Architectural Approaches to Help Circumvent the "Simulcast Roadblock" of IP Video Deployments

T. Cloonan, J. Allen, C. Ansley, R. Arnold, J. Brooks, M. Bugajski, C. Cheevers, T. Cotter, J. Howe, B. Hanks, D. Torbet, M. Uecker, & I. Wheelock

ARRIS

Abstract

Multiple System Operators (MSOs) throughout the Cable Industry are planning to roll out IP video services. Many MSOs hope to eventually migrate these IP Video Services to their DOCSIS 3.0 infrastructure in an effort to capitalize on many benefits, including the statistical multiplexing gains of channel bonding, the QoS that permits partnerships with other Internet content providers, and the simplicity of a single delivery system.

One particular challenge to this planned migration is created by the fixed amount of bandwidth available on the HFC plant to support simultaneous deployment of legacy services and the new DOCSIS-based IP video services during the transition. This "Simulcast Roadblock" problem will be described within this paper, and a toolkit of potential solutions will be identified.

INTRODUCTION

Background on IP Video Services

Multiple System Operators (MSOs) are beginning to plan architectures which will ultimately be used for the deployment and delivery of Internet Protocol (IP) Video services to their subscriber base. From the subscriber viewpoint, IP Video will manifest itself as a video delivery system that permits their video content to be distributed over their home IP network to multiple types of devices: IP STBs (with accompanying TV display devices), IP-enabled TV"s, game players, DVD players, handheld devices (such as tablets and smartphones), and PCs.

Different MSOs will move toward IP Video at different times and at different rates, and different MSOs will also choose slightly different architectures as they unveil these new services. However, the authors believe that once the transition to IP Video begins within any particular MSO network, it will likely occur quite rapidly. MSOs will work to offer competitive responses to the Over-The-Top Video (OTT) content providers who have begun to thrive on the Internet. In many ways, this OTT content has created a new challenge to the MSOs" legacy video delivery systems [Ins1].

By offering this new type of IP Video service to their subscribers, most MSOs are working to accomplish several important goals, including:

1) Gaining access to a broader audience through delivery to multiple screens in the home
2) Building a brand identity with the 15-30 year-old demographic (through their handheld devices)
3) Creating new means of further monetizing their high-quality video content with new subscription fees (for the various multiple screen types in the home)
4) Providing an opportunity to enter the growing "Internet advertising market" through directed advertising in IP-based videos
5) Ring-fencing their subscriber base by becoming the popular organizers and

aggregators of all IP video content (MSO-based and Web-based)

6) Reducing the high CPE costs traditionally associated with legacy STBs

Three Modes of IP Video Delivery

Three fundamentally different delivery paths are currently being architected and studied by MSOs, and the services associated with each delivery path may be rolled out at different times. These delivery paths include:

1) Internet-based, unicast delivery of MSO-owned VoD and Linear video content to subscribers not directly connected to the MSO‟s HFC plant. This is known as “off-net” delivery of the video content. Depending on where the content is actually hosted, this service could even be a cloud-based IP video service for the MSOs.

2) On-Net HFC-based, unicast delivery of MSO-owned VoD and Linear and Remote Storage DVR (RS-DVR) video content to subscribers directly connected to the MSO‟s HFC plant. This is known as MSO-managed or walled-garden or “on-net” delivery of VoD and Linear and RS-DVR video, because the MSO can better manage the link utilization and Quality of Experience associated with the service (since it is offered entirely on the MSO‟s links).

3) On-net HFC-based, multicast delivery of MSO-owned Linear video content to IP-based subscribers directly connected to the MSO‟s HFC plant. This is known as MSO-managed or walled-garden or “on-net” delivery of Linear video. The use of multicast delivery for Linear video provides bandwidth efficiency gains over the use of unicast delivery by reducing the number of stream replications appearing on the HFC plant.

Three Transport Methods for On-net Delivery

The actual on-net delivery of the content across the HFC network can be implemented in at least three different ways:

1) MPEG/IP Encap

The video content could be transmitted to the home using the legacy QAM-based MPEG-TS digital video distribution infrastructure (which is normally used to transmit video signals to legacy STBs). The in-home receiving device receives the MPEG-TS content, re-encapsulates it into IP packets, and then transmits those packets throughout the home IP network where they can be accessed and decoded by any IP-enabled device with the appropriate client

2) DOCSIS

The video content could be transmitted to the home using the legacy DOCSIS distribution infrastructure, with an advanced in-home cable modem (called a media gateway) receiving the IP packets. Depending upon the home video distribution chosen, the video packets may be stored in the media gateway to provide whole-home DVR functionality and time shift buffers, and/or processed to accommodate lower bit rate clients, and/or forwarded directly to a subtending client on the home network. Depending on how the MSO chooses to implement multicast Linear distribution, a multicast to unicast conversion of the IP streams might also be required within the media gateway before forwarding the packets through the home network. This multicast to unicast conversion is merely one of several important features, such as device discovery, encryption, digital rights management, and caching, which are found in media gateway devices. These media gateway devices have been called cable modems on “video steroids.” When used for Linear video, a DOCSIS video delivery service promises to offer all of the improved bandwidth efficiencies of legacy switched digital video services, because IP multicast streams will likely be transmitted to a service group only when one or more clients in that service group have joined multicast sessions for long-tail content.

3) Bypass

The video content could be transmitted to the home using a proprietary, bypass-enable EQAM with a subset of DOCSIS MAC functionality. The video streams would then be fed through the proprietary EQAM over a DOCSIS channel to a proprietary cable modem, which then forwards the video streams through the home network (assuming no media gateway functionality).

Each approach has its own list of potential issues. The authors believe that the negative issues for the proprietary bypass approach are quite constraining, and these issues have already been outlined in other papers (ex: [Clo1]). As a result, proprietary bypass approaches will not be considered within this paper, and the authors will instead focus on IP video solutions based on MPEG/IP Encap transport and IP video solutions based on DOCSIS transport. The authors believe that many MSOs will eventually want to chart a course that takes them to the particular solution that uses DOCSIS for IP video. Reasons for this belief include:

1) DOCSIS would provide the simplicity and low cost of a single, fully-converged infrastructure for all services; voice, data, and video.
2) DOCSIS channel-bonding provides statistical multiplexing gains allowing more video programs to be delivered to subscribers on a given number of channels. It is expected that many MSOs will also want to capitalize on the further benefits that can be realized if Variable Bit Rate (VBR) IP video streams are utilized instead of Constant Bit Rate (CBR) video streams in a channel-bonded IP Video environment. As an example, with four Downstream Bonded Channels in a particular Downstream Channel Set, it has been shown that an MSO can transmit ~25% more programs on those four Bonded Channels than when the four Channels are used in a Non-Bonded fashion (see **Fig. 1**). To account for the lower bandwidth capacity requirements for VBR IP Video on bonded channels (**Fig. 1**), we will utilize the average bandwidth of each video stream to determine the amount of capacity consumed whenever those IP video streams are flowing over a reasonably-sized Bonding Group with three or more DOCSIS Downstream Channels. However, we must assess a ~25% “No-Stat-Mux-Gain tax” whenever video programs are multiplexed together on small Downstream Channel Sets (or non-bonded channels). In particular, within this paper, it will be assumed that video programs which are multiplexed together on a non-bonded channels will each require an increased effective bandwidth capacity given by 1.25x(the average bandwidth value for the video stream). This “No-Stat-Mux-Gain tax” will account for the extra bandwidth headroom that must be made available in a single Downstream Channel to accommodate the burstiness and larger peak bandwidth-to-average bandwidth ratios of IP Video stream aggregates that do not contain many programs. Thus, we will assume that channel bonding systems with N bonded DOCSIS channels will support 25% more video programs than N-channel systems that do not support channel bonding (whenever N>=3). It should be clear that legacy MPEG-TS video delivery does not

provide bonding, and therefore it is subject to the 25% tax as well.

