ADAPTING ADAPTIVE STREAMING TO CABLE ACCESS

Xiaomei Liu
Cisco Systems

Abstract

Cable operators are increasingly embracing a new wave of video delivery to extend video services to multiple screens and off-net subscribers. The latest and the most important technology breakthrough to enable this new paradigm of video delivery is adaptive streaming. With wide support on client devices, adaptive streaming holds the promise of a unified video service delivery architecture for unmanaged devices and off-net subscribers.

Adaptive streaming successfully tackles the challenges of delivering video services over unmanaged networks. However, delivering video services using adaptive streaming at scale in managed cable networks is demanding.

This paper examines various technical challenges of adaptive streaming in managed cable access networks and then explores a variety of optimization options. It proposes upstream optimization with TCP ACK QoS prioritization and TCP ACK suppression in addition to simple upstream bandwidth expansion. For downstream optimization, the coupling of VBR video and adaptive streaming is proposed to reduce downstream bandwidth consumption and to improve video quality significantly. Finally, a flexible downstream QoS design with bandwidth protection, stream prioritization and video quality optimization is introduced.

INTRODUCTION

The popularity of Internet video streaming and over-the-top (OTT) video services shows no signs of abating. Cisco's Visual Network Index research predicts that 91% of all consumer IP traffic will be video traffic by 2014 [1]. Netflix, a leading OTT video streaming service provider, reached a milestone of 20 million subscribers at the end of 2010. Statistics indicate that Netflix video streaming traffic already represents more than 20 percent of total downstream traffic during peak times in the United States [2]. The initial success of OTT video providers, such as Netflix and Hulu, not only supports the business case of online video but also proves the feasibility of the underlying technologies.

Behind the scene of this undeniable success of streaming video lies the technical foundation: Adaptive Bit Rate (ABR) Streaming. Despite implementation variations from major ABR streaming vendors, such as Microsoft, Adobe and Apple, the architecture essentials of ABR streaming are the same. A brief description of adaptive streaming architecture is given in the following paragraphs while a thorough introduction is referenced [3].

An ABR video delivery system consists of servers, networks and clients as depicted in Figure 1. At the server side, a video asset is encoded at multiple video quality levels using different bitrates. The higher the encoding bitrate of an asset, the better the video quality. Video content encoded at a particular bitrate profile is further segmented into small fragments. Each fragment corresponds to a few seconds of video playtime. There is also a manifest file for each ABR video asset. The manifest file stores metadata of the asset, such as bitrate information and fragmentation boundary information. ABR servers include processing and packaging servers and origin servers. The processing and packaging servers convert linear and on-demand video input to video fragments and generate manifest files. The origin servers host the video fragments and manifest files. To support linear video delivery, the processing servers and packaging servers continuously ingest video content and output packaged fragments to origin servers. At the same time, the manifest files for linear streams are updated continuously.

