CHALLENGES DELIVERING MULTISCREEN LINEAR TV SERVICES

MULTICAST-ASSISTED ABR TO THE RESCUE

JOHN ULM, ENGINEERING FELLOW, BROADBAND SYSTEMS
# TABLE OF CONTENTS

OVERVIEW ........................................................................................................................................... 3

MANAGED VIDEO CONSUMPTION TRENDS ..................................................................................... 3

* Case Study – Linear vs. DVR trends...................................................................................................... 4

THE TRANSITION TO A MANAGED IP VIDEO SERVICE ..................................................................... 5

A Multicast Assisted ABR Architecture .............................................................................................. 7

VIEWERSHIP AND CAPACITY – CONTINUED RESEARCH ............................................................... 8

* Viewership – Total Viewers and Total Unique Programs................................................................. 9

ANALYTICAL MODEL – A CAPACITY ESTIMATION TOOL ............................................................... 12

CHANNEL CHANGE IMPACTS ON MULTICAST-ASSISTED ABR ..................................................... 17

* Channel Change Behavior – Live Empirical Data per SG .............................................................. 18

* Channel Change Behavior – Live Empirical data per Large 60-port CCAP chassis .................... 22

* Fast Channel Change – Unicast Capacity Requirements ............................................................ 24

CONCLUSIONS ..................................................................................................................................... 25

REFERENCES ........................................................................................................................................ 28
OVERVIEW

The cornerstone for MSOs IP Video migration is the ability of delivering fully managed Linear TV services to any screen: from tablets & Smart phones to IP STB & Smart TVs to gaming devices & PCs. To meet the challenge of a wide range of bit rates for varying screen sizes from thumbnails to 4K ultra-HD and varying network condition from off-net to on-net requires flexible Adaptive Bit Rate (ABR) protocols. However ABR protocols introduce their own challenges, including breaking the bandwidth capacity bank by requiring a unique stream per screen.

To make IP Video scale successfully, much industry attention has been placed on this problem and Multicast-assisted ABR has risen to be the solution of choice. To better understand the need and issues around implementing this solution, we analyze live viewing behavior across more than 100,000 STB for intervals over a month long. This allows us to accurately quantify viewership and multicast gains. With this knowledge, we validate a model that allows operators to accurately deploy Multicast-assisted ABR.

Our research has also led to the discovery of potentially significant issues around channel change behavior. We quantify the impact of spikes in channel changes that occurs several times per hour. Every channel change might lead to eight or more separate messages over the DOCSIS network. These massive channel change spikes can potentially swamp the network’s control plane. Also, any kind of fast channel change algorithm or personalized advertising could cause a spike in unicast traffic at the same time that actually doubles the required DOCSIS capacity over multicast alone.

In conclusion, our analysis details the significant bandwidth capacity benefits of Multicast-assisted ABR. This Multicast-based ABR solution will be required for operators to roll out IP Video to Multiscreen devices on a wide scale. But there are challenges around channel change events when implementing Multicast-assisted ABR; and we highlight how these might be addressed for IP Video migration to be successful.

MANAGED VIDEO CONSUMPTION TRENDS

Analyzing today’s viewership trends is informative and should apply to IP Video service as the subscriber is agnostic to the actual delivery scheme. Trends that may have an impact on viewership, such as nDVR and new, more user friendly UI are not limited to IP Video. A recent extensive case study on viewership behavior was examined in [ESHET_2014].
Case Study – Linear vs. DVR trends

This case study collected Linear and DVR data usage from more than 1000 homes over a period of a week. There was an average of 2.4 STB per home with Whole Home DVR capabilities and the latest User Interface available in all homes. Data collected includes number of viewers (STBs tuned to Linear content), number of unique Linear TV channels viewed, number of DVR recording sessions, and number of DVR playback sessions. A summary of this data is shown in Figure 1.

Looking at the viewership information, it is very clear that live Linear TV service (i.e. consuming Linear content in real-time) to the STB is still by far the predominant service. From a capacity management perspective, the key takeaways from this case study are:

- Linear is still the king with ~80% of viewers at prime time. No other service comes close.
- The potential advantages of multicast are evident from comparing the unique Linear TV views to the number of unique Linear TV channels (~75% saving).