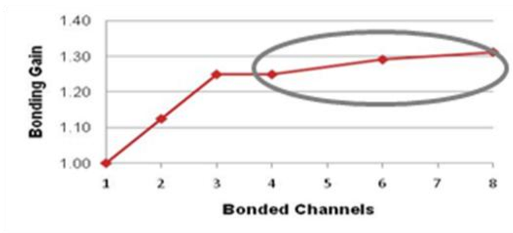

Fig. 1- VBR Stat-Mux-Gain (Bonding Gain) as a Function of the Number of Bonded Channels carrying 4 Mbps Streams

3) DOCSIS provides advanced QoS that is necessary to safely source IP video content from Internet content partners
4) Many DOCSIS CMTSs offer the high availability needed for future services
5) Per-channel pricing on DOCSIS CMTS equipment is dropping as demand for channels increases

As a result of these benefits, many MSOs are likely to pursue one of several paths in the future as they make their way toward an ultimate DOCSIS IP Video delivery solution. Two of these paths are outlined below.

Two Basic Transition Strategies

MSOs must carefully plan their paths as they migrate towards IP Video Architectures in the future. Two different paths are likely to be followed by different MSOs. One is the 2-Step Transition Strategy, and the other is the 1-Step Transition Strategy.

In the 2-Step Transition Strategy, MSOs convert their video delivery methods in two steps. They begin with their current video delivery methods, incorporating technologies such as analog transport of broadcast video, digital video transport of broadcast video, digital video transport of SDV, and digital video transport of VoD. They transition to Phase 2, IP in the Home, as they deploy an MPEG/IP Encap delivery system, using digital video transport of broadcast, SDV, and VoD over the HFC plant, and then converting to IP packets for transmission over the home network to low cost IP-based STBs or other IP-based connected devices. IP in the Home can be implemented as a replacement for the current in-home delivery system, or it can be implemented as an augmentation to the current delivery system. They finally transition to Phase 3, All IP, by moving their video streams to the DOCSIS infrastructure. The All IP Phase can be implemented as a replacement for the MPEG Delivery Phase delivery system, or it can be implemented as an augmentation to the MPEG delivery system. In other words, an MSO may choose to introduce the IPTV experience with an IPTV tier, which may or may not contain channels that overlap with the conventional tiers. Any overlap of the programming between MPEG and DOCSIS transport will require simulcasting of video streams on both the analog/MPEG infrastructure and the DOCSIS infrastructure.

In the 1-Step Transition Strategy, MSOs convert their video delivery methods in one step. They begin with their current video delivery methods, incorporating technologies such as analog transport of broadcast video, digital video transport of broadcast video, digital video transport of SDV, and digital video transport of VoD. They transition to the final All IP Delivery phase, Phase 2, when they move to the DOCSIS infrastructure. The All IP Delivery Phase could be implemented as a replacement for the current MPEG delivery phase, but in most real-world scenarios, there will likely be a window of time during which the MPEG and DOCSIS delivery systems are running in parallel. This requires simulcasting video streams on both the analog/MPEG infrastructure and the DOCSIS infrastructure.

In both the 2-Step Transition Strategy and the 1-Step Transition Strategy, there is a possibility that simulcasting of some video streams may be required on both the analog/MPEG infrastructure and the DOCSIS infrastructure. This simulcasting requirement could lead to challenging demands on the HFC plant bandwidth capacities, which could in turn become a roadblock to the deployment of IP video services. This paper will attempt to describe the potential Simulcast Roadblock problem and propose useful solutions.

PHILOSOPHIES ON THE FUTURE AND AN EXAMPLE OF THE SIMULCAST ROADBLOCK PROBLEM

Regardless of which Transition Strategy is selected, MSOs may run into the Simulcast Roadblock problem. The severity and magnitude of this problem will be different for each MSO. To some extent, the magnitude depends on which of two differing philosophies on IP Video bandwidth growth ultimately proves to be true. These two philosophies might be labeled the "zero-sum-game" philosophy and the "new-growth" philosophy. According to the "zero-sum-game" believers, MSOs who add MSO-managed IP video delivery to all three screens will not increase the total bandwidth consumption of their subscribers. Instead, they will merely ensure that a good portion of the IP video content viewed on PCs and hand-helds will be sourced from servers owned by the MSOs (instead of being sourced by Over-The-Top providers).

According to the "new-growth" believers, MSOs who add MSO-managed IP video delivery to all three screens and couple it with the high-end content in their content libraries (which is much better than the content in the content libraries of Over-The-Top providers) will create a new demand for bandwidth capacity that did not exist for the Over-The-Top providers in the past. This will ultimately increase the total number of eyes watching IP Video and increase the total bandwidth consumption of their subscribers. Recent deployments of iPad applications that can receive Live programming content from the MSOs" DOCSIS networks seem to be displaying this type of "new-growth" behavior, but it is still too soon to tell.

Regardless of which philosophy is correct and regardless of which Transition Strategy a particular MSO selects, the period of time when simulcasting is required may lead to challenges for the HFC plant bandwidth- especially if MSOs are reluctant to reduce the size of the existing service offerings. We will perform an example analysis below assuming the more conservative numbers of the zero-sum-game philosophy. However, it should be recognized that the magnitude of the simulcast bandwidth problem could be much larger if the new-growth philosophy proves to be at all valid in the future.

While it is not possible to present all of the different scenarios that MSOs are likely to encounter, it may be beneficial to examine the magnitude of the simulcast roadblock problem for a mythical MSO whose HFC plant and channel characteristics might be "typical." Defining a "typical" HFC plant for an analysis is always a challenging task, because the characteristics of different HFC plants can vary quite extensively from MSO to MSO, and the validity of any assumptions will undoubtedly be heavily debated. Nevertheless, for this paper, we will assume the following to be "typical" HFC plant characteristics for a mythical MSO in the future:

1. The conservative "zero-sum-game" philosophy will be assumed, so it is possible that bandwidth capacity demands would be even greater if the "new-growth" philosophy were used
2. The HFC plant is a 750 MHz plant that supports 115 channels
3. Each digital channel carries a 6 MHz, 256 QAM signal that supports 42 Mbps of raw bandwidth
4. For video distribution prior to the arrival of IP Video, no digital

broadcast programs are being used. Only analog Linear programs, digital SDV Linear programs, and digital VoD programs are being used prior to the arrival of IP Video. (Note: This approach may not be common, but it simplifies the analysis to assume that all legacy digital, Linear programs are delivered via SDV).

5. For IP Video distribution, no nailed-up programs are being used. Only Switched Digital multicast IP Linear programs and unicast IP VoD programs are used. (Note: It is possible that digital Linear programs could be nailed up in an IP Video distribution system, but it simplifies the analysis to assume that all IP Linear programs are delivered via Switched IP services).

6. Analog Broadcast programming is used to fill out the spectrum after all other services have been allocated their required number of channels. It is assumed that these Analog channels would only carry the basic tier programming content.

7. The number of selectable SDV programs offered by the MSO to their subscribers in the MPEG-TS pool = 250

8. The number of selectable Switched Digital multicast IP programs offered by the MSO to their subscribers in the IP Video over DOCSIS pool = 250

9. Required SDV (or Switched Digital multicast IP) Blocking Probability = 0.01%

10. Assume all programs transmitted over the HFC plant have been converted to MPEG4 encoding (whether used for MPEG-TS delivery or DOCSIS delivery)

11. The mix of MPEG4 programs on the HFC plant might be 50% TV-HD programs at an average of 8 Mbps, 30% TV-SD or PC-SD programs at an average of 1.5 Mbps, and 20% Handheld programs at an average of 300 kbps, which results in an average program bandwidth given by: (0.5)*(8 Mbps) + (0.3)*(1.5 Mbps) + (0.20)*(300 kbps) = 4.51 Mbps. For channel-bonded DOCSIS delivery systems, the bandwidth required per program will be that 4.51 Mbps value. For MPEG-TS delivery systems (where the stat-mux benefits of channel bonding are not provided), the "No-Stat-Mux-Gain" 25% tax must be applied, increasing the effective bandwidth per program to a value of (4.51 Mbps)(1.25) = 5.64 Mbps. (Note: This again includes a mix of traffic resolutions ranging from 8 Mbps for TV-HD to 300 kbps for Handheld video).

