bottom layer is the access network services layer which handles applications such as provisioning and subscriber management. The next layer is for applications that are more focused towards network engineers to maintain, manage and protect operator's internal network. Consumer services layer is where the new set of services or applications described above will be located. It is also easy to conceive that having this layered architecture and exposing the right hooks can facilitate development of 3rd party applications that can address a wide variety of application needs.

THE 3 FUNDAMENTAL VARIABLES TO PLAY WITH AND APPLICATION SWEET SPOTS

Compute, storage and networking are the three fundamental blocks that any application or service will require. These are also the three key knobs that vendors and service providers can play with to optimize the performance of applications and delivering the services that drive customer experience. There are wide variety of applications and services that operators are currently providing and/or plan to provide in near future. As shown in figure 2 these applications fall into several classes that can be optimized on one or more variables. For example, consider "policy management" application. This is primarily a control plane application and can be optimized by high CPU power, faster access to memory and persistent storage. Consider another example – VoD service. This services typically uses fixed assets that are preprocessed with multiple bit rates and stored in various storage locations. Delivering this service to end customers relies on substantial storage and network capacity with some need for compute power. The end customer experience (such as smooth access, higher quality video bit rate streams, engaging UI) can be optimized without breaking the bank of service provider (i.e. augmenting network capacity) by playing with the three variables.

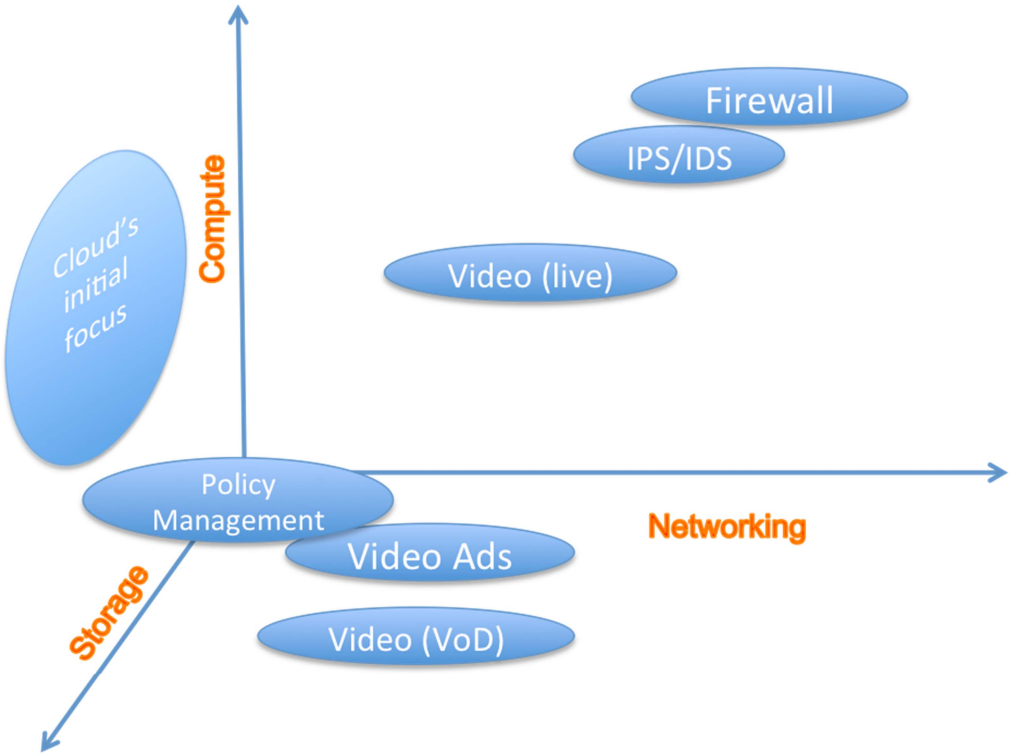


Figure 2: Three Fundamental Variables that can be optimized for applications or vice versa

How does Cloud Computing play a role?

Cloud computing platforms have been optimizing their offerings to enable applications running seamlessly on virtualized compute and followed it by making scalable storage simple to access and reliable. The fact that any fundamental block of operation often involves compute and storage required cloud providers to attach storage to compute and offer storage as a service. The volume of compute and storage in a cloud environment and associated transactions has found short falls in existing networking solutions. This has driven innovation in the networking space both in protocols (e.g. VxLAN, NVGRE, network orchestration) and hardware (dense 10G and 40G switches in 1RU form factor). However the innovation is primarily focused towards a dense, contained data center environment. The innovation is now expanding to address the issues in inter-datacenter traffic (also referred as east-west traffic).

A service provider environment is different when compared to a cloud service provider. In a service provider environment, the primary focus has been on networking to address the challenge of transporting more and more bits to many small and relatively sparse distributed end points.

Optimizing Applications for balancing the tradeoffs between the 3-axis

As show in figure 2, different applications have their own sweet spots in the 3-D space of compute, storage and networking. Some applications can be optimized and architected to change their coordinates in this space. For instance consider two key applications of a

traditional service provider: Video ads and "Video on demand". Both these applications can be hungry for the three resource categories. However if we carefully architect these applications and their delivery mechanisms, we can optimize the resource consumption and move them in the 3-D space. Videos can be preprocessed in one location at a defined time. They can then be distributed to various locations for caching (CDNs) during network's off peak hours. An intelligent controller can monitor requests and analyze trends to prefill video caches with likely movies or ads to locations closer to the requesting end points. Although CDNs have been performing these operations on static web content for many years now, the real application for service provider is ability to transfer these video during off peak hours to maximize the network utilization. This not only increases the network link utilization, but also reduces the need for expensive investment to build new links or upgrading to new technology.

CHALLENGES WITH CURRENT ARCHITECTURES

There are several challenges with current architectures. Network appliances are specifically designed for a service and very slow to change due to limited marketability. New services using network applicances require higher upfront investments and time, so it is difficutly for MSOs to try out a business model without significant capital and operational investment. This contrasts to Web content model where web content providers can try out new applications relatively faster and cheaper.

Silos: Typically an operator's network is a group of silos that can be based on services, technology, vendors, or protocols. Voice, Video and Data are three main silos that have existed for a long time. Voice has slowly transitioned over to Data and the walls between these silo entities are slowly blurring. Next with IPTV, as packet based transport and services become the predominant service, it has become easier than ever to imagine a single platform delivering these services to end customer.

Then next level of silo is based on technology. In the access network, until now DOCSIS has been the dominant technology of choice that runs on a HFC network. The DOCSIS access gateways (CMTS/CCAP) have been vertically integrated with all Layer 1-2, many Layer 3 and above functionalities. Some MSOs are deploying fiber deeper into the network and in some cases all the way to the home. With fiber to the unit or building, Layer 2 based technologies such as EPON or GPON can be used in delivering the data services. In certain cases, particularly MDUs, where Cat6 cabling is present, operators will use of Ethernet in the building to deliver the data services.

As operators are foraying into new services that are subscriber aware, the goal will be offer and maintain the same features across multiple access network technologies. Traditional way of tightly coupling these services to underlying access technology is not desirable.

Clearly with the use of several technologies to deliver the same services, vertically integrated systems tailored for each access technologies can soon become a bottleneck. Abstracting and keeping certain higher layer services independent from the transport technology can help in reducing the time to roll out new features or services both from vendors and MSOs.

Service semantics: Next challenge is coupling "Services" with underlying technologies. QoS is often the most commonly used feature in the access network to differentiate services. Variables such as peak traffic rate, sustained traffic rate, maximum burst size, latency can be controlled and used as distinguishing features in a service. In a DOCSIS

environment, service flows are used to distinguish traffic flows or services. Typically voice would have two service flows, different data tiers will use different service flows. In EPON, use of LLIDs or VLANs can be used as service differentiators. In GPON, combination of Alloc IDs, TCONTs, GEM port IDs can be used to distinguish service treatment.

It remains an open question whether all or a subset of these variables will continue to be the key levers of choice for MSOs to differentiate services in the future. This is especially a relevant question when considering the consumer appetite for feature rich software applications and seamless services, MSOs access network evolution towards smaller serving group sizes coupled with very high peak to average bandwidth ratios in the last mile access. Of these variables, differentiating on latency is likely to be a key lever, considering the use of cloud computing. Today, many web and mobile applications (most mobile applications use HTML 5), are merely thin clients, presenting an immersing user interface but behind the scenes they extensively rely on remote API calls. Minimizing the response times to these API calls will become a critical service.

Although it is hard to predict but given the natural evolution towards smaller serving group sizes and use of high capacity access technologies, it is unlikely that the access network continues to be the choking point or as constrained a pipe as it used to be when DOCSIS was first deployed. Table 1 and 2 shows the Average bandwidth per sub for different access technologies. For instance, consider the average bandwidth per subscriber when we started with DOCSIS 2.0 (0.08 Mbps per sub – see Table 1) and where we are today (see Table 2) - it is a dramatic difference. In the case of DOCSIS, serving group sizes are between 250 and 500 (note: there is a long tail distribuation with larger and smaller serving groups). In case of PON based access network, this serving group ranges between 32 to 128 subscribers. Most MSOs are also considering similar serving group sizes for DOCSIS in the near future. Normalized ratio of capacity to rate tiers offered is nearly 6-12 times higher in just DOCSIS.

Table 1: Average capacity per sub (Mbps) in DS

	500 subs per SG	250 subs per SG
D2.0	0.08	0.16
D3.0 4-ch	0.32	0.64
D3.0 8ch	0.64	1.28
DOCSIS 3.0	1.28	2.56

Table 2: Average capacity per sub (Mbps) in DS

	128 subs per SG	64 subs per SG	32 subs per SG
EPON	7.81	15.63	31.25
GPON	19.53	39.06	78.13
10G EPON	78.13	156.25	312.50
DOCSIS 3.1 (1x96MHz OFDM block +24 ch)	19.00	38.00	76.00
DOCSIS 3.1 (2x192MHz OFDM blocks)	36	72	144

SDN AND NFV FRAMEWORK

In NFV based architecture, the idea is to have a pool of compute, storage and networking resources that can be molded by software to serve multiple services. Combined with virtual networking and SDN where the control functions are managed by the software running in a centralized controller to initiate flows to orchestrate the VNFs running on COTS or VM, we can create a fully agile framework for provider infrastructure the industry has ever had. This will enable operators to:

- Build a service or add any features that offer value to customers and add new revenue streams for operators.
- Try a new service quickly, adapt quickly, deployed at any scale
- Optimize utilization of resources (consider the peak to average ratio of compute)

As shown in figure 3, the access technology can be abstracted from the services layer. By doing this, access technology vendors can

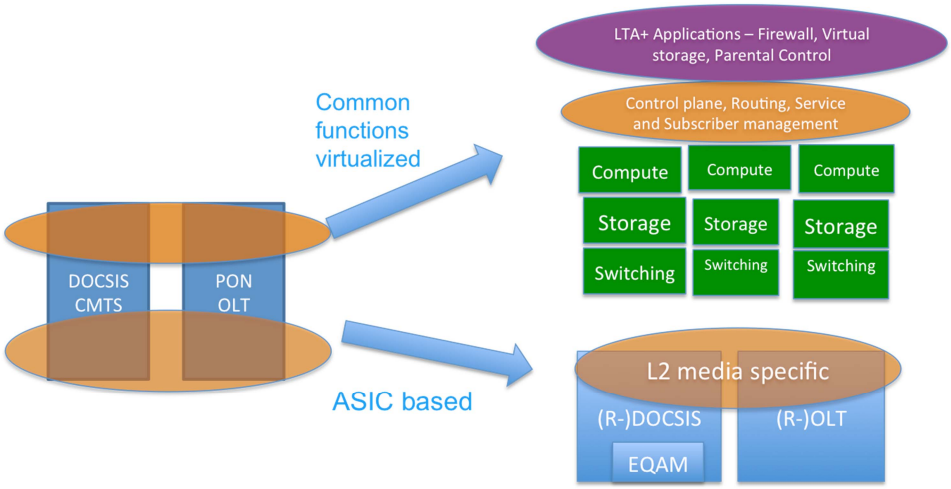


Figure 3: Separating Virtual functions in the Access Network

pack their ASIC based solutions effectively and only migrate their services, control/ management plane to a COTS platform. This will enable both operators and vendors to offer feature parity and reuse of certain components independent of the access technology thereby improving time to market for new services and features.

NFV offers a concept to migrate the hardware appliance based network functions to run on COTS and VM. Recall the example of network security application discussed in an earlier section, MSOs could leverage the NFV technology to run the network security on COTS and VM. The network security application is virtualized in Virtual Network Function (VNF) and deployed inside the network cloud infrastructure. MSOs can use SDN technology to control the flows from the customer to the network security VNF. This enables MSO to flexibly offer network security service in a much faster pace with little upfront investment and minimal physical changes. In comparison, the traditional model requires MSOs to deploy and integrate physical network appliances before serving customers.

<u>High Level Architecture for Virtual Home Network</u>

Figure 4 presents a high level architecture for a Virtual Home Network (VHN). Note, this is a use case study and will likely morph into a different architecture based on our ongoing learning and experiences. For details refer to [4]. This architecture includes:

- a Controller (VHNC) that contains user configurations and policies
- a set of Virtual Network Functions (VNF)
- a Packet Processor (VHNF) that performs functions such as a packet forwarding between multiple VNFs
- a VNF Manager (VNFM) managing the VNFs.

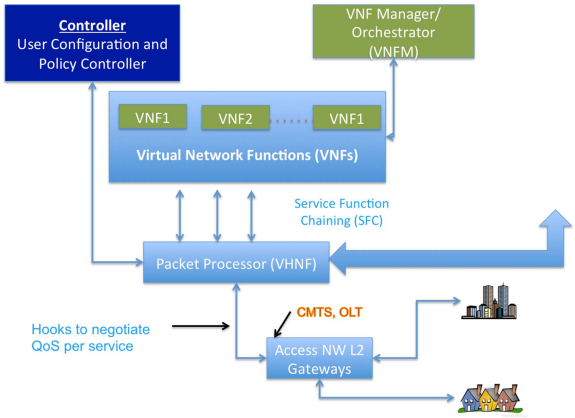


Figure 4: High Level Architecture for Virtual Home Network

There are five key interfaces for control, management and data path:

- The inteface between the Controller and the packet processor is used to exchange configuration and policy. For example, User A's living room TV has Committed Information Rate (CIR) set at 10Mb/s and must be protected by firewall function implemented in VNF.
- The interface between the Controller and VNFs is used to exchange configuration and policy. For example, VHNC could configure the firewall VHF to block any incoming ICMP messages to the User A's living room TV.
- The interface between VNF Manager and VNFs is used to exchange VHN management messages. For example, VNFM could instantiate a new firewall VNF when the current firewall VNF reaches certain capacity.
- The connection between the CPE and VHNF can be a common protocol agreed between CPE and VHNF. It could be Ethernet or any encapsulation technology such as GRE, PMIP or MPLS.
- The interface between the packet processor and VNF is defined by the Service Chaining Function protocol. IETF SFC WG is currently defining the specifications. VNF contains the service definitions and service logic. For example, Virtual Network Function 1 (VNF1) could be a parental control service and manage web filter rules configured by subscriber. Virtual Network Function 2 (VNF2) could be personal firewall that protects a home from botnet and intrusion. MSOs can scale VNFs horizontally to meet user demand. MSOs can also dynamically create VNF per subscriber only when the subscriber wants that service. For example, NSP initiates VNF1 for User X and VNF2 for User Y. In this model, MSO no longer updates CPE for service addition or modification. VHNC stores the user's service subscription. Each user may have different set of home services. For example: User A may have video service. User B may have VoIP service. VHNC contains the user's service subscription and interact with the VNF modules to provide proper services to users. VHNF is usually a device that is optimized for processing packets. It also implements the Service Function Chain function to forward user packets to proper VNFs. CPE is a simple access device that connects to the subscriber's devices at home to the MSO network.

Concerns on Virtual platforms and NFV:

A critically important metric is the packet performance, especially when VNFs are handling a lot of small packets. In a virtualized COTS based environment, there is a virtual switch or router that resides in the host operating system or hypervisor kernel to handle packets in and out of the server and in between Virtual Machines (VMs) in the same server. The virtual switches along with an SDN controller plays a big role in traffic steering and service chaining. Open Virtual Switch (OVS) is a good example that is widely adopted in NFV trials and Proof of Concepts (PoCs). However the packet performance using a purely software based OVS is considered inefficient and incurs high latency. Many vendors are working on a improved virtual switch that improves the efficiency of packet forwarding by up to 10x.

How fast can your physical network elements transmit and receive bit streams?

First, it is the raw speed of the servers and the switches connecting the servers to push packets in and out of the network. In a server, packet I/O is typically handled by a Network Interface Card (NIC). 10Gb/s NICs are very

common but for a high-performance NFV cloud, 40Gb/s or even 100Gb/s will be needed. There are new 25G and 50G standards being developed that provide a wide range of choices [3]. Inside the server, KVM and virtual switching layer play an important role (see question below).

How fast can your system handle packets?

As we currently envision, in a typical aggregation point such as a headend, each serving group in an access network will likely have 10Gbps Ethernet connectivity. Considering this as an example, on a 10G link the theoretical maximum rate of packet performance with 46-Byte packets and 1500-Byte packets is close to 15 million packets per second (pps) and 1 million pps respectively. A critically important metric here is the packet performance, especially when VNFs are handling a lot of small packets. As it turns out, it is a lot harder to achieve lossless small packet handling.

Table 3: Packet Processing Requirements

Packet Data Unit (PDU) in Bytes	Packet Handling Rate on a 10G link (Millions of packets/sec)	Calculations (38 byte IPv4 and Ethernet overhead)
46	14.881	10Gb/s / (84 B * 8 b/B)]
1500	0.813	10Gb/s/(1538 B * 8 b/B)
9200	0.135	10Gb/s/(9238 B * 8 b/B)

How fast can your system handle packets in a virtualized environment?

Even with a good foundation of raw throughput and packet performance, performance can be affected by the components that sit between the VM interface and the server's NIC, OS, and hypervisor. Multiple context switches and memory copies can happen before a received packet is delivered to the destined application (VNF), resulting in sub-optimal packet performance. Several solutions have been developed to over this limitation. With Data Plane Development Kit (DPDK), applications can link to to the DPDK library and call DPDK APIs to keep polling for packets. This eliminates interrupts for packet arrival. This solution can significantly enhance packet performance while still preserving hardware independence. But doing all packet processing in user space and constant polling still poses overhead to the CPU, impacting packet performance, latency, and performance of other applications on the same host.

Another alternative is to use hardware assist to bypass the kernel and use Single Root Input Output Virtualization (SR-IOV) to virtualize a single Ethernet port into multiple lightweight Virtual Functions (VFs) that can be associated with VMs. The communication between VMs and their corresponding VFs are through Direct Memory Access to eliminate lengthy copy operations.

SUMMARY

MSOs are seeing several transitions. We believe NFV and SDN can be used to realize synergies while addressing these transitions. NFV and SDN made a big promise to redefined networking and how MSOs manage and operate their network functions similar to managing and operating software applications. If they execute well to meet their promises, it will rationalize the current networking model, greatly simplify network management and operations. The ultimate success of this architecture will be when MSOs can start launching new applications and services with a short lead time, low upfront cost per user and ability to try out new services similar to the cloud service providers.

NFV and SDN are far from mature, but the potential benefits are enormous if the industry can work together and make it reality. A lot of progress has been made but more work is still required. Vendors are actively proving the feasibility and performance capabilities of running VNFs on COTS platforms. Architecturally, we need to define VNFs orchestration model and Service Function

Chaining among others. Operationally, both vendors and MSOs have to become familiar with the architectures, adopt new processes such as dev/ops to successfully transition.

References:

[1] Growth Architectures: Built to Last, Built to Launch, by Dr. Robert L. Howald. Published in Spring Technical Forum, NCTA 2014

[2] Mobile Apps and Revenue: http://techcrunch.com/2014/06/02/itunes-app-store-now-has-1-2-million-apps-has-seen-75-billion-downloads-to-date/

[3] New 25G and 50G: http://www.bitpipe.com/data/demandEngage.action?resId=1421690810_773

[4] IETF Draft - Problem Statements of Virtual Home Network: http://tools.ietf.org/html/draft-lee-vhs-ps-02, by Y. Lee and R. Ghai.

Designing Resilient DOCSIS Remote-PHY Solutions using High-Availability System Architectures

Daniel Lavender, Principal Architect, CableLabs
Karthik Sundaresan, Principal Architect, CableLabs

Abstract

High Availability (HA) has not been standardized as part of the CCAP and M-CMTS specification efforts and therefore implementation of these capabilities has been interpreted differently by each system implementer. This paper presents scenarios for applying High Availability (HA) systems concepts to the new DOCSIS Remote PHY architecture.

An overview of DOCSIS Remote PHY is presented to introduce the new architecture. Next, to set the context for high availability systems, models and tactics for generalized HA systems are presented. The document concludes by combining the concepts of Remote PHY and HA to suggest models and scenarios for creating High Availability (HA) systems supporting DOCSIS Remote PHY architectures.

DOCSIS REMOTE PHY (R-PHY) CONCEPTS

Traditional DOCSIS CMTS architectures are migrating to distributed architectures. The DOCSIS Remote PHY specifications describe an architecture that moves the PHY layer processing from the CCAP Core or CMTS Core to the edge of the network (e.g., node). This architecture model is generally referenced as Modular Headend Architecture version 2 (MHAv2). Figure 1 presents the high-level DOCSIS Remote PHY Architecture. The terms “Remote PHY” and “Distributed PHY” are synonyms and are used interchangeably throughout the documentation for MHAv2.

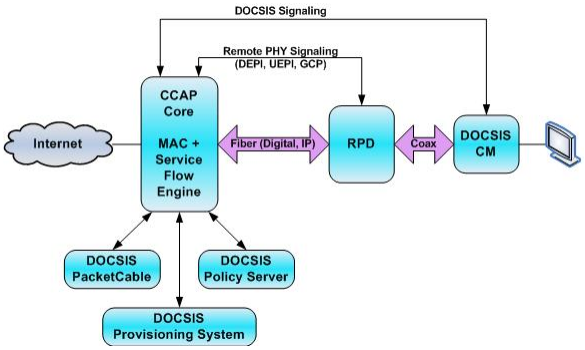


Figure 1 - DOCSIS Remote PHY Architecture
Source: R-PHY D03 specification

In the MHAv2 model, the PHY layer processing is moved from the CMTS/CCAP in the Headend to the node. MHAv2 allows a CMTS to support an IP-based digital HFC plant, meaning the Headend is connected to the Remote node via a digital fiber, a Layer 2 Ethernet link. In an IP-based digital HFC plant, the fiber portion utilizes a baseband network transmission technology such as Ethernet, EPON (Ethernet over Passive Optical Networks), GPON (Gigabit Passive Optical Network), or any layer 2 technology that supports a fiber-based layer 1. One of the common locations for a Remote PHY Device is the optical node device that is located at the junction of the fiber and coaxial plants.

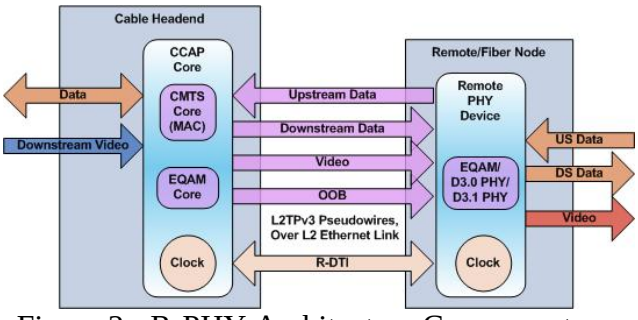


Figure 2 - R-PHY Architecture Components

The Remote PHY Device is connected to the CCAP MAC Layer using L2TPv3 (Layer 3) pseudowires, which tunnel the DOCSIS

payload from the CCAP Core to the Remote PHY device. MHAv2 allows multiple CCAP Cores, Video Cores or OOB Cores to connect to a series of Remote PHY Devices.

The Remote PHY architecture uses a combination of pseudowires for Upstream External PHY Interface (UEPI), Downstream External PHY Interface (DEPI), Out of Band (OOB) data, and timing for signaling. A timing solution referred to as R-DTI is used to provide timing services for functions such as DOCSIS scheduling across the CCAP Core and Remote PHY Device (RPD).

R-DEPI, the Downstream External PHY Interface, is the downstream interface between the CCAP Core and the RPD. More specifically, it is an IP pseudowire between the MAC and PHY in an MHAv2 system that contains both a data path for DOCSIS frames, video packets, and OOB packets, as well as a control path for setting up, maintaining, and tearing down sessions.

R-UEPI, the Upstream External PHY Interface, is the upstream interface between the RPD and the CCAP Core. Like R-DEPI, it is an IP pseudowire between the PHY and MAC in an MHAv2 system that contains both a data path for DOCSIS frames, and a control path for setting up, maintaining, and tearing down sessions. The R-OOB specification outlines multiple approaches to passing OOB (out of band) signals for MPEG Video distribution, through a Remote PHY system.

Remote PHY architecture relies on centralized software in the core—because PHY has minimal complexity, it is well suited to be distributed to the edge and into the node.

Architectural Advantages

The architectural advantages are:

- Because Remote PHY is compatible with legacy HFC plants, it can be deployed incrementally into existing plant architectures.
- Moving the PHY layer to the edge brings full IP closer to the end user (the subscriber), thus reducing complexities. And transitioning to standard IP switching and routing architectures enables simple changes to be made dynamically to delivered services.
- This model enables increased capacity over the HFC network because of SNR gains in the digital optical L2 network. This is beneficial for DOCSIS 3.1 technology deployments and their higher order modulations, which increase the available bandwidth on the network.

Operational Advantages

The operational advantages are:

- Operators are facing a space and power crunch within their Headends; Remote PHY architecture reduces the power consumption and space requirements in the Headend by moving part of the CMTS to the node.
- The digital optics are easier to maintain and simplify plant maintenance for the operator.
- Remote PHY architecture makes it easier for the operator to improve services that matter to the customer.
- Capital expenditures can be spread out over multiple periods of the company's financial year.

The Remote PHY architecture keeps the simple elements (PHY) remote and the complex elements (MAC) centralized. This architecture scales to the needs of the operator. It allows an Integrated CMTS and Distributed-CMTS line cards to be present in the same CCAP chassis.

For an optical access network based on linear optics, the Remote PHY Device is located at the hub. For an optical access

network based on digital optics, the RPD is located at the optical node.

The MHAv2 architecture permits RPDs to be managed by more than one CCAP Core. An RPD is controlled by exactly one "principal" CCAP Core and can communicate with zero or more "auxiliary" CCAP Cores, Video Cores, or OOB cores. An "auxiliary" core manages a subset of RPD resources, e.g., particular channels or RF ports. The principal and each auxiliary CCAP Cores establish their own GCP session and L2TPv3 control sessions with the RPD.

The term "CCAP Core" can refer to either the principal core or an auxiliary core. (Source: CM-SP-R-PHY-D03_150320, section 11.1). Redundant Remote-PHY Devices in the node with additional Remote-PHY failover capabilities in the CCAP Core is beneficial to MSO operations.

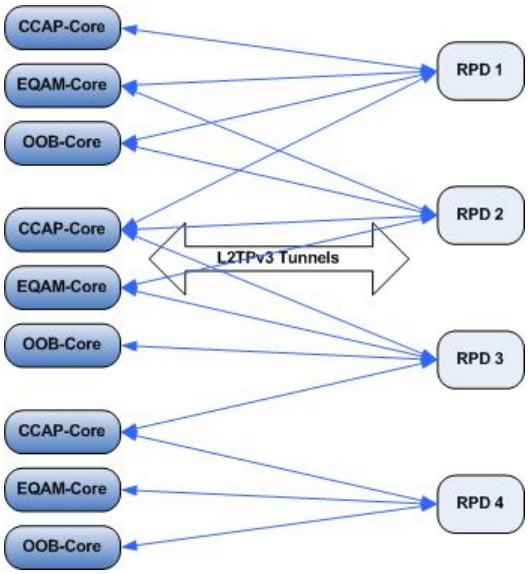


Figure 3 - R-PHY Architecture: Multiple RPDs Controlled by Multiple Cores

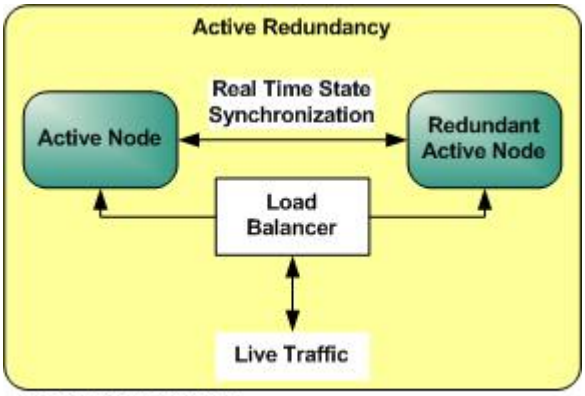
ARCHITECTURES FOR HIGH-AVAILABILITY SOFTWARE SYSTEMS

High Availability (HA) refers to the ability of resources in a computing system to remain available in the event of failures in the system. High availability systems are necessary as they enable business continuity and provide an expected and/or required level of availability to services provided to customers.

Engineering for HA Systems

From an engineering standpoint, key areas of focus to consider when designing high availability systems include:

- Eliminating single points of failure. This includes providing redundant components so that the failure of a primary component does not mean the entire system fails.
- Providing reliable failover between components. Failover between primary and redundant components requires cross-over connectivity and this can become a single point of failure if alternate paths are not included in the design. Four arrangements of redundant components are shown in the following figures (Active, Passive, N+1, Cold).
- Real-time performance monitoring. Detection of failing and failed components in near real time is critical for preventing failures and for reacting quickly to failures with automated recovery and/or manual intervention.

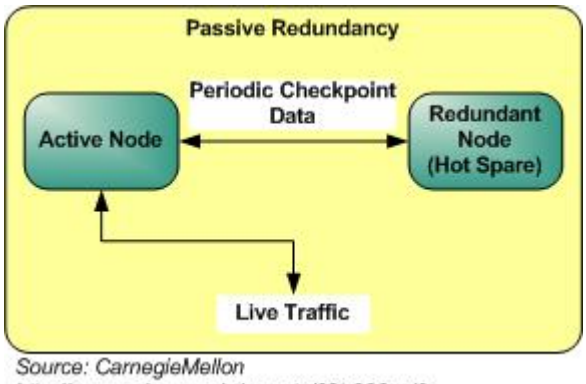


Source: CarnegieMellon
http://www.sei.cmu.edu/reports/09tr006.pdf

Figure 4 - Active Redundancy Failover Pattern

When designing systems conforming to the Active Redundancy Failover Pattern (see Figure 4), all active and redundant components (nodes) receive and process all inputs in parallel. This results in the active and spare components having an identical state at all times. In the event of a failure or urgently needed downtime (for example, to replace a component), this can occur in milliseconds.

Additionally, both the active and redundant components can operate in an online configuration that provides increased capacity. In the event of failure of a component, the system can continue to operate, albeit with reduced capacity. Failover is usually invisible to the clients because they are connected to the server using virtual IP addresses managed by the load balancer.

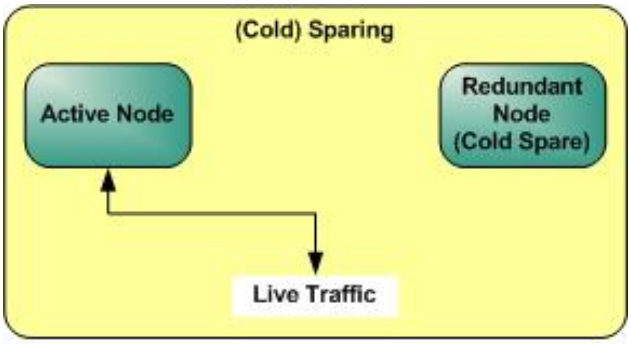


Source: CarnegieMellon
http://www.sei.cmu.edu/reports/09tr006.pdf

Figure 5 - Passive Redundancy Failover Pattern

When designing systems conforming to the Passive Redundancy Failover Pattern (see Figure 5), one of the components receives and processes all inputs. The redundant component, known as a warm spare, receives periodic updates that synchronize it with the active component. In the event of failure, small amounts of data or transactions may be lost due to latency in the synchronization processes.

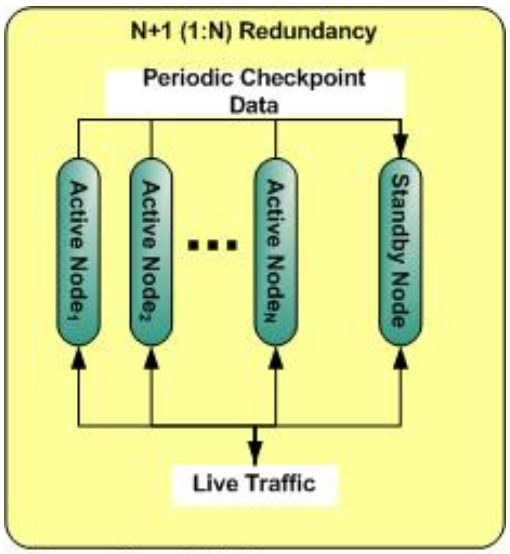
Nevertheless, failover can occur quickly, generally within a few seconds. This pattern provides a more cost effective solution for environments that don't need near real time failover without any loss of data. Failover in this scenario is visible to the users, as they typically must reconnect to the server.



Source: CarnegieMellon
http://www.sei.cmu.edu/reports/09tr006.pdf

Figure 6 - Cold Sparing Redundancy Failover Pattern

When designing systems that conform to the Cold Sparing Failover Redundancy Pattern (see Figure 6), one of the components receives and processes all inputs. The redundant component, known as a cold spare, is updated by applying periodic backups from the active node. These backups may be daily or more frequent, such as every five to ten minutes. In the event of failure, any required backups are applied to the cold spare, the processes on the cold spare are started, and the clients connect to the cold spare. When the active node is repaired, the process is reversed or alternatively, the original active node becomes the cold spare.



Source: CarnegieMellon
http://www.sei.cmu.edu/reports/09tr006.pdf

Figure 7 - N+1 Redundancy Failover Pattern

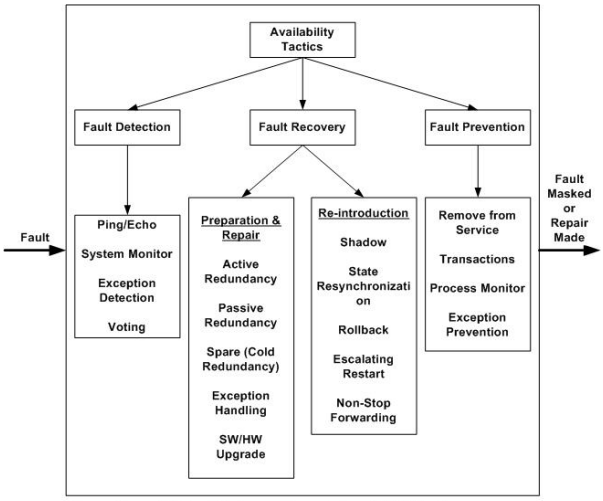
When designing systems that conform to the N+1 Redundancy Failover Pattern (see Figure 7), each of the 1–N active nodes operates as an independent device, receiving input, processing data, and providing output. Each of the active nodes periodically sends checkpoint messages to the standby node so it stays closely synchronized with the active nodes. If one of the active nodes fails, the standby node is immediately configured as the active node and begins processing as if it were the node it replaced. When the failed active node is repaired, it is brought online and the standby node resumes its function as the standby, first by re-synchronizing with the N active nodes.

Operational Considerations for HA Systems

Assuming that redundancy has been engineered into the system, the focus of the business continuity teams include:

- Fault Detection. This includes real time monitoring and oversight of the components with automated processes that detect and initiate failover.
- Fault Recovery. If a component fails, multiple paths may be followed depending on the type of failover (active, passive, N+1, cold spare) that was engineered into the system.
- Fault Prevention. This includes analyzing the monitoring data, constantly checking the system health parameters, or executing automated processes that prevent the occurrence of faults.

Figure 8 presents various approaches that operators can use for fault detection, fault recovery, and fault prevention.



Source: CarnegieMellon - http://www.sei.cmu.edu/reports/09tr006.pdf

Figure 8 - SEI Availability Architecture Tactics

Figure 8 depicts the flow of a fault (on the left) through detection, recovery, and prevention, exiting (on the right) with the fault masked or a repair made.

The following sections present additional descriptive narrative for items that are of primary interest to operators thinking about using DOCSIS Remote PHY systems.

Availability Tactics – Fault Detection

Fault detection is focused on monitoring the health and performance of the system and reporting the health and any anomalies to a control system. From an architecture perspective, isolating fault detection activities from recovery and prevention activities

increases independence of detection so it is not unduly influenced by the other activities. Three key aspects of fault detection are shown below:

- Ping/Echo. Pinging checks for a heartbeat in a system of component within a system. Periodically sending a message and receiving an event from a monitored system is a simple way to determine if a system is responding to input in the way that you expect. Lack of an echo from a ping request can be the first indicator that a system is experiencing difficulties.
- System Monitor. System monitoring is frequently implemented by installing an agent on a system or component that measures health-related operational parameters in real or near real time. If a measured parameter exceeds established thresholds, an exception or error message can be issued to a control system, which then takes recovery or preventative actions. Also, if the repair can be handled locally, the component can take immediate action (on its own) without direction from a controller.
- Voting. Generally used in real time systems that are in sensitive environments or performing mission critical functions. Frequently, three independent implementations of the same function vote on the next function or action and if two of the three voters agree, that action is taken. If the voters don't agree, then an exception can be raised and communicated to the control system.

Availability Tactics – Fault Recovery

Fault recovery is focused on getting the system back to a fully operational state. Four key fault recovery activities are described below:

- Preparation & Repair: Exception Handling. In general, every system fault is the result of a failure (or error) in some portion of the system. A popular approach for capturing the fault and reporting is through the concept of raising an exception. An exception captures an error event and provides additional details to a component or control system that are useful for triage and repair.
- Preparation & Repair: SW/HW Upgrade. This leverages the various failover models by upgrading the non-active system and failing over to the system with the new software/hardware. In the active-active scenario, one of the nodes is taken offline, upgraded and then brought back online.
- Re-introduction: State Resynchronization. This tactic periodically updates the state of the passive and cold spare nodes to the current state. By definition, the current state is the state at the time of a snapshot. Note, that as soon as a snapshot is taken it becomes stale, as the state of the active component is dynamically changing. However, the snapshot provides the ability to recover to a known state. The frequency of taking snapshots depends on the latency that is viewed as acceptable. The amount of latency that is deemed acceptable is often defined in a Service Level Agreement (SLA).
- Re-introduction: Escalating Restart. This tactic acknowledges the need to restart components in a known sequence. For example, a network node may need to be restarted prior to a related video playout server. Another example is restarting controllers before the child devices authenticate.

Availability Tactics – Fault Prevention

Two popular fault prevention activities are described below:

- Remove from Service. Periodically a system or component causes faults in

other systems and components (that is to say, it has gone rogue). In this case, take the offending system offline, perform maintenance and repairs and then bring the system back online. Another possibility is to remove a system from service altogether.

- Exception Prevention. As the saying goes "An ounce of prevention is worth a pound of cure." This means that identifying, isolating, and fixing faults before they negatively impact system performance is always preferable to waiting for downtime and unhappy customers.

APPLYING HIGH AVAILABILITY DISTRIBUTED SYSTEMS MODELS FOR DOCSIS R-PHY RELIABILITY

Combining the concepts of DOCSIS Remote PHY and systems engineering for High Availability leads us to recommendations for creating highly available Remote PHY systems. Keys to fault tolerance in Remote PHY include building protective redundancy into the system and including fault detection, recovery, and prevention functionality as part of the core design and implementation.

These are the different EQAM redundancy types that can be used to define different quality of service (QoS). Note: "Loss of service" means loss of set-top box synchronization.

- Redundancy Type I - User transparent, no loss of service.
- Redundancy Type II - Not user transparent, no loss of service.
- Redundancy Type III - Not user transparent, with momentary loss of service.
- Other – (Not defined in the EQAM doc). Not user transparent, significant duration loss of service.

Cable operators must at least support Type III redundancy: "Not user transparent, momentary loss of service." (For details, refer to CM-SP-EQAM-VSI-I01-081107, section 11.1.2.) These QoS definitions are directly applicable to Remote PHY and are closely correlated to subscriber satisfaction.

The CCAP Technical Report, "High Reliability and Redundancy Capabilities" (CM-TR-CCAP-V03-120511, section 5.2.3), provides guidance on developing HA capabilities that address hardware and network redundancy.

Architecture and Failover Scenarios

EQAM Redundancy Quality of Service types can be directly mapped onto the primary failover architecture models (Active, Passive, N+1, and Cold) described earlier. Table 1 summarizes the mappings and includes one additional variation for each of the active and passive failover architectures. The following sections elaborate on the information presented in the table.

Table 1 - EQAM QoS Redundancy Types and Architecture/Failover Scenario Mapping

EQAM QoS Redundancy Type		Architecture and Failover Scenarios					
		Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 4**	**Scenario 5**	**Scenario 6**
Type	**Description**	**Active-Active**	**Active-Passive**	**Active-Passive Remote**	**Parallel Path**	**N + 1 Redundancy**	**Cold Standby**
I	Transparent, no loss of service	Yes			Yes		
II	Not transparent, no loss of service	Yes			Yes		
III	Not transparent, momentary loss of service		Yes	Yes		Yes	
Other	Significant loss of service						Yes

Earlier, this document presented key architecture patterns that support high availability. These patterns are applied to Remote PHY in the following diagrams. The first five scenarios are preferred. Scenario 6 is included for completeness but is not recommended, as this scenario will result in significant loss of service during manual failover and restart. The scenarios are described below.

Active-Active Failover at the Node

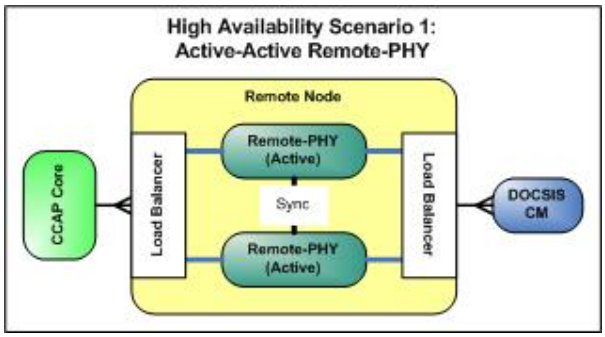


Figure 9 - Remote PHY Active-Active Failover Scenario

In the Active-Active failover scenario, both instances of Remote-PHY process signals and remain synchronized in real time using messaging between the devices. Load and virtualization is managed via the load balancers at both ends of the access point. When a failure occurs, all traffic is routed transparently through the remaining Active device by the load balancers. After the fault is repaired, the second device is brought back online and traffic is again routed through the device.

Active-Passive Failover at the Node

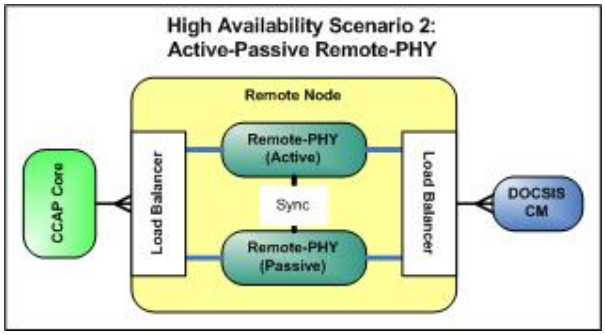


Figure 10 - Remote PHY Active-Passive Failover Scenario

The Active-Passive failover scenario routes all traffic through the Active Remote-PHY device via the load balancers at both ends of the access point. The Passive device is periodically synchronized with the Active device through shared messages. If the Active device fails, all traffic is routed transparently through the Passive device, which becomes the new Active device. After the fault is repaired, the device is brought back online and either becomes the active or the passive device.

Active Remote-PHY Failover at the Core

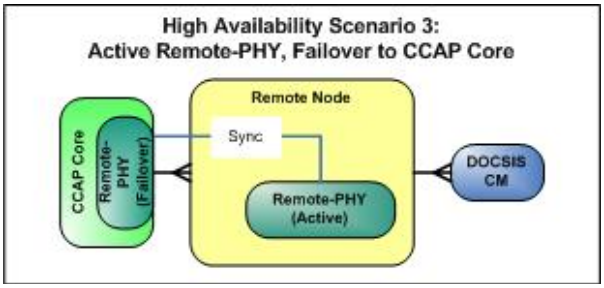


Figure 11 - Remote PHY Active-Remote Failover Scenario

The Active-Remote-PHY failover at the core scenario is a variation on the scenario described in Figure 10 (Active Passive Failover scenario). All traffic is routed through the Active Remote-PHY device. The Passive device is periodically synchronized with the Active device through shared messages. If the Active device fails, all traffic is routed transparently through the Passive device, which becomes the new Active device. After the fault is repaired, the device is brought back online and becomes the Active device.

Parallel Path Active-Active Failover at the Node

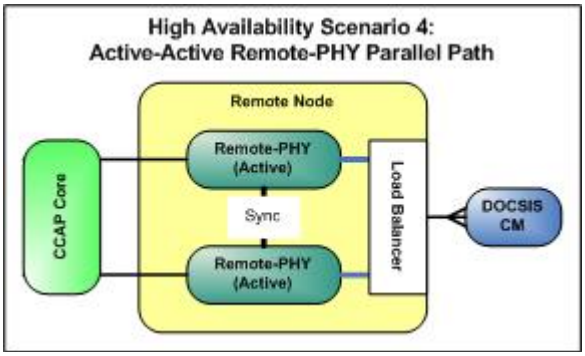


Figure 12 - Remote PHY Active-Active Parallel Path Failover Scenario

The Parallel Path Active-Active failover at the node scenario is a variation on the scenario described in Figure 9 (Active–Active Failover).

In this scenario, both instances of Remote-PHY process signals remain synchronized in real time through messages exchanged between the devices. Load and virtualization with the DOCSIS Cable Modems (CMs) is managed via the load balancers at the node. Between the node and the CCAP Core, link redundancy eliminates the need for a load balancer in the remote node. The CCAP Core in effect acts as the load balancer. If one of the devices fails, the Load Balancer routes all traffic transparently through the remaining Active device. After the fault is repaired, the second device is brought back online and traffic is again routed thru the device.

Active with N+1 Redundancy Failover at the Node

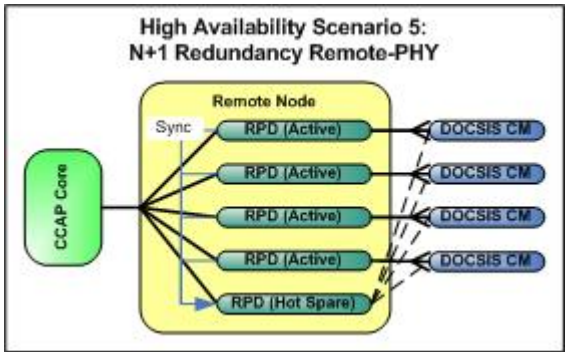


Figure 13 – Remote PHY N+1 Failover Scenario

The Active with N+1 Redundancy Failover scenario uses a hot standby spare (similar to the Active-Passive failover scenario) but instead of a one-to-one active to standby spare, this model has N active devices with a single spare available to all devices. In this example, N = 4 (plus the one spare). The Hot Spare (Passive) device is periodically synchronized with the Active devices through shared messages. If an Active device fails, all traffic is routed transparently through the Hot Spare device, which becomes the new Active device. After the fault is repaired, the device is brought back online and either become the active or the passive device.

Active-Cold Standby Failover at the Node

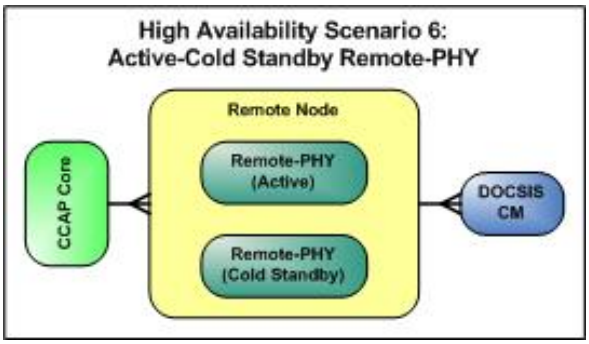


Figure 14 - Remote PHY Active-Cold Failover Scenario

The Cold Standby scenario is included for completeness but is not recommended. This failover model should be avoided because it typically results in hours to days of recovery effort applying backups, configuring the devices and networks and bringing the standby environment online. This model is generally used with batch processes that can tolerate significant downtime. An example of a system that could use this model is a data warehouse that only generates reports sourcing datasets that are loaded on a monthly basis.

SUMMARY

High Availability is directly applicable to the DOCSIS Remote PHY architecture and should be a part of any operator's design. HA is always a balance between competing needs including cost, availability, and latency. The solution you choose will be determined by your various constraints. Active-Active architectures provide the lowest latency but N+1 redundancy is versatile and not terribly expensive to implement. Cold spare redundancy is the lowest cost solution, but has the highest latency and generally the worst availability. In all cases, the balance between

cost, availability, and latency that best meets the business and customer needs is the best solution. Consider all the available options before you settle on a particular architecture.

In the communications services industry, N+1 redundancy is frequency used, especially for business critical application such as customer-facing services like streaming video and data. High Availability is especially important if your customers are business users.

High Availability is a rapidly evolving area and there are many opportunities to incorporate best practice recommendations into your HFC network.

REFERENCES

i. Chapman_John_PPT_MHAv2 SCTE (2013-08-30), John T. Chapman, Cisco Inc.

ii. Remote PHY Specification, CM-SP-R-PHY-D03-150320, Cable Television Laboratories, Inc.

iii. Realizing and Refining Architectural Tactics: Availability, Carnegie Mellon Software Engineering Institute, http://www.sei.cmu.edu/reports/09tr006.pdf

iv. Edge QAM Video Stream Interface Specification, CM-SP-EQAM-VSI-I01-081107, Cable Television Laboratories, Inc.

vi. CCAP Architecture Technical Report, CM-TR-CCAP-V03-120511, Cable Television Laboratories, Inc.

Distributed Trace For Video Systems

Michael Bevilacqua-Linn
Comcast

Abstract

Modern video delivery systems, especially those invoved in IP video, are large scale distributed systems. They are composed of many collaborating subsystems, written by different teams, often in different technologies. Comcast's IP VOD system, for instance, comprises dozens of different subsystems involved with manifest generation, licensing, encryption, caching, advertisement and other alternate content insertion, user interfaces, etc.

These systems have no central coordinator or statefull session manager and are often located in geographically disparate areas. An error can occur in a system several hops behind the player, causing a video to fail play. Correlating that error with the error observed by a user in a highly distributed environment is very challenging, as is determining how much latency is induced by various parts of the system.

This paper describes how a method for distributed trace, based around a protocol and library named Money, an annotation-based distributed trace protocol with roots in Google's Dapper and Twitter's Zipkin. We describe the Money protocol and how we're using it inside of Comcast for our video delivery systems.

INTRODUCTION

Distributed Video Systems

Distributed systems are a fact of life in modern system design. There are several reasons this is so:

- Subsystems can be scaled independently of one another. If built appropriately, this can be done horizontally on commodity hardware.
- Subsystems can be built by small, focused development teams, without the need for much coordination overhead.
- Subsystems can be built using a variety of architectures and techniques, which makes it possible to build a new subsystem using state of the art methods, rather than being weighed down by years (or decades) of history.

Of course, nothing in life is free. One of the primary costs that come with large scale distributed systems is that understanding the system as a whole is extremely challenging.

Developers of an individual subsystem may understand their subsystem and its interfaces, and may even have some understanding of subsystems that call them, and that they in turn call. However, it's unlikely that they could explain the system as a whole.

The below sketch represents a fairly standard, though incomplete, architecture for an IP VOD system. Comcast's full production infrastructure contains many more subsystems, however this sketch is sufficiently demonstrative of the complexities involved.

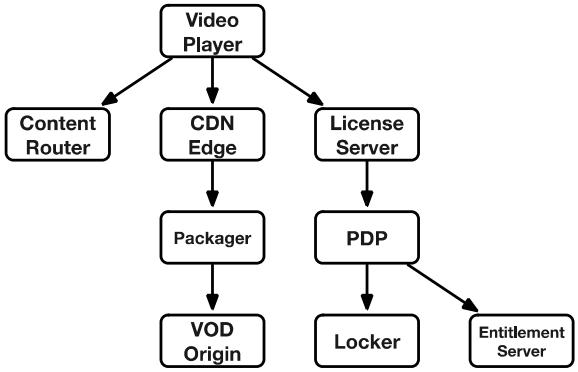


Figure 1 – Sample Architecture Sketch for IP VOD

The sample architecture contains several different classes of subsystems. First, there are players for every platform that video is delivered to, such as STBs, iOS, Android, etc. Second, there's a CDN comprised of a Content Router, Edge Nodes and Origin Nodes.

There's a Packager which produces ABR streams in various formats, and finally content protection infrastructure made up of a License Server, Policy Decision Point, which determines whether a given license request should be allows, a Locker which contains information on purchased assets, and an Entitlement Server which contains information on an account's entitlements.

Generally, no one team understands the system as a whole. Even when an architecture team is involved to provide oversight and higher level design, important, often gritty, details of system to system interaction are poorly understood.

In particular, two questions that are of great interest to engineering and operations teams are fiendishly difficult to answer in such a complicated system. The first is "Why did this video fail to play?", and the second "Why did this video take a long time to start?"

Let's examine failures first. Even assuming that every individual subsystem has a well defined set of error conditions that allow an operator of the subsystem to immediately identify and fix an issue, an assumption that already borders on fantasy, how is does an operator of the overall system determine that a particular failed playback was caused by a subsystem several network hops away from the player?

A common approach is to start by assuming that the player is at fault, forcing the team responsible for operating it to spend time figuring out which upstream system threw the error so that they can pass the bucket of water on. This process repeats itself, going deeper into the system, until the culprit is eventually found.

This is a slow, expensive process that generally involves ticketing systems, conference calls, and the occasional heated conversation.

Attempts to make troubleshooting easier generally follow a few paths. Sometimes, an attempt is made to pass detailed error information from the depths of the system to the front, which forces all subsystems in the path to understand deep implementation details of the subsystems they themselves rely on. This violates the encapsulation of those subsystems, and rarely ends well.

Another, more successful approach is to correlate metrics and log data from various systems using an existing identifier and a rough time period. For instance, it may be possible to pull together telemetry and log data for a given account, device combination over the course of a few seconds, and assume that they were all done on behalf of the same playback session.

The drawbacks to this approach are that they require subsystems to have access to the necessary identifiers, which may not be the case, not every system needs an account identifier to passed to it.

A more complete solution is to have the player create a unique session identifier for a video playback session and pass it through the call chain. Each subsystem then attaches the session identifier to all emited operational metrics, so that they can be correlated by data processing systems.

With appropriate data collection systems in place, this gives system operators a complete view across all the system as a whole, and makes it possible to determine where in the call chain an error occurred. However, a single session identifier makes it difficult to understand anything about individual service to service interactions within the session, they must be represented in an ad-hoc manner in emitted metrics, or known a-priori by operators.

This becomes more import when we're trying to understand where latency is induced in video is induced in video startup. While it's easy to record an overall timing from the player, it's difficult to understand which subsystems interactions are responsible for inducing latency, and how much, even if the individual subsystems are performance tested and their latency characteristics are well understood.

In real world situations, . In order to do so, timing metrics must be collected from every client-server interaction, from both the client and server sides. A single session identifier is insufficient for that level of granularity. In the next session, we'll examine a mechanism for tracing requests through a distributed system that does provided the needed granularity.

Tracing a Distributed System

The sequence of events in any complex request through a distributed system can be modeled as a graph of interactions between the individual subystems. Using the reference architecture in Figure 1, starting a session would looke something like Figure 2.

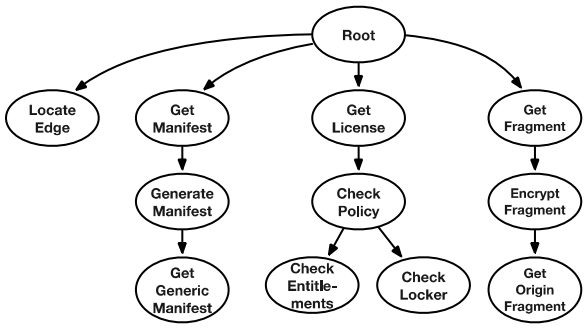


Figure 2 - Partial Call Graph For IP VOD

Here, every node in the graph represents an interesting interation between two subsystems, such as the player locating an appropridate CDN edge node from the content router, or the license server checking a Policy Decision Point to decide whether or not to allow a license request.

The goal of a distributed trace applied to IP video is to reconstruct the graph of calls that takes place to satisfy a particular video session so that useful telemetry can be attached to the individual nodes in the trace. In figure three, we add timing information in to the .

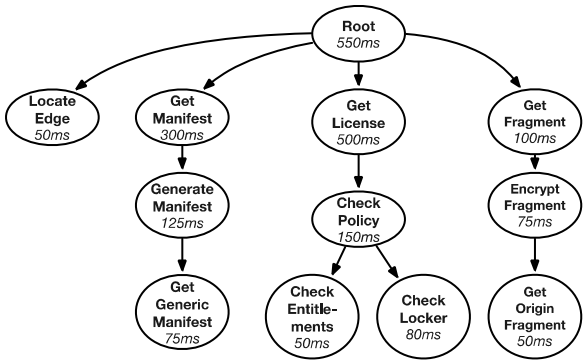


Figure 3 - Distributed Trace With Timings

Traces and Spans

The main abstractions in a Money distributed trace are Traces and Spans.

A trace is generally scoped to all operations and interactions that take place to satisfy some use interaction, in our case, starting a video and the ensuing series of fragment requests that take place to keep it going. The call graph in Figure 3 represents an entire trace.

A span represents some interesting interaction in a trace, generally between two subsystems in the distributed system, and associated metadata. In Figure 3, spans represent calls between systems which collaborate to start VOD playback.

A Minimal Implementation

The core of the Money protocol is small.

Trace identifiers are strings, which are expected to be GUIDs such as those specified in RFC-4122. Span identifiers are 64 bit signed integers, which are expected to be psuedorandomly generated at the creation of a new span.

Trace and Span identifers are passed from system to system on an HTTP header, `X-MoneyTrace`, which takes the following form, linebreaks are inserted for clarity:

```
X-MoneyTrace:
trace-id=$TRACE-ID;
parent-id=$PARENT-ID;
span-id=$SPAN-ID
```

Here, the `trace-id` field contains the GUID which identifies the overall trace, `trace-id` contains the 64 bit integer which identifies the current span, and `parent-id` identifies the parent of current span. It must be included so that a casual relationship between the two spans can be maintained throughout the trace.

The subsystem at which the trace is rooted, which will always be the player in our case, creates a Trace identifier. It then creates Span identifiers and corresponding an `X-MoneyTrace` headers for every request made to second level systems. For spans at the root, the `parent-id` and `span-id` are set to the same value, to indicate that the span is its own parent.

As the request progresses through the system, each subsystem must parse the `X-MoneyTrace` header and make its identifiers available as context for the subservice. When one subsystem makes a request to another, it creates a new Span identifier to encompass the new span, and creates a new `X-MoneyTrace` header with the appropriate Span and Trace identifiers.

On the return path, the `X-MoneyTrace` header is passed back as a response header. This makes it easy for systems which cache HTTP responses to keep track of the Trace that originally generated the response.

This small protocol is all that's required to participate in a Money distributed trace.

Data Collection and Processing

Once distributed trace data is being passed through the system, it can be used to tag existing log messages, or existing application level telemetry. We have found it useful to collect the following base set of telemetry across applications.

Metric	Description
span-id	ID for the current span, a Long
trace-id	ID for the current trance, a GUID of some sort

parent-id	ID for the parent, a Long
span-name	Operator readable name for the span
start-time	Start time for the span, logged as GMT
http-response	If the span encompasses an HTTP request/response, the response code
span-duration	The duration of the span, in microseconds
response-duration	If the span encompasses an HTTP request/response, the response duration, from the point of view of the application doing the logging
error-code	When the span did not complete successfully, an error code must be populated here
span-success	Whether or not the span completed successfully, the string “true” or “false”

A good starting point for metrics collection is to use existing application logs and emit a single log line containing span metadata as a set of key value pairs.

CONCLUSION

Modern IP video delivery systems tend to be highly distributed and complex. Operating such a system is fundamentally different from operating many traditional video delivery systems in that there is no centralized orchestrator that has a full view of the system.

We can get this overall view of the system using a simple distributed trace protocol to pass trace context through the system, which can then be used to tag existing telemetry and log data.

REFERENCES

Google Dapper - http://research.google.com/pubs/pub36356.html

ESTIMATING DOWNSTREAM PERFORMANCE AND DOCSIS 3.1 CAPACITY IN CAA AND DAA SYSTEMS

Michael Emmendorfer, Brent Arnold, Zoran Maricevic, Frank O'Keeffe, and Venk Mutalik
ARRIS

Abstract

Recently the most often asked questions by cable operators around the world surround the expectations of their "future" network performance and capacity when using DOCSIS 3.1 technology in CAA (Centralized Access Architecture) and DAA (Distributed Access Architecture) systems. Answering these questions can be complicated because there are so many different factors that must be examined to predict the "future" end-of-line (EoL) performance and resulting DOCSIS 3.1 capacity. Since MSO-to-MSO network architectures can vary widely, and even within a cable operator's own network architectures may vary, it is important that an understanding of performance impacting factors is identified to effectively predict EoL performance and DOCSIS 3.1 capacity.

This paper and underlying model will be the first published analysis of its kind, because it considers both current and future factors that determine end-of-line (EoL) performance, DOCSIS 3.1 capacity and/or CNR margin. The model is a comprehensive analysis based on real-world data of existing systems as well as predicting the performance of future DOCSIS 3.1 systems not yet created, all in an effort to estimate Downstream Performance and DOCSIS 3.1 Capacity in CAA and DAA Systems.

The model and paper identify four core areas impacting EoL performance: 1) Network Element and Configuration Settings, 2) Network Architecture and Topology, 3) Access Architecture (i.e. CAA & DAA), and 4) Network Condition and Performance.

These four areas influence EoL performance and ultimately D3.1 capacity and/or operator margin and will be examined, in detail in this paper. Specifically, the network element configuration settings such as transmitter MER, RF input levels, partial and full D3.1 spectrum loading, spectrum placement, and spectrum utilization of 750 MHz up to 1218 MHz. The network architecture and topology area examines parameters such as distance between facility and node, wavelength count, amplifiers cascade count, and CPE connection point. Access architectures such as CAA using amplitude modulated (AM) optics vs. future DAA using digital optics systems are examined. Network conditions that can degrade the performance and happen at any time and at any segment are also covered.

This paper will unveil for first time:

1) *The results of a new model, created by the authors of this paper, that predicts future downstream end-of-line (EoL) performance and DOCSIS 3.1 capacity in CAA and DAA systems.*
2) *That "all of the following factors" matter in predicting EoL performance including: 1) Network Elements and Configuration Settings, 2) Network Architecture and Topology, 3) CAA and DAA, and 4) Network Condition and Performance.*
3) *The variance in the maximum downstream EoL amplifier performance and DOCSIS 3.1 modulation possible, when operating CAA using AM optics with different network architectures and configurations.*
4) *Where CAA and DAA performance are differentiated, as this will be an option available later this decade.*

INTRODUCTION AND OVERVIEW

This paper sets out to perform the ambitious goal of estimating the future performance and capacity of DOCSIS 3.1 technology in CAA (Centralized Access Architecture) and DAA (Distributed Access Architecture) systems that in some areas do not exist. Specifically, our challenge was to develop an innovative and comprehensive model based on real-world data of existing systems, where available, as well as predict the performance of future DOCSIS 3.1 systems not yet created. The comprehensive results documented in this paper, for the first time in our industry, yield predictions or estimates of downstream performance and DOCSIS 3.1 capacity in CAA and DAA systems. We undertake this challenge because ever since the introduction of DOCSIS 3.1 technology the industry has wondered what level of performance and capacity could be achieved.

Recently, the industry is considering placement of DOCSIS 3.1 in two different classes of access architecture, CAA and DAA systems. Our industry is now questioning what level of performance and capacity could be achieved when using DOCSIS 3.1 technology in CAA using Amplitude Modulated (AM) optics versus future DAA using digital optics systems. The model can also predict where CAA and DAA performance are differentiated.

Finally, the model and paper define a set of network element configuration settings, network architecture and topology; access architecture use cases (i.e. CAA & DAA), and network condition and performance that are representative of the wide range of cable operator networks worldwide. The model and paper define a set of parameters that nearly every cable operator should find representative of their network architecture and topology that will then provide a prediction of their particular downstream performance estimate and possible DOCSIS 3.1 capacity in CAA and DAA systems.

Predicting or estimating the future performance of end-to-end systems and technologies that do not yet exist is difficult. We feel this is a useful undertaking by providing the most current information available in an effort to consider the potential of the downstream performance and DOCSIS 3.1 capacity in CAA and DAA systems. There is a lot of ongoing research studying bandwidth capacities for DOCSIS systems, performance end-to-end, and transitions strategies, please refer to additional areas of research found in the references [ALB] [CLO], and [MUT]. In the future, as complete DOCSIS 3.1 systems become available, we will likely publish other reports based on end-to-end lab results and ultimately real-world network deployment measurements (not just estimates as found in this paper).

The creators and authors of the model and this paper note the information provided is for educational purposes only and do not recommend sole reliance on this information for investment and operational decisions. The model values are predictions based on the combination of current and future product performance estimates, and may not represent actual performance. The information provided supersedes any previously issued estimates. The authors and ARRIS reserve the right to change these performance estimates without notice.

DOCSIS 3.1 OVERVIEW

This section will not review the details of DOCSIS 3.1 technology but rather cite some of the major technology benefits that will be used to maximize network capacity.

The section will highlight some of the requirements to support a given modulation scheme. If the reader would like to learn more about the features and functions of DOCSIS 3.1 technology, please review the CableLabs® DOCSIS 3.1 specification initially released in 2013, and subsequent releases in 2014. Additional presentations and papers published prior to the start of the DOCSIS 3.1 program defined many of the key attributes found in the current DOCSIS specifications. ARRIS released a series of presentations and papers published in the beginning of 2011 and the final release was February 2012, which defined many of the core features of what later became DOCSIS 3.1. [EMM1] [EMM3] In May 2012 at the NCTA conference, a paper published by John Chapman (Cisco), Michael Emmendorfer (ARRIS), Rob Howald (Motorola), and Shaul Shulman (Intel), referred to as the joint paper also called for another version of DOCSIS. [JOINT]. This paper too defined many of the core features for what later became DOCSIS 3.1.

DOCSIS 3.1 enables five core features that will allow cable operators to maximize, expand, and optimize network capacity. These are as follows:

- Adds Multicarrier Modulation Technology
  - Adds downstream OFDM (Orthogonal Frequency-Division Multiplexing)
  - Adds upstream OFDMA (Orthogonal Frequency-Division Multiple Access)

- Adds Error Correction Technology
  - Low-density parity-check (LDPC) codes (inner FEC) and Bose-Chaudhuri-Hocquenghem (BCH) codes (outer FEC)
  - Enables more spectrum capacity compared to legacy DOCSIS in similar SNR/MER

- Adds Multiple Modulation Profiles (MMP)
  - Allowing groups of customers to operate at highest capacity

- Expands Modulation Formats
  - Downstream up to 16384QAM and Upstream up to 4096QAM

- Expands Cable Spectrum Band Plan
  - Downstream may occupy 54 - 1218 MHz or up to 1.7 GHz
  - Upstream may extend to 5 - 204 MHz

The existing cable network downstream and upstream performance can support higher order modulation formats than those available today. [EMM1] [EMM2] [EMM3] The support of higher order modulations with the existing network may not be ubiquitous across the MSO footprint, or even within a serving group, as some segments of the network will differ in performance. [EMM4] This paper will show the use of higher orders of modulation to obtain more capacity with DOCSIS 3.1 over the “existing” Optical and Coaxial network. [EMM4]

Since not all users can utilize the same order modulations, the introduction of the use of multiple modulation profiles (MMP) is important. [EMM5] The use of MMP will allow groups of users the ability to reach the highest order possible, so that the network as a whole may be optimized and to maximize capacity and b/s/Hz. [EMM5] DOCSIS 3.0 and prior releases limited cable operators from increasing downstream and upstream network capacity or bits per second per Hz (b/s/Hz) for several reasons; first there was a

limit set to 256QAM modulation and no support for MMP. [EMM4]

Key Takeaway Regarding DOCSIS 3.1

The authors believe that the DOCSIS 3.1 era should reexamine our industry approach determining the minimum performances allowed for 64QAM and 256QAM and adding six dB of margin. Our industry essentially cares about meeting a minimum target, because the network was only as good as the weakest link; the worst value determined the modulation order for the entire segment.

The authors believe that the DOCSIS 3.1 era will allow us to look at the "maximum" performance of the end-of-line, and not just the minimum value of the weakest link. This is because DOCSIS 3.1 does not require all users to use the same modulation orders, as was the case with all previous versions of DOCSIS. The use of DOCSIS 3.1 will introduce a new tool called multiple modulation profiles (MMP). This is an important addition to DOCSIS and this should also change the way our industry looks at network performance and capacity planning; this may have a profound impact on overall operations. The use of MMP will allow groups of users the ability to reach the highest order possible, so that the network as a whole may be optimized and to maximize capacity and b/s/Hz. We should no longer assign a large value of six dB of headroom and should no longer care only about minimum performance values of the network.

This paper suggests the use of higher orders of modulation to obtain more capacity is possible with DOCSIS 3.1 over the existing optical and coaxial network and may increase as the cable operators evolve their network to smaller service groups or uses of different access architectures, as discussed in the next section. The authors also suggest DOCSIS 3.1 use of OFDM, new FEC, expanded modulations orders, and the use of multiple modulation profiles (MMP) are important features that will also enable operators to reach the maximum capabilities of the network.

ACCESS ARCHITECTURE OVERVIEW

This section will describe centralized and distributed access layer architectures that represent the location of the access layer network elements. [EMM2] If all intelligent network elements of the access layer reside in the MSO facility like a headend or hub, then this type of system will be called centralized access layer architecture. However, if any of the intelligent network elements of the access layer are located in the outside plant or MDU location, then this type of system will be referred to as distributed access layer architecture. [EMM2] Today, the industry simply refers to these two different classes of network access architecture as Centralized Access Architecture (CAA) and Distributed Access Architecture (DAA).

Centralized Access Architectures (CAA) using Amplitude Modulated (AM) Optics

The centralized access layer architecture requires the MAC and PHY functions to reside at a cable operator facility. CAA allows the OSP, such as nodes, to be relativity simple devices and the network is in many ways transparent.

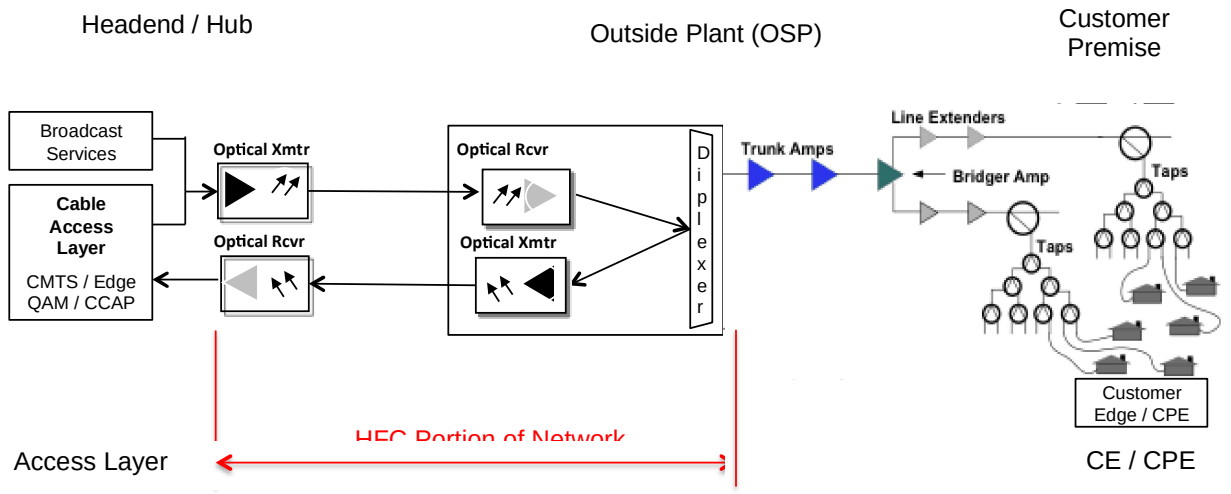


Figure 1: Centralized Access Architecture and HFC

Figure 1 is an illustration of an access layer network element over a transparent Outside Plant (OSP) to the customer edge / CPE with the HFC portion of the network highlighted. The HFC uses two technologies of optical transport in the return; amplitude modulated (AM) also referred to as analog return path and digital return, which may commonly be referred to as Broadband Digital Return (BDR), or simply Digital Return.

Amplitude Modulated (AM) or analog, is the optical technology used for optical forward path transmission of cable signals downstream. The advances in analog forward laser technologies enable transmission of the 54-1002 MHz of spectrum this is over 150 channels, each 6 MHz wide. This will expand to meet the DOCSIS 3.1 1218 MHz requirement. The forward path is layer 1 media converter style architecture and the optical transmission may be shared with multiple HFC nodes and may also carry many technologies transparently. There are two network architectures for the forward: Full Spectrum (illustrated in figure 2); and another called QAM Narrowcast Overlay, or simply Narrowcast Overlay (figure 3).

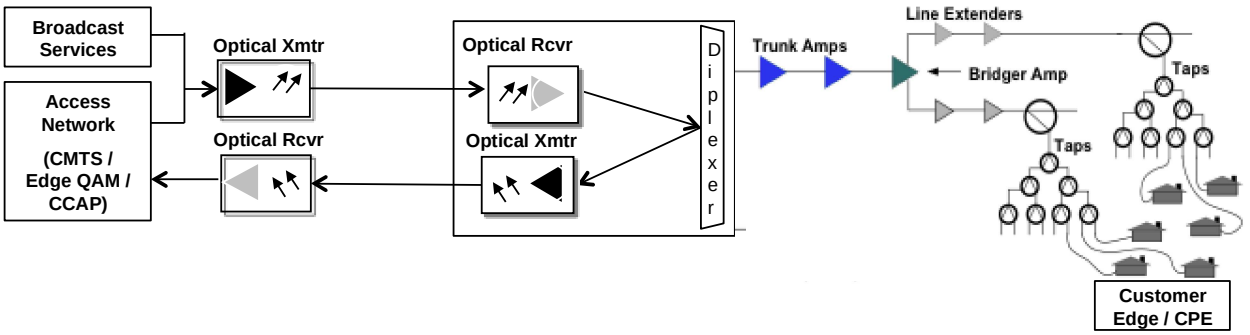


Figure 2: Hybrid Fiber Coax (HFC) with Full Spectrum and Node +N

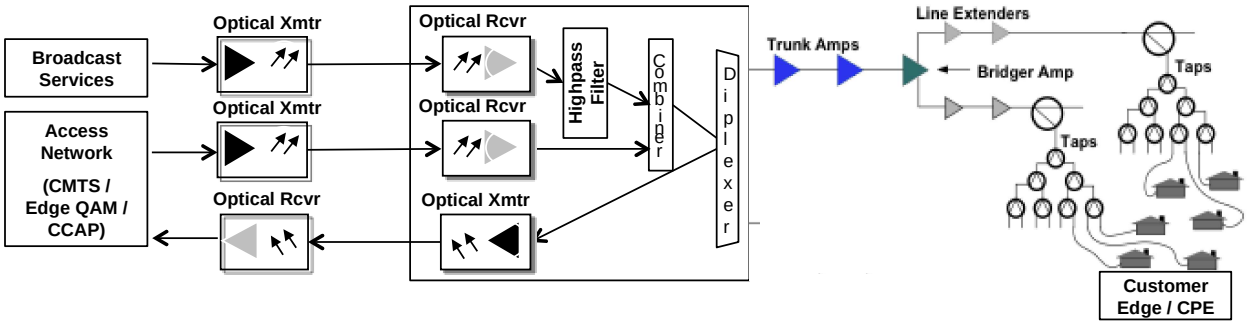


Figure 3: Hybrid Fiber Coax (HFC) with QAM Narrowcast Overlay and Node +N

The MSO serving area between headend and node will be in most cases less than 40 km; therefore this will be easily supported with HFC architecture. The support for extremely long distance to and from the node may be a factor for the HFC. The optical capabilities of HFC have many dependencies, variables, and trade-offs to determine the HFC optical link performance.

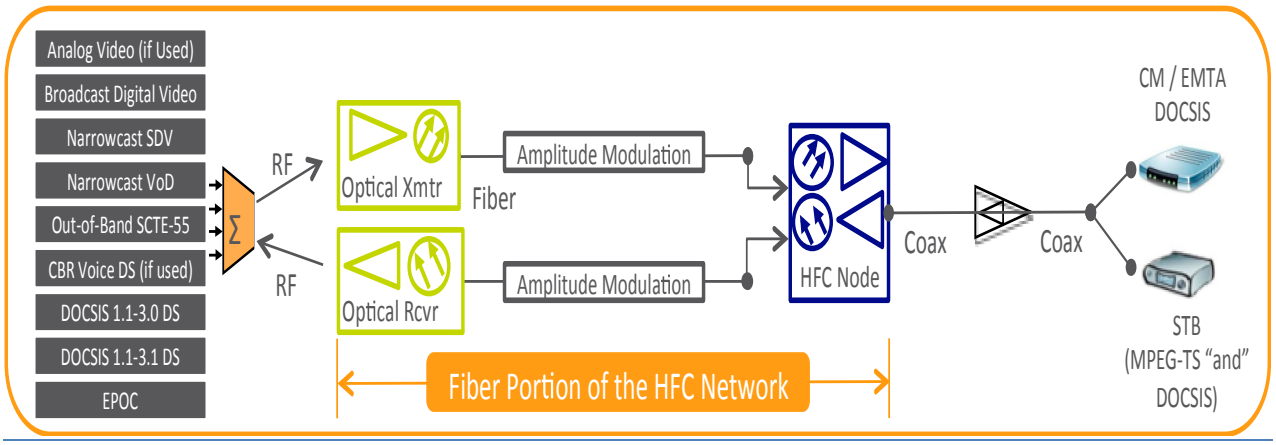


Figure 4: Amplitude Modulated (AM) Optics Flexibility

Amplitude Modulated (AM) core benefits include: (shown in figure 4)

- Transparency and flexibility of the MAC/PHY it carries
- Places least complexity in the node
- Enables CAA keeping complex software and hardware in the headend

Summary of Amplitude Modulated (AM) Performance Challenges and Limitations:

- CNR performance degrades with distance (Facility to Node)
- CNR degrades with when adding wavelengths
- CNR degrades with higher spectrum 750 MHz to 1.2 GHz
- CNR performance varies at different spectrum bands
  (so does remote PHY) below 700 MHz has best End-of-Line
- Partial Loading of D3.1 performs better than Full Spectrum
- Operationally requires rebalancing (in narrowcast overlay

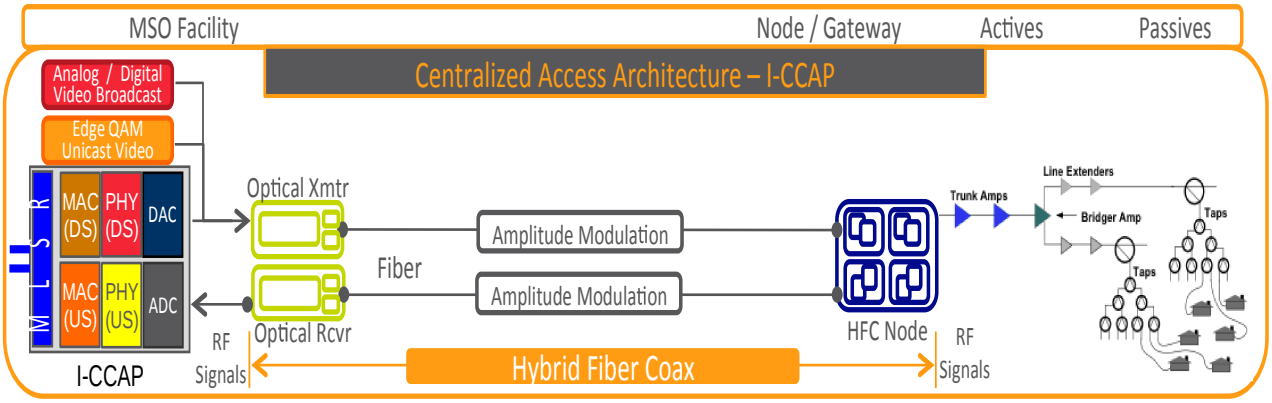


Figure 5: Centralized Access Architectures (CAA) using Amplitude Modulated (AM) Optics

Figure 5 illustrates the Centralized Access Architectures (CAA) using Amplitude Modulated (AM) optics used in today's cable network. This depicts the DOCSIS MAC and PHY layers in the I-CCAP located at the MSO facility and the transport of other services and technologies implemented in different network elements all connected and transported across hybrid fiber coax use of amplitude modulated optical transport.

Figure 6 identifies optical impairments that can impact AM optical systems. We have highlighted several areas from non-linear such as those found in single wavelength and multiple wavelength systems. The optical linear impairments include the fiber linear effects and impairments due to optical passives. In the 2013 NCTA paper we analyzed the Amplitude Modulation (AM) optics challenges and causes that impact overall performance; please refer to that paper for greater detail and understanding of the optical impairments found in single wavelength and multiple wavelength AM optical systems. [EMM4]

Highlights of the Optical Impairments

Single Wavelength or More

- SBS (Stimulated Brullion Scattering)
- SPM (self phase modulation)

Multiple Wavelength Challenges

- SRS:
  - Creates RF crosstalk. Most severe with large wavelength spacing and low RF frequencies
  - SRS induced CSO is a relatively minor secondary effect primarily when the BC load is maximum
- XPM:
  - Creates RF crosstalk
  - Most severe with small wavelength spacing, long fiber links and high RF frequencies
- 4WM:
  - Creates OBI between beats and the carriers
  - If unresolved impacts CCN
  - If severe impacts BER
  - Optical passives: Create CSO due to passives slope (dB/nm)

Long Distance Challenges and Support for High Frequencies

- Fiber Dispersion:
  - Creates composite second order (CSO)
  - CSO increases with longer fiber links and higher frequencies
- Optical Link budget: AWGN

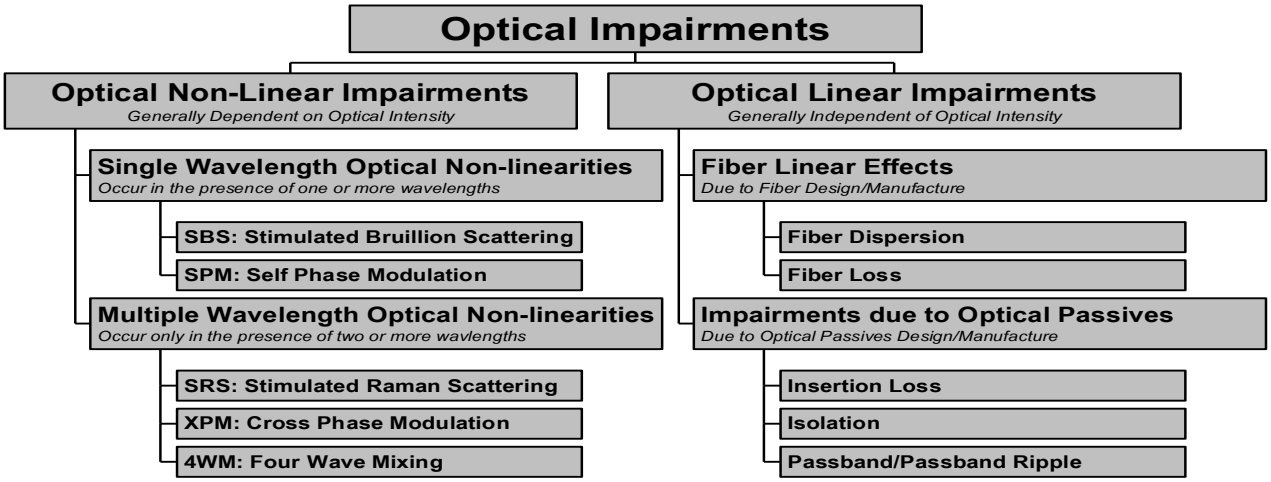


Figure 6: Identifying Optical Impairments for Amplitude Modulated (AM) Optics

Important Takeaway Regarding CAA using AM Optics

We compare single wavelength systems, multiple wavelength systems, several headend to node optical distances, and different amount of frequency or spectrum loading when using AM optics. Essentially the challenges we identified with CAA using AM optics above are examined in this paper. This downstream model and analysis measures the impacts of CAA using AM optics considering a vast set of use cases that represents the majority of MSO networks today and in the future. The differences in performance between CAA using AM optics and DAA using digital optics are documented in this paper.

Distributed Access Architectures (DAA) using Digital Optics

There is another Cable FTTN network architecture class that is not an HFC architecture or technology. This will extend the IP/Ethernet delivery network beyond the hub location to the node (or even MDU) where PHY layer or MAC/PHY layer processing would occur. “Distributed Access Layer Architecture” is when PHY layer or MAC/PHY layer processing takes place outside the headend or central office facility; this processing would take place in the Node, Cabinet, or Basement of MDU. [EMM2] Today, the industry simply refers to this as Distributed Access Architecture (DAA).

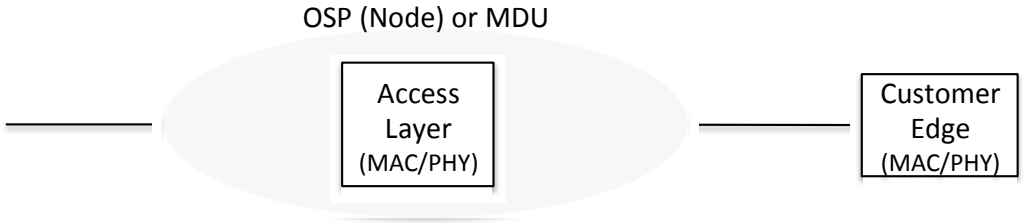


Figure 7: Distributed Access Layer Reference Architecture

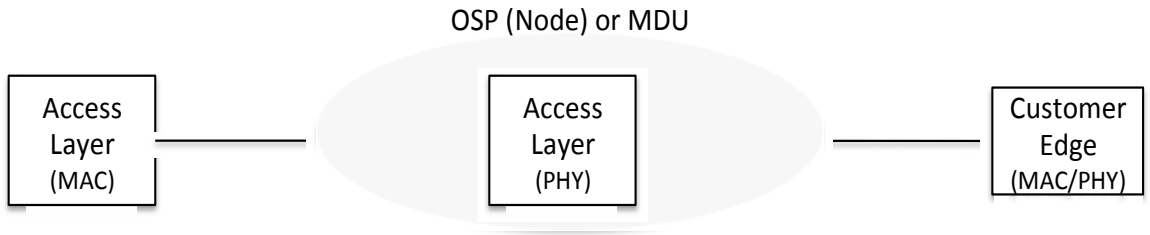


Figure 8: Partial Distributed Access Layer Reference Architecture

Figure 7 is an illustration of an access layer network element that is distributed in the OSP or MDU basement location; this moves the access layer closer to the customer edge/CPE. This architecture does not use the HFC optical network. However, it will use the coaxial cable network, going through amplifiers and passives. In figure 8, just a portion of the access layer is placed in the outside plant or MDU basement. [EMM2]

Moving from AM optics to digital optics for fiber-to-the-node (FTTN) will force us to place PHY or MAC/PHY access layer functions in the node. The use of digital optics is required and this will place new functions in the node and add or remove functions from the headend. Figures 9, 10, and 11 illustrate the functional layers and building blocks of MPEG-TS and DOCSIS MAC and PHY function, as these functions may be split between the headend and node in the future.

The industry terms and definitions used below for Distributed Access Architecture (DAA) were defined in previous papers. [EMM6] Excerpts below are from the "Side-by-Side Comparison of Centralized vs. Distributed Access Architectures" published at 2014 NCTA. They describe the three leading approaches for DAA. [EMM6]

1) Remote PHY (R-PHY): This places the full PHY layer including the FEC, symbol generation, modulation, and DAC/ADC processing in the node. This is analogous to the Modular Headend Architecture (MHA), but different in many ways; such as timing and support for extreme separation of the MAC and PHY layer and support for DOCSIS 3.1 would have to be written. This approach could be called Remote PHY Architecture (RPA). Please refer to Figure 9.

2) Remote - Access Shelf (R-AS): Places the entire "Edge QAM" MAC and PHY layer functions in the node. Video security and encryption may or may not be placed in the node. The lower DOCSIS MAC functions for scheduling and the entire PHY functions are placed in the node. This could be referred to as the R-AS. The M-CCAP packet shelf remains in the headend and performs the DOCSIS upper MAC functions while the M-CCAP Remote Access Shelf performs Edge QAM MAC and Lower DOCSIS MAC functions. Please refer to Figure 10.

3) Remote CCAP (R-CCAP): Places the entire upper and lower MAC and PHY layer functions in the node. This places the CMTS, Edge QAM and CCAP functions into the node. Please refer to Figure 11.

Remote PHY (R-PHY) Architecture

In figure 9 please refer to the definition above called Remote PHY (R-PHY). The architecture of using a CCAP MAC Shelf with a Remote PHY could be called Remote PHY Architecture (RPA), as this resembles in some ways the Modular Headend Architecture (MHA) defined by CableLabs.

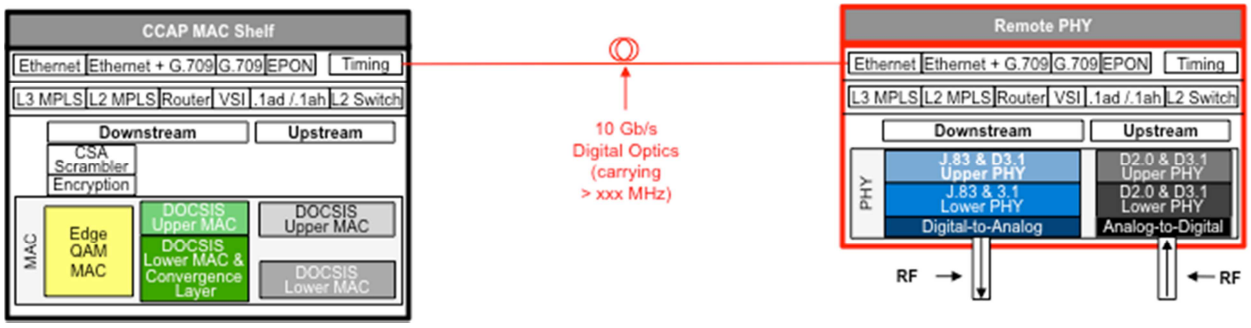


Figure 9: Remote PHY Architecture (RPA) Detailed View

Remote Access Shelf Architecture

In figure 10, please refer to the definition above called Remote Access Shelf (R-AS). This is very similar to the Modular CCAP architecture that defined a Packet Shelf containing the DOCSIS upper MAC functions and the Access Shelf (AS) containing the DOCSIS lower MAC and full PHY functions.

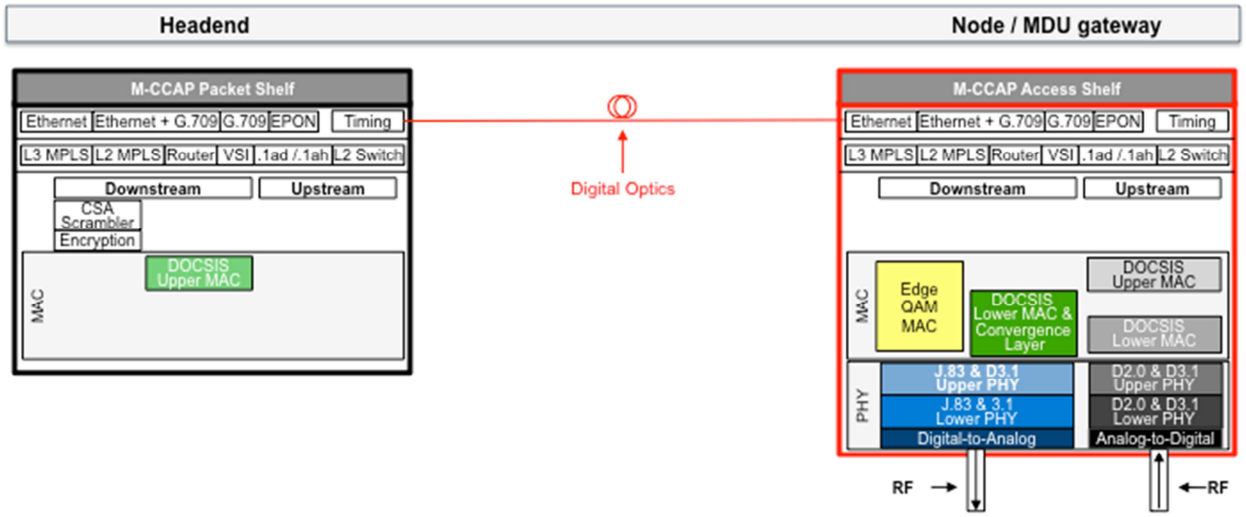


Figure 10: Remote Access Shelf (R-AS) Architecture Detailed View

Remote CCAP Architecture

In figure 11 please refer to the definition above called “Remote CCAP (R-CCAP)”. This is the entire CCAP in the node minus the CSA Scrambler and Video Encryption.

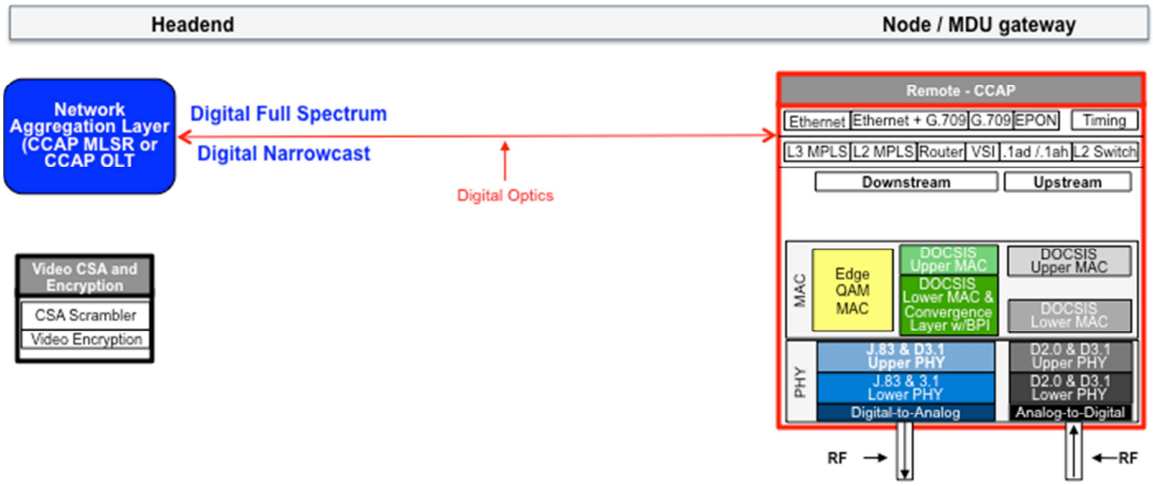


Figure 11: Remote CCAP (R-CCAP) Architecture Detailed View

The two illustrations in figures 12 and 13 represent the Remote PHY and Remote CCAP and for the purposes of this paper these are grouped together to be referred to as Remote Gadget.

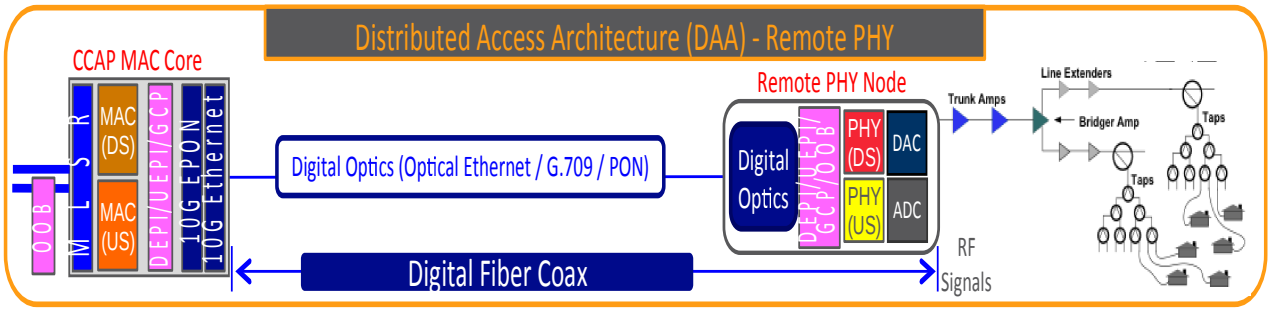


Figure 12: Remote PHY (R-PHY)

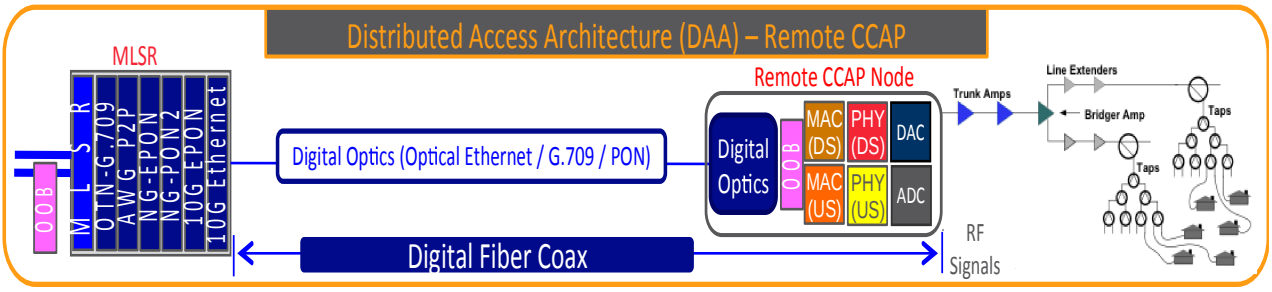


Figure 13: Remote CCAP (R-CCAP)

<u>Important Takeaways Regarding DAA using Digital Optics</u>

DAA will use a digital optical link to the node avoiding the use of amplitude modulated optics and the noise contribution that it adds to the overall cable system's end-of-line (EoL) performance. The figures above represent Remote PHY (R-PHY) and Remote CCAP (R-CCAP). Later in this paper our analysis will define these architectures as simply DAA or Remote Gadget. In either case digital optics between facility and node are used removing the noise contribution found in

CAA use of amplitude modulated optics. The relevance of using digital optics vs. amplitude modulated optics is a core part of this paper.

UNDERSTANDING THE PERFORMANCE IMPACTS TO END-OF-LINE (EOL)

The section above examined the centralized access architectures and distributed access architectures this paper will highlight key differences in both capabilities and performance. This section takes into account several areas that may impact end-of-line (EoL) performance, DOCSIS 3.1 capacity and/or operator margin. Each one of these areas is broken down into separate levers that are examined to account for a different parameter under test, which will yield a different result in many cases. The areas are described below and though the paper illustrates some of the results, our model has approximately 2000 outputs.

Network Element and Configuration Settings

- Network element and configuration settings of today are measured against future settings
  - Overall operating configuration settings (worst, intermediate, and best case)
- Transmitter MER for CAA CCAP and DAA CCAP
  - Transmitter MER CCAP settings (worst, intermediate, and best case)
- RF input levels
  - DOCSIS 3.1 at ~9 dBmV per 6 MHz (6 dB Down from analog)
  - DOCSIS 3.1 at ~15 dBmV per 6 MHz (not 6 dB Down from analog)
- Spectral width used 750, 1002, & 1218 MHz & impacts of Sub, Mid, and High-Split
- Partial or full spectrum loading of DOCSIS 3.1 vs. SC-QAM
- Spectrum band placement of DOCSIS 3.1 (regardless of CAA or DAA)

Network Architecture (Topology)

- Network architecture / topology Impact analysis from facility to home gateway
- Fiber distance between facility and node
  - Parameters examined facility to node distance of 15, 25, 40, 80 km
- Wavelength or lambda count and type of optical band used
  - Parameters examined 1, 4, 8, and 16 wavelength systems
- Amplifiers cascade depth
  - Parameters examined node +6 amps, +5, +4, +3, +2, +1, +0
- End-of-Line at point-of-entry to CPE or use of in-home wiring to CPE

Access Architecture (CAA and DAA)

- This model estimates the DOCSIS 3.1 network capacity using CAA and DAA
- This model also identifies conditions where CAA and DAA are differentiated
- DAA is termed Remote Gadget (Inclusive of Remote PHY or Remote CCAP)

Network Condition and Performance

- Network condition and performance can impact EoL performance & D3.1 capacity
- Equipment (CCAP source, AM optics & node (if used), amplifier, tap and CPE)
- Coaxial cable condition and type (express cable, distribution cable, drop cable)
- Use and condition of in-home wiring Network (not assumed in this version)
- Accounting for noise and additional Noise Margin

INTRODUCTION TO THE ARRIS DOWNSTREAM PERFORMANCE MODEL

Our model can be customized for any cable operator. The model takes into account the core areas impacting EoL: Network architecture (topology); network element configuration settings; and access architecture (CAA & DAA). The network architecture / topology examines parameters such as fiber distance between facility and node, lambda count, amplifiers cascade count, and spectrum (750 MHz, 1002 MHz, and plans to use 1.2 GHz). We can also assume network element configuration settings that can calculate the current configuration setting of today to support both analog video and digital and then future configuration setting that may be all digital, thus no analog. Finally, we can measure the impact of the access architecture, which examines the use of CAA and AM optics vs. DAA and digital optics impacts on EoL.

Model Objectives

Estimate Downstream Network Performance and D3.1 Capacity:

- Estimate the end-of-line (EoL) CNR performance to the last active
- Estimate maximum D3.1 modulation orders and thus maximum capacity

Estimate downstream performance considering variables the will impact performance:

- Network elements and configurations
- Network architectures
- CAA and DAA
- Network condition and performance

Model Methodology

In our model we used existing products and real-world configurations and/or those recommended to our customers. The model and paper's focus is on estimating or predicting the performance of DOCSIS 3.1 technology in CAA and DAA systems, all of which do not exist at this time. For the products that do not exist performance estimates have been placed in the model. Every effort is made by ARRIS design and implementation experts to estimate accurately the performance range of future D3.1 products used to predict future performance. These numbers have and will change as we learn more. As productions systems end-to-end become available we will evaluate the current model input with the data from our lab testing and then in field deployments.

Model Mechanics and Assumptions

The output of the model is a maximum value that could be attained under the assumptions and parameters of each use case while assuming a network in normal working order. Our model will provide an estimate of the maximum EoL as defined at the amplifier location for node +6 to +1 and amplifier 0 will be the HFC node or DAA remote gadget device. To determine the DOCSIS 3.1 modulation order a measurement at the RF connector of the CPE will need to be performed on a per subscriber basis, this is not currently possible because DOCSIS 3.1 end-to-end systems do not exists. The authors will not estimate the average or worst case CNR at the subscriber's home until end-to-end DOCSIS 3.1 systems can be tested in a lab and then field deployments.

The paper defines end-of-line (EoL) as the maximum estimated CNR at the amplifier. The EoL values will vary because of network element configuration settings, network architecture and topology, and access architecture (i.e. CAA & DAA).

The segment between the amplifier and the RF connector of the Dual Port D3.1 gateway may be less than or equal to the maximum estimated CNR amplifier values. The CNR at the CPE can vary based on

network condition and performance. The segment between amplifiers contains roughly about 750 to 1000 feet of coaxial cable with approximately five through taps and a terminating tap and from each tap a drop cable spanning approximate 75 to 150 feet or more connects the subscriber's home. In the model and paper we assume the DOCSIS 3.1 Dual-Port Gateway is located at the point-of-entry of the home to terminate the connection to the cable operator's access network. The IHN (In-Home Network) is not used for DOCSIS 3.1 in our current version of the model. The in-home network (IHN) is just a Local Area Network (LAN) to deliver cable operator services to data home networking and video rendering devices in the home. Figure 14 provides an illustration of the scope of the model and illustrates the architecture assumptions.

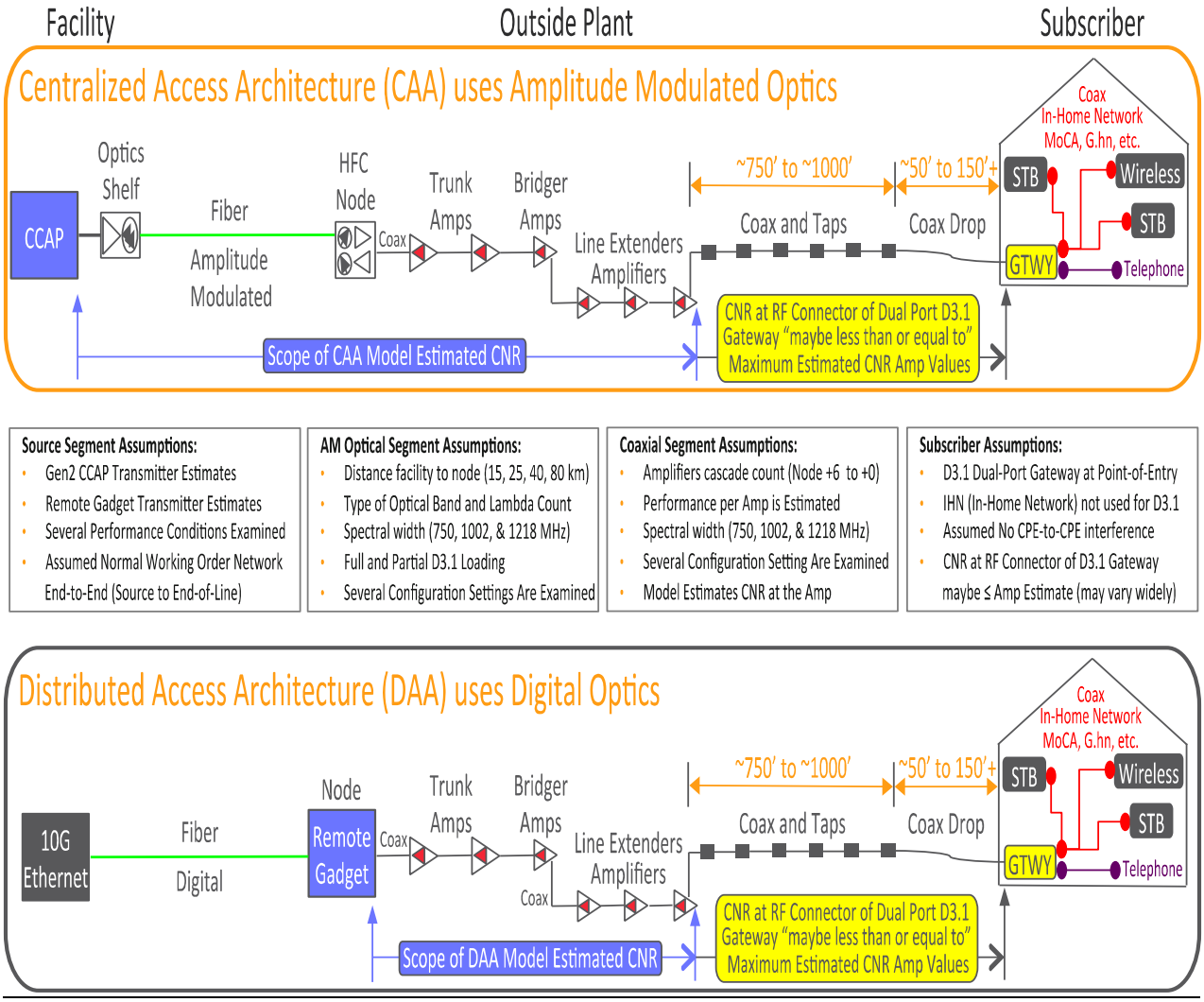
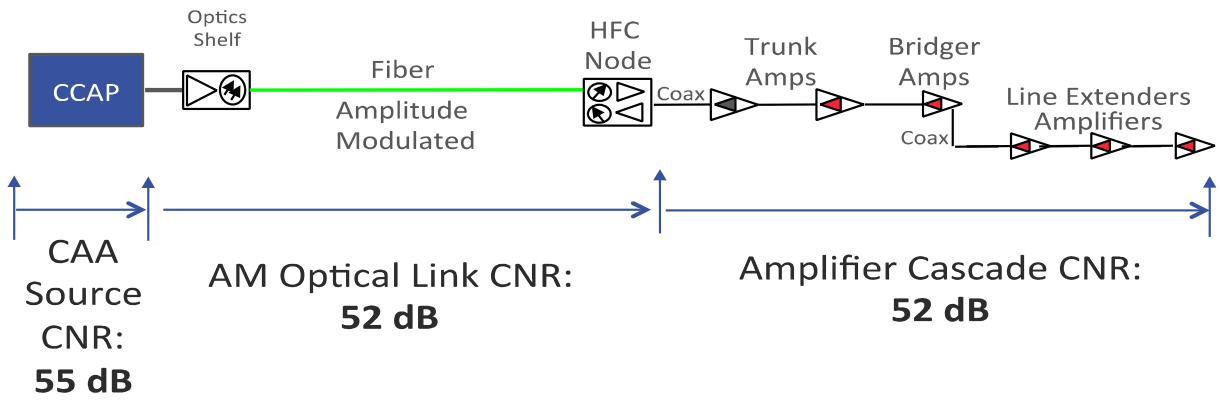


Figure 14: Diagram and Assumptions to Scope of DOCSIS 3.1 CAA & DAA Analysis

Carrier to Noise (CNR) and Signal to Noise (SNR) ratios both represent signal power to noise power ratios and are often used interchangeably, with the most recent trend to reserve SNR for a baseband and CNR for an RF band signal, as codified in DOCSIS 3.1 specifications. Similarly, Modulation Error Ratio (MER) represents average signal constellation power over average constellation error power and is often interchanged with SNR. In this paper, we refer to CNR in the spirit of DOCSIS 3.1 specifications.

In an HFC transmission system, it is often insightful to analyze of a cascade of variously performing system elements such as: signal source, optical transmission, coaxial plant RF amplifier chain, tap and coaxial distribution segment, drop cable and in-home coaxial network (if used) to list a few. Each of these can be individually characterized by the standalone link CNR. One way to track the CNR of the whole system exactly is to analyze the signal's progression from the start to the end of the system, and to add noise contributions along the way. However, a simplified approach, which makes assumptions of properly selected level ranges at each of the links, can use individual link CNR to estimate the end-to-end CNR.

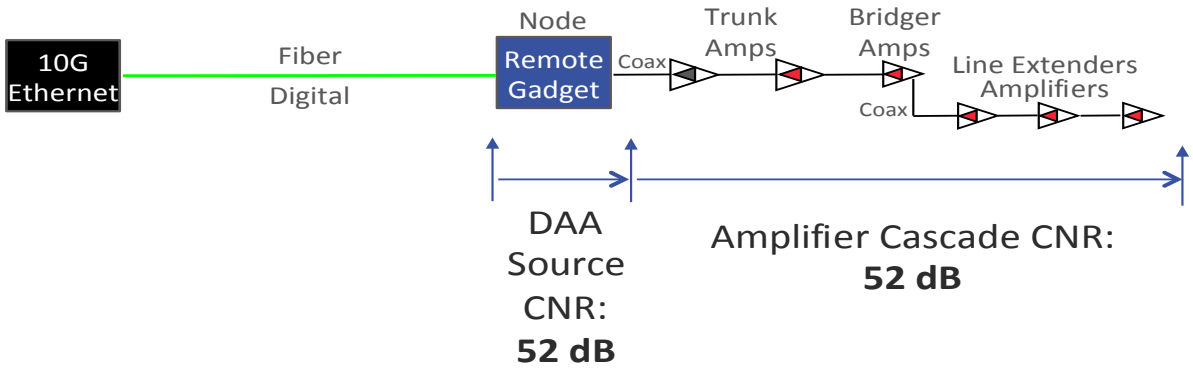
Figure 15 and figure 16 illustrate the approach the model uses to calculate CAA end-of-line CNR and DAA end-of-line CNR respectively. The difference as illustrated in both the network diagram and the equation is that CAA includes the AM optical link CNR contribution, while the DAA does not require an optical contribution to be added to the CNR equation. The values used are for illustration purposes only.



CAA EoL CNR Equation:

$$-10\log\left(10^{-\frac{55}{10}}+10^{-\frac{52}{10}}+10^{-\frac{52}{10}}\right)=\ 48\ dB$$

Figure 15: CAA End-of-Line CNR and Equation (values used are for illustration purposes only)



DAA EoL CNR Equation:

$$-10\log\left(10^{-\frac{52}{10}} + 10^{-\frac{52}{10}}\right) = 49\ dB$$

Figure 16: DAA End-of-Line CNR and Equation (values used are for illustration purposes only)

In figure 17, the DOCSIS 3.1 specifications defined the modulation and associated CNR at the RF connector of the cable modem. The authors are estimating the use of 8192QAM and 16384QAM, along with the required CNR. The required level for CM downstream post-FEC error ratio is defined as less than or equal to 10-6 PER (packet error ratio) with 1500 byte Ethernet packets.

CableLabs DOCSIS 3.1 Specifications (Downstream)			
Modulation	Coding	Required CNR at Connector up to 1 GHz	Required CNR at Connector 1 GHz to 1.2 GHz
512QAM	8/9 LDPC	30.5 dB [1]	30.5 dB [1]
1024QAM	8/9 LDPC	34 dB [1]	34 dB [1]
2048QAM	8/9 LDPC	37 dB [1]	37.5 dB [1]
4096QAM	8/9 LDPC	41 dB [1]	41.5 dB [1]
8192QAM	8/9 LDPC	45 dB	45.5 dB
16384QAM	8/9 LDPC	49 dB	49.5 dB

[1] CM-SP-PHYv3.1-I04-141218 - Table 7–41 - CM Minimum CNR Performance in AWGN Channel

Figure 17: DOCSIS 3.1 Modulation Capabilities with Required CNR at RF Connector

Model Estimates

The creators and authors of the model and this paper note the information provided is for educational purposes only and do not recommend sole reliance on this information for investment and operational decisions. The model values are predictions based on the combination of current and future product performance estimates, and may not represent actual performance. The information provided supersedes any previously issued estimates.

The authors and ARRIS reserve the right to change these performance estimates without notice.

ESTIMATING DOWNSTREAM D3.1 CAPACITY IN CAA AND DAA SYSTEMS

The CAA class of access architecture places the DOCSIS MAC and PHY layers in the cable operator's facility and uses AM optical technology, referred to as analog optics, to carry the DOCSIS signals to the node. The optical span between headend and node may vary widely from several kilometers (km) apart to over 100 km. The DAA class of access architecture places some or all of the DOCSIS functions in the outside plant, like a node or cabinet and use digital optical transport to the node, such as 10 Gigabit Ethernet. The use of CAA and AM optics is thought to lower overall performance compared to DAA systems that use digital optics. This section will predict downstream DOCSIS 3.1 performance in CAA and DAA systems.

The CNR estimates in the figures represent the maximum estimated CNR amplifier values under the defined parameters for each use case under study. This means that the EoL is the amplifier; this is the generally accepted meaning for the term EoL and it may also apply to the tap port as well. In normal working order systems the CNR measured at the amplifier should also be the measurement found at the tap port, which is between 100 to 1,000 feet away from the amplifier. In DOCSIS 3.1 the defined CNR value that determines the modulation order is the CNR value at the RF connector of the cable modem. In this paper, we state that the CNR at RF connector of dual port DOCSIS 3.1 gateway may be less than or equal to maximum estimated CNR amplifier values.

Network Element and Configuration Settings Estimates

The model and paper defines three operating configurations (worst case, intermediate case, and best case), as seen in figure 18. A major difference between them will be the settings assumed for all of the equipment. The worst case will assume that DOCSIS 3.1 will operate in CAA using AM optics 6 dB down from analog video services as it has since the inception of the service. This approach was successful at operating analog services that needed high performance as well as digital services that needed 256QAM. Cable operators may still operate the digital and DOCSIS channels at 6 dB down even without the presence of analog video services. For example a typical RF input levels of ~15 dBmV per 6 MHz channel analog and ~9 dBmV per 6 MHz channel digital, 256 QAM could be found in most cable operators deployments worldwide.

Segment	Worst Case Operating Configuration	Intermediate Case Operating Configuration	Best Case Operating Configuration
Description	Current Configuration Settings for AM optics and RF cascade at "6dB below analog" levels	Configuration Setting NOT Operating 6 dB below analog levels supporting for full spectrum D3.1	Highest Design Targets of Future Products
Source	Minimum Transmitter MER Target	+2 dB Above Minimum Transmitter MER Target	+3 to 4 dB Above Min. Transmitter MER Target
Optical Segment	QAM channels at "6 dB below analog" levels	Optimizing OMI for QAM Channels for Full Spectrum NOT 6 dB below analog levels	"Future" AM Product
RF Cascade Segment	QAM channels at "6 dB below analog" levels	Optimizing RF Levels for QAM Loading for Full Spectrum NOT 6 dB below analog levels	Highest Optimizing RF Levels for QAM Loading for Full Spectrum

Figure 18: Operating Configuration Assumptions

Worst Case Operating Configuration

This assumes that DOCSIS and digital services would continue to operate at an RF input level of 6 dB down with or without the presence of analog video services. In this worst case example, DOCSIS 3.1 would operate as it has for the last a decade at ~9 dBmV per 6 MHz channel where it only needed enough performance to achieve 256QAM. Additionally, the worst case example assumes minimum settings for CCAP source transmit MER, AM optics (if used), and amplifiers.

Intermediate Case Operating Configuration

This assumes that analog services which operated with a ~15 dBmV per 6 MHz channel analog is no longer present and the model assumes that DOCSIS 3.1 could operate with this RF input level. The intermediate case model also assumes a normal CCAP source transmit MER, AM optics (if used), and amplifiers, which we predict are achievable in real-world deployments.

Best Case Operating Configuration

This assumes that all of the network elements are operating at the highest design targets. Figure 19, defines parameters options and parameters used for the three Operating Configurations (worst case, intermediate case, and best case). This table will be used in the subsequent sections and the text in the table highlighted in red illustrates the key area under assessment. The resulting maximum estimated CNR amplifier value is captured in the line graph following the parameters table.

Category	Parameters Options	Parameters Used
Network Element and Configuration Settings	Overall Operating Configuration Settings (Worst, Intermediate, Best Case)	Under Assessment
	Transmitter MER CCAP Settings & RF input levels (Worst, Intermediate, Best)	Under Assessment
	Spectral Width used 750, 1002, & 1218 MHz (amount of spectrum in use)	1218 MHz
	Partial or Full Spectrum Loading of DOCSIS 3.1 vs. SC-QAM	Full Spectrum D3.1
	Spectrum Band Placement of DOCSIS 3.1 (regardless of CAA or DAA)	Full Spectrum D3.1
Network Architecture (Topology)	Facility to Node Distance (15, 25, 40, 80 km)	ALL
	AM Optical Wavelength or Lambda Count (1, 4, 8, or 16)	1 Wavelength
	Amplifiers Cascade Depth (6, 5, 4, 3, 2, 1, 0)	ALL
	End-of-Line at Point-of-Entry to CPE or Use of In-Home Wiring to CPE	Point-of-Entry to CPE
	CPE Type (Single-Port CM, Single-Port Gateway, Dual-Port Gateway)	Dual-Port Gateway
Access Architecture	CAA and DAA	Both
Network Condition and Performance	Network Equipment and Coaxial Condition / Performance	Normal Working Order
	Accounts for AM Optical and Plant Noise / D3.1 Spec CMTS & CM Noise Req.	Included
	Additional Noise Margin	None

Figure 19: Parameters for Network Element and Configuration Settings Estimates

The analysis results for the above parameters are illustrated in Figures 20, 21, and 22. The mode uses the network elements settings defined in figures 18 and 19. The line graphs headings located above the line values describe system under test. The CNR values at the amplifiers as well as the associated DOCSIS 3.1 modulation format defined assuming spectrum up to 1 GHz are defined in the graph and defined in figure 17. When considering spectrum between 1 GHz and 1.2 GHz, this will require a slightly higher CNR as defined in Figure 17.

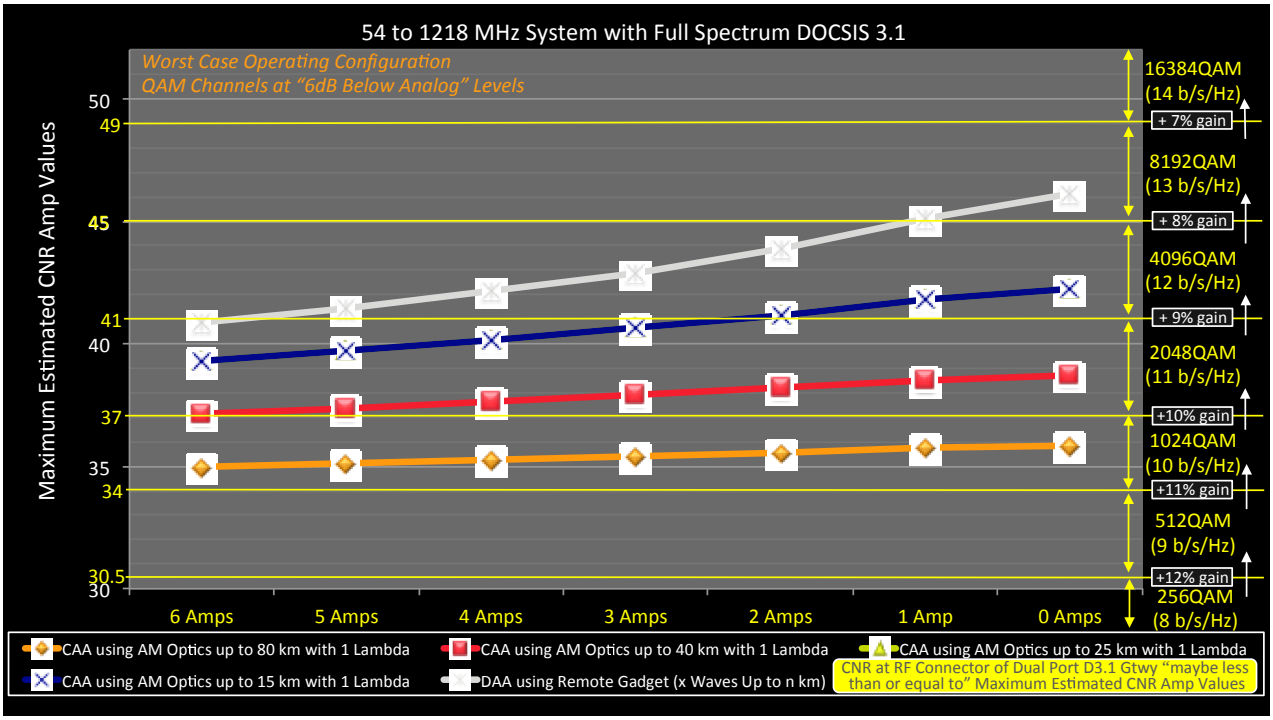


Figure 20: Worst Case Operating Configuration Settings – 1218 MHz

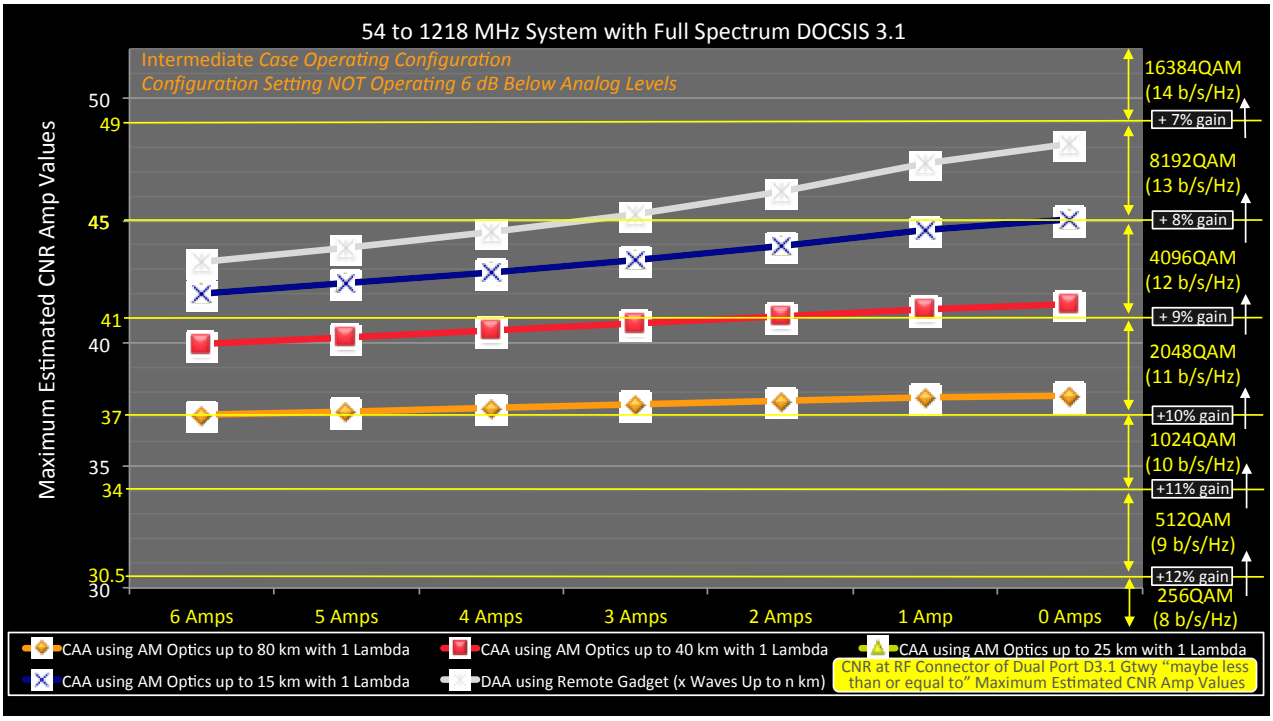


Figure 21: Intermediate Case Operating Configuration Settings – 1218 MHz

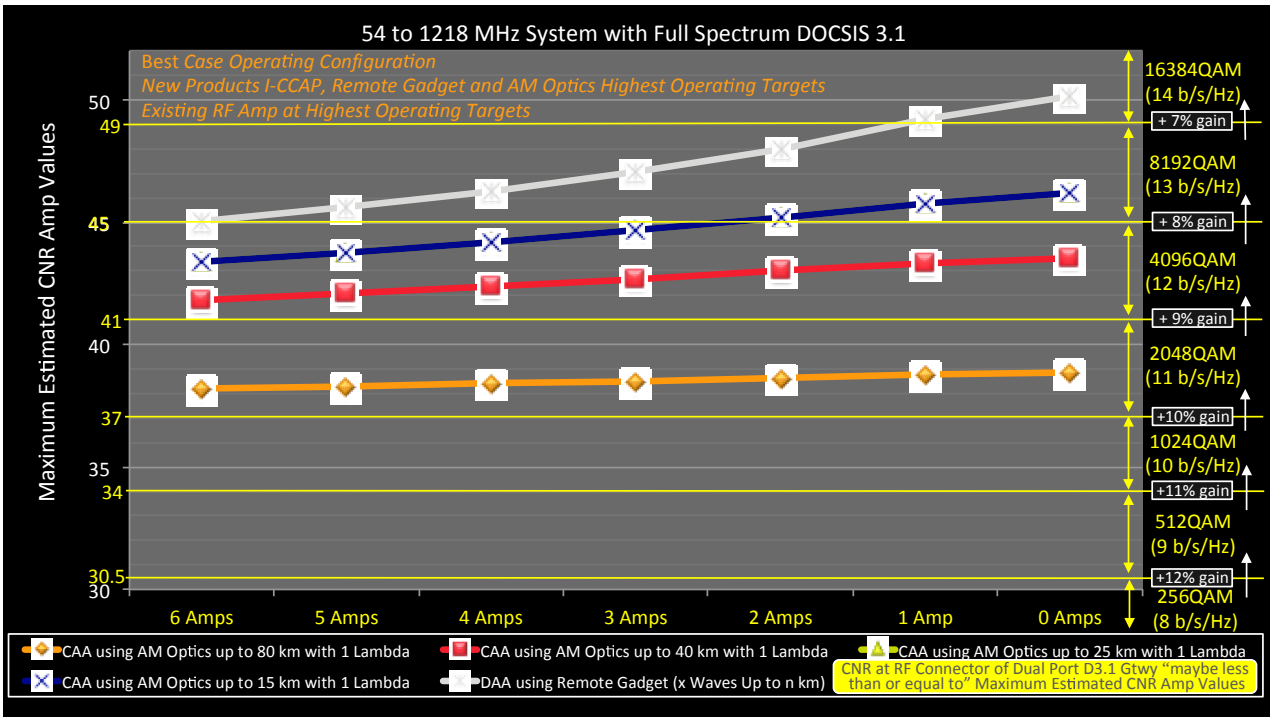


Figure 22: Best Case Operating Configuration Settings – 1218 MHz

Conclusions for Network Element and Configuration Settings Estimates

- Worst case operating configuration use is not ideal
  - Uses DOCSIS 3.1 in a 6 dB down from analog configuration (not ideal for all digital)
- Best case operating configuration use is not ideal
  - Uses the highest design targets of future products
  - Uses the highest optimizing RF levels for QAM loading for full spectrum
- Intermediate case operating configuration is recommended
  - Uses currently available product configured to support digital not 6 dB down
  - Uses average / moderate design targets of future products
  - The additional analysis will assume the intermediate case settings

A Key Takeaway from this Analysis

Considering network element and configuration settings estimates, the graphs clearly show a major variation in performance. If just considering the worst case versus the intermediate case for CAA architectures, there are major differences in end-of-line. When considering the element settings, specifically operating DOCSIS 3.1 at 6 dB down as currently implemented in today's DOCSIS network this will limit overall performance, as seen in figure 20. While operating the DOCSIS 3.1 with the same RF input level as analog, performance will improve at the end-of-line amplifiers, as seen in figure 21.

Spectral Width Analysis 750 MHz vs. 1002 MHz vs. 1218 MHz with Full Spectrum DOCSIS 3.1

Spectrum width or loading was identified as a performance impacting area with CAA using amplitude modulation optics; refer to figure 6 highlights of the optical impairments. In this section we examine the spectrum width and loading of full spectrum DOCSIS 3.1 up to 750 MHz vs. 1002 MHz vs. 1218 MHz to determine performance differentiation of CAA and DAA. Figure 23 identifies the parameters used in the analysis.

Category	Parameters Options	Parameters Used
Network Element and Configuration Settings	Overall Operating Configuration Settings (Worst, Intermediate, Best Case)	Intermediate
	Transmitter MER CCAP Settings & RF input levels (Worst, Intermediate, Best)	Intermediate
	Spectral Width used 750, 1002, & 1218 MHz (amount of spectrum in use)	Under Assessment
	Partial or Full Spectrum Loading of DOCSIS 3.1 vs. SC-QAM	Full Spectrum D3.1
	Spectrum Band Placement of DOCSIS 3.1 (regardless of CAA or DAA)	Full Spectrum D3.1
Network Architecture (Topology)	Facility to Node Distance (15, 25, 40, 80 km)	ALL
	AM Optical Wavelength or Lambda Count (1, 4, 8, or 16)	1 Wavelength
	Amplifiers Cascade Depth (6, 5, 4, 3, 2, 1, 0)	ALL
	End-of-Line at Point-of-Entry to CPE or Use of In-Home Wiring to CPE	Point-of-Entry to CPE
	CPE Type (Single-Port CM, Single-Port Gateway, Dual-Port Gateway)	Dual-Port Gateway
Access Architecture	CAA and DAA	Both
Network Condition and Performance	Network Equipment and Coaxial Condition / Performance	Normal Working Order
	Accounts for AM Optical and Plant Noise / D3.1 Spec CMTS & CM Noise Req.	Included
	Additional Noise Margin	None

Figure 23: Parameters for Spectrum Width Analysis 750 MHz vs. 1002 MHz vs. 1218 MHz

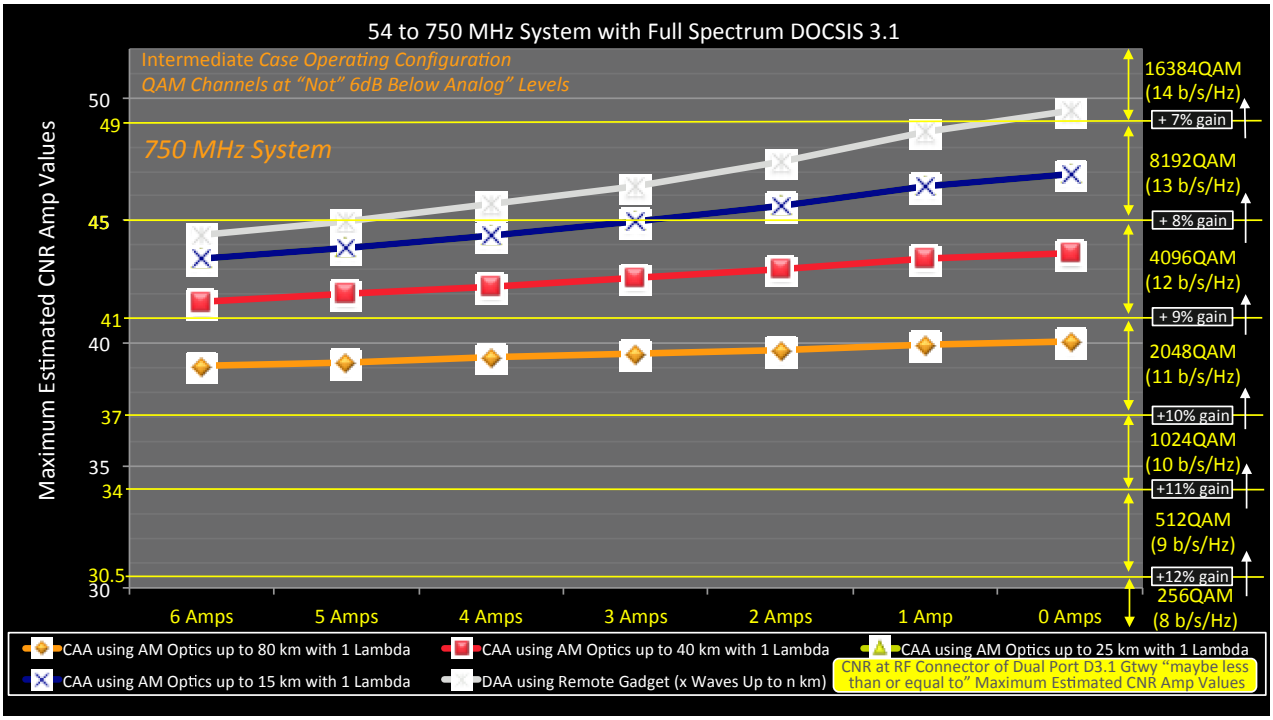


Figure 24: Spectral Width Analysis 750 MHz with Full Spectrum DOCSIS 3.1

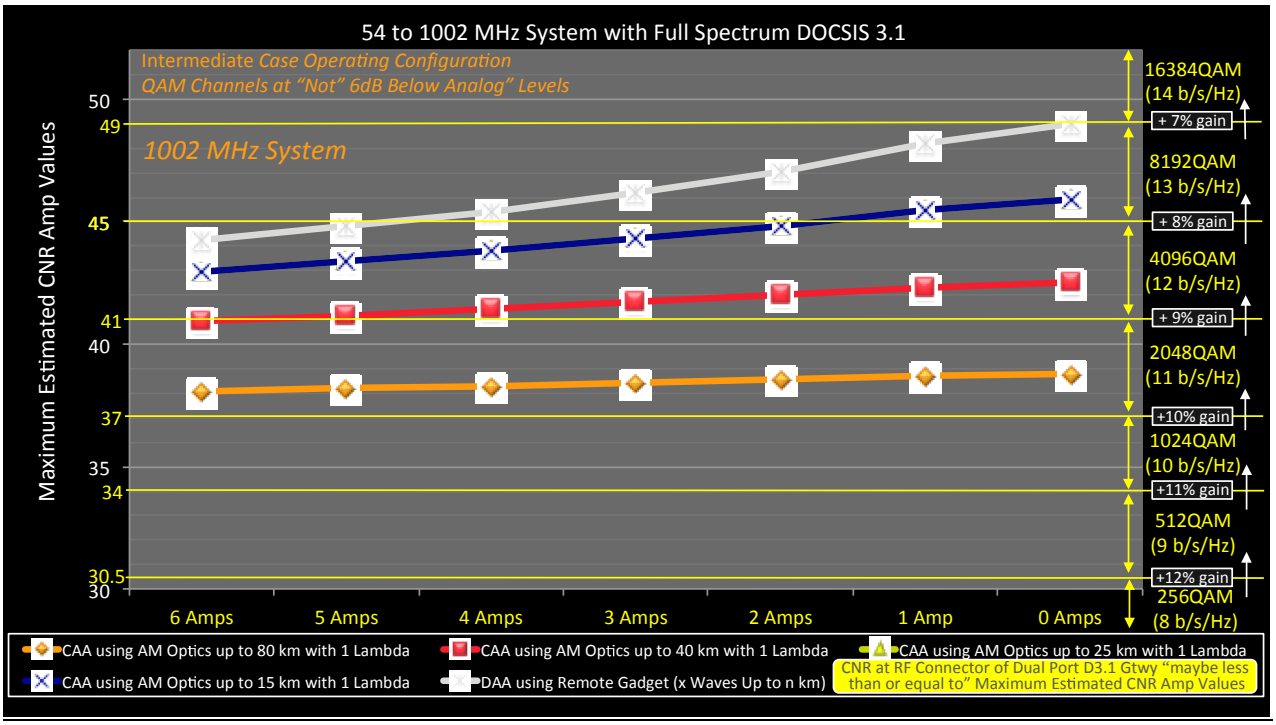


Figure 25: Spectral Width Analysis 1002 MHz with Full Spectrum DOCSIS 3.1

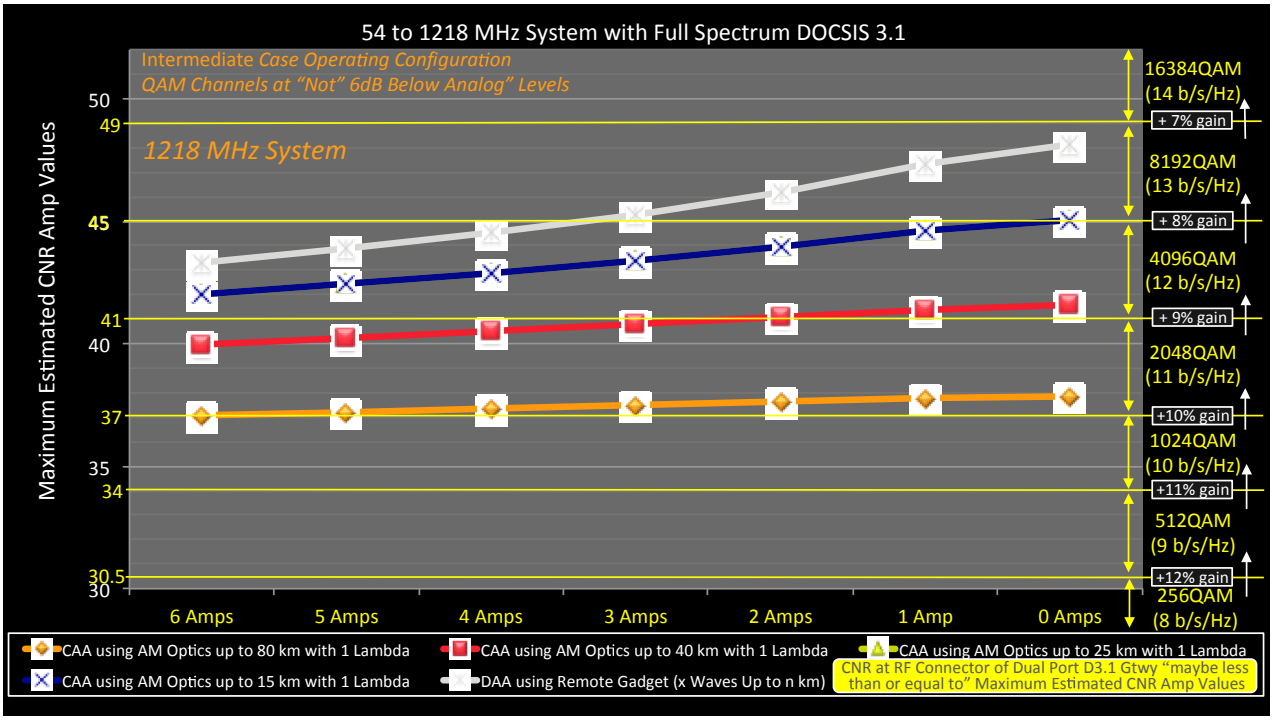


Figure 26: Spectral Width Analysis 1218 MHz with Full Spectrum DOCSIS 3.1

Additional Spectrum Level and Tap Component Analysis Considerations

Taps are the components with the most variability in passband characteristics, because there are so many different manufacturers, values, and number of outputs. Most were designed more than ten years ago, well before >1 GHz bandwidth systems were considered. One of the serious limitations of power passing taps is the AC power choke resonance. This typically is around 1100 MHz, although the notch frequency changes with temperature. Tap response resonances are typical from ~1050 to 1400 MHz. This is an important finding when leveraging the existing passives; therefore the use above 1050 MHz may not be predictable or even possible. Taps in cascade may affect capacity, thus additional testing is required [EMM1]

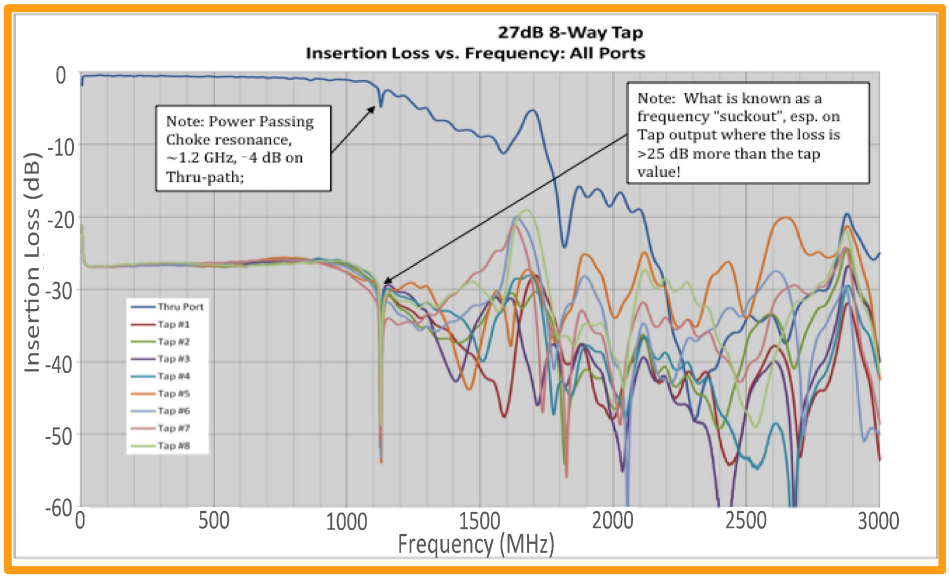


Figure 27: Spectrum Level and Tap Component Analysis Considerations

Use up to 1218 MHz Could be Difficult

- Tap performance can vary widely
- Tap performance may not yield up to 1218 MHz (above 1100 MHz could be a challenge)
- Modulation orders will likely be Lower than those used below 1 GHz
- AM optics and remote gadget perform better below 1 GHz
- Model estimates to the amplifier
- The authors have modeled up to 1218 MHz though use is not assured

Conclusions for Spectral Width Analysis 750 MHz vs. 1002 MHz vs. 1218 MHz

The tap performance as our industry approaches 1 GHz or even higher to 1.2 GHz could exhibit variation in performance depending on manufacture, model used, date of production, conditions in the field, and other factors. [EMM1] Assuming all parameters are the same except for spectrum load of 750 MHz, 1002 MHz, and 1218 MHz, figures 24, 25, and 26 show a minor decline in end-of-line as estimated at the amplifier, 2 dB delta comparing 750 MHz and 1218 MHz.

AM Optical Wavelength Count of 1, 4, 8, 16 with DOCSIS 3.1 Full Spectrum 1218 MHz

A major concern in the industry has been the uncertainty of performance of CAA using amplitude modulation optics when increasing the number of wavelengths. The performance may also be impacted when other parameters are assumed, such as full spectrum DOCSIS 3.1 to 1218 MHz and long distances. The AM optical concerns were raised earlier in the paper as well as other papers reference for more detail. What was unclear we estimate in end-of-line CNR to the amplifier values, assuming the parameters used in Figure 28.

Category	Parameters Options	Parameters Used
Network Element and Configuration Settings	Overall Operating Configuration Settings (Worst, Intermediate, Best Case)	Intermediate
	Transmitter MER CCAP Settings & RF input levels (Worst, Intermediate, Best)	Intermediate
	Spectral Width used 750, 1002, & 1218 MHz (amount of spectrum in use)	1218 MHz
	Partial or Full Spectrum Loading of DOCSIS 3.1 vs. SC-QAM	Full Spectrum D3.1
	Spectrum Band Placement of DOCSIS 3.1 (regardless of CAA or DAA)	Full Spectrum D3.1
Network Architecture (Topology)	Facility to Node Distance (15, 25, 40, 80 km)	ALL
	AM Optical Wavelength or Lambda Count (1, 4, 8, or 16)	ALL Under Assessment
	Amplifiers Cascade Depth (6, 5, 4, 3, 2, 1, 0)	ALL
	End-of-Line at Point-of-Entry to CPE or Use of In-Home Wiring to CPE	Point-of-Entry to CPE
	CPE Type (Single-Port CM, Single-Port Gateway, Dual-Port Gateway)	Dual-Port Gateway
Access Architecture	CAA and DAA	Both
Network Condition and Performance	Network Equipment and Coaxial Condition / Performance	Normal Working Order
	Accounts for AM Optical and Plant Noise / D3.1 Spec CMTS & CM Noise Req.	Included
	Additional Noise Margin	None

Figure 28: Parameters for AM Optical Wavelength with D3.1 Full Spectrum 1218 MHz

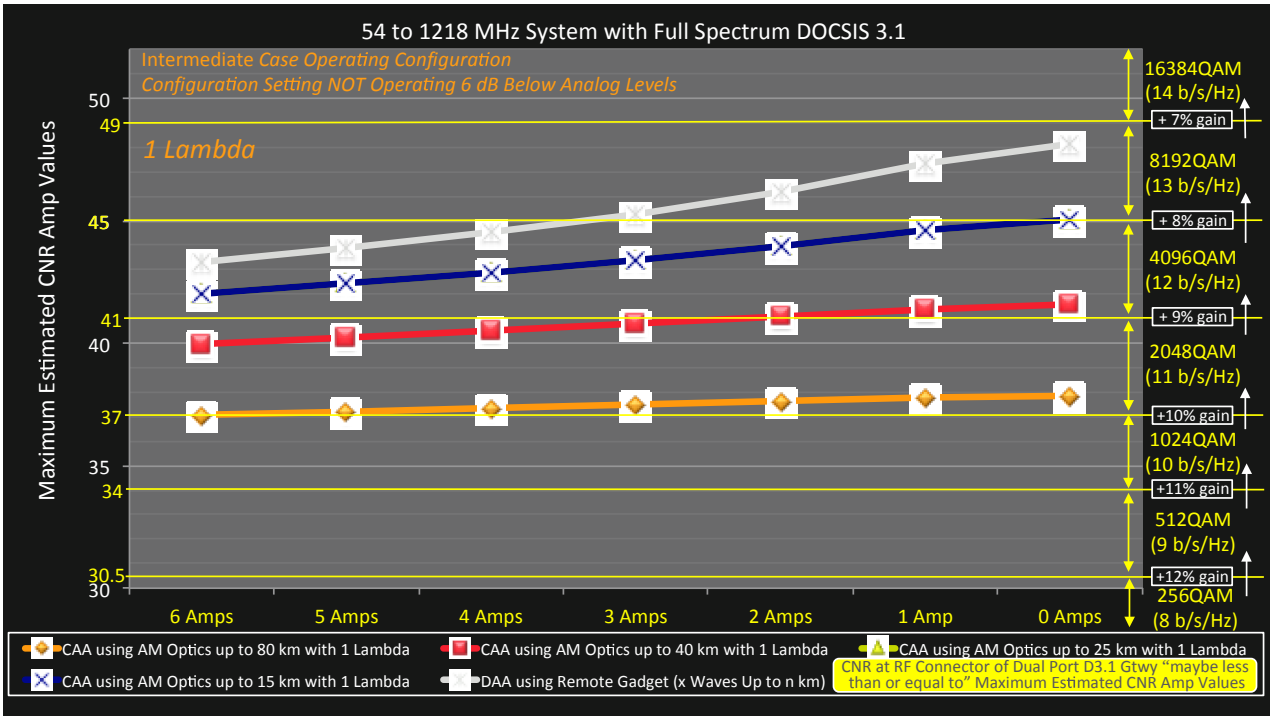


Figure 29: DAA vs. CAA with Amplitude Modulated (AM) Optical – One (1) Wavelength

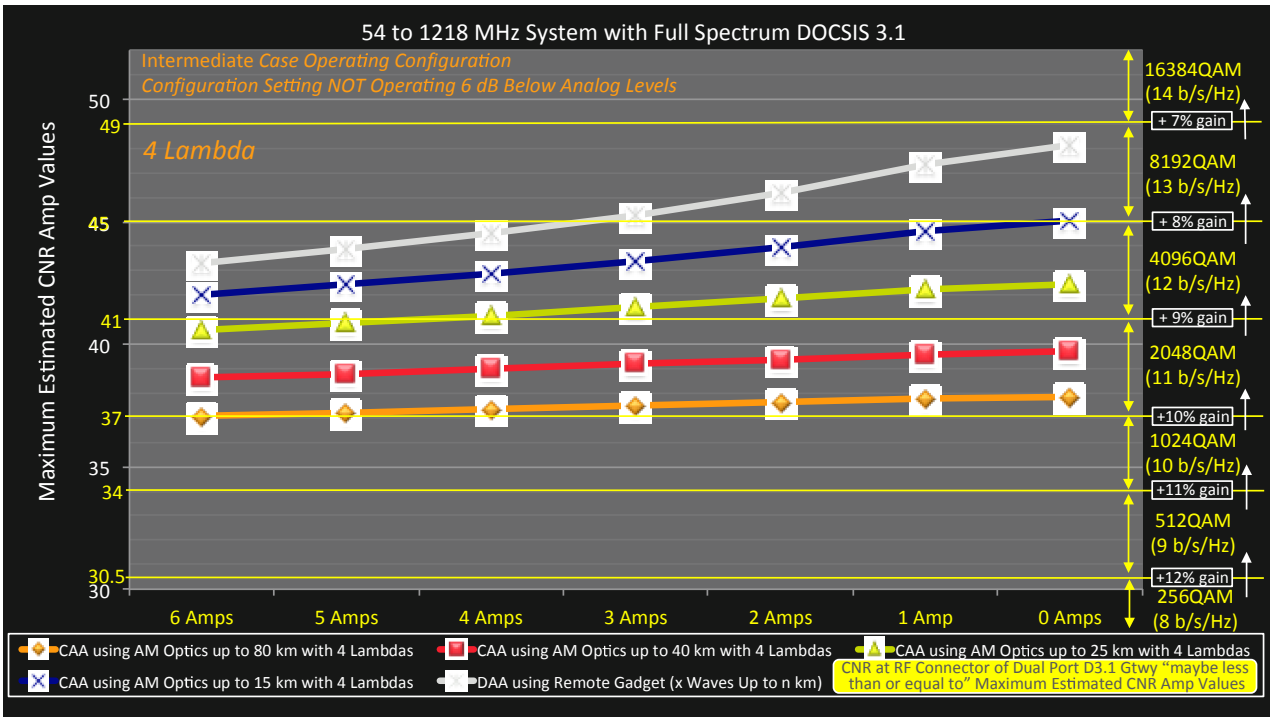


Figure 30: DAA vs. CAA with Amplitude Modulated (AM) Optical – Four (4) Wavelengths

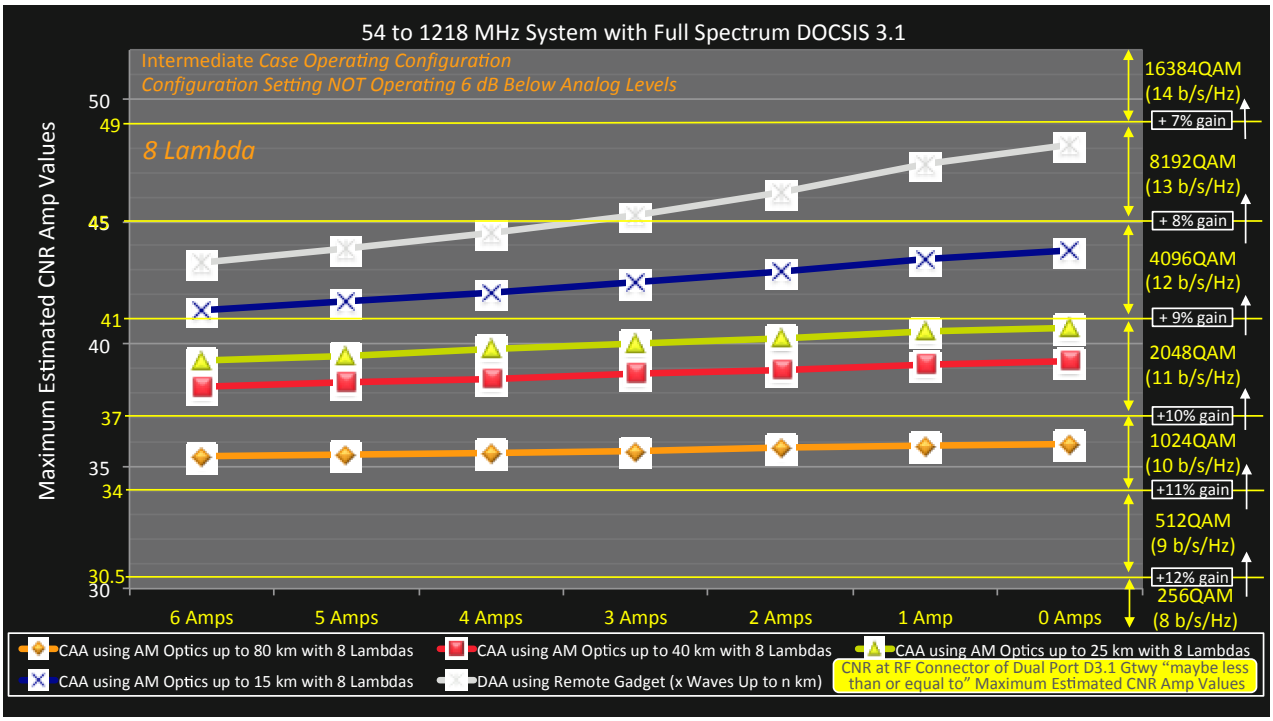


Figure 31: DAA vs. CAA with Amplitude Modulated (AM) Optical – Eight (8) Wavelengths

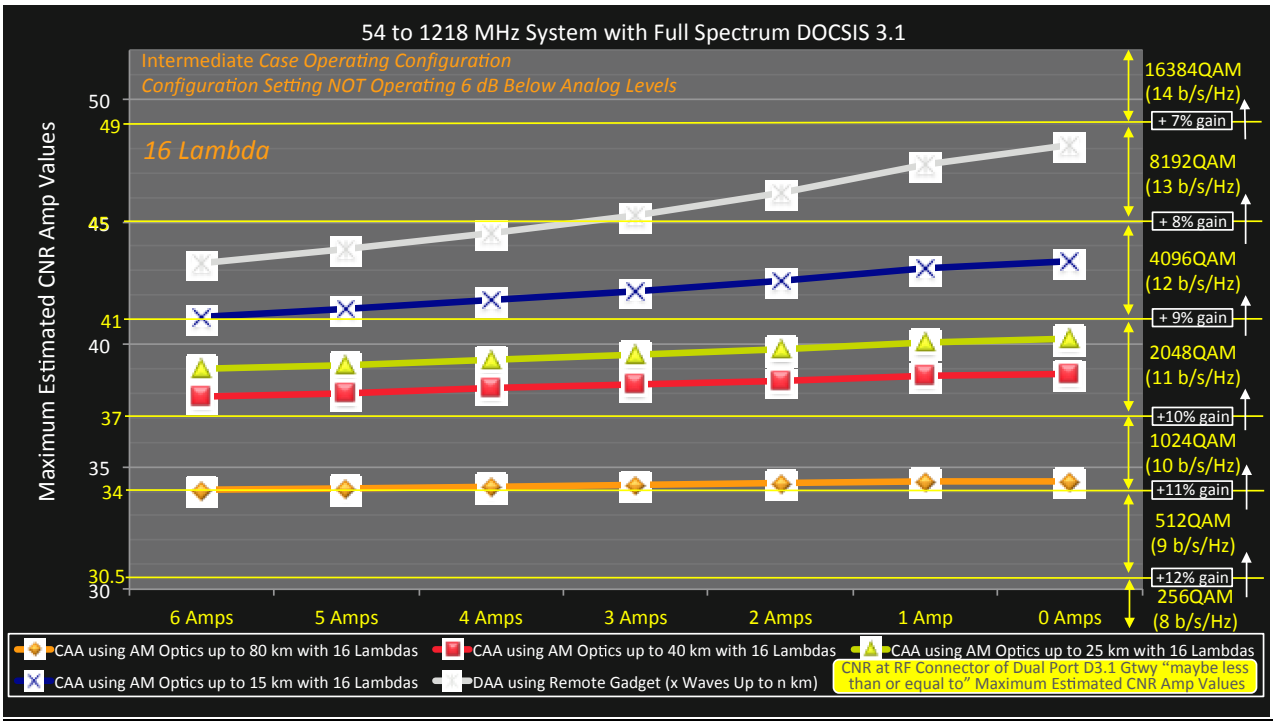


Figure 32: DAA vs. CAA with Amplitude Modulated (AM) Optical – Sixteen (16) Wavelengths

Conclusions for AM Optical Wavelength Count of 1, 4, 8, or 16
(Figures 29, 30, 31, and 32)

This single analysis captures the major differentiation that may exist in end-of-line CNR at the amplifier when comparing CAA and DAA systems. This may also mean a difference in DOCSIS 3.1 modulation seen at the cable modem. This list below captures the key highlights of this analysis.

1) There is a very long distance between facility and node using AM optics impacts modulation order more than any other factor, thus DAA "may" yield a 30% gain compared to CAA 80 km links (9-14 dB at the Amp with 16 waves).

2) Wavelengths (1 - 16) & Distance (25 - 40km) CAA using AM optics will support high order modulation and DAA "may" yield 0% to 18% capacity gain (1-9 dB at Amp with 16 waves).

3) Amplifier count DAA affects modulation order yielding a 0% to 8% gain (5 dB) and CAA has 0-2 dB improvement with fewer amps.

4) Overall the model and paper EoL estimates to the amplifier in CAA and DAA systems will set the maximum DOCSIS 3.1 modulation order possible, the difference in performance between CAA vs. DAA varies widely from DAA having 0% to 30% capacity gain over CAA systems.

Partial DOCSIS 3.1 with Placement Under 600 MHz and Remaining Spectrum SC-QAM

All of the analysis previously captured in this paper assumed that DOCSIS 3.1 would be used in full spectrum environments as this may represent end state architecture. However, our industry is a long way from full spectrum DOCSIS 3.1 and even the use of 1218 MHz. The parameters defined in this section examine the use of partial DOCSIS 3.1 spectrum and placement under 600 MHz with the remaining spectrum used for single carrier QAM (SC-QAM). In this analysis, the maximum spectrum used is 750 MHz or 1002 MHz as this represents many Cable operators current spectrum band plan. Figure 33 defines the parameters used in this section.

Segment	Partial Spectrum Case Operating Configuration	Partial Spectrum Case Operating Configuration
Description	Configuration Setting NOT Operating 6dB below analog levels supporting for 2-3 blocks of 192 below under 600 MHz	Configuration Setting NOT Operating 6dB below analog levels supporting for 2-3 blocks of 192 below under 600 MHz
Spectrum	54 – 750 MHz Full Spectrum DOCSIS 3.1	54 – 1002 MHz Full Spectrum DOCSIS 3.1
Source	+2 dB Above Minimum MER Transmitter Target	+2 dB Above Minimum MER Transmitter Target
Optical Segment	"Current" AM Optical Product Carrying 2 to 3 Blocks of 192 OFDM below 600 MHz	"Current" AM Optical Product Carrying 2 to 3 Blocks of 192 OFDM below 600 MHz
RF Cascade Segment	Optimizing RF Levels for QAM Loading for Full Spectrum NOT 6dB below analog levels	Optimizing RF Levels for QAM Loading for Full Spectrum NOT 6dB below analog levels

Figure 33: Operating Configuration Assumptions for Partial Spectrum Loading

Category	Parameters Options	Parameters Used
Network Element and Configuration Settings	Overall Operating Configuration Settings (Worst, Intermediate, Best Case)	Intermediate
	Transmitter MER CCAP Settings & RF input levels (Worst, Intermediate, Best)	Intermediate
	Spectral Width used 750, 1002, & 1218 MHz (amount of spectrum in use)	750 and 1002 MHz
	Partial or Full Spectrum Loading of DOCSIS 3.1 vs. SC-QAM	Partial D3.1 & SC-QAM
	Spectrum Band Placement of DOCSIS 3.1 (regardless of CAA or DAA)	D3.1 Under 600 MHz
Network Architecture (Topology)	Facility to Node Distance (15, 25, 40, 80 km)	ALL
	AM Optical Wavelength or Lambda Count (1, 4, 8, or 16)	4
	Amplifiers Cascade Depth (6, 5, 4, 3, 2, 1, 0)	ALL
	End-of-Line at Point-of-Entry to CPE or Use of In-Home Wiring to CPE	Point-of-Entry to CPE
	CPE Type (Single-Port CM, Single-Port Gateway, Dual-Port Gateway)	Dual-Port Gateway
Access Architecture	CAA and DAA	Both
Network Condition and Performance	Network Equipment and Coaxial Condition / Performance	Normal Working Order
	Accounts for AM Optical and Plant Noise / D3.1 Spec CMTS & CM Noise Req.	Included
	Additional Noise Margin	None

Figure 34: Parameters for Partial Spectrum and Spectrum Placement for D3.1

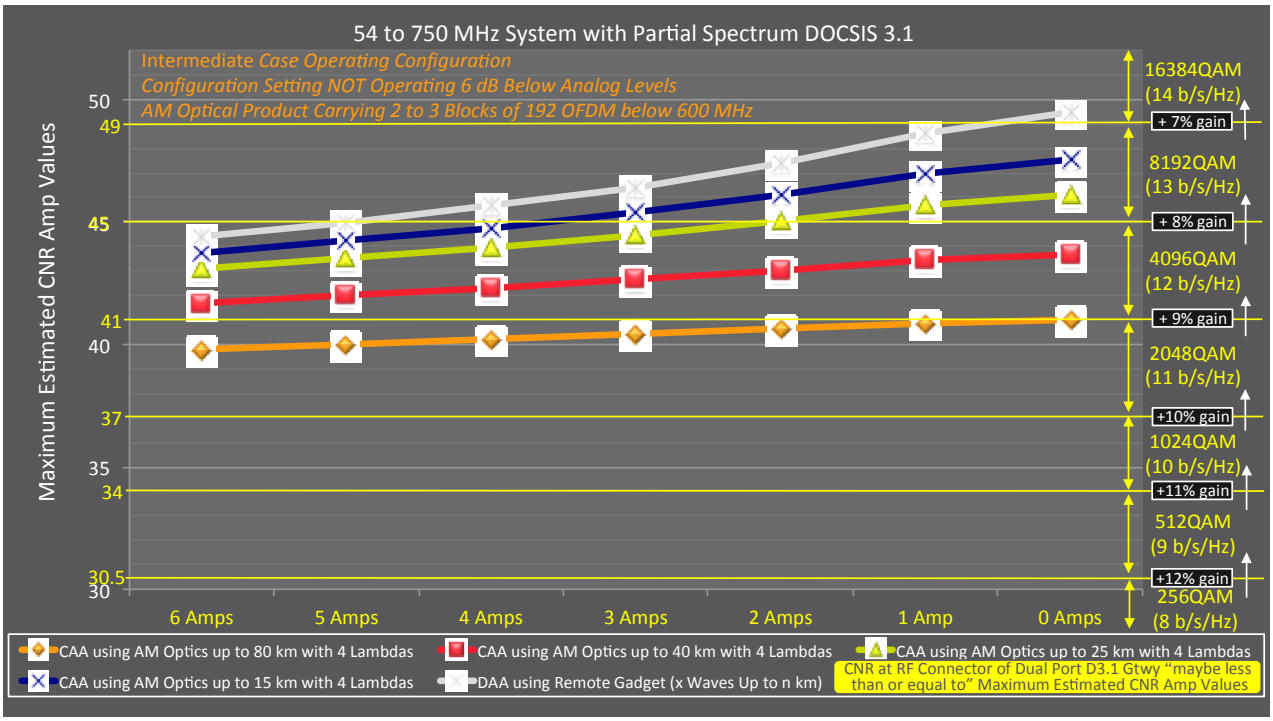


Figure 35: Partial Spectrum and Spectrum Placement for D3.1 – 750 MHz System

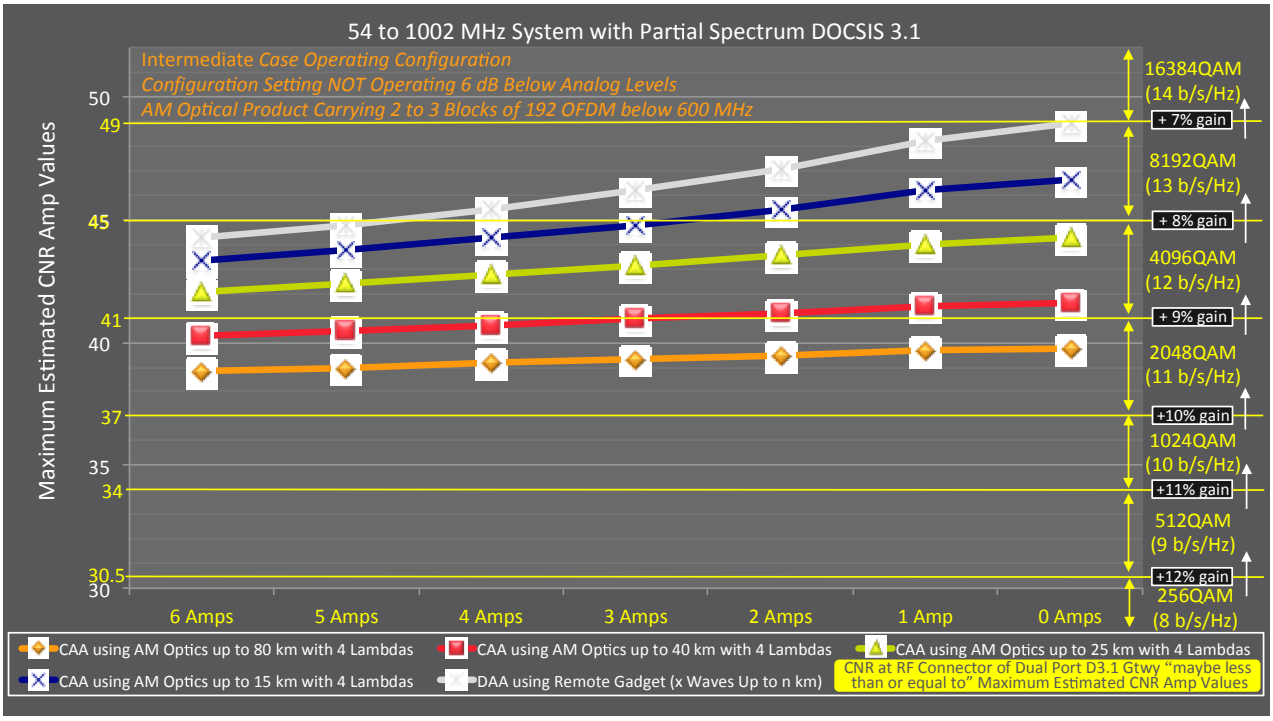


Figure 36: Partial Spectrum and Spectrum Placement for D3.1 – 1002 MHz System

Figure 35 Conclusions Using Partial DOCSIS 3.1 Spectrum in 750 MHz Systems

- 80 km links 2K QAM appears possible
- 40 km links 4K QAM appears possible
- 25 km links 4K to 8K QAM appears possible
- 15 km links 4K to 8K QAM appears possible

Figure 36 Conclusions Using Partial DOCSIS 3.1 Spectrum in 1002 MHz Systems

- 80 km links 2K QAM appears possible
- 40 km links 2K to 4K QAM appears possible
- 25 km links 4K QAM appears possible
- 15 km links 4K to 8K QAM appears possible

Conclusions for Partial Spectrum and Spectrum Placement for D3.1

- All MSOs will begin with partial spectrum DOCSIS 3.1
- Full spectrum D3.1 will not be used for a long time!
- Some MSOs may not upgrade spectrum or OSP to 1218 MHz
- Some MSOs will mine all they can out of 750 MHz and 1002 MHz Systems
- Some MSOs will wait for further analysis of actual spectral use above 1 GHz
- The best DOCSIS 3.1 spectrum is lower than 600 to 700 MHz
- Placement of DOCSIS 3.1 spectrum to take the place of analog video
- Analog needed the best spectrum and so does DOCSIS 3.1
- Both AM optics and remote gadget perform best at lower spectrum bands
- Using low frequency band and partial D3.1 spectrum loading maximizes CAA performance (compressing DAA gains)
- Perhaps DOCSIS 3.1 could replace location of analog video as it is removed

KEY SUMMARIES

The paper illustrated that network element configuration settings matter and can impact performance. The network element configuration settings such as transmitter MER, RF input levels, partial and full D3.1 spectrum loading, spectrum placement, and spectrum utilization of 750 MHz up to 1218 MHz can all influence the estimated EoL CNR at the amplifier and the maximum possible DOCSIS 3.1 modulations that may be supported at a given service group.

The paper also illustrated that network architecture and topology parameters matter and can impact performance. Network architecture and topology parameters such as distance between facility and node, wavelength count, and amplifiers cascade count can all influence the estimated EoL CNR at the amplifier and the maximum possible DOCSIS 3.1 modulations that may be supported at a given service group. The paper has illustrated that access architectures such as CAA using AM optics vs. future DAA using digital optics systems can both influence the estimated EoL CNR at the amplifier and the maximum possible DOCSIS 3.1 modulations that may be supported at a given service group.

The paper illustrates and states that the network conditions that can degrade the performance and happen at any time and at any segment, thus it is important to note that CNR at RF connector of dual port DOCSIS 3.1 gateway may be less than or equal to the maximum estimated CNR amplifier values and this value will determine the actual subscriber DOCSIS 3.1 modulations supported. The paper shows that tap performance in high frequencies approaching and above 1 GHz may be performance impacting.

The summary of the entire paper can be summarized in the following six figures. All

of these figures will assume an intermediate network elements configuration setting as described earlier in the paper. The paper and model has shown the partial DOCSIS 3.1 spectrum use and placement of two to three 192 MHz DOCSIS 3.1 blocks below the 600 to 700 MHz bands will yield the best results.

This paper also focused on the end-state architecture that would eventually use full spectrum DOCSIS 3.1 across the entire spectrum bands of either 750 MHz, 1002 MHz, or 1218 MHz. The following figures estimate the EoL CNR performance at the amplifier assuming full spectrum DOCSIS 3.1. The intermediate settings are used and these assume that analog services which operated with a ~15 dBmV per 6 MHz channel analog is no longer present. The model assumes that DOCSIS 3.1 could operate with the analog RF input level. Today DOCSIS 3.0 and prior systems as well as digital video operates at about ~9 dBmV per 6 MHz channel, known as 6 dB below analog. The intermediate case model also assumes a normal CCAP source transmit MER, AM optics (if used), and amplifiers, which we predict achievable in real-world deployments.

The following six figures illustrate the key differences in the estimated EoL CNR amplifier performance as well as the associated percentage of gain DAA – remote gadget systems could have when compared with CAA using AM optics. In all cases DAA has an estimated EoL CNR amplifier performance better than CAA. Each of the figures illustrate the percentage of gain DAA has against several CAA using AM optics use cases. The DAA – remote gadget percentage gain will be measured against the estimated EoL CNR measurements at the amplifiers 0 or node measurement. The figures also illustrate the maximum DOCSIS 3.1 modulation possible for a given use case or service group and the maximum percentage of increase DAA may have in DOCSIS 3.1 data capacity. These figures estimate the maximum performance possible as estimated at the amplifier and as stated before it is important to note that the CNR at RF connector of dual port DOCSIS 3.1 gateway may be less than or equal to maximum estimated CNR amplifier values. This paper illustrates the maximum estimated performance for a given use case or service group. The paper did not predict the any performance ranges at the subscriber homes in real-world deployments. The authors will wait until end-to-end DOCSIS 3.1 systems are available for lab and field examination, and we fully expect to revise this model and paper estimates using actual measurements when possible.

Figure 37 and Figure 38 summarizes the performance estimates for a 750 MHz system with full spectrum DOCSIS 3.1. Figure 39 and Figure 40 summarizes the performance estimates for a 1002 MHz system with full spectrum DOCSIS 3.1. Figure 41 and Figure 42 summarizes the performance estimates for a 1218 MHz system with full spectrum DOCSIS 3.1.

Figures 37, 39, and 41 summarize the estimated maximum EoL CNR amplifier performance as well as the associated percentage of gain DAA – remote gadget systems could have when compared with CAA using AM optic, several use cases or service groups are examined. Figures 38, 40 and 42 summarize the estimated maximum DOCSIS 3.1 modulation and maximum data capacity possible as well as the associated percentage of gain DAA – remote gadget systems could have when compared with CAA using AM optics, several use cases or service groups are examined. The maximum data capacity for CAA and DAA will vary depending on many factors, such as node +6 through node+0 (the node), figures 38, 40 and 42 illustrate the percentage range of DAA improvement when compared with CAA using AM optics.