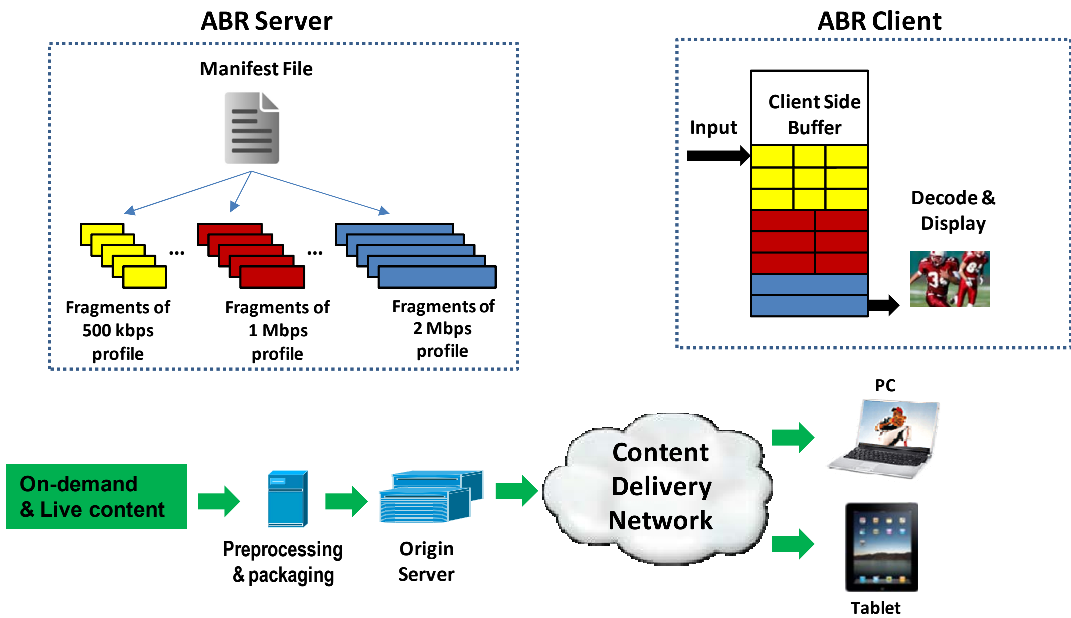

Figure 1. Adaptive Bit Rate Streaming

Since ABR was designed to support video delivery over unmanaged networks, ABR streaming presumes best effort networks without bandwidth guarantee and QoS protection throughout the ABR delivery process. Large scale deployments of ABR streaming typically employ Content Delivery Network (CDN) to reach clients across far reaching locations. Instead of having centralized ABR streaming servers to directly serve ABR clients, client requests are served by distributed CDN edge cache servers. The most popular content is cached at the CDN edge cache servers and can be served to clients immediately upon request. If content requested by a client is not in the cache of CDN edge servers, the CDN obtains the content from an origin server and then delivers the content to the requesting client. Given that ABR streaming utilizes HTTP for the delivery of video fragments, the standard HTTP caching capability of CDN can be readily leveraged without imposing special requirements on the CDN. The CDN system greatly reduces origin server loading while at the same time minimizing the bandwidth demands on regional and aggregation networks.

At the client side, an ABR client first obtains the manifest file of a video asset and then requests video content from streaming servers or CDN on a fragment by fragment basis. Each fragment is obtained via HTTP GET with the URL uniquely identifying the requested fragment. Depending on implementation, byte-range may also be used for fragment request. When the client requests a fragment, the client's desired bitrate profile is provided to the server. Based on multiple factors, such as client screen resolution, CPU load, power consumption, and network conditions, the client can adjust the bitrate profile requested to the server in real-time by sending a new profile request to the server. For the rest of the paper, the focus is on the bitrate adaptation to network conditions.

How does ABR streaming solve problems of video delivery over unmanaged or congested networks? The two major roadblocks to successful video delivery in unmanaged networks are unreliable packet delivery and lack of bandwidth guarantee for video traffic. Due to the fact that video content is highly compressed, any packet drop in a delivery network greatly impacts the end user experience. Common encoding technologies, such as MPEG, improve coding efficiency by removing temporal and spatial redundancy of video sequences. A video sequence is divided into groups of pictures. Each group has a reference frame and several other pictures that have decoding dependencies on the reference frame. A single packet drop can therefore impact a group of pictures if the dropped packet happens to be on the reference video frame. ABR streaming solves the unreliable network problem by first delivering video packets over the TCP protocol instead of the UDP protocol. TCP protocol is a reliable transport protocol that automatically repairs packet loss through retransmission. As a result, the video application layer benefits from a reliable transport and improved packet loss resilience.

In order to solve the second problem of non-guaranteed network bandwidth, ABR streaming applies an innovative approach dealing with network bandwidth variations. When networks experience congestion, an ABR client down-shifts the bitrate profile of requested video fragments to align with less available network bandwidth. Similarly, when the network load is light, the ABR client upshifts the bitrate profile of requested fragments to improve video quality. Video quality at the client side is self-adapting to network conditions such that end users can obtain the best video quality under the constraint of network resources.

An ABR client maintains a large client side buffer (e.g. 10-30 seconds). Network conditions are derived by monitoring the client side buffer fullness and the achieved throughput. Low buffer fullness implies network congestion while high buffer fullness suggests sufficient network bandwidth. To ensure good end user experience, the ABR client switches seamlessly from one fragment to another fragment even when the two fragments have different bitrate profiles. ABR video encoding guarantees fragment coding independency to support this seamless transition. This bitrate adaptation mechanism and smooth transition design cope with network congestion effectively and make video delivery over unmanaged network possible.

Just as important is the prevalence of ABR clients on Customer Premise Equipment (CPE) devices in consumer markets. With industry heavyweights Microsoft, Adobe and Apple all supporting ABR streaming, ABR streaming has become the technology of choice for many operators to reach unmanaged consumer devices.

In short, key advantages of ABR streaming are:

- TCP transport and reliable video delivery
- Bitrate and video quality self-adaptation to network conditions
- HTTP and efficient content caching with CDN
- Wide availability of ABR clients

ADAPTIVE STREAMING CHALLENGES IN CABLE ACCESS

Cable operators are under severe competitive pressure as consumers are choosing to view premium content on devices outside of the managed and limited TV experience. Cable operators are moving rapidly in response to these pressures and have begun to leverage the ABR streaming technology to extend video services to unmanaged CPE devices and to off-net subscribers (i.e. subscribers outside provider networks).