12. The initial HFC Channel Map for the "typical" 115-channel plant at the beginning of the transition is defined as shown below:

    Analog Broadcast = 75 channels

    SDV = 33 channels

    VoD = 4 channels

    HSD/VoIP = 3 channels

13. The Fiber Node size is 500 House-Holds Passed (500 HHP)

14. The Video take-rate will remain fixed at 60% of the HHP (500*0.6=300 Video subscribers per Fiber Node)

15. The Percentage of Video subscribers with Digital STBs (or IP Video Gateways) in their homes will remain fixed at 75% (300*0.75=225 subscribers with Digital STBs or IP Video Gateways per Fiber Node) … only these subscribers can access SDV and VoD services

16. The Percentage of Video subscribers who consume only Analog programming will remain fixed at 25%.
17. The HSD/VoIP take-rate will remain fixed at 40% of the HHP (500*0.4=200 HSD/VoIP subscribers per Fiber Node)
18. The Service Group size for each of the Narrowcast service tiers will begin the transition period as shown below:

    SDV Service Group size = 4 Fiber Nodes (300*4*0.75=900 SDV subscribers)

    VoD Service Group size = 4 Fiber Nodes (300*4*0.75=900 VoD subscribers)

    HSD/VoIP Service Group size = 4 Fiber Nodes (200*4=800 HSD/VoIP subscribers)
19. Busy-Hour/Busy-Day Utilization (as a function of SDV subscribers) for SDV Video Services = 60% (Thus there are 900*0.6 = 540 active viewer homes per SDV Service Group)
20. Busy-Hour/Busy-Day # of active SDV viewers per active viewer home = 1.5 (Note that 540*1.5 = 810 active viewers per SDV Service Group)… from a separate SDV usage analysis with program popularity curve defined by the Power Law with alpha=0.7 and with 250 selectable programs, this requires ~225 transmitted programs (as illustrated in **Fig. 2**).

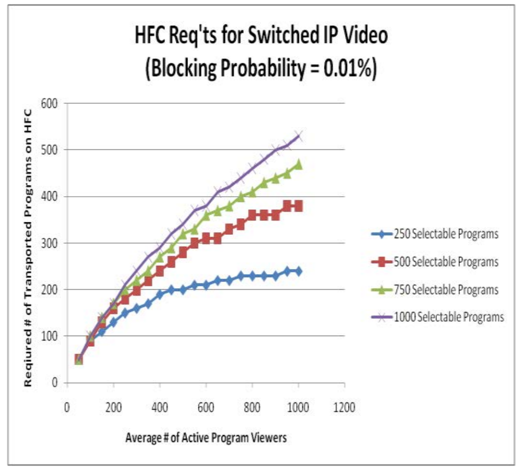

Fig. 2- Required Number of Transported Programs vs. Number of Active Program Viewers

(Note: The shape and height of the curves in **Fig. 2** are very sensitive to the assumptions made on the associated program popularity curves, and field data has indicated that the program popularity curves can vary quite extensively with the demographics of the service group and with the nature of the available programs. Thus, the program popularity curves can change, and the values in **Fig. 2** would also change).

21. For MPEG-TS transmission, if each viewer watches a non-bonded program at 5.64 Mbps, the total bandwidth transmitted = 1269 Mbps total… if each MPEG-TS channel provides 38.8 Mbps of usable bandwidth, this requires 1269/38.8 = 33 channels… For DOCSIS transmission, if each viewer watches a bonded program at 4.51 Mbps, the total bandwidth transmitted is 1015 Mbps total… if each bonded DOCSIS channel provides ~36 Mbps of usable bandwidth, this requires 1015/36 = 29 channels)
22. Busy-Hour/Busy-Day Utilization (as a function of VoD subscribers) for VoD Video Services = 3% (27 active

viewers per VoD Service Group… For MPEG-TS transmission, if each viewer watches a non-bonded program at 5.64 Mbps, the total bandwidth is 152 Mbps… if each non-bonded MPEG-TS channel provides 38.8 Mbps of usable bandwidth, this requires 152/38.8 = 4 channels… For DOCSIS transmission, if each viewer watches a bonded program at 4.51 Mbps, the total bandwidth is 122 Mbps… if each bonded DOCSIS channel provides ~36 Mbps of usable bandwidth, this requires 122/36 = 4 channels)

23. At the beginning of the transition, Busy-Hour/Busy-Day HSD Bandwidth per HSD/VoIP subscriber = 100 kbps (80 Mbps per HSD Service Group… adding 25% head-room, as many MSOs do, yields a need for 100 Mbps of HSD Bandwidth per Service Group… if each DOCSIS channel provides ~36 Mbps of usable bandwidth, this requires 100/36 = 3 channels)

24. DOCSIS HSD Bandwidth demands will increase by 50% every year, so the DOCSIS HSD Bandwidth demands on a yearly basis during the transition to IP Video over DOCSIS are described in **Fig. 3** below for five successive years:

	Avg HSD BW per Sub (kbps)	Avg HSD BW per Serv. Group (Mbps)	Avg HSD BW (w/ 25% headroom) per Serv. Group (Mbps)	# Reqd DOCSIS Channels per Serv. Group
Year 1	100	80	100	3
Year 2	150	120	150	5
Year 3	225	180	225	7
Year 4	338	270	338	10
Year 5	506	405	506	15

Fig. 3- HSD Bandwidth Trends

25. IP Video Gateways can access all of the DOCSIS Downstream Channels associated with IP Video… this eliminates the complications resulting from multiple viewers within a home (behind a single IP Video Gateway) trying to access different Linear multicast streams on different bonding groups when all of the IP Video Gateway tuners are already in use

26. No node-splits will be performed during the period of transition. (Note: This assumption may not be a valid assumption for many MSOs, but it will be valid for some).

27. Extra bandwidth required to support the transition from legacy MPEG-TS video to IP Video over DOCSIS will be obtained by reclaiming channels from the Analog TV service tier.

28. Static, separated application groups will be assumed (i.e.- HSD/VoIP traffic will use one set of DOCSIS channels and IP Video traffic will use another set of DOCSIS channels)

29. Static, separated IP Video tiers will be assumed (i.e.- VoD IP Video traffic will use one set of DOCSIS channels and Switched Linear (SDV) IP Video traffic will use another set of DOCSIS channels)

30. The transition will span a period of five (5) years

Using these assumptions, we can now determine the number of channels required when using the 1-Step Transition Strategy. Using this strategy, the MSO would continue to support its legacy MPEG-TS Video transmission system to its legacy STBs while simultaneously supporting the installation of a new IP Video over DOCSIS transmission system to its DOCSIS IP Video Gateways. As a result, simulcasting of the most popular video streams to both types of end-points is obviously required.