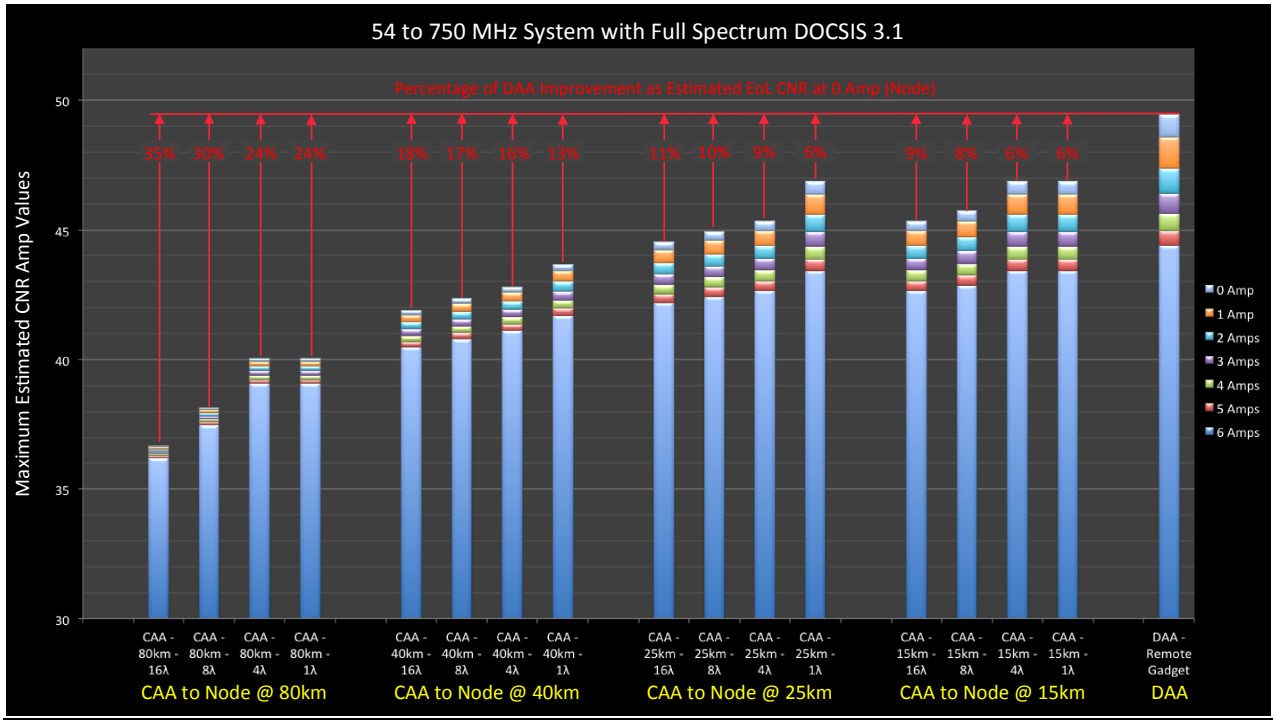


Figure 37: Estimated Maximum EoL CNR Amplifier Performance and Percentage of Gain DAA – Remote Gadget Against CAA using AM Optic Systems – DOCSIS 3.1 Full Spectrum 750 MHz

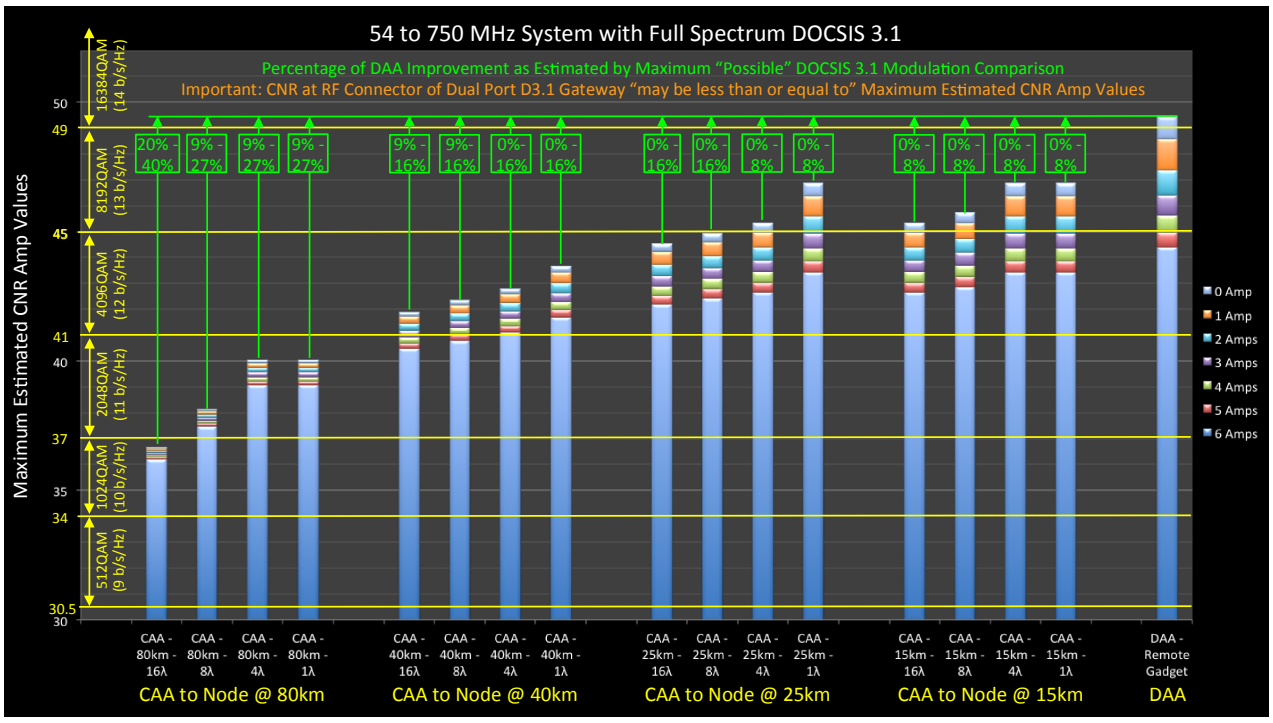


Figure 38: Estimated Maximum D3.1 Data Capacity Gain Possible and Percentage of Gain DAA – Remote Gadget Against CAA using AM Optic Systems – DOCSIS 3.1 Full Spectrum 750 MHz

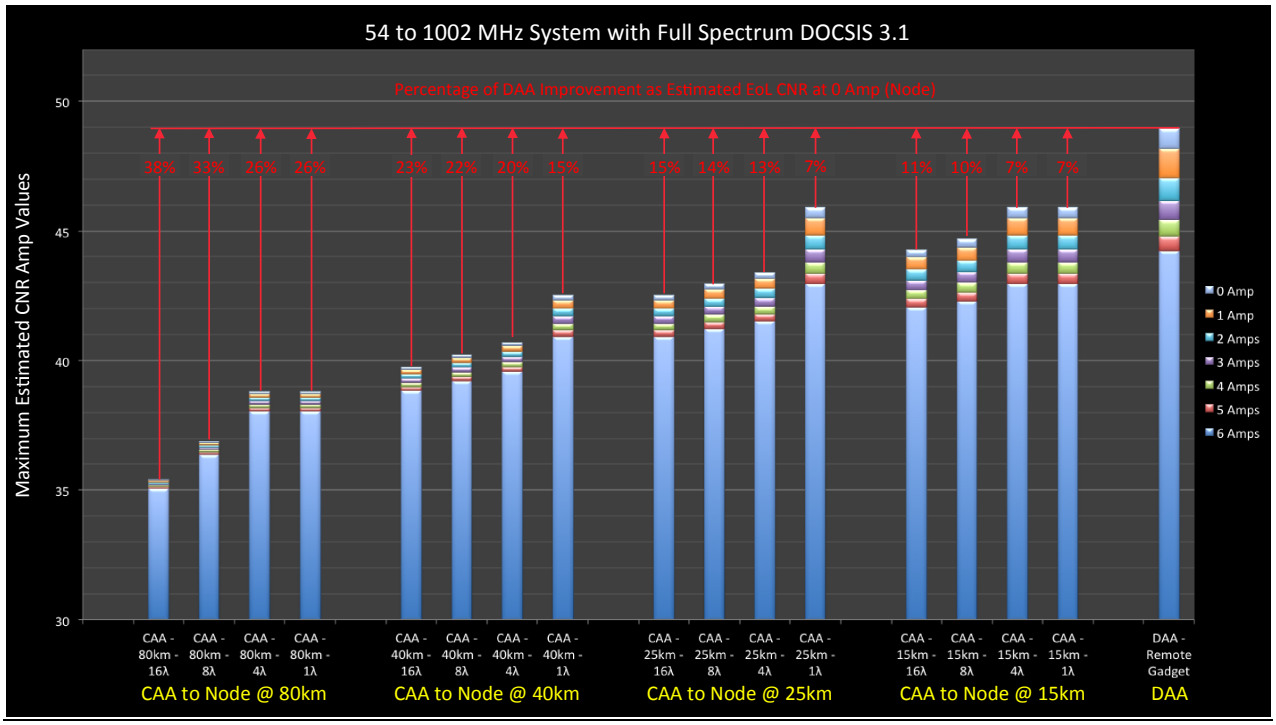


Figure 39: Estimated Maximum EoL CNR Amplifier Performance and Percentage of Gain DAA – Remote Gadget Against CAA using AM Optic Systems – DOCSIS 3.1 Full Spectrum 1002 MHz

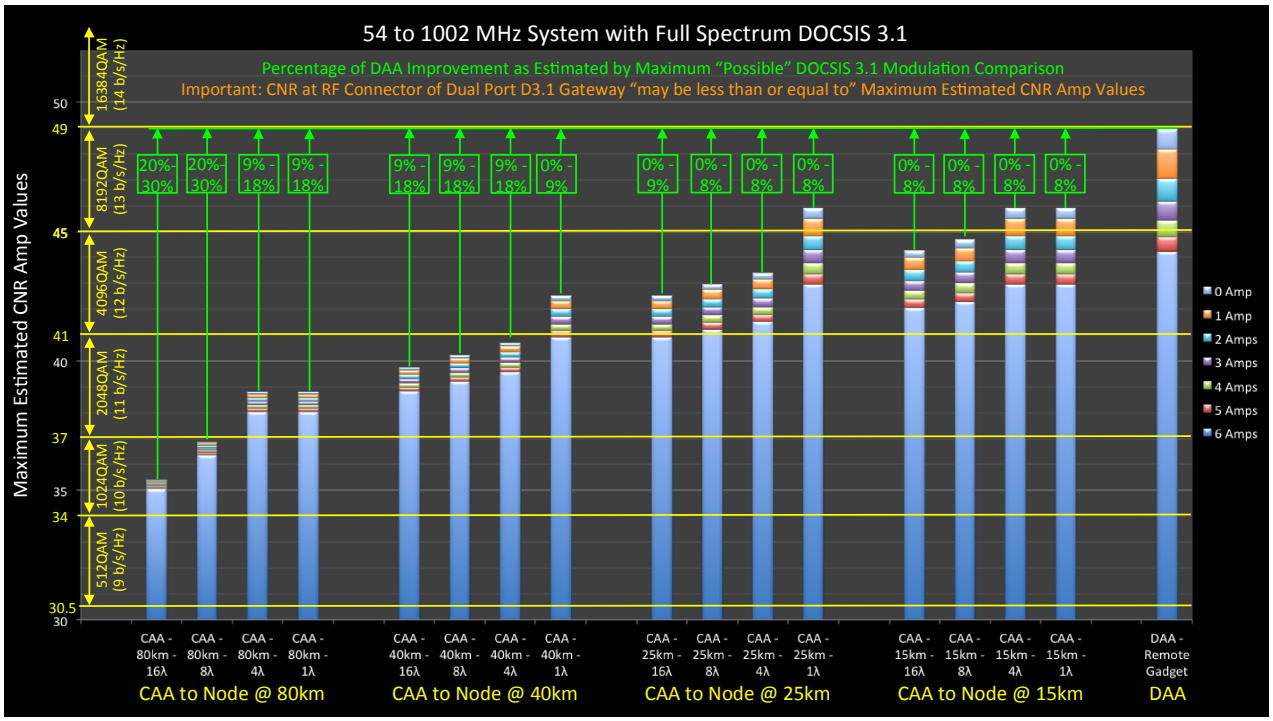


Figure 40: Estimated Maximum D3.1 Data Capacity Gain Possible and Percentage of Gain DAA – Remote Gadget Against CAA using AM Optic Systems – DOCSIS 3.1 Full Spectrum 1002 MHz

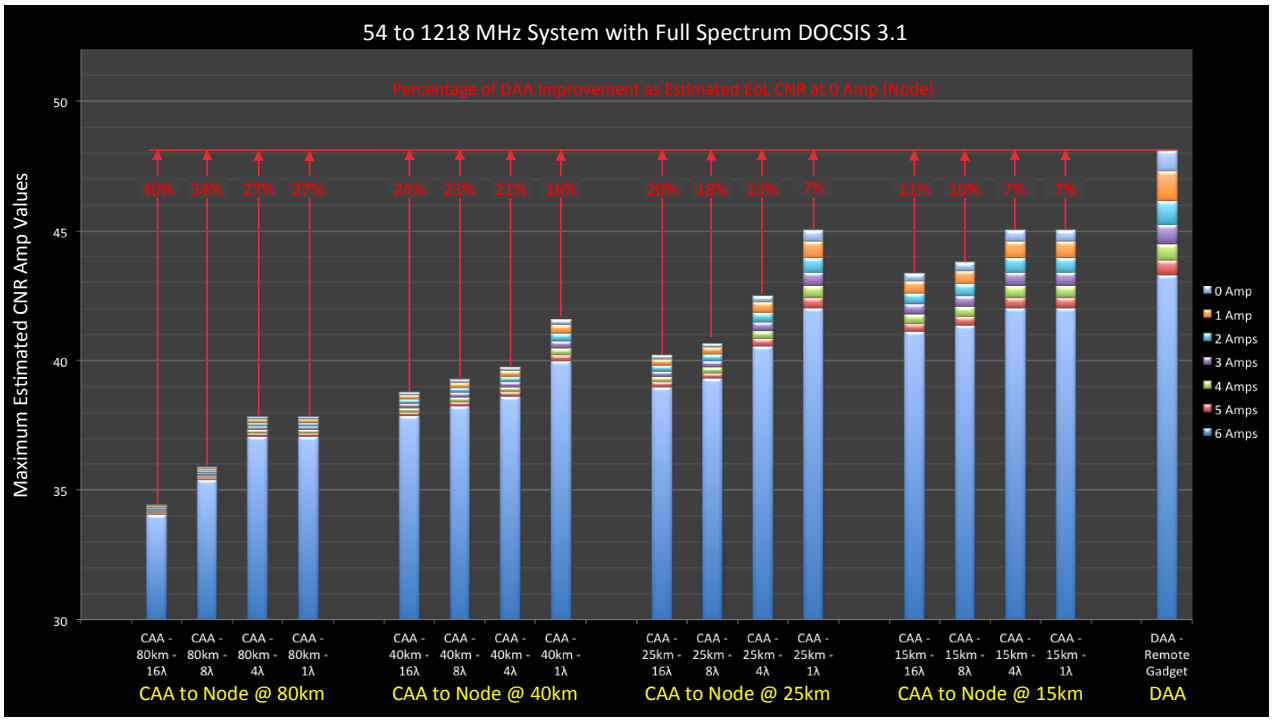


Figure 41: Estimated Maximum EoL CNR Amplifier Performance and Percentage of Gain DAA – Remote Gadget Against CAA using AM Optic Systems – DOCSIS 3.1 Full Spectrum 1218 MHz

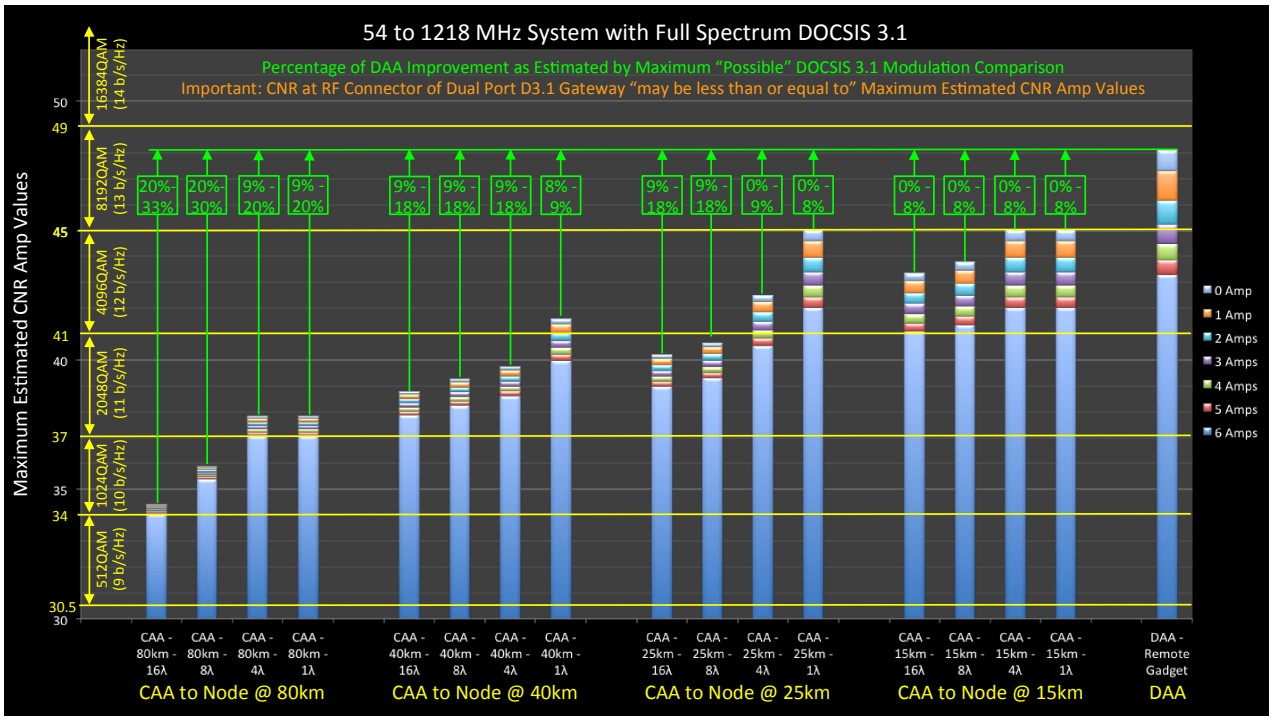


Figure 42: Estimated Maximum D3.1 Data Capacity Gain Possible and Percentage of Gain DAA – Remote Gadget Against CAA using AM Optic Systems – DOCSIS 3.1 Full Spectrum 1218 MHz

CONCLUSIONS

The previous section captures the key summaries of the model and paper. Cable operators may utilize different network configurations and have different network conditions resulting in overall performance variation. Considering the network architecture and topology variation that exist among cable operators around the world, we needed to define a set of parameters. Accounting for a wide variation in real-world network conditions and performance that is representative for all cable operators and subscribers served is nearly impossible. Since network condition and performance may also vary widely among cable operators, our model accounts for some noise conditions and defines a network under normal working order.

The best way to resolve a wide variation in both network architecture and actual system performance is to define several network element configuration settings, network architectures and topologies that we feel are representative of the vast majority of cable operator worldwide. The end-of-line (EoL) defined as the maximum estimated CNR at the amplifier. The segment between the amplifier and the RF connector of Dual Port D3.1 gateway "maybe less than or equal to" the maximum estimated CNR amplifier values, and this can vary as well based on network condition and performance.

When end-to-end DOCSIS 3.1 systems become available, additional lab testing and ultimately field data will determine the actual end-of-line performance. The authors believe that the DOCSIS 3.1 era will allow us to look at the "maximum" performance of the end-of-line and not just the minimum value of the weakest link. This is because DOCSIS 3.1 does not require all users to use the same modulation orders, as was the case with all previous versions of DOCSIS. The use of DOCSIS 3.1 will introduce the use of new tool called multiple modulation profiles (MMP).

Overall Key Findings in Order of Importance:

1) The maximum Downstream EoL performance estimates to the amplifier and resulting maximum DOCSIS 3.1 modulation order possible will vary widely between CAA and DAA systems based on many factors.

2) Network Architecture (Topology) matters:
    a. The distance between headend and node matters in system using CAA with AM optics, thus the longer the link the worse the performance, while DAA has no degradation
    b. The number of wavelengths matters in system using CAA with AM optics, thus the more wavelengths the worse the performance, while DAA has no degradation
    c. Number of amplifiers matters in estimating the EoL performance in either CAA or DAA, with DAA estimates show the largest EoL improvement as amplifier cascade is reduces
    d. Short optical spans between headend and node with few wavelengths yields the best performance for CAA using AM optics reducing the performance gap with DAA systems
    e. We recommend a dual port home gateway architecture to avoid using the in-home wiring network and to block noise impairments from entering the access network
    f. We recommend finding and removing CPE-to-CPE interferers

3) Network configuration settings matters: running D3.1 at 6 dB down from analog will limit performance / capacity

4) Spectrum usage such as up to 750 MHz, 1002 MHz, or 1218 MHz matters: the less amount of spectrum the greater the performance per MHz in in both CAA and DAA systems. The increased performance while occupying less spectrum will not offer more overall capacity when compared with 1002 or 1218 MHZ systems

5) Spectrum placement matters: in both CAA and DAA see benefits of using the lower spectrum band

6) Partial D3.1 spectrum loading matters: maximizes CAA performance reducing the performance gap with DAA systems

7) Network elements matter: D3.1 will have better equipment from CCAP to cable modem

8) The authors believe that the DOCSIS 3.1 era should reexamine our industry approach determining the minimum performances allowed for 64QAM and 256QAM. Our industry essentially cares about meeting a minimum target, because the network was only as good as the weakest link; the worst value determined the modulation order for that the entire segments. DOCSIS 3.1 and MMP this may not be a driver and knowing maximum is of interest

9) We should no longer assign a large value of six (6) dB of headroom with MMP

10) It's too early to know for sure the DOCSIS 3.1 capacity using CAA or DAA systems, as we need end-to-end DOCSIS 3.1 equipment and field measurements. The industry needs to measure the EoL at the drop and in the home for D3.1 CNR in current and future network architectures and configurations

Overall Estimates of EoL CNR at the Amp

1) The very long distance between the facility and a node using AM optics will impact modulation order more than any other factor, thus DAA "may" yield a 30% gain compared to CAA 80 km links (9-14 dB at the Amp with 16 waves)

2) Wavelengths (1 - 16) & Distance (25 - 40km) CAA using AM optics will support high order modulation and DAA "may" yield 0% to 18% capacity gain (1-9 dB at Amp with 16 waves)

3) Amplifier count DAA affects modulation order yielding a 0% to 8% gain (5 dB) and CAA has 0-2 dB improvement with fewer amps

4) Spectrum Band and Full Spectrum D3.1 for 750 MHz vs. 1.2 GHz have little impact to EoL (2 dB)

5) Overall the model and paper EoL estimates to the amplifier in CAA and DAA systems will set the maximum DOCSIS 3.1 modulation order possible, the difference in performance between CAA vs. DAA varies widely from DAA having 0% to 30% capacity gain over CAA systems.

ACKNOWLEDGEMENTS

The authors would like to thank the following individuals and test teams: John Ulm, Marcel Schemmann, Tom Cloonan, CCAP system test team, optics test team, amplifier test team, tap test team and technical publishing team.

REFERENCES

[ALB] Ayham Al-Banna et. al., "Measurement-Based EoL Stochastic Analysis and DOCSIS 3.1 Spectral Gain," in Proceedings, The NCTA Cable Show Spring Technical Forum, May, 2015.

[CLO] Tom Cloonan et. al., "Lessons from Telco and Wireless Providers: Extending the Life of the HFC Plant with New Technologies," in in Proceedings, The NCTA Cable Show Spring Technical Forum, May, 2015.

[EMM1] Emmendorfer, Michael J, S. Shupe, D. Cummings, T. Cloonan Contributors: Z. Maricevic, M. Schemmann, B. Dawson, V. Mutalik, J.Howe, A. Al-Banna, and F. O'Keeffe "Next Generation – Cable Access Network (NG-CAN), An Examination of the Drivers, Network Options, and Migration Strategies", NCTA, Chicago, IL ., June 14-16, 2011

[EMM2] Emmendorfer, Michael J, Shupe, Scott, Maricevic, Zoran, and Cloonan, Tom, "Examining HFC and DFC (Digital Fiber Coax) Access Architectures, An examination of the All-IP Next Generation Cable Access Network," 2011 SCTE Cable-Tec Expo, New Atlanta, GA, Nov. 14-17, 2011.

[EMM3] Emmendorfer, Michael J, Shupe, Scott, Cummings Derald, Cloonan, Tom, and O'Keeffe, Frank "Next Generation - Cable Access Network (NG-CAN), Examination of the Business Drivers and Network Approaches to Enable a Multi-Gigabit Downstream and Gigabit Upstream DOCSIS Service over Coaxial Networks", 2012 SCTE Canadian Summit Toronto, Canada, March 27-28, 2012

[EMM4] Emmendorfer, Michael J, and Cloonan, Tom "Examining the Future Evolution of the Access Network" NCTA, Washington, D.C., June 10-12, 2013

[EMM5] Emmendorfer, Michael J, The Evolution of Cable Access Technologies and Network Architectures, For This Decade and Beyond", 2013 SCTE Cable-Tec Expo, Atlanta, GA, October 21-24, 2013.

[EMM6] Emmendorfer, Michael J, Thomas J. Cloonan, John Ulm, and Zoran Maricevic "A Side-by-Side Comparison of Centralized vs. Distributed Access Architectures", 2014 NCTA, Los Angeles CA, April 29 - May 1, 2014.

[JOINT] Joint Supplier ("Mission is Possible: An Evolutionary Approach to Gigabit-Class DOCSIS", Authors J. Chapman, M. Emmendorfer, R. Howald, & S. Shulman, (Cisco, ARRIS, Motorola, and Intel), NCTA, Boston, MA May 21-23, 2012

[MUT] 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, The NCTA Cable Show Spring Technical Forum, May, 2015.

Evolution of CMTS/CCAP Architectures

Karthik Sundaresan
CableLabs

Abstract

The CMTS has evolved over time, becoming a CCAP supporting both high-speed data and video services. As bandwidth capacity needs grow rapidly, there is increasing pressure on the infrastructure, and new Distributed CCAP Architectures are emerging to address the need for smaller scope, more flexible options.

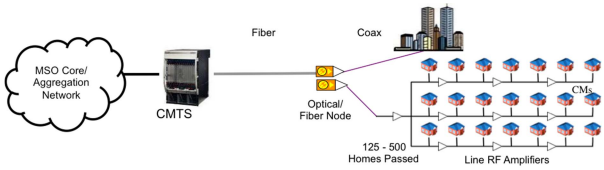
Distributed CCAP implementations can bring significant benefits in certain HFC network deployments. It will enable higher PHY layer performance for DOCSIS 3.1, and reduce space and power needs at the headend. The paper describes various CCAP architectures and how they handle data and video.

BASELINE ARCHITECTURE

Cable operators implement and deploy IP High Speed data, Linear Broadcast Video, Video on Demand, Voice, and various other integrated services to their end customers. The primary choice of access technology for this today is DOCSIS over a hybrid-fiber/coax (HFC) cable network.

Data Service

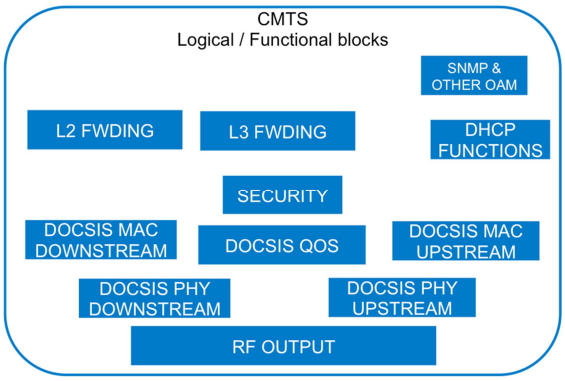
The DOCSIS system allows transparent bi-directional transfer of Internet Protocol (IP) traffic between the cable system head-end and customer locations over the HFC network. The Cable Modem Termination System (CMTS) is the central platform in enabling High speed Internet connectivity over the Cable HFC network. The CMTS platform has evolved over time along with the DOCSIS Specifications. The CMTS provides the MAC and PHY layer connection to the Cable Modem (CM) at customer premise.



DOCSIS Access Network

The CMTS consists of various logical functional components, at a high level these are as follows:

- DOCSIS PHY Layer
  - Upstream Receiver
  - Downstream Transmitter
- DOCSIS MAC Layer
  - Upstream MAC and Scheduler
  - Downstream MAC Processing
  - DOCSIS QoS
  - Security
- RF Output block
- L2 forwarding block
- L3 forwarding block
- IP Processing for DHCP
- SNMP agent / CLI, etc.

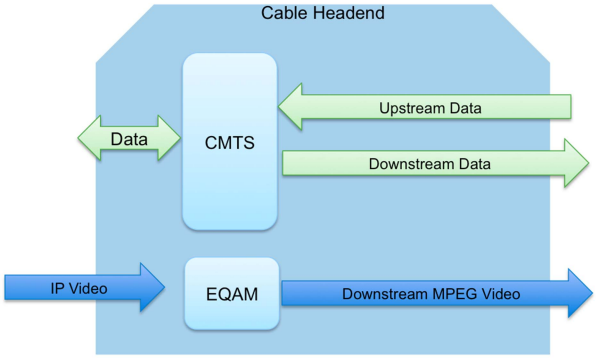


CMTS Functional Blocks

Video Distribution

The Video EdgeQAM is a key piece of equipment in any headend, hub site to enable video services. The role of the EQAM in the video-on-demand and switched-digital-video architecture is to receive an IP unicast or multicast stream containing MPEG transport

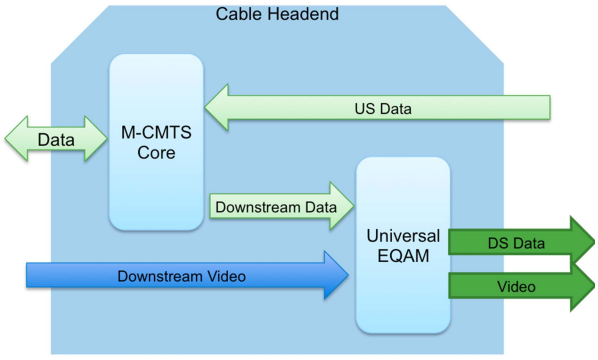
stream packets, and then produce that transport stream on one or more RF outputs for transmission over the HFC cable plant. The EQAM contains the multiple Gigabit Ethernet (GigE) data inputs. The content sent to an EQAM over the GigE interface is framed in an MPEG-2 transport stream (TS). An EQAM will support both Single Program Transport Stream (SPTS) and Multiple Program Transport Streams (MPTS). The EQAM can also re-stamp PCR timestamps for de-jitter processing, which helps reduce network impairments.



Baseline Architecture, Data & Video

MODULAR HEADEND ARCHITECTURE

Over time there have been many steps in the evolution of the CMTS platform. One of the first steps was the creation of the Modular Headend Architecture (MHA), which essentially separated out the DOCSIS downstream PHY layer out of the CMTS and moved it to a separate EQAM device.

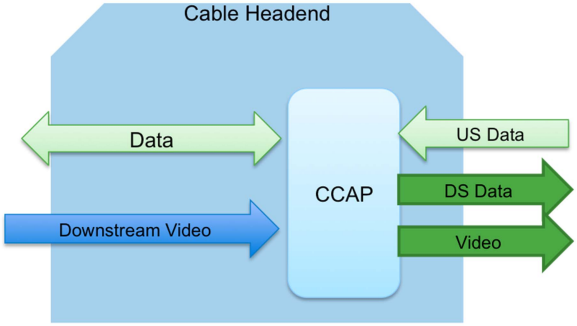


Modular Headend Architecture

A new interface called the DEPI (Downstream External PHY interface) was defined to support sending the data from the CMTS Core to the EQAM. The idea was to reuse the EQAM to modulate the bits on to the wire for both downstream DOCSIS data as well as MPEG video. The video EQAM now becomes a universal EQAM handling both video and DOCSIS data as inputs. The upstream receiver remains at the CMTS core. Since the DOCSIS MAC and PHY were separated in the MHA architecture, a new DOCSIS Timing Interface was introduced to keep the two devices closely synchronized. The Modular CMTS was essentially a two separate platforms as compared to the fully integrated CMTS.

INTEGRATED CCAP

The next big step in the evolution of CMTS platforms was the Converged Cable Access Platform (CCAP). The CCAP was intended to provide a new equipment architecture option for manufacturers to achieve increased Edge QAM and CMTS densities that MSOs require. The CCAP leverages existing technologies, including DOCSIS and also newer technologies such as Ethernet optics and EPON (Ethernet Passive Optical Network). The CCAP unifies the CMTS, Switching, Routing, and QAM functions at the headend, so that all data, video, voice functions can be handled over IP before conversion to RF or Optical signals. The CCAP eliminates the need for the combiner functionality in the headend.



Integrated CCAP

DOCSIS 3.1

DOCSIS 3.1 is the next generation in the evolution of DOCSIS and it brings some fundamental changes to the technology. DOCSIS 3.1 introduces Orthogonal Frequency Division Multiplexing (OFDM) as the new PHY layer technology, and allows for wide channels from 24 MHz to 192 MHz wide, moving away from the legacy 6 MHz sizes. DOCSIS 3.1 introduces Low Density Parity Check (LDPC) based Forward Error Correction (FEC).

DOCSIS 3.1 also introduces options for additional HFC spectrum, by allowing for expansion of the US Split (Mid-split or High-split). It also allows expansion of the downstream spectrum at the higher end. DOCSIS 3.1 LDPC FEC allows data transmission approaching the theoretical limits, and this enables 50% more efficient modulations such as 4096 QAM, harnessing more capacity on existing HFC networks. In the long run, DOCSIS 3.1 could support data rates reaching 10 Gbps downstream and 1 Gbps upstream.

CABLE OPERATOR CHALLENGES

Cable operators today face numerous challenges. The customer and market demands for higher bandwidth and data rates are increasing at a fast pace, and the demand for data and video services is forcing operators to continuously upgrade the plant capacity. As bandwidth demand continues to grow, facility space, power, and other related factors start becoming a concern. Below are some issues that factor into the discussion.

Analog optical networks between the hub and the fiber node:

- The length of optical link can be a limiting factor in managing the SNR performance.
- For DOCSIS 3.1 technology, capacity in the plant can be optimized without the analog optical noise floor.

Digital optics:

- Digital optics promise reduced OPEX as compared to analog optics. For digital optics, MSOs can use low-cost small form-factor pluggable (SFP) lasers instead of high-priced distributed feedback (DFB) lasers used in AM fiber. In this manner, they avoid the operations costs of maintaining rigorous performance on AM links as plant conditions change.
- The increased reach of the technology can be used to reduce facilities expense.
- Digital optics also provide a lower CAPEX for capacity growth. With a shift to digital optics, the throughput on the fiber becomes much greater (wavelengths can be packed much closer spectrally in the Dense Wavelength Division Multiplexing (DWDM) than with WDM spacing for analog optical signals. Typical spacing between wavelengths in AM mode is 100 GHz, which allows up to 40 wavelengths on a single fiber, whereas spacing for digital wavelengths is 25 GHz or less, which leaves room for 160 or more wavelengths. DOCSIS 3.1 modulations and/or capacity can be increased with no linear optics noise, and better SNR.

Facilities costs (space, power, HVAC):

- Increasing cost of distribution hub facilities cost is a concern for operators.
- As the power consumption in the hub escalates, distributed CCAP architectures allow reduction of facilities and power costs associated with hub and headend facilities.

Node + 0 architectures:

- There is a need to move to smaller service groups and to N+0 architectures to increase the bandwidth available to the user. The number of service groups needed might increase by a large factor with Fiber Deep HFC architecture and the

question is if the current CMTS platforms can keep up.

Low cost deployments:

- In some global markets, the need for low-cost smaller-scope deployment options are important to enable the success of cable broadband.
- Chinese MSOs in particular face unique challenges. Their plant architecture consists of a digital optical packet network (using point-to-point, EPON, GPON, etc.) to the MDU (Multi Dwelling Unit) and coax within the MDU. A centralized CMTS is not as economical in this architecture.

Smaller CMTS step functions:

- The industry needs smaller increments in cost and capacity when increasing capacity for just a few more users. The cost of adding a adding an entire new CMTS/line card for a minor increase in needs is sometimes cost prohibitive.

Flexibility and other benefits

- Distributed architectures move portions of the CMTS functionality into nodes; this gives operators flexibility and another option in their tool belt for deploying DOCSIS.
- There are also various benefits of a high rate digital backhaul which can be used for other services, such as wireless backhaul, business fiber extensions, and so forth.

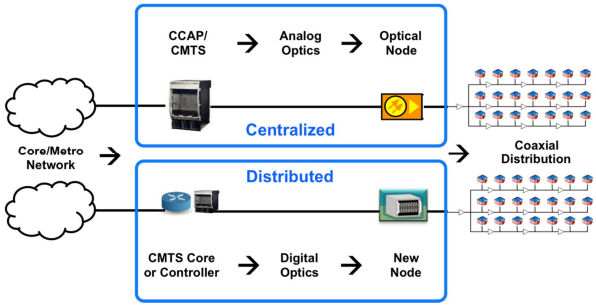
Timing:

- To get to the higher bandwidth channels allowed DOCSIS 3.1 in the upstream, an operator may be considering moving the US-DS split in the cable plant. This could be a disruptive change to the customers and would be a big effort for the operator. The same challenges exist for a Distributed Architecture deployment, which also will need to touch the fiber nodes.
- If the operators decides to make both changes, they could benefit by timing both the upstream split and deploying a distributed solution together.

As a result, MSOs today are investigating distributed CCAP solutions, in addition to leveraging digital optics in the access network. The question is whether the benefits of moving away from AM optics outweigh the costs and risks of placing processing and RF hardware in an environmentally hardened chassis on the outside plant or in an MDU basement.

DISTRIBUTED CCAP ARCHITECTURES OVERVIEW

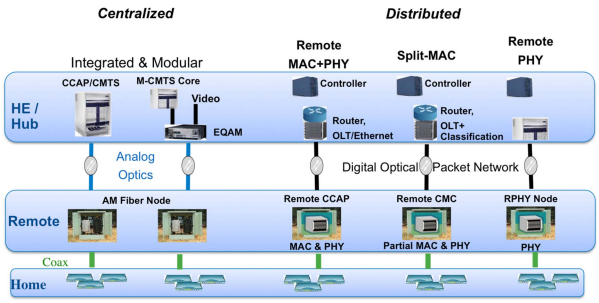
There are multiple distributed CCAP architectures are emerging in the marketplace. The basic idea is to distribute some or all of the functionality of the CMTS/CCAP down to a remote location, like the Fiber Node.



Digital Optics and Shifting of Functionality

There are 3 distributed architectures that have come forth so far. These are the Remote PHY, Remote MAC-PHY and the Split-MAC variations.

The concept behind Remote MAC-PHY is to move the entire CMTS/CCAP into the Remote node. The idea behind the Remote PHY is to split the CMTS between the MAC and the PHY Layers and move the PHY layer to the Remote Node.



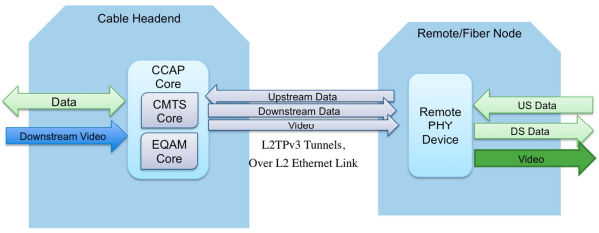
Centralized and Distributed Architecture Summary

The Split-MAC is in between the above two options, where some of the MAC functionality is defined and left at the headend and the remaining MAC and PHY functionality are moved to the Fiber Node. This option is defined in the C-DOCSIS System specification. The rest of this paper is focuses on the Remote MAC-PHY and Remote PHY options, as these have gained more traction in the American, European and China markets, whereas the Split-MAC has gained limited traction in the China market.

The foundation for Distributed CCAP Architectures is a digital optical plant, which essentially makes the connection between the head-end and the fiber node a Layer 2 Ethernet connection. In a digital HFC plant, the fiber portion utilizes a baseband network transmission technology such as Ethernet, EPON (Ethernet over Passive Optical Networks), GPON (Gigabit Passive Optical Network), or any layer 2 technology that would support a fiber-based PHY layer.

REMOTE PHY

The Remote PHY technology is also know as MHAv2 (Modular Headend Architecture version 2) as in many ways it builds on the original MHA architecture. MHAv2 uses a Layer 3 pseudowire between a CCAP Core and a set of Remote PHY devices. One of the common locations for a Remote PHY device is the optical node device that is located at the junction of the fiber and coax plants.



Remote PHY Architecture

In a Remote PHY System, the integrated CCAP is separated into two distinct components. The first component is the CCAP Core and the second component is the R-PHY Device (RPD). The CCAP Core contains both a CMTS Core for DOCSIS and an EQAM Core for Video.

The CMTS Core contains the DOCSIS MAC and the upper layer DOCSIS protocols. This includes all signaling functions, downstream and upstream bandwidth scheduling, and DOCSIS framing. The DOCSIS functionality of the CMTS Core is defined by the existing DOCSIS Specifications. The EQAM Core contains all the video processing functions that an EQAM provides today.

The Remote PHY Device contains mainly PHY related circuitry, such as downstream QAM modulators, upstream QAM demodulators, together with pseudowire logic to connect to the CCAP Core. The RPD platform is a physical layer converter whose functions are:

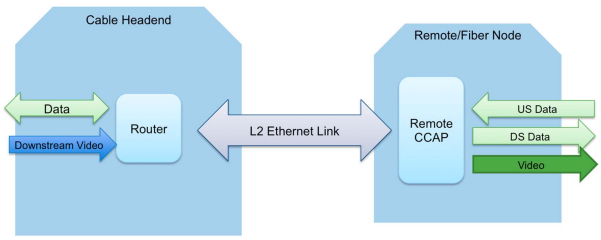
- To convert downstream DOCSIS, MPEG video and OOB signals received from a CCAP Core over a digital medium such as Ethernet or PON to analog for transmission over RF or linear optics.
- To convert upstream DOCSIS, and OOB signals received from an analog medium such as RF or linear optics to digital for transmission over Ethernet or PON to a CCAP Core.

REMOTE MAC PHY

The Remote MAC PHY technology moves both the DOCSIS MAC and PHY layers down to the Fiber Node. The connection between the Headend and the node is essentially a Layer 2 Ethernet connection. There are two options for this, which are different based on how video is handled. In both cases the data forwarding CMTS functionality is at the remote node. A compact CMTS is deployed at the fiber node and the CMTS NSI connects through the digital optical network back to the cable headend. For video, there are two options, as described below: Remote CCAP or Remote CMTS+Divided EQAM.

Remote CCAP

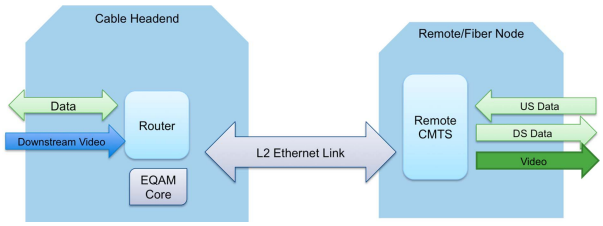
The RemoteCCAP term applies to an architecture where both the data and the video functions are moved to the Remote node. The CMTS functionality and the EQAM functionality are completely moved to the Fiber Node, and hence the term Remote CCAP. The Video and data transit the L2 Ethernet link like any other IP traffic. The Video needs to be encrypted to protect from unauthorized access at the Remote node.



Remote MAC-PHY Architecture: Remote CCAP

Remote CMTS + Divided EQAM

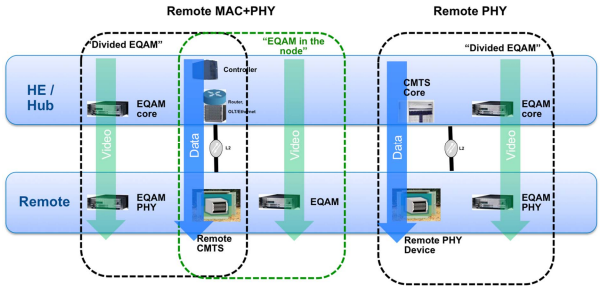
The Remote CMTS term applies to an architecture where only the data/CMTS functionality is moved into the remote node. The video/EQAM functionality is divided between the headend and the remote node, as in the Remote PHY architecture. The video MPEG packet processing is handled in the headend by an EQAM core device and the EQAM PHY inside the Remote CMTS handles the modulation of the video onto the wire.



Remote MAC-PHY Architecture: Remote CMTS + Divided EQAM

COMPARING DISTRIBUTED ARCHITECTURES

Due to the digital optical plant needed, any of the distributed architectures have PHY layer performance gains, which are very helpful in getting to the higher order modulations as defined in DOCSIS 3.1 (1024 QAM/4096 QAM). The digital fiber link can give 3–7dB gains, which can bump up the possible modulation order, increasing capacity.



Remote MAC-PHY and Remote PHY Architecture Comparison

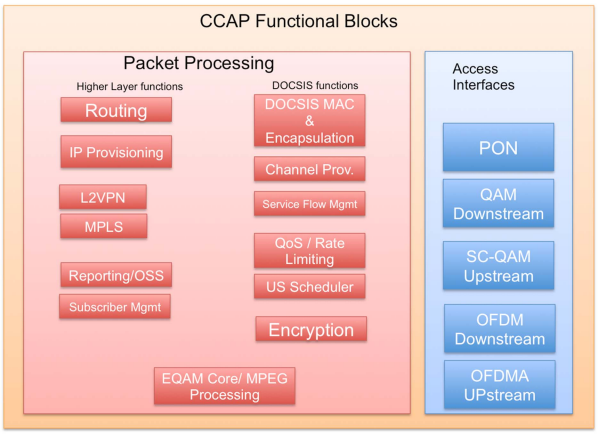
The Remote PHY architecture keeps the Remote Node device relatively simple, with only the PHY layer modulation in the Remote Node. There is also added communication between the CMTS MAC and PHY layers to setup the connections as needed, but this is needed only at startup.

In Remote MAC-PHY architectures the CMTS is condensed and sized for a node, although now with the MAC processing adds more intelligence to the device. But given that the CMTS design is known and has been used for a long time, the operators may prefer this familiar provisioning and management environment.

SYNERGIES WITH VIRTUALIZATION

As computation power available in the microprocessor platforms today increases, the cost of Common-Off-The-Shelf (COTS) servers is going down dramatically. This is leading to a huge boom in virtualization, where operators are virtualizing various functions within their networks. Considering this new functionality, which can be implemented in the customer premise, instead of implementing new ASIC designs and rolling out a new CPE device, operators now have the ability to create those services on the cloud and direct user traffic through each of those virtual functions.

Similar ideas on virtualization will affect the CCAP devices and their implementations. The main PHY layer functionality, which modulates the bits onto the wire need to be on the cable plant. But essentially all other functionality could become candidates for virtualization.



Virtualization Candidates within a CCAP

In a distributed CCAP architecture there are various pieces, which naturally gravitate towards virtualization. It is possible to run many of the CCAP packet-processing functions inside a virtual machine in the headend. This can start of with some of the easy candidates (labeled *Higher layer functions* in the diagram above) such as routing functions, IP/DHCP provisioning functions, OSS and reporting functions, and subscriber management functions.

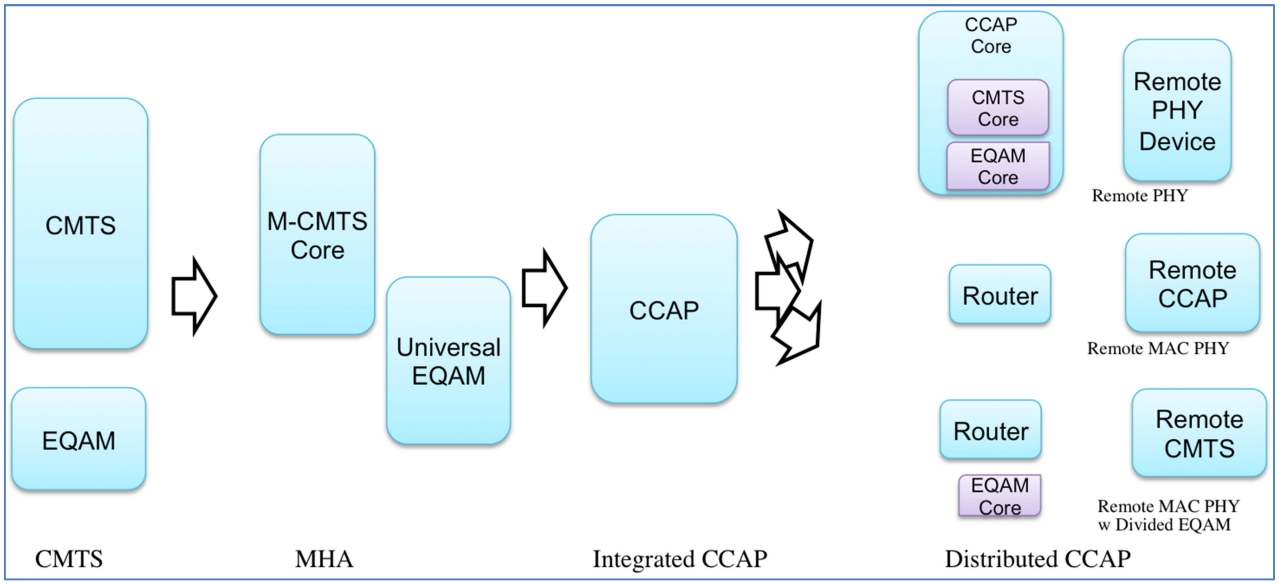
For a Remote MAC-PHY architecture, the above set of features may be all that can be achieved, as the remaining functions would already be implemented in the hardware at the Remote CMTS device. In the Remote PHY architecture, since the MAC level functionality is being implemented in the CMTS -Core at the headend, many of the higher-level functions as well some of the core DOCSIS functions (as labeled in the diagram) can be virtualized and be run from a remote data center. The main concerns at this point are the latency delays of such architecture. Whether those problems are surmountable remain to be seen. If the trend towards virtualization persists, and the latency problems are ironed out, one can imagine operators in the future running lightweight DOCSIS CCAPs in the cable plant. The CCAPs will act as a media converter, and all the intelligent functionality of the CCAP will be implemented and controlled from the cloud.

EVOLUTION OF HFC NETWORK

This paper has described the various CMTS/CCAP architectures seen by the cable industry so far. The Integrated CMTS/CCAP is a staple of the operators and that platform here is to stay for quite some time as it meets many of the current needs. Driven by specific future needs and competitive pressures in different parts of the plant, the cable operators will assess their need to go to Distributed CCAP architectures. The choice between

technologies is complex and involves numerous tradeoffs.

In the future, operators will be transitioning most of their services including video to run over IP/DOCSIS. The different CCAP architectures, either the integrated or the distributed platforms, will be used in in different parts of the network. The platform, which helps the operators gain the most bandwidth and operational efficiencies for the cost and help in the IP transition of services, will become the favored approach.



Evolution of the CCAP Architecture

HFC IMPROVEMENT FOR DOCSIS 3.1 EVOLUTION

Maxwell Huang
Cisco Systems

Abstract

The DOCSIS 3.1 PHY and MAC standards have specified the QAM modulation order as high as to 16384QAM, however, we could not go beyond 512QAM (or 1024QAM) order on the current HFC plant as the CNR performance is primarily limited by analog optics and long amplifiers cascade. Hence, to achieve the 10G cable access goal we will have to expand the cable spectrum to a very high frequency like 1794MHz. But the 1794MHz expansion in cable spectrum appears too challenging and leaves tremendous concerns…

The paper aims to explain the ideas of evolving the HFC plant to have the better CNR performance and hereby be able to support the high QAM modulation like 4096 QAM, so that we may just need to expand the cable spectrum to a relatively lower frequency like 1.44GHz to make HFC a 10G access network.

1.0 HFC Challenges in DOCSIS 3.1 Evolution

In order to provide the 10Gbps high speed data through the existing Hybrid Fiber Coaxial (HFC) network to be able to compete with other access technologies like 10G PON and EPOC, many of cable operators and CableLabs worked together for the standardization of the DOCSIS 3.1 (D3.1) which defines the specifications for the 4th generation of high-speed data-over-cable system primarily including PHY, MAC, MULPI, CCAP and CMSI etc.

The DOCSIS 3.1 evolution is to get the robust data throughput increase through two phases. The phase 1 aims for 7Gbps downstream (DS) data throughput and 1Gbps upstream (US) data throughput, and the phase 2 shoots for the 10Gbps DS throughput and 2.5Gbps US throughput. To get such robust increase in throughput, there are some major changes being introduced into the DOCSIS 3.1 evolution as follows.

- Involve the Orthogonal Frequency Division Multiplexing (OFDM) to improve the date rate per cable spectrum (bit / Hz) efficiency
- Involve the Low Density Parity Check Code (LDPC) for the OFDM FEC to further improve the bit / Hz efficiency
- Expand the bandwidth to the 1.218GHz in phase 1 and to 1.794GHz (could change later) in phase 2 to ultimately explore the cable spectrum for use of adding more QAM / OFDM channels
- Add the OFDM channels primarily above 1GHz, gradually shut down traditional analog channels and SC-QAM channels and replace with the OFDM channels

The DOCSIS 3.1 PHY and MAC standards have specified the QAM modulation order as high as to 16384QAM, however, we could not go beyond 512QAM (or 1024QAM) order on the current HFC plant as the Carrier to Noise Ratio (CNR) performance is primarily limited by analog optics and long amplifiers cascade (N+5, N+7 etc). Hence, one way to achieve the 10G cable access goal is to expand the cable spectrum to a very high frequency like 1794MHz.

Downstream Data Throughput (long amplifier cascade, increased OMI)	Band	MER	QAM Modulation	Date Rate (6MHz CH)	Total Data Throughput
	(MHz)	(dB)		(Mbps)	(Mbps)
Current Capacity	105 - 1002	33	256	38	5681
Capacity increase by OFDM w/ LDPC	105 - 1002	36	1024	49	7326
Capacity increase by OFDM w/ LDPC & Spectrum Expansion	258 - 1218	36	1024	49	7840
Capacity increase by OFDM w/ LDPC & Spectrum Expansion	500 - 1794	36	1024	49	10568

Table 1

The Table 1 shows the phased data throughput increase by the OFDM w/ LDPC and the spectrum expansion on top of the available Modulation Error Ratio (MER) performance primarily associated with CNR & distortion

performance on the HFC network. Note that the actaul data throughput shall be getting less when taking the guard band of OFDM channel into account.With the introduction of the OFDM w/ LDPC we can go to the 1024 QAM from today's 256 QAM to increase the data rate by approximately 29% with the same CNR as today.

But to achieve the 10Gbps cable access the cable specturm still needs to expand to an extremely high frequency like 1794MHz that appears too challenging as the loss at 1794MHz is huge (could go as high as 70dB), and therefore leaving tremendous challenges and concerns in chip and product development particularly at the reliability associated with significant power increase.

The band from 500MHz to 1794MHz contains 217 QAMs (in 6MHz bandwidth), it accounts for 0.7dB increase in the composite power at input of transmitter (Tx) compared to 185QAMs in a band from 108MHz to 1218MHz. Unless we use better laser which could be costly, the OMI to Tx will have to reduce 0.7dB accordingly, otherwise the increased power will hurt distortion performance.

We also need to increase the reverse spectrum to approximately 400MHz to achieve 2.5Gbps target upstream data throughput as shown in the Table 2.

Upstream Data Throughput (long amplifier cascade)	Band	MER	QAM Modulation	Date Rate (6MHz CH)	Total Data Throughput
	(MHz)	(dB)		(Mbps)	(Mbps)
Current Capacity	5 - 85	27	64	28	373
Capacity increase by OFDM w/ LDPC	5 - 85	30	256	38	507
Capacity increase by OFDM w/ LDPC & Spectrum Expansion	5 - 204	30	256	38	1260
Capacity increase by OFDM w/ LDPC & Spectrum Expansion	5 - 400	30	256	38	2502

Table 2

As a result, the upstream RF power from subscribers CPE's will have to increase according to the cable and Tap losses at 400MHz. Also, the reduction of the Noise Power Ratio (NPR) range because of the increased power loading on the upstream is the big concern against the current HFC architecture where sufficient operational headroom is needed to maintain against the variations from the analog fiber transporation and long cascade of amplifiers.

The following secions will explain the ideas of evolving the HFC plant to have the better CNR performance and hereby be able to support the high QAM modulation like 4096 QAM or even higher modulation, so that we may just need to expand the cable spectrum to a relatively lower frequency like 1.44GHz to make challenge down to a manageable degree but still accomplishing a 10G cable access network.

2.0 HFC Performance Impact By Different DOCSIS 3.1 Loadings

The Figure 1 depicts the phased data throughput increase through DOCSIS 3.1 evolution on a spectrum loading view. It also depicts complexity of the spectrum loading at the phase 1 as there could be a few interim upgrades going on. As DOCSIS 3.1 standards allow cable operators to gradually (rather than "folklift" reconstruction) shun down traditional analog channels and Single Carrier Quadrature Amplitude Modulation (SC-QAM) channels, and switch to OFDM channels, there will be the mixed loadings of analog, SC-QAM and OFDM channels coexisted in the same cable spectrum.

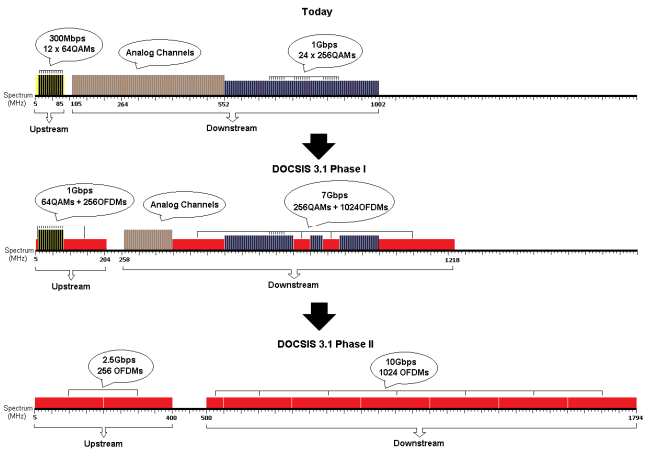


Figure 1

Following let's take some case studys to see if we have opportunities for the HFC performance improvement through the interim spectrums.

2.1 Case Study I _ "Heavy Analog" to "Light Analog"

For instance there is a HFC network where 74 analog channels are loaded from 54MHz to 550MHz and above 550MHz to 1002MHz 75 SC-QAM channels are loaded. Now for the sake of an interim throughput increase, a cable operator plans shutting down 48 analog channels to free up 288MHz band and filling in a 192MHz OFDM channel and a 96MHz OFDM channel. This upgrade scenario is from so called a traditional "heavy analog" to a "light analog".

The Figure 2 depicts two options in freeing up the band where analog channels are presently loaded and adding one 192MHz OFDM channel and one 96MHz OFDM channel at the freed band in addition to the OFDM channel added above 1GHz to 1.2GHz. i.e. option I is to shut down analog at the band from 264MHz to 552MHz, and option II is to shut down analog at the band from 105MHz to 393MHz.

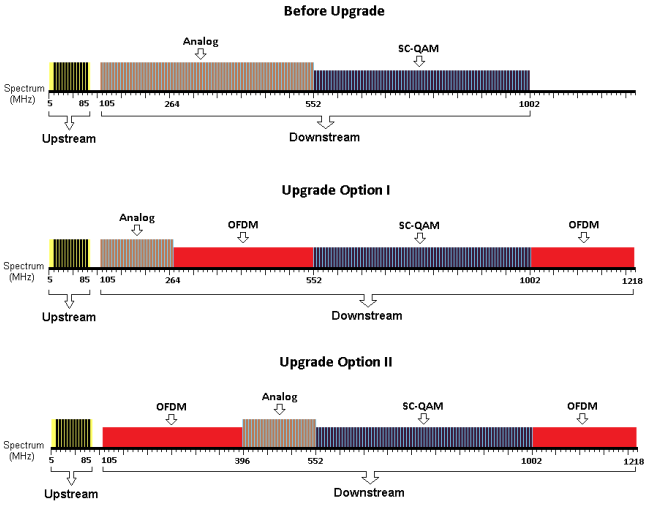


Figure 2

So, which is the better option from the HFC performance perspective? Followings will try to answer this question. Assume that the level of the newly added OFDM channels are 6dB down to those analog channels at the transmitter input, then we can expect the less composite power for this "light analog" compared to "heavy analog" at the input of forward optical transimtter and the output of nodes and amplifiers. Following table shows the calculated results of the Composite Power (CP) vs. channel loadings at transmitter input and node/amp output.

Scenarios	Level / Tilt at Fwd Tx, Node/ Amp	Loading Description	CP (dBmV)
Before Upgrade	15dBmV(QAM 6dB down), flat at Tx Input	108 – 552MHz Analog; 552 – 1002MHz QAM	34.68
	58dBmV(QAM 6dB down), 14.5dB tilt at Node/Amp Output		70.35
Upgrade option I	15dBmV(QAM 6dB down), flat at Tx Input	108 - 264MHz Analog; 264 - 552MHz OFDM; 552 - 1002MHz QAM; 1002 - 1218MHz OFDM	33.19
	58dBmV(QAM 6dB down), 18dB tilt at Node/Amp Output		68.85
Upgrade Option II	15dBmV(QAM 6dB down), flat at Tx Input	108 - 396MHz OFDM; 396 - 552MHz Analog; 552 - 1002MHz QAM; 1002 - 1218MHz OFDM	33.19
	58dBmV(QAM 6dB down), 18dB tilt at Node/Amp Output		69.15

Tabel 3

We can see that at the transmitter input both the options have the composite power (total OMI) 1.47dB less than that before upgrade which is because the newly added OFDM channels have the 6dB backoff relatively to the analog channels. In principle we can increase the OMI per channel by 1.47dB to improve the SNR by 1.47dB while unlikely degrading distortion performance at the transmitter side since the CP is ketp unchanged.

The added QAMs will introduce more noise - like intermodulation products and may therefore degrade the Carrier to Intermodulation Noise (CIN) performance a little bit in case new distortion products have higher level than existing ones. The CIN degradation may or may not affect MER performance depending where is applied. For purpose of simplicy this paper will not include MER degradation possibly introduced by the added QAMs into analysises.

At nodes and amplifiers side we can see the different reductions in the composite power between two options that the option I is reduced to 68.85dBmV and option II to 69.15dBmV. The difference comes from the tilt used for node /amp. The Figure 3 pictures how the tilted output of the nodes and amp makes difference in the composite power for the option I and option II.

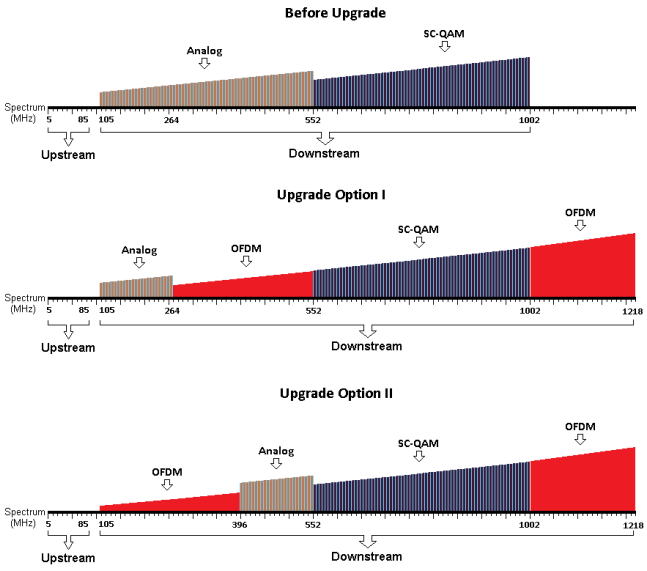


Figure 3

For composite power wise we could increase the node / amp output level by 1.5dB for option I and by 1.2dB for option II for the CNR improvement while maintaining the equivalent distortion performance at the node / amp output.

The option I sounds to be the better choice because if cable operators increase the OMI per channel at transmitter by 1.47dB for the SNR improvement, then the output levels of nodes and amps are automatically increased by 1.47dB, hence if we choose the Option I we can keep the CP on those nodes and amps' output to avoid hurting the distortion performance.

Is there something beyond composite power that matters? Having the QAMs (SC-QAM and OFDM) locate in the continuous bands should yeild the better CIN performance than that be spread out in discrete bands. And when the numbers of the analog channels is reduced from 75 to 30, the numbers of CTB and CSO shall have the significant reduction from approximately 200k down to 15k. The positive impact from a reduced number of analog channels may cancel the negative impact from an increasing number of QAM channels from distortion perspective, but let's get it verified before making conclusion.

2.2 Case Study II _ "Light Analog" to "All QAM"

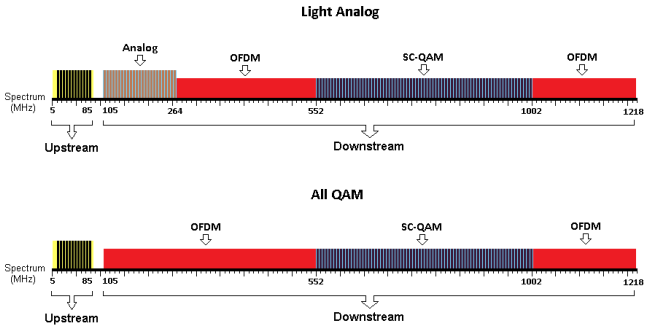


Figure 4

The Figure 4 illustrates an upgrade from a light analog to the All QAM. The All QAM refers to that the entire forword spectrum is loaded by either SC-QAM or OFDM channels but no analog channels. We can do same analysis for this upgrade scenario. The table 2 shows the difference in composite power between the "light analog" and "All QAM.

Scenarios	Level / Tilt at Fwd Tx, Node/ Amp	Loading Description	CP (dBmV)
Light Analog	15dBmV(QAM 6dB down), flat at Tx Input	108 - 262MHz Analog; 264 - 552MHz OFDM; 552 - 1002MHz QAM; 1002 - 1218MHz OFDM	33.19
	58dBmV(QAM 6dB down), 18dB tilt at Node/Amp Output		68.85
All QAM Option I	15dBmV(QAM 6dB down), flat at Tx Input	108 - 552MHz OFDM; 552 - 1002MHz QAM; 1002 - 1218MHz OFDM	31.67
	58dBmV(QAM 6dB down), 18dB tilt at Node/Amp Output		68.67
All QAM Option II	15dBmV(QAM 6dB down), flat at Tx Input	108 - 552MHz OFDM; 552 - 1002MHz QAM; 1002 - 1218MHz OFDM	31.67
	58dBmV(QAM 6dB down), 25dB tilt at Node/Amp Output		67.32

Tabel 4

As the remaining analog channels are replaced with OFDM channels, the CP at the Tx input is reduced by 1.52dB. Obviously, one should consider increasing the channel level to Tx by 1.52dB to improve the CNR. However, there is only 0.18dB reduction at the node output if the 18dB tilt is still being used. We need to find a way to get same amount of CP reduction at node and amp side to eliminate the concern at distortion performance degradation.

Fortunately, we can resolve the concern by increasing the tilt of node. The Table 2 also includes the All QAM option II which is to use the 25dB tilt instead by which one can have the 1.53dB CP reduction at node and amp outputs. By changing the tilt one can improve the CNR and unlikely causing the distortion degradation at nodes and amps.

Somebody may worry about the NF degradation for low band concerning the 25dB tilt being used. Yes it does affect the NF at equalizer's starting frequency of 52MHz, however, for the case discussed here at 105MHz the NF reduction is only about 0.3dB, and for above frequencies the degradations are getting even smaller. Since the optical link is usually the more domindant factor than amplifiers cascade for the link CNR performance, this 0.3dB degradation in node and amp's NF would merely make 0.1dB degradation for the overall link CNR performance.

2.3 Performance Impact By PAPR of OFDM

The Peak to Average Power Ratio is considered the primary disadvantage of the OFDM due to the fact that OFDM symbol has Gaussian amplitude distribution (due to its multicarrier nature). It is true, but mainly in a comparison to a single channel or a small number of channels. The PAPR impact from an OFDM channel with 192MHz bandwidth is at comparable level if compared it to the composite effect of the SC-QAM channels occupying the 192MHz bandwidth as well.

Furthermore, unlike single-carrier, OFDM offers ways of reducing peak-to-average power. One such method illustrated using this graph is called tone reservation. In this method a few (< 1%) of the tones are reserved to reduce the high amplitudes in an OFDM FFT. The results shown have been obtained by simulating the specific method given in the DVB-T2 specification. It is seen that the peak power of OFDM can be made to be less than four single-carrier channels at clipping probabilities of interest to cable applications. Another way is to implement the CCDR.

Nevertheless, the MER degradation had been noticed when node or amp with QAMs loaded is about to saturate. The PAPR effect drives the gainblocks in node or amp into non-linear region and therefore causing the unnegligible MER degradation. Particularily for All QAM loading, the overall PAPR impact shall increase in light of the OFDM's that are in replacement with analogs and added at newly expanded spectrum. Thus, some compression margin must be considered for node and amp's application to avoid MER degradation.

2.4 Reverse Analysis

The Composite Power (CP) of the reverse spectrum is continuingly increasing as more and more SC-QAM and / or OFDM channels are added into rerverse spectrum as driven by CAGR. On the following table the 26dB of link gain is assumed, we can see how the CP is increased over the reverse spectrum loadings.

Loading	Level at Rev Tx / Rx	CP (dBmV)
4 ch (5-42MHz)	17dBmV at Tx Input	23.02
	43dBmV at Rx output	49.02
8 ch (5-85MHz)	17dBmV at Tx Input	26.03
	43dBmV at Rx output	52.03
31 ch (5-204MHz)	17dBmV at Tx Input	31.91
	43dBmV at Rx output	57.91

Tabel 5

The CP increase analysis above is for a fixed Tx input level, however we actually care more about the Noise Power Ratio (NPR) range for the reverse band concerning cable temperature variation, reverse noise funneling effect, ingress noise and component variations etc. NPR stands for a dynamic level range in which the CNR or the MER or the Bite Error Ration (BER) can meet or exceed a given threshold. It is interesting to study how much NPR is reduced becaused of the CP increase and if there are ways to improve it.

For the data shown on the Figure 4 it comprises of low, mid, and high frequency splits (edge frequenies of 42MHz, 85MHz and 204MHz) using 64 QAM and 256 QAM channel loading, for an analog DWDM return transmitter, operating at +8dBm output power over a 16dB optical link (40km of fiber plus 8dB of passive loss).

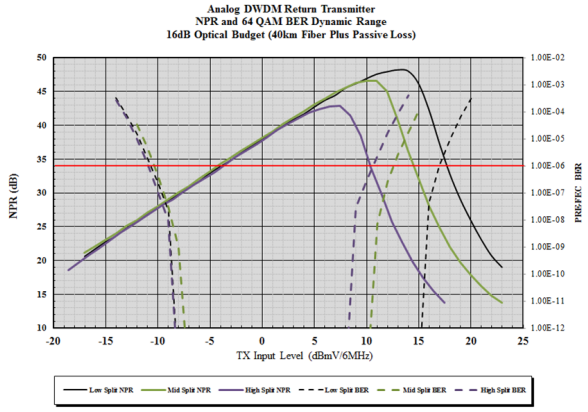


Figure 5

The plot shows the measured BER as a function of incremental RF input level. The left hand side of the curve is generally regarded as the "CNR" side as there is very little to no distortion occurring as RF levels are too low to cause it. The BER increases as RF level increases because the signal is increasing relative to the system noise floor. The right hand side of the curve is generally referred to as the "clipping" or "distortion" side of the curve as bit errors are being caused primarily by signal distortion due to overdriving the reverse gainblock in the analog return or A/D converter input in the digital return.

Looking at the NPR results of the three splits we can see the right side curve is shifted, in other words the dynamic range is reduced, because the increased CP makes mid split and high split enter the non-linear zone earlier than low split. We can see the reduced range is about 2dB bigger than the increased CP because of the nature of the non-linear clippng.

In optical networking, the amount of dynamic range for a given modulation format needs to be considered to ensure proper operation of the transmitter under fielded conditions. Typically, 12dB of operational headroom has been recommended for robust operation. Unless we can significantly imporve the noise or distortion performance for whole reverse system, but that is too challenging, the NPR reduction can not be avoided and appears to be a big concern against the DOCSIS 3.1 evolution. We should consider smaller node sizes and shorter cascades to reduce the amount of ingress noise and the impact of temperature and component variations, and using the analog DWDM lasers with tightly controlled over temperature.

3.0 HFC Performance Improvement By Fiber Deeper

In regard to the concerns mentioned in previous section we should look at reducing the number of the cascaded amplifiers which we call it the fiber deeper. Let's take a N+7 (a node plus 7 amplifiers) as the benchmark to see how much CNR can be improved by going for different scenarios of N+5, N+3, N+1 and N+0.

3.1 Forward Analysis

Majar Assumptions for CNR Analysis to the analog channels and the QAM channels:

- NF of Single Amplifier: 8dB
- Input Level To Each Amplifer: 8dBmV
- Wavelength: 1310nm
- OMI / analog channel: 3%

- OMI / QAM channels: 1.5%
- Nyquist noise bandwidth: 4MHz
- Symbol rate of QAM channels: 5.0569Msym/sec
- EIN of Receiver (Rx): 5.5 pA / Hz
- Optical Input Power to Rx: -2dBm
- Channel Loading: 79 analog plus QAMs to 1218MHz

Following let's see the End of Line (EOL) CNR comparison for Analog and QAM channels,

Scenarios	N+7	N+5	N+3	N+1	N+0
EOL Analog CNR	48.1	48.8	49.7	50.8	51.5
Δ (N+7)	0.0	0.7	1.6	2.7	3.4
EOL QAM CNR	43.4	43.7	44.1	44.4	44.6
Δ (N+7)	0.0	0.3	0.6	1.0	1.2

Table 6

If we change the optical input power to Rx to -6dBm while keeping other assumptions unchanged, we will get the following results.

Scenarios	N+7	N+5	N+3	N+1	N+0
EOL Analog CNR	45.6	46.0	46.4	46.9	47.2
Δ (N+7)	0.0	0.4	0.8	1.3	1.6
EOL QAM CNR	39.7	39.9	40.0	40.1	40.2
Δ (N+7)	0.0	0.1	0.3	0.4	0.5

Table 7

We can see that the overall CNR improvement for the -6dBm input to Rx is less compelling than that for the -2dBm input. That is because the optical link (consists of transmitter, fiber and node receiver) becomes a more dominant factor than the RF amplifiers cascade at the lower optical input. It means the analog optics is distance depedant (Rx getting lower optical power). But in the HFC fiber deep networks, Rx are expected to get lower optical input power because of the use of DWDM / CWDM supporting an increasing number of fiber deep nodes.

Per the previous study for respective scenarios of "Heavy Analog" upgrade to "Light Analog" and "All QAM" in the section 2.1, the OMI can increase 1.47dB and 1.52dB respectively and unlikely hurting distortion performance. In light of HFC networks are already on the way (or planned) converting to All QAM loading, we could assume the 3dB (≈ 1.47dB + 1.52dB) increase in OMI from "Heavy Analog" to "All QAM" upgrade, to see the combination effect of going for All QAM loading plus Fiber Deeper.

Majar Assumptions for CNR Analysis for All QAM loading:

- NF of Single Amplifier: 8dB
- Input Level To Each Amplifer: 8dBmV
- Wavelength: 1310nm
- OMI / QAM&OFDM channels: 2.12%
- Symbol rate of QAM channels: 5.0569Msym/sec
- Optical Input Power to Receiver: -6dBm
- EIN of Receiver: 5.5 pA / Hz
- Channel Loading: All QAMs from 105MHz to 1218MHz

Scenarios	N+7 (1.5%OMI)	N+7 (2.12%OMI)	N+5 (2.12%OMI)	N+3 (2.12%OMI)	N+1 (2.12%OMI)	N+0 (2.12%OMI)
EOL QAM CNR	39.7	42.3	42.5	42.8	43.0	43.2
Δ (N+7)	0.0	2.6	2.8	3.1	3.3	3.5

Table 8

Now the CNR improvement by combinaion effect of the OMI increase and the Fiber Deeper looks significant.

Distortion improvement is another major benefit from going for Fiber Deeper.

Majar Assumptions for Distortion Analysis:

- Channel Loading: 79 analog plus QAMs
- CTB of Single Node / Amp: 73dB
- CSO of Single Node / Amp: 65dB
- CCN of Single Node / Amp: 52dB

Scenarios	N+7	N+5	N+3	N+1	N+0
CTB	54.9	57.4	61	67	73
CSO	56	57.2	59	62	65
CCN	43	44.2	46	49	52

Table 9

The distortion improvement as a result of reducing the number of cascaded amplifiers looks significant. The distortion improvement is indeed helpful to go for the higher modulation of QAM. The Fiber Deep HFC network will be also helpful for other HFC performance improvement in terms of HUM, group delay e.t.c.

3.2 Reverse Analysis

We can do same analysis for the upsteam path to see how CNR be improved by reduing the number of cascaded amplifers. However, the upstream CNR of a cable network is calculated somewhat differently than the downstream CNR. In the forward path, the network branches out from a common point—say, a node. The worst-case downstream CNR is almost always through the longest individual cascade of amplifiers. In the reverse path, the network combines at a common point—the node, hub site, or headend. This results in a reverse funneling effect for system noise and impairments. Instead of calculating the CNR for a given cascade of amplifiers, the upstream CNR accounts for all the reverse amplifiers that are connected to a common point.

Assume the current HFC network is N+7 where has 30 amps in total connecting to node. Let's take the following assumption for the reverse CNR analysis

Majar Assumptions:

- NF of Amplifier: 13.5dB
- Input Level To Amplifer: 10dBmV
- Wavelength: 1310nm
- OMI / channel: 3%
- Optical Input Power to Receiver: -13dBm
- EIN of Receiver: 2.5 pA / Hz
- Symbol rate of QAM: 5.12 Msym/sec

Scenarios	N+7	N+5	N+3	N+1	N+0
Amp Number	31	21	9	2	1
Headend CNR	30.47	30.64	30.85	30.98	31
Δ (N+7, Original)	0	0.17	0.38	0.51	0.53

Table 10

The CNR improvement by the Fiber Deeper looks relatively small. Unlike downstream links the upstream links usually don't use optical amplifiers, therefore reverse Rx in headend get low optical input power unless optical link is short. Hence the optical link primarily determines the overall upstream CNR performance.

However, it is absolutely wrong to conclude that Fiber Deeper be not helpful for the upstream improvement. By the the Remote PHY or Baseband Digital Return (BDR) technology, we can make the link CNR performance indepedant with optic distance. It will yield significant improvement for the overall link CNR performance together with the CNR improvement on the amplifiers cascade shown in Table 11.

Amp Number	31	21	9	2	1
Amp Cascade CNR	39.67	41.37	45.04	51.58	54.59
Δ (31 Amps, Original)	0	1.7	5.37	11.91	14.92

Table 11

The Figure 5 is the plot showing the reverse MER data vs. incremental input levels to Tx tested in an optical link with different optical input power (+2dBm, -8dBm and -18dBm) to Rx. We can see that the MER keeps about 1:1 relationship with the Tx input level 1dB before entering the non-linear zone of Tx. As in principle the 1dB increase in OMI makes CNR 1dB better. Thus for some operating range, we can count on the better MER on the improved CNR.

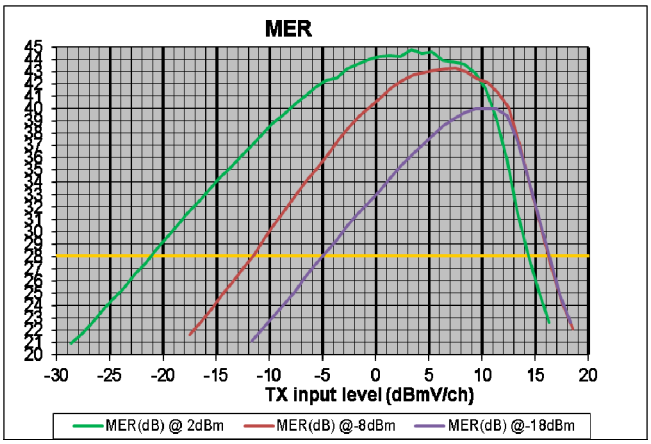


Figure 6

The above analysis is only from normal operation view, but in case there are some ingress noise the reverse CNR at the headend gets dramatically worse, for some cases it severerly impacts the BER and therefore reducing the data throughput. The prominent benefit from going fiber deeper is significantly reducing the probability and the degree of the ingress noise, hereby maintaining relatively stable BER performance through the high speed data rate.

More meaningfully, as the cable temperature variation and the probability of ingress noise are getting much smaller as a result of going fiber deeper, we can be safe to increase the reverse loading as the operational headroom is no longer remained that critical as before.

3.3 Bit / Hz efficiency Improvement By Fiber Deep And OMI Increase

By approaches of the OMI increase and the Fiber Deep we can see the 3.5dB CNR improvement in the Table 8 and 9dB CCN improvement in the Table 9 on the downstream, but considering the possibly bigger PAPR impact and worse CIN because more QAMs are added. Now let's assume 3dB MER improvement and see what we can benefit from a bit / Hz efficiency improvement view.

Table 12 shows that we can go 2048 QAM modulation on top of 39dB MER and therefore have approx 29% of the bit / Hz efficiency improvement. As a consequence, it requires the less cable bandwidth to achieve the 10Gbps DS rate.

Downstream Data Throughput	Band (MHz)	MER (dB)	QAM Modulation	Date Rate (6MHz CH) (Mbps)	Total Data Throughput (Mbps)
Capacity increase by OFDM w/ LDPC & Spectrum Expansion	500 - 1794	36	1024	49	10568
Capacity increase OFDM w/ LDPC & Fiber Deep	500 - 1680	39	2048	54	10620

Table 12

The cable loss at 1680MHz still appears high. We could consider Fiber Deeper architecture of N+1 or N+2 and shortening the cable lengths between node to amp (or amp to amp), so that the cable loss is reduced accordingly, hereby we can maintain relatively low design challenge for HFC product development.

The bit / Hz efficiency for the upstream is unexpected to improve if we keep using the analog optics as discussed in the section 3.2. Thus it sounds necessary to introduce the digital return technology like Remote PHY if we want to improve the bit / Hz efficiency for the upstream.

4.0 HFC Performance Improvement By Remote PHY

Heading to the Remote PHY architecture is definitely helpful for the HFC performance improvement. It is said that we can get the 3dB CNR improvement by moving to the remote PHY. More meaningfully, the optical link CNR of the Remote PHY system is no longer dependant on optical link distance.

In the analog optical transpotation system the noise contributions are from a number of factors in terms of Relative Intensive Noise (RIN) of laser; shot noise and thermal noise relevant to OMI, optical input power to Rx, photodiode's responsivity, noise current of amplifier etc; and also Amplified Spontaneous Emission (ASE) of EDFA when is being used.

However, in the Remote PHY digital transportation system, the link performance is distance independent – same MER for 0 km as for 100 km. The optical input power to Rx is no longer the key factor in determining the system CNR. The number of wavelengths used is not a factor since on/off keyed digital modulation only requires ~20dB of SNR; thus fiber cross-talk effects do not play a role in limiting performance in access-length links (<160 km).

The output of the receiver is no longer dependent on optical input power, which allows the operator to make modifications to the optical multiplexing and de-multiplexing without fear of altering node output levels. It is quite meaningful to support the fiber deep architecture.

4.1 Further Bit / Hz efficiency Improvement By Remote PHY

Let's see what happens on the bit / Hz efficiency improvement by the combination efforts of OFDM w/LDPC, All QAM loading, Fiber Deep and Remote PHY.

By moving to the Remote PHY and the Fiber Deep architecture the CNR performance is no longer limited by optical distance and cascaded amplifiers, but primarily determined by the reverse gainblock performance. In light of the reverse gainblock available today can meet the 39dB MER for more than 25dB of the input dynamic range, so let's assume the 39dB MER for the following analysis.

Upstream Data Throughput	Band	MER	QAM Modulation	Date Rate	Total
	(MHz)	(dB)		(Mbps)	(Mbps)
OFDMA w/ LDPC & Fiber Deep & Remote PHY	5 - 284	39	2048	54	2511

Table 13

Now it only requires expanding the reverse band to approximately 284MHz to be able to achieve the 2.5Gbps US rate goal. Compared to previously planned 400MHz, the output level from subscribers CPE's will have the significant reduction. A reverse band from 5MHz to 284MHz makes the less NPR reduction compared a band from 5MHz to 400MHz.

Now let's take 284MHz as the US edge, and with necessary crossover band needed for diplexer, we could determine 360MHz as the starting frequency of DS band. Similar to upstream now the MER performance is primarily determined by forward gainblock. In light of the forward gainblock available today can mostly meet the 40dB MER before entering non-linear region and the OFDM w/ LPDC can make another 3dB improvement, so let's assume the 42dB MER to do the following analysis.

Downstream Data Throughput	Band	MER	QAM Modulation	Date Rate (6MHz CH)	Total Data Throughput
	(MHz)	(dB)		(Mbps)	(Mbps)
OFDM w/ LDPC & Fiber Deep & Remote PHY	360 -1440	42	4096	59	10620

Table 14

It does make difference in the coaxial cable bandwidth required for achieving the target 10Gbps DS data throughput. The QAM numbers in a forward band of 360MHz to 1440MHz is even less than that of a band from 108MHz to 1218MHz, thus we could at least maintain the current OMI to forward Tx.

5.0 Summary and Conclusion

In addition to the introduction of OFDM with LDPC, we really can do something from the HFC side to help head to the target data throughput. We could take via several leaps as to expanding coaxial cable bandwidth, heading to Light Analog or All QAM loading, going to Fiber Deeper, and deploying Remote PHY.

The first leap introduced by the OFDM w/ LDPC makes approximately 29% increase based on the current infastructure, but there still remains the tremendous challenges to explore the cable bandwidth to 1794MHz and the concerns against the CPE's output level increase and the reduced operational headroom on the upstream. We will have to tremendously increase the gain, RF output power, bandwidth and power dissipation for

device and product to compensate the cable loss rolling out to 1794MHz DS and 400MHz US that makes the mission almost impossible. And if the long amplifier cascade story continues, the operational headroom reduction on the upstream will turn out to be a big concern against cable temperature variation, component variation, ingress noise and reverse funneling noise effect. We must consider other approach to lighten the concerns that have addressed.

Upgrading to "light Analog" or "All QAM" is something MSO's are going to, or plan to, do as driven by CAGR. We could take the opportunies to improve DS CNR by increasing the OMI to Tx as the second leap.

The third leap of going to the Fiber Deep architecture is the necessary evolution for HFC. With the Fiber Deep approach we can improve the bit / Hz efficiency with the improved CNR and distortion to require less cable bandwidth for achieving 10Gbps DS throughput; Have less concern against the reduced operational headroom on the upstream. However, the significant bit / Hz efficiency is unexpected on the upstream since the CNR performance is still primarily limited by use of the analog optics.

We could depoly the Remote PHY architecture as the fourth leap. Remote PHY is the significant value adder in helping imporve the system CNR and hereby requiring the even less cable bandwidth to achieve the target throughput. More meaningfully, as the Remote PHY is the optical distance indepedant, we can add multiplexing and de-multiplexing to support the Fiber Deep architecture without fear of degrading MER performance. And then together with Fiber Deep architecture the US MER performance is no longer dependant on optic distance and amplifiers cascade.

The Table 15 and 16 indicate the bit / Hz efficiency improvements by these movements to achieve 10Gbps DS and 2.5Gbps US throughputs but with the much lighter cable spectrum expansion.

Upstream Data Throughput	Band	MER	QAM Modulation	Date Rate	Total
	(MHz)	(dB)		(Mbps)	(Mbps)
Current Capacity	5 - 85	24	64	27	360
OFDMA w/ LDPC & Spectrum Expansion	5 - 204	30	256	38	1260
	5 - 400	30	256	38	2502
OFDMA w/ LDPC & Fiber Deep & Remote PHY	5 - 284	39	2048	54	2511

Table 15

Downstream Data Throughput	Band	MER	QAM Modulation	Date Rate (6MHz CH)	Total Data Throughput
	(MHz)	(dB)		(Mbps)	(Mbps)
Current Capacity	105 - 1002	33	256	38	5681
OFDM w/ LDPC & Spectrum Expansion	258 - 1218	36	1024	49	7840
	500 - 1794	36	1024	49	10568
OFDM w/ LDPC & Fiber Deep	500 - 1680	39	2048	54	10620
OFDM w/ LDPC & Fiber Deep & Remote PHY	360 -1440	42	4096	59	10620

Table 16

The reduced cable specturm expansion would help make the challenges for chip and product development down to a manageable degree. The DS bandwidth of 360 to 1440MHz is close to what we have today that helps maintain the composite power at Tx input and node / amp output; Compared to US bandwidth 5 to 400MHz, the 5 to 284MHz bandwidth makes the less reduction for the NPR range..

The DOCSIS 3.1, when getting the target data throughput achieved, has the good position to compete with the fiber to home technologies in light of its QoS, scailability, and much less CAPEX to rolling out the 10G cable access network.

6.0 Reference

(1) John T. Chapman, Mike Emmendorfer, Robert Howald, and Shaul Shulman "An Evolutionary Approach to Gigabit-Class DOCSIS, DOCSIS NG WP," May.21, 2012

(2) Steve Contra, "HFC Architecture Evolution"

HIGH DYNAMIC RANGE, VISUAL ADAPTATION, AND CONSUMER PERCEPTION: APPLYING VISUAL SCIENCE IN THE SEARCH FOR CONSENSUS ON HDR

Sean T. McCarthy, Ph.D.
ARRIS

INTRODUCTION

As momentum builds in the market transition to 4k Ultra HD, the emergence of High Dynamic Range (HDR) technology is drawing video engineers' attention to new issues that must be resolved if the full potential of a next-generation viewing experience is to be realized.

While 4k UHD provides the pixel density essential to enabling an immersive, big-screen TV viewing experience, HDR represents an opportunity to deliver a level of realism on video displays of all sizes that is far beyond what has been possible until now. By delivering greater contrast, increased luminance and an expanded color gamut, HDR vastly surpasses the SDR (Standard Dynamic Range) parameters that were standardized with ratification of Recommendation ITU-R BT. 709[1] 25 years ago.

Unlike 4k UHD, which has been a fairly straight-forward step in the evolution of display resolution built on the SDR foundation, HDR introduces a new paradigm where the dimensions of the new viewing experience must be defined in keeping with basic principles of the human visual response system. This will impact everything that's done in the creation and dissemination of video content from initial capture through production, post production, and processing for distribution.

Presently there are a number of approaches to HDR vying for traction in the marketplace. Much of the modeling that has gone into setting their parameters has focused on human contrast sensitivity, often in the context of non-broadcast use cases such as cinematic and episodic TV programming.

However, as various organizations pursue HDR standardization initiatives to facilitate market adoption, the priority must be on choosing solutions from a much broader perspective on visual response processes. Along with responsiveness to degrees of contrast, developers must consider human perceptual factors such as:

- Light and dark adaptation
- Brightness sensitivity
- Reaction to ambient light
- How color is perceived under different conditions
- Responses to frame-rate flicker that might be introduced with expansions in dynamic range

Moreover, planning for HDR must take into account the live TV broadcast environment. Content producers and distributors will have to determine how human visual response in the HDR environment will impact quality parameters for advertising, channel changes between HDR and SDR programming and the presentation of user interfaces, captioning and other textual and graphic elements. The impact of various HDR modes on bitrate and bandwidth requirements will also be an important consideration, especially for MVPDs.

In this paper, we look at the approaches to HDR and review the characteristics of human perception that will impact industry players' selection of HDR modes and their implementation in the multichannel viewing environment. Our goal is to provide basic information that will help manufacturers,

producers and distributors answer key questions such as "How bright a screen is bright enough?"; "How bright is too bright?"; "How quickly can a person adapt to changes in brightness, and what does that mean in terms of commercial insertion?"; and "How does the experience of HDR TV change with room lighting?"

THE EMERGENCE OF HIGH DYNAMIC RANGE IN TELEVISION

High contrast ratios between the darkest and brightest elements of images within a single frame and across sequential frames have long been applied in the cinematic viewing experience and to improve accuracy in the reading of medical images (Figure 1). In the case of theatrical displays, high levels of contrast are achieved in a dark-room environment where the highest luminance levels are well below that of even cathode ray tube TV displays but nonetheless deliver a much better viewing experience than can be achieved with the lower contrast levels used in traditional television programming.

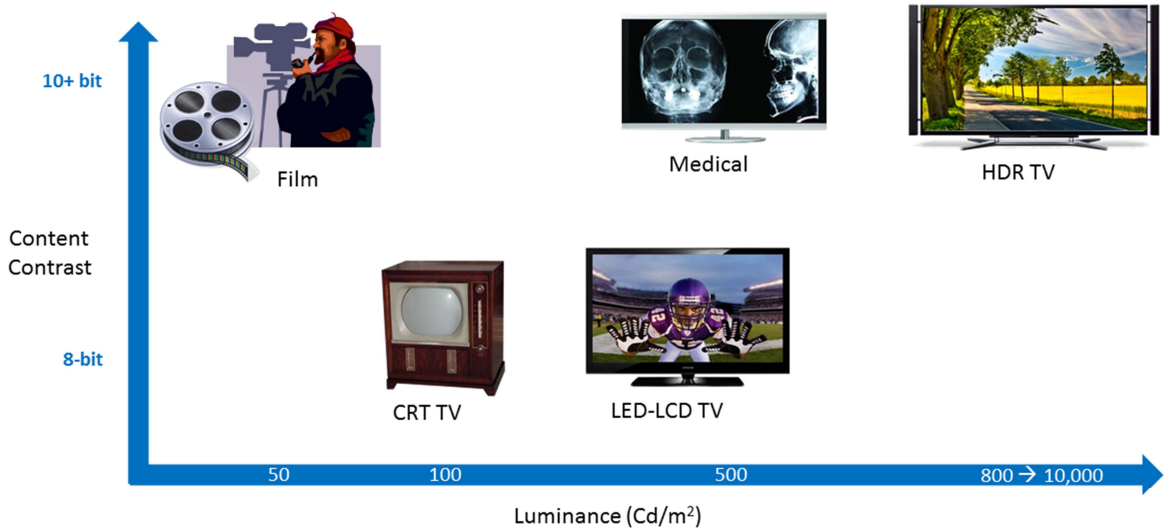


Figure 1. HDR television is part of the evolution of display and coding technologies and services.

Today these contrast limitations are not determined by the capabilities of display devices. Instead, they're a function of industry adherence to the parameters set by the ITU-R BT. 709, which were based on the fact that the typical luminance of CRTs used as reference displays for HDTV production is about 100 cd/m^2.

No such limitations were present with the introduction of HDR in medical imaging, which, by virtue of raising the peak luminance level up to 4,000 cd/m^2, more closely approached what we know today as HDR in terms of contrast range and luminance. (In actual practice, the Digital Imaging and Communications in Medicine[2] (DICOM) standard is employed in controlled lighting environments where sufficient contrast is achieved with a top luminance level around 600 cd/m^2.)

The television industry is now developing standards for HDR by specifying electro-optic

transfer functions (EOTF) and opto-electric transfer functions (OETF). [An EOTF specifies the non-linear mapping of illumination at the camera to digital code values. An OETF specifies the inverse non-linear mapping of digital code values to display luminance.] SMPTE ST-2084[3] specifies an EOTF and OETF that uses a technique of perceptual quantization[4] that is based on the same sophisticated model of human spatial contrast sensitivity[5] that is the basis of the DICOM standard. SMPTE ST-2084 supports peak luminance up to 10,000 cd/m^2 and is intended for displays used primarily for mastering non-broadcast content. An alternative EOTF/OETF has been proposed[6] that builds on the current SDR standards, ITU-R BT.18867 (EOTF) and ITU-R BT.709 (OETF), by adding a "knee" to handle luminance greater than 100 cd/m^2.

So far, the commercial introduction of HDR has been in conjunction with 4k UHD sets, most of which employ various iterations of LED LCD display technology. However, HDR is not intrinsically tied to 4k UHD and therefore could be used to greatly enhance the viewing experience with HDTV displays and, eventually, personal devices such as tablets and smartphones as well.

Along with higher luminance and contrast ratios, HDR goes hand-in-hand with an industry trend to broaden the color gamut beyond the 16.78 million colors supported by ITU-R BT.709. ITU-R BT.2020[8] has standardized a palette reaching 1.07 billion colors with 10-bit encoding and 68.7 billion colors with 12-bit encoding. In between these two levels is the DCI P3[9] color gamut, representing about 700 million colors, which has been in use with cinematic projections since 2007. Most providers of first-generation HDR technology are delivering 90 percent or more of the colors in the DCI P3 space.

STANDARDIZING HDR PARAMETERS

With a proliferation of HDR modes in play, the industry has made standardization of HDR a top priority, as evidenced by the efforts of the ITU, MPEG, Blu-ray Disc Association, SMPTE, ATSC, Ultra HD Forum, and UHD Alliance. As these initiatives progress toward what hopefully will be a unified approach to bringing this technology to market, it's essential that these groups look beyond the impact degrees of contrast have on human visual response to how the full range of perceptual processes are impacted by any given set of contrast, luminance and color parameters.

Addressing the Live HDR Imperative

It's also important to note that, as illustrated by Figures 2 and 3, the test case for setting specifications and best practices should be live broadcast TV. In the case of non-live, cinematic or episodic programming (Figure 1), the workflow for processing content entails capturing and encoding from a composed setting, an entire program, for delivery into post production processing.

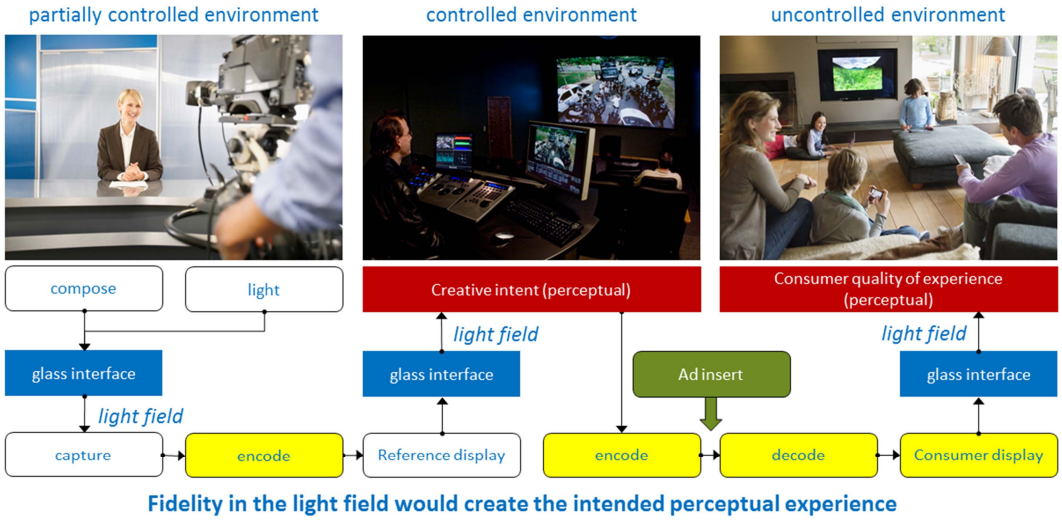


Figure 2. Non-live cinema and episodic programming enable experts to craft content to deliver the creative intent. Nonetheless, television is viewed in many different lighting environments that could impact the effectiveness of HDR. Moreover, advertisements and other interstitial content is usually inserted into programming. Those interstitials could be SDR, which could impact the overall consumer quality of experience.

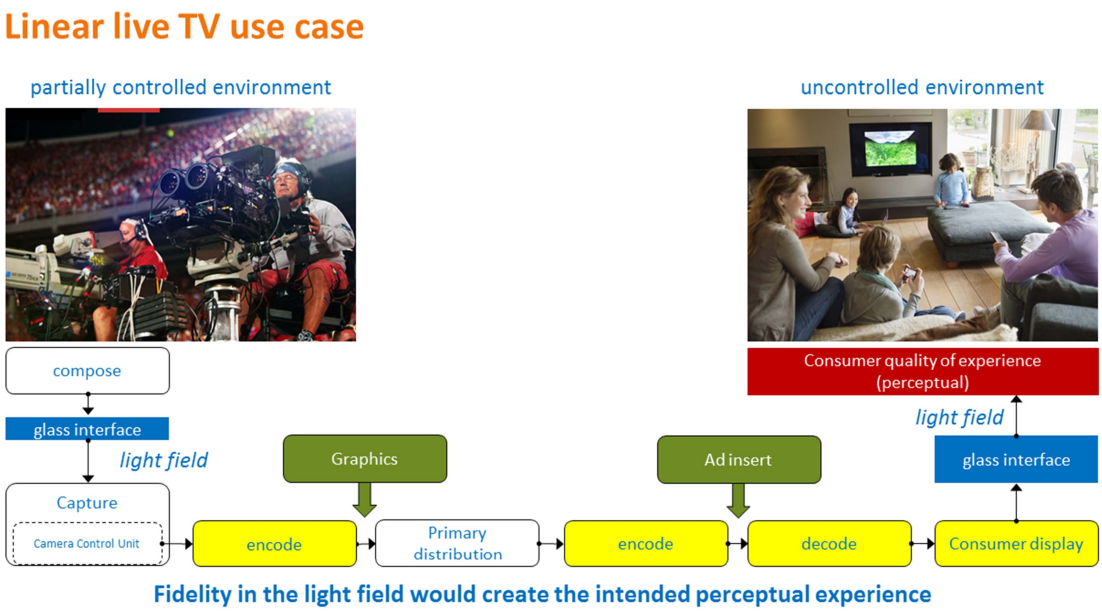


Figure 3. Live HDR television poses special challenges including management of real time graphics and logo overlays, management of venue lighting conditions, and management of HDR and SDR interstitials.

Post production editors have an opportunity to use reference monitors in non-real time to ensure the creative intent of the producers is captured in the formatting of the program for HDR. With the fidelity of the light field maintained from capture through

post production, the intended viewing experience is assured so long as that fidelity is maintained by the display in an uncontrolled viewing environment.

In contrast, with live broadcast the correspondence in fidelity between capture and viewing is dependent on decisions made in an uncontrolled environment at the point of capture. These decisions can produce unintended or intended consequences in the viewing experience, depending on how well the cameramen and the camera control unit operators understand the unique characteristics of HDR and to what extent those characteristics have been defined and adopted by the production community.

Production practices need to be adopted to ensure maximum creative benefits in the use of HDR extend all the way to make-up artists and set designers. Old ways of doing things can produce unrealistic results under the probing realism of HDR cameras.

Conforming Out-of-Program Dynamics to HDR

The standardization of HDR and its production practices also needs to be factored into advertising, graphic overlays, picture-in-picture applications and user interfaces to ensure a consistent, non-jarring viewing experience. Principles need to be developed based on the nuances of opto-physiological processes discussed in the ensuing section as well as basic common sense. For example, the industry does not want to repeat the experiences that led to the CALM (Commercial Advertisement Loudness Mitigation) Act by inserting HDR-enabled ads that don't conform to the parameters used with HDR programming.

Applying the Principles of Visual Science

There is a wide range of information to be applied from visual science in industry efforts to establish standards and practices relating to delivering the best possible HDR viewing experience to consumers. As discussed in the following section, these include:

- Processes of light and dark adaptation
- Effects of luminance and screen size on flicker perception
- The impact of luminance on color perception
- The impact of speed of adaptation on scene changes, program changes and commercials
- How light field and context effect perceptions of brightness and color

PROCESSES OF LIGHT AND DARK ADAPTATION

The human visual system adapts to 9 log units of light intensity[10]. Several physiological processes contribute to the human visual system's ability to adapt to this enormous range of light intensities. These include:

- Changes in pupil diameter
- The impact of luminance on retinal illuminance
- The impact on light absorption from the bleaching of photoreceptor photo pigments
- Speed of adaptation
- Contrast sensitivity

An understanding of these processes can be useful in addressing potential issues such as: sequential viewing of HDR and non-HDR content, maximum luminance and preservation of creative intent.

Changes in Pupil Diameter

Changes in the size of the pupil regulate the area through which light enters the eye by a factor of approximately 16. In fully dark-adapted conditions, the pupil of the eye has a diameter of approximately 8 mm. In fully light-adapted conditions, the pupil diameter reduces to approximately 2 mm.

The impact of a typical HD viewing environment, corresponding to a 30-degree stimulus on the size of the retinal image, is plotted in Figure 4 in accord with the Watson-Yellot[11] model. In this calculation a display having a luminance of 100 cd/m^2 may be expected to have a diameter of ~4.5mm. At 1000 cd/m^2, the pupil diameter can be expected to be ~2.5 mm, which would correspond to an approximately 3-fold reduction in pupil area.

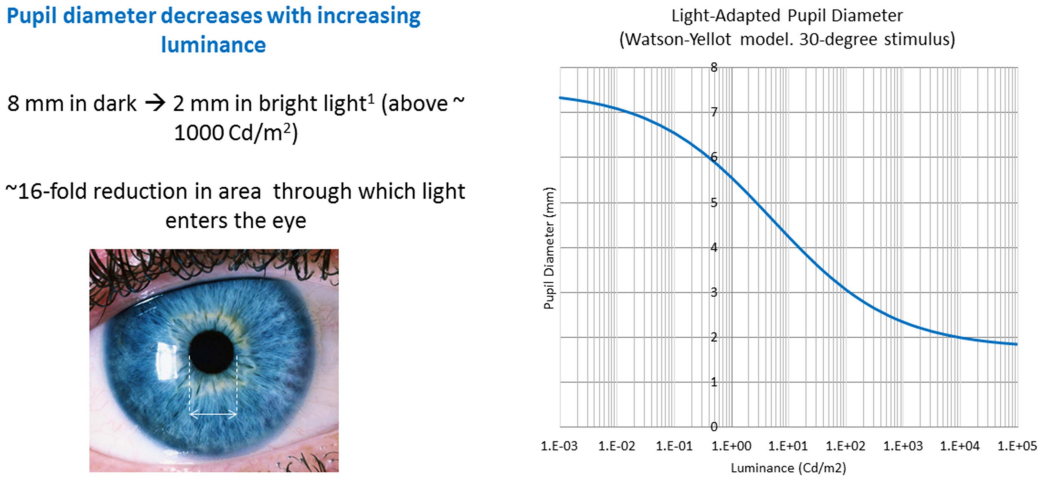


Figure 4. Illustration of the change in pupil size with luminance.

Between fully dark-adapted and light-adapted conditions, the pupil diameter changes in a graded manner. Exposure to bright light can reduce the area of the pupil, and consequently reduce retinal illumination, by approximately a log unit in a half second.

Pupil diameter can also be affected by ambient background light or lack thereof, which should be taken into account with the differences between cinematic and home viewing. Other factors affecting pupil diameter include the size of the visual stimulus, the position of the stimulus relative to the center of vision and the use of monocular or binocular stimuli, also all affect pupil diameter, all of which are considered in the Watson-Yellot model.

The Impact of Luminance on Retinal Illuminance

Pupil diameter as impacted by display luminance regulates the illuminance of the retina. Vision scientists use the term “troland” (Td), equal to luminance multiplied by pupil diameter, as a measure of retinal illuminance. Table 1 provides troland values for various luminance levels calculated using the Watson-Yellot model of pupil diameter for a visual stimulus of 30-degrees. (Note that between 10 and 1000 cd/m^2, luminance changes by a factor of 100 but retinal illuminance changes by a factor of approximately 30.)

Luminance (cd/m^2)	0.001	0.01	0.1	1	10	100	1000	10000
Pupil Diameter (mm)	7.3	7.1	6.5	5.6	4.2	3.1	2.3	2.0
Pupil Area (mm^2)	42.2	39.4	33.6	24.2	14.1	7.4	4.3	3.1
Retinal Illuminance (troland)	0.042	0.39	3.4	24	141	736	4309	31286

Table 1. Luminance, Pupil Size and Retinal Illuminance (troland)
(Pupil sizes predicted from Watson-Yellot model for 30-degree stimulus)

Both screen luminance and ambient background should also be considered when predicting pupil size and retinal illuminance. Although changes in pupil size account for a relatively small portion of visual adaptation, it is worth noting that reduction in pupil size also increases depth of focus and reduces glare.

Settings for HDR must take into account the impact of luminance on the retinal photoreceptors that determine illuminance in both bright and dark home environments and even outdoor situations in cases where the viewing experience is extended to handheld devices. The level of sensitivity and the speed of adaptation can vary considerably depending on those conditions.

The retina has four photoreceptor classes: rods; short-wavelength sensitive S-cones ("blue"); medium-wavelength sensitive M-cones ("green"); and long-wavelength sensitive L-cones ("red"). It is well known that rods are responsible for night vision and cones for daylight and color perception, but it is more accurate to think of visual adaptation in three categories that better reflect the gradual shift from rod-dominated vision to cone-dominated vision as light conditions brighten:

- Scotopic (below 0.001 cd/m^2) dominated by rods
- Mesopic (0.001 to 10 cd/m^2) mix of rods and cones
- Photopic (above 10 cd/m^2) dominated by cones

In darkened cinema theaters and home environments, mesopic-level adaptation might be a significant consideration. In bright home and mobile environments, photopic-level adaptation would be more typical.

Bleaching Adaptation

With the current generation of displays, we are just entering into a point where what is known as "bleaching adaptation" could become significant and impact a viewer's experience. This is the phenomenon that occurs when the eye adjusts to sustained periods of illumination. The bleaching impact will be an important consideration in determining what average and peak levels of HDR brightness should be in the context of temporal shifts in luminance.

Vision is initiated by activation of light-sensitive photopigments. Photoreceptors contain a high concentration of light-sensitive biological pigment called rhodopsin in rods and cone-opsin in cones. The concentration of excitable photopigments is reduced during sustained illumination in bleaching adaptation.

The photopigment is made up of a protein (opsin) and a vitamin-A based chromophore. The chromophore absorbs a photon causing it to change shape, which in turn changes the shape of the surrounding protein. The altered protein ignites a powerful cascade of biological reactions that results in the light-induced neural response.

Once the chromophore absorbs a photon, it can no longer respond to light and needs to be recycled, which takes some time. During illumination, photopigments are used up and the optical density of the photoreceptors decreases; i.e., photoreceptors absorb less light because there are fewer chromophores available to catch photons.

This is when bleaching adaptation occurs (Figure 5). Over a period of steady illumination, the photopigment concentration will reach a new steady-state level[12] when the rate at which chromophores absorb photons is balanced by the rate at which they are replenished through the retina's recycling program.

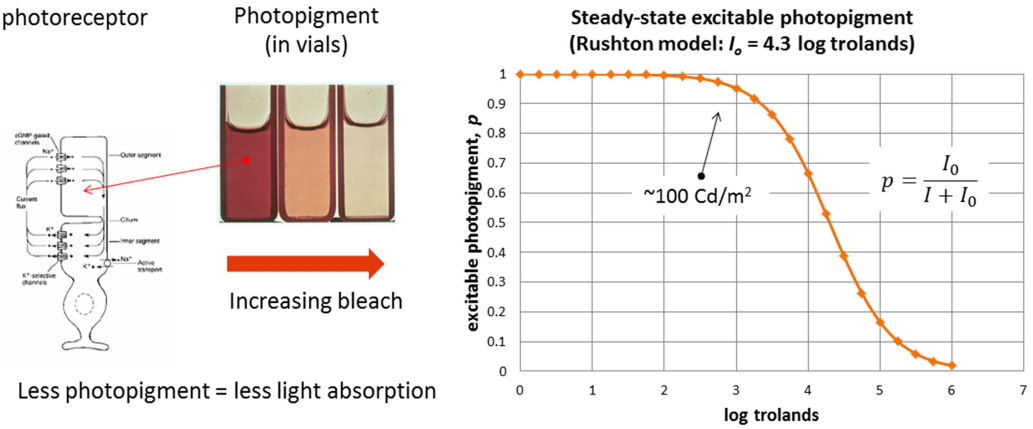


Figure 5. Illustration of the effect of luminance on photoexcitable photopigment if retinal photoreceptors.

Speed of Response and Adaptation

Rod-driven vision differs from cone-driven vision not only in terms of absolute sensitivity but also in terms of the speed at which light variations are signaled. The peak of the response of a dark-adapted rod to a flash of light occurs at approximately 120 msec[13].

The peak of a light-adapted rod response occurs at approximately 75 msec[13]. The peak of cones response occurs at approximately 20 msec[14]. The change in the time-scale of neural responses can be expected to play an important role in sensitivity to flicker and to motion judder.

Photoreceptors adapt to moderate non-bleaching step changes in increased illumination on the time scale of seconds[15, 16]. During this light adaptation process, the absolute sensitivity of photoreceptors decreases and the kinetics of responses increase in speed.

The time course of adaptation to decreased illumination (dark adaptation) depends on the intensity and duration of preceding stimuli[10]. If the preceding exposure was low enough that no significant photopigment bleaching occurred, then dark adaptation may also be measured on the time scale of seconds but slightly slower than light adaptation. For beaching adaptation, dark adaptation may take longer, on the order of tens of seconds to minutes. Dark adaptation also has two distinct phases: one is driven by recovery of cone sensitivity; the other even slower phase is driven by rod sensitivity.

In bright home and mobile viewing environments, both light and dark adaptation to modulations in illumination may be expected to proceed on a time scale measured

in seconds. In dark home and theater environments, rapid changes going back and forth from mesopic-level to photopic-level luminance might result in slower dark adaptation.

EFFECTS OF LUMINANCE AND SCREEN SIZE ON FLICKER PERCEPTION

Flicker perception, or the sensitivity to temporal changes across video frame sequences, has not been an issue at the various frame rates used with video mapped to the dynamic ranges of ITU-R BT.709 on traditional screens. Yet with the onset of HDR on large-screen displays, flicker sensitivity could become an issue affecting the rate at which content is captured as well as the display frame and refresh rates.

The same is true of judder, the perception of uneven or jerky video playback that arises from movement of objects, edges or detail from one frame to the next. Increases in contrast, sharpness of detail and motion speed can cause judder, especially with increased screen sizes, which, for any given viewing distance, have the effect of bringing the images closer to the viewer.

Figures 6 and 7 illustrate the impact of luminance and screen size, respectively, on flicker. The critical flicker frequency (CFF) is the temporal frequency at which flicker is perceived as a steady light. CFF may be predicted from luminance with the Ferry-Porter Law[10], which states that the CFF increases in proportion to the logarithm of luminance.

Flicker sensitivity also depends on the location of the stimulus on the retina and the size of the stimulus. The temporal frequency threshold for flicker in the center of vision is higher than in the peripheral at all luminance levels. And the temporal frequency threshold for large stimuli (such as from a UHD display or cinema screen) is higher than for smaller stimuli at all luminance levels. (Granit-Harper law[10])

Given that TV manufacturers have increased frame rates in newer models to 120 and even 240 Hz, there's no reason to expect flicker or judder perception to be caused by HDR display systems. However, content captured at the 24 frame-per-second (fps) rate used with films and episodic series could be problematic. If the need to accommodate HDR leads the motion picture industry to raise the capture frame rates to, say, 48 fps, distributors will have to make adjustments based on the impact higher rates will have on bandwidth requirements for HDR content.

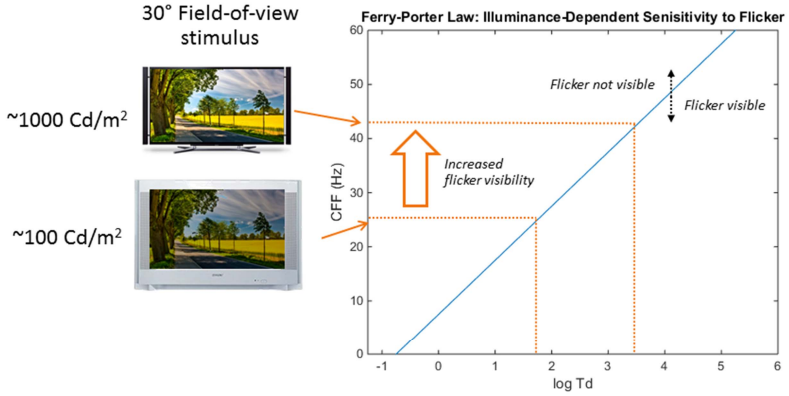


Figure 6. Illustration of the Ferry-Porter law. Brighter HDR displays may be expected to make non-smooth motion and flicker more noticeable.

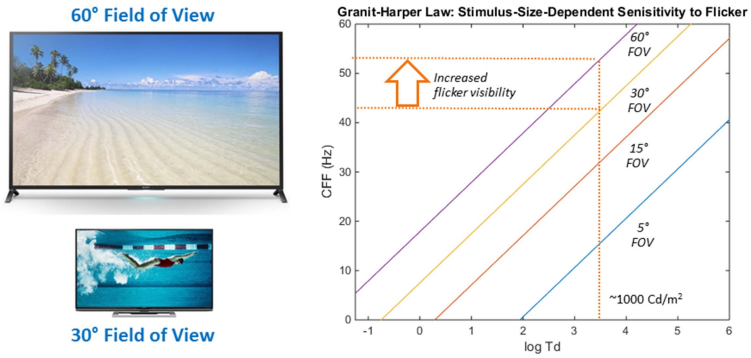


Figure 7. Illustration of the Granit-Harper law. Larger displays may be expected to make non-smooth motion and flicker more noticeable.

THE EFFECT OF LUMINANCE ON COLOR PERCEPTION

As luminance increases so does the ability of the human visual system to discriminate between colors at ever smaller gradations. As shown in Figure 8, color discrimination may be described in terms of MacAdam ellipses[17], a measure of the just-noticeable difference (JND) between colors. The size of MacAdam ellipses shrinks with increased luminance[18], which means that the JND decreases. Consequently, more bits would be needed to code color without introducing noticeable errors, particularly at high luminance. Thus, 10-bit encoding may be expected as a minimum bit depth for HDR for any color space (ITU-T BT.709, ITU-T 2020, or DCI P3).

Luminance also affects the perception of the hue (Figure 9). The Bezold-Brücke[19] effect describes the perception of two stimuli having the same wavelength but different luminance as different hues.

EFFECTS OF ADAPTATION SPEED ON PROGRAM CHANGES AND COMMERCIALS

As detailed in the preceding discussion, the speed of light and dark adaptation depends on the level of retinal illuminance, the duration of illumination, the kinetics of the changes in pupil size, the rate of change of photoreceptor sensitivity, the rate of change of excitable photopigment, and the state of overall visual adaptation (scotopic, mesopic, or photopic).

All of this has important implication for the impact of rapid local or global luminance changes in television programming when HDR and non-HDR content is presented sequentially to viewers.

Here, it's important that HDR-formatted commercials be pegged to the same parameter set for HDR programming. Conversely, non-HDR compliant commercials should not be placed with HDR programming.

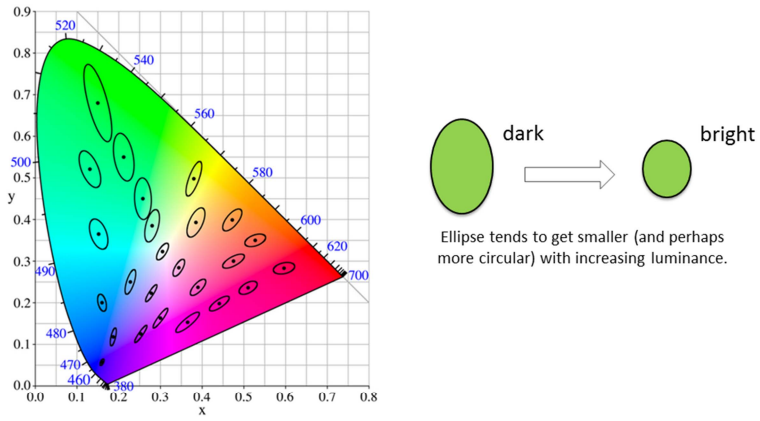


Figure 8. Illustration MacAdam ellipses that may be used to describe color discrimination.

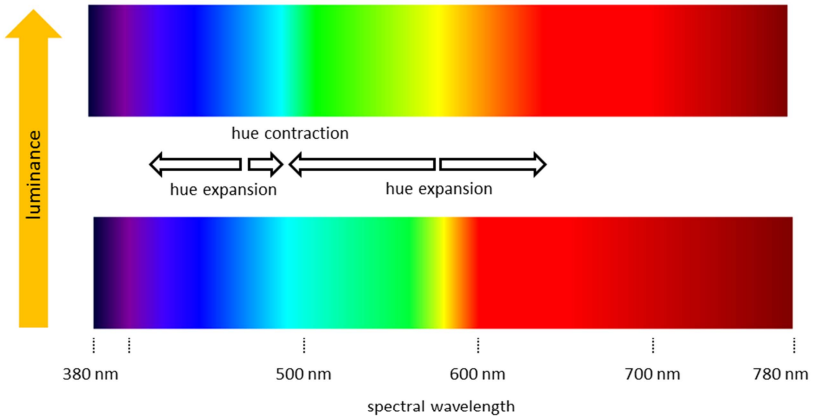


Figure 9. Illustration of the Bezold-Brucke effect in which luminance affects perception of hue.

THE EFFECT OF THE LIGHT FIELD AND CONTEXT ON BRIGHTNESS AND COLOR

Perceptions of brightness and color are not just a matter of luminance intensity. Variations in the light field across the frame can profoundly influence visual responses[20], as shown in Figure 10. (See also the highly compelling color illusion created by R. Beau Lotto[21]).

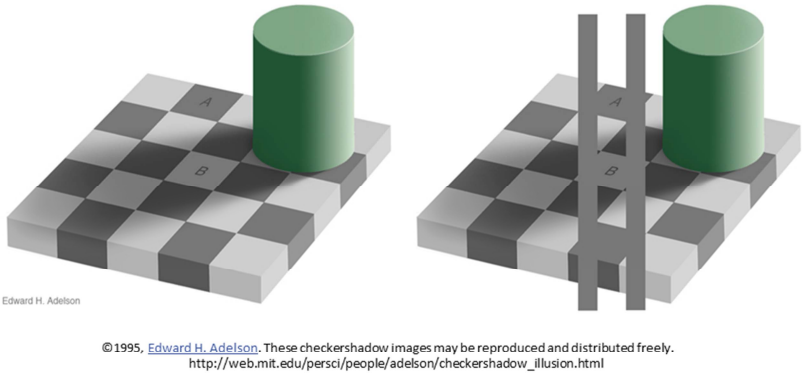


Figure 10. Illustration that the visual system is more than simply a logarithmic light meter. The squares A & B are identical though they a perceived to be very different.

Such nuances will come into play as content producers move to using HDR in the creation process. They will need to avoid relying on luminance specifications alone in ensuring their creative intent is conveyed to viewers.

CONCLUSION

HDR is bringing a welcome transformation in the television viewing experience to the benefit of producers, distributors and consumers, provided the industry is careful to map parameters to the realities of human perception. Specifically, there needs to be general understanding with respect to the following points:

The requirements of live broadcast programming should be the threshold for setting HDR parameters:

- Dynamic contrast and color ranges will have to deliver an optimal viewing experience without reliance on intervening post-production processes.
- On-site production teams will have to incorporate understanding of HDR to ensure fidelity to creative intent.

The different rates of light and dark adaptation could play a significant role in QoE

- Light adaptation is fast. Dark adaptation is slower
- Mixed SDR and HDR could impact QoE for ad insertion, scene changes and program selection.
- Bleaching adaptation could significantly slow adaptation to changes in program luminance.

Bleach fractions might begin to be significant at 1000 cd/m2.

- They definitely will have an impact at multiple thousands of cd/m2

- It might be beneficial to limit average scene luminance to low bleaching fractions
- It might be beneficial to limit the duration of high-luminance highlights to minimize after images

Adaptation could impact user interaction:

- Graphic overlays might need to be tailored to the expected light/dark adaption state
- GUI could easily be impacted

HDR should not be considered in isolation. It impacts:

- Color perception with implications for bit depth and color gamut
- Sensitivity to flicker as a function of screen size, frame rate and display refresh rate

REFERENCES

1) ITU-R BT.709-5, "Parameter values for the HDTV standards for production and international programme exchange", Recommendation ITU-R, BT Series Broadcasting service (television), 2002.
2) NEMA PS3.14-2011 "Digital Imaging and Communications in Medicine (DICOM) Part 14: Grayscale Standard Display Function", NEMA 2011
3) SMPTE ST-2084 "High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays" SMPTE. 2014
4) Miller, S., Daly, S., and M. Nezamabadi. "Perceptual signal coding for more efficient usage of bit codes", SMPTE Motion Imaging Journal 122, 52{59 (May 2013)
5) Barten, P.G.J., Contrast Sensitivity of the Human Eye and its Effects on Image Quality. SPIE Optical Engineering Press: Bellingham, WA, 1999.
6) Borer, T. "Non-linear Opto-Electrical Transfer Functions for High Dynamic Range Television" BBC R&D White Paper. WHP 283. July 2014
7) ITU-R BT.1886 "Reference electro-optical transfer function for flat panel displays used in HDTV studio production" ITU-R, BT Series Broadcasting service (television), 2011.
8) ITU-R BT.2020. "Parameter values for ultra-high definition television systems for production and international programme exchange." Recommendation ITU-R, BT Series Broadcasting service (television), 2014.
9) SMPTE ST-431-1 (DCI P3). "D-Cinema Quality – Screen Luminance Level, Chromaticity and Uniformity" SMPTE 2006.
10) Levin, L.A., Nilsson, S.F.E., Ver Hoeve, J., Wu, S., Kaufman, P.L., and Alm, A. (2011) Alder's Physiology of the Eye, 11th edition. Saunders – Elsevier.
11) Watson, A.B. and Yellot, J.I. (2012) "A unified formula for light-adapted pupil size." J. Vis 12(10):12
12) Rushton, W.A.H. & Henry, G.H. "Bleaching and regeneration of cone pigments in man." Vision Research 8, 617-6311968
13) Friedburg, C., Thomas, M.M. and Lamb, T.D.(2001) "Time course of the flash response of dark- and light-adapted human rod photoreceptors derived from the electroretinogram." J. Physiol. 534.1, pp. 217-242.
14) Van Hateren, J.H. and Lamb T.D. (2005) "The photocurrent of human cones is fast and monophasic." BMC Neuroscience, 7:34.
15) Cao, L.-H., Luo D.-G., and Yau K.-W. (2014) "Light responses of primate and other mammalian cones." PNAS. Vol. 11 No. 7

16) Korenbrot, J.I. (2011) “Speed, adaptation, and stability of the response to light in cone photoreceptors: The functional role of Ca-dependent modulation of ligand sensitivity in cGMP-gated ion channels “JGP vol. 139, No. 1, 31-56.
17) MacAdam, D.L. “Speed, adaptation, and stability of the response to lightOpt Soc Amer. Vol 32, No. 5. 1942
18) Pridmore, R.W. and Melgosa, M. (2005) “Effect of luminance on color discrimination ellipses: Analysis and prediction data.” Color Res. & App. Vol. 30, No. 3, 186-197
19) Pridmore, R.W. (1999) “Bezold–Brucke hue-shift as functions of luminance level, luminance ratio, interstimulus interval and adapting white for aperture and object colors” Vision Research 39, 3873-3891
20) Adelsen, E.H. “Checkershadow illusion.” 1995 (http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html)
21) Lotto, R.B. www.lottolab.com University College London

HYBRID CONTENT GENERATION WORKFLOWS: FROM MPEG-2 TS TO ADAPTIVE STREAMING AND ISO-BMFF

Yasser Syed, PhD and Alex Giladi
Comcast/Technology & Product: VIDEO Architecture Group
InterDigital Communications, Inc.

Abstract

A transformation is happening in the content network delivery systems of the cable / IPTV world, which is increasingly moving us from a traditional MPEG-2 TS [Transport Stream] push-based ecosystem, toward a pull-based unicast ecosystem. As this transformation is occurring, we still need to maintain a consistent service experience, including aspects such as linear programming, DVR, accessibility features, etc. An established set of mechanisms and operational procedures already exists, via the traditional MPEG-2 TS model. Rather than rebuilding these mechanisms and operational procedures again, a hybrid content generation workflow model can be utilized to convert these services. This paper examines how this conversion can occur, such that the source content & service data may originate, be conditioned, and processed in a traditional MPEG-2 TS environment – but ultimately delivered to the customer in one or more adaptive streaming formats.

INTRODUCTION

There are two concurrently-happening paradigm shifts marking the industry's transition from traditional QAM-based distribution to multi-screen adaptive streaming. The first shift is the underlying distribution model. While the industry's original digital video distribution model offered push-based multicast delivery, the emergence of multi-screen environments that rely on HTTP downloads, which inherently are pull-based.

The second paradigm shift the container format. The industry relied on MPEG-2 transport streams for nearly two decades, but use of fragmented ISO-BMFF [ISO - Base Media File Format] as a container format is on the rise. The two formats had completely different design goals – while MPEG TS were built for efficient low-latency real-time transmission, the ISO-BMFF is instead a media *storage* format. Segment-based delivery makes the use of ISO-BMFF feasible, and the departure from tightly synchronized MPEG-2 TS multiplexes makes advanced features possible – such as late binding, where audio and video components can be separately selected by the client and downloaded, then synchronized at the receiver.

Those two shifts are buttressed by one over-riding reality: The functionality and customer experience in a multi-screen environment needs to be a superset of the one to which they are accustomed, and not a radical departure. Moreover, many regulatory requirements apply – or are expected to apply – to multi-screen world same way as they apply to current cable systems. This requires a precise translation of existing functionalities into the new ecosystem.

Despite the tectonic shifts at the transport and container level, the basics have not changed as much: The bitstreams being decoded are still the same bitstream formats, regardless of whether they're packaged into ISO-BMFF boxes or MPEG-2 TS packets.

This paper examines how a hybrid system can be designed using MPEG-2 TS and DASH (Dynamic Adaptive Streaming over HTTP) constructs to create a resource-efficient and robust system. The system supports scalable and cost-efficient content delivery using one or more popular of adaptive streaming technologies. It also discusses constructs and processes that are being defined in standards bodies and consortia, including MPEG, SCTE (WG7) and the DASH Industry Forum (DASH-IF), all of which assist in the definition of and conversion between network distribution and storage, and end-client distribution.

THE MPEG-2 TS MULTIPLEX

The MPEG-2 Transport Stream is the backbone of all current cable, IPTV and satellite distribution networks. It entered the digital video marketplace two decades ago. MPEG-2 TS also accounts for a sizeable part of the adaptive streaming landscape, especially considering that the TS-based Apple's HLS (HTTP Live Streaming) is still considered a dominant adaptive streaming system . Beyond linear and on demand distribution systems, MPEG-2 TS is also used in physical media (Blu-Ray).

The MPEG-2 transport stream is a ubiquitous media carriage mechanism that can define parameters and interaction between multiple packetized elementary streams, to create a viewable media experience. Some of its features include:

- Tight encoder/decoder synchronization: The transport stream format provides enough information to assist the decoder in reconstructing the encoder clock
- Single clock use for all media delivery, decoding and presentation
- A single buffer model to allow reliable receiver implementation
- Multiplexing: A single MPEG-2 TS stream may contain multiple components (video, audio, and data) for a single program, as well as multiple unrelated programs
- Codec independence: While originally designed in early '90s together with MPEG-2 audio and video, MPEG-2 TS successfully carries newer formats, such as H.264, H.265, AAC and (E-)AC-3; moreover it can be easily extended to support newer generations of codecs.
- Fast channel acquisition and recovery in the stream or file, through repetition of the system information (SI) located in tables such as the Program Management Tables (PMTs) and Program Association Tables (PAT.) The underlying codecs are expected to repeat all sequence-level information – such as sequence and picture parameter sets (SPS / PPS) in H.264 and H.265 video codecs.

Over the past two decades, a significant amount of infrastructure was built that supports various broadcast, cable and IPTV use cases related to transport streams (see `Figure 1`.) In the case of North America, an example is SCTE 35, an in-band cueing and messaging mechanism for ad insertion, blackouts, and alternate content. SCTE 35 allows easy communication between content providers and service providers [2.] Traditional TS-based systems also provide in-band functionality for purposes such as CEA 608/708 closed captioning and active format description (AFD).

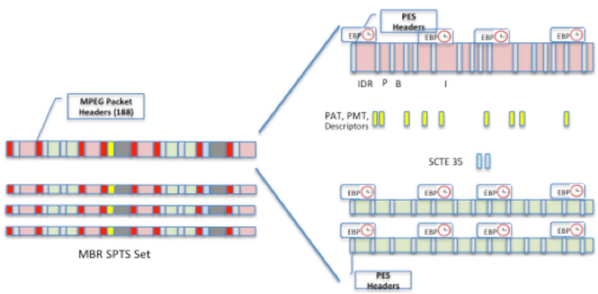


Figure 1: MPEG-2 TS STRUCTURES

Furthermore, the maturity of the technology enabled the development of a proven set of testing methodologies and equipment. Those developments validated media, enabled service monitoring, and shored up the troubleshooting mechanisms that can both isolate faults and ensure service quality .

By contrast, the shift to adaptive streaming shares more commonality with progressive HTTP download of video files, than with the traditional push-based delivery.

As opposed to the traditional and continuous stream, adaptive streaming methods segment each stream, and an end-client is offered a description of available segments – a *manifest*. The majority of the logic resides at the client device, and it is the client device that decides which of the available alternatives it can sustainably present, as described in the manifests. The alternatives – *representations* are listed in the manifest.

For the rest of this discussion, we will use terminology based on DASH, as it provides a very rich and precise vocabulary for manifest description (see `Figure 2`). In the DASH taxonomy, *representation* conceptually means single media (one or more elementary streams), comprised of multiple short *segments* (otherwise known as chunks.) An *adaptation set* is a set of seamlessly switchable representations, and a *period* is a time box within which representations exist. Each period can be viewed as an independent playlist (similar to a master m3u8 in HLS) or a completely different program, and a period border can be seen as a splice point. [Note that the fact that we are using DASH terminology here does not mean that we are limiting ourselves to DASH outputs.]

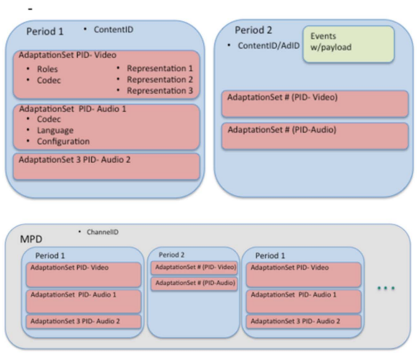


Figure 2: DASH MANIFEST STRUCTURE

The adaptive streaming market today is fragmented into several technologies, which are at their cores are very similar, usually varying only by container type. The dominant system is Apple's venerable HTTP Live Streaming (HLS), which is deployed in several different variations (protocol version 3 is very different from 6, for instance.) Next in popularity are Adobe's HDS (HTTP Dynamic Streaming), Microsoft's SmoothStreaming and MPEG DASH – the latter being a relatively new international standard for adaptive streaming. In practice if support for a reasonable crop of devices is needed this means a requirement for several "profiles," targeted to different devices classes. For context, a Netflix presentation from May 2014 mentions 6 profiles in use within its content generation workflow [8]. Each profile will invariably target a different crop of devices and use different adaptive streaming systems.
As far as segment formats are concerned, several options are used simultaneously. HLS uses MPEG-2 TS segments, as well as MPEG-1 audio elementary streams. HDS and SmoothStreaming use fragmented ISO-BMFF. Lastly, DASH allows both ISO-BMFF and MPEG-2 TS. The DASH-IF IOP (DASH Industry Forum / Interoperability Points) and DVB (Digital Video Broadcast) DASH groups use ISO-BMFF, while the SCTE (Society of Cable Telecommunications Engineers) recently standardized both MPEG-2 TS-based profile (SCTE DASH/TS) and ISO-BMFF-based profile (SCTE DASH/FF) as DVS 1202, DVS 1203 and DVS 1208.

Moreover, if the work of providing multi-rate transcoding is already happening, a convenient optimization would focus on generating content for traditional QAM-based distribution, using the same infrastructure.

Complicating things even further is the matter of multiple encryption schemes, which may also need to be supported. HLS alone comes with two incompatible segment encryption modes, and neither are suitable for MPEG-2 conditional access systems, which can be trivially retargeted to traditional QAM-based distribution.

This demand to produce a multitude of output formats is not necessarily an insurmountable problem – notice that the underlying audio, video and closed captioning bitstreams are the same. The change is only at container and encryption layers. Because at the core there are more things in common than are different, this immediately led us to establish the concept of a "common intermediate format" or CIF. With it, the generation of segments in specific container format (*packaging* or *multiplexing*) can be decoupled from the transcoder's processing of content.

MULTI-TIER CONTENT GENERATION WORKFLOWS

A simple approach to generating multi-rate content for ABR distribution, from a real-time MPEG-2 TS feed, can be summarized as (1) transcoding the source into the appropriate set of bitrates, (2) writing the generated segments to the origin server, and (3) letting CDNs [Content Delivery Networks] take care of the rest. Had this solution worked satisfactorily and scaled well, this paper would have been much shorter! As often is the case, the devil is in the details – details of the service being deployed, the types of coverage needed, and the technologies supported.

Having considered a multitude of profiles – and the resulting outputs – we arrived at a seemingly more complex solution, but one that is ultimately more scalable. It's a multi-tiered system that consists of (1) a transcoding tier, which outputs in an intermediate multi-rate format, and (2) a packager tier that generates the final content from this intermediate format. The third tier of this system is the content protection tier, which integrates with DRM systems used for the deployment. This tier encrypts the packager output into the required format, such that the key, provided later by the licensing server, can be used to authorize secure playback in the user device. This solution reduces the amount of variations required to support today's video marketplace.

The computational complexity of packaging is orders of magnitude smaller than that of a transcoder. This means that a high-density solution – a single physical server simultaneously handling a large amount of streams – is feasible, and can run on off-the-shelf, inexpensive hardware. This produces order-of-magnitude hardware cost reductions – a single, inexpensive, off-the-shelf server, rather than a rack of real-time encoders running on specialized hardware. Operational benefits abound, too, in the form of lower energy and space overhead. Moreover, the horizontal scaling characteristics of a multi-tier system are significantly better than these of a traditional "all-in-one" transcoder.

Efficient multi-packager deployment benefits from the use of multicast distribution within an MSO-operated internal network. The easiest way to integrate the existing infrastructure with the new hybrid multi-rate workflows involves the reuse of the (currently ubiquitous) MPEG-2 TS over UDP (User Datagram Protocol) multicast. An alternative

approach exists -- ISO-BMFF in conjunction with a reliable file multicast protocol (such as File Delivery over Unidirectional Transport, FLUTE). It provides a good solution for the wireless multicast case (e.g., LTE eMBMS, or Long Term Evolution / Evolved Multimedia Broadcast Multicast Service), but it, in our opinion, is an unjustifiably complex solution to the simpler problem of multicasting in a provisioned core network that is practically error-free. Hence, MPEG-2 TS became our container format of choice for the common intermediate format (CIF), and UDP multicast, with packets carrying a number of complete transport stream packets, is our multicast format of choice.

ADAPTIVE TRANSPORT STREAMS AS AN INTERMEDIATE FORMAT

The Adaptive Transport Stream (ATS) specification, SCTE DVS 1196, was created to satisfy the need for the CIF format. It provides backwards-compatible virtual segmentation, and adds some constraints to the MPEG-2 TS stream to allow for an easy conversion to adaptive streaming technologies (see `Figure 5`). The resulting single-program MPEG-2 TS stream is playable in its regular context, but the modifications allow the stream to be easily adapted into an adaptive streaming workflow.

ATS is built upon three main concepts: An *EBP structure*, which is a marker inside a transport stream packet; an *EBP descriptor,* which describes the multiplex as the part of in-band system information, and a *source description,* which is a DASH-inspired manifest that completely defines an *ATS set* – a set of multiplexes or/and files associated with the same content (see `Figure 3`).

The EBP structure is carried in adaptation field of a transport stream packet and is inserted at the transcoding stage. It adds the following:

- GOP [Group of Picture] conditioning of video and audio elementary streams, to make them segmentable, with the option to incorporate more than one segment size
- Segment adjustment for splice points
- Stream mark-up indicators to indicate segmentation points (e.g. virtual segments)
- A wall clock time indicator that accompanies each segment
- An adaptation field structure for stream mark-up, labeling, and time information (EBP structure)
- The creation and ability to segment-align multiple outputs from a single content source, so as to provide a multi-bit-rate set of streams as a source for different representation segments.
- Segmentation in encrypted domain is possible. Adaptation fields are never encrypted when a MPEG-2 CAS is used, hence segmentation is possible even without decrypting the incoming stream.

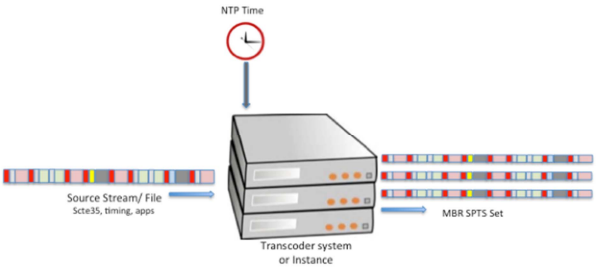


Figure 3: TRANSCODER ATS COMPLIANT STREAMS

The EBP descriptor in the PMT describes the element(s) of the ATS set carried in the multiplex. This description contains information such location of EBP structures for each elementary streams, its random access characteristics, and expected segment size. A more detailed description of the ATS set is provided within using source description

document, which is presented in more detail later.

Adaptive transport streams can be transmitted via IP multicast to various parts of the network, to bring them closer to CDNs and edge caches (see Figure 4). To use these types of streams for adaptive streaming purposes, manifests and indexes need to be created.

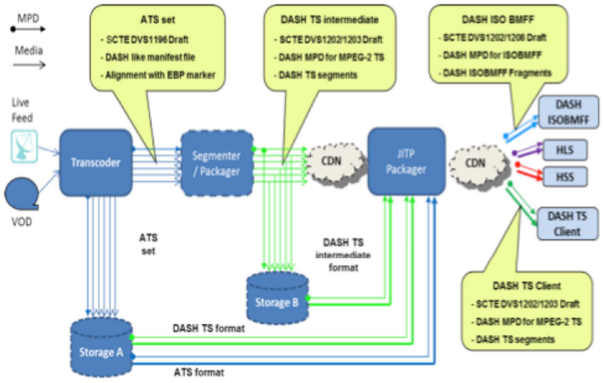


Figure 4: REFERENCE ARCHITECTURE DIAGRAM [7]

ATS SOURCE DESCRIPTION AS AN INTERMEDIATE MANIFEST FORMAT

An ATS set needs a detailed description format that brings together differerent files and multiplexes within the set and provides sufficient amount of information for generating content in the end format. Feature-richness of DASH makes it a perfect candidate for describing the multitude of representations contained in our intermediate format. It's a suitable base for the generation of client manifests used by different adaptive streaming systems. The purpose of source description is somewhat different from the one in adaptive streaming – a packager is told by the client explicitly which representation to chose. The packager needs to be able to select and transform intermediate format into the requested format; frequently the packager will also double as a manifest generator – providing the client with the manifest it will use for selecting the representation and requesting the segments.

Source description cannot really be called "DASH," as we have to allow for non-HTTP sources, such as UDP multicast and locally stored files – but it is as close as it can be. The selection of DASH Media Presentation Description (MPD) as a basis for the manifest format carries an interesting consequence: It makes it possible to mix UDP and HTTP into a single manifest, which means that an extended DASH/TS client can bring non-multicast resources over HTTP.

An ATS manifest needs to be designed that can be converted to an adaptive streaming manifest format – e.g. an HLS m3u8 or DASH MPD. DVS 1196 contains a definition of such a format, which is based upon DASH/TS – a format defined in SCTE DVS 1202 and 1203. The purpose of DVS 1202 is to establish container-independent constraints at the manifest and codec level that apply to both MPEG-2 TS (1203) and ISO-BMFF (1208) formats. DVS 1203 defines additional constraints, DASH MPD attributes and elements specific to an MPEG-2 TS format, while DVS 1208 defines additional constraints, DASH attributes and formats that are specific to ISO-BMFF profile and implemented through DASH-IF specification. This combination allows for the creation of a manifest for the intermediate format (using DVS 1202/1203) and the ability to convert this to a manifest for an ISO-BMFF file to an end client player (aka cloud-to-ground).

Intermediate format accommodates a superset of bitrates and formats needed to support traditional QAM distribution – thus MPEG-2 TS streams can be generated from it and be transmitted to existing set-top boxes, or used in legacy pay-per-view and VOD services delivered to QAM customers. Since our intermediate format is based on MPEG-2 TS, recreating a compatible stream is as simple as remultiplexing two-three

representations, and applying the appropriate conditional access system.

Using this hybrid structure offers several advantages for handling video services. It allows the adaptive streaming systems – packagers and clients – to use information acquired through known signaling mechanisms in MPEG-2 TS (e.g. SCTE 104 / SCTE 35), without creating new mechanisms to gather this information. Further, video and audio quality can be monitored and analyzed in the network using existing testing equipment, rather than building new equipment in the adaptive streaming environment to do the same thing. It can save on storage by storing one common intermediate format, and it can save on transcoding costs by transcoding once for both services.

PACKAGING TIER – CREATING FLEXIBILITY, ROBUSTNESS, AND SCALE

The packager described in Figure 4 fetches and converts CIF segments to end-client ISO-BMFF segments, and converts the CIF manifest into a manifest for ISO-BMFF-based adaptive streaming system, such as the DASH-IF IOP profile of MPEG DASH.

A simple packager implementation would rely on random access point signaling in MPEG-2 transport stream packets. Inserting random access points at fixed time intervals, and segmenting at every IDR frame, would suffice in this case. Despite our desire to see a single, universally-adopted adaptive streaming system and serve only it to all clients, this is far from becoming current marketplace realities. Thus dictates the need to support at least 2-3 of these systems. Different systems often require different segment lengths – e.g., HLS typically uses 8 second segments, while ISO-BMFF systems use 2 second chunks. The A-V delay is also an issue – for same video, the corresponding audio segment border would be in a different place in multiplexed MPEG-2 TS and in ISO-BMFF segments.

In a two-tier workflow implementation design, multiple packagers perform simultaneous segment generation at different locations. In order to allow a truly robust, scalable and distributed packager system, the packagers would be designed to be completely independent of each other, only implicitly synchronized by the PCR clock of their respective multicast inputs. It follows that in a fault-tolerant system, a failure of a packager instance should never affect the rest of the system.

Load balancing is another important consideration in package implementation design. It is typically handled internally at a CDN or cloud level, hence it is inadvisable, to perform crude DIY load balancing, using packager-specific URLs, and having the same session always bound to a single packager – which renders the overall system far more fragile and less scalable. As a result, we arrived at a requirement to have requests to same segment URLs return bit-identical HTTP response bodies, no matter which packager generated the segment. A client should be unaware which packager was serving it – this should be a load balancer decision and the load balancer should be able to change its decision at any arbitrary point in time.

The ATS format is the intermediate format of choice for a packager design – especially when mullticasts are needed or virtual segmentation is advantageous. An alternative and physically segmented format is the TS-based DASH/TS.

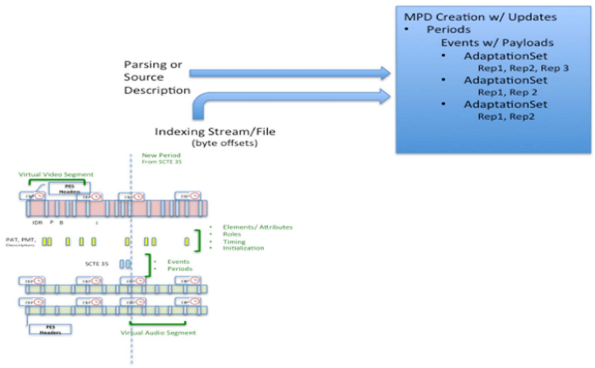


Figure 5: MANIFEST CREATION THROUGH MPEG-2 TS

In addition to segment generation, a just-in-time packager is expected to provide an up-to-date manifest upon client request. For this to happen, the just-in-time packager can use the intermediate format paradigm – using the DASH-like ATS source description.

Initial "bootstrapping" information needs to also make it to the new packager at its start-up time, by unspecified means. Once a packager has a sufficient knowledge of where the virtually segmented streams' location and characteristics, it then will maintain the manifest, reacting to the changes reflected in the incoming stream. These streams may be virtually segmented, adaptive transport streams or physically segmented DASH/TS.

In case of DASH/TS input, generation of an MPD is trivial – all that's required is a tight synchronization between the packager MPD and the origin MPD.

In some cases, an MPD needs to be generated and maintained by the expected changes in the stream, which can come directly from encoder. In the case of information propagated from the transcoder, a DASH MPD Update message can be used. This message contains a complete MPD within the stream, on PID 4, which was recently reserved by MPEG for precisely that purpose. MPD Update messages can be sent every time there is a change – e.g., switch to a different asset.

In the case of adaptive transport streams, maintenance of an MPD can be done either at the transcoder, or locally, at each packager. In both cases, new periods can be started at each splice point, or program changes can be indicated by the incoming SCTE 35 cue message. Out-of-band metadata or/and UPIDs in SCTE 35 segmentation descriptors can be used to derive period asset identifiers.

In case MPD is created automatically from the incoming MPEG-2 transport stream, translation of PMT and elementary stream information into MPD elements and attributes is relatively straightforward. Codec information can be derived from the PMT and – if needed – from parsing elementary streams.

As an example, the maximum segment duration and segment access point type can be learned from the EBP descriptor, and bandwidth values are also present in the adaptive stream PMTs.

Population of many of the adaptation sets and representation attributes also requires elementary stream parsing. Although some bits and pieces of information are available from (often optional) PMT descriptors, the expected level of detail in the MPD is higher than that. Profile/level information, if absent in the PMT, needs to be derived from the elementary streams.

ATSC caption service descriptors are used to signal carriage of CEA 608 / 708 closed captioning in MPEG-2 TS. These also need to be mapped into the MPD (for SCTE DASH and DASH-IF IOP 3.0) or into the m3u8 manifest for HLS.

Beyond language, audio descriptors can be used to derive signaling for associated services and broadcast / receiver mix. In case of MPEG DASH, a recent amendment to the specification created an interoperable way of signaling them

CONCLUSIONS

In this paper, a hybrid workflow system that allows reuse of existing infrastructure and creates a cost-effective backwards-compatible architecture for serving linear and on demand programming to large over-the-top audiences. It uses adaptive streaming technologies for segments using multiple container formats, without causing a large storage and transmission burden in order to support multiple formats. The system adds optimization, stability, scalability, and synchronization between transcoder systems and multiple packagers, using multicasting of MPEG-2 TS structures. Using the MPEG-2 TS structure and the manifest derived from this as a common intermediate format provides an elegant way to support today's fragmented adaptive streaming market, without being an undue burden. In the end, the hybrid delivery system takes advantage of the similarities in player decoding behaviors across adaptive streaming containers, and across the MPEG-2 TS and File format structures. Lastly, the concept of common intermediate format is extended to the player platform, to enable consistent browser playability. The paper describes an efficient way to convert MPEG-2 TS into ISO-BMFF-based adaptive streaming formats using the constraints defined in DVS 1196, DVS 1202, DVS 1203, and DVS 1208.

REFERENCES

[1] Jan van der Meer, Fundamental and Evolution of MPEG-2 Systems: Paving the MPEG Road, John Wiley & Sons, West Sussex, United Kingdom, 2014.

[2] ANSI/SCTE 35 2014, Digital Program Insertion Cueing Message for Cable.

[3] ANSI/DVS 1196 R1a 2015, Adaptive Transport Stream.

[4] ANSI/DVS 1202 R1a 2015, MPEG DASH for IP-Based Cable Services Part 1: MPD Constraints and Extensions.

[5] ANSI/DVS 1203 R1a 2015, MPEG DASH for IP-Based Cable Services Part 2: DASH/TS Profile.

[6] ANSI/DVS 1208 R1a 2015, MPEG DASH for IP-Based Cable Services Part 3: DASH ISO-BMFF Profile.

[7] ANSI/DVS 1206 R1a 2015, Reference Architecture for Cable ABR Services.

[8] Encoding At Netflix Media Meetup, May 2014- Netflix Digital Supply Chain, http://files.meetup.com/9842252/BAMM Meetup_NetflixEncoding_140521.pdf

Impact of CCAP to CM Distance in a Remote PHY Architecture

John T. Chapman, CTO Cable Access & Cisco Fellow, Cisco, jchapman@cisco.com
Gerry White, Distinguished Engineer, Cisco, gerrwhit@cisco.com
Hang Jin, Distinguished Engineer, Cisco, hangjin@cisco.com

Abstract

The Remote PHY architecture currently being specified by CableLabs and developed by multiple vendors has the potential to change the DOCSIS access network significantly.

The basis of the architecture is to move the PHY and the digital to analog conversion from the CCAP device in the hub to the optical node and replace the analog optical link between CCAP and the node with a digital link such as 10G Ethernet.

This removes the distance limitations imposed by the analog link and could in theory enable the CCAP core to be located in a remote data center thousands of kilometers from the CM. The DOCSIS protocol must continue to operate between CM and CCAP both within the normal 100 mile (160 km) DOCSIS range and over this new extended distance.

This paper looks at how this type of system could be deployed and the impacts of increasing the distance between the CM and the DOCSIS cores on throughput and latency seen by the end user.

The paper is in three parts: theory, simulation and test.

Part One explains the R-PHY approach and discusses some potential deployment scenarios. It looks at the impact on the DOCSIS protocol of moving the CCAP core to a remote data center and using an IP network to connect the core to the R-PHY node.

Part Two describes a simulation model developed to investigate the relationship between CM to CCAP distance and DOCSIS performance. The model has been used to investigate the impact of both latency and jitter in the network connecting the CCAP core to the Remote PHY and the paper will present these results

Part Three describes a series of experiments performed to validate the model and confirm the impact of extended distances in a real world deployment. The paper presents these results and compares them with the theoretical analysis and the simulated data.

This paper also discusses on how a Remote PHY based CMTS can operate at distances much further than an I-CMTS.

TABLE OF CONTENTS

TABLE OF CONTENTS **2**

INTRODUCTION **3**
- How R-PHY Works 3
- How R-PHY is Deployed 5
- Why R-PHY is Cool 6
- How R-PHY Compares 8
- Recap 11

PART 1: THEORETICAL DISCUSSION **12**
- The Case for a Centralized DOCSIS Scheduler 12
- Reference Network for Latency Analysis 13
- How The REQ-GNT Latency Works 13
  - *A Gut Check* *16*
  - *Poisson Distribution of REQ-GNT Times* *16*
  - *DOCSIS Pipelining* *17*
- Scenario 1: 0 to 100 Miles 18
- Scenario 2: Beyond 100 Miles 19
- Recap 20

PART 2: SIMULATION **22**
- Simulation Function Blocks 22
  - *Ingress Packets* *22*
  - *Queue Management* *22*
  - *Request Management* *22*
  - *Request Reception* *22*
  - *US Data Reception* *23*
  - *Scheduler* *23*
  - *MAP Builder* *24*
  - *CM MAP Process* *24*
- Simulation Results 24
  - *Algorithm 1: Baseline algorithm* *25*
  - *Algorithm 2: Proactive Grant Algorithm* *29*
- ReCap 32

PART 3: TEST & MEASUREMENT **33**
- Test Set Up and Components 33
  - *Traffic Profile* *34*
  - *Measurements* *35*
  - *Latency Added by Test Bed* *35*
  - *CIN Delays Tested* *35*
- Non-ideal Testing Conditions 35
  - *10% versus 90% Loading* *35*
  - *1 GE versus 10 GE* *36*
  - *IEEE 1588 based R-DTI Timing* *36*
- Baseline Results for Latency 36
  - *I-CMTS and R-PHY Downstream* *36*
  - *I-CMTS and R-PHY Upstream* *37*
- CIN Delay and Downstream Latency 37
  - *Average Downstream Latency* *38*
- CIN Delay and Upstream Latency 39
  - *Average Upstream Latency* *39*
  - *Minimum Upstream Latency* *40*
  - *Maximum Upstream Latency* *41*
  - *Upstream Latency Distribution for a Single CM* *42*
- Why the Measured Results were Better than Theory 42
- Throughput and Frame Loss 44
  - *Downstream* *44*
  - *Upstream* *44*
- Summary of Test Results 44
- Recap 45

UNIFIED THEORY OF US LATENCY **46**

CONCLUSIONS **47**

APPENDIX I – DELAY IMPACT OF NODE TO CORE DISTANCE **49**
- Transmission Time 49
- Switching Time 49

APPENDIX II – CCAP CONFIGURATION **50**

REFERENCES **51**
- *Related Papers by the Same Authors* *51*
- *Related CableLabs Specifications* *51*

ACKNOWLEDGEMENTS **52**

INTRODUCTION

How R-PHY Works

Remote PHY (R-PHY) is the next stage in the evolution of DOCSIS and video service delivery. It is a modification of the current CCAP (Converged Cable Access Platform) architecture in which the PHY component is removed from the CCAP platform and relocated into a separate Remote PHY Device (RPD). To better understand the R-PHY approach, we will compare it to the I-CMTS.

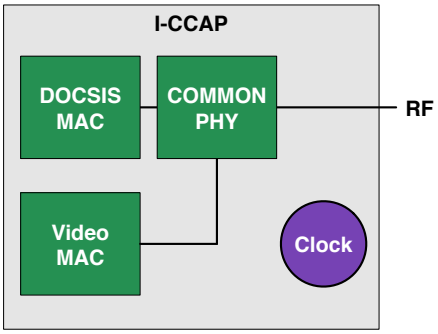


Figure 1 – Integrated CCAP

A very simplified view of a CCAP device is shown in Figure 1. There is a DOCSIS MAC that is a combination of classification, rate shaping and framing for both the upstream and the downstream.

There is also a video MAC that is a combination of encryption, signaling, and MPEG-TS multiplexing. Both the DOCSIS MAC and the video MAC connect into a common PHY. Today this PHY is QAM for DOCSIS 3.0 and video, but will expand to include OFDM for DOCSIS 3.1.

The same components – but arranged differently in Figure 2 – form an R-PHY system. The RPD is connected to the CCAP Core (CCAP minus the RF PHY) by an IP network. The combination of CCAP Core and RPD provide the functional equivalent of the integrated CCAP.

Essentially R-PHY takes the digital interface between the MAC and the PHY components in the CCAP and extends it over the IP network to the RPD using pseudowire technology. That means the I-CMTS and the R-PHY systems are using the same components and the same interfaces.

The resulting system is shown in Figure 3. It is very important to note that the R-PHY Device barely participates in the DOCSIS protocol. The RPD is like a layer 2 gateway device that moves packets from one side the other, providing basic PHY layer conversion

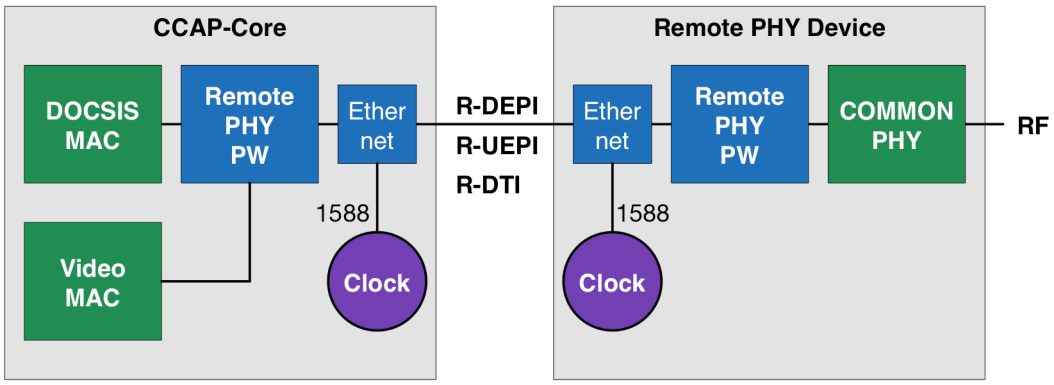


Figure 2 – Remote PHY CCAP

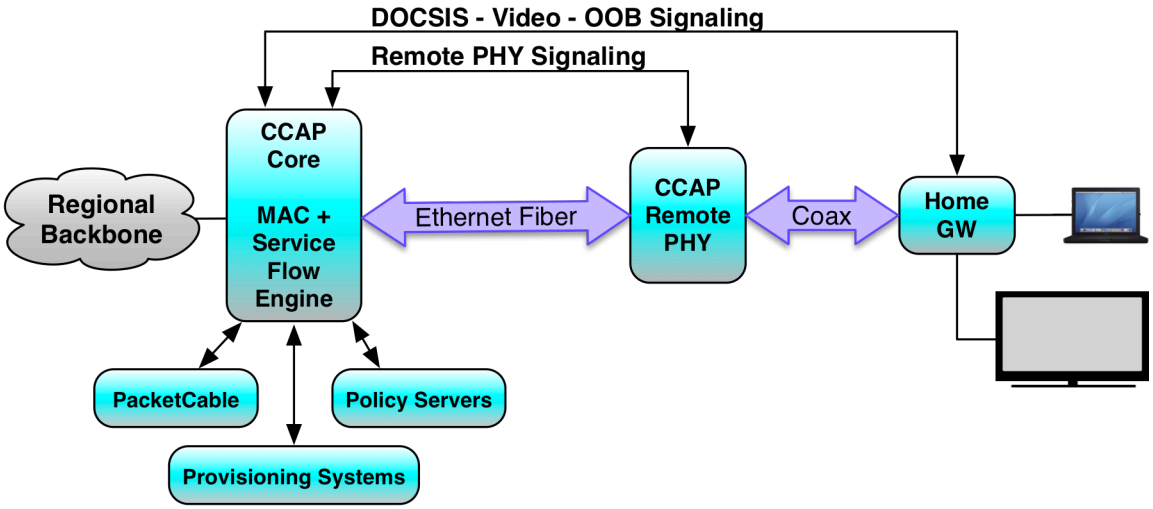


Figure 3 – Remote PHY System

and encapsulation services. There is no scheduler or DOCSIS framing in the RPD. The one exception to this rule is that the RPD is allowed to extract REQ messages out of the upstream so that the REQ messages can be sent with higher QoS than the data.

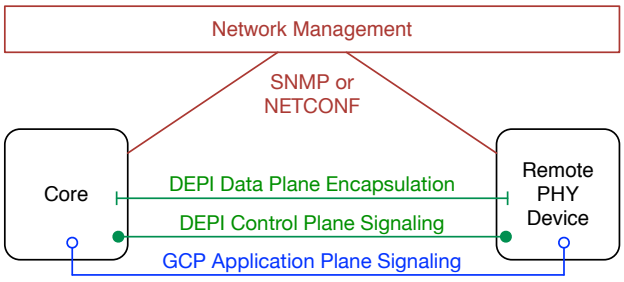


Figure 4 – R-PHY Signaling

The R-PHY architecture uses a set of standardized protocols for its operation as illustrated in Figure 4. The pseudowires (IP tunnels) between the Core and the RPD are managed with a protocol called R-DEPI (Remote Downstream External PHY Interface).

An extension to DEPI is R-UEPI (Remote Upstream External PHY Interface). R-DEPI is based upon the IETF protocol L2TPv3 (Layer 2 Tunneling Protocol version 3). The R-DEPI control plane is used for session establishment and tear down and to manage fail-overs for HA (High Availability). The R-DEPI data plane also includes S-BFD (Seamless Bidirectional Forwarding Detection) to perform data plane diagnostics. R-DEPI supports DOCSIS, Video, and OOB (Out of Band) all using the same techniques and protocols.

To program the ASICs and their corresponding functionality within the RPD, a separate protocol called GCP (Generic Control Plane) protocol is used. GCP is a protocol shim that carries application specific data elements, such as TLVs (type-length-value), over the GCP shim, using a transport such as TCP.

All components are also manageable through SNMP or similar evolved network management protocols such as NETCONF/Yang.

For a more in depth view of R-PHY refer to the References section at the end of this paper.

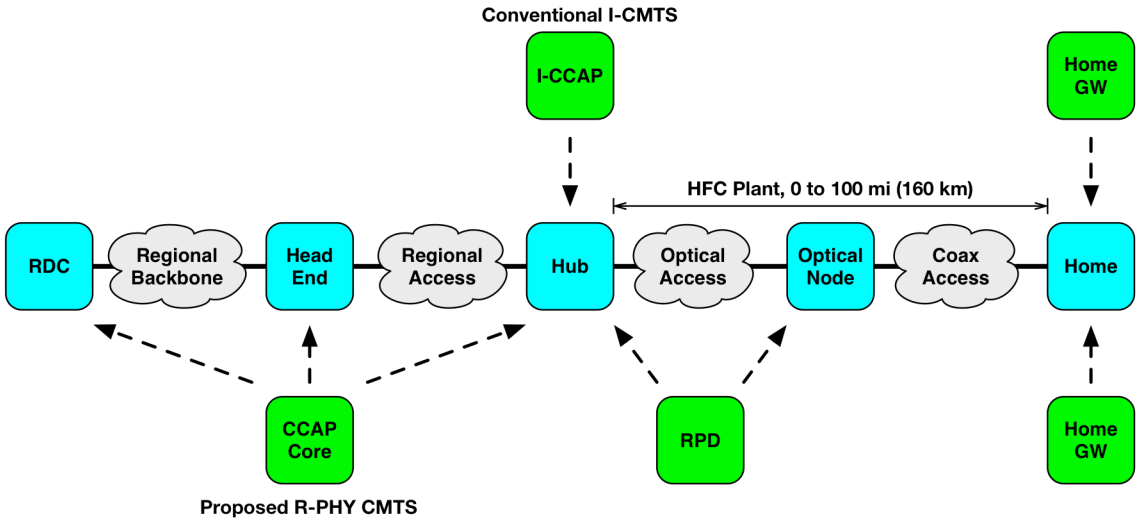


Figure 5 – R-PHY Deployment Model

How R-PHY is Deployed

A reference model for a typical plant of the Cable Operators is shown in Figure 5. A Regional Data Center (RDC) is connected to multiple head ends over a regional backbone network. The head ends in turn are connected to multiple distribution hubs over a regional access network.

Both the regional backbone and regional access networks are typically IP/Ethernet networks running over a digital optical transport. The distribution hubs are connected to optical nodes over the optical access network.

Historically the optical access network has been an analog network but in an R-PHY deployment this can also be IP/Ethernet over digital fiber. From the optical nodes coaxial cable carries an analog signal link to the cable modem at the subscriber premise.

In this reference architecture, the return path fiber from the optical node terminates at the Hub. For this reason, the I-CCAP has been located in the hub. In some systems where the hub has multiplexed and transported the return path fiber to the head end, the I-CCAP can be located at the headend. However, for this white paper, we will refer to the I-CMTS location as being at the hub.

One distinct advantage of the R-PHY architecture is that it removes the distance limitations imposed by the analog link and could in theory enable the CCAP core to be located in a remote data center thousands of kilometers from the CM. Thus R-PHY enables the CCAP-core to be located in the RDC, head end or hub. Similarly the RPD can be located in the Hub or optical node.

The DOCSIS protocol must continue to operate between the CCAP and home gateway over this extended distance. The intent of this paper is to explore the impact of the distance from the Core to RPD on DOCSIS performance and thus to determine which of these deployment scenarios can be used in practice.

Why R-PHY is Cool

An R-PHY System with a CCAP Core and a set of RPDs is functionally equivalent to an I-CCAP. Thus an I-CCAP can still be the best choice if the I-CCAP is located at the hub and feeding a classic analog HFC plant.

The deployment choice of an R-PHY system has to be based on providing value to the customer and solving customer problems that cannot be solved by a traditional I-CCAP. As such, the R-PHY system is positioned to address new and emerging market segments.

Here are a few of the value propositions of a CCAP Remote PHY system.

1. **Enables hub site consolidation**

   Because of the return path fiber limitation, each hub site requires one or more CCAPs. This is not a problem for large hubs where the high density of the CCAP is an advantage but in a small hub may result in underutilized CCAP platforms.

   For operators with lots of small hubs, the total number of partially used CCAP chassis can add up. When the CCAP Core is moved from the hub to the head end in a system with small hubs, better utilization can be achieved and far fewer CCAP chassis may be needed. This can lower the overall system capital costs and operating costs.

2. **Higher bit-rates for DOCSIS 3.1**

   If the DOCSIS 3.1 PHY is located in the optical node, then the transmission impairments from the optical path and the laser and receiver are eliminated.

   That will allow DOCSIS 3.1 to operate at one or two higher levels of modulation. This translates to a 10% to 20% increase in throughput for a given spectrum.

3. **Longer reach**

   Analog fiber cannot reach as far as digital fiber. This distance limitation may prevent deeper fiber deployments.

4. **Enables Ethernet to the Node**

   Replacing the analog optical links to the node with standard Ethernet links enables the operator to leverage the costs, scaling and product development of this much larger market e.g. migration from 1G to 10G to 100G.

5. **Efficient sharing of HFC fiber with FTTH and PON.**

   One of the new major investment areas for Cable Operators is PON (Passive Optical Networks). The current installation plan for PON is to run new fiber all the way from the hub to the home.

   Deployment expense could be lowered on the PON installation if new fiber was installed from the optical node to the home for the PON and backhaul was shared over the same fiber network with R-PHY.

6. **Enables sharing of commercial and residential plants**

   Currently, many cable operators have two plants they own and manage. One is an analog fiber plant for residential and the other is a digital fiber plant for

commercial.

If both plants were digital, these could be one common plant. That means that residential or commercial services could be sold and installed anywhere within the combined footprint of both plants. It also means that an investment in one plant is an investment in the other.

7. **Enables HFC fiber to become a full service IP network**.

   Currently the HFC plant is a semi-proprietary network that carries two specific services – DOCSIS and video.

   What if the fiber plant was an IP network? Now that same fiber could carry other services such as wireless backhaul and business services. This would increase the value and usefulness of the fiber plant.

8. **More SG per wavelength**
   In an analog optics system, one wavelength services one node and one SG (service group). In a digital system, the 10GE can service multiple SG since the 10GE can easily multiplex packets.

   If each SG was 5 Gbps, then a 10 Gbps backhaul could support two SG. If there is shared bandwidth across SG due to multicast, then up to four SGs could share one 10GE link. This is one forth the number of forward path lasers that are needed. This provides a 4x scaling difference.

9. **More wavelengths**

   Analog optic systems are typically limited to 16 wavelengths per fiber due to cross talk between the spectrum on each wavelength. Digital optic systems can go as high as 80 wavelengths.

   When deploying deep fiber with more SG, the extra wavelength capacity means that fiber does not have to be pulled all the way back to the hub. This provides a 5x scaling difference.

10. **Better scaling**

    With the PHY separated from the MAC, from the packet engine and from the backhaul, the PHY and MAC can be custom matched to provide a more efficient solution.

    For example, a given backhaul and MAC could support more SG on a DOCSIS broadband only system than on a DOCSIS broadband and Video system. A fixed MAC must be a compromise for all options. This provides a 4x or so scaling difference.

    Note: combining items 8-10 results up to a 80x scaling difference.

11. **Lower plant maintenance costs**

    HFC plants based on analog fiber should be calibrated using a sweep and balance technique twice a year. Digital plants with deep fiber and N+0 passive access do not need the same calibration. Digital optical nodes are simpler to install and operate.

12. **Lower optics costs (10G)**

    With the increase in use of 10GE optics in networking, 10GE optical transmitters and receivers should become less expensive than the analog optical transmitters and receivers.

    The circuit design that supports the

digital optics is also simpler and cheaper than the circuit design that supports the analog optics.

13. **Simpler fiber design rules**

    Analog optics have more complex link budgets as the signal can degrade with distance and the quality of the fiber. Digital systems have simpler design rules. This means that it should be easier and less costly to do field designs.

14. **Replace RF combing with switching**

    In today's hub, there is a large RF combining and splitting network. Ethernet switches could replace all this. This would be more compact and easier to troubleshoot.

    The RF to Fiber to RF conversion would be completely replaced again by Ethernet and Ethernet switches. This will simplify operations, reduce power, cooling and rack space.

15. **Enables CCAP virtualization**

    Virtualization refers to moving the packet forwarding engine and its surrounding services from a dedicated hardware appliance (such as an I-CMTS and moving it into the cloud to run on a generic server architecture.

    An I-CMTS has RF ports and RF ports cannot be virtualized. The R-PHY architecture solves this by moving the RF circuitry to a separate chassis. The remaining circuitry in the CCAP Core can then be virtualized.

How R-PHY Compares

CableLabs has defined an over-all solution umbrella called the Distributed CCAP Architecture (DCA). Within that architecture, there are different approaches to distributing the software and hardware between the core and the remote device. The three recognized approaches are:

- *Remote PHY*: The PHY chip is placed in the remote device. The rest of the hardware and all of the CCAP software remains centralized in a core device.
- *Remote MACPHY*: In a true remote MACPHY, the PHY, the MAC framer, the US scheduler, and half the MAC signaling is placed in the remote device, while the packet classification and switching remains centralized.
- *Remote CCAP*: The entire CCAP is placed in the remote device. There is no core entity other than a standard router.

In addition, a vendor may choose one approach for DOCSIS and a different approach for Video and OOB.

In this white paper, the Remote PHY approach is used consistently for DOCSIS, Video, and OOB. Here are the advantages of the Remote PHY architecture over other architectures.

1. **It's simple**

   The least amount of circuitry is placed in the RPD. This would be the PHY chip as it exists today. None of the CCAP software is placed in the RPD.

   The RPD does of course have a minimum amount of software to operate, but it is not subscriber aware.

2. **It works**

   R-PHY uses a MAC-PHY interface that was initially designed back in DOCSIS 2.0. Back then it was called DMPI, or DOCSIS MAC-PHY Interface. This later led to DEPI and UEPI.

   DEPI was used for M-CMTS (Modular CMTS). UEPI was used for DOCSIS 3.0 chips inside of most if not all I-CMTSs. So, R-PHY is a proven design that is being updated and extended across a longer network.

3. **Complete Interoperability**

   This is really the most important feature. All R-PHY CCAP Core vendors will be able to interoperate with all RPD manufacturers. Any CCAP core will work with any optical node vendor. CableLabs will drive this interoperability.

4. **Complete standardization at CableLabs**

   Only the Remote PHY community has worked with CableLabs to produce seven specifications that will completely describe the Remote PHY operation. Alternate architectures will be described by a technical report, but there will not be any requirements listed. The seven Remote PHY Specifications are:

   - CM-SP-R-PHY: Remote PHY System Specification
   - CM-SP-R-DEPI: Remote Downstream External PHY Interface
   - CM-SP-R-UEPI: Remote Upstream External PHY Interface
   - CM-SP-R-DTI: Remote DOCSIS Timing Interface (based on IEEE-1588)
   - CM-SP-R-OOB: Remote Out Of Band
   - CM-SP-R-OSSI: Remote PHY Operations Support System Interface
   - CM-SP-GCP: Generic Control Plane

5. **Centralized software**

   All CCAP software, including the upstream scheduler, is located centrally and made by the same vendor.

6. **Same feature velocity as I-CCAP**

   The CCAP software for R-PHY is the same software used in the I-CMTS. That means the much larger I-CMTS market will pay for the software for the much smaller and emerging DCA market.

7. **DOCSIS MAC can be upgraded centrally.**

   The DOCSIS MAC contains a huge amount of functionality (the majority in software) – much more than a normal MAC such as an Ethernet MAC – and it is something that changes on an ongoing basis.

   By having a centralized MAC, it can be easily upgraded. With a remote MAC, this is much harder to do, especially if it is a cost reduced hardware based MAC implementation.

8. **MAC and scheduler can be scaled as needed since they are central.**

   It is unknown at this time what the CPU impact will be for the DOCSIS 3.1 scheduler. The DOCSIS 3.0 scheduler also has a CPU impact that is proportional to the amount of data capacity it is managing.

To put the scheduler into the node means designing for the maximum compute resources plus engineering margin now even though those resources may not needed at deployment.

9. **Same consistent approach for DOCSIS, Video, and OOB**

   R-PHY uses the same consistent architecture, signaling, and pseudowire technology for DOCSIS, video and OOB. This provides architectural leverage and operational simplification.

10. **Wi-Fi, EPOC, Cloud-RAN and other access technologies use a similar approach.**

    Many other access technologies arrived at the same conclusion of putting the least amount of electronics and intelligence at the edge. Wi-Fi for example, uses centralized software algorithms for tuning RF performance. C-RAN (Cloud Radio Access Network) goes to the extreme of only putting DAC and ADC with RF at the edge and centralizing the rest. EPOC (EPON over Coax) used a remote PHY in a CMC (Coax Media Converter) with a centralized scheduler.

    The best solutions to difficult problems are usually simple solutions. Centralized solutions historically have been simper than distributed solutions. This is a form of architectural validation.

11. **DOCSIS and Video traffic are already encrypted on the fiber.**

    With a centralized DOCSIS and video MAC, all traffic from the CCAP Core to the CM will be encrypted. If the DOCSIS and video MACs, along with the DOCSIS and video encryption were moved to the RPD, then all the DOCSIS and Video content on the fiber would be in the clear.

    This would lead to needing a new encryption link from the core to the RPD such as IPsec. There would then be a conversion between one encryption scheme to the other at the RPD. This would require two security schemes to protect the data and video, doubling the chances of configuration errors and increasing the vulnerability.

    Why bother with twice the encryption cost, complexity and conversion when you can do it with just one encryption scheme by keeping the payload encryption centralized.

12. **Minimum components in RPD should yield best cost**

    By minimizing the components and the CPU compute requirements in the RPD, the RPD has the opportunity to become more economical. It is also the intent of the R-PHY architecture that the RPD become a single chip device.

13. **Minimum components in the RPD should yield the best availability.**

    Less hardware components means better MTBF. Less software features means less bugs, more stable software and fewer upgrades. All of these increase system availability.

14. **Supported by multiple silicon vendors**

    There is a lot of silicon talent in the industry that can design the PHY chips. However, it requires a different skill set and more investment money and time to

get the MAC right.

Recap

- Remote PHY is really a centralized software architecture
- Remote PHY CCAP and RPD systems will be standardized and interoperable.

PART 1: THEORETICAL DISCUSSION

The Case for a Centralized DOCSIS Scheduler

Since the beginning of the R-PHY architecture, there has been a technical debate as to where the upstream scheduler should be. Should it be in the CCAP Core where the rest of the software is or should it be in the optical node with the upstream PHY?

Perhaps we have been looking at this all wrong. Perhaps it is not a technical decision. Perhaps it is a business decision. If an R-PHY system could be built and could work well with the upstream scheduler at either location, are there compelling business reasons to pick one location over the other?

If there was just one really good reason to have a centralized upstream DOCSIS scheduler, that reason would be interoperability. If a centralized scheduler could be made to work, then the manufacturer of the CCAP Core and the RPD could be different.

Any CCAP Core vendor could connect to any RPD vendor. The RPD would not have any knowledge of the upper layers of DOCSIS nor care. The CCAP Core would dictate all system performance.

As a cable customer, do you really want your CCAP purchasing decision determining your optical node manufacturer? In reverse, do you want your purchasing decisions for your optical node to determine your CCAP vendor? Probably not.

If the CCAP Core came from one vendor and the RPD plus a remote scheduler came from another vendor, what is the probability the system would work well together? The upstream scheduler determines all the upstream QoS. That means the downstream QoS would be from one vendor and the upstream QoS from another vendor. Not a good scenario.

And who would you call when something goes wrong or you need a feature added? What kind of system performance, stability and upgrade path would you have with a CCAP Core from one vendor and RPDs from multiple other vendors on that same CCAP core? These are operational considerations that transcend minor differences in performance.

So, will a centralized upstream scheduler work? What is the definition of working well? The big difference in centralized and remote upstream scheduling is the REQ-GNT latency. Lets analyze REQ-GNT latency for that answer.

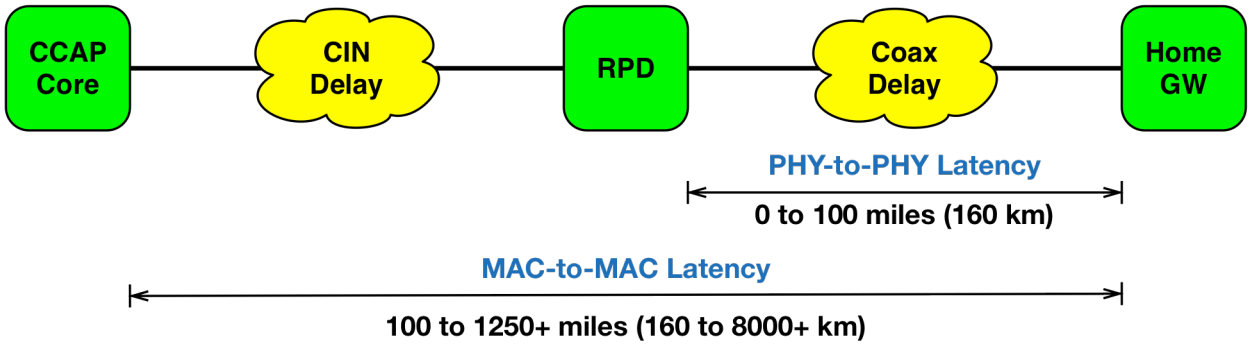


Figure 6 – R-PHY Network Latency Model

Reference Network for Latency Analysis

A simplified, conceptual view of an R-PHY network is shown in Figure 6. This model will be used going forward to discuss the impact of increased CCAP Core to RPD separation. The Converged Interconnect Network (CIN) has been introduced as a single entity to represent the CCAP Core to RPD link.