DVR viewership was second to Live Linear TV with ~15% of viewers at prime time; while VOD usage (<5%) as well as multiscreen usage (<1%) trail with very few of the total viewing eyeballs. Some additional insights from the data include:

1. At peak time (~6-8 PM) almost 60% of the STBs (1.4 STBs / Household) are consuming Linear TV channels.
2. Number of STBs consuming Linear TV channels changes dramatically throughout the day peaking at almost 60%, but is similar between days at a particular time of day.
3. As expected, number of unique Linear TV channels being watched is dramatically lower, at ~300 channels at peak.
4. Number of DVR recording sessions is also very high, at peak time with over 1000 concurrent sessions. Note – the gateway system supports up to six concurrent sessions.
5. DVR recording peaks are correlated with the linear peak with limited recording at off-peak.
6. Number of concurrent DVR play back sessions is significantly lower than the number of recording sessions and accounts to 9% of the STBs.
7. People are recording much more than they are actually watching.

THE TRANSITION TO A MANAGED IP VIDEO SERVICE

An IP Video service is nothing new in the cable space. Over the past few years operators have been offering VOD and Linear services to mainly secondary screens. IP Video is yet to be used as the vessel for the primary video service. The motivation to expand IP Video to the big screen is clear – first and foremost it enables CPE cost reduction, eliminating components such as the cable card, and QAM tuners. Transition to IP video also opens the path to leveraging consumer electronics like consoles and Smart TVs without relying on a STB. An additional driver is the opportunity for significant OpEx and CapEx saving by maintaining a single architecture for video service to all devices.

A baseline IP Video delivery architecture is depicted in Figure 2. With this Adaptive Bit Rate (ABR) architecture, Linear TV channels are transcoded to multiple profiles, segmented, encrypted and pushed to the Origin server. From the Origin server the segments are distributed via a CDN and made available over HTTP to ABR clients residing on STBs, Smart TVs, consoles, mobile devices or browsers, where the latter may consume the Linear service inside or outside the home, depending on content rights. The ABR client registers with the DRM application server and accesses it to get the keys to decrypt the content.

With the availability of new technology, namely home gateways with 16 or more downstream tuners as well as CCAP platforms, the transition to IP is now becoming feasible and cost-effective. The biggest challenge remaining is the bandwidth. Ubiquitous deployment of CPE and Gateways with DOCSIS 3.1 and HEVC technology is still several years away. Cost of massive node splits is still significant and proactively replacing all legacy STBs with IP STBs, needed to reclaim the spectrum of QAM Video, is many years away from being an economical solution. Moreover, high-speed data service keeps growing rapidly and is competing with IP video for the limited spectrum resources available. Thus it is important to operators to find other ways of managing the spectrum needed for the introduction of IP Video services.
Multicast or Unicast Delivery?

ABR is a unicast only video delivery system which uses the most amount of capacity. [ESHET_2014] continued with an analysis of live viewership usage to produce some initial guidelines for the number of DOCSIS channels required for both Unicast only and Multicast only IP Video systems. Those results are summarized in the table below:

<table>
<thead>
<tr>
<th># Tuners</th>
<th>Unicast (Max DS)</th>
<th>Multicast (Max DS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>250</td>
<td>17</td>
<td>8</td>
</tr>
<tr>
<td>500</td>
<td>35</td>
<td>10</td>
</tr>
</tbody>
</table>

It is clear that 50% saving on capacity can be achieved even at small service groups of 125 tuners (or ~35 subscribers). The saving grows with the service group size reaching over 70% at 500 tuners (~140 subs). Moreover, this data indicates that a basic Linear IP Video service is feasible with a minimal spectrum of 5 to 10 DOCSIS downstream channels. It should be emphasized that actual capacity requirement are highly dependent on the HD take rate. In the case of the data analyzed here the number of SD channels at peak time was roughly 2.5 the number of HD channels. Moreover, other aspects of the solution, like unicast traffic for enabling fast channel change, and
targeted ad insertion may dramatically increase the bandwidth requirements, if not addressed effectively.

**A Multicast Assisted ABR Architecture**

While the advantages of an Adaptive Bit Rate solution are well known [ULM_2013, ULM_2010], there is also a need for providing multicast delivery for the most popular content. This has led to a new hybrid system known as Multicast Assisted ABR (M-ABR) Architecture.

