The Big Network Changes Coming with 1+ Gbps Service Environments of the Future

A Technical Paper prepared for SCTE/ISBE by

Tom Cloonan
CTO - Network Solutions
ARRIS
2400 Ogden Ave, Suite 180
Lisle, IL 60532
630-281-3050
tom.cloonan@arris.com

Tushar Mathur
Sr. System Engineer
ARRIS
tushar.mathur@arris.com

Ruth Cloonan
Data Analytics Expert
Blue Opus
ruth.cloonan@gmail.com

Ben Widrevitz
Staff System Engineer
ARRIS
ben.widrevitz@arris.com

John Ulm
Engineering Fellow
ARRIS
john.ulm@arris.com

Table of Contents

Title	Page Number
Introduction	4
Traffic Engineering and 1+ Gbps Services	4
Utilization Levels and 1+ Gbps Services	10
TCP Performance Levels and 1+ Gbps Services	13
Symmetric Services and 1+ Gbps Services	18
RPHY Linear IP Video Delivery and 1+ Gbps Services	20
Conclusion	29
Abbreviations	30
Bibliography & References	31

List of Figures

Title	Page Number
Figure 1 - Downstream Average Bandwidth Trends	5
Figure 2 - Upstream Average Bandwidth Trends	5
Figure 3 - Downstream Billboard Bandwidth Trends	6
Figure 4 - Downstream & Upstream Billboard Bandwidth Trends	6
Figure 5 - Probability Density Function of Bandwidth of a Subscriber with Low Tavg and High Tmax	7
Figure 6 - Probability Density Function of Bandwidth for two Example Subscribers & their Aggregate Bandwidth	8
Figure 7 - Various Techniques for Increasing Utilization Levels in MAC Domains	11
Figure 8 - Impact of Various Techniques for Increasing Utilization Levels in MAC Domains	12
Figure 9 - Bandwidth in Each of the 8 Individual TCP Connections and the Aggregate Bandwidth	17
Figure 10 - Full Duplex DOCSIS (FDX) and Use of the HFC Plant Frequency Spectrum	19
Figure 11- Example of Extended Spectrum DOCSIS Operation	20
Figure 12 - Required # of Transmitted Video Programs vs # of Viewers for Switched IP Video	22
Figure 13 - Linear IP Video Delivery Example using IP Unicasting	24
Figure 14 - Linear IP Video Delivery Example using Always-On, Nailed-Up IP Multicasting	25
Figure 15 - Linear IP Video Delivery Example using Switched IP Multicasting (Non-Realizable)	26
Figure 16 - Linear IP Video Delivery Example using Switched IP Multicasting (Realizable)	27
Figure 17 - Linear IP Video Delivery Example using MAC Processing in the Fiber Node	28
Figure 18 - Linear IP Video Delivery Example using a Combination of Always-On, Nailed-Up IP Multicasting & Switched IP Multicasting	29

List of Tables

Title	Page Number
Table 1 - Average Bandwidth, Standard Deviation, & Coefficient of Variation vs. Service Group Size	9
Table 2 - Required Bandwidth Capacity, SC-QAMs, & OFDM Blocks for Future Service Group Sizes	10
Table 3 - Average Utilization Levels as a Function of Time	10
Table 4 - Maximum TCP Bandwidth (in Mbps) for Various RTT and PER Values	16

Introduction

Whether operating on HFC or PON or 5G infrastructures, future operator networks will undoubtedly be required to deliver service bandwidths in excess of 1 Gbps. Since these higher-SLA services are not like past services, many network attributes and network operational procedures must be changed to accommodate the new bandwidth levels. This paper will explore many of these changes.

Traffic Engineering and 1+ Gbps Services

Traffic engineering is an important area that will definitely be impacted by the arrival of Gbps services, because many of the traditional rules of thumb and formulae that have worked well for years may no longer be valid.

Consider Figures 1, 2, 3, and 4, which illustrate current consumption (average) and billboard (peak) bandwidth trends and extrapolations into the future for several anonymous, sampled Multiple System Operators (MSOs). The following observations can be made:

- Within Figures 1 & 2, Average Bandwidth growth is still rising exponentially in both the Downstream (36% CAGR) and Upstream (17% CAGR), but the growth rate for the Downstream seems to have slowed in the last couple of years
- In addition, these Average Bandwidth growth rates are much lower than the Billboard bandwidth growth rates (defined by the Nielson Law) shown in Figures 3 & 4
- Even if the Billboard bandwidth growth rates drop to lower rates (as predicted by several operators and illustrated with the green line of Figure 3), it is still expected that the Billboard bandwidths will grow at a much higher rate than the Average Bandwidths

Thus, the distance between maximum bandwidth levels (Tmax) and Average Bandwidth levels (Tavg) will continue to increase, creating an interesting scenario in which Tmax for a single subscriber will be much higher than Tavg for that subscriber.

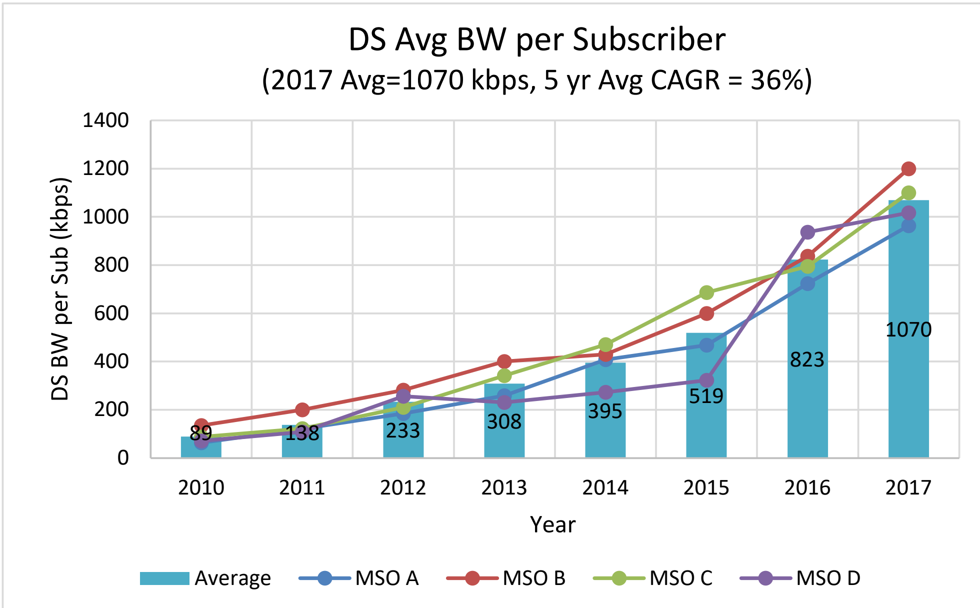


Figure 1 - Downstream Average Bandwidth Trends

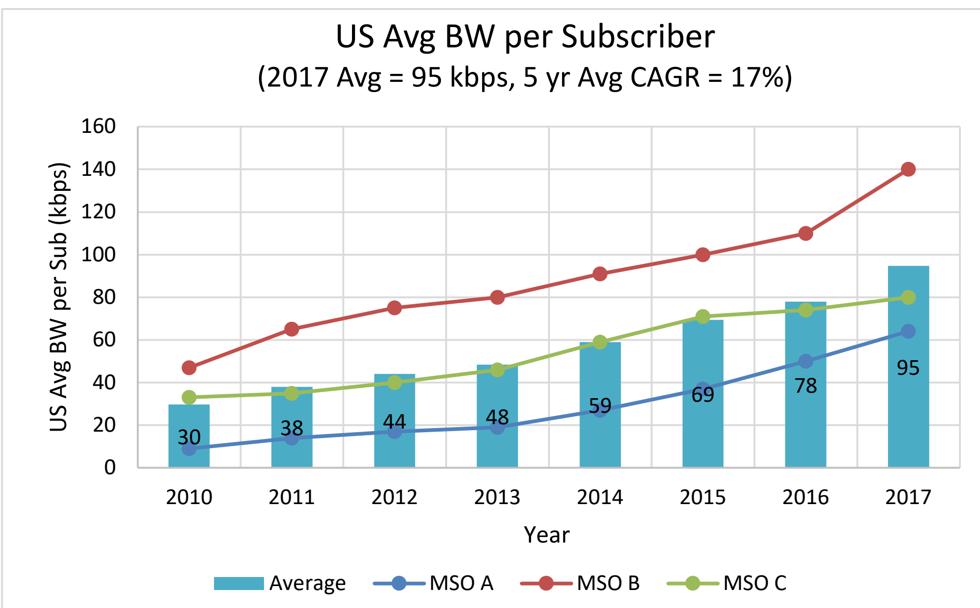
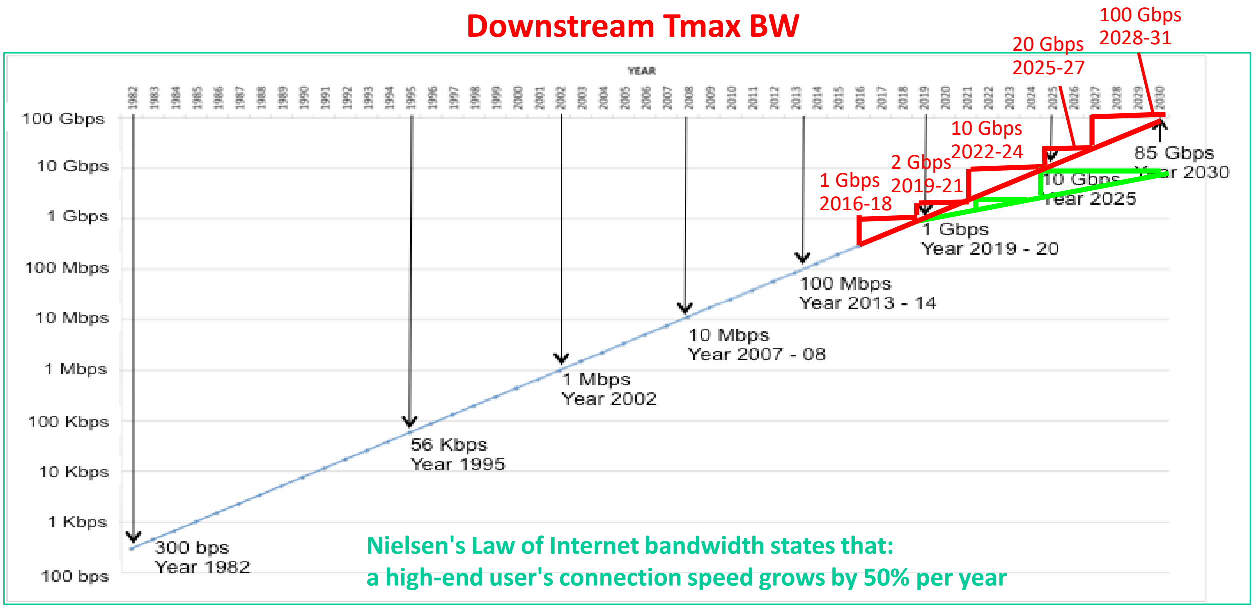