It is the delay imposed by the CIN that could impact the DOCSIS protocol as the CIN could be anywhere from a single Ethernet link to a multi-hop IP network.

There are two general scenarios for which R-PHY Latency can be evaluated.

1. For distances within the DOCSIS limits of 0 to 100 miles (0 to 160 km).
2. For distances greater than the DOCSIS limit of 100 miles (>160 km)

These two scenarios are very significant. The DOCSIS specification imposes a maximum operating distance from CMTS to CM of 100 miles (160 km) for DOCSIS 3.0. Note that for DOCSIS 3.1, the distance limit has been reduced to 50 miles (80 km).

The first scenario really compares that latency of an R-PHY system to the latency of existing DOCSIS I-CMTS systems.

The second scenario is a new market opportunity that only R-PHY can address. For the R-PHY scenario, the distance limitations that went into the 100 mile limitation are divided out into two distinct categories:

- PHY-to-PHY distance retains the 100 mile distance limitation and pertains to ranging and power levels. Since the CCAP PHY is in the RPD and stays in the hub or node, there is no increase in the PHY-to-PHY distance.
- MAC-to-MAC distance refers to software processes that are dependent upon the delay from the CCAP to the CM. The classic example is the upstream scheduler that is the main subject of this white paper. Other dependencies may include message time-outs.

Before examining if the upstream scheduler will work with R-PHY for each scenario, lets review how the upstream latency impacts the upstream scheduler.

How The REQ-GNT Latency Works

Downstream Latency

The downstream latency seen by DOCSIS packets between arriving at the

CMTS core and being received at the CPE is made up from the following components:

- CMTS packet downstream processing time
- CIN delay due to transmission time on links and switching time in intermediate nodes (refer to Appendix I – Delay Impact of Node to Core Distance)
- RPD downstream processing time
- Transmission delay over the analog network including transmission time, and PHY impacts such as FEC and inter-leaver delays
- CM downstream processing
- Transmission time on the CM to CPE link

Of these components, only the CIN delay varies as the distance between the CMTS Core and the CM change. Downstream DOCSIS operations on a packet are relatively straightforward (classify, apply QoS, add DOCSIS header, transmit) with the impact of CIN delay a linear function.

Thus the measured downstream latency should be equal to the CIN delay plus a fixed processing overhead as shown in Figure 7. The processing delay of 1.7 ms used in this plot was measured for 1518 byte packets using an integrated CMTS.

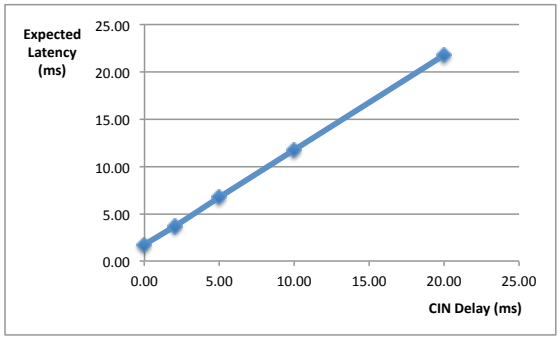


Figure 7 – Expected Downstream Latency vs CIN Delay

The CIN delay for a 100 mile plant is roughly 800 usec. The graph is calibrated in delay that can be later converted to distance.

Note that packets generally always see this delay in Figure 7 since the packets have to travel from the CPE to the RDC on their way to an Internet peering point. So, the significance here is that although the overall network delay is the same, the delay on the regional link is smaller and the delay that the CIN sees is greater as the CCAP Core moves away from the RPD.

Upstream Latency

The upstream latency seen by DOCSIS packets between being transmitted from the CPE and being transmitted into the core network from the CMTS core is much more complicated due to the DOCSIS upstream scheduling.

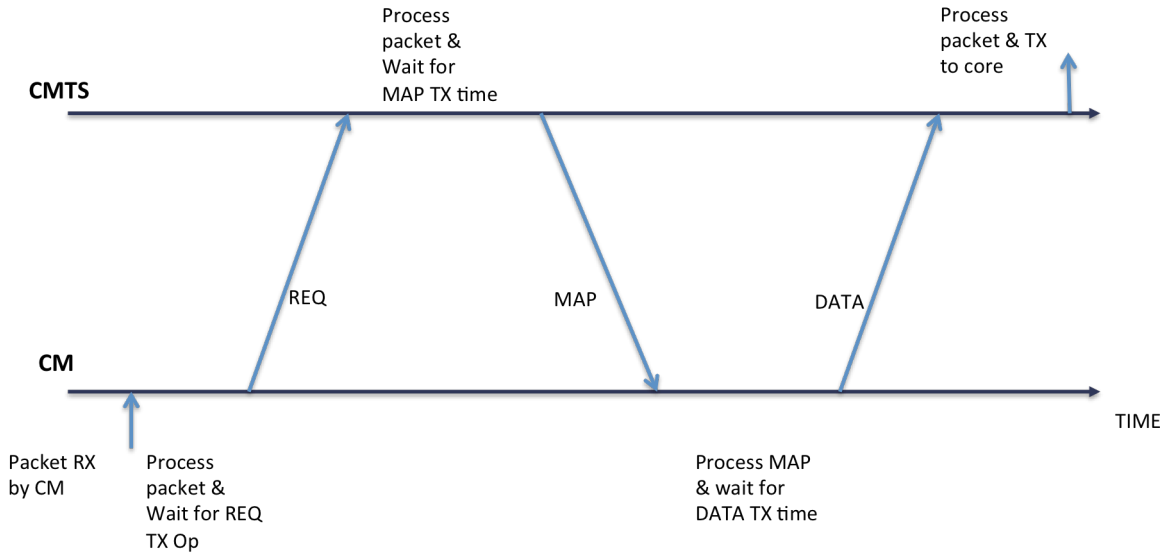


Figure 8 – Upstream DOCSIS Data Transmission

The upstream latency is made up from the following components and shown graphically in Figure 8:

- Transmission time on the CPE to CM link
- CM upstream processing
- CM waits for a transmit opportunity to send a request
- CM sends a bandwidth request to CMTS via RPD and CIN
- CMTS processes the Request and generates MAP with a data grant
- MAP transmitted to CM via CIN and RPD
- CM waits for transmit opportunity to send data
- Data packet transmitted to CMTS core via RPD and CIN
- Packet processed by CMTS and transmitted to core network

From Figure 8 it can be seen that approximately three transmissions traverse the CIN for each packet sent, the bandwidth request, the MAP and the data packet itself. (Approximately because requests are in bytes which could be part of a packet or multiple packets)

Thus the measured upstream latency would be expected to be equal to at least three times the CIN delay plus a fixed processing overhead plus a potentially variable overhead related to the DOCSIS protocol.

Of these components, only the CIN delay and possibly the DOCSIS protocol overhead vary as the distance between the CMTS Core and the CM changes. A first order approximation could assume the DOCSIS protocol delays to be invariant with CIN delay and model the upstream latency as a fixed offset plus three times the applied CIN delay as shown in Figure 9.

This uses 6ms as an approximation for the fixed offset and DOCSIS protocol delay (the 6ms value is based on measurements of upstream latency taken from I-CMTS systems).

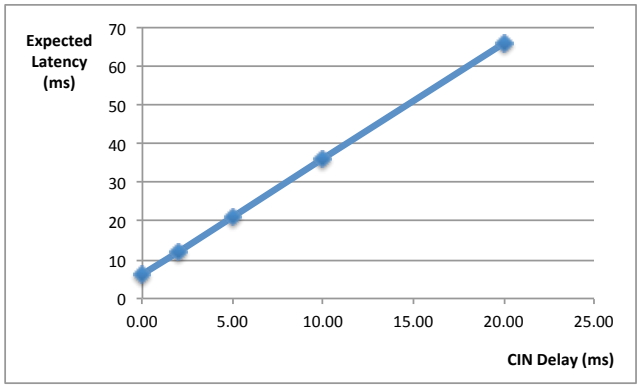


Figure 9 – First Order Approximation of Upstream Latency as a Function of CIN Delay

However the interaction between the CIN delay and the DOCSIS scheduling is potentially more complex than in the downstream case as some of the processing and wait times within the DOCSIS protocol can overlap with CIN transmissions. Thus Figure 9 should be considered to be a rough guide at best.

A Gut Check

Do upstream packets on a DOCSIS system really see a 6 ms delay? Does that make sense? Let's do a gut check by looking at the problem completely differently. Lets look at it from a CM viewpoint and how the CM interacts with MAPs.

Today, for DOCSIS 3.0 systems, a CMTS typically sends a bandwidth MAP to the CM roughly every 2 ms. This is shown in Figure 10. The CMTS also sends the MAP in advance of when it is needed. So at any one time, there is a one MAP in use and one MAP on deck that is about to get used.

Thus the most efficient a DOCSIS system gets is a REQ-GNT cycle that is every second MAP which is approximately 4 ms. This can vary quite abit because MAPs are rounded off to packet boundaries, so they can be longer or shorter than 2 ms. Also the REQ and GNT can be anywhere in the MAP which adds another +/- 2 ms of uncertaintly.

But, if we stuck with 4 ms as an average, and then allowed for 1 ms for queuing and process time in both the CM and CMTS, we would get the 6 ms of minimum upstream latency for a packet that is seen in the test results. This is a minimum latency as the REQ could see contention and have to be resent which would result in a much larger request latency.

Poisson Distribution of REQ-GNT Times

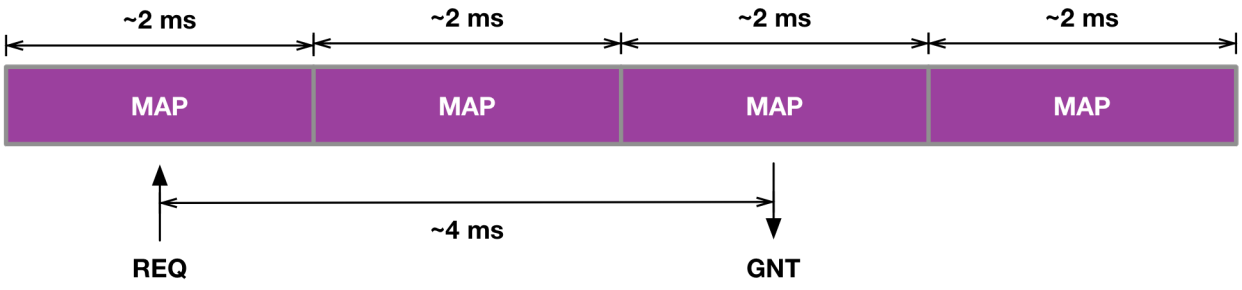


Figure 10 – MAP with one REQ-GNT cycle

In this basic gut check, the REQ-GNT delay is about 6 ms. However, this is really a minimum delay. During a loaded upstream, there will be contention on the REQ slots. This contention will cause the REQ-GNT delay to increase, perhaps well beyond the 6 ms.

The increase in delay might not happen if the requests are sent in a piggyback. However, since Piggy backing only occurs in tight bursts that are within 4 ms of each other, there can still be many restarts of DOCSIS flows.

One could argue that the increase in request-grant delay actually helps to throttle TCP connections during congestion. Delay is a natural part of the networking process. The trick for a protocol such as DOCSIS in a congested situation is to continue to work in a well-defined manner and do so at wire speed.

With 6 ms (or a number such as this) and random occurances of contention with random arrival times of REQs, the REQ-GNT delay most likely will follow an offset Poisson Distribution.

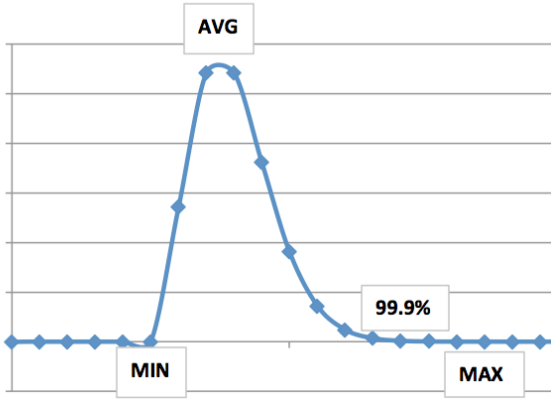


Figure 11 – Offset Poisson Distribution

In this distribution, there would be a minimum delay which is the distribution offset. This might be a value like 6 ms. Then there would be an average value, such as 12 ms. Then there would be a max value, like 50 ms.

The tail of this distribution may need to be modified to be more exponential to mirror the exponential backoff algorithm of the DOCSIS Request algorithms, but that is a refinement for another time. It is more the concept that is important right now.

DOCSIS Pipelining

How does DOCSIS work with a 6 ms or higher upstream packet latency? The answer is the DOCSIS does what all high speed

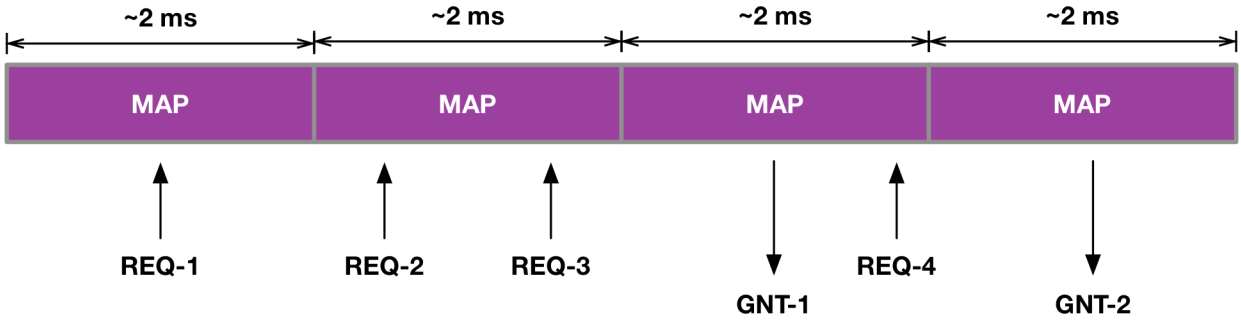


Figure 12 – MAP with pipelining

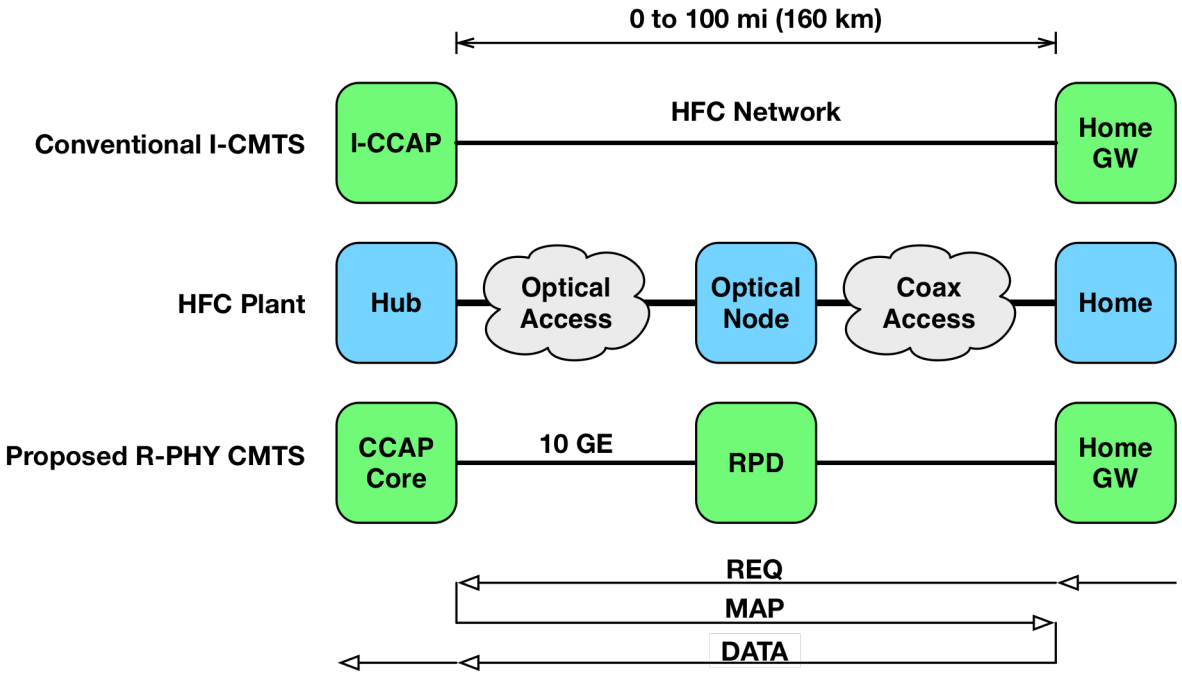


Figure 13 - Latency Scenario #1: 0 to 100 miles

systems with latency do – it pipelines.

DOCSIS allows for multiple outstanding requests. This is pipelining and is shown in Figure 12. The next operation is started before the prior one is complete. Further, the REQ and GNTs are byte based, so they can get larger to accommodate a build up of packets in the CM.

The net result is that once the DOCSIS system gets up and running, it is able to achieve wirespeed. In fact, it can achieve wirespeed for very large values of delay. DOCSIS needs to do this for an I-CMTS because on a busy upstream there can be lots of REQ latency due to contention, but the upstream still can operate at saturation.

Note that given this baseline minimum limit of around 6 ms based upon MAP usage, making the plant considerably smaller, such as putting the scheduler in the optical node, may not result in any signifigant performance improvement.

Even if the length of the MAP interval were made smaller, there is a MAP advance time that would dominate and may just require more MAPs between the REQ-GNT time so that the overall delay would be about the same. If there was any improvement, any real increase in latency would only be on the order of 1-2 ms.

Meanwhile, with or without such an improvement in latency, the throughput for either case would still be wirespeed due to the DOCSIS pipelining.

Scenario 1: 0 to 100 Miles

In scenario #1, the I-CCAP is located at the hub site because this is the point in the

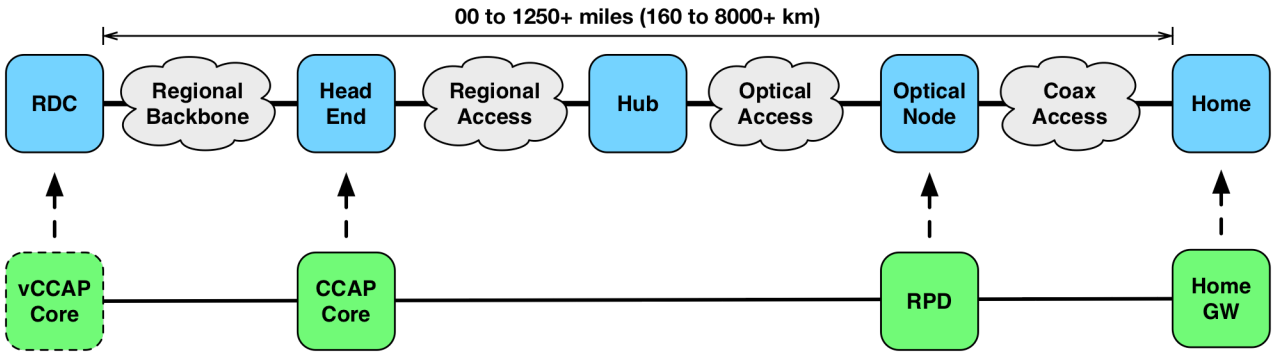


Figure 14 – Latency Scenario #2: Beyond 100 miles

network where the return path fiber terminates. The CCAP Core is also located at the hub site. This is shown in Figure 13.

This means the distance from the CM to the I-CMTS is the same distance as the distance from a CM to a CCAP-Core.

The same PHY is used for an I-CCAP as for the R-CCAP. It is just that PHY is located in a different location. In fact, the 10G MAC-PHY interface is also the same for both the I-CCAP and the R-PHY. This is by design. Both systems use the same chip sets and software protocols.

The REQ-GNT traffic still traverses the same path for both the I-CMTS and the R-PHY system. Despite the location of the PHY (central or remote) the REQ-GNT traffic remains essentially unchanged.

So if the network between the CCAP Core and the RPD were a perfect network with no congestion or jitter, then the round trip REQ-GNT latency should be the same for I-CMTS and R-PHY for DOCSIS networks under 100 miles (160 km). To restate, there is little to no impact due to network latency as that was already built into the DOCSIS network budget.

There is one concern at this point that is raised over network jitter. Network jitter is generally caused by queue build up, which can occur if the network is congested. In the CIN, this can cause an increase in latency.

To allow for that, the designers of DEPI/UEPI allowed the REQ and MAP traffic to be extracted from the data traffic and placed on a separate pseudowire that would allow a higher quality of service to be assigned to the REQ and MAP traffic.

Thus, in a properly configured network, even though the data traffic may see higher latency due to network jitter, the REQ and MAP traffic should not as they will not see that queue build up.

In summary, there is very little differences between and I-CMTS and R-PHY system. What differences that exist are quickly masked by the 6 ms latency of the DOCSIS upstream and corrected for with the DOCSIS pipelining.

Scenario 2: Beyond 100 Miles

So now it gets really interesting.

If DOCSIS has this scalable pipelining built into it that can address the MAC limitations, and the R-PHY architecture has

the same the PHY limitations, how high a latency could the REQ-GNT loop tolerate, and how far a distance could that distance be? What would that application be for?

One of the unique features of the R-PHY architecture is the ability to re-locate the CCAP-Core. An earlier example of how this feature can be used is hub site consolidation where the CCAP Core is placed at the head end and the R-PHY devices are placed at the hub or the node.

This allows for fewer CCAPs to be used in a system where there are many smaller hubs. The distance from the head end to the CM, though, may be larger than 100 miles.

Another emerging application for the CCAP Core is to virtualize it. When a CCAP Core is virtualized, then the software that was running on a dedicated hardware appliance such as the CCAP Core hardware is moved to a generic server environment in a data center. That data center may be located even farther way than the head end facility. These two deployment options are shown in Figure 14.

Note that a virtualized CCAP (vCCAP) system by definition must use a server complex for the forwarding plane. It is not sufficient to take a bunch of smaller CCAP devices and connect them together with network management to manage them as one large CCAP. That latter technique is defined by CableLabs as a CCAP Management Abstraction (CMA).

As the latency of the REQ-GNT loop gets excessive, so does the REQ-GNT delay. For example, for a CMTS to CM latency of 10 ms, the REQ-GNT delay could be 32 ms. (three times the delay plus 2 ms of the CMTS and CM). 10 ms is about 1250 miles (2000 km). As you will see in the simulation and test results, DOCSIS with no upstream scheduler modifications still works at these loop lengths and can saturate the upstream link.

However, it is possible to do better. Additional pipelining technology can be used by having a centralized scheduler and measuring the REQ-GNT loop,. Essentially, it becomes possible to proactively send grants to the CM.

This means that the CMTS can use predictive behavior and not wait for each REQ. The net result is that a 10 ms delay from CMTS to CM (2000 km) can result in only a 10 ms or so effective REQ-GNT time. This is incredible.

This means that at extremely long lengths, even in excess of 8000 km, it is possible to operate a centralized scheduler and not incur additional packet latency.

As reference, 8000 km is 1.7x the distance from San Francisco to New York City.

In the following simulation and test sections, we will look at the impact of proactive granting on performance.

Recap

- The upstream latency of a I-CMTS system is about 6 ms
- For the current DOCSIS operating limits of 0 to 100 miles, the additional latency introduced by R-PHY is almost insignificant compared to the existing DOCSIS latency.
- Pipelining techniques built into the DOCSIS protocol make DOCSIS highly tolerant of excessive latency.
- Making the REQ-GNT loop smaller than 100 miles by putting the scheduler

in the optical node may not result in any increase in performance

- R-PHY can actually operate with a central upstream scheduler at very large latency and loop lengths. Increases in REQ-GNT latency beyond normal one-way latency can be managed with advanced scheduler techniques.

PART 2: SIMULATION

To drive beyond theory, a simulation model was established for the DOCSIS US scheduler and simulations were run with typical R-PHY deployment scenarios to assess the impacts of the CIN delay on the latency of US traffic.

The model includes all the essential DOCSIS MAC function blocks of the system relating to the US traffic latency calculation.

Figure 15 is the high-level diagram showing all the function blocks included in the simulation model. A brief description is given for each function block as follows.

Simulation Function Blocks

Ingress Packets

The ingress packets at the network interface arrive randomly in time and sizes. The average throughput is calculated over time, and packet size and arrival interval are adjusted in such a way that the average throughput equals to the pre-configured one for the CM.

Queue Management

In the queue management block, all the ingress packets are time stamped for the purpose of the latency calculation. The packets in the queue are managed as a FIFO. In the simulation, max buffer size is set as 100 KB. Packets will be dropped if the buffer is full. This may occur in the cases where high or close to 100% CMTS loading is configured.

Request Management

The Request Management block manages all the requests: it checks the buffer length and sends the requests with proper requested sizes. It time stamps all the requests that are sent out. Whenever an updated MAP ACK time arrives, the Request management updates the outstanding requests accordingly to ensure no duplicated requests are sent for same packets.

Request Reception

Requests are processed in the request reception block. The CMTS MAP capacity (the total bytes that can be carried in each MAP 2ms interval) is checked, and if the request can be accommodated, the request

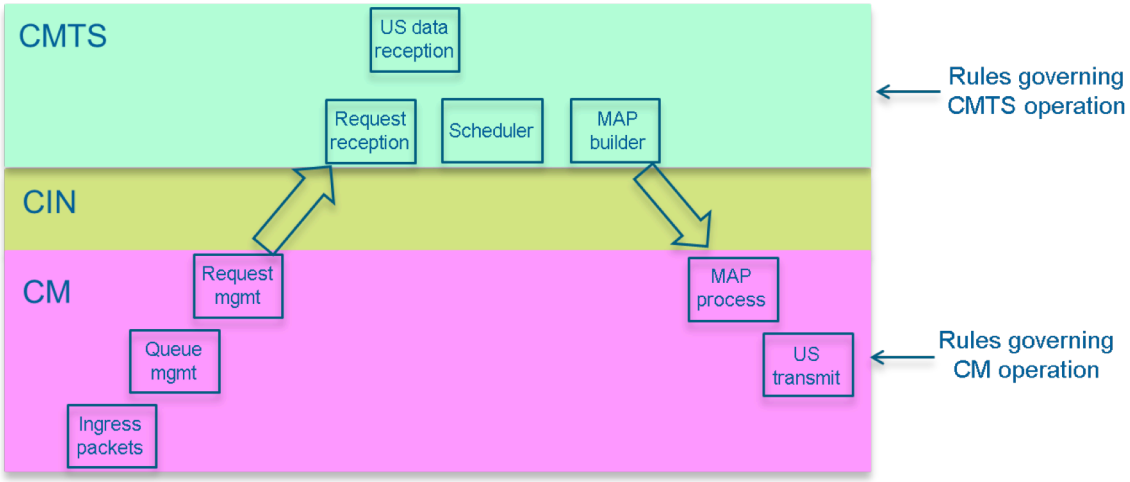


Figure 15 – Simulation Model and Function Blocks

will be granted. The actual grant time will have at least 1 MAP delay to simulate the process delay at the CMTS.

The granting could be partial if there is not enough capacity to fully accommodate the request. Moreover the grant could be a few MAP ahead if the next MAP runs out of capacity.

US Data Reception

All the normal data packets are processed in the Data Reception block. The US latency is computed for each packet by comparing the packet's ingress time stamps and the packet's arrival time at the CMTS.

The US latency is defined as the time difference between the packet ingress time at the network interface of the CM and the packet arrival time at CMTS.

Scheduler

The US scheduling algorithm is running in the scheduler block. For each request received, the scheduler computes the size and the time of a grant if a grant is warranted. The grant is based on both the requested sizes and the MAP capacities. Two scheduler algorithms are simulated:

1. Baseline algorithm

   In this algorithm, a grant is given only when a request is received. This request then grant literately follows DOCSIS MAC US scheduling protocol. In this request-then-grant process, the US latency consists of three parts:

   - the propagation delay of the request traversing the network from CM to CMTS,
   - the delay of the MAP building and the MAP traversing the network from CMTS to CM, and
   - the packet propagation delay from CM to CMTS.

2. Proactive grant algorithm:

   A proactive grant algorithm reduces the latency by activating mechanisms that help establish a pipeline of requests and grants. Depending on the actual mechanisms implemented, the proactive grant scheduling may result in extra BW requests from CMs, or extra grants to CMs, or a combination of both.

   The extra requests that are sent by a CM in addition to the normal requests are denoted as proactive-requests. With the proactive grant algorithm, a CMTS can give out grants to CMs based on received requests, received proactive-requests, or without receiving any requests.

   The grants that are given out by CMTS to CMs in the first case are denoted as reactive-grants; the extra grants that are given out by CMTS to CMs in the two latter cases are denoted as proactive-grants.

The logic behind the proactive grant algorithm is that these proactive grants will allow CMs to bypass the conventional steps of the request-grant mechanism, and send packets right away without going through the normal three stage process (request-grant(MAP)-send), so that US latency can be reduced.

To improve the utilization of the proactive grants, the following rules are observed when the CMTS gives proactive grants:

- Provide more proactive grants to CMs with higher US traffic
- The sizes and intervals of proactive grants are dynamically adjusted based on CM heuristic performance metrics (US utilization ratio, latency, etc.)
- Proactive grants and reactive grants (grants resulting from normal requests) are managed holistically, based on the channel utilization and the total bandwidth (proactive and reactive) assigned to a given CM.

MAP Builder

The MAP builder puts all the grants in place and adds the ACK time. The designated transmit time, which includes the MAP advance Time is added for each grant.

MAP advance time is computed as follows:

- MAP_advance_time = cm_t+cmts_t+CIN_t+ds_phy_t+hfc_t+ margin

Where

- cm_t is the process time required by CM
- cmts_t is the process time (MAP building) required by CMTS
- CIN_t is the propagation delay of CIN (one way)
- ds_phy_t is the processing time required by R-PHY
- hfc_t is the round trip delay of the HFC network
- margin is an extra safety margin

The purpose of the MAP advance time is to allow the CM to receive the MAP ahead of time and have enough time to prepare the US transmission at the designated time. The MAP advance time includes the time budget to allow for the HFC US transmission to arrive at the R-PHY in time.

Unless specified otherwise, all the simulation results are obtained with the following values:

- cm_t=1 ms;
- cmts_t=1 ms;
- ds_phy_t=2ms;
- hfc_t=1 ms
- margin=1ms

In the simulation, we also set us_phy_t = 2 ms, where us_phy_t is the time required for PHY to process US traffics. us_phy_t is required to compute packet US latency.

CM MAP Process

There are three main functions in the MAP Process block:

a) Check the ACK Time in each MAP received and update the outstanding requests accordingly;

b) If there is a grant in the MAP for a CM, the CM prepares the data to transmit at the designated time;

c) Remove the packets from the data queue if they are transmitted.

Simulation Results

The simulation is an event driven process, and runs with an emulated time clock. The time clock runs with a 0.2 ms time increment (it can be set to other time increments, 0.2 ms gives a sufficiently fine time granularity with a reasonable simulation speed).

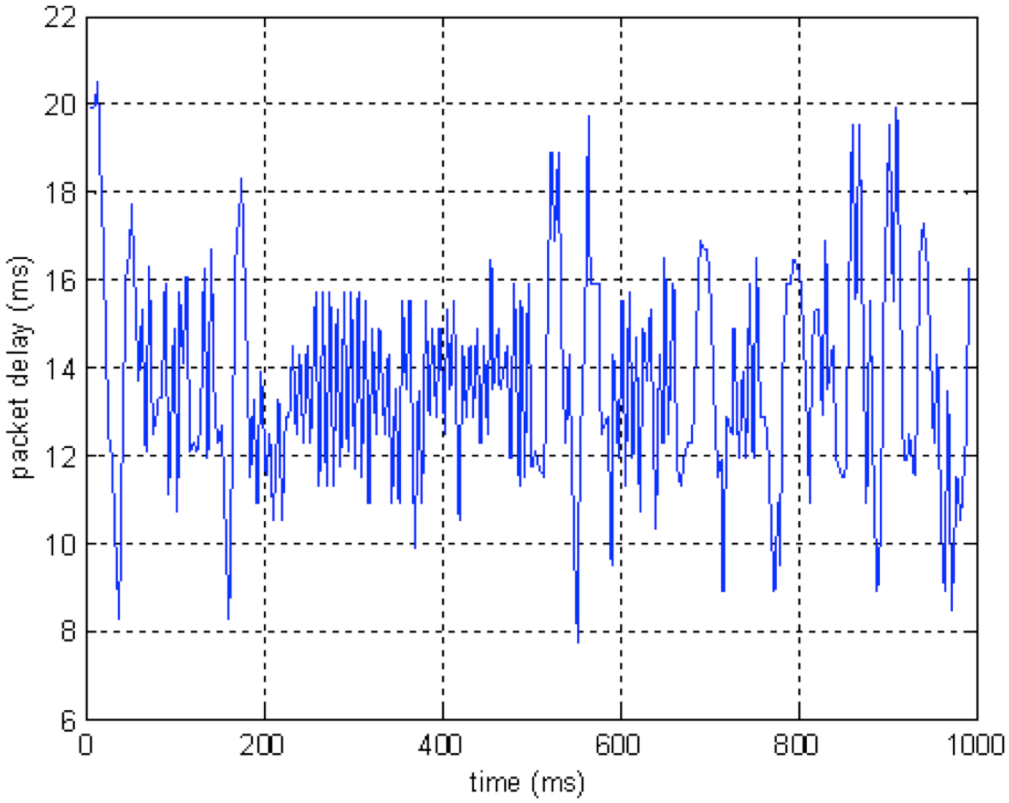


Figure 16 – CM Packet Delay, REQ-GNT scheduler, 2 ms CIN, 30 CM, 90% Loading

All the MAC function blocks execute at each time increment. The simulation runs for a period of one second. However, the US latencies are computed only with packets which arrived at the CMTS after 500 ms to avoid any data abnormalities in the beginning of the simulation that may be due to the simulation initial setup.

A total of 30 CMs are included in the simulation. All the CMs are set to have the same throughput and priority. In the case when the max CMTS capacity is reached (for example, for 90% loading case), the CMTS will throttle the throughput equally for each and every CMs. The bandwidth throttle is done using a first-in-first-serve algorithm, and the sequence of the CMs is selected randomly.

Algorithm 1: Baseline algorithm

There are two scheduler algorithms investigated in the simulation:

1. the baseline algorithm and
2. the proactive grant algorithm.

The simulation results for the baseline algorithm will be presented in this section to establish a baseline for performance (latency) comparisons.

The simulations are first run for the case with 2 ms CIN delay. 2 ms CIN delay corresponds to 400 km in distance and is a good representative value for the majority of deployment scenarios.

The total number of CMs = 30, with each CM having a throughput = 3.64 Mb/s.

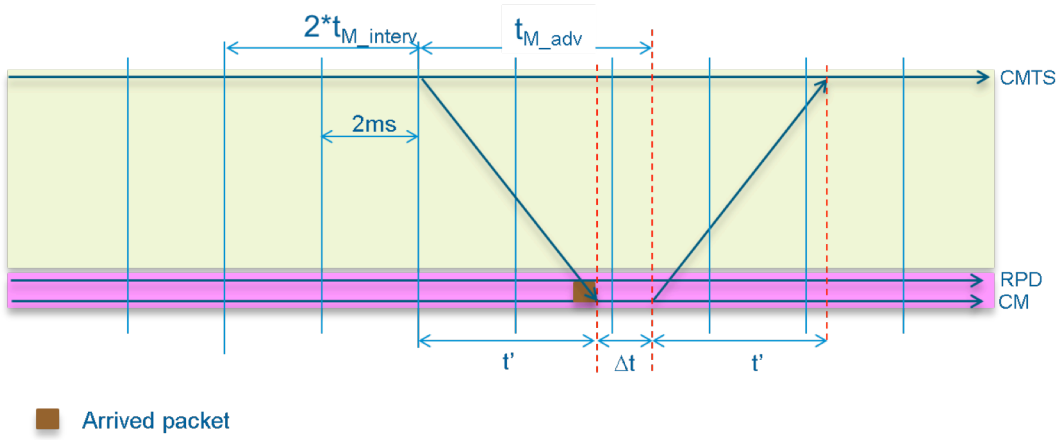
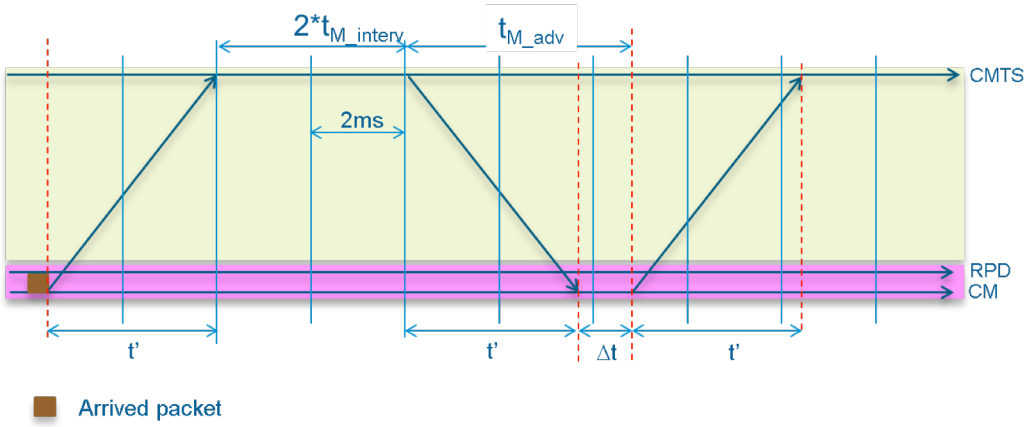


Figure 17 – The minimal US latency with a pipeline of request and grant

CMTS loading is set to 90%. The simulation results on US packet delay are given in Figure 16.

It is evident from the simulation results in Figure 16 that the minimal delay is 8 ms, and the max delay is about 20 ms. Minimal delay is achieved when the packets arrive immediately before a MAP is received, and the CM is able to send those packets using the grant contained in the MAP. In this case, the US latency is just the MAP advance time, which is illustrated in Figure 17. With 2ms CIN delay, the MAP advance time is 8ms

The maximum latency occurs when the packets go through the normal request-then-grant process and their request arrives at the CMTS right on the MAP time grid (actually just a bit after the MAP grid), causing the grant time postponed by 2 MAP intervals. The initial version of the simulation used does not take into account collisions, which would result in the loss of request packets. Thus the maximum delays will not include retransmission times.



where $t_{M_interval}$ is the Map interval(2ms), t_{M_adv} is the Map advance time, t' is the DS or US delay (assume symmetric). $\Delta t = t_{M_adv} - t'$.

Maximal US latency = $t' + 2*t_{M_interv} + t' + \Delta t + t' = 2*t_{M_interv} + t_{M_adv} + 2t'$

Figure 18 – The max US latency in a normal request and grant process

This is illustrated in Figure 18. With the system parameters values configured in the simulation and 2 ms CIN delay, the max US latency is 20 ms.

By averaging the delay over time and over all CMs, one obtains 13.5 ms mean US latency. 13.5 ms is close to the average of the min and max latency ((8+20)/2=14), which means the packets almost have the equal opportunity to be allocated right after it arrives at CM, or have to go through the normal request and grant process with the max latency, or anything in between.

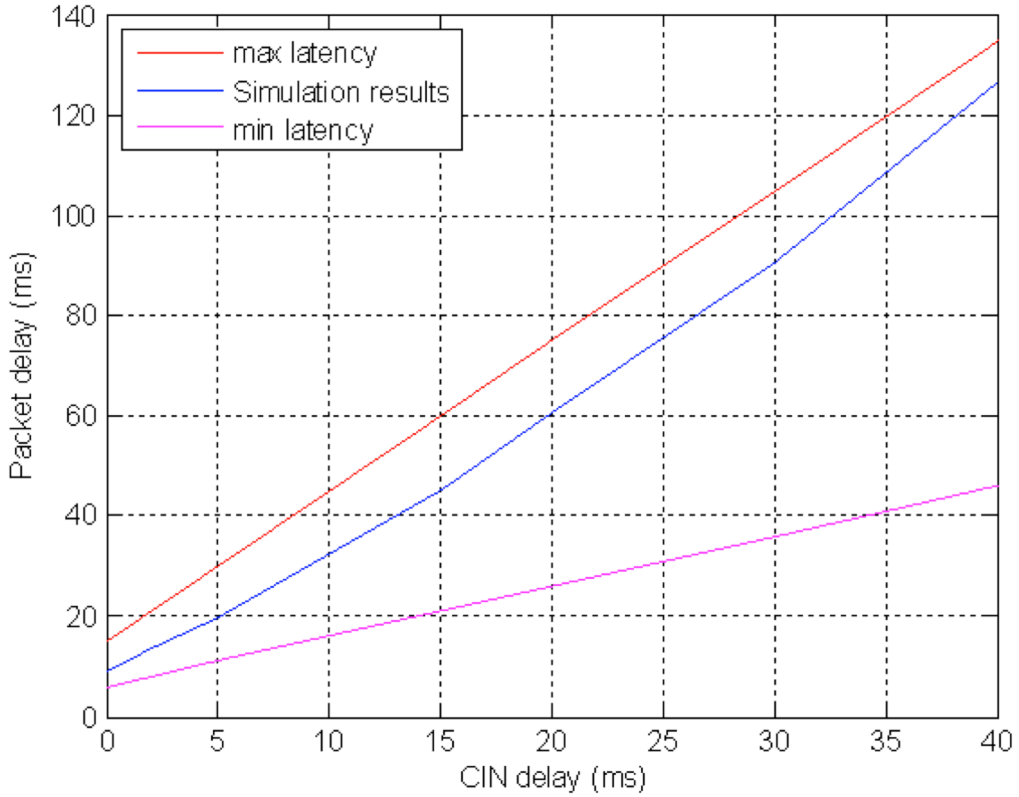


Figure 19 – US Latency vs CIN Delay, REQ-GNT Scheduler, 30 CM, 90% Loading

Simulations were run with other CIN delays. Figure 19 shows the US latency as a function of the CIN delay (CIN_t) for 30 CMs and 90% network loading. The min and max latencies are also plotted on the same figure. The min and max latencies are defined in Figure 17and Figure 18.

Here are the observations:

1. The simulation results falls between the theoretic max and min latencies for all the CIN delays.
2. The simulation results are closer to the max latencies, indicating that, with the baseline request then grant scheduler algorithm, majority of the packets go through the three stage request-grant-send process, and incur the max latency. This is particular obvious with high CIN delay.
3. Some of the packets go through the pipeline of request and grant, incurring less latency than the normal request then grant process. This is obvious because the simulation results are 5~10ms less than the max latency curve.
4. The latency in general follows the max latency curve with a fixed offset

   US latency = fixed offset + Max_delay = Fixed_offset+ $2*t_{M_interv} + t_{M_adv} + 2t'$

   Where t_{M_interv} is the map interval, and t_{M_adv} is the Map advance time, t' is the DS delay.

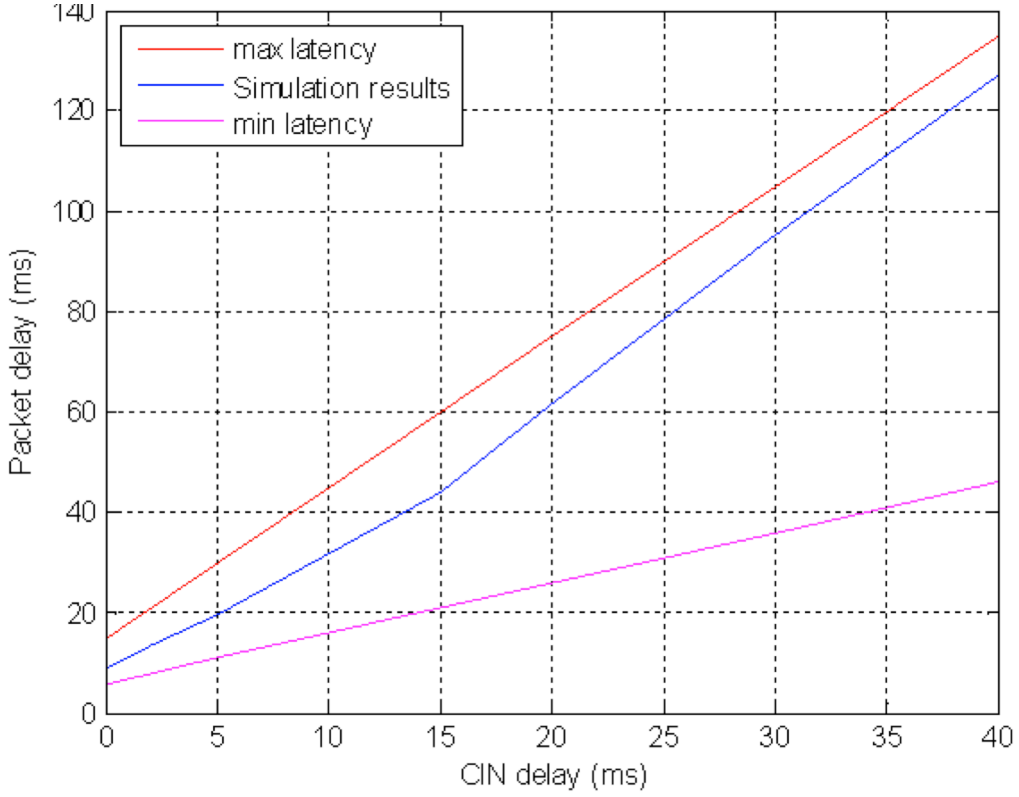


Figure 20 – US Latency vs CIN Delay, REQ-GNT Scheduler, 30 CM, 50% Loading

With the configured system parameters, we have then:

- US latency
  = fixed offset + 4 cm_t + cmts_t
  + ds_phy_t + hfc_t + margin
  + CIN_delay + CIN_delay
  + ds_phy_t + hfc_t/2
  = fixed_offset' + 3*CIN_delay
  ~= 2 + 3*CIN_delay

Three times the CIN_delay is expected, as in the normal request-grant-send three stage process, the packets needs to traverse the CIN three times, incurring 3X CIN delay. The fixed offset counts for other delay (MAP process time, and extra dwelling time in the queue, etc.)

This 3*CIN delay is also observed with different network loading. Figure 20 shows the US latency as the function of the CIN delay (CIN_t) for the case 30 CMs and 50% network loading. It is obvious that all the observations described above for 90% network loading are also valid here. The latency follows the same 3*CIN delay curve.

Algorithm 2: Proactive Grant Algorithm

In this algorithm, the scheduler proactively generates grants to CMs based on received pre-requests or even when no requests are received. Grants that do not result from normal requests are called proactive grants. The proactive granting algorithm needs to follow certain rules to optimize the overall performance (latency, BW utilization, etc.).

The basic idea of the proactive grant is to give extra grants to CMs to increase their

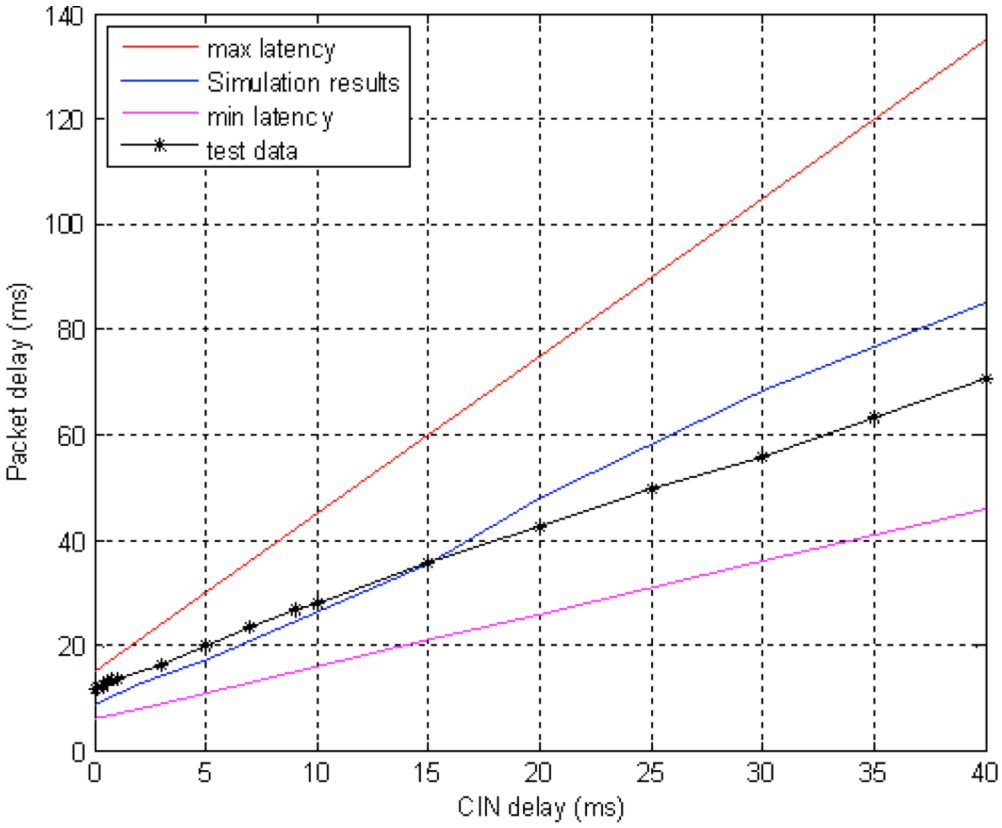


Figure 21 – US Latency vs CIN Delay, Proactive GNT Scheduler, 30 CM, 90% Loading

opportunities to send packets without going through the three stage process, thus the latency is reduced. It is critical to have the right sizes and intervals for the proactive grants so an effective pipeline of request and grant can be established and, at the same time, a high BW utilization is achieved.

Too many proactive grants will definitely help reduce the latency, but may cause some of the proactive grants to be un-used or partially used, leading to low BW utilization. So, the sizes and intervals of the proactive grants need be dynamically adjusted with heuristics based on the CM performance metrics and traffic pattern.

Figure 21 shows the simulation results computed with the proactive grant algorithm. The number of CMs included in the simulation is 30, and the network loading is 90%. All other system parameters are the same as before. By comparing with Figure 19 (the request then grant algorithm), it is evident that the proactive grant reduces the US latency significantly.

The mean delay curves lies roughly mid-way between the min and max latency curves, indicating that the packets almost uniformly fall between the two extreme cases of 'send out right away' and 'go through three stage process'.

The simulations show that the pipeline of the request and grant process is established for a considerable number of

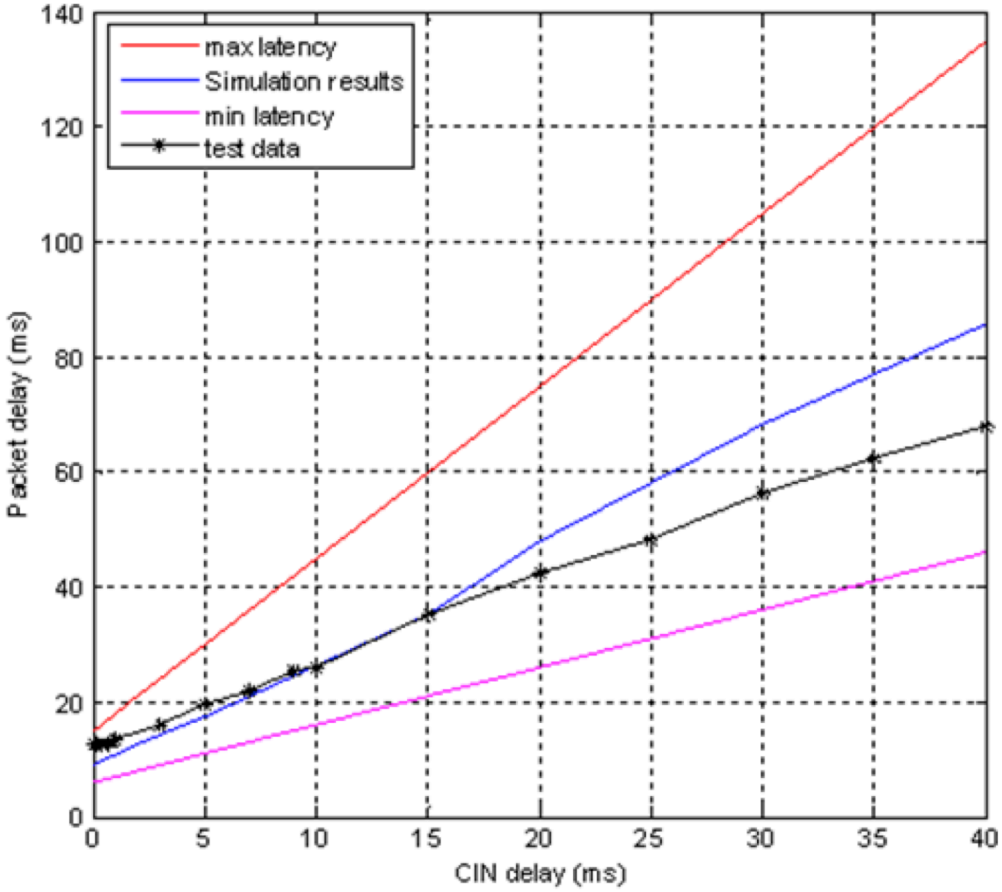


Figure 22 – US Latency vs CIN Delay, Proactive GNT Scheduler, 30 CM, 50% Loading

packets. This results from the extra grants from the proactive grant mechanism.

The test data with the same system configurations are also shown in Figure 21. It is evident that the simulation results agree well with the test data, particularly in the range of 0 ms to 20 ms CIN delay. This range covers the majority of the deployment scenarios, if not all.

Similar performances and agreements are observed with other simulation cases. Figure 22 shows the simulation results and test data for 50% network loading. Both the simulation results and test data show that the proactive grant significantly reduce the US latency.

By examining the simulation results and test data, one can derive the following relationship between the latency and CIN delay.

- US_latency
  = Fixed Delay Offet + 2*CIN_delay
  where the fixed offset is about 9 ms.
- Recall that the US latency follows in general 3*CIN delay curve with the request-then-grant process, the proactive grant scheduler algorithm eliminates 1*CIN delay from the US latency.

In an ideal case where all the requests and grants go through a pipeline, this could result in an US latency of 1*CIN delay (the min delay illustrated in Figure 17). The 1*CIN delay US latency is the theoretical low bound, one could approach to this low bound by designing better scheduler algorithms that fully exploit CM traffic patterns and heuristic performance metrics.

Needless to say, the scheduler algorithm needs to balance out the complexity of the algorithm, BW utilization and actual application QoS requirements.

ReCap

- The max US latency has a slope of three times the CIN delay, and the min US latency has a slope of one times the CIN delay
- The conventional request-then-grant scheduler follows the max US latency curve with a fixed offset.
- The proactive scheduler helps establish a pipeline of requests and grants through proactive granting and can significantly reduces US latency.
- Simulation results show that at least one equivalent CIN delay can be eliminated from the US latency by employing proactive scheduling algorithms.

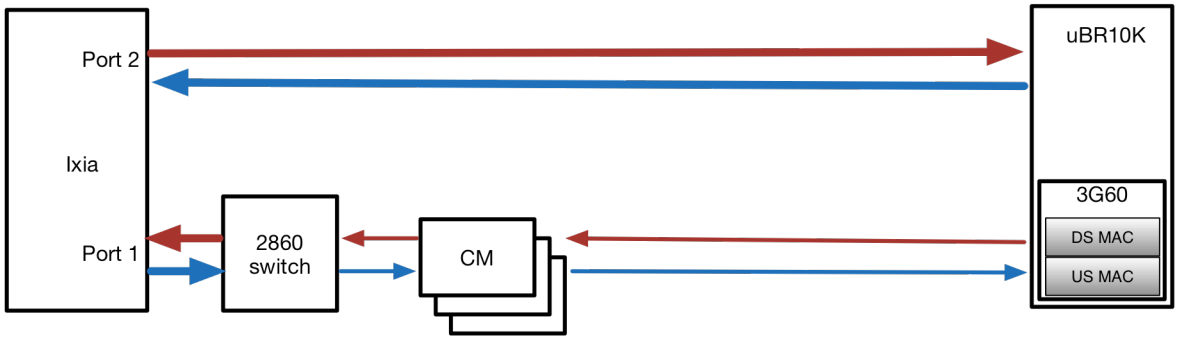


Figure 23 – Test Set Up I-CMTS

PART 3: TEST & MEASUREMENT

Following the simulation experiments we were able to obtain the equipment to generate some experimental data, which could verify (or otherwise) the previous results. To facilitate this a number of tests were run to measure the impact of increasing distance between the cable modem and the CMTS Core.

Test Set Up and Components

The test set up used was as shown in Figure 23 and Figure 24.

CMC

The Cable Media Converter (CMC) used for the testing is a C-DOCSIS (China DOCSIS) type 3 – remote PHY device.

A C-DOCSIS type 3 device such as the CMC contains the DOCSIS PHY with the CMTS Core system providing the higher layer functions including the DOCSIS MAC and all packet classification, QoS and forwarding.

The CMC is connected to the CMTS Core using an IP connection operating over a Gigabit Ethernet link with DEPI / L2TPv3 tunnels to encapsulate the DOCSIS traffic. This architecture is functionally equivalent to a Remote PHY architecture from a DOCSIS 3.0 perspective. It is the first generation of Remote PHY and is a pre-standards version

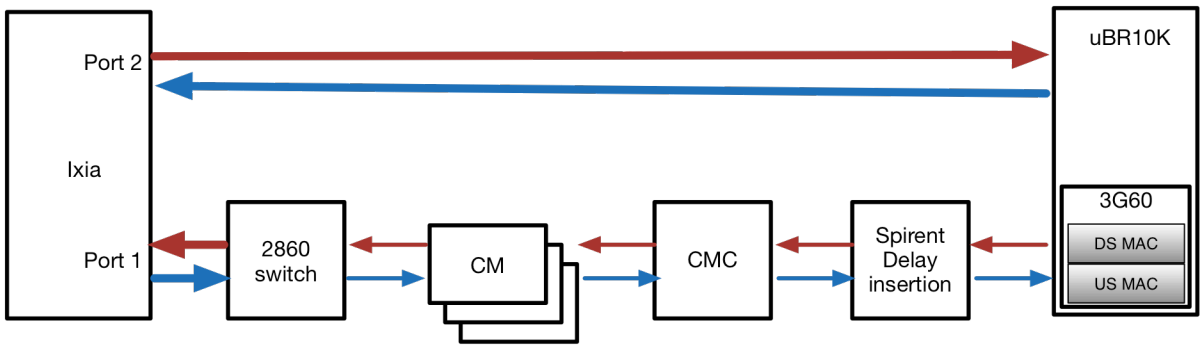


Figure 24 – Test Set Up R-PHY

of the work being done at CableLabs.

This test setup only supports DOCSIS 3.0 and does not support video or DOCSIS 3.1). It does however provide a convenient (and available) platform to investigate the impact of network latency.

CMTS Core

The CMTS core used was a Cisco uBR10012 with an MC3GX60V line card running the CMC controller and DOCSIS MAC software. The physical interface towards the MC was a Gigabit Ethernet. The same CMTS was used for both the Remote PHY and Integrated CMTS tests.

CIN Delay Insertion

A Spirent Aperto test platform was used to inject latency and jitter into the link between the CMTS core and the CMC providing a controlled environment, which was simple to modify during the experiments. Delay can be injected independently to the upstream and downstream using the test equipment but for the data set described in the paper upstream and downstream CIN delays were symmetric

Ixia Test Generator

An Ixia test generator with 10 Gigabit interfaces was used to generate traffic and measure latency. A consistent traffic profile, described in the section Traffic Profile was used throughout the testing.

Cable Modems

For each test there were a total of 30 DOCSIS 3.0 CMs online.

Ethernet Switch

A 2860 Ethernet switch was used to convert between the 10G port on the IXIA and the 1G Ethernet ports on the cable modems. The load on the switch was low enough that it was operating in a non-congested mode. The switch will add some minor amount of latency but is operating at a small fraction of capacity.

HFC

The test bed was a laboratory system with all the components co-located and used a minimal HFC plant with very short cable lengths. Thus transmission delays in the fibers and coaxial cables were essentially zero.

Traffic Profile

The following traffic profile was used for all of the testing (including the baseline):

- Downstream capacity was 16 channels modulated at 256 QAM.
- The downstream traffic was equally divided between each of the 30 DOCSIS 3.0 modems.
- A number of downstream load scenarios were tested varying between 10% and 90% of downstream capacity.
- Upstream capacity was 4 channels of 6.4MHz modulated at 64 QAM.
- The upstream traffic was equally divided between each of the 30 DOCSIS 3.0 modems.
- A number of upstream load scenarios were tested varying between 10% and 90% of upstream capacity.
- The number of frames transmitted in the downstream during each test run varied between 10,000 and 90,000 depending on load.
- The numbers of upstream frames transmitted during each test run varied between 2,100 and 18,900 depending on load.

- All traffic was best effort, 1500 byte packets

Measurements

Referring to Figure 23, upstream latency measurements were made between port 1 and port 2 of the Ixia tester. Downstream latency measurements were made between port 2 and port 1.

For each test the following results were recorded

- Total frames sent, received and lost for each modem
- Minimum, maximum and average latency for each modem.

Note that the measurements do not include the 95% of 99% population points. The maximum number reported is the 100% population point that is at the tail end of the Poisson distribution.

To keep the amount of data contained in this white paper, the reported results are the average seen by each of the 30 CMs. So, the minimum in this paper is actually an average of 30 minimum test results. The average latency reported is actually an average of 30 average latencies. Each of those averages are actually averages of multiple test runs.

So the average results are actually averages of averages.

Latency Added by Test Bed

Any added latency from the test components will be constant as the network delay is increased. As such it is accounted for in the baseline delay measurement.

CIN Delays Tested

The CIN delays used ranged from 0 to 40 ms of delay, representing networks in the range of 0 to 8000 km. The DOCSIS specification allows for delays of up to 0.8 ms between the CMTS and CM so that CIN delays up to 1 ms represent networks roughly comparable to potential I-CMTS deployments.

Non-ideal Testing Conditions

10% versus 90% Loading

Tests were run with a packet loading of 10% to 90%. The most accurate results of system are the results at 90% loading.

This is counter-intuitive because a higher system loading should result in more buffering and more latency.

However, it is the opinion of the authors that the test setup does not exactly model real world traffic. In the test setup, packets are released at an even rate. In the real world, low traffic on the network is when there are fewer CMs generating traffic less often. However, when a single CM generates the traffic, it occurs in a burst.

The net difference is that in the real world, a CM will begin using the DOCSIS piggyback grants even on a lightly loaded system. In this test set up with evenly spaced packets, piggybacking probably did not kick in until the loading was quite high.

	I-CMTS			R-PHY 0 ms			Delta		
Load	**Min**	**Avg**	**Max**	**Min**	**Avg**	**Max**	**Min**	**Avg**	**Max**
10	2.1	2.4	2.7	2.1	2.5	2.8	0.0	0.0	0.0
25	2.0	2.4	12.0	2.1	2.5	2.8	0.1	0.0	-9.2
50	1.8	2.3	12.1	1.8	2.3	2.8	0.0	0.0	-9.3
75	1.7	2.0	2.6	1.8	2.0	2.6	0.1	0.0	0.0
90	1.7	2.0	2.7	1.7	2.0	2.8	0.1	0.0	0.1

Table 1 – I-CMTS and R-PHY Downstream Latency

To illustrate this point, the log file reported the following results:

- In case of 10% usage, there are 2211 BW_REQ, in which, 50 are piggyback request (2%);
- In case of 90% usage, there are total 12205 BW_REQ, in which, 7792 are piggyback request (64%).

1 GE versus 10 GE

Future generations of CMTS Cores and RPDs will use a 10 GE interface. This test setup used a 1 GE interface. Thus, this particular test setup will contain transmission latency that will not exist in a product system.

IEEE 1588 based R-DTI Timing

This test system did not have the benefit of 1588 timing. Thus, extra engineering margin (MAP Advance Time) was used to account for the CIN delay.

With an R-DTI system, one should expect even better performance.

It is worth noting, however, that a R-PHY system could be made to work quite well without R-DTI.

Baseline Results for Latency

To create a baseline, two sets of tests where done.

1. The first was on an Integrated CMTS with 30 CMs.
2. The second was on the same CMTS with an R-PHY Device connected with 0 ms of CIN delay and the same 30 CMs.

The CMTS was the same CMTS in both cases and the software was the same. The line card that drove the R-PHY Device was a modified integrated CMTS line card with the PHY circuitry removed.

In theory, both these systems are similar since the components are the same – they are just packaged differently.

I-CMTS and R-PHY Downstream

The downstream latency for the two reference test setups for various system loadings is shown in Table 1. The differences between the two configurations are also shown.

In comparing the minimum and average downstream latency of an I-CMTS with a R-PHY system with 0 ms CIN delay, there are no appreciable differences.

	I-CMTS			R-PHY 0 ms			Delta		
Load	**Min**	**Avg**	**Max**	**Min**	**Avg**	**Max**	**Min**	**Avg**	**Max**
10	4.8	10.6	43	4.3	13.5	46	-0.5	2.9	3.1
25	4.7	9.5	45	3.4	11.8	44	-1.3	2.3	-0.6
50	4.6	10.4	50	1.7	12.5	48	-2.9	2.1	-1.5
75	4.7	9.8	51	1.7	11.5	50	-3.0	1.6	-0.8
90	4.6	11.3	63	1.7	11.8	57	-2.9	0.5	-5.5

Table 2 – I-CMTS and R-PHY Upstream Latency

The max value is a Poisson tail value. In this test case, it seems that queuing delays on the I-CMTS for two of the loading cases where higher with the I-CMTS than the R-PHY. These are likely test/queuing anomalies and not significant.

I-CMTS and R-PHY Upstream

The upstream latency for the two reference test setups for various system loadings is shown in Table 2. The differences between the two configurations are also shown.

The minimum latency for the I-CMTS upstream is in alignment with the theoretical latency of 4-6 ms.

The minimum latency for the R-PHY system as measured is better than the I-CMTS. This should not actually be the case and is the result of a scheduler enhancement. This is explained in section "Why the Measured Results were Better than Theory"

The average results for both systems show latencies in the 10 to 12 ms range. This is surprising at first, but is a result of buffering in a real world system.

The practical measurement point between the I-CMTS and R-PHY system with 0 ms of CIN delay would be the 90% data point of 0.5 ms. (refer to the Section "10% versus 90% Loading")

The maximum system delays that is a result of contention requesting is surprising high at 50 to 60 ms. It is surprisingly similar in the two systems as well. This can be attributed to the request backoff algorithms for both systems resolving themselves in a similar way.

It is worth noting that these high latencies are natural latencies for DOCSIS and illustrates how well DOCSIS works at these latencies.

CIN Delay and Downstream Latency

Having established the baseline data for the test bed with zero network delay the same tests were run for a range of network delays.

The average, latency for each CM was measured and these values used to calculate the average latency across the CM population as CIN delay was increased.

Average Downstream Latency

CIN delay (ms)	Average DS Delay		
	10% load	50% load	90% load
0	2.46	2.26	1.98
0.2	2.66	2.46	2.18
0.4	2.86	2.66	2.38
0.6	3.06	2.86	2.58
0.8	3.26	3.06	2.78
1	3.46	3.26	2.98
3	5.46	5.26	4.98
5	7.46	7.26	6.98
7	9.46	9.27	8.98
9	11.46	11.26	10.98
10	12.46	12.27	11.98
15	17.46	17.26	16.98
20	22.46	22.27	21.98
25	27.46	27.27	26.99
30	32.46	32.27	31.99
35	37.46	37.27	36.99
40	42.46	42.27	41.99

Table 3 – Downstream Latency with CIN Delay

Table 3 and Figure 25 show the downstream latency measured as the CIN delay was increased for the three load factors.

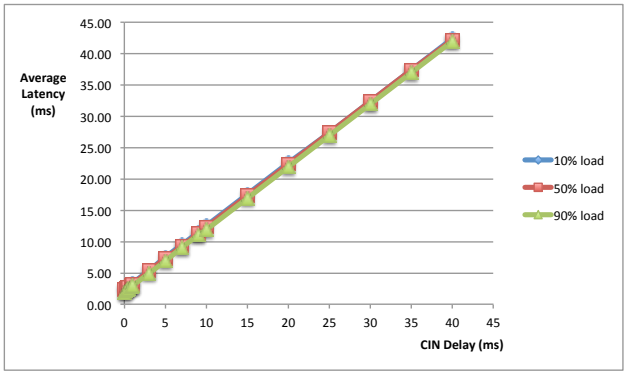


Figure 25 – Average Downstream Latency vs CIN Delay

The measured average latency was seen to be a small, almost constant (~2ms) addition to the imposed CIN delay. In effect it was the baseline delay + the CIN delay.

The downstream performance was seen to be very tolerant of CIN delays remaining linear up to 40ms of imposed CIN delay.

CIN Delay and Upstream Latency

Average Upstream Latency

CIN delay (ms)	Average US Delay		
	10% load	50% load	90% load
0	13.50	12.55	11.82
0.2	13.42	12.55	11.96
0.4	14.42	12.36	12.42
0.6	14.40	12.92	13.24
0.8	15.51	13.18	13.59
1	15.36	13.44	13.48
3	20.93	15.95	16.29
5	26.81	19.47	19.91
7	29.30	21.75	23.38
9	30.48	25.51	26.82
10	30.92	26.09	27.84
15	41.04	35.08	35.66
20	41.20	42.32	42.57
25	54.22	48.12	49.69
30	58.09	56.28	55.69
35	67.09	62.42	63.19
40	69.92	68.00	70.64

Table 4 – Upstream Latency with CIN Delay

Table 4 shows the upstream latency measured as the CIN delay was increased for each load factor. From 0 to 1 ms of CIN delay represents CM to CMTS distances approximately within the DOCSIS specification (0.8 ms is the DOCSIS limit). In this region, the R-PHY system exhibited latencies in the 11 to 15 ms range. For CIN distances beyond the DOCSIS limit the latencies varied from 16 to 70 ms.

Figure 26 shows the average upstream latency plotted against the one way CIN delay for CIN distances in the range of 0 to 200 km. The results show that for each way CIN delays in this range the behavior is relatively linear with a modest increase from the 12 ms baseline to between 13 and 15 ms at the 200 km (1 ms) distance.

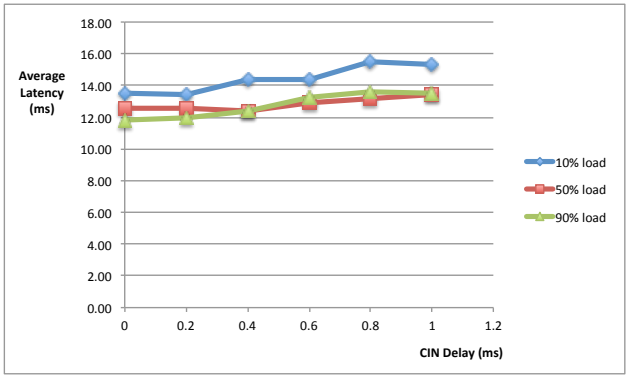


Figure 26 – Average Upstream Latency vs CIN Delay (0 to 200 km)

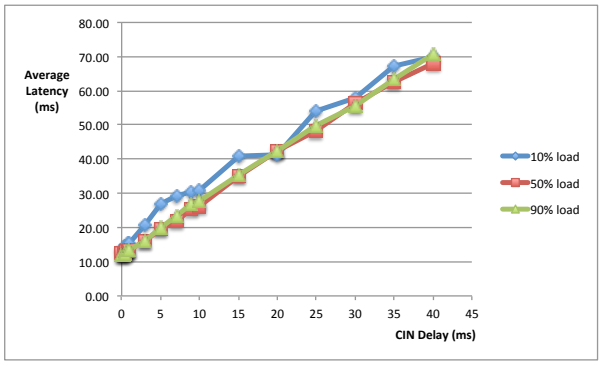


Figure 27 – Average Upstream Latency 0 to 8000 km

Figure 27 shows the latency measured for the entire range of tests. This extended the CIN latency to 40ms representing an 8000 km (each way) network. As seen from the graph the latency increases linearly with the CIN delay. It is interesting to note that the 10% load measurements show higher latencies than the 50% and 90% load at these CIN delays. This is due to the higher level of piggybacking at the higher utilization rate.

Minimum Upstream Latency

The minimum latency was measured at the three load factors and the average across the CM population calculated and shown in Table 5 and Figure 28.

Columns 2, 3 and 4 of Table 5 show the average minimum latency measured for the three load factors. Columns 5, 6 and 7 show the measured latency minus the CIN delay.

CIN delay (ms)	Average Min US delay			Average Min US delay - CIN delay		
	10% load	50% load	90% load	10% load	50% load	90% load
0	4.28	1.73	1.69	4.28	1.73	1.69
0.2	5.29	2.08	1.88	5.09	1.88	1.68
0.4	4.22	2.15	2.08	3.82	1.75	1.68
0.6	6.12	2.30	2.27	5.52	1.70	1.67
0.8	5.06	2.49	2.47	4.26	1.69	1.67
1	5.49	2.69	2.67	4.49	1.69	1.67
3	4.72	4.68	4.67	1.72	1.68	1.67
5	6.74	6.66	6.66	1.74	1.66	1.66
7	8.67	8.66	8.65	1.67	1.66	1.65
9	10.74	10.68	10.64	1.74	1.68	1.64
10	11.71	11.67	11.65	1.71	1.67	1.65
15	16.67	16.66	16.66	1.67	1.66	1.66
20	21.72	21.66	21.66	1.72	1.66	1.66
25	26.73	26.67	26.64	1.73	1.67	1.64
30	31.69	31.67	31.65	1.69	1.67	1.65
35	36.67	36.66	36.66	1.67	1.66	1.66
40	41.73	41.67	41.65	1.73	1.67	1.65

Table 5 – Minimum Upstream Latency vs CIN Delay

For the higher load factors, this is essentially a constant of approximately 1.7ms on top of the CIN delay.

For the 10% loading there is a distinct difference between CIN delays of less than 3 ms and higher delays. Above 3 ms we again see a roughly constant adder of approximately 1.75 ms. Below 3 ms the delta is much higher in the 4 to 5 ms region. Further study is needed to investigate this area.

The measured difference in US latency for a 0 mile plant to a 100 mile plant is 2.47 – 1.69 = 0.78 ms. This matches the 0.80 ms of CIN latency.

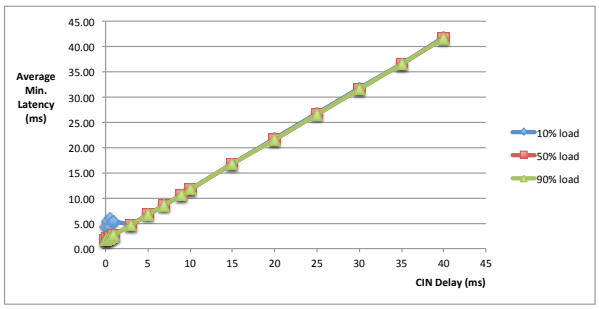


Figure 28 – Average Minimum Upstream latency vs CIN Delay

As can be seen from Figure 28, the minimum latency seen shows very little variation with load.

With the proactive granting strategy described previously the minimum latency occurs when a frame arrives at a CM and is sent "immediately" using an available proactive granted transmission opportunity. Thus the 1.7ms "adder" to the CIN delay represents the processing overhead.

Maximum Upstream Latency

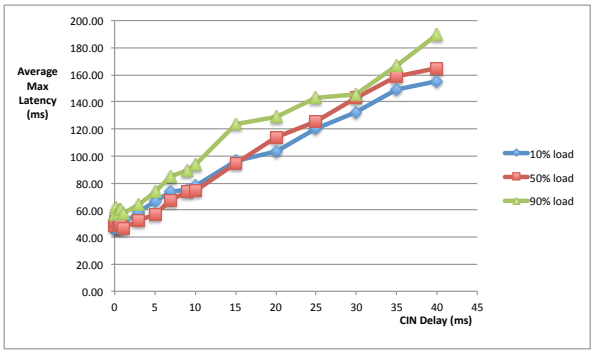


Figure 29 – Average and Maximum US Latency

CIN delay (ms)	Average Max US Delay		
	10% load	50% load	90% load
0	46.30	48.46	57.22
0.2	45.84	51.57	61.67
0.4	47.53	48.91	58.05
0.6	47.43	49.26	59.97
0.8	48.04	49.64	59.92
1	49.53	46.00	57.39
3	58.21	52.83	63.47
5	66.39	56.70	73.46
7	73.11	67.37	85.22
9	75.30	73.44	88.89
10	77.54	73.95	93.46
15	96.02	94.77	123.55
20	103.32	113.77	128.71
25	119.89	125.05	142.94
30	132.63	143.08	145.52
35	148.69	158.69	166.92
40	154.77	165.20	189.63

Table 6 – Average Maximum Upstream Latencies

The maximum latency seen across the CM population is shown in Table 6 and

Figure 29. The maximum latency is high even with a CIN delay of 0. This would be expected from the latency distribution seen in Figure 30 as the maximum latency is measured for those packets at the extreme edge of the Poisson distribution.

These values are included for completeness of analysis but are not statistically significant since they are at the tail end of the distribution.

The impact of CIN delay on maximum latency appears to be more significant than on average or minimum latency and to increase at approximately 3x the CIN delay. This may reflect the scenario in which no proactive grant is available for a packet so that it is subject to the 3x CIN delay (request, grant, data) discussed in the theory section. A more detailed analysis of this is planned as a future work item.

Upstream Latency Distribution for a Single CM

The upstream latency distribution was also measured for a single CM with a CIN delay of 0.6 ms (which is well within the limits of the DOCSIS protocol) and the results shown in Figure 30.

The distribution shows packet latencies varying between 5 and 45 ms even within the DOCSIS operating region. The DOCSIS protocol is clearly very tolerant of varying latencies as the CM was operating perfectly normally when the measurements were taken.

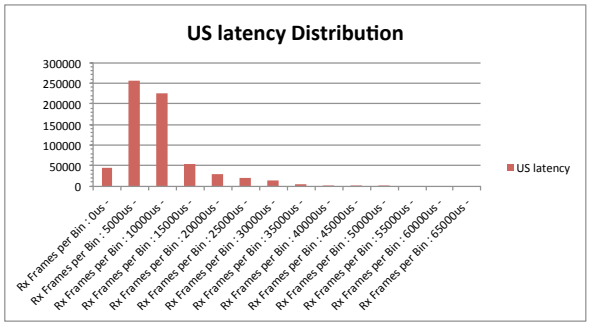


Figure 30 – US Latency Distribution

Thus within the 0 to 200 km operating region the additional 1 to 2 ms latencies due to R-PHY and CIN delay are well within the normal latency distribution seen for DOCSIS traffic.

Why the Measured Results were Better than Theory

Inspection shows that the measured latency was approximately twice the CIN delay. This was in conflict with the theory developed in the previous section on upstream latency, which concluded that a minimum of three times the CIN delay would be incurred as request, grant and data messages must traverse the CIN.

This was the result of pro-active granting. The ACK Time field in the MAP was modified creating a mechanism to generate extra requests and hence grants. This caused the CM to think that requests had been dropped and to resend them. In fact the requests were still in transit to the CMTS so that the effect was to generate additional grants to the CM as shown in Figure 31 – Extra Grants.

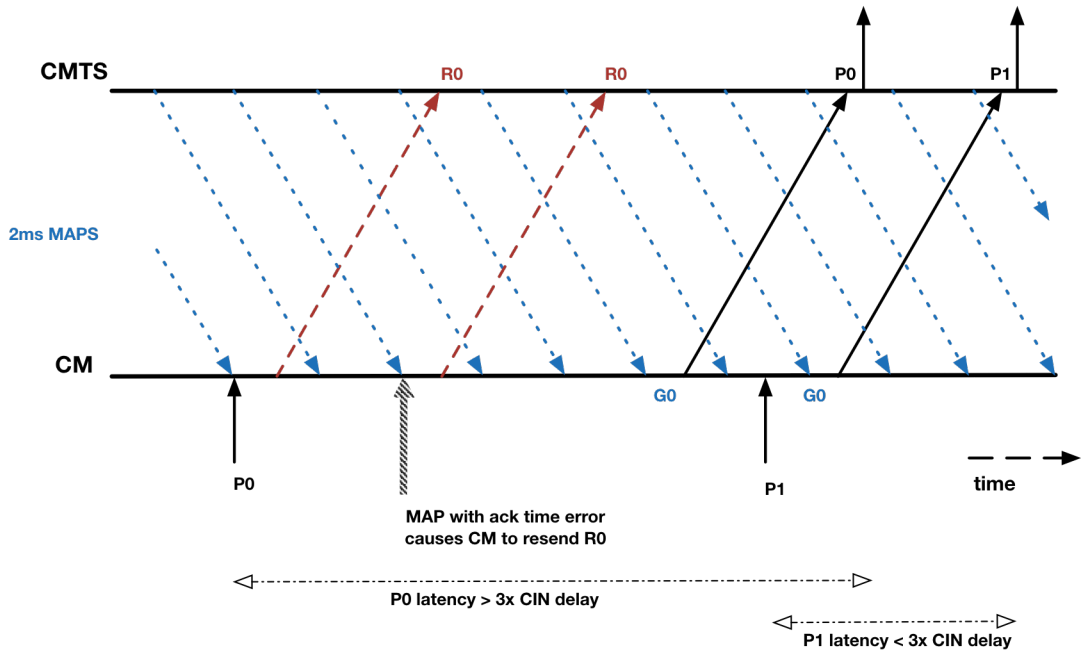


Figure 31 – Extra Grants

The sequence of events is as follows:

- Packet P0 is received at the CM
- Request R0 is generated asking for sufficient upstream bandwidth to send P0.
- A MAP is received with an ack time error. The CM looks at the ack time in the MAP and thinks that R0 has been lost. Thus it resends R0.
- The CMTS receives the first R0 request and responds with a MAP with grant G0
- The CM transmits P0 at the time specified in the map.
- The CMTS receives the second R0 request and responds with a MAP with a second grant G0
- The CM receives packet P1
- The CM transmits P1 at the time specified in the map for G0. Thus P1 has lower latency.

From the above it can be seen that the time to transmit P0 is in fact > 3x the CIN delay as expected (request + grant + data must all traverse CIN sequentially) but due to the additional grant P1 can be transmitted sooner as the request grant cycle is short circuited.

The simple act of generating additional grants to a CM had the effect of significantly reducing the latency impact of the CIN delay for upstream DOCSIS traffic.

The proactive granting applies to all the upstream data measurements in the paper. While the mechanism used was not practical for deployment it was much simpler to implement than changing the scheduler and served to illustrate the effectiveness of even a very crude proactive granting strategy.

Throughput and Frame Loss

This section contains additional test data that was generated and is included for completeness.

Downstream

The downstream tests were run at 10%, 50% and 90% load as described in the section Traffic Profile.

No downstream frame loss at any of the load profiles was seen as the CIN delay was increased suggesting that downstream throughput is not impacted by CIN delay even at the extreme limits of the test. This would be expected from a consideration of CMTS downstream operation as the CMTS core rate shapes the transmission of frames to the RPD to avoid overruns.

Upstream

The upstream tests were also run at 10%, 50% and 90% load as described in the section on Traffic Profile.

Table 7 shows that some frame loss in the upstream was seen, which is to be expected in a DOCSIS system. As seen from Table 7 and from Figure 32, this peaked at 0.04% for the 50% load case and dropped to 0 for the 90% load case.

Frame loss does not appear to increase as a function of CIN delay. The consistently higher frame loss during the 50% load testing appears to be anomalous and may be the result of problems during the test. Unfortunately we did not have time to repeat this test prior to publication but it warrants further investigation.

CIN delay (ms)	Upstream Frame Loss		
	10% load	50% load	90% load
0	0.001%	0.021%	0.000%
0.2	0.000%	0.021%	0.000%
0.4	0.002%	0.024%	0.000%
0.6	0.000%	0.032%	0.000%
0.8	0.001%	0.028%	0.000%
1	0.001%	0.040%	0.000%
3	0.013%	0.036%	0.000%
5	0.000%	0.041%	0.000%
7	0.000%	0.036%	0.000%
9	0.011%	0.036%	0.000%
10	0.004%	0.037%	0.000%
15	0.000%	0.037%	0.000%
20	0.000%	0.035%	0.000%
25	0.003%	0.036%	0.000%
30	0.001%	0.034%	0.000%
35	0.009%	0.037%	0.000%
40	0.009%	0.038%	0.000%

Table 7 – Upstream Frame Loss

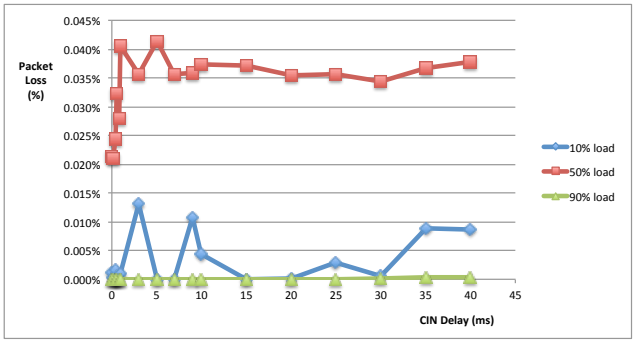


Figure 32 – Upstream Packet Loss vs CIN Delay

Summary of Test Results

The data provides a number of insights relating to a distributed architecture based on R-PHY:

- An R-PHY architecture can support 90% loading on both upstream and downstream links with minimal impact on packet loss for CIN delays of up to 40ms (corresponding to 8000km of fiber).
- Downstream latency increases linearly with CIN delay
- Proactive granting by the CMTS scheduler can reduce the upstream latency for best effort traffic.
    - Upstream minimum latency increases linearly at approximately 1x CIN delay (proactive granting working at its best)
    - Upstream average latency increases linearly at approximately 2x CIN delay (average proactive grant improvement seen during tests)
    - Upstream maximum latency increases linearly at approximately 3x CIN delay (no procative grant available seen during tests)

Recap

- This was a good first pass at testing and can form the basis for further and more in-depth testing.
- The R-PHY test data contained a rudimentary proactive granting scheme. The results without proactive granting should have shown a longer delay but were not available prior to publication. Conversely, a more optimized proactive granting scheme should result in average results that are closer to the shorter delays predicted by theory and simulation than were measured here.
- Upstream latencies followed a Poisson distribution offset by the minimum delay of the system.
- The latency of the system actually got lower with increased load. This is the opposite of what happens in a normal system. This is probably attributed to DOCSIS “piggy-back” grants kicking in which reduce contention grants. As such, until the test methodology is altered, it is the measured latency at 90% loading that should be the most significant.
- Low traffic density IXIA tests do not accurately represent a true low density network as IXIA testing sends evenly spaced packets rather than burst per CM. This prevents DOCSIS piggybacking from occurring. As such, the test setup should be modified to have a number of traffic bursts rather than evenly spaced packets.
- The R-PHY system for 0 to 100 miles worked very well (note that DOCSIS 3.1 now only requires 50 miles). The added delay for R-PHY versus I-CMTS seem to vary from 0.5 ms (from Table 2) to 2.3 ms (0.5 ms from Table 2 plus 1.7 ms from Table 4) which is close to the predicted one MAP interval time. This is compared to the average US latency of an I-CMTS which is on the order of 10 to 11 ms (Table 2).
- The R-PHY system with a centralized system worked at wirespeed for a 40 ms CIN latency

UNIFIED THEORY OF US LATENCY

A rudimentary formula that covers the observed behavior is shown below.

US packet latency = (Q x CIN Delay) + D ms + R ms

Where:

- Q is the impact of the CIN Delay. Q is 3 for reactive granting, and between 1 and 3 for proactive granting, depending upon algorithm efficiency.
- D = DOCSIS processing at CM and CMTS, including REQ backoff.
- R = Additional R-PHY processing.
- Values should all be Min, Avg, or Max

Measured latencies from Table 2 were:

CIN Delay: Min = 0, Max = 0.4 ms (D3.1) or 0.8 ms (D3.0), 40 ms (extended)

D: Min = 5, Avg = 10, Max = 50

R: Min = 0, Avg = 0.5 ms, Max = 2

Based upon these measure values, the CIN latency is insignificant for values below 1 ms and really only becomes significant for value of 2 ms (400 km) or more. It also shows that moving the scheduler to the RPD will have little to no impact.

Note: The measured values are based upon limited experiments and specific test patterns and circumstances and are subject to change.

CONCLUSIONS

One of the biggest features of the Remote PHY architecture is that the software – including the upstream scheduler – is completely centralized. This allows for complete interoperability between a CCAP core and third party RPD vendors. Imagine the challenges if your CMTS Core software came from one vendor and your upstream scheduler came from another vendor. Would it work? Who would you call when it did not work?

The potential technical challenge with centralizing the upstream scheduler has been the REQ-GNT delay. This white paper split the problem into two categories:

1. 0 to 100 miles (DOCSIS 3.0 spec)
2. 100 miles and beyond (beyond spec)

It then compared various R-PHY deployments against an I-CMTS deployment.

The impact of R-PHY on a system 0 to 100 miles is roughly zero to one MAP intervals or about 0 to 2 ms. This was confirmed by theory, simulation, and lab measurements. Meanwhile, the average upstream delay of the I-CMTS and R-PHY at 0 ms was found to be on the order of 10 ms. This is a result of REQ retries, CM and CMTS buffering, processing and reassembly.

The predicted and measured upstream packet latencies in systems less than 100 miles are 10x or more than the one-way latency from the CMTS Core to the RPD.

Thus, the latencies in systems less than 100 miles in radius tend to be dominated by MAP duration and packet buffering and reassembly rather than the REQ-GNT delay.

R-PHY can support the same high utilization of DOCSIS channels as an integrated system due to the pipelining inherent in the DOCSIS protocol.

A basic R-PHY system with an unmodified upstream scheduler will provide near-equivalent latencies to an integrated system for distances of up to 160 km (the existing DOCSIS distance limit).

The white paper took the research further to see if the CMTS core could be relocated from the hub site to a more central location.

Upstream packet latencies in systems that are much greater than 100 miles are dominated more by the CMTS Core to the RPD path latency. In fact, the REQ-GNT delay would normally scale by three times the system one-way latency.

However, with the use of a proactive upstream scheduling mechanism, an R-PHY system can provide latencies that are equivalent to an integrated system for distances well beyond the DOCSIS limit.

An R-PHY system was tested at the equivalent of 8000 km (5000 miles) from the CCAP core to the RPD (based upon 40 ms of CIN delay. This is the equivalent of 1.7 times the width of the USA.

The high tolerance of R-PHY to CIN delay will simplify virtual deployments by allowing the use of standard virtual machine environments.

Theory, simulation and measurement align to show that an R-PHY solution can be successfully constructed and deployed.

Further, deployment of a CMTS Core in a regional data center with Remote PHY Devices in fiber nodes with centralized

software is also a very practical architecture
up to significant CIN distances.

APPENDIX I – DELAY IMPACT OF NODE TO CORE DISTANCE

The added delay between the CMTS core and the Remote PHY Device has two components

- The transmission time over the transport links
- The switching time in any intermediate nodes

Transmission Time

Table 8 shows the transmission time for various link lengths based on the speed of light in fiber of $2*10^8$ $ms^{-1.}$

Distance (km)	Accumulated Transmission Time (mS)
100	.5
200	1
300	1.5
400	2
500	2.5
600	3
700	3.5
800	4
900	4.5
1000	5
2000	10
3000	15
4000	20
5000	25
6000	30
7000	35
8000	40

Table 8 – Transmission Time as Function of Distance

Switching Time

The network between the CMTS Core and the Remote PHY Device is expected to be a controlled network that is not oversubscribed. The switching time for current 10G Ethernet switches in the absence of congestion is < 2us for a 1518 byte packet.

Number of Switching Nodes	Accumulated Switching Time (mS)
1	.002
2	.004
3	.006
4	.008
5	.010

Table 9 – Ethernet Switching Time

Table 9 shows the accumulated switching time in a multi-hop network. This shows clearly that switching time can be expected to provide a minimal contribution compared to overall network delay in a well designed and managed network.

Thus the total added latency is approximately 1ms per 200km (ignoring switching time in nodes)

Note: DOCSIS one way latency limit of 0.8ms = 160km

APPENDIX II – CCAP CONFIGURATION

CMTS configuration

```
interface Cable8/0/0
 load-interval 30
 downstream Modular-Cable 8/0/0 rf-
      channel 0-3
 cable mtc-mode
 no cable packet-cache
 cable default-phy-burst 0
 cable map-advance dynamic 1000
      20000
 cable bundle 1
 cable upstream max-ports 4
 cable upstream bonding-group 800
  upstream 0
  upstream 1
  upstream 2
  upstream 3
  attributes A0000000
 cable upstream max-channel-power-
      offset 6
 cable upstream balance-scheduling
 cable upstream 0 connector 0
 cable upstream 0 frequency 15000000
 cable upstream 0 channel-width
      6400000
 cable upstream 0 power-level 2
 cable upstream 0 docsis-mode atdma
 cable upstream 0 minislot-size 1
 cable upstream 0 power-adjust
      continue 6
 cable upstream 0 range-backoff 3 6
 cable upstream 0 modulation-profile
      225
 cable upstream 0 attribute-mask
      20000000
 cable upstream 0 shutdown
 cable upstream 1 connector 0
 cable upstream 1 frequency 21400000
 cable upstream 1 channel-width
      6400000
 cable upstream 1 power-level 2
cable upstream 1 docsis-mode atdma
 cable upstream 1 minislot-size 1
cable upstream 1 power-adjust
      continue 6
 cable upstream 1 range-backoff 3 6
 cable upstream 1 modulation-profile
      225
 cable upstream 1 attribute-mask
      20000000
cable upstream 1 shutdown
cable upstream 2 connector 0
cable upstream 2 frequency 27800000
cable upstream 2 channel-width
     6400000
cable upstream 2 power-level 2
cable upstream 2 docsis-mode
     atdmacable upstream 2
     minislot-size 1
cable upstream 2 power-adjust
     continue 6
cable upstream 2 range-backoff 3 6
cable upstream 2 modulation-profile
     225
cable upstream 2 attribute-mask
     20000000
cable upstream 2 shutdown
cable upstream 3 connector 0
cable upstream 3 frequency 34200000
cable upstream 3 channel-width
     6400000
cable upstream 3 power-level 2
cable upstream 3 docsis-mode atdma
cable upstream 3 minislot-size 1
cable upstream 3 power-adjust
     continue 6
cable upstream 3 range-backoff 3 6
cable upstream 3 modulation-profile
     225
cable upstream 3 attribute-mask
     20000000
cable upstream 3 shutdown
cable sid-cluster-group num-of-
     cluster 2
cable sid-cluster-switching max-
     request 1
```

REFERENCES

Related Papers by the Same Authors

[1] John T. Chapman, "Remote PHY for Converged DOCSIS, Video and OOB", NCTA Spring Technical Forum, June 10, 2014

[2] John T. Chapman, "DOCSIS Remote PHY," SCTE Cable-Tec Expo, October 21, 2013

[3] John T. Chapman, Rakesh Chopra, Laurent Montini, "The DOCSIS Timing Protocol – Generating Precision Timing Services from a DOCSIS System", NCTA Spring Technical Forum, June 14, 2011

[4] John T. Chapman, Jeff Finkelstein, "Maximizing Legacy and DOCSIS 3.0 Bandwidth in the Forward Path using M-CMTS and EdgeQAM Technologies", SCTE Cable-Tec Expo, October 30, 2009 http://www.johntchapman.com/Legacy-and-DOCSIS3-with-M-CMTS-WP-%282009-10-30%29.pdf

[5] John T. Chapman, "Next Generation CMTS – An Architectural Discussion", SCTE Emerging Technologies, January 15, 2008 http://www.johntchapman.com/NG-CMTS-WP-071125.pdf

[6] John T. Chapman, "DOCSIS Timing Interface – Timing for DOCSIS Networks", National Institute of Standards and Technology (NIST) Keynote Address, March 10, 2007 http://www.johntchapman.com/NIST-jchapman-070310.pdf

[7] John T. Chapman, "Modular CMTS Tutorial", CableLabs Summit, December 2006 http://www.johntchapman.com/M-CMTS-Tutorial-Public-070220a.pdf

[8] John T. Chapman, Andrew Page, Carmelo Iaria, Glenn McGilvray, "Transitioning to M-CMTS", NCTA National Show, April 2006

[9] John T. Chapman, "Modular CMTS Architecture", NCTA National Show, April 3, 2005 http://www.johntchapman.com/ncta-2005-modular-cmts-wp.pdf

[10] Gerry White, "A Next Generation Architecture for Business Services Delivery Based on SDN and NFV", SCTE Cable-Tec Expo, 2014

[11] Gerry White, "Transforming Cable Infrastructure into a Cloud Environment", NCTA Spring Technical Forum 2012.

[12] Gerry White, John Ulm, "New Converged Access Architectures for Cable Services", NCTA Spring Technical Forum 2011

Related CableLabs Specifications

[13] "CM-SP-R-PHY: Remote PHY System Specification", Revision I01, CableLabs, 2015

[14] "CM-SP-R-DEPI: Remote Downstream External PHY Interface", Revision I01, CableLabs, 2015

[15] "CM-SP-R-UEPI: Remote Upstream External PHY Interface", Revision I01, CableLabs, 2015

[16] "CM-SP-R-DTI: Remote DOCSIS Timing Interface", Revision I01, CableLabs, 2015

[17] "CM-SP-R-OOB: Remote Out Of Band", Revision I01, CableLabs, 2015

[18] "CM-SP-R-OSSI: Remote PHY Operations Support System Interface", Revision I01, CableLabs, 2015

[19] "CM-SP-GCP: Generic Control Plane", Revision I01, CableLabs, 2015

ACKNOWLEDGEMENTS

The measurements described were part of a series of tests executed by the following Cisco engineers

- Atlanta and RTP teams: especially John Downey, Chris Leghorn, John Skrobko and Abhi Mahendiran.
- China Team: especially Yubin Chen and Yi Zhang

Many other Cisco engineers provided insight into the theoretical and practical subtleties of the DOCSIS protocol, which was invaluable in producing this paper.

INTRODUCING AN ABSTRACTION LAYER TO CONTROL THE USER EXPERIENCE, APPS AND INTEGRATE CONTENT ON ALL SCREENS

Albert Dahan, Founder and CTO, Metrological

Abstract

Multichannel video programming distributors (MVPDs) are facing increasing complexity as they seek to deliver pay-TV services, over-the-top (OTT) content, widgets, and third party applications that are integrated into a single experience and accessible across a growing number of devices. MVPDs want to accomplish this with the ability to instantly control the user interface (UI) and user experience across devices. The challenge is that each device behaves differently and requires device-specific solutions if MVPDs decide to solve this natively. A unified approach that enables MVPDs to simultaneously introduce new services and applications across devices is needed to simplify the process and reduce operational expenditures.

Resolving this issue demands a controversial approach. Rather than continuing to develop big, closed, monolithic software solutions, MVPDs can now use small, modular software components that connect new web-based and open source technologies with existing legacy ecosystems, to avoid being locked into proprietary systems. One way to resolve this issue is to use an abstraction layer that separates the MVPD backend, device specific middleware and operating system from the user interface, presentation and navigation layers on each device. This abstraction layer or framework should operate independently from the device and have its own mechanism in place to manage services and applications (apps) in real-time within the framework. This framework should also be able to support features that bridge the gap separating live TV and OTT content from features such as contextual apps and widgets, and merge all content on to one screen.

This paper will outline how to implement a managed framework approach that uses the browser as a generic integration point to run and control the user interface, apps and client services across connected devices. It will discuss how a cloud-based framework can be optimized and accelerated within a specific MVPD ecosystem and applicable set-top box (STB)/browser configuration to improve the user experience and minimize CPU and memory usage of the end device. Apps only need to be developed once to run on all devices and can be developed by anyone using an open Software Development Kit (SDK). This enables MVPDs and content providers to control their services and user experience in real-time, reduce costs, and rapidly introduce new apps and services.

BACKGROUND

Change is underway in the pay-TV industry as proprietary and hardware-based platforms are being replaced with more flexible IP delivery and software-driven solutions that are increasingly based on open source. The industry's evolution from offering video on a single TV screen to multi-screen delivery of live TV, OTT content and apps requires more agile solutions that can rapidly deliver new services to emerging devices.

With STB apps projected to reach 40 percent of pay-TV households by 2019 (source: ABI Research), TV apps represent the next logical extension to the app-driven

world that mobile subscribers have embraced. Bringing app content to the TV transitions the experience from a passive, lean-back experience into a lean-forward, proactive experience that offers more dynamic engagement, such as games, quizzes, voting, and personalized advertising.

An application framework or abstraction layer approach can bring integrated viewing experiences to the entire device ecosystem by using the browser as a generic integration point to run and control the user interface, apps and client services across connected devices.

With a device and software agnostic platform approach, MVPDs and content owners only have to develop an app once to deploy across all devices. This platform operates independently from the device itself, ensuring that all relevant hardware and software applications can be controlled and managed centrally. App platforms offer the flexibility to quickly configure and launch integrated app services, and control the user experience across devices in real-time. This abstraction layer enables new levels of personalized experiences, such as apps that present contextual content specific to the viewer's interests across devices.

By processing video data on this framework and on very low-powered hardware, MVPDs can yield the benefits of minimal CPU usage with accelerated memory performance. Using this technology, MVPDs with legacy STBs can now launch robust TV app platforms without needing to deploy next-generation STBs.

In the last few years, this type of framework or abstraction layer approach has become more prevalent, as an alternative to streamline app development and deployment for MVPDs. This technology lays the groundwork for MVPDs to develop a successful TV app strategy that goes beyond offering a collection of apps. Operators can deploy their own branded app store that complements the TV experience, including a relevant catalog of apps and widgets, with an underlying technical platform to configure, launch and operate it.

TRENDS DICTATING A NEW APPROACH

Technology drivers

Today's consumers demand delivery of video whenever and wherever they are, in a way that is seamless to access and consume. This puts pressure on MVPDs to combine live TV, video on demand and OTT/app services into a single user experience across devices.

MVPDs typically have a separate backend management structure for content source. Managing multiple, separate structures becomes more challenging as services are distributed to an increasing number of devices.

Offering OTT/app content in addition to live TV and on demand viewing increases the CPU and memory demands on the STB. This directly impacts overall STB costs and complexity for MVPDs with older generation STBs in their footprints.

Business drivers

A number of business drivers are causing the industry to look for a new way to develop and deploy TV apps including:

- The development costs of TV app stores include the number of apps, devices and customizations required. These costs are increasing dramatically due to the volume

of devices and networks on which the MVPD needs to have an app available.

- Time to market is a big cost determinant in the overall development of TV apps. The number of apps, device and network customizations required and back office complexities are creating costly time delays in getting apps and TV app stores deployed. The faster an app or app store can be deployed, the sooner the MVPD can monetize the service and see benefits such as stickiness and consumer satisfaction.

- MVPDs need to be able to productize all video offerings, including apps. But because MVPD backend systems aren't able to manage different assets agnostic to its heritage or distribution (live TV, on demand and OTT/apps), MVPDs are often unable to productize, distribute and deliver video in a way that meets consumer expectations.

EVOLUTION OF TV APP PLATFORMS

Historically, MVPDs have approached TV app deployment by developing native apps that run on STB middleware clients. This approach usually requires an MVPD to develop an app for each STB platform and multi-screen device that a consumer will use to access the service. This approach does not scale and is often too costly, requiring device specific deployment and maintenance.

In the past, some MVPDs used web-based apps for TV app development. However, using web apps instead of native apps causes each device to behave differently, requiring specific browser optimizations. This approach also lacks proper integration with the device middleware and MVPD backend components.

Today, each MVPD offers its own proprietary environment to support app publication, causing app publishers, such as content providers, to struggle to keep pace with app development. This means that each app needs to be specifically developed to run within this environment. This results in huge development costs for content providers that want to have a broad reach across all MVPD environments. It also causes some content providers to offer apps to a limited number of MVPDs.

MVPDs face the same issues as content providers. The development and maintenance costs required for several different versions of apps increases the operational costs, making it difficult to develop an attractive business case for TV apps.

TV app platforms have been moving back and forth between native and web-based technologies without great success. 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 will be reviewed in the next section of this paper.

Overcoming HTML5 drawbacks

HTML5**,** the fifth revision of the HTML standard, is a markup language used for structuring and presenting content for the Internet (source: Wikipedia). The HTML5 standard is emerging rapidly for tablets, phones and desktop environments. As is the case with evolving standards, MVPDs need to address a few key drawbacks in order to overcome issues that can arise in using HTML5.

HTML5 complexities can be minimized by first ensuring that the browser is up to date with the HTML5 features implemented correctly on the device. Using HTML5 combined with CSS and JavaScript gives developers an easy way to develop applications. HTML5 tools for profiling, debugging and rendering flaws are evolving to a level that can now be used on embedded devices.

Second, MVPDs need to ensure that the app framework addresses aspects such as Window Management, focus behavior, application lifecycle, security, compatibility layers or polyfills.

Lastly, using a unified abstraction between the browser and the app can help to further reduce complications. MVPDs need to move Application Program Interfaces (APIs) to the cloud as much as possible. This way APIs can be updated independently and without causing changes to the framework or the applications. Selecting a user interface framework that abstracts rendering across browsers and device types will simplify rendering across devices.

MANAGED FRAMEWORK APPROACH

Simply deciding to quit offering native apps does not solve the complexity. The device browser requires a mechanism that accelerates and optimizes the support of browser-based apps. Secondly, there's a need to create a controlled environment that separates the backend, device specific operating system and native client from the browser and the platform on which web apps can be developed in one unified way.

This mechanism or framework also needs to handle the integration with web apps and native device aspects such as device navigation, scaling and display, and other advanced features that require connection between the web app and the consumer device.

MVPDs are under competing pressure to lower the cost of service and to increase the OTT and app content and app capabilities on the STB. Today's more prevalent approach, which handles everything locally on the device, is not equipped to solve these two mutually exclusive demands on the STB. The way to solve this roadblock is to ease the demands on the STB by handling these new features outside the device.

Taking steps to make sure that technical platforms are open and easy to integrate will help make scaling and operating easier. Additionally, it will also help MVPDs to remain attractive to content providers and developers to expand app content and libraries to subscribers.

ELEMENTS OF A FRAMEWORK / ABSTRACTION LAYER APPROACH

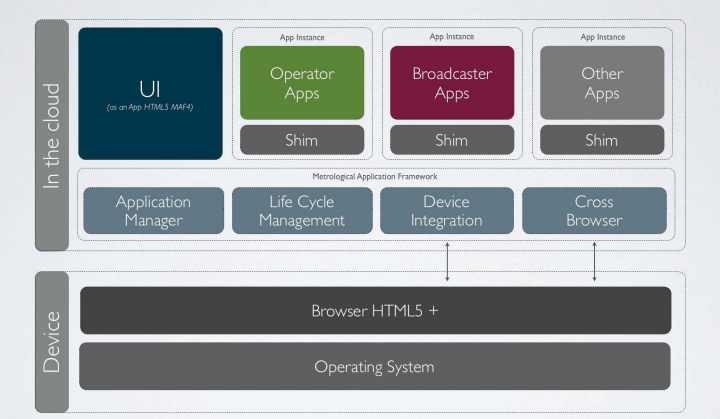


Figure 1: Sample Architecture for TV App Abstraction Layer

Definition and approach

This approach uses an abstraction layer that separates the MVPD backend, device specific middleware and operating

system from the user interface, presentation and navigation layers on each device. This abstraction layer operates independently from the consumer device, including a STB or any second-screen device, has its own browser onboard, and has its own mechanism in place to manage services and apps in real-time within the framework. This framework is also able to support features such as contextual apps and widgets, and merge all content onto one screen.

This modular approach is unlike today's self-contained apps that have key functionality built directly into the app itself. All applications are isolated in their own sandbox from a development and security process. This separates the user interface, EPGs, channel bar, notification, VOD , etc. into a shared resource, thus simplifying the lifecycle management of each app and speeding up the development process. Now key functionality can be used in a more modular way rather than being confined to single app. For example, for a self-contained app a channel lineup change means that the entire app needs to be updated. An abstraction layer approach enables MVPDs to create apps with specific content but to share these key functionalities such that updates are shared and can be performed automatically

Benefits over other approaches

The implementation of a managed framework approach that uses the browser as a generic integration point to run and control the user interface, apps and client services across connected devices offers several benefits:

- The cloud-based framework can be optimized and accelerated within a specific MVPD ecosystem and applicable STB/browser configuration to improve the user experience and minimize CPU and memory usage of the end device.

- Apps only need to be developed once to run on all devices.

- Apps can be developed by anyone using an open SDK. This enables MVPDs and content providers to control their services and user experience in real-time, reduce costs, and rapidly introduce new apps and services.

- The framework backend can support live TV, video on demand and OTT/app content with an integrated experience on the TV screen.

SYSTEM ARCHITECTURE

Technology requirements

The main requirement for implementing an abstraction layer is to ensure that devices have a secured/trusted browser environment that provides maximum support for HTML5 specifications. This includes specifications that are still in draft format. HTML5 is evolving rapidly and the browser on the device should be able to adapt to updates and changes accordingly. Most of the middleware capabilities can be mapped to HTML5 specifications or moved to the cloud.

Integration process

By enabling an application framework, device integration is simplified, lifecycle management is controlled and runtime is securely isolated. The key to leveraging the application framework is to (1) isolate features within an application and (2) expose services between applications. For example, video on demand and EPG are separate applications but search and recommendations

functions should be able to leverage both types of content within the app. The basics of an application framework should contain window and focus management, lifecycle management of applications, resource management, security, and cross application communication.

Furthermore, it's important for MVPDs to select a user interface framework that runs correctly and efficiently on embedded devices. Web application development should use the correct development tools to isolate memory leaks, hardware acceleration bottlenecks, and avoid high CPU and memory usage.

This approach uses the device browser. Therefore, the framework settings need to be optimized and the browser needs to be accelerated.

Additional points of integration that are required include:

- STB integration: enables app store navigation via the remote control unit, provisioning of channel tune information, exchange of channel change commands, etc.
- EPG integration: enables apps to run contextual to the live TV or IP video content played out.
- User management integration: handles user access, user profile information and other criteria.
- Billing integration: processes and handles app and in-app purchases via the MVPD invoice.

NEW MIDDLEWARE APPROACHES

Most new middleware architectures have a browser approach, but enabling HTML5 development can pose challenges for middleware testing. It is important to leverage HTML5 correctly. For example, in creating a rich user experience for devices that have OpenGL capabilities, MVPDs and developers should take a look at new standards, such as WebGL.

LEVERAGING OPEN SOURCE TECHNOLOGY

The use of open source components will help speed development on both apps and devices. Open source offers a faster time to market and enables MVPDs to tap into a big developer community to bring more apps and innovation to TV apps.

There are, however, downsides to using open source in the TV environment. Open source can pose a security risk because of its inherent transparent approach to the developer community. MVPDs should also be aware of the restrictions and legal risks associated with an open source approach. For example, the basic philosophy of making software source code available to developers and providing software to others with limited or no restrictions on its use can have legal impacts on intellectual properties, patents and revenue streams.

ABSTRACTION LAYER IMPLICATIONS FOR THE APP DEVELOPER COMMUNITY

Benefits of an open SDK

Using an open SDK is crucial in simplifying the development of apps. The SDK should come equipped with an open

source and lightweight user interface framework, to ensure performance and stability on any device. By developing on the SDK, MVPDs can ensure the application will run on any type of device.

An open SDK also circumvents the need to redevelop or 'reinvent the wheel' every time a new operating system is launched, and with pre-packaged app store content available, MVPDs can, in effect, hit the ground running.

And finally, an open SDK also offers a more cost effective means for distributing niche content. Using an SDK, developers can reach target audiences more effectively, rather than using a wider broadcast route.

NEW FUNCTIONALITY

This framework approach is able to merge all content sources and support new features that enhance the overall consumer experience. A key advantage of using an app framework approach is that it enables MVPDs to bring these new features and the same user experience to legacy STBs without touching the device. Examples of this type of functionality include:

- Device pairing: enables viewers to connect their mobile phone and use it as a remote control, a video game controller, a payment device, or an authorization mechanism that enables consumers to simply to swipe videos to their TVs for instant viewing.

- Contextuality: uses APIs that allow apps to recognize what the customer is watching and suggest contextually relevant internet content across devices. For example, while watching sports or a wildlife movie, contextual apps can display related content such as real-time scores or animal clips, integrating app or OTT content with the live TV experience. A contextual app can deliver "Red Button" functionality as well. The overlay or side-by-side app supports interactive program formats, such as live voting and sing-alongs, enriching the viewer's experience of a live broadcast.

- Social Media Integration: personalizes the TV experience, enabling consumers to share TV and app content with family and friends.

- Cloud based UX: supports centrally managed, cloud-based user interfaces that can work across multiple devices. This offers the potential for MVPDs to be able to perform a single update of the user experience that leads to an automatic, real-time update of the user interface across all devices.

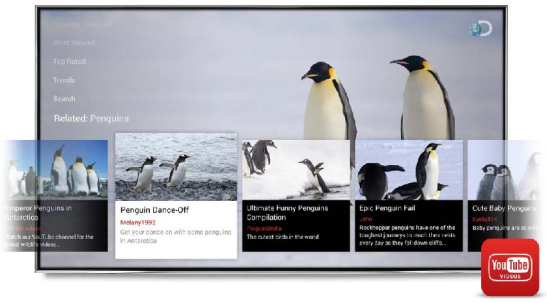


Figure 2: An Example of Contextual Viewing Experiences

Benefits for TV MVPDs

The framework or abstraction layer approach can be productized, as a fully customized TV app store with the ability to manage the user experience across devices. This enables MVPDs to easily add a suite of business management and business intelligence tools. A web-based CMS enables MVPDs to manage their own app stores and user experiences across devices in real-time.

The approach can be used to deploy an app store and user experience on legacy devices. The framework can deploy an app store using a cloud rendering service even if the device does not have a browser. In this case the app content will be delivered to the legacy device "as a live channel".

This means that legacy devices and browserless STBs don't need to be replaced in order to support TV apps stores or offer more integrated viewing experiences. In essence, these STB platforms are "refitted" by putting the user interface and OTT/app services in the cloud. This approach has a number of other benefits for MVPDs including:

- Enables a quick launch of MVPD branded app stores with a unified consumer experience across screens.
- Offers a device agnostic approach across an expanding device ecosystem.
- Enables content monetization that helps MVPDs to increase ARPU, customer loyalty and reduce churn.
- Provides an integrated experience that combines live TV, OTT, social media content and capabilities into a single consumer experience.

BENEFITS FOR CONTENT PROVIDERS

In addition to benefits for MVPDs, this approach also has benefits for content providers. Using an abstraction layer approach, programmers and other content providers can develop an app once, rather than multiple times to reach multiple devices, in order to connect and interact with consumers. The technology also offers more reliable delivery with mechanisms that ensure that content is provided with quality of service to any device.

DEPLOYMENT EXAMPLES

Liberty Global uses an application abstraction/framework approach that is currently live across three of their platforms, including a cloud-based browser rendering solution that is deployed on its Horizon platform.



Figure 3: Selection of Liberty Global Horizon TV Apps

All applications are developed using an SDK that runs directly across all platforms and is integrated with the app framework. For example, a hackathon that was organized with 100 developers in collaboration with Liberty Global, created 20 applications during the event using the open SDK.

Market results

UPC Hungary, a Liberty Global property, used a variation of this technology to bring online content to its pay-TV subscribers. The MVPD partnered with YouTube and other technology companies to deliver 20-plus apps to legacy STBs. This was the first time that an MVPD offered an online video service to its entire subscriber base using legacy STBs.

In this deployment, UPC Hungary is using a cloud-based interface and TV app framework, including more than 20 apps, that is being delivered over the MSO's managed network. The MVPD is offering the TV apps to its 520,000 subscribers as part of its platform, providing additional value for existing subscribers.

UPC Hungary subscribers are streaming more than a million minutes per day of YouTube content. 68% of subscribers have tried the TV apps service, and of that number, 83% have returned for additional views.



Figure 4: UPC Hungary YouTube App Usage

CONCLUSION

MVPDs need a mechanism that enables them to offer OTT/app content seemlessly integrated with the live TV and VOD offerings across a growing number of devices. An abstraction layer approach can be used to integrate these content sources with a single backend management system that streamlines app development and deployment. By separating the MVPD backend, device specific middleware and operating system from the user interface, presentation and navigation layers on each device, MVPDs can avoid having to develop native apps for each and every device.

This approach is deployable on legacy STBs to enable robust MVPD branded TV app store deployments. It also brings new functionality to TV apps, such as interactivity, contextual experiences, mobile device pairing, and social media integration that resonate with consumers. Global MVPDs are just beginning to reap the benefits of this technology that is engaging consumers and driving increased content and app usage on MVPD branded app stores.

Lady Giga Takes the Stage

Dr. Robert L. Howald
Vice President of Network Architecture
Comcast

Abstract

The insatiable appetite for rich content, ubiquitous connectivity, and sharing of experiences through social media has kept the battle for high-speed data supremacy alive and highly competitive. There does not appear to be an end in sight to the growth of IP-capable devices in the home. A single household may have a dozen or more devices with connectivity. Most of them are video capable and part of an emerging pattern of behavior of simultaneous use. "Lean Back" viewing is a decaying consumer segment.

With so many video-capable devices, today's "killer app" remains streaming video, and its impacts have been widely felt. And, video evolution itself is not complete, as 4kHD clearly shows. Meanwhile, the consumer market has barely scratched the service on the Smart Home, Telemedicine, or other potentially bandwidth-consumptive, non-media related applications. An end to IP data growth does not yet appear to be in sight.

So, while the debate continues about whether year-on-year capacity growth is permanent or becomes a gradual tapering over time, every service provider knows it is not sufficient to simply look at the demand side of capacity. It has long been considered table stakes to keep pace with persistent year-on-year data growth by adding IP resources. Operators have to simultaneously focus on the service speed offering, also referred to as "billboard speed." In this category, the latest market buzz that Operators are answering the bell for is delivering the Gigabit service tier.

In this paper, we will explore the delivery of a Gigabit from the operator's perspective. We will look at the marketing aspects that have driven the message into industry and consumer lexicon. We will compare the buzz created to realities of common usage, aggregate demand, and traffic engineering. We will consider the obstacles to delivery and consumption and the relationship to application benefits. We will describe the "layers" of the Giga-Ask, and identify the solution possibilities and service capabilities enabled by these solutions. We will take on the technical details of these solutions and project where they can take us down the road. Lastly, we will articulate the key messages that the Cable industry should focus on as Lady Giga takes the stage.

PERSISTENT AGGRESSIVE CAGR

Consumer data capacity demand has advanced steadily in a compounding fashion since the introduction of broadband Internet services. The commonly accepted Compound Annual Growth Rate (CAGR) of downstream traffic is 40-50% per year. Various approaches have been taken to chart the growth of downstream traffic as it relates to HFC capacity to guide network investment strategy, including the insightful Capacity Management Timeline approach, an example of which is shown in Figure 1 and described in [5]. The 1 Gbps capacity requirement, from a demand perspective, is based on a

hypothetical DOCSIS allocation and scenario in 2014, as identified on the chart.

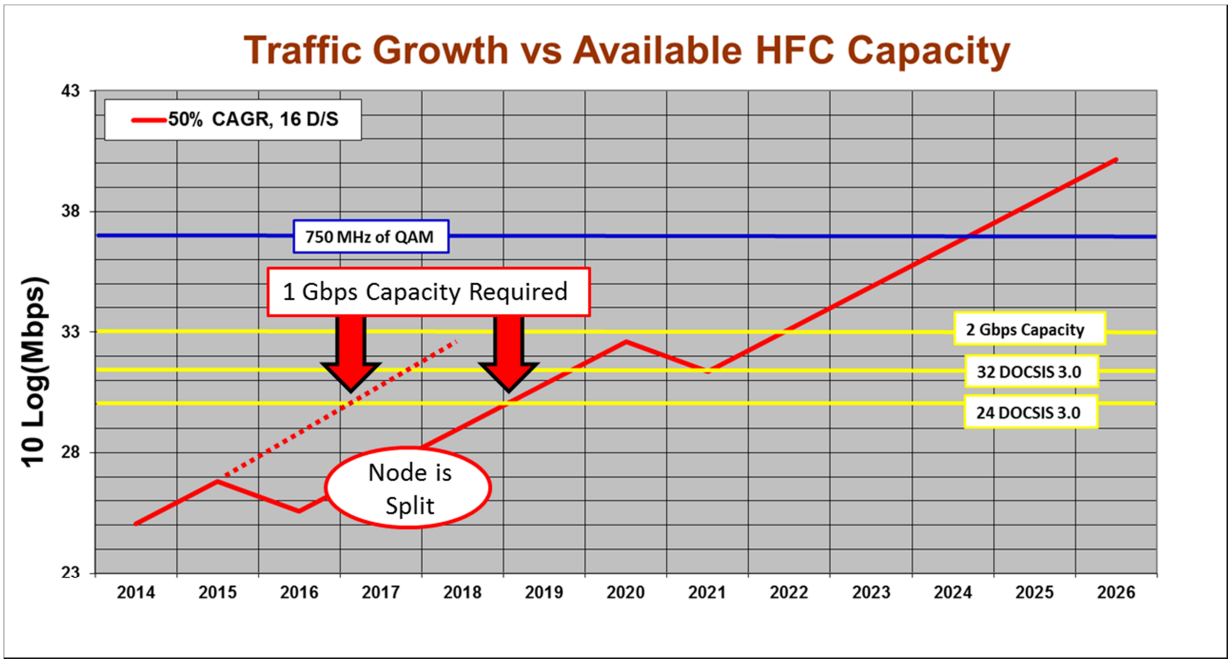


Figure 1 – Capacity Management Timeline Approach to Traffic Growth

It is sometimes more convenient in practice to consider growth in terms of Traffic Doubling Periods (TDPs), or how much time passes before the traffic is 2x the current amount. Conveniently, 40% is a doubling of approximately every two years (50% is about 21 months). Other convenient rules of thumb, in particular as we evaluate options for the upstream, are a TDP of 3 years for a 25% CAGR and 4 years for a 20% CAGR.

Doubling periods are useful because we can then easily frame the problem in terms of DOCSIS channels, CMTS ports, Headend combining, or nodes splits or segmentations. These are all known tools in the toolkit that are used regularly to manage capacity growth of HFC networks. And, they are also well-understood in terms of financial, facilities, and network bandwidth implications. There are other tools available, such as video encoding technology (HEVC) and the move to an all-IP network.

As an example of TDP use, Comcast has raised Internet data speeds on average once a year for over a decade to keep pace with capacity demand and the applications that require sustained broadband speeds, such as over-the-top (OTT) video and file sharing services. It wasn't that long ago that 10 Mbps was a tremendous breakthrough, one that enabled strong market share against DSL-based Internet providers in North America in the early years of DOCSIS.

Using the simple TDP rules above, going from 10 Mbps speeds to 100 Mbps speeds at 40% CAGR is just over 3 TDPs, or about 6 years. Coincidentally, DOCSIS 3.0, which enabled 100 Mbps of per-user capacity, was completed in 2006. This crude "10x" analysis applies also for the case of going

from 100 Mbps to 1 Gbps. The 100 Mbps target was a major consideration for the channel bonding approach used in DOCSIS 3.0, since single 6 MHz QAM slots were limited to about 40 Mbps. Just as DOCSIS 3.0 development began and completed in advance of the 100 Mbps expected need, the DOCSIS answer to Gigabit speed – DOCSIS 3.1 – has been developed in advance of the projected need, and appears to be arriving just in time to deliver on this emerging market target.

(Note that throughout the paper, we will use the word "Gigabit" as synonymous with Gigabit per second or 1 Gbps).

GO GIG OR GO HOME

In the data-centric world of Ethernet, cost effective connectivity advanced to support residential broadband speeds, but more so because the computer era drove data technology for businesses, and businesses had internal data needs before demand growth of Internet access was a driving factor. Today, business services Internet requirements tend to be more symmetric, while residential services tend to be more asymmetric. In short, Ethernet connectivity preceded and lay in wait of access speeds capable of using 100BaseT connections.

Eventually, as residential speed wars reached beyond 100 Mbps, the steady battle to claim the fastest Internet meant continually increasing top tiers….200 Mbps, 250 Mbps, 300 Mbps, and even 505 Mbps, despite modest impacts to the overall Quality of Experience (QoE) for most users. While not necessarily tied to user experience for the majority of Internet subscribers, there is certainly marketing value to being the fastest service provider in an era when the Internet is no longer a luxury, but an essential service at the top of the list of what consumers find important. And, of course, business advantage could be had in some cases by deploying a service speed that put a different access technology and architecture at a competitive disadvantage.

As speeds worked their way through the Nx100 Mbps range, the "Gig" moniker came into view. Now, rather than just bigger numbers, an entirely new language and branding opportunity was in play. As such, compared to other speed war thresholds, "Gig" arrived perhaps sooner, in part spurred on by new competitors taking advantage of the nature of the linguistic change, but also by marketing of fiber-based solutions as the answer to having a network that is deemed "future-proof." Google Fiber, though modest in actual customers served and limited to only a small number of select cities, has aggressively marketed FTTH deployments around Gigabit capability. Gigabit is showcased through a host of futuristic applications that presumably could drive CAGR trends long into the future.

On the Rebound

The economic recovery took many years to get underway, and only relatively recently has new construction become robust. Over that period of time, technology has only become more infused into everyday life through social media and a variety of other applications available on smartphones and tablets. Builders and property owners recognize the importance of High Speed Internet (HSI) to potential new residents. In fact, on surveying, it is the MOST important amenity selected [6]. Showing new potential residents that the property is committed to the best technology and Internet experience is a major part of a successful business model for property owners. As a result, these owners have become aggressive in marketing

the technology aspects of their properties to potential residents.

While there is a general awareness that Gigabit service has primarily marketing value today as opposed to being necessary for actual traffic demand, the "future proof" component of Gigabit has significant value, and for good reason. This is particularly so when considering properties that forge a bulk services agreement with an operator based on a multi-year contract. Given the CAGR trends discussed above, it makes sense that property owners prepare their property for what the Internet service demands will look like at the *end* of their contract, and to do what they can to prevent disruption due to rewiring on the property during the course of the contract.

Lastly, trends in new housing growth indicate that multiple dwelling unit (MDU) living, often in downtown cores and/or live-work-play model village-towns are a larger weighting of new homes than was the case prior to the real estate crash. The demographic in the MDU environments tends to be younger than the average US adult. This trend can be attributed to the economic recovery that is allowing 5+ years of underemployed but relatively recent college graduates to become fully employed and begin to seek better living conditions.

New customer opportunities are great news for cable companies. It has been quite some time since there has been any significant new construction. However, this emerging generation of potential new customers do not have longstanding ties to major operators. The "triple play" has less value, voice service is often mobile-only, and viewing habits weigh the convenience of access and devices on par with, if not above, content volume. High speed Internet is a must-have, while hundreds of cable channels of programming is not – and cable voice service is even less so. This allows data-only services to be sufficient for these consumers, which is a business model a wider range of service providers can support than just large full-service providers.

In summary, the emergence of a "Go Gig or Go Home" mindset has arisen out of a combination of:

- Real-estate growth weighted towards MDUs and young adults
- High profile Fiber-to-the-Home (FTTH) deployments in North America
- New competition focused on an "easier" to deploy data services model
- Everyday lifestyles with heavy reliance on technology and Internet access
- Years of uninterrupted CAGR and rapid change in technology
- Desire of builders and owners of complexes to market competitively to potential residents

LAYERS OF THE GIGABIT ASK

The DOCSIS 3.0 standard was issued in August of 2006. The most significant component that this 4th version of the standard introduced was the concept of "channel bonding." Bonding allowed multiple 6 MHz channels in the downstream to be used as a single "wideband" channel from a user's perspective. Whereby a user might have been assigned one of four downstream CMTS channels in DOCSIS 2.0, DOCSIS 3.0 allowed a user to receive traffic on all four channels at once. Therefore, instead of the approximately 40 Mbps of downstream peak rate of a single channel, now N x 40 Mbps was possible, with N as high as 32. While the network itself does not change actual capacity without adding new

channels altogether, any single user could be allocated more capacity and also would be able to achieve much higher service tier speeds.

Bonding in the upstream was also introduced in DOCSIS 3.0. Upstream channel widths vary and the upstream is a time-division-multiple-access (TDMA) approach for each upstream carrier, so the specifics of how bonding is done are very different. However, conceptually, upstream bonding serves the same purpose as the downstream – more bandwidth can be provided to an individual user, and higher speeds are possible.

Not coincidentally, N = 32 downstream channels means that a full-size downstream bonding group is more than a Gigabit of IP/DOCSIS capacity. By 2006, cable had been operating DOCSIS networks for about a decade, and the compounding growth of data traffic through the early era of broadband had become well known and understood. The projected need for a Gigabit was clearly on the horizon, and operators reacted by ensuring that the standard for their latest version enabled bonding enough channels to reach Gigabit capability.

While the above described the demand-side expectation driving DOCSIS 3.0 parameters, another competitive driver at that time was the relatively recent introduction of FTTH services in North America using Broadband Passive Optical Networks (BPON) technology (622/155 Mbps), and a few years later using Gigabit Passive Optical Networks (GPON) technology (2.5/1.5 Gbps).

We're Heeeeeere

Now, in 2015, the Gigabit that was on the horizon in 2006 is now officially at the doorstep. The DOCSIS technology put in place at that time, and the emerging DOCSIS technology – DOCSIS 3.1 – has been and is still being evolved and deployed with a keen awareness of competitive capacity and speeds. Capacity, speed and also fiber connectivity are all component parts of today's Gigabit discussion. A convenient way to describe the "asks" of a service provider in this new age is to view modern requirements as "Layers" as shown in Table 1.

Table 1 – "Layers" of the Gigabit Ask

- Is it a Fiber Network ?
- Is it a 1 Gig Network ?
- Is there a Path to a 1 Gig Network ?
- Is it 1 Gig Speed ?
- Is it 1 Gig Symmetrical Speed ?

As cable operators bring DOCSIS 3.1 to life to deliver Gigabit speeds, fiber-based providers have campaigned to link Gigabit to fiber solutions. Because of cable's long history of deploying fiber and pulling it deeper and deeper into neighborhoods, *cable operators alone are well-position to deliver Gigabit solutions over either coaxial or fiber last miles.* Many cable operators today already deliver Gigabit services over point-to-point Ethernet connections using fiber to business customers.

Nonetheless, it is quite common for a Request for Information (RFI) from a potential customer to begin with a preference for a fiber-to-the-premises (FTTP) network. There are multiple technology approaches available to fulfill this as a requirement, and all are Gigabit capable.

Beyond simply the infrastructure to install, the complex ownerships or homeowner associations responsible for telecommunications services, per the second

and third bullets in Table 1, want to understand whether the proposed system is a Gigabit network, or has an evolution path to become Gigabit capable. Specific fiber and Gigabit interest began in earnest in 2013, as the economy emerged out of the long recession, hiring ramped up, and new home construction took a turn for the better. Indeed, new construction "Greenfield" accelerated throughout 2014 and is projected to continue to grow. As mentioned, this represents an exciting new opportunity for operators, as business growth associated with adding new customers had been difficult to obtain during difficult economic years.

More recently, "Gigabit" has evolved to the next bullet in Table 1, to a speed or a *service rate* of 1 Gbps. This is where the possibility of gamesmanship in advertisement and marketing, as well as legal language, can make it difficult to compare solutions and offerings apples-to-apples. This often has to do with how providers define Gigabit given that such a speed is impossible to achieve at the payload level from a GbE port or practical WiFi connection, and also that "proof" of achieving the Gigabit threshold is not uniform across providers.

Lastly, Gigabit service rates have migrated to a considering Gigabit *symmetrical* service rates.

GIGABIT ASK: SOLUTION OPTIONS

First and foremost, it is important to note that *DOCSIS 3.1 over HFC enables all of the Gigabit scenarios* in Table 1. It does not, in most typical single family neighborhood cases, meet the "all-fiber network" criteria, although it can. DOCSIS is typically delivered, of course, over an existing coax – to-the-home (CTTH) solution. The requirements set forth in the DOCSIS 3.1 standard were developed with the understanding that it would operate over existing HFC plant today without any changes required, and also that Gigabit services were a primary objective.

It is the case, and has always been the case, that the coaxial capacity is capable of such high capacity services [5]. This is because, historically, the HFC network was designed to meet requirements that support a high number of analog video services. Analog video is very sensitive to channel impairments, and as such network requirements demanded very high CNR, very low distortion, and very wide bandwidth because of the inefficient way that analog video uses spectrum. Therefore, the HFC network was made to be broadband, extremely high fidelity, and extremely linear – characteristics that modern communications systems can exploit to tremendous capacity effect. DOCSIS 3.1 is simply the first version of the standard to ensure that this latent capacity is exploited fully and for the purpose of Gigabit. Cable operators have invested wisely and in a timely manner to keep pace with the growth of capacity and demands for higher speed, and DOCSIS 3.1 is the latest example of this.

We will detail some common variations emerging for HFC network evolution in the next section. First, however, we consider the fiber-based possibilities associated with Gigabit.

Gigabit Multi-Path

DOSCIS 3.1 is not limited to coaxial deployments, as technology based on SCTE 174 2010 [7], commonly referred to as RF-Over-Glass (RFoG), enables DOCSIS over FTTH networks. This SCTE standard's key value is that it defines an all-fiber optical distribution network (ODN), but which is

consistent with current cable services, back-offices, and practices.

All-fiber networks can also be implemented with point-to-point Ethernet – as mentioned, this is common today in cable networks for commercial customers – as well as Passive Optical Network technologies such as those previously described. Many cable operators have based their PON architectures on Ethernet Passive Optical Network (EPON). While both PON approaches have merit, EPON has been augmented by standardization that allows compatibility with cable provisioning, as well as other features that attract cable providers. Both PON technology families are implemented worldwide.

For PON, some North American deployments are capable of Gigabit, but a large percentage of existing FTTH deployments use BPON technology. BPON, of course, is unable to deliver Gigabit capacity or speeds, as it is based on 622/155 Mbps total capacity. GPON and EPON families are available that support Gigabit, however.

Ethernet support for Gigabit services is enabled by dedicated Gigabit Ethernet circuits, typically on a DWDM network, and no or minimal aggregation on the network side, depending on specific service terms. It can also be enabled via 10 Gbe circuits that enable more sharing of the circuit/wavelength.

Table 2 summarizes the architecture capabilities of various operator options when considering the features described in Table 1. Rather than compare HFC and DOCSIS variants to dated options for PON technology of more limited capability, a comparison based on emerging 10G EPON technology is used since this is what DOCSIS 3.1 aims to enable.

Note that Fiber-to-the-Premises (FTTP) is used to accommodate both SDU (fiber-to-the-home) and MDUs (fiber-to-the-building) within a single table in Table 2, using the color coded checkmark identified in the note at the bottom of the table.

Table 2 – "Layers" of the Giga-Ask Versus Architecture Options

	HFC-Based				FTTP		
	Standard		Fiber Deep (N+0) DOCSIS 3.1		SCTE 174 2010 DOCSIS 3.1		PON
Network Feature ↓	**DOCSIS 3.0**	**DOCSIS 3.1**	**Mid-Split 85 MHz**	**High-Split 212 MHz**	**Mid-Split 85 MHz**	**High-Split 212 MHz**	**10G EPON**
Fiber			√	√	√	√	√
Gbps Network	√	√	√	√	√	√	√
Fiber AND Gbps Network			√	√	√	√	√
1 Gbps Service		√	√	√	√	√	√
Symmetric 300 Mbps Service			√	√	√	√	√
Symmetric 1 Gbps Service				√		√	√

√ = Applies to MDU Only

HFC Evolutionary Variants

As mentioned in the previous section, there are several evolutionary architectures for HFC, some of which are called out in Table 2 and described in more detail below.

The "Fiber Deep" approach, which can be thought of as "Fiber Deepest" in that it is a passive coax last mile following the amplifier, is an approach described in [5] based on striking a balance between optimizing an N+0 network design for maximum coverage (hhp), but with considerations to the existing installed fiber, coax, and powering infrastructure. The core principles of this approach are shown in Figure 2.

Fiber deep conveniently addresses one of the nagging obstacles to service evolution of classic HFC – the currently defined boundary of allocated upstream spectrum. DOCSIS 3.1 itself – without new consideration of upstream spectrum – will add significant new capacity potential to the HFC upstream in two ways.

First, as in the downstream, the DOCSIS 3.1 standard calls out more spectrally efficient QAM formats. The increased spectral efficiency attributable to this alone is 66-100%.

Second, the DOCSIS 3.1 signal format based on OFDM will enable efficient use of formerly difficult spectrum below 20 MHz, which standard QAM/TDMA tools often found impossible to traverse. When combined, more than a doubling of upstream capacity is expected from these two components.

Upstream CAGR has historically been inconsistent, but generally significantly lower than the downstream. Using TDPs, for example, and a 20% upstream CAGR on a healthy upstream network today, upgrading

to DOCSIS 3.1 and implementing a node split would deliver capacity that lasts 4 + 4 = 8 years, where 20% CAGR means a doubling of traffic after four years.

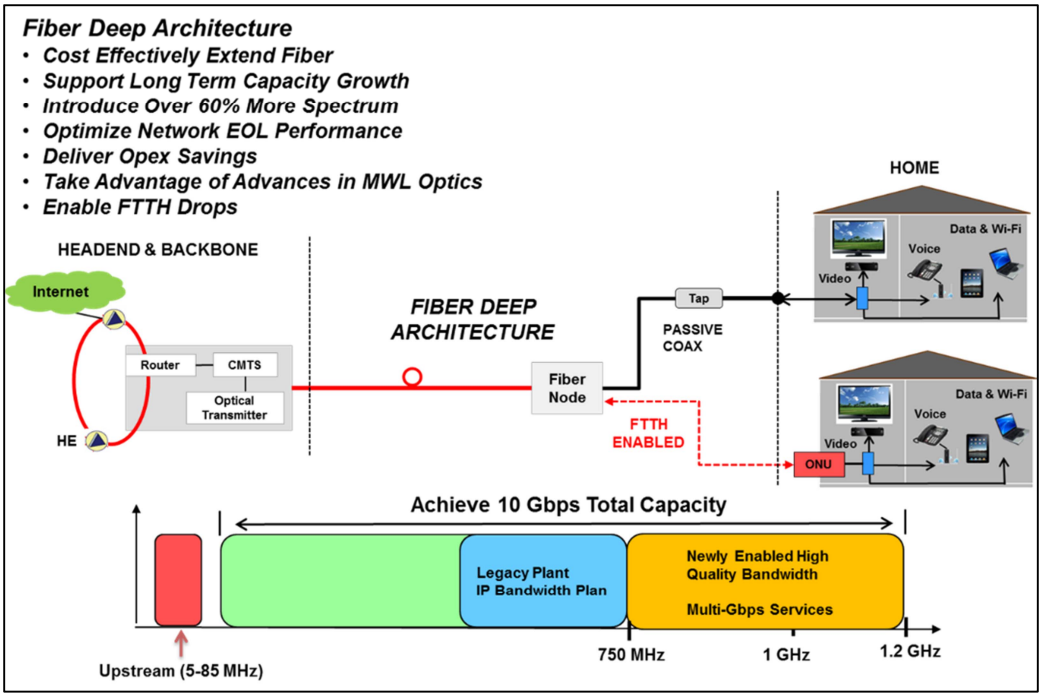


Figure 2 – Fiber Deep Principles [5]

Nonetheless, under continued growth of data traffic, the 42 MHz allocation may become insufficient over time, and perhaps more importantly it places speed limits on the upstream service tiers that cannot be solved by node splitting alone. When this spectrum becomes insufficient is relatively predictable as shown in the coarse TDP analysis above, but with the uncertainty that comes with what has been inconsistent upstream traffic growth year-on-year. The "Napster Moment" in the early 2000's, ushering in peer-to-peer file sharing services, is the reference example for a step function upstream growth scenario, but it has not been repeated since.

Figure 3 is a an example of analysis of upstream capacity [5] that represent the information content of Figure 1 in a different format, recognizing that a fixed CAGR analysis such as is shown for the downstream is less insightful for the upstream.

Returning to the topic of upstream spectrum allocation, the fiber deep approach, because it eliminates amplifiers, makes the expansion of the return to 85 MHz relatively simple in the network itself. Installation of an 85 MHz capable node is all that is required in the plant. There are other elements of the conversion (spectrum clearing of 54-108 MHz, home architecture, CPE out-of-band compatibility [OOB], HE signaling), but none that are technology barriers.

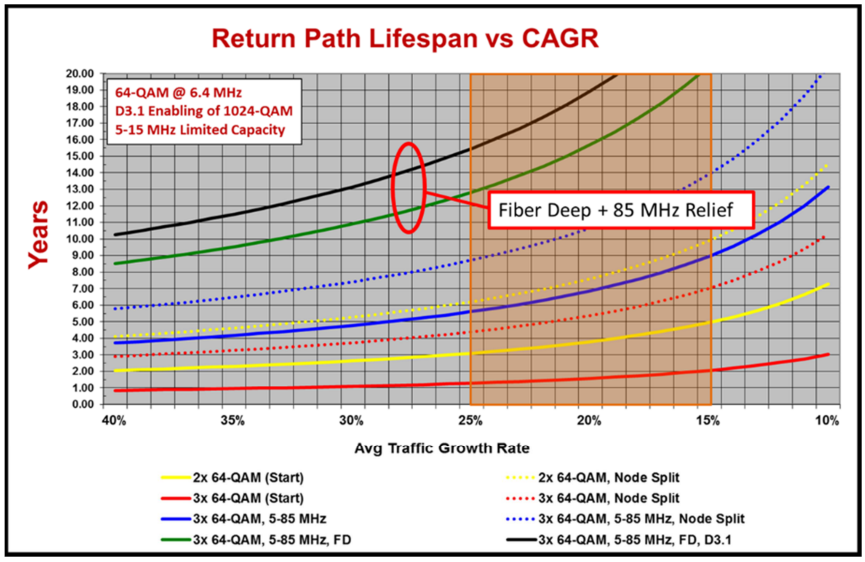


Figure 3 – Comparison of Options for Managing Upstream Growth

What 85 MHz provides is Nx100 Mbps services (i.e. 100 Mbps, 200 Mbps, and 300 Mbps) while allowing the continued support of the vast majority of STBs in the field that use a legacy narrowband downstream OOB channel for STB communications, such as VOD. What the 85 MHz return does *not* provide is a path to Gigabit capability in the upstream. It is simply not enough spectrum to do that. However, this is specifically why the DOCSIS 3.1 standard calls also for the possibility of a *higher* bandwidth return, with an upper edge of 212 MHz. This bandwidth and the corresponding definition of the OFDM and underlying QAM parameters enables Gigabit services. In the case of the high split, it comes with the trade-off of the loss of significantly more downstream bandwidth, as well as the inability to easily support OOB communications for most legacy STBs in the field.

In summary then, DOCSIS 3.1 with an 85 MHz mid-split enables asymmetrical services with Gigabit or multi-Gigabit downstream. It does so while preserving legacy video services, and with a very modest impact to downstream spectrum. The Internet service tiers for this case could be, for example 1G/300M (asymmetrical). With a 212 MHz upstream split, DOCSIS 3.1 will enable symmetrical Gigabit speeds, as well as asymmetrical services such as 1G/300M or 2G/1G. Table 2 summarizes the service coverage for each of these permutations of network and spectrum for different access architectures.

Revisiting Spectrum Allocation

While we have focused on the upstream implications above on the way to symmetrical Gigabit, note that the Fiber Deep approach enables extended downstream possibilities as well. By defining a single active driving the last mile coax, the potential to extend the downstream to 1.218 GHz per the DOCSIS 3.1 standard also exists. In so doing, DOCSIS 3.1 technology can deliver up to 10 Gbps of downstream capacity while simultaneously delivering over 2 Gbps of upstream capacity.

An element of the Fiber Deep architecture that ensures that the "up to" 10 Gbps and beyond is enabled fully and robustly is the emerging architecture approach based on distributed DOCSIS technology. Similar to how telco providers push digital fiber deep to squeeze the most capacity possible out of the very limited twisted pair last mile, cable operators can implement a logically similar architecture [4]. In the case of telcos, limiting the distance from the DSL Access Multiplexer (DSLAM) to the home was essential for reasonably useful broadband capacity, and they have been forced to continue to get closer to the home as part of the strategy to keep pace with CAGR.

By contrast, cable operators do not have such a prevalent last mile bandwidth constraint to overcome. Instead, however, by deploying distributed systems based on digital fiber, the end-of-line (EOL) SNR performance at the home is optimized. Linear fiber optics and RF amplifiers contribute SNR loss from the Headend CMTS transmission point to cable modem receiver in the home, and also from the home cable modem transmitter upstream to the HE CMTS. A Distributed Architecture with a digital optical basis, of which there are several possibilities, and a Fiber Deep system that eliminates amplifiers remove both of these RF channel performance degradations. The maximum DOCSSI 3.1 QAM profiles – the most spectrally efficient "MUST" format being 4096-QAM – are thereby most robustly enabled.

The 1.218 GHz extension also plays the important role of "replacing" the bandwidth that is lost to the downstream by a 212 MHz upstream high split. Of course, only DOCSIS 3.1 can use this bandwidth above 1 GHz. In this way, the 212 MHz architecture is connected to the all-IP transition based on DOCSIS 3.1 as cable's IP pipe. Looked at another way, even in traditional HFC networks (750 MHz, 860 MHz, or 1 GHz), which will exist for many, many years, use of the 212 MHz upstream bandwidth will *essentially* be tied to a full IP conversion – one that implements the IP delivery but that also retires QAM delivery. The latter removes the video simulcast that inefficiently uses spectrum, and allows the full all-IP conversion to take place.

CONTEMPORARY TRAFFIC SNAPSHOTS

"Who needs a Gigabit?" is a common question as the market matures around Gigabit solutions and consumers look for the value that the extreme speeds bring – asking themselves what is it that they actually get for the price. No doubt, one of the more popular applications initially will be the speed test itself, but consumers engaged in large peer-to-peer file sharing and commercial services customers are the most likely to notice the benefits of Gigabit in comparison to, for example, more common peak speeds such as 50 Mbps or 100 Mbps.

Commercial enterprises commonly have more symmetric demands, since the majority of the traffic is email, data, and file both to and from the enterprise. The size of the enterprise sets the bandwidth required for the enterprise to efficiently execute its business.

The residential space, however, is less likely to have symmetric demands. The file sharing user case is the nearest exception. In relative terms, it is how the early "Napster" days were turbo charged by broadband speeds for what was, at that time, large music files exchanged over dial-up connections. Over time, it is anticipated that perhaps the Internet of Things (IoT) and the associated

machine-to-machine (M2M) traffic may swing the pendulum back towards more symmetry of consumer traffic.

Gigafy Me

Outside of the measureable time savings that Gigabit may bring to an HD movie transfer, applications for Gigabit today are limited. Furthermore, with most consumer's receiving their data over WiFi, the bottleneck to the client device is often the wireless connectivity of the Home LAN. That is, even should a Gigabit get to the home, getting that Gigabit throughout the home has other constraints to consider.

Not signing on the Gigabit tier, and even the impacts of localized bottlenecks should not affect over the top (OTT) streaming video services in a properly engineered network if the wireless network is properly within range. Table 3 shows the range of average video rates as a function of video format and encoding technology [1].

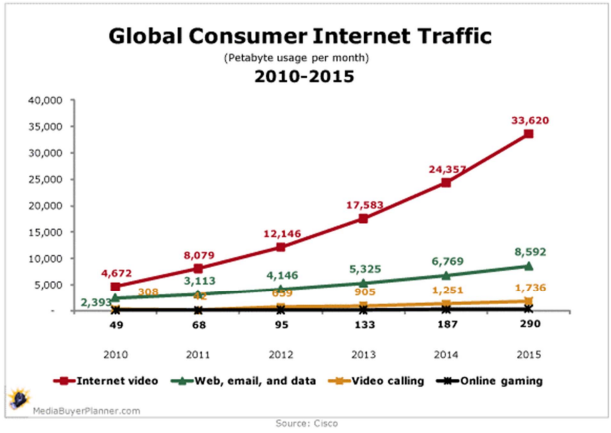
Table 3 – Video Streaming Rates

Resolution		Digital Compression Method and Bit Rate		
Resolution Terms	Frame Size / Scanning System / Frame Rate	MPEG2	H.264	HEVC
SDTV	480 / i / 30	3.7 Mbps	2 Mbps	1 Mbps
HDTV	720 / p / 30	6 Mbps	3 Mbps	1.5 Mbps
HDTV	720 / p / 60 (or 1080 / i /60)	12 Mbps	6 Mbps	3 Mbps
HDTV	1080 / p / 60	20 Mbps	10 Mbps	5 Mbps
4K UHDTV	4Kx2K / p / 60	80 Mbps	40 Mbps	20 Mbps
8K UHDTV	8Kx4K/ p / 60	320 Mbps	160 Mbps	80 Mbps

While volumes of streaming video from cable operators themselves, or from OTT providers have had a significant role in keep the CAGR train moving forward at a 50% per year clip for new capacity, even 1080p HD streams based on MPEG-4 (H.264) encoding are on the order of 10 Mbps average bit rate. Multiply this by 2-3x for action-packed movies during peak bursting periods. Thus, Gigabit service is not necessary for supporting a high quality streaming video experience. However, what is very critical for streaming services, because it is sustained rates that must be met for long periods of time (versus web browsing sessions, for example) is that operators have allocated sufficient capacity to handle the growing number of streaming users at peak busy hour (pbh). Operators that offer Gigabit service should have traffic engineered the network around well-understood rules guiding the distribution of peak speeds, concurrency of use, and total capacity. This fundamental traffic engineering tenet applies whether the capacity or speed number is followed by a "G" or an "M."

Figure 4 shows how the growth of streaming video has come to dominate the traffic demand [2,3]. The top figure shows actual data through 2013 and projections through 2105, while the bottom extends the projection through 2018, with the addition of file sharing. The file sharing application is insightful in that it represents an application associated with potential high growth upstream traffic, as common email and web-surfing applications are limited in upstream demand.

What is clear from the bottom chart of Figure 4 is that streaming video is expected to grow at a more rapid pace than file sharing or normal web use. As such, while market messages around symmetrical data services are growing in the residential space, the trends are headed to increased asymmetry of actual demand.



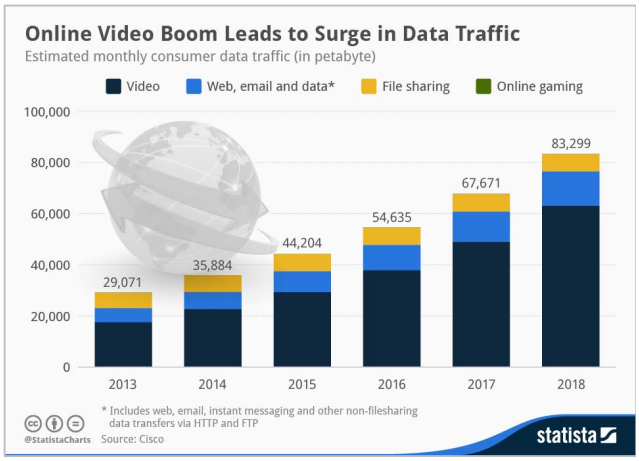


Figure 4 – Downstream Traffic Growth is Dominating Recent Trends [2,3]

Figure 5 shows the state of modern consumer traffic imbalance over the course of 24 hours. It is consistent at peak busy hour (pbh) with Figure 4 where video streaming tends to emerge as the dominant application in "primetime." Network engineering practices are built around managing through the peak periods. As noted in [1], in the early 2000's, as peer-to-peer file sharing became a dominant application of the time, asymmetry ratios during peak periods were closer to 2:1.

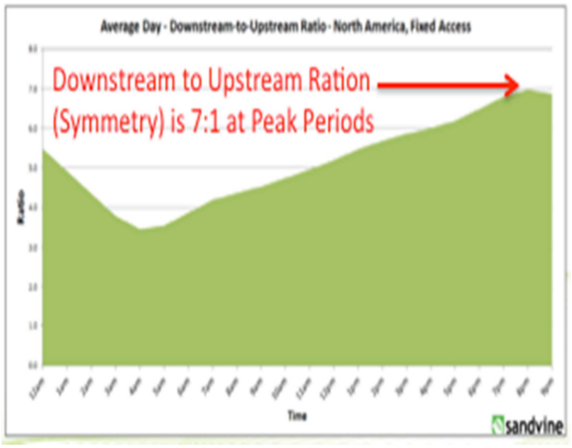


Figure 5 – Traffic Asymmetry has Increased with Streaming Video

STRIKING THE RIGHT TONE

While Gigabit is an important and relevant 2015 buzzword storming the marketplace, it is important to remember that cable operators foresaw this as far back as 2006 with DOCSIS 3.0, and upped the ante further in developing the DOCSIS 3.1 specification which began in 2012, with deployment beginning in 2016. Furthermore, DOCSIS 3.1 anticipated a multi-gigabit era in developing the requirements, ensuring that 10 Gbps of downstream capacity was possible on the coax in an all-DOCSIS 3.1 system. Even in the face of continued data growth, this creates a long-term, sustainable network, whose power is multiplied by the additional levers of spectrum re-allocation, business-as-usual service group splitting, distributed DOCSIS architectures, and an all-IP transition.

Cable operators have consistently raised capacity and speeds as demand has continued to grow. This is testament to the flexibility of the HFC architecture, since cable is well into the second decade of DOCSIS data

services which, all the while, has undergone compounding data growth year on year. It is important to articulate and balance the excitement of the Gigabit momentum with the reality of practical data demand downstream and upstream, and the proven ability that cable has shown in keeping ahead of this demand for the entire life of the broadband Internet.

It is rarely a good idea to be debating whether there is a need for higher bandwidth from the side of the more conservative perspective. The archives are littered with quotes of technology leaders of the past whose predictions are unattractively juxtaposed against the reality that has come to pass. However, the volumes of extremely granular data that operators have on actual usage and trends, and the track record of cable operators to meet the demand as it arises, lend unmatched credibility to matching broadband bandwidth and speeds to solutions that satisfy subscribers.

Importantly, while marketing and competition has affected perceptions around the capability of cable networks, it is clearly demonstrable that "Gigabit" is not a number tied to fiber-based access. The notion that fiber is necessary for Gigabit services is pure mythology. Indeed, even today's 750 MHz digital cable systems are delivering over 4 Gbps of traffic all day, every day. The allocation of those bits has simply weighted to video services historically, and mostly broadcast video services. That mix is constantly changing as both data technology and video technology become more efficient, and more and more services migrate to IP. It will be demonstrably and explicitly true that coaxial infrastructure enables Gigabit when DOCSIS 3.1 hits the field in 2016.

CONCLUSION

Gigabit is the latest target in the continuing race to the highest "billboard" speed. Applications for Gigabit today may be limited, but if the history of data growth is any guide, it will not be too long before "multi-Gigabit" is the word of the day, and eventually Gigabit becomes insufficient. CAGR simply always wins out to a fixed target if it persists unabated.

Talking about Gigabit now perhaps seems to be a massive leap forward, in part because we are officially leaving the "Megabit" era, and the shift in language can be imposing. However, DOCSIS capacity and speeds have been steadily marching forward year after year, keeping pace with the compounding way that demand has behaved, and since inception of the service. Gigabit is simply the next, logical emerging threshold. It is perhaps a bit ahead of its projected timeline based on CAGR because of market forces and the aforementioned language threshold, but with the persistent CAGR at 50%, this modest acceleration of timeline does not significantly alter the continuing evolution challenge cable operators continue to face head-on.

Not surprisingly given the flexibility of HFC, there are a range of options that meet the criteria of Gigabit, including media-specific requirements such as fiber only – something common in new developments – or coax only, which is often preferred in existing MDUs in order to the major intrusiveness that building rewiring can entail.

All in all, Lady Giga is not an intimidating presence to cable operators. Cable, in fact, is an ideal dance partner, with its message on Gigabit crystal clear:

- Speed is nothing new to cable. This is the industry that invented broadband Internet and put the "High Speed" in high speed Internet, displacing dial-up and racing past DSL.
- Gig is just the next speed tier that the 5th iteration of DOCSIS will accomplish.
- Cable operators have increased speeds since the inception of DOCSIS, at every turn meeting customer demand for increased bandwidth. "Gigabit" does not change this dynamic.
- Cable operators can deliver Gigabit over coax OR over fiber, and no other provider has this flexibility
- Cable operators are delivering Gigabit via FTTP today to thousands of businesses
- Cable operators are carrying multi-Gigabit to the home everyplace an all-digital system operates today
- Quality of Experience is more than just the speed to the home, it is also the speed to the devices within the home and access outside the home. Operators deploy leading edge WiFi, and taking advantage of a powered access network to deliver coverage outside the home

Cable welcomes Lady Giga to the stage. She may sound shiny and new, but for cable operators with 15+ years of DOCSIS and HIS under their belt, she is really just an old friend wearing a new dress.

REFERENCES

[1] Emmendorfer, Mike and Tom Cloonan, "Nielsen's Law vs. Nielsen TV Viewership fo Network Capacity Planning," Cable Show Spring Technical Forum, Los Angeles, CA, Apr 29-May 1, 2014.

[2]http://blog.getresponse.com/use-facebooks-new-video-features.html

[3]http://ipcarrier.blogspot.com/2014/02/asymmetrical-traffic-dominates-networks.html

[4] Howald, Robert, "Fueling the Coaxial Last Mile," SCTE Conference on Emerging Technologies, Washington, DC, April 3, 2009.

[5] Howald, Robert, "Growth Architectures: Built to Last, Built to Launch," Cable Show Spring Technical Forum, Los Angeles, CA, Apr 29-May 1, 2014.

[6]https://www.nmhc.org/Content.aspx?id=5870

[7] SCTE 174 2010, "Radio Frequency Over Glass Fiber-to-the-Home Specification"

LESSONS FROM TELCO & WIRELESS PROVIDERS: EXTENDING THE LIFE OF THE HFC PLANT WITH NEW TECHNOLOGIES

Tom Cloonan, Ayham Al-Banna, Mike Emmendorfer,
Zoran Maricevic, Frank O'Keeffe, John Ulm
ARRIS Group, Inc.

Abstract

This paper draws on lessons from the past (within the telecommunications space) to predict some of the new technologies that may be considered by Multiple System Operators (MSOs) as they move forward into a service provider world of the future. It is a future that will undoubtedly demand more and more bandwidth to be offered to subscribers over time.

After presenting some historical data on the evolution of telecommunications systems and some new traffic engineering information on bandwidth growth trends, the paper will attempt to identify the life-span of the current Hybrid Fiber/Coax (HFC) infrastructure through which most Voice, Video, and Data services will be provided in the future. Potential techniques for extending that life-span will also be explored. It will be shown that with appropriate management, the life-span of the current HFC network can likely be extended well into the 2030s or 2040s (or beyond).

The paper then explores some of the key technologies that may be utilized in concert with or in lieu of the existing HFC network during this period.

Most of the predictions in this paper will draw heavily on the historical lessons that can be learned from the network evolutions that have taken place within the Telco & Wireless industry during the past two decades. The paper will attempt to show why these lessons may be applicable to the future evolutions that are likely to take place in the Cable industry.

BACKGROUND

Introduction

Correctly predicting the future is a difficult, but critical task for any company. This is especially true when an industry is facing a time when transitions in technologies are being considered- with the potential end-of-life for one technology approaching and prospects for the birth of a new technology looming in the foreseeable future. Are there ways to extend the life of the existing technology? Are those extensions beneficial or not? If a transition is to take place, when should it take place? Should the transition be done quickly or gradually? Which technologies should be used during the transition? Which of many available technologies should carry the load in the future? All difficult questions that need to be answered as any industry approaches a "technology transition window."

Some in the Cable industry have argued that the HFC network that provides the access network backbone for MSOs is slowly beginning to approach one of those "technology transition windows." Some believe that the bandwidth demands of subscribers in the near future may exceed the capacities of their existing HFC infrastructure, and they are trying to prepare for that eventual occurrence.

Are these arguments that the HFC plant is approaching obsolescence correct? If so, will the HFC network need to be replaced in the 2010s? In the 2020s? In the 2030s? In the 2040s? Beyond that date? Should a transition to a new technology occur after DOCSIS 3.1

and Distributed Access Architecture deployments, or should the new technologies supplant DOCSIS 3.1 and Distributed Access Architectures. Should the new technologies be introduced in parallel with existing technologies? How should the transition be orchestrated- quickly or gradually? What technology or group of technologies will replace the current technologies? Will it be Passive Optical Networks (PONs) or Radio Frequency over Glass (RFoG) or Point-to-Point Ethernet? Will it be something else?

As stated above, these are all difficult questions to answer, but answering them in the right way is critical for every MSO. And it should not be shocking if different MSOs answer these questions differently and select different paths, because in many cases, the "correct" answer is heavily dependent on many different and interesting factors.

1) What is the starting point- i.e., what is the status of the MSO's current HFC infrastructure?
2) What is the desired ending point - i.e., what technologies does the MSO wish to use in the future?
3) How quickly can they transition? (Note: This usually becomes an economic question, requiring the MSO to perform a Business Case Analysis on the various transition plans).
4) What are the capabilities and costs of the different technologies under study?
5) What improvements are expected in the different technologies under study?

The authors will explore many of these topics and attempt to make predictions within this paper. However, it is clear that any predictions on the future require some amount of information to help guide those predictions. The authors decided to do what many researchers have done to predict the future, by looking at the past. Perhaps answers to these questions can be found by looking to history and exploring similar evolutions in history to see how they played out.

But are there any examples in history where large service providers using copper-based technologies found themselves at a point where bandwidth requirements of the future looked like they would stress the capabilities of their copper-based technology? If so, is there anything to be learned from the manner in which they dealt with the situation. The authors believe that the answer to both of these questions is yes. And because of that, the paper will now take a brief look at the recent history of the Telco & Wireless industry.

TELCO HISTORIES

A Brief History Of The Telco Industry's Voiceband Modem Evolution (1950's to 1998)

The wireline telephone systems of today's Public Switched Telephone Networks (PSTN) have deep roots in history. They are actually close descendants of the telegraph systems that also used wires to transmit signals across long distances in the early-to-mid-1800s.

Throughout the mid-1800s, many researchers were looking at new technologies that would ultimately be useful for telephone transmissions of voice signals across a wire. With the granting of patent number 174,765 by the U.S. Patent Office on March 7, 1876, Alexander Graham Bell's ideas were pushed forward as a seminal approach to the transmission of voice signals over wires. The first overhead telephone lines were set up between houses in Boston using iron and steel wires with an earth (ground) return. The use of copper wires for the telephone transmission became possible in 1877 with Thomas Doolittle's invention of a process to hard-draw copper into durable wires. From that point on, copper wires were used for most telephone systems. Alexander Graham Bell patented two-wire, twisted-pair circuits in

1881, leading to lower-noise transmissions. This permitted long-distance copper lines to be used to connect New York to Boston in 1884 and New York to Philadelphia in 1885. Bundling of these two-wire circuits inside of cables to carry many parallel lines became common-place by 1890. By 1900, most of the U.S. telephone system was connected using twisted-pair cables. Multiplexing of multiple phone conversations on a pair of wires, improving the insulation, and improving the cable sheath became the focus of many innovations during the 1880, 1890, 1900, and 1910 decades. While many other improvements have been made since then, it is interesting to note that (for all practical purposes) telephone wires between the telephone central office and the home have changed very little in the past 100+ years. (Note: By comparison, the higher-performance coaxial cables used within the Cable industry were only invented in 1942, so the Cable industry's coax will likely have a much longer life ahead of it if it matches the total lifespan experienced by twisted pair wires. One may wonder why it shouldn't experience a similarly long life with its increased performance levels relative to twisted pair). [COPP] [EASY]

The two-wire copper wires that make up the copper loop for Telcos have thus been used to transport voice signals for more than 100 years. However, a new type of signal began to appear on these copper wires in the 1950s (~60 years ago). It was during that decade that the U.S. Air Force's Ballistic Missile Early Warning System (later called SAGE) began to use telephone lines to transmit signals between radar stations in Canada and IBM 790 computers in the U.S. These signals needed to have spectra that could be carried by the traditional voiceband spectrum of 0-3.3 kHz (the passband of the telephone lines). As a result, modems (modulator-demodulator systems) were used to convert binary information from the computers into analog signals with appropriate amplitude and frequency characteristics. These original Air Force modems could transmit at data rates of 75 bits per second, and eventually achieved speeds of 750 bits per second. They were predominantly unidirectional or half-duplex transmissions, though.

In 1958, the bi-directional modem concept was commercialized by AT&T with the creation of the Bell 103 modem. This modem was based on Frequency-Shift-Keying (FSK) and permitted full-duplex transmission (the transport of signals in both directions at the same time). It also permitted a data rate of 300 bits per second (which was considered adequate for business and scientific applications of the time).

As researchers at universities began exchanging data between computers, the demand for modems grew. The demand for more modem bandwidth also began to grow. AT&T soon developed the Data Set 202 modem, which provided a higher data rate of 1200 bits per second. In 1962, they released the Data Set 201 modem with 2400 bit per second performance. (Note: The Data Set 201 modem was required to work on specially-conditioned leased or private lines). In 1965, Robert Lucky (of AT&T Bell Labs) invented adaptive equalization, which helped mitigate against channel frequency response issues. In 1968, a non-AT&T company called Milgo released the 4400/48 modem that could operate using proprietary protocols at 4800 bits per second over minimally-conditioned telephone lines. In 1971, Codex added many innovations using suppressed-carrier single side-band modulation with their 9600 modem that could operate at 9600 bits per second. That was followed by the world's first PC-based modem (the 300 bps 80-103A) was created by Dale Heatherington and Dennis Hayes in 1977.

Improvements continued throughout the next decade, and standardization became more

desirable over time so that modems from different vendors could communicate with one another. This standardization effort became especially important as more home-based subscribers began using modem technology with Tim Berners-Lee's introduction of the three fundamental World Wide Web protocols (HTML, URI's, and HTTP) in 1989. [WEBF]

By 1991, the ITU-T V.32 and V.32bis standards were created, permitting modem transmissions at data rates of 9600 bits per second and 14,400 bits per second, respectively. These standards made heavy use of new technologies such as echo cancellation and Trellis-Coded Modulation (TCM). Then the ITU-T V.34 standard created modems that could support data rates of 33,600 bits per second by capitalizing on silicon technology improvements and measuring the characteristics of the telephone line and tuning transmission parameters.

In the 1998-1999 timeframe, the ITU-T committees were putting the finishing touches on a new set of standards called ITU-T V.90 and ITU-T V.92. These two specifications utilized an incredible number of recent innovations in communication theory (such as Quadrature Amplitude Modulation (QAM), pre-coding, and spectral shaping), and the specifications were to be based on the best technology available at the time. The specifications drew heavily from two competing and proprietary modem designs that had been created a few years earlier. These included the K56flex proposal from Rockwell and Lucent and Motorola and the X2 proposal from US Robotics. V.90 modems would permit download speeds of up to 56,000 bits per second and upload speeds of up to 33,600 bits per second. V.92 modems would permit download speeds of up to 56,000 bits per second and upload speeds of up to 48,000 bits per second.

These incredibly high data rates supported across the analog wires propelled the speeds to values that would come close to the digital bandwidth capacity available on the digital DS-0 circuits that were utilized in the trunk connections between digital switching systems in different telephone central offices. Because of this and because of the complexity of the resulting modem chips, the V.90 and V.92 modem technology was also called "V.Last". This nickname was used because many of the researchers associated with the telephone line modem technologies believed that V.90 and V.92 might be the end of the line for the (then) forty-year-old modem technology improvements that had roughly begun back in 1958 with the Bell 103 modem definition. Many researchers believed that for the support of future broadband data services, copper telephone lines were essentially becoming obsolete. At that particular point in time, their view of the evolution of modem technology on copper telephone lines was probably driven strongly by the chart shown in Fig. 1a. There were two data rate growth spurts, with a long 20-year period of relative stagnation (from 1971 to 1991) between the growth spurts. Most were probably preparing for another period of stagnation. The only research looking at increasing the data-rate on the copper telephone lines proposed moving the transmission spectrum above the 3.9 kHz voiceband, and many researchers believed that the transfer functions of the transmission media above that voiceband limit were challenging at best. There would be spectral regions filled with amplitude and phase variations that would greatly distort any high data-rate signals. Voiceband modem designers were constrained in both power and frequency levels (as shown in Fig. 1b)- with power limits from the FCC and codecs and with frequency limits from codecs and ring detectors. This led to a belief that spectral regions above 3.9 kHz on copper lines were a "No-Man's Land" for broadband services, as

shown in Fig. 1b. [BRIT] [MODE] [NETH] [HIST] [DIAL] [56KB]

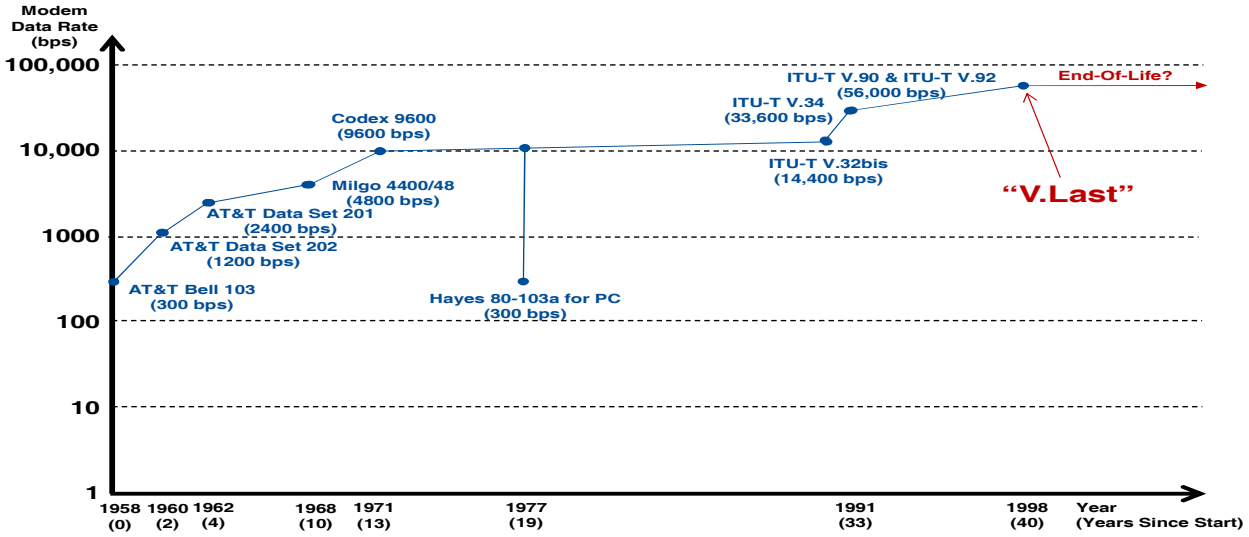


Fig. 1a- Telephone Line Voiceband Modem Data Rates vs Time

Due to these concerns about the end-of-life of broadband data services on copper telephone lines, many Telco executives and researchers began to look at new technologies that might be able to deliver the expected data rates of the future. The Telco industry had entered a window of time that was labeled as the "technology transition window" earlier in this paper. The entire Telco industry needed to develop a new game plan to ensure that they would not become irrelevant as an Internet Service Provider going forward into the future.

The steps and decisions made by various Telcos between 1998 and the present time will be the key focus of the remainder of this paper, because it is the belief of the authors that the Cable industry can learn from their successes and failures during their "technology transition window."

During the past twenty years, Telcos have been undergoing a lengthy transition in technologies, as they have slowly migrated from copper loop plants to fiber and/or wireless infrastructures. The decisions that they made during that twenty-year journey

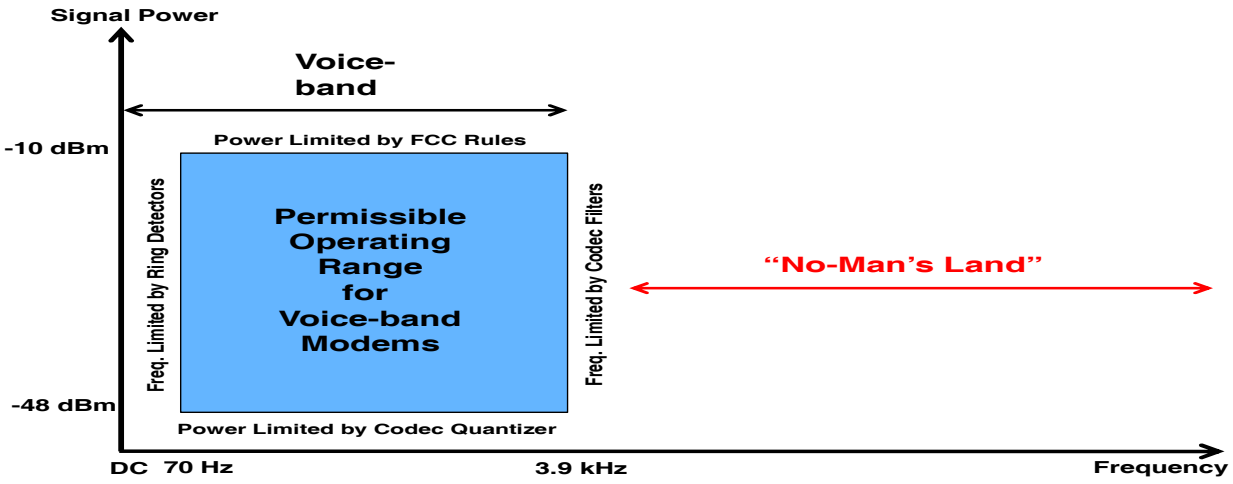


Fig. 1b- Old Views on Useable Spectrum for Copper Twisted Pair

may be instructive to MSOs as the Cable Industry begins to wrestle with questions about the longevity of its own existing technology infrastructure: the HFC network. The paper will begin by briefly looking at what Telcos have done between 1998 and the present time.

The Challenges That Faced The Telco Industry in 1998

The Telco Industry found itself in a complex situation during the 1998 time-frame. Many within the Telco industry believed that modem performance levels over copper lines may not see data rate improvements that would keep up with the bandwidth demands of subscribers. At the same time, the Telco Industry was encountering a new Service Provider competitor in the form of Cable MSOs. With the release of the DOCSIS 1.0 specification in 1997 and the first CableLabs certification of modems and CMTSs in 1999, high-speed data services with shared 30 Mbps channels (over HFC plants) were becoming a challenging competitor to telephony modem technology (which at the time, appeared to be limited to much lower ~56 kbps service levels).

Because of these challenging trends, Telco executives needed to re-invent their companies and had to make some important decisions on how to move forward. Although each operator might have followed slightly different game plans with slightly different technologies and slightly different schedules, the Telco industry (in general) could be said to have followed a three-pronged path. They ultimately decided to invest in three different last-mile technology paths for broadband data services going forward. Those paths included:

a) Broadband data services over fiber-based Passive Optical Networks (PONs)
b) Broadband data services over the growing Wireless (Cellular) infrastructure
c) Broadband data services using newly-evolving Digital Subscriber Line (DSL) technology running over the traditional copper telephone lines.

Between 1998 and the present time, each of these three Telco technology areas saw impressive change and growth. These changes are briefly outlined in the following three sections.

A Brief History Of The Telco Industry's PON Evolution (1998 to Present)

In the 1990s, PON was a relatively new point-to-multipoint fiber-based technology that was beginning to see the light of day- it was coming out of the labs and moving into early products and deployments. This timing was fortuitous for many Telcos who were hunting for their next-generation access technologies that might replace copper wire. PON became one of the key areas of focus for those involved in replacement activities. The evolution of PON was helped by early standards work. But it was somewhat slowed by a battle between two different standards efforts (one in ITU-T, and another in IEEE).