The architecture outlined in Figure 4 achieves both the goal of a unified architecture to reach all devices in and outside the home and the goal of minimizing the bandwidth requirements for a Linear IP video service. Three new components are added to the baseline architecture: a Multicast Server, a Multicast Client and a Multicast Controller. The role of the Multicast Server is to pull new Linear segments as they are made available in the CDN and deliver them over multicast. The Multicast Client serves two roles. It serves as a cache for segments arriving over multicast as well as a transparent proxy for requests coming from the ABR clients.

When an ABR client requests a segment, the Multicast Client will intercept the request, check if it can be fulfilled from the cache, and if not, pass it to the CDN. A request for a Linear TV channel segment not already cached in the ABR client can trigger an IGMP Join request to the appropriate multicast in order to start filling the cache. As such, in a typical situation, the first few requests for segments will be fulfilled via unicast whereas all following requests would be met by the cache being filled by the multicast.

Finally, the Multicast Controller serves multiple roles:

1. Collect viewership reporting from the Multicast Clients
2. Control the lineup being offered via multicast
3. Control the (proactive) caching on the Multicast Clients
4. Control the delivery and caching of ads in the Multicast Client

Note that some Multicast Controller roles are directly aimed at optimizing the Multicast assisted ABR service, and ensuring high efficiency compared to a pure unicast service.
VIEWERSHIP AND CAPACITY – CONTINUED RESEARCH

The previous findings were based on empirical data from a single location and single operator. We continued our research to look across a much broader cross section of the industry. Live viewership data was collected from over 150K STB from multiple operators, including a mix of large urban and suburban sites. About one third of these STB were multi-tuner DVR boxes. The total number of Linear TV channels offered for viewing varied from 335 to 615, of which 70 to almost 200 were HD channels respectfully.

All viewership data was sampled at one second intervals around clock, 24x7, for anywhere from 18 to 28 days. Our tool allowed us to organize the raw data from tuners into different sized groupings for analyzing behavior across varying Service Group (SG) sizes. Our primary focus was on SG sizes of 125, 250 and 500 tuners with enough variation around these sizes to observe any patterns.
**Viewership – Total Viewers and Total Unique Programs**

In determining bandwidth capacity requirements, it is critical to understand the viewership behaviors per Service Group (SG). Figure 5 provides two charts from two different operators. The key differentiator between them is the total number of Linear TV channels being offered for viewing. MSO #1 had a total of 335 Linear TV channels including 70 HD channels. MSO #2 had significantly more content available with 615 total TV channels including approximately 200 HD channels.

The charts represent data collected from dozens of different SG. Each SG has a data point for the maximum number of viewers seen across the several week sample period and the maximum number of unique TV programs across the same period. Note that the max viewers and max programs may occur on different dates & times. The SG sizes varied from 125 to 625 tuners with a large concentration around 125, 250 and 500 tuners. Note that since this study was done on an SDV system, the data was organized on a tuner basis. In an IP video world, we might expect an average of 4-5 ‘tuner equivalents’ per home (i.e. IP STB, tablets, Smart TV’s).

The data from the two MSOs look very similar; in fact, the viewership based on SG size tracks very closely between the two. Overall, max peak viewership appears to be a mostly linear relationship with SG size. However, for a given SG size, there was noticeable variation in the max peak viewership between SG, which might be on the order of 40% to 60%. If we add 10% margin for additional statistical variations, we have the following rough guidelines on max peak viewership per SG:

- **Guidelines for Max Peak Viewership per SG**
  - Up to 80% for small, 125 tuner SG
  - Up to 70% for 250 tuner SG
  - Up to 65% for larger 500+ tuner SG

The data points for the max number of unique programs viewed over the several week sample period was also similar between the two MSOs. However, the data definitely flattens out rather than continually rising with SG size.
The max number of unique programs peaked around 125 Linear TV channels for both. Notice that this is significantly less than the 335 and 615 total programs being offered for viewing. While the max number of unique Linear TV programs was similar, it turns
out that the bandwidth capacity between the two was different and is shown in Figure 6.