Despite the success of ABR streaming in OTT environments, deploying managed ABR based streaming video services at scale in provider networks has a few challenges. The first challenge is bandwidth consumption. ABR streaming may produce multiple bandwidth pressure points throughout cable networks, but the bandwidth bottleneck is most likely to be the cable last mile. In the downstream direction, the TCP based ABR streaming requires unicast transmission, so last mile bandwidth savings derived from efficient multicast delivery is not yet available. Delivering managed linear video services to ABR clients directly via unicast in cable access networks is not attractive, at least from a bandwidth perspective. Optimization of linear video delivery to ABR clients is an active area of research and innovations in the industry.

In the upstream direction, the upstream bandwidth required for ABR video delivery should not be overlooked. The dominating contributing factor of upstream traffic for ABR streaming is TCP Acknowledgement packets (ACKs). With delayed TCP ACK implementation, every two TCP packets downstream must have a TCP ACK packet upstream. This standard-based TCP behavior is defined by IETF RFC 2581. Assuming 60-byte TCP ACK packet size and 1500-byte downstream TCP video packet size, the TCP ACK overhead translates to approximately 2% of downstream bandwidth. For instance, if the total downstream bandwidth consumed by ABR video traffic is 300 Mbps, then 6 Mbps upstream bandwidth is required. A two-way delivery model is replacing the well-known one-way delivery model used in traditional cable video systems.

When one considers that hundreds of subscribers share the last mile upstream bandwidth of 27 Mbps (e.g. 64 QAM Annex B upstream) in a typical cable network today, this TCP ACK overhead is significant. Compounding the problem, cable plants are asymmetric and the upstream spectrum is extremely limited. DOCSIS 3.0 only supports upstream spectrum from 5MHz to 85MHz (about 300 Mbps per service group capacity) while downstream can reach 1 GHz (over 5 Gbps per service group capacity). Though the cable industry is actively exploring solutions to expand upstream capacity [4], the plant upgrade cost, the CPE cost and industry standardization process will limit the upstream capacity for the foreseeable future. Meanwhile, new upstream-intensive applications such as consumer telepresence, home monitoring and automation will significantly strain the DOCSIS upstream path.

In addition to the bandwidth requirements, an equally challenging fact is that ABR streaming video quality is sensitive to network congestion in both upstream and downstream directions. It is not surprising that downstream congestion causes ABR streams to downshift to lower bitrates and to reduce video quality. What about the impact of upstream congestion? Not so obvious is the effect of upstream congestion on ABR video quality. Even when the downstream has sufficient bandwidth to accommodate higher bitrate profiles of an ABR stream, any upstream congestion in the network may reduce TCP throughput and cause the ABR stream to downshift to a lower video bitrate profile and result in inferior video quality. The vulnerability of ABR streams to upstream congestion directly impacts end user experience and reduces bandwidth efficiency of cable access in the downstream direction.

UPSTREAM OPTIMIZATION

To improve upstream transport for ABR streaming, a multi-pronged approach is presented: leveraging upstream channel bonding to increase upstream bandwidth capacity, applying upstream Quality of Service (QoS) to prioritize ABR TCP ACK packets and enabling TCP ACK suppression to eliminate unnecessary TCP ACK packets.

Upstream Capacity Expansion

For deployments of managed ABR video streaming services in cable networks, DOCSIS upstream capacity must be carefully planned to accommodate the additional upstream bandwidth requirement for TCP ACKs. DOCSIS 3.0 introduces both upstream channel bonding and downstream channel bonding. Without channel bonding, an upstream channel capacity is limited to 27 Mbps using Annex B QAM64 modulation. With the latest cable modem channel bonding technology, up to four upstream channels can be bonded together to form a larger upstream pipe. The upstream capacity is expanded to over 100 Mbps, quadrupling the original upstream capacity.

Upstream Quality of Service (QoS)

The sensitivity of ABR video quality to upstream congestion makes upstream QoS critical to ABR streaming video. Given ample downstream bandwidth, operators desire to obtain the best ABR streaming video quality and minimize the impact of upstream congestion to video quality. The proposed design gives upstream TCP ACKs of ABR streaming higher priority over non real-time upstream traffic by applying standard based DOCSIS QoS. With DOCSIS, different priorities can be applied to service flows. TCP ACKs from ABR video can be classified to higher priority service flows to receive differentiated delivery service in the upstream direction.

Proof of concept work was conducted to evaluate the impact of QoS on ABR streaming video. In this proof of concept test, a Microsoft Smooth ABR video stream with two bitrate profiles (2.1 Mbps and 6 Mbps) was delivered over a DOCSIS 3.0 access network. The downstream peak capacity was about 300 Mbps with an 8-QAM (Annex B 256 QAM) bonded channel. The upstream peak capacity was about 100 Mbps with a 4-QAM (Annex B 64QAM) bonded channel. The upstream was congested by generated network traffic. The experiment recorded both bitrate profiles requested by the ABR client and ABR video fragment (chunk) download time.