Figure 2 - Upstream Average Bandwidth Trends



Source: http://www.nngroup.com/articles/law-of-bandwidth/

Figure 3 - Downstream Billboard Bandwidth Trends

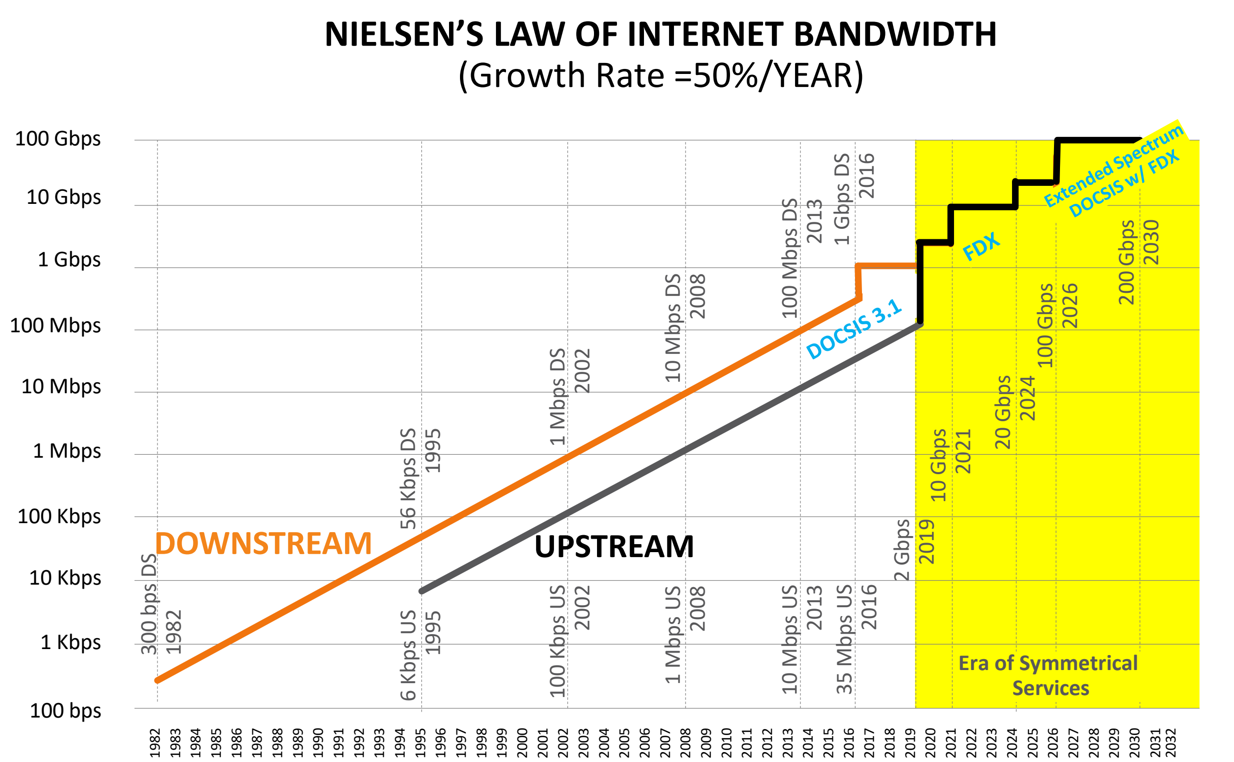


Figure 4 - Downstream & Upstream Billboard Bandwidth Trends

With much higher Tmax levels (exceeding 1 Gbps) and with Tavg levels that are relatively lower (relative to Tmax), it can be shown that the typical subscriber's transient bursts to maximum bandwidth levels are likely to occur for shorter windows of time and are also likely to occur much less frequently. To illustrate this point, let us consider the following contrived example. Assume that there exists a (somewhat strange) subscriber who only receives downstream traffic at one of two discrete traffic rates. That subscriber either receives bandwidth at a rate of 1 Mbps or 1 Gbps, and they never receive bandwidth at any other traffic rate. Assume also that we know that the subscriber has an Average Bandwidth given by Tavg = 2 Mbps.

Armed with this simple information about the subscriber, it is interesting to note that we can calculate the probability (the fraction of time) that the subscriber transmits at 1 Mbps and the probability (the fraction of time) that the subscriber transmits at 1 Gbps. This results from the fact that we have two equations in two unknowns if we draw the subscriber's transmission rates on a probability density function (pdf) diagram, as shown in Figure 5. For a pdf, we require that $\int pdf(x)dx = 1$. For a pdf with an average of Tmax = 2 Mbps, we also require that $\int x\, pdf(x)\, dx = Tavg$. We can solve these two equations to determine that the probability $P(1\ Mbps\ transmissions) = 0.999$. We can also solve these two equations to determine that the probability $P(1\ Gbps\ transmissions) = 0.001$. Thus, with low Tavg values and high Tmax values, it is clear that the probability of high bandwidth transmissions will be required to be quite small.

Figure 5 - Probability Density Function of Bandwidth of a Subscriber with Low Tavg and High Tmax

Thus, it should be clear that when subscribers do burst to their Tmax levels, the extremely high bandwidth levels are likely to perform most data transfers in a fairly short timeframe, so they will be on and off the network quite quickly. For example, a typical web page sized at 4 Mbytes = 32 Mbps would download at a 20 Mbps rate within 1.6 seconds, however that same 32 Mbit-sized web page would download at a 1 Gbps rate within 32 milliseconds.

As a result of this effect, it is also likely that the probability of having multiple maximum bandwidth bursts occurring simultaneously for many different subscribers becomes quite low as Tmax values rise and the probability $P(Tmax)$ values drop. For example, if a Tmax burst event for subscriber #1 has a low probability P_1, and if a Tmax burst event for subscriber #2 has a low probability P_2, then assuming the two burst events are independent, the probability of the two Tmax burst events occurring simultaneously has the even lower probability of $P_1 \times P_2$.

Consider the probability density functions shown in Figure 6 for two of our (contrived) example subscribers and for the aggregated bandwidth generated by those two subscribers when they share bandwidth capacity within a Service Group (SG). The aggregated bandwidth can be 2 Mbps (when both are receiving at 1 Mbps), 1001 Mbps (when one is receiving at 1 Mbps and the other is receiving at 1000 Mbps), or 2000 Mbps (when both are receiving at 1000 Mbps). The probabilities for each of these probabilistic events is also shown within the aggregated bandwidth probability density function, and it can be seen that the probabilities $P(2000\ Mbps) \ll P(1001\ Mbps) \ll\ P(2\ Mbps)$.

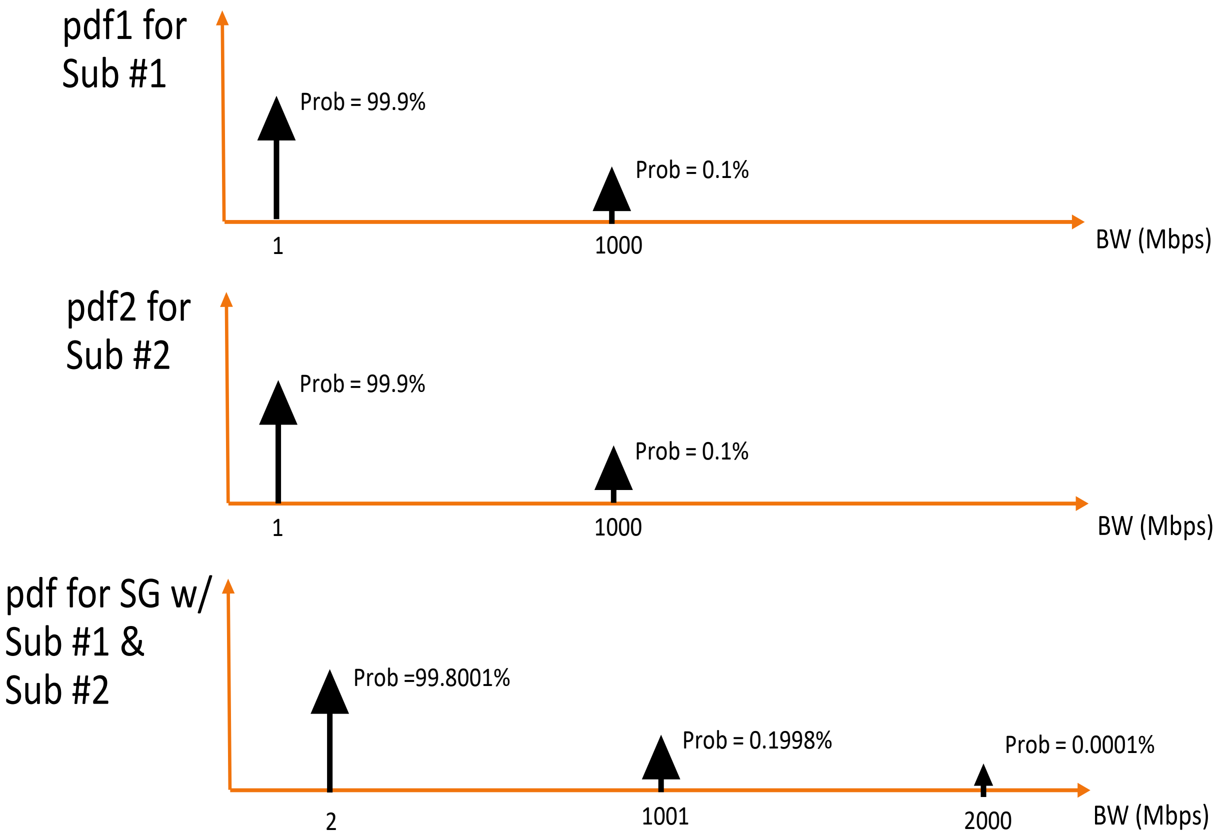


Figure 6 - Probability Density Function of Bandwidth for two Example Subscribers & their Aggregate Bandwidth

Carrying this idea even further, if we have Nsub subscribers within a Service Group and if each subscriber $'i'$ exhibits an Average Bandwidth level of Tavg and a maximum bandwidth level of Tmax with a burst probability given by P_i, then the Average Bandwidth within the Service Group will be given by Nsub*Tavg. However, the probability of ever seeing a Service Group bandwidth approaching Nsub*Tmax becomes very small, especially as Nsub is increased (since that probability is given by $P_1 {\times} P_2 {\times} \ldots {\times} P_{Nsub}$). This fact can also be characterized by calculating the Service Group's Average Bandwidth, Standard Deviation, and Coefficient of Variation ($CV\ =\ Standard\ Deviation/$ $Average\ Bandwidth$) for Service Groups of different sizes using data collected from the field today (where Tmax_max = 100 Mbps). For illustration, we also show the bandwidth associated with $(Tavg\ for\ SG) + (3{\times}\ Std.Dev.for\ SG)$. These results are shown in Table 1.

Table 1 - Average Bandwidth, Standard Deviation, & Coefficient of Variation vs. Service Group Size

Nsub for SG	Tavg for SG (Mbps)	Std. Dev for SG (Mbps)	Coef. of Var. for SG	(Tavg for SG)+(3*Std Dev for SG) (Mbps)
256	284.16	64.47	0.226879223	477.57
1024	1150.98	128.93	0.112017974	1537.766
4096	4460.54	258.00	0.057840479	5234.544

It can be seen that the Coefficient of Variation tends to be reduced for larger Service Group sizes, implying that the relative spread of bandwidths within the Service Group becomes smaller (relative to the average Service Group bandwidth). In particular, it appears that the Tavg value for the Service Group traffic grows linearly with Nsub, whereas the Standard Deviation for the Service Group traffic grows proportionally to the square root of Nsub (approximately). This is an important point, because it implies that much of the equipment within the headend (ex: CCAP Cores, Routers, Switches, etc.) that supports larger numbers of subscribers will be able to operate quite well even if they only support slightly more bandwidth capacity than $Nsub \times Tavg$. And, they will not typically be required to support anything close to a bandwidth capacity of $Nsub \times Tmax$.