<table>
<thead>
<tr>
<th># Tuners</th>
<th>MSO #1 Unicast</th>
<th>MSO #1 Multicast</th>
<th>MSO #2 Unicast</th>
<th>MSO #2 Multicast</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>10</td>
<td>6</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>250</td>
<td>17</td>
<td>8</td>
<td>19</td>
<td>11</td>
</tr>
<tr>
<td>500</td>
<td>35</td>
<td>10</td>
<td>35</td>
<td>15</td>
</tr>
</tbody>
</table>

Note – Figure 6 assumes that SD streams require 2Mbps and HD streams use 6Mbps. So while both cases had roughly the same number of programs being viewed, the multicast bandwidth capacity required for MSO #2 is 33% to 50% higher due to a higher mix of HD viewers and content. Remember that MSO #2 had approximately 200 HD channels available for viewing while MSO #1 had 70 HD programs.
The table in Figure 6 provides a summary of unicast and multicast capacity required for the two MSO case studies. For a 500 tuner SG, which might be 100-125 IP video subscribers, the multicast gains are extreme, on the order of 60% to 75% over a pure unicast based ABR system. Even for very small SG of 25-30 subs (~125 tuners); operators can achieve Multicast gains of 30% to 40%.

Note that these multicast bandwidth capacity numbers are only considering the Linear TV viewership portion. Additional capacity will be needed on top of this for VOD, nDVR and any additional capacity needed for a Fast Channel Change algorithm.

ANALYTICAL MODEL – A CAPACITY ESTIMATION TOOL

Live viewership data is invaluable. It lets us see what real people are actually viewing so we can measure things like actual multicast gains and the impact of channel changes. However, this empirical data also has its limitations. An extremely large set of data must be collected to understand the variations from SG to SG and MSO to MSO. The data is sampled at one instant in time and represents a single set of parameters. Finally, it measures existing systems; it can’t anticipate new systems that don’t exist yet.

There has been much work on analytical models over the last decade. Much of this has come out of the SDV world and has been morphed to support IP Video multicast. Some previous work in this area is highlighted in [ULM_2009, ULM_2012]. Analytical models provide control over any possible input and allow one to investigate a lot of “What-if” scenarios. It is not meant to replace the empirical data, rather it supplements it.

The basic outputs of an analytical model are shown in Figure 7. It shows the capacity requirement differences between all unicast delivery system (e.g. ABR) compared to an all multicast delivery system. The X-axis is the number of active IP Video viewers at any instant in time and the Y-axis is the required number of DOCSIS channels. This graph assumed a 50/50 mix of HD and SD viewers with video bit rates of 6Mbps and 1.5Mbps respectively. As can be seen from the graph, unicast bandwidth requirements grow linearly with the number of viewers. 160 IP video viewers require 16 DOCSIS channels.

There are two curves shown for the multicast bandwidth and it is very important to understand the differences between the two. In a multicast world, the number of unique streams required may vary based on which program is being watched by each viewer. The model is based on the Pareto principal where the probability of watching popular content is much higher than less popular programs. The resulting Pareto curve is the origin of the phrase “The Long Tail” to refer to content that is rarely viewed. The lower red curve represents the nominal or typical bandwidth required. So, this is the
average bandwidth and at any point in time one may be as likely above this or below this point.

While average bandwidth utilization is interesting and important, it does not provide a sufficient metric for system capacity planning. Our analytical model also calculates a “high water mark” that represents the maximum capacity required for 99.99% of the time. In the SDV world, this is sometimes referred to as 0.01% Blocking Probability. That is the probability of not having sufficient bandwidth and being blocked from watching a program is 0.01%. As can be seen in Figure 7 for 160 active viewers, IP Video bandwidth consumption will be just over 6 DOCSIS channels on average, but less than 10 DOCSIS channels 99.99% of the time. This compares with the 16 channels required by a pure unicast delivery system.