The results of non QoS-assisted best effort delivery and QoS-controlled delivery are compared in Figure 2. In the first case, all traffic was delivered best-effort without special QoS treatments. The ABR stream under test was only able to reach 2.1 Mbps. In the second case, TCP ACKs of ABR streaming were given higher QoS priority than other high speed data traffic. This time, the ABR stream reached 6 Mbps. As demonstrated, in the situation of congested upstream and no upstream QoS, only the lower bitrate profile and the lower video quality were achieved even if downstream bandwidth was abundant. When QoS and higher delivery priority were applied to upstream TCP ACKs, ABR video delivery achieved the higher video quality.

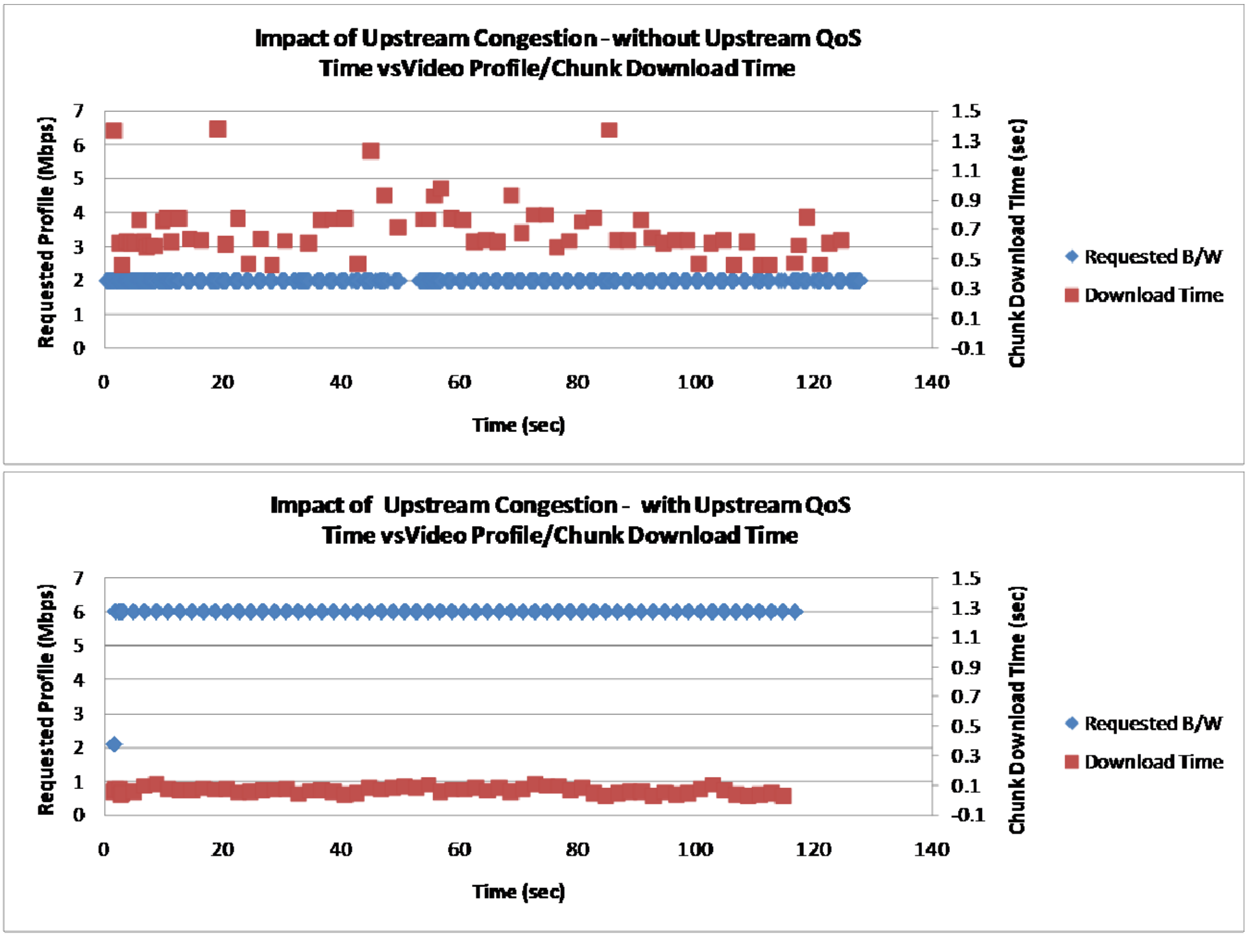

Figure 2. Applying QoS to Upstream TCP ACK

It is also interesting to observe the chunk download time in both cases. The chunk download time is the time a network takes to deliver an entire video fragment. Each fragment in the example was 2 seconds in video playing time. The size of the video fragment in the second case was larger than that in the first case because of the higher bitrate profile. Still, the chunk download time in the second case was much smaller compared with that in the first case. The reason has to do with the way video fragments are delivered in ABR streaming.