The key attribute of this transition plan is that the MSO will attempt to offer a full array of video services and a complete channel line-up to its customers in both the legacy MPEG-TS camp and the new IP Video over DOCSIS camp. As a result, it is typically assumed that a given home will be either receiving its video feeds through the MPEG-TS Video transmission system (with legacy STBs or Hybrid IP Video Gateway/IP-STB), or the home is receiving its video feeds through the IP Video over DOCSIS transmission system (with IP Video Gateway/IP-STB). In other words, the assumption is that no home utilizes both types of services. However, it is also assumed that a particular Service Group will have to support both types of homes (legacy STB-based homes and IP Video over DOCSIS-based homes). Finally, it is assumed that there will be a gradual transition in the number of homes that support IP Video over DOCSIS, with the IP Video over DOCSIS pool growing from zero homes in the beginning of the transition to 100% of the homes at the end of the transition. Thus, all of the homes that are added to the IP Video over DOCSIS pool are homes that have been removed from the MPEG-TS pool. Traffic Engineering can take advantage of this fact by appropriately adjusting the number of channels to match the expected number of SDV (Switched Linear) and VoD sessions for both legacy MPEG-TS service and IP Video over DOCSIS service. Due to the legacy-to-DOCSIS transition effect, the number of channels required for SDV and VoD sessions within the legacy MPEG-TS service will decrease while the number of channels required for switched video and VoD sessions within the IP Video over DOCSIS service will increase. However, there will typically be an increase in the total number of channels required for SDV (Switched Linear) and VoD programs being sent to the MPEG-TS service and the IP Video over DOCSIS service during the middle years of the transition. This is primarily due to the shape of the SDV curve in **Fig. 2**, which illustrates that the number of channels required to transport SDV services to a pool of subscribers does not rapidly decrease until the number of viewers within a given pool is reduced to a fairly small number. Thus, we need to move most of the MPEG-TS subscribers over to IP Video over DOCSIS service before we can experience a rapid drop in the number of required channels associated with MPEG-TS SDV service.

A Traffic Engineering analysis for the 1-Step Transition Strategy was performed using the simplifying assumptions listed above. This particular analysis assumed that the transition occurred slowly over a five-year period. During each of the latter four years, we assumed that the MSO converted 25% of the digital video homes from MPEG-TS STBs to IP Video Gateways running only the DOCSIS feature capability (no Hybrid MPEG/IP capability). After the fifth year, all of the digital video homes within the Service Group had therefore been converted into IP Video Gateway homes running IP Video over DOCSIS. The tables shown in **Fig. 4** below illustrate the number of channels required for each of the different Narrowcast service types carried over the HFC plant on a year-by-year basis while this 1-Step Transition Strategy is being carried out. The final row (in yellow) shows the total number of required Narrowcast channels, and the “hump” that occurs in years 2, 3,and 4 is quite apparent. It is this “hump” that could cause MSOs to struggle with the Simulcast Roadblock problem.

The “hump” in the last row of **Fig. 4** is a direct result of the need to simultaneously support legacy SDV MPEG-TS channels, legacy VoD MPEG-TS channels, Switched (Linear) IP Video over DOCSIS channels, and VoD IP Video over DOCSIS channels. It is interesting to note that the hump disappears by the end of the transition period, because by then all of the Linear MPEG-TS channels and VoD MPEG-TS channels have been removed from the Service Group.

If the MSO is unable to turn off Analog and/or Digital Broadcast channels during the middle portion of the transition period, then this simulcast tax could become a serious impediment to the deployment of IP Video over DOCSIS. As a result, it might be beneficial to explore techniques for reducing the amount of bandwidth required during the simulcasting period- i.e., techniques for getting "over the hump." These techniques can be beneficial even if simulcasting is not performed, because they can generically be used to help conserve channel bandwidth in an IP Video environment- with or without simulcasting. However, the techniques are definitely helpful if simulcasting is implemented by the MSO.

			Year		
Per Service Group	1	2	3	4	5
# Homes Accessing Legacy MPEG-TS SDV	900	675	450	225	0
# Homes Accessing Legacy MPEG-TS VoD	900	675	450	225	0
# Homes Accessing DOCSIS Switched Video	0	225	450	675	900
# Homes Accessing DOCSIS VoD	0	225	450	675	900
# Homes Accessing DOCSIS HSD	800	800	800	800	800

			Year		
Per Service Group	1	2	3	4	5
Legacy MPEG-TS SDV Bandwidth (Mbps)	1269	1184	1072	733	0
Legacy MPEG-TS VoD Bandwidth (Mbps)	152	114	76	38	0
DOCSIS Switched Video Bandwidth (Mbps)	0	586	857	947	1015
DOCSIS VoD Bandwidth (Mbps)	0	30	61	91	122
DOCSIS HSD Bandwidth (Mbps)	80	120	180	270	405

			Year		
Per Service Group	1	2	3	4	5
Legacy MPEG-TS SDV Bandwidth (# channels)	33	31	28	19	0
Legacy MPEG-TS VoD Bandwidth (# channels)	4	3	2	1	0
DOCSIS Switched Video Bandwidth (# channels)	0	17	24	27	29
DOCSIS VoD Bandwidth (# channels)	0	1	2	3	4
DOCSIS HSD Bandwidth (# channels)	3	5	7	10	15
Total Required Narrowcast Channels (# channels)	40	57	63	60	48

Fig. 4- Traffic Engineering Tables for a 1-Step Transition Strategy

TOOLS FOR SOLVING THE SIMULCAST ROADBLOCK PROBLEM

There are actually several potential tools in the toolkit that can be employed to help reduce the total amount of bandwidth capacity used by IP Video over DOCSIS feeds. Many of these tools would likely be implemented in the CMTS, but some would be implemented in other IP Video sub-systems.

Bandwidth reclamation

Some traditional techniques for salvaging bandwidth capacity on the HFC plant may come in handy when IP Video simulcast issues begin to develop. A common technique that might be employed by many MSOs is the simple bandwidth reclamation process. In this process, channels previously being used for legacy service are retired and the channel is re-injected into the MSO channel map with a new service association. A typical example has some basic tier, legacy analog channels being retired. These retired legacy channels can still be fed into the subscriber homes if they are replaced by digital feeds and DTA deployments within the homes… or the retired legacy analog channels could merely be removed from the program list without replacement, leading to a reduction in the available programming to those analog subscribers). In either case, the retired legacy channels could then be re-assigned to carry DOCSIS IP Video services.

Service Group Size Reductions

Another traditional technique for creating extra bandwidth capacity within a service group is the use of service group size reductions. This process can be carried out using either node segmentation or node splits.

With node segmentation, each of the four coaxial distribution legs emanating from a fiber node is assigned its own set of upstream and downstream channels (instead of sharing a single set of upstream and downstream channels between all four distribution legs). This results in each coaxial distribution leg having a quarter of the subscribers managed by the original fiber node.

With node splits, the homes passed by a single fiber node are re-assigned to two fiber nodes, with each of the new fiber nodes having half of the subscribers as the original fiber node.

In both cases, the number of subscribers sharing a set of upstream and downstream channels is reduced. Since the amount of bandwidth capacity required for narrowcast services (ex: SDV, VoD, HSD) is related to the number of subscribers sharing the bandwidth, these service group size reductions provide an effective means of reducing the bandwidth capacity required to satisfactorily support those narrowcast services. If the number of subscribers sharing a set of narrowcast services is cut in half, then the number of required VoD and HSD channels can roughly be cut in half. If the number of subscribers sharing a set of narrowcast services is cut in half, then the number of required SDV channels can roughly be reduced according to the non-linear curves shown in **Fig. 2**. As a result of the narrowcast channel count reductions that result from the service group size reduction, narrowcast channels are freed up and can be used by the new DOCSIS IP Video services. Unfortunately, SDV and IP multicast achieve their most impressive channel gains for larger groups of users, so a tradeoff exists between reducing the demand for unicast services by reducing the number of subscribers and reducing the efficiency of the SDV model by retreating from the most efficient service group size. One possibility may be to combine nodes for the SDV/IP multicast groups, while narrowcasting the DOCSIS HSD and VoD groups.

Connection Admission Control

Connection admission control (or CAC) is a powerful feature available in many CMTSs that can help ensure a good quality of experience for DOCSIS HSD subscribers and DOCSIS IP Video subscribers whenever the available bandwidth capacity becomes scarce.