One of the key challenges is to align our analytical modeling tool with our empirical data. While the analytical model is based on active IP Video viewers, we’ve seen that the empirical results presented so far are based on the physical number of tuners in a Service Group. Then within a sampling of many SG, each SG might have varying numbers of active IP video viewers as well as different mixes HD/SD viewers. So first we’ll take a look at mapping Tuners to Viewers as shown in Figure 8.
The chart on the top of Figure 8 is the unique program data from MSO #1 in Figure 5 sorted by the number of Tuners in a Service Group. As can be seen for 500 tuner SG, the maximum number of unique programs seen by each SG varied from 88 to 129 unique programs, almost a 50% swing. As we analyzed the empirical data, we wanted to understand the relationship between unique programs and maximum number of viewers rather than tuners. This is now shown in the second chart on the bottom of Figure 8. As can be seen, there is a much stronger correlation between unique programs.
and max viewers than there is between unique programs and tuners in a SG. At 260 viewers, the variation is only from 109 to 129 unique programs, less than 20% swing between them.

The next step in validating the analytical model is to correlate it to these results. All of the parameters from the empirical data in Figure 8 were entered into the analytical model. The only parameter left to adjust are the Pareto parameters. As explained in [ULM_2009], the Pareto curves are based on the Zipf function where the key input is the exponential parameter, Z. Note that as Z increases, the Long Tail becomes smaller and multicast gains increase.

![Max Programs Viewed by # of Max Viewers](image)

**Figure 9 – Correlating Analytical Model to Empirical Data, Z=1.125**

Earlier SDV research often used Z=1 or even used values of Z less than one. Our research looked at various values of Z to find the best one to fit the data. The resulting value of Z=1.125 provided the best fit and is shown in Figure 9. Note that the analytical model results are the High Water Mark so should be just higher than all of the empirical data results. Had we used a value of Z=1, then the analytical model results would have increased by almost 15% for the larger number of viewers. This might be acceptable to some if they would like some additional margin in their system.

Our research into the optimum value for Z with other empirical data sets has shown that the value may actually vary higher to Z=1.2 or even Z=1.3. Again, larger values of Z imply smaller Long Tail and more multicast gain. For the data set in Figure 9, using Z=1.25 would reduce the model results by 12%-13% for larger numbers of viewers. Note that the parameter Z becomes more important for larger SG with many more viewers. For smaller SG, the impact of choosing the value for Z is much less.
With an analytical model that is correlated to the live viewership data, we can now explore various tradeoffs to understand the potential impacts in shifts due to one or more parameters. This is the real power of the analytical model. Two examples are given below.

In the first example shown in Figure 10, the amount of unicast based On-Demand traffic is varied from 0% to 100%. On-Demand traffic might be VOD, nDVR or ABR based multiscreen clients. This example assumes a 50/50 mix of HD and SD viewers; a total of 200 HD programs available for viewing and 400 SD programs.

As can be seen for 240 active IP video viewers, the bandwidth required can vary by a factor of 2:1 from 12 to 24 DOCSIS channels. A chart like this will be useful for an operator considering a switch to nDVR where On-Demand traffic may jump from 5%-10% for VOD usage to 25% to 40% unicast in an nDVR system.

For the next example shown in Figure 11, the number of HD programs available for viewing is varied for an HD only system. In this example, the number of HD programs available is varied from 70 HD programs up to as many as 800 HD programs. As can be seen, the impact becomes much more substantial for large SG with many additional viewers. For 300 active viewers, it also reaches a 2:1 swing in bandwidth required.
These examples help illustrate how a couple parameters can cause giant swings in the bandwidth required for IP video. Remember that these charts are for the High Water Mark and meant to show the maximum bandwidth required 99.99% of the time across dozens and dozens of different SG. In general, as operators have smaller SG, the relative swings in bandwidth required become much smaller.

**CHANNEL CHANGE IMPACTS ON MULTICAST-ASSISTED ABR**

The overriding goal of an M-ABR system is to provide bandwidth capacity savings compared to a pure, unicast only, ABR system. Remember, that M-ABR is a hybrid system where video streams are first received as unicast content directly from the CDN, and then later potentially switched to multicast delivery based on decisions by the Multicast Controller. Any capacity analysis of an M-ABR system needs to account for the mixed unicast/multicast traffic.

Our research into the area has found that Channel Change (CC) events are fairly synchronized in time, up to six times per hour. This corresponds to the programs ending
on half hour intervals and the two commercial breaks in between. It turns out that these CC events can cause a spike in the system that greatly effects capacity requirements.