Unlike traditional video delivery, an ABR video fragment is always delivered at the current data rate of the network regardless of the bitrate profile of the fragment. In other words, a fragment with a bitrate profile of 6 Mbps will be delivered at 300 Mbps if the network throughput is 300 Mbps for the stream. Due to this bursty nature of ABR video delivery, the chunk download time is inversely proportional to the TCP throughput. The higher the TCP throughput, the less time is needed for a fragment to download. In the first case, upstream congestion and TCP ACK packet loss significantly reduce the TCP throughput; therefore, the chunk download time is longer. In contrast, QoS prioritization allows much higher TCP throughput in the second case. The increased TCP throughput and shorter chunk download time cause the profile adaptation mechanism of the ABR client to choose a higher bitrate profile. Since ABR streaming supports multiple TCP sessions per video stream, the TCP throughput refers to the aggregated TCP throughput for the stream.

This proof of concept study clearly demonstrates the importance of upstream QoS in delivering better quality of ABR video streams and utilizing downstream bandwidth more efficiently.

ACK Suppression

For a better understanding of TCP ACK suppression, DOCSIS upstream delivery is briefly reviewed here. In order to transmit a packet upstream in a DOCSIS network, a cable modem must request bandwidth from a CMTS. The CMTS then grants the bandwidth and schedules the packet delivery. The cable modem waits for its scheduled timeslot before it transmits the packet. This cycle is referred to as the request-and-grant cycle. Without any optimization, TCP throughput is limited by the request-and-grant cycle, because the modem can send only a single TCP packet in the upstream direction for each request-and-grant cycle.

One optimization technique to improve upstream throughput in DOCSIS is called concatenation. Concatenation allows a cable modem to combine multiple upstream packets in a single upstream transfer. Concatenation is applicable to all traffic types and is not limited to TCP traffic. The concatenation becomes even more efficient with DOCSIS 3.0 upstream channel bonding, when continuous concatenation is used and the concatenation occurs at sub-packet boundaries. All small packets, including ABR ACKs, benefit from concatenation and the improved upstream efficiency.

The other optimization technique to improve upstream bandwidth efficiency is specific to TCP traffic and is called TCP ACK suppression. In TCP transmission, each ACK packet contains an acknowledgment number acknowledging the last contiguous byte received successfully. All prior bytes are considered acknowledged. ACK suppression takes advantage of this cumulative nature of the TCP acknowledgement scheme and removes unnecessary TCP ACKs in the upstream direction. In the example shown in Figure 3, when the cable modem receives the first TCP ACK from the CPE, it sends a request for bandwidth equivalent to one TCP ACK. When the grant arrives, the cable modem already has three ACKs from the same TCP flow queued up. It is only necessary for the cable modem to send ACK#3 to acknowledge the receipt of all three TCP ACK packets. This scheme of sending only the last TCP ACK in the queue decreases the bandwidth consumption in the upstream direction.

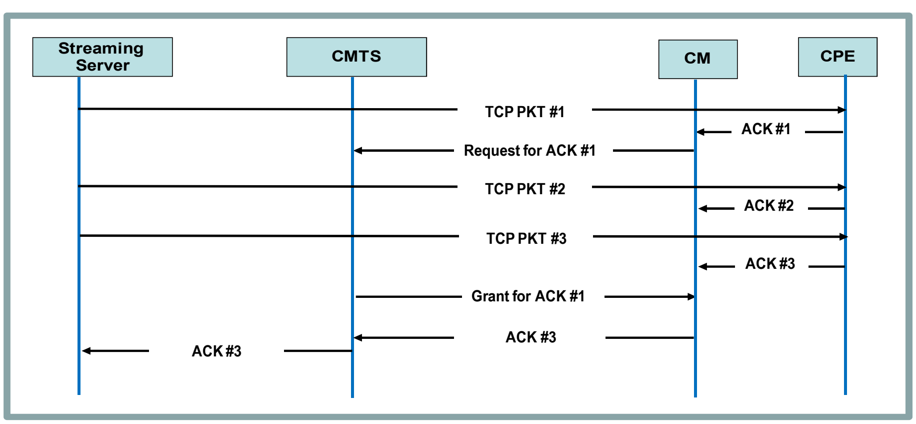

Figure 3. Upstream TCP ACK Suppression

TCP ACK suppression, when enabled in cable modems or residential gateways, has the potential to reduce upstream bandwidth required by ABR streaming. From Figure 3, it is apparent that the effectiveness of TCP ACK suppression is highly dependent on how fast TCP ACK packets are accumulated. In the above example, there are three ACK packets accumulated in the DOCSIS request-and-grant interval. Only one out of three ACK packets needs to be sent upstream. Therefore, an ACK suppression rate of 67% is achieved. At the other end of the spectrum, if there is only one ACK packet arriving in the request-and-grant interval, the ACK suppression rate will be zero.