One of these standards efforts was originated by an organization of mostly Telco service providers and vendors. This group was called the Full Service Access Network (FSAN) working group, and they then partnered with the ITU-T organization to publish the G.983.1 spec on January 13, 1995. It described an Asynchronous Transfer Mode-based (ATM-based) transport system over fiber, and it was called ATM PON or APON. This APON system permitted operation of a single PON over ~20 km with up to 32 Remote Nodes. It supported 155 Mbps service in the downstream and 155 Mbps service in the upstream. On a single fiber, the APON would operate at 1.55 um wavelengths for

downstreams and 1.3 um wavelengths for upstreams. Dual fiber systems were also permitted and would operate at 1.3 um wavelengths for both downstream and upstream transmissions. [OPTI]

With the failure of ATM to gain widespread market acceptance, the specification was quickly modified to create a more Ethernet-friendly version known as Broadband PON or BPON. It supported both 155 Mbps service and 622 Mbps service in the downstream and 155 Mbps service in the upstream. When shared across 32 Remote Nodes, this could yield an average bandwidth of 155 Mbps/32 = 4.8 Mbps per subscriber (a bandwidth that was 85 times greater than that which was achievable by copper wire modem technology). [FREN]

FSAN and ITU-T then began work on a new specification aimed directly at providing service for the highly-coveted 1+ Gbps market. This specification (known as G.984.x or GPON) began development in 2001 and produced final documents in the 2003/2004 time-frame. This permitted first interops of equipment to occur in January of 2006 and first equipment deployments to occur in 4Q07. The GPON specification permitted a 20 km physical reach (60 km with special optics) out to 128 Remote Nodes, and it supported 622 Mbps downstream/622 Mbps upstream operation, 1.25 Gbps downstream/1.25 Gbps upstream operation, and 2.5 Gbps downstream/1.25 Gbps upstream operation. On a single fiber, the GPON would operate at 1.48-1.5 um wavelengths for downstreams and 1.29-1.33 um wavelengths for upstreams. Heavy deployments of GPON began in 2008 and 2009. [TELC]

In 2005, the FSAN and ITU-T began work on another new specification aimed at providing 10 Gbps service. The first specification to result from this effort was the ITU-T G.987 or XGPON1specification. XGPON1was released in the 2009/2010 time-frame, and it permitted a 20 km physical reach (40 km with special optics) to 128 Remote Nodes, and it supported 10.3125 Gbps downstream/1.25 Gbps upstream operation. On a single fiber, the XGPON1 would operate at 1.575-1.58 um wavelengths for downstreams and 1.26-1.28 um wavelengths for upstreams. [HOOD]

XGPON1only gained small market acceptance, so efforts quickly began within FSAN and ITU-T on its successor specification, which is known as ITU-T G-989 or NG-PON2. The first release of the first document for this specification occurred in 2013, and releases of other documents are expected in the future. NG-PON2 supports many different modes of operation based on Time and Wavelength Division Multiplexing (TWDM). Typical modes of operation in early deployments may be supporting 9.953 Gbps downstream/9.953 Gbps upstream operation. However, future versions (based on WDM) may support 40 Gbps downstream/9.953 Gbps upstream operation. It is designed to support up to 256 Remote Nodes, but normal operation will likely see 32-64 Remote Nodes in service. Reaches may range from 20 km to 60 km, depending on split ratios. On a single fiber, the NG-PON2 would operate at 1.596-1.603 um wavelengths for downstreams and 1.524-1.544 um wavelengths for upstreams. These systems are not yet available. [G989]

In parallel with the FSAN and ITU-T specification efforts, the IEEE has a similar effort underway to create a parallel set of standards which are unfortunately incompatible with the FSAN/ITU-T standards.

The IEEE Ethernet in the First Mile (EFM) working group began their PON specification effort in January 2001, with an aim to create a method of delivering 1 Gbps Ethernet service over point-to-multipoint fiber systems. Since they were based on Ethernet from the start, these systems are typically called Ethernet

PON or EPON systems. The resulting specification was called IEEE 802.3ah or EPON, which was released in 2004. It supported 1.25 Gbps service in the downstream and 1.25 Gbps service in the upstream. It is designed to support up to 64 Remote Nodes (although more may be possible). Reaches may range from 10 km to 20 km, depending on split ratios. On a single fiber, the EPON would operate at 1.48-1.5 um wavelengths for downstreams and 1.26-1.36 um wavelengths for upstreams. These systems began to see real deployments in 2005. [KRAM] [TECH]

The IEEE's second PON specification effort was begun in 2006 (as the 10G-EPON Task Force) and led to the creation of the IEEE 802.3av or 10G-EPON specification, which was released in 2009. A related specification defining system interoperability requirements was the IEEE P.1904.1 or SIEPON specification, which was released in 2012. Yet another related specification defining longer reach optics was the IEEE 802.3bk specification, which was released in 2013. It supported 10.3125 Gbps service in the downstream, and 10.3125 (or 1.25) Gbps service in the upstream. It is designed to support up to 64 Remote Nodes (although more may be possible). Reaches may range from 10 km to 20 km, depending on split ratios. On a single fiber, the EPON would operate at 1.575-1.58 um wavelengths for downstreams and 1.26-1.29 um wavelengths for upstreams. These systems were beginning to see real deployments in 2014, and the IEEE is beginning considerations on the next-generation of EPON standards that will likely use WDM techniques to offer even higher bit-rates. [10GE] [ETHE]

With all of the above PON options, the quoted bit-rates are all "raw bit-rates" that would have to be reduced if one wanted to account for the overhead from various forms of line encoding (8B/10b or 64b/66b were common) and MAC framing. Thus, the actual bit-rates available to subscribers would have to be reduced from these "raw bit-rates."

It should come as no surprise that the Telco markets have been bifurcated, as different Telcos have decided to follow different paths with different roll-out strategies. In general, North American and European and South American markets have tended towards the use of the FSAN and ITU-T standards, whereas China and other APAC markets have tended towards the use of the IEEE standards. [HARA]

Nevertheless, one can track the general bandwidth trends provided by the evolving PON technologies. These trends are captured in Fig. 2.

A Brief History Of The Telco Industry's Wireless Broadband Evolution (1998 to Present)

The rapid evolution of improvements found in cellular phone technology and the related wireless broadband technologies during the past 35 years have been legendary. The pace at which improvements have taken place in cellular/wireless broadband technology is quite remarkable, and it has spawned an array of technologies whose names look like alphabet soup.

Due to the large number of cellular/wireless variants that have evolved over time within different regions of the world, the paper will not attempt to track every evolutionary step. Instead, the authors will track large steps to see how this form of Telco service evolved over time.

Cellular phone technology can trace its humble origins to the first generation (1G) analog systems that began to be deployed in the late 1970's and early 1980's. NTT operated the first analog cellular system in 1979. Soon after that, NMT, AMPS, and TACS were three early cellular systems that

found deployments in northern Europe, North America/Latin America, and UK/China/Japan, respectively. Frequency modulation and Frequency-Division-Multiple-Access were common in many of these systems. Directional antennae and frequency re-use also became common technologies within these early systems. By 1993, it became apparent that extra bandwidth on the analog AMPS networks could be used to allow the low-bit-rate transmission of IP packet data to cellular phones (primarily for messaging). The maximum data-rate supported on these Cellular Digital Packet Data (CDPD) networks was ~19.2 kbps. This data-rate was much lower than the 56 kbps data-rates that were provided by telephone line modems of the time.

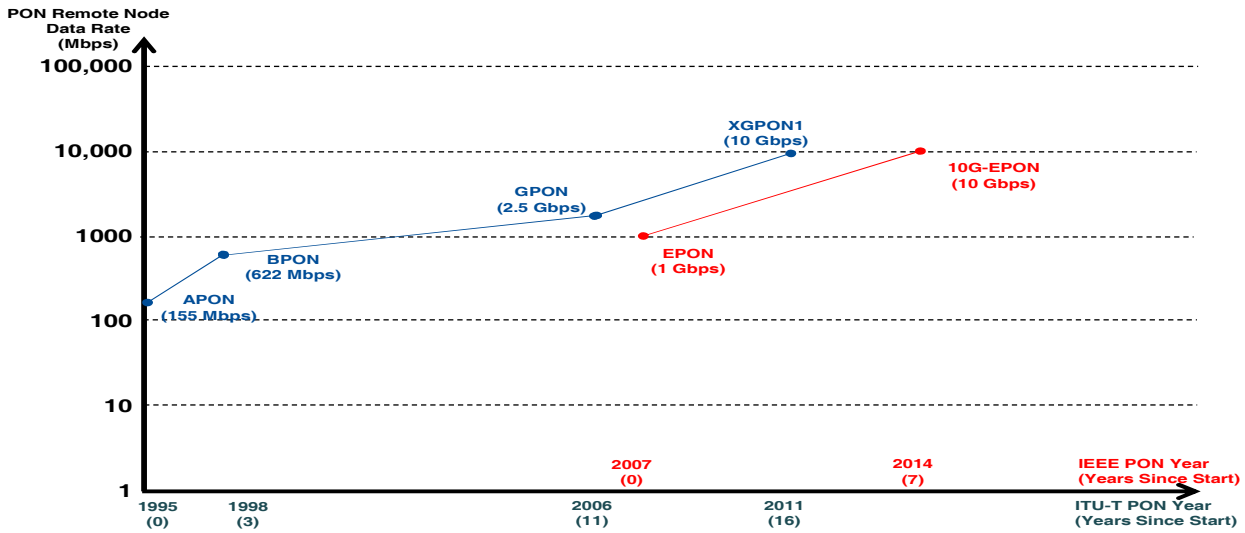


Fig. 2- PON Maximum Downstream Data Rates vs Time

Second generation (2G) cellular systems were introduced in the late 1980's. This new technology helped to transition the cellular market from analog to digital modulation techniques (ex: Time Division Multiple Access and Code Division Multiple Access (CDMA)). Many different types of systems were defined, including Global System for Mobile Communicattions (GSM), IS-54, IS-95 (aka CDMA One), IS-136, and Personal Digital Cellular (PDC). Improvements in the spectral efficiencies of these 2G cellular systems and increases in spectral widths (created by expanded frequency license sales) permitted much higher data-rates and more simultaneous phone calls to be carried over the cellular frequencies. With the advent of 2.5G cellular systems, GSM added General Packet Radio Service (GPRS) support in ~2000 to permit packetized data to be transmitted at the higher data-rates of 56 kbps. Then further enhancements led GSM systems to develop Enhanced Data Rates for GSM Evolution (EDGE), which provided even higher data-rates of 236 kbps.

In 2001, Telcos deployed third generation (3G) cellular systems that began to focus on providing more efficient packet transfers over the wireless interface. The ITU (in conjunction with the 3GPP project) helped to define the International Mobile Telephone-2000 (IMT-2000) standard to help guide this effort. The ETSI organization in Europe created the Universal Mobile Telecommunication System (UMTS) standard from the IMT-2000 work. In the U.S., the IMT-2000 work was modified to create the cdma2000 standard. High-Speed Packet Access (HSPA) was added to the UMTS systems to permit data transmission at data-rates of at least 200 kbps. 384 kbps was a

common data-rate for the HSPA systems. A similar data transmission variant known as Evolution Data Optimized (EV-DO) was included in cdma2000 systems. [1G4G]

The 3rd Generation Partnership Project (3GPP) project carried the torch into the fourth generation of cellular systems (known as 4G). The 4G systems would focus on unifying the different technologies that had developed in the cellular space, and it would also attempt to provide higher data-rates, lower latencies, and greater levels of security. Given the name Long Term Evolution (or LTE), the spec opted to utilize Orthogonal Frequency Division Multiplexing (OFDM) and Multiple Input-Multiple Output (MIMO) antenna technologies. The first LTE systems were deployed in Stockholm and Oslo on December 2009, but they were not fully compliant with the 4G specs, and were therefore called 3.9G systems. More spec-compliant 4G systems began to be deployed in the 2011-2014 time-frame. [EVOL] These systems offered typical downlink data-rates of 5-12 Mbps, which permitted them to be utilized for the transmission of IP Video streams. [LTES]. Recent augmentations to the spectrum by some Telco service providers are permitting subscribers to experience 30 Mbps peak downlink data-rates (as of 2014). [PEAK]

The 3GPP project standardized a new specification known as 3GPP Release 10 in March of 2011. That specification is also called LTE-Advanced, and will likely serve as the definition for the next generation of broadband wireless services. LTE-Advanced will include many new features, including Coordinated Multi-Point transmission and reception (aka CoMP), MIMO techniques, bandwidth up to 100 MHz, cognitive radio, and Self-Optimizing Network (SON) techniques, and coordinated management of macrocell and picocell and femtocell transmissions to optimize performance. [LTEA]

The general bandwidth trends provided by the evolving cellular/wireless broadband technologies are illustrated in Fig. 3. (Note: There are many variants that could be added to this figure which would change the shape of the curve. However, this curve is adequate for providing general trends).

A Brief History Of The Telco Industry's DSL Evolution (1998 to Present)

While Telcos focused much of their post-1998 investment on other technologies (such as PON and Cellular), there was still a lot of wired Telco plant already deployed in the world. This infrastructure could be utilized to deliver broadband services if enabling technological break-throughs could be found to permit more bandwidth to be transmitted. Techniques to break through the 56 kbps modem bandwidth ceiling were needed. Fortunately, there was ground-breaking innovation being pumped into the old telephone line modem technology- even after 1998. Most of this work was focused on a new technology that came to be known as DSL technology.

The primary difference between DSL and its predecessor copper line technology (dial-up modems) is that DSL attempted to operate without interfering with traditional voice signals that are still operating in the voiceband spectrum below 3.9 kHz. Thus, a voice call and Broadband Internet service could co-exist on the copper loop. In addition, DSL transmissions do not require a telephone connection to be established (and paid for), so it was viewed as an "always-on" technology that could be used continuously without fear of creating large phone bills.

As a result of this basic change, DSL permitted (and required) its loop signals to be bandpass signals residing above 3.9 kHz in the spectrum. These bandpass signals only

needed to take on this form for transit across the copper telephone loop, because equipment in the Telcos Central Office would immediately convert the DSL signals into traditional Internet signals (Ethernet or ATM over Synchronous Optical Network or SONET technology) for ultimate transmission through the Internet. This fundamental change in approaches opened the door to many innovations, because the bandpass signals were not constrained to operate within 3.9 kHz of spectrum.

Since the higher-frequency DSL signals cannot pass through loading coils found in many telephone loops, the loading coils are usually removed to permit DSL service to operate. (Note: The removal of these loading coils can make voice service unusable over long distance loops, so there are distance limits placed on the reach of DSL loops. This eventually led to the creation of DSL Access Multiplexers, or DSLAMs, that are placed closer to homes, cutting down on the overall length of most telephone loops).

The first generation of DSL essentially began in 1988, when a group of engineers at Bellcore filed a patent application for a broadband technology that placed the broadband signals on a twisted pair loop above the 3.9 kHz voiceband. The patent proposal argued that the downstream and upstream bandwidth capacities could be asymmetrical, since most data might be transmitted towards the home rather than away from the home.

By 1991, a de facto standard for High-speed Digital Subscriber Line (HDSL) technology was completed by Joe Lechleider of Bellcore, drawing on many of the ideas

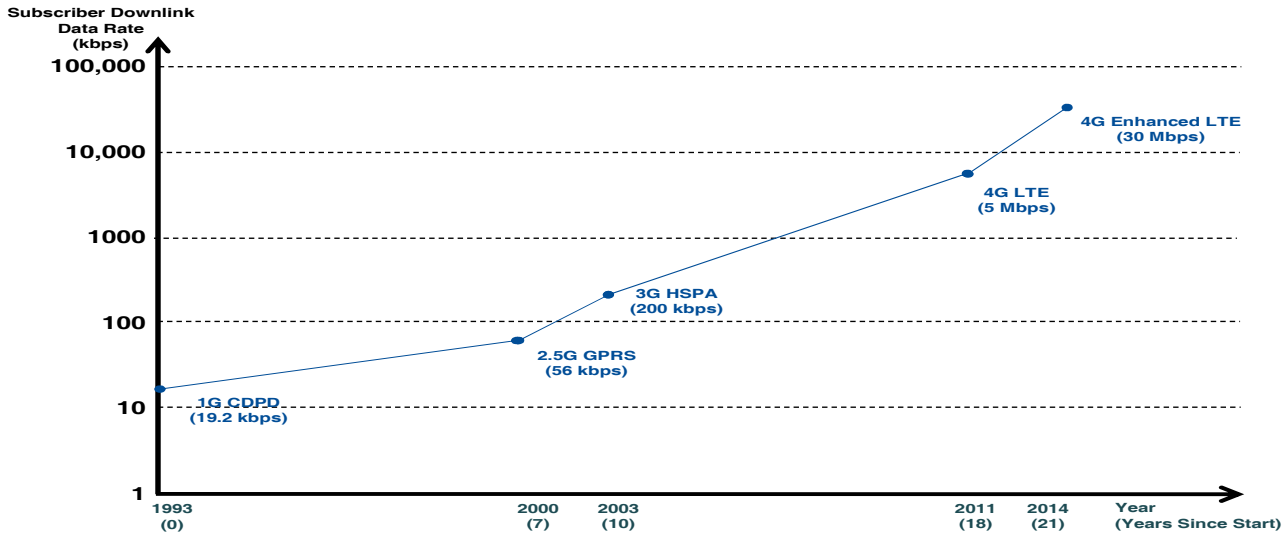


Fig. 3- Wireless Broadband Downlink Data Rates vs Time

(such as line equalization) from the Integrated Services Digital Network (ISDN) and T1 technologies that had been used to transport digitized voice signals across the Telcos long-distance telephony networks. It was a two-line system that transported 800 kbps on each line, offering a total of 1.6 Mbps of bandwidth capacity for two-mile loops (limited mostly by crosstalk). It was deployed to a relatively small number of subscribers (since it was limited by the undesirable requirement for two lines), and it was rapidly replaced by the next generation of DSL technology known as Asymmetric Digital Subscriber Line (ADSL).

ADSL had several different and competing proposals for modulation technologies. Carrier-less Amplitude Phase Modulation (CAP) and QAM were proposed by some companies, whereas Discrete Multi-Tone

(DMT) was proposed by other companies. Eventually, bake-offs showed that DMT performed better by being able to adjust each of its narrow carriers to conform to the varying Signal-to-Noise Ratios in each portion of the loop spectrum. In 1993, John Cioffi's company (Amati) was the first company to produce a working 6 Mbps ADSL product. By 1995, the ANSI T1.413.1995 standard existed, and 6.4 Mbps downstream/800 kbps upstream ADSL modems were being deployed using a 1.1 MHz spectrum, as shown in Fig. 4. An ITU spec for ADSL, known as ITU-T G.992.1, was also ratified in 1999.

Improvements to the ADSL technology eventually led to the ADSL2 spec known as ITU-T G.992.3. These ADSL2 modems operated at 12 Mbps (in the downstream) and 1 Mbps (in the upstream) using new technologies such as improved modulation formats and bonding of multiple lines. Deployments began in 2003.

The ADSL2+ spec (known as ITU-T G.992.5) was then ratified in 2003. These modems operated at 24 Mbps in the downstream and 1 Mbps in the upstream. It accomplished this by doubling the downstream frequency band to be 2.2 MHz wide, and it also added crosstalk mitigation techniques. Deployments began in 2005.

A new spec using slightly different technologies was then created. It was called Very-high-bit-rate Digital Subscriber Line (VDSL), and became the ITU-T G.993.1 spec in 2004. It offered 55 Mbps downstream and 15 Mbps upstream service. Early deployments began in 2004.

By 2005, improvements to the VDSL technology led to the VDSL2 spec known as ITU-T G.993.2. These VDSL2 modems operated at 55 Mbps (in the downstream) and 30 Mbps (in the upstream) for long-reach loops. Alternatively, they can be operated at 100 Mbps (in the downstream) for short-reach loops. VDSL2 utilizes up to 17 MHz of spectrum on the twisted-pair loop. Vectoring technologies (to mitigate crosstalk) and bonding technologies (to use multiple lines into a house) were used within this technology. Early deployments began in 2008. [DIGI] [DSLT] [CIOF]

In 2014, a new DSL set of standards known as ITU-T G.9700 and ITU-T G.9701 were adopted. They are also known as G.Fast. G.Fast is being designed to support a 150 Mbps to 1 Gbps service over loop lengths of 250 meters or less. It can utilize up to 212 MHz of spectrum on the twisted-pair loop, which is much more spectrum than previous DSL spectra. (It is much more spectrum than the early ADSL systems, as shown in Fig. 5a). It also uses Time Division Duplexing (instead of the Frequency Division Duplexing techniques used in previous DSL versions). Deployments of G.Fast are expected to begin in 2016. [GFAS]

When comparing the different DSL spectral efficiencies and spectral widths of ADSL vs. G.Gast, it becomes clear that Telcos have made much more use of spectral width increases and have relied less on spectral efficiency improvements. In fact, they seemed to be willing to accept drops in spectral efficiencies over time (probably due to the lower SNRs found in the regions of the spectrum where they expanded into). However, to continually remain competitive and extend their spectral width, it appears that the Telcos were quite willing to sacrifice network reach. This is somewhat illustrated within Fig. 5b, which includes data from several sources. [WHAT] [GFAS] [VDSL]

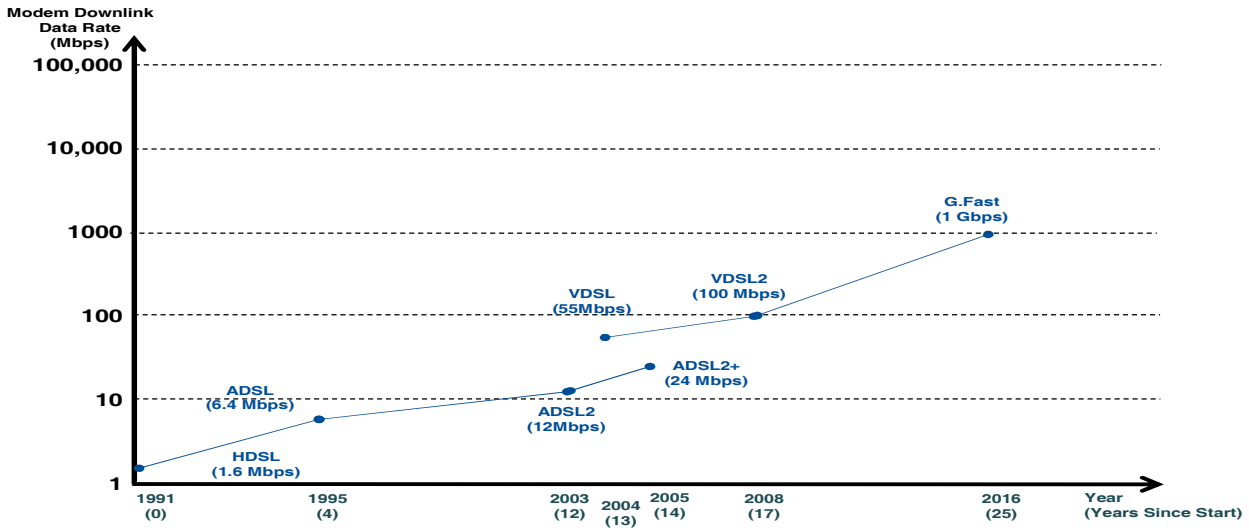


Fig. 4- DSL Downlink Data Rates vs Time

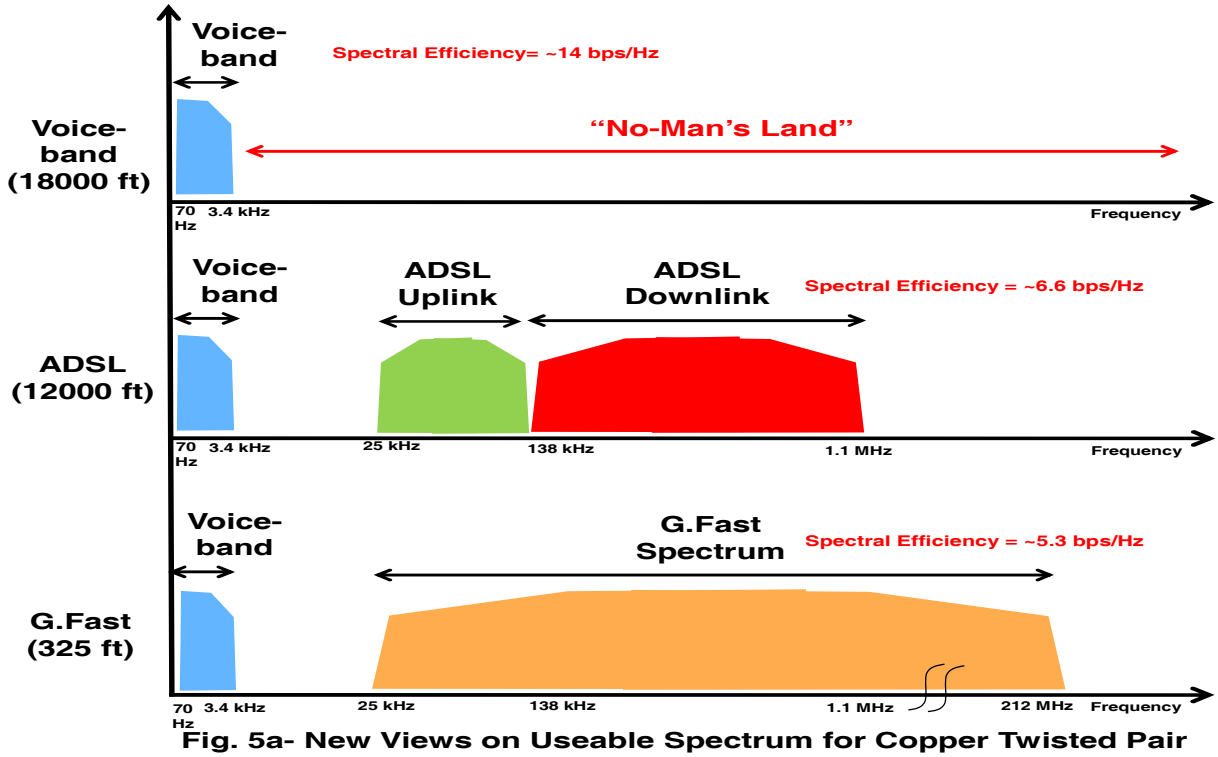


Fig. 5a- New Views on Useable Spectrum for Copper Twisted Pair

DSL ultimately became the predominant broadband technology deployed in Europe and many other parts of the world. Fig. 4 illustrates the bandwidth trends that led to this success. The growth in data-rates since 1998 are remarkable- especially for a twisted-pair copper line infrastructure that was predicted to be near its end-of-life when 56 kbps V.Last technology became available in 1998. The 2016 G.Fast data-rates of 1 Gbps will be approximately 17,857x faster than the V.Last data-rates of 56 kbps. In the 25 years from 1991 to 2016, DSL experienced an average CAGR in data-rates of 29.37%.

It is interesting to compare this CAGR to the CAGR values of other technologies that were studied. These comparisons are shown in Table 1. The data-rate CAGRs range from

Technology	DS Spectral Efficiency	DS Spectral Width	DS Bit-Rate	DS Reach
ADSL (1995)	1.3 bps/Hz	1.1 MHz	1.5 Mbps	18,000 ft
ADSL (1995)	5.8 bps/Hz	1.1 MHz	6.4 Mbps	12,000 ft
VDSL (2004)	7.8 bps/Hz	7 MHz	55 Mbps	1,000 ft
G.Fast (2016)	4.7 bps/Hz	212 MHz	1 Gbps	325 ft

Fig. 5b- Spectral Efficiency and Spectral Width and Bit-rate and Reach for Various DSL Technologies

~14% (for Voiceband Modems) to ~42% (for Wireless Broadband). GPON and DSL both experienced data-rate CAGRs of ~29%. In all cases, definite improvements were seen in the different technologies. It is intriguing to see that DSL provided an average data-rate CAGR of 29% per year, even after Voiceband modems had reached their end-of-life.

Technology	Years	Total Data-rate Increase	Average CAGR
Voiceband Modems	40	186.67	13.97%
GPON	16	64.52	29.75%
Wireless Broadband	21	1562.50	41.93%
DSL	25	625.00	29.37%
	Table 1- Data-rate Growth Rates		

TRAFFIC ENGINEERING FOR THE CABLE SPACE

A Quick Prediction Of Cable Traffic Engineering For The Future

When will the Cable Industry's current delivery infrastructure (the HFC plant) reach its end-of-life? Is the Cable Industry at a point where it is ready to enter a "technology transition window" similar to the one experienced by Telcos in the 1998 time-frame? Will Cable MSOs need to immediately embrace new technologies to replace their HFC Plant? Or will they find new techniques to extend the end-of-life of their existing HFC plant in ways that mimic what Telcos were able to do with DSL technologies on their copper loops?

These are all difficult, but important questions that need to be answered. In this paper, the authors will attempt to analyze the decisions made by Telcos over the past 17 years to predict what Cable MSOs might do in the coming future.

However, before looking at past Telco decisions, it would be beneficial to first perform some traffic engineering calculations to predict the amount of bandwidth capacity that will be required within Cable Service Groups of the future. One can then compare those predicted bandwidth capacity requirements to the bandwidth capabilities provided by the HFC plant (with various modifications), and guesses can be created on the expected life-span of the existing HFC Plant.

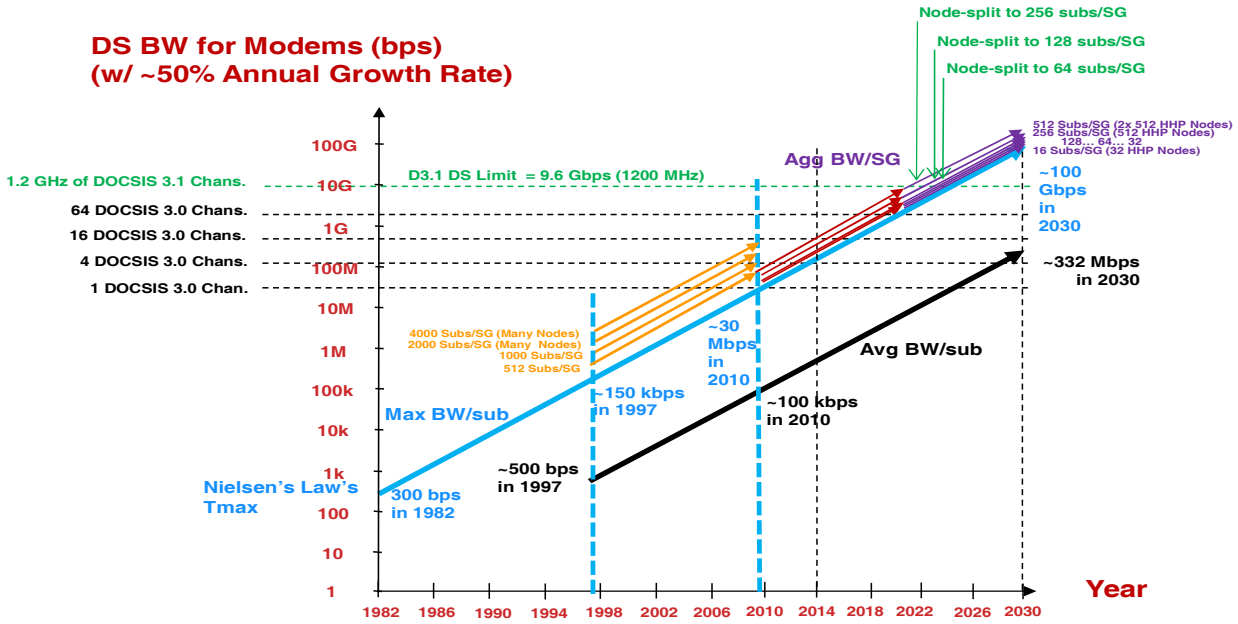


Fig. 6- Bandwidth Trends with 50% CAGR in Future

Recent studies have shown that the required bandwidth capacity for a particular service group can be estimated by knowledge of the Maximum Per-Subscriber Bandwidth Offering (Tmax), the Average Busy-Hour Per-Subscriber Bandwidth Consumption (Tavg), and the number of subscribers (Nsub) in the service group. [CLOO]

Within that study, it was shown that to provide reasonably good Quality of Experience levels to subscribers within a service group, the bandwidth capacity requirements (C) for the service group could be roughly specified by the following formula:

$$C >= Nsub*Tavg + 1.2*Tmax. \quad (1)$$

It should be noted that other more complex (and more accurate) formulae are also under study [EMM1], but this simple formula will be utilized within this paper to make some key points.

Using formula (1) and predictions about Nsub and Tmax and Tavg trends into the future, one can plot the required bandwidth capacity levels for various sized service groups.

This is done in Fig. 6 for a future where Tmax and Tavg continue to grow with a 50% CAGR (as they have done for the past decades). This is also done in Fig. 7 for a future where Tmax and Tavg grow with a lower 30% CAGR. This lower 30% CAGR is predicted by some MSOs who believe that IP Video streaming created the last big boost in bandwidth over the last several years and that no new applications will likely show up in the near future to continue pushing that 50% pace. Other MSOs still believe that the CAGR will continue at a 50% pace, driven (perhaps) by an upcoming move to 4K video streaming and then to Virtual Reality streaming.

In either case, the plots of Fig. 6 and 7 both illustrate that the HFC plant still has quite a bit of life left in it. In general, MSOs could use DOCSIS 3.1 across a 1.2 GHz plant and sustain the growth rates (with the assistance of a few node-splits) all the way until the mid-2020's (for a 50% CAGR) or the late-2020's

(for a 30% CAGR) before the highest Service Tier hits the limit.

the college students of today will help to invent those impossible-to-identify

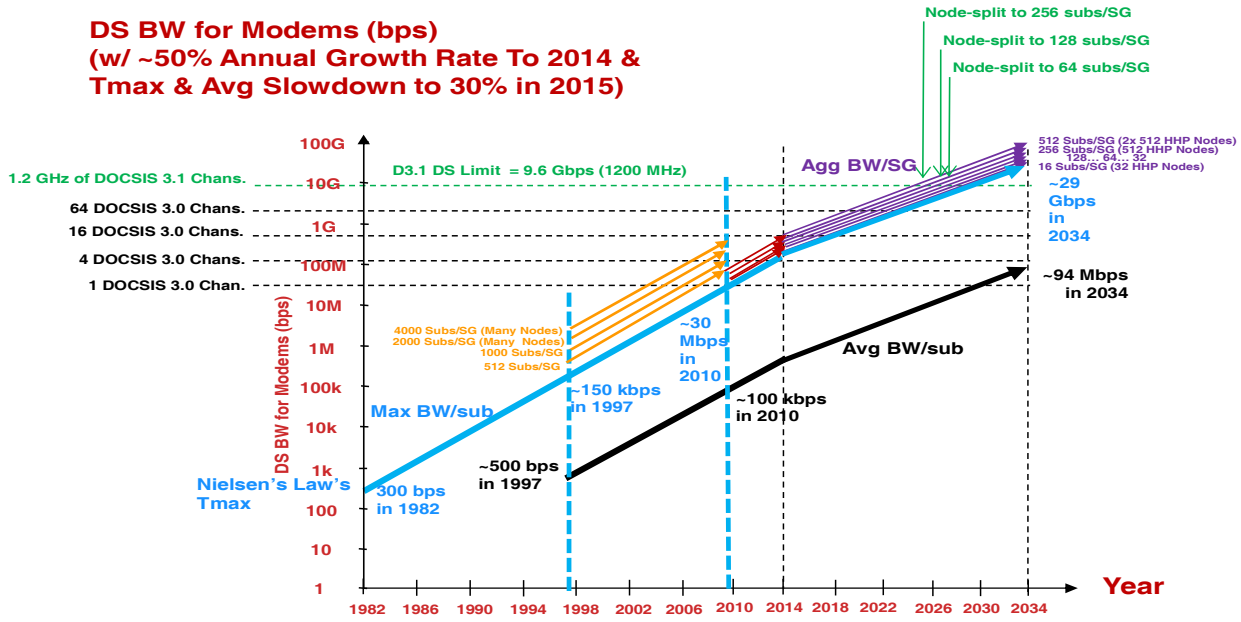


Fig. 7- Bandwidth Trends with 30% CAGR in Future

There is quite a bit of useful information provided by the plots in Figs. 6 and 7. The plot has the Downstream Bandwidth displayed in a logarithmic fashion on the y-axis, and it has the years ranging from 1982 to 2030 on the x-axis

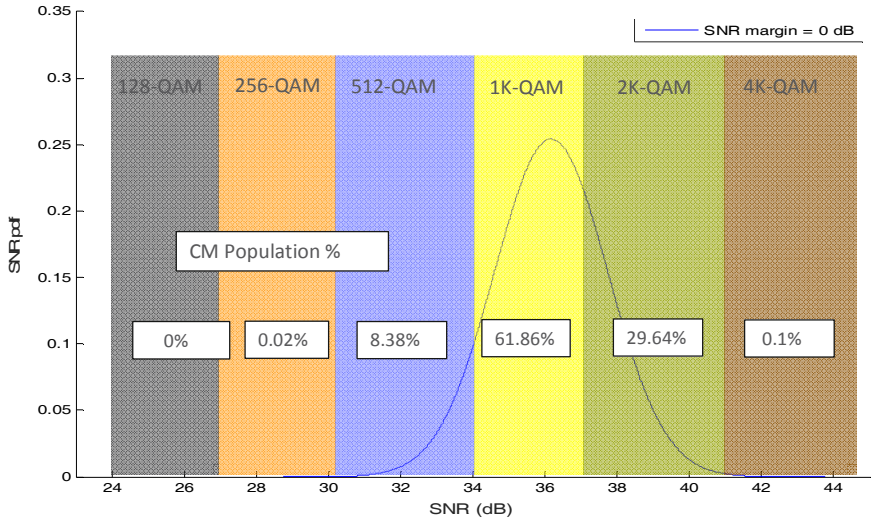
The first curve to explore is the light blue curve, which shows Nielsen's Curve that identifies the expected Tmax values on a year-by-year basis. This curve illustrates that MSOs will likely have to provide higher and higher Tmax values to their subscriber pool on a yearly basis. If these trends continue at a 50% CAGR, then Nielsen's Curve predicts that the Tmax value for a high-end modem may be on the order of 100 Gbps by 2030. The 30% CAGR predicts that the Tmax value may approach 29 Gbps by 2030. One may wonder what applications could possibly require that kind of bandwidth 15 years from now. The honest answer is that the authors do not know what those applications will be—nobody does. But it is probably fair to say that applications. It is, however, quite feasible that these applications may include 4K/8K IP video streaming, machine-to-machine applications, IoT applications, virtual reality systems, and futuristic holographic display systems.

The second set of curves of interest (in Figs. 6 and 7) that should be explored is the curve set illustrating Average Bandwidth Consumption rates as a function of time. The dark black curve illustrates the approximate average per-subscriber bandwidth consumed by a single subscriber during the busy-hour period of time (8pm-9pm). This curve displays an average value using contributions from both active and inactive users, so the average values (Tavg) end up much lower than the Tmax values. In fact, the Tavg values for a single subscriber tend to be ~300 times lower than the corresponding Tmax values within a given year! However, with a 50% CAGR, the plotted values will grow to be ~332 Mbps by 2030. With a 30% CAGR, the plotted values grow to be ~94 Mbps by 2030.

While a curve for a single subscriber is interesting, it is not very useful for traffic engineering analyses that must examine and predict aggregate traffic patterns for many subscribers in a shared pool (like a Service Group). The single subscriber curve can, however, be scaled upwards in a linear fashion to yield the average bandwidth for a larger pool of subscribers that can share the DOCSIS HSD channels within a Service Group (SG). When that scaled bandwidth is added to 1.2*(Tmax), the authors end up getting the amount of bandwidth predicted by formula (1). The orange, maroon, and purple curves display the scaled-up bandwidth requirements for many different sized Service Groups. It should be noted that two adjacent, parallel lines of any particular color (orange, maroon, or purple) are exactly one node split apart from one another in terms of the bandwidth capacity required to support the Narrowcast services.

would be available if the MSO were to donate ~100% of the 1.2 GHz HFC network bandwidth to DOCSIS 3.1 operation. The resulting bandwidth capacity of ~9.6 Gbps assumes above-average Downstream SNRs averaging ~40 dB and permitting an average DOCSIS 3.1 modulation order of 2048QAM with 8.8 bps/Hz spectral efficiencies across ~1100 MHz of the 1200 MHz spectrum. Fig. 8a shows how this level of performance is possible for *clean* plants in today's HFC network. [ALB1] (Note: It is the belief of the authors that providing an isolation between the access network and the home network can help ensure that many connections to homes could achieve 40 dB SNRs or better).

This level of bandwidth capacity (9.6 Gbps) would support most Narrowcast services until the time-frames indicated by the vertical green arrows within the figures (when the purple curves rise above the 9.6 Gbps threshold



Spectral efficiency (bps/Hz)							
	8K:	5.628	6.432	7.236	8.040	8.844	9.648
	4K:	5.216	5.961	6.706	7.451	8.196	8.941

Fig. 8a- Histogram of SNRs on Typical HFC Plants & Corresponding QAM Order

The dashed horizontal lines mark various bandwidth levels for different numbers (and types) of DOCSIS QAM channels. The green, dashed, horizontal line is the maximum amount of useable bandwidth capacity that

level). As can be seen in Fig. 6, this scenario would carry MSOs on their current HFC plant until the mid-2020's (for CAGRs of 50%). In Fig. 7, this scenario would carry MSOs on their current HFC plant until the late-2020's (for CAGRs of 30%).

It is important to note that this traffic analysis is based on Nielsen's Law. As a result, it is focused on the highest speed tier offering (sometimes referred to as the Billboard rate or tier). A study [ULM1] carried out with several large MSOs shows that the number of subscribers within the top Billboard tier actually account for 1% (or less) of the existing cable customers. Another important discovery from this study is that the lower bandwidth Service Tiers are also growing at proportionately lower CAGRs. This is shown within the plot of Fig. 8b.

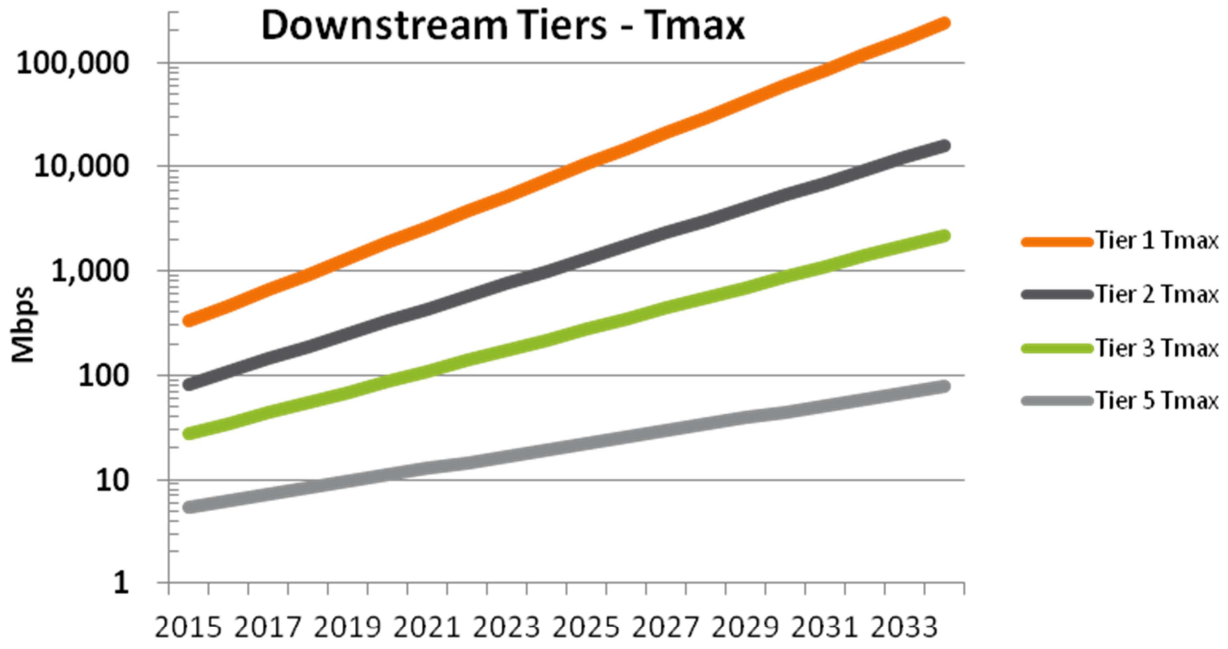


Fig. 8b- Tmax Trends for Different Service Tiers

While the orange Top Tier (Tier 1) is growing at 50% CAGR, it represents only 1% or less of the customer base. Meanwhile the lowest tiers (green & gray) are experiencing lower CAGRs, and they also account for the vast majority of the customers (in the range of 75%-90%+).

From this perspective, the vast majority of subscribers will likely be able to remain on the HFC plant for a much longer period of time than Performance Tier subscribers, while the Performance Tiers (e.g. 1% to 20% of subs) will likely hit the available bandwidth ceiling in the next decade or two. This is similar to what has been seen in the Telco space. While new DSL technology has come along (e.g. VDSL, G.Fast), this is typically being deployed to a limited number of elite customers, whereas the majority of the customer base continues to run on the older DSL technologies. Similarly, the latest and greatest wireless technologies are also rolled out slowly to a limited number of customers for a number of years before the aggregate masses are given access to the new technologies.

So, turning back to Fig. 6 and Fig. 7, one may wonder whether the green-arrow epochs

in the late 2020's (where the required bandwidth exceeds the HFC bandwidth capacity) marks the the true ending of the HFC plant? Will that force MSOs to enter a "technology transition window" prior to that time to permit a transition to new technologies and away from the HFC plant by the late 2020's? Is there anything that can be done to extend the life of the HFC plant beyond this late 2020 time-frame? The authors will attempt to answer these question in the following sections.

LESSONS LEARNED FROM THE TELCO AND WIRELESS PROVIDERS

Identifying Categories Of Lessons

In many of the previous sections, the authors outlined a brief history of various Telco and Wireless service providers. The focus was on the technologies that they used during their "technology transition windows" and the bandwidth capacities that they were able to achieve using those technologies.

In this section, the focus will change. Instead of looking at technologies, the authors will focus on some of the lessons that might be learned by looking more closely at what decisions were made by the Telco and Wireless providers as they evolved their networks.

There are both obvious lessons and subtle lessons that can be found if one keeps an open mind. In particular, there may very well be many actions taken by by Telco and Wireless providers as they transitioned their networks that may be candidates for use by MSOs as they move forward into their own future (which will undoubtedly include some modifications to their networks). Some of these ideas may be interesting to MSOs. On the other hand, some others may not be as palatable or practical. But the authors will outline all of them, because sometimes even some of the most outlandish ideas may prove to be valuable tools when push comes to shove and novel solutions are required.

Thus, the following sections will list many of the potential lessons observed by the authors as they researched the history of the various Telco and Wireless providers. Each of these lessons will be placed into one of three categories:

- Category A: Lessons with a high probability of being used by MSOs
- Category B: Lessons with a medium probability of being used by MSOs
- Category C: Lessons with a low probability of being used by MSOs

Lessons learned from the Telco and Wireless providers' past actions will be mapped into each category, and the proposed lesson will be coupled with a potential action that could be considered by MSOs. Obviously, the particulars of the past Telco and Wireless actions will need to be modified to be compatible with the future circumstances of MSOs, but the resulting observations should nevertheless prove to be interesting.

Category A: Lessons With A High Probability Of Being Used By MSOs

There were several lessons learned from Telco and Wireless providers that appear to be quite adaptable to the future that MSOs may face. These lessons are outlined within this section.

Lesson A1: Use Orthogonal Frequency Division Multiplexing (OFDM) and Low Density Parity Check (LDPC) Codes. This permits higher spectral efficiencies that approach the Shannon Limit for transmission systems. These types of approaches have been

used for years in both Wireless and DSL systems.

Obviously, these techniques are already being planned for future DOCSIS 3.1 systems, and therefore will be used by MSOs.

Lesson A2: Assign different modulation orders to subscribers with different Signal-to-Noise Ratios (SNRs) to optimize throughput. This has been done for years in both DSL and Wireless.

With the advent of Multiple Modulation Profiles in DOCSIS 3.1, these techniques are also being planned for future use within DOCSIS 3.1 systems. As a result, they will undoubtedly be used by MSOs in the very near future.

Lesson A3: Maximally utilize Moore's Law improvements in silicon and new linearizers/equalizers to permit signal reception with lower SNR and higher distortion. These types of improvement led to the continual bandwidth increases that have been evident in PON, Wireless, and DSL systems over many years.

Obviously, these improvements in the available technologies are difficult to predict, but there is a reasonably high probability that they will continue to occur into the future. It is not so much a question on "if" the industry will see advances, it is more a question on how fast and when those advances will be available. Thus, MSOs should plan to capitalize on them. As an example, it seems quite possible that continuous improvements in equalization and non-linear distortion cancellation techniques may permit even more advanced corrections of complex high-frequency distortions in the future. This may permit higher and higher frequencies to be transmitted and received on the HFC network. Fiber limits due to nonlinear optical noise and "copper plant" limits due to nonlinear amplifiers may be able to be pushed aside by new pre-processing and post-processing algorithms.

Lesson A4: Use IP Video as a transport technology for television programming. Telcos received greate benefit from migrating their infrastructure from analog voice to Voice over IP (VoIP). They have also adopted IP Video transport over DSL and PON for its benefits. The benefits of IP Video include the use of a single back-office infrastructure for delivery to all devices, the ability to transmit across any type of network (to endpoints both inside and outside the home), the statistical multiplexing gains of packet-based delivery on high-bandwidth channels, the availability of high-compression, low-bit-rate video codecs (such as H.264 and HEVC), and the increased bandwidth efficiency gains resulting from Adaptive Bit-Rate (ABR) algorithms which can dynamically change video resolutions.

These approaches are likely to be utilized by MSOs to transmit IP Video signals over DOCSIS channels in the future.

Lesson A5: Migrate appropriately-selected customers to Fiber-To-The-Premise (FTTP) service as their growing bandwidth needs necessitate this alternative service. Telcos have successfully targeted new technologies to selecte customers, while keeping the bulk of their customers on older, existing technologies as long as possible to maximinze their return on investments.

FTTP promises much higher bit-rates to subscribers in the long-run, and Telcos have migrated high-end users to FTTP technologies for years as their bandwidths grew to exceed the capacities of twisted-pair wires.

These FTTP approaches are likely to be the end game for most service providers-including MSOs in the distant future. FTTP will therefore be a part of every MSOs plan as

they migrate into the future. While the current HFC infrastructure can support most customers needs into the late 2020's and beyond, an eventual move to these FTTP technologies is likely to be a part of the deep MSO future. Unlike Telcos, MSOs have the benefit of being able to choose from several permissible FTTP technologies, including Traditional RFoG, Extended-Spectrum RFoG, and PON. Each approach has its own advantages and disadvantages.

GPON and EPON currently offer 10 Gbps Downstream capabilities with a promise of increased bandwidth capacities in the future. xPON technologies also offer scaling benefits due to their use in both the MSO and Telco spaces.

Traditional RFoG (with its DOCSIS 3.1 modulation orders carried inside of 1.2 GHz or 1.7 GHz spectral widths) could offer 10-15 Gbps Downstream capabilities and permits interoperability with existing customer-premise equipment.

Extended-Spectrum RFoG would eliminate the 1.7 GHz boundary of DOCSIS 3.1, and it would utilize higher spectral widths that could permit (say) 25-50 Gbps Downstream capabilities in the future. The higher bit-rates provided by Extended-Spectrum RFoG could be realized in both the Downstream and Upstream directions due to the fact that separate lamdas are used to transport the two types of signals. With separate RFoG wavelengths allotted for Downstream and Upstream transmissions, it may be possible to fill an 8 GHz Downstream spectrum and a1.2GHz Upstream spectrum with high bps/Hz signaling, including or exceeding 1024QAM. This would yield up to ~10Gbps in the Upstream concurrent with up to ~50Gbps in the Downstream.

Ultimately, cable operators might want to consider this type of Hybrid-PON (HPON) architecture that can simultaneously support a mix of these different FTTP technologies. This would offer the MSOs the best of all worlds and the maximum amount of future-proofing.

Lesson A6: Modify home networks to help increase the capacity of the access network. Telcos showed a strong willingness to do this for their DSL deployments, where they found it necessary to (for example) install DSL filters into the resulting home networks.

Approaches such as these may be applicable to MSOs in the future. For example, recent analyses have shown that the maximum achievable head-end to amplifier on some HFC plants may support very high modulation orders including 4096QAM while some HFC plants may support a maximum 1024QAM to the amplifier. [EMM2] However, the propagation of the signals from the head-end to the cable modem (through the home network) can oftentimes degrade the SNRs to much lower levels that will only support 1024QAM (or even less). [EMM2] [ALB2] These scenarios are captured in Fig. 9a and Fig. 9b. The resulting conclusion is that MSOs (like Telcos) may find benefits from physically separating their access network from their home networks. This may be accomplished using 2-port gateways at the portal into the home for the DOCSIS systems of the future.

Lesson A7: Use different access technologies in parallel to match appropriate bandwidth capacity to each subscriber. Telcos have been doing this in extreme ways ever since they introduced cellular network services (for mobile subscribers) and fiber network services (for high-bandwith subscribers) on top of their already-existing twisted-pair services (for normal-bandwidth subscribers). In many regions of the world, Telcos are managing all three of these services within the same region. They are even looking to expand to another

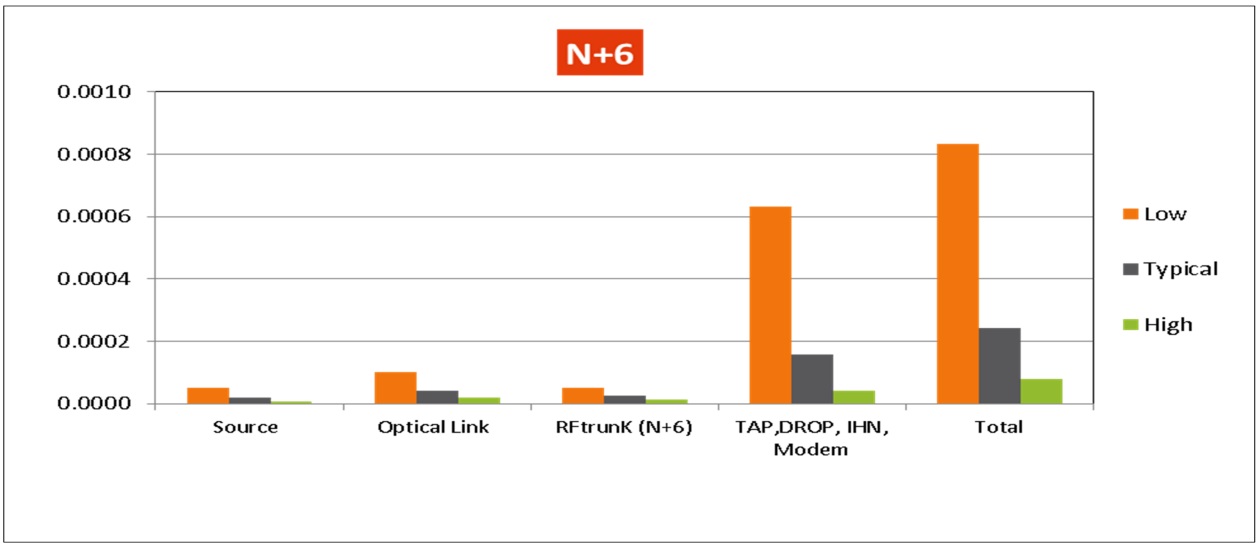


Fig. 9a- 1/CNR (linear scale) for different segments composing the N+6 network (centralized architecture)

form of more localized wireless service through the use of LTE-U technologies. While the parallel use of these technologies leads to increased operational complexities, the use of centralized, software-based management platforms has simplified the management of these various access technologies.

operating for many years to come. In addition, they may be managing new fiber-based solutions (using PON or RFoG) as well as new wireless-based solutions (using WiFi or LTE-U). Each service will be paired with appropriate subscribers based on the bandwidth needs of each subscriber. An extreme version of this approach (as applied

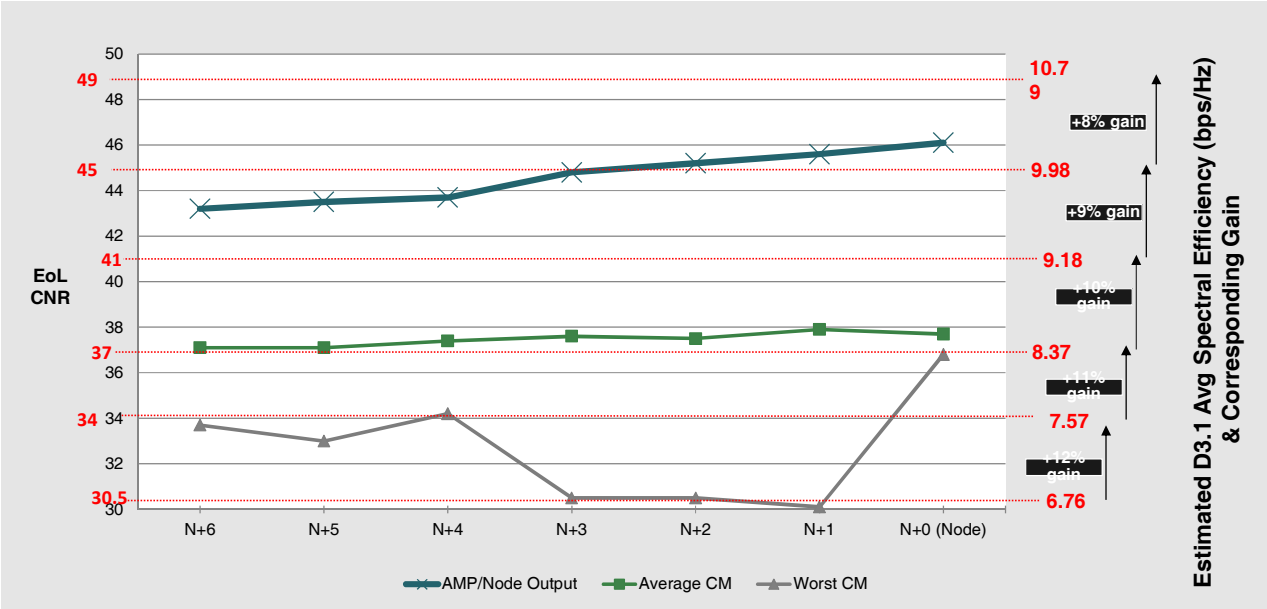


Fig. 9b- EOL CNR & Spectral Efficiency at various points in HFC Plant

MSO's may very well find themselves in a similar predicament as they move forward in time. They will likely have the HFC network

by an MSO) might include the use of Coax, Fiber (PON or RFoG), and WiFi/LTE-U all entering the home in parallel, as shown in Fig. 10. While extreme, it is nevertheless quite

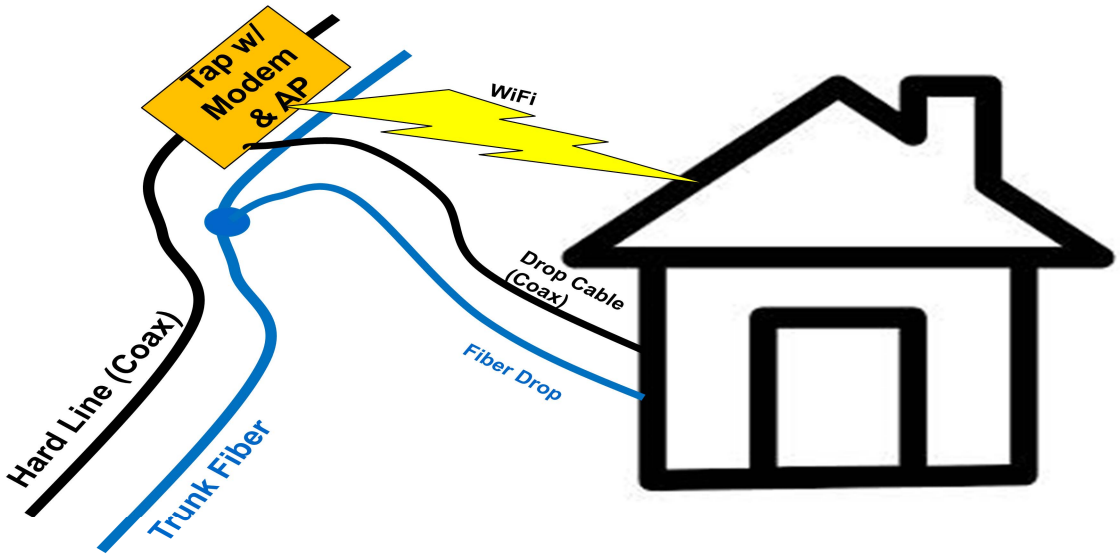


Fig. 10- Parallel Access Technologies Into The Home

possible that the natural evolution of the Cable plant could lead to a situation where all three of these technologies will be available to the MSO for driving bandwidth into the home. DOCSIS Provisioning of EPON (DPOE) and DOCSIS Provisioning of GPON (DPOG) obviously help to manage several of these different technologies using a common DOCSIS management infrastructure. Perhaps a DOCSIS Provisioning of WiFi (DPOW) definition will further expand the benefits of that common DOCSIS management infrastructure in the future.

Lesson A8: Use centralized, software-based management of network infrastructure platforms (ex: IMS & SDN). Telcos moved in this direction with their Intelligent Network (IN) efforts for fixed networks. They continued to support this approach with the development of the IP Multimedia Subsystem (IMS) to support Internet traffic on the traditional telecommunication networks. They then began showing early interest in Software Defined Networks (SDNs) to manage all network elements from a single logical control element. All of these approaches share some common elements. They all are based on open standards; they all attempt to subdivide control and dataplane functionality into separate sub-systems; and they all attempt to permit providers to more easily manage solutions from multiple vendors.

Portions of the IMS protocols (ex: SIP) have already been utilized by MSOs. In addition, SDN approaches are currently being studied by MSOs, and the MSOs may find similar benefits in their application of these types of technologies in the near future. These centralized, software-based management systems could be useful in managing the complexity of multiple access technologies or multiple Remote Gadgets in the future.

Lesson A9: Use Distributed Access Architectures to help deliver services. Telcos have used this approach quite successfully with DSL services. DSLAMs were placed outside of central offices to provide the final connections to the subscriber. High-bitrate fiber connections were typically used for connectivity to the DSLAMs.

MSOs may choose to move in this direction as well. Studies on various Remote Gadgets or Distributed Access Architectures (ex: Remote CCAP, Remote PHY, Remote MAC+PHY) are already under way.

Lesson A10: Use Carrier-Class equipment to ensure high service availability. Telcos have always taken high availability to be a keystone of their wired service offering. The wired telephone network has always been considered a "lifeline service" whose operation was critical to the survival of subscribers during life-threatening events.

MSOs are also working to build a Carrier-Class network. These efforts will undoubtedly pay off as subscribers are likely to continue to demand that level of service in the future.

Category B: Lessons With A Medium Probability Of Being Used By MSOs

There were several lessons learned from Telco and Wireless providers that may be interesting to MSOs in the future. Some MSOs may find these medium probability lessons to be interesting. Others may not find them useful. These medium probability lessons are outlined in this section.

Lesson B1: Expand channel spectrum to maximize channel capacity. In their DSL bandwidth expansions, it seems that Telcos realized early on that a K times increase in spectral width will oftentimes yield higher overall bandwidth capacities- even with a corresponding 1/(K+ε) decrease in power spectral density, implying that the total channel power is held constant. (Note: The ε value in the power spectral density formula above is included to account for any additional signal power loss due to increased signal attenuation or due to amplifier/active element distortions sometimes found in the higher frequency portion of the spectrum where the spectral expansion usually takes place).

A similar application of this rule in the MSO space would imply the use of Extended-Spectrum DOCSIS 3.1. This would ignore the current limits of 1.2 GHz or 1.7 GHz on DOCSIS 3.1, and extend the spectrum beyond those limits. For example, if a DOCSIS 3.1 system is already operating with (say) 4096QAM in 1 GHz of spectrum, then the application of a K=4x spectrum increase (with ε=0) would result in a spectrum expansion of 4x to 4 GHz of spectrum. However, if total signal power levels are maintained at the same level, then the power spectral density reduction would be ~6+ dB (in power per MHz). The ~6 dB reduction might therefore require the MSO to reduce their modulation order by two levels from 4096QAM (with a12 bps/Hz raw spectral efficiency) to 1024QAM (with a 10 bps/Hz raw spectral efficiency). This yields a 10/12 reduction in spectral efficiency. Even with this decreased modulation order, the gains from the spectral expansion out-weigh the losses from the lower modulation order, and the overall increase in bandwidth capacity would be given by 4*(10/12) = 3.33x, as shown in Fig. 11a. This increase would obviously be decreased for values of ε greater than zero (due to increased attenuation or distortion in the new spectrum), but nevertheless there might still be a sizable increase in bandwidth capacity resulting from the spectral expansion (as experienced by the Telcos). Whether this effect is beneficial to MSOs or not would undoubtedly be a function of the cost of the upgrade as well as.

Historically, MSOs have resisted plant upgrades that require them to expand the plant spectrum. Even as they are currently discussing spectral expansions for DOCSIS 3.1, most MSOs seemingly avoid consideration of spectral widths that might exceed 1.2 or 1.7 GHz. Telco history implies that MSOs may want to consider using this Extended-Spectrum expansion in all three of the media on which they will likely be transmitting signals in the future- on the HFC plant, on FTTP plant, and over Wireless.

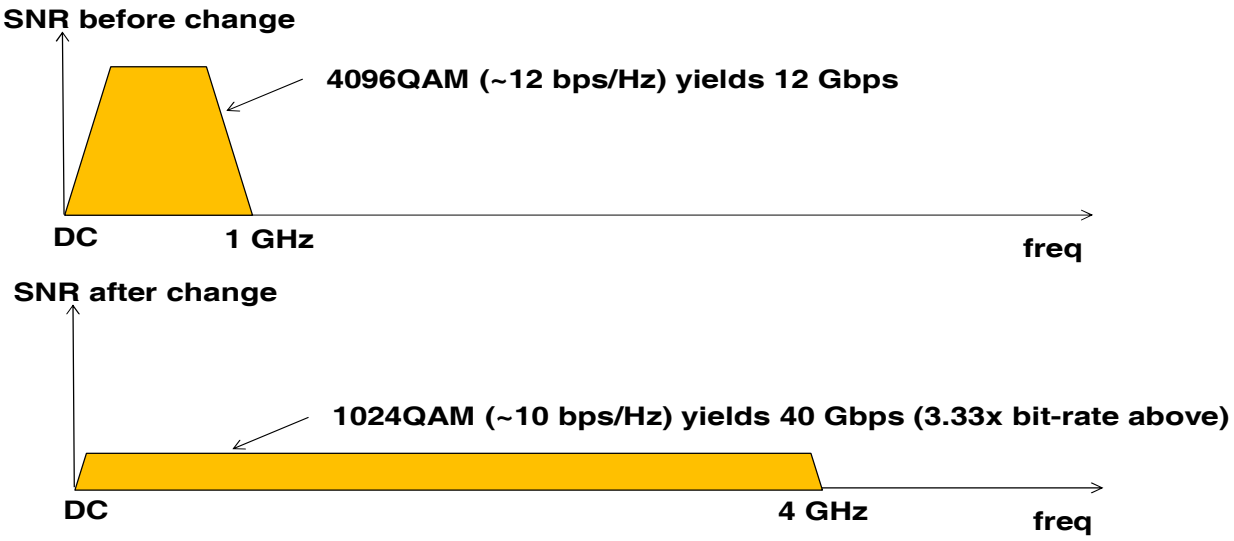


Fig. 11a- Benefits of Spectral Expansion

Lesson B2: Increase the reach of digital fiber runs and decrease the reach of the twisted-pair (copper) access network to decrease attenuation and maximize SNR and increase QAM Modulation levels. This practice was very common in the DSL space, and it was often implemented in conjunction with the spectrum expansion described in the previous lesson. Over a 21-year window of time, Telcos reduced their copper loop reach from 18,000 feet to 325 feet, while oftentimes adding more Ethernet or PON fiber length to reach the deeper DSLAMs. Each of the distance reduction operations enabled another round of spectral expansion and bandwidth capacity improvements to be supported on the existing twisted-pair infrastructure.

MSOs are already employing this technique (to some extent) with Fiber Deep architectures that extend the reach of the fiber and decrease the reach of the coaxial connections within the HFC network. However, they could choose to go to more extreme fiber depths (and shorter coaxial runs) to better mimic the actions of the Telcos. For example, Fig. 11b shows the potential of a Very Deep Fiber architecture with shared coaxial runs that might be limited to coaxial distances of less than 1000 feet.

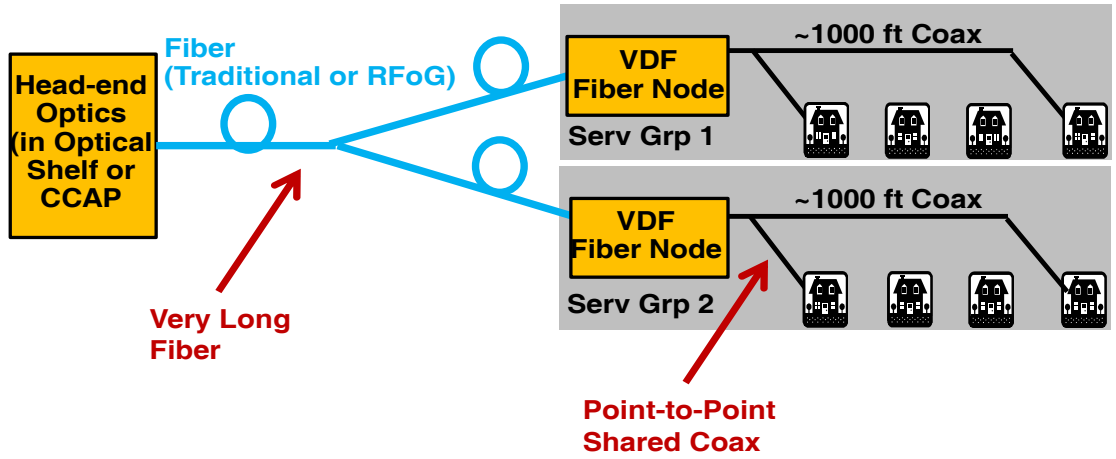


Fig. 11b- Very Deep Fiber (VDF) Shared Coax Technology

Lesson B3: Use "Selective Subscriber Shedding" to move heavy users to alternative transport technologies. This approach proposes to identify heavy users and move them off of the primary infrastructure onto an alternative infrastructure. This approach intrinsically leads to improvements in the performance of the primary transport technology. It is beginning to be used by Wireless providers as they create alternative technologies (picocells and femtocells) to carry heavy localized traffic without congesting the capacity on the primary macrocells in their Wireless infrastructure. The creation of the picocells and femtocells is essentially being done to protect the performance of their primary infrastructure investment (the macrocell). It basically allows heavy users to be moved to the small cells and reduces congestion on the marcrocells.

It is possible that MSOs may also want to adopt this policy for protecting the service level on their primary infrastructure investment (the HFC network) going forward. In particular, MSOs may want to provide an alternative infrastructure (such as a FTTP system or a Fiber-To-The-Curb (FTTC) system based on RFoG or PON) to which they can connect their high-usage and/or high-end Service Level Agreement (SLA) subscribers. If these subscribers are serviced by the alternate infrastructure, then their traffic is removed from the general HFC network, leaving the remaining customers with higher bandwidth capacities that are available to be shared by a smaller number of people. If all of the high-end SLA customers are removed from the network, this also gives the added benefit that the required bandwidth capacity for the service group (as given by Eq. 1) is reduced, because the Tmax term in the formula would be reduced to the next lower SLA level. This fact is illustrated in Fig. 12a.

Fig. 12b shows the benefits that might be obtained from the careful application of Selective Subscriber Shedding. If subscribers are only shed to alternative technologies when available HFC bandwidth capacities become overloaded and if the service tiers have different Tmax values as shown in Fig. 8b, it becomes clear that MSOs could offer services to a large percentage of their subscribers on their existing HFC plants deep into the 2030 time-frame. [ULM1]

Lesson B4: Use automated handoffs to permit efficient movement of traffic between wireless and wired technologies. MSOs have now begun to support Voice over LTE (VoLTE) to Voice over WiFi (VoWiFi) handoffs that give subscribers better performance levels inside of buildings and also protect their primary infrastructure (i.e., the macrocell) from congestion. [HAND]

MSOs may want to provide similar technologies to their subscribers. Once WiFi is provided by the MSOs, this convenient service will undoubtedly help reduce subscriber churn. It can be offered for voice, video, and data services.

Lesson B5: Focus roll-outs of new wired technologies on aerial plant systems where costs are minimized. Telcos have oftentimes focused their new feature rollouts on the easier-to-modify aerial plant systems, and later decided whether to continue the rollouts on buried plant. It is a convenient and cost-effective way to quickly roll out services-especially those that require new hardware.

MSOs have followed similar paths, and may want to use the same tricks in their own networks within the future. They can perform the required hardware modifications for future rollouts of Deep Fiber systems or Remote Gadget systems or FTTP systems or other systems on aerial plant systems first wherever permissible. Costs are reduced by following this approach. Once proven, the MSOs can then decide if they want to extend the changes to non-aerial systems (based on Return On Investment calculations).

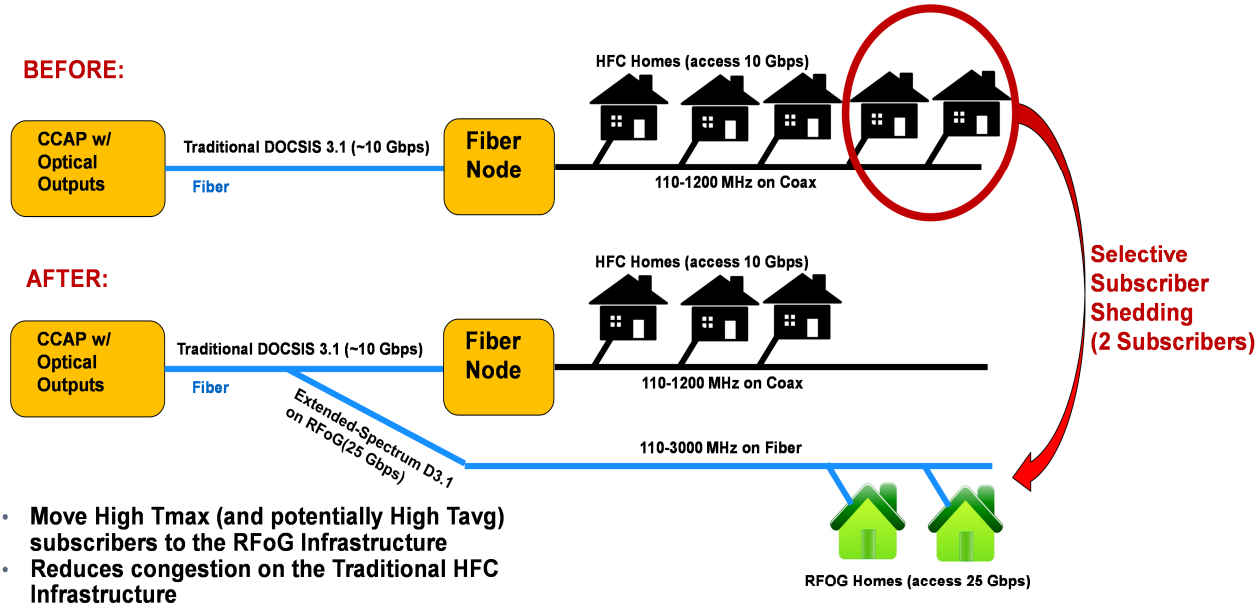


Fig. 12a- Selective Subscriber Shedding Example

<u>Lesson B6</u>: Use envelope-tracking, power-efficient amplifier sub-systems. These approaches have been optimized for applications within the LTE space, and Wireless vendors have placed great focus on distortion-correction within these systems. [JHAM]

These types of approaches may become applicable to cable head-end and access equipment in the future, because reductions in power consumption will become more and more important to MSOs over time.

<u>Category C: Lessons With A Low Probability Of Being Used By MSOs</u>

There were several lessons learned from Telco and Wireless providers that may be considered improbable (or even "crazy") for most MSOs to consider using in the future. Nevertheless, in the spirit of being open to all potential ideas and since it is feasible that some MSOs may find these low probability lessons to be interesting to explore, the lessons will be outlined in this section.

<u>Lesson C1</u>: Use fiber-to-copper media converters that are positioned extremely close to the homes to minimize twisted-pair (copper) attenuation. Telcos have been moving their DSLAMs and their G.Fast

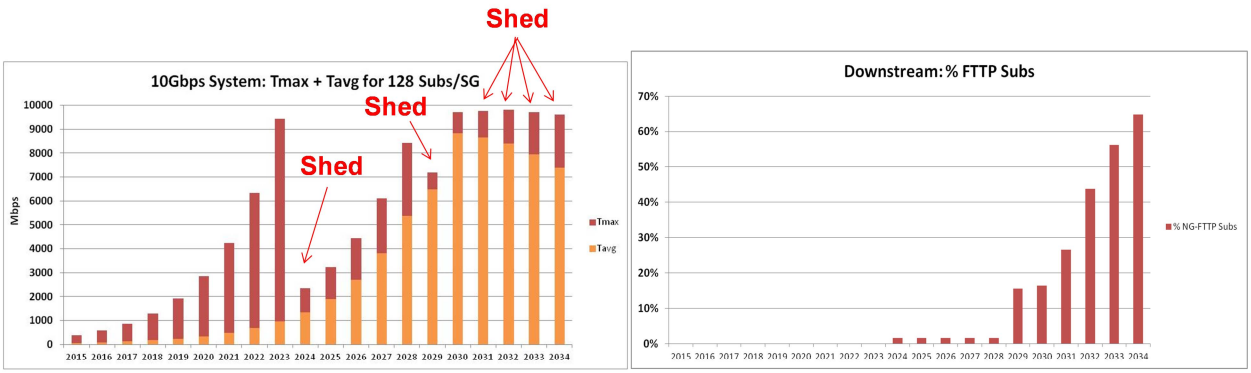


Fig. 12b- Bandwidth Required in HFC Service Group & Shed Subs in FTTH Service Group vs. Time

Distribution Points closer and closer to the home over time in an attempt to minimize signal loss and distortion in the twisted-pair portion of their transmission system. These moves have been costly, but the ability to continue to provide acceptable bandwidths to subscribers have warranted the moves, and the moves intentionally eliminated the expensive step of running new fiber connections all of the way to the house. As they make plans to move to their next generation of equipment (G.Fast), Telcos will be placing Distribution Points 300-600 feet from each house, and a single Distribution Point may only service 8-16 homes.

MSOs have also been moving their Fiber Nodes closer and closer to homes over time. However, Very Deep Fiber systems with coaxial reaches as short as 600 feet are seldom discussed or considered by most MSOs as they picture their future HFC networks in the 2020's. Perhaps those type of systems should be considered as MSOs move into the future. While this approach may seem extreme, it is feasible that some MSOs may find that business case analyses dictate that a similar path be followed for their HFC plants. In particular, for non-aerial plants in densely-populated urban areas where a plant re-build could require expensive city permits, the concept of continuing to re-use much of the already-deployed HFC coaxial infrastructure through backyards may prove appealing. This could lead MSOs to move Fiber Nodes much closer to the homes and having fewer and fewer homes per Fiber Node. The extension of the fiber required to support these Very Deep Fiber (VDF) systems could be accomplished in any one of several ways. It could be done by new trenching through existing easement areas. It could also be done in a less disruptive fashion using new technologies such as those that propose to core out the existing coax and create a conduit from the remaining plastic jacket where new fibers can be pulled [DEEP]. Whether MSOs will ever follow the lead of Telcos and have Fiber Nodes positioned within 600 feet of the houses and supporting only 8-16 houses is unclear. However, Very Deep Fiber (VDF) systems may very well find applications in the Cable space (for some MSOs) in the future.

Lesson C2: When necessary, draw electrical power from subscriber homes to power outside plant network equipment that is connected via copper connections to the homes. This approach (also known as reverse powering) is a controversial topic that is being considered by Telcos as they plan deployments of their G.Fast Distribution Points that are very close to the houses. The issue results from the fact that it is difficult to deliver power to the large number of Distribution Points that need to be positioned throughout the serviced neighborhoods. Whether subscriber-supplied power will be maketable or not is still up for debate. Those who argue that it may be marketable state that subscribers are already willing to supply the electrical power for their CPE equipment, and the Distribution Point could be envisioned to be an extension of the CPE equipment.

Only time will tell if this approach is marketable to consumers or not. If it is, then MSOs who choose to deploy Very Deep Fiber (VDF) systems could start to consider using electrical power delivered from the houses and over the coax as a source of power for their Very Deep Fiber Nodes.

Lesson C3: Use bonding of multiple physical connections into each home to boost the per-subscriber bandwidth capacity into the home. This approach to delivering a higher bandwidth service to homes has been common-place within the Telco space for many years. Two or more physical twisted-pair lines were oftentimes pulled into homes or businesses to offer multi-line voice service. In addition, ISDN used bonded twisted-pairs to offer higher bandwidth capacities. Different physical DSL lines have also been bonded by the Telcos in

the past to increase total per-subscriber bandwidth capacity. In all of these cases, Telcos found it beneficial to physically pull a second line into the home to offer the required bandwidth capacity. [2LIN] [ISDN] [BOND]

MSOs have already realized the benefits of a limited amount of bonding through the use of channel bonding in DOCSIS 3.0 systems. That form of bonding permitted two or more channels on a *single* coax to act as a single logical channel, increasing the total bandwidth capacity injected into the home. However, DOCSIS 3.0 channel bonding was typically limited to operation on a single physical coax entering the home.

To better mimic the behaviors of the Telcos, MSOs would have to be willing to drop two or more coaxial cables into a single home. Each additional drop coax added to a home would obviously increase the Operation Expenses associated with the installation, but once installed, it would provide another spectrum's worth of bandwidth capacity into the home, as shown in Fig. 13.

Alternatively, one coax into the home could be dedicated to Downstream bandwidth while the second coax into the home could be dedicated to Upstream bandwidth, which would be quite compatible with Extended-Spectrum RFoG systems of the future. Another form of potential bonding that could be explored by MSOs in the future might be a bonding group created by the combination of the existing coaxial cable drop and a Fiber drop and a WiFi drop running in parallel into the home. Steering of packets across the three links of different technologies would have to be managed by a "super CCAP" box either in the head-end or in a Remote Gadget node. (See Fig. 14). Whether or not this type of Cable broadband delivery into the home is desirable to MSOs will undoubtedly be determined by business case analyses that look at the cost and the ultimate Return On Investment.

<u>Lesson C4</u>: Use point-to-point (un-shared) copper connectivity into each subscriber's home to maximize each subscriber's bandwidth capacity. This is not a new innovation recently discovered by the Telcos; this was the standard mode of operation since telephone service deployments were initially deployed in the 1800's- one home, one twisted-pair loop. (Note: Party lines did share twisted-pair loops, but they were not the norm for most Telco

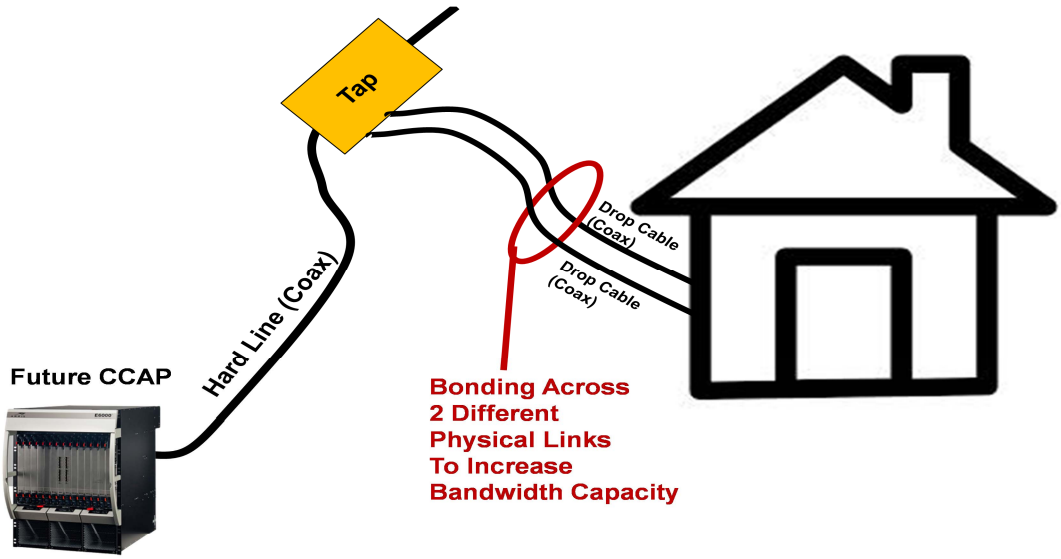


Fig. 13- Bonding Across Multiple Physical Links

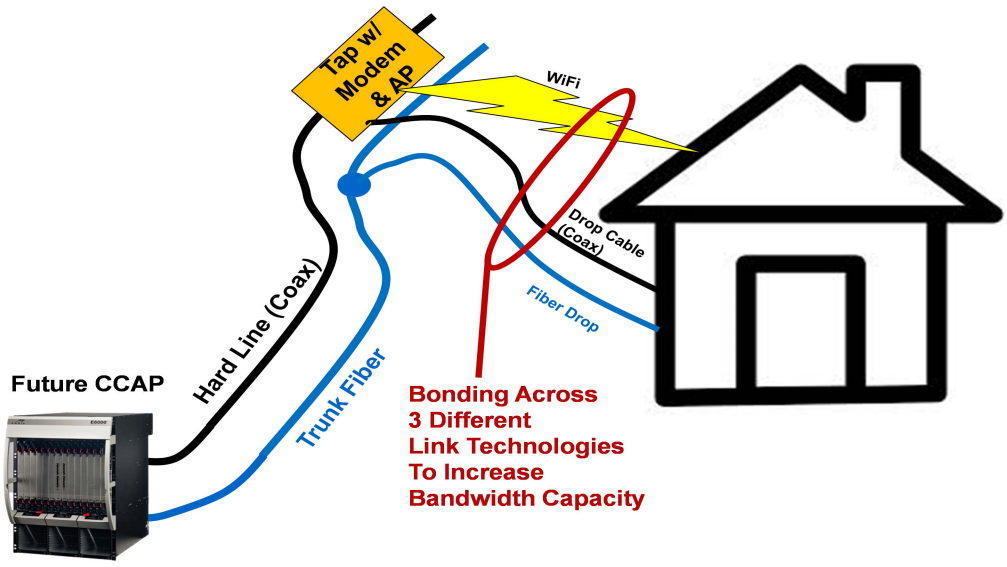


Fig. 14- Bonding Across Multiple Link Technologies

deployments). The disadvantage of un-shared copper connectivity was that many twisted-pair loops had to be deployed. However, the benefit of un-shared copper connectivity is that the bandwidth into a home is fully useable by the subscribers within that home. Thus, as the required per subscriber bandwidth capacity became larger and became more difficult to pass over a twisted-pair connection, Telcos were aided by the fact that the bandwidth on a single twisted-pair connection only needed to satisfy a single subscriber home. In addition, a "home-run" connection directly from the DSLAM or Distribution Point into an in-home modem minimizes signal degradation and micro-reflections.

MSOs may resist transitioning to this extreme form of point-to-point architecture that utilizes un-shared connections. It is very different from the current shared approach that has characterized HFC architectures since their inception. The transition would probably require large Operation Expenses to be committed to the transition. However, if MSOs did want to consider this approach as a means of capitalizing on that last 200 feet of coaxial drop cable that is already buried underground in the backyards of most homes, then they could choose to do it in at least one of two ways.

The first technique could use a "Drop from the Node" architecture that extends every point-to-point drop coax all the way back to a modified multi-drop Fiber Node, and the modified multi-drop Fiber Node would have to have a larger number of RF ports than is common today. (See Fig. 15).

This approach might be called the Very Deep Fiber (or VDF) point-to-point technology, and it is probably most applicable when an MSO has finally deployed a Node+0 architecture that eliminates all amplifiers within the coaxial portion of the network. For example, if future multi-drop VDF Fiber Nodes were moved deeper into the network and had (say) 32 homes connected to each multi-drop Fiber Node, then the multi-drop VDF Fiber Node would have to support 32 distinct and separate RF ports, where each RF port could feed an un-shared, point-to-point coaxial link to exactly one home. Each VDF Fiber Node could be assigned its own pair of lambdas on the fiber (for downstream and upstream transmissions), or RFoG OBI

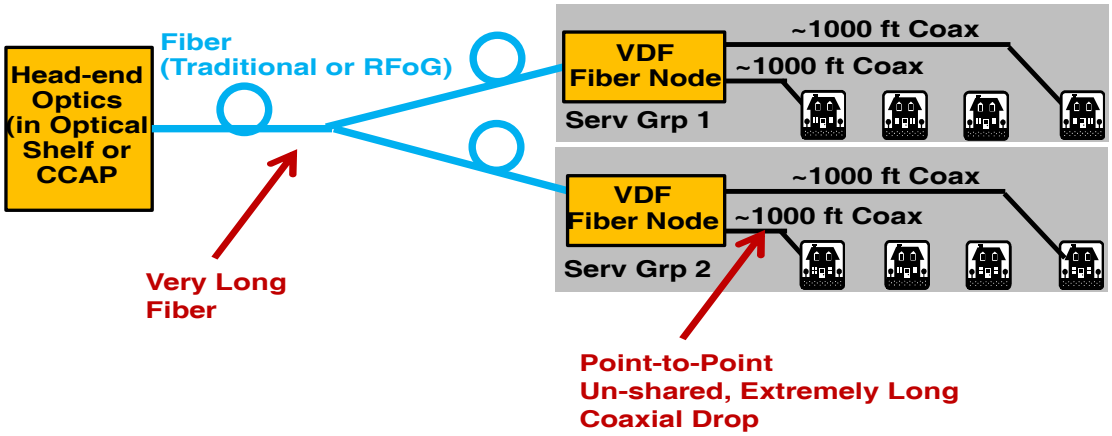


Fig. 15- Very Deep Fiber (VDF) Point-to-Point Technology

Mitigation techniques (such as Hybrid-PON technology) could be employed in the upstream direction. RFoG could then be used to permit multiple VDF Fiber Nodes to share a pair of lambdas on the fiber (which helps to conserve lambdas on the fiber). This VDF approach would obviously require the investment of a coaxial pull from each tap back to the multi-drop VDF Fiber Node (or the installation of a coaxial bundle from the multi-drop VDF Fiber Node to each of the tap locations). This VDF technique would have longer coaxial runs than a typical drop coax, so greater signal attenuation could become an issue with the VDF approach that could limit overall bandwidths.

The second technique circumvents the attenuation issues caused by the long coaxial runs of the first technique. This second technique could use a "Fiber To The Tap" (FTTT) architecture that delivers the fiber all the way to the tap in the backyard. (See Fig. 16).

The approach might be called the Extremely Deep Fiber (or EDF) point-to-point technology. It would obviously require the investment of a fiber pull through the neighborhoods and would require the replacement of taps with small EDF Fiber Node-like devices. Each EDF Fiber Node could be assigned its own pair of lambdas on the fiber (for downstream and upstream transmissions). Alternatively, RFoG Optical Beat Interference (OBI) Mitigation techniques (such as Hybrid-PON technology) could be employed in the upstream direction and RFoG could then be used to permit multiple EDF Fiber Nodes to share a single downstream

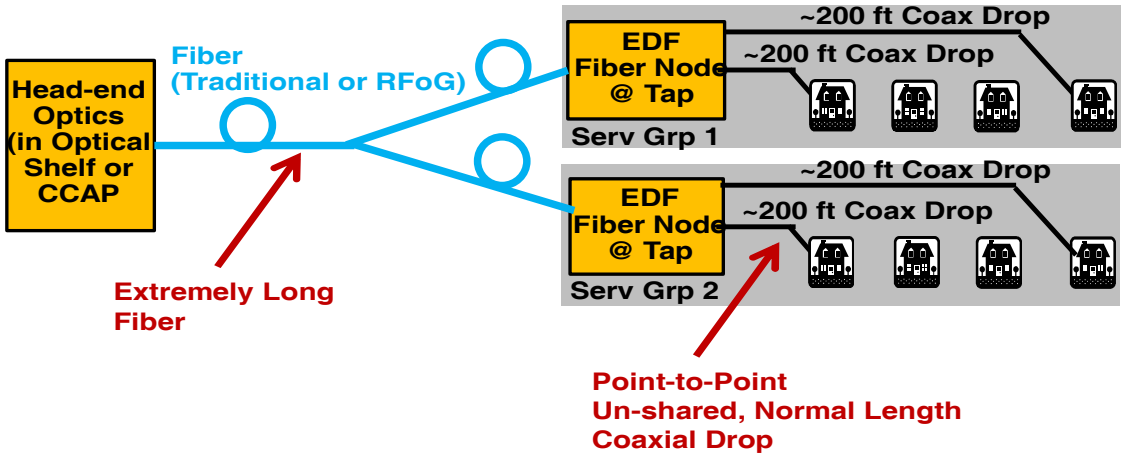


Fig. 16- Extremely Deep Fiber (EDF) Point-to-Point Technology

lambda and a single upstream lambda on the fiber (which helps to conserve lambdas on the fiber). [MUTA]

Although this EDF approach may not be appealing to some MSOs, other MSOs may find the idea interesting. As a result, the authors decided to briefly explore the point-to-point distance limits of typical coaxial drop connections in a little more detail. Several simplifying assumption can be made. These include:

1. The only sources of noise in the EDF-based point-to-point coaxial drop to the home is the Johnson-Nyquist thermal noise floor at the receiver and the receiver noise itself [JOHN]
2. The system temperature is 290K
3. The total power injected into each point-to-point link from the EDF Fiber Node source is maintained at a fairly reasonable level of (say) 50 dBmV with a CNR of 45 dB.
4. The transmitted power is uniformly distributed over the occupied spectrum (i.e. a flat Power Spectral Density). This is not necessarily the optimal PSD pattern as tilt added to the spectrum could help, but it does greatly simplify the calculations.
5. The coaxial drop cable is assumed to be RG-6. 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-d loss in the dielectric.
6. There is a "Coding Loss" of 4 dB associated with the particular physical layer channel code used (DOCSIS 3.1 OFDM). This is the difference between the performance (information capacity) of a perfect code, given by the Shannon-Hartley formula, and the performance of the channel code used. While highly optimized codes (e.g. Turbo Codes and some long-codeword LDPC variants) can in theory come within a fraction of a dB of the Shannon-Hartley limit, a more typical coding loss of 4 dB is appropriate when considering DOCSIS 3.1.
7. The receiver's input stage has a noise figure of 5 dB. Of course, low-noise amplifiers are available with much better noise figures than this, but 5 dB would be achievable using relatively low-cost components.
8. The receiver's demodulator has an implementation margin of 3 dB. 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.
9. The physical layer has a "Temporal-Spectral Efficiency" of 90%. This refers to the loss of capacity associated with guard bands, pilots, signaling channels, and cyclic prefixes.
10. There is a capacity loss of 5% associated with framing overheads and MAC messaging.
11. The calculated capacity is the total bitrate in either direction; it may be split in any ratio between upstream and downstream traffic.

Using these simplifying assumptions, the resulting SNRs and supported modulation levels and resulting bandwidth capacities can be calculated for different spectral widths and different distances of typical RG-6 drop cables. The resulting bandwidth capacity is shown in Fig. 17a. The corresponding Modulation Order (in bps/Hz) are also shown in Fig. 17b, where the QAM order can be calculated from the Modulation Order using the formula QAM Order = $2^{(\text{Modulation Order})}$.

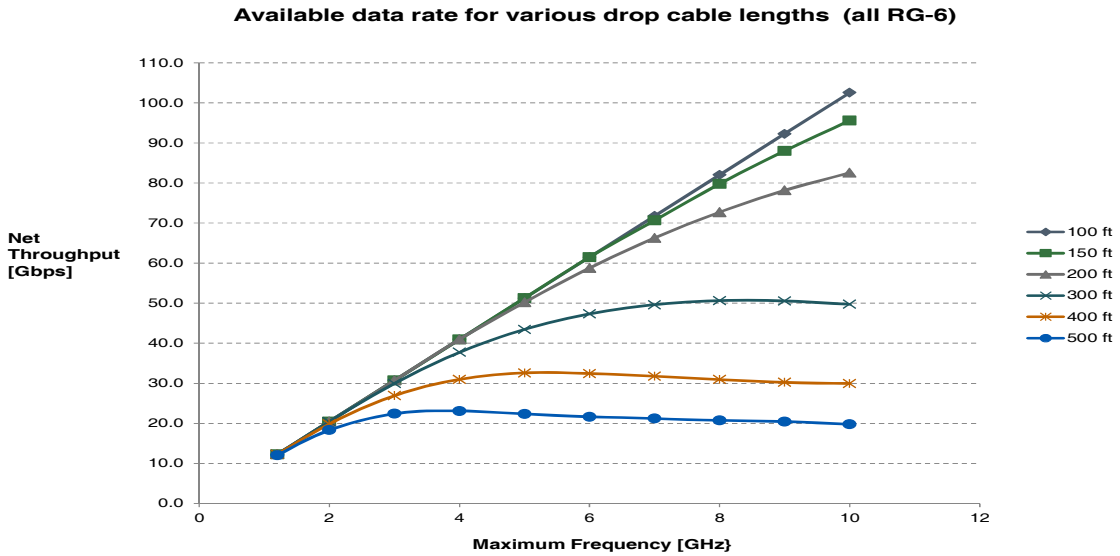


Fig. 17a- RG6 Coax Throughput Capacity vs Frequency (for Various Distances)

From these figures, it can be seen that (for the simplifying assumptions described above) there is a definite bandwidth capacity benefit (along with an unspecified cost) associated with the repeated doubling of the spectrum. This benefit is realized until the Modulation Order drops to a value of 1 (which corresponds to BPSK modulation).

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 10 GHz-wide OFDM spectrum would apparently be capable of transmitting ~20 Gbps of throughput from the EDF Fiber Node to the home. A 300 ft. length of RG-6 carrying a 10 GHz-wide OFDM spectrum would apparently be capable of transmitting ~50 Gbps of throughput to the

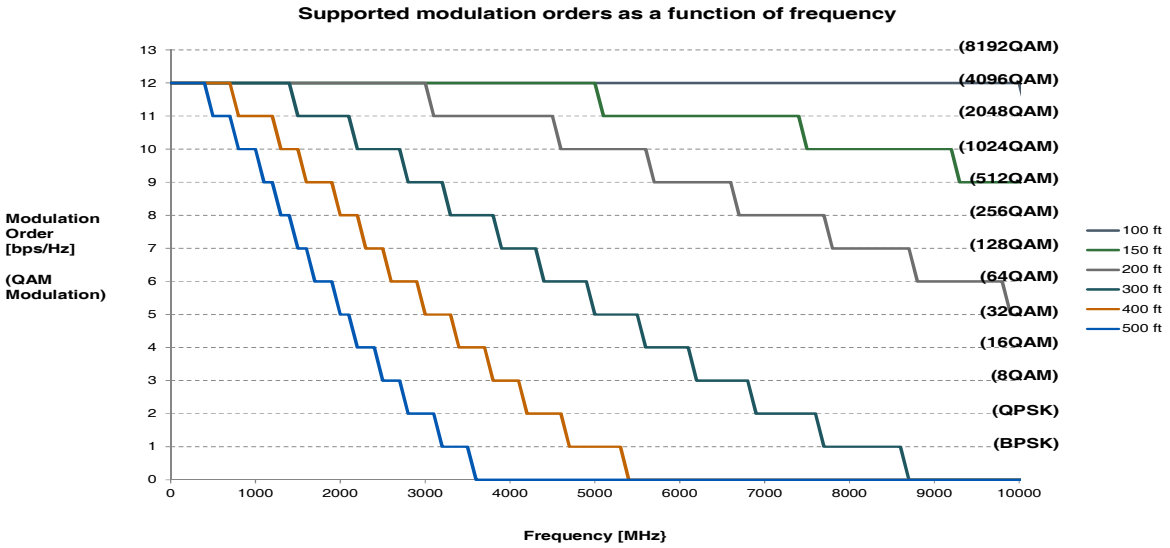


Fig. 17b- RG6 Coax Modulation Order vs Frequency (for Various Distances)

It is compelling to note the relatively high throughputs that seem to be achievable across

home. A 200 ft. length of RG-6 carrying a 10 GHz-wide OFDM spectrum would apparently be capable of transmitting ~80 Gbps to the

home. A 100 ft. length of RG-6 carrying a 10 GHz-wide OFDM would apparently be capable of transmitting ~100 Gbps to the home. While second-order effects would likely reduce these bandwidth capacities to some extent, the message is clear. Support for relatively large throughputs are likely to be possible on reasonable lengths of drop cables into the homes.

Lesson C5: Offer Wireless-only and Wireless+Wired connectivity to subscribers inside of their homes. Telcos obviously moved to the support of Wireless connectivity many years ago with their cellular infrastructure, and they are augmenting their offerings with small cell offerings (ex: picocells and femtocells) as well as WiFi offerings. Many of their subscribers are "cutting the cord" on wireline service and eliminating the need for a wired infrastructure into their homes, choosing instead to use only the wireless cellular connections. Other subscribers are using both the wireless and the original wired infrastructure to receive services inside of the home.

While the current HFC wired infrastructure can support most customers needs into the late 2020's and beyond, an eventual move to these Wireless technologies is likely to be a part of every MSO's future. However, instead of using cellular phone and LTE technologies, MSOs may opt to use WiFi (or similar LTE-U) as their primary method of delivering the Wireless service. This may make sense, as all wireless providers are gradually moving to small cell (ex: picocell or femtocell) technologies in the future.

WiFi can be used in a similar fashion by MSOs who may have an almost ubiquitous coverage area in urban areas. This coverage can be provided by Access Points in every home gateway and potentially by Access Points placed in Fiber Nodes and amplifiers and taps along their HFC plant runs. This WiFi coverage can be further enhanced through the use of Very Deep Fiber (VDF) or Extremely Deep Fiber (EDF) technologies. When placed in taps, this WiFi technology could also be used to provide wireless drop capabilities that augment the normal coaxial drop cables. The channel bonding techniques described earlier could then be used to provide bonding across the coaxial and WiFi connections. [ALB3]

Lesson C6: Use Self Organizing Network concepts to manage multiple overlapping Wireless networks. Telcos are beginning to spend considerable time studying Self Organizing Network (SON) concepts to dynamically and automatically manage the transmission frequencies and power levels and directionality of each antenna in their 4G Wireless networks. SON technologies can be used to help with self configuration, self optimization, and self healing operations. These technologies will undoubtedly be beneficial for the management of small cell environments where overlapping antenna coverage areas are common.

MSOs may be able to capitalize on these same SON technologies within their dense WiFi networks. This will become especially true as higher numbers of antennae become available on future access points, paving the way for more directionality and more Multi-User Multiple Input-Multiple Output (MU-MIMO) arrangements are utilized. SON technologies will undoubtedly help manage the many overlapping WiFi networks that will exist in the home and the neighborhood of the future. [SONT]

CONCLUSIONS

What Has Been Learned?

There are many lessons that MSOs can potentially learn from the past history of Telcos. These lessons include:

- Lesson A1: Use Orthogonal Frequency Division Multiplexing (OFDM) and Low Density Parity Check (LDPC) Codes
- Lesson A2: Assign different modulation orders to subscribers with different Signal-to-Noise Ratios (SNRs) to optimize throughput
- Lesson A3: Maximally utilize Moore's Law improvements in silicon and new linearizers/equalizers to permit signal reception with lower SNR and higher distortion
- Lesson A4: Use IP Video as a transport technology for television programming
- Lesson A5: Migrate appropriately-selected customers to Fiber-To-The-Premise (FTTP) service as their growing bandwidth needs necessitate this alternative service
- Lesson A6: Modify home networks to help increase the capacity of the access network
- Lesson A8: Use centralized, software-based management of network infrastructure platforms (ex: IMS & SDN)
- Lesson A9: Use Distributed Access Architectures to help deliver services.
- Lesson A10: Use Carrier-Class equipment to ensure high service availability
- Lesson B1: Expand channel spectrum to maximize channel capacity
- Lesson B2: Increase the reach of digital fiber runs and decrease the reach of the twisted-pair (copper) access network to decrease attenuation and maximize SNR and increase QAM Modulation levels
- Lesson B3: Use "Selective Subscriber Shedding" to move heavy users to alternative transport technologies
- Lesson B4: Use automated handoffs to permit efficient movement of traffic between wireless and wired technologies
- Lesson B5: Focus roll-outs of new wired technologies on aerial plant systems where costs are minimized
- Lesson C1: Use fiber-to-copper media converters that are positioned extremely close to the homes to minimize twisted-pair (copper) attenuation
- Lesson C2: When necessary, draw electrical power from subscriber homes to power outside plant network equipment that is connected via copper connections to the homes
- Lesson C3: Use bonding of multiple physical connections into each home to boost the per-subscriber bandwidth capacity into the home
- Lesson C4: Use point-to-point (un-shared) copper connectivity into each subscriber's home to maximize each subscriber's bandwidth capacity
- Lesson C5: Offer Wireless-only and Wireless+Wired connectivity to subscribers inside of their homes
- Lesson C6: Use Self Organizing Network concepts to manage multiple overlapping Wireless networks

In several ways, the MSOs are entering a period of time that may be similar to the period of time that Telcos have just navigated. The HFC plant has a long life left in it, but MSOs will still need to carefully utilize the resources of their HFC plant to extend its life and maximize the return on this already-deployed investment. And as they extend that life, MSOs will also need to begin phasing in new technologies for their future.

MSOs will undoubtedly be able to use at least some of the tactics and techniques adopted by Telcos over the past several decades. These tactics and techniques permitted Telcos to extend the life of their twisted-pair plant while they also introduced

new wireless and fiber-based technologies. The DSL techniques employed by Telcos on their twisted-pair loops allowed them to extend the life of their existing plant for at least twenty years beyond its originally-predicted end-of-life. If MSOs have similar levels of success with their HFC plants while they are introducing new technologies, then the MSOs may be able to extend the life of their HFC plants well into the 2030's or 2040' or beyond. It is quite likely that many MSOs will find themselves in a future environment similar to the environment of many Telcos today.

The HFC, RFoG, and HPON systems are likely to include some Extended-Spectrum DOCSIS solutions that can go beyond the 1.2 GHz spectral limits that are oftentimes discussed for DOCSIS 3.1 today. In fact, it is quite possible that many 10's of Gbps of bandwidth capacity could be provided using Extended-Spectrum DOCSIS solutions on all of these infrastructure types (HFC, RFoG, and HPON) in the future .

The Extended-Spectrum expansions permitted on the legacy HFC infrastructure will obviously be a function of the investmen

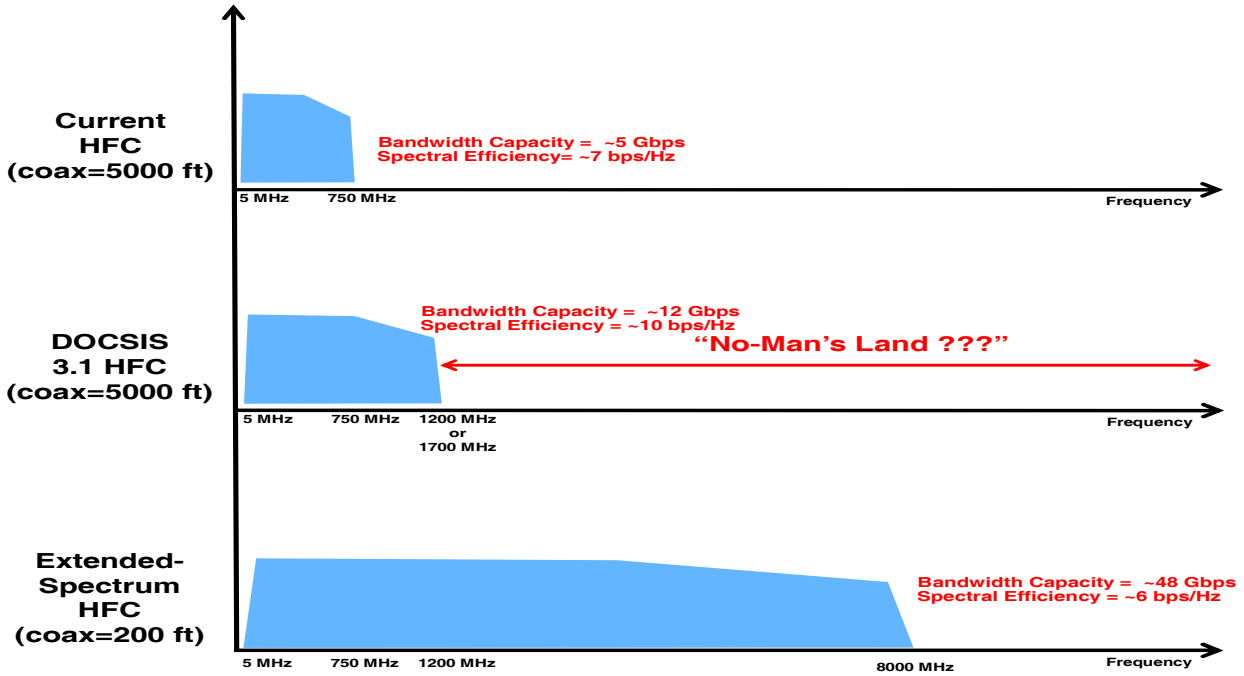


Fig. 18- Possible Future Views on Useable Spectrum for HFC Plant

This futuristic world may see many medium-level SLA subscribers and low-level SLA MSO subscribers still being serviced from the legacy HFC infrastructure, while some high-level SLA subscribers will likely begin to be serviced on new FTTP infrastructures (based RFoG or HPON or PON). The high-level SLA subscriber may also be moved to new Very Deep Fiber HFC and Extremely Deep Fiber HFC infrastructures.

levels that MSOs are willing to pour into the existing HFC plant. If they invest in it in a slow, continuous fashion (the way that Telcos did with their copper loop plant), then it is quite conceivable that the HFC plant of the future will contain very, very deep fiber runs and very, very short point-to-point coaxial drops to each home. These types of investments (if considered) could lead to an HFC evolutionary strategy described in Fig. 18. (Note: The expansion of HFC spectrum in Fig. 18 is quite similar to that which was seen for Telcos in Fig. 5a).

PON systems will also be utilized by MSOs in the future. These may be the technology of choice for the support of many business services subscribers. It may also be used by many MSOs for their residently subscribers as well. With PON systems also scaling to many 10's of Gbps in the future, these PON systems will also provide great bandwidth capacities for many FTTP subscribers.

In parallel with these HFC and RFoG and HPON and PON infrastructure improvements, a much more advanced WiFi or LTE-U Wireless network is also quite likely to develop within the MSO's network. This will be a dense system of Access Points that provide convenient connectivity to many of the subscribers in the future.

All of these paths mimic paths that have been taken by Telcos over the past several decades. As a result, it is quite possible that some of the lessons described within this paper may prove to be quite useful to MSOs in coming years.

REFERENCES

[10GE]http://en.wikipedia.org/wiki/10G-EPON

[1G4G]http://www.cs.ucsb.edu/~mturk/Courses/CS290I-2012/misc/1Gto4G.pdf

[2LIN]http://electronics.howstuffworks.com/two-line-phone.htm

[56KB]http://en.wikipedia.org/wiki/56_kbit/s_modem

[ABRA]http://www.mitr.p.lodz.pl/evu/lectures/Abramczyk3.pdf

[ALB1] Ayham Al-Banna et. al., "The Spectral Efficiency of DOCSIS 3.1 Systems," in Proceedings, SCTE Cable Tec Expo '14, SCTE, September, 2014.

[ALB2] Ayham Al-Banna et. al., "Measurement-Based EoL Stochastic Analysis and DOCSIS 3.1 Spectral Gain," in Proceedings, The NCTA Cable Show Spring Technical Forum, May, 2015.

[ALB3] Ayham Al-Banna, "WiMAX Links and OFDM Overlay for HFC Networks: Mobility and Higher US Capacity," in Proceedings, The NCTA Cable Show Spring Technical Forum, May 2010.

[BOND]http://www.netservers.net.uk/connectivity/bonded_adsl

[BRIT]http://www.britannica.com/EBchecked/topic/387079/modem

[CIOF]http://www.comsoc.org/files/Publications/Magazines/ci/hist-comm/2011-may.pdf

[CLOO] Tom Cloonan et. al., "Simulating the Impact of QoE on Per-Service Group HSD Bandwidth Capacity Requirements," in Proceedings, SCTE Cable Tec Expo '14, SCTE, September, 2014.

[COPP]http://www.copper.org/applications/telecomm/consumer/evolution.html

[DEEP]http://deepfibersolutions.com

[DIAL]http://www.tschmidt.com/writings/POTS_Modem_Impairments.htm

[DIGI]http://en.wikipedia.org/wiki/Digital_subscriber_line

[DSLT]http://www.broadband-forum.org/downloads/About_DSL.pdf

[EASY]http://easytel.com/services/history.asp

[EMM1] Mike Emmendorfer et. al, "Nielson's Law vs. Nielson TV Viewership for Network Capacity Planning," in Proceedings, The NCTA Cable Show Spring Technical Forum, April, 2014.

[EMM2] Mike Emmendorfer et. al, "Estimating Downstream Performance and DOCSIS 3.1 Capacity in CAA and DAA Systems," in Proceedings, The NCTA Cable Show Spring Technical Forum, May, 2015.

[ETHE]http://www.ethernetsummit.com/English/Collaterals/Proceedings/2012/20120222_T1-1B_DAmbrosia.pdf

[EVOL]http://www.ece.gatech.edu/research/labs/bwn/surveys/ltea.pdf

[FREN] Lou Frenzel, "What's the difference between EPON and GPON Optical Fiber Networks?", Electronic Design Magazine, Jan. 6, 2014.

[G989] ITU-T G.989.1 "40-Gigabit-capable passive optical networks (NG-PON2): General requirements", ITU-T, March 2013.

[GFAS]http://en.wikipedia.org/wiki/G.fast

[HAND]http://www.lightreading.com/mobile/carrier-wifi/t-mobile-turns-up-volte-to-wifi-handoff/d/d-id/710774

[HARA]http://www.eetimes.com/document.asp?doc_id=1272066

[HIST]http://www.historyofcomputercommunications.info/Book/1/1.1-BeginningsModemCompetition-CodexMilgo56-67.html

[HOOD] Dave Hood and Elmar Trojer, Gigabit-Capable Passive Optical Networks, John Wiley & Sons, Inc., 2012.

[JHAM] Junghyum Ham et. al., "A CMOS Envelope Tracking Power Amplifier for LTE Mobile Applications," in Journal of Semiconductor Technology and Science, Vol. 14, No. 2, April, 2014, http://www.jsts.org/html/journal/journal_files/2014/04/year2014volume14_02_13.pdf

[ISDN]http://en.wikipedia.org/wiki/Integrated_Services_Digital_Network

[JOHN]http://en.wikipedia.org/wiki/Johnson%E2%80%93Nyquist_noise

[KRAM] Glen Kramer, Ethernet Passive Optical Networks, McGraw Hill, 2005.

[LTEA]http://en.wikipedia.org/wiki/LTE_Advanced

[LTES]http://www.verizonwireless.com/mobile-living/network-and-plans/4g-lte-speeds-compared-to-home-network/

[MODE]http://inventors.about.com/library/inventors/blmodem.htm

[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, The NCTA Cable Show Spring Technical Forum, May, 2015.

[NETH]http://www.nethistory.info/History%20of%20the%20Internet/netsnmods.html

[OPTI]http://www.co.it.pt/seminarios/webcasting/itcbr_09_03_05.pdf

[PEAK]https://gigaom.com/2014/11/21/why-t-mobile-subscribers-in-21-cities-are-seeing-faster-speeds/

[SONT]http://en.wikipedia.org/wiki/Self-organizing_network

[TECH]http://www.digitimes.com.tw/tw/B2B/Seminar/Service/download/0519711180/DTF_9711182-3.pdf

[TELC]http://www.scribd.com/doc/205528065/Itu-Pon-Past-Present-Future#scribd

[ULM1] John Ulm et. al., "Is Nielsen Ready to Retire? Latest Developments in Bandwidth Capacity Planning," in Proceedings, SCTE Cable Tec Expo '14, SCTE, September, 2014.

[VDSL]http://www.analog.com/library/analogDialogue/archives/34-05/vdsl/index.html

[WEBF]http://webfoundation.org/about/vision/history-of-the-web/

[WHAT]http://whatis.techtarget.com/reference/Fast-Guide-to-DSL-Digital-Subscriber-Line

Leveraging IP ABR Video Delivery for Linear QAM TV

Weidong Mao
Comcast Cable

David Brouda
Imagine Communications

Abstract

As video service providers increase deployment of managed IP video service to various consumption devices (IPTV set-tops, tablets, smartphones, etc.) there has been an increase in unicast (HTTP) video delivery mechanisms using H.264 Adaptive Bit Rate (ABR) video encoding. At the same time, the MPEG-2-only legacy QAM set-top population is declining, leading to a tipping point in which the population of set-tops that are HD H.264-capable outnumbers the legacy devices.

The confluence of HTTP ABR/H.264-enabled devices and H.264 QAM set-tops leads to an interesting decision point: should service providers manage two core video platforms (IP and QAM) as many are doing today, or should these platforms converge at least at some level? The answer is "Yes." That is, the core should converge to HTTP, but QAM delivery to set-tops still need to be supported.

INTRODUCTION

Digital linear video distribution over QAM has been increasingly deployed over IP networks since the mid-to-late 2000's when edge QAMs with Gigabit Ethernet interfaces were first deployed for VOD and then broadcast video. While many edge QAMs are capable of performing multiplexing of MPEG-2 Transport Streams there are still dedicated platforms that groom and ad insert MPEG-2 and MPEG-4 video streams to desired bit rates for packaging on QAMs. This packaging might be in the form of Constant Bit Rate (CBR) or Capped Variable Bit Rate (VBR) in order to be delivered as Single Program Transport Streams (SPTS) for Switched Digital Video (SDV) or the packaging involves the multiplexing of more than one video stream into a Multi-Program Transport Stream (MPTS). Typically MPTS are not only multiplexed bit-for-bit but also statistically multiplexed to take advantage of variability in the encoded video across multiple programs such that there is an increase in efficiency without a decrease in video quality.

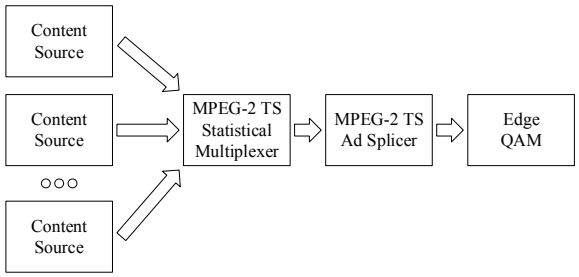


Figure 1 - Traditional Statistical Multiplexing & Ad Splicing

In addition to grooming for statistical multiplexing the function of ad splicing also takes place, see `Figure 1`; the platforms that perform these functions tend to be proprietary hardware platforms in which operational densities have not kept pace with technological innovations found in COTS (Common Off-The-Shelf) computing platforms.

QAM set-top boxes were originally deployed in support of Standard Definition (SD) MPEG-2 video. Subsequently High Definitional (HD) capable MPEG-2 set-tops were deployed followed by H.264-capable HD set-tops. While H.264-capable HD set-tops have been deployed for a number of

years most operators have not deployed much if any H.264 HD linear and VOD content for QAM video due to incompatibility of the content with the MPEG-2-only set-tops in the majority of customer homes. Driven by the desire to reclaim HFC bandwidth and reduce storage and network cost, the HD MPEG-2-only set-tops are starting to be replaced with H.264 capable HD set-tops. A tipping point will soon be reached at the vast majority of multichannel service providers that all the HD QAM set-tops are H.264 capable. This will allow conversion of existing linear or VOD MPEG-2 HD content to H.264 HD delivery. H.264 offers better quality and lower bit rates, but the traditional methods of statistical multiplexing and ad splicing are more complex in H.264 than MPEG-2; increased complexity directly correlates to increased cost and lower physical densities for processing. Service providers therefor have been challenged by a dilemma in which they want to save bandwidth but the cost and complexities to do so have not yet outweighed the benefits.

MULTICAST CORE NETWORK

In order to reduce operational complexities many service providers migrated their digital video network to using IP connectivity over Gigabit Ethernet. With encapsulating MPEG-2 Transport Streams in UDP/IP for delivery the best method for linear video distribution was multicast. IP multicast allowed operators to receive a linear channel at one site and distribute it simultaneously to many sites. Multicast IP networks were built using off-the-shelf networking hardware, which allowed service providers to migrate from their legacy point-to-point video distribution networks built on proprietary hardware platforms.

In support of redundancy, many service providers built two parallel IP network paths carrying duplicate multicast signals in case of network or video equipment failure. At the same time these same service providers also provide IP backbone and regional network for Internet access to their customers. Some providers keep these networks completely segmented from the core into their regional headends while others converge their networks. Segmented networks essentially are more costly to deploy, operate, and maintain while converged networks can be more complicated to operate particularly in the presense of multicast traffic. One operational complication is that multicast does not offer easy-to-implement security mechanisms in order to prevent multicast traffic from traversing networks where it may be administratively forbidden.

Some examples of administrative reasons for preventing the multicast from traversing specific networks might be: the network does not have sufficient bandwidth to carry all available multicast groups; without careful administration the multicast flows could easily make the network segment unusable due to congestion. Some multicast groups may be needed on the network but there are hosts or users on the network that are not permitted to receive traffic for security or business reasons. While Layer 2 and Layer 3 methods exist to solve these problems these methods are not easy to implement and configuration and operational errors can easily cause harm to the network or put business interests at risk.

As was mentioned, many operators will duplicate their video network paths in order to create a level of redundancy that protects from physical network failures. Since multicast implementations are a real time transmission of MPEG Transport Streams over UDP/IP there is no ability for receiving devices to requests retransmission of data that might a have been lost by the network. While both standard and proprietary methods in support of UDP transmission have been developed to support retransmission of lost data or recovery via adding Forward Error Correction, most

operators have not implemented these solutions and continue to operate primary and redundant video delivery network paths to maintain the desired level of Quality of Service (QoS). Receiving devices are responsible for determining which multicast signal is the better one to use; methods for making this determine can vary from implementation. In comparison, the duplications of equipments and network paths incur added cost and complexity with custom design that does not exist in general Internet technologies and standards.

In addition to lack of native resiliency in the real-time transmission of MPEG TS in UDP/IP there is also a lack of security. As was mentioned, when multicasting MPEG TS there can be operational difficulties to prevent one or more multicast streams from reaching an administratively forbidden destination; multicast can be disabled on a network segment but if some but not all multicast traffic is required to be present it can lead to operational complexities. Conditional Access and Digital Rights Management (DRM) are valid solutions to this problem but often do not scale well inside an operator core video network and these solutions are often operational complex.

A more modern solution leveraging HTTP unicast adaptive video delivery is ready to solve the challenges of deliverying linear video for QAM H.264 HD set-tops.

HTTP ADAPTIVE BIT RATE

Multi-screen video delivery has grown substantially in recent years. Rather than leverage any existing processing or delivery platforms in place for QAM, service providers have built parallel delivery platforms. These new platforms have taken advantage of COTS hardware and new methods for ad insertion that do not require proprietary hardware platforms to implement. However, other than leveraging the same content sources as are used for QAM delivery, multi-screen video platforms have been built in parallel to the traditional network (`Figure 1`).

Multi-screen linear video is encoded as H.264 and packaged and delivered over HTTP to consumer devices through a Content Delivery Network (CDN). Adaptive Bit Rate (ABR) is used to ensure a smooth playback experience for the consumer regardless of varying network conditions.

The video formats (or profiles) set as outputs of the ABR Transcoder have traditionally been targeted at small screens, as such the resolutions of the encoded video tend to be less than that of the original content source.

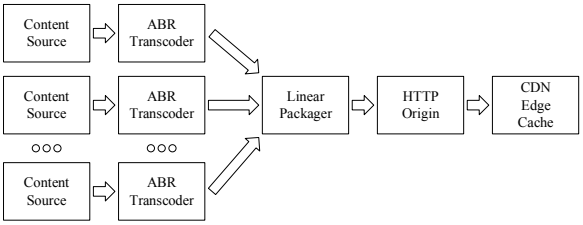


Figure 2 – Linear HTTP ABR Delivery

However as service providers consider supporting IP set-top boxes or higher-resolution displays such as computers or tablets it becomes necessary to retain the resolution of the content source but still deliver a seamless user experience during sub-optimal network conditions. As such, ABR is still a necessity.

As service providers have started to deploy cloud DVR (cDVR) services or linear streaming to big screens (connected TVs, game consoles, etc.) the resolution and bit rates have increased beyond what was originally deployed for small mobile screens. As the number of delivery formats (multiple delivery protocols and different DRMs) has increased service providers have turned to using Just In Time Packagers (JITP). JITP

ingests a single common format inside the service provider network and converts it to a device-specific format for delivery to a specific consumer platform. A similar approach can be taken with QAM such that a common linear input ABR format can be converted for delivery to a QAM set-top box.

As was discussed previously, UDP transmission of video is not protected against packet loss. Additionally, managing the flow of multicast traffic over specific router-to-router links can be challenging. With HTTP being based on TCP there are inherent retransmission capabilities. Widely deployed methods also exist to manage the traffic flows around the network with HTTP; CDN technology can be utilized in order to deliver content data only when requested. Other widely used IP network techniques for scalability and redundancy such as load-balancing, clustering, and virtualization are applicable for HTTP/TCP where UDP falls short.

A significant business challenge for service providers had been caused by technical limitations; securing multicast content on a core network has proven not only to be difficult but operationally complex. Securing that same content transmitted over HTTP can be as simple as using HTTP with TLS connections but added layers of security using software-based DRM are also feasible. Beyond a blanket protection of the content, the use of DRM and CDNs would allow service providers to easily manage and prevent where specific content is not authorized for distribution due to business constraints.

Even with HTTP based ABR IP video delivery in the core network that can be shared for both IP and QAM H.264 HD linear traffic, edge QAMs still require UDP/IP multicast carrying MPEG-2 Transport Streams as their inputs; separate or multiplexed audio and video segments delivered through ABR IP video at the core still need to be converted to MPEG-2 SPTS or MPTS format before delivering to the edge QAM devices. Linear content requires ads to be inserted before streamed to edge QAMs. It is also desirable to leverage the benefits of statistical multiplexing but without the burden of proprietary hardware platforms.

HTTP AD INSERTION

Standards-based methods have already been developed and deployed to simplify ad insertion in the HTTP realm as compared to ad splicing in Transport Streams. A common approach for HTTP ad insertion places additional stream processing responsibility on the ABR Transcoder and Linear Packager. The ABR Transcoder must detect SCTE-35 messages in-band and properly align the encoding frame structure such that a complete Group-of-Pictures (GOP) exists before and after the temporal locations of the ad (and not spread between the content and ad). The linear packager must also recognize ad placement locations and the boundaries created by the transcoder and so it can identify replaceable segments in the playlist it creates for the down-stream devices.

Now that the heavy lifting to cleanly mark where an ad starts and stops has been taken care of by the transcoder and packager any down-stream device need only evaluate the playlist and insert advertisements downloaded over HTTP that are similarly formatted with clean frame boundaries.

So, all of this exists, nothing is *new*. How do we get from an HTTP delivered ABR stream sitting on a CDN cache to an edge QAM that does not speak HTTP or understand ABR? The solution is simple: create a multiplexing gateway platform to convert between HTTP ABR and UDP/IP Transport Streams.

HTTP ABR TO UDP GATEWAY

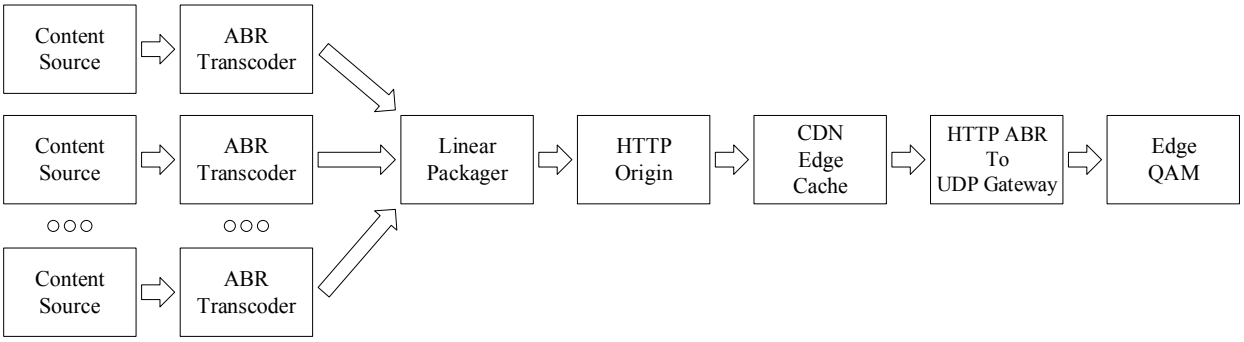


Figure 3 - HTTP ABR To QAM Gateway

The new HTTP ABR To UDP Gateway must act like an HTTP ABR client to ingest content including acting like an HTTP ABR ad insertion client. In the case of SPTS UDP/IP output the gateway simply needs to retrieve the designated ABR profile(s) from the CDN Edge Cache, perform the necessary HTTP ad insertion using the ad content profile that is less than or equal to the bit rate of the linear content ABR profile and then stream it to the edge QAM over UDP/IP as a Transport Stream. In the process any PID and Program Number remapping of the ad content should also be performed by the gateway.

MPTS output is a bit more complicated. While ad insertion is taken care of in the HTTP domain as previously described there is still a desire to take advantage of statistical multiplexing. However, doing so in proprietary hardware is not a scalable or desireable solution for many service providers. Luckily, the ABR transcoder and packager have already done the heavy lifting by creating a variety of bit rates for the same resolution video and marking the location of ads. Having the Gateway evaluate the ABR playlist by selecting an appropriate input bit rate profile more eloquently performs statistical multiplexing decisions compared to those previously performed in proprietary hardware.

Example

10 HD H.264 streams are configured to be multiplexed by the Gateway and transmitted as a 38.8 Mbps UDP/IP multicast stream which will be received by downstream edge QAMs. Simplistically as CBR encoded streams the average bit rate would be 3880 kbps. In this basic example (`Figure 4`) the highest encoded profile of each HD stream is actually 5500 kbps with additional profiles of 3750 kbps, 2750 kbps, and 2250 kbps. ABR content is packaged and placed on the CDN and delivered to the edge cache. The statistical multiplexing gateway retrieves the content from the edge cache as files over HTTP. With content encoded at fixed segment intervals (i.e. 2 seconds) the gateway is able to add up the segments from each of the 10 streams and ensure that they are always less than 38.8Mbps as shown in `Figure 4`.

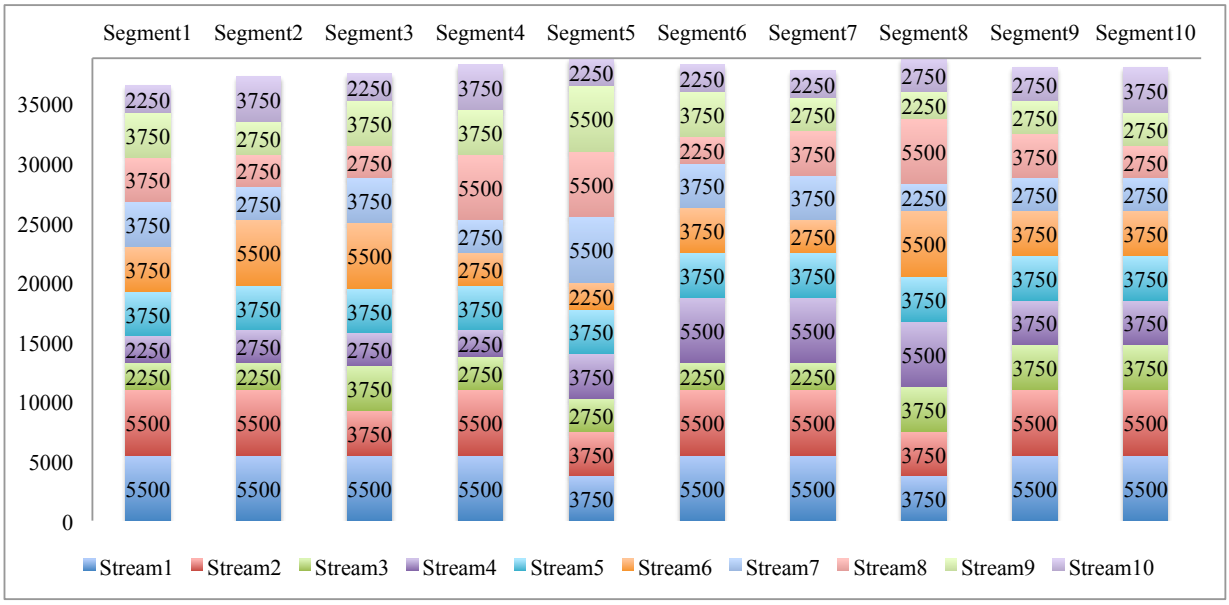


Figure 4 - Example ABR Stat-Mux

Increasing the number of profiles (bit rates increments) will allow for higher picture quality permitting the statistical multiplexing decisions to be more granular. Additionally, metadata regarding picture quality can be provided from the ABR transcoder downstream to the Gateway that will allow more intelligent multiplexing decisions about upcoming segments and which profiles to incorporate into the output.

Ad boundaries will add to the variability of the statistical multiplex since they will create scene changes (fade to black) and variable length segments. As ad content is transcoded offline in non-real-time it can be prepared with additional profiles that allow for more granular rate shifting during ad breaks which could potentially occur concurrently across multiple channels in a multiplex.

Implementation

The HTTP ABR to UDP Gateway acts like an ABR client and transmits the output as an MPEG-2 Transport Stream over UDP/IP. The profile decision process is different than a consumer-device ABR client as decisions are not made on available network bandwidth at the input but rather available bandwidth on the multiplexed output and is done in conjunction with multiple ABR streams simultaneously. These processes are not computationally complex and do not involve manipulation of the video payload like traditional statistical multiplex or ad splicing. While there is some basic manipulation of MPEG-2 Transport Stream data (Program Number remapping and PID remapping), and while concern must be given to the timing of multiplexing packets it is the expectation that all of these functions can be implemented on COTS computer server hardware at scale and can potentially be virtualized.

CONCLUSION

With the confluence of the increasing deployment of high resolution and bit rate HTTP ABR/H.264 and decline of MPEG-2 HD set-tops service providers can converge their core video distribution network to HTTP ABR, increasing bandwidth efficiency by eliminating redundant multicast core video networks for linear video delivery to H.264 HD QAM set-tops. Content security is increased and QoS is natively provided. Not

only does using HTTP delivery reduce the overall network management but also it unifies ad insertion into a common platform and simplifies statistical multiplexing at the same time. HTTP ABR delivery is the next evolution of the multichannel service provider's linear broadcast core video network.

MEASUREMENT-BASED EOL STOCHASTIC ANALYSIS AND DOCSIS 3.1 SPECTRAL GAIN

Ayham Al-Banna, David Bowler, Xinfa Ma
ARRIS Group, Inc.

Abstract

Achieving the promised DOCSIS 3.1 capacities is a function of the network architecture and its EoL performance. This paper analyzes the network End of Line (EoL) performance theoretically and stochastically based on location-aware real-world plant measurements. Computer simulations and laboratory experiments are also used to support the field measurements in characterizing the performance of different network segments (e.g., source, fiber links, coaxial cascades, and tap/drop/in-home network/modem) with an ultimate goal of identifying the network pieces that can potentially dominate the network EoL performance. The performance metrics of the network segments are then combined to yield an EoL performance range for multiple network architectures (e.g., centralized N+6/ N+3/ N+0 and distributed N+0/ N+3/ N+6 architectures). Finally, DOCSIS 3.1 spectral efficiency analysis is performed for the various network architectures given their EoL performance ranges.

1. INTRODUCTION

Traffic demand has been continuously grwoing over the past two decades triggering signifcant research efforts to find ways to augment the capacity of Hybrid Fiber Coax (HFC) networks [1] [2] [3] [4] [5] [6]. As MSOs prepare their networks for DOCSIS 3.1, they keep wondering about the potential capacity gain that can be obtained as they deploy DOCSIS 3.1 in their networks. While the question may sound simple, the answer can potentially change from one MSO to another as various MSOs could have different networks in terms of:

1. Architecture: fiber link length, number of wavelengths on the fiber link, number of amplifiers in cascade, coaxial cables length, etc.
2. Components performance, age, and supported spectral limits
3. Plant noise and interference
4. Subscribers' in-home network (IHN) architecture
5. MSO maintenance practices
6. Others

It is important to note that the network EoL performance depends on the above parameters. In particular, estimating the network capacity offered by DOCSIS 3.1 requires detailed study of the network in question, where the performance of different segments of the network (source, fiber portion, coaxial portion, in-home network, CM performance) are characterized using field measurements.

Previous analyses were performed to estimate the potential DOCSIS 3.1 capacity gain based on measurements collected from 20 Million CMs on Comcast network [7]. Since the number of data points is large, the theory of large numbers holds true such that the analysis would represent the potential average capacity gain across the whole network fairly well. The DS and US results of those analyses were slightly updated and are summarized in Table 1 and Table 2, respectively. Figure 1 and Fig. 2 are also included to show the DS and US DOCSIS 3.1 spectral efficiency analysis with MSO Signal

to Noise Ratio (SNR) operating margin of 0 dB, respectively [7].

The analysis presented in this paper is a continuation of the study summarized above, where network-specific analysis is performed. In particular, instead of performing the analysis based on measurements collected via large number of CMs located in different regions with various network architectures, the goal of the analysis included in this article is to perform DOCSIS 3.1 spectral gain analysis for a network of a specific configurable architecture. The initial thought was to obtain a probability density function (pdf) that describes the SNR performance of each of the network portions (e.g., source, various fiber links, multiple coaxial cascades/topologies, in-home network) through characterization based on theory, computer simulations, and field measurements.

\Table 1. Average DOCSIS 3.1 DS spectral efficiency and percentage improvement over DOCSIS 3.0

MSO SNR Operating Margin (dB)	D3.1 DS Spectral Efficiency (bps/Hz)		% improvement over D3.0 in DS Direction	
	4K FFT	8K FFT	4K FFT	8K FFT
0	7.6106	8.212217	20%	30%
1	7.36889	7.951399	16%	26%
2	7.129489	7.693074	13%	22%
3	6.899551	7.44496	9%	18%
4	6.688912	7.217669	6%	14%

Table 2. Average DOCSIS 3.1 US spectral efficiency and percentage improvement over DOCSIS 3.0

MSO SNR Operating Margin (dB)	D3.1 US Spectral Efficiency (bps/Hz)		% improvement over D3.0 in US Direction	
	2K FFT	4K FFT	2K FFT	4K FFT
0	6.65631	7.047349	60%	70%
1	6.41872	6.795801	55%	64%
2	6.1879	6.551421	49%	58%
3	5.957387	6.307366	43%	52%
4	5.721412	6.057529	37%	45%

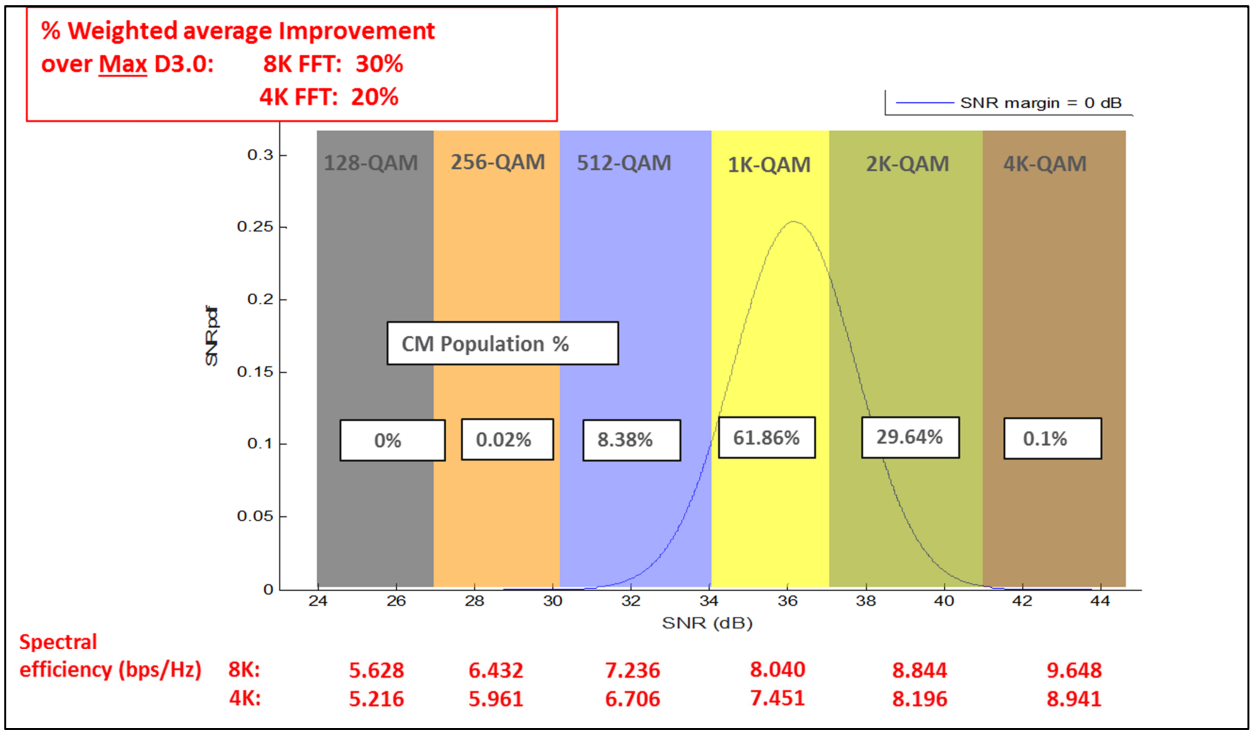


Figure 1. DOCSIS 3.1 Spectral efficiency analysis
with MSO SNR operating margin = 0 dB (DS Direction)

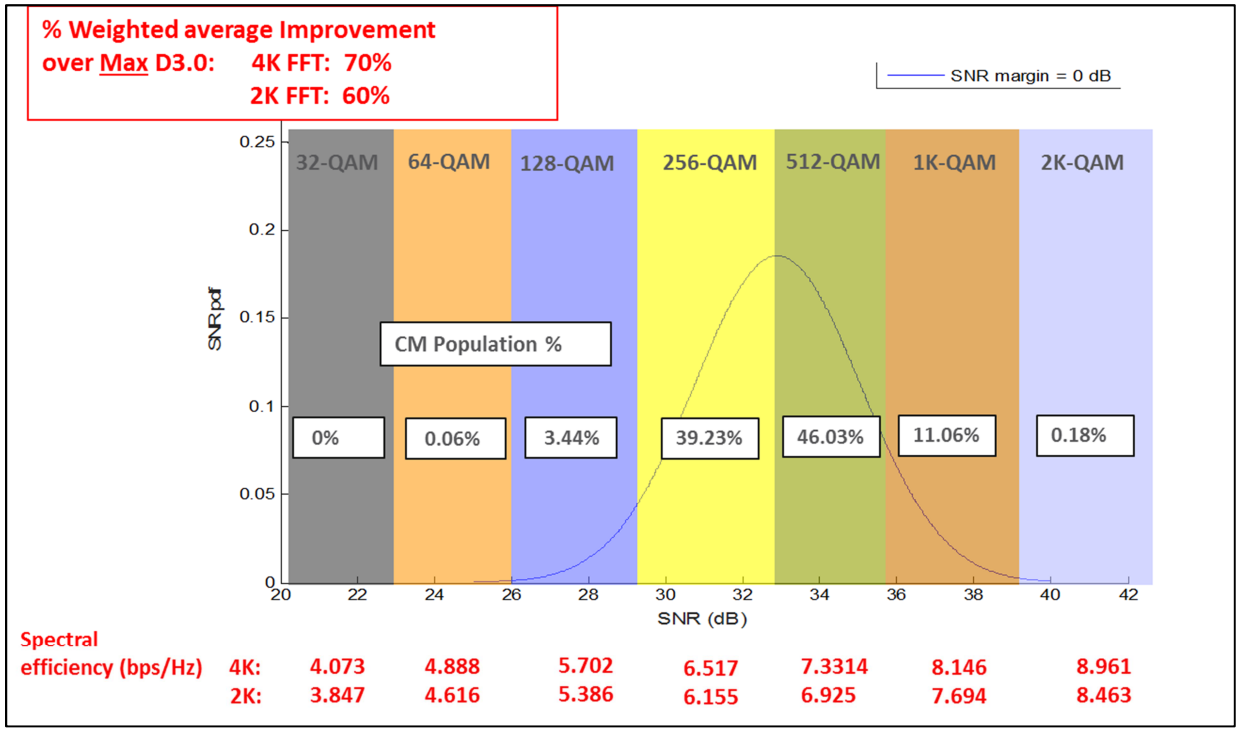


Figure 2. DOCSIS 3.1 Spectral efficiency analysis
with MSO SNR operating margin = 0 dB (US Direction)

The measurements that form the basis of the analysis presented this article were collected in collaboration with one of ARRIS' partner MSOs. What's unique about these collected measurements is that they are *location-aware* in the sense that the location of the CM reporting those measurements is known within the network topology. The analysis, which will be described later in this document, is not solely based on Modulation Error Ratio (MER) measurements but also takes Codeword Error Rate (CER) measurements into consideration due to the fact that MER measurements alone may not be sufficient to perform comprehensive analysis in some scenarios that will be explained later in the article. Other measurements (e.g., Tx power, Rx Power) were also collected to contribute to the analysis.

Performing an analysis on the collected field measurement revealed new facts manifesting a real-world twist to the above initial theoretical-oriented approach, where a network-specific pdf is generated based on convolving multiple pdfs. The results will show how the performance of multiple CMs on the same network (even at similar topology location) can be very different. In particular, the analysis of this paper will present scenarios where some portions of the network can actually dominate the whole network performance. Therefore, changing the network architecture (reducing the cascade length or moving to distributed architectures) does not necessarily add performance gain in those situations. In a nutshell, it will be shown that different CMs at similar locations in the same network can perceive the change in network architecture differently.

This paper is organized as follows. Section 2 describes the theoretical foundation for the analysis. Different cascade depths are characterized in section 3, where MER-based stochastic EoL performance analysis and potential D3.1 spectral efficiency gains analysis are presented. Section 4 shows the FEC analysis and relates that back to the MER-based analysis presented in Section 3. Different network segments are characterized and system EoL performance is generated for multiple architectures along with their corresponding DOCSIS 3.1 spectral efficiency gain in Section 5. Section 6 covers network optimization techniques that can be deployed to identify and resolve the performance limitations of HFC networks. Finally, Section 7 concludes the paper.

2. THEORITICAL FOUNDATION OF MER ANALYSIS

Assume that CNR_{EoL} represents the EoL Carrier to Noise Ratio (CNR) value (in dB) of a network that is composed of multiple segments (e.g., source, fiber, coax, tap/drop/In-home network/modem), as illustrated in Fig. 3. It can be shown that CNR_{EoL} is given by [8]

$$CNR_{EoL} = -10\log\left(10^{\frac{-CNR_S}{10}} + 10^{\frac{-CNR_F}{10}} + 10^{\frac{-CNR_C}{10}} + 10^{\frac{-CNR_{TDHM}}{10}}\right), \quad (1)$$

Where CNR_s, CNR_F, CNR_C, and CNR_{TDHM} represent the CNR value, in dB, of the source, Fiber, Coax, and Tap/ Drop/ Home-Network/ Modem (TDHM) pieces, respectively.

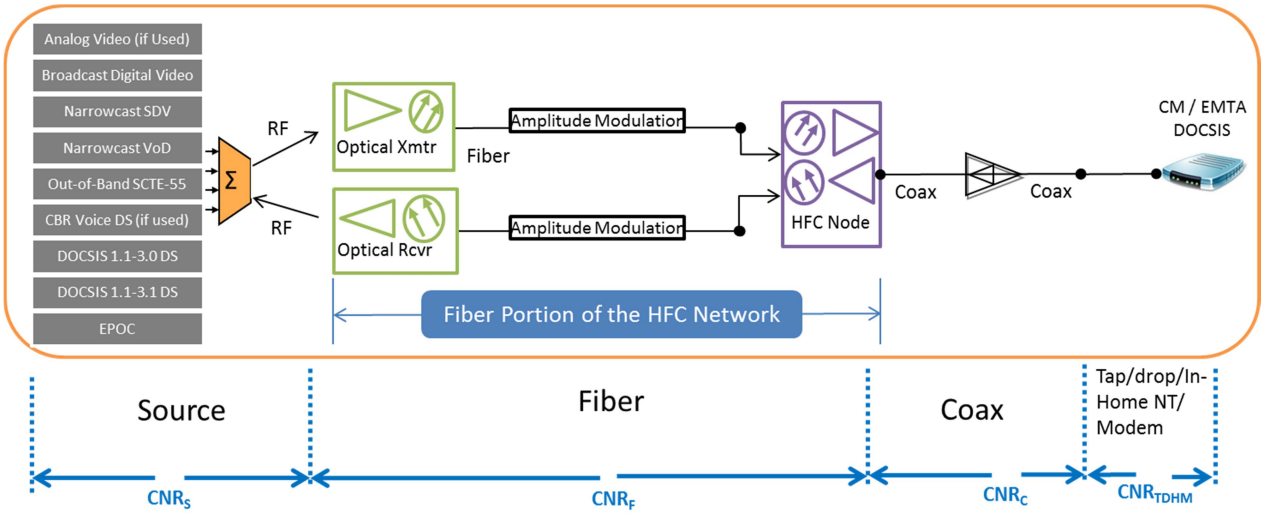


Figure 3. Different portions composing an HFC network

Since the focus of the article is to characterize different portions of the network efficiently, the article proposes a novel approach for estimating the CNR performance of different network portions solely based on the data measured at the CM. In particular, given the CNR values measured by CMs located in different places in the network, as shown in Fig. 4, the CNR of different portions of the network can be back-calculated using the following equations, which are derived from (1)

$$CNR_{optical} = -10\log\left(10^{\frac{-CNR_{node}}{10}} - 10^{\frac{-CNR_{source}}{10}}\right)$$

$$CNR_{RF1} = -10\log\left(10^{\frac{-CNR_{AMP1}}{10}} - 10^{\frac{-CNR_{node}}{10}}\right)$$

$$CNR_{TDHM2.1} = -10\log\left(10^{\frac{-CNR_{M2.1}}{10}} - 10^{\frac{-CNR_{AMP2}}{10}}\right), \quad (2)$$

Where $CNR_{optical}$, CNR_{RF1}, $CNR_{TDHM2.1}$ are the CNR values, in dB, as follows

- $CNR_{optical}$ is the CNR of the optical link,
- CNR_{RF1} is the CNR of the RF segment between node and AMP1 outputs
- $CNR_{TDHM2.1}$ is the CNR of the tap/ drop/ home-network segment that connects to Modem M2.1.

Additionally, $CNR_{M2.1}$ is the CNR measured at the input of CM M2.1 and CNR_{source}, CNR_{node}, CNR_{AMP1}, CNR_{AMP2} are the CNR values, in dB, at the output of the source, node, AMP1, and AMP2, respectively. Note that while CNR_{node}, CNR_{AMP1}, and CNR_{AMP2} in Fig. 4 could be measured in the field, it is not practical to do that. Therefore, the analysis avoids the need to measure those values in the field by introducing a novel methodology to estimate those values based on measurements collected by CMs at their corresponding locations. In particular, at any given place in the cascade, it can be safely assumed that the quality of the signal in the distribution network is equal to or better than the best signal quality observed at any modem within that section of the cascade. For example, the CNR value at the

output of AMP 1, CNR_{AMP1}, can be safely assumed to be equal or higher than the CNR values measured by CMs M1.1 and M1.2. Moreover, CNR_{AMP1} has to be higher than or equal to CNR_{AMP2}. Given that CNR_{node}, CNR_{amp1}, and CNR_{amp2} can be estimated via the approach described above, the performance of different portions of the network can be approximated using (2). That is, it will be possible to calculate $CNR_{optical}$, CNR_{RF1}, CNR_{RF2}, CNR_{RF3}, $CNR_{TDHMx.x}$, etc.

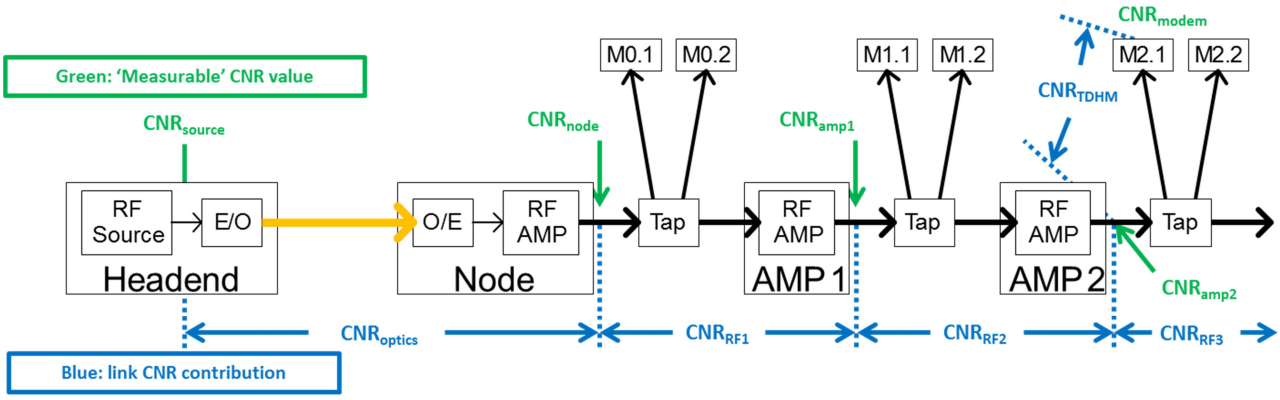


Figure 4. Estimating the performance of different portions of the network based on CM's CNR measurements

3. MER STOCHASTIC ANALYSIS & POTENTIAL DOCSIS 3.1 SPECTRAL EFFICIENCY

The MER analysis presented in this section is based on the concepts introduced in the previous section with field measurements collected from CMs of given locations. It is important to indicate that the analysis assumes that the measurements reported by the CM represent true MER values (i.e., not estimates of the channel CNR), which is true for most CMs. A sample set of collected measurements for a service node X1, where a fiber node is followed by 6 amplifiers in cascade (i.e., N+6) is shown in Table 3 with corresponding MER pdfs shown in Fig. 5. The X1 service group contained more than 500 customers. Note that the number of channels included in each row of Table 3 does not represent the number of unique channels servicing CMs in that segment but rather represents the absolute sum of all channels on all CMs in that segment (i.e., a channel can be counted multiple times). In particular, the total number of channels is equal to the number of measurements obtained by CMs in that segment because each CM provides a single measurement per channel and therefore one channel can correspond to multiple measurements each sourced from a different CM.

Observing the pdfs in Fig. 5, it is noted that there is large variations in MER values for CMs located at any cascade depth. The analysis in this section will try to study these curves and identify the sources of MER variations. For instance, MER measurements across different channel frequencies were collected as displayed in Table 4, where the results show that while the frequency of the channels whose MER were collected do contribute to the MER variation, those variations are minor and do not explain the range of MER variations shown in the pdfs.

Table 3. Reported MER measurements of CMs located in different cascade depths within service Node X1 (N+6 network)

N+x	Number of Channels	Number of Modems	Average MER (dB)	MER Stdev (dB)	Max MER (dB)	Min MER (dB)
0	45	10	37.66	0.46	38.7	36.8
1	181	35	37.88	1.13	42.1	30.1
2	546	124	37.48	1.08	41	31.8
3	507	104	37.59	1.36	42	30.5
4	708	143	37.41	0.90	41.6	34.2
5	179	37	37.12	0.61	38.8	33
6	401	81	37.09	1.25	41.4	33.7

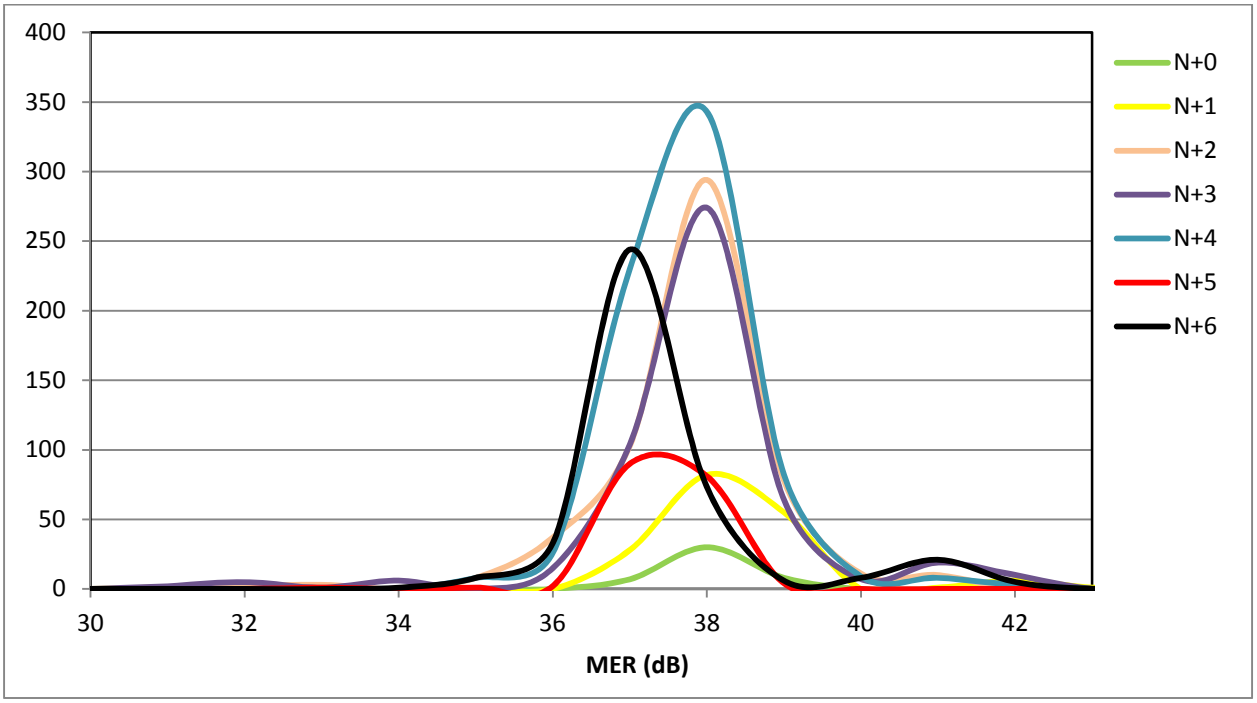


Figure 5. Distributions of MER values of a collection of CMs located in different cascade depths in Node X1 service group (N+6 network)

Table 4. Reported MER measurements vs channels frequencies

Channel Frequency (MHz)	Number of Channels	Average MER (dB)
603	361	37.44
615	363	37.24
609	351	37.32
627	341	37.31
621	325	37.11
633	333	37.19
639	339	37.10
645	348	36.94

Investigating the MER distributions shown in Fig. 5 brings up an interesting observation. The pdfs of MER values reported by CMs located at N+0/N+1/N+2/N+3/N+4 are similar. Moreover, there is only a slight shift for the pdfs corresponding to CMs at N+5/N+6 locations. One would naturally presume that deeper cascades will be manifested with a clear shift in the pdf as the location changes. To understand the underlying causes of the observed network behavior, more detailed MER analysis is performed. First, given that most of the pdfs looked similar; one potential reason for this situation could be that the optical link is limiting the performance (i.e., the bottle neck). Therefore, in order to verify or defy this assumption, the CNR value at the output of the optical link needs to be estimated using the MER values measured by the CMs.

However, before performing detailed CNR analysis for the optical link, it is worth noticing that CMs measure MER values that are calculated at the slicer. Those MER measurements include the modem's 'noise' contributions to the signal as it is received, demodulated, and decoded. Multiple CMs were characterized and it was found that the MER value reported by the CM tends to be close to the CNR value measured at the input of the CM when the CNR value is moderate (< 40 dB), where the background noise from the plan dominates the performance. On the other hand, when CNR values are high (> 40 dB), then the reported MER values will tend to be few dBs below CNR measured at the input of the CM mainly because the CM will be the dominant source noise contributions in this case. Several CMs were characterized to find an approximate mapping between the channel CNR value measured at the input of the CM and the CM's reported MER value so that comprehensive CNR analysis can be performed based on the available CMs MER values. For example, the maximum MER value of 41.4 dB as listed in Table 3 would correspond to a channel CNR of 43.2 dB. The rest of the analysis in this section is based on channel CNR values which correspond to the modems' measured MER values.

Given the CMs' measured MER values and the MER-CNR mapping, it is now possible to calculate the CNR value at the output of the fiber link (i.e., at the output of the fiber node) as originally intended. Recall that a CM located after the 6^{th} amplifier reported an MER value of 41.4 dB, which is equivalent to a CNR value of 43.2 dB at the input of that CM. Therefore, it is reasonable to assume that the CNR value at the output of the last amplifier (6^{th} amplifier) is at least 43.2 dB. Based on field experience, it is assumed that the CNR performance of the RF amplifier can range between 53 dB and 58 dB. Therefore, the CNR at the output of the fiber node considering the N+6 architecture of the

network under investigation must be at least equal to the following

$$CNR_{node} = -10*\log(10^{(-43.2/10)}-(6*10^{(-53/10)})) = 47.5 \text{ dB (Maximum)}$$
$$CNR_{node} = -10*\log(10^{(-43.2/10)}-(6*10^{(-55.5/10)})) = 45.1 \text{ dB (Nominal)}$$

$$CNR_{node} = -10*\log(10^{(-43.2/10)}-(6*10^{(-58/10)})) = 44.2 \text{ dB (Minimum)}$$

The above CNR_{node} values suggest that the performance of the optical link is comparable to the quality of the signal source and therefore it is concluded that the fiber link is not the performance limiter in this particular case. This can be observed in Fig. 6, which shows the above CNR_{node} values and the distribution of CNR values of node X1 on the same graph using a linear scale. Given that the fiber link and the 6-amplifier cascade are not the source of the performance limitation, it is very likely that the cause of performance degradation causing overlapping pdfs with wide MER measurement variations is an issue related to one or combination of the following: tap, drop, home network, or modem (TDHM).

Detailed CNR analysis is performed to verify that the one or combination of TDHM is the likely cause of performance degradation in this case. In particular, the minimum/average/maximum MER values reported by the CMs were identified for each cascade section as shown in Fig. 7. Note that while the figure shows reported MER values, the analysis was performed using CNR levels. Figure 7 shows that the MER value at the output of the 6th amplifier (AMP6) must be at least 41.4 dB because that value was reported by multiple CMs in that section of the cascade. The MER at the output of AMP5 must be at least 41.4 dB which is equal to the minimum MER at the output of AMP 6. Note that the CMs in the cascade section following AMP5 reported a maximum MER value of 38.8 dB. Since the MER value at the output of AMP6 (i.e., 41.4 dB) is larger than the maximum MER value reported by any CM in the cascade following AMP5 (38.8 dB), then the former was used to indicate the minimum MER value observed at the output of AMP5.

In a nutshell, the minimum MER value at the output of any amplifier (say AMPx) is equal to the maximum of 1) MER value at the output of the next amplifier (i.e., AMPx+1) and 2) the maximum reported MER value in the cascade section following AMPx. Following the same logic, the MER value at the output of amplifiers AMP4, AMP3, AMP2, and AMP1 must be at least 41.6 dB, 42 dB, 42 dB, and 42 dB, respectively. Similarly, the MER value at the output of the fiber node must be at least 42 dB. Note that while Fig. 7 and the description above use MER values, the actual analysis was performed using CNR values.

The analysis proceeds further in effort to estimate specific values for CNR at the output of the amplifiers instead of estimating lower limits as was shown in Fig. 7. In particular, the CNR is back-calculated based on the CNR value at the output of the next amplifier assuming an average amplifier performance of 55.5 dB. That is, the CNR at the output of AMP5 is back-calculated using the CNR at the output of AMP6 (43.2 dB) assuming AMP6 performance of 55.5 dB as follows

$$CNR_{AMP5\text{-}back\text{-}calculated} = -10*\log(10^{(-43.2/10)}-(10^{(-55.5/10)})) = 43.5 \text{ dB}$$

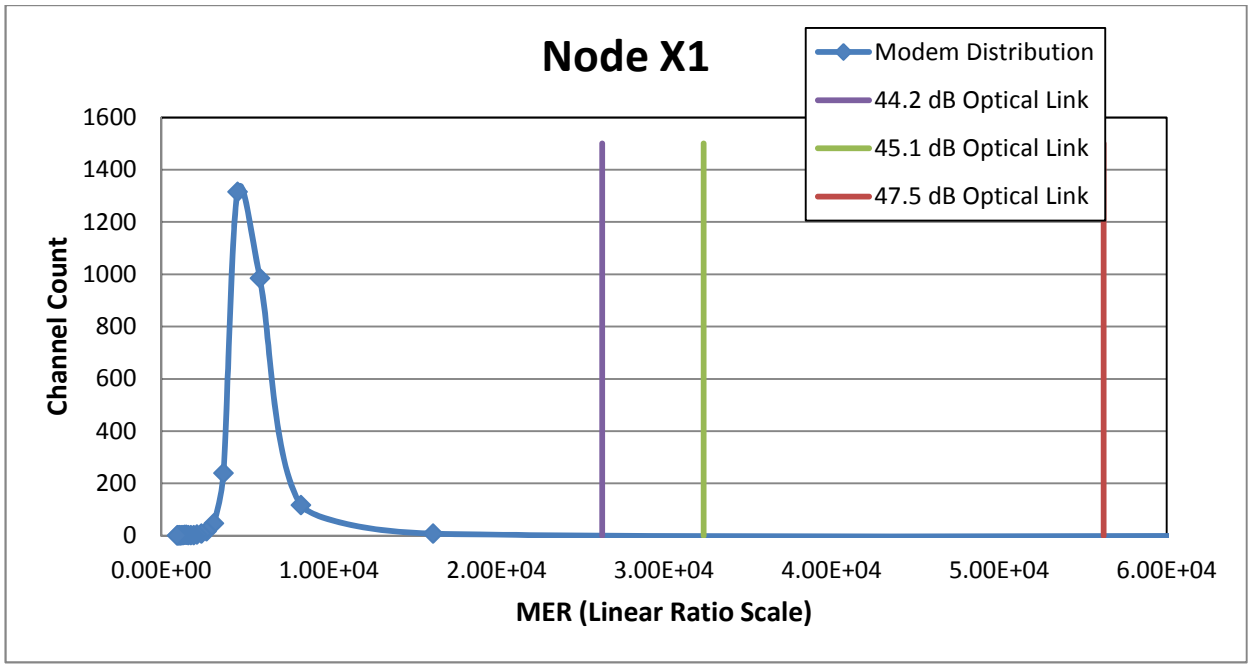


Figure 6. Node X1 CNR distribution and optical links CNR values showing that the optical link is not likely the source of performance degradation in Node X1

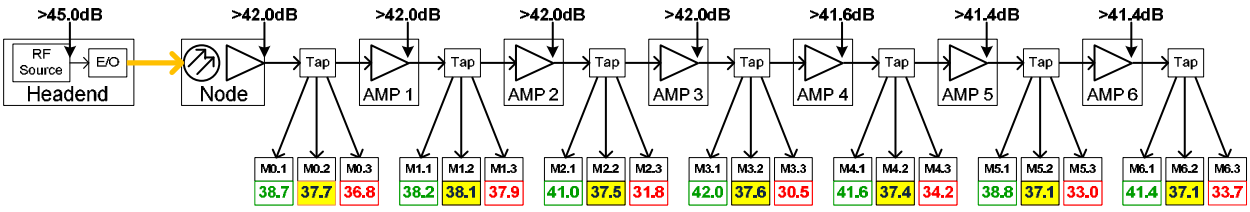


Figure 7. Minimum/Average/Maximum reported MER values by CMs in different cascade sections

The final estimated CNR value at the output of AMP5 is the maximum of two values, where the first value is the back-calculated CNR value above and the second value is the CNR that is equivalent to the maximum measured MER in the cascade section following AMP5 (i.e., CNR=MER = 38.8 dB). Therefore, the CNR at the output of AMP5 is estimated to be max(43.5, 38.8) = 43.5 dB. The CNR values at the output of the other amplifiers/node are calculated sequentially as summarized in Table 5.

Note that the CNR values at the output of different amplifiers within the cascade, which are estimated in Table 5, represent the achievable performance delivered to the tap within that cascade section. Therefore, comparing the CNR value at the input of the CM with that which is estimated at the amplifier/node output (i.e., delivered to the tap) can characterize the performance of the piece(s) between those two end points.

Specifically, comparing the tap CNR values with the average and minimum CNR values reported by CMs in corresponding locations, as summarized in Table 6, can show that at each point in the cascade:

- There is 9.3 dB - 15.1 dB difference in CNR between the worst CM and the amplifier/node serving that modem
- There is 6.1 dB - 8.4 dB difference in CNR between an average CM and the amplifier/node serving that modem

The above results suggest that, for significant number of modems in this service group, the link performance is dominated by one or combination of the following network pieces: tap, drop, in-home network, and the modem (TDHM).

It is critical to realize that these CMs whose link performance is dominated by one or a combination of the TDHM pieces will not likely perceive a large benefit of plant upgrade where N+0 and/or distributed architectures are implemented as a strategy to enhance the performance and increase the network capacity. In order to put this in perspective, DOCSIS 3.1 spectral efficiency analysis was performed for CMs with highest, average, and minimum performance using the assumptions listed in Table 7. In this spectral efficiency analysis, Additive White Gaussian Noise (AWGN) is assumed, the spectral efficiency is calculated for several SNR Gaussian distributions with the following mean values (dB): 30.5, 34, 37, 41, 45, and 49 as shown in Fig. 8. The analysis assumed a Gaussian SNR distribution with 1.18 dB standard deviation to match the MER field measurements. This analysis, depicted in Fig. 9, was performed in a similar way to that provided in [7], which also assumed that the average SNR was reduced by 0.25 dB to compensate for pilot boosting. DOCSIS 3.1 CMs are assumed to be robust to phase noise such that the optional QAM16K modulation order can be supported. The summary of the analysis is shown in Fig. 10.

Table 5. Estimating the CNR value at the output of different amplifiers within the HFC cascade

Parameter	Note	N+6	N+5	N+4	N+3	N+2	N+1	Node (N+0)
Number of Modems		81	37	143	104	124	35	10
Estimated Modem CNR (dB)	Max	43.2	38.8	43.7	44.8	42.3	45.1	38.7
back-calculated Node/AMP output CNR (dB)			43.5	43.7	44.0	45.2	45.6	46.1
Estimated Node/AMP output CNR (MAX[,]) (dB)		43.2	43.5	43.7	44.8	45.2	45.6	46.1

Table 6. Comparing average and minimum CM CNR values with serving amplifier/node CNR value

Parameter	Note	N+6	N+5	N+4	N+3	N+2	N+1	Node (N+0)
Estimated Node/AMP output CNR (MAX[,]) (dB)		43.2	43.5	43.7	44.8	45.2	45.6	46.1
Estimated Modem CNR (dB)	Mean	37.1	37.1	37.4	37.6	37.5	37.9	37.7
CNR Delta from Node/Amp (dB)	Mean	6.1	6.4	6.3	7.2	7.7	7.7	8.4
Estimated Modem CNR (dB)	Min	33.7	33	34.2	30.5	30.5	30.1	36.8
CNR Delta from Node/Amp (dB)	Min	9.5	10.5	9.5	14.3	14.7	15.5	9.3

Table 7. DOCSIS 3.1 parameters assumed for the spectral efficiency analysis

Parameter	Assumption Value
Channel size	*Synchronous* 192 MHz with 190 MHz active spectrum
Subcarrier spacing	25 kHz
FFT size	8K (8192)
FFT duration	40 usec
Subcarriers in 192 MHz	7680
Active subcarriers in 190 MHz	7600
Guard band (2MHz total)	80 subcarriers
Continuous Pilots	88
Scattered pilots	60
PLC subcarriers	16
CP duration	2.5 usec
NCP subcarriers	48
Effective FEC code rate	0.8785

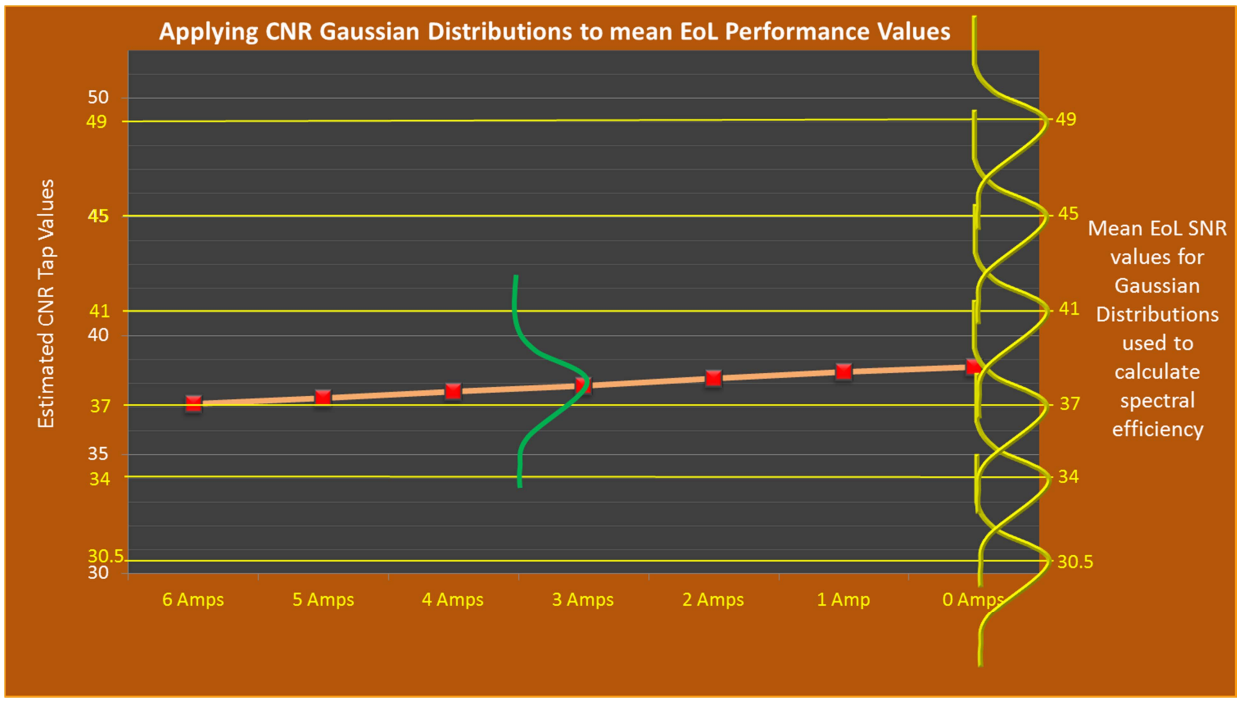


Figure 8. Calculation of DOCSIS 3.1 spectral efficiency for different CNR values: Gaussian CNR distributions are assumed with 1.18 dB standard deviation. Analysis was performed for several mean values: 30.5, 34, 37, 41, 45, and 49 dB

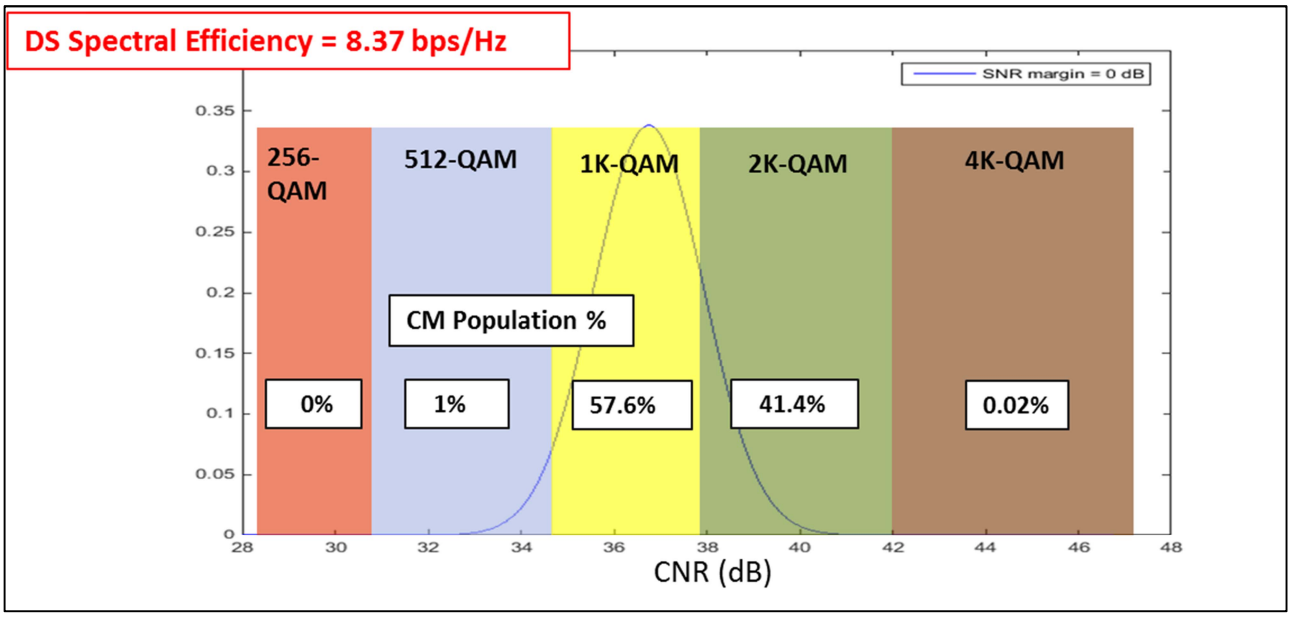


Figure 9. DOCSIS 3.1 DS spectral efficiency analysis for Gaussian CNR distribution with mean value of 37 dB and 1.18 dB standard deviation (with 0.25 dB pilot boosting compensation). Net spectral efficiency is 8.37 bps/Hz for mean SNR of 37 dB

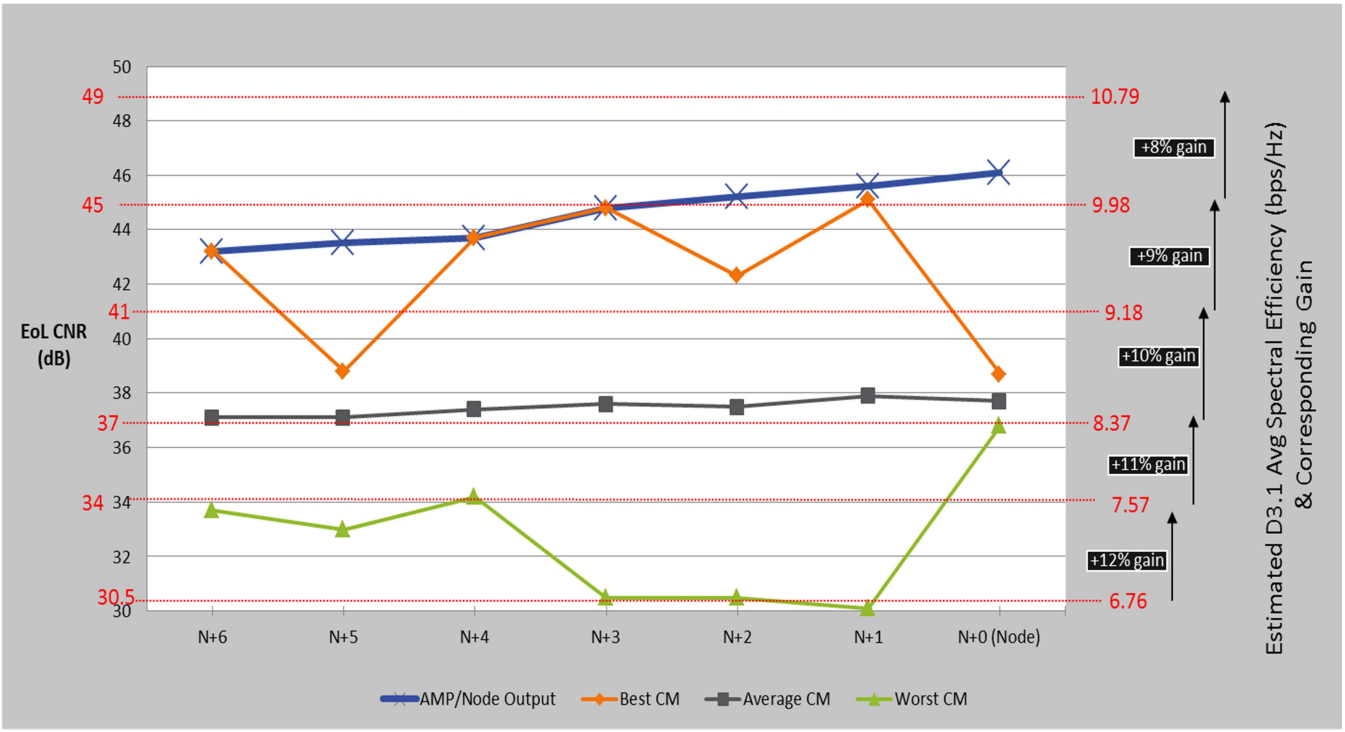


Figure 10. Spectral Efficiency for CMs with worst/average/best CNR performance based on EoL CNR values corresponding to MER values collected from CMs in service node X1 (N+6 network)

It can be noted from Fig. 10 that the lowest performing CMs gain very little benefit as the architecture is upgraded to smaller cascades or distributed architecture. In fact, the performance drops in some cases! Average performing CMs in this service group achieve marginal benefit while the best performing CMs get to their best capacity potential as plant upgrades are implemented. Moreover, it is critical to observe that subscribers located in the same cascade section can see very different performance due to differences in the performance of the TDHM pieces that correspond to their link. Observe that all CMs at the N+0 location performed similarly suggesting that they have comparable drop/tap performance.