Connection admission control is the important functionality that checks how much bandwidth is currently being utilized on a particular Downstream Service Group before a new service request is honored. The connection admission control algorithm compares that bandwidth to a threshold to determine if there is enough spare bandwidth capacity to permit a new service flow to be added to the Downstream Service Group. When the service flow is used to transport the packets for an IP Video stream, then the connection admission control algorithm is actually determining if there is enough spare bandwidth capacity to permit a new IP Video stream to be passed through a particular Downstream Service Group. This functionality ensures that a set of active IP Video streams do not over-subscribe the available bandwidth capacity within a Downstream Service Group. A rejected service flow will manifest itself as a message on the IP Video STB or client indicating that the user should try accessing the channel later.

Intelligent Load-Balancing

Load Balancing is a CMTS feature that identifies when the traffic loads on the Downstream Channels or Downstream Channel Sets are not evenly distributed. Operation with an unequally-distributed traffic load is undesirable, because it can lead to exceptional Quality of Experience levels for subscribers on the lightly-loaded Channels and unacceptable Quality of Experience levels for other subscribers on the heavily-loaded Channels. Experience has shown that it is prudent to re-distribute the traffic loads to create a more equal distribution of traffic loads and a more uniform (and hopefully acceptable) Quality of Experience level for all subscribers.

Once an uneven traffic load is identified by the Load Balancing algorithms within a CMTS, the CMTS can re-distribute the traffic load across the Downstream Channels of Downstream Channels Sets so that heavily-loaded Channels experience somewhat lighter loads, and lightly-loaded Channels experience somewhat heavier loads. This adjustment helps ensure that the bandwidth of all channels is evenly utilized.

Client-Controlled Adaptive Streaming for Unicast Programs

Adaptive Streaming is an interesting approach for trading off between IP Video bandwidth requirements, client device processing requirements, and the Quality of Experience levels for the subscribers. Adaptive Streaming techniques are being developed by most of the Video Delivery solutions on the IP Video market (ex: Adobe, Microsoft, and Apple), and MSOs may also decide to develop their own algorithms within their proprietary clients. All of these Adaptive Streaming solutions tend to recognize the fact that there may be times when a particular resolution video feed being sent to a subscriber is not the ideal resolution at that instant in time.

Sometimes, the video resolution (and associated bit-rate) may need to be temporarily reduced to adjust for heavy network congestion or to adjust for a heavy, transient processing load being placed on the client device. This reduction in video resolution will reduce the bandwidth demands on the shared network resources and will also reduce the processing load on the client device that must render the video. However, the resultant video display will also be lower quality.

At other times, the video resolution (and associated bit-rate) might be allowed to be temporarily increased to react to a lighter level of congestion on the network or a lighter processing load at the client device. This increase in video feed resolution will increase the bandwidth demands on the shared network resources and will also increase the processing load on the client device that must render the video. However, the resultant video display will be higher quality.

Most Adaptive Streaming algorithms are based on the client monitoring the arrival rate of MP4 fragments or the client‟s processing load to trigger changes in the requested stream resolution. The different stream resolutions can be accessed by having the client perform HTTP GETs for MP4 fragments from different content files with different resolutions whenever Adaptive Streaming adjustments are required. As a result, in its traditional form, Adaptive Streaming is best suited for unicast feeds- either unicast VoD feeds or unicast Live feeds. When managed by the clients, it is not well-suited for IP multicast feeds.

<u>Head-End-Controlled Adaptive Streaming for Unicast Programs</u>

While client-controlled adaptive streaming is the most well-known form of adaptive streaming, there is another form of adaptive streaming that may provide even more benefits to MSOs. This alternative form will be called head-end-controlled adaptive streaming within this paper. The goal behind head-end-controlled adaptive streaming is similar to the goal behind client-controlled adaptive streaming: it is a technique for dynamically adjusting the resolution and bit-rate of individual video streams in an effort to ensure that the video streams are successfully delivered in a timely fashion to the video clients, even in the presence of network congestion. Head-end-controlled adaptive streaming does not consider or adjust resolutions and bit-rates as a result of changes in the processing loads placed on the clients. In essence, it assumes that the clients have adequate processing resources to render the video streams that they have requested.

Head-end-controlled adaptive streaming requires an omniscient control element in the MSO head-end to monitor and manage the bandwidth for each of the video flows being transmitted over the HFC plant. This omniscient control element could instantiated in one network element or a combination of network elements such as the CMTS, a policy server, an application manager, an IP Session Resource Manager, or some other intelligent node in the MSO head-end. In general, the omniscient control element should be cognizant of several things that could impact IP video Quality of Experience, including:

1) The topology of the HFC plant (including which DOCSIS channels are accessible by which video subscribers)
2) The available content and utilization levels of each of the IP Video servers
3) The real-time congestion levels on the HFC plant
4) The attributes of each video subscriber (such as the maximum processing capability, the video screen size, the subscription level, and the priority)
5) The policy rules that the MSO would like to enforce during periods of congestion (such as which subscribers or which screen sizes should be adaptively throttled first)
6) The triggering rules that the MSO would like to use to initiate the adaptive throttling of video stream resolutions and bit-rates

All dynamic streaming client devices feature a sizeable play-out buffer which “buys” time that may be needed for effective, pro-active intervention on the flow of adaptive streams by the head-end-control element. The size of these buffers vary from 7 to 70 seconds across the available HTTP delivery schemas, and some of this buffer time can be used to remedy any delay issues caused by the use of a head-end-based control element.

Once the need for throttling is identified by the omniscient control element, there are many ways to manage the throttling. For example, the omniscient control element in the head-end could communicate with either the video client or the video server to initiate the throttling functions.

Omniscience and the enforcement of intelligent policy rules and triggering rules are the keys to making a head-end-controlled adaptive streaming system work well. When using head-end-controlled adaptive streaming, the ability to recognize congestion, the ability to identify the causes of the congestion, and the ability to determine how to intelligently throttle a sub-set of the video streams should lead to improved Quality of Experience and improved bandwidth utilization within the end-to-end IP Video solution.

VBR Channel Bonding with Stat-Mux Gains

The effect of using bonded VBR streams has already been mentioned within the traffic engineering analysis above, but its value is worth mentioning again. If MSOs are willing to use VBR encoding for the future video streams and are also willing to send those video streams across a channel-bonded environment with at least three or four bonded channels, then they will experience an increase in the total number of programs that they can transport across their HFC plant (when compared to the number of video streams that could be transported if CBR encoding and un-bonded channels were utilized). [Bug1]

As an example, simulations with a specific set of twenty-one 20 Mbps, MPEG4 video streams have shown that the move from CBR encoding on non-bonded channels to VBR encoding on non-bonded channels helped reduce the total number of required channels from 11 channels to 7 channels. The move from VBR encoding on non-bonded channels to VBR encoding on bonded channels helped to further reduce the total number of required channels from 7 channels to 4 channels. These numbers included some "breakage effects," whereby the high-bandwidth, 20 Mbps video streams did not nicely fill the available bandwidth of the channels. But nevertheless, the overall effect is still quite significant.

What further improves the performance of VBR in an adaptive streaming environment is the buffering provided inside of the IP video client device. The buffers bring a new dimension of elasticity into content streaming. The duration of delivery of each fragment can be shifted, „stretched-out‟ or accelerated by the throttling mechanism during periods of congestion, and as long as the buffer is not "starved" by this process, all video sessions will be uninterrupted.

Fig. 5 and **Fig. 6** illustrate how the same fragments are being successfully delivered to the client device in both uncongested and congested downstream channels. The packet transfer curves may look differently in these two instances but the same fragments are successfully delivered to the client devices in time for smooth playout.