To quantify the effectiveness of TCP ACK suppression, an experiment was carried out. An ABR video stream (Microsoft Smooth) was delivered over a DOCSIS 3.0 network with controllable available bandwidth. The efficiency of TCP ACK suppression was measured as the percentage of TCP ACKs that were suppressed by the cable modem. Traffic was captured by a Wireshark at both the client side and at the server side. The Wireshark capture was then analyzed to identify the suppressed ACK packets. From the results shown in Figure 4, it is clear that the efficiency of TCP ACK suppression is directly related to the available downstream bandwidth. When the downstream available bandwidth is 10 Mbps, the ACK suppression efficiency is only 30%. However, if the available downstream bandwidth is higher, namely 50 Mbps, the ACK suppression efficiency is 70%.

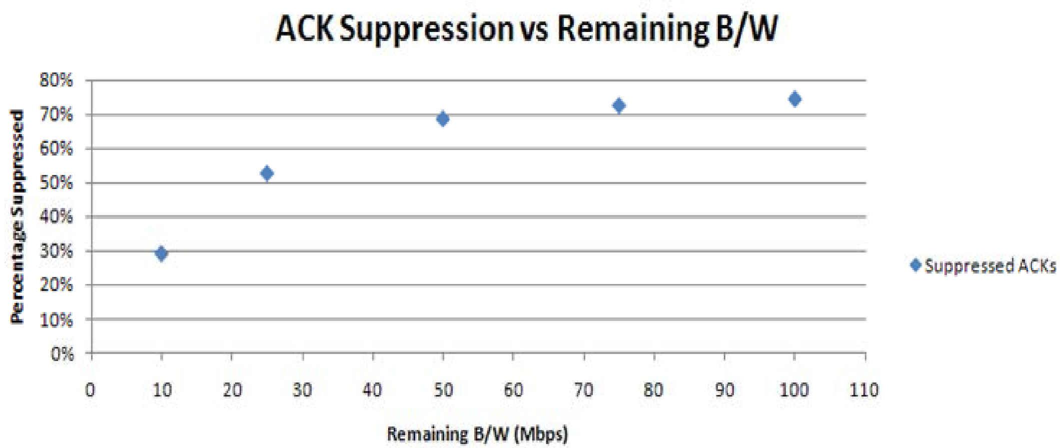

Figure 4. TCP ACK Suppression Efficiency

Taking a closer look at how ABR streams are delivered will provide insights into the ACK suppression efficiency results. ABR video fragments are delivered via HTTP/TCP. The fragment delivery speed is proportional to the TCP throughput of the transport. The faster the fragments are delivered, the more TCP ACKs are accumulated in each request-and-grant interval and thus the better TCP ACK suppression efficiency. Based on the test results above, it is clear that TCP ACK suppression is more effective when downstream bandwidth is higher.

In service provider networks, however, the TCP ACK suppression efficiency is further complicated by other factors. First, end-to-end network conditions may also impact TCP throughput. The proof of concept work described earlier is simplified to only consider the last mile cable access as the bottleneck for ABR video delivery. This is likely the case when CDN and content edge caching are used to facilitate the ABR video delivery. However, if edge caching is not involved, the delivery bottleneck can be anywhere in the network. The bottleneck reduces the end-to-end TCP throughput and can negatively impact TCP ACK suppression efficiency. Furthermore, when multiple ABR clients compete for last mile bandwidth, the available bandwidth for each client decreases. Consequently, the TCP ACK suppression efficiency can be reduced due to the lowered TCP throughput. Since the average download speed of the ABR fragments varies greatly throughout each day, it is difficult to accurately quantify the overall impact of ACK suppression.

DOWNSTREAM OPTIMIZATION

Two downstream optimization approaches are proposed. The first method lowers bandwidth consumption in the downstream direction by utilizing more efficient video encoding. The second approach optimizes bandwidth distribution under the constraint of existing network capacity to obtain best end user experience.

Variable Bitrate Video and ABR Streaming

The bandwidth advantages of Variable Bit Rate (VBR) video encoding over Constant Bit Rate (CBR) video encoding is well understood. With comparable video quality, VBR encoding saves 40%-60% bandwidth over CBR video [5]. However, VBR is typically used only in broadcast service in traditional cable video delivery. Since traditional cable video is transmitted over a narrow bandwidth pipe (e.g. 38 Mbps Annex B QAM channel), delivering VBR video requires MPEG statistical multiplexing to squeeze multiple VBR streams into the constant bandwidth pipe. Multiple drawbacks of MPEG statmuxing, such as high cost, long latency and video quality degradation, have prevented the adoption of VBR video in narrowcast video services such as switched video and on-demand video.