Traffic Engineering studies are currently under way to create better models to determine the actual required bandwidth capacities, and those results will be presented in the future. Due to the complex nature of the new 1+ Gbps services, these studies will likely need to be performed using a "bottom's up" approach that accounts for the contributions and statistics for both Average Bandwidth levels and burst bandwidth levels for each of the subscribers within the network. Early results of this work led to the realization that the "ARRIS QoE-based Traffic Engineering Formula" developed in [CLO1] is still quite valid for small service group sizes with $Nsub \leq 400$- even in the era of 1+ Gbps services. That formula is given by:

$$Required\ Bandwidth\ Capacity \geq Nsub \times Tavg + 1.2 \times Tmax\_\mathrm{max} \qquad (1)$$

where Tmax_max is the largest of the Tmax values. Larger Service Group sizes (with $Nsub > 400$) will require new formulae that are being researched. However, using Equation (1), we can find the required bandwidth capacity and DOCSIS 3.0 SC-QAM channel counts DOCSIS 3.1 OFDMA channel counts that typically-sized Service Groups of the future with typical bandwidths might require. It can be seen in Table 2 that DOCSIS 3.1 OFDMA blocks will definitely help provide the required bandwidth capacity, because the use of lower-efficiency SC-QAMs would consume a large portion of the Hybrid Fiber Coaxial (HFC) network spectrum.

Table 2 - Required Bandwidth Capacity, SC-QAMs, & OFDM Blocks for Future Service Group Sizes

Nsub for SG	Tavg for sub (Mbps)	Tmax_max for sub (Mbps)	Required BW for SG (Mbps)	Required # SC-QAM Ch's @ 36 Mbps ea.	Required # OFDM Blocks @ 8 bps/Hz
50	10	1000	1700	47	1.1
100	10	1000	2200	61	1.4
200	10	1000	3200	89	2.1
400	10	1000	5200	144	3.4

Utilization Levels and 1+ Gbps Services

The Required Bandwidth Capacity formula described in Equation (1) is a useful tool for calculating the approximate bandwidth levels required within a Service Group. A related equation can be derived if one considers the general meaning of the terms contained within Equation (1). In particular, the first term of Equation (1) is the Average Bandwidth passing into the Service Group. The second term represents the amount of headroom bandwidth required to support temporary bandwidth bursts in excess of the Average Bandwidth that may occur within the Service Group. The sum of the two terms is the total Required Bandwidth Capacity.

As a result, it should be clear that one can obtain the Average Utilization Level within a Service Group by dividing the first term by the sum of the first and second terms. In equation form, we find that:

$$Average\ Utilization\ Level = (Nsub \times Tavg)/(Nsub \times Tavg + 1.2 \times Tmax\_max\ ) \qquad (2)$$

We can create a table of “typical” values for the future to get a feel for how Average Utilization Levels will vary as a function of time. If we assume that Nsub drops as a function of time and if we assume that both Tmax and Tavg grow by a CAGR of 50%, then we can create the following table showing how Average Utilization Levels may change with time.

Table 3 - Average Utilization Levels as a Function of Time

Year	Nsub for SG	Tavg for sub (Mbps)	Tmax_max for sub (Mbps)	Avg BW for SG (Mbps)	Required BW Capacity for SG (Mbps)	Average Utilization Level
2017	400	1.00	1000	400	1600	25.0%
2019	200	2.25	2250	450	3150	14.3%
2021	100	5.06	5063	506	6581	7.7%
2023	50	11.39	11391	570	14238	4.0%

Table 3 illustrates several interesting trends that will likely occur as operators move into the future. First, it is quite apparent that the Average Bandwidth within the Service Groups of the future will likely grow, but expected Node Splits of the future will likely keep this Average Bandwidth growth to a relatively slow rate.

Required Bandwidth Capacity within the Service Groups of the future will also grow, but rapidly increasing values of Tmax will cause this growth rate to be quite fast.

With a slow growth rate in Average Bandwidths and a fast growth rate in Required Bandwidth Capacities, the Average Utilization Level is expected to drop quite rapidly as operators move into the era of 1+ Gbps services. In particular, for the example shown in Table 3, the Average Utilization Levels may drop from 25% in 2017 to only 4% in 2023.

This low Utilization Level may be viewed negatively by those who are responsible for paying the costs of the ever-increasing Required Bandwidth Capacities, because it appears that the extra Bandwidth Capacity is mostly being used to provide headroom for infrequent bandwidth bursts that exceed the Average Bandwidth levels within the Service Group.

If this low Utilization Level is deemed to be undesirable, then there are several network design techniques that can be utilized by the operator to reduce the severity of this low Utilization problem. These design techniques include:

- Using multiple Remote PHY or Remote MACPHY devices per Service Group (which effectively increases the Nsub value within the Service Group, which increases the Average Utilization level)
- Using MAC-based Nodes with sub-tending Remote PHY nodes (which also increases the Nsub value within the Service Group, which increases the Average Utilization level)
- Using Selective Subscriber Migration (which moves the highest Tmax subscribers off of the HFC plant and gives them Fiber To The Home optical feeds. The resulting subscribers have a much lower Tmax_max value, which increases the Average Utilization level).

These three techniques are clearly illustrated within Figure 7.

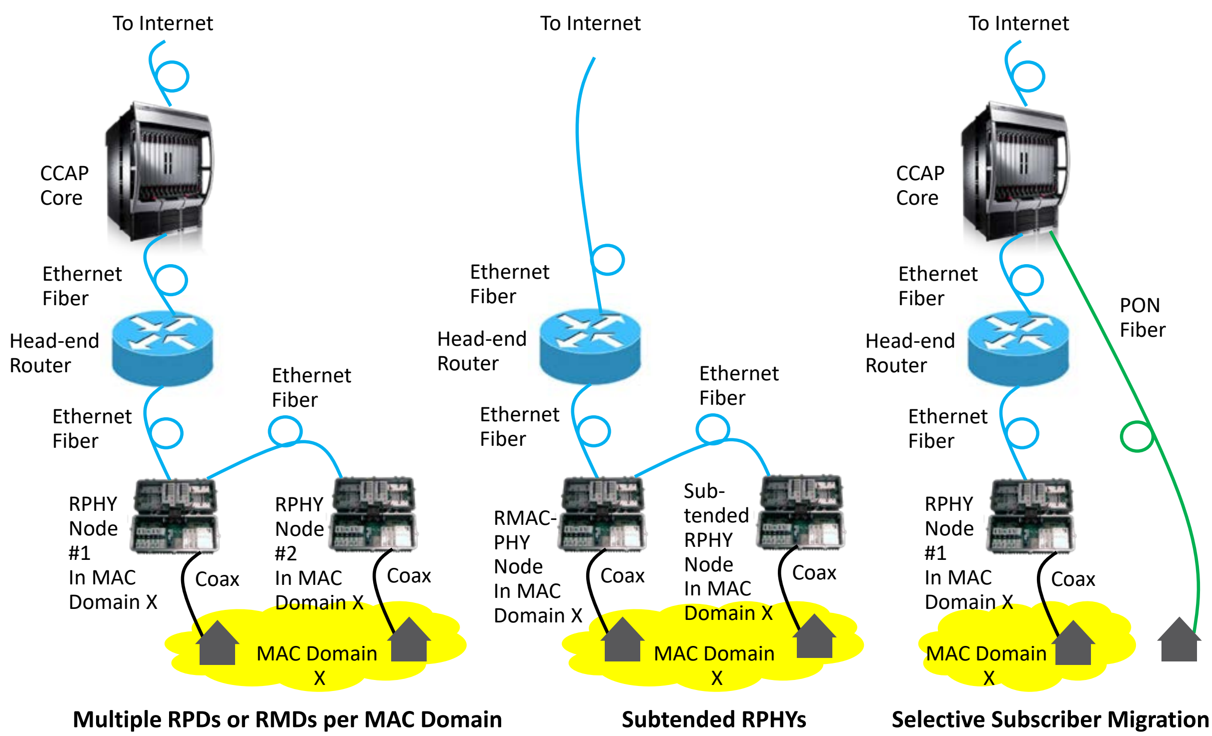


Figure 7 - Various Techniques for Increasing Utilization Levels in MAC Domains

Figure 8 illustrates (using bandwidth candlestick diagrams) how these various techniques help increase the Utilization Levels with example numbers. Within each candlestick diagram, the orange color indicates the magnitude of the first term ($Nsub \times Tavg$) within the ARRIS QoE-based Traffic Engineering Formula of Equation (1), which is the static bandwidth. The gray color indicates the magnitude of the second term ($1.2 \times Tmax\_max$) within the ARRIS QoE-based Traffic Engineering Formula of Equation (1), which is the QoE headroom bandwidth. As stated above, the Average Utilization Level is given by Equation (2). Within the candlestick diagrams, this is given by the height of the orange color divided by the height of the orange + gray color. In all three scenarios, the changes increase the Average Utilization Levels by about a factor of two. It should be apparent that combinations of these changes (ex: Multiple RPDs per MAC Domain plus Selective Subscriber Migration) can produce even higher Average Utilization Levels.

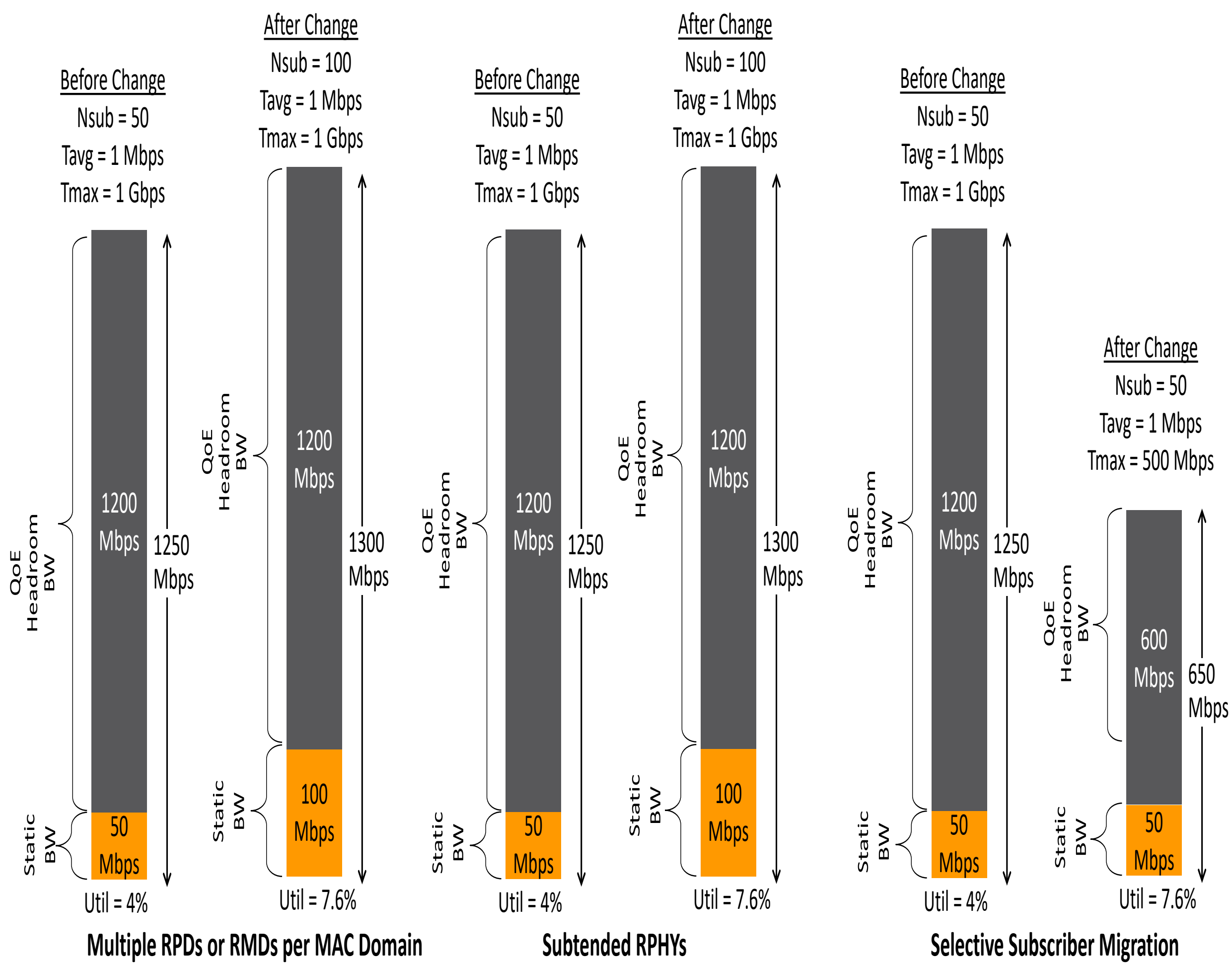


Figure 8 - Impact of Various Techniques for Increasing Utilization Levels in MAC Domains

TCP Performance Levels and 1+ Gbps Services

Even with very high bandwidth capacities available within the links of the network, the performance of 1+ Gbps services can still be limited by the operation of the Transmission Control Protocol (TCP). TCP is the underlying protocol used to manage the flow of data for most of the data transmissions across the Internet. It includes complex flow control and congestion control algorithms that were put in place to ensure that multiple users can interact and share the Internet bandwidth capacity in a fair and equitable fashion.