**Channel Change Behavior – Live Empirical Data per SG**

To get a handle on CC behavior, we analyzed empirical data from over 150K STB from multiple MSOs and sampled continuously across several weeks. For analysis, we evaluated the data at SG granularity (i.e. 500-600 tuners per SG) and then at the CCAP chassis level (i.e. ~35K tuners per 60-port CCAP).

Figure 12 shows the maximum number of CC events per second for dozens of different SG for both MSO #1 and MSO #2. The SG size varied from 125 to 625 tuners with clusters around 125, 250 and 500 tuners.

The empirical data shows that a 500 tuner SG might see as many as 45 CC events within a single second over the several week sample period. To put that in perspective, each CC event may cause an IGMP Join and an IGMP Leave operation. These in turn may trigger multiple DOCSIS messages to manage the multicast service flows. All told, there may be 8 or more control messages generated by each CC event. While the overall bandwidth required for these control messages will be relatively small, their impact on the control plane processing inside the CCAP may be more significant and needs to be considered.

Perhaps the larger observation from these results is that there is significant variation from SG to SG. It is on the order of 2:1 and was seen with both data sets. This large variation gave us concern so we dug a bit deeper to understand the CC behavior better. Next, we investigated the SG to SG variation for a set of SG that was basically the same size. This included 59 SG from MSO #1 and 64 SG from MSO #2 all in the range of 550 to 600 tuners with an average of 585 tuners per SG. Figure 13 shows the maximum CC event over the several week sample period for each of the different SG. For MSO #1, the worst case SG saw a Max CC event with 34 CC per sec while the SG with the lowest Max CC peak was only 14 CC per sec. The average SG Max CC peak was 24 CC per sec. MSO #2 was very similar with the worst case SG that had a peak of 33 CC per sec; lowest SG peak of 8 CC per sec and an average peak of 21 CC per sec.

Note that even with fairly consistent SG sizes, a significant variation in Max CC peak between SG can still be seen. This collaborated what we were seeing in Figure 12.
**Figure 12 – Max Channel Changes per Sec for MSO #1 & #2**

<table>
<thead>
<tr>
<th># Tuners</th>
<th>MSO #1 Max CC / sec</th>
<th>MSO #2 Max CC / sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>250</td>
<td>35</td>
<td>18</td>
</tr>
<tr>
<td>500</td>
<td>45</td>
<td>34</td>
</tr>
</tbody>
</table>
The previous results only look at the maximum CC peak across a several week period. We then took a closer look at one of the worst SG to understand the daily variations that occur in Max CC events. This is shown in Figure 14.

Empirical data was collected over 28 consecutive days from SG #36. On a day to day basis, there are still variations on the order of 2:1 for Max CC events. The worst day saw a peak of 29 CC per sec, while the minimum daily peak was only 12 CC per sec. The average daily peak roughly split these at 20 CC per sec.
Finally, we looked closely at the instantaneous behavior around the worst case CC events. Figure 15 shows the timeline for four of the worst case events that happened over more than 100 SG being sampled over several weeks.

SG #40 (green) had exhibited the highest peak at 34 CC events per second. However, the preceding and following seconds only had 7 and 4 CC events respectively. So this was truly a 1 second spike in the system and not a sustained event.

SG #49 (red) exhibited the largest number of CC events in a 2-second window (i.e. 50 CC per 2-sec). In the time immediately following this CC burst, it then drops off quickly to 14 and then 4 CC per sec. SG #36 (orange) had the largest 4-second window of 71 CC events. Finally, SG #32 (purple) is an example of a CC event that is spread over 4-5 seconds, but its worst 1-second peak is just 22 CC per sec.
So, our summary of observations from 600 tuner SG empirical data includes:

- **SG To SG Variations**
  - MAX to Min ratio of ~2:1
  - MAX Peak is ~50% more than Average SG Peak
- **Daily SG Variations**
  - MAX to Min ratio of ~2:1
  - MAX Peak is ~50% more than Average SG Peak
- **Instantaneous Burst**: 
  - Peak lasts for 1-2 seconds; then drops substantially

### Channel Change Behavior – Live Empirical data per Large 60-port CCAP chassis

Because of the significant variations from SG to SG, day to day and even second to second, we analyzed the impact of CC events at the CCAP chassis level to try to determine how much statistical gain, if any, we might experience. To emulate a 60-port CCAP chassis, the empirical data was organized in groups of 35K tuners.