DOCSIS 3.0 video delivery introduces a new VBR delivery model by providing a wideband transport pipe and network statmuxing [6]. Wider pipes with DOCSIS 3.0 channel bonding and skinner streams with advanced video coding eliminate the need of MPEG statmuxing. A large number of VBR streams are statistically multiplexed naturally and efficiently by the DOCSIS transport. To address concerns about guaranteed VBR video delivery via DOCSIS network statmuxing, several methods have been proposed [6]: mixing VBR video and best effort HSD data in a converged DOCSIS pipe, applying VBR admission control and error retransmission. Despite these improvements, VBR video delivery using unreliable transport such as UDP/RTP over DOCSIS is still considered by many as unguaranteed delivery.

As ABR streaming is gaining traction in managed video services, new possibilities to maximize the potential of VBR bandwidth savings are on the horizon. The use of ABR streaming and VBR encoding not only solves VBR network delivery reliability issues, but also enhances the ABR streaming with improved bandwidth efficiency. In a typical ABR streaming implementation, each video quality profile is explicitly signaled using bitrate value, such as 500 kbps, 1 Mbps, 2 Mbps streams. If we consider these bitrates as average bitrate instead of constant bitrate, VBR encoding can be applied to each video quality level instead of CBR encoding (Figure 5). VBR is a superior encoding choice as it naturally keeps the video quality constant while varying the encoding bitrates.

Although typical ABR video implementations today have clients signal absolute bitrate values to identify video quality profiles, the video quality levels can alternatively be represented using relative terms. In the upcoming MPEG Dynamic Adaptive Streaming over HTTP (DASH) standard [7], relative quality ranking is used instead of absolute bitrate and VBR can be supported easily.

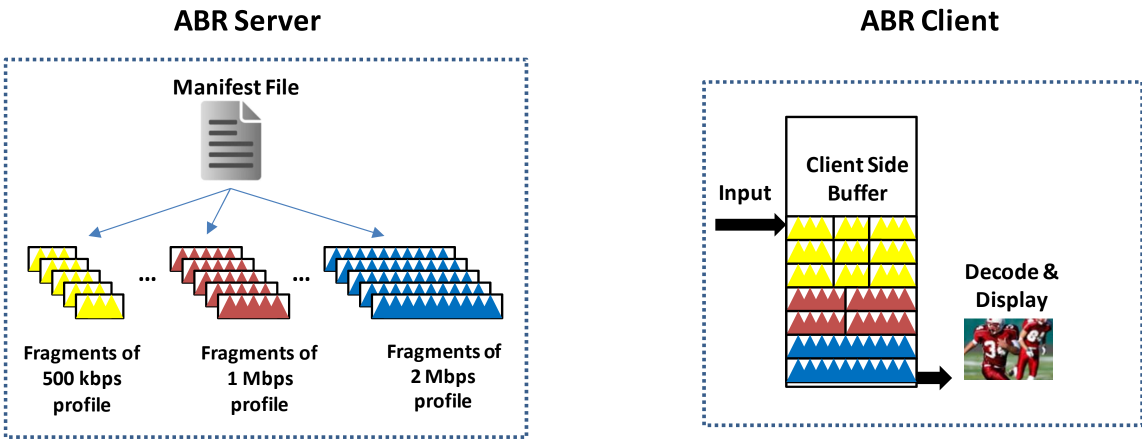

Figure 5. ABR Streaming with VBR Encoding

Thanks to TCP delivery and ABR streaming, VBR bitrate spikes are now handled gracefully and VBR delivery is guaranteed. In UDP based cable IPTV systems, the video delivery rate in the last mile is the same as the video consumption rate at the client side decoder. Namely, an 8 Mbps HD stream is delivered by a server at 8 Mbps to a client, excluding allowance of limited network jitter in the order of 100 ms. In contrast, ABR video is delivered at the maximum speed the access network permits. An ABR video fragment at 2 Mbps bitrate profile could be bursted at a speed of 100 Mbps in the last mile. ABR video fragments are downloaded to ABR clients, just like files. Therefore, the bandwidth variation of VBR streams is masked out by this fragment download operation and is no longer critical. In fact, it is the size or the average bitrates of video fragments that matter to the network. A more efficient encoding will yield a smaller size for each fragment. When fragments are smaller in size, they in turn consume less bandwidth and contribute to higher bandwidth utilization.

The powerful combination of ABR and VBR makes it possible to reap the bandwidth benefits of VBR video without dealing with the complexity of VBR rate variations in transmission pipes.

Downstream QoS

Given that ABR streaming adapts bitrate profiles to network conditions, are QoS and network controls still necessary? While ABR streaming provides a technique for OTT video providers to overcome bandwidth congestion issues in unmanaged networks, more can be done for service providers who own the network infrastructure. Taking advantages of network control to provide superior video experience to on-net subscribers is a key differentiator for service providers.