Many different flavors of TCP have been proposed and utilized over the years. Each one has a focus on a particular set of performance attributes within high-speed data connections. One of the more popular TCP algorithms in use within many Internet servers today is TCP CUBIC.

The CUBIC algorithm was developed in the 2005 time-frame, and it was optimized to perform well with high-bandwidth connections in long, large-latency networks. The congestion control algorithm within CUBIC was designed to make changes to connection bandwidth levels (as a result of detected packet loss or sudden packet delays) in a much less aggressive fashion than many of the congestion control algorithms that preceded it, such as TCP New Reno, TCP BIC, etc. Generally, TCP's throughput is dictated by packet loss and Round Trip Time (RTT), but CUBIC manages the throughput by only the packet loss. Another important change introduced with CUBIC was the use of timers instead of TCP Acknowledgements to trigger increases in connection bandwidth levels (or congestion window), which resulted in it performing well for both low-latency and long-latency networks.

As the name of the algorithm suggests, the congestion window function of CUBIC is a cubic function:

$$W_{cubic} = C \times (t - K)^3 + W_{max}$$

where C is a window scaling factor, *t* is the elapsed time from last window reduction, $W_{max}$is the window size just before the last window reduction, and $K = \sqrt[3]{W_{max}\beta / C}$, where β is a constant multiplicative decrease factor applied for window reduction at time of the packet loss.

β and *C* are the control knobs in this formula. The two values are also exposed to the end users for TCP CUBIC configuration. An optimal value of the knob variables is described as β = 0.8 and C = 0.4, by the inventors of CUBIC [RHE1].

As can be seen in the congestion window growth formula, CUBIC is independent of RTT and only depends on the packet loss times, unlike many of its predecessor congestion algorithms. Because of these beneficial changes, CUBIC was selected to be used in Linux kernels quite a few years ago. Since many servers within the Internet are based on Linux kernels, this implies that CUBIC is responsible for managing data flows for many of the client-server sessions operating over the Internet. As a result, the authors have opted to focus on CUBIC's TCP implementation within this section.

Two questions related to CUBIC's impact on 1+ Gbps data flows were studied by the authors. These included the impact of typical network delays as well as the impact of typical HFC packet error rates. Each of these is discussed below.

Network Delays during Connection Slow-Start: Once a TCP connection is up and running, network delays are not as problematic for CUBIC as they were for other TCP implementations, because of CUBIC's novel use of timers instead of TCP Acknowledgements.

However, CUBIC still uses Acknowledgements to pace its growth of the Congestion Window and bandwidth levels during the "Slow-Start" operation that occurs when a TCP connection is first established. This early-stage use of Acknowledgements may have an impact on how quickly a TCP connection can achieve its desired 1+ Gbps data rates.

It can be shown that for any TCP connection, the maximum TCP throughput level at any instant in time is given by:

$$Maximum\ TCP\ Throughput = \frac{min(Congestion\ Window\ Size, Receive\ Window\ Size)}{RTT} \quad (3)$$

where RTT is the Round-Trip Time for data transmissions, the Congestion Window Size (cwnd) is an internal state variable maintained by TCP at the source of the transmissions, and that Receive Window Size (rwnd) is an internal state variable maintained by TCP at the destination of the transmissions.

Under normal circumstances, the Receive Window Size is usually quite large, so the Maximum TCP Throughput is limited by $Congestion\ Window\ Size/RTT$. In essence, the TCP transmitter cannot send more than a Congestion Window's worth of packets until an ACK arrives for the first packet that was sent. TCP tends to increase the size of the Congestion Window when packets are flowing smoothly (without any packet drops or sudden packet delays), and that increase permits the TCP connection to transmit data at increasing data rates. Longer RTT's tend to reduce the data rates, and shorter RTT's tend to increase the data rates.

During Slow-Start, the Congestion window is increased by one Maximum Segment Size (MSS) every time an ACK arrives. It can thus be noted that the number of Acknowledgements that come back every RTT is double the number that came back during the previous RTT window (assuming no packet loss). If we assume that the MSS is 1500 bytes (12000 bits), then we can develop a formula indicating the amount of time '*T*' that it takes during Slow-Start to increase the size of the Congestion Window to a level that permits a particular desired bandwidth level. This formula is given by Equation (4):

$$Desired\ Bandwidth = \left(2^{\left(\frac{T}{RTT}\right)}\right) \times \frac{12000\ bits}{RTT} \quad (4)$$

RTT can vary a lot depending on the geographical locations of the clients and servers, but a "typical" RTT value might be between 20 msec and 100 msec. Using these typical RTT values, we can plug in the Desired Bandwidth of 1 Gbps and calculate the amount of time *T* that would be required for the Congestion Window to be increased (during the Slow-Start period) to a value that would permit 1 Gbps service to flow. When RTT = 20 msec, the total time *T* for the Slow-Start Congestion Window to grow to support 1 Gbps is given by 214 msec. When RTT = 100 msec, the total time *T* for the Slow-Start Congestion Window to grow to support 1 Gbps is given by 1.3 sec.

Thus, it should be clear that for short 20 msec RTTs, the Congestion Window will likely grow to permit 1 Gbps rates in a fairly short window of time (214 msec). However, for longer 100 msec RTTs, the Congestion Window will take a relatively lengthy 1.3 seconds of time to grow to permit 1 Gbps rates, and this lengthy time period may not be ideal for some applications. As a result, future TCP congestion control algorithms may need to be modified to eliminate this potential problem associated with slow bandwidth growth rates to 1+ Gbps service levels.

Packet Error Rates: Even if there is adequate bandwidth capacity to support high 1+ Gbps TCP connection rates within a network, there is another TCP effect that can cause the actual bandwidth experienced by users to be much lower than the bandwidth capacity. This effect is a result of the actions taken by the congestion avoidance algorithm of TCP.

The congestion avoidance algorithm responds to dropped packets (detected through unacknowledged TCP packets) by assuming that the dropped packets resulted from overloaded and congested buffers within intermediate network elements. This assumption leads the congestion avoidance algorithm to reduce the Congestion Window at the source, which in turn reduces the actual bandwidth of the connection (since bandwidth is roughly given by $Congestion\ Window/RTT$). These types of algorithms help ensure that the Internet remains operational during periods of congestion.

However, in a lossy network, the packet drops may also be caused by noise that corrupts the packets. Thus, noisy networks will experience packet drops, and those packet drops will also cause the TCP congestion control algorithms to throttle the bandwidth of the connection. As a result, it is important to determine the maximum bandwidth levels that might be permitted by a TCP connection that might be propagating over a noisy HFC plant or over noisy cables within a home.

To perform this analysis, the authors again focused on the operation of the CUBIC TCP algorithm. Simulations were performed using a CUBIC model in the NS2 simulator [WEI1], and various RTTs were simulated and various packet error rates were injected into the simulation to model a lossy HFC plant or lossy home network. In the end, the actual TCP connection performance was monitored for each of the RTT values packet error rate values.

The results are illustrated in Table 4, where the maximum TCP connection bandwidth (in Mbps) is displayed for different combinations of RTT (along the top) and Packet Error Ratio (PER) along the left side.

Table 4 - Maximum TCP Bandwidth (in Mbps) for Various RTT and PER Values

DS PER	RTT → / DS BER (w/ 1500 byte pkts)	10ms	20ms	40ms	60ms	80ms	100ms
10^{-6}	8×10^{-11}	1420.95	984.78	788.51	701.19	650.51	606.49
10^{-5}	8×10^{-10}	337.85	214.12	154.67	123.93	101.69	89.49
10^{-4}	8×10^{-9}	116.26	62.86	38.74	29.21	25.52	22.66
10^{-3}	8×10^{-8}	33.33	18.22	10.29	6.95	5.23	4.28
10^{-2}	8×10^{-7}	9.37	5.17	2.87	1.97	1.55	1.29

Within this table, it can be seen that only one of the grid elements (RTT = 10 msec, DS PER = 10^{-6}) actually supports 1+ Gbps service levels. All of the other grid elements are limited to actual TCP bandwidth levels that are lower than 1 Gbps. This illustrates how limiting TCP congestion control algorithms can be in the presence of noise-induced packet errors. The results imply that very low packet error rates and short RTTs must be maintained on the network to ensure high 1+ Gbps TCP throughput levels on any single TCP connection. This can be quite challenging because:

1) Reducing packet error rates to this level may require upgrades to the plant to reduce plant noise.
2) RTT of 10 msec or less can only be provided by close servers, so TCP connectivity to distant servers would not allow the 1+ Gbps service levels on the TCP connection. Most data would have to be cached in caching servers very close to the subscribers.

There will undoubtedly be pressure on MSOs and content providers to provide improvements like those described above as the world heads toward 1+ Gbps services. However, there are also likely to be a few other changes that will take place in the future. For example, the above results were obtained using CUBIC's TCP algorithm. New TCP algorithms are always being architected and designed to accommodate the new requirements as the Internet evolves, so perhaps a new variant of TCP will be developed that is more amenable to packet errors and longer RTTs and which will permit 1+ Gbps bandwidth levels even in the presence of these challenging environments.

In addition, there is another factor that must be taken into account. The above simulations assumed that the 1+ Gbps service level was going to be offered over a single TCP connection and that the single TCP connection was required to support the entire 1+ Gbps bandwidth. Fortunately, many applications in the world today do not operate in this fashion, and instead opt to transmit data between servers and a single client using more than one TCP connection. For example, Peer-to-Peer applications such as BitTorrent tend to use many TCP connections to deliver their content. In fact, even traditional web pages are usually delivered using multiple TCP connections to deliver the desired content.

The use of multiple TCP connections to deliver content provides a major benefit to the bandwidth levels for the application. In particular, it can be shown that an application which uses N parallel TCP connections to deliver content to a client will roughly result in an N times speed-up in the bandwidth levels associated with that application (when compared to the bandwidth levels that would have been achieved with a single TCP connection).

This fact was verified in the simulator, where an application was run with 8 parallel TCP connections. The particular network conditions were selected to be PER = 10^{-4} and RTT = 10 msec. With a single TCP connection, the maximum bandwidth was found (in Table 3) to be 116.26 Mbps. However, with 8 parallel TCP connections, the total bandwidth was increased to be 917.7 Mbps (which is close to 8 times the bandwidth of the single TCP connection). The actual bandwidth found in each of the 8 individual TCP connections is indicated in Figure 9. The benefits illustrated in Figure 9 may lead more and more applications to take advantage of parallel TCP connections in the future of 1+ Gbps service levels.