Figure 16 shows the daily peak CC variations seen at the chassis level. The max peak across the 28 sample days was 280 CC per second. The lightest daily peak was only 69 CC events with an average daily peak of 171 CC per second.
Figure 17 shows the worst case instantaneous CC events at the 60-port CCAP chassis level. The various worst case windows are:

- Max 1-sec window = 280 CC per sec
- Max 2-sec window = 465 CC per 2-sec
- Max 4-sec window = 698 CC per 4-sec
One interesting observation from this data is that while the max peak lasts 1-2 seconds, there is 3-5 seconds of residual CC events at roughly half the peak. As we analyze these events further it will be interesting to see if these are being driven by channel surfers.

So, what kind of statistical gains do CC events experience when aggregated over a 60-port CCAP chassis? The single worst case SG had a max peak of 34 CC per sec. Extrapolating 60 SG @ 34 CC per sec would suggest a max of 2000 CC per sec at the chassis level. But we see this is a factor of 7.3X too high! Even if using the average SG max peak of 24 CC per sec suggests 1440 CC per sec at the chassis level. This is still a factor of 5X more than the 280 CC per sec that was actually experienced. So the conclusion is that the daily and instantaneous variations in CC events help attribute to a sizable statistical gain for the large 60-port CCAP chassis, on the order of 5X:

- Guidelines for Max Peak CC per sec for 60-port CCAP chassis
  - ~300 CC events per sec for 1-2 sec
  - ~150 CC events per sec for following 3-5 sec

**Fast Channel Change – Unicast Capacity Requirements**

As stated earlier in the paper, it is important to understand the potential bandwidth capacity impacts from the additional unicast traffic generated by CC events. Additional simulations of the empirical data were run to simultaneously analyze the multicast viewership and the CC data. The total capacity required was calculated assuming that the CC event required an average of 2 seconds of unicast traffic before switching to a multicast stream. The results of this run are shown in the table below.

As can be seen in the table of Figure 18, the unicast impact of CC events grows linearly with the size of the SG. While the impact is just a single additional DOCSIS channel for 125 tuner SG, it doubles with 250 tuner SG and then quadruples with 500 tuner SG. For 500 tuner SG (e.g. 100-125 IP video subs), the unicast CC traffic is adding an additional 40% to the pure all-multicast bandwidth capacity requirements.

<table>
<thead>
<tr>
<th># Tuners</th>
<th>MSO #1 Unicast</th>
<th>MSO #1 Multicast</th>
<th>MSO #1 Mcast + CC</th>
<th>CC Delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>10</td>
<td>6</td>
<td>7</td>
<td>+1</td>
</tr>
<tr>
<td>250</td>
<td>17</td>
<td>8</td>
<td>10</td>
<td>+2</td>
</tr>
<tr>
<td>500</td>
<td>35</td>
<td>10</td>
<td>14</td>
<td>+4</td>
</tr>
</tbody>
</table>

*Figure 18 – Impact of 2-second Fast Channel Change (Unicast)*

In a separate case study that analyzed using SG in the 550 to 600 tuner range, the multicast capacity required was 130 unique streams compared to 360 unique streams.
needed for pure all-unicast ABR system. In looking at the impact of a 3-second Fast Channel Change algorithm, the max number of unique streams increased by 50% to 194 streams compared to the all-multicast scenario.

So, the capacity requirement impact of CC events can be significant. The impact becomes very dependent on any Fast Channel Change algorithms used and is very sensitive to the duration of the unicast bursts during the CC event.