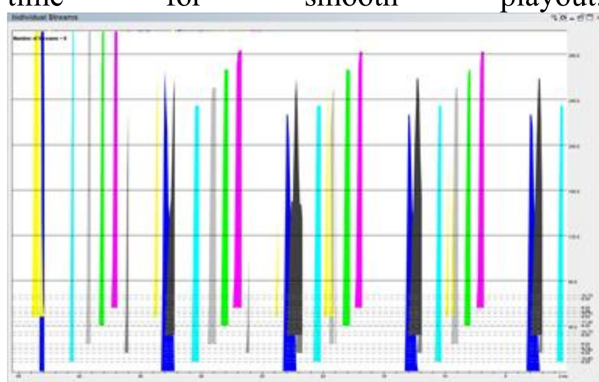

Fig. 5- Streaming of Fragments in an Uncongested Network

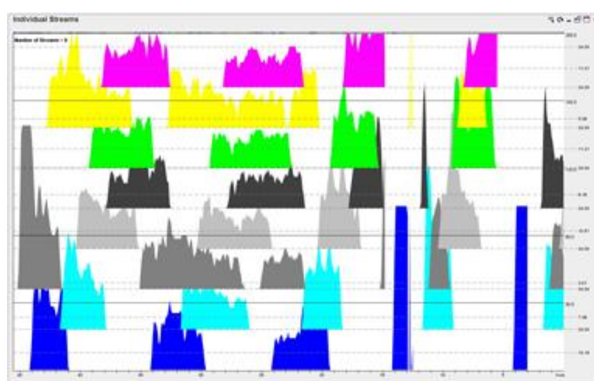

Fig. 6- Streaming of Fragments in a Congested Network

It is therefore possible to achieve a very efficient usage ratio on the HFC bandwidth (**Fig. 7**), measured by the Peak to Average Ration (PAR) of the composite (sum) of all streams. Such a high efficiency is not possible with the legacy MPEG2TS set-tops, because legacy MPEG2-TS set-tops do not tolerate much delay, jitter or wander within the stream without performing expensive, time consuming, quality-affecting transcoding for stat-muxing (typically 3 or 4 into one) at the head-end.

Our studies showed that a simple rule can be devised to calculate how many fragmented VBR streams can be reliably sent to the serving area over bonded DOCSIS3.0 channels. For calculations, the Average Bit Rate value must known for each of the streams. This value is typically set in the encoder for the live streams, and the actual average bit rate value can be easily calculated for VoD streams that are stored in servers. The Actual Average Bit Rate value of any VoD asset (called ActBR) can be derived by dividing the VoD asset size (in MB) by the duration of the asset (in seconds). The result needs to be then converted to Mbps (multiplying by 8) and is comparable to the ABR numbers of live streams.

The simple rule for calculating the required bandwidth inside of a <u>quad-bonded downstream</u> pipe (as shown on **Fig. 7**) is as follows:

(*Sum of ABR values of all live streams + sum of ActBR of all VoD streams) x 1.1*.

The 1.1 overhead factor, or 10% "tax" is a result of many simulations using the actual MSO content and allows the system to support fast „trick-play‟ on a limited number of devices, and also allows the system to support fast admission/start of several new IP video sessions.

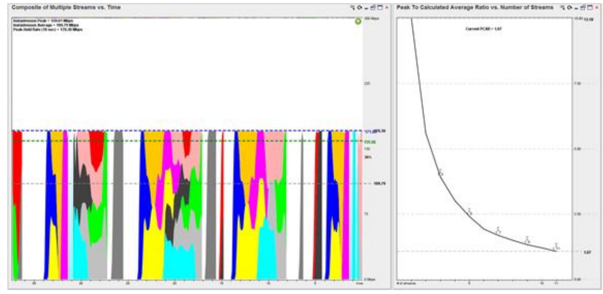

Fig. 7- Composite bit rate and Peak to Actual Average (decreasing as more Streams are added and combined) inside a Fully Loaded Quad-bonded Downstream

For a set of <u>octal-bonded downstreams</u>, simulations showed that the overhead factor ("tax") can be lowered to 1.05 (5%) as effects of self-averaging of VBR streams are more pronounced in the "fatter" delivery pipe.

<u>Hybrid Media Gateways with Both QAM and DOCSIS Receivers</u>

If the in-home gateway that receives the IP Video over DOCSIS signals has enhanced functionality that permits it to simultaneously receive both DOCSIS signals and MPEG-TS signals, then that enhanced gateway offers MSOs many new opportunities to deal with the Simulcast Roadblock issue.

One technique for reducing the number of simulcasted channels between the MPEG-TS tier and the DOCSIS tier is to only simulcast a fraction of the programs within the DOCSIS tier until all of the homes in a service group have been equipped with media gateways. For example, during the transition period, an MSO may choose to only transmit VoD streams and special interest Switched IP programs (ex: foreign language programming) to their media gateway subscribers over the DOCSIS feed. Those VoD streams and special interest Switched IP programs would not be sent over the MPEG-TS feed. As a result, the special interest Switched IP programs would not even be available to subscribers who still have legacy STBs that only receive MPEG-TS feeds. This could be used to encourage subscribers to sign up and pay an extra fee to get the new media gateway access which provides access to the special interest programming. Using this technique, the broadcast and SDV programs would only be moved to the DOCSIS tier when all subscribers in a service group were equipped with media gateway boxes, and the programs would no longer have to be transmitted on the MPEG-TS tier after that transition. It is possible and even probable that MSOs may actually leave a small set of broadcast or SDV programs in the MPEG-TS tier after the transition to support legacy STBs or DTAs, or to honor content or franchise contracts still requiring some analog or clear QAM channel transmissions. However, the availability of a hybrid media gateway can be used to greatly reduce the amount of simulcasting that occurs during the transition period.

Media Gateways with Large Numbers of DOCSIS Receivers

A subtle, but important problem that can lead to an increased demand for IP Video bandwidth can result if IP Video unicast feeds (often used for VoD service) and IP Video multicast feeds (often used for Switched IP service) are distributed across the channels in a Downstream Channel Set in an undesirable fashion. To give an example, consider the scenario shown in **Fig. 8**.

The IP Video streams are distributed across four Downstream Channels, but each media gateway has only two channels dedicated to accessing the IP Video streams. Media gateway x is accessing multicast streams A and B on Downstream Channels 1 and 2, respectively, while media gateway y is accessing multicast streams C and D on Downstream Channels 3 and 4, respectively.

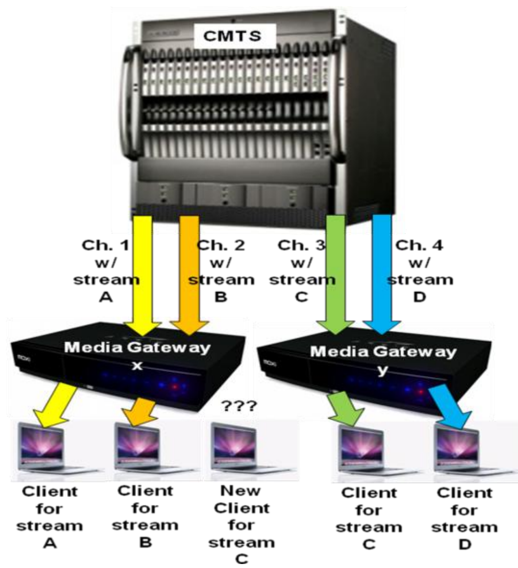

Fig. 8- Scenario with Multiple Video Multicast Streams on Different Channels

If a new video client behind the first media gateway (x) were to request the opportunity to Join and view multicast stream C, then the CMTS would have to perform some rearrangements to give media gateway x access to Stream C. The CMTS could (for example) create another replication of Stream C on either Downstream Channel 1 or Downstream Channel 2 to make it readily accessible to media gateway x, but this approach would tend to waste bandwidth on the HFC plant by transmitting multiple replications of the same Stream C content on the HFC plant. Alternatively, the CMTS could use DCC or DBC messages to force the media gateway to channel change its first receiver from Downstream Channel 1 to Downstream Channel 3 and then move already-flowing multicast video stream A to Downstream Channel 3 where the desired multicast stream C is found. However this channel change and flow movement could create a temporary disruption to the in-process video rendering of the video stream A. In addition, the movement of video stream A from Downstream Channel 1 to Downstream Channel 3 could have a domino effect if other media gateways were also receiving multicast video stream A, because their receivers would also be forced to change channels, and any flows besides video stream A that are being carried over Downstream Channel 1 to one of those media gateways would also have to be moved. Thus, many layers of problems can be identified with this proposed movement of stream A.