To verify the above results, a laboratory experiment was created with an N+3 network as shown in Fig. 11. The network had 4 amplifiers, 62 modems, 93 active channels and represented 15 km of fiber and over 5,700 feet of coaxial cable. The CNR values that correspond to the collected MER measurements from the CMs are summarized in Table 8, where it can be observed that CNR variations throughout the network are minimal and the CNR does not increase significantly as the cascade depth decreases. This is especially true when the cascade (i.e., rigid cable) part of the HFC does not dominate the performance, which is true for most real-world networks.

Tracing back to the field measurements, many service groups were investigated and few sample results are included here (e.g., service nodes X2, X3, and X4). Some nodes showed similar performance to node X1's performance, where one or combination of the TDHM pieces dominated the performance. For example, the node X2 with summary measurements listed in Fig. 12, Table 9, and Table 10, show that the performance for large percentage of CMs in this service group is dominated by the one or combination of the TDHM pieces. Specifically, the pdf's are overlapping and the tables show that at each point of the cascade:

- There is 5.3 dB - 10.1 dB difference in CNR between the worst CM and the amplifier/node serving that modem
- There is 2.7 dB - 9.4 dB difference in CNR between an average CM and the amplifier/node serving that modem

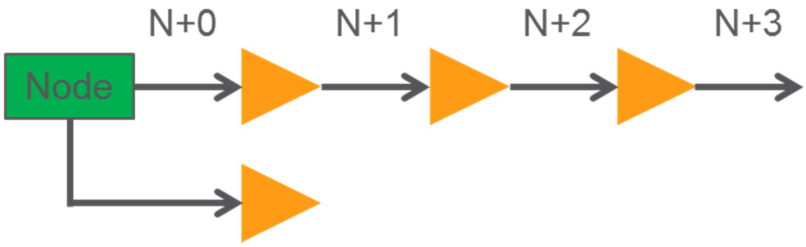


Figure 11. Laboratory experiment to investigate CNR vs. Cascade length

Table 8. CNR values measured at CMs located at different cascade depths in network shown in Fig. 11

N+x	Number of Channels	Average CNR (dB)	CNR Stdev (dB)	Min CNR (dB)	Max CNR (dB)
0	17	44.5	0.63	43.40	45.40
1	36	44.1	0.81	41.10	45.40
2	21	44.3	0.49	43.20	45.30
3	19	43.7	0.57	42.50	44.60

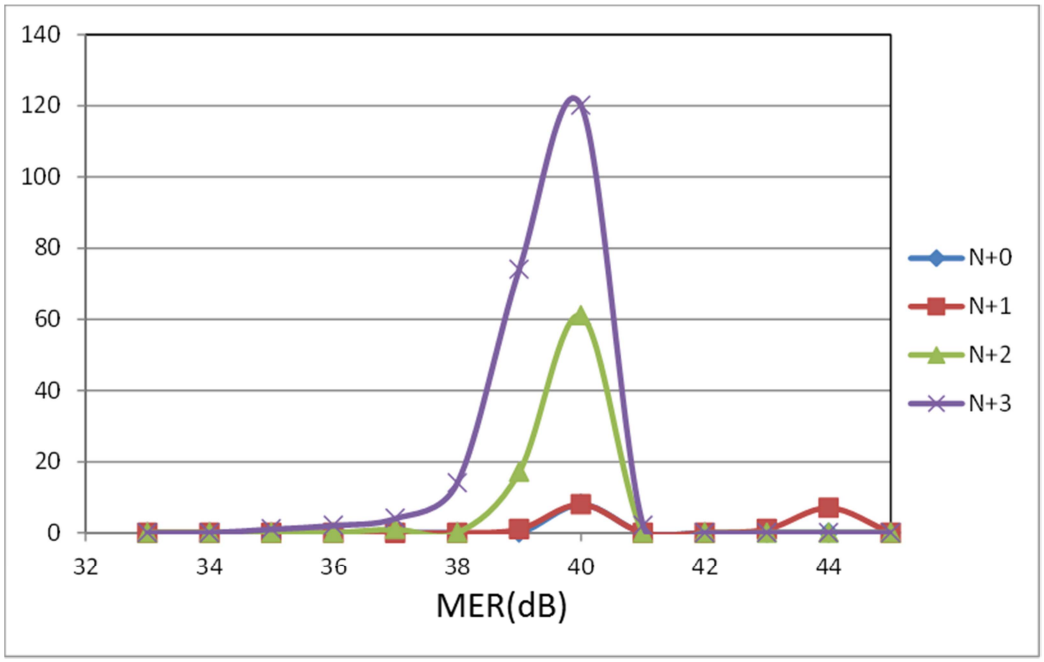


Figure 12. Pdf of MER values collected from CMs in service Group X2 (N+3 network)

Table 9. Summary of CNR values corresponding to MER values collected from CMs in service Group X2 (N+3 network)

N+x	# CMs	# Ch	CNR (dB)			
			Min	Avg	Max	σ
0	1	8	39.3	39.55	39.8	0.27
1	3	17	38.1	41.26	43.9	2.22
2	16	79	36.8	39.31	40.0	0.55
3	35	217	34.7	38.92	40.8	0.77

Table 10. Summary of CNR Analysis for service Group X2 (N+3 network)

Parameter	Note	N+3	N+2	N+1	Node (N+0)
Number of Modems		35	16	3	1
Estimated Modem CNR (dB)	Max	41.9	40.5	48.2	40.1
Back-calculated Node/AMP output CNR (dB)			42.1	42.3	49.1
Estimated Node/AMP output CNR (MAX[,]) (dB)		41.9	42.1	48.2	49.1
Estimated Modem CNR (dB)	Mean	38.9	39.4	42.9	39.7
CNR Delta from Node/Amp (dB)	Mean	3.0	2.7	5.3	9.4
Estimated Modem CNR (dB)	Min	34.7	36.8	38.1	39.4
CNR Delta from Node/Amp (dB)	Min	7.2	5.3	10.1	9.7

Several other service groups showed different behavior than node X1's performance, where the performance was either not completely dominated by the TDHM pieces or the performance was very good throughout the whole node. For instance, node X3 with measurements shown in Fig. 13 and with an analysis summarized in Table 11 indicates that while part of the performance is dominated by one or combination of the TDHM pieces, other factors affecting the performance are not far behind. This can be concluded from the analysis summary, where at each point in the cascade.

- There is 2.1 dB – 6.3 dB difference in CNR between the worst CM and the amplifier/node serving that modem
- There is 1.3 dB – 3.3 dB difference in CNR between an average CM and the amplifier/node serving that modem

Another service group example where the node performance was different from node X1's performance is node X4 with measurements summarized in Table 12, Table 13, Fig. 14, and Fig. 15. Observe that the analysis summary shows that the performance for the majority of CMs in this service group is very good, where *only* the lowest performing CMs are dominated by one or combination of the TDHM pieces. This is because at any point in the cascade

- There is 1.7 dB – 6.9 dB difference in CNR between the worst CM and the amplifier/node serving that modem
- There is 1.1 dB – 1.8 dB difference in CNR between an average modem and the amplifier/node serving that modem

The above observations suggest that the majority of the CMs are *not* dominated by the TDHM pieces. The degradation of about 2 dB at most is contributed to other factors like the shortage in the RF power levels where the low RF levels tend to limit the CNR levels as shown on the left side of Fig. 15.

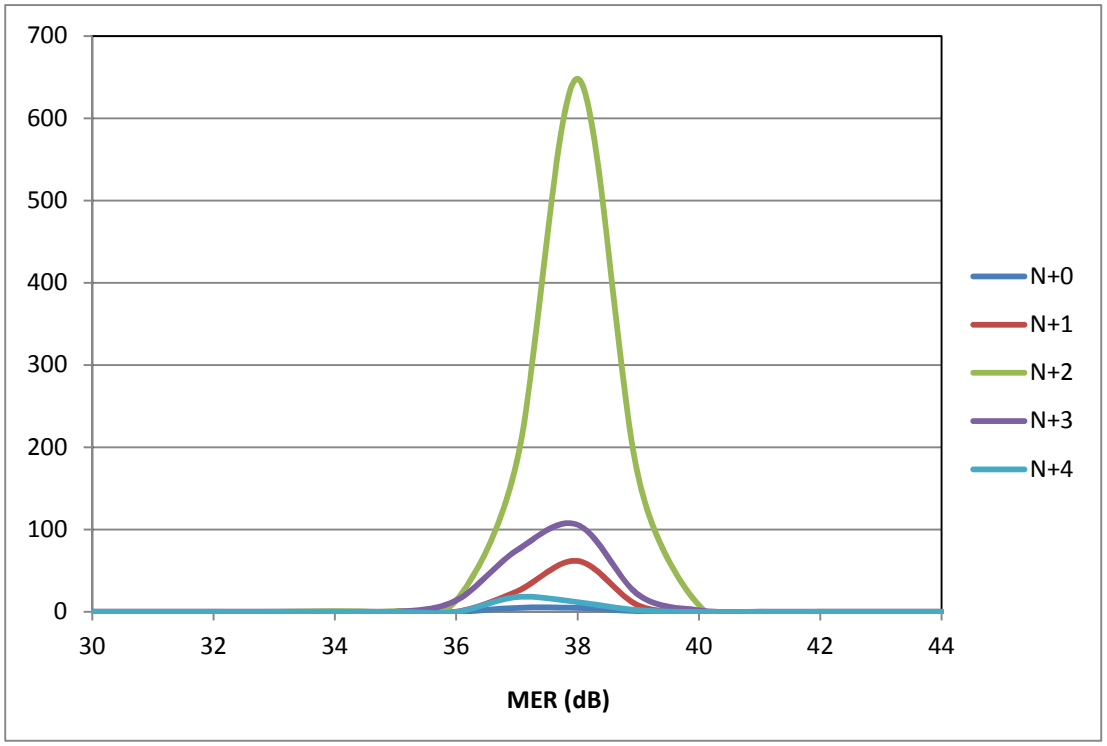


Figure 13. Pdf of MER values collected from CMs in service Group X3 (N+4 network)

Table 11. Summary of CNR Analysis for service Group X3 (N+4 network)

Parameter	Note	N+4	N+3	N+2	N+1	Node (N+0)
Number of Modems		8	38	201	18	3
Estimated Modem CNR (dB)	Max	38.4	39.5	40.3	38.2	38.1
back-calculated Node/AMP output CNR (dB)			38.5	38.7	40.4	40.6
Estimated Node/AMP output CNR (MAX[,]) (dB)		38.4	38.6	40.3	40.4	40.6
Estimated Modem CNR (dB)	Mean	37.1	37.3	37.6	37.5	37.3
CNR Delta from Node/Amp (dB)	Mean	1.3	1.3	2.7	2.9	3.3
Estimated Modem CNR (dB)	Min	36.3	35.4	34	36.3	36.6
CNR Delta from Node/Amp (dB)	Min	2.1	3.2	6.3	4.1	4.0

Table 12. Summary of CNR values corresponding to MER values collected from CMs in service Group X4 (N+3 network)

N+x	# CMs	# Ch	SNR (dB)				RF Power (dBmV)			
			Min	Avg	Max	σ	Min	Avg	Max	σ
0	5	40	37.3	37.9	38.2	0.2	-1.5	4.7	15.5	5.8
1	4	18	35.0	37.1	38.8	1.0	1.1	3.2	6.0	1.2
2	44	226	32.9	37.1	38.8	0.8	-9.9	-1.5	12.1	5.4
3	41	208	31.7	36.8	38.6	1.1	-9.7	-1.8	11.0	5.6

Table 13. Summary of CNR Analysis for service Group X4 (N+3 network)

Parameter	Note	N+3	N+2	N+1	Node (N+0)
Number of Modems		41	44	4	5
Estimated Modem CNR (dB)	Max	38.6	38.8	38.8	38.2
back-calculated Node/AMP output CNR (dB)			38.7	38.9	39.0
Estimated Node/AMP output CNR (MAX[,]) (dB)		38.6	38.8	38.9	39.0
Estimated Modem CNR (dB)	Mean	36.8	37	37.1	37.9
CNR Delta from Node/Amp (dB)	Mean	1.8	1.8	1.8	1.1
Estimated Modem CNR (dB)	Min	31.7	32.9	35	37.3
CNR Delta from Node/Amp (dB)	Min	6.9	5.9	3.9	1.7

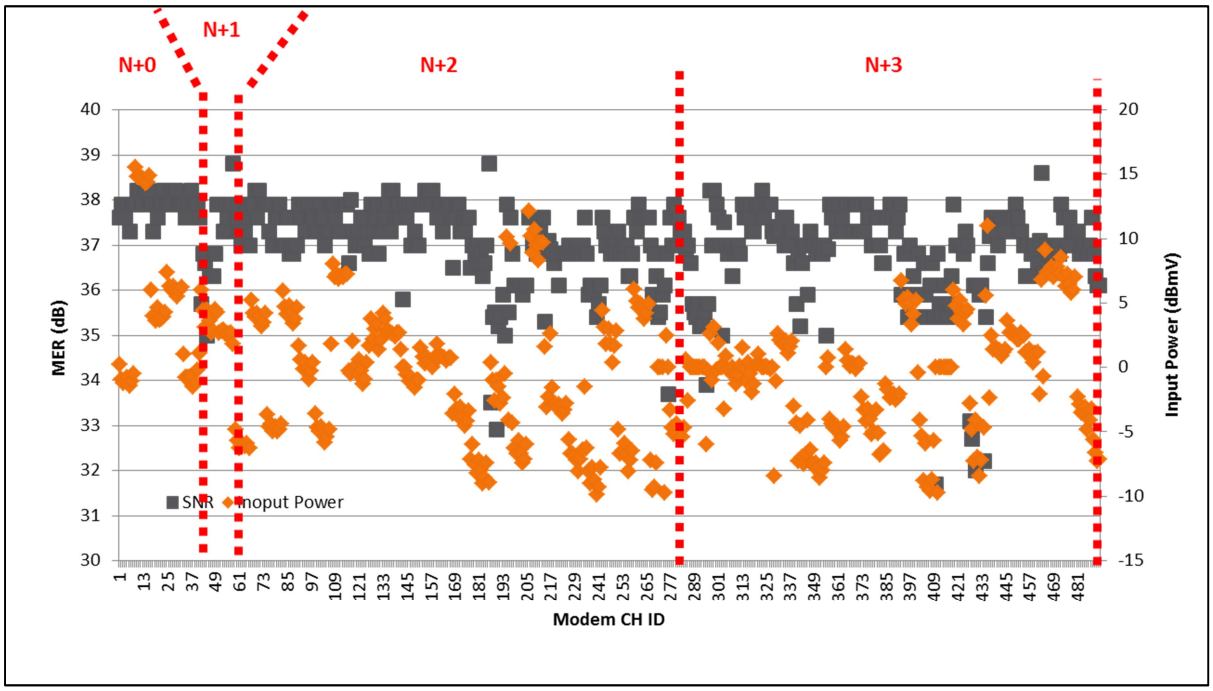


Figure 14. Cascade depth vs. RF power levels and MER values collected from CMs in service Group X4 (N+3 network)

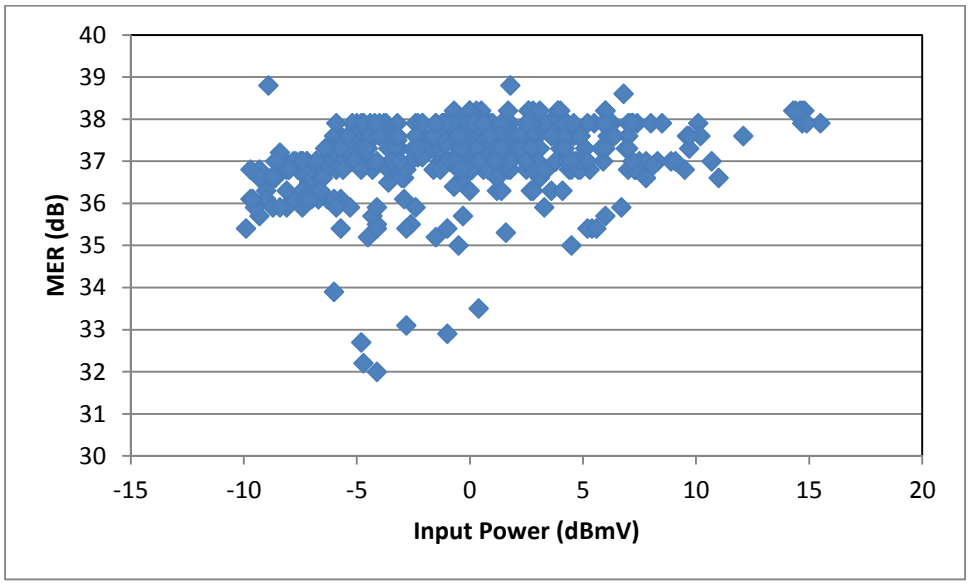


Figure 15. RF power levels vs. MER values collected from CMs in service Group X4 (N+3 network) showing RF levels limiting CNR performance

4. FEC ANALYSIS

It should be understood that for the EoL performance and spectral efficiency analyses to be comprehensive, other field measurements beyond MER values may need to be considered. This is due to the fact that average MER measurements reported by the CMs may not reflect the plant actual CNR when the CNR values are high and also MER measurements may not reflect all types of

plant noise and distortion. For example, the infrequent and irregular impulse noise events may not drop the average MER value reported by the CM. Similar behavior can also occur with optical/electric distortions as will be explained later in this section.

There are many parameters that can be collected in addition to the MER values to yield more comprehensive analysis. For instance, the RF signal levels measured at the CMs located in different parts of the cascades can help reveal plant issues and/or sources of performance degradation in some cases as was explained in the previous section. Another example of helpful parameters that can be collected via CMs is FEC codeword error statistics before and after FEC decoding is applied. FEC statistics are of significant importance as they indicate the actual performance level of the CM. In particular, the effect of any plant issue causing performance degradation and/or affecting customer service can be observed via collecting these counts.

It is important to realize that codewords errors can actually occur with CMs reporting high MER values. For example, Table 14, Fig. 16, and Fig. 17 show the error statistics for CMs within node X1 service group, where pre- and post-FEC events actually occurred with MER values making the two parameters look independent in those cases.

Table 14. Number of channels with pre- and post- FEC CWs corresponding to error threshold values in node X1

CER	# of channels with Pre-FEC	# of channels with Post-FEC
0	13	21
1E-10	0	0
1E-9	0	0
1E-8	1	0
1E-7	5	1
1E-6	2	594
1E-5	1661	2059
1E-4	1003	10
1E-3	0	0
1E-2	0	0

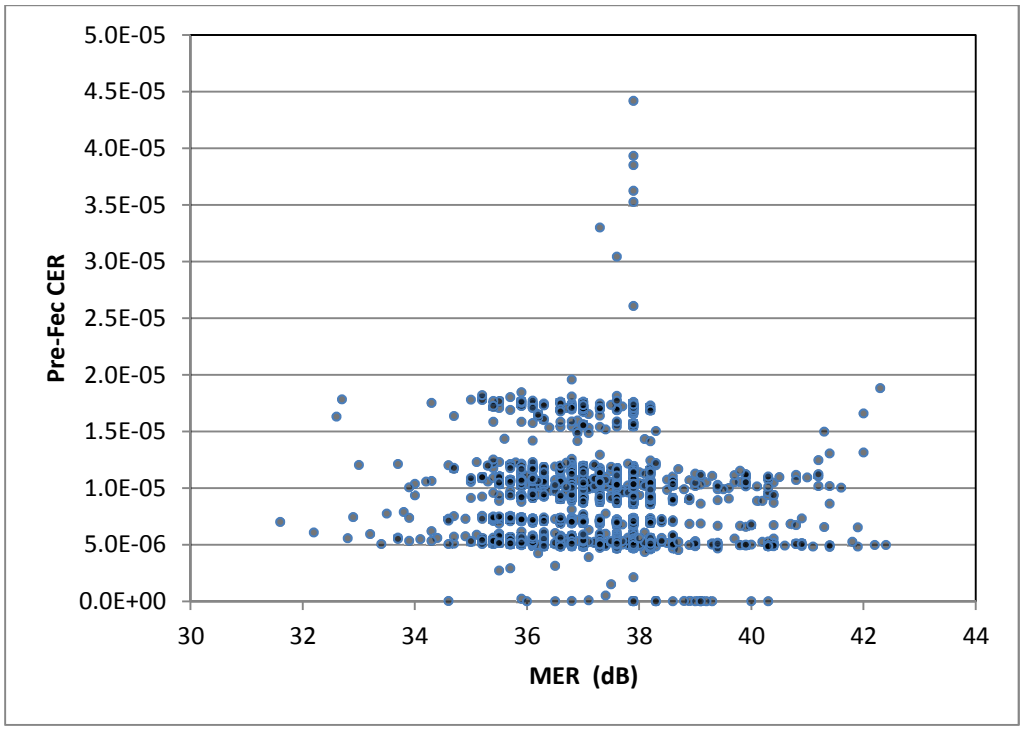


Figure 16. Pre-FEC counts vs. CNR levels corresponding to MER values reported by CMs in Node X1

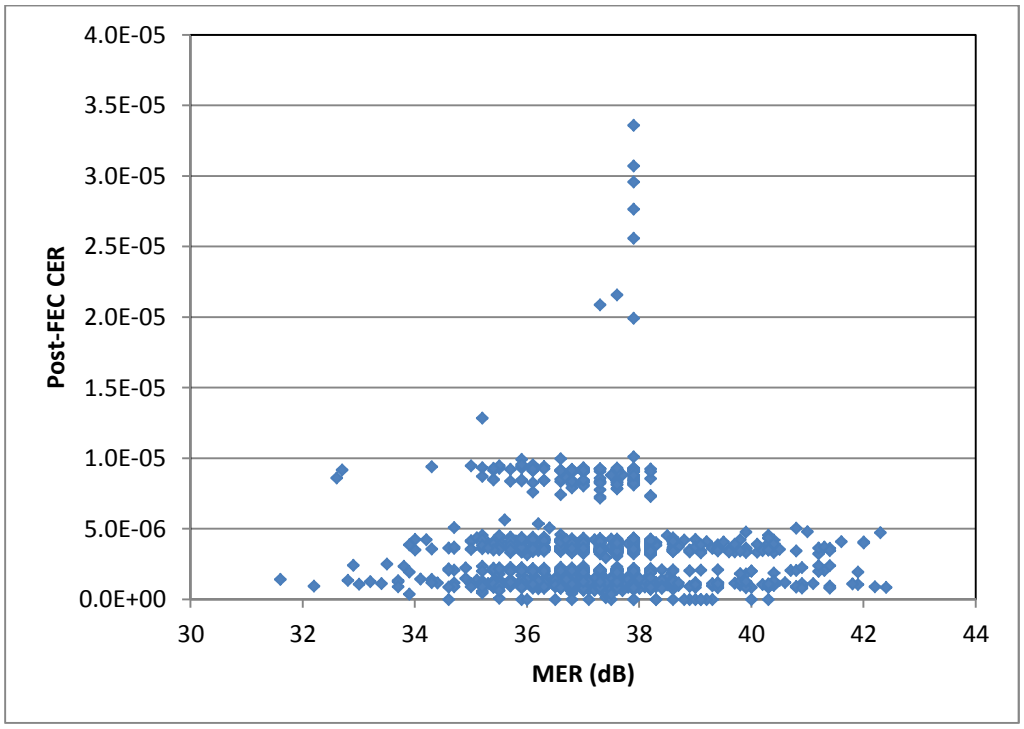


Figure 17. Post-FEC counts vs. CNR levels corresponding to MER values reported by CMs in Node X1

It was mentioned earlier that impulse noise events can cause such a scenario where FEC codeword errors are observed with high MER values. Another plant event that can cause this is distortion. In order to verify this, a laboratory experiment was conducted, where an amplifier was configured to introduce distortion products. The received signal is demodulate and decoded using a Sunrise Quadrature Amplitue Modulation (QAM) analyzer and a DOCSIS CM. The statistics reported by the Sunrise analyzer and the DOCSIS CM are summarized in Table 15, which actually shows correlation with the results observed in Fig. 16 and Fig. 17, where FEC errors can actually occur with high reported MER values (e.g., 40 dB).

This suggests that node X1 could have had distortion events, which were service impacting due to the fact there was significant number of channels with nonzero post-FEC errors. Similar to the previous nodes analyzed, the MER performance of this node is limited by one or combination of the TDHM pieces. However, due to the impact of distortion on the ability of the modem to correct codeword errors, it is likely that the optical modulation index (OMI) of this transmitter would need to be lowered in order to support higher orders of modulation.

Similar FEC statistics were collected for the three nodes discussed in the previous (i.e., X2, X3, and X4 with results summarized in Tables 16, 17, and 18 for service nodes X2, X3, and X4, respectively. Note that the events that caused pre-FEC errors in these service groups were not service impacting because it can be easily observed from these tables that most of the errors were corrected by the FEC decoder yielding insignificant number of channels with nonzero post-FEC errors.

Table 15. Distortion can cause pre- and post- FEC CWs event with high MER values

Sunrise QAM Analyzer @825MHz			Modem 1 @825MHz			Modem 2 @831MHz		
MER (dBc)	Pre-FEC BER	Post FEC BER	MER (dBc)	Pre-EFC CER	Post FEC CER	MER (dBc)	Pre-EFC CER	Post FEC CER
37.7	1.0E-04	1.0E-06	37.5	7.0E-02	5.4E-04	36.4	1.2E-01	2.2E-03
40.2	2.0E-05	5.0E-08	39.5	1.3E-02	1.8E-05	38.6	2.5E-02	6.6E-05
42.0	5.0E-07	0.0E+00	40.8	4.0E-04	3.0E-07	39.8	8.0E-04	5.0E-07

Table 16. Number of channels with pre- and post- FEC CWs corresponding to error threshold values in node X2

CER	# of channels with Pre-FEC	# of channels with Post-FEC
0	191	308
1E-10	0	0
1E-9	0	0
1E-8	11	0
1E-7	100	0
1E-6	7	1
1E-5	0	3
1E-4	2	1
1E-3	2	0
1E-2	0	0

Table 17. Number of channels with pre- and post- FEC CWs corresponding to error threshold values in node X3

CER	# of channels with Pre-FEC	# of channels with Post-FEC
0	1	1301
1E-10	0	0
1E-9	0	0
1E-8	2	0
1E-7	1051	2
1E-6	248	13
1E-5	16	13
1E-4	14	6
1E-3	4	1
1E-2	0	0

Table 18. Number of channels with pre- and post- FEC CWs corresponding to error threshold values in node X4

CER	# of channels with Pre-FEC	# of channels with Post-FEC
0	5	461
1E-10	0	0
1E-9	0	0
1E-8	1	0
1E-7	143	0
1E-6	309	1
1E-5	4	0
1E-4	24	24
1E-3	2	2
1E-2	0	0

The above measurements and analyses show that the potential capacity gain or delivered service level in a network should not be taken for granted when performing MER-based analyses. In particular, it is important to collect FEC measurements to verify that the MER values that represent the base of the analysis actually reflect the real plant situation of the network under question.

5. PERFORMANCE OF DIFFERENT NETWORK SEGMENTS, NETWORK EOL PERFORMANCE, AND POTENTIAL DOCSIS 3.1 SYSTEM CAPACITY

The performance analysis that was presented in section 3 characterized the coaxial portion of the network (i.e., cascade, TDHM). The analysis in this section builds on top of that analysis to characterize the performance of the optical links. Recall that the analysis in section 3 proposed a methodology to estimate the CNR value at the output of the fiber node for some example cases. Large numbers of service nodes were analyzed offline for the network in question and it was found that many of the nodes were able to achieve about 45 dB CNR at the output of the fiber node (similar to node X1's node output CNR estimated value in Table 5).

Assuming a node's output CNR of 45 dB and a typical QAM signal source CNR range of 48-50 dB, then the performance of the fiber links is estimated to be about 47-48 dB, which is in agreement with published numbers [9] and is comparable to the quality of the QAM signal source. Computer simulations were performed to validate the results, where C-band optics were simulated with +9 dBm launch power, 1GHz video loading with optical modulation index of 33%, and receiver power set to -1dBm to 1 dBm. The results of the computer simulations are displayed in Fig. 18, which show that one wavelength over 30 km fiber link has about 48 dB CNR which agrees with the above conclusions.

Given the analysis presented thus far in this article, the performance of different portions of the N+6 network in question can be deduced as shown in Table 19, where the performance ranges for different segments (source, optical link, RF trunk, tap, drop, in-home network (IHN), modem) were extended to accommodate wider selection of equipment performance, network parameters, and time/temperature variations. In fact, over the course of collecting field measurements and analyzing them, it was observed that MER typically varies by 1- 3 dB per week, over time and temperature while RF power typically varied by 1 – 5dB per week, over time and temperature.

To pictorially depict the importance of performance ranges listed in Table 19, the noise contributions from each of these network portions are shown in Fig. 19, which plots the inverse of linear CNR values. Observe that the noise contributions from the TDHM pieces are significant for low performing CMs in this scenario.

The performance of an N+3 and N+0 networks can be inferred based on the values presented in Table 19. For instance, the length of the AM fiber link may increase as the cascade length decreases. Additionally, more wavelengths may be used on the same fiber link to service more fiber nodes. Both of these variations may decrease the performance of the fiber link as assumed in Table 20 for an N+3 network. Note that the cascade length decreases from 6 to 3 which shifts the cascade performance range in Table 19 by about 4 dB to yield the cascade performance range in Table 20. Using the same logic, the performance range for different portions of an N+0 network can be summarized as shown in Table 21.

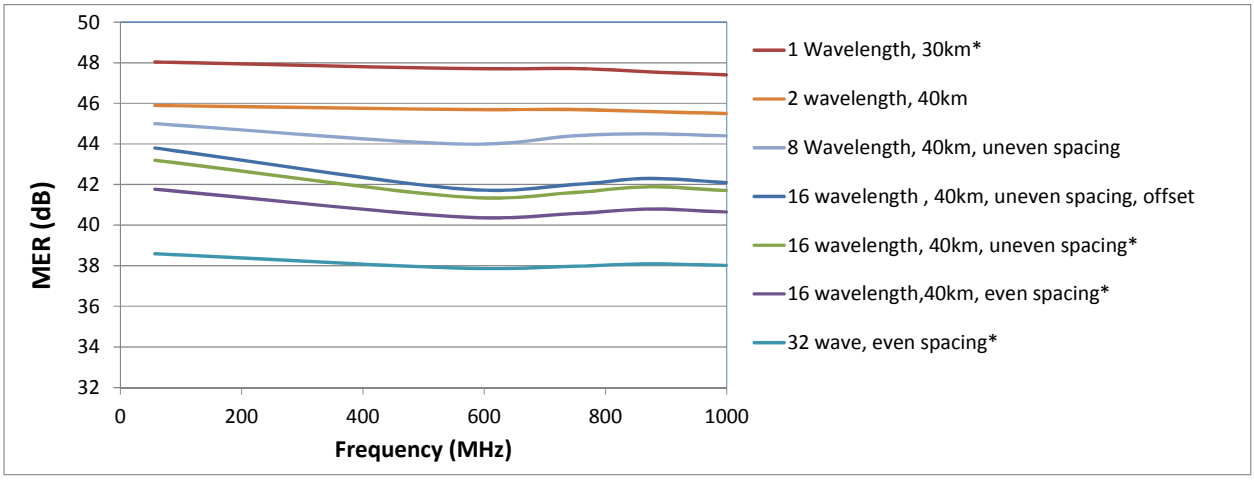


Figure 18. Computer simulations of different optical links scenarios

Table 19. Performance ranges for different segments of an N+6 network (centralized architecture)

Segment	Low	Typical	High
Source	43 dB	47 dB	51 dB
Optical Link	40 dB	44 dB	47 dB
RF Trunk (N+6)	43 dB	46 dB	49 dB
Tap, Drop, IHN, Modem	32 dB	38 dB	44 dB

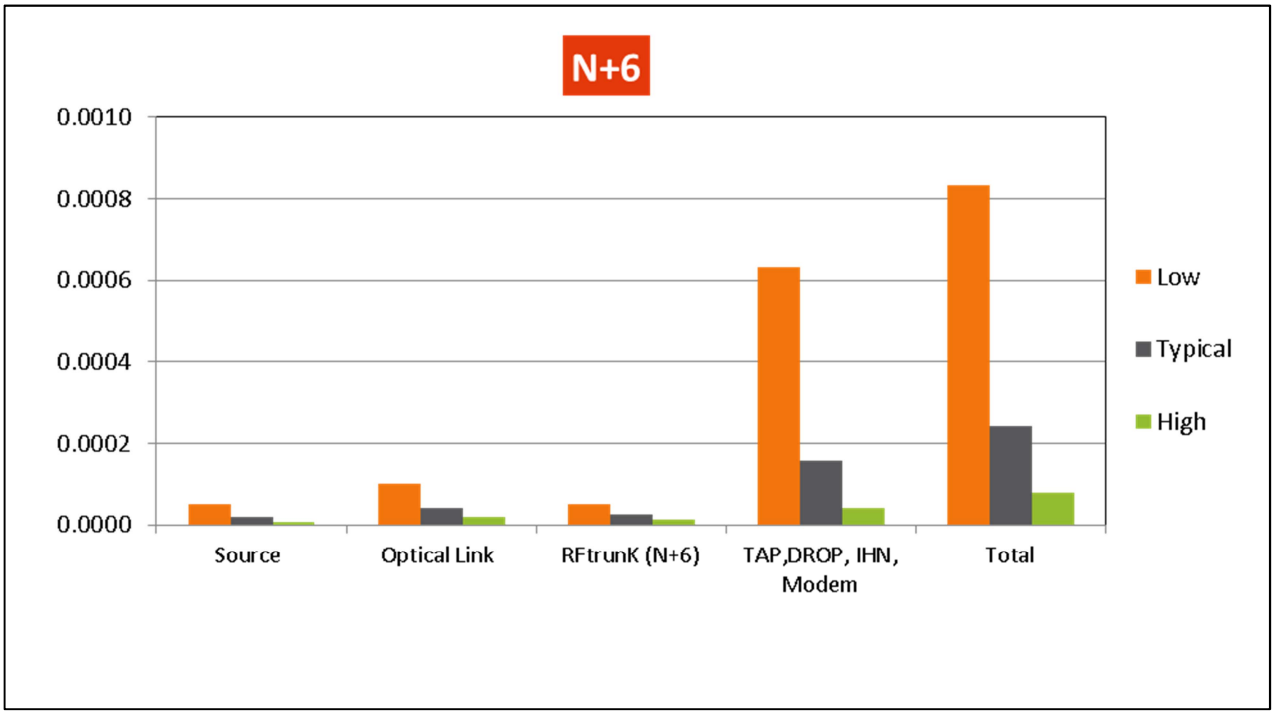


Figure 19. 1/CNR (linear scale) for different segments composing the N+6 network (centralized architecture)

Table 20. Performance ranges for different segments of an N+3 network (centralized architecture)

Segment	Low	Typical	High
Source	43 dB	47 dB	51 dB
Optical Link	39 dB	43 dB	47 dB
RF Trunk (N+3)	47 dB	50 dB	53 dB
Tap, Drop, IHN, Modem	32 dB	38 dB	44 dB

Table 21. Performance ranges for different segments of an N+0 network (centralized architecture)

Segment	Low	Typical	High
Source	43 dB	47 dB	51 dB
Optical Link	38 dB	42 dB	46 dB
RF Trunk (N+0)	N/A	N/A	N/A
Tap, Drop, IHN, Modem	32 dB	38 dB	44 dB

The above tables can be used to infer performance ranges for distributed architectures in N+6, N+3, and N+0 networks, as summarized in Tables 22, Table 23, and Table 24, respectively. Note that the ranges stay the same except for the source range, which has been de-rated to compensate for outdoor operation, higher phase noise, potential time/frequency locking, etc. Observe that the optical range is not a factor in the performance analysis in this case.

Table 22. Performance ranges for different segments of an N+6 network (distributed architecture)

Segment	Low	Typical	High
Source	43 dB	46 dB	48 dB
Optical Link	N/A	N/A	N/A
RF Trunk (N+6)	43 dB	46 dB	49 dB
Tap, Drop, IHN, Modem	32 dB	38 dB	44 dB

Table 23. Performance ranges for different segments of an N+3 network (distributed architecture)

Segment	Low	Typical	High
Source	43 dB	46 dB	48 dB
Optical Link	N/A	N/A	N/A
RF Trunk (N+3)	47 dB	50 dB	53 dB
Tap, Drop, IHN, Modem	32 dB	38 dB	44 dB

Table 24. Performance ranges for different segments of an N+0 network (distributed architecture)

Segment	Low	Typical	High
Source	43 dB	46 dB	48 dB
Optical Link	N/A	N/A	N/A
RF Trunk (N+0)	N/A	N/A	N/A
Tap, Drop, IHN, Modem	32 dB	38 dB	44 dB

The performance measures for the different network pieces with ranges summarized in the above tables can be combined to yield an EoL CNR performance *range* for different network architectures as listed in Table 25. Observe that each of the networks has a performance range and not a single value! This means that a CM in a network (e.g., N+0) that is expected to have superior average performance can actually perform worse than a CM in another network (e.g., N+6) that is expected to have inferior average performance and vice versa. For instance, a CM in an N+0 network with performance dominated by one or combination of the TDHM pieces will perform worse than a CM in an N+6 network where the performance is not limited by the TDHM pieces.

For all network architectures, it is important to realize that CMs with no performance issues in the TDHM pieces will have EoL performance at top of the range and will *always* achieve higher capacities than CMs at the bottom end of the range with performance degradation caused by issues in the TDHM pieces. The CNR ranges, listed in Table 25, and their corresponding noise contributions (inverse of linear CNR) are shown in Fig. 20 and Fig. 21, respectively.

Table 25. EoL CNR Performance ranges for different network architectures

Architecture	Low	Typical	High
Centralized (N+6)	30.8 dB	36.1 dB	41.0 dB
Centralized (N+3)	30.8 dB	36.5 dB	41.4 dB
Centralized (N+0)	30.8 dB	36.2 dB	41.4 dB
Distributed (N+6)	31.4 dB	36.8 dB	41.7 dB
Distributed (N+3)	31.5 dB	37.1 dB	42.2 dB
Distributed (N+0)	31.7 dB	37.4 dB	42.5 dB

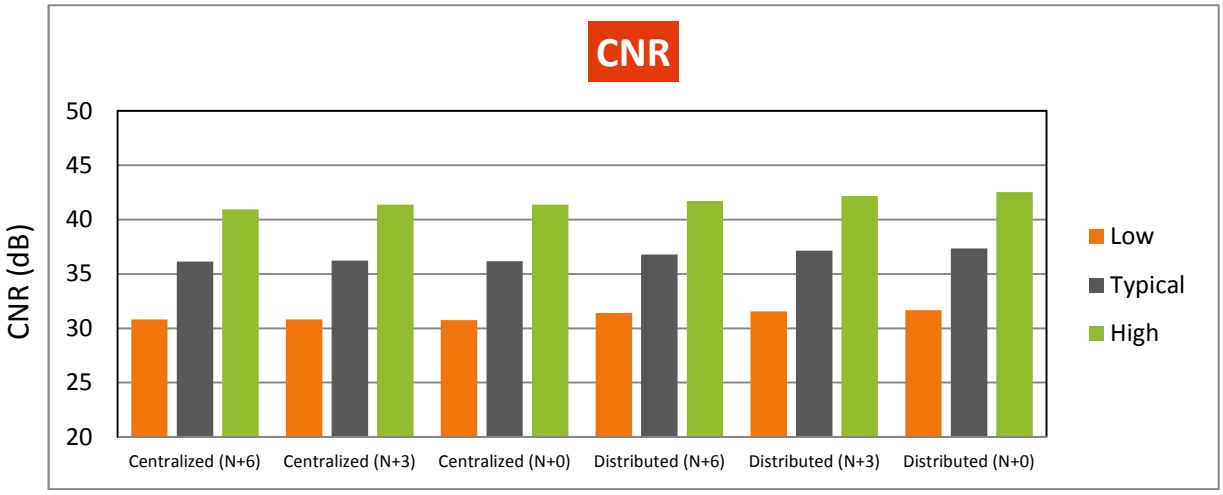


Figure 20. EoL CNR Performance ranges for different network architectures

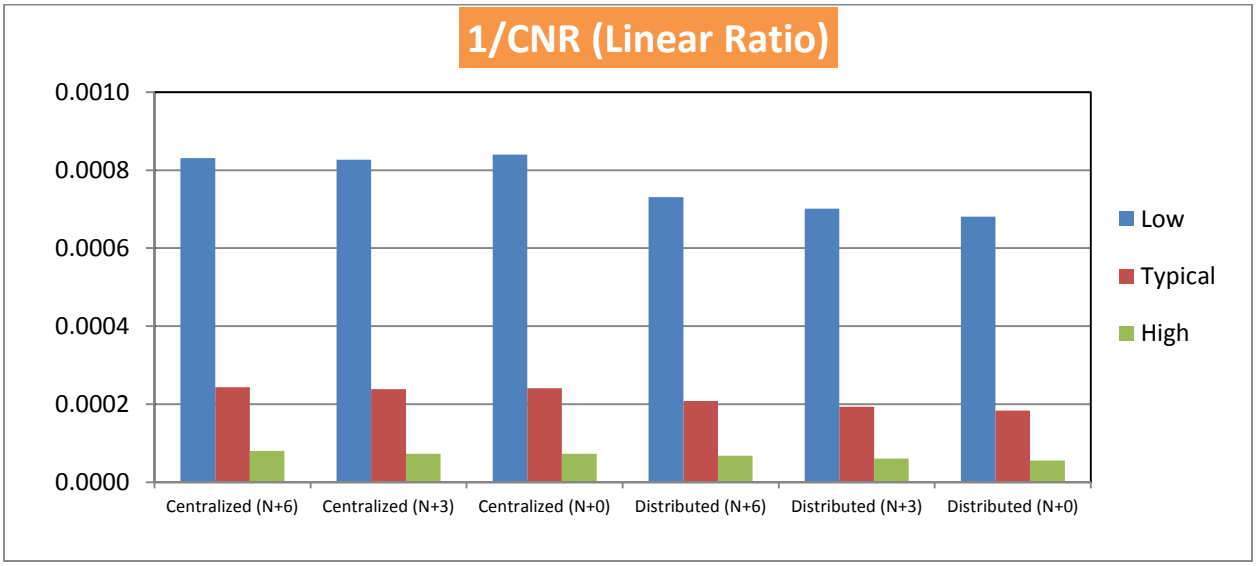


Figure 21. Noise contributions (1/EoL CNR linear scale) for different network architectures

The DOCSIS 3.1 spectral efficiency analysis for different architectures is performed as described earlier with the assumptions listed in Table 7. DOCSIS 3.1 spectral efficiency is calculated for certain CNR values, as shown in Fig. 22, where each CNR value is assumed to represent the mean value of a Gaussian CNR distribution with 1.18 dB standard deviation. Observe that low performing CMs do always exist regardless of the network architecture. In particular, poor performance for TDHM pieces can dominate the network performance for CMs affecting by those pieces for each and every network architecture.

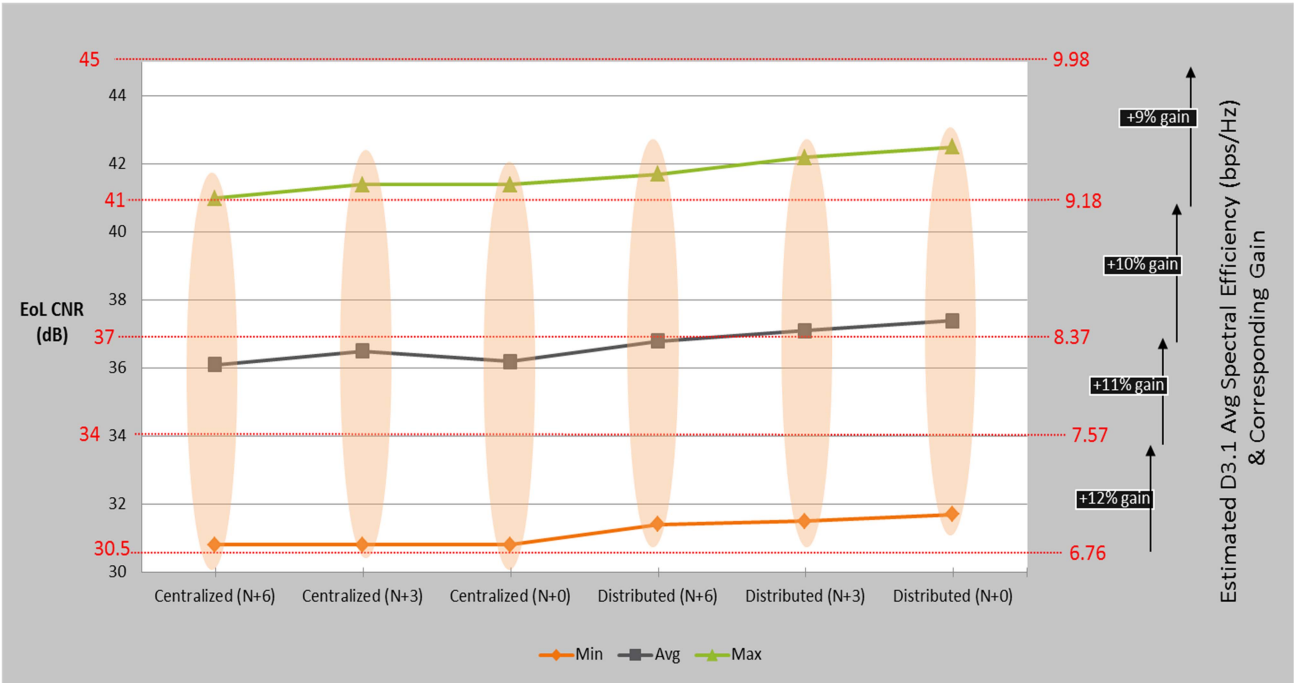


Figure 22. DOCSIS 3.1 Spectral efficiency corresponding to
EoL CNR distributions/ranges for different network architectures

6. NETWORK OPTIMIZATION

One of the key benefits of DOCSIS 3.1 is the ability to operate at significantly higher orders of modulation than DOCSIS 3.0, but in order for this to occur, it is also necessary to operate the HFC plant at much higher CNR levels. In many cases, there will be modems that will require plant maintenance, repairs or upgrades in order to realize DOCSIS 3.1 throughput gains. In the majority of the nodes that were evaluated in this analysis, it was found that the dominant limitation in CNR performance was found in the final portion of the network: the tap, the drop, the in-home network, or the modem itself.

Each element in the final portion of the network presents a unique set of challenges and resolution techniques. In this analysis, it was found that the tap was the least likely source of degradation in this portion of the network. Tap issues were most readily identified by evaluation of the receiver and/or transmitter equalizers in the modems. In this analysis, the drop was most commonly found to be dominant limitation.

This finding emphasizes the importance of performing drop certification when installing DOCSIS 3.1 subscribers and continuing to monitor this portion of the network with proactive HFC monitoring tools. The in-home network was found to be the second most common source of performance limitation, with particular emphasis on the impact on upstream performance, stressing the value of deploying a 2-port RF gateway at the home PoE in order to isolate in-home networks from the common portion of the HFC plant. The cable modem was found to be the primary limitation only slightly more often than the tap. In most cases, it was found that resetting the modem resolved the issue.

7. CONCLUSIONS

The article analyzed different network segments based on real-world field measurements. In this study, it was found that the link segment composed of the Tap, drop, Home network, and CM (TDHM) pieceparts is one of the main factors that can dominate the performance of the whole network. CMs that are affected by these situations will perceive plant upgrades negatively compared to other CMs that do not suffer from these issues. Cable modems whose performance is dominated by the TDHM pieces will always have poor performance regardless of the network architecture. In particular, it was shown that reducing the coaxial cascade length and/or moving to distributed network architecture will not likely increase the capacity of those CMs affected by issues in the TDHM pieces. The analysis also showed that CMs that do not have these issues can experience very high spectral efficiencies in existing networks and gaining more benefit as coaxial cascades get shorter and/or distributed architectures are considered. In a nutshell, the analysis showed that the performance of two CMs on the same network and located in similar network topology locations can be very different.

The analysis showed that the EoL performance of a network is best represented with a range of values, rather than a single number. In particular, it was shown that the performance of an N+6 network spans a 10 dB range (i.e., from 31 dB to 41 dB) with an average EoL CNR of 36.1 dB. The performance range shifts to higher average values as cascade length decreases and/or distributed architectures are deployed. For instance, the EoL performance range of an N+0 distributed architecture network is estimated to be between 32 dB and 43 dB with an average EoL CNR of 37.4 dB.

Understanding the impact of the final segment of the network, including the TDHM pieces is crucial for the efficient investment of time and resources in the HFC plant. In particular, deployment of a 2-port RF gateway at the PoE for network isolation, continued drop certification practices in order to identify degradation and verify performance, and the use of Proactive Network Maintenance (PNM) features and proactive HFC monitoring tools in order identify and resolve issues before they become service affecting will help MSOs to address the performance limitations observed in this analysis. Identification and resolution of the sources of performance limitations within the HFC plant is the key to optimizing the network in order to realize the full potential of DOCSIS 3.1 and maximize the throughput gains that will be achieved.

ACKNOWLEDGEMENT

The authors would like thank their colleagues, Dr. Tom Cloonan, Brian Basile, Xiang He, Shaoting Gu, and Clarke Greene for their valuable comments and contributions to the paper.

REFERENCES

[1] Ayham Al-Banna, “WiMAX Links and OFDM Overlay for HFC Networks: Mobility and Higher US Capacity,” in Proceedings, The NCTA Cable Show Spring Technical Forum, May 2010.

[2] Ayham Al-Banna, “Architecting the DOCSIS® Network to Offer Symmetric 1Gbps Service Over the Next Two decades,” Cable Tech Spring Expo, (May, 2012).

[3] Tom Cloonan et. al., “Simulating the Impact of QoE on Per-Service Group HSD Bandwidth Capacity Requirements,” in Proceedings, SCTE Cable Tec Expo ’14, SCTE, September, 2014.

[4] Tom Cloonan et. al., "Lessons From Telco & Wireless Providers: Extending The Life Of The Hfc Plant With New Technologies," in Proceedings, The NCTA Cable Show Spring Technical Forum, May, 2015.

[5] Mike Emmendorfer et. al, "Estimating Downstream Performance and DOCSIS 3.1 Capacity in CAA and DAA Systems," in Proceedings, The NCTA Cable Show Spring Technical Forum, May, 2015.

[6] 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, The NCTA Cable Show Spring Technical Forum, May, 2015.

[7] Ayham Al-Banna et. al., "The Spectral Efficiency of DOCSIS 3.1 Systems," in Proceedings of SCTE Cable Tec Expo '14, September, 2014.

[8] David Large & James Farmer, Broadband Cable Access Networks: the HFC Plant, Morgan Kaufmann publisher, Nov., 2008.

[9] 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," in Proceedings of SCTE Cable Tec Expo '13, October, 2013.

ACRONYMS AND ABBREVIATIONS

AMP	Amplifier
AWGN	Additive White Gaussian Noise
CER	Codeword Error Rate
CM	Cable Modem
CNR	Carrier to Noise Ratio
DS	Downstream
EoL	End of Line
FEC	Forward Error Correction
HFC	Hybrid Fiber Coax
IHN	In-Home Network
MER	Modulation Error Ratio
OMI	Optical Modulation Index
pdf	Probability density function
PNM	Proactive Network Maintenance
QAM	Quadrature Amplitude Modulation
SNR	Signal to Noise Ratio
Rx	Receive/Receiver
TDHM	Tap/Drop/In-Home-Network/Modem
Tx	Transmit/Transmitter
US	Upstream

MONITORING THE VARIABLE NETWORK: IMPACT OF DOCSIS® 3.1 ON PLANT UTILIZATION METRICS

Niki Pantelias
Broadcom Corporation

Abstract

In order to make good decisions about traffic engineering and capital expenditures in the access network, an operator must have a clear understanding of the capacity and utilization of the plant. A heavily utilized plant segment may require a node split or modification of channel allocation to support customer demand. Accurate assessment of network utilization allows potentially expensive upgrades to be targeted where the need is greatest and return on investment is the highest.

DOCSIS® 3.1 introduces new sources of uncertainty into the metrics for capacity and utilization. The use of multiple profiles with variable data rates increases performance for modems with high SNR, but also means that the capacity of the access network is not fixed; instead, it may vary over time depending on traffic pattern, mix of active modems, system configuration, plant conditions, and other factors.

This paper offers operators a clear understanding of the impact of new DOCSIS 3.1 features on current methods of measuring downstream plant utilization and capacity. Topics addressed include: use cases for multiple profiles with variable data rates, MAC-layer tools and methods for determining what modems can use a given profile, detailed analysis of the accuracy operators can expect from the DOCSIS 3.1 utilization MIBs for various configurations and traffic mixes, explanation of factors that impact accuracy and how operators can configure the system to address them, and suggestions for incorporation of this information into network planning.

BACKGROUND

DOCSIS 3.1 introduces many new capabilities for cable data transmission systems. The Orthogonal Frequency-Division Multiplexing (OFDM) technology used by the DOCSIS 3.1 downstream offers a wide range of parameter settings to address specific plant conditions. One important new feature of the DOCSIS 3.1 downstream is the ability to use different modulation profiles to reach different groups of cable modems.

Profile Definition

A "profile" in the DOCSIS 3.1 downstream is a set of assignments of modulation orders to OFDM subcarriers on a given channel. Profiles are formally defined by DOCSIS MAC Management Messages carried on the channel. A single OFDM channel can have a bandwidth of up to 192 MHz, and may contain in excess of 7000 data-carrying subcarriers. In the DOCSIS 3.1 MAC protocol, an OFDM Channel Descriptor (OCD) message defines these and other parameters of the channel. A different message called a Downstream Profile Descriptor (DPD) is used to define each downstream profile available on the channel. A DPD message assigns a modulation order to each data subcarrier on the channel. Supported modulation orders for data subcarriers range from 16 QAM to 4096 QAM (with a provision for higher modulation orders in the future). A subcarrier may also be "zero bit loaded" in a particular profile. In principle, every subcarrier could have a different modulation order from the subcarriers adjacent to it. This high level of flexibility allows profiles to be configured to avoid very narrow bands of interference or impairment.

A DOCSIS 3.1 Cable Modem Termination System (CMTS) allows up to 16 different profiles to be configured on a given channel. The profiles are identified in their respective DPD messages by numerical values, but in verbal communication, including the DOCSIS specification language, they are often referred to by letters, e.g., "Profile A," "Profile B," and so forth. "Profile A" is configured so that it can be received by all modems and is used for initialization and for distribution of certain broadcast MAC messages, such as MAPs and UCDs. Other profiles have non-specific uses but are expected to use higher modulation orders than Profile A in at least part of the spectrum. These profiles would be used to carry unicast or multicast data traffic more efficiently than Profile A. A DOCSIS 3.1 CM has the ability to receive up to four different profiles – Profile A plus up to three additional profiles; in addition, it is capable of conducting tests (but not passing traffic) on a fifth profile.

Use Cases for Profiles

Use of multiple profiles can be helpful whenever a downstream channel impairment impacts different modems on the plant in different ways. Some example impairments are:

- Low SNR. This may occur, for example, when a modem is located deep in a home network, behind a number of splitters with high loss. The result is a lower-than-expected signal level at the modem, even though the signal level at the entry to the home meets operator specifications. A profile addressing this case would use lower modulation orders across the full spectrum. On today's cable plants, SNR variations of 10 dB or more have been reported across different modems. This could motivate operators to provision multiple profiles with a range of modulation orders covering the range of SNRs on the channel.
- Tilt variations. Drop cables are one source of spectrum tilt due to their small diameter. Modems connected to drop cables of different lengths will see different tilt. A profile aimed at modems with higher tilt would use lower modulation orders at higher frequencies compared to one aimed at modems with lower tilt.
- Diplex filter rolloff. A long cascade of amplifiers, each with a high-order diplex filter at its input and output, will result in a wider rolloff region with higher loss compared to a short cascade. This case is similar to the case of tilt variation and is addressed by using lower modulation orders in the rolloff region, possibly with zero-bit-loaded subcarriers at the outermost frequencies.
- Geographically localized interferers. If a portion of a plant is located near a source of interference, such as a cellular tower, modems near that source may suffer from interference that does not affect modems farther from the source. A profile aimed at modems near the source might use zero-bit-loaded subcarriers at the relevant frequencies, whereas other profiles reaching modems far from the source could have nonzero modulation orders on these same subcarriers.

In all cases, profiles are a tool for avoiding the impairment in question, not eliminating it. Modems that are unaffected by any impairments will operate with the highest instantaneous throughput. Modems experiencing one or more impairments can use different profiles to share the same channel, albeit with reduced capacity due to the use of lower modulation orders or of zero bit loading for some subcarriers.

Assignment of Profile to Modems

When a new modem joins the network, it is necessary to determine what profiles will be

the best fit for that particular device. Is this modem a "low SNR," "average SNR," or "high SNR" device, and is it being affected by interference, tilt, or some other impairment?

One way to determine this is to examine the modem's per-subcarrier RxMER (Receive Modulation Error Ratio). For each subcarrier, the value measured at the modem can be compared to the RxMER required to receive the modulation order used for that subcarrier by a particular profile. If sufficient margin is available on all subcarriers, it may reasonably be expected that the modem can receive that profile successfully.

However, this comparison may not tell the whole story. The use of FEC (Forward Error Correction) on the downstream channel offers some amount of error correction capability, meaning that a modem might be able to use a profile even if its RxMER is too low on a small number of subcarriers. Unfortunately, it is difficult to predict exactly what combinations of RxMER will give the desired post-FEC performance. Sometimes the easiest way to find out whether a modem will support a particular profile is simply to try it – ideally in a way that does not disrupt user traffic if the experiment fails.

The DOCSIS 3.1 specification provides new MAC Management Messages incorporating all of the above approaches.

To gather RxMER information from a modem, the CMTS can use the ODS (OFDM Downstream Spectrum) message exchange. In response to an ODS-REQ message from a CMTS, the modem sends an ODS-RSP containing its measured RxMER per subcarrier for all non-excluded subcarriers on the channel. The measurement is performed on pilot or preamble symbols, which will occur on every non-excluded subcarrier at various points in time, so valid measurements can be made even for subcarriers on which the modem has not actively performed demodulation of data.

An operator or an external application may also collect RxMER information directly from the modem through the DOCSIS 3.1 OSSI interface by reading the SNMP MIB object `CmDsOfdmRxMer`.

The OPT (OFDM Downstream Profile Test) message exchange goes further by asking a modem to provide information that is specific to a given profile. In the OPT-REQ message, a CMTS may ask a modem to provide either or both of two types of information:

- RxMER comparison. The CM performs the per-subcarrier comparison described above, with the CMTS providing the target RxMER and required margin for each modulation order.
- FEC codeword testing results. The CM attempts to receive actual traffic on the profile – either in the form of "test codewords" artificially generated by the CMTS, or "live" traffic destined for other modems on the plant. The modem keeps counts of the total number of codewords received and codeword errors (if any), but does not pass any received data to its MAC layer.

The modem reports the results of these operations to the CMTS in the OPT-RSP message.

Results of the ODS and/or OPT exchanges may be evaluated by the CMTS itself or by third-party applications that interface with the CMTS. The specification does not prescribe exact algorithms for use of these results, since it is expected that profile evaluation will be an active area of vendor innovation.

Once a decision is made to assign a particular profile to a modem, the CMTS initiates an exchange of DOCSIS MAC

Management Messages to implement this assignment. For initializing modems, Registration messages may be used. Once a modem is operational, a DBC (Dynamic Bonding Change) exchange is used.

Profile Use in Packet Transmission

Once a profile has been assigned to a modem, the CMTS transmitter may use that profile to send traffic to the modem. As previously described, a modem may have up to four profiles enabled for reception of data. To avoid any issues with packet ordering or duplication, the CMTS is required to transmit all packets for a given Service Flow on a single downstream profile. If the Service Flow uses channel bonding, a single downstream profile must be used on each of the channels being bonded. If multiple Service Flows share the same resequencing DSID, all of these flows will have the same set of assignments of profiles on channels.

Typically, unicast Service Flows will use the highest-throughput profile currently assigned to the destination CM. For Group Service Flows (multicast flows), the selected profile will be one which is assigned to all CMs receiving the flow.

At the CMTS, Quality of Service (QoS) decisions are made to determine the order in which packets are to be transmitted. The basis for these decisions is unchanged in DOCSIS 3.1 – i.e., packets are classified to Service Flows and must pass through per-Flow rate-shaping before being presented to the downstream scheduling engine, which delivers packets for transmission based on relative priority and other QoS parameters. Importantly, profile assignment is not a QoS parameter and has no impact on QoS decisions. The rate, priority, and other QoS-related properties of a Service Flow are not altered by the fact that the Flow may be assigned to a lower- or higher-throughput profile.

Once QoS decisions have been made and a packet is ready for transmission, the CMTS modulator assigns subcarriers using the modulation orders specified by the DPD message describing the profile on which the packet will be sent.

Thus, the transmitter's use of profiles, and hence the mix of modulation orders used on the plant, is based entirely on the current mix of active traffic as managed by the downstream QoS and scheduling engines.

ESTIMATION OF CAPACITY AND UTILIZATION

With the optimizations possible through using multiple profiles come many challenges. Among these challenges is the question: What is the "speed," or capacity (maximum throughput) of the channel? If this value is approximately given by:

$$\begin{aligned}&\text{speed in bits per second} =\\ &\quad(\text{modulation order per subcarrier}\\ &\qquad\qquad\times \text{subcarriers per symbol}\\ &\qquad\qquad\times \text{symbols per second})\\ &\qquad\qquad- \text{overhead}\end{aligned}$$

then it can clearly be seen that using multiple profiles with different modulation orders gives multiple results for capacity. The impact of this can be substantial. The difference between 16 QAM and 4096 QAM is a factor of three, and different profiles may have different numbers of data-carrying subcarriers per symbol due to the use of zero-bit-loaded subcarriers to account for different interferers.

Similarly, how can the utilization of such a channel be measured? Typically this would be given by an equation such as:

$$\text{Utilization} =$$

$$\frac{\text{bits carried over a given time interval}}{\text{speed in bits per second} \times \text{interval duration}}$$

but since this relation includes the speed value from above, it suffers from the same dependency on modulation order.

Thus, a non-trivial question, both for designers and users of DOCSIS 3.1, is how to estimate downstream channel capacity and utilization in a meaningful way in the presence of multiple profiles. Because these metrics feed operator decisions about capital expenditures, it is vital to understand where the numbers come from and how they can be interpreted.

Normalization and the Reference Algorithm

The DOCSIS 3.1 OSSI addresses the first question in a simple, if limited, way: the "speed" of the downstream channel (as reported in the `ifSpeed` MIB object) is taken to be the throughput that could be achieved if all traffic on the plant used the profile with the highest throughput. This choice was made so that the behavior of `ifSpeed` for DOCSIS 3.1 would be consistent with previous DOCSIS versions, as well as with other interfaces such as Ethernet network ports, in the following ways:

- Reported `ifSpeed` is a constant (not time-varying) for a given channel configuration.
- Reported `ifSpeed` represents a maximum achievable rate; it is understood that actual throughput might approach but will never reach this value due to various factors (e.g., per-packet overhead on Ethernet interfaces).

One difference from other interfaces' treatment of `ifSpeed` is that DOCSIS 3.1 accounts for overhead due to FEC and other sources (discussed in more detail below).

Another important difference is that the actual speeds achieved on the channel may be much lower relative to the reported value when compared with interfaces such as DOCSIS 3.0 or wired Ethernet. In this respect, a DOCSIS 3.1 channel is more like a WiFi home networking channel, where the quoted speed can be achieved for devices close to the wireless access point, but distant devices may see significantly lower throughput.

This brings us back to the question of how the operator can extract meaningful metrics for channel utilization, so as to understand how heavily loaded the channel is and how loading might change under various projected scenarios.

To address this, the specification uses the concept of normalization and introduces a reference algorithm for making use of it.

To see how this concept is applied, it may be useful to imagine an example system using Profile A and Profile D. If Profile A is configured for 8 bits per symbol per subcarrier, and Profile D is configured for 12 bits per symbol per subcarrier, then all other things being equal (and discounting overhead for purposes of this example), the maximum possible throughput using Profile A would be two-thirds of that possible using Profile D. An inverse statement could also be made that a certain number of bytes sent using Profile A will consume 50% more of the channel than the same number of bytes sent using Profile D.

The above relationship can be used to estimate channel utilization in the presence of packets from both Profiles A and D. Bytes sent on Profile A "count" towards utilization 50 percent more than bytes sent on Profile D. If the `ifSpeed` of the interface is the throughput of Profile D in units of bits per second, then the fraction of the channel which

is utilized over some time interval T can be expressed as

$$\text{Utilization} = \frac{(\text{bytes sent on Profile D during } T) + 1.5(\text{bytes sent on Profile A during } T)}{\left(\frac{\text{ifSpeed}}{8}\right) \times (\text{duration of } T)}$$

In a real system, more than two profiles will probably be in use, and the relationship between their rates is probably not a simple fraction. Nonetheless, it is possible to generalize the above equation by defining a "normalization factor" for each profile:

$$\text{NF}(\text{Profile } P) = \frac{\text{ifSpeed}}{\text{throughput of } P}$$

By definition, the normalization factor of the highest-throughput profile on the channel is 1; other profiles will have normalization factors larger than 1.

For a set of Profiles P, the fraction of the channel which is utilized during some time interval T is then expressed as

$$\text{Utilization} = \sum_{P} \frac{(\text{bytes sent on } P \text{ during } T) \times \text{NF}(P)}{\left(\frac{\text{ifSpeed}}{8}\right) \times (\text{duration of } T)}$$

This equation is reflected in the reference algorithm given in the DOCSIS 3.1 OSSI specification for calculating utilization. It behaves as we would expect of such a metric. It reflects the fraction of the channel that is occupied, even if the resulting throughput in terms of bytes or bits is reduced due to the use of lower-throughput profiles. For instance, if traffic from a lower-throughput profile completely fills the channel, the actual rate will be lower than the rate reported in ifSpeed, and the resulting channel byte counts will reflect this, but the utilization value of 100 percent will correctly indicate that the channel is full and cannot carry any additional bytes.

How accurate is this formula? It turns out that it gives excellent results for most practical cases, but there are potential sources of inaccuracy due to non-ideal aspects of the system. Some can be accounted for in calculation, while others are dynamic and not directly visible.

To understand these items, it is first useful to study the sources of overhead on a DOCSIS 3.1 channel, particularly those related to FEC.

FEC Encoding on Multi-Profile Downstreams

Like previous versions of DOCSIS, the DOCSIS 3.1 downstream makes use of FEC to provide error correction capability. The cost of this capability is a loss in channel capacity due to the overhead of the parity bits that FEC adds to the data stream. In DOCSIS 3.1, a combination of block codes is used. An LDPC code contributes most of the error correction capability (and overhead), while a BCH outer code surrounds each LDPC block. A total of 1,968 parity bits are provided for each block of up to 14,232 information bits. The combination of information and parity bits is termed a "codeword." Codewords may be "shortened" by providing fewer information bits, but the number of parity bits per codeword is constant for these codes, so the efficiency of shortened codewords is lower than that of full-length ("full") codewords. Specifically, the efficiency of a full codeword is about 87.8 percent, while the efficiency of a shortened codeword can be virtually any number lower than this (as codewords down to one information bit are allowed).

Data packet sizes are, of course, variable, and in any given queue of packets to be transmitted, packet boundaries will almost never align with full codeword boundaries. To avoid constant codeword shortening, "streaming" of packets across codeword boundaries is typically performed. As much MAC-layer data as will fit is placed into a codeword, and if this results in only a partial packet being sent, the remainder of the packet is carried over to the next codeword. Two bytes at the beginning of each codeword are used as a "header" which provides a pointer for locating the first packet boundary within the codeword.

If only a single profile were in use, packets could be streamed continuously into codewords and codeword shortening would never be needed. However, the use of multiple profiles introduces new complexity. In order to properly decode an LDPC/BCH codeword, a receiver must be able to correctly receive all bits of the codeword. But it is the nature of a multi-profile system that not all receivers can decode all profiles. If a codeword were allowed to contain bits from packets being sent on different profiles, some receivers would not be able to decode it, and these receivers would lose the entire codeword, including the portion sent on profiles they can receive.

This problem is solved by assigning codewords to profiles. When the transmitter wants to send a particular packet, it creates a codeword and assigns that codeword to use the profile of the packet to be transmitted. Other packet data using the same profile can also be included in the codeword, but if no such data is available, the codeword must either be filled with MAC-layer "stuff bytes" or shortened. It cannot be filled with packet data using other profiles.

This ensures that each receiver can fully decode all codewords it needs, but how does the receiver know which codewords use which profiles, and where in the data stream they are located? DOCSIS 3.1 solves this by devoting a few subcarriers out of each OFDM symbol to communicating "Next Codeword Pointers," or "NCPs." NCPs indicate the start location of each codeword by means of a pointer to a subcarrier number, and also state which profile the codeword uses. NCPs use a very low modulation order (configurable from QPSK through 64 QAM) and are placed on subcarriers which can be received by all modems. When the transmitter creates each codeword, it also creates a corresponding NCP on the symbol during which the codeword begins.

NCPs consume subcarriers that might otherwise be used for data, so they contribute to the total overhead lost from the channel. The number of NCPs may vary from one OFDM symbol to the next, depending on how many codewords are included in that symbol. This means that the amount of NCP overhead may also vary. On most channels, the amount of overhead is 3 percent or less.

Throughput Calculation for a Single Profile

Besides FEC and NCP overhead, there are several other items that must be included if overhead is to be accounted for in calculating throughput.

On every OFDM channel, some subcarriers are dedicated to functions other than carrying data, and the transmitter must avoid using these subcarriers. PLC (PHY Link Channel) and continuous pilot subcarriers are defined in the OCD message and remain in a fixed location once defined. These contribute to channel overhead, but do not vary from symbol to symbol.

In contrast, scattered pilots occur on a different set of subcarriers every symbol, in a pattern that repeats every 128 symbols. Every potentially data-carrying subcarrier on the channel will contain a scattered pilot at some

point. When that happens, the subcarrier is not used for data and whatever number of bits might normally be carried on it is lost from the total available on the symbol. The exact amount of loss this represents depends on the modulation order assigned to that subcarrier on whatever profile might have been used on that subcarrier, had it been used. Of course, this is impractical to predict, and so this overhead is difficult to quantify on an instantaneous basis. However, it can be accounted for in a calculation that assumes that same profile is used for an extended time period (at least 128 symbols).

Based on all of the above items, and with careful study of the PHY specification and references, it is possible to calculate the post-overhead maximum throughput using a single profile to a very high degree of accuracy. A few assumptions must be made:

					Formula
symbol size	40	usec			
cyclic prefix	1.5	usec			
total symbol duration	41.5	usec			symbol size + cyclic prefix
symbols per second	24096				1/total symbol duration
subcarrier spacing	25	kHz			1/symbol size
channel width	96	MHz			
number of subcarriers	3840				channel width / subcarrier spacing
guard subcarriers			40		constant
PLC subcarriers			16		
continuous pilot subcarriers			24		48 * (channel width/192)
total exclusion band width	8	MHz			
excluded subcarriers			320		exclusion band width / subcarrier spacing
usable subcarriers	3440				number of subcarriers - (guard+PLC+pilot+excluded)
data bits/subcarrier/symbol	10				
raw data bits per symbol	34400				usable subcarriers * data bits/subcarrier/symbol
scattered pilots factor (apprx)	99.22%				127/128 (constant)
pre-NCP data bits per symbol	34131				raw data bits * scattered pilots factor
first-pass full codewords per symbol			2.106852		pre-NCP data bits per symbol / full codeword bits
NCP bits/subcarrier/symbol	4				
NCP subcarriers for floor(codewords)			36		(48/ NCP bit loading) * [floor(first-pass codewords)+1]
NCP subcarriers for ceiling(codewords)			48		(48/ NCP bit loading) * [ceiling(first-pass codewords)+1]
interpolated average number of NCP subcarri			37.3		NCP subcarriers for floor + (codewords - floor(codewords))* (NCP subcarriers for ceiling - NCP subcarriers for floor)
post-NCP average number of data subcarriers			3402.7		usable subcarriers - interpolated average number of NCP subcarriers
post-NCP average data bits per symbol			33761		post-NCP subcarriers * data bits/subcarrier/symbol * scattered pilots factor
NCP overhead			1.1%		(pre-NCP data bits - post-NCP average data bits)/pre-NCP data bits
FEC efficiency for MAC data			87.75%		MAC-layer bits per full codeword / full codeword total bits (constant)
MAC-layer average bits per symbol			29627		post-NCP average data bits per symbol * FEC efficiency for MAC data
Average MAC-layer throughput			**714**	**Mpbs**	MAC-layer average bits per symbol * symbols per second
FEC constants:					
full codeword total bits	16200				parity bits + information bits (constant)
codeword parity bits			1968		constant
full codeword information bits			14232		constant
codeword header bits				16	constant
MAC-layer bits per full codeword				14216	information bits - codeword header bits (constant)

Table 1: Simplified Calculation of MAC-Layer Throughput for a Single Profile

- The system will be observed over a relatively long interval (say, seconds or more), guaranteeing that symbol-to-symbol variations are averaged out.
- The system is fully loaded with packets for a single profile so that the transmit queue never underruns and the transmitter never needs to switch profiles.
- Full codewords are always used (although the specification does not require it, a quality transmitter implementation is expected to use full codewords under the conditions described).

The resulting calculated throughput describes the amount of DOCSIS MAC-layer data that can be carried on a single profile, accounting for PHY-layer overhead, but not subtracting out MAC-layer overhead such as DOCSIS® MAC headers. This is the value that would be used in the formulas given above for the normalization factor and channel utilization.

The complete details of such a calculation are beyond the scope of this paper. However, a simplified example is shown in Table 1. The example uses a 96 MHz wide downstream channel with an exclusion band of 8 MHz, which might occur if it were necessary to avoid some other service using a 6 MHz carrier (allowing a 1 MHz guard band on each side). For simplification, the example assumes a constant bit loading of 1024 QAM (10 bits per symbol) on all subcarriers. It also uses approximations for NCP and scattered pilot overhead, which have high validity if the system is observed over many symbols. The interpolation approach used for NCP overhead gives excellent results if the number of codewords per symbol is not too close to an integer, as in this example.

IMPACT OF MULTIPLE PROFILES

Multiple Profiles and Convergence Layer Functionality

In a real system, more than one profile will be used at a time. In DOCSIS 3.1, the downstream convergence layer is primarily responsible for implementing multi-profile transmission.

As previously discussed, each FEC codeword is assigned to use a particular profile, but packet boundaries rarely align with codeword boundaries. What should a transmitter do if it does not have enough packet data available on a particular profile to fill a codeword? This could happen either because the packet(s) in the queue are too short, or because, after filling some number of codewords in the streaming fashion previously described, there is a partial packet remaining that does not fill a codeword.

In this situation, a transmitter has two choices: it can shorten the codeword or it can wait for more packets to arrive for the profile in question. Either option has drawbacks. Codeword shortening has a high cost in efficiency: a 64-byte packet sent in a codeword by itself is about 20 percent efficient. If even a small fraction of packets is sent in this way, the loss for the channel as a whole may be considered significant. However, waiting for more data adds latency to the packets which are already available. Waiting might also violate the QoS principles being enforced at higher layers of the CMTS. While the transmitter is holding a packet on one profile, packets that may have lower QoS priority are being sent ahead of it on other profiles. The fact that a user happens to be on a profile for which little other traffic is available should not be allowed to alter the QoS guarantee to that user.

The solution adopted by DOCSIS 3.1 is a hybrid of these two choices. The transmitter is allowed to hold data for a certain small amount of time in hopes of receiving more data to fill a codeword. Within that window of time, packets from different profiles may

be transmitted in a different order than they were originally queued (although ordering must be maintained within a single profile). Once the maximum wait time is reached, the transmitter is expected to shorten codewords, giving up efficiency in exchange for meeting the desired latency bounds. A typical latency bound is 200 microseconds, although this value can vary, as will be discussed in more detail below. The intent is that the latency bounds are small enough so as not to interfere with QoS guarantees, although they may be large enough to impact latency budgeting for some services.

Figure 1, taken from the DOCSIS 3.1 PHY specification, shows a sample system architecture used for describing and specifying (though not necessarily implementing) this functionality. Packet forwarding and QoS, including shaping to the instantaneous channel rate, are performed in the leftmost blocks of the diagram. Packets are queued "single file" for transmission and passed to the convergence layer, which contains per-profile buffers that are used to hold data while waiting to fill a codeword. The wait time is expected to be constrained to match the configurable latency targets. The "codeword builder," shown within the PHY block, is the primary decision-maker. It makes a "best guess" tradeoff to decide which profile to use for each codeword, attempting to maximize efficiency without exceeding latency targets.

Importantly, there are no specification requirements stating when a codeword builder sends shortened codewords rather than full codewords. An implementation is free to send shortened codewords at any time, for any reason. As an example, when a channel is lightly loaded, an implementation may intentionally send shortened codewords to increase channel robustness (shorter codewords are more robust since there are more parity bits relative to the number of information bits). As another example, an implementation may choose to prioritize latency over efficiency by sending packets

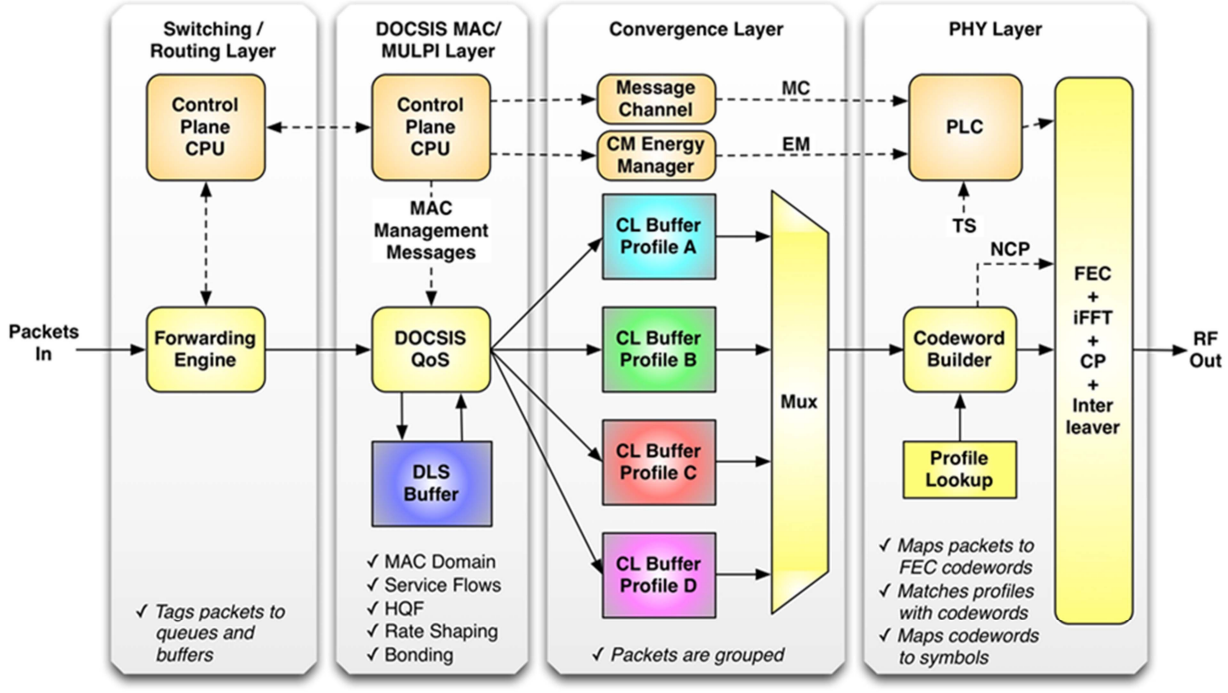


Figure 1: Sample System Architecture for Multi-Profile CMTS Downstream Transmitter (from DOCSIS PHY specification)

with less wait time when the channel is lightly loaded. In both cases, since the channel is not full, codeword shortening may be seen as a good use of otherwise unutilized capacity on the channel. If the load increases, the codeword builder can decide not to perform these types of codeword shortening. In other words, the bandwidth used by such cases of codeword shortening can be recovered if it is needed for actual data traffic.

Thus, it is expected (though not stated in the specification) that under conditions of heavy load, a codeword builder will use mostly full codewords. This assumption is built into the utilization reference algorithm, which contains no allowance for overhead due to codeword shortening. If the assumption is incorrect, then the reference algorithm will be less accurate than we would like. Thus, it is important to study conditions under which this assumption might not hold.

Codeword Builder Behavior Under Worst-Case Conditions

First consider a case in which the system is 100 percent loaded with traffic, all for a single profile. As already discussed, in this case it is expected that the codeword builder will virtually always use full codewords. In a steady-state condition, the buffer for the profile is always full, so full-size codewords can be created with no wait time.

Next, consider a case in which the system is 100 percent loaded with traffic, but this time for two different profiles. In this case, the codeword builder behavior is affected by the very-short-term traffic mix at the input to the convergence layer, since this will determine the fullness of the two convergence layer profile buffers. At one extreme, packets for the two profiles could arrive in alternating fashion – for instance, one packet for Profile A, immediately followed by a same-sized packet for Profile B, and so forth. In this case, the steady-state condition for the codeword builder would be that both buffers are equally full, and so full codewords can always be created. Each profile has to wait up to the duration of one codeword while the other profile occupies the channel, but this time is short compared to the latency targets.

At the other extreme, packet distribution for the two profiles could be extremely unbalanced at the codeword builder input; one profile – say Profile A – could dominate, with only short, infrequent, widely-spaced packets for Profile B. This situation is actually quite plausible. If the only traffic active on Profile B is, say, a Voice-over-IP flow, the packets on Profile B will be short and the QoS engine will ensure that they are spaced according to the packetization interval of the service. This interval (10 or 20 milliseconds) is much longer than the typical sub-millisecond latency target of the codeword builder. Even for services such as best-effort downloads and web browsing, the traffic shaping function of the DOCSIS QoS will ensure that packets for the flow are not clumped together, but are more-or-less evenly spaced.

Thus, any time a profile carries traffic for only a small number of Service Flows, a "widely spaced packets" scenario is likely to occur.

In this scenario, the codeword builder will primarily create full codewords for the dominant profile – say, Profile A. When a packet for Profile B arrives, it might attempt to wait for more data on Profile B, but that data will not arrive in time due to the packet spacing enforced by the QoS engine. The codeword builder will be forced to send only shortened codewords for Profile B.

The shortening of codewords in this case represents an actual loss of capacity for the system due to the use of multiple profiles. Had the traffic been sent on Profile A instead, it could have been included as part of a full codeword. However, with the traffic

occurring as isolated packets on a different profile, there is no option but to use lower-efficiency shortened codewords. In other words, the bandwidth consumed by codeword shortening cannot be recovered if the shortening is dictated by the traffic pattern and latency targets.

The worst-case loss for this scenario occurs when the spacing of packets on Profile B is infinitesimally larger than the latency target for Profile B. In this case, the codeword builder waits up to the latency target, then releases a codeword containing a single packet – then immediately thereafter, a new packet arrives, which in turn is delayed up to the latency target.

Now suppose that additional profiles are configured on the system, up to the maximum of 16. The worst case here occurs when one of the profiles – say Profile A – dominates the traffic, and all of the other 15 profiles exhibit the worst-case behavior described in the example above for Profile B. In this case, during every time interval equal to the duration of the latency target, 15 codewords must be shortened.

With this worst-case pattern in mind, it is possible to calculate the capacity lost due to required codeword shortening for a variety of parameter sets.

Table 2 illustrates the complete calculation for an example case. In this example, to

							Formula:	
	Average MAC-layer throughput (full codewords)		714	Mbps				
	Full codeword FEC efficiency for MAC data		87.75%				MAC-layer bits per full codeword / full codeword total bits	
	Average bit rate not counting FEC overhead		814	Mbps			MAC-layer throughput / full codeword efficiency for MAC data	
	FEC constants:							
	full codeword total bits	16200					parity bits + information bits (constant)	
	codeword parity bits		1968				constant	
	full codeword information bits		14232				constant	
	codeword header bits			16			constant	
	MAC-layer bits per full codeword			14216			information bits - codeword header bits (constant)	
	shortened codeword length (MAC bits)	608					76 bytes * 8 bits/byte (constant)	
	interval	200	usec					
	bits in interval	162705	bits				interval * average bit rate not counting FEC overhead	
Profile A interarrival time	200	usec	Profile A cws in interv		1		interval / interarrival time	
Profile C interarrival time	200	usec	Profile C cws in interv		1		interval / interarrival time	
Profile D interarrival time	200	usec	Profile D cws in interv		1		interval / interarrival time	
	total shortened codewords in interval				3		sum of shortened codewords over all profiles	
	parity+header bits for shortened codewords in interva				5952		total shortened codewords * (parity + header bits per codeword)	
	MAC data bits for shortened codewords in interval				1824		total shortened codewords * MAC-layer bits per codeword	
	total bits for shortened codewords in interval				7776		parity+header bits + MAC data bits	
	bits for profile B in interval	154929					bits in interval - total bits for shortened codewords in interval	
	Profile B codewords in interval	9.563534					bits for profile B / full codeword total bits	
	Profile B parity+header bits in interval	18974					Profile B codewords * (parity + header bits per codeword)	
	Profile B MAC data bits in interval	135955					Profile B codewords * MAC-layer bits per codeword	
	Total MAC data bits sent during interval		137779				Profile B MAC data bits + MAC data bits for shortened codewords	
	Total MAC data bits possible with all full codewor		142779				bits in interval * full codeword FEC efficiency for MAC data	
	Lost capacity due to codeword shortening		**3.5%**				(MAC data bits possible - MAC data bits sent) / MAC data bits possible	

Table 2: Sample Calculation of Codeword Shortening Capacity Loss Using Worst-Case Traffic

simplify the calculation, it is assumed that all profiles have the same bit rate. Although unlikely, this could occur on a real channel if the profiles targeted similar SNR but different exclusion bands. Four profiles are assumed to be in use, each with the same parameters as in Table 1. All four profiles have a latency target of 200 microseconds; thus, 200 microseconds is the interval over which the analysis is performed. It is assumed that Profile B dominates the traffic and uses only full codewords. The other profiles carry only a single 64-byte packet plus DOCSIS headers (76 bytes total) every 200 microseconds, requiring the use of a single shortened codeword once per interval.

Starting from the pre-FEC throughput of the profiles, the calculation determines the total number of shortened codewords and full codewords which are sent on a fully loaded channel during the analysis interval. This, in turn, defines the total number of FEC parity and codeword header bits. Subtracting these bits from the total available during the interval gives the number of bits which carry MAC-layer data. The difference between this value and the amount of MAC-layer data that could be carried using only full codewords represents the lost capacity on the channel as a result of using multiple profiles with a worst-case traffic pattern. In this example, the lost capacity amounts to about 3.5 percentof the total channel bandwidth.

In interpreting this result, recall that it is based on a worst-case traffic pattern in which traffic across profiles is extremely unbalanced, and the interarrival time of packets on a lightly-loaded profile is essentially equal to the profile's latency target. Although a pattern of widely-spaced packets is probably typical for lightly-loaded profiles, it is unlikely that the interarrival time will be so consistent. Sometimes it will be larger, in which case codewords will be shortened less often. Sometimes it will be smaller, in which case some aggregation of data packets can take place, resulting in shortened codewords that are longer and hence more efficient than those shown in the analysis.

Thus, the result of a calculation such as that shown in Table 2 should be taken as an upper bound on lost capacity which is very unlikely to be reached in a real system observed over a typical time period (seconds or more).

So how can the operator know what actual capacity loss might be taking place in a real system? Ideally, it could be reflected in reported utilization statistics. However, as previously mentioned, this loss of capacity is not accounted for in the reference algorithm for calculating utilization.

Importantly, a real-world implementation of a codeword builder may have little or no ability to account for this loss. The reasons may vary with implementation, but some considerations are:

- The codeword builder algorithm may consist of a large number of decision-making rules, and thus it may lack the ability to determine a "reason" why a particular rule set resulting in certain codeword sizes was invoked at any given instant.
- The codeword builder algorithm may need to use predictive techniques in order to meet latency targets; for example, it cannot wait until the last possible instant to decide to shorten codewords for all 15 different profiles, since 15 codewords will not fit in one OFDM symbol. Determining whether a different prediction would have given a "better" result is beyond the scope of an implementation.

Thus, it is entirely possible that capacity could be lost without any way of making this visible to the operator.

The DOCSIS OSSI specification allows a CMTS to deviate from the reference algorithm when reporting utilization by incorporating any vendor-specific information it may have that points to actual capacity loss in the system due to codeword shortening. Vendor innovation in this area might lead to CMTS reporting that improves upon the reference algorithm.

Regardless, the best way for an operator to cope with this potential capacity loss is probably to avoid it in the first place. This is most easily done by configuring profiles so that traffic is relatively balanced among them. Although this does not guarantee that codeword shortening will never be needed (since even when the long-term traffic mix is balanced, the very-short-term packet mix presented to the input of the convergence layer could be virtually anything), it does prevent a pattern of repeated codeword shortening such as that studied in the above analysis. Over time, occasional intervals requiring codeword shortening are likely to be balanced out by those in which traffic aggregation allows full codewords to be used so that the net capacity loss becomes negligible.

If it not possible to avoid potentially problematic configurations, the next best thing is for the operator to understand the contributing factors and their potential impact. It has already been seen that traffic imbalance across profiles is a major "red flag." Other contributors include the number of profiles, latency targets, and channel bandwidth. The methods of the previous analysis can be used to explore the effects of these factors.

Analysis of Multiple Configurations

The DOCSIS 3.1 specification recognizes the relationships described above and offers recommendations for latency targets based on channel bandwidth and number of profiles. These recommendations are shown in Table 3, which is taken from the DOCSIS 3.1 MULPI document (table 7-12 in that document).

It is recognized that when channel bandwidth is narrower, fewer full codewords can be transmitted in a given amount of time, making it harder for the codeword builder to achieve a certain latency without codeword shortening. Thus, latency targets are higher for narrow channels. Higher targets are also suggested for larger numbers of profiles.

In order to make it possible for an operator to deliver latency-sensitive services using configurations with higher targets, the specification suggests that some profiles have lower latency targets than others. Modems receiving a low-latency service are assigned profiles from among the limited subset of low-latency profiles, while other modems can be assigned profiles selected from the full set.

Total CMTS Profiles	**Minimum Profiles per Latency Target**	**Max Latency (us) based upon OFDM Channel Bandwidth (MHz)**			
		24	48	96	192
4	1	600	400	200	200
	3	800	400	200	200
8	2	800	400	200	200
	6	2400	1600	800	400
16	4	1600	1200	800	400
	12	3200	2400	1600	800

Table 3: Codeword Builder Latency (from DOCSIS MULPI specification)

The values in Table 3 are only recommendations, and vendors or operators may choose to configure different values. In particular, for some combinations, the latencies given are quite large. Values greater than a few hundreds of microseconds present challenges for low-latency services and may also impact DOCSIS performance in general. Operators may choose to observe these

latency targets, configure more aggressive targets at a cost in efficiency, or avoid combinations for which the table gives high recommended values.

To analyze cases in which different profiles have different latency targets, the calculation of Table 2 can be done using an analysis interval equal to the least common multiple of the latency targets. Profiles with targets shorter than the interval will have multiple codewords shortened during the interval.

It is also possible to incorporate different rates per profile by incorporating a normalization factor. Such calculations are beyond the scope of the current paper.

Table 4 summarizes the results of calculations performed for various combinations of number of profiles, channel bandwidth, profile throughput (assumed to be the same across all profiles), and latency targets. Where two different targets are given, the dominant profile is assumed to use the larger value. The worst-case traffic pattern described above is assumed for these calculations. This is only a small number of possible options, but should give an idea of the type of results that can be expected.

Several points can be gleaned from the data in Table 4:

- Increasing latency targets beyond those recommended in MULPI has serious consequences under worst-case traffic conditions, often tripling the loss due to codeword shortening.
- Low-rate channels present the greatest challenge for a latency vs. efficiency tradeoff. Performance is improved if the channel bandwidth is increased, and also if the average bit loading on the channel is increased – i.e., if

channel bandwidth after subtracting exclusion bands (symbol size)	bit loading (assumed constant across all subcarriers)	profile throughput (all profiles)	number of profiles	latency targets (usec)	worst-case % lost capacity
24 MHz (20 usec symbol)	10	160 Mbps	4	per spec (600/800)	4.4 %
				200/600	8.7 %
			8	per spec (800/2400)	4.8 %
				200/1200	14.8 %
			16	per spec (1600/3200)	6.2 %
				400/1600	17.6 %
96 MHz (40 usec symbol)	10	781 Mbps	4	per spec (200 for all)	3.2 %
			8	per spec (200/800)	3.5 %
				200 for all	7.5 %
			16	per spec (800/1600)	2.5 %
				200/800	7.2 %
192 MHz (40 usec symbol)	8	1262 Mbps	4	per spec (200 for all)	2.0 %
			16	per spec (400/800)	3.1 %
				200 for all	9.9 %
	12	1888 Mbps	4	per spec (200 for all)	1.3 %
			16	per spec (400/800)	2.1 %
				200 for all	6.6 %

Table 4: Summary of Results of Capacity Loss Calculation for Selected Parameter Sets

the channel is dominated by modems using a higher modulation order.

- The best performance in both latency and efficiency is achieved with smaller numbers of profiles. Not only do these cases have the lowest potential capacity loss in the presence of worst-case traffic, but in a real system, it is easier to balance traffic across profiles when there are only a few of them, making a worst-case traffic pattern much less likely.

RECOMMENDATIONS

Based on the above discussion and analysis, it is recommended that operators:

- Resist the urge to configure profiles that are usable by only one or a few modems on the plant. Although this may appear to offer better performance by allowing use of modulation orders more tailored to individual devices, it could have the opposite effect by creating suboptimal traffic patterns and forcing codeword shortening that reduces channel efficiency. An ideal profile set would have similar amounts of traffic on each profile, with enough different Service Flows to make aggregation likely at the codeword builder.
- Become familiar with the MULPI recommendations for profile latency targets. Ideally, configurations which result in high recommended latency targets should be avoided. If this is not possible, the operator should be prepared to tolerate either high latency (if recommended targets are observed) or low efficiency (if more aggressive targets are configured).
- Be aware that capacity loss due to codeword shortening may not be reflected in reported system statistics. In particular, operators should look for imbalance in traffic load across profiles, especially in systems using higher numbers of profiles and/or narrower channels. If these conditions are present, reported utilization values may contain inaccuracy of as much as a few percent.
- Rely on more than just a single statistic when planning node splits or plant upgrades. Utilization statistics remain an excellent guide, but they may vary over time depending on which modems are active, even if the plant has been configured to avoid hidden capacity loss. When utilization nears a value of interest, operators should monitor other statistics, such as those reflecting buffer fullness or packet loss, to determine whether a channel is overloaded.
- Incorporate the concept of normalization into network planning. When a node is split, subscribers added, SLAs increased, and so forth, it should be recognized that devices using lower-throughput profiles will impact the network more than devices using higher-throughput profiles, even if the SLA to both devices is the same.

SUMMARY AND CONCLUSIONS

This paper has offered an overview of how multiple modulation profiles may be defined and used in a DOCSIS 3.1 system. The concept of normalization was introduced, showing that the contribution to channel utilization of bytes sent on a profile is proportional to the reciprocal of the profile's maximum throughput. In examining the details of FEC codeword overhead and scheduling of profiles, it became clear that there are certain cases where inefficiencies can result in lost channel capacity. This occurs when some profiles are lightly loaded and carry few Service Flows, so that the operation of DOCSIS QoS results in infrequent, widely-spaced packets on those profiles. This traffic condition forces the transmitter to use less-efficient shortened codewords, reducing the bandwidth available

for carrying data. This lost capacity may not be reflected in reported utilization statistics.

It is important for the operator to be aware of the conditions under which this capacity loss can occur, ideally to configure the plant so as to avoid it, or, if this is not possible, to know when it might be taking place so that it can be accounted for in network evaluation and planning.

The best performance is achieved when small numbers of profiles are used on wide channels with high rates. When the number of profiles is larger or channel bandwidth is smaller, a tradeoff between latency and efficiency must be made. The DOCSIS 3.1 MULPI specification recommends latency values which can be expected to give good efficiency. If more aggressive latency targets are configured, efficiency can be expected to drop sharply.

For any configuration, the chance of capacity loss is minimized if traffic is distributed more-or-less evenly across profiles. This maximizes the chance that enough packet aggregation will occur to prevent forced codeword shortening. Profiles that serve only a few modems are the most likely to require codeword shortening that results in reduced capacity.

With an understanding of the factors at work and some attention to system configuration and monitoring, operators can deploy multiple modulation profiles with high efficiency and account for the impact of this feature on network utilization.

ACRONYMS AND ABBREVIATIONS

CM – Cable Modem
CMTS – Cable Modem Termination System
DBC – Dynamic Bonding Change
DOCSIS® – Data Over Cable Service Interface Specifications
DPD – Downstream Profile Descriptor
FEC – Forward Error Correction
MAC – Media Access Controller
MAP – Bandwidth Allocation MAP
Mbps – Megabits per second
MHz – Megahertz
MIB – Management Information Base
MULPI – MAC and Upper Layer Protocols Interface
OCD – OFDM Channel Descriptor
ODS – OFDM Downstream Spectrum
OFDM – Orthogonal Frequency-Division Multiplexing
OPT – OFDM Downstream Profile Test
OSSI – Operations Support System Interface
PHY – Physical layer
PLC – PHY Link Channel
QAM – Quadrature Amplitude Modulation
QPSK – Quadrature Phase-Shift Keying
QoS – Quality of Service
RxMER – Receive Modulation Error Ratio
SLA – Service Level Agreement
SNMP – Simple Network Management Protocol
SNR – Signal to Noise Ratio
UCD – Upstream Channel Descriptor

REFERENCE ACQUISITION

DOCSIS® specifications: Cable Television Laboratories, Inc. “CableLabs - Specifications” [Online] Available: http://www.cablelabs.com/specs

OPERATIONAL CHALLENGES OF 85MHZ DEPLOYMENT

Carol Ansley
ARRIS Group, Inc.

Abstract

Upstream expansion past the current 45MHz or 65MHz to increase available upstream bandwidth has been discussed for many years, but now it's about to become a reality. The standards are all finished and hardware is becoming available. But before a successful real world deployment can happen, a plethora of operational challenges must be assressed.

This paper gives an overview of the causes for concern, presenting set top box sensitivity testing results, as well as other characterization tests. Results from testing of existing taps are used to explain potential neighborhood effects of expanded upstream deployment that could pose operational headaches if not addressed proactively.

OPERATIONAL CHALLENGES OF 85MHZ DEPLOYMENT

Changing an HFC network from a 42MHz split to an 85MHz split, or an even higher frequency split such as 200MHz, presents significant challenges in the home and local access networks. These challenges result from the dramatic change in RF level experienced in the home and local access network at the frequency band shifting from downstream usage to upstream usage. This paper discusses the effects of the higher power levels in the home as well as in the local network, and presents some coping strategies to allow the network upgrade to proceed without requiring a wholesale upgrade of in-home equipment.

Why is there an issue?

Ideally, the change of the HFC frequency split would not cause any issues. In-home equipment, though capable of tuning to the frequencies now dedicated to upstream use, would have no reason to tune to that band once the downstream channels are removed. But, the tuners within settop boxes, cable modems, and cable-ready televisions are not perfect. The presence of significant energy in their originally designed reception band even outside of the nominal tuned channel can cause distortions due to automatic gain control (AGC) circuitry.

An AGC circuit exists to protect a device's internal tuner from too much energy in the incoming signal and to boost the signal when it is too low. Many AGC circuits integrate the power received over a wide bandwidth, not just the narrow channel currently desired, and adjust the level into the tuner accordingly. When the AGC circuit detects that the incoming signal power is too high, it will attenuate the signal to protect the tuner. Harmonic distortion or similar issues caused by a powerful upstream signal above 54MHz in the tuner may also contribute to the problem, but AGC effects appear to account for most of the issue.

The upstream transmit signal from an 85MHz cable modem (CM) is potentially quite powerful. As an example, in DOCSIS 3.1, the transmit level of a minimally sized 24MHz signal can be up to 56dBmV relative to 6MHz. As is shown in the figure below, the transmitted signal will generate reflections within each splitter that the signal crosses.

The isolation of the ports in the splitter sets the amount of reflection that will be seen on the other ports. Common inexpensive splitters may have as little as 20dB of isolation from port to port. Better quality splitters may have more loss, resulting in less reflected energy, but at the DOCSIS3.0 or 3.1 power levels, some reflected energy is to be expected. That reflected energy, as is shown

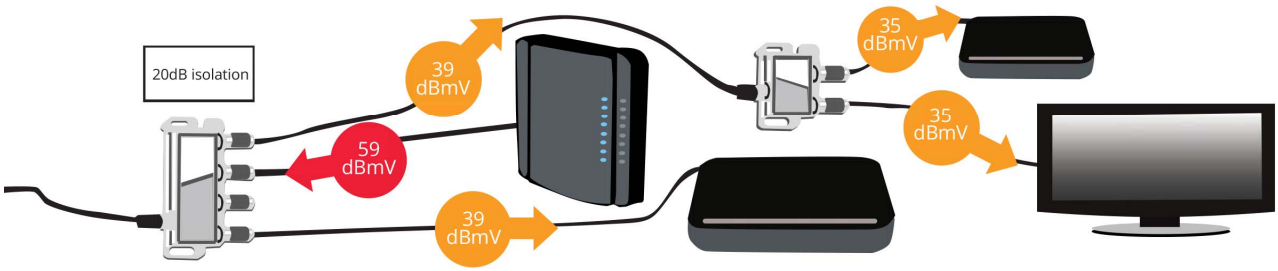


Figure 1 - Showing Reflected Signal in the Home

in the diagram, travels back through the rest of the home network.

When the reflected signal reaches a device, its internal AGC may react. Using the earlier example of a 56dBmV signal reflected with a loss of 30dB, the reflected signal could reach a directly connected set top box at 26dBmV, a much higher level than the typical 0dBmV downstream video signal within the home. When an AGC detects a high power level, it may attenuate the entire downstream signal. If the attenuation is too high due to 85MHz US reflections, the desired downstream video signal will be degraded or lost. The amount of additional attenuation present between the reflected signal and the affected device will vary depending upon the splitter arrangement within the home.

ARRIS tested a collection of set top boxes to determine what level of reflected signal tended to cause problems. Results of the testing are shown in the figure below.

The devices under test were tuned to a 256 QAM video channel at 111 MHz, with a full video channel load present up to 945 MHz. A block of upstream signals was added between 54.3 MHz and 74.5 MHz. The block was varied in level to discover at what level video degradation occurred. As would be expected, the level of the video signal as well as the level of the upstream signal both influenced the results. The resulting rule of thumb is that the difference between downstream video level and the 85 MHz upstream level must be less than 25 dB at the F-connector of the legacy device to avoid problems in the majority of set top boxes.

The problem of unexpected reflected energy

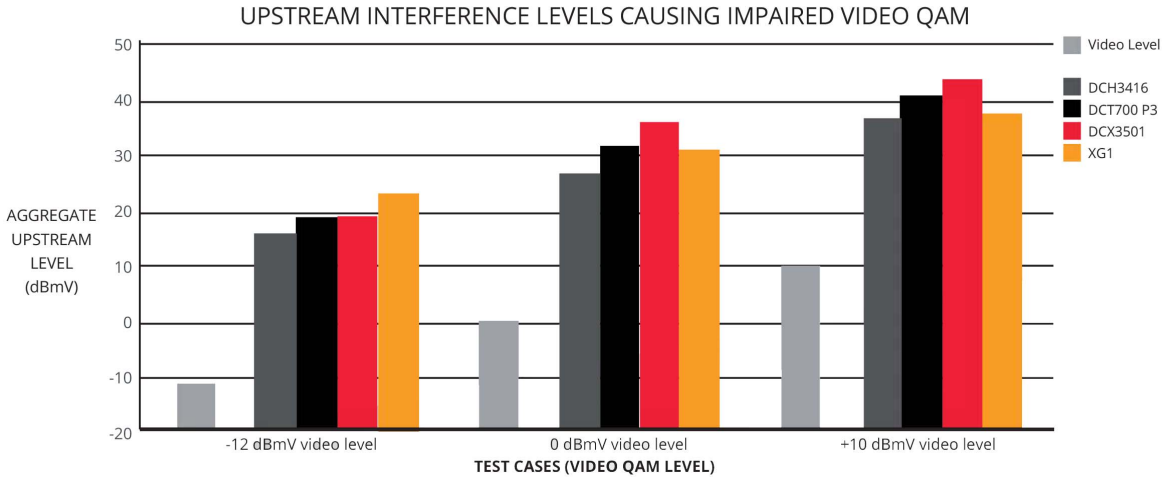


Figure 2 - Interference Testing Results

in the downstream tuner's receive band is not limited to the home where a new 85MHz CM is deployed. ARRIS also tested the performance of some taps. The test setup annotated with sample test results is shown below.

The internal circuitry of a tap is equivalent to a directional coupler tapping the trunk line with a splitter dividing that tapped energy to the tap's F-connector ports. When a signal is sent from one home into a port on a tap, reflections are generated in the splitter circuitry within the tap. Another home that shares a tap's internal splitter can receive a high level of reflected energy. In testing, reflections with as little as 24dB of loss were seen. Since a set top box in another home will still have drop cable loss as well as in-home splitter loss to reduce the reflected signal's power, the possibility of degraded performance is not as high as it is in the original home, but if the incoming video level is low, the reflected signal could still cause problems.

Diagnosing 85MHz Related Issues in the Field

Operationally, the worst problems to troubleshoot are the intermittent ones. Problems caused by AGC overload due to upstream transmissions below 85MHz will be intermittent, showing up only when the new CM is transmitting. A report of occasional video break ups can have many possible causes, and 85MHz reflections add another possible cause. As noted above, it is even possible that the subscriber reporting the problem does not have a D3.1 modem themselves; it could be their neighbor's new modem that is affecting their home through reflections at a shared tap. Aside from the splitter contributions, the hysteresis timing of the set top box's AGC circuit can also affect whether or not a problem is seen. If the AGC circuit has a slow hysteresis, then the shorter CM bursts may not even be detected by the AGC. But if the AGC is triggered, a slow

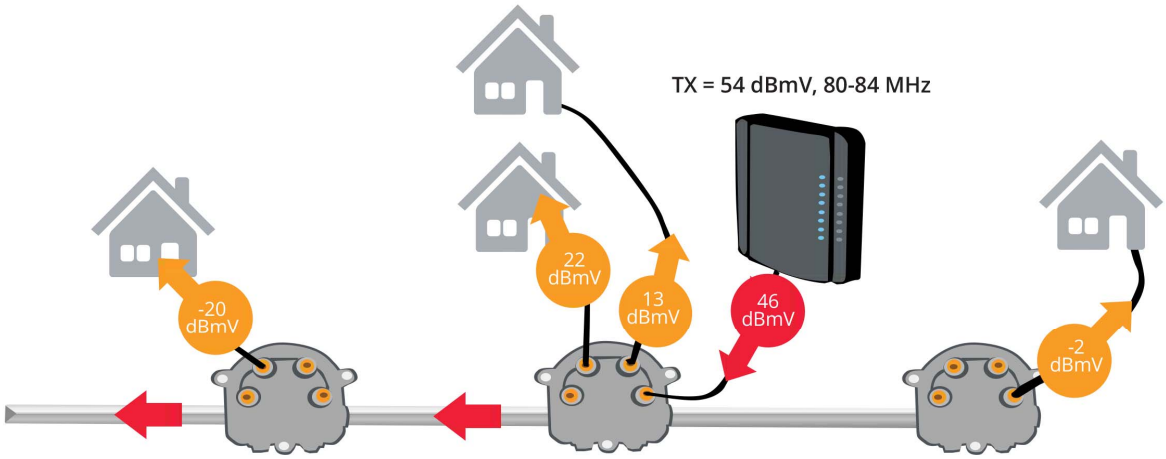


Figure 3 - Example Levels Seen Across Tap Ports

release may cause the problem to be visible to the affected customer even after the CM has ceased transmissions.

Because of the overlap in the use of video services with High Speed Data (HSD) services that has been shown in other studies, subscribers are likely to be watching

television at the same time that other people in their home or their neighbors are likely to also be using HSD services. If a reflected signal issue exists, it may be seen during prime time in the evening, but may disappear the following monring. A technician who arrives at the home may not see any video degradations and, unless trained to also look for 85MHz modem issues, may dismiss the problem as a temporary interruption in the plant.

Preventing 85MHz Related Issues

Several alternatives exist to remediate a home where a new 85MHz modem has been installed; the sensitive home equipment can be protected individually or addressed by remediating the home overall.
One approach to preventing in-home issues is to use a two-port cable modem for 85MHz deployments as shown below.

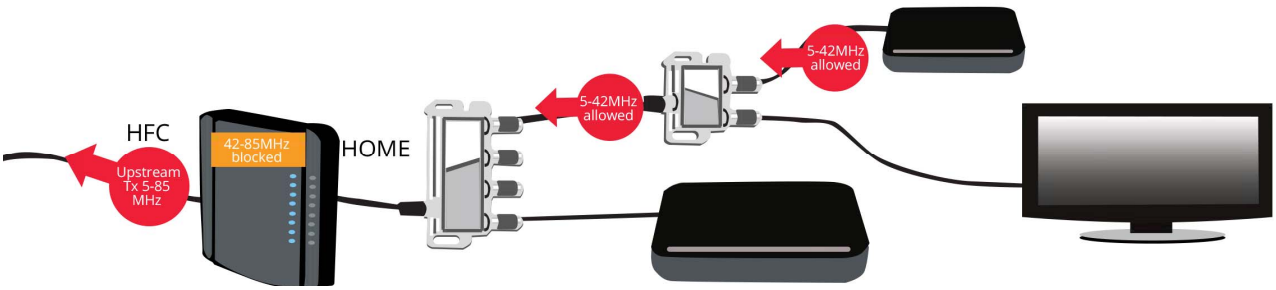


Figure 4 - Two Port Gateway Solution

One port addresses the hybrid fiber coax (HFC) network and the other port addresses the home network. Using this device eliminates the possibility of interference in the home by directing all 85MHz upstream transmissions directly onto the HFC network. Upstream transmissions in the legacy US band from within the home must pass through the CM from the home port to the HFC port.

Alternatively, if a traditional single port CM configuration is used, then filters that block signals between 42MHz and 85MHz should be added to each set top box or other cable receiving device or alternatively to a plitter port leg to protect multiple devices.

Note that a simple high pass filter is probably not sufficient, since many set top boxes have internal DOCSIS modems or other OOB means to transmit upstream. A simple high pass filter would change those two-way devices into one-way devices. A bandpass filter can be used to just block the offending transmissions between 42MHz and 85MHz while still allowing upstream transmission in the original upstream band.
Alternatively, a new splitter with internal filtering could be placed at the top of the home network, similar to common practice with EMTAs today as shown in the figure below.

The new splitter could have a port with additional filtering appropriate for an 85MHz CM. The additional filtering would prevent reflections in the 42-85MHz band from reaching the other ports used for the home.

Preventing Issues in the Local Access Network

A more challenging situation is present if Operations determines some tap ports do not have limited port-to-port isolation and will tend to pass 85MHz energy from one port to another. One option is to direct installers to place bandpass filters on other ports of a tap potentially affected when the new modem is installed. A potential drawback to this approach is that if a neighbor attempts a self

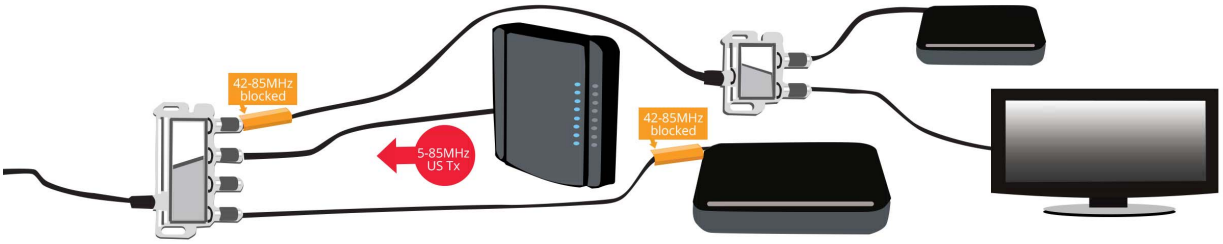


Figure 5 - Filter Placement to Prevent 85MHz Feedback

install of an 85MHz modem at a later time, it will fail unless the filter in the tap port is removed.

Another possibility is to replace the tap faceplate or even the entire tap as a part of the 85MHz plant upgrade. This solution makes the most sense if the tap is to be replaced anyway to enable 1.2GHz downstream service. Some MSOs are considering expanding their downstream services above 1GHz. The current taps in the field generally do not provide good performance in this band. If the tap is to be replaced anyway, also upgrading the isolation to prevent 85MHz crosstalk would be an ideal combination.

continue using the 54-85MHz frequency band for downstream services, but ensure that their new modems will be future proof against a later 85MHz plant upgrade.

Unfortunately, problems can be caused by an 85MHz modem even when it operates below 42MHz if it does not have a built-in switchable filter. Testing has shown that at least some DACs used for CM transmitters have a noise floor that pops up during transmissions. The levels seen were fairly low, for example -16dBmV. Because this noise would have to pass through the same port-to-port isolation loss discussed earlier that is seen, it is unlikely to affect digital

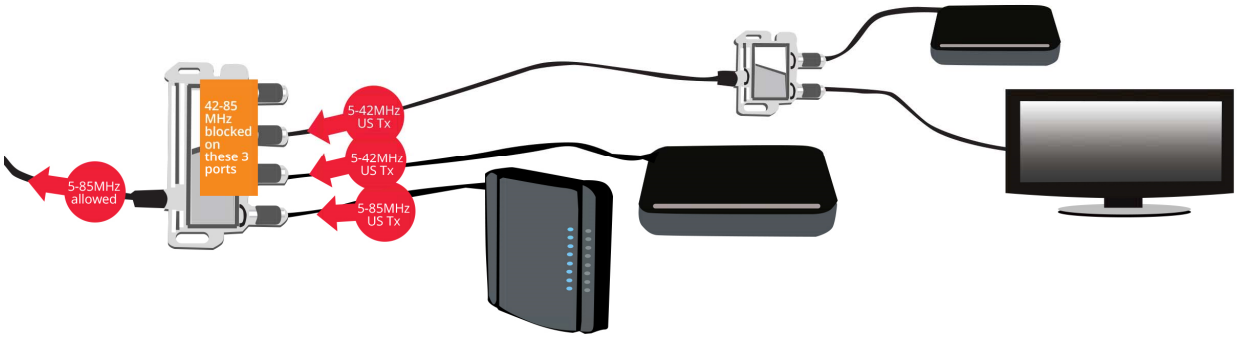


Figure 5 - Splitter with Embedded Filters to Prevent 85MHz Feedback

Advance Deployment of 85MHz Modems before Plant Upgrades

Some MSOs are considering deploying 85MHz modems now in advance of later 85MHz plant upgrades. They would like to

carriers, but it may cause distortion to analog video signals in the 54-85MHz band. Analog video signals can begin to show distortion with SNR of less than 40dB. If the incoming video signal is at 0dBmV, then the reflected jump in the noise floor could be around -

36dBmV. For lower video levels, the likelihood of problems would increase.
To avoid this problem, deploying a device with a built-in switchable filter is preferable.

Conclusion

Consumer demand for increased downstream bandwidth is also driving increased demand for upstream bandwidth. The latest DOCSIS standards allow or require support for upstream operation above the current operating 5 to 42MHz or 5 to 65MHz upstream bands. Successful expansion of the upstream band requires that actions be taken to prevent problems with legacy equipment already deployed in the field.
A two-port cable modem or gateway provides the most operationally friendly option, but other options can also provide acceptable performance. Improving the port-to-port loss in splitters in the home as well as within taps supporting homes with new 85MHz modems can prevent problems from showing up after the new modems are deployed.

PNM APPROACHES FOR IN-PREMISE WIFI NETWORKS – APPLYING LESSONS LEARNED FROM DOCSIS PROACTIVE MAINTENANCE PRACTICES

Scott Helms
ZCorum

Abstract

By leveraging the existing capabilities in wireless equipment, providers can develop processes to determine an optimal channel layout for Wi-Fi devices and help elevate the home Wi-Fi connection. The results can mitigate many of the problems associated with management of a Wi-Fi network and provide a better experience for subscribers.

This paper provides insight into Wi-Fi diagnostics for proactively detecting concerns as well as rectifying them. Proactive diagnostics for both Carrier Wi-Fi and standard in home service are covered. In addition, measuring and improving install quality for preventative management will be covered in this paper.

OVERVIEW

The adoption of PNM (Proactive Network Maintenance) for DOCSIS systems has increased over the last several years with many operators reporting success at getting ahead of plant problems. This has been done by mitigating and then correcting problems before the customer can detect them. This increases customer satisfaction and reduces churn. The success of PNM we see in DOCSIS cable plants leads us to examine the other parts of the service offering that could benefit from similar approaches.

WiFi networks, especially the in premise networks, are an area that needs attention and is growing very rapidly. There are estimated to be more than 7 billion **new** WiFi devices between now and the end of 2017 according to Sys-Con. (Afshar) Privately we've been told by operators that anywhere from 20-40% of all inbound tech support calls directly relate to the in home LAN and the vast majority of those are WiFi network issues. The expansion of CableWiFI as part of the strategy for large MSOs is also driving the expansion of WiFi as part of the core service offering and adding new challenges for delivering good performance. As clients with low RSSI connect to an access point the service degrades somewhat for the other clients, even if their RSSI is high. CableWifi clients often place that kind of strain on in home gateways as customers access those access points from outside the customer premise.

PNM precepts:

1. Collect data from network elements. In this case we can only collect information from the gateways and not usually the normal WiFi clients.

2. Automatically discern problems and where possible apply measures to mitigate the problem.

3. Provide reporting so that operational staff can prioritize and response to issues. In some cases the mitigation already applied will be sufficient.

4. Work with a wide range of currently deployed and new equipment.

COLLECTION OF DATA AND PROBLEM DETECTION

There are several methods of collecting data from gateways and these will vary by manufacturer and sometimes model.

1. TR-069 via TR-181
2. SNMP via CLAB-WIFI-MIB
3. SNMP via vendor specific MIB
4. Syslog parsing

For our field testing the devices selected only supported a vendor specific MIB for WiFi diagnostics. Our current testing is ongoing and currently covers a few hundred customers. We are currently gathering data every 15 minutes and storing the information in a database for offline analysis. We focused purely on the 2.4 GHz band for WiFi because we only had a few devices using 5 GHz bands and none showed any errors or contention.

We focused on gathering:

1. SSID name and current state of operations
2. Firmware version on the gateway
3. Currently used WiFi Channel
4. Which modes of 802.11 were being used
5. The output power of the AP (percentage and in DBM)
6. How often the gateway scanned to see if it should change to another channel.
7. The number of times the wireless AP itself had reset
8. Retransmit counters
9. Interference detection
10. Transmit Error Counters:

- no association error count
- RTS and CTS failures
- non data frames detected and FCS (CRC) errors

For the client focused stats:

1. Count, MAC, and status of all clients
2. Current time each client has stayed connected (Authenticated and Associated)
3. Received and transmitted packets per client
4. AP to client transmission failure counts
5. Up and down data rates from the AP to each client
6. The RSSI for each client

We have found many of the problems we expected to. For example, low RSSI clients have poor performance and nearly half of the clients we examined had periods of very poor RSSI and transmission rates. (See Figure 1, 2)

Errors were detected on at least one of the 2.4 GHz WiFi channels by 57% of the gateways and some of the most commonly used channels were the most problematic. Channel 11 for example showed interference on 42.86% of the devices tracked. (See Figure 3)

We also discovered some things that surprised us. About 12% of the gateways in our test group found interference on the channel they were using, but did not change that channel automatically as they had been configured to do. This occurred even though the devices could see other channels that were not impinged. This could have been because the device was actively sending or receiving data when it looked at its interference state, the user had hard coded the channel in the web interface, or for some other reason.

We saw tremendous swings in RSSI and data rates that were apparently unrelated to interference on the channel. These appeared to be the result of customers moving around their home and hitting spots of no or greatly reduced signal.

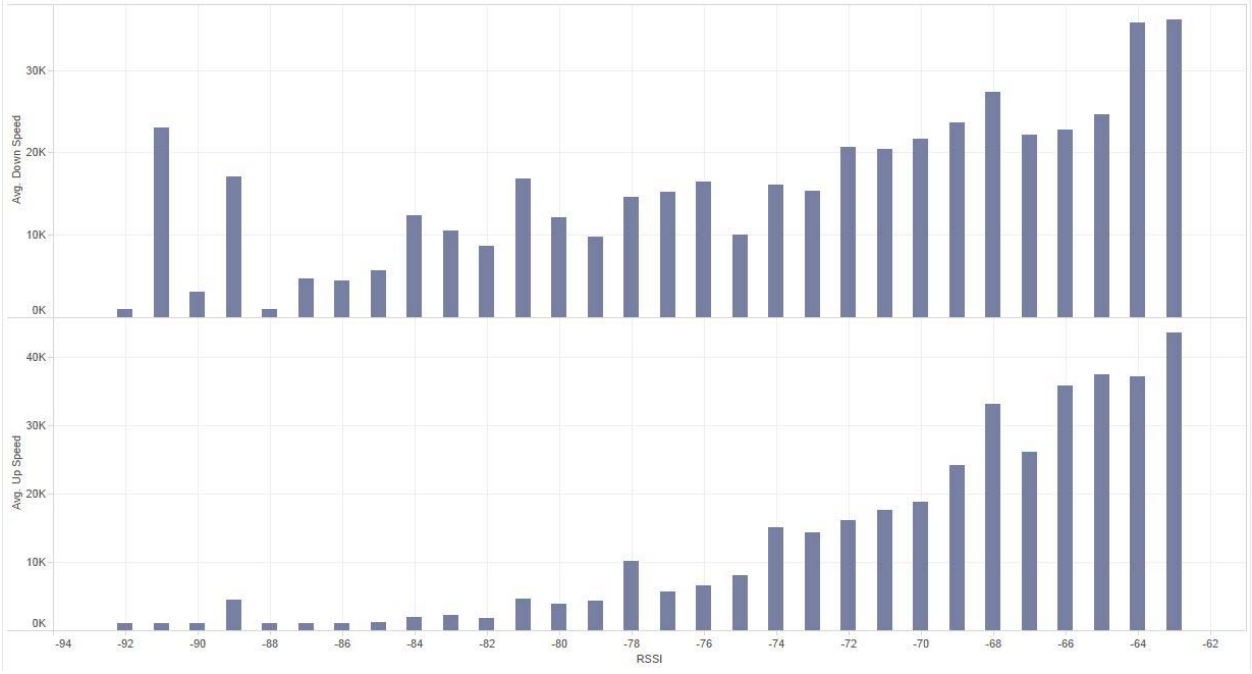


Figure 1 Graph of Low RSSI

Avg RSSI	Min RSS	Max RSSI	Std Dev
-61.94	-92	-1	14.9596

Figure 2 RSSI

Channel	1	2	3	4	5	6	7	8	9	10	11
% of AP's Inteference	17.05%	17.05%	23.04%	13.82%	17.51%	25.81%	31.34%	29.03%	27.65%	34.56%	42.86%

Figure 3 Common channel with Interference

AUTOMATED PROBLEM MITIGATION OR RESOLUTION

The next step was to proactively correct or mitigate the issues detected. Finding the devices that see errors on specific channels and then instructing those devices to move to new channels is straightforward and in most cases this resolved the problems we saw. We have a few cases where it did not and those are still under investigation.

Automatically correcting for poor RSSI clients by changing channels was effective in a few (~8%) cases, but will generally require moving the gateway to a more central location in the home or installing an additional WiFi AP. One additional item that we were able to detect and correct was a version firmware version that correlated to higher error counts.

REPORTING OF ISSUES FOR CORRECTION

Reporting on these problems is relatively straightforward, but the most impactful reports we found were looking for the top 10% of gateways with high error counts and/or retransmit counters. (See Figure 4)

The graph in Figure 5 displays tracking of individual client RSSI to determine if additional access points need to be installed. Tracking each client RSSI should be done both for troubleshooting and to see if we need to reposition the access point or consider installing a new secondary access point to ensure good service. It's critical to track RSSI both at the time of install, based on the installer's test client, and over time.

MAC	Time	WiFi TX Retransmit	No Association Errors	BAD FCS Count	RX Bad PLCP Count
xxxxxxxxxxxx	2014-07-19 17:02:07	26025	905	449583	526385
xxxxxxxxxxxx	2014-07-19 17:02:16	52	110	421393	1260174
xxxxxxxxxxxx	2014-07-19 17:02:17	0	60	18887576	67110359
xxxxxxxxxxxx	2014-07-19 17:02:19	0	50	74168	191856
xxxxxxxxxxxx	2014-07-19 17:02:23	1964	49	582823	2432109
xxxxxxxxxxxx	2014-07-19 17:02:23	1398	141	478979	652086
xxxxxxxxxxxx	2014-07-19 17:02:24	176570	940	357194	237959
xxxxxxxxxxxx	2014-07-19 17:02:24	0	1	0	0
xxxxxxxxxxxx	2014-07-19 17:05:54	78373	1855	2701035	1287425
xxxxxxxxxxxx	2014-07-19 17:05:54	0	0	0	0
xxxxxxxxxxxx	2014-07-19 17:05:56	41890	191	41479	47262
xxxxxxxxxxxx	2014-07-19 17:05:56	20809	349	883291	1012911
xxxxxxxxxxxx	2014-07-19 17:05:57	37813	900	5113959	1571170
xxxxxxxxxxxx	2014-07-19 17:05:57	364	858	10877	390711
xxxxxxxxxxxx	2014-07-19 17:05:59	148109	2619	346791	136623
xxxxxxxxxxxx	2014-07-19 17:06:00	0	31	179	923
xxxxxxxxxxxx	2014-07-19 17:06:01	0	90	5624	44563
xxxxxxxxxxxx	2014-07-19 17:06:02	72216	657	167477	48416
xxxxxxxxxxxx	2014-07-19 17:06:03	4575	517	283392	346111
xxxxxxxxxxxx	2014-07-19 17:06:04	0	30	122672	1315494
xxxxxxxxxxxx	2014-07-19 17:06:04	0	120	162792	2245441
xxxxxxxxxxxx	2014-07-19 17:06:04	0	90	2393908	5041171
xxxxxxxxxxxx	2014-07-19 17:06:05	0	90	621575	1473963
xxxxxxxxxxxx	2014-07-19 17:06:05	118327	2577	1348379	3063571

Figure 4 Error Report for Top 10% of Gateways

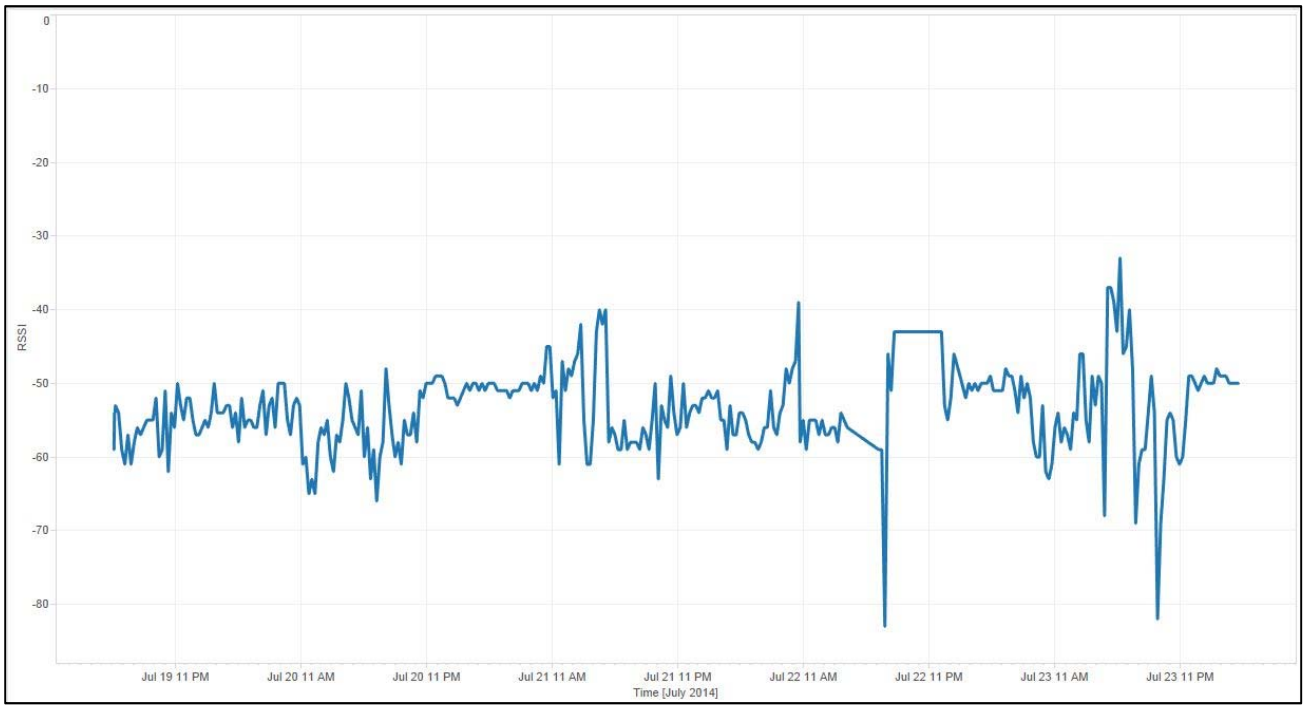


Figure 5 Graph of Single Client RSSI

BIBLIOGRAPHY

Afshar, Vala. "The BLOG." 12 February 2014. A. Huff Post TECH. 2 July 2014. <http://www.huffingtonpost.com/vala-afshar/50-incredible-wifi-tech-s_b_4775837.html>

PRACTICAL METHODS TO VALIDATE ULTRA HD 4K CONTENT

Sean McCarthy, Ph.D.
ARRIS

Abstract

Ultra High Definition television is many things: more pixels, more color, more contrast, and higher frame rates. Of these parameters, "more pixels" is much more mature commercially and Ultra HD 4k TVs are taking their place in peoples' homes. Yet, Ultra HD content and service offerings are playing catch-up. We don't yet have enough experience to know what good Ultra HD 4k content is nor do we know how much bandwidth to allocate to deliver great Ultra HD experiences to consumers. In this paper, we describe techniques and tools that could be used to validate the quality of uncompressed and compressed Ultra HD 4k content so that we can plan bandwidth resources with confidence. We will describe the statistical methods we use to validate Ultra HD 4k content, and will present some of our results. We will also explore the impact of high-efficiency video coding (HEVC) compression on the statistics of Ultra HD 4k content. The data and analysis we present are intended to provide tools and data that could be used to optimize bandwidth allocation and design Ultra HD 4k service offerings.

INTRODUCTION

Only a decade ago, high definition HD was the big new thing. With it came new wider 16:9 aspect ratio flat screen TVs that made the living room stylish in a way that old CRTs couldn't match. Consumers delighted in the new better television experience. Studios, programmers, cable, telco, and satellite video providers delivered a new golden-age of television. HD is now table stakes most places, and where that is not yet the case, it will be soon enough.

Yet now, before we hardly got used to HD, we are talking about Ultra HD (UHD) with at least four times as many pixels as HD. In addition to and along with UHD, we are getting a brand new wave of television viewing options. The Internet has become a rival of legacy managed television distribution pipes. Over-the-top (OTT) bandwidth is now often large enough to support 4k UHD exploration. New compression technologies such as HEVC are now available to make better use of video distribution channels. And the television itself is no longer confined to the home. Every tablet, notebook, PC, and smartphone now has a part time job as a TV screen; and more and more of those evolved-from-computer TVs have pixel density and resolution to rival dedicated TV displays.

Is all that resolution going to make a difference to consumers? If yes, what bandwidth will 4k UHD programming need? Those are two big questions our industry is exploring with respect to planning UHD services; yet they are not independent questions.

4k UHD is still new enough in the studios and post-production houses that 4k-capable cameras, lenses, image sensors, and downstream processing are still being optimized. Can we be sure yet that the optics and post processing are preserving every bit of "4k" detail? On the distribution side, could video compression and multi-bitrate adaptive streaming protocols change the amount of visual detail to an extent that it could conceivably turn "4k" quality into something more like "HD" or even less?

If the 4k content we have available today for bandwidth and video quality testing does not truly have a "4k"-level of detail, then we could go astray and plan for less bandwidth

than we might need for future 4k UHD services. If the 4k content we have available today is truly "4k", then we should also want to be sure that we do not over compress and turn 4k UHD into something less impressive.

Indeed, during our UHD 4k testing, we have found several candidate test sequences that appeared normal to the eye but turned out to have unusual properties when examined mathematically. Such content could lead to wrong conclusions when planning for UHD 4k bandwidth and services.

In this paper, we present mathematical techniques to help answer the question "How 4k is it?" Our method examines 4k UHD video to see if it has a statistically expectable distribution of spatial detail as a function of 2-dimensional spatial frequency. The benchmark for our statistical expectations is drawn from numerous studies of the statistics of natural scenes.

Our main objective in writing this paper is to describe methodology that might be useful in helping to decide which 4k UHD content should be included in the video test library intended to be used for bandwidth and video quality planning.

SOURCES OF 4K UHD CONTENT

There are many online places from which to obtain 4k content that could be considered for testing purposes. Industry-focused sources include the European Broadcasting Union[1,2] (EBU), CableLabs[3], and blenderfoundation[4]. Stock footage typically intended for promotional projects, but which might also be considered for testing purposes, is available online sites such as Shutterstock[5], NYC B.Roll[6], NatureFootage[7], and others that can be found by searching keywords such as "4k stock." Video-sharing sites such as YouTube[8] and Vimeo[9] host compressed 4k UHD content that could be candidates for testing certain kinds of 4k UHD services.

CAMERA CONSIDERATIONS

The quality of 4k content depends on the quality of the camera, the particulars of the post processing such as filtering and compression, and the skill of the camera operator and crew.

4k-capable cameras available today range from consumer camcorders to cream-of-the-crop professional 4k-cameras that are used to create premium cinema and television content. Even some smartphones boast 4k cameras.

Lens quality and image sensor size are key issues in 4k capture. Obviously, consumer and prosumer grade cameras should not be expected to have the top-quality lenses and image sensors found in high-end professional cameras. Yet, even in high-end cameras one needs to consider the interaction between the lens and image sensor. At this point in time, the image sensors found in many 4k cameras are larger than HD image sensors. As a result, depth-of-field tends to be shallower. Background and foreground details that are out of the plane of focus can be softer than they are in HD. Depth-of-field can be increased by decreasing the aperture, but at the expense on less light which can result in noisier video because of sensor noise. Longer exposure times could improve the amount of light captured, but then motion blur could become an issue. All of these opto-electrical items are capable of producing 4k content that has less spatial detail in the subject matter and more noise than would otherwise be expected. More important, such content could lead to wrong conclusions about the amount of bandwidth that will be needed to deliver great 4k experiences to consumers.

COMPRESSION CONSIDERATIONS

Video compression works mainly by strategically reducing the amount of spatial detail in video. Each compressed video frame

is predicted from previously stored frames as much as possible. Whatever is unpredictable is packaged as a residual signal and sent to decoders, but not before the residual is further refined by being converted into a signal having less numerical precision through a technique called quantization. In the MPEG family of compression standards (MPEG-2, AVC/H.264, and HEVC/H.265), quantization has the effect of preferentially reducing the spatial details that are associated with higher spatial frequencies. In addition, AVC and HEVC employ spatial blurring filters that reduce the noticeability of spatial discontinuities between coding blocks. Both frequency-sensitive quantization and spatial blurring can reduce the kind of fine detail that 4k aims to show off. Without that "4k" detail in our test content our bandwidth predictions could be off when the next even-better generation of 4k camera arrives.

ADAPTIVE STREAMING CONSIDERATIONS

Over-the-top streaming services have taken the lead in delivering 4k content to consumers. Such services typically employ one of several adaptive streaming protocols that enable video to play smoothly even when the consumer's bandwidth fluctuates significantly. Each adaptive-streaming video player senses its available bandwidth and requests a segment of programming that fits within its capabilities. If enough bandwidth is available, a 4k-capable video player would select lightly-compressed high bitrate segments having full 4k resolution (3840x2160). More restricted bandwidth could force selection of more aggressively compressed versions of the content though still at full 4k resolution. Even more restricted bandwidth can force selection of aggressively compressed sub-4k resolution (for example, 1920x1080 or 960x540 etc.). The modulation of the resolution and the compression level mean that 4k adaptive-streaming services might not always be fully 4k. Thus, any test content derived from such a source could impact how test results should be interpreted.

SPATIAL FREQUENCY

An image is normally thought of as a 2-dimensional array of pixels with each pixel being represented by red, green, and blue values (RGB) or luma and 2 chrominance channels (for example, YUV or YCbCr). An image can also be represented as a 2-dimensional array of spatial-frequency components as illustrated in Figure 1. The visual pixel-based image and the spatial-frequency representation of the visual image are interchangeable mathematically. They have identical information, just organized differently.

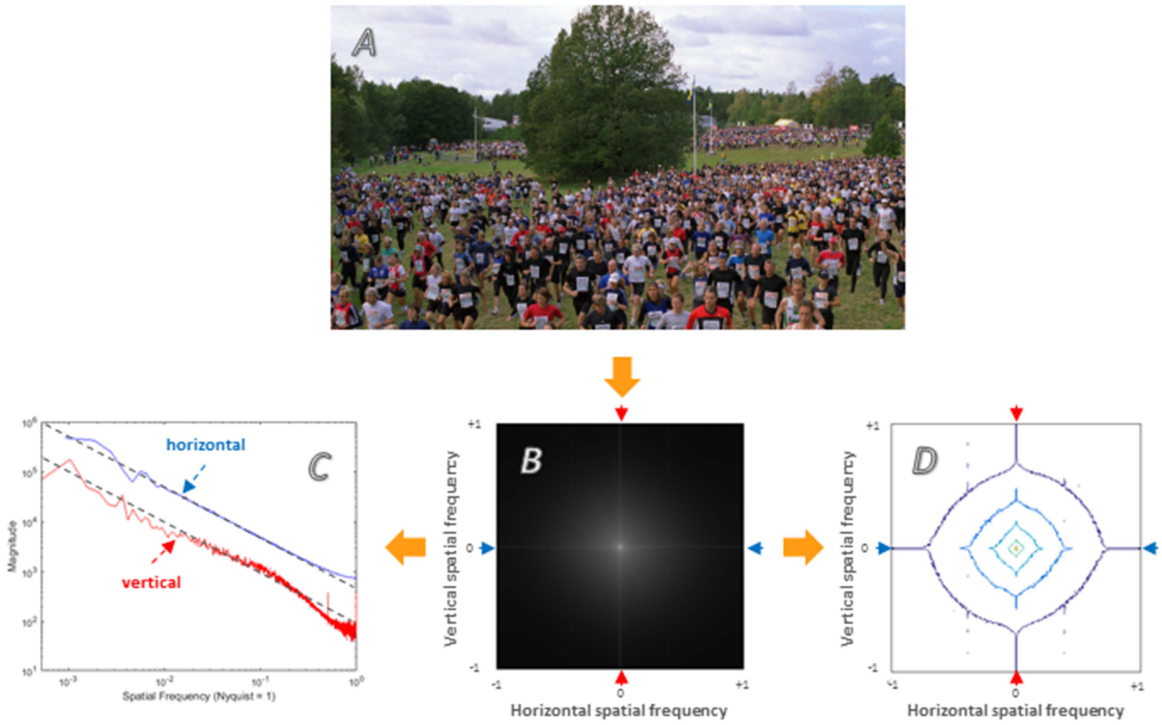


Figure 1. Illustration the representation of an image in terms of spatial frequencies. The visual pixel-based image (***A***) can be represented as a 2-dimensional array of complex numbers using Fourier transform techniques. The absolute-value of the complex numbers is shown as a 2-dimensional magnitude spectrum (***B***) in which brighter areas correspond to larger magnitude values. (Note that the log of the magnitude spectrum is shown in ***B*** to aid visualization. The horizontal and vertical frequency axes are shown relative to the corresponding Nyquist frequency (±1).) The magnitudes of the main horizontal and vertical spatial frequency axes are shown in ***C*.** The main horizontal spatial frequency axis corresponds to zero vertical frequency (blue arrows in ***B***), and the main vertical spatial frequency axis corresponds to zero horizontal spatial frequency (red arrows in ***B***). (Note that the magnitude spectrum is mirror symmetric around the 0,0 point (center of ***B*** & ***D***) along the main horizontal and vertical axes. Thus only the values from 0 to 1 (Nyquist frequency) are shown in ***C***.) The dashed lines in **C** indicate the 1/spatial frequency statistical expectation for natural scenes (the 1/spatial frequency appears as a line in the log plot). A contour map of the log of the magnitude spectrum is shown in ***D***. Contour maps provide useful *gestalts* of the overall 2D magnitude spectrum. The data shown in ***B***, ***C***, and ***D*** were obtained by averaging the magnitude spectrum of individual frames over 250 frames (5 seconds) of the SVT CrowdRun 2160p50 test sequence. (All video processing and analysis discussed in this were performed using MATLAB[10] and ffmpeg[11].)

Spatial-frequency representations of images can further be represented by a magnitude component and a phase component. The magnitude component, called the magnitude spectrum, provides information on how much of the overall variation within the visual (pixel-based) image can be attributed to a particular spatial frequency. (Spatial frequency is 2-dimensional having horizontal and vertical parts.) The phase component, called the phase spectrum (not shown), provides

information on how the various spatial frequencies interact to create the features and details we recognize in images.

For the purposes of this study, we find it useful to use contour maps of the log of the magnitude spectra (as shown in Figure 1.) to create a *gestalt,* the 2-dimensional spatial frequency composition of images.

STATISTICS OF NATURAL SCENES

Images of natural scenes have an interesting statistical property: They have spatial-frequency magnitude spectra that tend to fall off with increasing spatial frequency in proportion to the inverse of spatial frequency[12]. The magnitude spectra of individual images can vary significantly, but as an ensemble-average statistical expectation, it can be said that "the magnitude spectra of images of natural scenes fall off as one-over-spatial-frequency." This statement applies to both horizontal and vertical spatial frequencies. Examples of images adhering to this statistical expectation are shown in Figure 2.

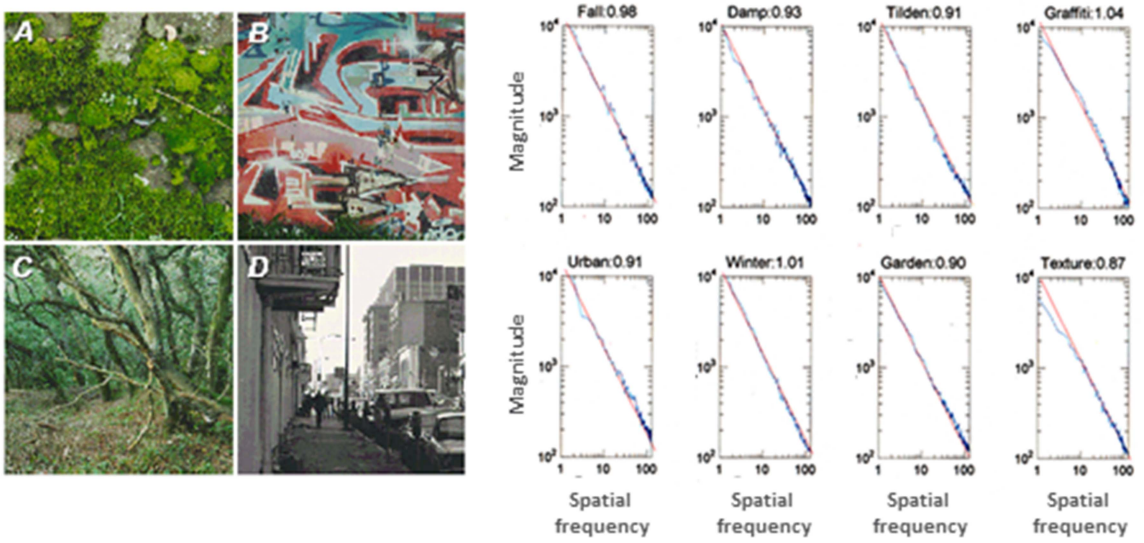


Figure 2. Examples of adherence to the 1/f statistical expectation. These 8 image series created and analyzed by McCarthy & Owen[13] illustrate the well-established statistical expectation that the magnitude spectra of images of natural scenes tend to be inversely proportional to spatial frequency. (Note that inverse proportionality appears as a straight line in the log plots shown.) The images shown in ***A***, ***B***, ***C***, and ***D*** are representatives of the "Texture" (close-ups of grass, sand, etc.), "Graffiti" (urban art), "Tilden Park" (woodland scenes), and "Urban" (San Francisco street scenes) image series, respectively. Other image series are "Fall" (colorful Fall foliage), "Damp" (puddles and environments where one might expect to find newts), "Winter" (snow scenes in New England), and "Garden" (UC Berkeley botanical garden).

Note that "natural-scene" images are not limited to pictures of grass and trees and the like. Any visually complex image of a 3-dimensional environment tends to have the one-over-frequency characteristic, though man-made environments tend to have stronger vertical and horizontal bias than unaltered landscape. The one-over-frequency characteristic can also be thought of as a signature of scale-invariance, which refers to the way in which small image details and large image details are distributed. Images of text and simple graphics do not tend to have one-over-frequency magnitude spectra.

A BENCHMARK FOR SPATIAL DETAIL

In this paper, we leverage the one-over-frequency statistical expectation to see if it holds for 4k UHD content we have considered for use in our lab tests. Examples of 4k (2160p50) video sequences that largely adhere to the one-over-frequency statistical expectation are shown in Figure 3. Examples of candidate UHD 4k (3840x2160) videos that violate the one-over-frequency statistical expectation are shown in Figure 4.

According to the results shown in Figures 3 & 4, the test sequences shown in Figure 3 remain viable candidates to be used in experiments to explore UHD bandwidth planning; but we would exclude the test sequences represented in Figure 4 from our UHD 4k test library.

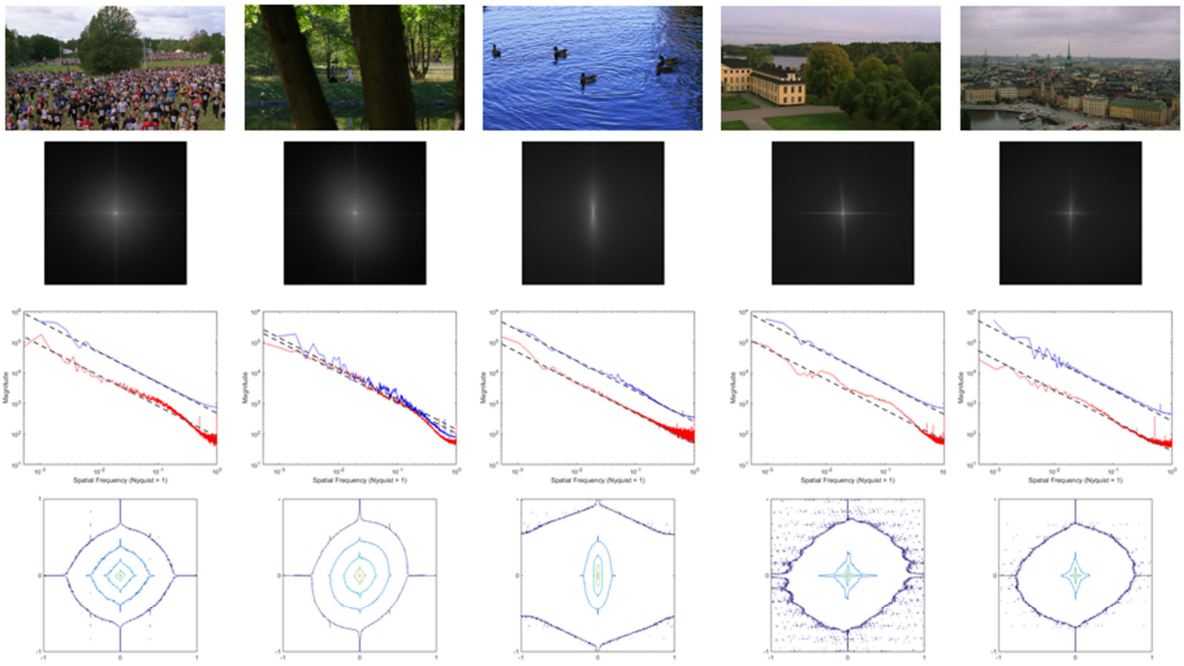


Figure 3. UHD 4k test sequences that have statically expectable magnitude spectra. The SVT UHD 4k test sequences (3840x2160 at 50 frames per second) "CrowdRun", "ParkJoy", "DucksTakeOff", "InToTrees", and "OldTownCross" are shown left to right in columns. The corresponding visual (pixel-based) image, log of the magnitude spectrum averaged over 250 frames (5 seconds), main horizontal and vertical axis components, and contour map of the log average magnitude spectrum are shown top to bottom in each column. Note that each of the sequences can be well-described by the one-over-frequency statistical expectation (dashed lines in the plot on the third row from the top); though there are some subtle deviations from statistical expectation (see Figure 5). Note also that the contour maps provide concise distinguishing information about each image sequence.

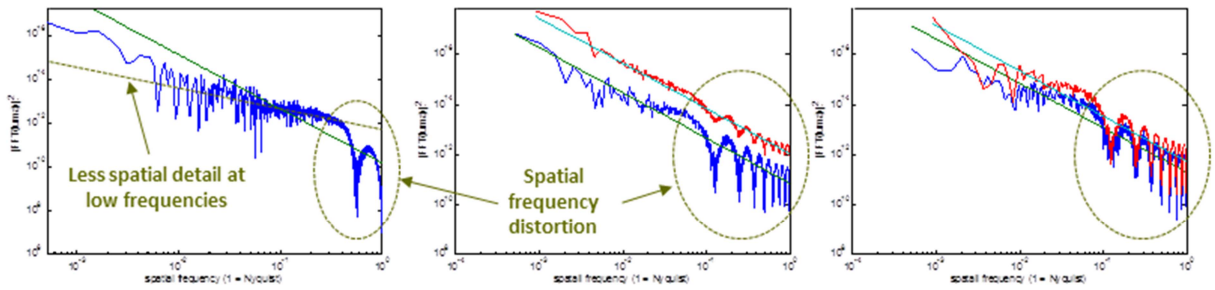


Figure 4. Example of candidate UHD test sequences that do not have statically expectable magnitude spectra. Some UHD 4k test sequences that we have obtained from various sources that appeared normal to the eye were found to have spatial magnitude spectra that were inconsistent with statistical expectations. Typical deviations from statistical expectations included: notch-like frequency distortions; excessive or diminished high or low frequency spatial detail (non-one-over-frequency behavior); and extraneous noise (see Figure 5).

The test sequences shown in Figure 3 are broadly in line with statistical expectations, however, they do show subtle deviations as illustrated in Figure 5. These deviations are mainly the presence of isolated narrow-band noise-like distortions and mild loss of high-frequency high spatial detail. The sole reason we present Figure 5 is to illustrate a method of scrutinizing candidate UHD 4k test content to an extent not possible with the eye alone. (It should be noted that the SVT UHD 4k test sequences shown in Figures 3 and 5 were produced 10 years ago, long before the emergence of UHD 4k as a consumer service, and thus were on the cutting edge of UHD 4k research and development.)

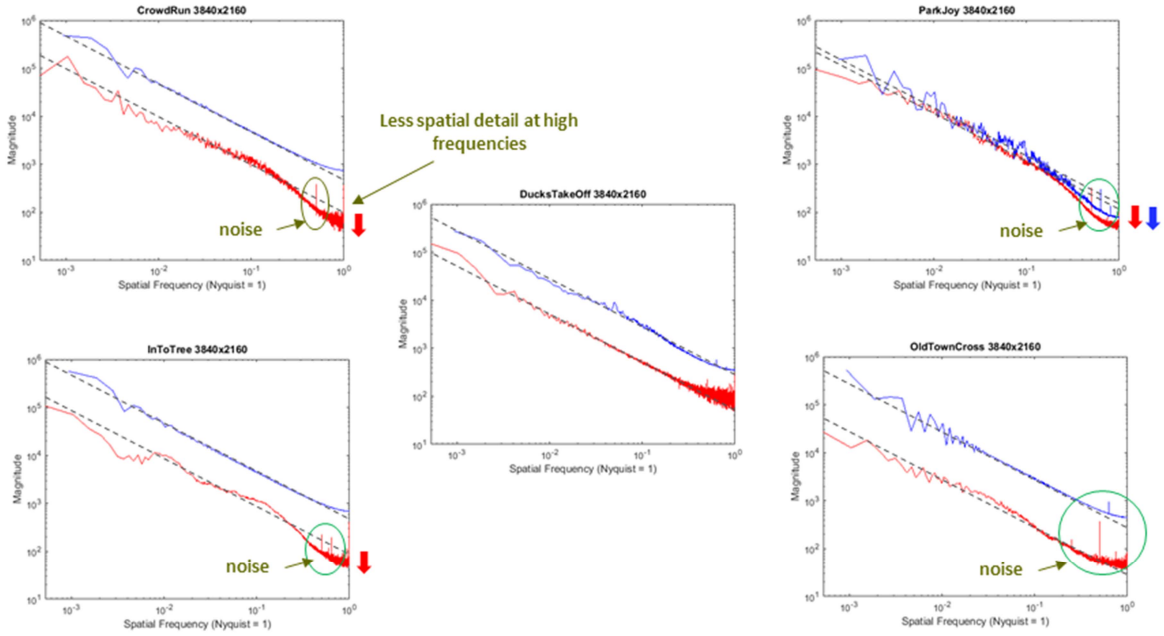


Figure 5. A closer scrutiny of subtle deviations from the statically expectable magnitude spectrum. Although most UHD 4k candidate test content matches the one-over-frequency statistical expectation in general, some sequence do show subtle deviations. As illustrates for the

SVT UHD 4k test sequences, these subtle deviations typically take to the form of extraneous noise that show up as isolated peaks and less-than-expected levels of high-frequency spatial detail (indicated by arrows pointing down). Note that the "DucksTakeOff" sequence meets statistical expectations particularly well.

EFFECTIVE RESOLUTION

A key feature of adaptive streaming protocols is the inclusion of reduced-resolution versions of content in order to provide uninterrupted video service even when a consumer's available bandwidth is significantly curtailed. Although compressed at resolution less than full 4k resolution, the content seen by a viewer is upconverted to 4k resolution by either a set top box or the television display itself. In this way, the effective resolution is less than the displayed resolution.

UHD 4k displays have such high resolution, and upconversion algorithms have become so good, that it is sometimes difficult to see by eye if a particular video is pristine full resolution or if some upconversion has occurred in the preparation of the content.

Figure 6 illustrates a method of analyzing the effective resolution of "4k" (3184x2160) resolution test content more quantitatively than can be done by eye. It is well-known that a reduced effective resolution correlates to a loss of high-frequency spatial detail. This loss could, in principle, be evident by inspecting the main horizontal and vertical axes of the magnitude spectrum. We find that is not always the case. Modern rescaling algorithms are very sophisticated and the difference between lowered effective resolution and full resolution can be subtle. Instead, we find that the contour maps of the log of the magnitude spectrum are a much more sensitive indicator of effective resolution. A test for accepting candidate UHD 4k test content into a master library could be along the lines of determining the average radius of the outermost contour, accepting test content only when a certain radius threshold is exceeded.

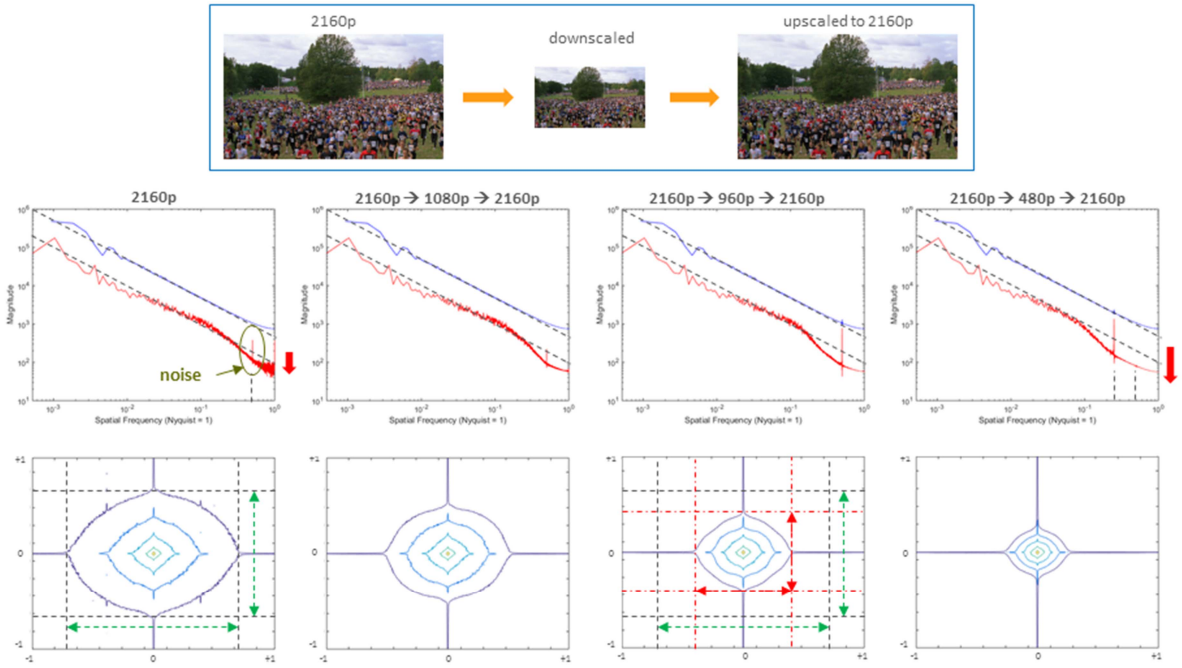


Figure 6. An example of using contour maps of magnitude spectra to examine effective spatial resolution. Some UHD 4k candidate test content could have lower effective resolution than full UHD 4k (3840x2160). We simulate such a situation by downscaling and then upscaling back to 3840x2160 resolution using ffmpeg. From left to right, the downscaled resolution is: unaltered 3840x2160; 1920x1080; 960x540; and 480x270 as an extremum. Note that examination of only the main horizontal and vertical axes of the magnitude spectrum (middle row) reveals some differences, most notably some reduction in high-frequency spatial detail and shift in the narrow-band noise; but these details are too subtle to make confident decisions, particularly when the original full resolution content is unavailable for comparison. The contour maps of the log of the average magnitude spectrum provide more clear-cut evidence. The contour levels are that same for all columns. Thus the constriction of the contours towards the center indicated that the magnitude spectrum narrows (loses high-frequency spatial detail) thus quantifying the reduced effective resolution. This is, of course, expected. The significance of this figure is that is illustrates that the contour map method can be a sensitive measure of effective resolution of candidate test video.

EFFECT OF HEVC COMPRESSION

Video compression changes the amount of spatial detail in video, but the extent to which spatial detail is lost depends of the content itself and the aggressiveness of compression; i.e. the target bitrate.

In Figure 7 we demonstrate that our method of evaluating test content provides a way of testing the effective resolution of HEVC compressed content. Note that our method indicates that effective resolution is more sensitive to compressin for some kinds of content compared to other kinds of content. As such, our contour-map method could be used to optimize the selection of compressed UHD 4k content for testing purposes in terms of both the intrinsic image characteristics of content and the impact of bit rate. In this way, our contour-map method could serve as a content-independent method of measuring effective resolution and thus selection of useable UHD 4k test content.

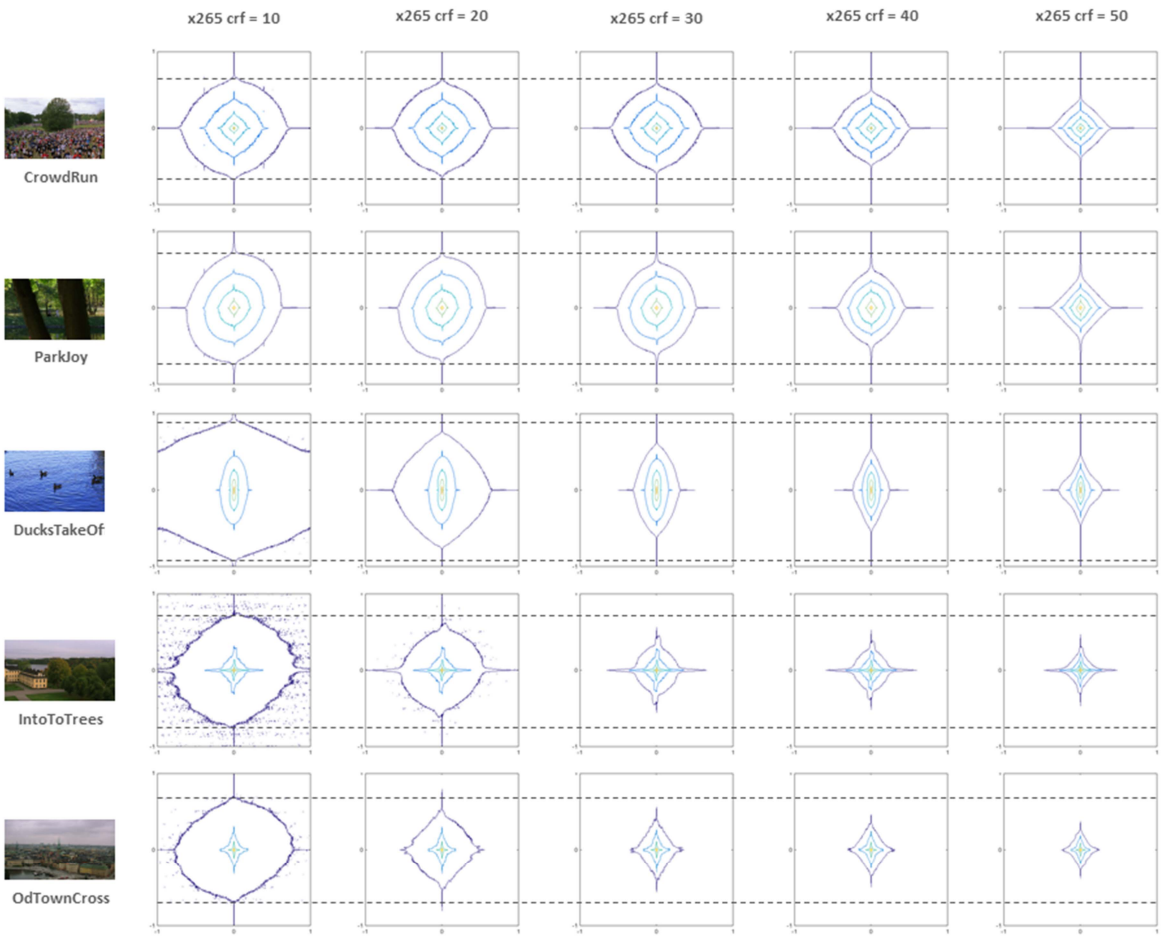


Figure 7. An example of using contour maps of magnitude spectra to examine the effect of video compression. Shown here are the contour maps of the average (250 frames, 5 seconds) of the log magnitude spectrum of each of the SVT UHD 4k test sequences compressed with HEVC to various extents. We used the libx265[14] library with ffmpeg to perform the HEVC compression. The ***crf*** value noted at the top of each column indicates the value of the constant rate factor (***crf***) parameter used in the ffmpeg libx265 command line. Smaller values of ***crf*** created more lightly compressed video. Video compressed with a ***crf*** value of 50 is typically very heavily artifacted. Video compressed with a ***crf*** value of 10 produces contour maps that are very similar to those for uncompressed video (see Figure 3). Note that the impact of the ***crf*** value is content dependent. For "CrowdRun" and "ParkJoy", the ***crf*** values below ~30 do not a have a major impact on effective resolution. On the other hand, a noticeable change in effective resolution is evident for a ***crf*** value of 20 for "IntoToTrees", "DucksTakeOff", and "OldTownCross". (The dashed lines provide a reference for the radial extent of outer contour of the lightly compressed and uncompressed versions of the video.)

DISCUSSION & CONCLUSIONS

The objective of this paper was to present techniques that might be useful in evaluating UHD 4k video sequences that could be candidates for testing related to planning UHD 4k products and services. Selection of test content that is not representative of anticipated UHD 4k programming – including future UHD 4k programming that will be available when the end-to-end UHD 4k ecosystem has been optimized – could lead to wrong conclusions about what bandwidth and level of video quality would be needed.

We show in this paper that ensemble-average statistical expectations related to spatial frequency magnitude spectra of images of natural scenes can be used as a benchmark of comparison to address the question: "How 4k is it?" We find that major deviations from statistical expectations can be considered grounds for excluding content from a test library. (Though perhaps some synthetic compression busters should be retained to stress compression equipment and distribution services.)

We also find that examination of contour maps of the log of the magnitude spectra are sensitive indicators of effective resolution. Contour maps that indicate a lack of 4k-level effective resolution can be considered grounds for excluding content from a test library. The main horizontal and vertical components of the magnitude spectrum seem to be good probes for detecting added noise and gross distortions; but they are not highly sensitive probes of effective resolution.

Significantly, our contour-map method is also a sensitive content-independent probe that can be used to evaluate compressed content for inclusion in UHD 4k video test libraries to be used for planning UHD 4k video quality and bandwidth.

REFERENCES

1) CableLabs 4k Resources. www.cablelabs.com/resources/4k/
2) EBU UHD-1 Test Sequences. https://tech.ebu.ch/testsequences/uhd-1
3) SVT High Definition Multi Format Test Set. tech.ebu.ch/webdav/site/tech/shared/hdtv/svt-multiformat-conditions-v10.pdf
4) blenderfoundation "Tears of Steel". mango.blender.org
5) ShutterStock. www.shutterstock.com
6) NYC B.Roll. www.nycbroll.com
7) NatureFootage. www.naturefootage.com
8) YouTube. www.youtube.com
9) Vimeo. www.vimeo.com
10) The Mathworks. www.mathworks.com
11) ffmpeg. www.ffmpeg.org
12) D.J. Field. "Relationship between the statistics of natural images and the response properties of cortical cells." J. Opt. Soc. Am. A. Vol. 4, No. 12 1987
13) S. McCarthy and W. G. Owen. Personal communication. University of California, Berkeley.
14) x265. www.x265.org

RDK + CA/DRM: ENHANCING INTEROPERABILITY FOR ROBUST REVENUE SECURITY

Petr Peterka
Verimatrix

Abstract

Open source platforms such as RDK offer many advantages, including the benefit of having a common implementation instead of a common specification; however, they are not without challenges, such as security threats and interoperability with proprietary components such as conditional access system (CAS) or digital rights management (DRM).

In this paper, the author explores how to optimize CA/DRM integration in the RDK bundle and ensure the revenue security of RDK applications.

The pay-TV industry relies on many moving parts to put an end-to-end television distribution system together. The set-top box (STB) on its own is an expensive custom-designed piece of consumer electronics. The Reference Design Kit (RDK) is trying to ease the cost burden of designing and integrating the STB by providing a common reusable middleware platform.

As a joint venture between several operators (Comcast Cable, Time Warner Cable, and Liberty Global), the RDK open source platform is well-poised for future growth and adoption. The platform currently has more than 200 licensees but will certainly add more in coming months as it continues to expand its capabilities. Although RDK was initially designed for the needs of North American operators, it is becoming allocable world-wide. Consider for example RDK Management's recent announcement that RDK now features support for the Digital Video Broadcasting (DVB) standard[1], which is widely used throughout Europe and other parts of the world.

UNDERSTANDING THE BENEFITS

RDK has become an increasingly popular pre-integrated software bundle as it offers many advantages, including the ability to support a common implementation instead of a common specification, which can accelerate the deployment of next-gen video products and services. According to RDK Central, using the bundle can also "enable TV service providers to standardize certain elements of these devices, but also to easily customize the applications and user experiences that ride on top."[2]

There are also cost benefits – RDK is a $0/royalty free commercial source code license – as well as the ability to enhance collaboration through full transparency into the source code. As an open source code initiative, the RDK enables community members to add new features and resolve issues on their own schedule and make contributions back to the community. Such an approach enables multichannel video programming distributors (MVPDs) to focus, innovate, and differentiate at the application and

[1] http://rdkcentral.com/rdk-management-expands-dvb-support-for-operators/
[2] www.redkcentral.com

services layer.

OUTLINING THE CHALLENGES

However, RDK is not without challenges, such as security threats and interoperability with proprietary components such as conditional access system (CAS) or digital rights management (DRM). Content protection in the form of a CAS client and DRM client must be supported by RDK in order to realize its benefit, however today's proprietary integrations with CAS clients are not scalable and create forked versions that are difficult to keep in sync with the main trunk RDK code. Once a forked version becomes stale and starts missing new features that have been developed after the forked version, the RDK benefits quickly diminish.

It is also worth noting that while RDK takes advantage of the open-source nature of RDK-M managed software development, the OSS license obligations need to be carefully managed when integrating with a proprietary implementation of a CAS or DRM client. Specifically, the abstract CAS APIs integrated into RDK must have an OSS license that allows CAS or DRM clients to be plugged in without putting OSS obligations on the CAS vendor that they could not comply with.

RDK AND REVENUE SECURITY

CA/DRM plays an important role in an RDK bundle. Typically, an MSO is providing a pay-TV service and needs to protect that service and the associated revenue. While, subscribers are mainly interested in premium content such as Disney, HBO or ESPN, content owners are focused on protecting their content from theft and illegal distribution. This is why MSOs have used CAS in their content distribution networks for decades.

Even though a CAS client is typically a very small component for an RDK bundle, it interacts with several other modules within RDK as well as the overall device such as a set top box (STB), DVR, HDMI dongle or a connected TV. There are several main interfaces:

- **Interaction with the content tuner/demultiplexer –** typical liner and on-demand television channel is encoded in MPEG-2 Transport Stream (TS) format. Systems-on-chips (SOCs) and middleware such as RDK provide tuning and de-multiplexing capability to optimize the channel tuning and channel changing experience. When a channel is encrypted by a CAS system, the TS carries additional messages such as EMMs and optionally EMMs. These need to be provided to the CAS client in order to properly decrypt the content. Newer systems utilize other formats such as MPEG-DASH, which has an equivalent mechanism to carry CAS or DRM specific data.

- **Interaction with a hardware root of trust –** most premium content requires high level of security robustness. It can be achieved through several methods but most often the system relies on unique hardware identity and hardware-based keys. Such functionality can be

exposed through a standardized key ladder or Trusted Execution Environment (TEE). Additionally, content is typically decrypted in hardware for performance and security reasons; therefore, additional access to these hardware resources is required.

- **Interaction with the application layer –** it is desirable to isolate the CAS functionality from the application layer as much as possible but some high-level APIs are needed. (This will be addressed in more detail later in the paper.)

- **Access to IP stack** - as many operators are migrating to all-IP architectures, IP-based CAS systems are more suitable than traditional 1-way systems. IPTV CAS utilizes the 2-way IP connectivity back to the MSO head-end in order to perform device authentication, entitlement and key retrieval at the time protected content is being accessed.

- **Forensic watermarking –** premium content such as UHD, requires additional level of protection in the form of forensic watermarking. Several methods of embedding the mark exist. Technologies embedding a mark into the decoded video require access to the decoder.

The CAS/DRM cannot be included in the open source as CAS vendors require a strict control over the CAS client implementation to guarantee integrity and authenticity. Furthermore, intellectual property included in the secure key management protocol and other cryptographic algorithms needs to be protected. If CAS client code was included in the open source, hackers could include "back doors" or gain too much knowledge about the system which could help them in defeating overall security.

Ideally, the CAS client should be isolated from the middleware and application functionality and minimally intrusive on the rest of the functionality. Therefore, abstracting out the main functions without exposing unnecessary details of the CAS system is highly desirable.

Furthermore, the MSO may want to retain the choice of with which CAS vendors they want to partner. A well-designed abstraction layer allows them to relatively easily upgrade one CAS client version with another or even replace one CAS technology for another.

MITIGATING THREATS TO REVENUE SECURITY

Isolating the CAS functionality also helps eliminate any negative impact on the RDK functionality, evolution and innovation while preserving the ability to quickly integrate a CAS system of choice without proprietary integration into RDK or requiring special forked versions of RDK. At the same time, any CAS client that will be integrated with RDK can be properly isolated through a set of abstract APIs whose OSS license allows a plug-in of a proprietary CAS client without contaminating it with undesirable OSS obligations.

Even though security and content protection shouldn't be exposed to the user in order to keep the user interface as simple as possible, there are some activities that need an interaction between the application and the CA client.

Tuning/Navigation

Tuning, channel change or on-demand asset selection are the main activities initiated by the user. When a protected channel or asset is selected, the proper entitlement and decryption must occur. When everything works well, the user shouldn't even see the security activity that is triggered in the background. The main steps shown on the diagram below (as application use case #1) are originated by the application (e.g. EPG, VOD library, PVR playback) and request the tuning (or more generally content selection) module which tunes to the indicated channel, starts parsing the associated metadata and determines whether this content needs any interaction with the CAS subsystem at all. Content metadata (e.g. MPEG PMT tables) may indicate presence of CAS or DRM information (e.g. ECMs, DASH PSSH). This is a point of interaction with the CAS client as this information needs to be passed in for CAS processing. Once the CAS function completes, assuming that the user is authorized to watch the selected content, the decryption module is properly armed.

Of course, behind the scene, the CAS client will use the TCP/IP stack to communicate with key management servers in the head-end to obtain content decryption keys, and with the SOC security functions to achieve desired level of robustness.

The scenario is very similar between tuning to a live channel, playing from a local or network DVR, or selecting a movie or TV episode from a VOD library.