**CONCLUSIONS**

Multicast-assisted ABR has risen to be the solution of choice for scaling IP Video delivery systems over cable. To better understand the need and issues around implementing this solution, live viewing behavior was analyzed across more than 100,000 STB for intervals over a month long. This allowed us to accurately quantify viewership and multicast gains and achieve some guidelines. Guidelines for viewership:

- Guidelines for Max Peak Viewership per SG
  - Up to 80% for small, 125 tuner SG
  - Up to 70% for 250 tuner SG
  - Up to 65% for larger 500+ tuner SG

While IP Video bandwidth requirements based on empirical data resulted in:

<table>
<thead>
<tr>
<th># Tuners</th>
<th>MSO #1 Unicast</th>
<th>MSO #1 Multicast</th>
<th>MSO #2 Unicast</th>
<th>MSO #2 Multicast</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>10</td>
<td>6</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>250</td>
<td>17</td>
<td>8</td>
<td>19</td>
<td>11</td>
</tr>
<tr>
<td>500</td>
<td>35</td>
<td>10</td>
<td>35</td>
<td>15</td>
</tr>
</tbody>
</table>

Note that MSO #2 had a significantly expanded program offering of more than 600 Linear TV channels with approximately 200 of them HD. This is roughly twice that of the MSO #1 data set.

While the empirical data from Live viewership is invaluable, it has limitations so our research validated our analytical model against the Live data. We were able to show that we can accurately model real live results. Our studies showed that a Pareto exponent of $Z=1.125$ worked best for this data set, but might have to varied from $Z=1$ to $Z=1.3$ depending on other circumstances. With a validated analytical model, some examples were shown of varying key parameters such as multicast/unicast mix and HD.
programs available for showing. Results show that a wide swing in these parameters could impact bandwidth requirements by a factor of 2:1.

Our research has led to the discovery of potentially significant issues around channel change behavior. The results of our empirical studies of Live viewership analysis shows:

- SG To SG Variations in Max CC
  - MAX to Min ratio of ~2:1
  - MAX Peak is ~50% more than Average SG Peak
- Daily SG Variations in Max CC
  - MAX to Min ratio of ~2:1
  - MAX Peak is ~50% more than Average SG Peak
- Instantaneous Burst for Max CC:
  - Peak lasts for 1-2 seconds; then drops substantially

While the SG to SG variations were extreme for Max CC, the analysis then looked at the impact of Max CC at the large 60-port CCAP chassis level. This resulted in the following guidelines:

- Guidelines for Max Peak CC per sec for 60-port CCAP chassis
  - ~300 CC events per sec per chassis for 1-2 sec
  - ~150 CC events per sec per chassis for following 3-5 sec

With a better understanding of Max CC behavior, the analysis then investigated the impact of additional unicast capacity needed for Channel Changes around these events. The bandwidth requirement appears to scale linearly with SG size and might require an additional 40% capacity increase over a multicast-only system with 500 tuners (e.g. 100-125 IP video subs).

In conclusion, our analysis details the significant bandwidth capacity benefits of Multicast-assisted ABR. This Multicast-based ABR solution will be required for operators to roll out IP Video to multiscreen devices on a wide scale. But there are challenges around channel change events when implementing Multicast-assisted ABR that operators will need to address.

**RELATED READINGS**

- PAPER: Transcoding Choices for a Multiscreen World
- PAPER: Efficient Content Processing for Adaptive Video Delivery
MEET ONE OF OUR EXPERTS: John Ulm

John Ulm is a Fellow of the Technical Staff within the ARRIS Network Solutions CTO group. In this role he has been investigating strategic technical directions for multi-screen services and broadband bandwidth expansion. Recent activities include research into Multi-screen Adaptive Streaming IP Video solutions; next generation CCAP architecture; next generation HFC architectures; and new HFC protocols including DOCSIS 3.1 and IEEE 802.3bn EPoC.

John’s two+ decades in the Broadband industry began as designer, architect and MAC protocol developer at LANcity, pioneering the industry’s first cable modem systems. He was one of the primary authors for the Cable Industry’s DOCSIS 1.0/1.1 specifications that drove cable modem success during its early days. He also spent time as a Network Processor architect for Nortel and as a senior technical consultant to the Broadband industry with YAS Corp.

John holds a BSEE and MSEE from RPI and has multiple papers and patents to his name.
REFERENCES


2) [ULM_EXPO_2013] – “Unmanaged ABR – How to Control those Unruly Teenagers”; J. Ulm, A. Eshet, N. Radian, SCTE Cable-Tec Expo, Fall 2013


4) [ULM_EXPO_2012] – “Managing Bandwidth Growth for Migration to ALL IP Video; A Case Study for Bandwidth Modeling Results”; J. Ulm, SCTE Cable-Tec Expo, Fall 2012