All of the problems mentioned in the previous paragraph can be easily circumvented if all of the media gateways were to be designed with more receivers. Broadband or full-band capture capabilities are becoming more accessible with the advent of higher-speed digital-to-analog converters, so many media gateway devices will be increasing the number of received channels in the future. In a perfect world, the media gateways would always have enough receivers to be able to tune to the number of DOCSIS Downstream Channels (IP video and HSD) that might be accessed by a typical home during the busy hour (which would typically be less than the total number of DOCSIS channels). As a result, it is clear that media gateways with larger numbers of DOCSIS receivers can help reduce multicast replications and can also help reduce the required bandwidth for IP video on the HFC plant.

Intelligent Assignment of Multicast Linear Programs to Specific Channel Sets

The problem outlined in the previous section illustrates how multicast video streams can cause exacerbated HFC bandwidth issues whenever extra replications of the multicast stream must be created to provide access to media gateways that may not be tuned to the channel on which the multicast video stream is being transmitted. As illustrated in the previous section, increasing the number of receivers on the media gateways can help to minimize this problem. However, if media gateways or modems are used that do not have these increased receiver counts, then other solutions must be found.

Another solution that could help minimize this extra multicast replication problem is described in this section. Basically, MSOs might decrease the number of extra multicast replications (and the amount of required HFC bandwidth) if they assign their most popular multicast streams to a specific set of channels in the HFC plant. In fact, it might be ideal to actually limit the most popular multicast streams to being placed on single, non-bonded service flows, while less popular multicast streams can be placed on bonded service flows to capitalize on statistical multiplexing gains. The theory behind this approach is that the popular Linear multicast programs are likely to be viewed by many subscribers, so if all of those programs are available on the same set of channels, then there is a high probability that a media gateway that is tuned to that set of channels to receive one popular multicast stream might also be asked (by another client in the home) to collect another popular multicast stream. If that additional multicast stream happens to appear on the same channel set as the first multicast stream that is already being received within the home, then extra replications are unlikely to be required.

As an example, the scenario shown in **Fig. 8** would not have been problematic if all of the program streams (A, B, C, and D) were all available on Channel 1, because both media gateways would have had the capability to be tuned to Channel 1. When the new client behind media gateway x requested access to program stream C, media gateway x would have easily accessed it on Channel 1 with the popular programs all assigned to Channel 1.

Intelligent Mixing of Compressible Traffic with Non-Compressible Multicast Linear Traffic

In most of the early-stage IP video deployment scenarios currently being planned by MSOs, the multicast Linear video streams may not be adjusting their resolutions and bit-rates the way that unicast video streams might. (Note: Unicast video streams are expected to capitalize on techniques such as adaptive streaming right away, but complications in the delivery of multicast video streams may preclude the use of adaptive streaming in early deployments. As a result, the multicast streams might be considered to be “incompressible” streams during these early deployments). As a result, channels carrying only multicast Linear streams would not be able to throttle down the bit-rate on the different streams if the aggregate bit-rate from the multiple multicast streams ever burst to higher-than-normal levels.

In an effort to provide some ability to absorb and adapt to the transient bit-rate bursts that might develop on a channel carrying multiple multicast Linear streams, there may be a benefit to including some “compressible” streams within the channel. Compressible streams on a channel can include:

1) Unicast IP video streams with very deep receiver buffers
2) unicast IP video streams with adaptive streaming capabilities enabled
3) unicast HSD streams whose data can be delayed to lower the overall bit-rate.

Simulations were created to explore the effectiveness of this technique. In particular, a high-utilization (93.6%), 4-channel bonding group was simulated with different mixes of multicast Linear IP video traffic and unicast IP video traffic. The buffers for the unicast IP video traffic were designed to hold up to 10 seconds of video and the buffers for the multicast IP video traffic were designed to hold only 500 milliseconds of video. The percentage of errored video seconds was monitored, where an error occurred within a given second if a video packet was not available in the receiver‟s buffer at the time the packet was needed for rendering. The output of this simulation is shown in **Fig. 9**. It can be seen that with 15% (or more) of the video program bandwidth being unicast, the percentage of errored video dropped to negligible levels. Thus, it is apparent that this approach can be quite effective.

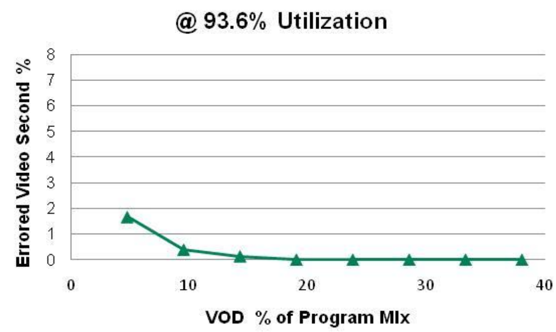

Fig. 9- Effect of Video Program Mix on % Errored Seconds

<u>Acceptable Buffering of Unicast Video Streams at the Client</u>

The examples in the previous section illustrate the generic benefits of using deep packet buffers in the video client. These buffers can act as “shock absorbers,” providing IP video packets to the rendering engine whenever needed. If the buffer depth is made large enough, then the IP video delivery system can flywheel through transient periods of congestion leading to delayed packets and dropped packets (whose re-transmissions are delayed).

Low-bandwidth recovery mechanisms for corrupted multicast Linear programs

Linear multicast IP video streams will often be terminated at high-resolution TVs, so video impairments due to lost packets would undoubtedly reduce the viewer Quality of Experience level. Unfortunately, the use of multicast will typically require the Linear video streams to be transported using UDP instead of TCP, so re-transmission of lost packets (if implemented at all) will have to be handled by an upper-layer protocol.

If an upper-layer protocol packet retransmission scheme is developed, then it could be problematic, because the loss of a single multicast packet on the HFC plant prior to the arrival at the multiple receivers who have joined the multicast would cause each and every one of those receivers to initiate a separate re-transmission of the packet. These semi-synchronized requests for re-transmission could lead to a semi-synchronized burst of re-transmitted packets, which could cause a transient bandwidth spike on the HFC plant, which would be undesirable.

Several techniques have been proposed to reduce this problem. One technique requires the receivers to use a randomized delay before sending their re-transmission requests, which should serve to spread out the re-transmissions. Another technique requires the head-end server to recognize the first request for a re-transmission and to re-transmit the requested packet using multicast so that all of the receivers would receive the single re-transmitted packet. The other requests for re-transmission of the same packet would be ignored, since it can be assumed that they would have been satisfied by the single re-transmission.

Temporal Spreading of DVR Program Tuning Changes at the Half-Hour Epochs

Whenever a program tuning change occurs, there are often bursts of video packet transmissions that occur as each client attempts to pre-load its receive buffer. At half-hour epochs, many program tuning changes can occur in a short window of time, causing the video packet transmission bursts to become semi-synchronized. Unfortunately, DVR devices may produce a large percentage of the semi-synchronized program tuning changes at the half-hour epochs. One way to help minimize this transient problem is to ensure that DVR boxes managed by the MSO use a randomized delay before sending their tuning requests, which should serve to spread out the bursts at the half-hour epochs.

Intelligent Pre-Loading of Media Gateway Memory with VoD Content That is Statistically Likely to Be Viewed

HSD channels and IP video channels will likely cycle through periods of high utilization and low utilization throughout every day, and in most cases, the peak periods of high utilization will typically be at the same time for both the HSD channels and the IP video channels. (Note: This peak period usually occurs between 7pm and 10pm). The real HFC bandwidth challenge for MSOs will usually occur during the periods of high utilization, so most of the techniques for reducing the channel bandwidths will be aimed at performing their duties during the peak utilization periods.

Obviously, this implies that other types of activities can be supported during the low utilization periods. Perhaps techniques that help reduce the bandwidth during the peak utilization periods can take advantage of the extra bandwidth available during the low utilization periods.