FLOW ID	Average Throughput (Mbps)
1	117.61
2	124.09
3	115.02
4	112.56
5	104.58
6	114.21
7	111.31
8	118.28
Aggregate BW	**917.70**

Figure 9 - Bandwidth in Each of the 8 Individual TCP Connections and the Aggregate Bandwidth

Symmetric Services and 1+ Gbps Services

Operators are continually being presented with new challenges. Over time, there has been more and more marketing pressure for operators to provide so-called "Symmetric Services." In a Symmetric Services environment, the Upstream Tmax value and the Downstream Tmax value must be similar in value. Ideally, the Upstream and Downstream Tmax values are equal. This type of service offering has become more popular as PON service providers have begun to offer it in recent years.

Providing a Symmetric Services offering has traditionally been a challenge for MSOs, because their traditional HFC plant was originally architected using Frequency-Division Duplex techniques to be asymmetrical, offering much more bandwidth capacity in the Downstream direction (using the 54-1002 MHz range of the spectrum) than in the Upstream direction (which was limited to using only the 5-42 MHz range of the spectrum in North America). Prior to the arrival of DOCSIS 3.1, this yielded a theoretical Downstream bandwidth capacity of ~5.6 Gbps and a theoretical Upstream bandwidth capacity of ~144 Mbps. As can be seen, this is quite asymmetrical.

The arrival of DOCSIS 3.1 has improved the situation to some extent. DOCSIS 3.1 operators can (for example) choose to use a High-Split HFC network with the Upstream spectrum contained within a 5-204 MHz range and with the Downstream spectrum contained within a 258-1218 MHz range. Assuming 10 bps/Hz useable spectral efficiencies during operation, this can yield a theoretical Downstream bandwidth capacity of ~9.6 Gbps and a theoretical Upstream bandwidth capacity of ~2 Gbps. While these are much higher bandwidth capacity levels, they are still asymmetrical.

An asymmetrical HFC network environment of this nature is therefore challenged when asked to provide Symmetrical Services, and the challenge becomes even more difficult in a 1+ Gbps service world where operators may be called upon to provide a 1 Gbps Downstream × 1 Gbps Upstream service offering. In the future, these numbers may grow to be 2 Gbps Downstream × 2 Gbps Upstream, or 5 Gbps Downstream × 5 Gbps Upstream, or even 10 Gbps Downstream × 10 Gbps Upstream. It should be clear that these higher levels of Symmetrical Service could be a challenge for today's asymmetrical HFC network.

A potential solution to this problem is currently being developed. The resulting solution employs Full Duplex DOCSIS (FDX) capabilities. While the details of this solution are beyond the scope of this paper, the basic idea is to permit the Downstream spectrum and the Upstream spectrum to overlap and utilize the same portion of the spectrum at the same time for both Downstream and Upstream transmissions, as shown in Figure 10. This form of operation will require the use of complex technologies such as Echo Cancellation to permit the simultaneous transmissions of signals in both directions.

Consider a system in which the Downstream spectrum is permitted to operate from 104-1218 MHz and the Upstream spectrum is permitted to operate from 5-684 MHz. Assuming 10 bps/Hz useable spectral efficiencies during operation, this can yield a theoretical Downstream bandwidth capacity of ~11.1 Gbps and a theoretical Upstream bandwidth capacity of ~6.7 Gbps. These much higher bandwidth capacity levels can begin to permit Symmetrical Services- even for a system requiring a 5 Gbps Downstream × 5 Gbps Upstream service offering.

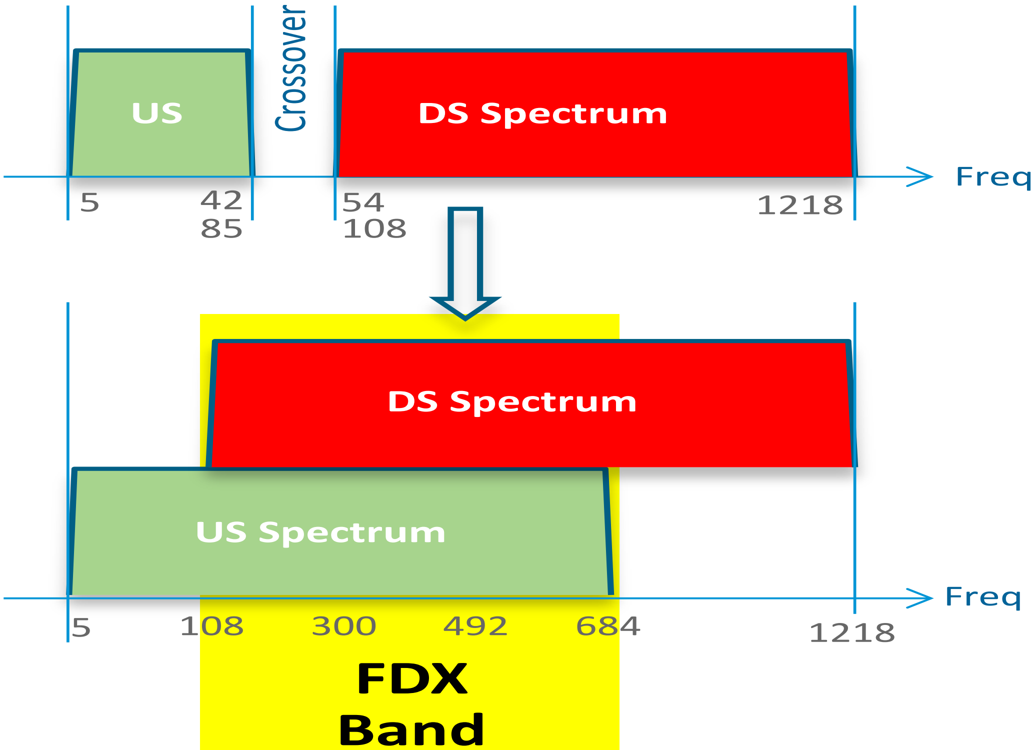


Figure 10 - Full Duplex DOCSIS (FDX) and Use of the HFC Plant Frequency Spectrum

Higher bandwidth Symmetrical Service offerings may require the use of other solutions that are currently being studied. For example, a 20 Gbps Downstream × 20 Gbps Upstream service offering might require the use of Extended Spectrum DOCSIS techniques, which permit DOCSIS 3.1 and FDX operations to extend beyond 1218 MHz (and even beyond the optional 1794 MHz limit mentioned within the DOCSIS 3.1 specification). These types of Extended Spectrum DOCSIS solutions are illustrated in Figure 11, and they may become practical in the 2020 decade. [CLO2]

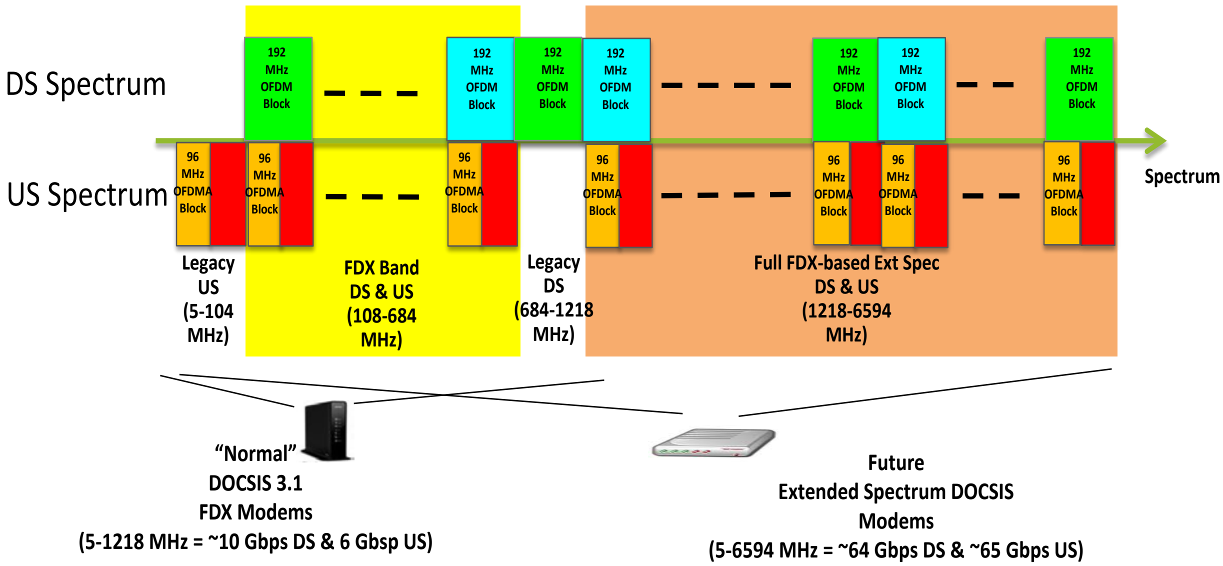


Figure 11- Example of Extended Spectrum DOCSIS Operation

In general, new concepts are continually being developed that should permit Symmetrical Services — even in the 1+ Gbps service environments of the future.

RPHY Linear IP Video Delivery and 1+ Gbps Services

The transition to IP Video is a popular transition that MSOs have been planning for years. Most operators foresee many benefits resulting from this transition, but they have been waiting for the correct timing (technological improvements, bandwidth availability, etc.) to introduce these changes.

IP Video over DOCSIS can yield many benefits, including:

- Convergence of all services over a single DOCSIS infrastructure platform
- The increased number of programs that can be offered due to the Statistical Multiplexing Gains that result from DOCSIS 3.0 Channel Bonding (which are not available when using MPEG video transport)
- The increased number of programs that can be offered due to the higher spectral efficiencies of DOCSIS 3.1 OFDM (which are not available when using MPEG video transport)
- The channel efficiencies that result from Switched Digital Video-like operation provided by the dynamic Service Flows and load-balancing of Service Flows and the use of IP Multicast streams within the CCAPs

There are two general types of IP Video Services that can be offered to IP Video subscribers, and each of these two types can be implemented in at least two different ways:

- Video on Demand (VoD) IP Video Service

a. Unicast VoD IP Video Service — using IP Unicast, whereby Unicast DEPI tunnels can be utilized between the CCAP Core and the Remote PHY Nodes in a Remote PHY environment

b. Multicasted VoD IP Video Service — using IP Multicast- which might be valuable if the VoD Service Group needs to span multiple Remote PHY Nodes in a Remote PHY environment, whereby Multicast DEPI tunnels can be utilized between the CCAP Core and Remote PHY Nodes

- Linear IP Video Service (Linear)

a. Nailed-Up, Always-On Linear IP Video Service — Nailed-Up Linear using IP Multicast, whereby Multicast DEPI tunnels can be utilized to send the feed to multiple Remote PHY Nodes in a Remote PHY environment

b. Switched Linear IP Video Service — Switched Linear using IP Multicast, whereby Multicast DEPI tunnels can be utilized to send the feed to multiple Remote PHY Nodes in a Remote PHY environment

Thus, it can be seen that IP Multicast and Multicast DEPI tunnels (for Remote PHY environments) may be called into service for at least three out of the four IP Video environments described above. It is therefore instructive to consider some of the side effects of using RPHY Multicast DEPI tunnels for IP Video delivery between the CCAP Core and the Remote PHY Node as the network moves towards the higher DOCSIS 3.1 channel bandwidths associated with the 1+ Gbps services of the future.

According to the Remote PHY Specification [RPH1], the Multicast DEPI support is “meant for the replication of an entire QAM or OFDM channel to multiple RPDs.” Each QAM or OFDM channel is associated with a single unique pseudo-wire within the Multicast DEPI tunnel. It should thus be clear that the QAM channel or OFDM channel associated with a particular pseudo-wire MUST carry exactly the same information (i.e.- video programs) to all of the destination RPHY Node endpoints of the Multicast DEPI pseudo-wire.