Initialization/Provisioning

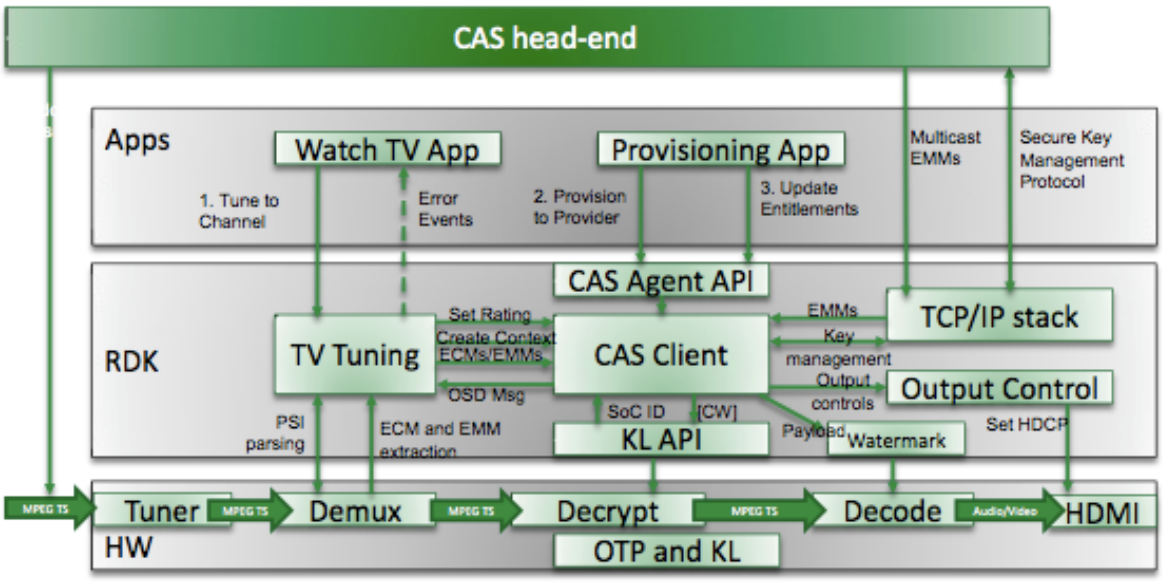
STBs or more generally user devices capable of consuming MSO-originated content may be leased by the operator following the typical traditional model, or could be retail devices such as connected TVs, retail STBs or dongles, etc. In the leased model, the STB and the CAS client may be preconfigured to connect and register to the operator with very little or no interaction with the user or the application. A retail device may be connected to one of several MSOs without the device manufacturer knowing the specifics. In this case, it may be the application that provides configuration information to the CAS client to provision to a specific MSO key management or license server.

This is shown on the diagram below as application use case #2.

Entitlements

CAS systems have different ways for obtaining entitlements. In traditional systems they are delivered using EMM messages that are demultiplexed from the MPEG 2 Transport Stream and passed to the CAS client. In IP-based systems, the CAS client contacts the head-end key management server and requests entitlements (and corresponding keys) over the 2-way IP channel. Most of this functionality can be hidden from the application but under certain

circumstances, the application may trigger an entitlement refresh action. The reason may be that the user subscribed to a new channel but the CAS client has not proactively requested the corresponding keys, yet. This is shown on the diagram below as application use case #3.



On-screen Display (OSD)

Even though CAS clients typically do not have a user interface, there are scenarios where it needs to provide some status to the user or even solicit input form the user.

For instance, when a user tunes to a channel that it is not entitled for, the decryption and eventual rendering of the content will fail. An application should know the reason so that it can provide adequate information to the user, show instructions how to sign up for that channel or tune to an alternative channel.

Another example is based on events that may be controlled by the user such as parental control or pay-per-view purchases. If these events require a user action, such as providing a security PIN code, the CAS client must have an ability to communicate with the right application.

When a specific channel or on-demand content is forensically watermarked, the use may need to be notified. This is another instance where the CAS client needs to provide additional information to the user via the application layer.

THE POWER OF STANDARDIZATION

There are different methods to standardizing on RDK. One way is to create an OSS implementation and share it with all vendors who need that functionality. All users of OSS benefit from a single implementation centralize code management process and testing

responsibilities. Another, possibly more traditional method is to create a standards organization which publishes a specification based on contribution of interested parties.

There are several standards organizations that help the TV industry. The main international one is MPEG but there are several others at the national level, including SCTE, CEA, CableLabs, etc.

One area that has been historically covered by a proprietary integration between a SOC vendor and the CAS vendor is what is generically called hardware root of trust. This may include unique chip identifier, access to unique chip root keys, a proprietary key ladder which implements a complex key hierarchy that is specific to each CAS vendor, and actual content decryption by secure hardware. There has been progress made in the last several years to standardize this functionality in order to assist with CAS interoperability. Such approaches make client downloadability more feasible, provide options for running multiple CAS clients on the same device or to replace an old version of a CAS client with new one.

An example of such standards is ETSI OMS KLAD[3] and its North American extension SCTE 201[4]. This allows chip vendors to manufacture CAS-independent SOCs that can be used in devices that need the ability to download a CAS client either during the device manufacturing process or later when the device is installed in consumer's home. Such a solution may also require a Trust Authority (TA) that generates and manages the chip root keys and provides them to all CAS vendors who end up using those chips. Such companies exist today and provide this role to the industry today.

Another standard that is very beneficial to the CAS interoperability goal is DVB Simulcrypt[5]. When an operator runs their pay-TV distribution system in the Simulcrypt mode, two or more CAS systems can coexist on the network. Some devices integrate or download CAS from vendor A while other devices may integrate a solution from CAS vendor B. Such systems have been successfully deployed around the world.

These standards when supported by RDK will make integration with CAS clients much faster and almost plug-and-play.

When an in-the-field or over-the-air downloadability of a specific CAS client is required, the actual download mechanism and the run-time environment need to be standardized as well. This is a very complex and difficult task but good progress has been done by the industry especially by GlobalPlatform[6]. The most relevant aspect of the GP activities is related to Trusted Execution Environment (TEE). The TEE architecture splits the SOC execution into a trusted world and untrusted world. Security sensitive

[3]http://www.etsi.org/deliver/etsi_ts/103100_103199/103162/01.01.01_60/ts_103162v010101p.pdf

[4]http://www.scte.org/documents/pdf/Standards/ANSI_SCTE%20201%202013.pdf

[5]http://www.etsi.org/deliver/etsi_ts/103100_103199/103197/01.05.01_60/ts_103197v010501p.pdf

[6] https://www.globalplatform.org/mediaguidetee.asp

computation such as ECM processing, content decryption and content watermarking is performed by a trusted application (TA) running in the TEE while non-secure parts of the CAS client run in the rich execution environment (REE).

The REE part would primarily implement the mapping of the Abstract CAS API, which is used by the RDK internal modules to communicate with a CAS client, to the specific implementation of the CAS client running as a TA in the TEE. By securely downloading a new TA, the CAS functionality can be upgraded or even replaced by a different implementation. GlobalPlatform standardizes the APIs for communicating between the REE and the TEE as well as access to cryptographic functions inside the TEE to make portability of TAs between SOCs as transparent as possible.

TEE has its own security architecture including secure boot, TA authentication and decryption, as well as HW root of trust. Additionally, the TEE may provide access to the OMS key ladder discussed above.

LOOKING FORWARD

From what has been discussed so far, it is clear that a significant amount of progress has been made already, RDK with all its OSS components is stable and enjoys real deployments. Standards such as SCTE, GlobalPlatform, MPEG-DASH have released stable versions of their specifications. What is still missing are the abstract CAS APIs and HW security abstraction.

The HW security abstraction can be modelled after the SCTE OMS standard and expose HW functions such as “get chip identifier”, “host authentication”, “set descrambler” and “get random number”.

Tuner and demultiplexer modules of RDK will need APIs to “create a tuning session” such that multiple channels can be processed simultaneously. Within each context, CAS specific data, such as ECMs and EMMs, will be passed to the CAS client via “send CAS data” API, and finally the descrambling will be activated/deactivated by “start session” and “stop session”, respectively.

Application oriented APIs will include “CAS initialization” to complete proper provisioning into a service, “purchase asset” and “update entitlements” related to access VOD or linear channels, “set parental rating” to limit what can be watched, and “get message” for any information that needs to be communicated to the user.

The detailed object model of these APIs is certainly more complex that the outline above but several companies spent significant amount of time to put a concrete proposal together that is being evaluated by the RDK community.

It is not difficult to see how the three seemingly antagonistic worlds of open source software, international and national standardization, and proprietary content protection can come together to assist in rapid deployment of a complex ecosystem.

Realizing Packet/Optical Network Convergence and Efficient Router Interconnect through Packet-aware Optical Transport

Chris Liou
Infinera Corp., 140 Caspian Ct, Sunnyvale, CA 94089, USA.

Abstract

In conventional network architectures, where IP/MPLS operates as a client layer over an optical transport layer, routers are employed throughout the WAN to provide a multitude of packet-centric bandwidth management functions, including packet services termination, aggregation, switching, QoS and transport.

New emerging high-speed packet management functions at 100GbE rates are emerging within next-generation optical transport systems that combine WDM, OTN and packet functionality in a way that provides cable operators with more cost-effective tools for optimizing packet traffic, increasing network resource efficiency, and reducing the total cost of the combined network.

This paper compares and contrasts the current PMO with an architectural approach based on packet-aware OTN/optical transport, discusses typical use cases, and presents the benefits this provides operators.

INTRODUCTION

The growth in packet traffic resulting from IP-based services, various cloud initiatives, and the shear growth in data/video-centric services is creating new strains on the operator's infrastructure as the challenge of growing or maintaining ARPU continues. As new competitive forces gain more traction in the marketplace, additional pressure is placed on network operators to not only differentiate their product offerings to gain and retain customers, but also ensure that the network infrastructure employs the best technologies and solutions for enabling the right combination of scalability, reachability, survivability and economics.

This presents a challenge for the network architect who must not only understand the relevant networking technologies applicable to the MSO market, but also consider new advances in technologies when exploring the best way to achieve optimal networking economics. One important technology advance that is driving network upgrades is the introduction of 100Gb/s coherent optics and the resulting fiber capacity that it enables. The upgrade from 10Gb/s to 100Gb/s transmission is vital for supporting the growth in network traffic and is being broadly deployed today.

While the use of 100Gb/s Layer 0 WDM technologies is a broadly accepted strategy for fiber capacity upgrades, there are many options for building the next layers of networking. Two methods used today include building a router-based L2/L3 layer directly over 100Gb wavelengths and building a router-based layer over a converged OTN/WDM layer. In both of these models, topology amongst routers is generally defined using IP link sizes commensurate with the underlying "wavelength" service, which is either 100GbE in the case of IP over WDM or a mix of 10/40/100GbE in the case of IP over OTN/WDM. All packet-level functions (statistical multiplexing, aggregation, QoS traffic management, switching) is relegated to the router layer, and the optical transport layer is providing router interconnect via transparent pipes.

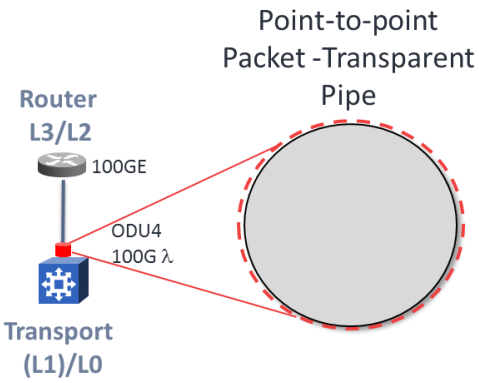


Fig 1 – Traditional approaches leverage optical transport layer for L0/L1 functions only, with all packet traffic processing performed solely by routers. The transport layer is oblivious to packets.

With new innovations emerging within the transport layer that integrates high-speed packet capabilities along with traditional L1/L0 functionality, however, new architectural options are available for building cost-efficient, scalable networks. The convergence of packet and optical transport capabilities into an integrated system, along with Transport SDN, is giving rise to a new set of tools that can help improve resource utilization across both layers, eliminating over-provisioning of router links and allowing the transport layer to cost-efficiently offload certain transport functions currently performed by routers.

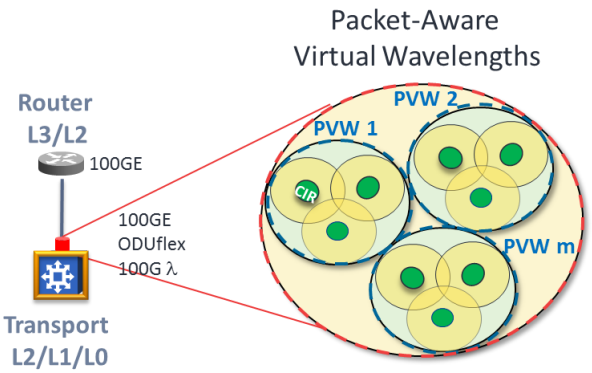


Fig 2 – Packet-aware transport employs PVWs to add packet processing, enabling more flexible packet-centric transport functionality that improves overall network cost and offloads routers from basic packet transport functions.

This paper discusses the architectural capabilities facilitated by a packet-aware transport layer that combines L2/L1/L0 functionality and compares this against current architectures. It introduces the concept of a Packet-aware Virtual Wavelength (PVW) and discusses how it plays a role in evolving the transport network, as well as the flexibility and savings it can bring to the combined IP/MPLS and optical transport network layers.

PMO CHALLENGES

The current approach for building MSO core networks generally involves the construction of a router layer that runs over an optical transport layer that is either based on Fixed or Reconfigurable OADM technologies for steering analog wavelengths between routers or on converged OTN/WDM technologies that digitally switch and groom router port traffic on to optical wavelengths.

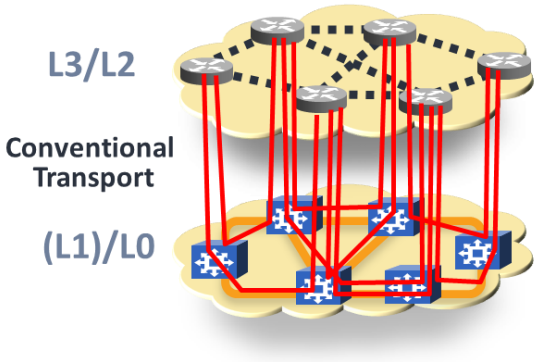


Fig 3 - Conventional architecture relies on routers to perform all packet functions. IP links and topology are rigidly coupled to wavelengths. This presents scale challenges when attempting to bypass transit routers and achieve fuller mesh connectivity between edge routers

The choice in optical transport technology is influenced by a number of factors beyond cost and fiber capacity – other considerations such as protection and restoration capabilities,

operational complexity, reliability, and speed of service turn-up play an important role.

The cost associated with delivering traffic across the network is predominantly influenced by the amount of bandwidth required at both the transport layer as well as the IP/MPLS layer. At scale, these costs are generally dominated by the total number of router ports required along with the total number of wavelengths required. The choice of network architecture to support a network's traffic demands involves determining what type of traffic to handle at which layer, identifying the appropriate service recovery architecture to protect against failures, and understanding where opportunities for "arbitrage" exist, i.e., determining the most appropriate layer to transport bits with an end objective of minimizing the cost/bit metric, while still meeting service SLA objectives.

Lack of Cross-Layer Coordination

Traditionally, networks are often planned, engineered, and deployed layer by layer. One organization generally determines the best solutions for the router layer to support all the L2/L3 packet-based services, and a separate organization generally determines the transport solution, focusing on L0/L1, using inputs from other organizations to determine overall traffic forecasts. Additionally, each organization may have strategies for protecting against network failures. This often gives rise to inefficiencies, such as in the form of over-provisioning, non-optimal placement of traffic flows, or non-optimized use of available resources at each of the network layers.

Inefficient Network Resource Utilization

One of the reasons for networking inefficiency can be attributed to where packets are managed and where certain networking features such as protection are accessed. Traditionally, all packet functions from service termination and adaptation to traffic management and aggregation and even fast protection against network failures is performed within the router. While this may provide operational convenience, it does not leverage the lower-cost capabilities of the optical transport layer and can lead to inefficient usage of router resources.

One common practice for reducing the costs at this layer consists of leveraging the transport layer to enable more router bypass and minimize intermediate packet hops, thereby reducing the transit router tax [1]. This approach, however, runs into challenges as the network migrates to 100G transmission, as direct connectivity to destination routers requires a dedicated port and dedicated wavelength service. Full mesh connectivity between edge or service routers using dedicated ports would eliminate the need for transit routers but is often impractical and economically non-viable beyond small networks. In large networks, optical bypass opportunities generally focus on network hot-spots where there is sufficient amount of persistent traffic to warrant express, while other sites interconnect via a set of transit core routers.

Over-provisioning due to FRR Protection

The reliance on routers to perform protection is another source of inefficiencies, as spare bandwidth in the router layer required for potential network failures also translates into spare bandwidth in the transport layer. Over-provisioning of the router layer by a factor of 2x to support router-based protection schemes like Fast Reroute (FRR) is not uncommon. As an alternative or supplementary approach, some networks employ 1+1 protection at the optical layer, which can achieve sub-50ms protection transparent to the IP/MPLS layer, but which requires 100% dedicated protection bandwidth at the optical layer. Other multi-layer recovery strategies to help increase the

utilization of router links include combining optical wavelength restoration along with MPLS FRR [2], but with sub-50ms protection still resident within the router layer, only so much reduction in overprovisioning can be achieved, as some spare router bandwidth must be available for failover scenarios, even if just for high priority traffic, such as Expedited Forwarding (EF) or Assured Forwarding (AF) DiffServ traffic.

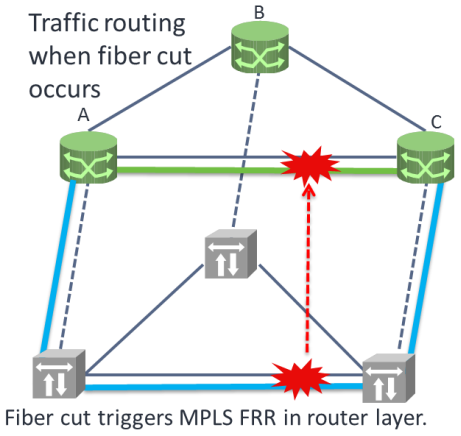


Fig 4a – with MPLS FRR, a fiber cut triggers protection at higher router layer, consuming spare resources at both the router and optical layers.

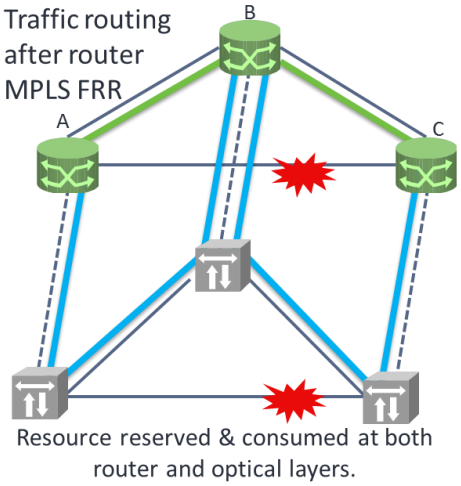


Fig 4b – Traffic from A to C is restored via router B in sub-50ms, but requires over-dimensioning of IP links, which also requires corresponding spare optical capacity.

With advances in the optical transport layer, however, new alternatives to relying solely upon routers to host basic packet transport, aggregation, multicast, and protection schemes are now emerging. These new tools are enabling architects to more efficiently manage traffic internally within the transport layer, thereby reducing the amount of traffic that needs to be processed by routers.

PACKET AWARE TRANSPORT

State of the art converged L1/L0 transport systems today combine sub-wavelength granularity circuit switching fabrics (typically OTN) along with WDM transmission, enabling not just multi-service and multi-rate transport services over the optical infrastructure, but also efficient switching and grooming of services into wavelength containers. These platforms can also support new advanced, deterministic sub-50ms shared mesh protection (SMP) schemes [3,4,5]. Such systems provide operators with the ability to decouple bandwidth services from the analog transmission layer, and facilitate a much more flexible operational model for dynamically provisioning transport bandwidth services, when compared to traditional transponder/muxponder and ROADM based approaches.

The next evolutionary phase of the optical transport network involves adding packet-awareness – the ability to perform Ethernet and MPLS packet processing (e.g., classification, metering, traffic management, QoS) at 100GbE speeds (and lower), mapping packet flows based on configuration rules into multiple flexibly sized ODUflex circuits, and expressing these individually encapsulated packet flows within the transport network to their end destinations.

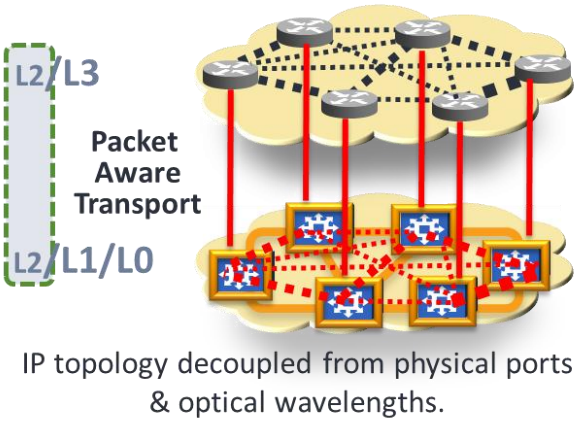


Fig 5 – packet-awareness in the transport layer converges L2/L1/L0 functionality, enabling more flexible and bandwidth-efficient router interconnect, as well as natively supporting Ethernet E-LINE/E-LAN/E-TREE services. L2 technology in both the router and transport layer enables better cross-layer resource coordination.

This convergence of L2/L1/L0 functionality transforms the transport network from one that delivers only 10/40/100G transparent wavelength services between routers to one that can natively support flexible packet-transport interconnect solutions without necessarily having to pass traffic back up to the router, thereby offloading routers and router ports from having to handle L2 transport functions and the associated traffic, and instead focus on higher-level L2/L3 service delivery. The new fundamental networking building block this enables is a **Packet-aware Virtual Wavelength** (PVW).

Packet-aware Virtual Wavelengths Overview

PVWs enable the transport layer to now become fully aware of the packet traffic coming from the routers while providing the performance and predictability of switched sub-wavelength granularity circuits. With Ethernet packets serving as a common technology denominator between the router and optical layers, greater synergies can be realized between both network layers. In addition to packet processing and traffic management fnctions typically found in routers and switches, PVWs can map individual packet flows into individual ODUflex circuits at the optical layer and apply integrated L1 and L2 QoS mechanisms to concurrently manage packet and circuit characteristics. Flexible packet mapping can be performed based on address fields such as VLAN IDs or MPLS labels, and can be further enhanced through the use of ACLs to perform policy-based forwarding.

As in packet systems, PVWs can apply different CoS models, such as flow-based and class-based queuing, and apply them on a per-client and per network-facing interface. Internally, MPLS-TP Pseudowires (PWEs) can provide a scalable encapsulation mechanism for the packet flows, before being mapped into ODUflex container circuits for optical transport. With both L1 and L2 functions localized, the amount of packet oversubscsription can be well coordinated with the size of the ODUflex circuit in increments of 1.25Gb/s, creating the cross-layer glue between packets and circuits and ensuring a proper over-subscription ratio is maintained for optimal balance between performance and economics. As more or less dedicated bandwidth is needed for the mapped packet flows, the ODUflex circuits can be hitlessly resized.

The seamless integration between L2 and L1 also provides a vehicle to differentiate the transport of packet services, even between the same source/destination routers. Routers that are adjacent from the L3 control plane perspective can use PVWs to distinguish single-hop traffic, based on some packet attributes, such as priority bits. High and low latency transport paths, for example, can be used to distinguish different application flows coming from a single router port. This provides operators with simpler way to control how transport network assets are utilized for supporting specific packet traffic flows.

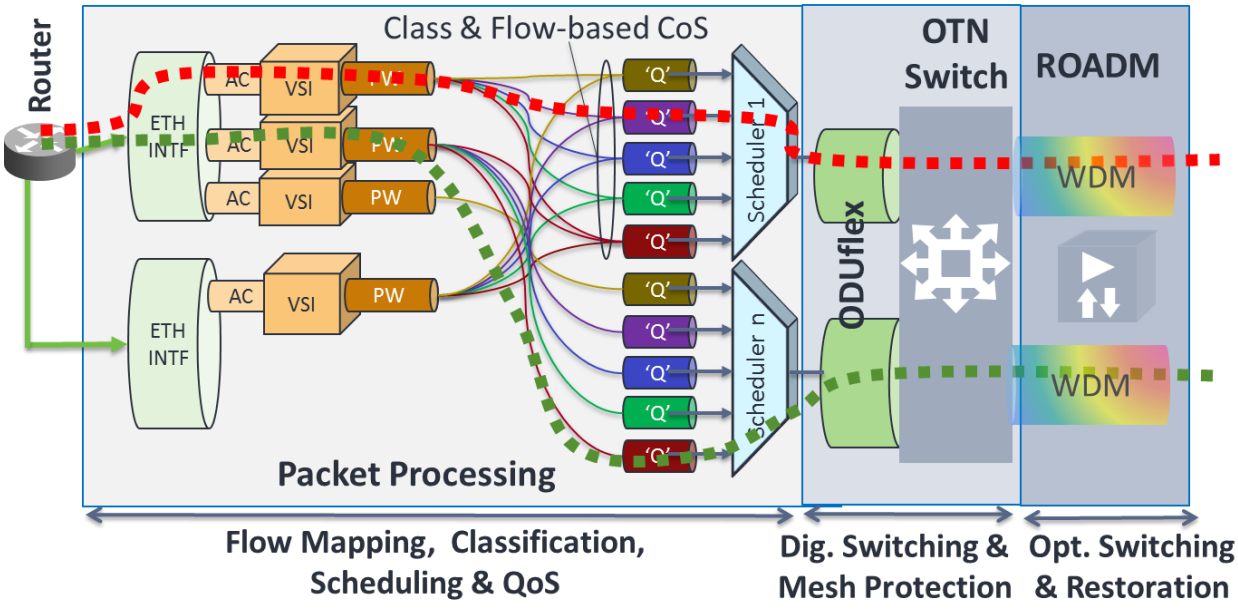


Fig 6 - Integration of L2 packet processing functions along with L1 OTN switching/grooming and L0 WDM ROADM switching creates a highly flexible packet-transport and service delivery platform. Highly differentiated transport capabilities can be offered for packet flows from routers, as well as native MEF Ethernet services.

Flexible Packet Transport Services

The flexible mapping of packet flows from the router into PWEs and then into ODUflex circuits also gives rise to new point-to-point (P2P), point-to-multipoint (P2MP) and multipoint-to-multipoint (MP2MP) packet transport services (including Carrier Ethernet services), delivered natively by the transport network. In addition to E-LINE type services (EPL, EVPL), PVWs can enable E-TREE and E-LAN based services without relying on an external system. Both Ethernet UNI and ENNI based services are supported, with full support for untagged traffic as well as C-VLAN and S-VLAN tagged traffic.

Standards-based L2 traffic management capabilities within the PVW facilitate the support for MEF 2.0 compliant services. Packet metering, queuing and scheduling along with standard CoS support and configurable ingress and egress bandwidth profiles provide a full complement of packet QoS capabilities for hosting native high-speed Ethernet services directly on the transport platform, up to 100GbE.

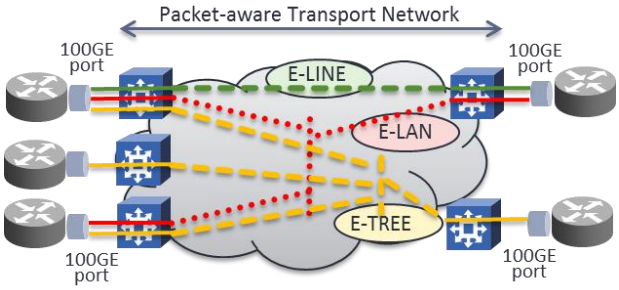


Fig 7 – The transport network can cost-effectively offload routers for high-speed Ethernet services, including E-LINE, E-TREE, and E-LAN.

Unlike conventional packet-switched networks for delivering packets to their destination, however, PVW-based networks intrinsically provide fine-granular, guaranteed dedicated bandwidth directly between the service termination sites, as the transport network fabric is based on switched ODUflex circuits. This ensures low-latency transport of traffic and guaranteed jitter for packet services, compared to packet-switched networks, since packet flows within PVWs

are not subject to congestion or contention from any flows other than the ones mapped into these transport tunnels at the source location.

Since separate PVWs can be used for different flows originating from a common router interface, operators can leverage this as a new tool for engineering differentiated, bandwidth efficient, SLA-enforceable packet transport services. PVWs provide operators with controls over typical L2 QoS parameters as well as L1/L0 circuit-centric bandwidth metrics (size, latency, transport layer path diversity, etc.). Oversubscription can be controlled per PVW, as well as the protection or restoration policy as part of a multi-layer protection strategy.

For P2MP and MP2MP services such as E-TREE and E-LAN, PVWs facilitate the creation of isolated transport overlay meshes, interconnecting multiple service termination sites. Multi-point EVCs can then be established that can perform the MAC-based learning and forwarding needed for ETREE and ELAN services. This packet transport capability can be leveraged not just for hosting high-speed MEF 2.0 multi-point services within the transport layer, but also for simplifying and reducing the cost of implementing IP multicast based services, as described later.

Transport-layer Protection of Packet Services

The convergence of L1 and L2 technologies into a packet-aware transport network also provides a rich set of flexible protection schemes at both the packet and OTN layers, and can significantly reduce the protection bandwidth requirements at the router layer. State-of-the-art packet-aware transport systems can offer ITU G.808.3 and G.ODUSMP compliant Shared Mesh Protection (SMP) schemes that provide resiliency against network failures in under 50ms for PVWs. Fast SMP schemes provide the performance guarantees typically associated with 1+1 protection schemes, but with much better bandwidth efficiency, by utilizing shared protection bandwidth and sophisticated algorithms for ensuring sub-50ms performance times for all impacted SMP-protected PVWs. Priority and preemption capabilities further enhance the differentiated transport protection schemes available.

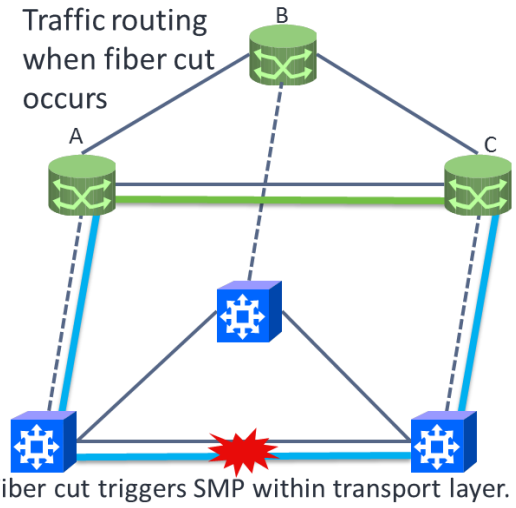


Fig 8a – with transport-layer SMP, the fiber cut is reacted to locally within 50ms.

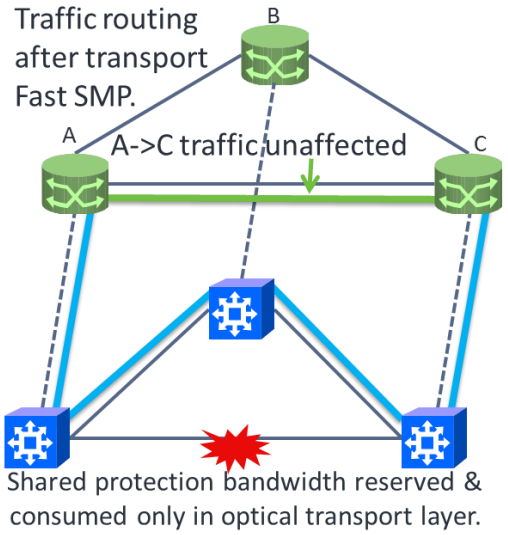


Fig 8b – since the SMP protection restores traffic within 50ms, the A->C traffic is restored without any impact to the LSP path. This reduces LSP churn and reduces over-dimensioning of IP links for protection bandwidth.

Fast SMP gives network operators a new alternative to MPLS FRR for protecting packet services, thus reducing overprovisioning of bandwidth at the router layer [5]. A proper mix of schemes at both the router and transport layers can be utilized, depending on use-case scenarios. For example, in a long-haul core backbone, larger LSPs or flows aggregated into PVWs can be mesh protected at the transport layer, closer to where network failures occur, while smaller flows can still be protected at the router layer, based on the operator's own traffic mix and service protection requirements. Operators that use PVWs to distinguish different levels of business services by mapping them to individual SLA-managed PVWs can apply different protection strategies for further differentiation.

Additionally, optical layer restoration leveraging ROADMs can provide operators with yet another layer of network resiliency by restoring optical capacity at the wavelengths and super-channels granularity. Such analog restoration schemes generally take several seconds, and can complement the sub-50ms performance of digital protection schemes when additional resiliency is required.

CABLE NETWORK APPLICATIONS & USE-CASES LEVERAGING PACKET-AWARE TRANSPORT

With the features, flexibility and economic benefits packet-aware transport networking can offer, cable operators can leverage this new set of capabilities across multiple applications. Some key use-case applications are described below.

Transport Layer Master Head-End Packet Aggregation

In metro and regional cable networks, it is common to find hub-and-spoke packet traffic patterns as traffic is aggregated upstream from multiple head-ends or hubs to a master head-end. With a conventional optical transport network, the packet-layer aggregation function is performed in routers, sometimes before passing traffic to the optical transport layer. In other scenarios, dedicated wavelengths are used to hub the routers from multiple sites along a ring back to the hub site, where the wavelengths terminate into a larger router.

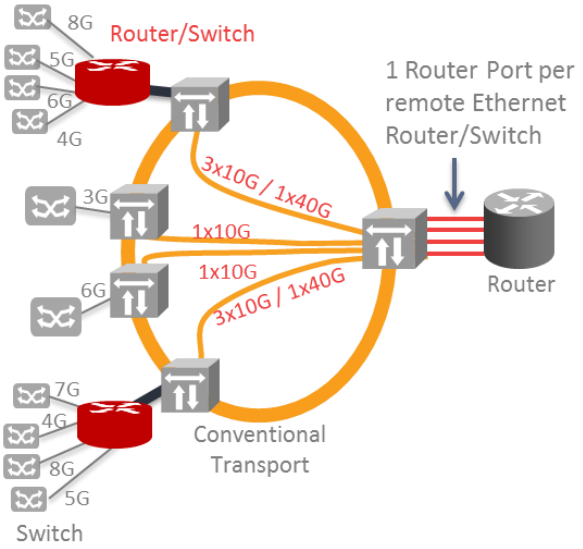


Fig 9a – in a conventional architecture, router ports aggregate packet traffic and transport the packets over dedicated wavelengths back to the hub location.

Packet-aware transport offloads the packet aggregation function from the router and handles it natively within the transport layer, thereby reducing router ports or potentially eliminating the aggregation router altogether. Multiple 10Gb/s ports can be aggregated locally at edge locations, while at the master head-end, packet traffic from multiple sites can be aggregated and handed off to a core router. Additionally, with the PVWs being configurable in 1.25Gb/s increments, transport bandwidth can be efficiently allocated and resized, based on actual packet

traffic load, instead of burning an entire optical wavelength.

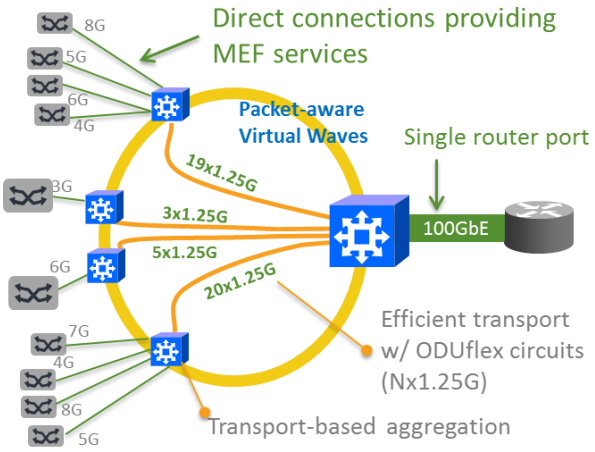


Fig 9b – in the packet-aware transport architecture, aggregation is handled natively in transport layer, and an appropriate amount of optical bandwidth is allocated, eliminating stranded capacity. In this example, the aggregation router is no longer needed.

Cost-efficient, Scalable LSP Protection

In multi-layer long-haul core networks that overlay IP/MPLS over an optical transport layer, the combination of packet-aware transport and Fast SMP can enable a much more cost-efficient traffic-engineered backbone through various router offload and bypass tools.

With MPLS-TP packet aware transport, LSPs can be mapped and aggregated into appropriately dimensioned ODUflex circuits at the optical layer, creating express lanes through the multi-layer network and bypassing intermediate LSRs. Transport of LSPs between the edge routers can further be differentiated through the use of separate ODUflex circuits for different groups of LSPs, each of which may have different latency characteristics and which may employ different protection schemes. New protection schemes such as Fast SMP can be used in lieu of MPLS FRR to provide sub-50ms protection capabilities, but with better bandwidth efficiency, resulting from reduced over-dimensioning at the router layer. By aggregating multiple LSPs into Fast SMP protected ODUflex circuits, a scalable end-to-end recovery scheme is enabled while simultaneously reducing LSP churn.

High-speed Pay-as-you-Grow Ethernet Service Delivery

A packet-awareness transport system can natively support MEF 2.0 services for high-speed interfaces, and offload higher-layer routers/switches. High-bandwidth Ethernet services that are hosted on routers/switches consume resources at both network layers, and incur interconnect costs associated with transporting the service from the router port to the optical transport layer.

By hosting services such as fractional 100GbE EVPL or EPL services directly off the transport platform, the cost of utilizing router/switch resources and the interconnect cost for those services are eliminated. Furthermore, an optimal amount of optical capacity can be allocated in increments of 1.25Gb/s to transport MEF services efficiently without stranding capacity, as can be the case when PVWs are not utilized, and an entire wavelength is allocated to transport an Ethernet service (e.g., 50Gb/s) that is less than the optical transport wavelength (e.g., 100Gb/s). PVWs also provide coordinated cross-layer configuration of L2 QoS parameters along with L0/L1 bandwidth characteristics for Ethernet services. Customer Ethernet services can be individually classified into ODUflex circuits, each of which can be switched to different destinations, and which can offer different latency, diversity, and protection/restoration characteristics. Hitless resizability of the PVW ensures the proper amount of capacity is provided for the supported Ethernet services, facilitating bandwidth-efficient pay-as-you-grow services for customers.

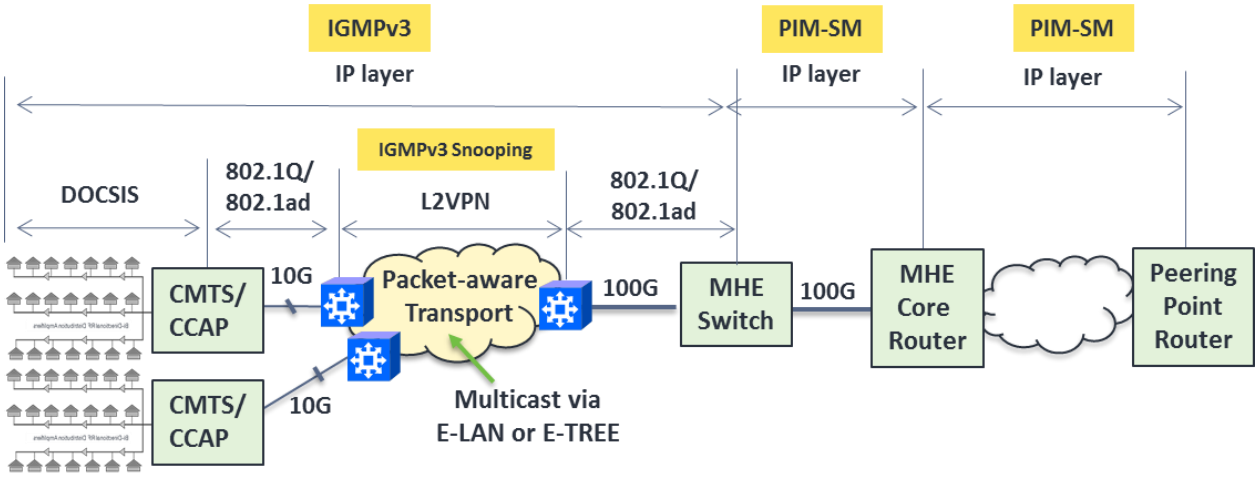


Fig 10 – IPTV multicast based on embedded E-LAN/E-TREE functionality and IGMP snooping within the transport layer can simplify the network by eliminating the aggregation router at CMTS/CCAP sites.

Transport Layer IPTV Multicast

Multi-cast IPTV is an increasingly strategic broadcast digital video delivery method for many cable operators as they make the transition to all-IP based service delivery. Conventional IPTV architectures leverage the router infrastructure to perform all the multi-cast functions – the combination of one or more routers at the master head-end and at each of the regional head-ends interconnected with static wavelengths generally provides support for IGMP/PIM and handles the multi-cast control plane and data plane functions.

With configurable E-LAN and E-TREE services natively available within the packet-aware transport network, however, packet flows can be multi-cast within the transport layer instead of within the router. This can eliminate the need for routers at the regional head-end performing packet aggregation from multiple local CMTS/CCAPs, and additionally can help enhance the utilization of the router port at the master head-end by relieving it from performing the multi-cast function. IGMP snooping enabled in the transport layer for each ELAN/ETREE instance ensures proper establishment of multi-cast forwarding entries downstream of the master head-end router to the CMTS/CCAP.

SUMMARY

The growth in bandwidth requirements on MSO networks necessitates exploration of new technologies and networking architectures in order to scale networks economically. The integration of packet-awareness into the converged L0/L1 network enables an evolution of the optical transport network from one that delivers point-to-point packet-agnostic bandwidth pipes to one that can deliver more dynamic, flexible packet transport services. This not only enables the transport network to natively host packet services such as E-LINE, E-LAN, and E-TREE services, but also facilitates more flexible and bandwidth-efficient interconnect solutions for routers, reducing the total cost of the network. The features enabled by packet-aware transport networks create a set of new networking tools available to MSOs, and include:

- Flexible mapping of packets into PVWs based on packet header fields, including labels & IDs, or additional

fields such as address or packet classification information

- Coordinated cross-layer configuration of packet & transport QoS parameters
- Per-PVW protection and restoration options, including sub-50ms SMP, serving as an alternative to MPLS FRR
- Integrated P2MP and MP2MP multicast forwarding for E-TREE and E-LAN support
- Hitless resizability of the optical bandwidth associated with PVWs to accommodate elastic scale-up and scale-down packet traffic

With these tools, many cable network applications can be more efficiently and cost-effectively implemented as compared to today's conventional approaches.

ACKNOWLEDGEMENTS

The author would like to acknowledge Gaylord Hart for reviewing this paper and providing insightful comments.

REFERENCES

[1] Network Cost Savings from Router Bypass in IP over WDM Core Networks, S. Melle et alia, OFC/NFOEC 2008.

[2] Multi-layer Restoration – The Impact on the Optical Layer, M. Gunkel, OFC 2014.

[3] Draft Recommendation ITU-T G.808.3., "Generic protection switching - Shared Mesh Protection"

[4] Draft Recommendation ITU-T G.ODUSMP, "OTN protection switching - Shared Mesh Protection"

[5] Novel Design of G.ODUSMP to Achieve Sub-50 ms Performance with Shared Mesh Protection in Carrier Networks, W. Wauford et alia, ECOC 2013.

Recent Advancements in Audio – How a Paradigm Shift in Audio Spatial Representation & Delivery Will Change the Future of Consumer Audio Experiences

Jeffrey C. Riedmiller and Nicolas Tsingos
Dolby Laboratories Inc.

Abstract

Recent advancements in spatial audio processing, distribution and delivery are now capable of bringing more lifelike, scalable and interactive audio experiences to consumers across a wide range of devices and applications than ever before. This paper provides the reader with a high-level tutorial on the latest research and development supporting the creation, distribution and delivery of immersive and personalized audio experiences to consumers.

BACKGROUND

Over the past 70+ years, broadcast television sound has evolved from providing a single *complete* (monophonic) audio program to having the capability of supporting a combination of 5.1-channel and stereo sound tracks, including multiprogram support for delivering alternate languages and/or descriptive audio. While the digital transition enabled efficient delivery of multiple audio programs to consumers, television sound continues to be created, distributed and delivered as one or more *complete* audio programs just as it has been for decades. The broadcast standards community (including the bitstream syntax utilized for consumer delivery) typically refers to these audio service types as *Complete Main* services. *Complete Main* services contain a complete audio program including dialogue, music and effect elements *mixed* together in either 5.1- or 2.0-channel format *prior* to being broadcast to consumers. Moreover, additional languages and/or descriptive programming are also *mixed* with dialogue, music and effects elements *prior* to being broadcast as *Complete Main* service types as well. Today, it is quite common for television services to include two audio programs (i.e. two elementary compressed bitstreams multiplexed with video and other service information). The main (primary) audio program in the native language is typically offered in 5.1-channel format while the secondary language is typically provided in stereo. Descriptive audio programming, when available, typically replaces the secondary language with native language descriptions pre-mixed over a stereo version of the native language soundtrack – restricting the availability of secondary language programming.

The primary and secondary audio program examples above provide viewers with two audio ***presentations*** in two-dimensional (5.1) and one-dimensional (stereo) channel-based formats respectively; where the primary program is offered in 5.1 (two-dimensional) and the secondary program is offered in 2.0/stereo (one-dimensional). The term ***presentation(s)*** is a key term associated with next-generation consumer audio experiences that aim to offer a greater degree of ***personalization*** under viewer control along with providing non-native speaking, hearing-impaired and visually-impaired audiences with the highest quality experience (only available with the primary audio program today).

To address the currently limited set of audio presentations (and related experiences), recent advancements in the development and

deployment of spatial audio [1][14][15] systems are poised to make the personalization of audio a practical reality for content creators, programmers, MVPDs (Multichannel Video Programming Distributors) and consumers alike. Additionally, these advancements also provide a substantial step forward in terms of ***immersiveness*** by enabling sound to be represented and delivered to consumers in three-dimensions. Immersive sound research and formats are fundamentally focused on providing a more accurate representation of the spatial attributes of an audio program (in a higher-dimensional space) while providing listener experiences that engage our sensory system in a more natural way – allowing sensory systems to process information more similarly to how listeners experience the natural world by providing more realistic and natural auditory cues to the listener including advanced binaural for a headphone listener.

IMMERSIVE SOUND – OBJECT-BASED AUDIO

The object-based audio paradigm provides the foundation for immersive sound and consists of a set of monophonic audio elements (tracks) and tightly coupled metadata representing the object's position in three-dimensional space. The positional metadata is generated with a great enough frequency to enable dynamic object movement and precise localization during playback. At playback time, an audio rendering engine utilizes this metadata to map each of the audio objects to a single or a combination of speaker outputs to achieve the desired spatial effect. Importantly, each object's position is represented as a coordinate in three-dimensional space that can also leverage overhead and/or upward-firing speaker configurations in addition to advanced 3D headphone rendering. For instance, ITU-R BS.2051 [10] defines up to 22 primary channels that can be used for playback. In the home, we expect 5.1 + 2 overhead speakers or 7.1 + 4 overhead speakers to be the most widely adopted 3D loudspeaker configurations.

Today's channel-based audio paradigm along with the emerging object-based audio paradigm are alternative methods for describing sound events. Channel-based audio expresses the audio scene in terms of loudspeakers utilizing pre-determined (nominal) positions and characteristics. Object-based audio describes sound events in a way that is independent of loudspeaker configurations/layouts and positions. The spatial accuracy, homogeneity and resolution possible with object-based audio enable new possibilities for the sound designer, mixer and director.

In traditional channel-based audio mixing, sound elements are mixed together using fixed speaker positions and angles assumed to be identical in every playback environment. For instance, 5.1-channel audio productions are created utilizing the left, right, center, left surround, and right surround speaker layout conforming to ITU-R BS.775. However, 5.1-channel playback in consumer environments typically deviates in terms of speaker distance and angles relative to the listener's position from the ITU-R BS.775 recommendation. This deviation has a substantial impact on the consumer's ability to experience the soundtrack in a manner that was intended (and heard) by the program producer during production. On the other hand, object-based audio mixing enables ***individual*** or ***groups*** of sound elements to be described in a three-dimensional space (rather than being forever assigned to a specific channel/speaker location) and delivered to the playback device where they are ***rendered*** based on the active speaker layout in use, including the number and position of speakers in 1-, 2- or 3-dimensions. The benefit of this approach is that by decoupling audio mixing and spatial representation from a single fixed channel-based layout, the audio soundtrack/experience

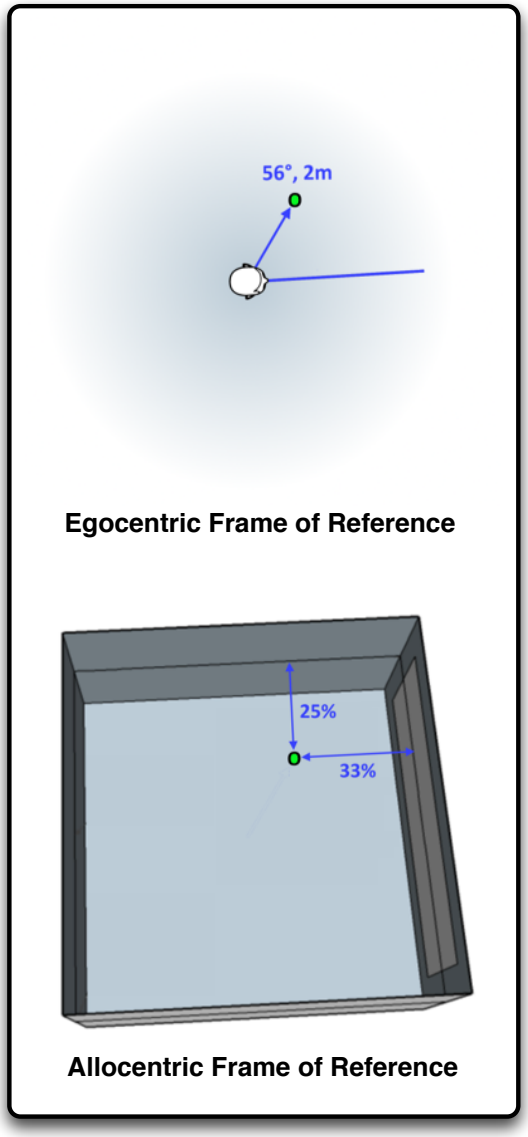
can scale optimally across any speaker layout in use today or in the future. In other words, the creative process utilizing objects mixes for the 'space' rather than to a fixed speaker location – again, which is assumed to be identical in every playback environment.

Coordinate System & Frame Of Reference

When designing a system to enable individual audio elements (objects) to be represented in 3-dimensional space, a frame of reference is required. Typically a system designer will consider the distinction between an *egocentric* (observer) or an *allocentric* (environmental) reference as a fundamental consideration.

An egocentric frame of reference represents (or encodes) an audio object's location relative to the position of the observer (or 'self'). An egocentric reference is commonly used for the study and description of perception.

An allocentric frame of reference represents (or encodes) an audio object's location using a reference location and direction relative to other objects in the environment. An allocentric reference is better suited for a scene description that is independent of a single observer's position and when the relationship between elements in the environment is of interest.



The rationale for choosing a frame of reference should consider the following:

- Provide optimum capture of artistic intent
- Provide optimum translation across a variety of listening environments.
- Be consistent with the production tools utilized to capture artistic intent.
- Provide consistent and predictable behavior across a wide listening area.

In general, audio mixers naturally approach the creative process in allocentric terms that is further reinforced by the fact that

audio panning tools having been implemented with an allocentric frame of reference for decades (i.e. the screen, room left, room back, etc.). As an example, the mixer's thought process for rendering audio objects follows "this sound should be on-screen", "this sound should be off-screen and a ¼ of the way from the left to the right wall", etc. Thus, audio object movement in space is defined in relation to the playback environment, e.g. a fly-over from the center of the screen, up across the ceiling and ending at the center of the back wall. Using an allocentric frame of reference, all of these relationships are preserved.

In contrast, utilizing an egocentric frame of reference can, for example, yield results where an object placed on the side wall of a large rectangular mixing stage ending up on the rear wall of a more square space during playback. Moreover, with three-dimensional egocentric audio frameworks (which includes distance) an object on the rear wall of a small mixing stage could end up well within the audience area of a large playback space such as an auditorium.

Artistically, the location of an audio object is most important when associated with an on-screen visual object. Using an allocentric reference and representation for every listening position, and for any screen size, individual sound objects can be described at the same relative position on the screen and optimally reproduced across a wide range of room sizes and shapes found from cinema to home environments.

Considering the rationale behind the use of an allocentric frame of reference, an audio object's position is preferably defined as a unit room where each object(s) 3-dimensional coordinates (x,y,z) in the range of [-1,1] x [-1,1] x [-1,1], correspond to the traditional panpot controls found in mixing consoles - left/right, front/back and by extension to 3D bottom/top. This is commonly referred to as a cartesian room normalized coordinate system, where X= -1 is full left, X=1 is full right, Y= -1 is front, Y=1 is back, Z=0 is traditional surround plane, Z=1 is overhead plane. This can be extended to Z= -1 to include a floor plane (below the listener). This defines and describes a normalized room hemicube/cube as in Figure 2.

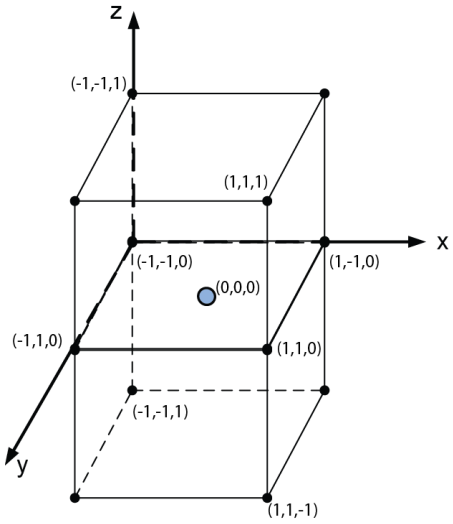


Figure 2 : Object Position Coordinate System

Modern surround-sound cinema and object-based consumer delivery systems use an allocentric frame of reference. ITU-R BS.2051 [10] provides mapping coordinates for predefined channels and corresponding loudspeaker positions associated with advanced sound reproduction systems.

Object-based Audio Rendering

An audio object rendering engine is a mandatory requirement for playback products and applications that support object-based audio for immersive and personalized audio experiences. An audio renderer converts a set of audio object signals with associated metadata to a different configuration of audio signals – e.g. speaker feeds, based on the metadata, **AND** a set of control inputs derived from the rendering environment and/or user preference.

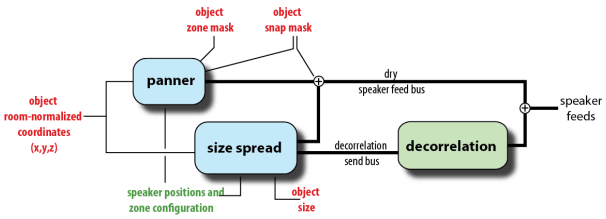


Figure 3 – Object-based Audio Renderer

At the core of rendering is a panning algorithm (see Figure 3). Most panning algorithms currently used in object-based audio production attempt to recreate the original auditory cues during playback via amplitude panning techniques [5][6][7][8] where, a gain vector [**Gi**] is computed and assigned to the source signal for each of the *n* loudspeakers in use The object signal s(t) is therefore reproduced by each loudspeaker as Gi(x,y,z)*s(t) in order to recreate suitable localization cues as indicated by the object <x,y,z> coordinates.

The design of panning algorithms ultimately has to balance tradeoffs among timbral fidelity, spatial accuracy, smoothness and sensitivity to listener placement in the listening environment, all of which can affect how an object at a given position in space will be perceived by listeners [8][9]. For instance, Figure 4 illustrates how different speakers maybe utilized among various rendering (panning) algorithms to place an object's perceived position in the playback environment.

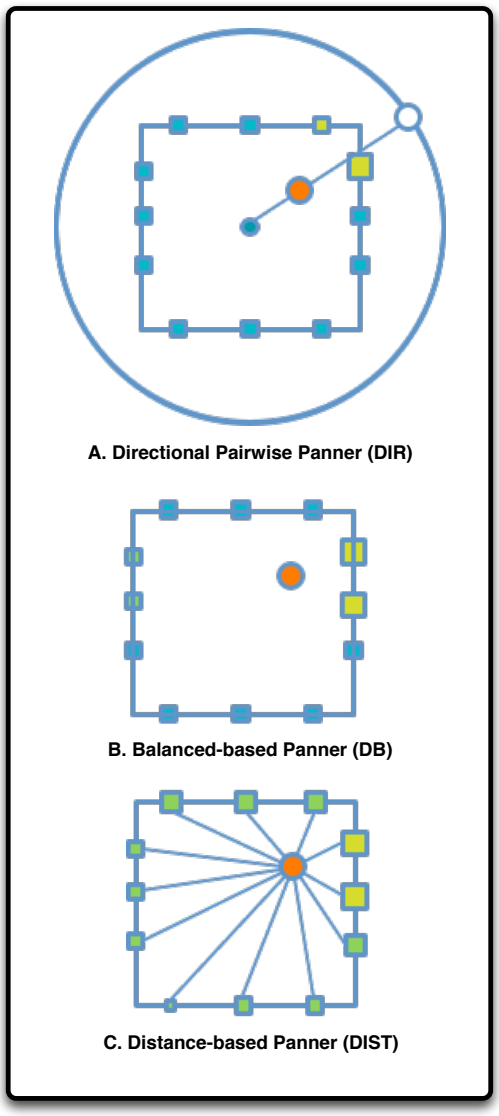


Figure 4 – Common Panning Algorithms

Directional pairwise panning (DIR) (see Figure 4 (a)) is a commonly used strategy that solely relies on the directional vector from a reference position (generally the sweet-spot or center of the room) to the desired object's position. The pair of speakers 'bracketing' the relevant directional vector is used to place (render) that object's position in space during playback. A well-documented extension of directional pairwise panning to support 3D loudspeaker layouts is vector-based amplitude panning (VBAP) [7] which uses triplets of speakers. As this approach only utilizes the direction of the source relative to a reference

position, it cannot differentiate between object sources at different positions along the same direction vector. It can also introduce instabilities as objects are panned near the center of the room. Moreover, some 3D implementations may constrain the rendered objects to the surface of a unit sphere and thus would not necessarily allow for an object to cross inside the room without going 'up and over'. As a side note, directional panning solutions can also create sharp speaker transitions as objects approach the center of the room. Where, a small movement an object's position would not always translate into a small variation in loudspeaker gains [Gi] for objects moving (panning) *through* the room/space.

The 'dual-balance' panning algorithm is the most common approach used in 5.1/7.1-channel surround productions today (Figure 4 (b)). This approach utilizes left/right and front/back panpot controls widely used for surround panning. As a result, dual-balance panning generally operates on the set of four speakers bracketing the desired 2D object position.

Extending to three-dimensions (e.g., when utilizing a vertical layer of speakers above the listener) yields a "triple-balance" panner. It generates three sets of one-dimensional gains corresponding to left/right, front/back and top/bottom balance values. These values can then be multiplied to obtain the final loudspeaker gains:

$$Gi(x,y,z) = Gx_i(x) \times Gy_i(y) \times Gz_i(z).$$

This approach is fully continuous for objects panned across the room in either 2D or 3D and makes it easier to precisely control how and when speakers on the base or elevation layer are to be used.

In contrast to the directional and balance-based approaches, distance-based panning (DIST) [5] (Figure 4(c)) uses the relative distance from the desired 2D or 3D object location to each speaker in use to determine the panning gains. As a result, this approach generally utilizes all of the available speakers in use rather than a limited subset which leads to smoother objects pans but with the tradeoff of being prone to timbral artefacts.

It should be noted that both 'dual balance' and 'distance-based' panning support smooth object pans in the sense that a small variation in an object's position will translate to a small change in loudspeaker gains.

Object-based Audio Rendering Requirements

Any object rendering algorithm used in professional or consumer playback applications must address (at a minimum) the following high-level functional requirements:

1. Support smooth panning for object rendered inside and through the room – including the two-dimensional surround plane and the three-dimensional cube/hemicube described above.
2. Support for rendering utilizing a traditional two-dimensional surround speaker layout even when overhead speakers are present. e.g. objects with z = 0 will render across the 2D speaker layout while objects with z = 1 will only render via overhead/upward firing speakers.
3. Implement an approach that minimizes the number of speakers used to reproduce an object – to preserve timbral fidelity.
4. Support additional creative control (via metadata) including the ability to dynamically mask out some of the speakers (or zones of speakers) when an object is being rendered across specific loudspeaker configurations
5. Be power preserving across speaker layouts: $\sum i\ Gi^2 = 1$.

Supporting the requirements above, the authors recommend a triple-balance implementation for deriving loudspeaker gains from object coordinates. The triple-balance approach expresses speaker gains **Gi** as the separable product of three panpot controls: left/right, front/back, top/bottom and is a direct extension of the commonly available panning tools on mixing consoles. The following background and supporting equations provide a representative example of how a simple panning algorithm would be implemented using a simple sine/cosine law. Other panning laws are also possible.

An indicative example for computing a one-dimensional (stereo) set of rendering gains utilizing an audio object's x-coordinate in [-1,1] could be derived as follows:

$$G_left = \cos((x+1)/2.0*\pi/2)),$$
$$G_right = \sin((x+1)/2.0*\pi/2))$$

An indicative 3/2/0 example (L/C/R/Ls/Rs) would be (2D rendering using <x,y> coordinates in [-1,1] x [-1,1]) :

A) Set all gains to 0.0

B) Compute left/right panpot for front and back speakers as in the previous stereo example:

```
G_ls = cos((x+1)/2.0*π/2)
G_rs = sin((x+1)/2.0*π/2)

if (x<=0.0f)
{
 G_l = cos(-x*π/2);
 G_c = sin(-x*π/2);
}
else
{
 G_c = cos(x*π/2);
 G_r = sin(x*π/2);
}
```

C) Combine with front/back panpot:

```
float c = cos((y+1)/2.0*π/2);
float s = sin((y+1)/2.0*π/2);

G_l *=c;  G_r *=c; G_c *=c;
G_ls *=s;  G_rs *=s;
```

D) Normalize power to 1.0 by dividing all gains Gi by $sqrt(\Sigma i\ Gi^2)$

Following the same principle, these examples can be easily extended using a third dimension for elevation (height).

Artistic Control Metadata

While the use of a consistent core rendering algorithm is desirable, it cannot be assumed that a given solution will always deliver consistent and aesthetically pleasing results across different playback environments. For instance, today the production community commonly remix the same soundtrack for different channel-based formats in use worldwide, such as 7.1/5.1 or stereo, to achieve their desired artistic goals for each format. With several hundred audio tracks competing for audibility, maintaining the discreteness of the mix and finding a place for all the key elements is a challenge that all theatrical/TV mixers face. Achieving success often requires mixing rules that are deliberately inconsistent with a physical model or a direct re-rendering across different speaker configurations.

To solve this problem, we introduce additional metadata used to dynamically reconfigure the object renderer [11] (Figure 3). For instance, the metadata can be used to "mask out" certain speaker zones during playback or use only a single loudspeaker to render an object for improved timbral fidelity. Alternatively, metadata can also be provided to render wide objects, covering a large spatial extent. Implementation of wide objects

requires the use of decorrelation processing in order to recreate the proper cues for perceptual size.

PERSONALIZED SOUND & ACCESSIBILITY

Audio personalization is a new concept that aims to enable viewers to choose and, when appropriate based on the intent of the content creator, tailor the experience to their unique needs or preferences. Personalization can include selection of languages, alternative commentary, visual descriptions along with the ability to adjust the level (and in some cases the spatial) relationship between the dialogue-based and music/effects audio object elements. Extending the object-based audio paradigm to support the professional interchange and consumer delivery of both individual audio objects and groups of audio objects (with accompanying metadata) will provide the building blocks for enabling the personalization of audio programming [12]. It will also ensure that the entire audience, including native and non-native speaking, hearing and visually impaired, are provided with a premium (including an immersive) experience typically only available with the primary language audio program today. As a simple example, the primary object-based audio building blocks (including metadata) for a program can be defined as represented in Figure 5.

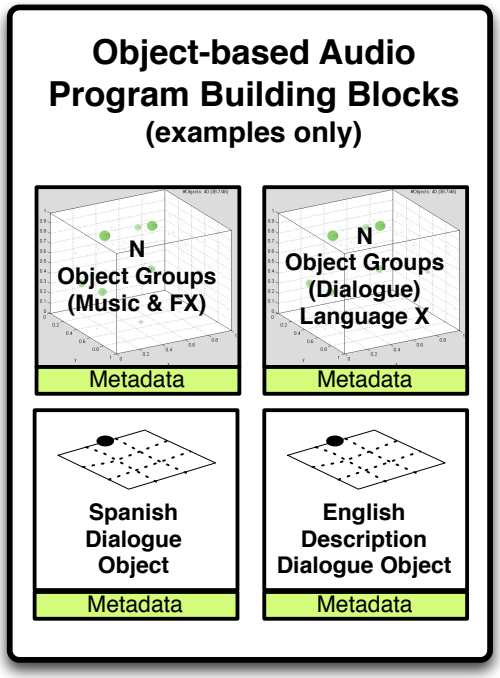


Figure 5 – Objects & Object Groups

The Music & FX object group (as the name implies) contains a set of spatial audio objects and metadata representing only the music and sound effects for the program. As represented in Figure 5, this object group when being simultaneously rendered with the English Dialogue object group would represent a single immersive *presentation* for the consumer. While the combination of the Music & FX object group and the Spanish dialogue (represented here as a single object rather than an object group with spatial attributes including spatialized effects - reverb, etc.) represents a second immersive *presentation.* A third *presentation* is further represented by the simultaneous rendering of the Music & FX group, English object group and the English description object. Again, these elements can be rendered based on the preferences of the consumer/viewer.

This approach provides more efficient delivery and higher quality experiences for non-native speaking and visually- and hearing-impaired audiences by supporting a fundamentally different approach to

programming distribution. Instead of distributing complete mixes of each type (for example, hearing- or visually-impaired), this approach distributes the program building blocks as discrete objects or object groups each containing music and effects, main dialogue, alternate dialogue, dialogue descriptions, etc. These program building blocks are then dynamically rendered during playback, possibly with consumer adjustment (personalization – including adjustment of the relative level between dialogue objects and the Music & FX object group – this address the needs of the hearing impaired audience). This approach also improves efficiency by reducing the overall datarate required for delivering the complete set of possible presentations.

A FLEXIBLE CONSUMER DELIVERY FORMAT FOR IMMERSIVE & PERSONALIZED AUDIO

Enabling efficient and flexible delivery of immersive and personalized programming to consumers also requires a new approach for consumer delivery. The AC-4 format standardized in ETSI TS 103 190 [13] provides a state-of-the-art and efficient mechanism for delivering both personalized and immersive audio experiences in object-based or traditional channel-based paradigms.

The AC-4 bitstream consists of a sequence of synchronization frames as shown in Figure 6. Each synchronization frame begins with a synchronization word and ends with a CRC word. The synchronization word enables a decoder to easily identify frame boundaries and begin the decode process while the CRC word allows a decoder or other system component to detect the occurrence of bitstream errors and invoke concealment as necessary. The portion of the AC-4 synchronization frame containing all of the audio essence and metadata is referred to as the raw AC-4 frame.

The high-level AC-4 bitstream structure shown in Figure 6 supports the carriage of audio and metadata substreams described by the Table-of-Contents (TOC) bitstream element. The arrows indicate the order of arrival of the associated information in the AC-4 bitstream.

AC-4 Table of Contents Element

The TOC bitstream element contains all of the information necessary to describe and enable immersive and personalized experiences including support for future extensibility. The *Presentation Info* element contains information describing one or more substreams to be decoded and presented simultaneously. Substreams can be thought of as single decodable units representing a specific channel, group of channels, a specific object or a group of objects. An important concept to note here is that a single substream may be shared across multiple presentations.

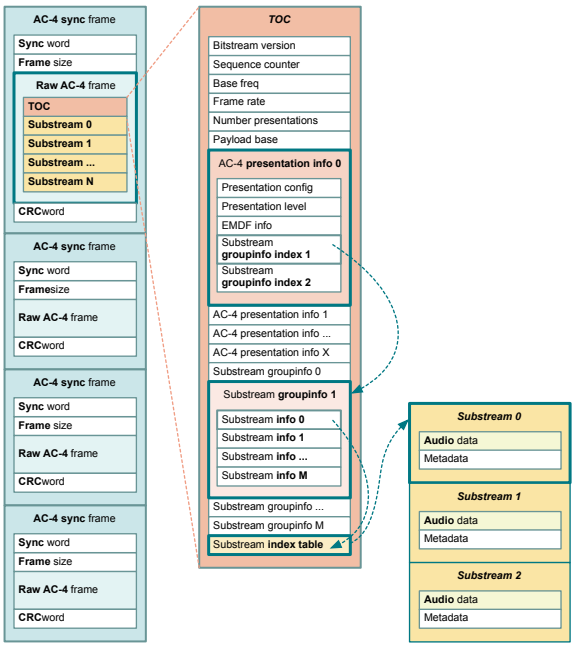


Figure 6: AC-4 Bitstream Syntax – TOC

AC-4 Presentation Information

The AC-4 Presentation Information element (AC-4 presentation info) contains the following sub-elements:

Presentation Config - Defines the configuration of the *Presentation* made up of a group of substreams (such as Music-and-Effects plus Dialogue, or Complete Main plus Associated Audio).

Presentation level - Defines the minimum level of the decoder that is able to decode the *Presentation*. For more information on the decoder levels refer to the AC-4 standard. [13]

EMDF info - Defines the properties of an Extensible Metadata Delivery Format (EMDF) substream associated with the *Presentation*.

Substream group info index – An integer that points to specific Substream Group Info fields dependent on the configuration of the *Presentation Config*.

AC-4 Substream Group Info

The *substream group info* element contains one or more substream info elements containing substream numbers that range from 0 to M. Each substream number points to an index in the substream index table. In the case of a single substream, as needed for a Complete Main channel-based configuration, only the first field in the list of substream info elements is used and the number contained in the substream info points to the respective substream via the substream index table. In the case of coded objects or object groups – for example, eight – the substream info list could be filled with eight substream info elements that point to the eight different objects represented by eight substreams. The substream index table is a table which maps substream numbers to the entry point for a specific substream in the AC-4 frame. All of the payload information such as audio essence and metadata is carried in substreams. When *Substreams* and *Presentations* are being described, each substream in an AC-4 bitstream is given a substream ID corresponding to an integer starting from 0. The first substream is referred to as S0, the second as S1, and so on. The first Presentation info element is referred to as P0, the second as P1, and so on.

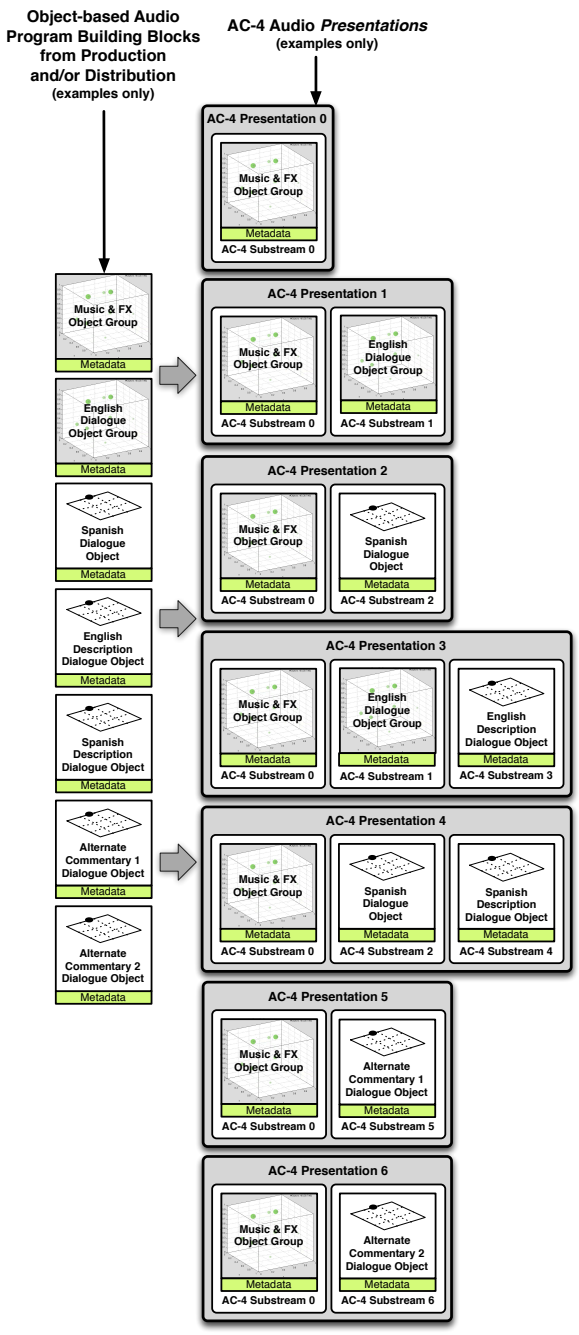


Figure 7 – Object-based building blocks mapped to AC-4 substreams & presentations (example only)

The object and object group building blocks - created during production - shown in Figure 7 are each mapped and carried in the AC-4 bitstream as an individual substream. These substreams are logically grouped into *Presentations* prior to being multiplexed for delivery. The *Presentations* expressed in Figure 7 represent a total of 7 unique *Presentations* carried within a single AC-4 bitstream. This approach provides more efficient delivery of alternate language and descriptive services (including alternate commentary) ALL in an immersive (3D) format. Note the use of the same Music & FX substream is shared across multiple presentations for optimum efficiency.

The example stream in Figure 7 provides the end user (consumer) with 7 unique experiences including:

1. Immersive (3D) Music & FX
2. Immersive (3D) English
3. Immersive (3D) Spanish
4. Immersive (3D) English with English Description
5. Immersive (3D) Spanish with Spanish Description
6. Immersive (3D) with Alternate Commentary 1
7. Immersive (3D) with Alternate Commentary 2

A typical broadcast bitrate for delivering all of the immersive experiences listed above would be <600kbps. In contrast, delivering the same experiences with deployed systems today would require pre-mixing all of the source elements on a presentation-by-presentation basis (prior to encoding) and simulcasting several independent audio elementary streams to viewers. This comes with an excessive cost in terms of total bitrate. (Which in this example is, ~ 2.7 Mbps – assuming 384kbps for each of the 7 presentations available – with the additional restriction of only providing 5.1-channel surround rather than immersive sound in 3D)

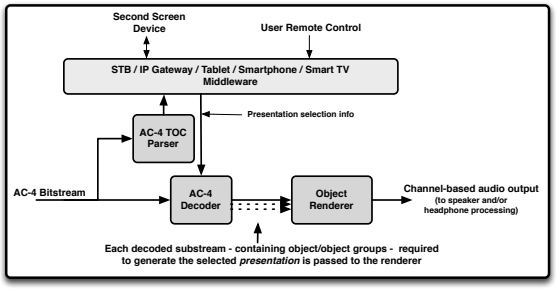


Figure 8 – Representative example of the primary processing blocks in an object-based audio subsystem of a playback device

Playback devices are required to support both traditional channel-based audio (stereo and multichannel) and object-based audio sources. A representative (simple) playback device is pictured in Figure 8. The TOC parser determines the number and type of *Presentations* available in the incoming AC-4 bitstream. It then extracts personalization metadata and decides which audio objects and/or object groups to simultaneously decode based on user selection or will automatically select the default presentation signaled in the bitstream. The playback device only decodes the substream(s) necessary for the selected presentation. The decoder produces a set of audio elements, object-audio metadata, and personalization metadata for use by the object rendering block.

Personalization metadata is further processed by the middleware in the playback device. The middleware presents the different personalization options to the user either via an on-screen user interface or by communicating to a second-screen (e.g. tablet) application. Either a remote control or touch-screen device can be used to customize the audio program. This customization can be as simple as selecting between the different audio presentations or as advanced as selecting the individual audio elements to be presented along with configuring the spatial position and level of each audio object or object group upon rendering. The level and range of audio personalization available is defined by the metadata authored by the content creator and/or distributor.

ADVANCED METADATA

Metadata has long been a fundamental component of the legacy AC-3 system that, for the first time, enabled scalable and consistent playback across the most common devices used at the time. However, as device capabilities, form factors, distribution paths and user applications have expanded over the past 20 years, the need for a more robust and richer set of metadata for this purpose has become increasingly apparent.

To address the device and application needs for today – with the capability to extend in the future - the AC-4 format supports an expanded and enriched set of metadata in many new categories including:

- **Presentation**: logical grouping and labeling of sub-stream/program elements to control playback mixing and rendering
- **Loudness**: ITU-R relative and speech-gated loudness information/control and regulatory information for automated adaptation of processing throughout the delivery and playback chain
- **Advanced dynamic range control**: multichannel and multi-band dynamic range information/control, including downmix and rendering loudness control
- **Dialogue enhancement:** support for both legacy (pre-mixed channels) and next-generation (music-and-effects plus dialogue) program sources
- **Spatial representation**: support for both channel-based and dynamic object-based programming
- **Dynamic rendering control**: speaker and headphone rendering control

- **Program/bitstream identification and synchronization**: for second-stream and second-screen applications
- **Interactive controls:** for personalization including language replacement, visual description and dialogue enhancement

All metadata generation and carriage within the AC-4 system takes advantage of a secure authentication mechanism that ensures robust and reliable delivery throughout the distribution and re-distribution pathways. This mechanism validates both the source and validity of the metadata as trusted. Metadata failing the authentication can be ignored or replaced by an authorized broadcaster.

To address future needs, metadata extensibility is critical to ensure that a low-friction pathway for distributing/delivering new metadata is always available. The Extensible Metadata Delivery Format (EMDF) syntax in AC-4 provides a structured and extensible container for a collision-free and open pathway for additional information (for example, third-party metadata, third-party application data, and so on) to be carried in AC-4 bitstreams and throughout the Dolby Audio System. This makes it easy for content creators and broadcasters to take advantage of this capability.

This following section provides an overview of the metadata parameters (and their application) essential for enabling next-generation immersive and personalized experiences utilizing the AC-4 format. Note: the metadata described in this section is not exhaustive and thus only focuses on the essentials for object-based rendering including personalization. The full set of metadata is described in ETSI TS 103 190 [13].

In general, the primary purpose of the object audio metadata is to:

- Describe the composition of the object-based audio program
- Deliver metadata describing how objects should be rendered
- Describe the properties of each object (for example, position, media classification, etc.)

Within the AC-4 format, a subset of the object audio metadata fields is essential to provide the best audio experience and to ensure that the original artistic intent is preserved. The remaining non-essential metadata fields are used for either an enhanced playback applications or aiding in the transmission and playback of the program content.

Essential Object Metadata for Immersive & Personalized Audio

Metadata critical to ensure proper rendering of objects and provide sufficient artistic control include:

- Object type / assignment
- Timing (timestamp)
- Object position
- Zone / elevation mask
- Object width
- Object snap
- Object divergence
- Number of presentations
- Default presentation
- Number of substreams per presentation
- Substream content type
- Language
- Gain

Object Type / Object Assignment - To properly render a set of objects, both the object type and object assignment of each object in the program must be known. For

spatial objects, two object types defined for current object-based audio production.

- Bed objects - This is an object with positional metadata that does not change over time and is described by a predefined speaker position. The object assignment for bed objects describe the intended playback speaker, for example, Left (L), Right (R), Center (C) … Right Rear Surround (Rrs) … Left Top Middle (Ltm).
- Dynamic objects - A dynamic object is an object with metadata that may vary over time, for example, position.

Timing (timestamp) - Object audio metadata can be thought of a series of metadata events at discrete times throughout a program. The timestamp indicates when a new metadata event takes effect. Each metadata event can have, for example, updates to the position, width, or zone metadata fields.

Object Position - The position of each dynamic object is specified using three-dimensional coordinates within a normalized, rectangular room. The position is required to render an object with a high degree of spatial accuracy.

Zone / Elevation mask - The zone and elevation mask metadata fields describe which speakers, either on the listener plane or height plane of the playback environment shall be enabled or disabled during rendering for a specific object. Each speaker in the playback environment can belong to either the screen, sides, backs or ceiling zones. The mask metadata instructs the renderer to ignore speakers belonging to a given zone for rendering. For instance, to perform a front to back panning motion, it might be desirable to disable speakers on the side wall. It might also be useful to limit the spread of a wide object to the two-dimensional surround plane by disabling the elevation zone mask. Otherwise, objects are spread uniformly in three-dimensional; and will fire ceiling speakers. Finally, masking the screen would let an overhead object be rendered only by surround speakers for configurations that do not comprise ceiling channels. As such, zone mask is a form of conditional rendering metadata.

Object Width - Object width specifies the amount of spread to be applied to an object. When applied, object width increases the number of speakers used to render a particular object and creates the impression of a spatially wide source as opposed to a point source. By default, object width is isotropic and three-dimensional unless zone masking metadata is used.

Object Snap - The object *snap* field instructs the renderer to be reproduced a single loudspeaker. When object snap is used, the loudspeaker chosen to reproduce the object is typically the one closest to the original position of the object. The snap functionality is used to maximize timbral accuracy during playback. A common usecase for the snap-to-speaker parameter is to create near-screen/wide pairs of objects along the side walls by using objects which snap to the proscenium speakers (e.g. in cinema exhibition). This is particularly useful for music elements, e.g. to extend the orchestra beyond the screen. When re-rendered to sparser speaker configurations (e.g., legacy 5.1 or 7.1), these elements will be automatically snapped to left/right screen channels. Another use of the snap metadata is to create ``virtual channels", for instance to re-position the outputs of legacy multichannel reverberation plug-ins in 3D.

Object Divergence - *Divergence* is a common mixing technique used in broadcast applications. It is typically used to spread a Center channel signal (for example, primary dialogue) across the speakers in the screen plane instead of direct rendering to the center

speaker. The spread of the Center channel signal can range from all center (full convergence), through equal level in Left, Right, and Center speakers, to full divergence where all the energy is in the Left and Right speakers with none in Center speaker. Regardless of how the center signal is spread, full convergence or full divergence, the spatial image of the center signal remains consistent. The object divergence field controls the amount of direct rendering of the object compared with the rendering of two virtual sources spaced equidistantly to the left and right of the original object using identical audio. At full convergence, the object is directly rendered as it would be normally. At full divergence the object is reproduced by rendering the two virtual sources.

Number of Presentations - The *number of presentations* parameter indicates the number of presentations represented in each AC-4 frame. This parameter can be updated on a frame basis for supporting dynamic insertion/deletion of substreams during distribution and/or within consumer playback devices.

Default Presentation - The *default presentation* parameter identifies which presentation shall be utilized in consumer playback devices that do not support personalization.

Number of Substreams per Presentation - The number of substreams per presentation parameter indicates the number of substreams (carrying audio essence and metadata) associated with each *presentation* indicated in the bitstream.

Substream Content Type - The *substream content type* parameter indicates the type of essence contained in the substream. e.g. dialogue, music, FX, etc.

Substream Language - The *substream language* parameter indicates the language the substream contains or is to be associated with during rendering playback.

Gain Value - The *gain value* parameter expresses a gain to be applied during rendering for normalization and/or when mixing with associated substreams.

Additional (Non-Essential) Object Metadata

Additional metadata critical to ensure proper rendering of objects and provide sufficient artistic control include:

- Object not active
- Object priority
- Screen scaling
- Object “is mutable”
- Minimum & Maximum gain values
- Positional constraints

Object Not Active - An object may not be “on” or active throughout the lifetime of a program. The object not active field indicates that the object’s audio essence is silent or should not be rendered. The object not active field can be used by encoders, decoders, and renderers to reduce processing complexity for objects that are inactive.

Object Priority - Each object has an associated ranking of importance, object priority. The priority is not an absolute ranking of objects but rather relative a normalized value. The higher the object priority value, the more important that object’s contribution to the content is. The object priority field can be used by encoders, decoders, and renderers to reduce processing complexity, bit rate allocation, or rendering quality for objects with lower priority in constrained systems.

Screen Scaling - In cinema auditoriums, the Left and Right speakers are almost always aligned with the left and right edges of the screen. However, in consumer playback environments the position of the Left and

Right speakers can vary greatly from being adjacent to the screen to being wider, sometimes much wider, than the screen. As a result audio and visual events that are co-located on the edge of the cinema screen may not be aligned on consumer screens. To address this problem, screen scaling metadata is needed to help a renderer scale the X- and Z-positions of objects that appear onscreen. In order to work correctly, several pieces of information are required. As such screen-scaling metadata encompasses several metadata fields including:

- Master screen size ratio - The master screen size ratio represents the ratio of the screen width to the distance between the Left and Right speakers in the mastering studio.

- Use Screen Reference - The Use Screen Reference field indicates that a renderer should use the screen size as the position reference rather than the room dimension (or speaker positions) when panning the associated object. The X- and Z-positions should be scaled according to playback screen size. This will ensure that visual and audio images are aligned.

Object "is mutable" - The *object is mutable* parameter indicates whether the audio object can be completely muted during playback – under creative control.

Minimum & Maximum Gain - The *minimum and maximum gain* parameter indicates (and limits) the amount of boost/cut a consumer is able to control (in terms of level) when rendering the associated object during playback.

Positional Constraints - The *positional constraint* parameter indicates (and limits) the range of spatial re-positioning a consumer is able to control when rendering the associated object during playback.

Intelligent Loudness Metadata

The following discussion highlights the essential loudness-related metadata parameters supported the AC-4 system. Intelligent Loudness metadata provides the foundation for enabling automatic (dynamic) bypass of cascaded (real-time or file-based) loudness and dynamic range processing commonly found throughout distribution and delivery today. Intelligent Loudness metadata is supported for both channel- and object-based audio representations.

Dialogue Normalization Level – This parameter indicates how far the average dialogue level is below 0 LKFS.

Dialogue Channel – This parameter indicates whether Left, Right and/or Center channels of the substream contain dialogue.

Loudness Practice Type - This parameter indicates which recommended practice was followed when the content was authored or corrected. For example, a value of "0x1" indicates the content author (or automated normalization process) was adhering to ATSC A/85. A value of "0x2" indicates the author was adhering to EBU R 128. A special value, "0x0" signifies that the loudness recommended practice type is not indicated.

Loudness Correction Dialogue Gating Flag - This parameter indicates whether or not dialogue gating was used when the content was authored or corrected.

Dialogue Gating Practice Type - This parameter indicates what dialogue gating practice was followed when the content was authored or corrected. This parameter is typically 0x02 – "Automated Left, Center and/or Right Channel(s)". However there are values for signaling manual selection of dialogue, as well as other channel combinations.

Loudness Correction Type - This parameter indicates whether a program was corrected using a file-based correction process, or a real-time loudness processor.

Program Loudness, Relative Gated - This parameter indicates the overall program loudness as per ITU-R BS.1770-3. In ATSC geographies, this parameter would typically be -24.0 LKFS, for short-form content as per ATSC A/85. In EBU geographies, this parameter would typically indicate -23.0 LKFS (LUFS).

Program Loudness, Speech Gated - This parameter indicates the speech-gated program loudness. In ATSC geographies, this parameter would typically be -24.0 LKFS for long-form content as per ATSC A/85.

max_loudstrm3s - This parameter indicates the maximum short-term loudness of the audio program measured according to ITU-R BS.1771.

max_truepk - This parameter indicates the maximum true peak value for the audio program measured according to ITU-R BS.1770.

loro_dmx_loud_corr - This parameter is used to calibrate the downmix loudness (if applicable), as per the Lo/Ro coefficients specified in the associated metadata and/or AC-4 bitstream, to match the original (source) program loudness.

ltrt_dmx_loud_corr - This parameter is used to calibrate the downmix loudness (if applicable), as per the Lt/Rt coefficients specified in the associated metadata and/or AC-4 bitstream, to match the original (source) program loudness.

Note regarding the loudness measurement of objects: The AC-4 system [13] supports loudness estimation and correction of both channel-based and object-based (immersive) programs utilizing an extension to the ITU-R BS.1770-3 recommendation (currently under study in ITU-R Working Party 6B). Moreover, the loudness estimation element of the system dynamically applies a spatial gain for each object or object group derived from each object's or spatial object-group's positional information carried in the program's metadata. Channel-based immersive formats are also supported. Additionally, the AC-4 system supports the carriage (and control) of program loudness at the presentation level across program interchange, distribution, and consumer delivery. This ensures any *Presentation* (constructed from one or more AC-4 substreams) available to the listener will maintain a consistent loudness.

CONCLUSION

There are few areas evolving as rapidly as consumer media. Changes in consumer behavior, market dynamics, the regulatory environment, and technology are unfolding faster than ever before.

This paper outlined the latest developments in the evolution of creating and delivering more lifelike audio experiences, flexible/efficient delivery of programming that addresses the needs of the non-native speaking, hearing-impaired and visually-impaired audiences and the most essential metadata parameters necessary to enable these experiences. Thereby meeting the needs of next generation broadcast by:

Making audio personal. Viewers will be able to tailor the experience to their preference by leveraging audio Presentations and compositional control provided by the AC-4 system. This will increase audience engagement and provide new commercial opportunities for programmers and MVPDs.

Making audio more immersive. The AC-4 system has been designed from the ground up to efficiently carry both channel-based and

object-based immersive programming. Leveraging the described object-based program building blocks and system components fosters continued innovation in device playback technologies and future-proofs content against the unpredictable evolution of playback device capabilities and environments.

ACKNOWLEDGEMENTS

The authors would like to thank the following individuals for their contributions to this and the entire body of research and development in the areas of immersive and personalized sound worldwide; Scott Norcross, Tim Carroll, Sean Richardson, Elmer Musser, Steve Silva, Roger Charlesworth, Freddie Sanchez, Sripal Mehta, Prin Boon, Jonas Roden, Mike Ward, Christophe Chabanne, and Aaron Master. We appreciate your support, dedication and shared vision for next generation audio.

REFERENCES

[1] The term "Spatial Audio" refers to audio processing and reproduction techniques that enable the perception of audio essence in space and thereby in context (e.g. with accompanying picture).

[2] C. Robinson, N. Tsingos, and S. Mehta, "Scalable format and tools to extend the possibilities of cinema audio," SMPTE Motion Imaging Journal, Nov. 2012.

[3] F. Rumsey, Spatial audio. Taylor & Francis US, 2001.

[4] R. K. Furness, "Ambisonics – an overview," in AE^{S} 8th International Conference, Washington, D.C., 1990.

[5] T. Lossius, P. Baltazar, and T. de la Hogue, "DBAP–distance-based amplitude panning," in Intl. Conf. on Computer Music (ICMC), Montreal, 2009.

[6] G. Dickins, M. Flax, A. McKeag, and D. McGrath, "Optimal 3D-speaker panning," Proceedings of the AES ^{1}6th international conference, Spatial sound reproduction, Rovaniemi, Finland, pp. 421–426, April 1999.

[7] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," J. of the Audio Engineering Society, vol. 45, no. 6, pp. 456–466, June 1997.

[8] D. Kostadinov, J. D. Reiss, and V. Mladenov, "Evaluation of distance based amplitude panning for spatial audio." in Proc. of ICASSP 2010, pp. 285–288, 2010,

[9] N. Tsingos, C. Q. Robinson, D. P. Darcy, and P. A. Crum, "Evaluation of panning algorithms for theatrical applications," Proceedings of the of panning algorithms for theatrical applications," Proceedings of th^{e} 2nd Intl. Conf. on Spatial Audio (ICSA), Erlangen, Germany, February 2014.

[10] ITU-R BS.2051. Advanced sound system for programme production. http://www.itu.int/rec/R-REC-BS.2051-0-201402-I

[11] C. Robinson, N. Tsingos, Cinematic Sound Scene Description and Rendering Control, SMPTE 2014 Annual Technical Conference

[12] J. Riedmiller, S. Mehta, N. Tsingos, P. Boon, "Immersive & Personalized Audio: A Practical System for Enabling Interchange, Distribution and Delivery of Next Generation Audio Experiences", SMPTE 2014 Annual Technical Conference

[13] ETSI TS 103 190, 'Digital Audio Compression (AC-4) Standard'

[14] Dolby Atmos Specifications http://www.dolby.com/us/en/technologies/dolby-atmos/dolby-atmos-specifications.pdf

[15] Dolby Atmos
http://www.dolby.com/us/en/technologies/dolby-atmos.html

REMOTE PON NETWORK PERFORMANCE

Edward Boyd
Tibit Communications

Kevin A. Noll
Time Warner Cable

Saifur Rahman & Nagesh Nandiraju
Comcast Cable

Fernando Villarruel
Cisco Systems

Abstract

Ethernet Passive Optical Network (EPON) systems have been successfully deployed worldwide for high-speed access networks and are enjoying a growing adoption by cable operators.

However, MSO access networks have challenges beyond the capabilities of traditional PON solutions. First, many operators have fewer trunk fibers to the node available for PON than required for a large number of subscribers. Second, Cable operators often find that their current access networks reach farther than the distances supported by traditional PON solutions. Greenfield deployments offer little relief because they are sometimes located far from the operator's existing head end or hub sites.

In an attempt to simplify maintenance, reduce operational costs, and improve reliability, many operators prefer a centralized network deployment model which reduces the number of head ends and hub sites and aggregates many customers over a large service area. Collapsing the head ends and hub sites in favor of a more centralized model naturally leads to longer distances between the network aggregation systems and the customers. This escalates the difficulties related to reaching those customers using traditional PON architectures.

Operators need a way to support a longer reach network that uses existing PON standards, technology, and systems. One approach to this problem can be gleaned from the network topology used in HFC architectures in use today. A remote NODE can provide an enclosure for an active PON device to extend the reach, minimize the fibers required to serve subscribers, and allow for large centralized systems.

This paper compares and contrasts several solutions for implementing a PON in a NODE-based architecture. The solutions examined here will include a simple PON extender, a remote optical line terminal (OLT), and a distributed PON MAC/PHY solution. Each solution is examined for its impact on trunk fiber efficiency, equipment cost, power, subscriber count, physical space, PON efficiency, physical layer performance, packet jitter, and packet delay. The operator requirements for the remote PON network are discussed to highlight the differences between these solutions and their impacts to the service opportunities.

THE REMOTE PON NETWORK

What is a remote PON network?

Ideally, PON Networks are fully passive without any active electronics in the outside plant with an OLT located in Headend or Hub Site. These Local PON networks can be serviced by large commercial OLT shelves in environmentally controlled offices.

Adding more headends or hub-sites to service FTTP deployments beyond the reach of the Local PON network is not cost effective. Additionally, air conditioned cabinets in the outside plant are expensive to build and maintain. While higher power optics can be used to increase the split ratio or the distance of a PON networks, these devices are not common and are therefore more expensive Longer Reach optics are also more difficult or impractical for higher speed PON like 10Gbps or future speeds beyond 10Gbps.

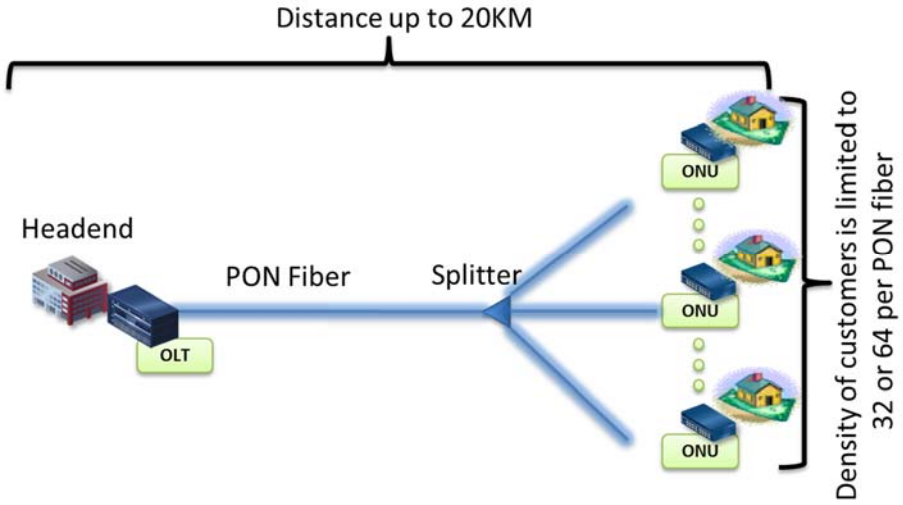


Figure 1: Local PON Network Limitations

Remote PON Networks cannot be serviced passively from an OLT in the Headend or Hub site. Remote PON Networks are required when the distance from the headend or hub site exceeds the PON specification or the density of customers requires more trunk fibers than available in the MSO network. Adding large numbers of trunk fiber can be prohibitively expensive for the operator. PON Systems have been designed for 1 to 32 or 1 to 64 split ratios at 20 kilometers using a fixed set of wavelengths. Since MSO networks have traditionally serviced longer loops and higher densities, the need for a remote PON is very common. For example, a high density MDU with fiber to the apartment would require many fibers to the MDU. A Remote PON originating at the MDU could have a higher split ratio and minimize the fiber required by the MSO.

A Remote PON Termination Device (RPTD) is needed to connect the Remote PON with the trunk network to the headend. The ideal Remote PON termination Device is hardened for the outside plant, small enough to fit in an amplifier or node housing, and uses the same standard PON optics for the Remote PON network as the Local PON networks. PON or WDM can be used on the trunk fiber to increase the density and increase the distance.

This paper will look at multiple solutions for the Remote PON Termination Device and compare them based on cost, efficiency, reliability, and network performance.

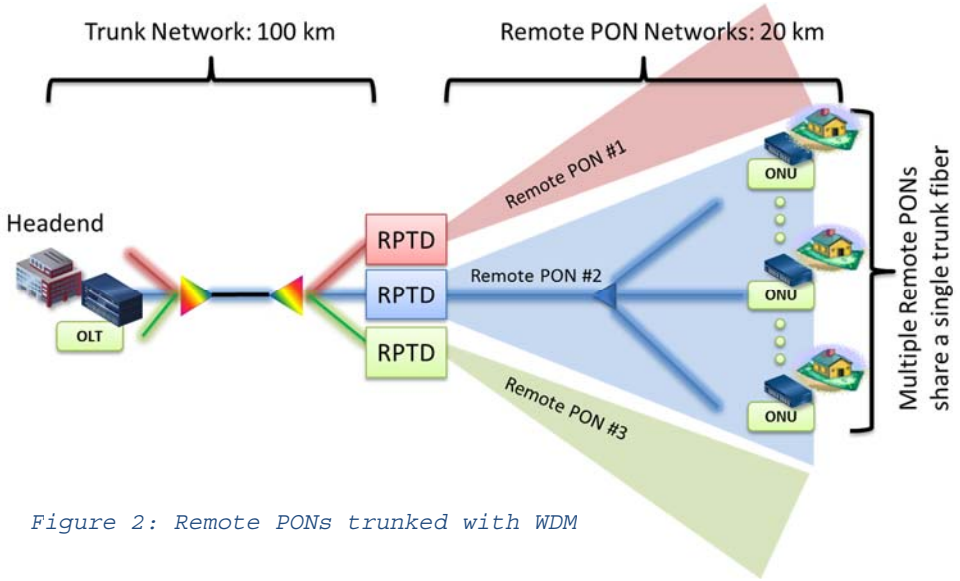


Figure 2: Remote PONs trunked with WDM

Figure 2 shows a typical application of the RPTD to service 3 distant Remote PON networks using 3 WDM wavelengths.

This paper will look at 3 topologies for backhauling the Remote PON networks. The three topologies are needed to support long distance, high density, or high bandwidth requirements.

Topology 1: PON-to-PON

A headend located PON can be used to feed a number of remote PON networks. This solution provides a very high density aggregation of subscribers. A 1:32 split ratio Trunk PON could support 32 1:32 Remote PONs. A single fiber from the hub site could support 1024 customers (32x32) in this configuration. If 1:64 split ratios could be supported on the Trunk PON or Remote PON, the subscriber count could be up to 2048 or 4096 customers from a single headend fiber.

Since this solution uses PON optics and supports 32 or 64 devices, it is lower cost and uses fewer fibers than the WDM Trunk

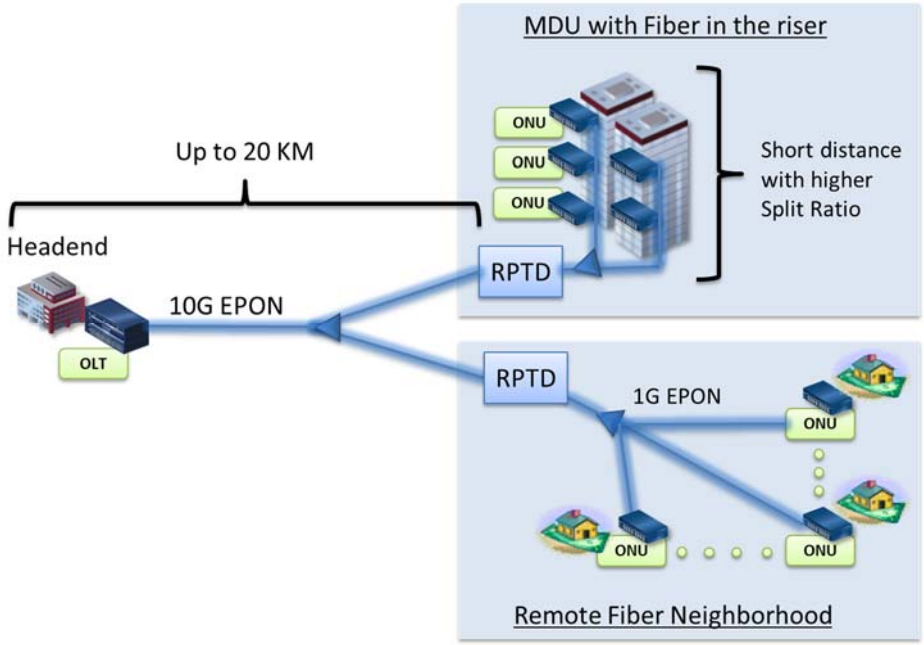


Figure 3: PON-to-PON Application

topologies. Of course, the WDM uplink topologies provide significantly more bandwidth.

The PON-to-PON topology could be a very cost effective solution for connecting MDUs with fiber to each apartment. A Remote PON terminating in the basement or outside the MDU would have shorter reach requirements so a 1:32 trunk to 1:64 remote PONs is possible. The operator could connect multiple buildings or MDU complexes to the trunk fiber and significantly reduce the number of fibers needed to service the subscribers.

The PON-to-PON could be used to aggregate multiple lower speed PON neighborhoods to a higher speed trunk PON. Low cost 1G ONUs would be used in the home while higher bandwidth 10G EPON is used in the trunk network.

In addition to the bandwidth limitations, the PON to PON topology has distance limitations compared to the WDM trunk topologies. Each PON is limited to 20km so the maximum reach of this solution is 40km if both PONs are at their full reaches. Because of the bandwidth distance limitation, a PON-to-PON topology would be more common used for high density solutions and less likely for long reach applications.

Topology 2: WDM-to- Single PON

Point-to-Point CWDM or DWDM optics can be used to provide a long reach and multiplexed trunk fiber to a Remote PON network. A pair of wavelengths for every PON would be appropriate when the distribution of the subscribers is beyond the reach of a Trunk PON or there is a need for more bandwidth to the Remote PON networks. The cost in the Trunk is increased due to the pair of WDM optics and OLT port required for each Remote PON network. While the PON-to-PON shares a 10G EPON uplink for a set of Remote PONs, the WDM-to-Single PON provides an 10Gbps uplink to each Remote PON. With high power WDM optics, it is possible to reach distances of 100 km. A single port device has the potential to be small and located close the subscribers so a larger split ratio (64 or greater) maybe possible.

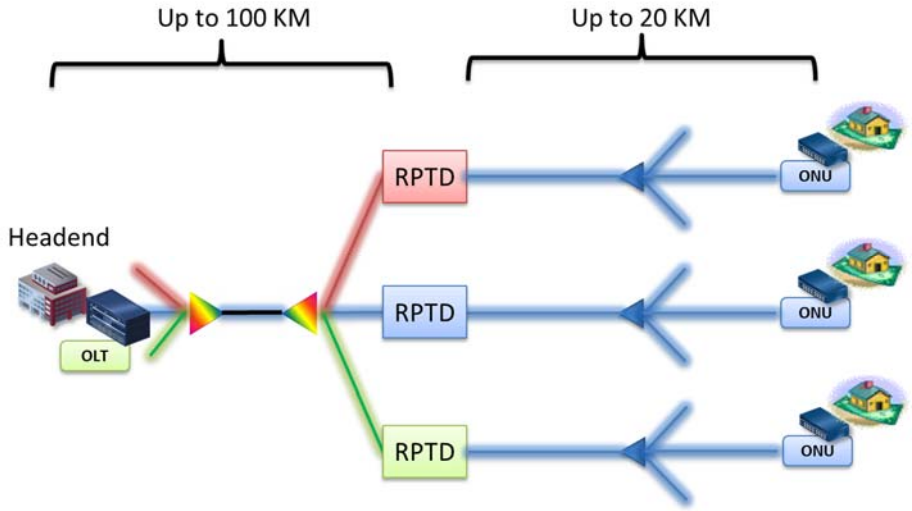


Figure 4: WDM-to-Single REMOTE PON

Topology 3: WDM-to-Multiple PONs

In many situations, the MSO will need to connect more than 32 or 64 subscribers at a remote location. In this case, a device that supports more than a single PON off of a single

wavelength pair would be beneficial. By supporting 2 to 8 PONs off a RPTD, an operator can significantly reduce the power, space, and cost of a single WDM wavelength and OLT port per Remote PON. The uplink from the RPTD could have one or more 10G Ethernet links or a higher speed 40G/100G Ethernet trunk connection.

The aggregation of Remote PONs onto higher speed or fewer Trunk links allows an operator to conserve the Trunk fiber and reduce the cost. The ideal solution would allow the operator to start with a single uplink and flexibly add uplink bandwidth from the RPTD to respond to higher customer take rate or higher bandwidth demands.

This type of solution can be used at a new housing community outside the reach of the Headend site. In some cases, a MSO may choose this type solution to reduce the space or eliminate equipment in small remote hubsites.

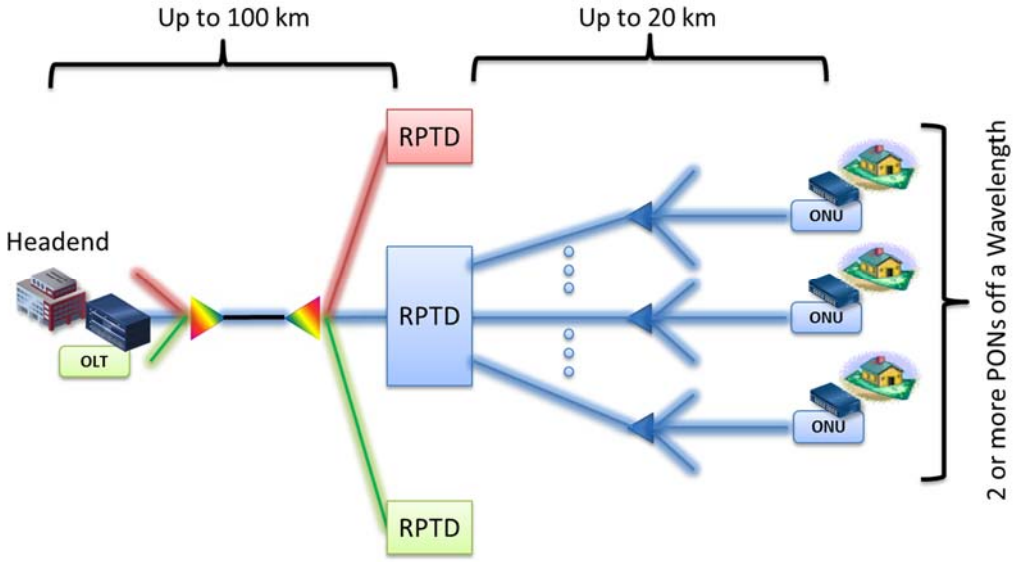


Figure 5: WDM to Multiple PONs

SOLUTIONS FOR THE REMOTE PON TERMINATION DEVICE

PON Repeater (PON Extender)

A PON Repeater or PON Extender is a physical layer device designed to use a single PON MAC layer across fiber networks. The purpose of the Repeater is to extend the distance beyond the typical 20Km reach and to conserve fiber on the trunk side. In general there have been two different basic approaches for PON Repeaters: OEO and packet repeater.

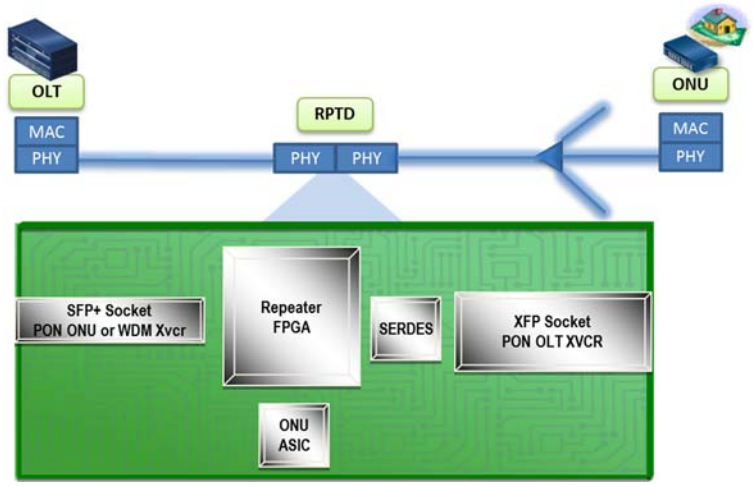


Figure 6: Anatomy of Repeater/PON Extender based RPTD

The OEO approach is a layer 1 device which converts optical signals into the electrical domain for amplification and then converts back to the optical domain at a different wavelength. While the downstream signal is easy to implement with this approach due to the fact that it is a continuous transmission, the bursted upstream is more problematic to implement since information from the MAC to identify slot boundaries is not available. In addition, supporting coexistence of different PON rates such as 1Gbps and 10Gbps poses another challenge. The OEO needs to properly discern the length of upstream burst and prevent false detection. On the trunk side, the bursted upstreams are combined into a continuous signal. To date there has not been a commercially available OEO Repeater that supports coexistence of different PON rates on the same port. The Repeater also does not allow data rates on the trunk side to be converted to a different rate on the remote side since it is intended to be a pass through device.

An alternative to the strict OEO layer 1 approach is to implement using a packet repeater where the entire PON packet is recovered and then a copy of the packet is regenerated. This approach is more complicated as the entire packet has to received and regenerated. This approach can add delays that pose potential issues with violating the PON delay and jitter requirements. With either of these approaches because the PON MAC is not terminated in the extender, the OLT DBA has to be tweaked to allow for the longer distances. MSOs need to work with the OLT provider to enable these functions as they are beyond the PON specifications. MSOs also need to be aware of upstream performance issues due to extended grant times for upstream transmissions. To minimize this performance impact, OLT vendors will need to make further changes to the DBA which can be a time consuming and costly endeavor.

Controlling the clock jitter and wander is another challenge with the Repeater solutions. Since EPON is loop-timed, the clock frequency of the all downstream and upstream signals in

and out of the Repeater must match the OLT downstream. The jitter and wander added by a long fiber on the trunk network can cause a violation to the tight jitter and wander requirements of the Remote PON network.

Figure 6 shows a block diagram of a possible PON Repeater solution. While some PON Repeaters only support Point-to-Point transceivers on the trunk side, this paper assumes that the PON Repeater can support either PON or Point-to-Point WDM transceivers on the trunk side. On the PON side, a burst SERDES and PON OLT transceiver are required. PON Repeaters often contain an ONU ASIC so the device can register with the OLT and provide some management capability. The functionality in the FPGA is based on the amount of PON physical layer included. An OEO device will have a smaller and simplier FPGA than a packet repeater that includes the Forward Error Correction and line decoding functions. In all cases, the FPGA must have a fixed delay through it, pass clocking, and provide a multiplexor path for the ONU ASIC to the Trunk network.

Remote OLT

Figure 7 shows a block diagram of a Remote OLT. The central office shelf has been shifted to the location of the Remote PON. The OLT MAC Layer, traffic management, switching, subscriber management, and other functions require a significant number of large ASICs and memories for a 10G PON system. The Headend equipment is significantly simplified to an Ethernet aggregation switch for the topologies with WDM on the trunk network.

The Remote OLT can overcome the performance and distance issues of the Repeater. However, creating a smaller and hardened OLT for the outside plant is a significant challenge. There are several challenges for putting an OLT in a node. The node has limited power, space, and is environmentaly challenging. Many of the components in a headend OLT are not designed for industrial (-40° to 100° C) or extended (-40° to 120° C) temperature range as forced air cooling is ready available and the expected temperature are much lower in the headend. Higher temperature components can increase the cost and limit the choices of

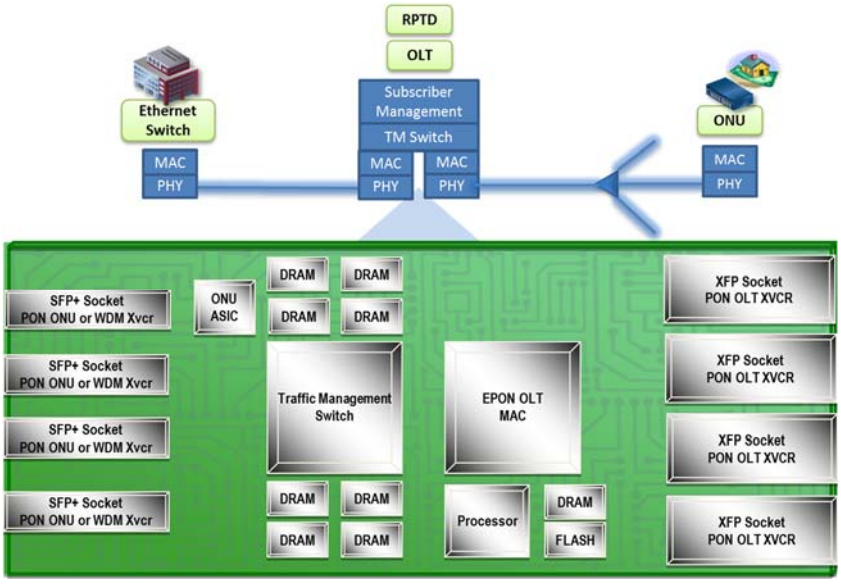


Figure 7: Anatomy of Remote OLT based RPTD

components. Even a scaled down OLT with just a handful of PON ports will have much higher power requirements than a Repeater. A typical node provides up to 140W of power and an OLT could consume most or all of the electrical power budget due to having the full functionality of OLT in the node. Space is also a precious resource inside a node housing especially one that is also hosting a traditional HFC node.

Even with the challenges noted, a Remote OLT has several important advantages. Due to the PON having to work with shorter loop distances, higher splits can be supported with low cost optics. The Remote OLT can improve IP layer performance as DBA can be tweaked to improve polling and the OLT can aggregate PON bandwidth into the trunk side.



Figure 8: Legacy OLT Functions

Distributed OLT

The Remote OLT's significant performance improvement is desireable but the power, space, and cost of the solution makes it impractical. The third option is a Distributed OLT Architecture. To describe this option it is necessary to understand the anatomy of a common OLT. An OLT is made up of the following functions: optical engine (including both forward lasers and return receivers), Media Access Control (MAC, enforcing synchronicity and end point access policies to the shared PON medium), layer 2 switch fabric (providing up and down link connectivity to the outside world, dissecting and combining data streams to and from multiple MAC endpoints), and management software (providing policy guidance for provisioning , control, and monitoring of the PON system). See `Figure 8`.

In a Distributed OLT each function of the OLT can be considered independently to address the needs of a particular topology or service mapping. Considering functions separately allows a network to be built and scaled around their natural strengths. See `Figure 9`. The optics have been considered separately in small form factor pluggables, so that is not new. The MAC, layer 2 switch, and controlling software on the other hand have been considered one box, with no visibility to their innate flexibilities.

The MAC for instance in the effort of creating an efficient box, which services as many PON ports as possible, would span many service groups at a time, which is not ideal for very granular applications of PON. The layer two switch embedded into a box leverages the low cost and flexibility of what is now a commodity part in a very competitive market. The control and management mechanism can reside either locally or remotely allowing flexibility of operations alongside parallel services that might be present.

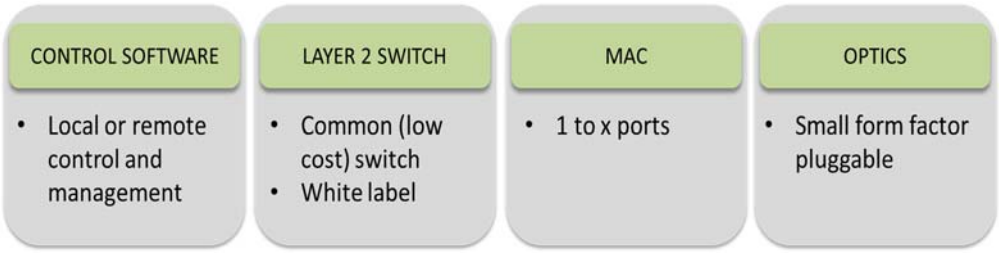


Figure 9: Modularized Functions of the OLT

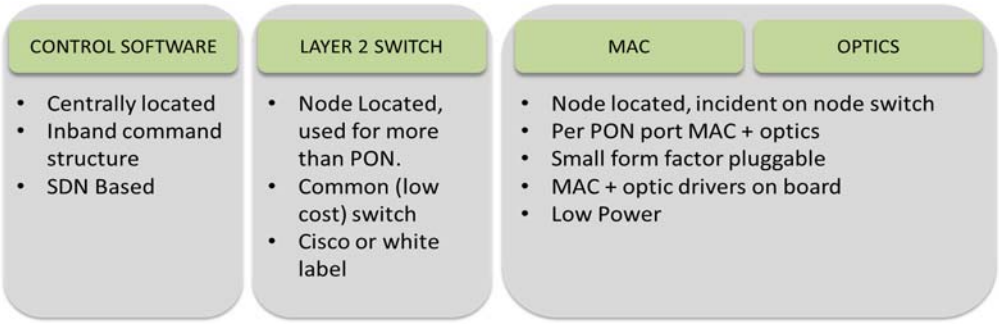


Figure 10: Distributed OLT, MAC and Optics Integration of OLT

For the cable space, which could favor a granular approach to PON deployments, one application of OLT modules is the combination of the MAC function and the optical engine. Allowing a one to one correspondence between a PON and its optical and MAC service. See Figure 9. In particular when taking into consideration the ubiquitous nature of the HFC node and the evolution towards nodes as platforms hosting various services beyond HFC it makes sense to have a PON option that can scale per-PON-port in the presence of layer two switching which might already be present due to other applications running through a node.

One notable item in the OLT structure (see Figure 10) is a centrally located PON software suite. This control module is no longer bound to the location of the OLT box, but can instead be a logical abstraction of the OLT at a central location. On one hand this fits particularly well in an evolution to an SDN based application of PON services. This approach needs a robust communication mechanism between the central command location and the remote switch and OLT.

It is then conceivable that a PON application specifically tailored for cable is just another orchestrated service, using Ethernet transmission and tagging, alongside the other services offered through the HFC node. These services can be applied via the insertion of a pluggable optic and MAC. This allows new PONs to be created as needed.

Distributed OLT Solution

The breakdown of the OLT allows for a smaller solution in the outside plant and a higher density solution with more functionality in the Headend. Figure 11 shows the anatomy of the solution using off-the-shelf PON components. The shelf processor, traffic management, and subscriber management has been removed from the Remote OLT to significantly reduce the size and power requirements.

DOCSIS Provisioning over EPON (DPoE) introduces the concept of "Virtual Cable Modems" for the management of ONUs. The ONUs in a DPoE system are simple devices with only layer 2 management. The DPoE software in the headend translates layer 3 management to a simpler layer 2 management for the ONU. This layer of simplification allows low cost ONUs used in the rest of world to work with a DOCSIS based backoffice similar to a Cable Modem. For a Distributed OLT, a "Virtual OLT" can be imagined to simplify the OLT requirements of the RPTD. The RPTD can be managed by layer 2 OAM and the majority of the OAM to ONUs can be passed through from the Headend to the ONU. Just like the messaging from the DPoE virtual cable modem to the ONU, a cable industry specification for the layer 2 OAM from the DPoE system virtual OLT to the OLT MAC in

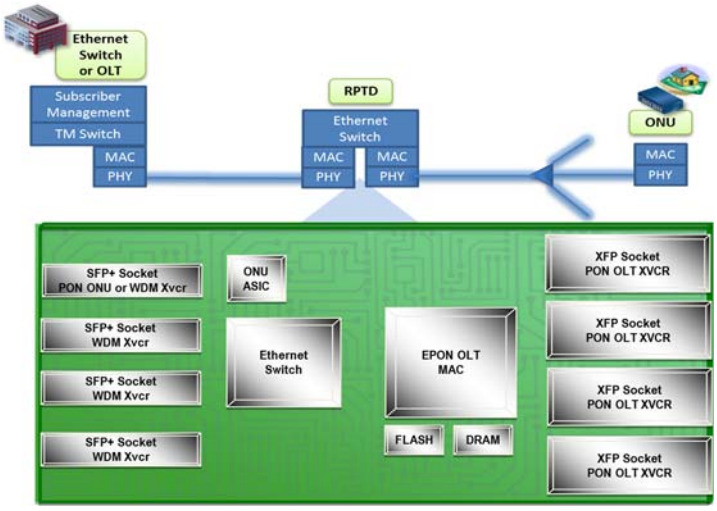


Figure 11: Anatomy of Distributed OLT based RPTD (fixed board)

RPTD would allow for interoperability and a clear definition of the required functionality.

To simplify the switching requirements, the downstream traffic could be shaped and and switched in VLANs for transport to the RPTD. The RPTD will not require buffering for shaping or complex switching decisions. A VLAN tag is removed and mapped to the downstream LLID. There is no need to pass the downstream SLA or switching rules to the RPTD. In the upstream direction, the Dynamic Bandwidth Allocation needs to be configured with the upstream SLA. The switching rules are passed directly to the ONU for the mapping of user traffic to the Logical Links (LLIDs). The RPTD can shape the upstream traffic based on the DBA and add a VLAN tag based on the upstream LLID. In the upstream, the SLA and LLID to VLAN mapping will be configured from OAM.

In the headend, the traffic management, switching, and subscriber management can be centralized in a system supporting many remote or local PON networks. The centralization and abstraction of these functions allows for greater flexibility, lower cost, and easy upgradeability. It would be much more difficult to upgrade or modify these functions if they were located in the outside plant RPTD. In addition, the cost is higher and reliability is lower for functions hosted in the outside plant. For this reason, all functions that can be virtualized or hosted in the headend should be done there. The functions in the RPTD should be limited to the translation of VLANs to PON links.

Looking forward, it is conceivable that a headend box could support remote access for point-to-point Ethernet, PON, wireless, DOCSIS, or other access technologies. It is also conceivable that this function could be in the cloud and further from the edge of the network. The management of the OLT MAC can be done in a management VLAN through many hops of Ethernet switches. A new shelf is not required to realize the Distributed OLT in the short term. An OLT shelf today could be designed with a blade of point-to-point Ethernet ports that either bypasses the PON OLT MAC or is designed without it. The OLT shelf or DPoE software would need modification to send the new virtual OLT OAM messages.

The Modular Distributed OLT

Since the Distributed OLT allows for a simplified interface from point-to-point Ethernet to Ethernet PON, it is possible to imagine a more densely integrated solution

with even greater flexibility. In the Modular Distributed OLT, the EPON OLT MAC function is simplified to a single channel and integrated into a pluggable module. The industry has already seen the integration of the EPON ONU MAC and GPON ONU MAC functions into SFP optics modules so it is reasonable to assume that the same integration would be possible with a simplified OLT MAC function. Ideally, the form factor of this pluggable would be SFP+ so it is interchangeable with the WDM Optics modules but other form factors such as XFP and QSFP could be benefical as well.

single WDM uplink and a single PON in addition to other configurations with many PONs and uplinks.

This approach has the benefits of the Remote OLT while not encumbering the solution with all the challenges of placing the full set of OLT functions in the node. Keeping the lower layer MAC with the optics in the node preserves the timing and optical link budget for the local loop while lowering the power and space impact of the upper layers (traffic management functions, device management and IP). The use of VLAN between the node and headend OLT preserves the function that each side of the

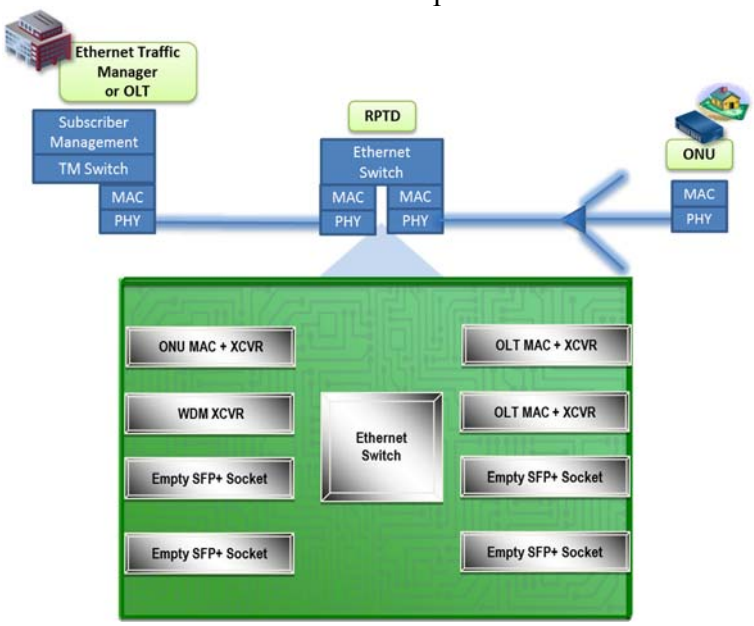


Figure 12: Anatomy of a Modular Distributed OLT based RPTD

The simplification of the PON OLT and ONU functions to pluggable modules allows a simple Ethernet Switch to become the only fixed onboard device for an RPTD. This allows for a low initial install cost and significant flexibility on the number of Remote PONs and uplink Trunks. The 8 port switch in Figure 12 could have any number of Remote PON OLT modules or WDM uplink modules until all 8 ports are filled.

The majority of the cost is in pluggable WDM or PON modules so it scales with the bandwidth and subscriber count. The 8 port switch shown will be cost effective with a distributed architecture performs while allowing the other end to independently function. For example, downstream traffic management can be done in the headend while still allowing the MAC to distribute the data on the appropriate flow to the ONU.

Having the more complex functions in the headend also allows better scaling of those components across multiple remote nodes. In the future, improved efficiencies via next generation silicon and/or alternative architectures such as network virtualization can be done in the headend while the

components in the node remain the same. This allows operators to leverage next gen technologies without the associated high touch upgrades in the node.

PERFORMANCE ANALYSIS

Systems for Analysis

The Remote OLT, Distributed OLT, and Modular Distributed are fundamentally the same for performance analysis. They move the EPON MAC Layer and EPON PHY Layer closer to the Remote PON. For simplicity, the analysis will refer to these flavors of RPTD solutions as the Distributed solution. Clearly the Remote OLT is not a Distributed solution but it will have the same data plane performance for delay, jitter, and efficiency since the location of the management, shaping, and switching doesn't change the performance on the PON network.

The EO Repeater and Packet Repeater solutions will have a small delay difference in the upstream. The Packet Repeater will have additional delay for the maximum size of a 1 Gbps burst slot. The Packet Repeater will also add additional inefficiencies in certain configurations. A detailed analysis of the Packet Repeater is beyond the scope of this paper. To simplify the analysis, both the EO Repeater and Packet Repeater will be the Repeater solution. It should be noted that the numbers correspond to the EO Repeater and the best case for Packet Repeaterand.

DBA Modifications

Many papers have been written on modifications to the Dynamic Bandwidth Allocation (DBA) algorithms of the OLT as a way to mitigate the performance impacts of longer reach and higher density PON networks. Some have suggested predictive algorithms for granting. While this method may decrease delay, it can dramatically decrease the efficiency. Other suggestions such as a Multi-Thread DBA [3] are generally good ideas for improving the performance and the delay in the PON upstream.

Modifications to the DBA algorithm are not considered here. The delay and efficiency issues explored in this paper are fundamental (i.e. minmum delay, MAC round trip time, efficiency of many bursts) to expanding the PON MAC over higher density or longer reach networks. The DBA for the system is a request and grant, weighted round robin, and assumes multi-threading (i.e. multiple grants in flight).

PON-to-PON High Density Deployment

The PON to PON topology can be cost effective for aggregating high density networks. While it allows for extending the reach from 20 km to 40 km, it is not expected to be used for the longer reach due to the distance limitation of 40 km. The PON to PON network is a low cost solution for aggregating many subsribers from MDUs or neighborhoods when trunk fiber is limited.

In some situations, it will make sense to connect multiple 1G EPONs to a single 10G Trunk EPON as a cost effective aggregation scheme. Since the Repeater is a physical layer construct, it is not possible to connect multiple 1G EPONs to a 10G EPON through a Repeater. 10G EPON has a different frame format for Encryption and Discovery so a single data rate conversion isn't possible. The Distributed model can support 10G to 1G EPON connection since the MAC layers are terminated.

Since both solutions can support 10G EPON Trunk PON and 10G EPON Remote PON, the analysis will be focused on this solution. The biggest difference between the Distributed and Repeater solution is the isolation of the PON overhead. In a Repeater Solution, all of the Remote PONs and the Trunk act as a single PON. The OLT at the trunk will need to request and grant to all of the ONUs and their Logical Links (LLIDs). The amount of PON

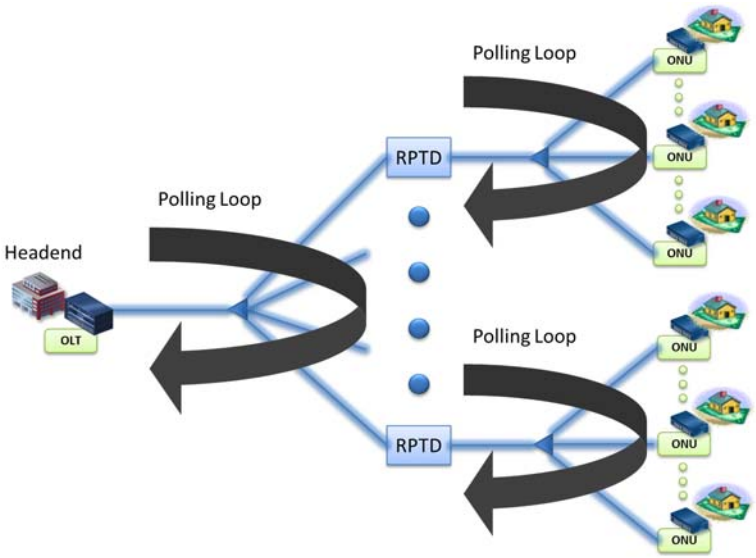


Figure 13: Parallel Polling in the Distributed Solution

overhead for requesting, granting, and burst overhead multiplies by the number of Remote PONs.

In the case of the Distributed Solution, the OLT at the trunk will request and grant to each of the Remote PON Termination Devices which in parallel will request and grant to their ONUs. The PON overhead is isolated to the PONs and doesn't increase as the number of Remote PONs increases.

Figure 13 shows how the polling (grant to report queue status) is isolated and performed in parallel for the Trunk and Remote PON networks in the Distributed Solution.

The isolation of the Remote PON networks significantly reduces all of the overhead for the PON.

The upstream burst overhead of 10G EPON is described in [2]. The upstream overhead for 128 ONUs or LLIDs is taken as a reference. A PON with 32 ONUs and 4 LLIDs each will act like 128 single LLID ONUs. For comparison, a 2ms DBA cycle time will be considered. With the Repeater solution, the upstream burst overhead increases with the number of Remote PONs. Table 1 shows the calculation of the overhead and bandwidth based on the size of the PON MAC domain.

Remote PONs	1	2	4	8	16	32
ONUs per Remote PON	32	32	32	32	32	32
Total ONUs	32	64	128	256	512	1024
LLIDs per ONUs	4	4	4	4	4	4
Cycle Time (ms)	2	2	2	2	2	2
FEC Encoding Overhead (%)	12.38	11.87	10.85	8.81	4.73	0
Codeword Quantization Overhead (%)	0.54	1.08	2.16	4.32	8.64	17.28
Time-Quantum Rounding Overhead (%	0.09	0.18	0.36	0.72	1.44	2.88
Idle Prefix Overhead (%)	0.08	0.16	0.32	0.64	1.28	2.56
Burst Mode Transmission (%)	3.48	6.96	13.92	27.84	55.68	111.36
Guard Band Overhead (%)	0.51	1.02	2.04	4.08	8.16	16.32
Control Overhead (%)	0.43	0.86	1.72	3.44	6.88	13.76
Net Transmission Overhead (%)	17.52	22.14	31.38	49.86	86.82	160.74
Net Throughput (Gbps)	8.248	7.786	6.862	5.014	1.318	0

Table 1: PON Overhead for large PON MAC domains

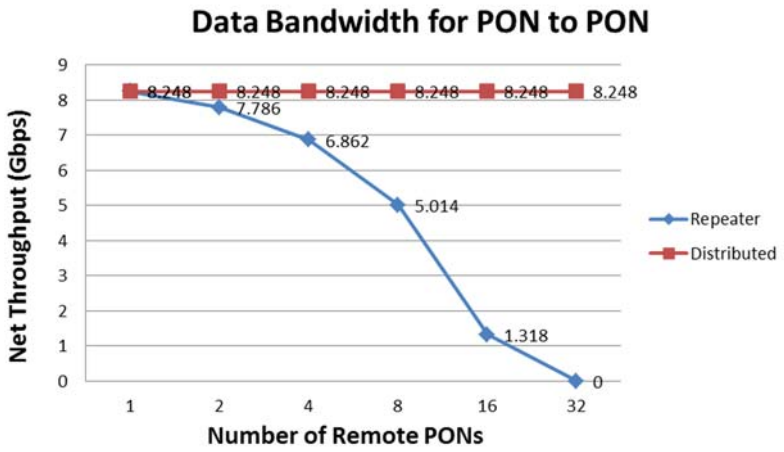


Figure 14: Data Bandwidth for PON-to-PON

With the Distributed Solution, the isolated PONs allow the overhead to be held constant. Figure 14 shows the effect on data throughput based on the number of Remote PONs, each with 32 ONU with 4 LLIDs per ONU.

The upstream burst overhead penalty is significant for the Repeater Solutions. 8 Remote PONs reduces the bandwidth to 5 Gbps and 32 Remote PONs would be consumed by the burst overhead.

The delay and jitter of the network will increase with the number of Remote PONs for both solutions. The PON-to-PON scheme is not the best solution for high bandwidth or low delay/jitter. In the case of the Repeater solution, the single scheduler allows for a low jitter and low delay output. In the case of the Distributed solution, the polling rate or cycle time can be reduced to reduce the delay and jitter.

The requirement of equal speed for the Trunk PON network and the Remote PON network along with the efficiency penalty make the Repeater solution impractical for a high density outside plant aggregation network.

WDM-to-Single PON

The WDM-to-Single PON topology has a single Remote PON with a WDM uplink to the Headend site. The cost of a pair of WDM optics for each Remote PON makes this topology more expensive than a PON-to-PON solution with a single OLT port and ONU optics modules. There is significantly more bandwidth and better delay/jitter in this topology. Since there is a single Remote PON for both Distributed and Repeater solutions, the efficiency is the same.

The biggest difference between the distributed PON solution and the PON repeater when using WDM to a single Remote PON is the upstream delay. In the distributed solution, the EPON upstream control frames are confined to the PON fiber network. In the repeater solution, the PON protocol frames must travel across the PON and the trunk fiber networks. Since the trunk fiber is a majority of the propagation delay, this has a significant impact to the EPON MAC and upstream delay. Figure 15 illustrates the difference between the two solutions.

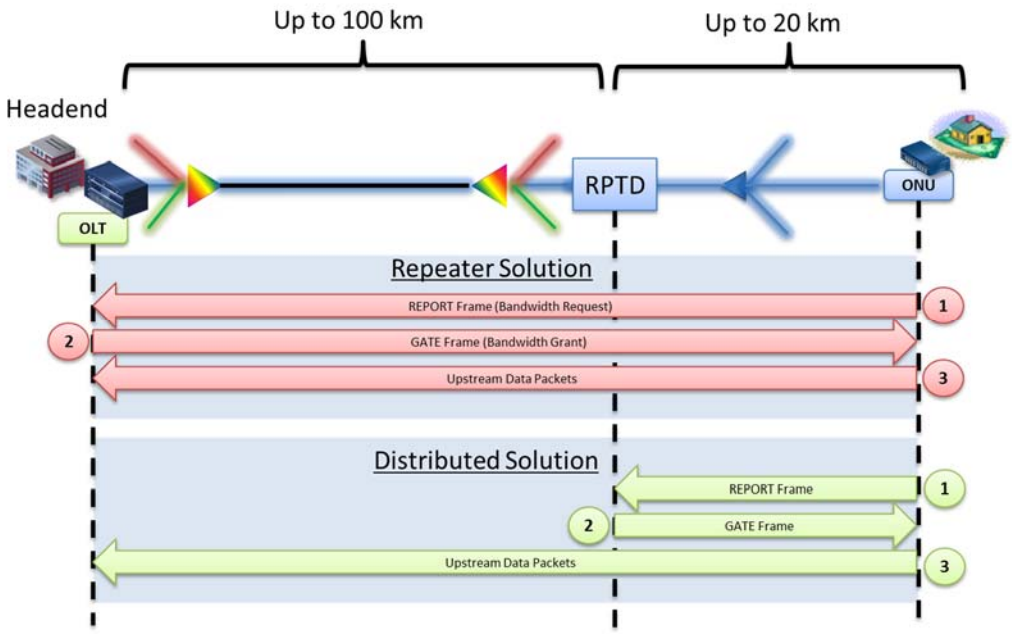


Figure 15: Upstream Data Flow in WDM to Single PON Topology

When an upstream frame arrives at an ONU, a REPORT frame will be sent upstream from the ONU MAC to the OLT MAC. The REPORT contains the bandwidth request for the upstream frame. The OLT MAC responds to the REPORT frame with a GATE granting the ONU an upstream slot for the data. The ONU transmits the data upstream to the OLT. In the Repeater solution, the GATE and REPORT frames must propagate across the trunk network and the PON network. In the case of the Distributed solution, the GATE and REPORT frames are limited to the PON network.

The IEEE 802.3 EPON PON standard is defined for a MAC layer Round Trip Time propagation delay of up to 1 ms. The major portion of the delay is in the fiber network. A 20 km link has a one way propagation dely of 100us so the round trip time is roughly 200us on the fiber. The OLT/ONU MAC and PHY layers normally require 50us. A 20 km EPON OLT MAC would normally see a 250 microsecond round trip propagation time. Since the Repeater solution increases the MAC layer propagation delay dramatically, the IEEE PON standard limit will be reached. At 75 km of trunk network, the IEEE limit will be reached and the Repeater solution is no longer viable. Since the Repeater will also add delay, a more practical limit is around 60 km. It should be noted that while the standard supports up to 1 millisecond of delay that the OLT devices are not required to support 1ms and many support much less. An OLT device that supports up to 500µs would be limited to a limited to 25 km of trunk fiber. Since the distributed solution doesn't extend the MAC layer across the trunk fiber, there is no round trip time impact. The distributed solution can work up to the limits of the trunk fiber transceivers and could go beyond those limits with multiple stages of transceivers.

In addition to the distance limit, the Repeater solution shows has a much greater minimum upstream delay over the Distributed solution. The downstream delay should be very similar or the same on the Repeater or Distributed system. The REPORT and GATE frame going over the trunk effectively triples the transmit delay over the trunk network. The minimum upstream delay for a 10 km Remote PON as a function of the trunk fiber length is shown in .

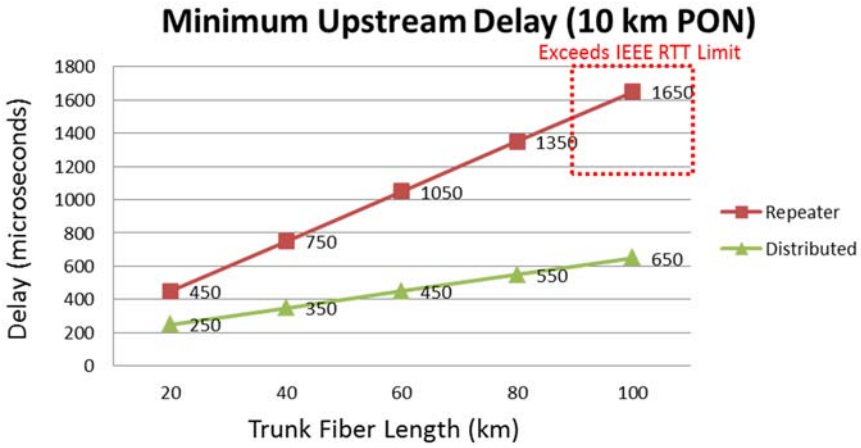


Figure 16: WDM to Single 10km Remote PON Minimum Delay

The Repeater shows a much higher delay over the Distributed solution. The delay for the Repeater is almost triple (650µs versus 1.65ms) and 1 ms greater at 100 km. With a 20 km Remote PON, the results (see Figure 17) show a similar difference in upstream delay. Of course, the Repeater is limited in distance by the IEEE round trip of 1ms so the longest trunk fiber distance may not be possible.

The additional delay for the Repeater solution also has an effect on the maximum upstream data rate for the Repeater solution. The EPON ONU must buffer data while it waits for the response from the OLT. In a 10G EPON Since ONUs have a fixed amount of buffer, an increase often results in a drop in maximum data rate upstream. For example, a 20 km local PON or will require buffer for 250µs of OLT response time. In the Repeater with a long trunk fiber, the OLT response time could increase 4 times. The ONU must buffer more data and burst in large blocks to handle the delay. In the case of the Distributed Solution, the delay from the OLT MAC to the ONU MAC is the same as a local PON. There is no need for additional ONU buffering on the Distributed solution.

Figure 18 shows the additional ONU buffer in kilobytes as a function on Trunk fiber length to transmit upstream at the same peak data rate. The Repeater will require an additional megabyte of buffer for an 80 km trunk fiber while the Distributed solution doesn't require any additional buffering to keep the same

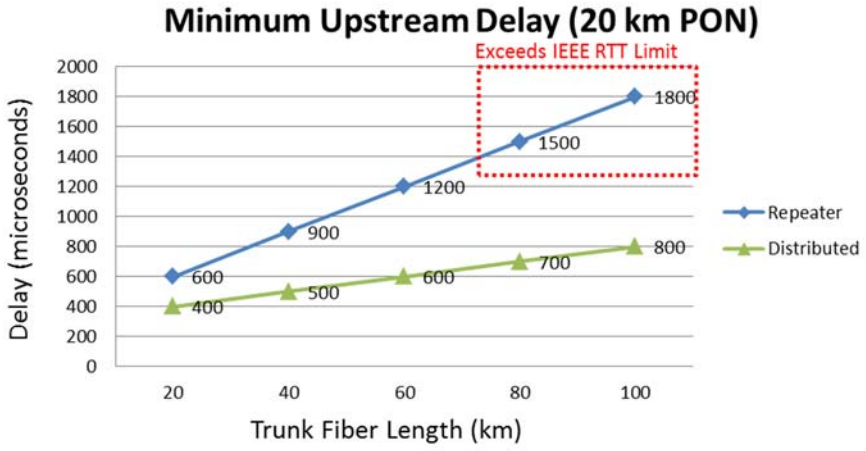


Figure 17: WDM to Single 20km Remote PON Minimum Delay

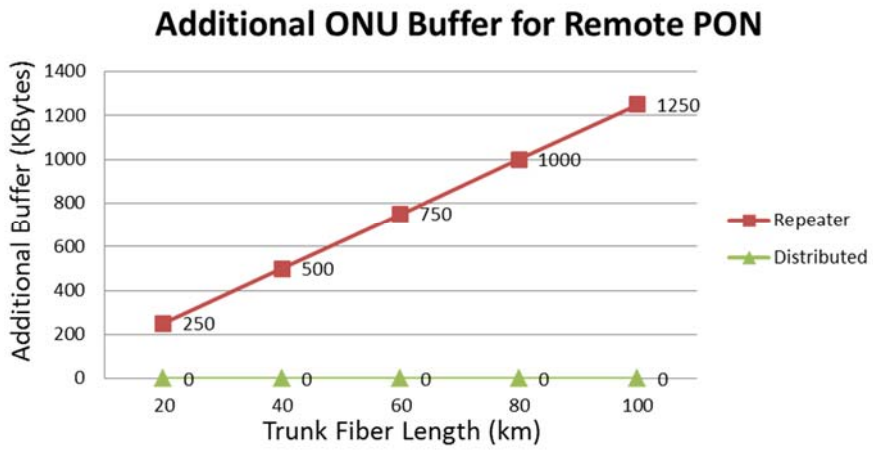


Figure 18: Additional 10G ONU Buffer to support Remote PON

maximum data rate. The buffer size for all upstream LLIDs should be increased to accumodate the additional delay. ONUs on a Repeater based Remote PON should be configured with a larger buffer than those on a Local PON. Of course, the data buffering in an ONU is a significant part of the cost and most ONUs are already configured to use all of the available buffering. If the buffering can't be increased, the maximum data rate will drop. Since most applications do not require the maximum upstream data rate of 8.7Gbps for 10G EPON, many ONUs only have buffer to support a lower maximum data rate for a single ONU.

In Figure 19, the maximum upstream data rate for an ONU is shown as a function of Trunk fiber length. A 4 Gbps maximum rate ONU (or single LLID) and an 8.7 Gbps ONU (or single LLID) are considered in this graph. Since the Distributed Solution doesn't increase the OLT to ONU MAC delay, it doesn't show any change in delay as a function of the trunk fiber length. The Repeater based solution will decrease the maximum upstream burst rate by up to 50% depending on the length of the trunk fiber.

For this reason, ONUs on a Repeater Remote PON will require either a different configuration or will support limited services compared to ONUs on a local PON.

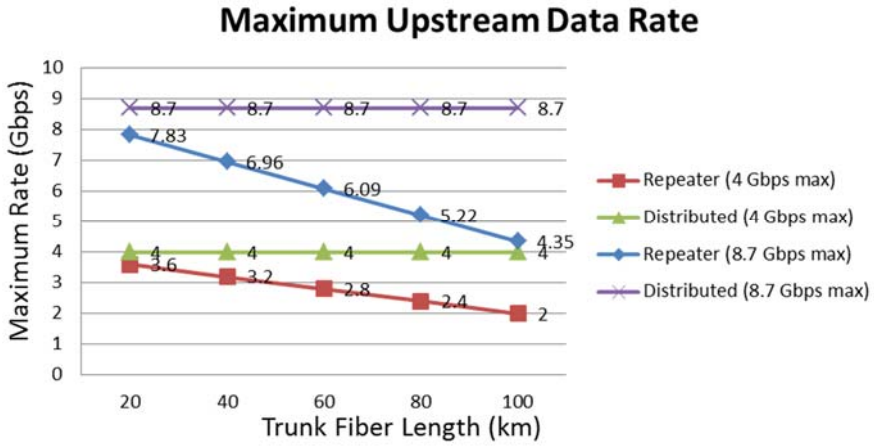


Figure 19: WDM to Single PON Maximum Upstream Data Rate

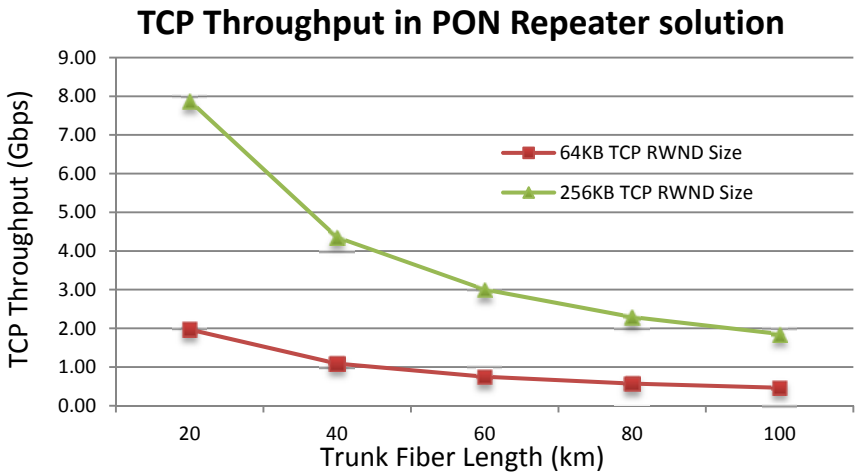


Figure 20: Repeater TCP Throughput Penalty

In the PON repeater architecture, due to the additional propagation delay that is added to the request and grant time, there is an impact to the TCP throughput in the downstream. shows the TCP downstream throughput of an ONU with two different TCP Receive Window sizes – 64k and 256k. As can be noticed from `Figure 20`, with the 20km trunk fiber distance (ONU to OLT distance), each 10G ONU can receive a theoretical TCP throughput of about 8 Gbps. Clearly with the increase in the trunk fiber distance between the PON repeater and OLT, the TCP throughput degrades from 8 Gbps to sub 2Gbps. The degradation is less dramatic when TCP receive window size is 64k but the overall TCP throughput is lower. TCP stack in older operating systems used to only support 64k receive window size. But most modern operating systems (Windows 7, 8, Mac OS-X) use window scaling which let TCP use higher than 64k receive window size.

WDM-to-Multiple PON

The WDM to multiple PON architecture allows a single wavelength uplink to connect to multiple PONs. A single Remote PON Termination Device may have multiple WDM uplinks with multiple Remote PONs each. Using a single WDM transceiver to service multiple Remote PONs reduces the cost, power, and size for networks requiring higher density than the WDM-to-Single PON topology.

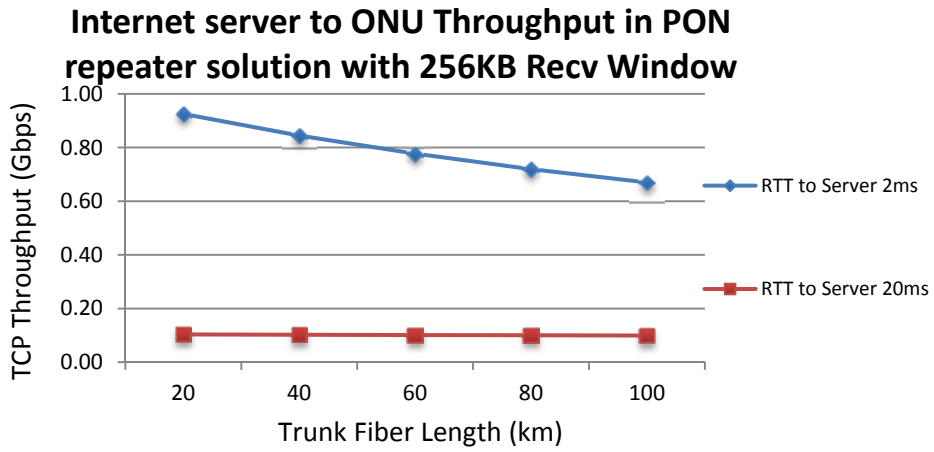


Figure 21: Internet Server to ONU Throughput

In the case of the Repeater, the multiple Remote PONs act as a single PON to the uplink wavelength. The FPGA in the RPTD acts a splitter. This solution increases the number ONU/LLIDs on a single OLT like the PON-to-PON solution. As shown in the PON-to-PON analysis, the combining of multiple PONs on a single MAC severely impacts the data throughput. The WDM-to-Multiple PON topology has the same long reach and MAC delay as the WDM-to-Single PON topology. For the Repeater solution, the WDM-to-Multiple PON has the lower efficiency of the PON-to-PON with the delay/jitter penalty of the WDM-to-Single PON.

In the case of the Distributed Solution, the Remote PON networks are isolated to avoid the upstream overhead penalty. The OLT MAC layer is located at the Remote PON so the ONU buffering, discovery size, and delay is the same as a Local PON Network. There is no need for special ONUs and there is no distance limit for the trunk network.

Figure 22 shows the Distributed Modular RPTD for the WDM-to-Multiple PON. The RPTD is an Ethernet Switch with sockets for OLT, ONU, or WDM transceiver modules. The operator can specify the RPTD by total port count of the Ethernet switch and then choose the number of OLT PON modules or WDM uplink transceivers based on the service requirements. The Remote PON networks will have the same performance as the local PON network plus the one way delay on the trunk fiber. In the upstream direction, the RPTD may see congestion. It is not desireable to have a switch with large external memories that requires configuration of SLAs. To prevent overflow, the switch should provide a priority based flow control or other mechanism to meter the traffic from the modules. There are multiple solutions to this issue but it is beyond the scope of this paper.

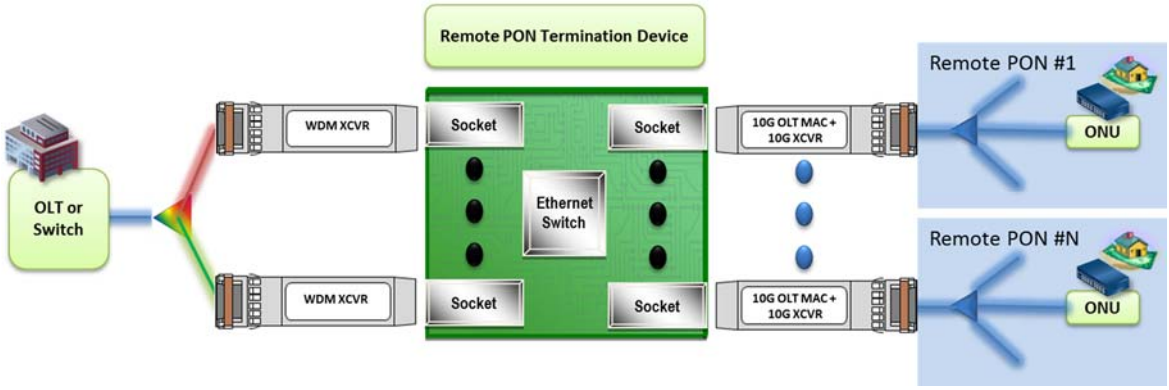


Figure 22: The Modular Distributed OLT solution for WDM-to-Multiple PON

CONCLUSIONS

Remote PONs can be very beneficial to MSOs. They allow the operator to run fiber with very high efficiency or connect customers at very long distances. An operator can reach customers without adding hubsites and could potentially consolidate hubsites. A Remote OLT is not always practical due to the complexity that it places in the node.

We introduced the concept of a Distributed OLT architecture that places the EPON MAC in the outside plant and centralizes the other OLT functions in the Headend. The Modular Distributed OLT provides a more flexible version of the Distibuted OLT for the RPTD. Multiple topology choices can provide high bandwidth, low delay, long reach, or high density.

Fundamentally, the Remote PON Termination Device can be designed as a remotely located EPON Physical Layer (e.g. PON Extender, EO Repeater, or Packet Repeater) or an EPON Media Access Control Layer (e.g. Remote OLT, Distributed OLT, or Distributed Modular OLT). The Repeater solutions with the Remote PHY expand the EPON MAC layer over a larger number of users or a much longer distance. The expanded MAC layer in the Repeater solution limits the distance, lowers the efficiency, and significantly increases the delay. The remote MAC Layer solution has the same distance and user count as a Local PON network so it can use the same ONUs with equivalent performance as a Local PON network.

Table 2 highlights the differences between the possible solutions for the Remote PON Termination Device. Based on these results, it makes sense for the industry to study and standardized a Distributed OLT architecture and Modular Remote PON Termination Device.

REFERENCES

[1] Glen Kramer, Ethernet Passive Optical Networks, McGraw-Hill, 2005

[2] Rajesh Roy, Glen Kramer, Marek Hajduczenia, Henrique J. Silva, “Performance of 10G-EPON”, IEEE Communications, November 2011

[3] Bjorn Skubic, Jiajia Chen, Jawwad Ahmed, Biao Chen, Lena Wosinska, Biswanath Mukherjee, “Dynamic Bandwidth Allocation for Long-Reach PON: Overcoming Performance Degradation”, IEEE Communications, November 2010

	FPGA PON Repeater (PON Extender)	Remote OLT	Distributed (Modular) OLT
Power	Low	High	Low
Size	Small	Large	Small
Relative Cost	Low	High	Low
Total Reach (Trunk Fiber Limit)	60 Km [or less]	100 Km+	100 Km+
Impacts PON Timing Budget	Yes	No	No
Require Larger ONU Buffer	Yes	No	No
Supports Any Ethernet Speed Uplink	No	No	Yes
Supports High Density of ONUs	No	Yes	Yes

Table 2: RPTD Solution Summary

SDN As A Matchmaker For Remote Phy Architecture

Alon Bernstein, Sangeeta Ramakrishnan
Cisco Systems

ABSTRACT

The Remote Phy (RPHY) architecture facilitates the separation of MAC (Media Access Control) resources from Phy (Physical layer) resources. The follow-up question is how do we pair them up?
The most straightforward method is to view the Remote Phy Device (RPD) in the RPHY architecture as a fixed "satellite" of the MAC and use static configuration for pairing them up. But this would only be a starting point. This paper will outline how a dynamic and flexible separation of these resources facilitates new options for OPEX savings as well as new load sharing and availability models.
The paper will also outline an SDN architecture to guide the implementation of these new capabilities.

OVERVIEW

The traditional view of RPHY is that it's satellite architecture. This view sees the RPD as an extension of a CCAP core (not an independent entity) with a static association of the RPD to CCAP core.

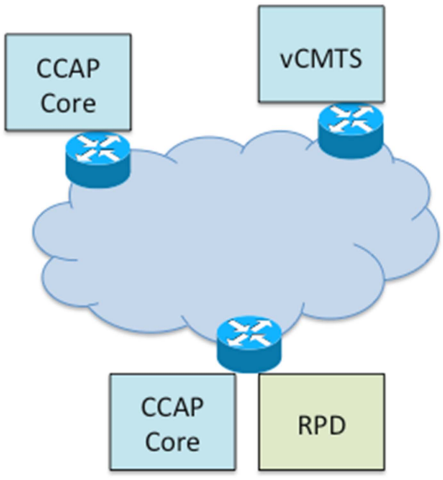


Figure 1 Dynamic association on MAC to RPD

Our point of reference is that we have a pool of Phy resources and a pool of DOCSIS MAC resources that can be paired up in any way and based on any kind of policy. Figure 1 shows a number of resources distributed at various sites in the operator network. The RPD refers to a Remote PHY node or a Remote PHY shelf. (Throughout the paper we will continue to use the term RPD to apply generally to either a RPHY node or a RPHY shelf). The CCAP-core refers to an entity that contains all CCAP functionality but the PHY as defined in the Remote Phy Architecture Document (see ref[4]). The vCMTS refers to a virtualized version of CCAP-core where the MAC functions, including scheduling and MAC level processing, have been virtualized. As shown in Figure 1 the RPD can be paired with a MAC port on the co-located CCAP core but it can also be paired with a CCAP core that is remote or with a virtual CMTS (vCMTS) in a remote or local data center. Because the MAC can be implemented on either a CCAP or vCMTS we will refer to the term "MAC" throughout this document as the resource that implements the DOCSIS MAC functions) instead of referring to a CCAP core/CMTS/vCMTS.
Since various instances of MAC functions may be available and may reside in different locations, a method to determine how to pair a RPD with a MAC resource is required.
The most obvious pairing is distance based, i.e. connect the RPD to the nearest MAC resource. As this paper will demonstrate there are other options and choices that can be made. The following sections will outline a few use cases for an intelligent mapping approach.
The actual "pairing" itself is done by means of controlling the L2TPv3 tunnel endpoints that are used for the DEPI and UEPI protocols.

PHYSICAL CMTS AND VCMTS

From an RPHY architecture perspective the vCMTS is a MAC resource just like the physical CCAP and the choice between the two is a policy-based decision. For example the physical CMTS ports can be paired first and only if the core runs out of ports then vCMTSs can be spun up and dynamically connected to the RPD.

LOAD BALANCING

A common issue with capacity planning is that enough equipment has to be installed to support a worst-case or near worst-case traffic load. With flexible mapping of resources its possible to optimize the network resource usage.

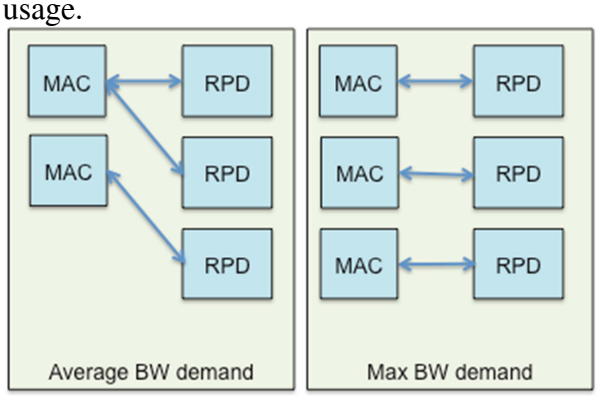


Figure 2 Balancing load

Figure 2 depicts a case where a single MAC resource can handle the load of two RPD during average traffic loads, but only one RPD when traffic peaks. A new MAC can either be a vCMTS that is spun to accommodate the load or a centralized physical CCAP that is designed to handle the overflow traffic from several remote nodes.

AVAILABILITY

The RPD can be connected to an active and a standby MAC resource. While this is a standard mode of operation, the concept of "matchmaking" can be used to have the redundant port in a different box than the active one. In fact, the redundant instance may be intentionally placed in a different data center to allow for geo-redundancy. See ref [2] for a general discussion of redundancy for NFV that is applicable to RPD as well.

PLACEMENT MODULE

The location of the RPD in relation to the MAC is an important piece of information when deciding how to pair the two. There are several methods that can be used to identify the location of MAC and RPD:

1. Pre-configuration: as part of the installation process of a CCAP device or an RPD its location is stored in a database. In the case of vCMTS the location of the data center is sufficient.
2. DHCP servers usually assign IP address pools based on physical location. Inspecting the IP address assigned can be mapped to a physical location.
3. Both (1) and (2) can be combined and cases of mismatch can be flagged. This will provide an extra layer of assurance that the mapping is correct.

Once the location of the endpoints is known the placement module can use the location information in combination with other data and policies to decide which RPD and MAC to pair. As mentioned in the use cases above the additional data can be the role of the MAC port (active/standby), bandwidth load and more.

LOCATION OF SERVICES

Placement of MAC and RPD resources need to take into account the location of other services such as Deep Packet Inspection engines, carrier grade NAT and more. The reason is that we don't want to create "tromboning" of data through the network. Consider the following example with a caching service:

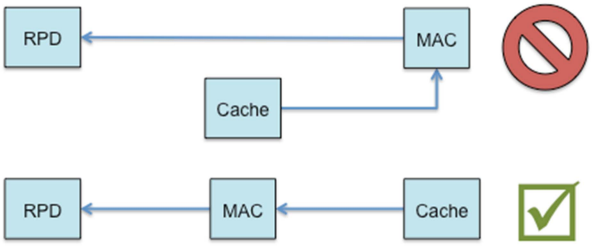


Figure 3 Caching service example

If the cache is placed near the RPD and the MAC is remote (e.g. a distributed cache and a centralized CCAP core) then traffic from the cache will flow all the way up to the MAC only to go down to the RPD as depicted in the top of `Figure 3`. Clearly the placement of the MAC functions needs to take into account the caching service and placed after the cache (the bottom of `Figure 3`).

TRAFFIC ENGINEERING

When the RPHY and the MAC are co-located there is little need for traffic engineering (TE). When the two are remote its critical that the path between them is engineered to have minimal delay and jitter.
Traffic engineering is well understood and the same engineering that is used for other use cases can be used for engineering the path between the MAC and the RPD.
Over the years traffic engineering has been perceived as relatively complex and not easy to manage. Fortunately the introduction of new TE methods (such as Segment Routing - see ref [3] for a discussion of use cases) in combination with SDN greatly simplifies and automates TE creation.

SDN ARCHITECTURE

Previous sections outlined the main components and functions needed for matchmaking: RPD, MAC, Placement, services and TE. The next step is to bind them all together in an SDN framework

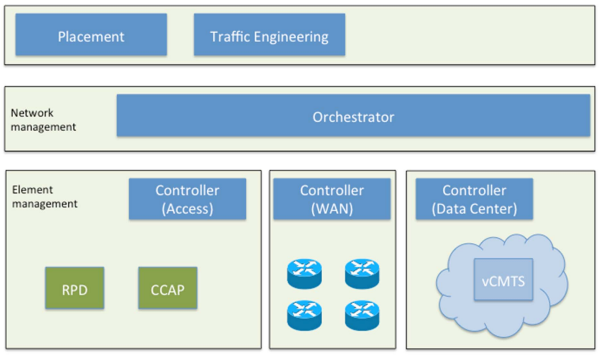


Figure 4 SDN framework for matchmaking RPD and MAC

As depicted in Figure 4 we can fit the placement and TE modules on top of a general purpose SDN/orchestration framework. Furthermore, even the placement and TE modules are fairly generic and can apply to many other use cases.
Let's explore this base architecture and how it relates to other use cases that require placement and TE.
A typical network is segmented into domains. These domains are typically segmented on organizational as well technological boundaries (for example the access, WAN and data center domains in Figure 4).
While this division to domains helps in "dividing and conquering" the complexity of a service provider network it leads to siloes and lack of simple end-to-end provisioning. The orchestrator, or as some prefer to call it the "controller of controllers" solves this problem. It can manage a complex distributed transaction across all these domains in order to provide end-to-end functions. Note that even the domains themselves may be fairly complex networks and might require their internal orchestration, but for the sake of simplicity we assume a simple two level hierarchy. The orchestrator takes upon itself the traditional TMN (Telecommunications Management Network, see ref [1]) role of "network management" and the controllers assume the role of element managers. This can be confusing as sometimes controllers are branded differently for different domains, but they might all still be built from the same building blocks and similar concepts. Its also

worth noting that while the "purist" SDN view as represented by OpenFlow may be relevant in the data center, it is less so in other areas of the network and irrelevant to pairing the MAC and RPD since OpenFlow is very limited to creating network paths.

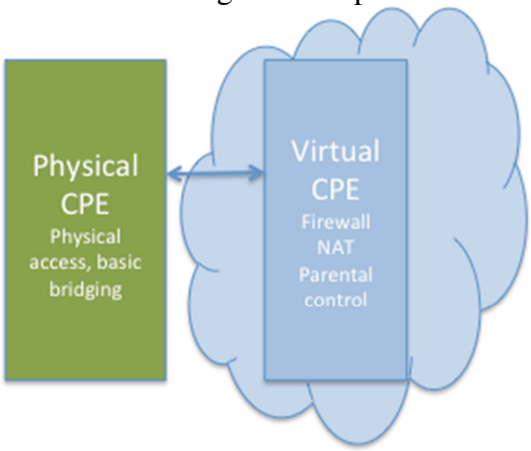


Figure 5 Virtual CPE use case

What other use cases can use a similar infrastructure? Consider the case of a virtual CPE. In a nutshell a vCPE refers to a system where the customer has a very simple physical CPE in the premise and all the advanced services (e.g. NAT, parental controls, firewall) are implemented in the cloud. Without getting into too many details on vCPE we can observe similar themes to the RPD/MAC use case since both have a relation between a physical device and virtual one. The issue of placement appears in the vCPE use case as well because the same policies that govern the placement of a vCMTS relative to an RPD apply here as well - whether distance based, or availability based or load based. The same is true for Traffic Engineering as well. The CPE may be a business CPE with a service level agreement on QoS that requires TE. The same TE solution that applies to the CPE case can be used for engineering a path between RPD and a MAC resource.

There will be customizations needed for each use case. As an example, the CPE use case may use GRE or IPSEC to connect the virtual instance to the physical one. RPD and MAC are connected via L2TPv3, but one can see how the majority of functionality remains the same.

Finally this is meant only as an example and other use cases can leverage this infrastructure as well.

In summary the solution for pairing RPD and MAC resources does not need to be a highly unique solution to solve only the RPD/MAC use case. It can build upon a foundation that is relevant to other use cases and could integration in general.

CONCLUSIONS

RPD can be dynamically and in a highly flexible way be assigned to various MAC resources, whether a local CCAP device in the hub, a remotely located CCAP or a vCMTS. The choice can be policy based and can be changed along with changing network dynamics.
Furthermore, the SDN infrastructure used to achieve this agility does not need to be built from the ground up for the specific use case of pairing a MAC resource with an RPD resource. Instead it builds upon the same SDN infrastructure that is used to enable a host of other network use cases where pairing of resources is needed.

REFERENCES

1. "Telecommunications Management Network". ITU-T Recommendation series M.3000
2. Availability for NFV, Alon Bernstein, INTX 2015
3. https://tools.ietf.org/html/draft-filsfils-spring-segment-routing-use-cases-01
4. Remote Phy Architecture, cablelabs doczone

SDNized CABLE ACCESS NETWORKS

Karthik Sundaresan
CableLabs

Abstract

Software-Defined Networking ideas are showing a lot of value in the networking industry. SDN can transform cable access networks and simplify MSO operations.

This paper introduces an SDN architecture for Cable access networks. Data Models and South Bound Protocols are foundational tools upon which an SDN architecture is built. The controller can configure and manipulate network devices dynamically while the operator focuses on developing the services to be delivered. This architecture will apply to any underlying access technology such as DOCSIS, EPoN etc.

INTRODUCTION

Software defined networking is an evolutionary concept whose time has come. In computer programming, the focus shifted from initially writing machine-level code, then moving on to assembly language code and then progressively moving on to higher-level languages. That way programmers didn't have to worry about machine level details and instead focused on developing the algorithms and applications.

Similarly, networking is transitioning now from using manual device level commands (via CLI, etc.) for setting up the network device and the path, towards higher levels of abstraction in an SDN world, where the network and the services on top of it can be programmatically configured in a software-defined way.

SDN ideas are showing a lot of value in the networking industry. Networks are being made more programmable and are becoming virtualized. While there has been significant focus on SDN in core and data center networks, it will also provide significant benefits in the access network. It will enable quicker deployment of new services, increase flexibility for MSOs, and reduce operational expenses.

BENEFITS OF SDN PARADIGMS IN AN ACCESS NETWORK

SDN concepts have taken root in Data center environments and other self-contained networks where the operator has full control of every device. The Cable Industry can make use of the new paradigms to make the access networks more efficient.

The DOCSIS Access Network is a great success story for the Cable Operators. IP High Speed Data (IPHSD) is the most important of services for the customers. Operators face a lot of challenges as the access network grows in size and the services offered over them increase day by day. Operators are managing huge numbers of devices, thousands of CMTSs and millions of Cable Modems, Set top boxes and eMTAs. The challenges of provisioning these devices, managing services, and creating interconnects becomes mind-boggling for an operator, requiring lots of people to manage and debug a network.

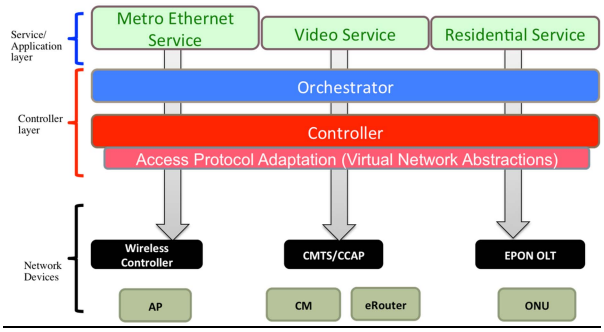
SDN techniques bring a new order to the chaos of service provisioning/management and streamline the MSO operations on a DOCSIS network. An SDN Controller becomes the device which talks to all the MSO network devices to configure the devices and to provision services on them. Today, operators use a disparate set of tools to manage the services on a network, including a separate vertical for provisioning and managing each new type of access

technology, be it DOCSIS, EPoN or WiFi services. Services over DOCSIS need individual config files for each Cable Modem, which doesn't scale well as the number of service tiers increases. It especially doesn't scale when provisioning a service like an L2VPN circuit, where each CM needs a unique config file.

The SDN Controller will implement a common set of protocols to talk with different network devices. This allows the MSO to start focusing on creating better services to run as 'Applications' above the SDN Controller and let the controller configure or program each network device appropriately.

SDN ARCHITECTURE FOR CABLE ACCESS NETWORKS

SDN architecture for cable networks can be described in 3 layers. At the center, there is the SDN Controller. As described before this device is aware of and can talk to each of the network devices in the domain. The protocol used could be an industry standard protocol or it can be a specific protocol used by a particulare network device. This layer is shown as the Access Protocol Adaptation layer. The Adaptation layer essentially abstracts out the individual pieces of the network. The SDN controller takes the commands from the Application or service Layer and translates it to the specific devices in the network.



SDN Architecture for Cable Access Networks

Different parts of the network controlled by various SDN Controllers; the orchestrator is a higher-level entity that is used to orchestrate the services across those domains. On the top of it all sit the various applications the operator wants to deploy. In the access network, this could be IP HSD service, or L2VPN service, or setting up access for Third-party internet Access as a service.

This architecture allows for any device or access technology to be used in the access network and the same applications/services will be implemented seamlessly over each of those access networks. This is the true power of the network abstraction, which comes from the introduction of SDN and a controller device in the network.

The Access protocol Adaptation sub-Layer of a Controller is used to talk to each of network devices and it takes the form of a South-Bound Plugin on a controller device. The North-Bound programming interface used by the applications to talk to the controller becomes an NB API exposed by the controller.

The service/business logic interfaces are commonly referred to as "north-bound" of the controller, and the device-specific interfaces are referred to as "south-bound". The SDN controller exposes NB APIs to the applications; these are the "north-bound" APIs from the perspective of the SDN controller. The SDN controller enables programmatic interfaces to each of the network elements it controls, and this is the "south-bound" API.

DATA MODELS AND SOUTH BOUND PROTOCOLS

The two main enablers for a software programmable network are

- The configuration and statistics information, which is exchanged

between the SDN controller and each of the network devices
- The protocol, which carries the needed information.

An Information model or the data model needs to be developed for every network device that needs to be programmed. This data model will include all the data elements, like device settings or service settings, which need to be read by or written from a controller, to enable the devices and services across the network. It will include the elements needed to configure the device on boot-up (day-zero config). More importantly, the data model will also include the data elements which represent the creation of services over those network devices (dynamic-config).

The southbound protocol itself is independent of the data it carries back and forth. A suitable protocol choice would need to address both the day-zero config as well as the dynamic-config operations. The day-zero config requires a lightweight protocol, which is responsive to the needs of dynamic nature of setting up and tearing down of services.

For data modeling, YANG is the choice of the networking industry. NetConf and RestConf are the recommendations in terms of the protocol choices to carry the data back and forth from a controller to a network device. Any HTTP-based mechanism would also be a good choice as it is a simple stack that can be implemented by a variety of devices.

The Network Configuration Protocol (NetConf) is a network management protocol developed and standardized by the IETF (RFC 6241). It provides mechanisms to install, manipulate, and delete the configuration of network devices. Its operations are realized on top of a simple remote procedure call (RPC) layer. The NetConf protocol uses an Extensible Markup Language (XML)-based data encoding for the configuration data as well as the protocol messages. The protocol messages are exchanged on top of a secure transport protocol.

YANG is the data modeling language used to model configuration and state data manipulated by the NetConf protocol, remote procedure calls, and notifications. It is a "human-friendly" modeling language for defining the semantics of operational data, configuration data, notifications, and operations.

RestConf is a REST (Representational State Transfer) like protocol running over HTTP for accessing data defined in YANG, using data stores defined in NetConf. Currently it is an IETF draft that describes how to map a YANG specification to a RESTful interface. The REST-like API is not intended to replace NetConf, but rather to provide an additional simplified interface that follows REST-like principles and is compatible with a resource-oriented device abstraction.

USE CASES

IP HSD Service

High Speed Data (HSD) service is one of the most common services offered by an operator to end-users today. The current workflow for HSD using customer-provided cable modem is outlined as follows. The Customer Service Representative (CSR) is responsible for gathering customer and device information and entering that information into the BSS/OSS system. The current approach requires the CSR to be a middleman between the customer and network components such as BSS/OSS. One of the biggest issues faced by operators today is the static configuration of services using the CM configuration file. Anytime a new service is to be added, the operator needs to update the configuration file

and reboot the CM. This is not scalable, causes outages, and also causes many mistakes.

An SDN controller can automate the tasks of a CSR. It will be an easy transition to an operation where managing CM configuration files becomes a thing of the past. All CMs will boot up with a basic configuration file, which essentially gets a CM online at a very low data tier. This allows the user to get online and pick his set of services from an MSO web portal. This portal interacts with the SDN controller, takes the list of services and translates them to commands to set up the services on the DOCSIS network. All the DOCSIS Service Flows, classifiers, and drop rules are set up dynamically by the SDN controller communicating with a CMTS. Essentially everything in the CM configuration file can be configured dynamically at run time. One of the big benefits of this method of operation is that whenever a user wants to change their level of service, they can do so at anytime.

L2VPN Service

MSOs offer L2VPN services to Business customers. L2VPN service is used mainly by businesses to communicate across geologically separate campuses. The process of ordering an L2VPN service on the DOCSIS network and setting up the network path from the cable modem is as follows. Currently, a new L2VPN service order requires a CSR to enter information into the BSS/OSS system. Upon receiving such an order, a network engineer will then configure the NSI path and will create a configuration file to be downloaded to the corresponding cable modems. One major issue with the current workflow model is that it comprises two disparate processes; the edge/core side where a pseudo-wire is configured and the access side where the CMTS-CM-CPE are located. This L2VPN setup takes time since separate organizations create and respond to work tickets, and this makes the process error-prone since it involves the manual exchange of data. An SDN controller improves the workflow by "orchestrating" the two sides of the network and eliminating manual interventions. The SDN controller can be used for such a service order and setup. Similar to the HSD use case, a web portal can replace the CSR and a corresponding SDN application configures the L2VPN path on the NSI side.

The SDN Controller improves the process by dynamically associating a service flow to an L2VPN pseudo-wire. This means that a cable modem does not need to be rebooted in order to activate an L2VPN service. The DOCSIS service flow would be brought up as a standard flow using the controller to talk to the CMTS via a southbound protocol (this could be PCMM) and the NSI side pseudo-wire would be set up using other commands from the controller to the appropriate device elements. The controller would bind the DOCSIS service flow to the pseudo-wire automatically, thereby achieving a dynamic setup of an L2VPN service.

Other Use Cases for SDN in the access network

The Cable operator handles various other services like setting up Internet Access for Third Party operators, Lawful Intercept for law enforcement agencies, Voice services, and Video services. In each of these services there are manual steps taken by a network engineer to set up some part of the service. All of these are candidates for automatic configuration, provisioning and management using a centralized SDN controller.