One possible approach might attempt to pre-load VoD video content into DVR memories within the media gateway during the low utilization periods so that the VoD would not need to be transmitted over the HFC plant during the peak utilization periods. The trick is to try to predict which particular VoD-based video content might be viewed in the near future by a particular home. Information such as the successive episodes of a series previously watched in VoD can provide hints that help identify likely VoD candidates for the future.

Intelligent Pre-Loading of Media Gateway Memory with Linear Content That is Statistically Likely to Be Viewed

The content pre-loading methods described in the previous section can be applied (with slight variations) for other applications. For example, many MSOs are interested in ensuring that their IP video subscribers will have an above-average viewing experience. This will typically require support for features like:

1) Fast Linear feed channel changes (when the viewer tunes from one multicast Linear program feed to another multicast Linear program feed)
2) Start-over features that permit the viewer to re-start a Linear program feed if they join the program within a half hour (or so) of the feed"s starting time
3) Trick-mode features that permit the viewer to pause, rewind, and fast-forward (up to the present time) a Linear program feed

There are many ways to provide the features listed above, but one common technique involves the launching of unicast IP video feeds to each viewer requesting the use of those features. This launching of separate unicast IP video feeds to each viewer can lead to a significant increase in the bandwidth capacity required for IP video delivery over the HFC plant. This would obviously exacerbate the Simulcast Roadblock problem. As a result, finding alternatives to the launching of separate unicast IP video feeds would be beneficial.

One alternative method requires the media gateway to pre-cache the Linear video streams as they are distributed via efficient multicast IP video feeds, so that the content is locally available in the home whenever the viewer initiates a channel change or start-over or trick-mode operation. This pre-caching obviates the need for the launching of additional unicast IP video feeds. Depending on the storage technologies (ex: DRAM, Flash, rotating disk) used in the different models of media gateways, there may be different memory speeds (which dictate the number of IP video streams that can be simultaneously written in parallel), different endurances (repeatable write cycles, which dictate the lifetimes for the media gateways), and different capacities (which dictate the total number of IP video streams that can be stored).

Since there will be these finite storage limits, the most challenging part of pre-caching multicast Linear IP video streams is determining which Linear streams to pre-cache and when to pre-cache them.

For example, a simplistic design might attempt to pre-cache every multicast Linear IP video feed that is passing down to the service group, but this design could suffer from many problems. Accessing all multicast Linear IP video feeds could place costly requirements on the number of tuners that must be supported in the media gateway. It could also place costly requirements on the speed and endurance and size of the storage technology that would be holding all of those streams. These added costs may not be acceptable to the MSO.

Thus, a carefully-planned design may be required for the pre-caching algorithms used in media gateways. As an example, one might consider a design that only pre-caches Linear IP video feeds during periods of time when HFC bandwidth congestion is expected, because use of pre-caching at other times may not be beneficial. There may be more than enough spare bandwidth to permit the use of unicast IP video feeds to provide fast channel changes, start-overs, and trick-modes during those periods of low HFC bandwidth congestion.

A carefully-planned design may also use history-based heuristics within each media gateway to try to predict the Linear IP video feeds that are likely to be accessed within each half-hour window. These heuristic algorithms might (for example) search back through the records to identify the particular Linear IP video feeds that were viewed by that home during the same half-hour window of time in previous weeks to try to predict the Linear IP video feeds that might be viewed during the same half-hour window this week. If that Linear IP video feed is requested by other subscribers within a service group, then it probably makes sense for the media gateway to pre-cache that feed.

Data analysis on real-world viewing habits was performed in an effort to determine the efficacy of heuristics-based algorithms of this nature. One algorithm used the most popular programs watched by this particular home in the same half-hour window 7 days ago and 14 days ago. This algorithm then added the most popular programs watched by the entire service group in the same half-hour window 7 days ago and 14 days ago. The results (describing the cache miss ratio) for media gateway caches of different sizes are illustrated in **Fig. 10**.

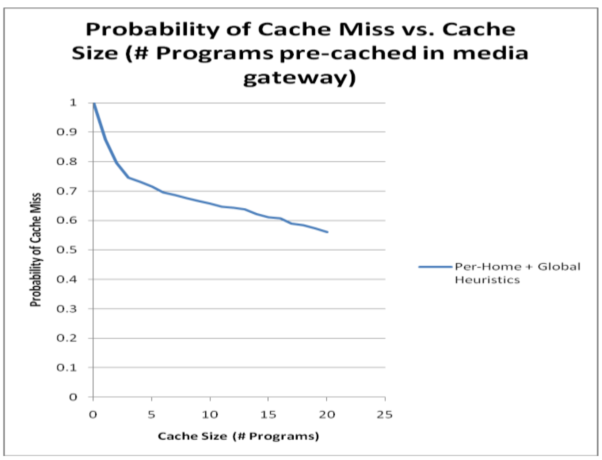

Fig. 10- Probability of Cache Miss vs. Cache Size (# Programs in media gateway)

As can be seen from the plot, these heuristics-based algorithms can give some assistance in the attempt to predict the Linear IP video programs that will be viewed in a particular half-hour window. If the media gateway can pre-cache up to 20 programs and the programs are selected using the aforementioned heuristics-based algorithm, then the data analysis predicts that the pre-cache programs would yield a 55% cache miss ration (=45% cache hit ratio). If the programs that are viewed during a particular half-hour period within a home are already pre-cached in the media gateway 45% of the time, then there is a good chance that channel change, start-over, and trick-mode events will find their desired content already stored in the gateway about 45% of the time. This could greatly help minimize the Simulcast Roadblock problem, because the use of pre-caching will reduce the total amount of required HFC bandwidth. In general, the additional unicast IP video feeds that would be launched if the content was not pre-cached in the gateway would not be launched if the pre-caching is performed in the media gateways.

CONCLUSIONS

In the first half of this paper, we outlined the various transition paths that MSOs might follow as they begin to migrate their networks to support new IP video services. The paper has also described and quantified (via an example) some of the challenges that MSOs may face as they try to deploy both legacy video service and IP video service on a single HFC plant infrastructure. In general, MSOs will need to find intelligent ways to navigate through bandwidth challenges that may face them as they begin to deploy both types of services. A particular problem known as the "Simulcast Roadblock" was defined to be the HFC bandwidth challenges that may result if MSOs try to transmit the same video streams in both their legacy MPEG-TS delivery system and their new IP video delivery system.

In the second half of the paper, we outlined a lengthy list of tools that MSOs can use to help them reduce the amount of HFC bandwidth required as they begin to deploy IP video services. These tools can be used to help solve the "Simulcast Roadblock" problem or they can be used in a general fashion to simply reduce the amount of bandwidth required to deliver a given subset of IP video program feeds to a service group.

These tools can be used in isolation or they can be used together to create a blended end-to-end solution for the IP video delivery problem. It is unlikely that any MSO will choose to use all of the tools mentioned in the second half of this paper, but MSOs may choose to use a sub-set that best suits their needs and the constraints created by their own HFC plant.

It is hoped that MSOs will find this toolkit to be a valuable resource as they begin their transition towards the deployment of IP video delivery systems in the upcoming future.

REFERENCES

[Bug1] M. Bugajski, “Use of Variable Bit Rate Video in Digital CATV Interactive Services,” in Broadband, Vol. 31, No. 2, August, 2009.

[Clo1] Cloonan, Tom, “Bullpen: DIBA Reconsidered,” Communications Technology, January 1, 2009, http://www.cable360.net/ct/strategy/businesscases/33178.html.

[Ins1] “The US Market for OTT Video Services”, Instat Research, August 2010

[Pin1] Pinson, Margaret, et. al., “HDTV Subjective Quality of H.264 vs. MPEG2, with and without Packet Loss,” http://www.its.bldrdoc.gov/pub/n3/video/ieee_10.pdf.