If there are ANY differences in any of the signals (i.e. video programs) on a QAM or OFDM channel at any of the RPHY Nodes being fed by a particular Multicast DEPI pseudo-wire, then that RPHY Node receiving the different signals cannot receive the same pseudo-wire as the other RPHY Nodes, and a separate pseudo-wire must be set up for the different QAM or OFDM channel feeds to that particular RPHY Node. As will be shown below, this is an important point that may drive many architectural decisions.

In past studies on Service Group bandwidth requirements, it has been shown that the decision on whether it is optimal to transmit a particular set of video program using IP Unicast or Switched IP Multicast or Always-on IP Multicast is a function of several factors. These factors include the popularity of the programs (defined by the alpha value (α) for the assumed Power Law Distribution, where higher alpha values imply a smaller number of very popular programs), the number of viewers, and the number of available programs in the content library.

Note: The Power Law Distribution can be defined as follows. For a list of programs numbered $i = 1 \; to \; i = N$, the probability that a particular subscriber selects a particular program number (i) is given by:

$$P(i) = \frac{i^{-\alpha}}{\sum_1^N i^{-\alpha}}$$

where the summation is taken from program $i = 1$ to program $i = N$. For the Power Law popularity curve, α = 0.7 is a typical value that will be used in this example.

An example scenario for a system with various numbers of available programs in the content library is shown in Figure 12. We will focus on the 250 program case with blocking probabilities of 0.01% (identified by the bottom blue plot within Figure 12).

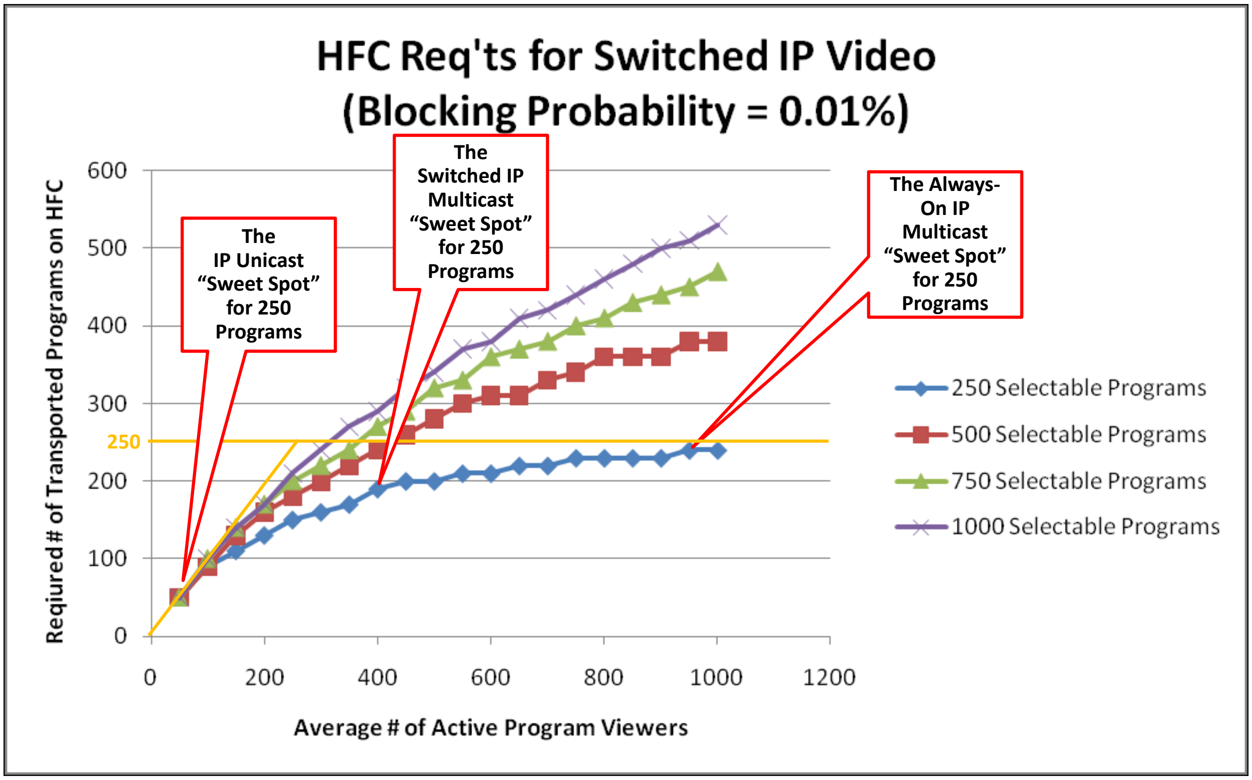


Figure 12 - Required # of Transmitted Video Programs vs # of Viewers for Switched IP Video

Within this figure, it can be seen that for a small number of Active Viewers (< ~100), the MSO might as well utilize IP Unicast, because there are little savings from using IP Multicast. If the MSO's IP Multicast solution is robust and low-cost, then there is no harm in using IP Multicast in this range. Within this figure, it can also be seen that for a large number of Active Viewers (> ~900), the MSO might as well utilize Always-On IP Unicast, because there are little savings from using Switched IP Multicast. If the MSO's IP Multicast solution is robust and low-cost, then there is no harm in using IP Multicast in this range In the middle region of the Figure (where the number of Active Viewers is between 100 and 900), there is clearly a region where Switched IP Multicast can yield improvements in the bandwidth required by the Service Group over both unicast and Always-On IP Multicast. Thus, when considering the Bandwidth savings within the Service Group, the use of Switched IP Multicast can always yield some improvements.

The above analysis is only focusing on the problem from the point-of-view of the Service Group bandwidth requirements - i.e. it attempts to answer the simple question, "How much spectrum is required to carry the Linear video signals on the RF coaxial connection that runs between the Fiber Nodes and the subscriber homes?"

There is another topic and point-of-view that probably should also be analyzed within an RPHY environment, and that is the topic of the CCAP Core processing and input/output requirements. This alternative analysis would attempt to answer the question, "How much processing power and input/output bandwidth is required on the CCAP Core in the headend or hub to support the Linear video signals?" In actuality, both of these topics (Service Group bandwidth and CCAP Core bandwidth) need to be considered together.

To illustrate the point, let us consider this particular topic area of Remote PHY systems in more detail by using a simple example. Assume that an operator has converted to a Fiber Deep Remote PHY environment, and the resulting headend contains 500 Remote PHY Fiber Nodes with 50 subscriber homes per Remote PHY Fiber Node. This results in a total of 25,000 subscriber homes attached to the headend. Assume also that the operator has 250 Linear video programs within their content library and assume that each video program consumes an average of 9 Mbps of bandwidth in the DOCSIS pipe. Note the 9 Mbps average might result from a blend of Standard Definition and High Definition and 4K content. Obviously, the total Service Group bandwidth required to continuously transmit 250 Linear video programs at 9 Mbps each would be 2250 Mbps (2.25 Gbps). However, there are many ways to architect a system that delivers this 2.25 Gbps stream of Linear bandwidth from the CCAP Core to each of the 500 Remote PHY Fiber Nodes. Techniques include:

- Using "VoD-like" IP Unicasting of a single Linear video program stream from the CCAP Core to each of the active subscribers viewing a Linear video program. This technique is used today by many MSOs to deliver Linear video to many second screen devices (PCs, smart-phones, and tablets) over DOCSIS. There are 25,000 subscriber homes in the example head-end, so if each subscriber home contains ~2.5 people, that results in 62,500 people being provided with service.

  If 60% of them (i.e. 37,500 people) are actively watching Linear video content in a particular busy-hour window of time, that implies that there would be (62,500) × (60%)=37,500 IP Unicasted Linear video programs (consuming (37,500) × (9 Mbps) = 337.5 Gbps of aggregate bandwidth) sent from the CCAP Core to the group of Remote PHY Fiber Nodes. Each of the 500 Fiber Nodes will (on average) receive 1/500th of this aggregate bandwidth, which results in ~675 Mbps of Linear video bandwidth consumed by the subscriber homes in each Service Group. (See Figure 13).

  It should be clear that the packets for the Linear programs can be multiplexed in with regular High-Speed Data packets on a channel set going to a particular Fiber Node.

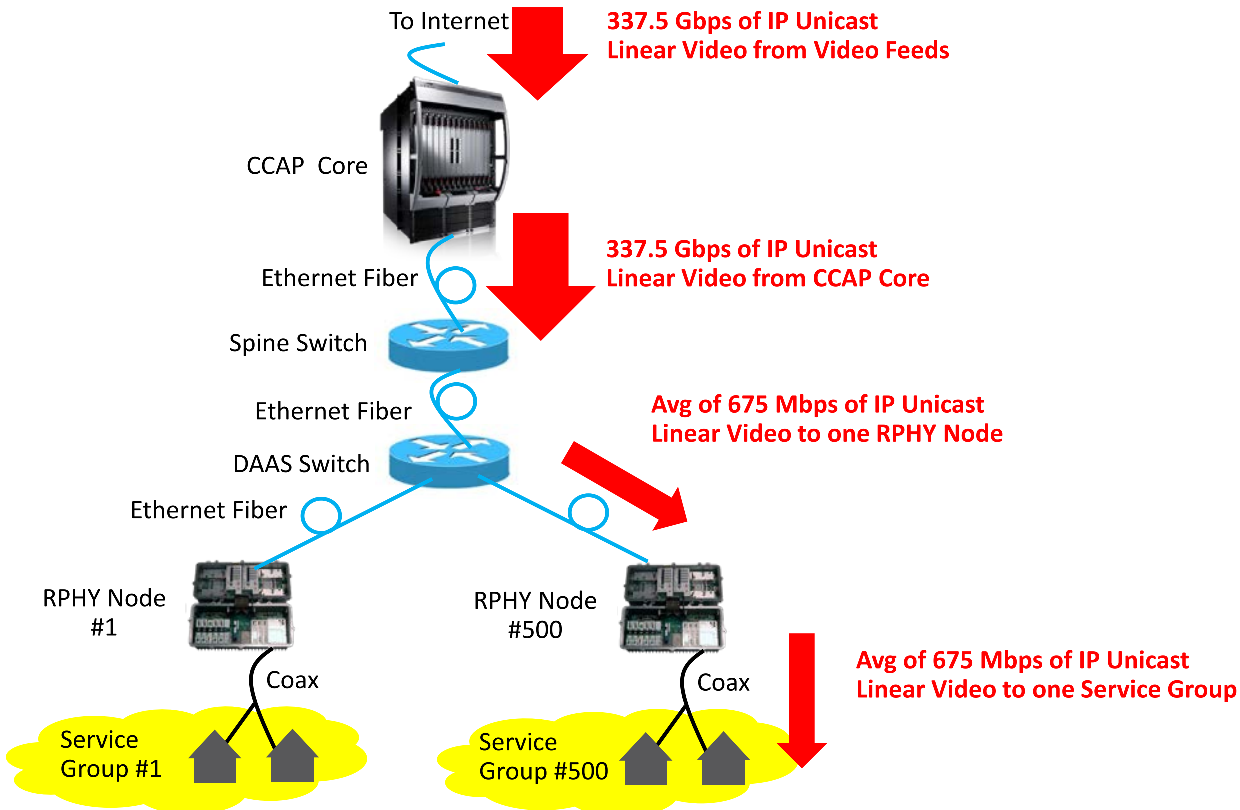


Figure 13 - Linear IP Video Delivery Example using IP Unicasting

- Using Always-On, Nailed-Up IP Multicasting of a single 2.25 Gbps streamed multiplex containing all 250 Linear video programs from the CCAP Core to all of the 500 Remote PHY Fiber Nodes (using the Spine/DAAS Switch Network to replicate the single stream from the CCAP Core to all 500 Remote PHY Fiber Nodes). This implies that only (250) × (9 Mbps) = 2.25 Gbps of aggregate bandwidth would be processed and transmitted from the output of the CCAP Core. The Spine/DAAS switch network would replicate this bandwidth to each of the Remote PHY Fiber Nodes.