DOCSIS3.1 Profile Management

There is another class of applications that can be enabled by an SDN infrastructure which includes features that are part of the CMTS but not directly tied to the DOCSIS

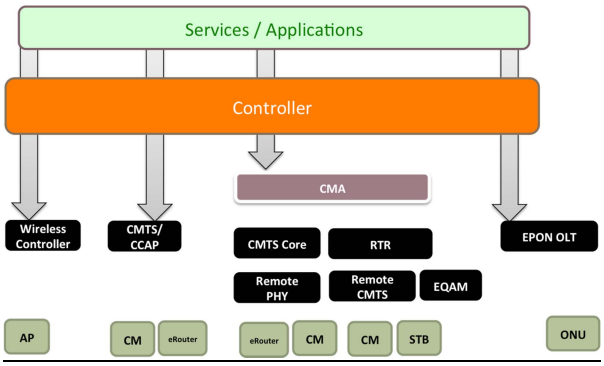
protocol. DOCSIS 3.1 introduces many new features into the access network. These include Variable bit loading and the use of multi-profile downstreams and upstreams, upstream probes to check the quality of the upstream OFDMA signal, etc.

The configuration, initiation logic and compute processing needed to optimize some of these functions (e.g., DS Profile setup or Load Balancing of CMs) are not a core part of the DOCSIS MAC and PHY layers itself. This allows such functionality to be moved out of a CMTS and implemented as an "application" running outside the CMTS. This application can communicate with the CMTS to gather the needed information, process the data, and make intelligent decisions to set up CMTS as needed. The various applications implemented externally communicate with the CMTS to optimize the overall network performance.

To realize this idea, again the basic steps are to develop the data models and a protocol to convey that information back and forth. As an example, to leverage the new OFDM PHY to its maximum benefit, different subcarriers use different modulation orders. Optimizing the downstream profiles will allow a downstream channe to operate with a lower SNR margin, potentially allowing a channel to operate at an overall higher throughput. The logic to achieve this can be external to a CMTS and enable innovation. For an operator, it allows uniform operation of such algorithms across different CMTS platforms. An SDN controller with the appropriate data models to represent the data, and the protocols to configure a CMTS and a northbound API to the application will enable such applications.

CCAP MANAGEMENT ABSTRACTION

The DOCSIS access networks can be integrated into an SDN world in many ways. The CMTS platform itself is evolving in the access network, from large CMTS devices to CCAP devices that are designed to provide increased EQAM and CMTS densities in a combined platform. There are also various Distributed CCAP Architectures being developed by the Cable industry. These architectures for the data transport side leave a router at the head-end and allow the CMTS MAC & PHY layers, or just the CMTS PHY Layer, to be moved down to a remote fiber node. Similarly, on the video side, all the EQAM functionality or the EQAM PHY layer can be moved to remote node.



CCAP Management Abstraction (CMA)
Distributed CCAP Architectures

These various CMTS architectures can be tied into the MSO's SDNized network. The various pieces can be controlled and managed via a CCAP Management Abstraction (CMA) Layer. The CMA abstracts the various components of this distributed system into a single, cohesive CCAP platform for the OSS/BSS systems. The CMA is an application whose function is to create a container around a number of CMTS physical and virtual functions, and enable them to be managed and provisioned from existing back-office systems just like an integrated monolithic CCAP is today. Colloquially the CMA is also known as a Virtualized CCAP

APPLYING SDN ARCHITECTURE TO OTHER TYPES OF ACCESS NETWORKS

The SDN architecture described here for DOCSIS Access networks can be easily

applied to other Access network technologies and pieces of network equipment on the MSO network. Take, for example, an EPoN deployment by an operator: the EPoN systems can be provisioned using the DPoE specifications. An SDN controller could essentially play that role of the DPoE System and a virtual Cable Modem, to take in DOCSIS Provisioning commands and translate them to EPoN specific commands, and send them to an existing EPoN OLT system.

In this way, legacy equipment with disparate provisioning systems can all be controlled via a Single SDN controller. Again, the key would be to create the needed data models for EPoN systems to make sure the application/service intents can be translated appropriately to the technology which is being configured. The protocol to talk from an SDN controller to the OLT could be something like RestConf or some other legacy protocol supported by the OLT, as long as the SDN controller can support that southbound protocol. This can apply to other access technologies as well, e.g., GPoN deployments, new services on top of DOCSIS or PON deployments, Wi-Fi AP provisioning and management, etc.

CONTRIBUTIONS TO OPENSOURCE CONTROLLERS

The industry needs to focus outward to the open source community, as that's where all the innovation and thought regarding SDN architecture are happening. One example is the Opendaylight (ODL) open source controller, which is an open platform for network programmability to enable SDN in networks of any scale. ODL software is a combination of components including a fully pluggable controller, interfaces, protocol plug-ins and applications. In the Helium release, the ODL platform has added a PacketCable MultiMedia plugin, which essentially enables the ODL controller to communicate to an existing CMTS platform using the COPS interface and the PCMM data protocol. CableLabs led the software effort to contribute the southbound protocol plugin and needed data models to the ODL code base. Similar efforts will be needed to enable support and management of Cable technologies by open source controllers.

FUTURE DIRECTIONS FOR EVOLUTION OF THE ACCESS NETWORK

The SDN architecture described here forms the foundation for how the MSOs will deploy and manage their networks in the future.

Once an SDN Controller is in the middle of the network, talking to the various devices and network components, then operators can start focusing on developing better applications. As newer access technologies are added, they will be integrated into this framework and the existing services will apply to them seamlessly. The customer will be happier with the same service being available on their home network or, even if they are on the road roaming, as the SDN controller will be able to set up the same services across different networks.

New concepts, such as service function chaining, are taking root within the industry. Different network functions and applications are being virtualized and being run on COTS hardware in the cloud. The SDN controller and orchestrator again are in the middle of that architecture, helping direct the traffic from the customers/access network to the appropriate VMs and services in the cloud. The SDN paradigms will bring simplicity to the MSO network operations by abstracting out the complexity of the individual network devices and their configuration.

Securing Remote PHY Infrastructure

Pawel Sowinski, Gerry White
Cisco Systems, Inc.

Abstract

The Remote PHY (R-PHY) architecture represents the next stage in the evolution of DOCSIS and video service delivery as defined in a family of specifications under development by Cable Television Laboratories (CableLabs®). Products can be expected in the near future and operators need to understand the issues with the new technology, and start planning how to deploy it.

Unlike previous versions of DOCSIS this new architecture significantly impacts the HFC access infrastructure. The CCAP PHY components migrate into R-PHY devices (RPDs) located at the edge of the IP network, which is also extended deeper into the outside plant. To enable this evolution, existing linear optical links are converted into standard Ethernet connections and the traditional fiber nodes are replaced with IP enabled R-PHY devices.

Thus the R-PHY architecture requires MSOs to deploy a large number of IP networking devices into inherently unsecure portions of the network, such as pole mounted nodes and remote cabinets. This extension of IP deeper into the plant exposes the network to a set of security threats so that the infrastructure and the RPDs must incorporate critical security measures to protect the network, the RPDs and the customer data.

The paper outlines a comprehensive approach to ensure the security of distributed IP networks in insecure locations, using R-PHY as a specific example, taking into account the unique network, protocol, and application characteristics of R-PHY systems. The authors assess the scope of security threats, propose mitigation techniques to address the identified vulnerabilities and recommend security requirements for individual network components. In particular, the paper details such procedures as the secure authentication of RPDs to prevent unauthorized access to the MSO's IP network and approaches to secure R-PHY control and data connections.

In summary, the paper demonstrates why, where, and how standards-based, distributed IP networks can be secured in a cost effective and interoperable manner using R-PHY as a specific example.

INTRODUCTION

Until recently, the conventional wisdom perceived network security as a complicated subject, which was historically understood and tackled by only well-trained and experienced experts. This view is changing in todays, post Snowden world as security awareness takes a prominent place in the global zeitgeist. Not a single week passes without news of significant cybercrimes by various perpetrators such as thieves/hackers, disgruntled employees, organized criminals, terrorist organizations, or even state-sponsored groups. New attack vectors and threats are identified every day.

A recent survey of governments, businesses, and individuals in the U.S., China, Russia, and India found that more than 88% of respondents believe that cyberspace threats are significant. While many respondents feel comfortable with online banking and shopping, more than 69% are not comfortable with sharing identity and personal data online.

This is a valid concern—the latest Internet Crime Report by the Internet Crime Complaint Center shows an increase in cybercrimes, as thieves seek out personal data and other valuable information for their own advantage or as hackers disrupt operations of businesses for various reasons.

In fact, the threat of cyberattacks has been elevated to one of the nation's most pressing security, economic and safety issues. In this year's State of the Union speech president Obama pushed Congress to bring cybersecurity legislation to fruition in order to combat emerging attacks. "No foreign nation, no hacker, should be able to shut down our networks, steal our trade secrets, or invade the privacy of American families, especially our kids," Obama said in his address. "We are making sure our government integrates intelligence to combat cyber threats, just as we have done to combat terrorism."

As developers of equipment and operators of networks we have the responsibility to react to this rapidly changing environment and actively work to provide security for vulnerable components such as R-PHY Devices.

A lot parallels can be drawn between R-PHY devices and Internet of Things (IoT) devices. Earlier this year the Federal Trade Commission released a report on critical security and privacy issues related to IoT technology. The FTC's report proposes security guidelines for manufacturers of IoT devices. The FTC hopes that "by adopting the best practices we've laid out, businesses will be better able to provide consumers the protections they want and allow the benefits of the Internet of Things to be fully realized".

The FTC's press release outlines a number the guidelines for companies developing Internet of Things technology from which we selected four (out of total six) summary recommendations best fitting R-PHY security concerns. The FTC report recommends to:

- "Build security into devices at the outset, rather than as an afterthought in the design process;
- When a security risk is identified, consider a "defense-in-depth" strategy whereby multiple layers of security may be used to defend against a particular risk;
- Consider measures to keep unauthorized users from accessing a consumer's device, data, or personal information stored on the network;
- Monitor connected devices throughout their expected life cycle, and where feasible, provide security patches to cover known risks."

While R-PHY deployments will never reach even a small fraction of the scale of IoT, which already has over 25 billion devices, the FTC's guidelines are equally relevant to Remote PHY security.

SCOPE

The paper primarily focuses on the digital networking aspects of R-PHY system security, especially those elements that impact external network or programming interfaces of the RPDs and the MSO systems with which they interoperate.

We decided that other security concerns fall out of scope of the paper, for example, how to ensure the physical security of RPDs or how to deter R-PHY device theft. We expect that manufacturers of RPDs will individually incorporate effective elements of physical security in RPDs such as tamper proof enclosures or the ability to raise alarms if the device's access doors are open by unauthorized person.

Those features however, in most cases do not require standardization or directly impact other systems and therefore are less interesting for a discussion on the security of an IP network infrastructure built for physically insecure locations.

REMOTE PHY

Remote PHY (R-PHY) is the next stage in the evolution of DOCSIS and video service delivery. It is a modification 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) by an IP network. The combination of CCAP Core and RPD provides the functional equivalent to 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 in Figure 1 and Figure 2.

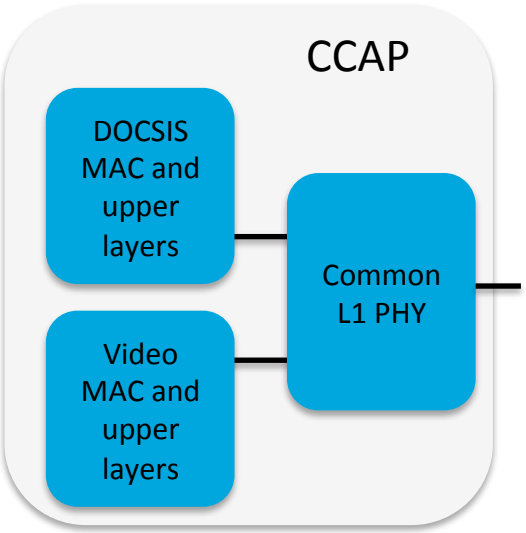


Figure 1 - Integrated CCAP

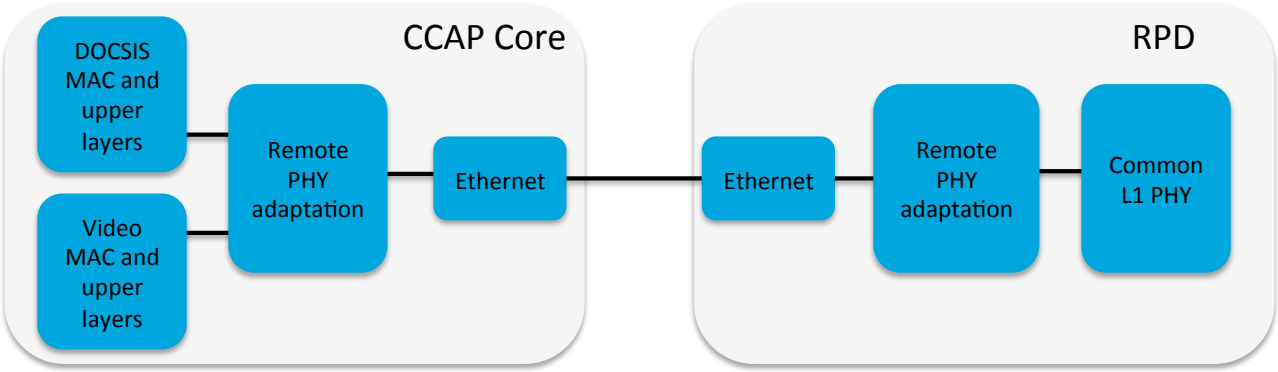


Figure 2 - CCAP with Remote PHY

This has the effect of moving the digital to RF conversion from the CCAP Core, located in the head end or hub, to the RPD located deeper in the network, such as in a fiber or optical node. Standard digital optics (e.g. Ethernet) can be used for the CCAP to RPD link in place of the analog optics previously used. RF over coax or analog fiber is used for the RPD to CM link, which is now much shorter. This extends the all digital IP network deeper into the plant providing advantages in lower cost, simpler operation and better performance. The migration of the physical layer from the CCAP to the RPD is transparent to the external DOCSIS infrastructure and CPEs so that these can be used unchanged as shown in Figure 3.

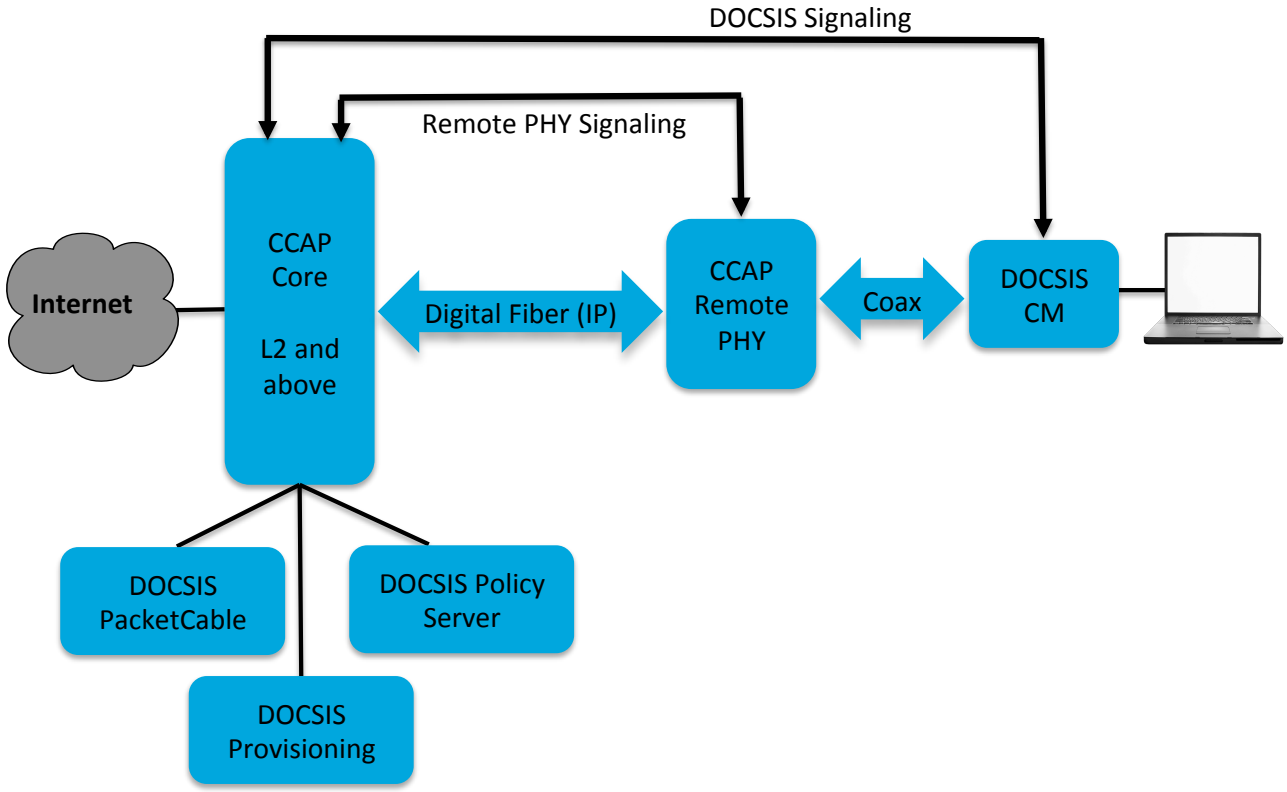


Figure 3 - RPD Deployment Environment

For a more in depth view of R-PHY please refer to [R-PHY SCTE] and [R-PHY NCTA].

R-PHY SECURITY ENVIRONMENT

In the next several years Cable Operators could potentially install hundreds of thousands of RPDs in their networks. The majority of them will be deployed as Fiber Node replacements and will be installed in inherently unsecure locations, such as pole mounted nodes and remote cabinets. The fibers which are today carrying analog HFC signals will be converted to digital use as primarily 10-gigabit Ethernet links carrying IP

data. This extension of IP deeper into the plant exposes the network to a set of security threats so that the infrastructure and the RPDs must incorporate critical security measures to protect the MSO's network, the RPDs and the customer data. Before attempting to solve this problem we will first examine how introduction of R-PHY changes the MSO's network and assess the newly introduced security threats.

As shown in Figure 4 an integrated CCAP is deployed in a head end or hub location with analog fiber / coax links to the DOCSIS CMs and video STBs. In either head end or hub deployments the CCAP is placed in a physically secure location, typically in protected network equipment buildings in a hub. Physical access to the CCAP equipment is granted only to authorized personnel. Network access to the CCAP is secured with a wide range of features including multilevel access, authorization controls and traffic control features such as source verify, access control lists and network domain isolation.

The optical node, CMs and STBs are in unsecure (from the operator's perspective) customer premises so that additional security mechanisms are required. Thus CMs must comply with the DOCSIS security specifications and video devices must comply with one of the proprietary video security schemes in use. In both cases authentication of the end device is required and traffic on the CCAP to CPE links is protected by encryption. The optical node is a physical layer optical to coax converter, is transparent to the security mechanisms used and is less susceptible to security threats. The trusted network boundary is aligned with the boundary of the HFC plant, which carries analog RF signals and inherently limits exposure to active threats. While some operators have seen isolated attempts at theft of service based on "perfect" replication of the CM hardware, such attempts are relative easy to identify and defend against.

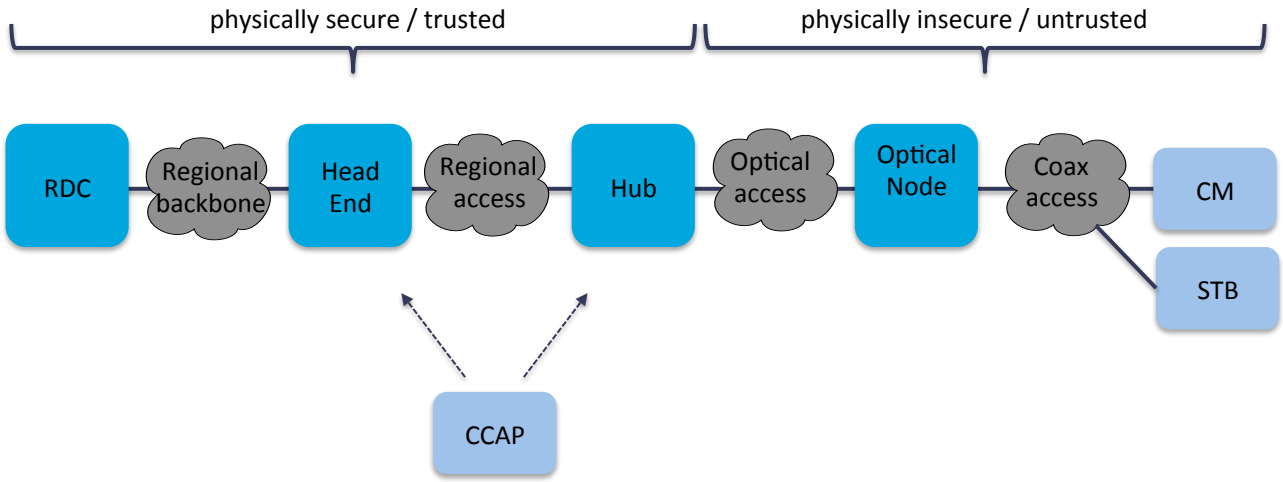


Figure 4 - CCAP Deployment Options

As shown in Figure 4 the deployment options for an R-PHY solution are more diverse. With the transition away from analog optics the CCAP Core may be deployed at a Regional Data Center (RDC) deeper in the network or in the same head end and hub

locations as an integrated CCAP. All of these are physically secure locations. An RPD may be deployed in secure head end or hub locations but may also be deployed much deeper in the network in an inherently unsecure location such as a street cabinet or a fiber node. With the deployment model when both CCAP Core and RPD are within the trusted part of the network the requirements for authentication or security between these devices can be relaxed or made optional. The combination of CCAP Core plus RPD provides the same functionality as an integrated CCAP so that the existing DOCSIS and video security to the CMs and STBs remains in place with traffic in the untrusted portion of the network encrypted as with the integrated CCAP.

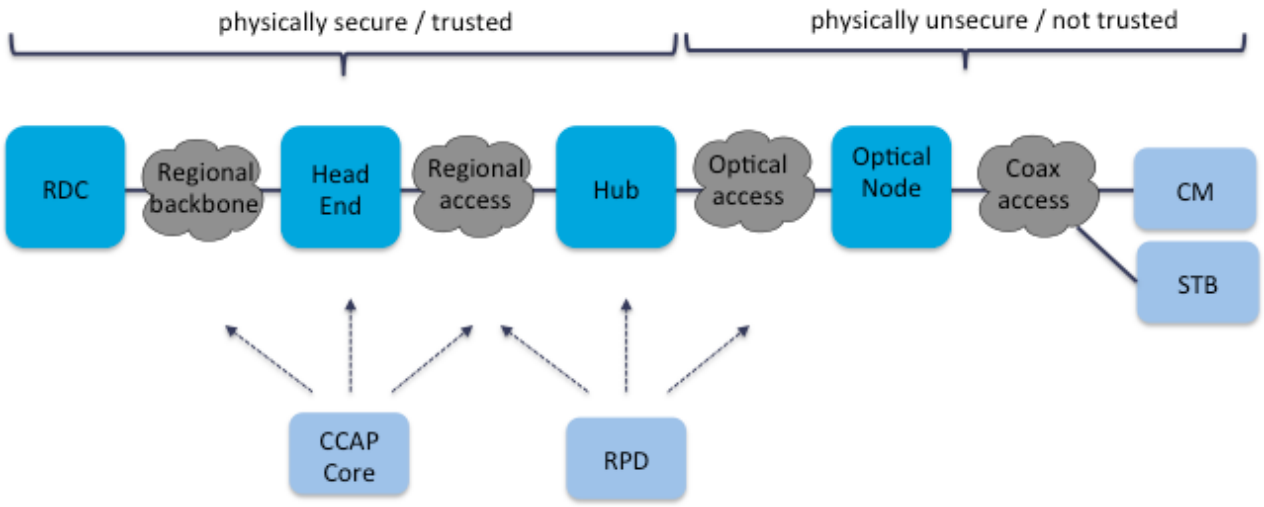


Figure 5 - R-PHY Deployment Options

When the RPD is deployed outside the physically secure domain such as in a street cabinet or optical node the security exposure is raised. The DOCSIS and video traffic is still encrypted between the CCAP core and the CMs & STBs so that it is not vulnerable to attack but the RPD represents a potential vulnerability. The RPD must be connected to the MSO IP network so that it can access not only the CCAP Core but also IP devices such as DHCP servers and network management systems. Thus the RPD requires access to the trusted network and so presents a potential point of ingress for attacks on the MSO network and services. An analysis of the threats and options for their mitigation follows.

ASSESSMENT OF SECURITY THREATS

Remote PHY equipment, in particular the RPDs, represents an attractive target for cyber attackers. RPDs become a part of the critical infrastructure through which the cable operators deliver services representing the bulk of their revenues. Assuming the success of the technology, it's conceivable that at some point the largest share of US broadband and video will be delivered with the aid of R-PHY technology. It is a distinct possibility that in a few years time an RPD will participate in the transport of internet traffic between our homes and the rest of the digital world. As mentioned earlier, attackers can easily gain physical access to RPDs because the majority of RPDs will be installed in

unsecure locations and include externally accessible Ethernet ports.

Unauthorized Network Access

In the opinion of the authors, unauthorized network access represents the most pressing security threat resulting from introduction of the R-PHY. Since the RPDs are attached to the Converged Interconnect Network (CIN), not directly to the CCAP Core, an attacker impersonating an RPD may gain access to the CIN, the MSO internal network and the devices and systems interconnected by it.

As a result of the unauthorized access to the MSO network, the attacker may be able to further exploit the vulnerabilities of the network to steal data or service, conduct cyber-espionage, eavesdrop on and tamper with customers' data or even disrupt the MSO's operations by launching Denial of Service (DoS) attacks or by other methods.

As countermeasures, the RPDs and the network they attach to must provide means for effective secure authentication and authorization of RPDs onto the MSOs network. Given the scale of RPD deployments, the processes and provisioning of these functions must be automated and minimize any configuration effort.

Customers' Data Privacy

The RPDs do not collect or maintain any subscriber data. However the customer traffic passes through the RPD. The RPD provides purely PHY layer functions; it does not perform any function that requires examination or modification at MAC or higher layers. The customer's traffic delivered via DOCSIS and video service is already effectively protected by BPI+ encryption or video encryption schemes.

Could a very sophisticated attacker posing as an RPD somehow compromise the integrity of those schemes? Conservative assessment leads to the conclusion that it is impossible to completely eliminate such a possibility. However, the authors are not aware of any documented history of significant breaches of these techniques. The threat to the integrity or privacy of encrypted customer data passing through the RPD link is low when the operators take advantage of established encryption techniques.

Theft of Service

The vulnerability towards service theft is similar to threats to Customer's Data Privacy. Because the service data is protected at a higher level and passes through the RPD opaquely, we believe that the RPD's exposure to theft of service appears to be low.

Denial of Service Attacks and Service Disruption

An RPD attaches to the MSO network through one or more high capacity links, typically 10 Gb/s Ethernet. This creates the potential for the RPD H/W platform to become a potent tool for origination of Denial of Service attacks on the operators' network. A "hijacked" RPD or a device impersonating an RPD could be used to inject a high volume of traffic into the MSO network and facilitate a DoS attack on network infrastructure e.g. DHCP, DNS servers or CCAP Cores.

Another possibility is disruption of network operation through the injection of routing or control protocol packets, for example "duplicate" IGMP joins. The countermeasures to these types of threats include securing access to the RPD itself, mandating secure software update, secure boot technology and effective network access authentication.

Other Threats

The authors believe that the outlined security threats represent the most serious risks associated with a R-PHY deployment. In the rapidly changing security environment new threats will undoubtedly emerge. While we cannot predict the future precisely we believe that the best approach is for the R-PHY architecture to adopt a defense in depth strategy, which can be upgraded over time to address emerging threats.

PROPOSED SOLUTIONS

The previous sections of the paper have described the security vulnerabilities raised by RPD deployment outside the trusted network. Fortunately security technologies and practices are available to provide a viable alternative to physical security. The nature and use of these options is discussed in depth in the paper.

As with many complex problems the simplest approach is to divide the problem into a number of components and address these one by one. In this case we have broken the problem into the following sub problems as shown in Figure 6:

1. Protecting the integrity of the trusted network by controlling access to it from the RPD
2. Requiring mutual authentication between the CCAP Core and the RPD and protecting the control plane transactions between them.
3. Protecting the data plane from the CCAP Core through the RPD to the CM
4. Protecting the software integrity of the RPD

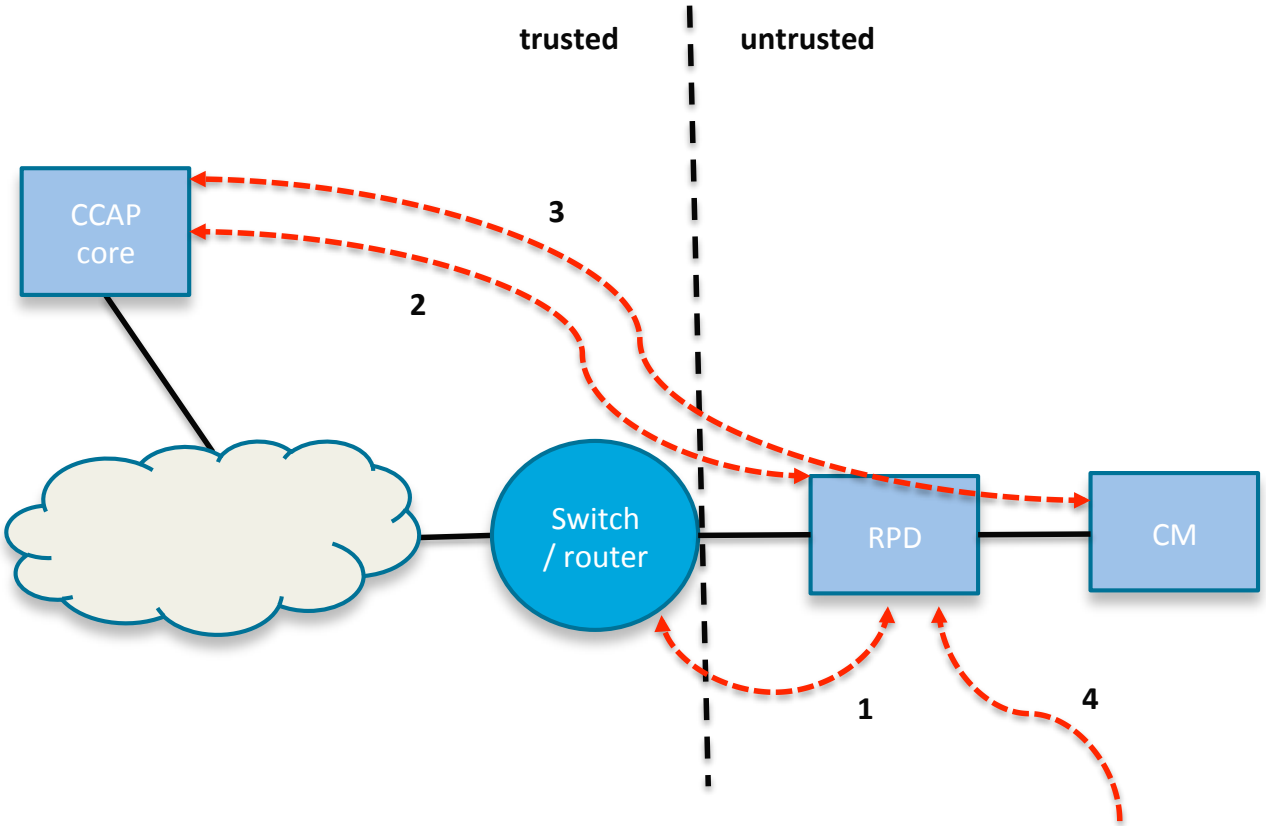


Figure 6 - Problem Set

The following sections of the paper will address these items in turn.

Protection of the Trusted Network

As mentioned earlier the RPD may be deployed in an unsecured location (refer to Figure 5) and is based on standard network technologies such as IP and Ethernet. Thus the RPD represents a point from which an attack could be launched on the MSO trusted network. In order to protect against this it is necessary to control access to the trusted network by an RPD. Fortunately similar problems have been addressed for Ethernet and WiFi networks and standard solutions developed which can be leveraged. In this case we have chosen to use the 802.1X Port Based Network Access Control standard developed by the IEEE [IEEE 802.1X] and supported by virtually all Ethernet equipment vendors .

802.1X Authentication

802.1X is a port based access mechanism used with Ethernet switches. Each port on the switch may be either fully open, fully blocked or allow a subset of packets only e.g. for device management. When the RPD is first connected to a switch port it will receive restricted or no access to the trusted network (depending on switch configuration). 802.1X messages are exchanged to identify the RPD and unblock the port if the RPD is recognized as authorized for the network.

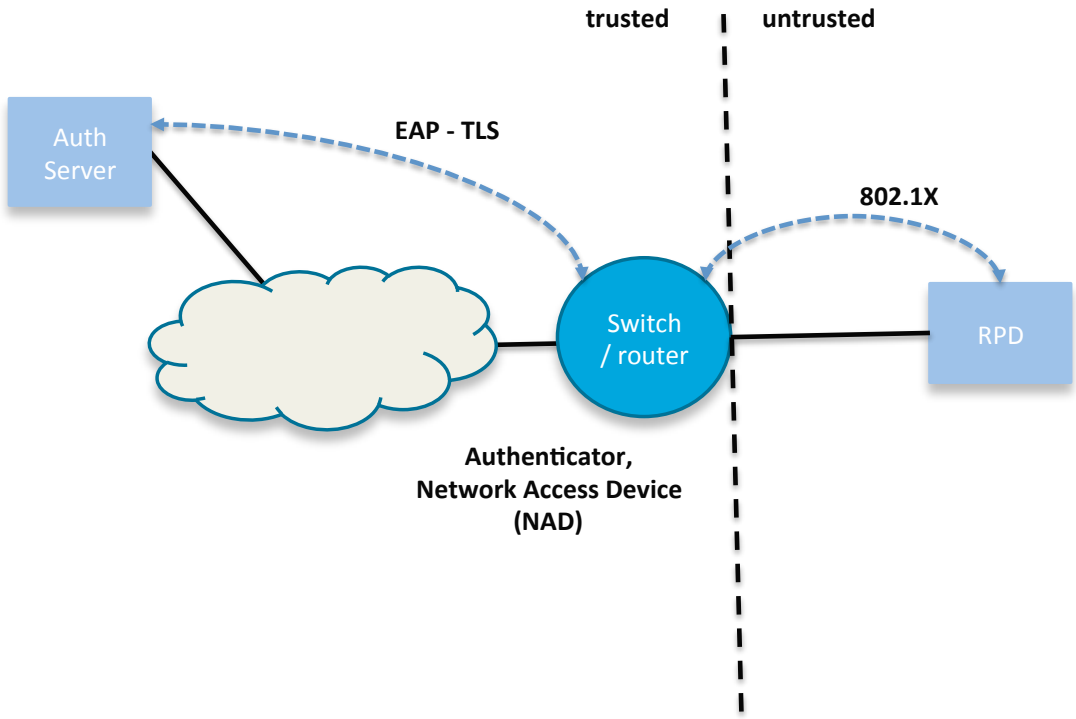


Figure 7 - 802.1X Authentication

An RPD deployment using 802.1X is shown in Figure 7.

There are three components involved in the process:

- **The Supplicant** is the device wishing to gain access to the network, in this case it is the RPD.
- **The Authenticator** is the Network Access Device (NAD), the switch or router, located at the edge of the trusted network to which the RPD must connect to in order to gain network access. The NAD itself is within the trusted domain and acts as the gatekeeper for the trusted network
- **The Authentication Server** contains the database and policy used to control access.

802.1X uses EAP, the Extensible Authentication Protocol [IETF RFC 3748], to enable authentication using a centrally administered Authentication Server. It defines EAP encapsulation over LANs (EAPOL) for communication between the supplicant (RPD) and the authenticator (NAD), which are connected directly over Ethernet. The authenticator and the authentication server exchange EAP messages over an IP network using Radius or Diameter.

Message Exchanges

The supplicant (RPD) presents credentials to the authenticator (NAD), which forwards these to the authentication server for verification. The authentication server determines the validity of the credentials and informs the NAD whether the supplicant is allowed to access the trusted network.

The message exchange used to establish authentication is shown in Figure 8.

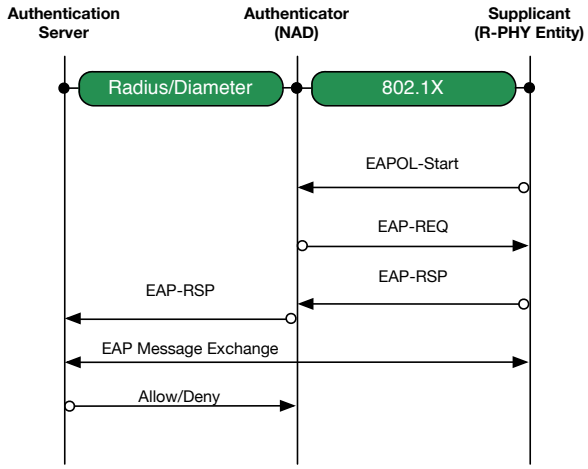


Figure 8 - 802.1X Message Exchange

The RPD informs the Authenticator of its existence using an EAPOL-start message. The authenticator generates an EAP-REQ to ask for credentials. The RPD returns these in an EAP-RSP, which is forwarded to the authentication server. The RPD and authentication server exchange EAP messages (via the authenticator) and on completion of the process the server informs the authenticator and the RPD of the result. At this point if the RPD is allowed access the Authenticator opens the switch port.

Security Credentials

In this case the credentials used are X.509 certificates issued by CableLabs, which are very similar to those used to authenticate cable modems. This enables both operators and equipment vendors to leverage existing processes for certificate deployment and validation.

MACsec

In its basic form 802.1X controls network access for a single host connected to a switch port. This is shown in Figure 9 as scenario 1. However once the port has been opened for an authenticated device it is also open for any

other devices on the same physical port. This can be problematic in some topologies such as when multiple hosts are connected via a daisy chain or through a hub as shown in scenarios 2 and 3. It is also susceptible to a device masquerading as valid by spoofing the MAC address of an authenticated system and to man in the middle attacks by inserting a rogue device between the RPD and the NAD.

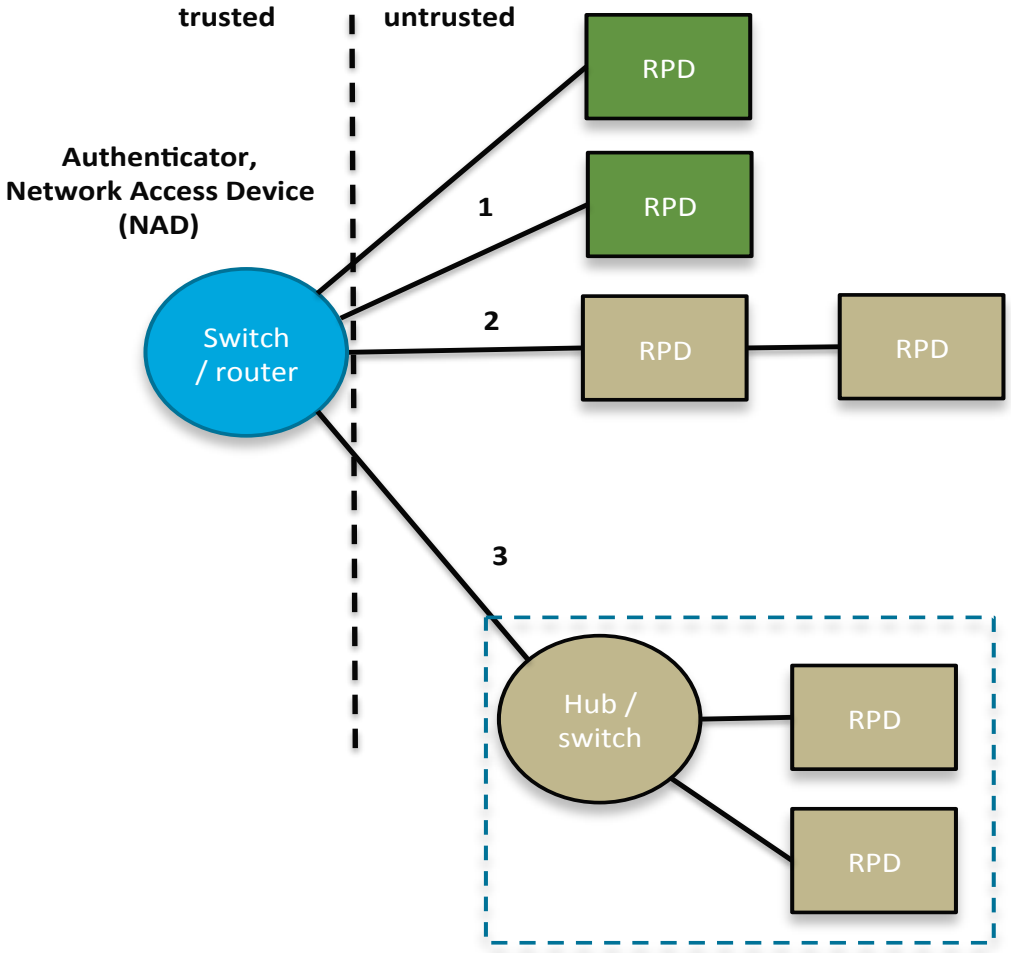


Figure 9 - RPD Deployment Topologies

To address these cases an additional mechanism, 802.1AE (MACsec) is required.

For each device connected to a physical port on the NAD MACsec defines a virtual port. A connectivity association is established between the virtual port and the MAC address of the device. Each device is authenticated via the EAP mechanism as previously described.

Security keys are derived for the connectivity association from the keying material in the EAP exchanges. Thus each connectivity association (and hence attached device) has its own encryption keys, which are used to secure messages between the NAD and the device using AES128 encryption. This enables multiple devices per physical port to be supported and prevents address spoofing or man in the middle attacks.

Mutual Authentication and Control Plane Protection

A second piece of the security architecture is mutual authentication between the CCAP Core & RPD. This is required to resolve the following risks:

- RPD could launch DOS attack on CCAP
- RPD control traffic could be hijacked to deny network access to CMs
- RPD could be replaced to enable a man in the middle attack of subscriber traffic
- A spoof CMTS could attach to the RPD for CM traffic capture

Mutual Authentication

Both the CCAP Core and the RPD are issued with X.509 certificates by Cablelabs and both have securely stored private keys. The Internet Key Exchange protocol version 2 (IKEv2) is used to facilitate mutual authentication based on these keys and certificates. IKEv2 is an IETF defined standard that is widely deployed to secure communications across the Internet. It is documented [RFC 7296] elsewhere and will not be covered in any detail in this paper.

IKEv2 provides the following important functions for the CCAP core to RPD interface as shown in Figure 10:

- Authentication of the two endpoints. The CCAP Core and the RPD exchange certificates provided from the Cablelabs certificate infrastructure. Each device (CCAP and RPD) can independently verify the authenticity of the certificate it has received from the other using standard mechanisms [ref RFC 5280]. Messages signed by the private key of each device and decoded using the public keys (from the exchanged certificates) ensure that each device is who they claim to be.
- Generation of security keys. Following the authentication exchange shared keying material is generated which is then used to secure future exchanges.
- Negotiation of a set of security profiles. Each profile defines which traffic is to be protected by the profile and the cryptographic algorithms to be used. For the RPD to CCAP interface two profiles are used.
  - For the Generic Control Plane (GCP) traffic used in general RPD configuration.
  - For the L2TPv3 control plane used for tunnel control.
- Periodic updates to security keying material as needed over time.

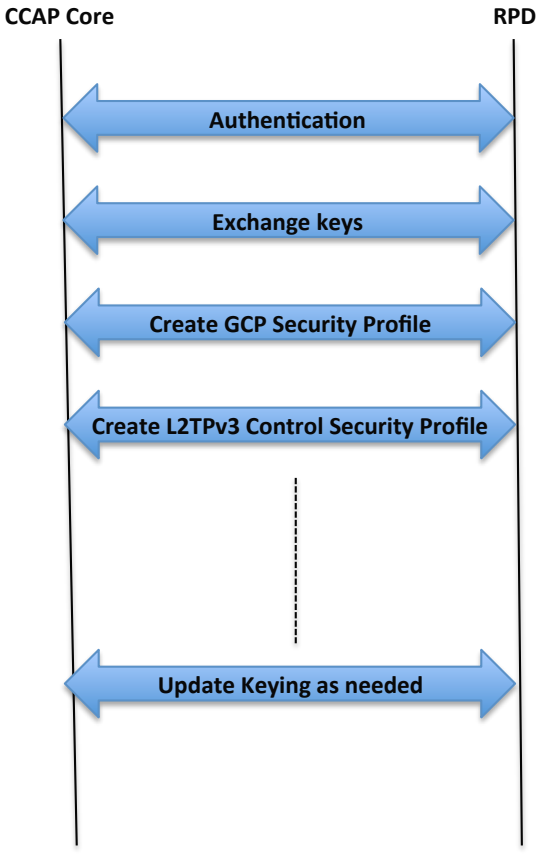


Figure 10 - IKEv2 Functions

Control Plane Protection

During the IKEv2 message exchanges the CCAP Core and RPD negotiate a security profile. This defines the traffic to be protected and the algorithms and keying material to be used. Internet Protocol Security (IPsec) is then used to secure the control plane by authenticating and or encrypting the payload. As with IKEv2 IPsec is a widely deployed Internet standard.

An operator can choose whether to select encryption and authentication or use authentication only. Authentication without encryption leaves the payload open to inspection but still prevents tampering with commands and requires fewer resources. This is shown in Figure 11.

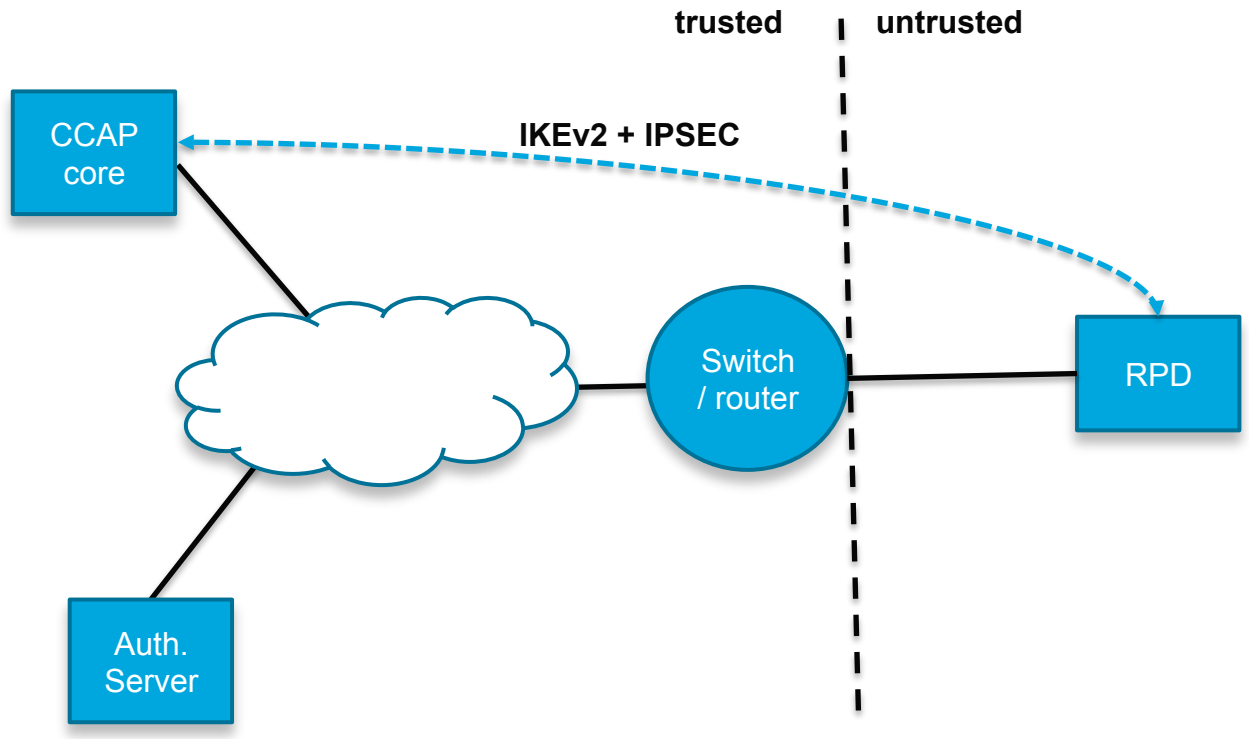


Figure 11 - Control Plane Protection

Data Plane Protection

In an R-PHY deployment the data plane traffic is carried over IP/Ethernet links between the CCAP Core and the RPD where it is converted to RF for transmission to the CM or STB. The CM and STB have always been in locations outside the operators control and the mechanisms to secure this traffic based on DOCSIS BPI and video encryption algorithms are widely deployed and well understood.

The RPD simply provides a physical layer conversion of the data plane from digital to analog with no impact on the end to end security mechanisms, which can continue to operate unchanged. This is a major advantage of the R-PHY over more complex distributed

architectures working at higher layers of the protocol stack.

Figure 12 illustrates the set of security mechanisms used to protect an RPD based network.

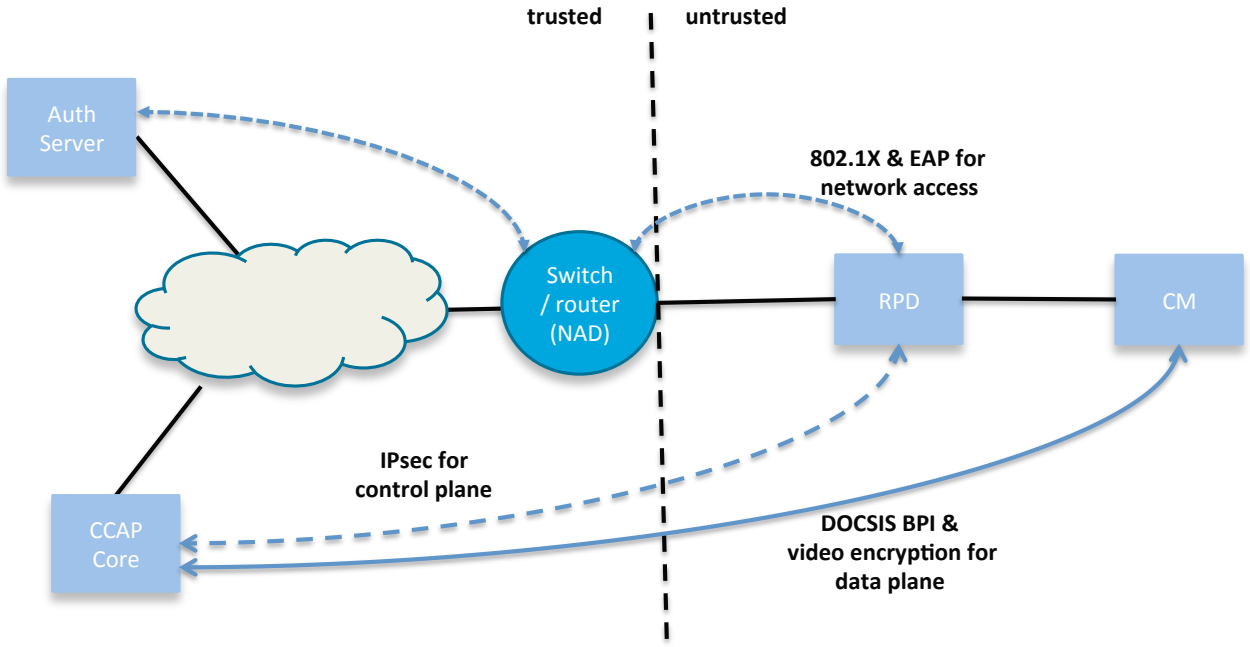


Figure 12 - RPD Based Network Security Architecture

Integrity of RPD Software

Another layer of security provides protection for the software running on the RPD from unauthorized modification. If an attacker was able to replace or modify the software running on the RPD it would open the possibility of misusing the RPD in ways the paper explained earlier.

Two functions provide the foundation for protection of the software integrity. These are Secure Software Download (SSD) and Secure Boot. Additional features augment them by limiting exposure to unauthorized modification during software execution.

Secure Software Download (SSD)

Software upgradeability is one of the most basic requirements for any computing or networking system. Cable operators need to periodically download new versions of software onto their RPDs to add new features, to fix bugs … and sometimes to address security vulnerabilities. Authenticating the source and verifying the integrity of downloaded code is fundamental to the security of R-PHY architecture.

The principles, including code signing requirements for RPD secure software download, and many implementation details have been adopted from the DOCSIS 3.1 Security Specification. Broadly speaking,

with respect to secure software download; the RPD assumes the functions of a DOCSIS Cable Modem.

It is envisioned that such an approach will allow the operators to reuse the majority of the OSS infrastructure already deployed for CM software and security certificate management to perform equivalent functions for RPDs.

To download an RPD software image securely, the RPD vendor and/or MSO will digitally sign the image using the appropriate code verification certificate (CVC) and place the image on a Software Download server.

The operator can trigger the download of the software image via parameters obtained by the RPD from the CCAP Core in the control protocol, or directly via SNMP commands.

After an RPD downloads a software image, it validates the image by verifying that the included CVC chains to the Root CA Certificate trust anchor, and by checking the image's digital signature. If this validation is successful, it installs the software image for operation.

Secure Boot and Software Execution

To complement secure software download, RPD vendors may choose to implement a set of features to further ensure that the integrity of the executed software has not been compromised. Secure Boot offers the prospect of a hardware-verified, malware-free bootstrap process that can improve the security of many system deployments.

Secure Boot technology ensures that the first code executed on a hardware platform is authentic and unmodified, establishing a root of trust in the system to build upon. This sets the foundation for establishing a full chain of trust to validate all levels of software executing on the hardware platform, including multiple phases of the boot process and the final executable image. The Secure Boot chain of trust is often attached to an immutable hardware anchor which provides secure device identity. [802.1AR-2009] describes a standard for a secure device identifier (DevID) which is a cryptographic identity bound to a device and used for assertion of the device's identity. IEEE 802.1AR compliant devices can rely on CableLabs issued RPD device certificates explained later in the paper and the executable image signing procedures of SSD.

Reliance on a secure, hardened hardware based anchor eliminates or at least raises the difficulty bar for bit-by-bit cloning of an RPD along with its identity, a scheme sometimes exploited with DOCSIS Cable Modems.

To date, multiple platforms and vendor companies have implemented secure boot techniques, including Linux, Apple, Microsoft, Cisco and MicroSemi. Secure boot implementations are generally considered proprietary as platforms and operating systems vary in requirements. RPD vendors may decide to choose different computing platforms for their RPD products. For those reasons, and because Secure Boot does not involve any external interfaces we feel that Secure Boot implementation choices should be left to individual vendors rather than be mandated by a standard specification.

Run Time Integrity and the other secure coding processes prevent many security attacks, for example Day Zero attacks. Some of the Run Time Integrity technologies that could be used in an RPD include Object Size Checking, Address Space Location Randomization, and executable space protection. We highly recommend that

vendors include these or similar mechanisms for added protection.

Security Certificates

R-PHY security protocols rely on digital certificates for establishing device identity, for authentication of key exchanges between RPD and NAD and between RPD and CCAP Core as well as in verification of the software downloaded to the RPD. Like BPI+ certificates, the R-PHY certificates are based on [X.509] standard version 3. In fact, the R-PHY digital certificates scheme mostly reuses the new Public Key Infrastructure (PKI) defined in [DOCSIS SECv3.1].

R-PHY brings two straightforward extensions to the new CableLabs PKI certificate hierarchy. The CableLabs Device CA Certificate, which in DOCSIS is used to sign Cable Modem CA Certificates, will be utilized to issue RPD Device Certificates.

A new CableLabs Service Provider CA Certificate will be created to serve as a root of trust for certificates issued to MSOs or CCAP Core vendors. The Service Provider CA will be hosted by CableLabs or an approved 3rd party which issues AAA Server Device Certificates or CCAP Core device certificates to approved manufacturers.

The hierarchy of the CableLabs PKI Certificates is shown on Figure 13.

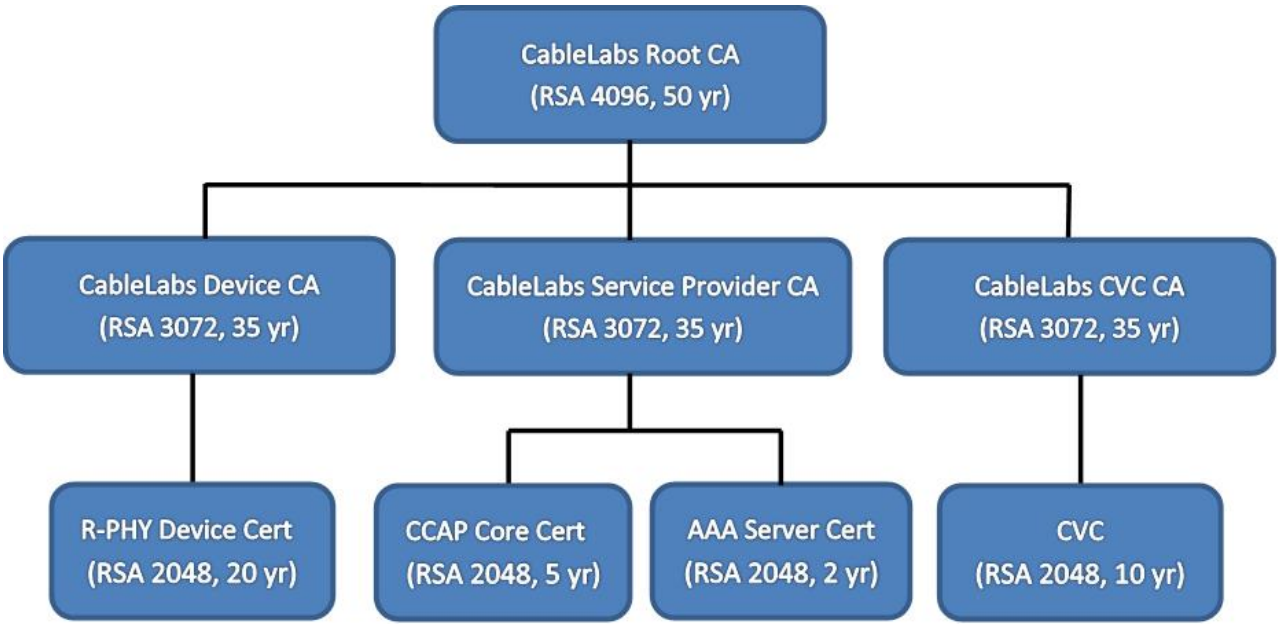


Figure 13 - CableLabs New PKI Security Certificate Hierarchy with R-PHY Extensions

The CVCs adopted to validate the RPD downloaded software images (SSD) do not require any modification to the CableLabs PKI. The CableLabs Root CA will be utilized as a trust anchor for issuing and validating CVC CAs and CVCs exactly as currently specified in [DOCSIS SECv3.1],

The reliance on the DOCSIS certificate scheme in the R-PHY environment offers a number of self-evident benefits. Not only are the involved parties familiar with the BPI+ scheme, but also the existing infrastructure and procedures can be easily extended to support R-PHY. For example, the requirements for device certificate storage and

certificate management, including the certificate revocation procedures can be directly reused.

Leveraging the well-established and verified DOCSIS methods will cut down on development time and the usual risks associated with the adoption of a new technology. The relatively straightforward expansion of the DOCSIS security certificate scheme onto the R-PHY architecture, an technology considered adjacent to DOCSIS, proves the vision of authors of the original and subsequent generations of DOCSIS Security specifications.

Secure Network Management

The R-PHY operational paradigm dictates that the RPD principally operates under the control of the CCAP Core, which provides all the interfaces to the operators' OSS systems. From the perspective of runtime operation, the distributed CCAP should have the same look and feel as integrated CCAP. This can be accomplished through a GCP connection, secured through mutual authentication as explained earlier.

In addition to runtime management from the CCAP Core, the RPD itself will also support local network management interfaces. The scope of the functionality enabled through these interfaces is limited to debug, offline HFC plant maintenance and the management of a few standalone RPD features. RPDs will support an SNMP agent and may also support other types of NM interfaces such as local or network enabled console line interface or an HTTP interface.

As in any modern networking device, these interfaces must be secured with encrypted transport, access control lists and access authentication and multi-level privilege authorization. The RPD manufactures need to follow well-known practices and principles for securing network management access to devices.

Attention needs to be devoted not only to particular network management interfaces, but also to the system a whole. The operating system or system software in an RPD can have instances where the TCP/IP or UDP stack has ports open for services or purposes not used in production devices; examples include messaging clients/servers, debug and testing ports, etc. For security purposes, these ports need to be closed when the RPD is deployed in the network. TCP and UDP ports dedicated to these purposes must be closed by default. Similar requirement are applicable to physical Ethernet ports. Finally, the RPD needs to cryptographically protect all sensitive data held in the system, including passwords and security keys.

CONCLUSION

An R-PHY architecture provides sufficient advantages to the MSO so that we can be certain it will be deployed over a significant portion of the cable infrastructure. As RPDs are deployed in locations with no physical security they could create a security vulnerability for the network. The RPD is an intelligent software enabled device based on Ethernet and IP technology and thus could provide an attractive point of entry into the network.

Security threats are increasing, general public awareness of cybercrime is at an all time high and internet security has moved beyond the technical realm to the highest levels of media and government. The paper has described some of the major threats that an R-PHY architecture could enable including unauthorized network access, breaches of customer privacy, theft of service and denial of service attacks.

Fortunately there are a number of solutions that can be applied to secure the network. The paper has described a defense in depth strategy based on proven standard approaches to resolve the problems so that:

- Access to the trusted network is protected using IEEE 802.1X and MACsec.
- The RPD and the CCAP Core are mutually authenticated using X.509 certificates using IKEv2.
- The control connection between the CCAP core and the RPD is protected with IPsec.
- The data plane traffic is protected with existing DOCSIS and video encryption.
- The software integrity of the RPD is protected using secure software download and secure boot technology.

Validating security solutions can be a very complex task but all of these mechanisms have been proven in telecommunication industry deployments and have been the subject of in depth security analysis. Thus we are confident that they provide the best options for effective R-PHY security. In addition to their security credentials they also provide a cost effective option and a fast time to market by leveraging existing standards and silicon.

We believe that using the methods described an R-PHY based architecture can be deployed while maintaining stringent levels of network security. We also believe that the outlined defensive features can evolve along with the threat environment.

The mechanisms have been described using R-PHY as the primary example but can be generally applicable to any IP device located in an insecure location whether this is an RPD, an Ethernet switch or an OLT.

Acknowledgements

The authors would like to sincerely thank the following people who provided ideas, as well as valuable insight into both DOCSIS specific and general internet security issues and protocols, which was instrumental in producing this paper: **Nancy Cam-Winget, John Chapman, Dan Hegglin, Stuart Hoggan and Brian Weis**.

Abbreviations

AAA Authentication Authorization Accounting
AES Advanced Encryption Standard
BPI Baseline Privacy Interface
BPI+ Baseline Privacy Interface Plus
CA Certificate Authority
CCAP Converged Cable Access Platform
CMTS Cable Modem Termination System
CIN Converged Interconnect Network
CM Cable Modem
CPE Customer Premises Equipment
CRC Cyclic Redundancy Check
CVC Code Verification Certificate
DevID Device Identifier
DHCP Dynamic Host Configuration Protocol
DNS Domain Name System
DOCSIS Data-Over-Cable Service Interface Specifications
DoS Denial of Service
EAP Extensible Authentication Protocol
EAPOL Extensible Authentication Protocol Over LANs
FTC Federal Trade Commission
GCP Generic Control Plane
HFC Hybrid Fiber-Coaxial
IEEE Institute of Electrical and Electronics Engineers
IETF Internet Engineering Task Force
IGMP Internet Group Management Protocol
IKEv2 Internet Key Exchange Version 2
IOT Internet Of Things
IP Internet Protocol
IPsec Internet Protocol security
L2TPv3 Layer 2 Transport Protocol version 3
LAN Local Area Network
MAC Media Access Control
MSO Multiple Systems Operator
NAD Network Access Device
OLT Optical Line Terminal
OSS Operations System Support
PHY Physical Layer
PKI Public Key Infrastructure
R-PHY Remote PHY
RPD Remote PHY Device
RDC Regional Data Center
RF Radio Frequency
RFC Request For Comments
SNMP Simple Network Management Protocol
SSD Secure Software Download
STB Set-top Box
TCP Transmission Control Protocol
UDP User Datagram Protocol

References

[R-PHY SCTE] John T. Chapman, "DOCSIS Remote PHY", SCTE Cable-Tec Expo, 2013.
[R-PHY NCTA] John T. Chapman, "Remote PHY for Converged DOCSIS, Video and OOB", NCTA Spring Technical Forum, 2014.
[R-PHY System] CableLabs, "DOCSIS Remote PHY Specification, D03", 2015
[DOCSIS SECv3.1] CableLabs, "DOCSIS 3.1 Security Specification, CM-SP-SECv3.1-I01-141023"
[IEEE 802.1AR] IEEE Std 802.1AR™-2009, Secure Device Identity
[X.509] ITU, Recommendation X.509 (10/12), 2012

SFC in the DOCSIS Network

James Kim
Cable Television Laboratories, Inc.

Abstract

Service Function Chaining (SFC) is a concept that has been around for a while. Newer technology advancements in the SDN and NFV space now give operators a method to implement SFC architecture in the network. SFC also gives the operators two main benefits; flexibility in offering new services to customers and data path manipulation.

CableLabs' Open Networking Project, which includes participants from cable operators, the vendor community, and CableLabs staff, has been researching SFC and trying to identify the necessary DOCSIS components to enable SFC.

This paper looks briefly at SFC and the research work currently happening in IETF and then the use cases relevant for the DOCSIS network. Finally, the paper examines different methods of introducing SFC into the DOCSIS network, identifies the benefits for each, and provides a recommendation for cable operators.

INTRODUCTION

Service Function Chaining is a dynamic working group within IETF that has made significant progress over the years. Some of the active work areas within IETF include defining the problem statement that SFC intends to address or solve, defining the SFC architecture, and identifying different use cases for SFC.

The concept of SFC is that service functions or device functions that are embedded in network devices (router, switches, etc.,) can now reside in the core or in a data-center, and can be tied into the customer's service as needed. This allows operators to offer additional services to the customer as required and permits customers to choose which services they want or do not want on their account. The SFC also allows operators to redirect customer traffic to a certain path in the SFC domain as needed. For example, a Carrier-grade Network (CGN) function can reside in the SFC domain, and operators can redirect customer traffic to the CGN function before traversing out to the Internet.

Service functions can include but are not limited to:

- Firewall
- Parental Control
- Video Optimizers
- CGN
- Traffic Engineering
- Policy Management
- Load Balancers

This list of services can be revised when new services are introduced or older or unnecessary services are no longer needed.

SFC ENCAPSULATION METHOD

Several methods exist when implementing SFC in the network. Some of the methods are using Network Service Header (NSH), Generic Network Virtualization Encapsulation (Geneve), and OpenFlow (OF).

NSH

NSH is an IETF draft. The important notion with NSH is the metadata as the metadata carries important information pertaining to that frame/packet. The metadata identifies the different service chains a frame has to traverse

through. An encapsulation is added to the header to be forwarded in the SFC domain. The encapsulation ensures that the original frame will be forwarded to the SFC domain regardless of the underlying transport. Additional information regarding NSH can be found at IETF's website and also at OpenDaylight's wiki page.

GENEVE

Geneve is another method to implement SFC and, like NSH, is an IETF draft. Geneve came from network virtualization use cases where a virtual tunnel is created between two endpoints. The original intent of Geneve was to design an encapsulation technique that is flexible enough to meet the current virtualization use cases and also works with future use cases and needs. A key concept with Geneve is control plane independence from the tunnel endpoints. Another key concept is the ability to insert metadata into the frames (similar to NSH). While the creation of Geneve protocol was driven by network virtualization, the Geneve protocol can be used for SFC. Per Bruce Davie with VMware, Geneve takes the best of VXLAN, NVGRE and STT and puts it into a single protocol and encapsulation. Additional information regarding Geneve can be found at IETF's website.

OPENFLOW (OF)

Another method is using OF to implement SFC. The OF table can be used to identify packets or frames and take various actions. The OpenStack community is currently looking at using OpenFlow to enable service chain but the project is still in the incubation stage. Additional information regarding the OF specification can be found at Open Networking Foundation and information regarding the OpenStack project can be found at OpenStack's website.

Other methods of implementing SFC exist and newer methods could be developed in the near future. Each of these methods has pros and cons and will need to be reviewed in more detail at that time.

THE DOCSIS NETWORK

As stated in the introduction, SFC allows operators to easily insert or remove services. Figure 1 below shows a high-level diagram of SFC in the DOCSIS network and the data center. The two SFC edge boxes (shown in green) are the entry and exit points of the SFC domain. The various work happening within IETF and other SDOs defines the SFC domain. In Figure 1, it is assumed that there is an SDN controller for the DOCSIS network and a separate SDN controller for the data center where the SFC function resides. The orchestrator resides above the SDN controller and ensures end-to-end coordination. This model aligns with other standards, such as ETSI and MEF, looking at the orchestrator function.

Figure 1 also shows a use case in which CM1 might prefer to have Service 1, Service 2 and Service 3, whereas the CM2 would only want Service 4 and 5.

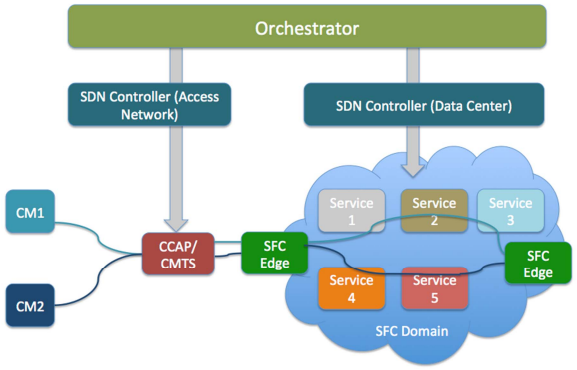


Figure 1 – Overview of SFC in the DOCSIS and Core Network

Figure 2 shows a high level description and the starting point of SFC. From the DOCSIS network, the SDN working group has identified four areas in which SFC can

start. The four starting areas for SFC are the application, the home gateway, the CM and the CMTS.

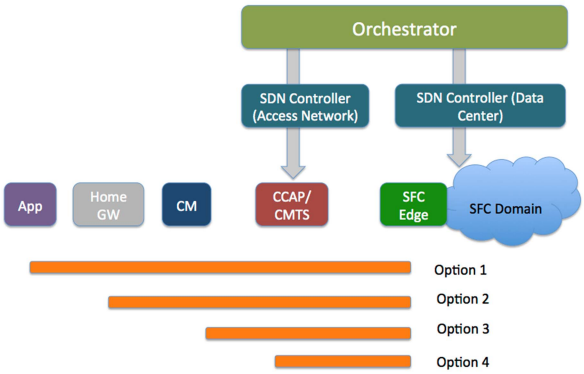


Figure 2 – Different starting points of SFC

The four options discussed next will look at SFC from the upstream direction only (from CM to CMTS). The SFC requirements for downstream have not been investigated at this time.

APPLICATION AS THE STARTING POINT

The application can be the starting point on the SFC. There are two methods from the application. The first is the overlay method which creates a tunnel (or a virtual connection) from the application to the SFC edge. In this scenario, everything from the application would be sent to the SFC. The DOCSIS network would not be aware of the application contents and there would be no classification done on the DOCSIS Network (CM/CMTS). This method would not work if the application needed to leverage the DOCSIS QoS mechanism or QoS mechanism from the CMTS to the SFC edge.

The second scenario is to mark the application so that the DOCSIS network and core network is aware. An example of this would be an MSO-provided VOIP service. The application would mark the traffic so that proper QoS could be applied on the DOCSIS network, and from the NSI side of CMTS to the SFC edge. Other MSO-provided applications, such as Video On Demand or home security, can leverage the QOS mechanism on DOCSIS and the rest of the network.

HOME GATEWAY AS THE STARTING POINT

The next starting point option is to initiate SFC from the home gateway or router. Within the home gateway option, there are several methods of implementation. The first is to use an overlay mechanism such as VXLAN or GRE from the home gateway to the SFC edge. This is similar to the first method when starting SFC from the application. The DOCSIS network and the path from the CMTS to the SFC edge would not be aware and all traffic from the home gateway would be sent to the SFC edge.

The next method is to have service chain(s) or traffic destined to the SFC edge marked and create a separate DOCSIS Service Flow. All other traffic not destined to the SFC edge would be on the primary service flow from the cable modem. This requires coordination between home gateway and the DOCSIS network so the DOCSIS network is fully aware that certain traffic is destined to the SFC edge.

CM AS THE STARTING POINT

The third option is to start at the CM. This option first appeared to be ideal, but once we started to investigate further and tried to identify the requirements from the CM, it became evident that initiating from CM was not ideal. Implementing SFC in the CM would not scale and there is no added value. The implementation that can be done on the CM can also be done on the CMTS (as described next).

CMTS AS THE STARTING POINT

The last option is to start SFC from the CMTS. Similar to the application and the home gateway, there are different methods of implementing on the CMTS. Figure 3 shows the three different methods of implementation from the CMTS.

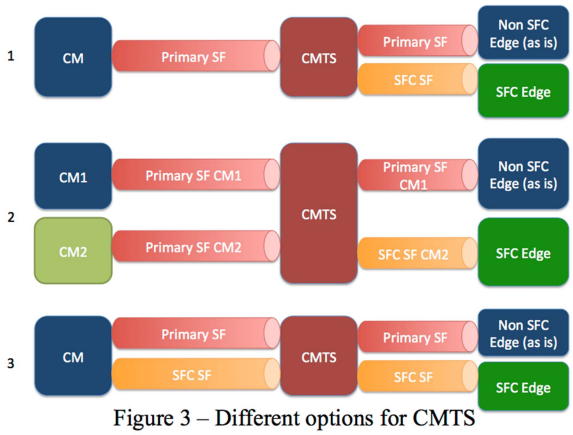


Figure 3 – Different options for CMTS

Figure 3 (CMTS option 1) shows the first method of CMTS initiating the service function. There would be one service flow in the DOCSIS network and all traffic, whether it is destined to the SFC edge or not. Once the traffic arrives at the CMTS, the CMTS would identify and separates traffic that is destined to SFC edge. This method is similar to having the CMTS perform a deep packet inspection on each frame/or packet as it arrives and leaves the CMTS.

The next method is per CM classification (as described in Figure 3, CMTS option 2). In this method, the CMTS identifies which CMs are destined for the SFC edge and which CMs are not destined for the SFC edge. Once identified, the CMTS can mark all traffic from that CM that is destined and send it to the SFC edge. The other CMs not destined for SFC edge will be routed/switched as is.

The last method (CMTS option 3) is to have multiple service flows per CM. The traffic destined for the SFC edge would be on a separate service flow and CMTS should be able to send that traffic to the SFC edge. The frames or packets not destined for SFC edge would use the primary service flow.

The first method of having the CMTS do the work in identifying the traffic type and marking will not scale. There are just too many CMs off the CMTS and each CM can have significant traffic depending on the number of users. The burden put on the CMTS might be too much at this time.

The second and third methods might be better choices, depending on the use cases and how operators want to deploy SFC. If the operator prefers to direct all traffic from the CM to the SFC domain, then the second option might be preferred. If the operator prefers to be more granular, then the third option might be preferred.

MARKING THE TRAFFIC

The above section discusses marking a customer traffic. The CMTS or the home gateway need to mark the traffic destined for the SFC edge, and the SFC edge needs the capability to read the marking. An example of marking the traffic might be to use the key field on the GRE header (if a tunneling mechanism is used) or to use the Geneve options field if using Geneve as the encapsulation method. If an operator is running an MPLS network, an MPLS shim header can also be used between the PE and the SFC edge.

RECOMMENDATION

Each starting point has pros and cons. The application starting point is very granular and there is doubt about whether the traffic should go through the SFC domain or not. Starting SFC from the application is useful when the operator has full control over the application. The downside to this method is that it is limited at this point; however, as MSOs start to expand into the home network, using an

overlay technique may be the easiest to implement. Unfortunately, it does not take full advantage of the DOCSIS network and the rest of the MSO's network. Further research is needed where the applications leverage the DOCSIS and the core network, not the preferred network.

As stated before, starting from the CM doesn't scale. Legacy CMs might not support it, and using it does not offer any advantages.

Finally, a combination of using the home gateway and the CMTS might be the best option available right now. More research is needed to identify what is needed between the home gateway and the CMTS, but this option allows SFC to take advantage of the DOCSIS network and the core network.

REFERENCES

1. Service Function Chaining (SFC) Architecture, draft-ietf-sfc-architecture-07, https://datatracker.ietf.org/doc/draft-ietf-sfc-architecture/

2. Network Service Header, draft-quinn-sfc-nsh-07, https://datatracker.ietf.org/doc/draft-quinn-sfc-nsh/

3. VM Ware Blog on Geneve, VXLAN and Network Virtualization Encapsulations, http://blogs.vmware.com/cto/geneve-vxlan-network-virtualization-encapsulations/

4. Geneve: Generic Network virtualization Encapsulation, draft-gross-geneve-02, https://datatracker.ietf.org/doc/draft-gross-geneve/

5. IETF RFC 2784, Generic Routing Encapsulation.

6. OpenStack Neutron, Service Function Chaining using OpenFlow, https://blueprints.launchpad.net/neutron/+spec/service-function-chaining-using-openflow

The Edge Resource Center: Leveraging NFV and SDN for High Availability/High Performance Network Functions

Jeff DeMent, Erich Arnold, and Mircea Orban
ARRIS Group, Inc.

Abstract

Service providers around the world are embracing Network Function Virtualization (NFV) and Software Defined Networks (SDN) as enabling technologies to help increase service velocity and reduce costs. While some network functions may be suitable for deployment in a centralized data center, others (such as content caching and distributed denial of service attack (DDOS) detection and mitigation) are most effectively deployed at the network edge in order to meet the necessary latency targets and avoid degrading the user's quality of experience. This paper explores the concept of an Edge Resource Center, a small-scale virtual machine infrastructure (VMI) co-located with the CCAP devices in a cable head-end that supports high availability and high capacity SDN/NFV resources.

INTRODUCTION

Software Defined Networks (SDN) and Network Function Virtualization (NFV) are creating enormous interest as enabling technologies to help increase service velocity and reduce costs. SDN helps increase service velocity by adding a level of programmability to the network that's not typically found in purpose-built network elements. This programmability typically comes in the form of abstract RESTful web-service style application program interfaces (APIs) for configuring, monitoring, and managing network elements and services, allowing service provider and third party developers to quickly create new network applications. This promises the opportunity for increased automation of the service provisioning and service management functions, including the potential for end user self-provisioning, resulting in faster and more accurate deployment of new services.

Network Function Virtualization aims to reduce network operator CAPEX and OPEX by implementing network functions on commercial off-the-shelf (COTS) servers rather than on dedicated hardware appliances. NFV takes some of the technologies developed over the last decade to enable large scale multi-tenant data centers– namely virtualization of compute and storage resources, in many cases implemented with open source software – and applies them to network functions that have typically been implemented with ASICs, specialized packet processors, and embedded software. Using SDN or other techniques, multiple virtual network functions (VNFs) can be linked together into a *service chain*, applying the appropriate functions to each user's traffic in the proper order. Standard data center virtual infrastructure management software can be employed to create a dynamic elastic network service infrastructure where common compute, storage, and network resources are allocated to specific network functions based on demand.

One of the value propositions for SDN and NFV deployments proposed for cable MSOs is the opportunity to move network service functions onto standard virtualization infrastructure in large centralized data centers. While some network functions may be suitable for deployment in a centralized data center, others (such as content caching and

DDOS detection and mitigation) are most effectively deployed at the network edge in order to use network bandwidth more efficiently and to meet the desired latency targets to avoid degrading the user's quality of experience. This paper explores the concept of an Edge Resource Center (ERC), a small-scale virtual machine infrastructure (VMI) co-located with the CCAP devices in a cable head-end that supports high availability and high capacity SDN/NFV resources, along with the enhancements needed to allow the CCAP device to participate in the distributed NFV service chain.

PROBLEM STATEMENT

The motivation for MSOs to develop an architecture for network services based on SDN and NFV is well-documented: costs to satisfy exponentially increasing bandwidth demand are growing faster than service revenues, making rapid development and introduction of new revenue-generating services a necessity.

To meet the network operational goals needed to address this business challenge – increased service velocity, reduced equipment cost, and lower operational complexity – the industry is looking to technologies and solutions that have been successful for the large-scale cloud computing providers. Some of the desired characteristics are:

- A cloud-based architecture using standard IT virtualization technology
- Use of standard COTS servers to replace multiple network appliances, reducing costs
- Service scalability, including the ability to locate network functions at different locations
- Elastic compute and network capability for efficient resource utilization with changing service demands
- Network programmability, based on RESTful APIs, for rapid service creation and increased automation
- Open standards and interfaces, to avoid vendor lock-in
- Co-existence and compatibility with existing network equipment

This is an aggressive list and it remains to be seen whether many of these benefits can be realized. Some of the challenges to be addressed are listed below.

Packet latency: Diverting each user packet to a centralized data center and forwarding it through multiple service functions, each of which inspects the packet and makes forwarding decisions, can add significant latency and impact the user's perceived quality of experience for latency-sensitive applications such as gaming. When the service functions are running in virtualized environments, the packet latency can be further degraded by multiple hops through virtual switches and guest O/S stacks.

Software complexity/reliability: Mission-critical services need to be engineered to the same "5-9s" level of reliability as the existing access network equipment such as modern CCAP platforms. Development of fault-tolerant software has always been a challenging exercise, even in a single-vendor homogenous hardware environment. Integration of loosely-coupled software modules from multiple vendors into a highly-available service offering adds significant complexity to the modules themselves and to the SDN controller/orchestrator and associated applications responsible for managing the service.

Congestion avoidance and mitigation: Adding a centralized programmable SDN controller with a global view of network operating conditions would seem to simplify the job of steering user traffic around congested points to keep services operating

efficiently. However, programmability can be a double-edged sword; SDN applications (potentially from different vendors) can choose conflicting recovery actions, resulting in congestion spread rather than abatement. Techniques to isolate congestion and apply consistent mitigation policies are needed.

The remainder of this paper provides an overview of some of the candidate SDN and NFV technologies under investigation and introduces the Edge Resource Center as an architectural model for addressing these challenges.

TECHNOLOGY BACKGROUND

SDN Overview

SDN traces its history back through at least three decades of research and experimentation in programmable networks in the telecom (e.g., Advanced Intelligent Network) and data networking fields [1], but has really gained traction with the development of the OpenFlow™ protocol in 2007-2008 and its subsequent standardization by the Open Network Foundation. This generation of SDN attempts to break the dependence on network equipment vendors (and the resulting long development times) to implement new network features by completely separating the control plane (e.g., policy and routing) from the forwarding plane. This allows the control plane to be moved to general purpose computing platforms, using OpenFlow technology to program the forwarding tables in the now simplified (and theoretically cheaper) data plane forwarding equipment.

In recent years the SDN focus, particularly in service provider networks, has shifted from complete decoupling of the control and forwarding planes, to augmenting existing networks with additional the programmability needed to rapidly design, prototype, test, and deploy new network services. This requires programmatic access to network configuration and monitoring functions as well as programmable forwarding. Accordingly, additional protocols beyond the OpenFlow protocol are now considered as part of the SDN toolset: NETCONF and SNMP for configuration and management, PacketCable™ Multimedia (PCMM) for policy and QoS in the cable access network, and others.

SDN Architecture

The main distinguishing feature of the SDN architecture is a software-based SDN controller, which discovers (or is configured with) the network topology, provides abstract "northbound" APIs to applications for network control and management, and communicates forwarding, configuration, and policy instructions to the network elements it controls. A number of both open source and proprietary SDN controllers have been developed over the past few years. Currently one of the more popular controllers is from the open source OpenDaylight project, referred to from here on as ODL (figure 1).

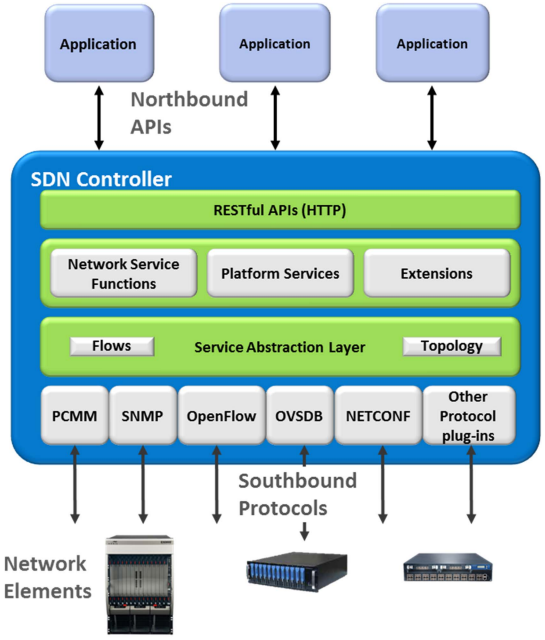


Figure 1: OpenDaylight Controller Architecture

The ODL controller consists of an extensible model-driven service abstraction layer that performs the mapping between API requests from applications (or internal higher-level services) and the southbound protocols that execute the requests via interactions with network elements such as switches (physical and virtual), routers, and CCAP devices. Southbound protocols are implemented as Java™ plug-ins, allowing new protocols to be added without requiring frequent new releases. ODL's capability to support legacy protocols such as SNMP and PCMM, as well as vendor-specific protocols, is an important feature for interoperating with existing cable network infrastructure.

Early SDN proponents favored a logically centralized SDN controller (although possibly distributed across multiple compute nodes for increased capacity and/or redundancy), on the theory that a centralized controller with global knowledge of the network topology and current conditions could help optimize packet forwarding decisions in the network elements. While this may be feasible for smaller networks, it becomes impractical for large geographically dispersed service provider networks with a variety of equipment types. SDN architectures for service provider networks today generally include multiple domain-specific controllers, for example, an access-network controller, data-center controller, and core network controller as illustrated in figure 2. These controllers may then be coordinated by an orchestration layer or "controller of controllers."

The prototype ERC includes a highly available OpenDaylight SDN controller cluster that controls the CCAP devices, local switches, and virtual switch infrastructure for the entire head-end. It's anticipated that at some point in the future the ERC controller may contain an interface to an orchestrator or higher-level controller that coordinates the activities of all head-ends and other network regions.

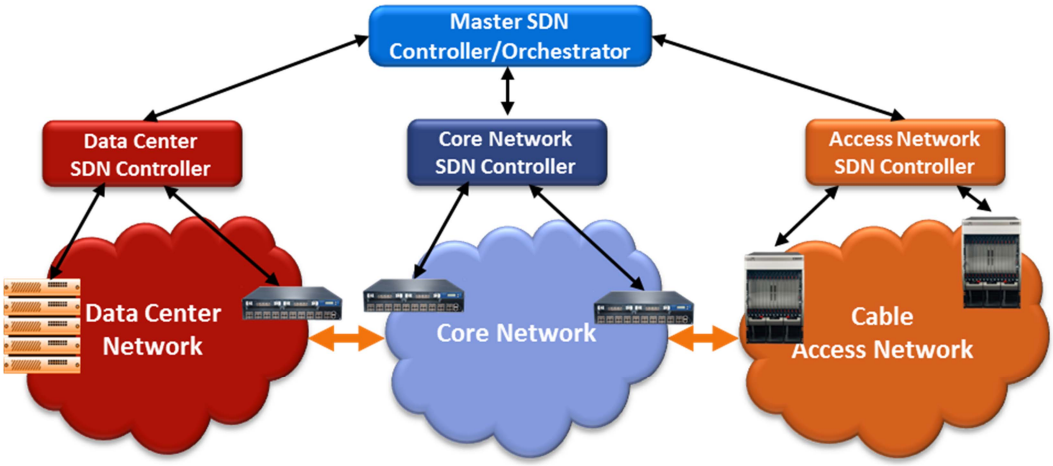


Figure 2: Distributed Domain-specific SDN Controllers

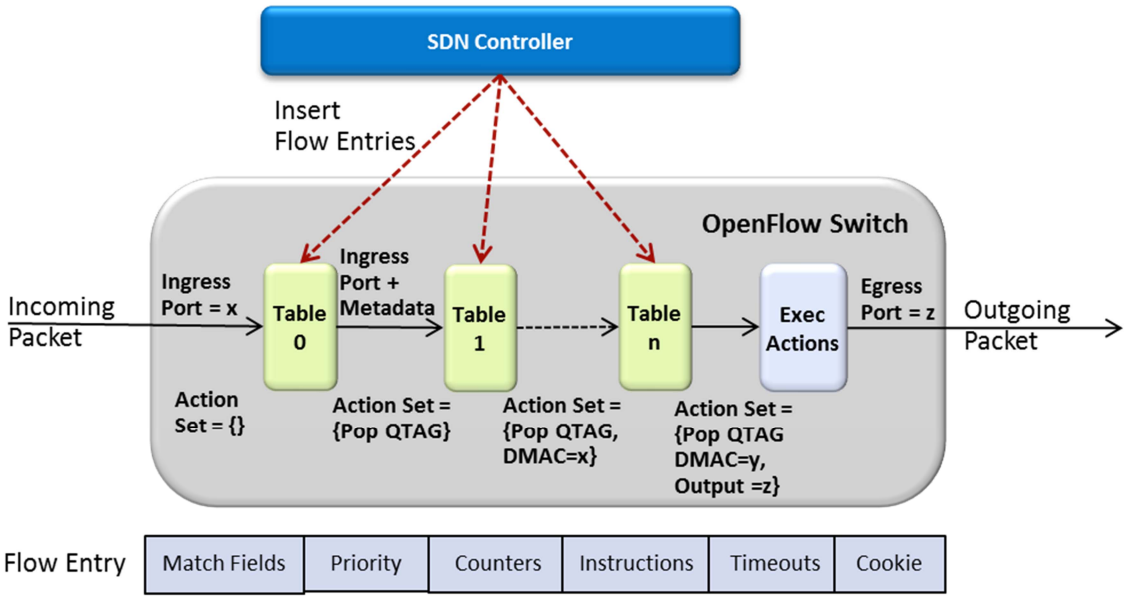


Figure 3: OpenFlow Switch Model

OpenFlow Protocol

The OpenFlow protocol provides a standardized interface for a control process (e.g., an SDN controller) to program flow tables in switches and routers [2]. The OpenFlow switch model is comprised of one or more flow tables, with each table entry containing a matching rule (set of header fields, including wildcards) used to match against incoming packets, along with a set of actions to perform on matching packets such as output port selection, header field overwrites, and counts to increment (figure 3). Later versions of OpenFlow technology allow multiple tables to be chained together, with packet metadata (intermediate results) passed from table to table to be used as additional matching criteria, and the resulting actions accumulated until the end of the chain. The OpenFlow switch initiates a secure connection to its designated controller to advertise capabilities and accept flow programming commands.

While OpenFlow technology has been proposed as a replacement for the embedded switch/router control plane, we propose here using it to augment the existing control plane in cable access network elements. Adding OpenFlow capability to a CCAP device allows for fine-grained dynamic traffic steering of packets at the network edge. For example, reacting to a detected DDOS attack by steering all packets destined for the attack target through a stateful firewall that scrubs out the attack packets while allowing the valid traffic to pass. OpenFlow protocols can also be used to program virtual switches to direct flows between network functions in the ERC.

Other SDN Protocols

Several existing protocols have also been repurposed for use in SDN, particularly for support of service provisioning via the SDN controller. We have already mentioned PCMM, a cable-specific protocol for configuring dynamic DOCSIS® service flows with appropriate QoS attributes, and SNMP for network management, particularly status and statistics collection. Both are implemented in the prototype ERC.

Two other SDN-related protocols that are being considered for future use in the ERC are

NETCONF and BGP-Link State (BGP-LS). NETCONF is an IETF®-originated protocol for modifying configuration data in network elements. In particular it differentiates the "running" (active) configuration data store from the "startup" configuration data store and "candidate" configuration data stores that may become active when committed. NETCONF data stores are modeled by the YANG data modeling language, which can be directly mapped to XML. The NETCONF protocol is based on a remote procedure call (RPC) model.

BGP-LS is a set of extensions to the well-established Border Gateway Protocol to allow BGP-speaking network elements to share link state and traffic engineering information with external agents such as an SDN controller. Its main purpose is to enable topology discovery in the SDN controller [3].

Network Function VirtualizationOverview

NFV aims to change the network operator cost structure by moving from the current practice of deploying a separate hardware-based network appliance per function to a model with all software-based functions on running COTS hardware. With NFV, each server is capable of performing multiple roles by leveraging standard IT industry compute, storage, and network virtualization techniques [4].

The ETSI NFV Industry Specification Group has taken the lead in publishing a number of documents, including general requirements, a reference architectural framework, use cases, and best practices for performance and portability. The ETSI NVF architecture includes a standard virtual machine infrastructure (VMI) capable of hosting one or more virtualized network functions (VNFs), each of which may also have an element manager to manage it. The architecture is agnostic to the actual virtualization technology (e.g., hypervisor) employed.

The reference framework also includes an NFV management and orchestration component (known as MANO) which is responsible for lifecycle management of the VNFs [5]. The MANO is responsible for normal FCAPS functions. It also has the responsibility for understanding the operational characteristics of the VNF (called key performance indicators, or KPI) and scaling the VNF up by adding resources (for example, CPU or memory) or creating new VM instances, or scaling down by reallocating resources or removing VM instances. The Orchestrator component of the MANO also integrates with the operator's OSS/BSS system.

The prototype ERC deploys a highly available OpenStack® cluster to implement the MANO virtual infrastructure management component. VNF management (e.g., dynamic scaling of VNF capacity) and orchestration are topics for future research.

VMs and Docker™ Containers

Virtual Machine technology has been widely deployed in data centers for Web-based applications and is the basis for NFV. VM technology is usually implemented with a hypervisor that logically partitions a server's physical resources – CPUs, memory, disk space, and network interfaces – and allocates the resource slices to individual VMs. Each VM contains its own operating system instance (called the guest O/S), allowing for different operating systems to be running in different VMs on the same physical server. These extra layers of software add overhead and can be the source of additional packet latency for data plane VNFs.

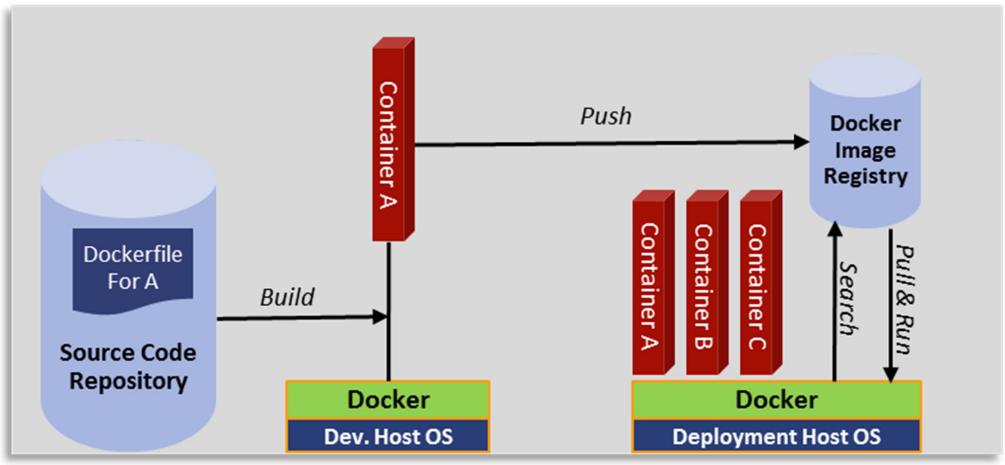


Figure 4: Example Docker Build and Deployment Model

Docker containers are independent but complimentary technology that has become popular recently. The Docker architecture is an open platform for distributing and deploying cloud applications based on the Linux container model [6]. Docker containers are used to provide an isolated runtime environment for the application that encapsulates all components, libraries, and configuration dependencies. Docker containers are hardware and platform-independent, allowing applications packaged as Docker containers to run on many different platforms (both physical and virtualized) and improves portability across different operating systems. Figure 4 illustrates the Docker build and deployment model for applications packaged in Docker containers.

A Docker container differs from a virtual machine in that it only encapsulates the application and its dependencies, but not the guest operating system. This makes it lighter weight than a VM – it runs as a process in user space on a host operating system – but still provides isolation from other applications running on the same platform. Note that the actual deployment platform for the application may be either a physical server or a VM itself.

EDGE RESOURCE CENTER OVERVIEW

An Edge Resource Center (ERC) is a small-footprint virtual machine infrastructure (VMI) co-located with a group of CCAP devices in a cable head-end that provides highly available/high performance SDN/NFV resources (figure 5). Locating these resources at the network edge helps to control packet latency for network functions requiring minimal and/or predictable latency to provide users with the desired quality of experience. It can also help reduce network bandwidth utilization by keeping traffic that can be processed locally in the head-end rather than back-hauling it to a centralized data center, only to be forwarded back to the originating head-end.

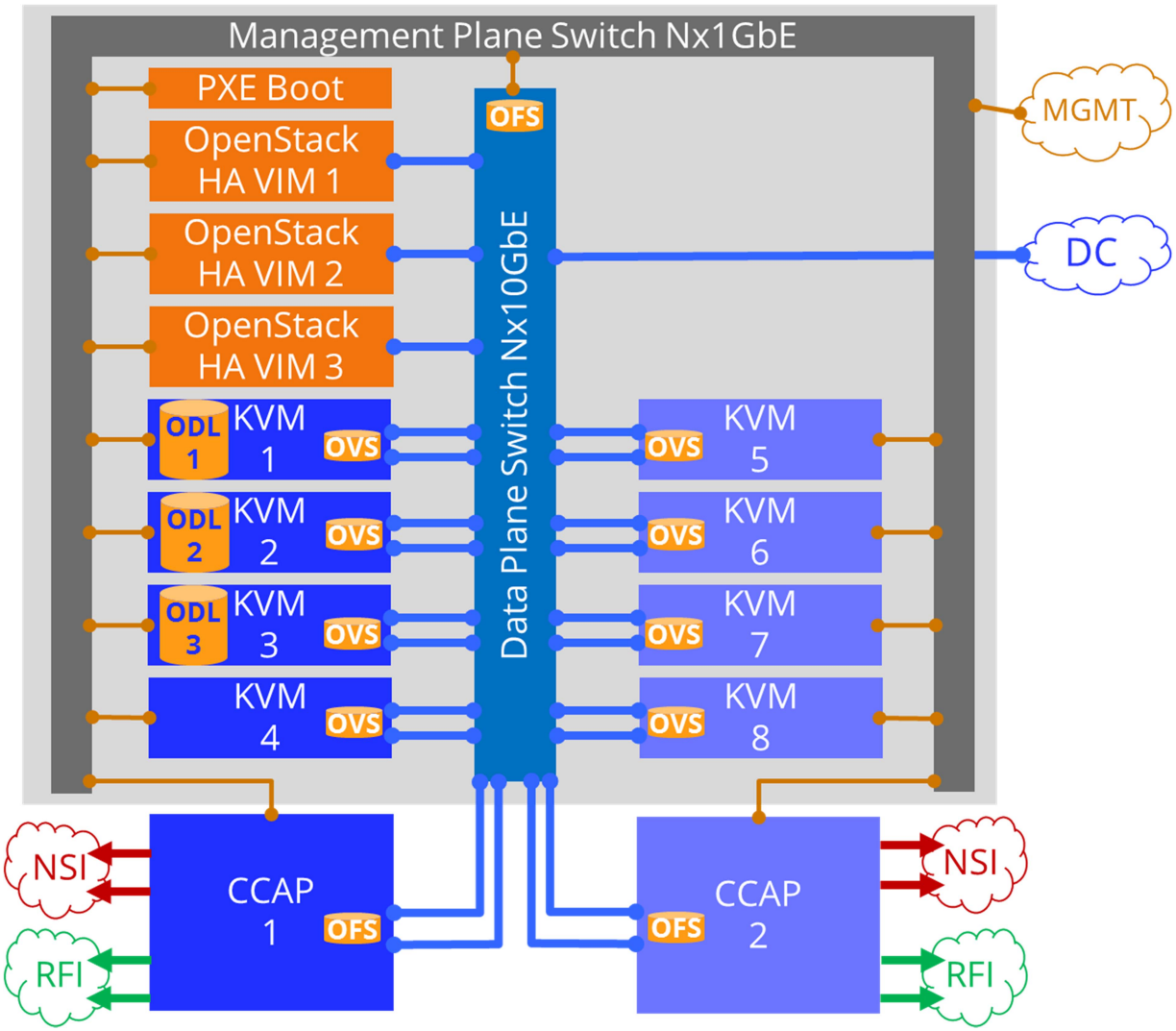


Figure 5: Edge Resource Center Model – Scaled for 2 CCAP devices

ERC Functional Components

The primary components of the ERC are:

COTS Server Array: The ERC hosts are low cost COTS servers running the KVM (Kernel-based Virtual Machine) hypervisor that provide high capacity elastic compute and network resources. Each server runs an Open vSwitch (OVS) instance that provides layer 2 switching between VMs in the same server as well as access to external networks, and can be controlled via OpenFlow technology. This configuration supports a large population of lightweight VMs that provide local subscriber services such as virtualized firewalls, deep packet inspection (DPI) filters, edge caches, and virtual CPE. The servers typically provide only minimal storage resources, but can be upgraded to support edge services such as content caching that have more significant storage requirements.

OpenStack VIM: The ERC VMI is managed by a local OpenStack Virtual Infrastructure Manager in a high-availability cluster controlling the KVM hypervisor hosts (compute/storage nodes) in a distributed redundant configuration. OpenStack software

provides a web-based dashboard for creating and removing VM instances, configuring virtual networks, and managing virtual storage volumes. It also includes an image service for managing VM images with different guest operating systems and configurations.

OpenDaylight SDN Controller: An OpenDaylight SDN controller (ODL) cluster is also hosted in a high-availability VM configuration on the KVM hosts, to support dynamic network configuration and automation. The ERC ODL configuration includes a PCMM protocol plug-in to support dynamic service flows with specified QoS attributes in the cable access network, as well as OpenFlow protocol support for dynamic service chaining and SNMP for configuration and monitoring.

Switched Management Network: An Nx1 Gbps switched management network connecting all devices.

Switched Data Plane Network: An Nx10 Gbps switched data plane network connecting the servers and CCAP devices, with the potential to connect to remote data centers. The data plane switch also utilizes OpenFlow technology.

CCAP Devices: The CCAP devices provide the termination point for the cable access network. The ERC CCAPs include a prototype OpenFlow switch capability to augment their standard routing and switching capability.

Service Availability and Performance

One of the primary goals of the ERC is to support virtual network functions at the same "5-9s" level of reliability as the existing CCAP platforms. Achieving this level of reliability requires that both the network function itself and the supporting infrastructure services – i.e., the OpenStack VIM and the OpenDaylight SDN controller – operate at this level of reliability.

Since the underlying server hardware does not typically meet this target, software redundancy techniques are needed to meet this goal. Both OpenStack and ODL are deployed in the ERC as 3-instance clusters, using existing open source clustering and data replication technologies. Interfaces to applications employ virtual IP addresses that can float between cluster instances.

High availability for individual network functions running on the ERC is up to the VNF developer. Different redundancy models are used depending on whether the VNF is stateful or stateless. Stateless VNFs are commonly deployed in an active/active model using load-balancing. A failure of a VNF instance using this model can affect as little as a single transaction, with remaining instances taking over the load until a new VNF instance can be created. Availability for stateful VNFs is a little more complicated, typically requiring multiple instances of the VNF in active/backup arrangements, with state data replicated or check-pointed from the active instance to the backup instances. A failure of the active instance in this case requires interaction with the SDN controller to reconfigure the traffic flow towards the backup instance, with a resynchronization phase later when the failed VNF instance is restored. The same clustering and replication technologies employed by OpenStack software and ODL may be used by VNFs using either model.

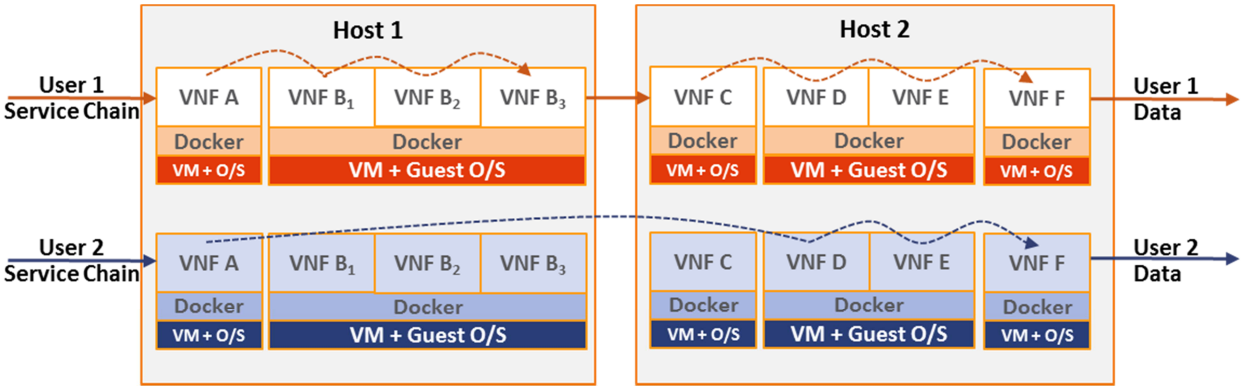


Figure 6: The VM Container Ship Model

Key performance characteristics for any deployed network function are bandwidth and packet latency. In virtualized environments, delivering predictable bandwidth and packet latency depends on many factors: the capability of the underlying hardware (e.g., number of CPU cores, threading capability, memory and I/O bandwidth), the performance of the virtualization software and O/S stacks (including the virtual switch implementation), and the loading factors of other functions sharing the platform to name just a few.

The ETSI NFV Industry Specification Group has published a set of best practices for performance and portability of NFV data plane workloads to help guide network operators in making appropriate choices for virtualization infrastructure [7]. The recommendations include suggested requirements for the architecture and capabilities of both the underlying hardware and hypervisor/host O/S. Also included are proposed templates for specifying the capabilities of a compute host and its virtualization infrastructure and for specifying the requirements of a VNF virtual machine image. The virtual infrastructure manager or orchestration system could then match VNF image requirements against host capabilities when deciding where to locate the VNF.

The VM Container Ship Model

One of the models under evaluation in the prototype ERC to help manage both performance and availability of VNFs is the use of a combination of VMs and Docker containers to group related VNFs or multiple components of a single logical VNF. In this model, the VM acts as a "container ship" running the related Docker containers as independent processes on the VM's guest O/S (figure 6).

This model can have several advantages.

- The related VNF components share a single guest O/S image and can use more efficient inter-process communication techniques than switching between VM images through the virtual switch.
- The related VNF components can comprise a single "fault group" with a *shared fate*. This means that the failure of a single VNF component fails the group. This can simplify the design of the VNF components since each component does not have to be designed to deal with the failure and recovery of its dependent components; instead the entire VM instance can be destroyed and restarted.

- Multiple small VNFs in a per-subscriber service chain could be co-located in a single VM image. This "vertical slicing" of related VNFs at the per-subscriber level can limit the impact of a VM failure to a single subscriber and take advantage of the improved performance and simplified recovery strategy described above.

USE CASES

Two possible use cases for ERC-based services are discussed below: DDOS detection/mitigation and network DVR playback over IP. The DDOS protection use case illustrates the use of SDN with the OpenFlow protocol for dynamic service chain creation, while the nDVR playback case illustrates the use of SDN with PCMM for dynamic bandwidth allocation in the cable access network.

DDOS detection/mitigation

MSO-provided DDOS protection is a good candidate for deployment using an SDN and NFV-based architecture in the ERC. A typical implementation for DDOS protection today in an MSO network might be to connect a dedicated DDOS appliance supporting a small number of subscribers in the data path between the Internet and the CCAP device serving the subscriber. The DDOS appliance examines all traffic to and from the subscriber, mostly just passing it through except for the rare times when an attack is in progress.

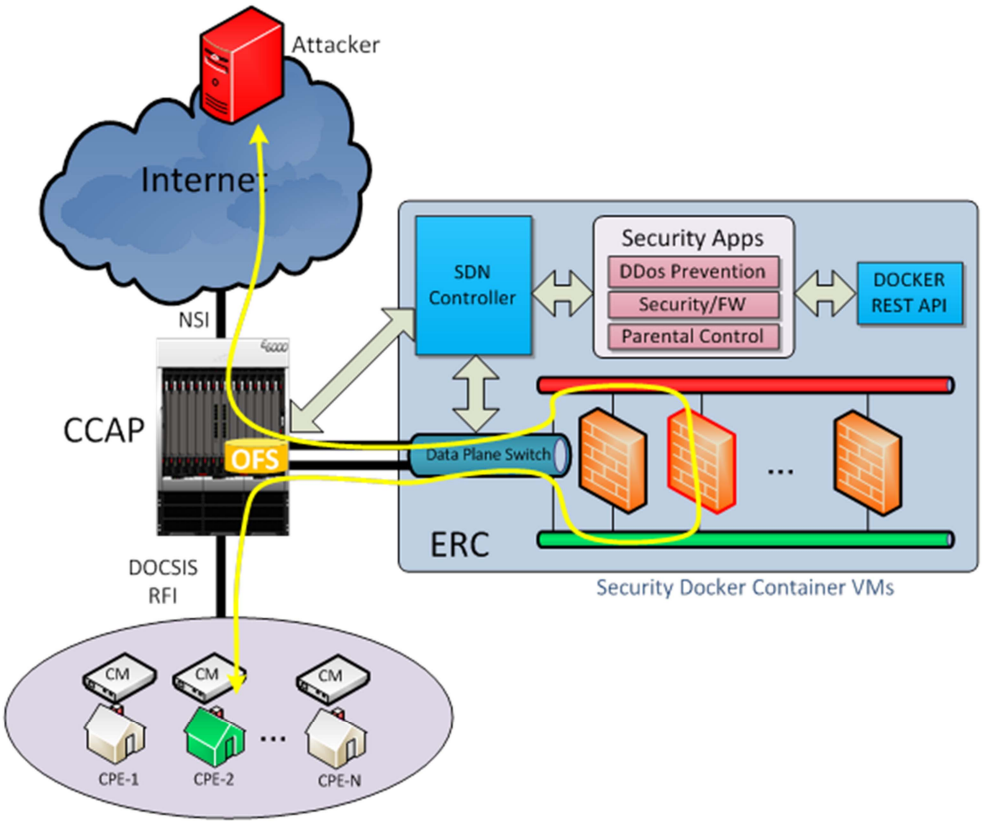


Figure 7: Example DDOS Detecton/Mitigation Service in ERC

An SDN/NFV approach using ERC resources (figure 7) instead would have 2 components:

- A DDOS detection application implemented as an SDN application, using OpenDaylight northbound APIs to monitor traffic statistics and derive usage patterns for DDOS protection subscribers.
- A virtualized stateful firewall network function deployed in the ERC as a Docker container or as a standalone VM. The firewall can be dynamically created when a DDOS attack is detected or, for faster response or simplified load balancing, a firewall instance can be statically created for each subscriber.

The dynamic DDOS protection service works as follows.

1) The user logs into a self-service application to provision the DDOS service. Depending on the details of the service, the IP address or addresses to be protected may be derived from this request, or might be manually entered by the subscriber (other provisioning scenarios are possible, not considered further here).
2) The self-service application notifies the DDOS detection application of the subscriber IPs to be protected. The DDOS detection application begins monitoring traffic statistics for the designated IP addresses via the ODL instance in the head-end serving the user. All data traffic continues to be passed straight through the network to/from the user device.
3) When the onset of a DDOS attack is detected, the DDOS application requests a stateful firewall instance to be allocated in the ERC. Using the OpenFlow protocol (via ODL), the application inserts flow entries in the serving CCAP device to forward all traffic to and from the attacked IP addresses through the stateful firewall VNF.
4) The firewall VNF monitors traffic in both directions to identify legitimate downstream traffic, passing it back through the CCAP device to the user. Attack traffic is discarded at the firewall.
5) Once the attack has ended, the DDOS application removes the CCAP flow entries for that subscriber, and traffic flow resumes directly from/to the user without going through the stateful firewall VNF. The firewall VNF may be destructed at this point or may just go idle, waiting for the next attack on that subscriber.

Network DVR Playback

The IP nDVR playback use case (figure 8) uses the PCMM capability of the ERC SDN controller to provision a new service flow across the DOCSIS network with desired QoS characteristics for the user's playback session. The new service flow serves 2 purposes: (1) allocating the required DOCSIS bandwidth for the video playback session depending on the video quality (SD, HD, or UHD) and (2) ensuring the video playback is not counted in the user's normal High Speed Data service counts.

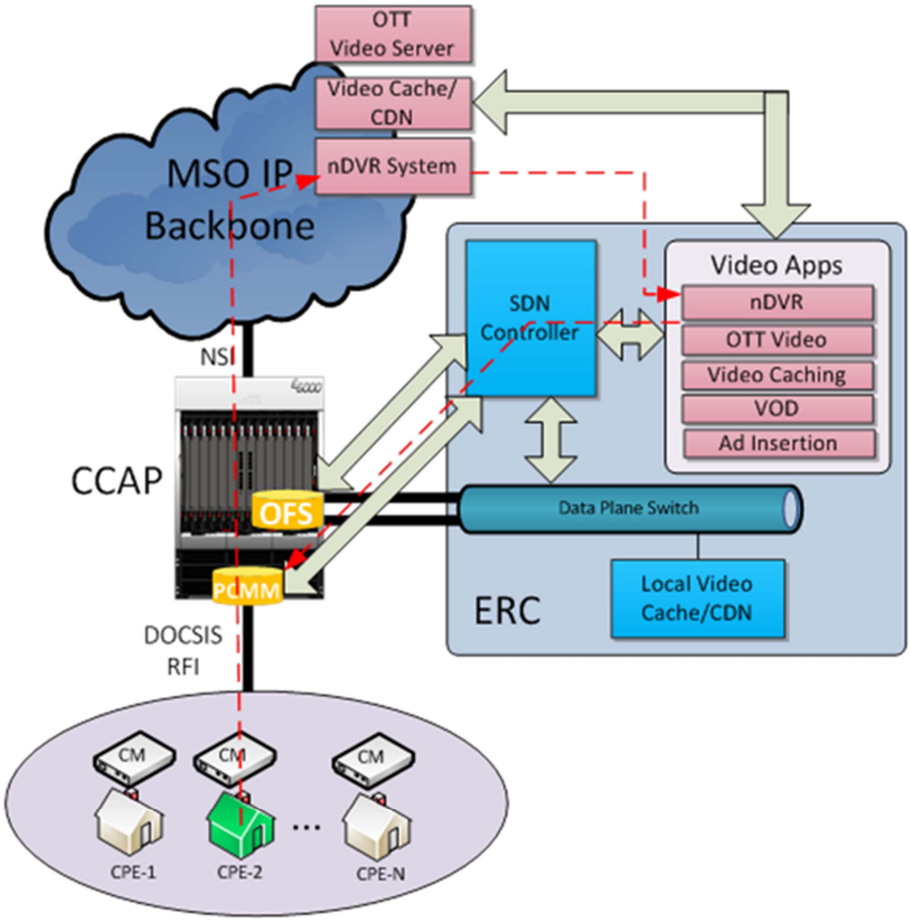


Figure 8: ERC nDVR Playback Example

The nDVR playback operation is fairly simple.

1) At the time of the playback request, the nDVR server notifies the nDVR application in the head-end ERC of the user's request, including the destination IP address and port number to be used for playback and the required bandwidth.
2) The local nDVR application uses the ODL PCMM API to provision a new service flow for the playback session, including the classifiers (IP address & port number) to identify the playback traffic and the required QoS settings.
3) The downstream playback session IP packets are classified to the newly created downstream service flow by the CCAP device serving the subscriber.
4) At the end of the playback session, the nDVR server notifies the local nDVR application to remove the associated service flow.

CONCLUSIONS & FUTURE WORK

Network operators have embraced SDN and NFV as critical technologies for improving service velocity and reducing costs, but some network functions have latency and availability requirements that may be difficult to meet using a standard centralized IT cloud infrastructure. Locating these functions at the network edge can be an effective strategy to ensure predictable performance that provides users with the

desired quality of experience. It can also help reduce network bandwidth utilization by keeping traffic that can processed locally in the head-end rather than back-hauling it to a centralized data center, only to be forwarded back to the originating head-end.

We propose deploying these services in an Edge Resource Center (ERC), a small-scale virtual machine infrastructure co-located with the CCAP devices in a cable head-end/hub that supports high availability, high capacity SDN/NFV resources. The ERC includes high-availability OpenStack and OpenDaylight clusters and supports a large population of lightweight VMs that provide local subscriber services such as NFV firewalls, DPI filters, edge caches, and virtual CPE. Adding OpenFlow capability to the CCAP devices takes advantage of their native packet classification functions to add programmable dynamic service chaining capability to the ERC.

The ERC supports a deployment model that groups related VNFs or VNF components packaged in Docker containers onto a single VM that acts as a "container ship" to help offset the overhead introduced by the hypervisor and vSwitch for packets that traverse multiple VNFs. The VM container ship also acts as a fault group for the related VNF components to help simplify VNF failure and recovery strategies.

As SDN and NFV technologies are evolving rapidly there will surely be new capabilities needed in the ERC. The following are just a few of the capabilities being considered for future study:

- Adding additional service chaining capabilities such as the Network Service Header (NSH) being standardized in the IETF SFC work group for passing packet metadata between VNFs;
- Adding NETCONF or other protocols to add additional support for automating service provisioning;
- Integrating the ERC with an orchestration system, or "controller of controllers", to support distributed service chains with reach beyond the head-end; and
- Adding data analytics capabilities to facilitate new applications in the ERC.

REFERENCES

[1] J. R. a. E. Z. N. Feamster, "The road to sdn: an intellectual history of programmable networks," *ACM SIGCOMM Computer Communication Review,* vol. 44, no. 2, April 2014.

[2] N. McKeown and others, "OpenFlow: Enabling Innovation in Campus Networks," *ACM SIGCOMM Computer Communication Review,* vol. 38, no. 2, April 2008.

[3] H. Gredler, J. Medved, S. Previdi, A. Farrel and S. Ray, *North-Bound Distribution of Link-State and TE Information using BGP,* IETF draft draft-ietf-idr-ls-distribution-10.

[4] ETSI Industry Specification Group for NFV, "Network Functions Virtualisation – Introductory White Paper," October 2012. [Online]. Available: http://portal.etsi.org/NFV/NFV_White_Paper.pdf.

[5] ETSI GS NFV-MAN 001 V1.1.1, *Network Functions Virtualisation (NFV); Management and Orchestration,* 2014-12.

[6] "What is Docker?," [Online]. Available: https://www.docker.com/whatisdocker/. [Accessed 4 March 2015].

[7] ETSI GS NFV-PER 001 V1.1.2, *Network Functions Virtualisation (NFV); NFV Performance & Portability Best Practises,* 2014-12.

The Evolution of Cable Network Security

Matt Tooley, NCTA

Matt Carothers, Cox Communications

Michael Glenn, CableLabs

Michael O'Reirdan, Comcast

Chris Roosenraad, Time-Warner Cable

Bill Sweeney, Comcast

Abstract

The challenges of protecting networks from cyber threats have evolved from computer viruses and spam plaguing desktop computers to advanced persistent threats and cloud-based multi-gigabit distributed denial of service attacks. This paper examines the evolution of the attack vectors to the DOCSIS™ networks and their associated attack surfaces. It then surveys the tools that are in the DOCSIS standards today that can be used to mitigate the threat vectors. Further it examines the evolution of cable networks technologies such as software defined networking and big data. Finally it then proposes methods to integrate software-defined networks and big data to improve the overall robustness and security of the cable networks.

CABLE NETWORK ATTACK SURFACES

We define an attack surface as the exposed, exploitable vulnerabilities that an adversary can leverage to help them accomplish their objectives. In other words, every point that a system may interact with the outside world is part of its attack surface, and each presents some amount of risk. Intuitively, the "smaller" the attack surface, the less likely a vulnerability will exist on a system that an adversary can exploit. Examples of the attack surface in the real world include (Northcutt n.d.):

- Open ports on servers with software accepting and processing packets on those ports
- Services available inside a firewall
- Software that processes incoming data, email, XML, files, and industry-specific custom data exchange formats (EDI)
- Interfaces, SQL, web forms
- Employee access to sensitive information that may be socially engineered

The attack surfaces can be further broken down into:

- Software Attack Surface – vulnerabilities exposed through applications, databases, and operating systems.
- Human Attack Surface – human vulnerabilities exploited through social engineering, human error, violation of trust by an insider, and/or employee attrition.
- Hardware attack surface – vulnerabilities existing in hardware architectures,

microcode, firmware and BIOS code specific to hardware.
- Network Attack Surface – combining vulnerabilities in a coordinated way through the network to achieve greater results than leveraging the vulnerabilities in isolation.

Software Attack Surface

We define an attack surface as the exposed vulnerabilities that an adversary can exploit to assist them in meeting their objectives. Their objections can include confidential data theft, data destruction, system or hardware destruction using software, denial of service, data or service corruption (including mis-information, inaccurate data and propaganda) and others. An important point is that vulnerabilities must be exposed to the threat actor to be used in their attack. Insecure default system configurations can create a large attack surface for adversaries.

The software attack surface includes software in consumer products that attach to the network, the customer premise equipment (CPE) such as the cable modem, the home gateways and all the other equipment in the home that connects to the network –computers, printers, smartphones, tablets, thermostats, garage door openers, and web cams. Within the operators network the software attack surface encompasses all the networking equipment such as the cable modem termination system (CMTS), and servers on the network such as the DNS and NTP servers.

Human Attack Surface

Humans are a fundamental part of the attack surface. Any device that a human interacts with can be considered part of the attack surface (Dark Reading 2012). The wide spread adoption of an always-on Internet - via its integration into a broad range of devices - has only increased the human attack surfaces as humans interact with the Internet in ever increasing ways. Humans by nature are going to continue to provide a vulnerable attack surface. Education and training can reduce the vulnerability, but at the same time attackers have improved their social engineering techniques.

Hardware Attack Surface

There seems to be an emerging form of attack which relates to the compromise / attack of hardware either at the point of manufacture or prior to delivery to the customer. These are likely to be a part of the cable infrastructure since these are pervasively deployed technologies. The best recent example of this is the Equation Group recently called out by Kaspersky and also the activities of the Tailored Access Operations (TAO) group at the NSA, which was highlighted in various documents released by Snowden.

Network Attack Surface

The network itself provides an attack surface. The sheer scale of the network infrastructure of Internet Service Providers (ISP) and their massive customer base present an incredibly attractive attack surface (Mistry 2014). Large-scale deployments of homogeneous systems create a multiplication effect when a vulnerability is discovered in those homogeneous systems. The significant available aggregate bandwidth in an ISP's network when combined with a common, widespread vulnerability, make ISPs targets for a form of attack known as a distributed denial of service (DDoS) attack that will be discussed later in this paper.

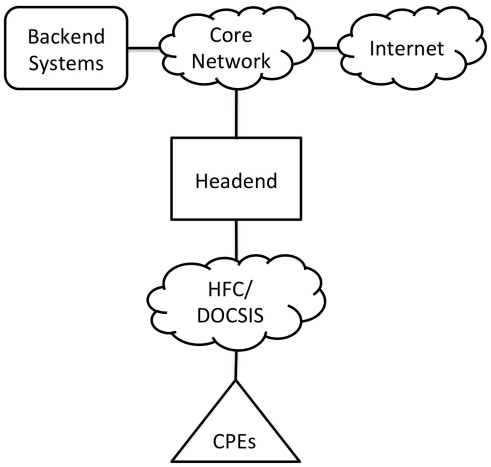


Figure 1 Cable Network

The cable network is no exception. Cable networks have a range of entry points. In addition to the home, businesses and hosting environments, the operators network infrastructure itself provides a number of entry points that include the routers and switches along with the backend processing systems.

ATTACK VECTORS AND COUNTERMEASURES

The term attack surface is used to describe the potential surface areas for the attack vectors. For this paper we define an attack vector as the method a threat actor uses to accomplish their objectives. Over the years we have seen an increase in the number of attack vectors in use with the widespread adoption of the Internet as shown by the total volume of abuse tickets logged by one operator.

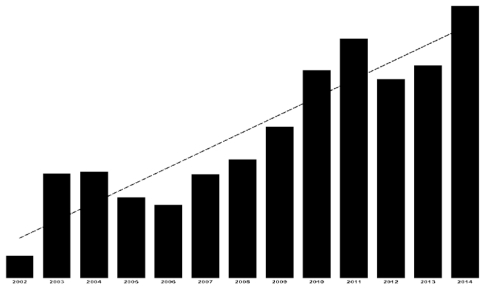


Figure 2 Total Abuse Ticket Volume 2002-2014

The abuse attack vectors can be categorized into two broad categories – penetration attacks and denial of service (DoS) attacks. Penetration attacks are those attacks that use a delivery mechanism to transport a malicious payload to the target host or system to gain unauthorized control (IT Law Wiki n.d.). According to US-CERT "a denial-of-service or distributed denial-of-service attack is an attempt to prevent legitimate users from accessing information or services." (United States Computer Emergency Readiness Team 2013). The most common types of DoS attacks rely on flooding a network with network traffic and overwhelming the victim's system. In addition, malicious software that prevents a legitimate user from using the intended device or service falls into this category.

Penetration Attacks

As an attack vector, the two primary forms of penetration attacks are unauthorized malicious software and unauthorized access.

Viruses & Spyware

The malicious software includes spyware, viruses, and Trojan horses, with the US-CERT team describing spyware as a type of malicious software that collects information from a computing system without the users consent (United States

Computer Emergency Readiness Team 2013).

The first computer virus is believed to be the Creeper Virus (Chen and Robert 2004) that was first detected in the early 1970s and was used to infect DEC PDP-10 computers. Until the mainstream adoption of Internet, most malware was limited to impacting only the host computer system and the user's ability to interact with the system. In some cases, once the malicious software was successfully loaded on the system, it denied the user interaction with the system.

The mainstream adoption of the Internet and the resulting expanded attack surface area, led to an increase in the malware infecting computer systems as shown in Figure 3. Note the data was not normalized for factors such as subscriber growth and improved detection capabilities. Therefore, the overall growth may not be as pronounced as shown in figure, but even without normalizing the data we can discern a continued upward trend in malware detected.

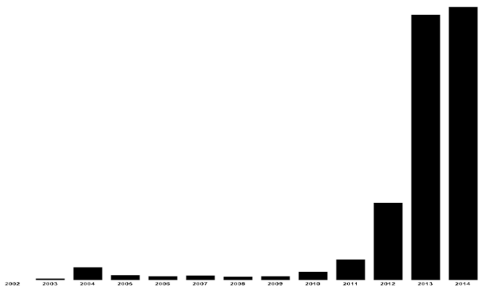


Figure 3 Malware Detected Trouble Tickets 2002-2014

One of the primary countermeasures used against viruses and spyware has been antivirus and malware detection programs.

Phishing

Attacks using the human attack surface leveraging email are known as phishing. Phishing is the attempt to acquire sensitive information such as usernames, passwords, and credit card details by masquerading as a trustworthy entity or by installing backdoors such as keystroke loggers. Phishing tries to fool the target into unknowingly revealing sensitive information to the threat actor or handing over the information by using social engineering, convincing the target to install malicious software, unknowingly download malicious software from website, go to a malicious website using their browser, or to call a number.

Phishing countermeasures include (United States Computer Emergency Readiness Team n.d.):

- Consumer awareness and training
- Shutting down phishing host sites
- Web browser toolbars
- Strong Authentication and Authorization
- Anti-virus software

Botnets

As malware evolved, additional features were added that allowed malware to go from operating in an isolated instance to multiple instances working together through coordinated command and control (C2) systems. A collective of malware infected devices operating under a unified C2 system through the Internet is commonly referred to as a botnet (robot + network). The botnet itself became a new attack vector, since the botnet could be used to send spam or participate in forms of attacks such as a distributed denial of service attacks at greater scale and with larger impact than infected machines working in isolation. The individually

infected bots in the botnet communicate with a command and control center for instructions to allow them to work in coordinated fashion.

In addition to the countermeasures used against viruses and spyware, as noted in the Anti-bot Code of Conduct for ISPs (ABCs for ISPs) (Communications Security, Reliability and Interoperability Council (CSRIC) 2012), botnets can be detected by monitoring the network for communications with known botnet command and control centers and then taking action to block the communication to the command and control and/or notifying infected end-users so that they can take the necessary actions to remediate the infected machine.

In the early days of botnet mitigation, some ISPs would identify C2 servers and concentrate on blocking communications to or taking down C2 servers. This approach was far easier for ISPs than trying to notify customers and assisting them in mitigating and remediating large numbers of infected subscriber devices. Threat actors recognized this weakness in their architectures and designed redundant, multiple layers of C2 servers; increased the sophistication of C2 protocols; and designed peer to peer C2 protocols to obfuscate C2 communication paths and to make the C2 architecture more resilient.

It should be pointed out that ISPs are not the only Internet ecosystem player that can effectively notify users and assist in mitigating and remediating infected devices. Banks, search engines, and e-commerce providers all have the ability to identify, notify, mitigate and remediate their customers and devices with malware infections. As demonstrated in coordinated law enforcement botnet takedowns, these ecosystem players can be more effective in remediating botnet infections with their customers than ISPs.

Denial of Service Attacks

The second category of attacks is the denial of service attack. As the name implies, a DoS attack is an attack that prevents a legitimate user from accessing information or service. DoS attacks come in many forms – email spam, network flood, server overload, resource exhaustion and others. Attacks sources can be concentrated or distributed, leveraging amplification techniques to maximize their effects.

Email spam

The widespread adoption and use of the Internet resulted in a corresponding widespread adoption and use of e-mail. With the widespread adoption of email came the introduction of a new threat vector – spam. Spam is not limited to email, as spam refers to the use of an electronic messaging system such as email or Instant Messaging (IM) to send unsolicited messages. Email became the vehicle of choice for those who send spam (i.e. spammers), because of the ability to reach a large target audience for very low cost. Spam is reported to compose some 80 to 90 percent of all the email. (M3AAWG 2014). Spam without effective filtering tools results in a denial of service attack, as email inboxes become filled with junk mail and the network transit links similarly become filled to capacity with junk mail. For network operators with limited Internet transit, email spam can consume large portions of network capacity and effectively impact the bandwidth available on their network for other purposes.

As shown in Figure 4, the amount of deliberate spam sent by legitimate businesses via e-mail has decreased over time while the amount sent by malware-infected PCs has risen.

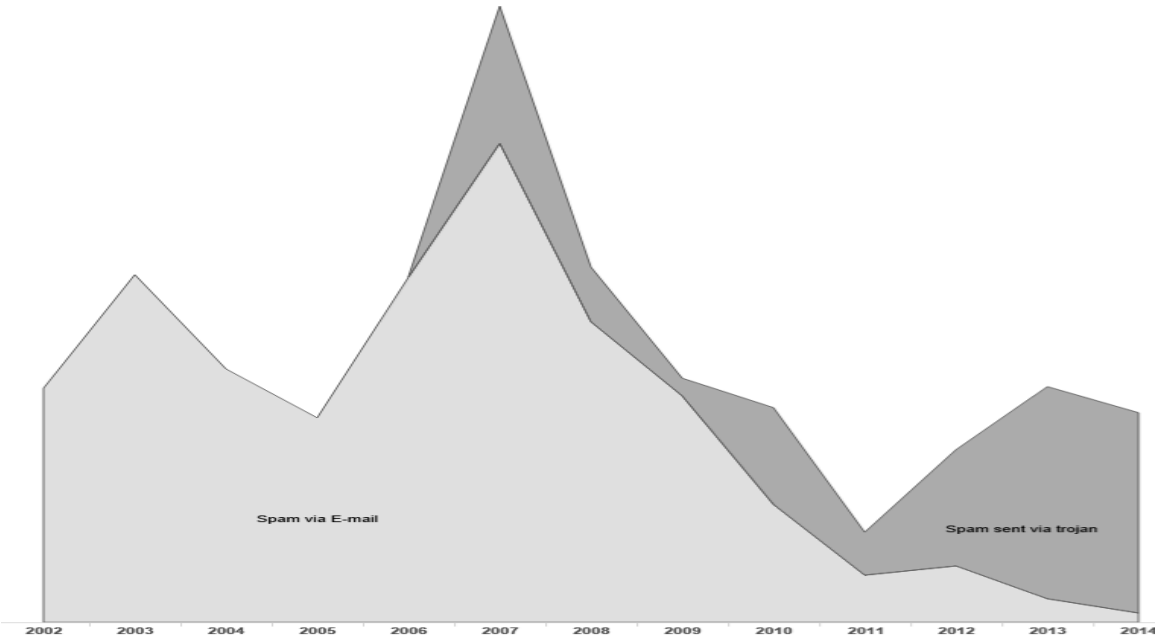


Figure 4 Spam via Email vs. Spam via Bot

The Messaging Malware Mobile Anti-Abuse Working Group (M[3]AAWG) has published a series of best practices and countermeasures for mitigating spam. Included in the best practices is the blocking or management of the well-known SMTP port, port 25 (M3AAWG 2005), to prevent unauthorized use of the SMTP relay function in the email servers.

Network DoS Attacks

One of the earliest forms of a DoS attack using the network attack surface is the ping flood attack, in which the attacker overwhelms the victim with Internet Control Message Protocol (ICMP) echo request (ping) packets. Attackers developed the Distributed Denial of Service (DDoS) attack as a way around the countermeasures for DoS flood attacks (ping, TCP SYN, UDP, and MAC address).

Distributed Denial of Service Attacks

One of the earliest noted forms of a DDoS attack is the Smurf attack (Wikipedia contributors 2015). A Smurf attack is where a large number of ICMP packets with the intended victim's spoofed source IP address are broadcast to a computer network, resulting in a large number of ICMP responses getting sent to the victim (Wikipedia 2015).

As noted in Akamai's 2014 State of the Internet Report (Akamai 2014), both the volume of DDoS attacks and the average bandwidth per attack continued to increase. Akamai also reported that attackers continued to favor force over technique aided by mass exploitation of web vulnerabilities, millions of Internet Of Things (IOT) devices, and successful botnet building.

The growth in botnets provided a platform for more sophisticated DDoS attacks leveraging an installed base of bots. The botnets created an attack vector that has the potential to harness a large number of computers working together in

a coordinated fashion. By harnessing the aggregate bandwidth of the network using infected hosts, attackers dramatically increase the scale and magnitude of the attacks. DDoS attacks occur at all seven layers of the OSI stack (National Cybersecurity and Communications Integration Center 2014). The FCC CSRIC IV Working Group 5 on Remediation of Server Based DDoS Attacks created a complete taxonomy of different types of DDoS attacks and mitigation or remediation best practices (Communications Security, Reliability and Interoperatbility Council 2014). There are four major categories of DDoS attacks:

- Direct and reflective volumetric attacks
- Application layer attacks
- State exhaustion attacks, and
- Control plane attacks

The US-CERT team lists a number of countermeasures for mitigating large scale DoS/DDoS attacks (United States Computer Emergency Readiness Team 2013) that include:

- Layer 3 & 4 line-rate access control lists (ACLs) and rate limits
- Layer 4-7 Packet Screens, session limits, and SYN cookies

Distribute Reflective Denial of Service Attack

The magnitude of the DDoS attack vector can be further increased using a reflection amplification technique resulting in a Distributed Reflective Denial of Service (DRDoS) attack (United States Computer Emergency Readiness Team 2014). The reflection amplification is achieved by first sending a set of requests that get reflected. The attacker sends forged requests of some type; typically the source address in the IP headers is forged or spoofed, to a very large number of computers that will reply to the requests. The amplification in the attack is achieved by sending a request that results in a response much larger than the request. A good example of this is a DNS request. The DNS request is a small number of bytes (e.g. ~60 bytes), but can result in a response that is much larger (e.g. >1000 bytes).

The DNS Amplification attack (United States Computer Emergency Readiness Team 2013) is a popular form of a DRDoS attack, in which attackers use publically accessible open DNS servers (aka open resolvers) to flood a target system with DNS response traffic. The attacker sends a DNS lookup request to an open DNS server with the source address spoofed to be the target's address. The DNS server then sends the DNS record response it is sent instead to the target. The attacker's DNS request is typically submitted to request as much zone information as possible to maximize the amplification effect.

The US-CERT team has identified 11 different potential attack vectors for UDP protocols that include DNS, NTP, and SNMPv2 (Rossow 2014) that can be used for DDoS/DRDoS attacks.

DDoS attacks require a different set of countermeasures than simply blocking the traffic on an abused TCP or UDP port because some implementations of protocols may send the response using same IP port on which it was received. Instead of blocking the port, a widely used countermeasure will check for the spoofed IP addresses as described in IETF BCP 38 (Senie and Ferguson 2000) and BCP 84 (Baker and Savola 2004).

Man-in-the-Middle

Yet another attack vector is the Man-In-The-Middle attack (MITM). In a MITM

attack the adversary secretly relays and possibly alters the communications between two parties who believe they are directly communicating with each other. MITM the middle attacks can be performed using a number of techniques:

- Network taps that mirror to offline analysis system
- Re-routing the traffic
- Masquerading as the intended destination for an open WiFi or cellular access point

Over the last couple of years there have been widespread reports of persistent Internet surveillance by governments by using a network tap (optical or electrical tap) to mirror/copy all the traffic to an offline storage and analysis system. (Gorman and Valentino-Devries 2013) These taps were upstream from cable companies and performed on the Internet backbone.

These widespread reports have prompted a response by the Internet community to recommend encrypting all Internet communications. In 2014, the Internet Architecture Board (IAB) released a statement recommending that encryption should be deployed throughout the Internet protocol stack (Internet Architecture Board 2014). Earlier in the year, in May 2014, the IETF did a technical assessment and declared pervasive monitoring is an attack on privacy with the release of BCP-188/RFC 7258, Pervasive Monitoring Is an Attack, (Farrell and Tschofenig 2014). To mitigate the attack on privacy, RFC 7258 recommends that encryption should be authenticated where possible and that all newly designed protocols should prefer encryption. A second IETF RFC was released that advocates for the use of "Opportunistic Security" (Dukhonvi 2014). Opportunistic Security recommends upgrading from cleartext to unauthentic session encryption when authenticated encryption is not available. The concept here is that some encryption, even if unauthenticated, is better than no encryption.

It should be noted that adding encryption and authentication techniques can cause system availability to decrease and creates a larger attack surface for denial of service attacks. Disrupting the authentication or encryption mechanisms and the larger packets sizes and CPU/memory requirements to implement these technologies increase the ability of threat actors to disrupt the underlying protocol or service.

MITM attacks can be performed in a number of ways. One method is to use a technique known as "route hijacking" where the attacker intentionally announces an illegitimate route advertisement using the Border Gateway Protocol (BGP) to other networks. The rogue route can result in the traffic getting diverted away from its intended route. Once diverted, the traffic can either result in the destination being unreachable and creating a DoS attack, or it can be diverted through additional networks resulting in MITM attack. There are a number of counter-measures described in the NIST publication SP-800-54 (Kuhn, Sriram and Montgomery 2007) with prefix filtering being the most basic mechanism for protecting BGP routers from routing disruptions (Kuhn, Sriram and Montgomery 2007, 42).

A third form of MITM attacks can occur as a result of masquerading or spoofing. Attackers masquerade as open public WiFi access points by spoofing the WiFi SSID and as destination websites by spoofing the hostname using a technique known as DNS spoofing. A DNS spoofing attack uses forged DNS information to supply the victim with an illegitimate IP address that takes them to the attack site instead of the intended site. As a

countermeasure to DNS spoofing, the IETF developed a suite of specifications, Domain Name System Security Extensions (DNSSEC) (Wikipedia contributors 2015) to enable DNS clients to authenticate the DNS data.

Countermeasures Summary

Attack Surface	Attack Vector	Countermeasures
Software	Virus/Spyware	Detection
Human, Software	Phishing	Consumer awareness, Site shutdowns, Browser toolbars, Authentication & Authorization, and Anti-virus software
Software, Network	Spam	Port Blocking Access Control Filtering
Software, Network	Bots	Network monitoring Anti-virus software
Network	DoS/DDoS	Filtering, Access Control Network Ingress Filtering
Network	Man-in-the-Middle	BGP Prefix filtering, DNSSec, Encryption, secure WiFi access points

CABLE STANDARDS

The DOCSIS™ standards employ a number of features that can be used as part of deploying countermeasures for the threat vectors discussed in this paper.

Cable Modem Protocol Filters

DOCSIS supports two IP protocol-filtering methods. The initial method specified in DOCSIS 1.0 is specified in RFC 4639 (Woundy and Marez 2006) and supports filtering at the IP layer and the logical link control (LLC) layer and only supports IPv4. With the release of DOCSIS 1.1, as part of adding support for IPv6, the concept of the Upstream Drop Classifier (UDC) was introduced. UDCs enhanced the filtering capabilities by adding support for IPv6 and rules that were mutually exclusive of whether it was TCP or UDP. The inclusion of UDC further reflects that upstream filtering is best done as close as possible to source in the cable modem (CM) itself, while downstream filtering is best done in the cable modem termination system (CMTS) or further up in the network.

The filters can be statically configured as part of the cable modem provisioning process or they can be dynamically configured. They can filter on any combination of:

- IP Protocol – TCP, UDP
- Source IP Address
- Destination IP Address
- Source Port
- Destination Port
- Type of Service (TOS)
- Direction

Source Address Verification

The DOCSIS specification (CableLabs 2014) includes a feature known as Source Address Verification (SAV). The SAV feature per the specification defaults to

"enabled" on the CMTS and ensures that the customer premise equipment (CPE) located behind a cable modem cannot spoof addresses in order to obtain access to services or disrupt the services to others. When the SAV feature is enabled, the CMTS drops any upstream packets whose IP source address has not been assigned by the operator, including those whose source IP address has already been assigned to another device.

The feature provides an implementation of network ingress filtering of upstream traffic as described in the IETF's BCP-38 (RFC 2827). In addition to preventing DDoS traffic using spoofed source IP addresses, the feature also prevents theft of service by users statically assigning an IP address not assigned to the cable modem.

Baseline Privacy Plus (BPI+)

Baseline Privacy Plus (BPI+) is one of the DOCSIS security specifications. BPI+ provides two key security features – data privacy over the cable network and unauthorized access to the network's RF media access control (MAC) layer services. BPI+ provides data privacy by encrypting the packet or user data in the DOCSIS frames. Cable modems ship with an X.509 certificate installed in them. During the modem registration process, the CM must send its X.509 certificate to the CMTS for authentication. The authentication process prevents forged and cloned modems from gaining access to the network services.

PacketCable™ Multimedia

The PacketCable Multimedia Specification (PCMM) (CableLabs 2011), provides a dynamic quality-of-service (QoS) framework for dynamically changing the provisioned bandwidth on a per IP-flow basis. The framework supports dynamically creating a "gate" that has attributes that describe the packet filter similar to those of the static cable modem filters as well as QoS attributes such as bandwidth allocated for the IP-flow.

PCMM can be used as a countermeasure for blocking or rate limiting of threat vector traffic such as e-mail spam or DDoS traffic that is originating from a CPE behind a cable modem.

IETF

In addition to security elements in the DOCSIS and PacketCable specifications, the IETF specifications include many security elements. Two key areas for securing the network reside with securing the BGP and DNS. Both of these protocols are considered critical for the operation of the Internet. If either service is disrupted, it can have regional or global effects on Internet traffic.

Border Gateway Protocol

BGP is the routing protocol for border routers to exchange routing information between autonomous systems (AS) on the Internet (IETF 2006). As a countermeasure for a DDoS attack, the routers can be configured to null route (aka black hole) the attack traffic. The effect of the null routing is to drop all the packets of the attack traffic. A variation of a black hole is to sink hole the traffic. When sink holing the traffic, instead of null routing and dropping the attack traffic, the attack traffic is sent to a capture device where the data can be used for further analysis.

BGP Flowspec is another technique operators use to distribute granular router ACLs to a large number of routers using BGP. BGP Flowspec can be used to automate distribution of blocking or rate limiting ACLs to mitigate DoS/DDoS

attacks in minutes instead of manually configuring ACL on routers that can take hours or days.

DNSSEC

The IETF enhanced the DNS protocol with a set of security extensions, DNSSEC. DNSSEC was designed to protect application from using forged or manipulated DNS data as the result of a MITM attack. To validate DNS queries, all the responses are signed with a digital certificate. By checking the digital certificate, the DNS resolver can check that the information is the same as what the domain owner published. These extensions have been adopted by a number of major ISPs but there is considerable resistance to further adoption.

APPLICATIONS OF EMERGING TECHNOLOGIES FOR CYBER COUNTERMEASURES

A number of technologies are emerging that may prove to be valuable, as they mature, in providing new, effective countermeasures. Included in this list of emerging technologies is:

- BGP Security (BGPSec)
- Big Data
- Software Defined Networking
- Network Virtualization Function

BGPSec

The IETF is developing an extension, BGPSec, to secure the BGP protocol. Currently the BGP protocol is vulnerable to what is known as "route hijacking" that can lead to a MITM attack. Route hijacking occurs when a network or AS advertises that it can route to select destinations that it is not authorized to advertise. However the adoption of this set of extensions is likely to be a very long process and is not guaranteed.

BGP security involves several important aspects of verification:

- Route origin verification – ensuring that the route announcement was made by the authorized party
- Path verification – ensuring that the route the data traversed was not maliciously manipulated or affected by human error

BGP security involves several aspects for successful deployment and wide spread implementation:

- An accurate database of authorized routes
- A secure method for publishing authorized routes globally
- A scalable and secure method to verify route origin
- A scalable and secure method to verify the path the data took to the end user and that the path was the best or most appropriate path

Not all of the above requirements must be met at once. For example, route origin verification alone would be a big step forward for securing the BGP protocol.

Several methods for securely publishing route information have been proposed with the most prominent method being Resource Public Key Infrastructure (RPKI) (American Registry for Internet Numbers n.d.). Route Origin VERification (ROVER) (Gersch 2012) is an alternative method that may be more scalable.

Big Data

As a result of the previously discussed reports of the pervasive monitoring and

the Internet community's response to work to secure all Internet communications by encrypting all traffic, this will have an impact on security operations and network management (Moriarty and Morton 2015). Some current security techniques will need to evolve and adapt to effects of encryption, while others will cease to work effectively.

The term "Big Data" has been described to mean data sets so large or complex that processing and storage applications are inadequate. The large data sets create challenges in the analysis, storage, and visualization across the data set. Big Data is one technology that can be leveraged to improve the passive monitoring capabilities as part of adapting and evolving to the realities of trend towards ubiquitous encryption for Internet traffic.

Operators often have multiple sources of threat feeds. The threat feeds typically includes DNS data, multiple C2 feeds, multiple black/white lists, and indicators of compromise (IOC) feeds. The threat feed information may or may not be stored in dedicated repositories as the feed information is often specific for the supporting network appliances. The data feeds result in large data sets and therefore are often not retained for long periods of time due to the limits of the conventional data storage technology. Big data technology can be applied to the multiple threat feeds to create a single repository with the aggregated data. The data in the single repository can then be analyzed and correlated to identify trends and other threat patterns for the application of an appropriate counter-measure.

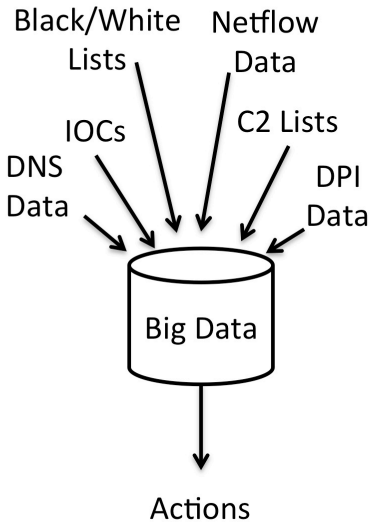


Figure 5 Big Data - Threat Aggregation

SDN

Software-Defined Networking (SDNs) is an approach to networks that decouples the networks data or forwarding plane from its control plane to provide a centralized view of the overall network. SDNs enable the automation of network services through a policy-based decision framework such as PacketCable Multimedia and 3GPP's Policy-Charge Rules Function (PCRF) to orchestrate which network traffic goes where. Figure 6 illustrates the SDN logical architecture.

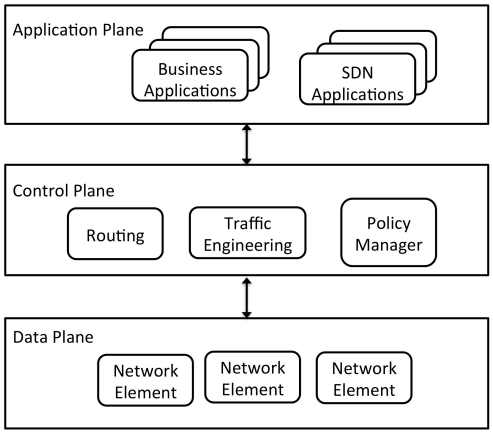


Figure 6 SDN Logical Architecture

Network function Virtualization (NfV) decouples network functions from proprietary hardware appliances to virtualize network services. NfV delivers the virtualized infrastructure upon which an SDN's control plane software, L2/L3 data plane software, and network services can run.

The NfV technology can be extended to the CPE device(s) in the home to create a virtual CPE where most of the CPE functions are pulled into the operator's network. The CPE device then acts as a simple layer-2 device for providing network connectivity and all the network services functions such as NAT, firewall, routing, VPN, filtering, etc. being delivered by the virtual CPE in the headend or service provider's cloud.

The combination of NfV with SDN (NfV/SDN) provides two key features for cybersecurity countermeasures:

- Elastic scaling of the network functions
- Ability to dynamically re-allocate packet processing across the SDN

SDN + NETWORK SECURITY

The NfV/SDN architecture can be enhanced with Big Data to provide a means for operators to automate network countermeasures. This architecture can be used to build and deploy a suite of network security SDN applications. In addition to the two previously mentioned key features of the NfV/SDN architecture, it also helps operators to mitigate the impacts the threat vectors discussed earlier by leveraging the inherent pseudo-wire techniques (e.g. L2VPN, DOCSIS service flows, PCMM Gates) to filter, block, isolate or reroute suspect traffic.

The virtual CPE as part of this architecture will allow for better security and operational control for the operator. Better control is provide due to the fact that the operator has better visibility and control of the various state information in the virtual CPE such as the NAT tables, local routing and VPN, and packet filters. This will enable the operator to more quickly address and mitigate network issues.

A key element in SDN architecture is the SDN controller. Conceptually the SDN controller communicates to various access networks elements in the cable networks headend (e.g. CMTS, CCAP) and provides a platform for developing technology-agnostic network applications (e.g. L2VPN application connecting end-points across the underlying networks). SDN improves the process for dynamically binding an IP-flow to pseudo-wire (e.g. DOCSIS service flow or PCMM gate in the DOCSIS network and VLAN in the core) to isolate the suspect traffic from other traffic.

As shown in Figure 7, we are proposing two new SDN applications – 1) Security and 2) a BGPSec cryptographic processor.

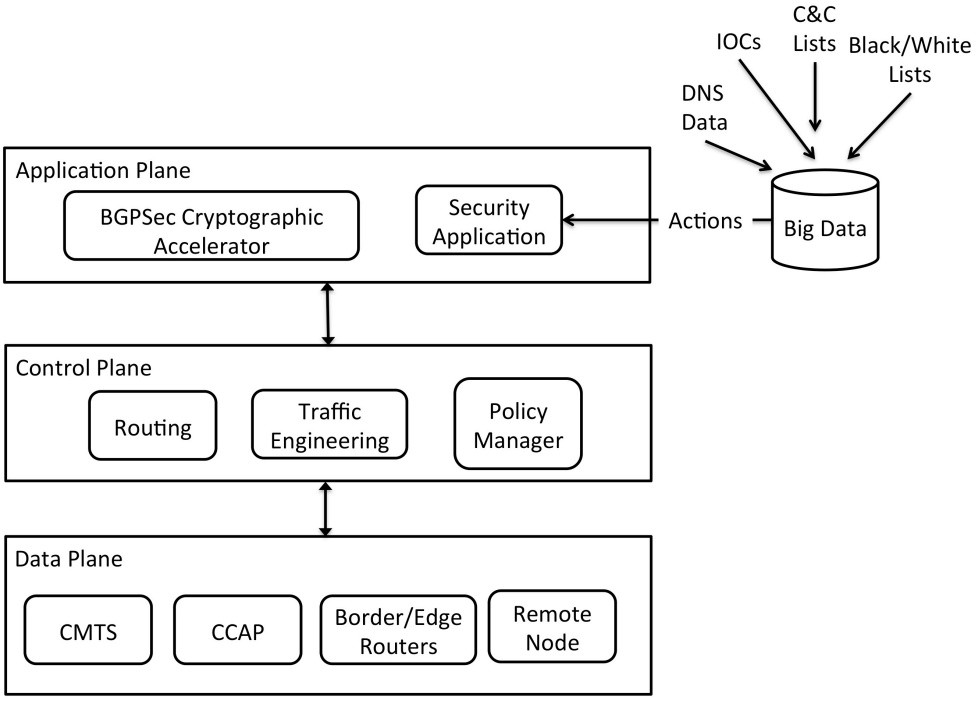


Figure 7 Security Apps in Cable SDN Architecture

This security application would ingest the analyzed and correlated threat feed actions from the Big Data system along with traffic flow information occurring on the network and orchestrate a set of network counter-measures that include:

- Blocking or rate limiting malicious traffic as close to the source as possible
- Routing suspect traffic to a sinkhole or black hole
- Routing the traffic to the data scrubbing system
- Dynamically adding bandwidth
- Add CPU and memory resources in the controller plane to mitigate resource attacks
- Dynamically adding a pseudo-wire for the suspect traffic to provide a level of isolation
- Dynamically isolating and routing critical device traffic (such as medical device traffic) to authorized locations to minimize the expose of the data and device attack surface

The second SDN application is a dedicated BGPSec cryptographic processing application. BGPSec is a cryptographically intense protocol (Small Business Innovation Research n.d.). BGPSec requiring two levels of cryptography. First, each of the RPKI objects must be verified and second, it requires the router to sign and verify each BGP message that it sends (Goldberg 2014). The volume of cryptographic processing can fluctuate due to network conditions. The capability to dynamically add additional cryptographic processing (e.g. CPU) can reduce the network convergence time when there is a network change. In addition, the use of the NfV/SDN architecture allows standard off-the-shelf CPUs to be used instead of custom cryptographic processors.

The addition of the BGPSec cryptographic SDN application can help to reduce the cost of deploying BGPSec by

eliminating the need for dedicated cryptographic processors in the routers. The broad adoption of BGPSec and the ability to authenticate advertised routes will greatly reduce or eliminate accidental and malicious route hijacks and BGP MITM attacks.

SUMMARY

In this paper we have surveyed the current network threat landscape in terms of both the network attack surfaces and the associated network attack vectors. We further illustrated how the NfV/SDN architecture can be integrated with Big Data to allow operators to address real-time threats in an automated fashion. An enhanced NfV/SDN architecture was proposed that can be enhanced with Big Data to provide a means for operators to automate network countermeasures.

BIBLIOGRAPHY

Akamai. "[state of the internet] / security - Q4/2014 Report." 2014.

American Registry for Internet Numbers. *RESOURCE PUBLIC KEY INFRASTRUCTURE (RPKI).* https://www.arin.net/resources/rpki/ (accessed March 31, 2015).

Baker, Fred, and P. Savola. "Ingress Filtering for Multihomed Networks." March 2004. http://www.rfc-editor.org/bcp/bcp84.txt (accessed April 2, 2015).

CableLabs. *Data-Over-Cable Servicee Interface Specification DOCSIS 3.1 - Security Specification.* Specification, CableLabs, 2014.

CableLabs. *PacketCable(TM) Specification Multimedia Specification.* specification, CableLabs, 2011.

Chen, Thomas, and Jean-Marc Robert. *The Evolution of Viruses and Worms.* 2004. http://vxheaven.org/lib/atc01.html (accessed 3 9, 2015).

Communications Security, Reliability and Interoperability Council (CSRIC). *Final Report: U.S. Anti-Bot Code of Conduct (ABCs) for Internet Service Providers (ISPs) (A Voluntary Code).* Federal Communications Commission, 2012.

Communications Security, Reliability and Interoperatbility Council. *Working Group 5 Remediation of Server-Based DDoS Attacks Final Report.* Federal Communications Commission, 2014.

Dark Reading. *Minimizing the Attack Surface Area: A Key to Security.* 03 23, 2012. http://www.darkreading.com/analytics/security-monitoring/minimizing-the-attack-surface-area-a-key-to-security/d/d-id/1137387 (accessed 03 16, 2015).

Dukhonvi, V. "Opportunistic Security: Some Protection Most of the Time." *IETF.org.* December 2014. http://www.ietf.org/rfc/rfc7435.txt.pdf (accessed March 30, 2015).

Farrell, S., and H. Tschofenig. *Pervasive Monitoring is an Attack.* May 2014. http://www.rfc-editor.org/rfc/rfc7258.txt (accessed March 27, 2015).

Gersch, Joseph. *ROVER- BGP Route Origin Verification via DNS.* April 2012. https://ripe64.ripe.net/presentations/57-ROVER_RIPE_Apr_2012.pdf (accessed March 31, 2015).

Goldberg, Sharon. *Why Is It Taking So Long to Secure Internet Routing.* Vers. Volume 12. acmqueue. September 11, 2014.

http://queue.acm.org/detail.cfm?id=2668966 (accessed March 24, 2015).

Gorman, Siobhan, and Jennifer Valentino-Devries. "New Details Show Broader NSA Surveillance Reach." *The Wall Street Journal*, August 21, 2013.

IETF. *RFC 4271 - A Border Gateay Protocol 4 (BGP-4).* January 2006. http://www.ietf.org/rfc/rfc4271.txt (accessed March 20, 2015).

Internet Architecture Board. *IAB Statement on Internet Confidentiality.* November 14, 2014. https://www.iab.org/2014/11/14/iab-statement-on-internet-confidentiality/#more-7754 (accessed March 27, 2015).

IT Law Wiki. *Penetration Attack.* http://itlaw.wikia.com/wiki/Penetration_attack (accessed 03 17, 2015).

Kuhn, Rick, Kotikalapudi Sriram, and Doug Montgomery. *Border Gateway Protocol Security.* National Institute of Standards and Technology, NIST, 2007.

M3AAWG. *M3AAWG Email Metrics Program: The Network Operators' Perspective - Report #16 - 1st Quarter 2012 through 2nd Quarter 2014.* M3AAWG, 2014.

M3AAWG. *Managing Port 25 for Residential or Dynamic IP Space Benefits of Adoption and Risks of Inaction.* M3AAWG, 2005.

Manadhata, Partyusa K. "An Attack Surface Metric." PhD Thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, 2008.

Manadhata, Pratyusa K, and Jeannette M Wing. *Attack Surface Measurement.* IEEE Transactions on Software Engineering, 2010.

Mistry, Bipin. *DDoS Attack left "Sweden not Working".* 12 2014, 2014. http://www.securitybistro.com/?p=9047&utm_content=infosecurity&utm_source=twitterfeed&utm_medium=twitter (accessed 03 16, 2015).

Moriarty, K., and A. Morton. "Effect of Ubiquitous Encryption." Vers. 1. *IETF Network Working Group.* March 2015. http://www.ietf.org/id/draft-mm-wg-effect-encrypt-01.txt (accessed March 30, 2015).

National Cybersecurity and Communications Integration Center. *DDoS Quick Guide.* January 29, 2014. https://www.us-cert.gov/sites/default/files/publications/DDoS%20Quick%20Guide.pdf (accessed March 17, 2015).

Northcutt, Stephen. *Security Laboratory: Defense in Depth Series.* http://www.sans.edu/research/security-laboratory/article/did-attack-surface (accessed 03 16, 2015).

Open Web Application Security Project (OWASP). *Attack Surface Analysis Cheat Sheet.* 7 18, 2014. https://www.owasp.org/index.php/Attack_Surface_Analysis_Cheat_Sheet (accessed 3 11, 2015).

Rossow, Christian. "Amplification Hell: Revisiting Network Protocols for DDoS Abus." *Internet Society.* February 23, 2014. http://www.internetsociety.org/sites/default/files/01_05.pdf (accessed August 11, 2014).

Senie, D., and P. Ferguson. *RFC 2827 - Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing.* May 2000. http://tools.ietf.org/html/rfc2827 (accessed March 18, 2015).

Small Business Innovation Research. *Cryptographic Acceleration for Border Gateway Protocol Security (BGPSE...* http://www.sbir.gov/content/cryptographic-acceleration-border-gateway-protocol-security-bgpsec-1 (accessed March 24, 2015).

United States Computer Emergency Readiness Team. *Alert (TA13-088A) - DNS Amplification Attacks.* July 22, 2013. https://www.us-cert.gov/ncas/alerts/TA13-088A (accessed March 18, 2015).

—. *Alert (TA13-088A) DNS Amplification Attacks.* March 29, 2013. https://www.us-cert.gov/ncas/alerts/TA13-088A (accessed March 17, 2015).

—. *Alert (TA14-017A) UDP-based Amplification Attacks.* March 07, 2014. https://www.us-cert.gov/ncas/alerts/TA14-017A (accessed March 17, 2015).

—. *Security Tip (ST04-015) Understanding Denial of Service Attacks.* 02 16, 2013. https://www.us-cert.gov/ncas/tips/ST04-015 (accessed 03 17, 2015).

—. *Spyware.* 02 06, 2013. https://www.us-cert.gov/security-publications/spyware (accessed 03 17, 2015).

—. *Technical Trends in Phishing Attacks.* https://www.us-cert.gov/sites/default/files/publications/phishing_trends0511.pdf (accessed Mach 18, 2015).

Wikipedia contributors. *Domain Name System Security Extensions.* Vers. 650332715. March 7, 2015. http://en.wikipedia.org/w/index.php?title=Domain_Name_System_Security_Extensions&oldid=650332715 (accessed March 24, 2015).

—. *Smurf Attack.* Wikipedia, The Free Encyclopedia. January 26, 2015. http://en.wikipedia.org/w/index.php?title=Smurf_attack&oldid=644239924 (accessed March 24, 2015).

Wikipedia. *Smurf Attack.* The Free Encyclpedia Wikipedia. 01 26, 2015. http://en.wikipedia.org/w/index.php?title=Smurf_attack&oldid=644239924 (accessed 03 16, 2015).

Woundy, R., and K. Marez. "RFC 4639 - Cable Device Management Information Base for Data-Over-Cable Service Interface Specification (DOCSIS) Compliant Cable Modems and Cable Modem Termination Systems." *IETF.org.* December 2006. http://www.ietf.org/rfc/rfc4639.txt (accessed March 24, 2015).

The Next Evolution in Cable: Converged, Distributed and Virtualized 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

As Cable operators are venturing into multiple access technologies, it has become quite clear that the next frontier in our industry is to evolve towards a Distributed Architecture. Whether that is by distributing just the PHY, or distributing both MAC and PHY, the trend is quite clear from both vendors and operators. This trend is intractably linked to two additional evolutions: Convergence of the various access technologies, and Virtualization of certain access 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 topics of Convergence and 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
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. If not fully utilized, 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.

Proportionally few HFC networks have been deployed to operate up to 1 GHz, although all equipment available today can support the use of spectrum up to 1 GHz and even 3 GHz for some components.

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.

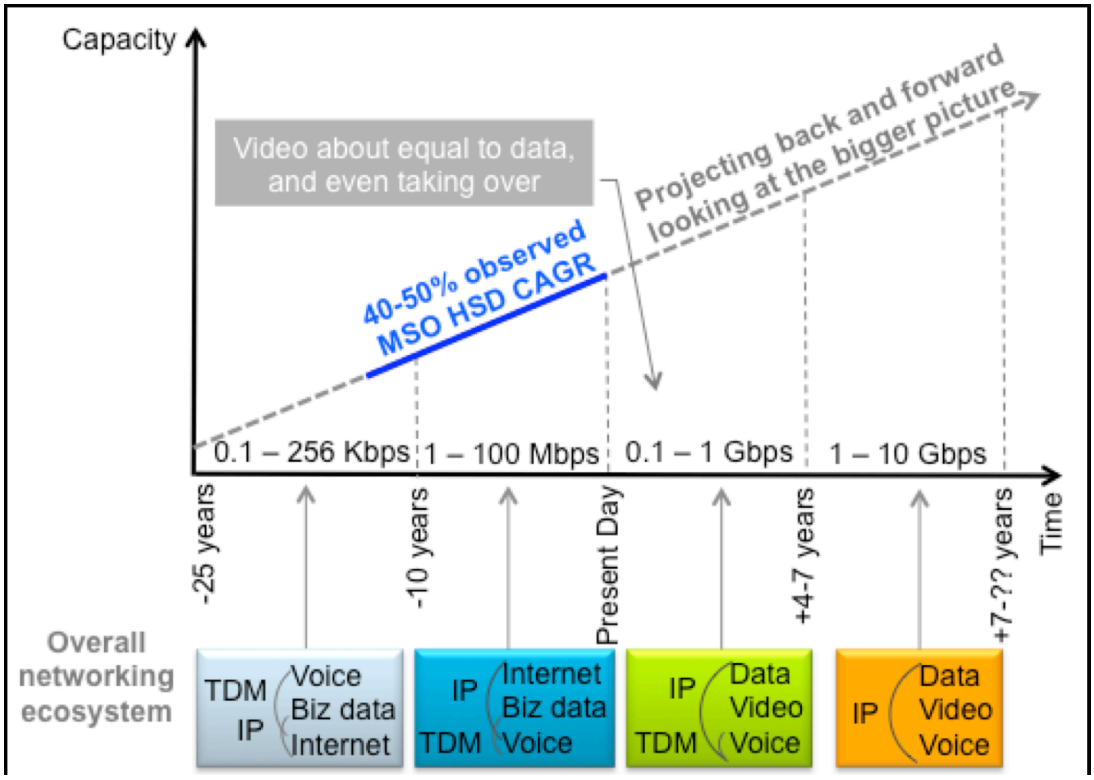


Figure 1: Example of HSD capacity utilization over time

Finally, the growth in HSD services continues. Network operators have 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, but the demand has continued to increase at the same relentless rate.

How does this compare to other operator's data services and a longer period? As shown in Figure 1, projecting the MSO's HSD service growth back in time to when Internet services started as shown in the diagram, 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 MSOs 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 will be required to provide HSD services in the range approximating 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 services, 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

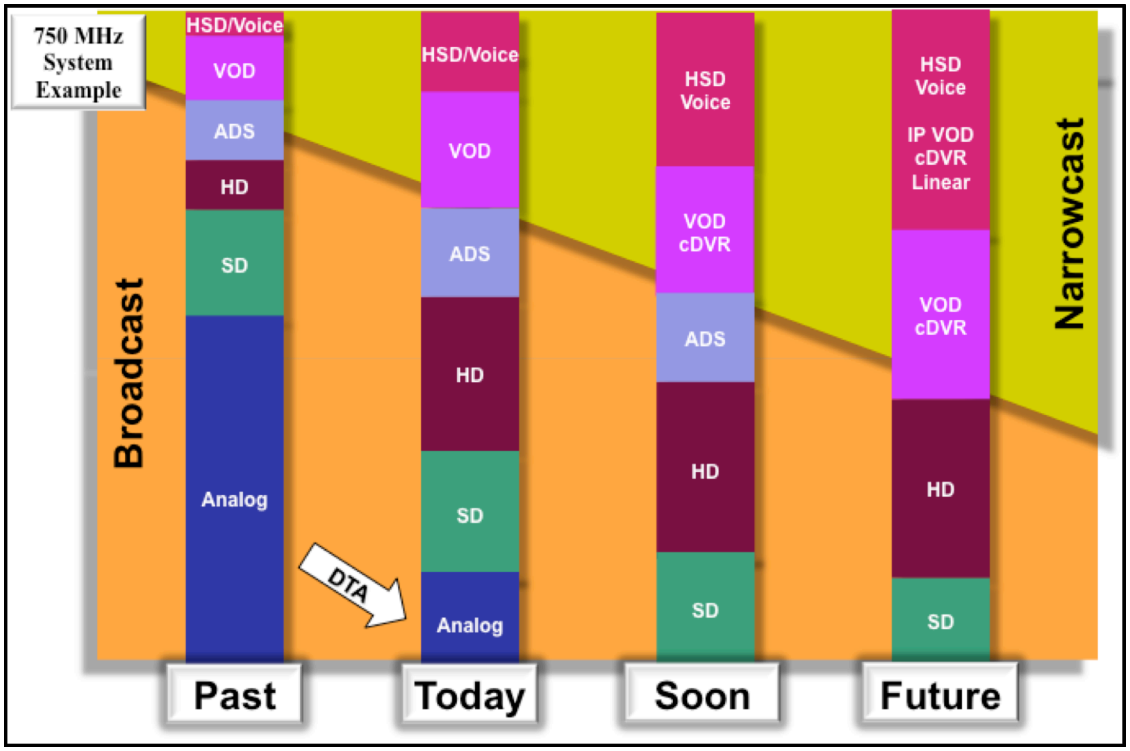


Figure 2: Example of narrowcast service growth over time

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 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 could be very significant.

However, these are difficult tools to take advantage on the network since proportionally relatively few legacy set-tops still support AVCs and VBR, especially the latter. 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 services.

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 will continue to be deployed. Moreover, this increase in simultaneous use of advanced, IP video services while maintaining legacy services will be especially impacting over time as its penetration increases.

All of the above, coupled with the success experienced by MSOs in recent with business services, 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. NOTE: For further details on the DOCSIS 3.1 technology and its implementation, see the DOCSIS 3.1 Symposium materials from the SCTE Expo 2013 show.

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 4 and even 8 channels, and soon requiring well beyond that.

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

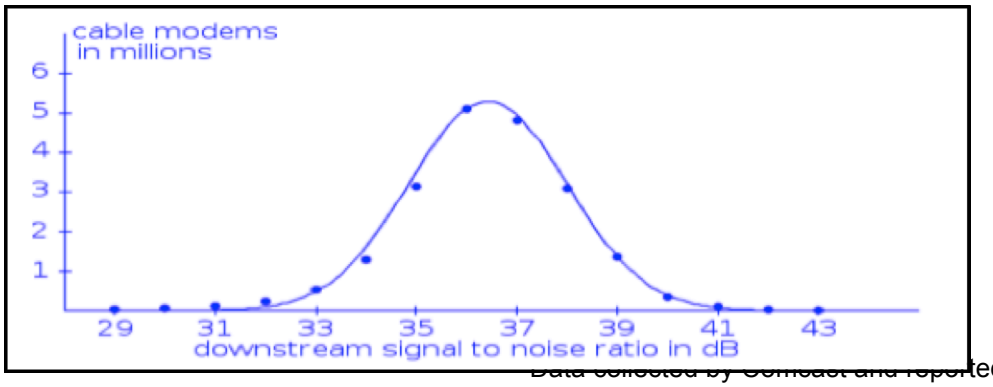


Figure 3: Example of downstream SNR for a large population of cable modems

in SNR.

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

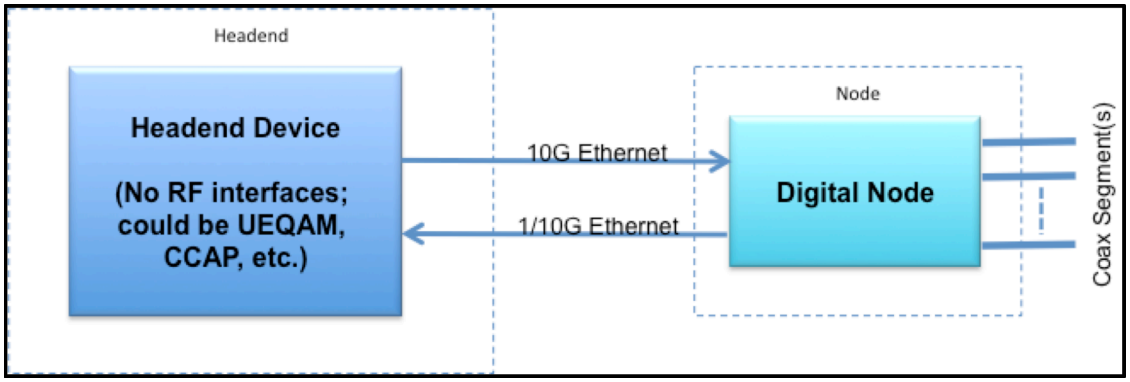


Figure 4: Digital Forward – High Level Architecture

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

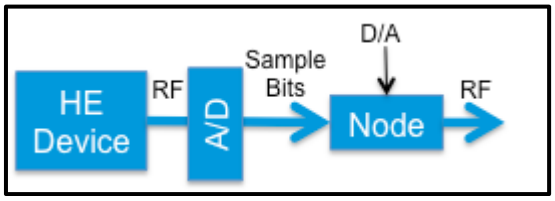


Figure 5: Block diagram for Option 1

- 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

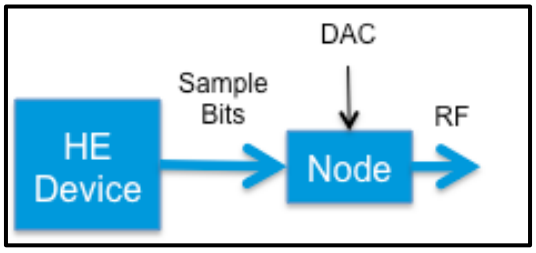


Figure 6: Block diagram for Option 2

- 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

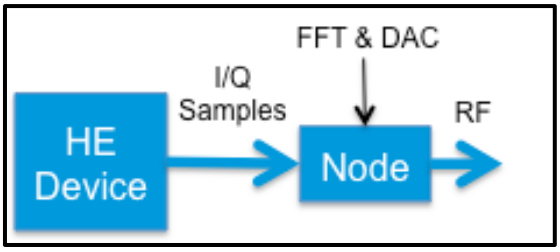


Figure 7: Block diagram for Option 3

- 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

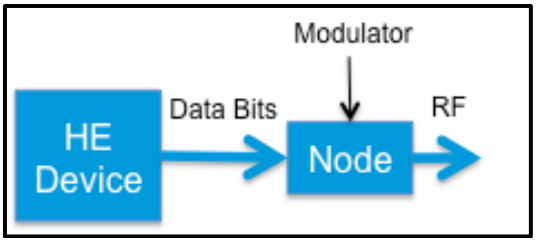


Figure 8: Block diagram for Option 4

- 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

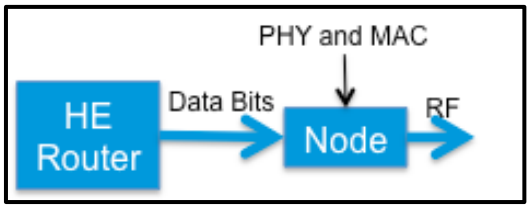


Figure 9: Block diagram of Option 5

- 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.

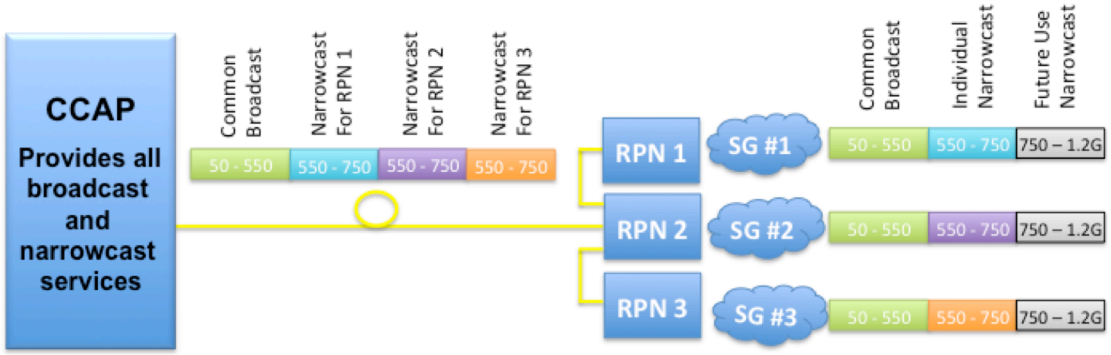


Figure 10: Reuse of broadcast capacity across multiple RPNs

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.

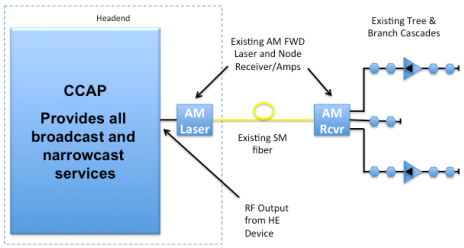


Figure 11: Single traditional HFC node

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.

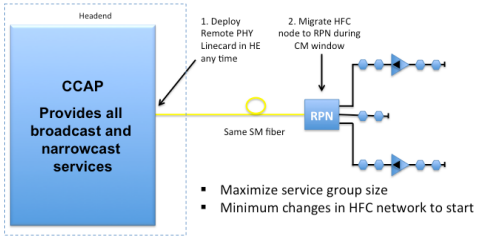


Figure 12: RPN deployment step 1

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.

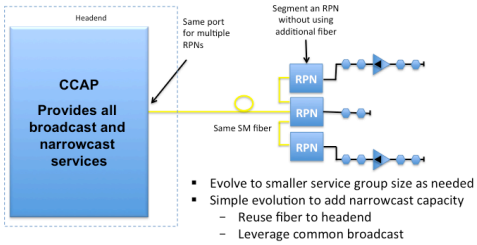


Figure 13: RPN deployment step 2

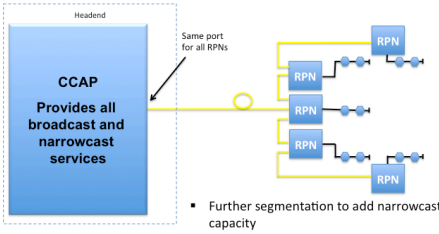


Figure **14** shows how further segmentation could take place by replacing the remaining amplifiers in the network with RPNs.