  Each of the 500 Fiber Nodes will continuously receive and transmit this entire 2.25 Gbps streamed multiplex of bandwidth onto the HFC plant. (See Figure 14). It should be clear that the packets for the Always-On, Nailed-Up Linear programs must be isolated to a separate channel set to ensure that the IP packets within that channel set are the same for every Fiber Node. They cannot be multiplexed in with regular High-Speed Data packets going to the Fiber Node.

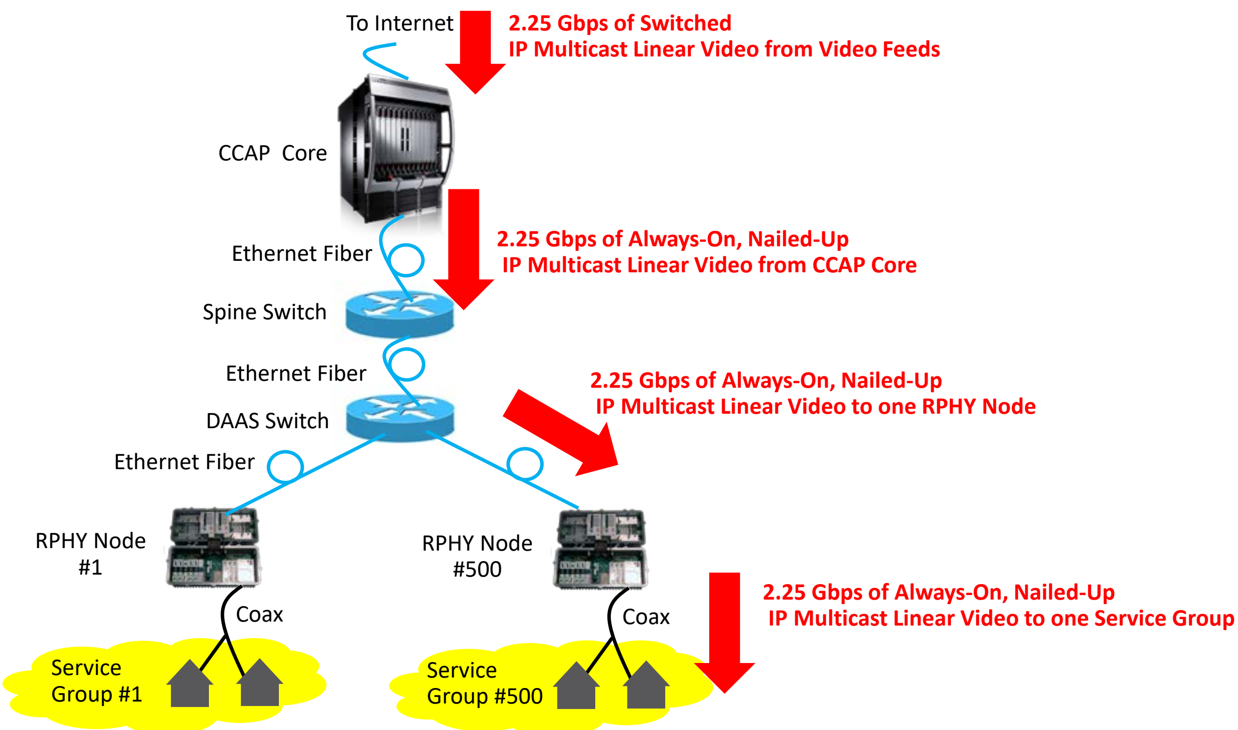


Figure 14 - Linear IP Video Delivery Example using Always-On, Nailed-Up IP Multicasting

- Using Switched IP Multicasting of only the actively-viewed programs from the CCAP Core to the active viewers within all of the 500 Remote PHY Fiber Nodes (using the Spine/DAAS Switch Network to replicate each active video program stream from the CCAP Core to whichever ones of the 500 Remote PHY Fiber Nodes have active viewers for that particular video program stream). With up to 37,500 active viewers supported by the headend, it is practically guaranteed that all 250 programs will need to be processed and transmitted from the CCAP Core, resulting in a total of (250) × (9 Mbps) = 2.25 Gbps of aggregate bandwidth transmitted from the CCAP Core.

  Assume that each Remote PHY Fiber Node has (50 subscriber homes) × (2.5 people per subscriber home) × (60% viewing activity) = 75 actively viewing subscribers during the busy-hour window of time. Based on the blue 250 program curve of Figure 12, it can be seen that a Service Group with 75 active viewers would require ~75 programs to be transmitted, which would imply an average Linear video bandwidth of (75) × (9 Mbps) = 675 Mbps. (See Figure 15).

  At first glance, this solution seems ideal, because it keeps the bandwidth low in the CCAP Core and it keeps the bandwidth low in the Remote PHY Fiber Node's Service Group. Unfortunately, this solution is not easily implementable, because the set of active programs being viewed in one Remote PHY Fiber Node would typically be different than the set of active programs being viewed in another Remote PHY Fiber Node. Since the multiplex of IP packets injected into a particular SC-QAM or OFDM channel for a particular Fiber Node have to be created in the MAC, this implies that a different multiplex on a different channel pseudo-wire would have to be uniquely constructed for each Remote PHY Fiber Node by the CCAP Core's MAC processing functions. This ends up implying that a unique and (most likely) different channel set (ex: ~675

Mbps/36 Mbps ≈ 19 SC-QAM channels on ~19 pseudo-wires) containing an average of ~675 Mbps must be created for each of the 500 Remote PHY Fiber Nodes.

Note: We could also use of fraction of a single 192 MHz OFDM channel and a single pseudo-wire for carrying the 675 Mbps of Linear video content to each one of the 500 remote PHY Fiber Nodes. Thus, a total aggregate bandwidth of ~(500) × (675 Mbps) = 337.5 Gbps must be processed and transmitted from the CCAP Core to service the Linear program content library containing the 250 programs. (See Figure 16). That unfortunately places a large processing load and interface load on the CCAP Core.

This problem can be alleviated if the MAC processing is moved from the CCAP Core into the Remote PHY Fiber Node (creating a Remote MACPHY solution). In this case, the headend could source 2.25 Gbps of IP Multi-casted Linear video content to each of the Fiber Nodes, and the Fiber Node's MAC processing functions can parse through the video content and only process and forward (via IP Multicast) the ~675 Mbps of viewed Linear content onto the HFC plant of the Service Group. (See Figure 17).

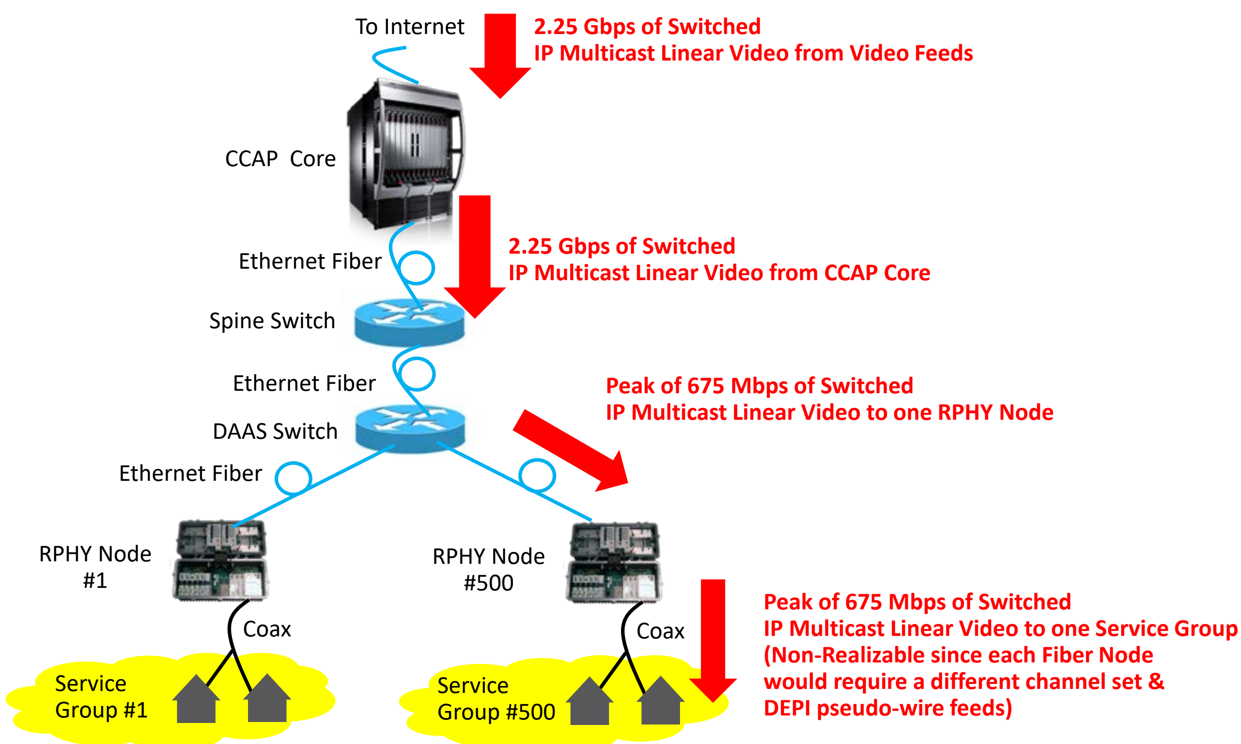


Figure 15 - Linear IP Video Delivery Example using Switched IP Multicasting (Non-Realizable)

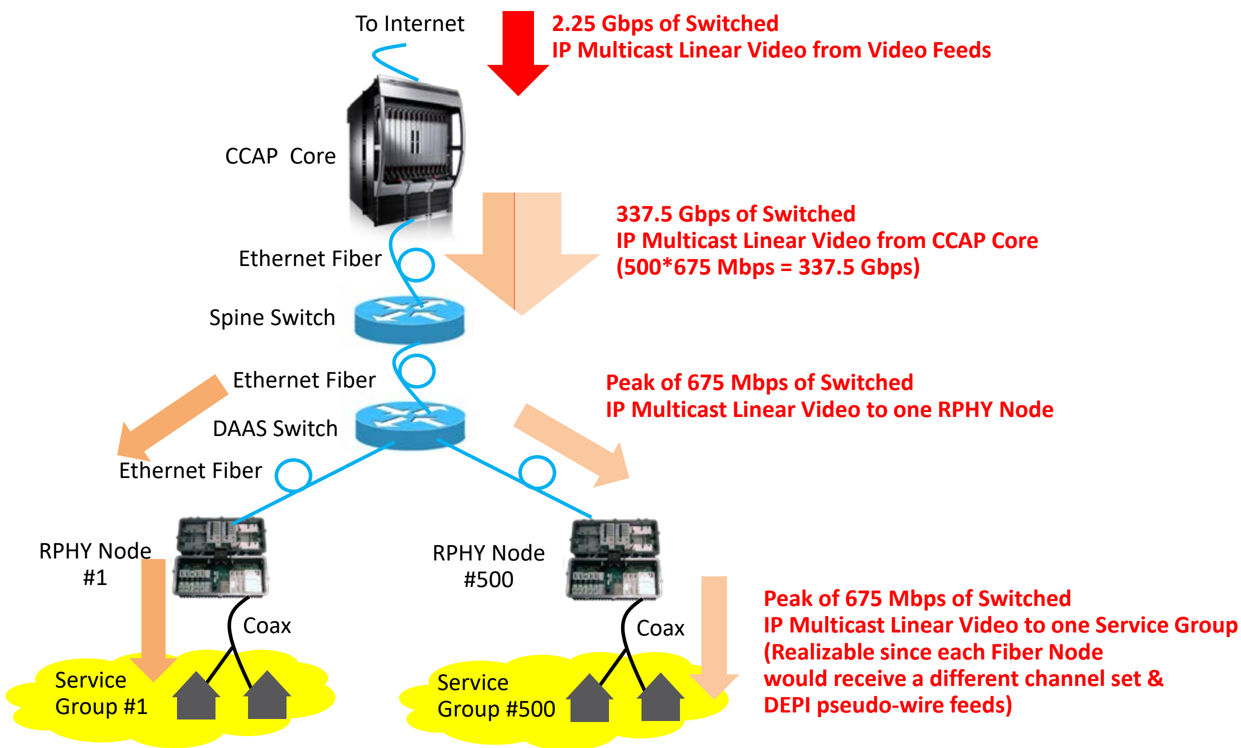


Figure 16 - Linear IP Video Delivery Example using Switched IP Multicasting (Realizable)