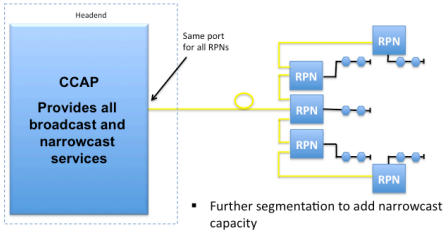


Figure 14: RPN deployment step 3

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.

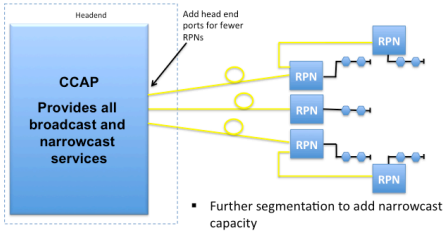


Figure 15: RPN deployment step 4

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.

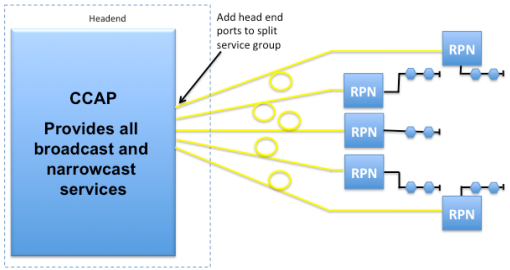


Figure 16: RPN deployment step 5

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.

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
ARRIS Group

Abstract

Many forward looking thinkers in the the Cable industry believe that a move of HFC to an all fiber network is necessarily a move to a Passive Optical Network (G/EPON). However such a change would be quite disruptive and would bring with it challenges of maintaining two separate systems - one for the vast majority of customers now served by HFC and the other for those to be served by PON networks - for a very long time. For starters, only a small percentage of the total network could be converted to all fiber each year, even if all available resources are dedicated to this conversion activity alone. Furthermore, HFC traditionally has been cost effective because of large service group sizes per different offerings (HSD, VOD, VoIP), while PON link budgets generally limit one to far smaller service groups thus leading to much higher startup cost.

What if there was a better way to go all fiber? This ideal way would enable the HFC and the all fiber systems to coexist, with the same equipment at the home and in the headends. It would enable MSOs to continue deployment and activation with the same set of tools and personnel. Finally it would provide a substantial increase in bandwidth, capacity and reliability to last the next several decades.

Recent advances in RFoG technology enable just that. Now Cable MSOs can seamlessly migrate to an all fiber system while increasing capacity, conserving critical infrastructure and enhancing reliability.

Why is now the right time? Up until now, a wider adoption of RFoG was severely limited by a particularly deleterious effect called Optical Beat Interference (OBI). Even very modest amounts of OBI have a severe effect on not only the upstream throughput but on the downstream throughput as well. Recent advances in technology have enabled the complete elimination of OBI thereby unlocking the true potential of fiber.

In this paper, we provide critical insights into the innovations that enable OBI Free RFoG transmission. We will the discuss intrinsic capabilities of what we call Hybrid PON (HPON) technology, explain how this technology works with existing HFC analog and QAM video and D3.0 and D3.1 signals while also being completely transparent to the myriad of traditional PON standards such as the 10G EPON, 1G EPON, GPON and XGPON1.

INTRODUCTION

The broadband industry is fully aware of Nielsen's Law, which has popularized the "Billboard Internet Speed". Nielsen's Law of Internet Bandwidth simply states that a high-end user's connection speed grows by 50% per year. This means it doubles roughly every 21 months. It turns out that this "Law" has held fairly consistently for three decades now, having started with 300-baud phone modems in 1982. Nielsen's Law is important because the highest data speed offered is a determining factor for sizing the network.

The research of Dr. Thomas Cloonan, CTO of ARRIS [1,2] combined with Nielsen's Law is captured in Figure 1. The chart shows the growth from 1982 through the present and projected to the year 2030. This data, referred to as Cloonan's Curve, also reflects the historical 50% CAGR as does Nielsen's Law. Based on this, the highest Billboard Tier will out grow HFC capacity within the next 5-8 years. These predictions

have caused cable operators to seriously consider when to start a migration to Fiber To The Premise (FTTP).

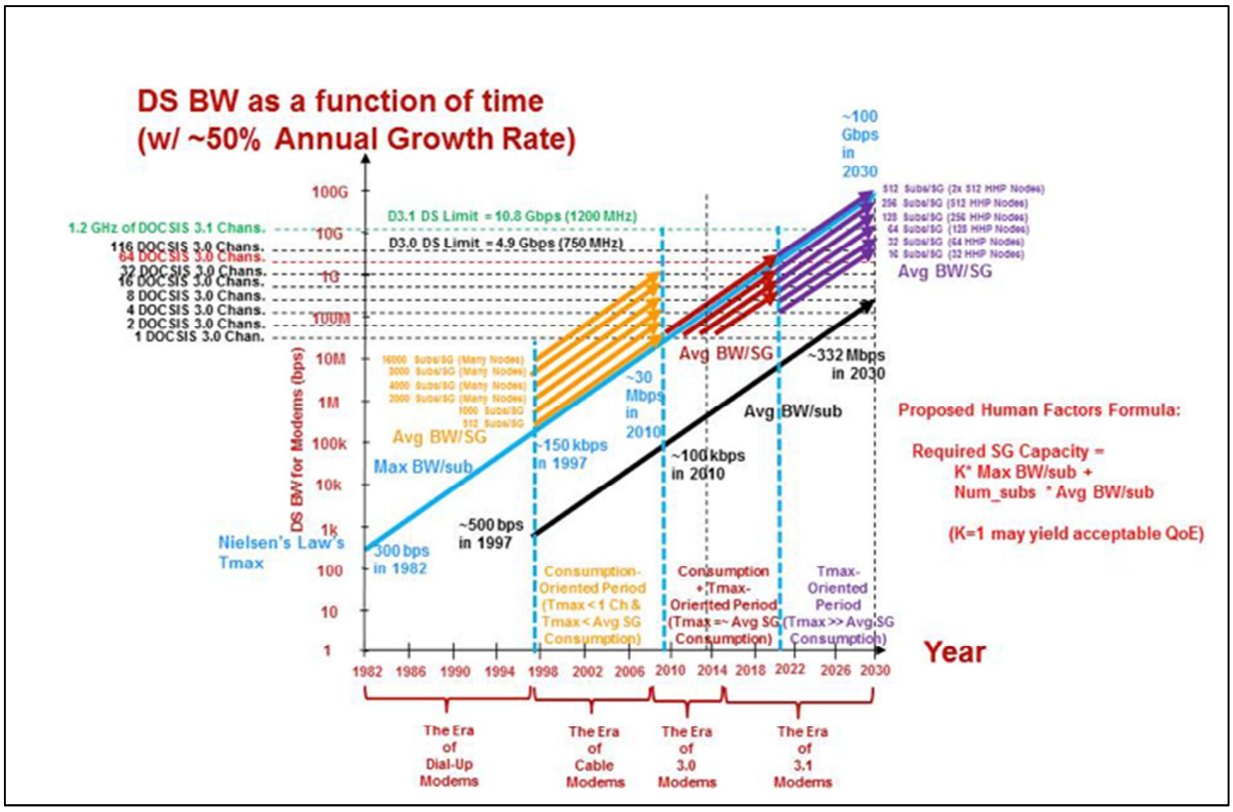


Figure 1: DOCSIS HSD Downstream Traffic Engineering predictions with 50% CAGR in future

As networks reach these critical inflection points many MSOs consider a move to all fiber as a good investment, unlocking the potential of the fiber and the ancillary benefits associated with increased reliability and lower power consumption such networks bring to bear. Competitive pressure from Telcos such as the FiOS and persistent demands from new housing developers all add to the crush of all fiber deployments.

Many today however equate a move to all fiber also with a move to binary-modulated PON based architectures. In reality, there are multiple options, ranging from the aforementioned binary-modulated PONs such as GPON and EPON and a new Hybrid PON (HPON) architecture that is fully compatible with existing HFC and completely transparent to any of the RF physical layer options including GPON/EPON.

We begin this paper with a discussion of all things involving a move to all fiber. While Greenfield builds are increasingly all fiber, the case of converting an existing HFC plant to all fiber is also then considered. The process of converting HFC plant to all fiber plant can be a long duration process, involving newer construction techniques. There will most likely be multiple logical stop points that are sufficiently fiber deep to let demand accumulate before a final push to all fiber is undertaken at the last mile. Since the process of going all fiber is a long duration event, it is critical to have a PON technology that is compatible with the current infrastructure at the headend, in the field and at the homes, has a long shelf life and a simple migration path. Finally we discuss HPON technology and how it satisfies all the requirements given above and aids the migration to all fiber.

TRANSITIONING TO AN ALL FIBER NETWORK IS A JOURNEY NOT AN EVENT

We begin with an interesting thought experiment: if indeed "money is no object", how long it would take to transition a nation to an all fiber infrastructure?

With a population of roughly 320 million, USA has approximately 130 million households. Great majority of those (~120 million) are reached by cable. These are referred to as "households passed" (HHP) or "homes passed" (HP). Pockets of HHP are served by HFC nodes, to a tune of approximately 400 HHP per node, thus there are a total of around 300,000 nodes in the country. An HFC node is oftentimes equivalent to an autonomous Service Group (SG) whose members share into services comprising HSD, telephony or VoD, for example.

While in a typical coax architecture of N+6 there are around 25 actives per node, a rule of thumb in the industry is that there is a fiber node per 5 plant miles in a typical Cable plant. Comcast for example, [3] with 628,000 plant miles and 125,000 confirms this rule of thumb. Furthermore, the typical distance from the hub to a node is around 10 miles. If these statistics are expanded and projected to a national level, there are about 1.5 million of coax plant miles. (300,000 nodes in the nation times ~5 plant miles per node results in ~1.5 million coax plant miles).

The buildup of this amazingly-capable communication infrastructure did not happen overnight – it took roughly 40 years (1950-1990) to complete the first wave of predominantly coaxial-cable plant, followed by ~20 years (mid 1990s to present) of intensive laying of fiber, predominantly for the trunk fiber lines, the ones connecting hubs to nodes. This historical perspective lets us arrive to the construction rate of approx. 37,500 miles of coax plant built per year, in the first 40 years of cable TV network buildout [3,4].

Another approach to take is to look at what a typical fiber construction crew can build per day / week / month, and extrapolate from there. In an urban setting, a crew may do a mile per day of an aerial plant, or it may do half-that if an underground plant. (A separate splicing crew would follow the construction crew to do all the fiber splicing and connecting). These assumptions give us about 2.5-5 miles per week, at 5 workdays per week. If we look at a blended aerial/underground rate of ~4 miles per crew per week, for a typical construction season that lasts 6 months or 26 weeks – it would result in ~100 miles per crew per year.

To complete a coax network to an all-fiber network transition over, say a period of 20 years, it would take ~75,000 miles of rebuild per year, and require ~750 fully-dedicated construction crews to do so.

Verizon FiOS deployment for example took roughly 8 years to build fiber to roughly 18 million homes [5,6]. Using our per-node statistics from before, the 18 million HP number corresponds to ~45,000 nodes and represents roughly 300,000 miles, constructed over 8 years. From these two numbers we get a construction rate of 37,500 miles of fiber route per year, which, interestingly enough, is exactly the same as the estimate of rate with which the first wave of coax construction took place! So if the Verizon FiOS buildout rate were projected onto the whole nation, it would take around 40 years to fiberize the nation.

During the transition, a legacy HFC network has to be operated, side-by-side, next to an all-fiber network. Setting up, connecting and maintaining the network and provisioning of services doubles – if HFC were to remain HFC and FTTP deployments exclusively go the way of a traditional PON technology. It is therefore essential to secure a technology that enables a harmonious coexistence of HFC and FTTP while not excluding the options of deploying traditional PON services for residential or business customers as needed. In this paper, we introduce just such a

technology, which we refer to as Hybrid PON (HPON).

Many MSOs have considered:

1. Capping coax plant growth and switching to fiber-only for the new builds
2. Success based fiber deployment for businesses that need it and are willing to pay for it
3. Selective buildout to a few very high-demand users, who need and can pay for higher service rates. As a result of this, in addition to satisfying the high bandwidth users, their former service group members get a fairer access to the newly available bandwidth

A "Fiber-to-the-Curb" (FTTC) approach enables all the three scenarios described above. With a "mini-node" (or ONU) located at the curb - say at a typical tap location - fiber route will pass by virtually all potential customers, so the whole area is described as a "fiberhood". All of the standard triple-play services are available at all the homes-passed, by the virtue of the existing coax drops. The new HPON architecture and technology to be described in this paper will enable very high capacity bandwidth offerings over standard coax [15].

For those customers who require more, a separate or additional fiber-drop is drawn from the mini-node to the customer premise, to provide a dedicated high-speed link, of a multiple Gigabit Symmetrical data rate. But this step needs to be taken only as needed, with options to connect any and all of the homes-passed [17]. This approach provides high capacity overall, preserves the option of serving high data customers with fibers and obviates the immediate need for fiberizing the last 200 feet or so of coax - which would be a huge cost and time saver – and upgrading it as part of a "success based" schedule.

An interesting technological option becoming more economically available nowadays comprises "coring" the existing hardline coax cables by removing the inside conductor and the dielectric foam, and then using the outside conductor as the conduit for the fiber optic cable. Such hardline cables typically terminate at the curb [14]. This approach provides for a fast and no-mess construction, with considerably lesser chance of blowing through a neighborhood electrical or a gas line.

TRAFFIC ENGINEERING FOR THE NEXT TWO DECADES

Research by co-author J. Ulm [7] of several large North American MSOs reveals that only a fraction of a percent of their customers actually subscribe to the Top Billboard Service Tier.

The study provides a breakdown of subscribers per Service Tiers. This is shown in Table 1. Anywhere from 76% to 94% of the subscribers fell into the most popular Common and Economy Tiers. The remaining subs fell into the Performance Tiers.

Subscriber Distribution	MSO #A	MSO #B	DS Speed	CAGR
Top Tier, Billboard speed	0.2%	0.1-1%	300 Mbps	50%
Performance Tiers	24%	5-18%	75	32%
Common, Popular Tiers	64%	50-72%	25	26%
Economy Tiers	12%	15-40%	5	15%

Table 1: Subscriber Mixes across the Service Tiers

In addition to the subscriber distribution across the different service tiers, the study also uncovered that the service tiers were all growing at different rates! While the Top Billboard Tier was following Nielsen's Law with a 50% CAGR, the Common Tiers had a growth rate of half that. But this actually makes sense. The operators are the ones actually controlling the growth rates. The Top Billboard rate is for bragging rights, yet only impacts a fraction of the customer base. The 50% CAGR results in CPE equipment being obsolete for that tier every 2-3 years. A small price to pay (or may be not) for top bragging rights.

Meanwhile, operators try to get as many years as possible out of the CPE investments for the majority of their customers in the Common and Economy Tiers. The lower CAGR for these tiers means that CPE equipment will remain viable for 5-8 years before being relegated to customers in lower tiers, making for a very financially successful business plan.

Figure 2 shows the predicted growth rates for the various Service Tiers over the next 20 years. If we consider 10Gbps to be the approximate capacity limit of HFC utilizing DOCSIS 3.1, it shows that the Top Billboard Tier still runs out of capacity in 6-8 years. However, the Performance Tiers have a 12-15 year runway before exhausting HFC capacity. Finally, the Common and Economy Tiers (e.g. 76%-94% of subs) are happily content with HFC capacity for 20+ years. In fact, if their growth rates stall then the majority of HFC subs might live on HFC "Forever".

The other interesting fact from this study is that traffic engineering for the next decade will be dominated by the Service Tier burst rate as defined by the DOCSIS Tmax parameter. This means that once below ~500 subs per SG, further SG size reductions have diminishing benefits. This is very important in an FTTP strategy. Many PON architectures are optimized for 32 or 64 ONU, while traditional DOCSIS systems handle eight times as many devices. What this means is that the PON architecture will require 8 times as many OLT ports as needed for CCAP.

HPON – AN FTTP TRANSFORMATION

A cable operator considering a significant FTTP investment to every subscriber in the near term may be providing capacity at significant expense for a need that might not materialize for decades, if ever as shown in Figure 2. The Hybrid PON (HPON) architecture allows the operator 'success based' steps towards FTTP, with more modest near term investment. Network capacity grows as it is needed and targeted at only the subscribers that actually need it.

A good first step is to push fiber a bit deeper (e.g. down to 250 Homes Passed) and upgrade the HFC plant to 1GHz or 1.2GHz and maybe move to an 85MHz upstream. While DOCSIS 3.1 will work in the existing HFC plant and give operators significant benefits; this upgrade helps enhance D3.1 capabilities. For about 20% of the cost of pulling FTTP to all subscribers, the operator now has an HFC plant with the equivalent capacity of 10G/1G EPON!

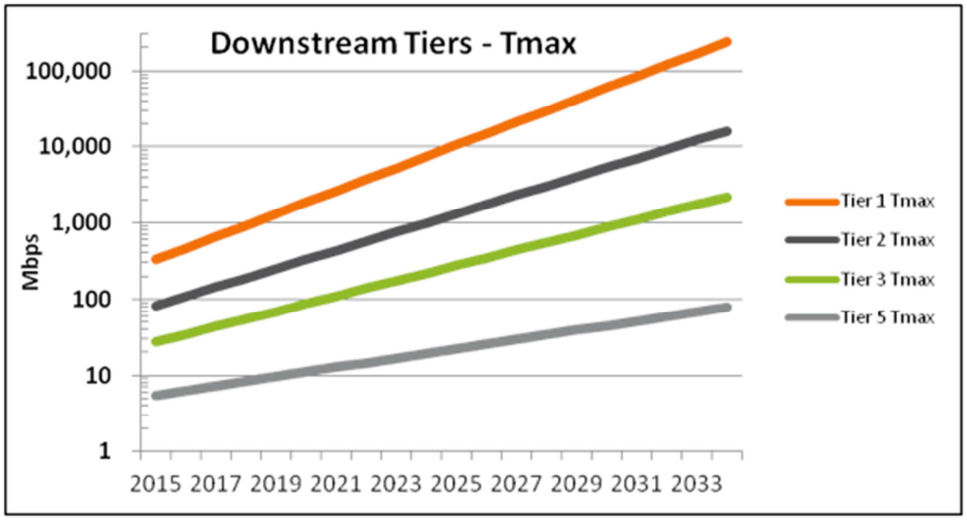


Figure 2: Multi-Service Tiers Growth

As was shown in the traffic engineering analysis in Figure 2, this relatively small investment in HFC has effectively doubled its capacity from ~5Gbps for 750MHz to 10Gbps for 1.2GHz D3.1; and will still be used for 20+ years by the majority of the Operator's subscribers. Plenty of time for the Operator's to recoup their investment.

Eventually, the Top Billboard Tier will out grow HFC, and the operator will need to move that small handful of subscribers to FTTP. With the HPON strategy, every time fiber is pushed deeper, sufficient dark fiber is pulled with FTTP everywhere as the end game. In this case it is pulled from the nearest optical node to the premise. These are all costs that would have been part of a full FTTP upgrade anyway, but have now been deferred in time. Once the first FTTP user is lit up, then the rest of the neighborhood is ready for FTTP service, with only the final fiber drop needing to be pulled. Also, operators can push optical nodes deeper at this time and eliminate even more of the HFC active components.

The HPON strategy allows operators to incrementally enhance all of their existing HFC plants in parallel and roll their FTTP evolution on a success driven basis. Doing a massive full FTTP migration would have forced the operator to this one serving area at a time and stagger this over a very long period. This means some areas may be under-served with HFC when they have Top Billboard Tier customers that need FTTP sooner.

COMPARING HPON AND TRADITIONAL PON CAPACITIES

Part of the attractiveness of HPON is that it is agnostic to the underlying technologies being transported over the FTTP, and will support a simultaneous mix of these technologies. So, one customer can get 10G/10G EPON if that's the service needed while other customers get OBI-free D3.1.

Note that HPON-based D3.1 has some interesting inherent advantages. Because the downstream and upstream are carried on separate wavelengths, their spectrums can overlap. In the downstream, spectrum might range from 54-1218MHz. This allows all legacy signals (e.g. STB OOB, FM music band) to be carried in their original spectrum and avoid the headaches caused moving the upstream split on existing HFC.

HPON-based D3.1 has even greater benefits in the upstream. With OBI-free operation and enhanced SNR (which is detailed later in the paper), the upstream can operate freely at 204MHz with up to 4096-QAM modulation. This is almost a 2.5G PHY rate and nets almost 2 Gbps of data capacity. Most importantly, the data rate is sufficiently high to offer a Gbps upstream service tier. This is something that 1G EPON, GPON and 10G/1G EPON cannot claim.

The bottom line, if we normalize HPON-based D3.1 to traditional PON PHY Rate terminology, it would be roughly equivalent to a 6G/2.5G PON. Taking a look at other possible D3.1 configurations, the downstream and upstream tables show how its capacity would align with various traditional PON technologies.

EPON SCHEDULER EFFICIENCIES

We now look at how the standard CMTS scheduler in an OBI free HPON architecture might match up with an EPON scheduler over a wide range of parameters including large Service Groups.

Previous research into EPON efficiencies [8,9] shows that the EPON Scheduler efficiencies are affected by several key parameters: ONU count, LLID per ONU, and the Grant Cycle time. These analyses were done for a very limited number of ONU/LLID since most people deploying PON were looking at using 32 or maybe 64 ONU per PON.

Downstream Spectrum	Nominal Data Capacity	PON Equiv
32 '3.0' + 96MHz OFDM	2 Gbps	GPON, 2 x 1G EPON
32 '3.0' + 2x192MHz	5 Gbps	2 x GPON, ½ 10G
32 '3.0' + 4x192MHz	8.7 Gbps	10G EPON, XG-PON1, NG-PON2
12-24 x 192MHz	~20-40Gbps	NG-PON2

Table 2: Mapping HPON D3.1 Downstream Options to PON Equivalents

Upstream Spectrum	Nominal Data Capacity	PON Equiv
85 MHz OFDMA	750 Mbps	1G/1G 10G/1G EPON, GPON
RFoG 204 MHz OFDMA	1.8 Gbps	EPON w/ 10G/1G co-exist, XG-PON1, NG-PON2 (2.5G)
RFoG 500 MHz OFDMA	~5 Gbps	EPON w/ 10G/1G co-exist
RFoG 1.2 GHz OFDMA	~10Gbps	10G/10G EPON, NG-PON2

Table 3: Mapping HPON D3.1 Upstream Options to PON Equivalents

The size of the PON SG was generally limited by the optical power budget, which is very sensitive to the distance. So maybe the service provider would have 32 ONU per PON at 20km.

Some recent advances may allow operators to deploy PON over longer distances and with larger fan-out. Technologies like Remote OLT and PON extenders can decouple the trunk link budget from the ODN fan-out. Conceivably, PON can support 128 or 256 ONU per SG as well.

However, PONs are a pure polling system whose overheads increase with SG size and Grant Cycle Time (GCT). While 10G EPON upstreams promise to make data transmission 10 times faster, the truth is that the PHY overheads such as laser turn on & off times may be the same as 1G upstream or just marginally better. Early PON systems also tended to be focused on best effort data services. As more latency sensitive services like voice and gaming are deployed, then Grant cycle times must be increased and additional polling overhead is incurred. The impact of all these overheads for large SG is an area for further research and will be detailed in subsequent papers.

While talking about upstream EPON efficiencies, it is also important to note the impact of 1G/10G upstream coexistence. It turns out that "10G EPON" actually comes in two flavors: 10G/1G and 10G/10G EPON. They both have 10 Gbps downstreams, but their upstream PHY rates are 1 Gbps and 10 Gbps respectively. The wavelengths for the two flavors are very close so they share a common optical receiver in the OLT. This means that the scheduler must treat them as a single SG and only allow one to transmit at a time.

The problem this creates is very similar to that seen in the WiFi world when 802.11g was introduced and had to co-exist with 802.11b devices. When the slower devices transmit, it takes a proportionately longer time to send its data. When a 1G upstream transmits a packet, it takes ten times longer to transmit as the same size packet from an ONU with a 10G upstream.

Some operators may want to deploy 10G/1G initially because they are more cost effective. Later, 10G/10G devices are deployed into the field. However, the co-existence issue might cripple upstream capacity. With a 50/50 traffic mix between 1G and 10G upstreams, there might be less than 2 Gbps of upstream capacity available. The impact of 1G & 10G coexistence is another area for further research and will be detailed in subsequent papers as well.

WHAT IS IN A NAME? PASSIVE OPTICAL NETWORK OR POWERED OPTICAL NETWORK

Cable operators requirements tend to have long distances and limited trunk fibers. While typical distances from the headend to the node are around 20 to 40 km, some of the longer links approach 60 to 80 km. As networks have grown manifold since the trunk fibers were laid out many decades back, the trunk fibers are always at a premium.

Typical Passive Optical Networks (PONs) serve anywhere from 16 HHPs to 128 HHPs with an average of around 32 HHP. There is therefore a deep relationship between the trunk distance traversed from the headend to the splitting location and the number of fiber splits that can be supported by a PON network. For example, a typical 24 dB PON network can support 20 km with 32 splits or 10 km with 64 splits or 2 km with 128 splits.

Because of the typical Cable operator distance requirements and the scarce trunk fibers, it is nearly always the case that PONs will always have remote OLT (Optical Line Terminator) or PON Extender locations very close to the fiber splits. These remote OLTs or the PON extenders typically perform OEO type of an operation to enhance the splits and the reach and many times also do Wavelength translations and aggregation to more fully utilize truck fibers. Therefore none of the commonly accepted PONs are really passive in the true sense and are really Powered Optical Networks.

	GPON	1G-EPON	XGPON1	10G/1G-EPON	10G/10G-EPON	
Spec	ITU-T G.984	IEEE 802.3bk-2013	ITU-T G.987	IEEE 802.3bk-2013	IEEE 802.3bk-2013	units
PON Rate	2.488G / 1.244G	1.25G/1.25G	10G/ 2.5G	10G/1.25G	10G/10G	Gbps
DS λ	1480 - 1500	1480 - 1500	1575-1580	1575-1580	1575-1580	nm
US λ	1310 ±20	1310 ±50	1260 - 1280	1310 ±50	1270 ±10	nm
Split Ratio	1:128	1:64	1:128	1:64	1:64	
Max Reach	20 (A, B, C) 40 (B+, C) 60 (C+)	PX10: ≥10 (1:16) PX20: ≥10 (1:32), ≥20 (1:16) PX30: 20 (1:32) PX40: 20 (1:64)	20, 40	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)	km

Table 4: Various Traditional EPON Standards

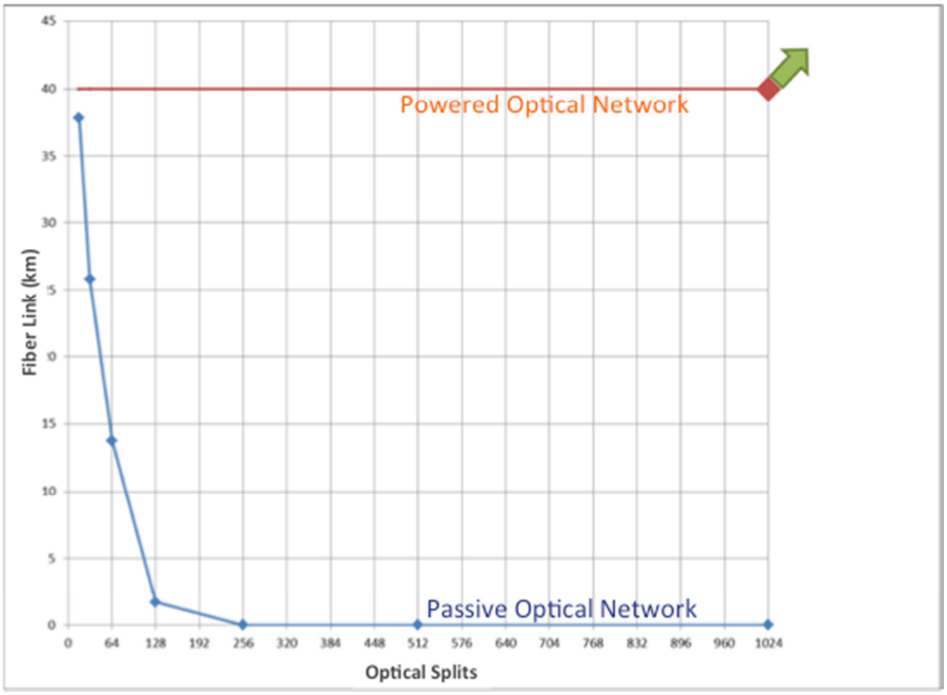


Figure 3: Illustrating Fiber Link vs. Fiber Splits for the passive and Powered Optical Networks

Typically, however this is not a serious problem. This is because if this were a Greenfield design, it is always possible to provide power for the remote OLT locations or to the PON Extender locations. Brownfield designs will already have power due to existing Coax network. While remote OLTs typically require large amounts of power and specialized Cabinets with heat exchangers, the PON Extenders, typically require more modest amounts of power and can typically be strand mounted within existing optical node housings. However it is always the case that typical PON systems require cabinets in one way or another to accommodate fiber splicing.

The Hybrid PON architecture described in this paper uses technology that eliminates OBI and facilitates the complete reuse of existing DOCSIS infrastructure. In addition, it enables data capacity that exceeds 10G EPON in the upstream and downstream while it simultaneously enables vast numbers of fiber splits covering up to 1024 HHP. Yet since the system has minimal processing power that is characteristic of traditional PONs, it requires very minimal power. The ability to reach 512 HHP for example requires less than 75W, which is about half of the power consumed by a typical 4X4 Node in deployment today.

The significance and benefit of a move from a processing intensive PON OLT operation to a minimal processing transparent HPON architecture cannot be overemphasized. In case of the traditional PON architecture, each iteration of speed increase would have to result in a change in the OLT and the PON extender as well as the ONU at the house. And we have seen that letting myriad forms of PONs to be in existence could lead to efficiency reductions. However in the HPON architecture, the downstream is completely transparent to the modulation PHY formats and the upstream is completely OBI free and a well-designed system is also transparent to the upstream PHY formats. Similar is the case for the ONU unit in the house. Therefore a vast majority of the devices remain in the network unaffected as improvements in speed; capacity and throughput get worked out over the next several decades.

OPTICAL BEAT INTERFERENCE

Optical Beat Interference (OBI) is a profound issue in RFoG reverse path and affects the system in debilitating ways [10].

In a traditional HFC system such as the one shown above, since the CMTS scheduler allows multiple CPEs to burst at the same time at different RF frequencies, there are no collisions in the upstream. This is not the case when the HFC system gets replaced with a traditional all optical RFoG system.

In an RFoG system, upstream bursts from the CPEs connect to their own Optical Network Units (ONUs). The ONUs detect the burst and then turn on. When they thus operate, they are all nominally on the same optical wavelength. In reality each laser is at a slightly different wavelength due to manufacturing tolerances of the lasers.

If these wavelengths are appreciably different, or if near identical wavelengths burst at different times, there is minimal interaction along the upstream. However, when the near identical wavelength ONUs simultaneously burst they create strong interaction in the receiver, this is called Optical Beat Interference (OBI).

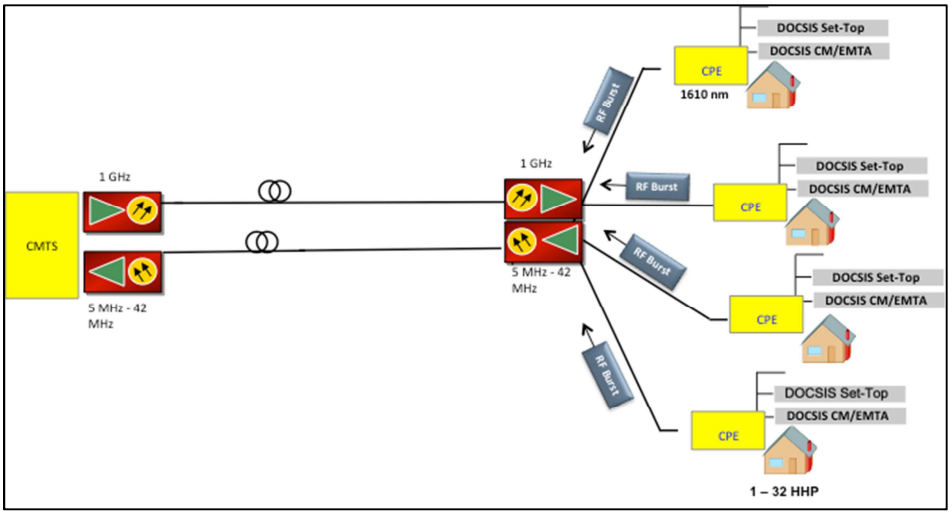


Figure 4: Illustrating a typical HFC Network

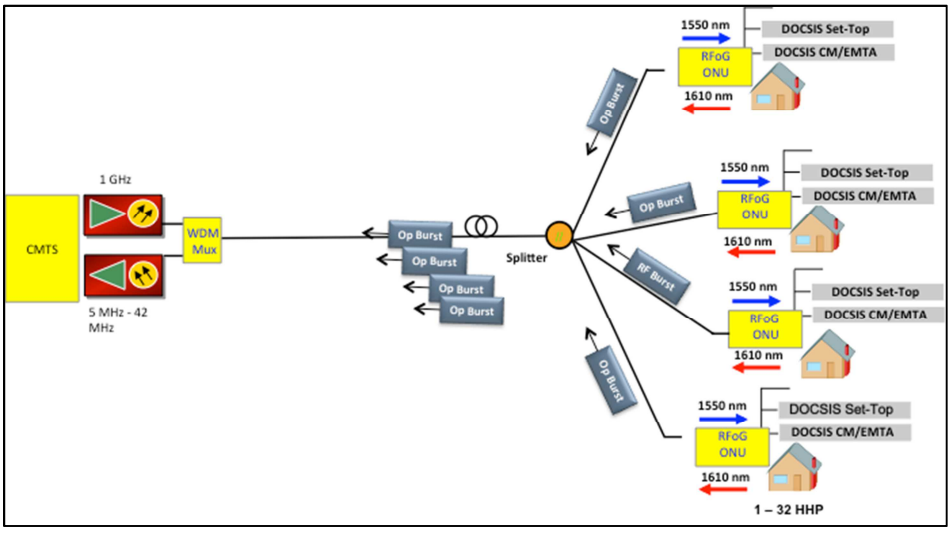


Figure 5: Illustrating a typical all fiber traditional RFoG network

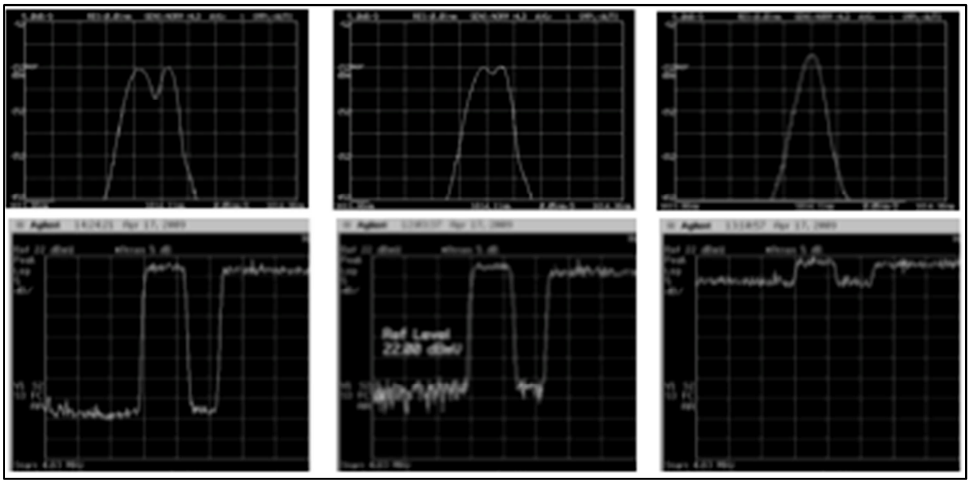


Figure 6: Illustrating OBI as a function of wavelength difference

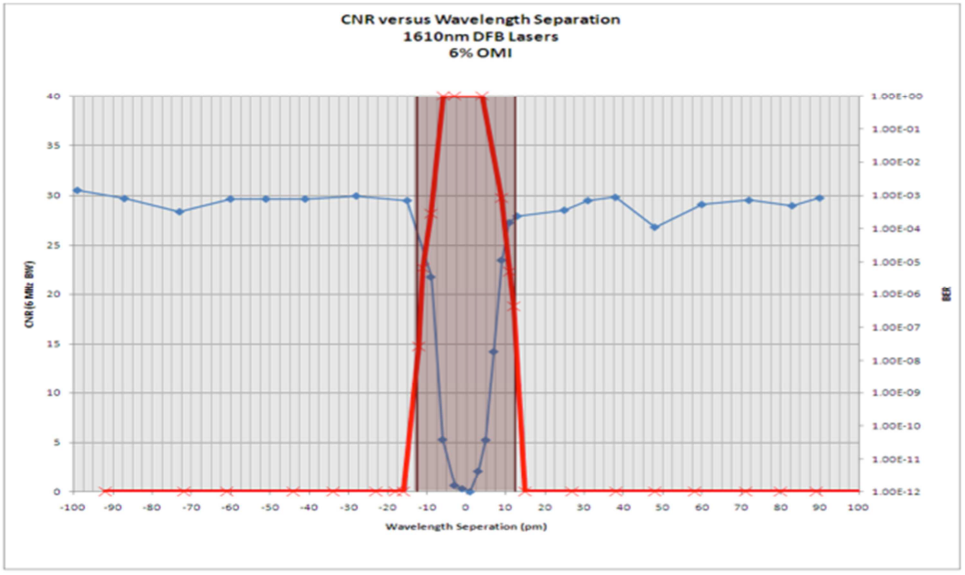


Figure 7: Illustrating SNR and BER under OBI conditions

OBI is a result of the heterodyning of the two (or more) closely spaced wavelengths present on the same detector. Heterodyning results in the down-mixing of the optical frequencies of the two or more lasers into the RF domain, appearing as wideband noise.

The resultant noise has the ability to impact the signal integrity of upstream communication channels, most notably on the receiver where the OBI has occurred, but can also affect the upstream DOCSIS service group via the introduction of noise into the RF combining network presented to the CMTS.

In the above shows a measured result of SNR and BER as a function of optical wavelength separation of two ONUs. When the ONUS wavelengths are far apart, the SNR and BER are good. However, when the wavelengths are within around 0.0125 nm of each other, there is a precipitous drop of SNR and a consequent rise of BER, corresponding to the picture shown above depicting the optical and RF spectra above.

It is important to realize that the 12 pm is around 10GHz of optical frequency, where typical 1610 nm lasers operate at around 200THz. Based on these test results it is seen that an OBI event wipes out the entire upstream SNR. In the case where there are more than 2 ONUs that can be simultaneously operational, such as when there are 4 DOCSIS 3.0 bonded channels, the net effect is that the an OBI event between any two ONUs wipes out the transmission of any other ONUs that were transmitting at the same time, even if they did not really contribute to the OBI. While it will be shown that this effect is already very severe in a 4 channel DOCSIS 3.0 case, it will be much worse when the number of bonded channels increase such as with the 85MHz D3.0 or with the 204MHz D3.1, where dozens of ONUs can simultaneously operate.

OBI probability depends upon the probability of two or more ONUs operating at the same time, and the probability of two or more ONUs having wavelengths within the OBI zone. The probability of two or more ONUs being on simultaneously is based on the Utilization of the upstream and is given by an application of the Binomial distribution, while the probability of the wavelengths being within the OBI zone is given by a ratio of the OBI zone relative to the entire region over which the optical wavelengths may be distributed. In modern laser designs, the wavelength range is typically less than 2.5 nm, with the OBI region be around 0.025 nm, the probability of OBI occurring is pretty high.

In reality however, the probability of the OBI occurring is much higher that the already high probability show above because of start up drift of a laser.

Since the ONUs only turn on when there is a burst of RF, the laser within the ONU is off until such time as it perceives a burst of the RF. Then, when the laser turns on, it starts absorbing the bias current applied to it and puts out optical power. All through the initial time of its turn on, until it reaches steady state, the temperature of the laser chip keeps increasing, with a consequent increase in the optical wavelength. All of this happens in a time scale of about a 100 to 200 us, and is fundamental to the operation of the laser. Since the time scale is so very short, it is not therefore possible to correct for the fast wavelength drift. The laser ultimately settles to a steady state value of about 0.5 nm higher than where it initially began operation.

The net result of this start up drift, which is a characteristic of the dynamic upstream system, is that the probability of OBI occurrence is increased by about two orders of magnitude, rendering the upstream system vulnerable to severe OBI for all meaningful upstream utilizations and a significant reduction in upstream throughput. What this means is that eliminating OBI needs to be treated as a strategy and cannot be solved by simplistic and opportunistic changes in optical wavelengths.

Eliminating OBI therefore requires the ONU wavelengths to be separated by at least 0.5 nm, preferably around 1 nm so that they never overlap. The only practical way of doing this is to DWDM like optics in the 1610nm region. Such a solution besides being expensive is also operationally impractical. This is because there would be fewer than 20 wavelengths possible in a typical SCTE specified 1610+/-10nm band. Therefore, there would have to one receiver per 20 ONUs and furthermore, the operator would have to map each household to one of the ONU wavelengths. The larger numbers of receivers thus needed would also increase the thermal noise in the system thereby drastically reducing the dynamic range of the upstream in addition to adding to the overall system cost. Finally, sparing costs would be prohibitive as would the system upgrades would be unintuitive and expensive.

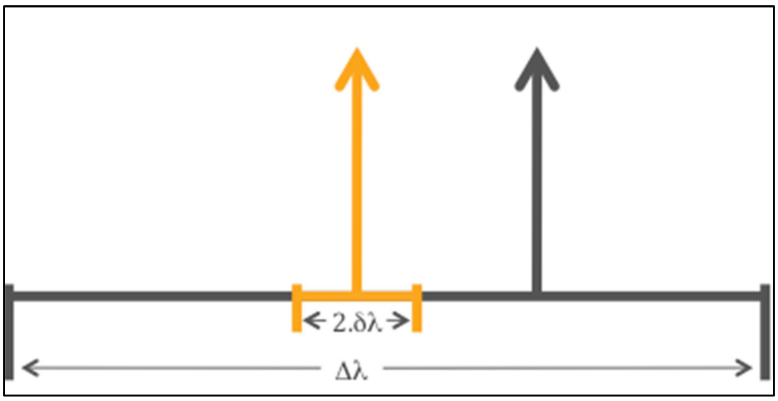


Figure 8: Evaluating the probability of OBI under static conditions

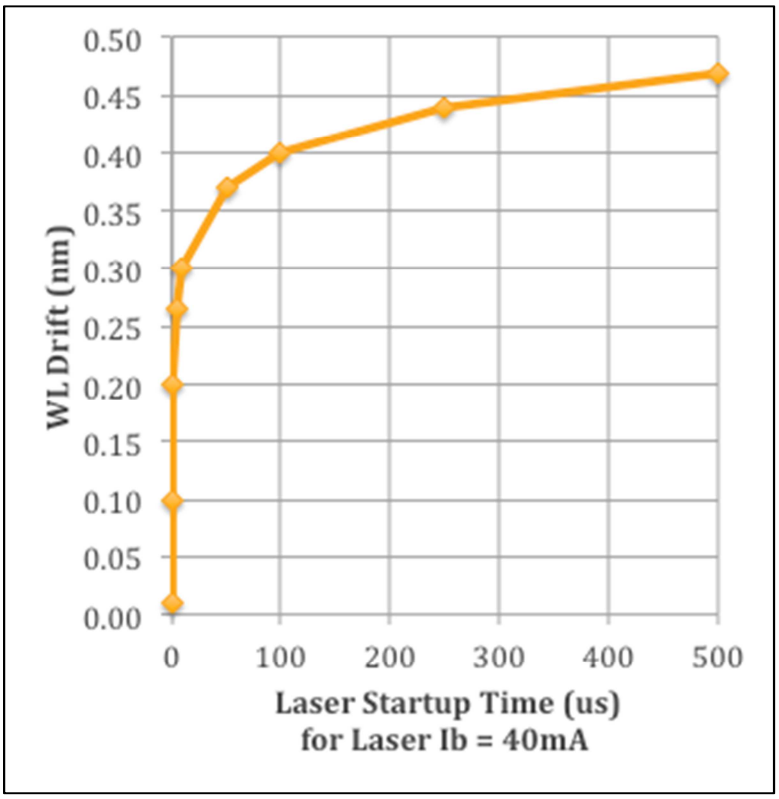


Figure 9: Illustrating the start up drift of an uncooled laser

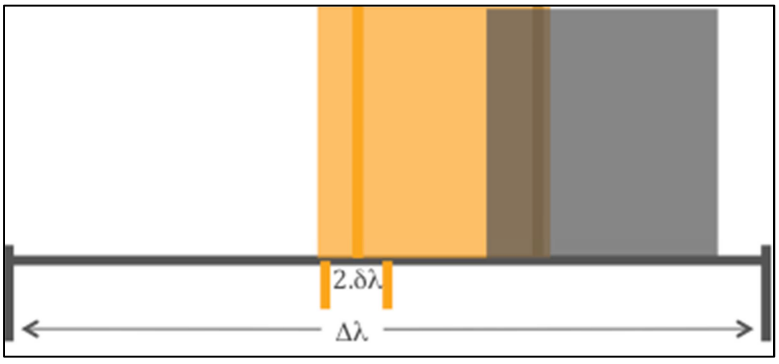


Figure 10: Evaluating the probability of OBI under burst mode conditions

THE ACCELERATED RFOG TEST

Since the dynamic nature of the wavelength drift is the dominant factor in eliminating OBI, it is easy to miss the exceedingly deleterious OBI effects if the test were done in a very simplistic static manner Therefore it is essential to evaluate OBI and potential elimination methods in the dynamic environment of a CMTS test bed.

The figure above represents a typical CMTS test bed. 32 CPEs feed 32 ONUs, which are combined in a splitter and go thru 20km of fiber. These are then connected to a receiver and a transmitter thru a WDM at the headend. A QAM256 channel is combined at the output of the receiver to act as a probe channel to be able to measure MER, BER and constellation diagram. Any impact on the probe channel is solely due to the OBI effects (recall that when OBI occurs, entire upstream is splattered by the OBI induced noise). The forward transmitter is then connected to the other end of the WDM and receives its HSD signals from the CMTS. The CMTS also receives the burst signals from receiver thus completing the loop. Both the CPEs and the CMTS are run thru a traffic generator, in this case thru a ByteBlower. An optical spectrum analyzer is connected thru an optical splitter as are RF spectrum analyzers to measure the upstream RF spectrum and the probe MER, BER and constellation. One can monitor the Code Word Error Rate (CER) before and after FEC on the CMTS as well as the frame loss on the ByteBlower as a measure of system performance.

When the CMTS and the ByteBlower are activated, the CPEs and the ONUs burst in the upstream creating OBI, which impacts the optical, RF and the CMTS domains.

Presented in Fig. 12 is a representative optical spectra of 8 ONUs with varying wavelength separation. After the CMTS is activated, the OSA is put on a max hold so we could clearly see the effects of startup drift.

Presented in Fig 13 is a representative optical spectra with Max hold on an OSA after just 2 minutes of the CMTS being activated. The smear of the optical wavelengths indicates the effects of the startup drift and the reason and the cause of exceedingly high amounts of Obi observed in real RFoG upstream systems. It is for this reason that opportunistic wavelength movement schemes of Obi mitigation do not suppress Obi effectively. And we have already discussed that meaningful wavelength separation is expensive, cumbersome and difficult.

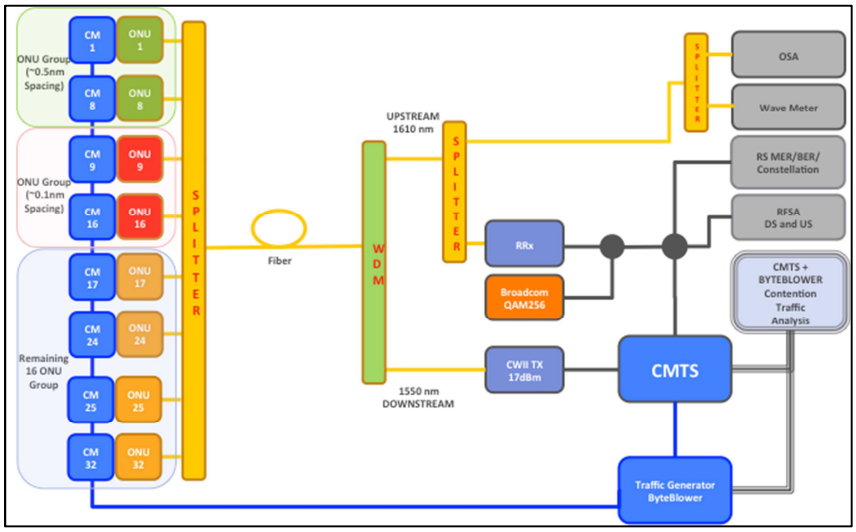


Figure 11: Comprehensive test bed to evaluate the effects of OBI

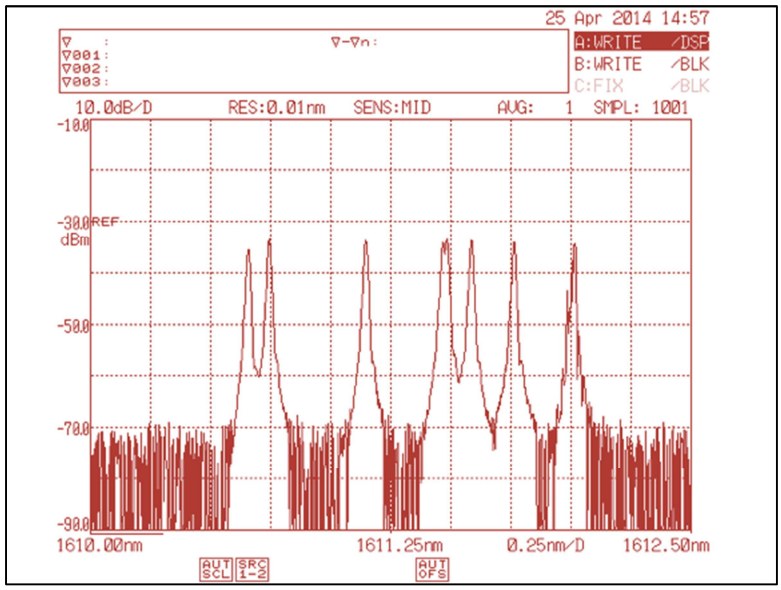


Figure 12: Measured optical spectrum under static conditions for traditional RFoG

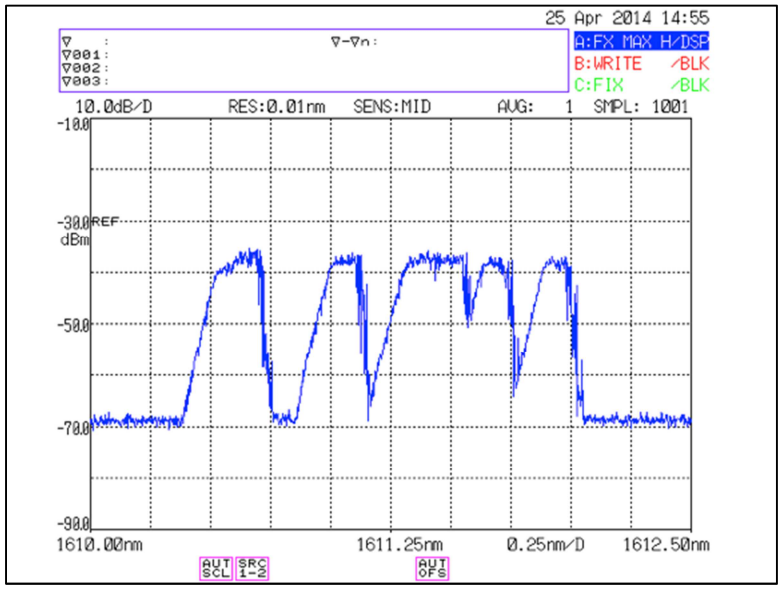


Figure 13: Measured optical spectrum under burst mode operation on Max Hold for traditional RFoG

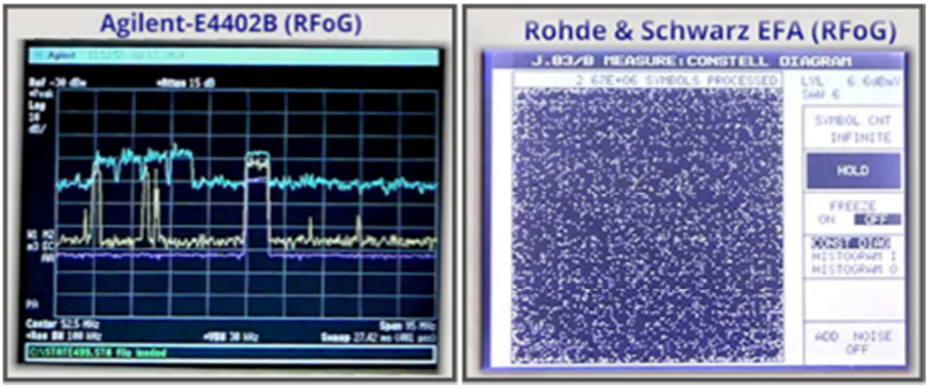


Figure 14: Measured Upstream RF Spectrum & constellation of probe channel with OBI in traditional RFoG

Presented above is the RF spectrum of upstream and a constellation diagram of the probe channel of a typical 32 ONU test bed described earlier for a utilization of 50% after a 2-minute test. It is seen that the spectrum is dominated by the OBI noise and the constellation is completely dark.

With such poor spectrum, the frame loss at the ByteBlower exceeds 20% and both the UDP and the TCP traffic at the upstream are severely affected. Since the TCP traffic at the upstream is severely affected, the TCP traffic of the downstream is affected as well.

The practical customer impact of such severe degradation is a significant impact on an Ookla speed test and a loss of significant voice quality on a VoIP connection. Over the Top services such as Netflix are affected as well due to accumulating TCP acknowledgements that are unable to make their way back to the Netflix servers in a timely manner.

Tests show that even modest amounts of utilization produce significant OBI and these contrive to affect a precipitous drop in a customer measured Ookla speed test as show above.

ELIMINATING OBI

We have previously seen that eliminating OBI at the ONUs is not possible with simplistic and opportunistic wavelength shifting methods. Eliminating OBI by using wavelength selective ONUs is expensive, cumbersome and impractical. An analysis of eliminating OBI by implementing a TDM CMTS scheduler is shown in the next section. Besides severely limiting throughputs this approach is no guarantee against OBI due to non-DOSCIS channels and due to malicious users. It is therefore essential to find a new way to eliminate OBI that is ONU and CMTS independent and immune to malicious users.

A complete elimination of OBI unlocks the full potential of the optical fiber and enables up to 10Gbps of upstream capacity and up to 20Gbps or more of downstream capacity with appropriate optical amplification. An active splitter configuration enables such OBI elimination with a very modest power requirement of about 5W.

Furthermore, the active splitter overcomes the upstream and downstream splitter losses thus enabling long distances and high splits as well as enabling the aforementioned high capacities. The large throughputs can also allow larger service groups which can help eliminate substantial upstream and downstream RF combining and splitting in the headends thus saving valuable headend space and reduce power consumption of headends.

With the same test configuration described earlier, but with the passive splitter replaced by the active splitter, the CMTS test described earlier is repeated to verify OBI elimination.

It is seen in Fig 17 that the RF spectrum is clean and the probe constellation is clear, thus establishing the OBI elimination even in the burst mode CMTS test bed.

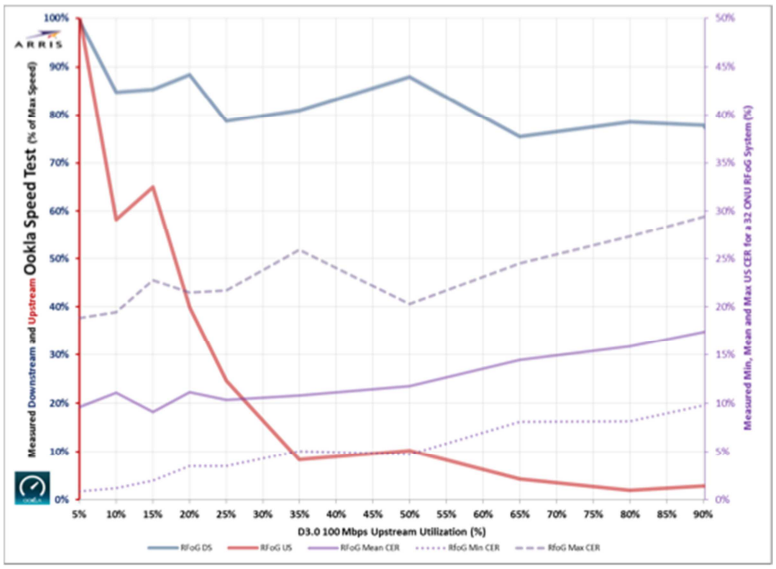


Figure 15: Measured speed test results under OBI for traditional RFoG

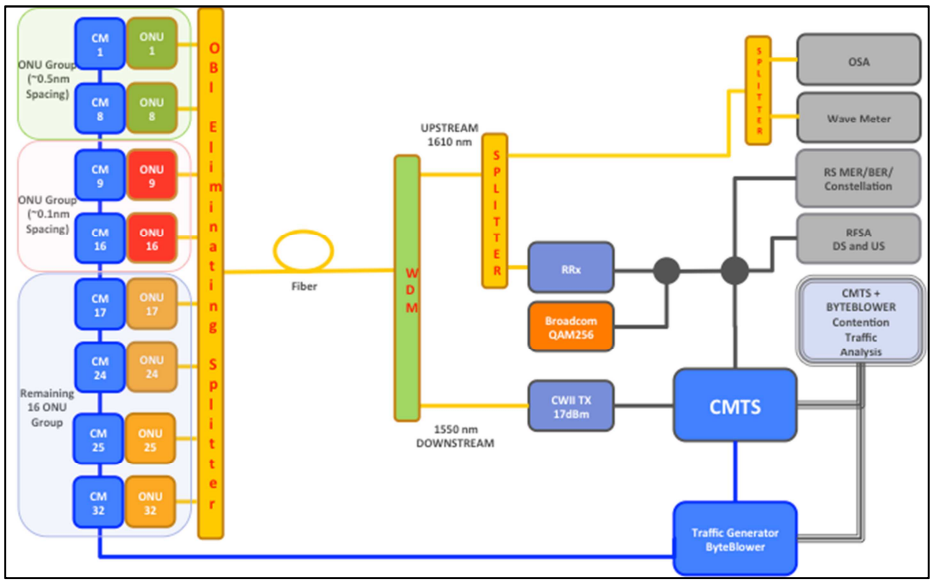


Figure 16: Comprehensive test bed to evaluate the effects of OBI with the OBI eliminating active splitter

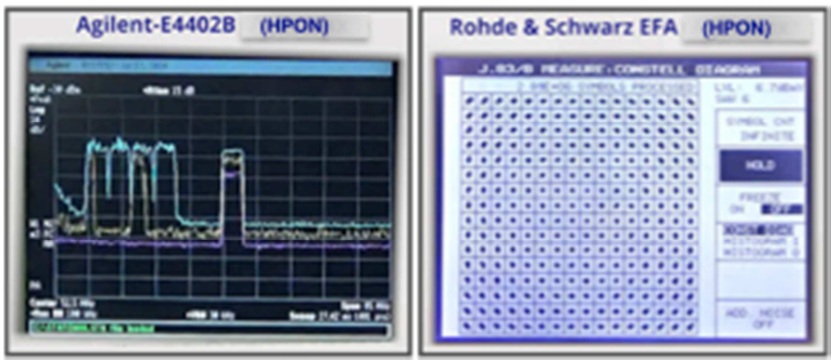


Figure 17: Measured RF Spectrum of the Upstream and the constellation of the probe channel under OBI conditions on an active splitter with OBI eliminated

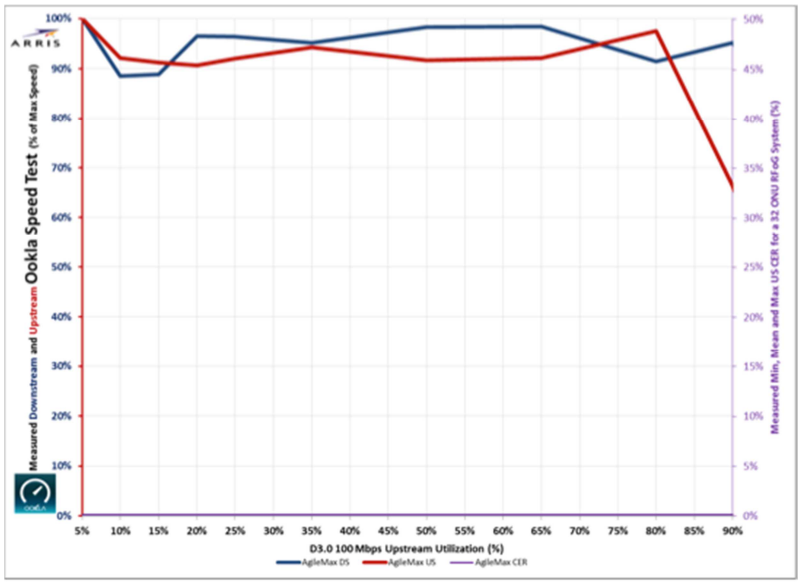


Figure 18: Measured speed test results under OBI for with the OBI eliminating active splitter

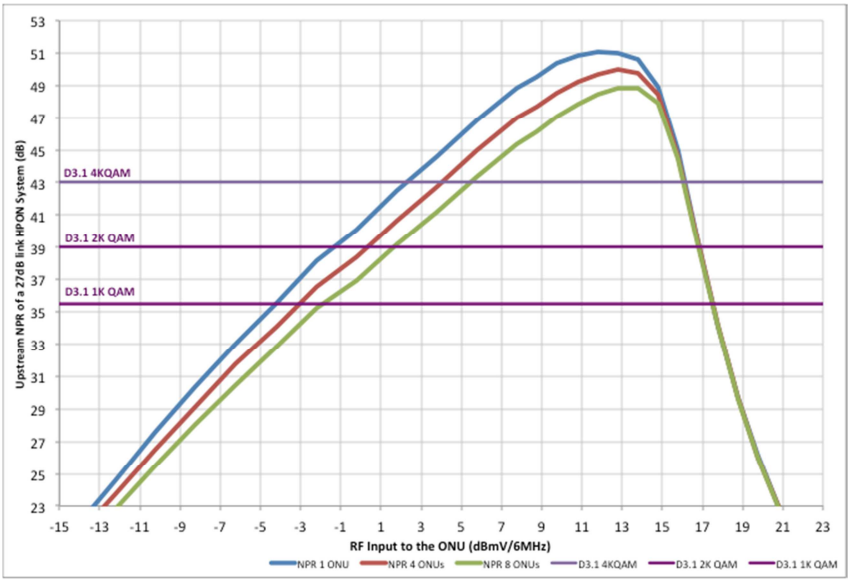


Figure 19: Measured NPR of a 25km, 27dB Link Budget 204 MHz HPON upstream system

Repeating the Ookla speed test confirms that the upstream and downstream capacity is unaffected. The Vonage connections on the VoIP system are also clear.

To further establish the complete elimination of OBI and the multiple benefits of the active splitter concept, NPR tests at 85MHz were done with progressively higher numbers of ONUs always on a 30dB RFoG upstream reverse path.

As can be seen from the measured data above, the NPR of the active splitter RFoG system is excellent with minimal noise side degradation with multiple ONUs on simultaneously [11]. This important test illustrates the fact that complex DOCSIS 3.1 modulation formats may be used for upstream transmission and that even if dozens of ONUs simultaneously operate as is expected in the DOCSIS 3.1 environment, the transmission rate is unaffected. Similar testing in the

downstream with present day ONUs and lower cost Direct modulated transmitters over similar distances yield better than 41dB MER, sufficient for 1K QAM operation to 1.2GHz. Since HPON upstream is in the 1610 nm band and the downstream is in the 1551 nm band, the Active Splitter allows a completely transparent path for all the other remaining wavelengths to pass through, thus enabling true RFoG and traditional GPON/EPON coexistence. An optimized multi wavelength plan enables high capacity upstream and downstream wavelengths to be used thus conserving trunk fiber. With this in place, one can deploy high performance, high capacity OBI free RFoG along with GPON and EPON for a truly Hybrid PON architecture [12,13].

OBI MITIGATION ATTEMPTS IN THE CMTS

We have seen in previous sections that eliminating OBI at the splitter location is effective and efficient. This gives the operator ultimate freedom to choose any best of breed ONU and CMTS from any vendor. This approach is vital to a viable and successful deployment of the HPON architecture.

As we have shown earlier, other PHY layer OBI mitigation attempts at the ONU are proprietary, cumbersome, impractical, expensive and do not scale with larger Service Group sizes.

Single TX CMTS Scheduler OBI Mitigation

Perhaps the most common method of mitigating OBI recently has been by using a specialized CMTS scheduler that only allows one DOCSIS cable modem to transmit at a time. This is conceptually simple and has had some success in the very early days of RFoG when most of the operators used DOCSIS 2.0.

While this approach eliminates the possibility of OBI from DOCSIS modems it does not address the possibility that OBI might occur when another service such as Legacy Settop Box (STB) Out-of-band (OOB) signals transmit at the same time as the DOCSIS modem. This is completely out of control of the CMTS scheduler.

It turns out that STB OOB traffic is relatively light so it would have minor impact on the DOCSIS traffic and show up as a slightly increased error rate. However, as RFoG becomes more successful and the DOCSIS utilization increases there is now a risk that the STB OOB traffic may be starved if at least one of the DOCSIS upstreams is being used every time the STB OOB tries to transmit. While this problem may not have been noticed yet, it is something that the operators need to be cognizant about.

In a DOCSIS 3.0 environment, multiple cable modems may transmit simultaneously. Therefore the opportunity of OBI is considerably more and approaches near certainty for increasing upstream channels, increasing modems and increasing utilization. Continuing this approach to mitigate OBI by the CMTS Scheduler in DOCSIS 3.0 environment allows only one cable modem may transmit at any one time. In addition to being a specialized software that is unique to RFoG, an operator who has been reliably running a CMTS Scheduler for years across millions of HFC subscribers must now use different software, even if it is from the same vendor. And as we will show, this approach is exceedingly inefficient and has very limited throughput, defeating the purpose of its development.

DOCSIS 3.0 CMTS Scheduler Efficiencies

The Active Splitter based OBI solution allows operators to use the identical CMTS Scheduler that they are using across millions of other subscribers. These schedulers have been optimized to deliver maximum performance across multiple channels with guaranteed Quality of Service (QoS).

Meanwhile, a Single TX OBI CMTS scheduler (a.k.a. TDM CMTS scheduler or the

"RFoG aware Scheduler) has made a conscious decision to limit transmission across all DOCSIS channels to a single modem. This means that the scheduler is starting to trade-off between total capacity and QoS. In the early RFoG days with very few DOCSIS upstream channels and limited utilization, this was an effective trade-off. However, as shown below, as the DOCSIS utilization increases, and the number of bonded RF channels increase from 4 to 8 and above this method becomes exceedingly inefficient and self-limiting.

Our analysis first looks at scheduler behavior for DOCSIS 3.0 systems. We assume that any DOCSIS 2.0 modems have been removed as they can significantly degrade performance of a Single TX OBI scheduler approach with multiple upstream channels. An example of a Single TX scheduler in a D3.0 system with four US channels would take a 1000 byte packet and fragment this into four simultaneous transmissions from this modem of 250 bytes plus DOCSIS overheads.

Obviously, sending nothing but extremely large packets would result in decent efficiencies. However, this would not be indicative of real world behavior, especially if it was a large Service Group with hundreds of modems. Research has shown that the upstream traffic is very bi-modal, with pre-dominantly small packets around 64 bytes and large packets near the Ethernet maximum of 1522 bytes. Some analysis of live broadband systems shows that a mix of 70% small packets and 30% large packets is reasonable.

Note that with this mix the average packet size is about 500 bytes, and that 91% of the data is actually carried by the large packets, even though they are only 30% of the total packets. Our results for a DOCSIS 3.0 system are shown in Fig 20. The number of upstream channels are varied from one to eight channels with each channel being 6.4MHz and 64-QAM.

Looking at the results in the Figure, the first observation is that both schedulers have the same efficiency for a single channel. This makes sense since nothing special needs to be done to avoid OBI. Also, the efficiency of ~88% is the result of the small packets being mixed in. Had there only been large packets, then the efficiency would have been in the 98% or 99% range.

As the number of DOCSIS 3.0 upstreams are increased, a standard CMTS scheduler with a splitter-based PHY Layer solution can schedule freely across all channels and won't see any efficiency degradation. It remains constant no matter how many upstream channels are deployed.

A Single TX CMTS Scheduler becomes more inefficient as channel count increases. If you think about a 64-byte packet fragmented across an eight-channel system, there is only 8 bytes of payload in each channel and the DOCSIS PHY burst Overhead starts to dominate.

At two upstream channels, the difference between the schedulers is a relatively small 13%. As the system capacity increases to four upstream channels, the Standard CMTS scheduler now provides ~40% more capacity than the Single TX OBI scheduler. Finally, at eight upstream 3.0 channels, the Standard CMTS Scheduler is providing twice the capacity of the Single TX OBI Scheduler. This would map to ~200 Mbps for a Standard CMTS Scheduler vs. ~100 Mbps for the Single TX CMTS Scheduler.

In addition to these DOCSIS data rate efficiencies, it should also be noted that a Single TX CMTS Scheduler would not be able to use any contention based Request (REQ) mechanisms either. This means that for large SG it may not scale as well and have increased latencies compared to a Standard CMTS Scheduler.

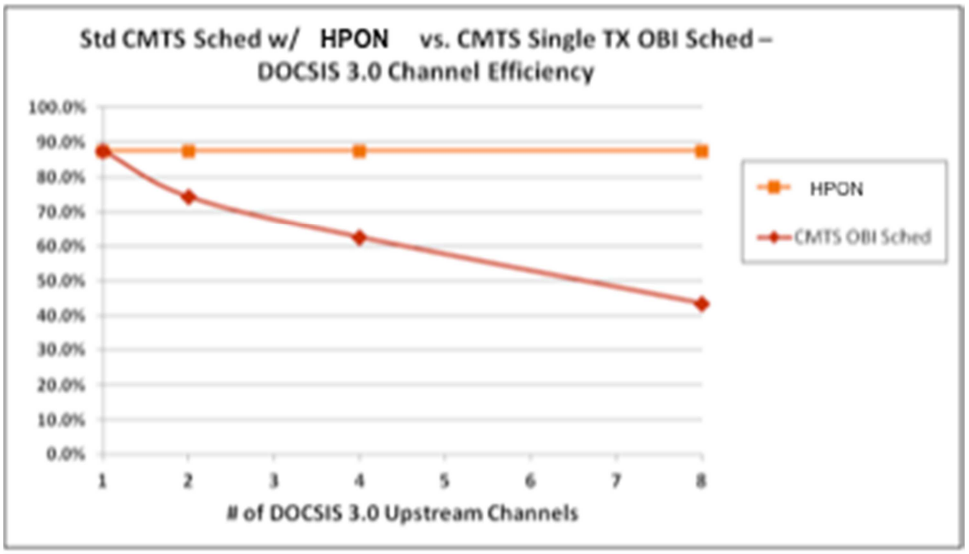


Figure 20: CMTS Scheduler Efficiencies – Std Sched over HPON, Single TX OBI Sched

DOCSIS 3.1 CMTS Scheduler Efficiencies

While the DOCSIS 3.0 results are very intriguing, perhaps the most important factor in scaling is the ability to deploy DOCSIS 3.1 capabilities over HPON architecture. D3.1 brings many potential technology improvements, but perhaps the most significant is the wide channels enabled by OFDMA. The first D3.1 products can potentially support up to a pair of 96 MHz OFDMA channels on a 204 MHz upstream.

In addition to the wider channels, the HPON-based D3.1 systems will have much improved SNR, which also enables 4096-QAM modulation in the upstream.

From an OBI perspective, the D3.1 OFDMA channels represent a serious challenge, as there may be dozens of different modems transmitting simultaneously on different subcarriers. To provide a handle on the potential impact, it is valuable to understand some basics of the D3.1 OFDMA channel.

The DOCSIS CMTS Scheduler is completely based on Mini-slots, for both D3.0 and D3.1. All transmit assignments, a.k.a. Grants, are a multiple number of Mini-slots. For a D3.0 ATDMA channel, Mini-slots are defined as a unit of time (e.g. 32 byte times) and is thus one-dimensional. For a D3.1 OFDMA channel, a Mini-slot is two-dimensional in both time and frequency.

The D3.1 Mini-slot is defined to be 400KHz wide in frequency. This maps to 8 subcarriers if using 2K FFT or 16 subcarriers if using 4K FFT. In the time dimension, the Mini-slot can vary from 120 to 360 microseconds. This maps to 6 to 18 symbol times for 2K FFT or 6 to 9 symbol times for 4K FFT. The symbol depths were chosen to allow interleaving in the upstream.

In D3.1, a Frame Time is a collection of Mini-slots that are all synchronized and transmitted at the same instant of time. The total number of Mini-slots per Frame Time is now a function of the total upstream spectrum available to the OFDMA channels. This is shown in Table 5.

The number of payload bytes per Frame Time can vary significantly. If the objective is to minimize the total bytes per Frame Time to reduce the possibility of OBI, then that leads to the selection of 2K FFT with only six symbol times per Mini-slot. The last row of Table 5 shows how many payload bytes are available in a Frame Time for this

configuration and various spectrums. Note that this is best case for OBI and could increase up to a factor of six larger than this.

The number of payload bytes per Mini-slot can vary as well, and might fall in the range of 50 to 400 bytes. For the scenario above (i.e. 2K FFT, 6 symbols per Mini-slot), a small 64 byte packet will consume 2 Mini-slots and a large 1522 byte Ethernet packet might be ~30 Mini-slots. As can be seen from Table 5, a D3.1 system can potentially have dozens of simultaneous transmitters.

To eliminate OBI in D3.1 with a Single TX CMTS Scheduler becomes extremely problematic. If a modem has a single 64-byte packet to transmit, they must consume the entire Frame Time in order to eliminate OBI. Even transmitting large 1522 byte packets, or concatenations with 4500 bytes or 9000 bytes is small compared to the total bytes available in a Frame Time, especially the 33KB Frame Time for 204MHz @ 4096-QAM. The CMTS Scheduler efficiency results with the same packet size mix of 70% small, 30% large packets are shown in Figure 21.

D3.1 Mini-slot Efficiency	42 MHz	65 MHz	85 MHz	204 MHz	204 MHz Enhanced
Total Mini-slots per Frame Time	~82	~140	~190	474	474
Modulation	1024-QAM	1024-QAM	1024-QAM	1024-QAM	4096-QAM
Bytes per Frame Time	4483	7862	10,774	27,608	33,178

Table 5: DOCSIS 3.1 Upstream OFDMA Frame Parameters

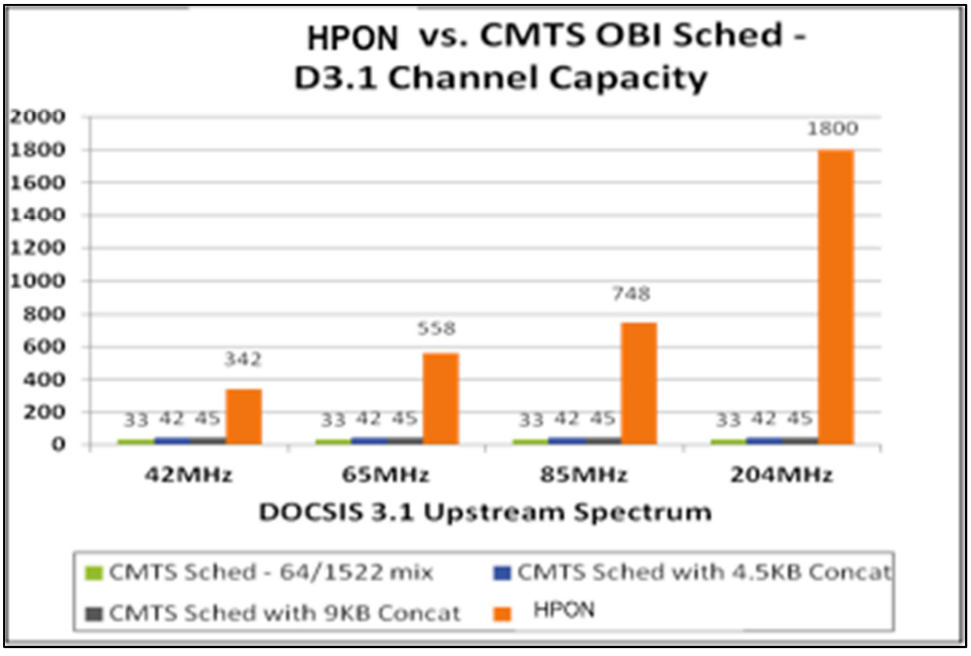


Figure 21: D3.1 Scheduler Efficiencies – Stud Sched over active splitter, Single TX OBI Sched

As can be seen in Figure 21, the Single TX CMTS Scheduler cannot even come close to filling a 42MHz OFDMA channel. In fact, it gets higher capacity using D3.0 ATMDA whose Mini-slots are shorter in time.

Meanwhile, the Standard CMTS Scheduler operating over HPON cannot only achieve full D3.1 OFMA utilization, it can also take advantage of the improved upstream SNR to operate at 4096-QAM. With HPON enabling overlapping upstream and downstream spectrum, then 204MHz operation becomes a reality and HPON-based D3.1 can deliver a 1.8Gbps upstream data rate (almost 2.3Gbps PHY rate). This is more than enough to offer a true 1Gbps upstream service tier to cable customers.

FUTURE OF BINARY AND ANALOG OPTICS

At 1550 nm RFoG/HPON is more suitable for SMF-28 fiber access networks than binary modulation. This surprising result is due to the twin effects of the superior capability of QAM schemes to take full advantage of available bandwidth and convert it to higher throughput relative to the binary formats and the effects of fiber dispersion on links where fundamental limits can cap NRZ binary modulation formats to around 20 Gbps throughput at 25 km and fewer at greater distances. This paper indicates that HPON when used with the DOCSIS 3.1 standards can support greater than 40 Gbps and distances up to 40 km on a single wavelength.

Attainable SNR of the optical link is a function of frequency and distance [11, 16]. 40 Gbps throughput can be obtained with around 4 GHz of bandwidth.

For a binary system the throughput was simply given as:

$$bin_bps(L) := 10 \cdot \sqrt{\frac{120}{L}}$$

This is plotted as markers for the 3 fiber lengths at their respective Nyquist bandwidths. Whereas the binary modulation permits operation at higher bandwidth (compared to a dual sideband AM modulated signal around a carrier) it does not provide more throughput.

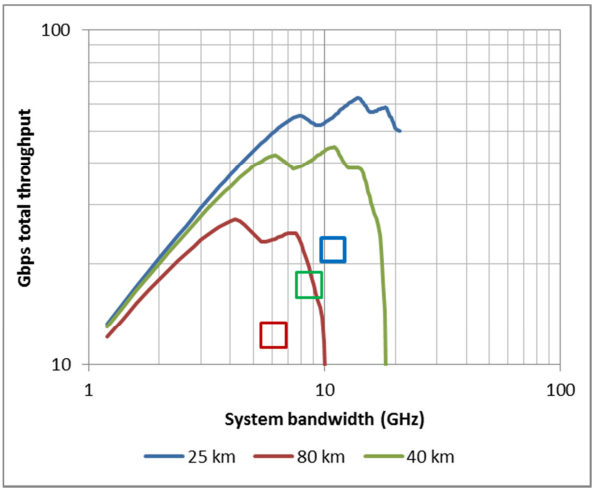


Figure 22: Illustrating the Capacity of an RFoG link with OFDM/LDPC vs. Binary Modulation

The addition of an EDFA in the headend, as needed for RFoG, increases the available downstream bandwidth in the system from 10 Gbps (PON) to more than 40 Gbps (RFoG). As shown in the figure even more throughput can be available with wide bandwidth receivers. However this paper is limited to ONUs that can be realized with currently available low cost components and the discussion is therefore be limited to around 5 GHz receiver bandwidth. It is further concluded that due to dispersion a binary transmission format cannot reach the same throughput as an RFoG system with EDFAs, neither at 25 km reach nor at 40 km reach.

RFoG Receiver performance

Estimated capacity is shown below as a function of receiver power with 4 GHz receive bandwidth at 25, 40 and 80 km. At -6 dBm receive power almost 40 Gbps is

achieved, the performance saturates above –3 dBm to 40 Gbps.

These estimates are for a multi-carrier RF modulated signal (OFDM with LDPC) with a modulation index set to the same value as for current QAM256 operation. With peak-average power reduction (PAPR) the throughput can be increased by around 10% beyond these results. Margins needed for current DOCSIS 3.1 impementations can reduce throughput by around 10%.

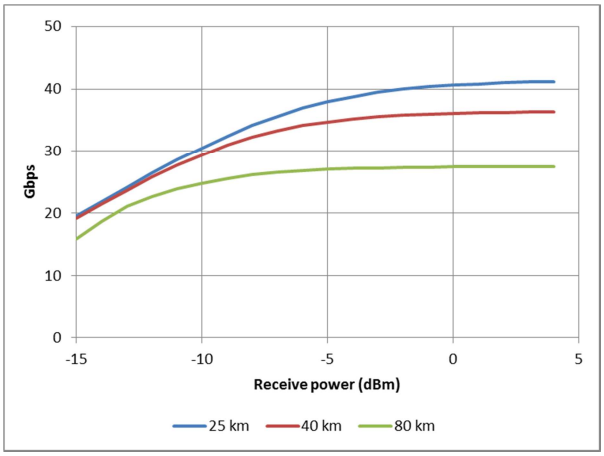


Figure 23: Illustrating System Throughput in Gbps as a function of receiver input power

RF amplification of the received signal to 4 GHz or higher is readily available with low-cost commercial MMICs. Thus the receive side of the ONU can be realized with a subset of the components used in a 10 Gbps PON ONU. The required components can satisfy reduced bandwidth and gain requirements. The transmit side of the ONU that will be proposed here can be based on a regular directly modulated DFB laser with 1 GHz of bandwidth. Given the availability of wideband lasers and amplifiers that is not significantly more complex than RFoG ONUs with less upstream bandwidth.

Modulation Formats and Throughput

In the access part of the plant the fiber loss budgets are relatively small compared to long haul telecommunication. In RFoG the number of forward wavelengths is also small compared to long haul communication systems. By using optical amplification in the access plant, permitted for one or a few wavelengths on a fiber, the SNR can be increased so much that analog modulation formats, in spite of their shortcomings, readily obtain 4x the throughput of binary modulation formats given practical receiver bandwidths available today using low cost components. Binary modulation formats with optical amplification could permit larger service groups, however due to traffic engineering, the size of service groups needs to be limited unless much more than 10 Gbps forward capacity can be offered, that is not available today in low cost binary data processing receiver and electronics and its application is limited by fiber dispersion. Therefore the binary modulation formats do not benefit from optical amplification in the access. It can thus be concluded that, unlike in long-haul telecommunication, analog modulation formats are superior to binary modulation formats in the access plant. This should come as no surprise, in most short distance bandwidth limited systems analog modulation formats are used (DSL, CAT-6 10 GbE cables, WiFi and of course DOCSIS).

While it may sound contradictory to classify a fiber system as a bandwidth limited system it is fiber dispersion and the OE conversion (receiver) and associated processing electronics that do pose a practical bandwidth limitation in fiber systems. In fact this bandwidth limitation is acknowledged by 40G PON manufacturers resorting to 4 wavelengths each operating at 10 Gbps rather than attempting to run one wavelength at 40 Gbps. Reality is that aggregate throughput levels are expected to increase to a level that far exceeds the baseband bandwidth of low cost receiver electronics that are used with simple on-off keying on standard SMF fiber. Current low-cost analog technology on the other hand can already provide a transparent

pipe with the required throughput capability today.

FUTURE UPSTREAM CAPACITY

In the upstream SNR is evaluated accounting for laser RIN, laser power and loss budget, the upstream bandwidth and the modulation index that can be allowed per channel given a total bandwidth, an additional dynamic window to accommodate uncertainty in transmitter OMI setting and the optional use of PAPR (Peak to Average Power Reduction) in the transmission format.

The effective modulation index just under clipping induced BER is 20% for regular RF modulated systems and around 32% for RF modulated systems with PAPR (Peak Average Power Reduction) methods applied.

The return bandwidth can vary, up to 1 GHz should not significantly affect return laser and driver cost, this primarily affects the choice of diplex filters. The attainable throuput is analyzed [11] and provides up to 10 Gbps upstream capacity. In a 204 MHz return system more than 1.5 Gbps of upstream capacity can be attained. It should be noted that in an RFoG system the forward and reverse bandwidths could be overlapping such that 1 Gbps symmetrical operation can be offered in a system with an 85 MHz split in the in-home network.

CONCLUSIONS

We began this paper with a premise that a move to all fiber need not necessarily be a move to one of the traditional binary PON architectures. Since a move to all fiber is a long duration process spanning several decades maintaining disparate end-to-end systems could be cumbersome and expensive.

In this paper, we presented an innovative Hybrid PON architecture that enables a true coexistence of traditional HFC and traditional PONs (such as GPON and EPON) along with an OBI Free high capacity HPON system that is capable of supporting analog, broadcast video, D3.0 and D3.1 formats.

We also discussed the true nature of OBI and its debilitating impact on tests that affect customer perception - such as the Ookla Speed Test. We also discussed the various significant inadequacies of some approaches that seek to just mitigate OBI rather than fundamentally eliminate it and the glaring inefficiencies of others that could eliminate OBI but at a significant loss of throughput that defeats the purpose of their deployment.

A deployment of the OBI Free HPON system as described in this paper maintains all options to deploy traditional binary PON architectures for business customers and very high data consumers, while enabling a well-ordered migration to FTTH by supporting traditional HFC as well. Thus HPON satisfies the anticipated growth in traffic demand while providing MSOs with plant and equipment investment protection well into the future.

REFERENCES

1. Predictions on the Evolution of Access Networks to the Year 2030 & Beyond; T. Cloonan, M. Emmendorfer, J. Ulm, A. Al-Banna, S. Chari, The Cable Show NCTA/SCTE Technical Sessions Spring 2014
2. Neilsen's Law vs. Nielsen TV Viewership for Network Capacity Planning, Mike Emmendorfer, NCTA 2014
3. Comcast hybrid fiber and coax plant: http://business.comcast.com/docs/default-source/brochures/the_comcast_network.pdf?sfvrsn=0
4. Time Warner Cable fiber plant: http://www.carrierethernetnews.com/articles/828971/time-warner-cable-business-class-enhances-ethernet/
5. Verizon FiOS FTTP deployment: en.wikipedia.org/wiki/Verizon_FiOS
6. www.bloomberg.com/news/articles/2015-02-05/verizon-to-sell-landlines-towers-in-two-deals-for-15-6-billion
7. Is Nielsen Ready to Retire? Latest Developments in Bandwidth Capacity Planning; John Ulm, T. Cloonan, M. Emmendorfer, J. Finkelstein, J.P. Fioroni; SCTE Cable-Tec Expo 2014
8. 10 G-EPON: Drivers, Challenges, and Solutions; Glen Kramer, ECOC 2009
9. Performance of 10 G EPON; Rajesh Roy, Glen Kramer, Marek Hajduczenia, Henrique J. Silva, Topics in Optical Communications, IEEE Communications, November 2011
10. Beat It! Handling OBI In RFoG Systems; Ayham Al-Banna, Larry Spaete, Amarildo Vieira, Sebnem ZorluOzer, Marcel Schemmann, Tom Cloonan, Dean Stoneback, Venk Mutalik, NCTA 2014
11. Partnership for Extended Capacity: DOCSIS 3.1 with RFoG; Venk Mutalik, Marcel Schemmann, Ayham Al-Banna, Zoran Maricevic, NCTA 2014
12. When Wavelength Collide Chaos Ensues: Engineering Stable and Robust Full Spectrum Multi-Wavelength Networks; Venk Mutalik, Fernando Villaruel, Marcel Schemmann, Kwang-Yi Wu, SCTE 2011
13. Value of Transparency in the Evolution of the HFC Networks; Venk Mutalik, Marcel Schemmann, Brent Arnold, SCTE 2013
14. Coax Core Ejection technology: http://deepfibersolutions.com/applications/
15. Ayham Al-Banna et. al., "Measurement-Based EoL Stochastic Analysis and DOCSIS 3.1 Spectral Gain," in Proceedings, The NCTA Cable Show Spring Technical Forum, May, 2015.
16. Mike Emmendorfer et. al, "Estimating Downstream Performance and DOCSIS 3.1 Capacity in CAA and DAA Systems," in Proceedings, The NCTA Cable Show Spring Technical Forum, May, 2015.
17. Tom Cloonan et. al., "Lessons from Telco and Wireless Providers: Extending the Life of the HFC Plant with New Technologies," in Proceedings, The NCTA Cable Show Spring Technical Forum, May, 2015

TIME SCHEMES IN REMOTE PHY ARCHITECTURE

Hang Jin and Yubin Chen
Cisco Systems

Abstract

The Remote DOCSIS Timing Interface (R-DTI) is the timing specification standardized recently for R-PHY architecture. R-DTI leverages on IEEE1588v2 with a special profile for DOCSIS applications. This paper explains how R-DTI works and what are the requirements on the time/frequency synchronization between CCAP core and RPD to ensure proper DOCSIS operations. Two operation modes that are defined in R-DTI, the Node_Slave mode and Node_Master mode, are explained and their operations are illustrated. Tests are done in Cisco R-PHY Lab to validate R-DTI under various network conditions. Test results show that, with the commercial R-DTI solutions, the timing and frequency synchronization between CCAP core and RPD can be achieved within 5 minutes in all the test cases and CM can operate normally without any performance degradation after the synchronization is achieved.

INTRODUCTION

Remote PHY (R-PHY) has been defined in Modular Headend Architecture v2 (MHAv2) (reference 1). It promises higher capacity and lower operation cost by moving the PHY from the CCAP at headend to the optic node and connecting CCAP core and node through digital fiber. The connection between the node and CM is still DOCSIS and over coaxial, but with much short distance. This results in improved signal quality and reduced OPEX. The specifications of R-PHY architecture have been standardized by CableLabs and participating vendors (reference 2).

Since PHY locates in optic node, separated from CCAP core, timing/frequency synchronization are required between node and core to ensure proper operation. The conventional DOCSIS DTI timing is not appropriate for R-PHY application. CableLabs R-PHY specification has defined R-DTI (remote DOCSIS timing interface) as the timing protocol for the timing/frequency synchronization between CCAP core and node in R-PHY architecture. It leverages the IEEE 1588-2008 (1588v2) protocols and defines specific profile for DOCSIS R-PHY application (reference 3).

Deploying 1588v2 in R-PHY architecture meets with lots of challenges due to the stringent timing/frequency sync requirements of DOCSIS PHY, MPEG video and OOB (out-of-band) signals. In this paper, we introduce the timing scheme in R-PHY architecture: what are the timing/frequency synchronization requirements for CCAP core and node to support various services (data, video, and OOB), how 1588v2 works in R-PHY architecture, and what the design challenges and solutions are. Also include are the Lab test data on R-DTI performances.

TIME/FREQUENCY SYNCHRONIZATION REQUIREMENTS

Figure 1 shows the simplified R-PHY architecture emphasizing the Remote DOCSIS Timing Interface (R-DTI). In R-PHY architecture, the PHY of CCAP is moved to the node, called as R-PHY device (RPD), and the rest of the CCAP is called CCAP core. CMTS core and the RPD are usually located in two different physical

locations, and their operations require they synchronize in both timing and frequency. Timing and frequency synchronization between CCAP core and RPD are achieved through the precision time protocol (PTP, IEEE 1588v2) as defined in R-DTI.

The requirements for timing and synchronization of the DOCSIS system come from the following areas:

1) DOCSIS physical level timing requirements
2) Time/frequency synchronization between CCAP core and R-PHY device (RPD)
3) Precision timing services like mobile backhaul, FCC911

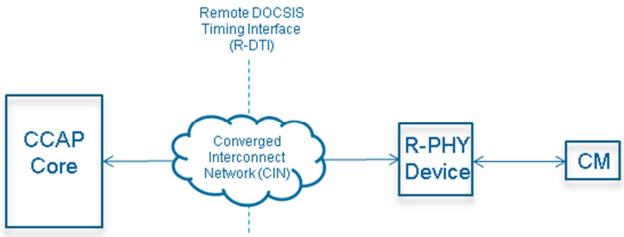


Figure 1. High level R-PHY Architecture

Figure 2 shows the R-PHY's clocking architecture. The time and frequency recovered from the 1588 DPLL/NCO will drive the DS PHY and US PHY. It should meet the clock quality specifications as follows.

- Physical level requirements of DOCSIS 3.0/3.1 PHY, EQAM, OOB and NDF/NDR:
  - DOCSIS reference clock (10.24MHz or 204.8MHz)
    Accuracy: +/-5ppm
    Jitter or phase noise: DOCSIS DRFI spec for D3.0 and PHY spec for D3.1
    Frequency drift rate: 10ppb per second for DOCSIS data
    10ppm per hour for QAM video
  - Time
    Time jitter: SYNC jitter must be less than 500ns (DOCSIS 3.0)
- Time/frequency synchronization between CCAP core and RPD
  - Accuracy requirements: time +/- 1ms; frequency: +/- 200ppb
- Precision timing services like mobile backhaul, FCC-911
  These requirements are product specific. R-PHY system could use DTP protocol to support advanced timing services.

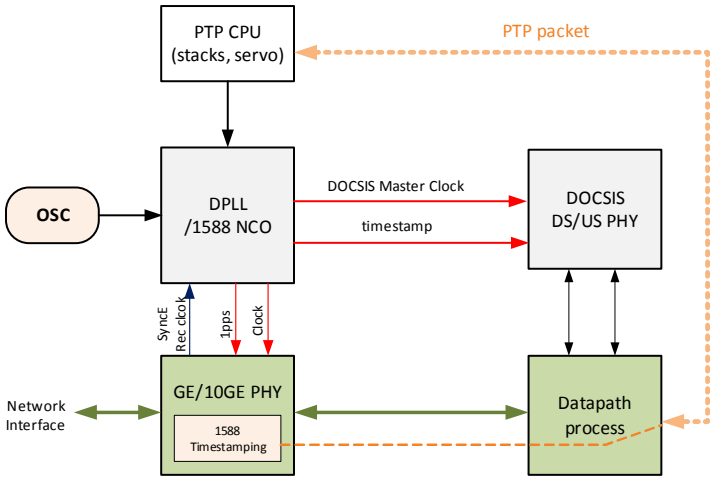


Figure 2. R-PHY Clocking Architecture

HOW 1588v2 WORKS IN R-PHY NETWORK

R-DTI adopts IEEE 1588v2 with a frequency and time & phase profile customized for R-PHY applications. IEEE 1588v2 provides time synchronization between two nodes across a packet network without mandating all intermediate nodes being replaced, hence becomes the best tool for the R-DTI implementation. There are two operation modes defined in R-PHY architecture.

1) Node_Slave operation mode

Figure 3 is the timing architecture for the Node_Slave operation mode. RPD acts as the timing slave, and its clock is driven by an external master clock. The master clock may collocate with the CCAP core, thus the CCAP core provides with the master clock (figure 3(a)), or in general, the master clock locates in the cloud, and the CCAP core itself is a slave to this network based grand master (GM) as well (Figure 3(b)). In the latter case, the CCAP core will be a Boundary Clock: it acts as a slave synchronized to the network GM but serves as the master clock to the RPD. In both cases, in the Node_Slave mode, RPD always acts as an Ordinary Clock (OC) synchronized to the master clock coming from CCAP core. This will ensure all network devices (RPDs and the CCAP core) have a single clock source which could be derived from the network 1588 GM. DOCSIS operation requires all devices are in a single clock domain.

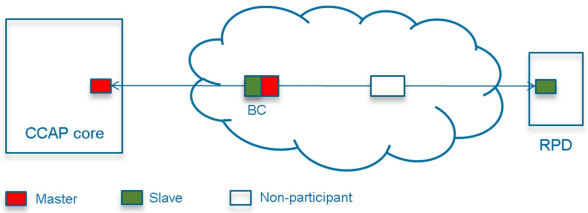


Figure 3(a). Master clock located with CCAP core

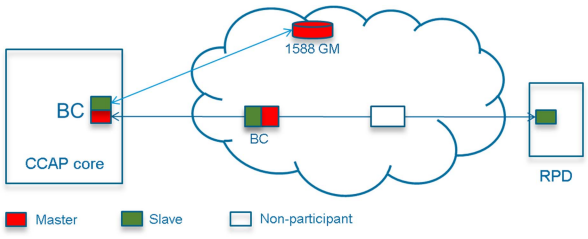


Figure 3(b). Network based 1588 GM

Figure 4 shows the basic synchronization message exchanges in the Node_Slave operation mode. It is evident that the time and frequency synchronization are achieved by the 1588 client located in RPD, as specified in IEEE 1588v2 (reference 4).

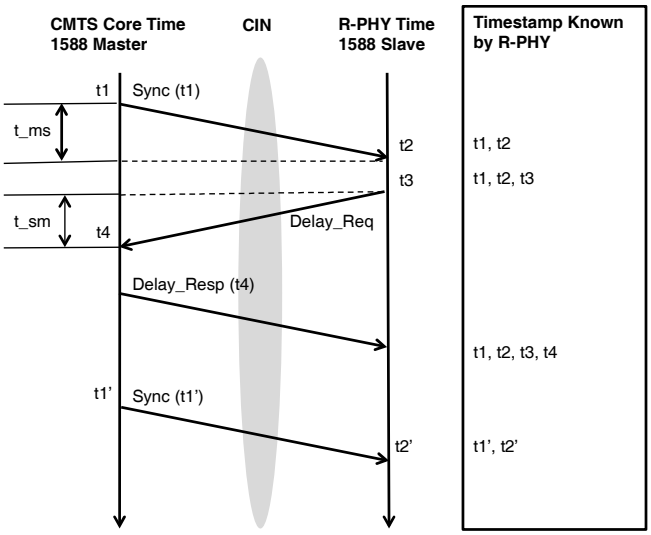


Figure 4. Synchronization message exchanges in the Node_Slave operation mode

2) Node_Master operation mode

The other operation mode defined in R-DTI is the Node_Master mode. Shown in Figure 5 is the timing architecture for the Node_Master operation mode.

In the Node_Master operation mode, the R-PHY acts as a 1588 Master, while CMTS Core acts as a 1588 Slave.

In this mode, the CCAP core needs to generate multiple clock islands, one for each RPDs that it connects to. The RPD distributes its frequency and phase information through 1588 synchronization messages.

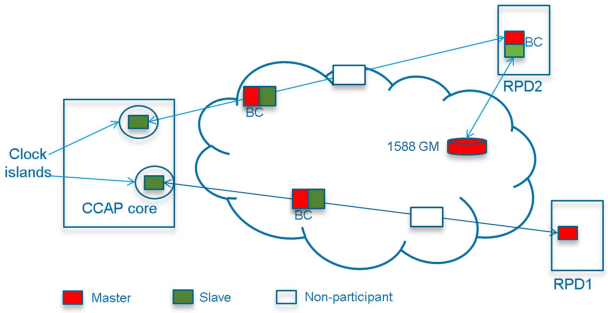


Figure 5. Operation Mode – Node_Master

CMTS Core obtains the frequency and phase information from the synchronization messages and runs a phase calibration process to track the RPD time and frequency. The timing and frequency information for each RPD will be kept in the clock island corresponding to that RPD. Any messages to a RPD (for example, MAC scheduling) with timing stamps will be sent out with the 'local timing' in its corresponding clock island.

The local timings in the clock islands need be updated continuously through IEEE 1588 messages. Figure 6 shows the basic synchronization message exchanges in this case between the CCAP core and the RPD. In this case the time and frequency synchronization are achieved by the 1588 client located in CCAP core.

Please note: in this case, the local clock and timing at the CCAP core will not be adjusted to synchronize with any particular RPD. The timing/frequency synchronization between CCAP core and RPDs are logic through the clock islands.

In the Node_Master operation mode, the local clock at RPD may be in a free run mode with an internal clock source such as an oscillator, or RPD may synchronize to a 1588 GM in the network. In the latter case, the R-PHY act like a 1588 BC – it is a slave clock to the 1588 GM in the network and acts like a master clock to the CCAP Core. As described before, the local clock at the CCAP core does not need be adjusted to synchronize to the RPD, as the clock synchronization between CCAP core and RPD are logic and through the clock islands in this Node_Master operation mode. The CMTS core may be in a free run mode or synchronized to an external clock source rather than RPD's.

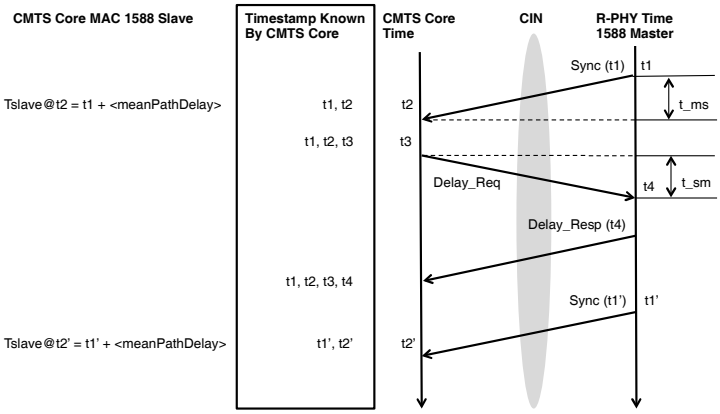


Figure 6. Synchronization message exchanges in the Node_Master operation mode

R-DTI specifies the RPDs must support both Node_Slave and Node_Master operation modes. The CCAP core must support the Node_Slave mode, and may support the Node_Master mode.

IEEE 1588 PROFILE FOR R-PHY

IEEE 1588v2 provides time synchronization between two nodes across a packet network without mandating all intermediate nodes being replaced, hence becomes the best tool for the R-DTI implementation. The accuracy requirements of the PTP synchronization are application-dependent. Sub-microsecond accuracy is required for typical telecom applications (3G/LTE, etc.). The accuracy requirements for DOCSIS operations are much lower (see the next paragraph).

Therefore, R-DTI leverage the existing IEEE-1588v2 protocol but with tailored profile for DOCSIS applications. In other words, R-DTI is a specific 1588 profile for DOCSIS R-PHY application. The same PTP stacks and hardware components can be used at both master and client sides. Often, the R-DTI client has the capability to be configured to operate in two different modes: a) DOCSIS R-PHY Mode; b) Advanced Timing Mode.

1) DOCSIS R-PHY mode

In this mode, the profile definition considers the timing and frequency requirements for proper operation of DOCSIS, video, OOB, and NDF/NDR in R-PHY architecture. Precisely, the 1588 profile for R-PHY calls for the following requirements:

- Phase alignment <= 1ms
- Frequency accuracy <= +/- 5ppm
- Frequency drift <= 10ppb/s and 10ppm/hr
- Phase jitter < 500ns
- Lock time from cold start < 5 minutes
- Worst time and frequency offset between two nodes (cold start): 10ppm in frequency, and arbitrary time offset in time

These requirements are customized for R-PHY application, ensuring proper DOCSIS operations.

When IEEE 1588 client is used in R-PHY (either at CCAP core or RPD), it servo algorithm should be tailored for this RPHY profile. The timing and frequency synchronization should be realized in two phases: the initial acquisition phase where the phase, frequency can be adjusted arbitrarily, and the steady phase where disruptive time adjustment shall not be allowed, and the frequency adjustment slope shall be less than 10ppb/s (and 10ppm/hr). The acquisition phase shall be completed in less than 5 minutes, as illustrated in Figure 7.

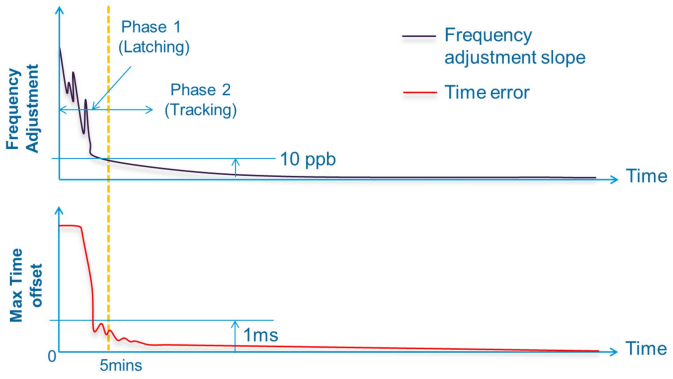


Figure 7. IEEE 1588 client servo algorithm in two phases.

2) Advanced Timing Mode

In the advanced timing mode, the output clock quality could be significantly improved to meet the mobile backhaul or DTP requirements. As such the servo could require a much longer time to converge, leading a much longer clock settle time.

SIMULATION AND TEST RESULTS

To evaluate timing and frequency synchronization performances of R-DTI in various CIN scenarios, commercial R-DTI solutions have been tested in Cisco R-PHY Lab. Figure 8 and Figure 9 show the two test setups. Test Setup-A is for testing R-DTI client performance under various CIN conditions. Different CIN conditions are modelled with different network PDV. In Setup-B, the recovered time and frequency out of the R-DTI client evaluation board were fed into the DOCSIS 3.0 R-PHY node. The purpose of test Setup-B is for evaluating the impact of R-DTI on actual R-PHY operations.

1) Test Setup A

Equipment used in test Setup-A includes:

- PTP client evaluation board from R-DTI solution vendors. The servo algorithm is specific for R-DTI
- 1588v2 grand master, TP-5000
- Calnex's Paragon-X, it is used for simulating the PDV of CIN and measuring the recovered clock quality
- Raken TCXO E6615LF on the PTP evaluation board

2) Setup B

Setup-B is based on Setup-A but contains additional equipment for evaluating the impacts of R-DTI on R-PHY operation:

- Cisco R-PHY system, includes uBR10K, R-PHY node(CMC), and cable modems
- IXIA for simulation of the user traffic
- Switches for connecting CMTS core, R-PHY, and IXIA

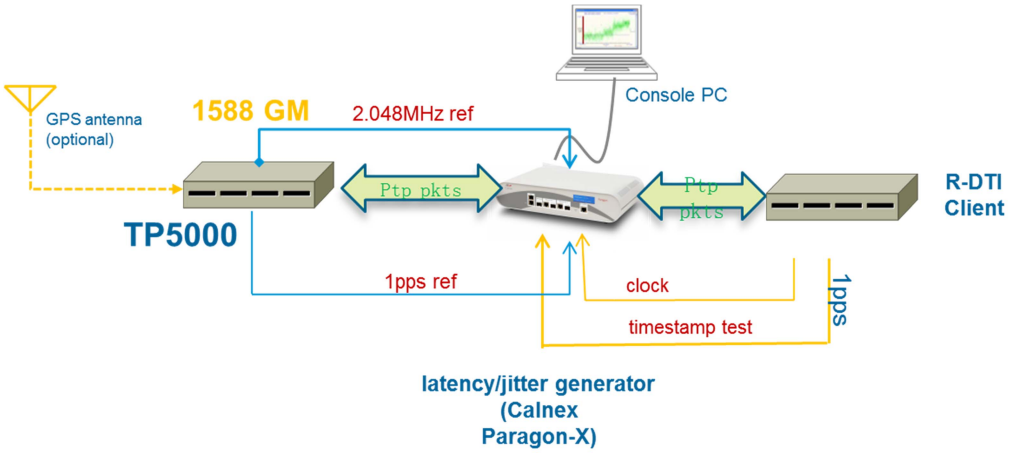


Figure 8. Test Setup-A

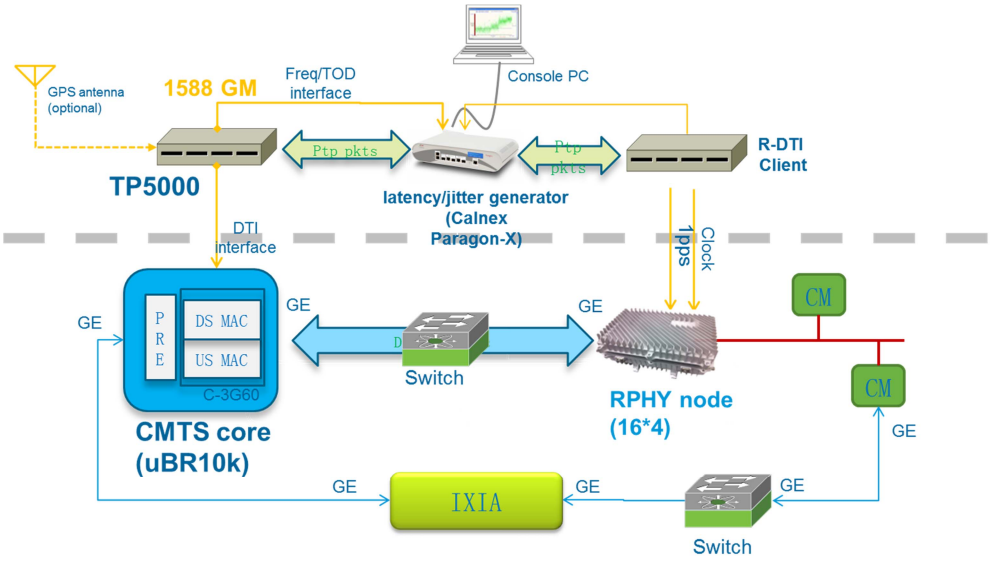


Figure 9. Test Setup-B

In both test Setup-A and Setup-B, R-DTI is configured as follows:

- 1-step 1588, unicast, 16pps sync packet rate
- R-DTI profile in PTP evaluation board

3) R-DTI Test results with Setup-A

Test cases in Setup-A were created to provide sufficient coverages on network conditions and address some key issues with R-DTI. As IEEE 1588 servo algorithm is tailored for R-PHY application, which has a much relaxed timing accuracy requirement compared to typical telecom applications, one expects much faster timing and frequency convergences in all test cases. Calnex Paragon-X is used to simulate various CIN network conditions through various packet delay variation (PDV), and the time/frequency synchronization performances in different phases are measured for each of the test cases.

The simulated networking conditions covered:

- Gaussian random packet delay variation with different peak, mean, or variation
- Symmetric & asymmetric delay between master and client
- Different traffic load, link re-route, random noise, network congestions, dynamic floor/mean delay
- 1x, 10x, 20x, 50x, 100x of ITU-T reference PDV pattern of Metro Ethernet, defined in G.8261 TC12~17
- Measured PDV data from some service providers' network
- Measured PDV data from Cisco routers
- RPD using R-DTI synchronization in different work status

Table 1 lists all the test cases, and the convergence times and time/frequency accuracies achieved after the time/frequency lock is declared in each case.

Table 1. Test Cases

TC#	Description	Lock time (s)	Time error after lock (us)	Frequency offset after lock (ppb)
1	Symmetric PDV, Gaussian p2p 5ms	**240**	**179**	359
2	Symmetric PDV, Gaussian p2p 10ms	**240**	171	601
3	Asymmetric PDV, -uplink Gaussian p2p 10ms -downlink Gaussian p2p 7ms	**240**	138	617
4	Dynamic and asymmetric PDV -uplink Gaussian p2p 10-7-5ms -uplink Gaussian p2p 2-2-2ms	**240**	343	587
5	Dynamic and asymmetric PDV -uplink Gaussian p2p 7-5-3ms -uplink Gaussian p2p 2-1-2ms	**240**	285	1134
6	Dynamic and asymmetric PDV with variable floor and mean delay	**240**	407	816
7	G.8261 TC12-17, 1 times magnitude	**240**	233	726
8	G.8261 TC12-17, 20 times magnitude	**240**	193	715
9	G.8261 TC12-17, 100 times magnitude	**240**	615	1645

10	Operator A tested PDV data	**240**	4	18

It is evident that the R-DTI client can converge and meet the R-DTI timing and frequency accuracy requirements within 5 minutes for all the test cases. Please note: some of the test cases render very disruptive CIN conditions and could be considered as worst cases. In all the test cases, the initial frequency offset is set to 10ppm, the worst case per DOCSIS specification. When the lock is declared (within 5 minutes), the servo still works to improve the timing and frequency accuracy but with no disruptive time adjustment (phase jump) and caps the frequency adjustment slope to less than 10ppb/s. Figure 10 and Figure 11 show the time and frequency converge process for some of the test cases.

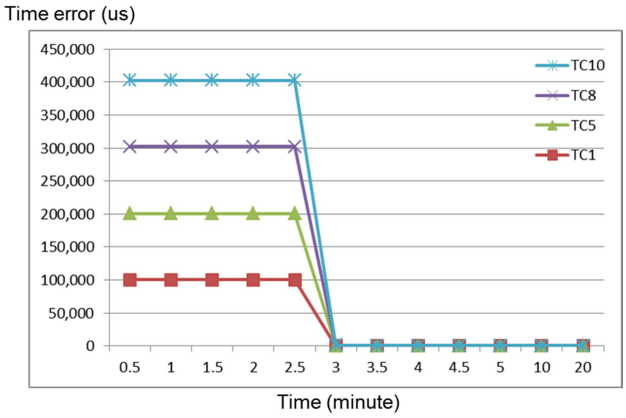


Figure 10. Timing errors with time

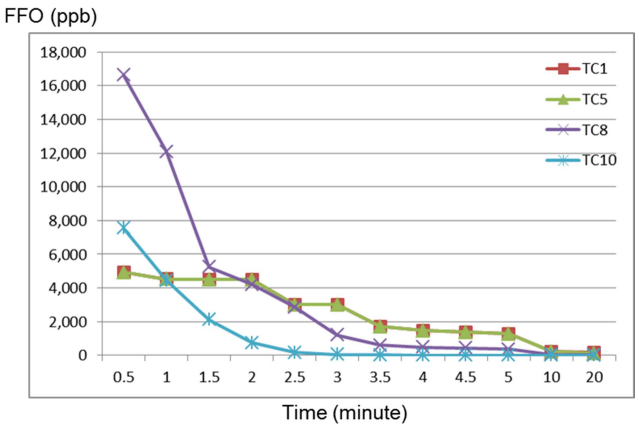


Figure 11. Frequency errors with time

3) Test results in Setup-B

With Setup-B, the impacts of R-DTI on actual R-PHY system operation are evaluated. The same test cases used for Setup-A are used for test Setup B. The evaluation metrics include the following items:

- phase noise of the local clock with the reference clock driven by the recovered clock via R-DTI
- CM initialization
- CM throughput and latency
- Timing holdover during clock outrage/glitch
- Timing and frequency convergence during PTP SW restart (warm-restart/switchover).

a) Phase noise

DOCSIS DS/US has stringent requirements on reference clock's phase noise. To meet the phase noise requirement, an external VCXO is required. This VCXO is then driven by a reference input that comes from the recovered clock via R-DTI.

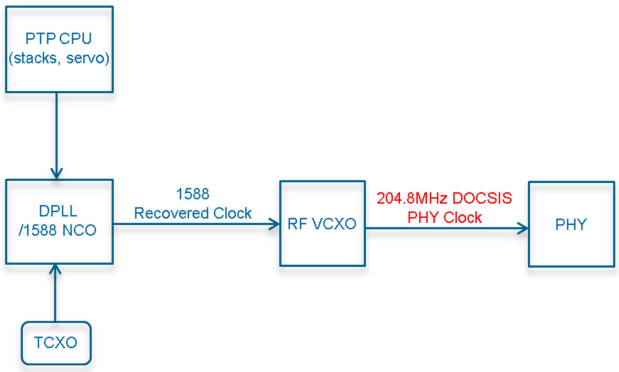


Figure 12. RF reference clock from R-DTI

Figure 12 show the clock architecture. The phase noises are measured for both R-DTI acquisition stage (before lock) and steady stage (after lock), and the measurement results are shown in 13 and 14. Compared to the phase noise performance of the VCXO with a 'good' reference clock, the phase noise performance of the clock driven by the recovered clock via R-DTI has no degradation after the lock is declared; indicating that the R-DTI has little impacts on the phase noise performance of VCXO after the lock is achieved. There are some slight degradations on phase noise in the frequency range below 60Hz in the acquisition stage. This wouldn't be an issue as the system is not expected to start normal operation in the acquisition stage. Moreover, phase noise below 60Hz has little impacts on system performance and is not even specified in DOCSIS PHY specs.

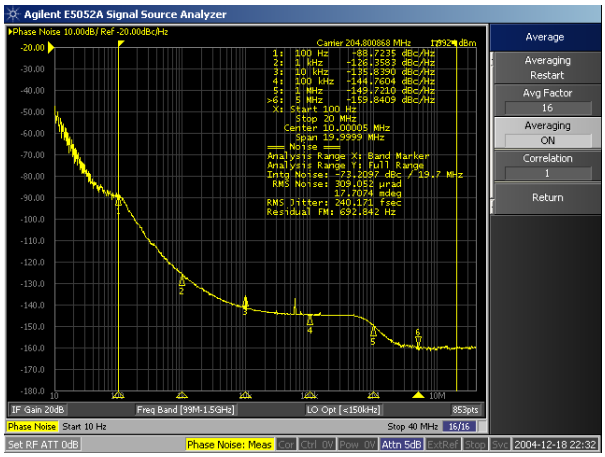


Figure 13. RF VCXO output clock phase noise during 1588 acquisition

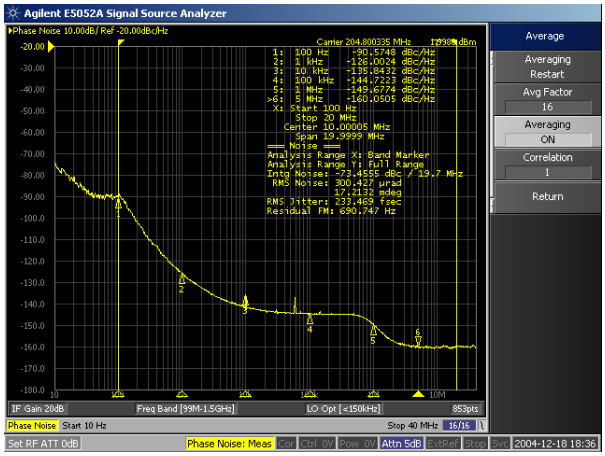


Figure 14. RF VCXO output clock phase noise after 1588 lock declared

It is evident that the phase noise of the RF reference clock is stable and can meet DOCSIS requirements after the 1588 client lock is declared (5 minutes after power up).

b) CM initialization

There are no visible differences observed on CM initialization once the time/frequency lock is declared on RPD.

c) CM throughput and latency

Compared with I-CMTS system, there are no visible differences of the throughput and latency observed with the same test conditions once the time/frequency lock are declared on RPD.

d) Timing holdover

Test the R-DTI client holdover capability during clock outrage/glitch by disrupting 1588 link of RPD on purpose. It is observed that CM can continue work properly for up to 1.5 hours with typical commercial TCXO.

e) Timing and frequency convergence during PTP SW restart (warm-restart/switchover)

There are no visible impacts observed on CM performances during PTP SW restart (warm-restart) or PTP GM switch, the output time/frequency keeps stable without any visible differences on performance.

SUMMARY

Lab tests on R-DTI performances have been completed for various CIN conditions in Cisco R-PHY Lab. The CIN conditions are modeled with different network PDV through the network emulator (Calnex's Paragon-X). Test results show that commercial R-DTI solutions can achieve timing and frequency converge within 5 minutes under all the test conditions and with arbitrary initial phase offsets. After the convergence, the timing and frequency accuracies meet the requirements specified in R-DTI specs and ensure proper DOCSIS operations. Little impacts are observed on RF VCXO clock phase noise once the time/frequency lock is declared (within 5 minutes). No visible differences are observed on CM throughput and latency once the time/frequency is converged.

With a typical TCXO, once the time/frequency are converged, R-DTI has the capability to maintain the required timing and frequency accuracy for up to 1.5 hours without a single 1588 messages exchanges. This capability will ensure the continuous normal DOCSIS operation during 1588 link outrage/glitch (1588 GM switchover, clock source HA switch, etc.) .

REFERENCES

1. John Chapman, DOCSIS Remote PHY, SCTE 2013-08030
2. John Chapman, Remote RPHY Technical Report for Converged DOCSIS, Video and OOB, Technical Report, Oct 16, 2014
3. John Chapman, et al, Remote DOCSIS Timing Interface, CM-SP-R-DTI-D02-141216, Dec 16, 2014
4. IEEE Std 1588 2008, Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control systems, July 2008.

Virtual Business CPE

Kevin Luehrs
CableLabs

Abstract

Evolution of network-based services that has come to be known as cloud computing gave rise to Software Defined Networking (SDN) and Network Function Virtualization (NFV). These technologies are migrating from data centers to data communications networks, resulting an a paradigm shift in structures and management of networks.

SDN and NFV allow for more flexible and agile deployment of networks and network services, and can facilitate innovation of services. CableLabs is applying Lean Startup methods to develop and validate data models and APIs exposing open source network controller functions for business services applications. The APIs enable use of OpenDaylight open source controller platform to configure virtual switches as virtual business CPEs.

VIRTUALIZATION OVERVIEW

Cloud Computing and Virtualization Technologies

The term ‘cloud computing’, and the use of the term ‘cloud’ to refer to a remote server network or the Internet in particular, has been in use less than 10 years but its use has become widespread. Useage of the term cloud computing varies but it typically refers to the provision of services from servers in one or more remote data centers accessed via the Internet or private wide-area network instead of from a local server or network.

Development of cloud computing in the mid-2000s by companies such as Google, Amazon, and Microsoft is important to the computing and networking industries for several reasons. One is cloud computing enabled service providers to resolve or prevent problems such as space limitation, electric power limitation and overheating in local or on-premise facilities caused by continually increasing the number of servers needed to provide services to a growing customer base. Moving demand to servers in remote data centers alleviated demand on the local infrastructure while continuing to provide services, such as e-mail service and file storage service, to customers.

Another important outcome of development and optimization of cloud computing was application and evolution of software defined networking (SDN). Software defined networking as it’s currently known derived from many years of research, specification, and implementation to make network behavior software programmable.

Key characteristics of SDN are separation of the data plane and control plane, and centralization of control plane functions under a single software controller via a well-defined application programming interface (API). The controller communicates to network elements via the API to establish data paths and configure network functions and rules such as packet forwarding and dropping.

SDN was not developed specifically for cloud computing applications but it lends itself well to the environment where programmability of network elements provides flexibility and ease of configuration and management. This is especially useful for dynamic environments like data centers, where demand for service fluctuates over time.

Rapid growth of network traffic driven in large part by the increase in the number of

mobile devices sending and receiving data creates challenges for managing traffic and network resources. SDN provides a means of improving manageability of a network. By providing a centralized point of control in the network, SDN allows a more global view of the resources and traffic. Programmability of control functions acting on network elements through APIs help enable the automation of network engineering and resource allocation. Automation of network engineering is a tool that can help networks scale with increasing traffic demand.

Virtualization in data centers or in a network can be done for hardware or software components of the system. Hardware virtualization refers to initiating virtual machines or self-contained instances of an operating system and software applications on a hardware platform emulating an independent compute, storage or network hardware device.

Virtual machines allow the rapid creation of compute, store or network resources when the workload demands it and the rapid deletion of the resources when they are no longer needed. VMs also allow more efficient use of servers and other hardware devices by adding functions to existing servers without having to find space, power and heat control for another physical server. VMs also enable compute resources to be rapidly cloned in any server location.

Network Function Virtualization (NFV) is another technology enabled by, and potentially leveraged by, cloud computing. NFV involves implementing network functions such as routing, switching, tunneling, packet tagging, Service Level Agreement monitoring, and load balancing to name a few as software functions that can be run independently or as a group on industry standard servers.

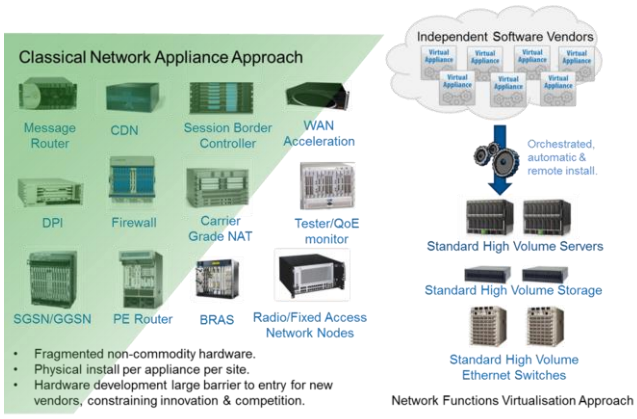


Figure 1 Vision for Network Functions Virtualisation (ETSI NFV ISG – NFV White Paper)

Separating functions provided by networking equipment into modules that can be instantiated when and where they are needed, and deployed on commodity storage, computing or data forwarding platforms enables a high degree of flexibility in configuring a network and allows changes to be made quickly.

Other benefits of virtualizing network functions include increased rate of innovation, opportunity for more centralized control and monitoring, and reduced capital expense.

SDN and NFV allow data centers to evolve away from collections of purpose-built hardware platforms confined to a limited number of racks in a particular facility to virtualization of compute, store, and network functions that can be cloned, chained and distributed into as many commodity hardware platforms as needed. This creates a new paradigm for flexibility in service deployment, ease of management and opportunity for rapid development of services.

SDN Controllers

Cloud computing and networking technologies developed for use in data centers for IT industry are making a transition to telecommunications. Emergence of access technologies such as the cable industry's DOCSIS 3.1 data over cable specification,

Ethernet Passive Optical Network (EPON), and EPON Protocol over Coax (EPoC) brings network speeds to market that make use of SDN and NFV for data communications across service provider networks feasible.

As SDN and NFV technologies are rapidly evolving, so too are open source SDN controller projects. An SDN controller is a software framework that abstracts and exposes command and control of real or virtual network elements to applications through a common set of application programming interfaces (APIs). The controller is the central point of control for the network when network control has been separated from the data plane.

Commercial SDN controllers are available from several technology providers such as Big Switch, Plexxi, Brocade, Hewlett-Packard, Juniper and Cisco. Commercial controllers offer support from the manufacturer and are designed to integrate well with the manufacturer's networking hardware and software, as well as with open source components and controllers.

Open source controllers, like other open source software, are generally developed in a collaborative environment and software developers share code changes with other developers. Code is made available to the developer community for download often under an open source license, and members of the open source developer community can modify the source code and publish the changes back to the community as a fork in the code. Open source development enables rapid development and the opportunity for innovation by drawing on the experience and skills of a relatively large and diverse pool of developer resources. Open source development, including new features, feature enhancements, and bug fixes can happen at a pace dictated only by the availability of developers to work on the code.

Open source controllers include Beacon, Floodlight, POX, Open Contrail, and OpenDaylight. Floodlight and OpenDaylight are both written in Java, both are supported by large communities of developers and both have a large number of functions that can be combined to create a controller environment optimized for a particular network. Most open source controllers support OpenFlow, an open standard protocol for controlling data flow in a network. Unlike the other open source controller options, OpenDaylight provides a Service Abstraction Layer (SAL) that manages plug-in modules for several protocols used to communicate with and control network devices, including Netconf, OVSDB, SNMP, and others in addition to OpenFlow.

Orchestration

As network configurations become complex and span multiple networks and in some cases multiple service providers, a system monitor is needed to maintain information about services and elements and to control and manage systems and groups of systems. In cloud computing this function is performed by the orchestrator.

The orchestration layer abstracts network controller functions from the applications layer. In turn the controller abstracts network elements and connections from the orchestration layer. These relationships are shown in Figure 2.

VIRTUAL BUSINESS CPE PROJECT

Accelerating SDN and NFV Readiness Using Lean Methodology

Established network solutions vendors as well as startups including SDN controller providers are making data center and networking SDN and NFV solutions available now and adoption has begun worldwide. However the rate of deployment is limited by logistics of the change to a new architecture and lack of understanding about SDN and

NFV and their implications. Technology media company TechTarget recently reported results of a survey of IT companies worldwide about their priorities for 2015. Of responding companies 15% reported plans to implement SDN and 27% plan to implement network virtualization. Lack of understanding of SDN and NFV technology was cited as a reason for delaying SDN and NFV deployments.

CableLabs with support from its member cable companies initated a project in 2014 that uses a Lean Startup approach to quickly learn how to apply SDN and NFV to data networking using business services as a set of use cases to focus efforts. Business service providers recognize that SDN and NFV have potential to improve efficiency of their operations and facilitate product innovation and development of new services for their business customers. The project is focused on SDN and NFV on open source implementations and selected OpenDaylight as the open source controller for a reference platform.

The Lean Startup method defined by entrepreneur Eric Ries optimizes capital and human creativity to drive creation of successful output by leveraging lean manufacturing techniques. Lean Startup methodology includes rapid scientific experimentation, early feedback from stakeholders, focus on essential features and rapid adaptation to feedback and changing conditions.

The business services sector of a service provider's business includes a wide range of services including Ethernet services, cellular backhaul, virtual private networks (layer 2 and layer 3), cloud services, security services and more. To make progress on developing SDN and NFV solutions for business services using 'rapid scientific experimentation' a method is proposed to develop a prioritized list of business services use cases then focus on developing standard data models and application programming interfaces (APIs) for the network services for each use case, leveraging available open source controller platforms. The APIs expose functions of the OpenDaylight controller required to implement network services, using virtual or actual network resources for each use case.

Network services APIs can then be shared with the open source community to be standardized for use by business services application vendor companies. In this way the open standard APIs will help enable the vendor community develop business services SDN applications and virtual network functions for service providers, leveraging the open source controller.

By focusing on each use case as a "thin slice" of the communications stack and developing and validating the data model and API for each slice, the challenge is subdivided into manageable segments. By submitting the APIs and data models to the open source community optimizes the APIs for the open source controller platform(s) and makes the APIs available to application developers.

Figure 2 illustrates the thin slice concept. Here the use case shown is MEF Services, which refers to Service Definitions specified by Metro Ethernet Forum such as Ethernet Private Line service. The open source controller is shown as OpenDaylight.

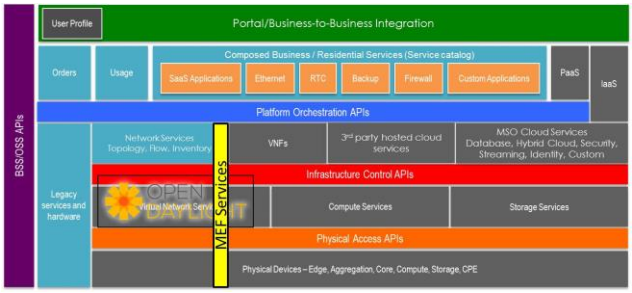


Figure 2 CableLabs CableCloud Architecture with Thin Slice

The network services APIs, working through the open source controller, assist business services applications with

configuring network resources such as virtual switches for necessary connectivity and service level agreement (SLA) enforcement such as burst rate control, throughput limiting, and MTU size enforcement. Much of this configuration can be considered SDN functionality.

Where virtual switch configuration is not sufficient to satisfy business services objectives additional capabilities can be added by appropriate virtual network functions (VNF). For example, many SLAs require enforcement and monitoring of performance parameters such as throughput rate, delay and jitter. A virtual network function may reside in the cloud, in a network element, or divided between the cloud and network elements. Figure 3 illustrates this concept with Fault Management and Performance Management VNFs in the network element as well as FM and PM VNFs in the cloud.

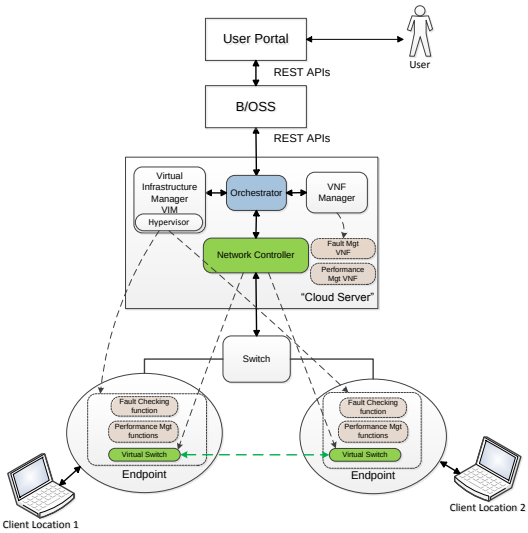


Figure 3 Ethernet Private Line Service Using Performance Management and Fault Management VNFs

Business services deployments use a Network Interface Device (NID) at the customer's premises to terminate the service provider's network and interface the service provider's network to the client's network. Figure 3 shows a generic 'endpoint' network element that represents the role of a NID. The NID performs several functions supporting the delivery of business services including bandwidth management, quality of service, packet tagging, and SLA monitoring such as loopback testing and other performance management functions and detection of faults in the network.

Like other elements of a network, a NID can be implemented using SDN and NFV to realize benefits previously mentioned. Indeed, MEF specifies Virtual NID (vNID) object definitions, and technology providers have begun announcing Virtual NIDs or Virtual Business CPEs but they are as yet immature and not yet fully leveraging the flexibility and automation potential of SDN and NFV.

Open source controller platforms are still evolving. Some Virtual Business CPE or vNID vendors are starting to use some aspects of open source controllers in their products but the potential of the open source controller platform is not yet being realized.

As open standard APIs for business services are made available and optimized for open source platforms and technology suppliers leverage them to develop applications and virtual network functions for business services, the virtual business CPE will also become a key part of the virtualization of business services.

CableLabs' goal is to help establish the cable industry as a leader in open source NFV innovation, specification, and integration. CableLabs has established several projects within its CableCloud initiative to move the state of NFV technology forward. The Virtualized Business CPE project is following the 'thin slice' approach and developed data models and APIs for provisioning of Ethernet Private Line service using OpenDaylight. The project developed a development and demonstration platform illustrated in Figure 3, using Raspberry Pi small form factor computing platforms running virtual switches

to emulate the NID. Using a simple UI to initiate the process, the APIs call OpenDaylight to configure the virtual switches in the Raspberry Pis to establish a connection between them. The platform validates the APIs and proves the concept of creating virtual business CPEs using APIs and OpenDaylight.

The initial use case or 'thin slice' was Ethernet Private Line service. The project will continue by expanding the data models and APIs to include support for additional business services.

CableLabs believes in participating in open source projects to enable and facilitate vendors developing business services solutions for service providers such as the cable operators. This includes submitting work to open source projects for continued development and eventual release to business services application developers.

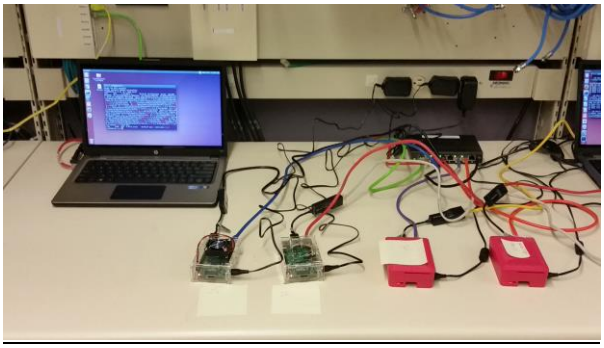


Figure 4 CableLabs Virtual Business CPE Development and Demonstration Platform

REFERENCES

Conversation with Eric Schmidt hosted by Danny Sullivan, Search Engine Strategies Conference. August 9, 2006. http://www.google.com/press/podium/ses2006.html

Network Functions Virtualisation (NFV) – Introductory White Paper. SDN and OpenFlow World Congress. October 22 – 24, 2012. http://portal.etsi.org/NFV/NFV_White_Paper.pdf

Muhammad H. Raza, Shyamala C. Sivakumar, Ali Nafarieh, and Bill Robertson. A Comparison of Software Defined Network (SDN) Implementation Strategies. 2nd International Workshop on Survivable and Robust Optical Networks (IWSRON). In Procedia Computer Science 32 (2014) 1050 – 1055.

OpenDaylight Technical Overview. http://www.opendaylight.org/project/technical-overview

Katherine Finnell. News recap: 2015 global SDN implementations vary. TechTarget. March 2015. http://searchtelecom.techtarget.com/news/2240242047/News-recap-2015-global-SDN-implementations-vary

Eric Ries. (2011) The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. Crown Business. 336 pages.

Metro Ethernet Forum (MEF). (2008). *Technical Specification: MEF 6.1 Ethernet Services Definitions – Phase 2.*

Metro Ethernet Forum (MEF). (2013). *Technical Specification: MEF 7.2 Carrier Ethernet Management Information Model*.

Virtualizing the Home Network

Michael Kloberdans
CableLabs

Abstract

One of the areas of focus in the Cable industry has been a virtualized home network. This paper proposes a road map with an analysis and recommendations, which can assist service providers to develop their technology strategies to meet today's and tomorrow's changing needs. Specifically, this paper will address:

- *Encapsulation*
- *Control Protocols*
- *Multiple Home Routers*
- *Virtual Function Technologies*
- *Operational Concerns*

BACKGROUND

CableLabs has been analyzing how cloud-based network architectures can help service providers enhance their service offerings and provide new features with faster time-to-market and reduced costs. This paper summarizes an analysis directed towards virtualizing the home network.

TRANSPORTING DATA STREAMS

Encapsulation is a means to transport home network information across the access network as shown in Figure 1.

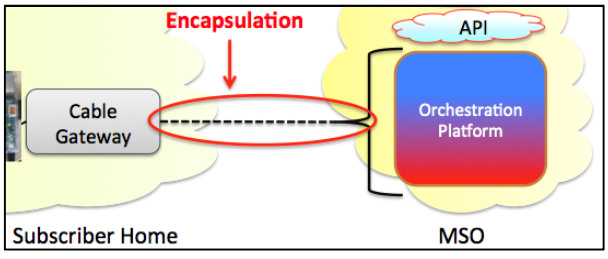


Figure 1: Encapsulation

Encapsulation requires both marking data streams at the home gateway and transporting those streams between the home gateway and the service provider cloud. Virtualization allows one to think in terms of services, and not just devices. We mark packets in order to apply a policy or service to those packets, such as VoIP, Parental Control, etc. A policy is therefore applied to a stream of packets that qualify for a specific type of service from the service provider.

One exciting opportunity offered by virtualization, is to enable new services to be applied to groups of devices in ways that are not practicable using traditional approaches. For example, a subscriber can access an MS service provider 'service portal', and define which data devices in the home should be part of a particular service group. For example, the subscriber may choose to place two tablets, an Xbox gaming unit and a printer to be gathered in the Kid's service group. Those devices would then have a specific firewall policy (e.g. allowing Xbox traffic), a specific parental control policy (e.g. time limits and restricted web access) and perhaps even specific virus protection software that excels at gaming threats. Data stream marking is critical in identifying, not only a subscriber, but also service groups that have multiple devices in each group. It is then possible to apply multiple firewalls and other services to a single residence, again, based on service groups and not just a residence location.

Data stream marking can be done at layer-2 if tunneled using VLANS, or newer approaches such as IEEE 802.1ah (mac in mac). Layer-3 can be used to identify service groups, both IPv4 and IPv6, especially if tunnel technologies are used. While one could use a single tunnel for each service group in a

home, this method doesn't scale well and is not popular with service providers. Leveraging unused or little used IP header fields in the 'inner' IP header of a tunnel is an option for classifying group data flows, however, straying from best practices always has consequences and leads to proprietary protocols, which is not the goal of virtualization.

Tunnel endpoints terminate at a layer-3 device at the edge of the home to an aggregation router in the service provider network. The home device may be an eRouter (e.g. Cable Modem and router in a single box) or it may be the subscriber's router behind the Cable Modem (CM). Another idea for marking group flows is a mapping of a pre-assigned value that is not intrinsically tied to a device and placed in a table in the service provider cloud. One could associate the value 347, as an example, to indicate that the flow requires BW on demand, VoIP support and business strength firewall. A nice benefit of a seemingly arbitrary value is the security feature of not exposing VLAN or MAC information outside the home. Below is a diagram showing all the areas discussed in this paper.

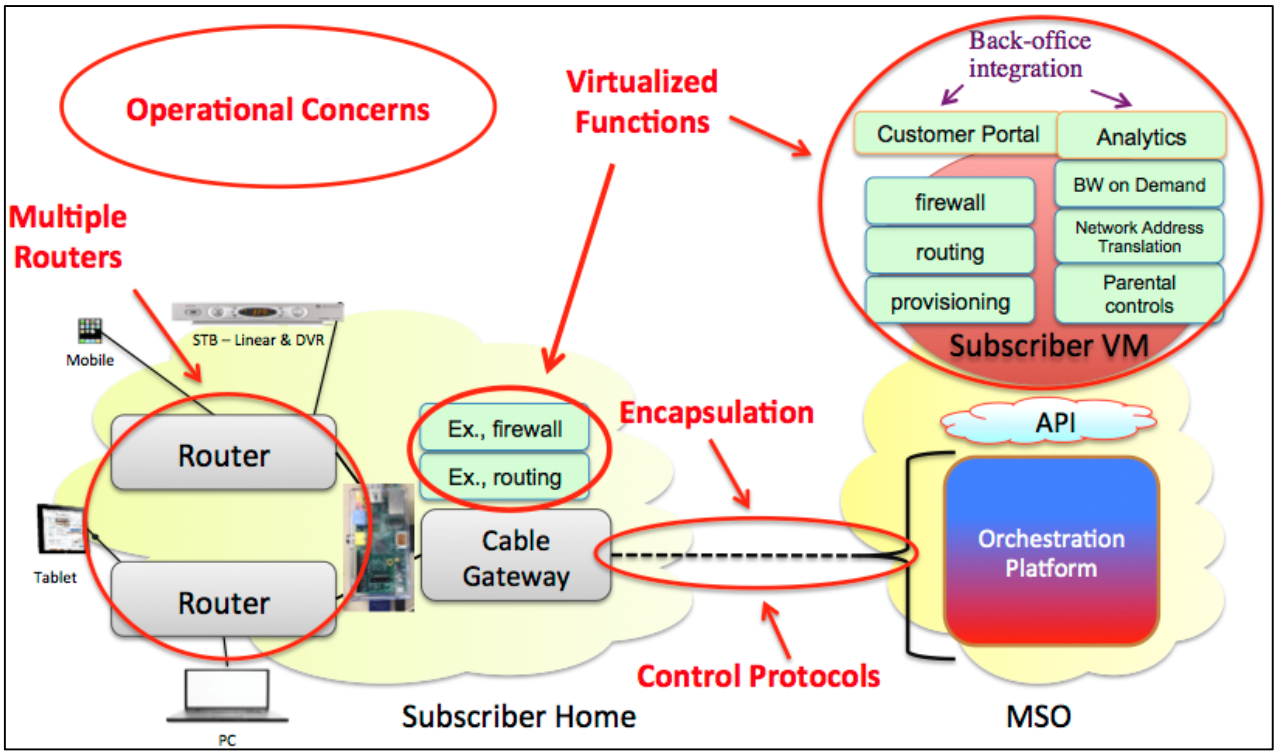


Figure 2: Challenges in the Virtualization Problem Space

Even though they may be considered individually, all areas in the problem space are interrelated. Thus, security, control protocols, latency, etc. are part of the encapsulation analysis. We've discussed marking data flows for Service Function Chaining (SFC), and now need to focus on transport protocols to complete the encapsulation analysis. As mentioned previously, there are various tunneling protocols available to transport data flows between the home gateway and the service provider cloud. Most of these tunnel protocols are well known, and there isn't need to discuss them here, other than to warn of scalability issues. Scalability is the driver to have only one tunnel to a subscriber's home.

A new protocol is also being developed by the IETF called Segment Routing. This protocol

has the advantage of 'steering' traffic from a source to a destination. While there is great promise with this protocol, it is still draft and only supports IPv6, making it unsuitable as a stand-alone solution for service providers.

CONTROL PROTOCOLS

The Cable Gateway must be both provisioned and maintained, which is the function of control protocols. The challenge here is supporting both legacy devices, which might be quite old, while embracing new devices. Because it is unreasonable to expect legacy devices to add support for new protocols, SNMP and TR-069 will continue to be used during the transition to smart devices. TR-069 is the preferred alternative to SNMP to avoid the need to support more than one legacy protocol.

MULTIPLE HOME ROUTERS

A decade ago, most homes had a single router. This device had ports that provided IPv4 addresses (DHCPv4 & NAT), switch ports, and perhaps WiFi service. Today, many homes have more than one router often merely to extend the WiFi range in their home as end devices are increasingly leveraging the ease of connectivity and mobility that WiFi provides. However, instead of buying a WiFi extender, many people have simply bought another router, not understanding the difference. Now, there may be two boxes providing firewalls (which is not optimal, but the system still works), and also NAT. Two NAT sources can often dole out private IPv4 addresses from the duplicated IP pool space of 192.168.1.x. Now there is a real chance of encountering duplicate IP addresses. Even if there is only one IP address given to a device, only one of the routers will recognize that address as being leased, and that router may not be the default gateway. This leads to network connectivity issues in the home and is difficult to troubleshoot, especially remotely.

The story gets more complicated as other routers covertly enter the home, disguised as something else, like a Chromecast or Roku USB stick. These are actually layer-3 devices. And the hardware and software used for home security is often a layer-3 device. We can now see that many 'devices' are actually L3-aware devices. Why is that an issue? In a word, visibility! We lose the visibility of important layer-2 information when a router shields it. A layer-3 device will act as a proxy for protocols such as DHCP and ARP requests. The router will substitute its own MAC address, again as a proxy, for these protocols thereby obfuscating the layer-2 information we need to see in order to troubleshoot problems. Without the visibility of these end devices, Service Providers cannot determine which devices are the cause of (for example) exceeding data caps, or help with troubleshooting when a subscriber calls with a problem. Without visibility, we cannot offer new or enhanced services.

One of the largest gains of virtualizing the home network is to enable techniques that allow ISPs to see the layer-2 connectivity information for devices in the home. This makes it possible to logically extend service groups that span multiple routers as illustrated in Figure 3.

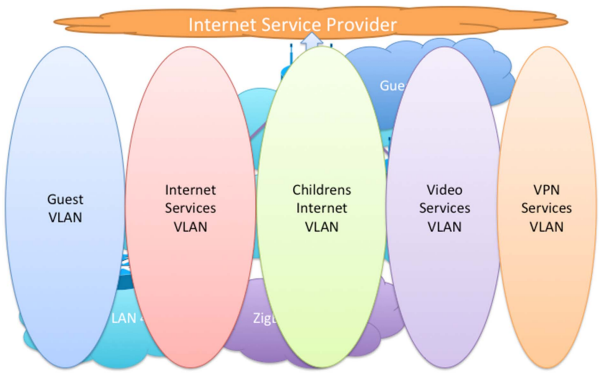


Figure 3: Example Service Groups Spanning Multiple Routers in the Home

Troubleshooting can be greatly enhanced if the service provider has information about

devices connected to the home network. For example, a subscriber calls the service provider help desk to report that a tablet cannot connect to the wireless radio on the router. The help desk technician sees that there are two tablets on the premise, and one is connected. They can know with confidence that the Router's WiFi is working and explain that the problem is most likely with the setup of the second tablet. This can also eliminate sending a service truck to the residence.

Home routers are becoming very sophisticated with new protocols such as IEEE 802.11AC, dual band radios, Service Discovery using UPnP, multiple guest SSIDs, beam forming, NAT implemented in hardware, etc. These modern routers need more features to allow the visibility needed in a multi-router home. Protocols such as GRE can tunnel layer-2 information to the service provider's virtual cloud infrastructure to enable analytics which help drive innovative new services.

Some Service Providers want to virtualize typical router functions, such as DHCP and NAT, and provide that service from the cloud. The benefit is the visibility, at layer-2 of the devices needing these services. Another benefit in providing typical home gateway services in the cloud is the time it takes to update existing, or create new services. If ten thousand home gateways rely on the ISP for DHCP services, for example, the updated DHCP function is replaced in one location in the cloud, eliminating the task of pushing out an upgrade to those thousands of devices. The speed of implementing a 'flash' upgrade is vastly reduced compared to the phased approach of delivering the change across thousands of end devices. Equally important is the time saved in testing. Because the feature is virtualized, there is no need to test the feature on multiple vendors' products, and multiple hardware and firmware versions for each of the vendor's products. Thus, a new or upgraded feature can be implemented in only a few weeks instead of typically 12 – 24 months. This results in OPEX savings, increases reliability and speed to market.

VIRTUALIZED NETWORK FUNCTIONS

A Virtualized Network Function (VNF) is a traditional network node feature that has been implemented in software with well defined, inputs and outputs instead of proprietary hardware. Common examples are NAT, Firewall, Parental Controls and other, newer features such as VoIP, bridge conferencing and WAN Accelerators. Open interfaces are required to enable the VNF module to be used in conjunction with other VNFs and management and orchestration systems originating from multiple vendors. This is the objective of industry bodies such as the ETSI NFV Industry Specification Group.

Because network node functions are now implemented in software, they are no longer restricted to the physical location of a piece of hardware. The previous section described some benefits of placing VNFs in the Service Provider's cloud. However, it is also perfectly acceptable to place the VNFs on the home gateway, for example to reduce Latency.

Latency: optimum placement of the VNF may be important for certain applications. For example, if two gamers were competing in the same residence (an extreme example), latency would be minimized if the home gateway provides the bridging or routing.

If the same function were moved to the ISPs cloud, the path between the gamers would have to extend from one gamer, to the ISPs nearest node of support and back again to the other gamer in the same home. The signal propagation through copper, or even fiber, would add latency to the path, as well as processing the signal at the ISP and again when received by the home gateway.

Operational Impact: There are significant cost and operational implications for the Service Provider to implement virtualization. Specifying and deploying VNFs requires new operational processes as well as investment to purchase the server infrastructure. Whether a network feature is provided on the home gateway or in the cloud, the feature hasn't changed for the subscriber. The subscriber has not perceived any new value even though the home hardware has changed and a sizable investment has been made in the provider's infrastructure. It is true that the subscriber now has access to a portal that lists new and enhanced services, but the basic network functions provide the same experience as the previous solution. The Service Provider reaps the benefits of faster time to market streamlined operations which should improve customer experience.

Fault Protection: If the VNFs associated with the home network reside in the service provider cloud it is important to consider what happens when the WAN link to the home goes down. Even minimal network features would no longer be available. Take for example a neighbor who visits a home that has lost their WAN link to the ISP. The neighbor brings their tablet and wants to print a picture. That simple action is thwarted because the tablet cannot get an IP address for the home network because the DHCP services are hosted in the ISP's cloud. If routing decisions are made in the ISP's cloud, then all communications between service groups (e.g. VLANs) in the home cease when the WAN link goes down. One solution to this issue is to have a latent basic set of network functions that still reside in the home gateway. They become active when triggered by some logic which determines that the WAN link has been down for a reasonable mount of time, and it is time to provide basic services in the home. Solutions are beginning to appear that address this issue.

Another promising approach is to provide a hybrid solution where basic VNFs are placed on the home gateway for fault tolerance and cost reasons, and the more sophisticated VNF functions are located in the service provider's cloud. Not all VNFs are equal in their demands for CPU and/or storage resources. A basic firewall VNF could easily be placed on the home gateway and a superior firewall offer could be service provider cloud-based. The home firewall would supply most of the everyday needs and the ISP advanced firewall could handle the tougher cases where CPU and storage demands exceed that of the home gateway. The author believes this hybrid approach is the future of VNF placement because it makes sense for the subscriber by providing fault tolerance and for the ISP because is saves costs. In this vision the service provider cloud extends into the home network with the gateway having similar capabilities to a cloud-based server albeit with a cost-sensitive approach to its' design.

CONCLUSION

There are significant benefits for service providers to reduce cost and improve customer experience by virtualizing the home network, but there are also challenges. Some of the more important challenges have been examined in this paper and are being addressed by collaborative work being undertaken at CableLabs and in conjunction with the wider industry.

REFERENCES

Chris Donley, SDN & NFV: Moving the Network into the Cloud, *http://www.cablelabs.com/sdn-nfv/*

ETSI : Network Functions Virtualisation http://www.etsi.org/technologies-clusters/technologies/nfv

VIRTUALIZING VIDEO PROCESSING FOR SCALE, AGILITY, AND PERFORMANCE

Glen Griffith, Eric Rosier, Brad Medford
Ericsson, Inc.

Abstract

Video headends are currently built in separate silos for linear broadcast, video on demand (VOD), and multiscreen video services. These service silos, combined with a mix of deployment infrastructures - dedicated hardware, private cloud data centers, and public clouds - as well as a mix of vendors have created a complex and inflexible platform for the processing and distribution of video content. This environment increases the cost and time to market for delivery and monetization of video services.

A virtualized video headend based upon open and well-defined APIs provides a platform for expansion and greater monetization of video services and simplifies both the management and operation of video processing across a complex mix of infrastructures and vendors. This type of video processing platform improves service agility by decreasing the time to market for new services, leveraging performance benefits of different deployment technologies, and providing both operational and capital cost benefits.

This paper describes the scope and functionality of a virtualized video headend, identifying and detailing the key requirements and the relation to pertinent technologies such as cloud, network functions virtualization (NFV), and software defined networking (SDN). A reference model for a virtualized headend is presented and the associated benefits and challenges are and discussed.

INTRODUCTION

The pay TV industry is undergoing a period of dramatic change brought on by a 'digital revolution' in video - video content, now available in digital form can be delivered over an increasingly accessible public internet to a growing array of consumer devices, especially mobile devices. Demand for video delivered to mobile devices is expected to grow by around 45 percent annually through to 2020, when it will account for roughly 55 percent of all mobile data traffic [1].

These changes affect all participants in the pay TV value chain, from content creators to broadcasters to pay TV operators to consumers. All players in this value chain are pressured to introduce new video services of higher quality to more consumer devices in a shorter amount of time and for lower costs. Networks for production and delivery of these TV services have traditionally been designed and built as static facilities, sized for peak expected loads, with a fixed topology. Expansion of these networks requires additional capital expenditure, so upgrades are typically infrequent and incremental in nature.

In recent years, the IT industry has undergone a significant evolution towards software and virtualization of resources. NFV and SDN are being incorporated into service provider networks to leverage virtualization and cloud technologies and to create flexible, adaptable network infrastructures for delivery of services. The principles and technologies of NFV and SDN can be extended to video processing in order to bring the same benefits enjoyed by the IT and communications industries. Video processing virtualization – the separation of the management and video

processing functions - is applicable to a substantial part of the video processing chain and brings similar benefits as the ones demonstrated in the IT domain.

This paper focuses on the virtualization of video processing functionality for participants in the pay TV value chain, describing the functionality and scope of a virtualized headend, explaining the benefits, describing how it might be constructed, and identifying critical issues that must be addressed.

TODAY'S VIDEO PROCESSING ENVIRONMENT

Video headends - the portion of pay TV networks which acquire and process video content to be delivered as video services - have evolved in piecemeal fashion. Video headend systems have traditionally been designed and optimized for delivery of linear broadcast television services - the primary type of video services delivered. These systems have been static facilities designed to perform a specific set of functions and sized for a given amount of video traffic. The growth of VOD services in the 2000s saw the introduction of a new type of video processing infrastructure to support the processing of file-based video content. The systems to support this 'offline' processing were deployed as separate video headends, often from a different set of vendors. The rise of multiscreen video services in recent years has created the need for another type of video processing for delivery of video streams to connected devices such as PCs, smartphones, and tablets. This need was filled by deployment of yet another video headend infrastructure, separately provisioned and managed. The traditional pay TV video headend grew dramatically in scope and complexity from the original linear broadcast processing platforms, consisting now of a broad mix of video processing technologies and vendors and requiring as much IT management expertise as video.

Video processing technology has undergone radical changes over the last several years – compression technologies have evolved to the point that multiple standards have been defined and used for a given function. For example, compression technology went through multiple generations of changes - MPEG2 to AVC/H.264 and standard definition (SD) to high definition (HD) for both - in the last 10 years, with the introduction of HEVC on the horizon. Transition to IP-based transport as an alternative to ASI has also brought about extensive architectural changes in the headends themselves and has become widespread, creating additional complexity in the transport formats to be supported.

Another factor driving complexity is the increasing number of transport technologies that must be supported. While MPEG2 transport stream continues to be the predominant protocol, cable, terrestrial, and satellite networks require several different modulation schemes. This, in combination with regional variations, such as the use of Switched Digital Video in US cable, as well as the use of CBR, VBR and ABR techniques have created a complex matrix of use cases.

Additionally, vendors' systems have traditionally been non-interoperable. This results in managing each vendor's equipment separately. The number of Operations & Management (O&M) systems is not only complex but can negatively impact reliability, increasing the Mean Time to Restore (MTTR) system operation.

For these reasons, current generations of video headends have struggled to gracefully cover all use cases and have typically only supported a subset of those required. This is the case in the vast majority of video processing equipment present in networks today.

As a result of these trends, traditional video headends are a mix of different systems built over time to integrate the different advances in video services. The following diagram illustrates today's video headend environment:

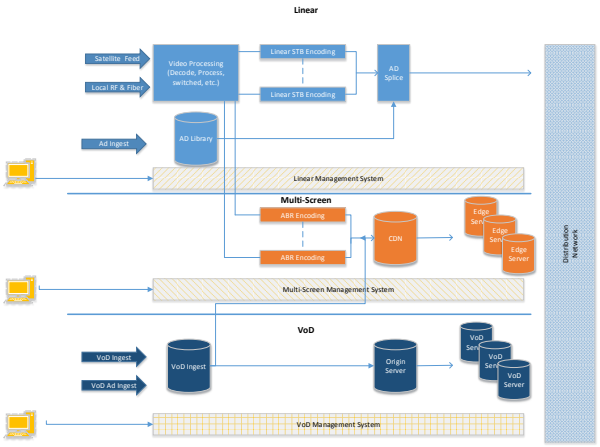


Figure 1: Traditional Video Headend Architecture

The current model for video headends is characterized by multiple silos based on the required services, processing platforms, and delivery networks. Different appliances are needed for different classes of services (linear, video on demand, and multiscreen), as well as for the different deployment technology and transmission networks.

THE VIRTUALIZED VIDEO HEADEND

A virtualized headend is more than just the implementation of video processing functions in software so that these functions can be virtualized and deployed on datacenter infrastructures. The concept of virtualization for video processing is chiefly related to the separation of the management of video processing operations from the implementation of the video processing functions on a mix of deployment platforms - from purpose-built hardware, Commercial Off The Shelf (COTS) hardware, private cloud, and public cloud infrastructures. A virtualized video headend is capable of managing video processing operations across all available deployment platforms so that each platform can be leveraged for its unique advantages to deliver optimum processing performance in terms of quality, cost and efficiency. This approach creates a flexible pool of resources on which video processing operations may be deployed. A virtualized headend is essentially the application of NFV and SDN to video processing.

The following diagram illustrates the logical layers of a virtualized headend.

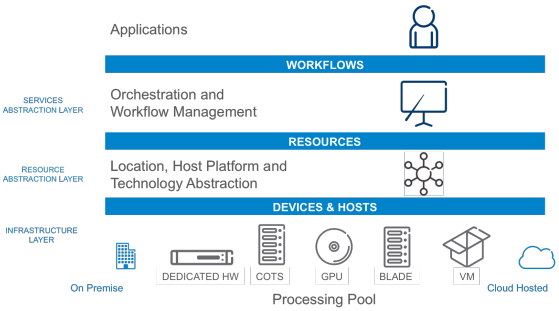


Figure 2: Virtualized Headend Logical Layers

The Infrastructure Layer is where the deployment platforms for the processing components reside. This layer abstracts specifics of the various deployment platforms and publishes its capabilities to the Resource Abstraction Layer. The Infrastructure Layer provides the following:

- Video processing hardware and software
- Video data pipeline - Media Inputs, Media Transformation, Media Outputs
- Low-level Analysis
- Network redundancy ("If this feed is not available, then use another one")

The Resource Abstraction Layer models the processing capabilities, resource requirements, and available resources of the Infrastructure Layer and publishes its capabilities to the Services Abstraction Layer.

This layer provisions and engages appropriate resources to carry out the processing operations directed by the Service Abstraction Layer. The Resource Abstraction Layer provides the following:

- Redundancy of devices and servers
- Resource allocation ("How many of these can I have?")
- License management
- Northbound API for orchestration of processing tasks
- Configuration of system elements, either individual or bulk
- Monitoring of elements ("Is my equipment healthy?")

The Services Abstraction Layer provides an end-to-end service view of video processing and distribution flows from content ingest to delivery, modeling these flows from the perspective of input and output video characteristics and engaging with the Resource Abstraction Layer to effect provisioning and deployment of required processing components on the Infrastructure Layer. The Services Abstraction Layer provides the following capabilities:

- Management interface for defining video flows
- Modeling of video flows based on source and output video characteristics
- Interaction with a Resource Abstraction Layer to effect video processing required by defined video flows
- Workflow management for execution of video processing and distribution rules: **if** this **then** that
- Analytics for management of Service Level Agreements and Key Performance Indicators purposes

Capabilities of a Virtualized Headend

A virtualized headend must meet many requirements in order to optimize video quality and efficiency of delivered video services while providing the operational flexibility of a virtualized environment:

Support for all types of video processing deployment platforms

Video headends consist of a range of deployment platforms. Different types of deployment platforms excel at different types of video processing. Processing of live, broadcast-quality video, for example, requires real-time performance, which is often best supported by purpose-built hardware. Processing of file-based video for a VOD library may be done offline, which lends itself to software processing on COTS or cloud infrastructure where more processing time can be used. Picture quality continues to be the number one rated consumer feature in terms of importance [2], so a virtualized video headend must leverage the advantages of each type of processing platform to deliver the highest quality result for a given resource budget. The following diagram illustrates how a virtualized video headend can harness the strengths of different processing platforms:

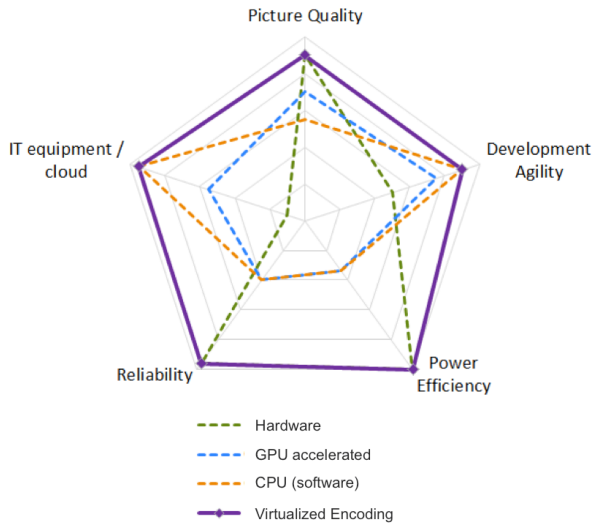


Figure 3: Virtualized Processing Optimization

Support for broadcast video requirements

For a broadcaster or a TV service provider, the video headend is the heart of their operation, providing the video content which is the lifeblood of their business. A virtualized video headend must meet the same rigid requirements as traditional broadcast video headends in the areas of availability, content security, low latency, and video encoding quality.

Converged Management Across Service Types

Converged management across service types is important, because it reduces complexity and simplifies operation of the video headend. Traditional video headends consist of separate infrastructure silos for linear, on-demand, and multiscreen services, each with its own management system and user interface. A virtualized video headend consolidates the management of video processing for all services into a single platform, creating a service-level view of video processing flows.

Separation of video processing workflow management from video processing implementation

This capability is central to increasing service velocity and making the video headend 'future-proof'. Orchestration and management of the video processing workflows must be independent of the processing components to allow changes in processing infrastructure driven by advances in technology and the business/commercial environment.

Support for multiple licensing schemes

A virtualized headend introduces new levels of flexibility in video processing workflows. New licensing models are required to take full advantage of this flexibility. At least four licensing modes must be supported:

1) Pay Per feature: Optional base license for an appliance plus licenses per feature, according to different parameters (SD stream, HD, Video Quality, etc.)
2) 'All you can eat': do what you want with the equipment, within the limits of performance of the hardware
3) Per capacity: Use bandwidth (or physical characteristics of transmission) as a dimensioning factor
4) Per use: Pay when you use the system. This is the common licensing scheme of temporary, cloud-based encoding instances

A virtualized headend will require an abstract model of equipment license and unified license management encompassing and aggregating these different models.

Integration with other network virtualization technologies

Broadcasters and TV service providers are actively deploying virtualization technologies, such as private cloud infrastructure, software-defined networking, and network functions virtualization to provide dynamic platforms for deployment of network services. Leveraging these technologies maximizes their operational efficiency and allows video processing to build on the strong foundation of these virtualization trends.

Ability to optimize video processing flows per a processing 'budget'

A virtualized video headend must model the capabilities and resources of the underlying video processing infrastructure and be capable of engaging those processing resources best suited for the processing operations within a set of constraints or 'budget' for that operation.

Ability to dynamically allocate logical network 'slices' to meet the processing demands of video service flows
A virtualized video headend models video processing flows as a collection of virtual functions which can be chained together to create a network slice for the ingest, processing, and delivery of a video service.

Ability to dynamically scale processing resources with demand
A virtualized video headend creates an elastic pool of resources from the available infrastructure and is able to dynamically scale processing functions up and down based of the demand from video services.

Service-Oriented Architecture
Embracing Service-Oriented Architecture principles - loosely coupled components, each fulfilling its service contract – enables a virtualized video headend to more easily integrate with other systems to extend virtualization and orchestration beyond the video processing domain.

Open and standards-based
A virtualized video headend based on open standards enables more rapid development and allows use of innovative new capabilities from the open software community or third party companies. An open architecture, with published APIs provides a platform to develop new capabilities on the virtualized video headend framework.

Virtualized Headend Reference Model

ETSI have defined a reference model for NFV management and orchestration [3]. The following diagram extends this model to include the media processing domain and provides a reference framework for a Virtualized Headend:

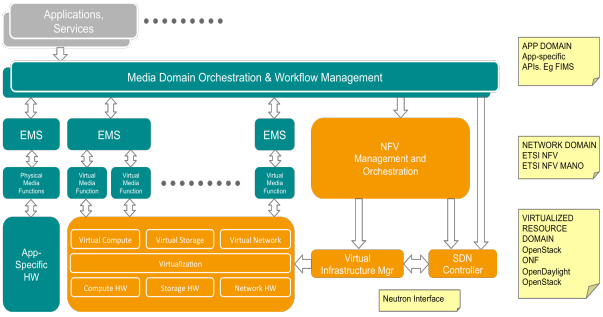


Figure 4: Virtualized Headend Reference Model

Components in green are the components directly involved in providing the processing and management capabilities of the Virtualized Headend; components in orange represent networking and datacenter infrastructure; components in gray are the application and service components which define the video services to be delivered.

Applications and services such as video content management and pay TV middleware define the services to be delivered as well as the parameters and constraints for these services. These service parameters are communicated to the Media Domain Orchestration & Workflow Management layer via a mechanism such as APIs or in-band metadata.

The Media Domain Orchestration & Workflow Management layer distills the service level parameters into required media processing and network functions and orchestrates the instantiation and provisioning of these functions via NFV Management and Orchestration and media Element Management Systems (EMS). Media Domain

Orchestration & Workflow Management may also interact directly with SDN Controllers to request network connectivity for physical functions.

NFV Management and Orchestration configures, coordinates and manages applications, services and underlying virtual and physical infrastructure connected over one or more networks including SDN. Its goal is to ensure the optimum use of virtual resources to meet the required levels of service quality.

NFV Management and Orchestration instantiates required virtual network functions through Virtual Infrastructure Managers and SDN Controllers. VIMs instantiate Virtual Network Functions via a virtualized compute, storage, and network infrastructure.

Media Domain Orchestration & Workflow Management orchestrates provisioning and configuration of the required headend functions on the Virtual Network Function VMs through the EMS components. If physical headend functions are required, Media Domain Orchestration & Workflow Management orchestrates the provisioning and configuration of these functions via the appropriate EMS.

Benefits

A virtualized video headend provides a number of benefits:

Increased Service Velocity
Future success for pay TV operators will depend on the ability to introduce new services and features in a timely manner. Consumers have become accustomed to frequent upgrades from app and web-based service providers and expect the same level of service from their pay TV providers. In a virtualized headend the addition of new services or channels becomes more a matter of software configuration than the installation and interconnection of hardware devices. Furthermore, through orchestration and workflow management the software process can be tested and refined and once deployed, can result in new services being rolled out in significantly less time.

Lower Capital Expense
As the complexity of video networks has grown so has the associated equipment costs. A virtualized video processing environment lowers capital expense by shifting from a traditional appliance model where processing equipment is purchased then amortized over multiple years to a utility model where video processing resources are consumed on demand. The virtualized headend is also able to manage the video processing infrastructure more efficiently, maximizing the processing output of the available resources thereby lowering overall capital costs.

Lower Operational Expense
Traditional video headends require separate management and monitoring systems for different services. The growth in complexity of video processing environments has driven up the operational expense of managing these environments. A virtualized headend is able to decrease this operational expense through converged management of video processing across all delivered services.

Green
A virtualized headend provides an environmental benefit through more efficient use of infrastructure, thereby lowering power consumption and reducing required cooling.

Future-proof
A virtualized headend protects the investment made in it against obsolescence. Separating the management of the video processing from the

implementation allows processing implementations to evolve independently, allowing advances in processing technology to be introduced while maintaining a consistent management layer.

Better overall picture quality
By leveraging the performance benefits of different deployment technologies to exploit the strengths of each, aggregate picture quality across all delivered streams is maximized.

Challenges

Transitioning from today's traditional video processing architectures to a fully virtualized video headend will present substantial challenges for pay TV operators. Some of these challenges are listed below:

Legacy Infrastructure
A flash cut to a completely new video processing platform is not feasible – existing revenue-producing services must be supported; some level of support for legacy processing equipment from a virtualized video processing platform will be required.

Content Security
Video content is amongst the most valuable and expensive data an operator will acquire, and content licensing agreements have very strict security requirements. Content security is critical and must be covered according to the expectation of the content owners. As video processing networks become virtualized and employ more COTS and cloud infrastructure, content security must be maintained. Assuming the video processing network is a private network is no longer sufficient to ensure security. Safeguards against unauthorized access must be in place and a content chain of trust must be maintained for virtualized headends to meet the rigid content security requirements.

System Availability
Linear broadcast video processing infrastructures are the heart of a pay TV operator's system, with commensurate SLAs and uptime requirements. Traditional IT infrastructures are not built to meet these rigid availability and failover requirements. A virtualized headend must meet these stringent requirements across all types of deployment platforms.

Dependence on IT technologies
A virtualized headend places IT technologies in the critical path for all pay TV services – if the IT infrastructure has a fault, it can now affect delivery of video services as well as control and management systems.

Required skill sets
A shift to virtualized video processing represents a significant change in the required skills for workers within the video processing environment. Traditionally, workers involved in managing the video headend have had specialized skills specifically related to the video domain. A virtualized headend will require as much IT and datacenter management expertise as video.

Key Technologies and Standards

Video Processing Capabilities in General Purpose Processors

Video, driven by the widespread adoption of video-enabled applications and increasing video quality expectations, will dominate data traffic going forward. General purpose CPUs are becoming much more efficient at video processing to cope with their changing workload. Mainstream silicon vendors are addressing this demand. Both Intel and AMD have released several iterations of processor

architectures which embed dedicated video encoding and decoding on the processor core. Intel have also released multiple generations of Intel® AVX (Advanced Vector Extensions). AVX is an instruction set extension supported by both Intel and AMD x86 processors which is designed to improve performance of Floating Point-intensive applications, such as image and audio/video processing.

Encoding Software Tools/Middleware

Microprocessor manufacturers have enhanced their CPU and GPU architectures to accelerate video encoding and transcoding. In order to maximize performance of these architectures, video processing middleware toolkits such as Open CL and FFMPEG have been developed to provide efficient access to the hardware.

Open Computing Language (OpenCL) is a framework for writing programs that execute across heterogeneous platforms consisting of central processing units (CPUs), graphics processing units (GPUs), digital signal processors (DSPs), field-programmable gate arrays (FPGAs) and other processors. OpenCL includes application programming interfaces (APIs) to control the platform and execute programs on the compute devices. OpenCL is an open standard maintained by the non-profit technology consortium Khronos Group. It has been adopted by Apple, Intel, Qualcomm, Advanced Micro Devices (AMD), Nvidia, Altera, Samsung, Vivante, Imagination Technologies and ARM Holdings.

FFmpeg is an open-source software project that produces libraries and programs for processing multimedia data. FFmpeg includes an audio/video codec library (including HEVC), an audio/video container mux and demux library, and a command line program for transcoding multimedia files. FFmpeg supports a very broad range of audio and video codecs and file formats. Numerous commercial products and services use FFmpeg, including YouTube, Zencoder, and Handbrake as well as many popular game products.

Mezzanine Format Standardization

The growth in the use of mezzanine formats for distributing content between broadcasters and TV service providers has led to a need for standardization. Broadcasters do not want to produce a version for each pay TV operator, and pay TV operators do not want to receive a different format from each broadcaster.

Several standardization bodies are currently working in this area:

- The Advanced Media Workflow Association (AMWA) has defined a set of specifications that build on MXF to provide a vendor-neutral exchange format for finished programs.
- In February 2014 a Joint Task Force on File Formats and Media Interoperability was announced by its sponsors, the North American Broadcasters Association (NABA), Advanced Media Workflow Association (AMWA), Society of Motion Picture and Television Engineers (SMPTE), International Association of Broadcast Manufacturers (IABM), American Association of Advertising Agencies (4A's), and Association of National Advertisers (ANA). The European Broadcasting Union (EBU) is participating as an observer.
- The Digital Production Partnership (DPP) is an initiative formed by the UK's public service broadcasters. The group has defined a UK-wide standard for finished program delivery, which is a defined subset of AMWA AS-11.

Network Functions Virtualization (NFV)

NFV is an initiative to virtualize the network services that have traditionally been carried out by proprietary, dedicated hardware. The goal of NFV is to decouple network functions from dedicated hardware devices and allow network services that are

now being carried out by routers, firewalls, load balancers and other dedicated hardware devices to be hosted on virtual machines (VMs).

FIMS

In 2010 the Advanced Media Workflow Association (AMWA) and the European Broadcasting Union (EBU) jointly set up a project to develop standards for a framework to implement a media-friendly Service-Oriented Architecture (SOA) with the aim to provide a flexible and cost-effective, reliable, and future-proof solution for production of broadcast content. The resulting project is the Framework for Interoperable Media Services (FIMS). FIMS describes a vendor-neutral common framework for implementing Interoperable Media Services using a SOA-based system, supporting interoperability, interchangeability and reusability of media-specific services. It should allow best of breed content processing products to be integrated with media business systems.

Example Use Cases

The following serve as examples of how a virtualized headend might operate in practice to address video processing challenges:

Introduction of an LTE Broadcast service for an existing video content source

In this scenario, an operator has acquired mobile broadcast rights for a very popular channel. The operator wishes to add this channel to its LTE broadcast service, which requires HEVC encoding. The operator edits the headend configuration for the channel, adding a new LTE broadcast output with desired quality level and processing budget. The virtualized headend provisions an instance of HEVC multiscreen encoding to support the required output profiles and also provisions input and output network flows as well as CDN storage capacity to deliver LTE broadcast service.

Temporary broadcast of a popular sports tournament

In this scenario, an operator wishes to establish a set of channels – available in linear SD and HD as well as multiscreen formats – for broadcast of games from a popular sporting tournament taking place over a 3-month period. The operator enters configuration information - video source, desired output formats, and quality levels, as well as the broadcast schedule - into the virtualized headend management interface. The virtualized headend provisions required processing resources from the pool of available resources, selecting the appropriate processing platforms to provide the optimum result. The virtualized headend also interoperates with available SDN controller infrastructure to provision network capacity for ingress and egress video flows.

Import of a new content provider's VOD movie catalog

In this scenario, an operator has obtained video on demand movie rights for a major studio's entire film catalog, which includes all titles in both SD and HD as well as streaming rights to connected devices. The operator has launched a marketing campaign promoting the new VOD content. Video from the studio will be delivered in high bitrate mezzanine format. All video content must be ingested, processed, and ready for VOD playout by the target launch date. After the operator configures the requirements of the processing operations, the virtualized headend determines the amount of processing and network resources required to successfully encode the source video files into all desired output formats in the time allowed, storing the content in the appropriate locations for playout.

CONCLUSION

Advances in network function virtualization and software-defined networking can be applied to the video processing domain to create a virtualized video headend. A virtualized headend abstracts the management of video processing operations from the infrastructure implementation of these functions. Such a virtualized headend must be capable of leveraging multiple types of video processing deployment platforms ranging from purpose-built hardware to COTS hardware to both private and public cloud infrastructures in order to optimize picture quality per a given processing budget. By abstracting the management of the video processing operations from the implementation, the virtualized headend can dynamically provision and allocate processing resources to create video processing 'network slices' in response to the demands of video applications and services. This optimizes overall video processing performance by leveraging the strengths of each type of processing platform and improves cost efficiency by maximizing the use of the available infrastructure and simplifying operations.

REFERENCES

[1] **Ericsson Mobility Report**
February 2015
http://www.ericsson.com/res/docs/2015/ericsson-mobility-report-feb-2015-interim.pdf

[2] **Ericsson ConsumerLab Report**:
TV and Media 2014
http://www.ericsson.com/res/docs/2014/consumerlab/tv-media-2014-ericsson-consumerlab.pdf

[3] **ETSI GS NFV-MAN 001 V1.1.1 (2014-12)**
Network Functions Virtualisation (NFV); Management and Orchestration

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.