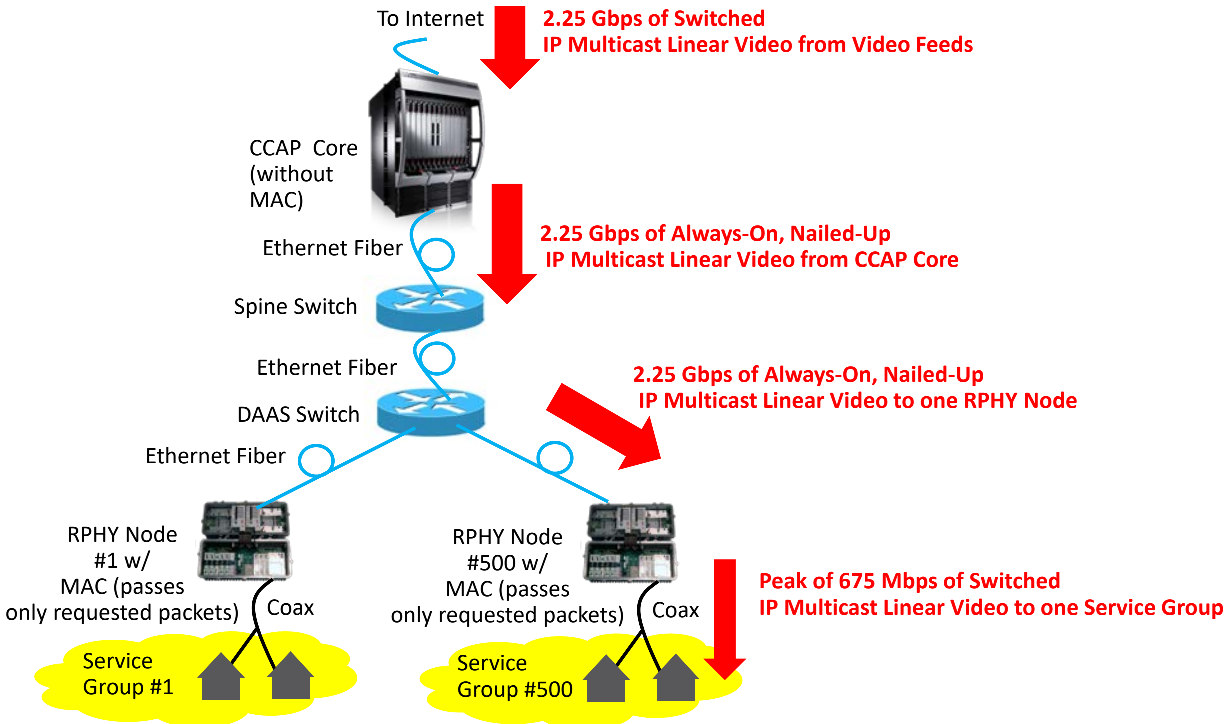


Figure 17 - Linear IP Video Delivery Example using MAC Processing in the Fiber Node

- Using a combination of Always-On, Nailed-Up IP Multicasting for the most popular portion of the Linear video content library along with Switched IP Multicasting for the low-popularity remainder of only the actively-viewed programs. These two sets of Linear video streams would be transmitted from the CCAP Core to each Remote PHY Fiber Node. A single multiplex of the most popular programs would be sent from the CCAP Core and replicated by the Spine/DAAS switch network to go to all 500 of the Remote PHY Fiber Nodes. Then a separate multiplex of low-popularity Switched IP Multicast Linear video programs would be sent separately to each of the 500 Remote PHY Fiber Nodes. This gives a nice compromise solution.

  As an example, if the 60 most-popular programs were included within the Always-On, Nailed-Up IP Multicasted multiplex, that would imply that the CCAP Core would generate a single multiplex of (60) × (9 Mbps) = 540 Mbps containing the 60 most-popular programs. This could be carried in a channel set and pseudo-wires comprising 540 Mbps/36 Mbps = 15 SC-QAM channels. The remainder of the viewed Linear programs associated with a particular Fiber Node would consist of ~(675 Mbps – 540 Mbps) = 135 Mbps of low-popularity programs that could be sent directly to a particular Fiber Node in a Switched IP Multicast feed on a channel set and pseudo-wires containing 135 Mbps/36 Mbps = ~4 channels. Thus, a total of ~675 Mbps and ~19 SC-QAM channels and pseudo-wires would be used to transmit the Linear video programs to a particular Remote PHY Fiber Node. OFDM channels could also be used.

  At the CCAP Core, a total of (540 Mbps + 500 × 135 Mbps) = 68.04 Gbps of Linear video content must be processed and transmitted out of the CCAP Core. (See Figure 18). This represents a great reduction in the CCAP Core processing requirements relative to the previous solution. It should be clear that the packets for the low-popularity programs can be multiplexed in with regular High-Speed Data packets on a channel set, but the high-popularity programs must be isolated to a separate channel set to ensure that the IP packets within that channel set are the same for every Fiber Node.

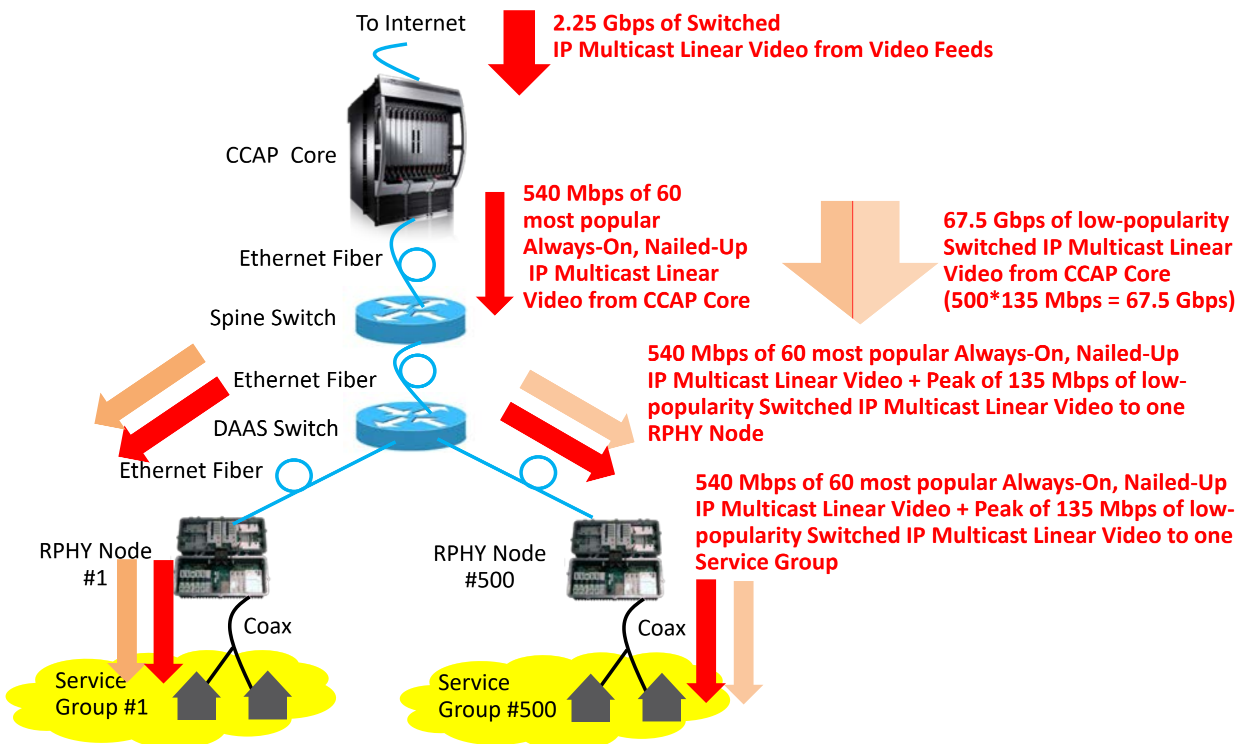


Figure 18 - Linear IP Video Delivery Example using a Combination of Always-On, Nailed-Up IP Multicasting & Switched IP Multicasting

Conclusion

Within this paper, the authors have studied several issues that may need to be addressed as operators move into a future world supporting 1+ Gbps service tiers. The areas that were studied included Traffic Engineering, Utilization Levels, TCP Performance Levels, Symmetrical Services, and Remote PHY IP Linear Video Delivery. Issues were identified, and potential solutions to the issues were proposed. At a high-level, the transition into the 1+ Gbps service environment will require adjustments, but most of the issues seem to have reasonable solutions.

Abbreviations

Avg	Average
BW	Bandwidth
CAGR	Compound Annual Growth Rate
CCAP	Converged Cable Access Architecture
DAAS	Distributed Access Architecture Switch
DEPI	Downstream External PHY Interface
DOCSIS	Data Over Cable System Interface Specification
DS	Downstream
FDX	Full Duplex DOCSIS
Freq	Frequency
Gbps	Giga bits per second
GHz	Gigahertz (one billion Hertz)
HFC	Hybrid Fiber Coax
HD	High Definition
Hz	Hertz
IP	Internet Protocol
MAC	Media Access Control
Mbps	Megabits per second
MHz	Megahertz (one million Hertz)
MSO	Multiple System Operator
MSS	Maximum Segment Size
OFDM	Orthogonal Frequency Division Multiplexing
PER	Packet Error Ratio
PHY	Physical Interface
PON	Passive Optical Network
QAM	Quadrature Amplitude Modulation
RF	Radio Frequency
RPHY	Remote PHY
RTT	Round Trip Time
SC-QAM	Single Carrier Quadrature Amplitude Modulation
SCTE	Society of Cable Telecommunications Engineers
SD	Standard Definition
SG	Service Group
SLA	Service Level Agreement
Tavg	Average Bandwidth Traffic Rate
TCP	Transmission Control Protocol
Tmax	Maximum Bandwidth Traffic Rate
Tmax_max	Maximum of all Maximum Bandwidth Traffic Rates
UEPI	Upstream External PHY Interface
US	Upstream

Bibliography & References

[CLO1] T. J. Cloonan et. al., “Simulating the Impact of QoE on Per-Service Group HSD Bandwidth Capacity Requirements,” in SCTE Cable Tec Expo ’14, September 2014, Society of Cable Telecommunications Engineers.

[CLO2] T. J. Cloonan et. al., “Using DOCSIS to Meet the Larger BW Demand of the 2020 Decade and Beyond,” in SCTE Spring Technical Forum ’16, 2016, Society of Cable Telecommunications Engineers.

[RPH1] Remote Downstream External PHY Interface Specification CM-SP-DEPI-I06-170111, CableLabs.

[RHE1] I.Rhee, L.Xu, “Cubic: A New TCP-Friendly High-Speed TCP Variant”, http://www4.ncsu.edu/~rhee/export/bitcp/cubic-paper.pdf.

[WEI1] D. X. Wei, P. Cao, “NS-2 TCP-Linux: An NS-2 TCP Implementation with Congestion Control Algorithms from Linux”, in proceedings of ValueTool'06 -- Workshop of NS-2, Oct, 2006.