DOCSIS access networks provide two types of traffic delivery mechanisms. One is the Committed Information Rate (CIR) service with guaranteed delivery. The other type is the Best Effort (BE) delivery. A DOCSIS service flow can be delivered by means of CIR, BE, or a combination of both. In general, traffic without bandwidth guarantee and admission control is opportunistic traffic.

Although ABR video does not require CIR protection, there is no guarantee of video delivery quality. Network congestion may force an HD ABR client to receive video at a low-bitrate profile, e.g. 1 Mbps, even if higher bitrate profiles are offered. Usually, such a low bitrate stream may have unacceptable video quality for HD devices. Delivering content with unacceptable quality not only negatively affects users' perception of the quality of an operator's video service, but also wastes precious network bandwidth. It is beneficial to deliver ABR video with a CIR floor to protect the video with minimum acceptable quality. This minimum acceptable video quality or bitrate should be directly related to the client screen size and content type. If the network is not able to provide even the bare minimum acceptable video quality, the new ABR video request should be rejected by admission control and proper feedbacks can be sent to clients. Meanwhile, the design allows the ABR video to move to higher bitrate profiles when network bandwidth is available for better video quality. The ABR traffic beyond the CIR protection threshold is treated as BE traffic in DOCSIS access.

Opportunistic video traffic can be further optimized in cable access. DOCSIS access is designed to provide different QoS priorities to different types of traffic. Providing higher priorities to ABR streaming video over other types of non-realtime traffic is straightforward and can be easily accomplished. Marking managed ABR streaming traffic with a different DSCP value and then applying a higher DOCSIS priority to it in the last mile will prioritize managed ABR streaming against other traffic during network congestion. What is more complicated is traffic prioritization among ABR clients themselves. Research [8] shows that when two ABR clients competing for available bandwidth, the bandwidth allocation between the two clients are not deterministic and the video quality profile oscillates. Future research involving a larger number of ABR clients may shed additional light into the client behavior. More interesting questions are: What is the desired bandwidth allocation among competing ABR clients? How can the network facilitate better bandwidth allocation?

To obtain superior end user video experience, not all ABR clients should be treated equally. For instance, an ABR client is serving video content to an HD device at a video bitrate profile of 4 Mbps. Another ABR client is serving video content to a SD device at the same video bitrate profile of 4 Mbps. Downshifting the HD ABR client to a lower bitrate profile (e.g. 2 Mbps) will negatively impact user experience while the effect of downshifting the SD client to the next bitrate profile is less noticeable. Under network congestion, it is desirable to provide higher priorities to ABR streams at bitrate profiles critical to end user viewing experience.

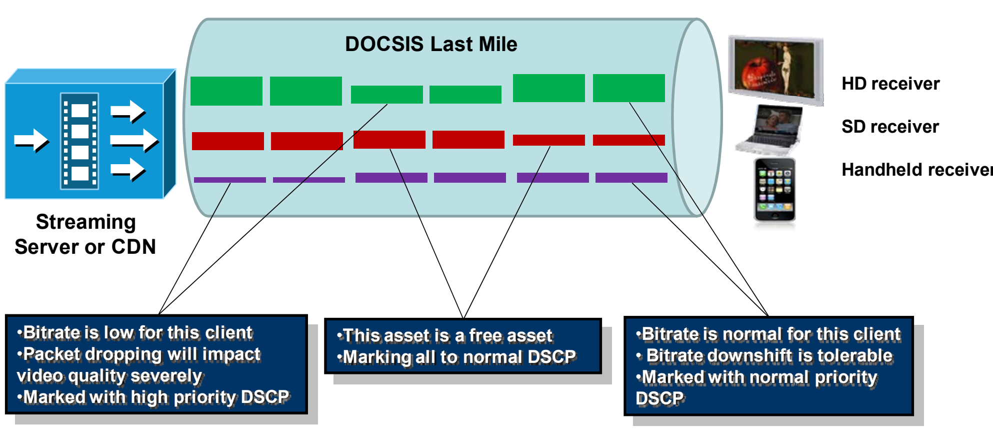

Figure 6. ABR Stream Prioritization in Downstream

The proposed QoS scheme is illustrated in Figure 6. When an ABR client requests a video segment via HTTP GET, it indicates to the network that it is operating in a normal or a critical condition through DSCP marking of the TCP traffic. An ABR client running in a bitrate profile critical to end user video experience is in a critical condition. Traffic with different markings can be sent using different TCP sockets. The DOCSIS network just provides a higher priority to traffic marked in critical conditions to minimize the impact of video quality degradation. Taking this prioritization scheme one step further, operators can potentially apply business rules. For example, free assets and free service may not be allowed to use the critical condition priority. The proposed scheme prioritizes ABR streams among ABR clients and optimizes end user viewing experience.

SUMMARY

The next wave of video entertainment is coming our way. With technology advancements in adaptive streaming, video content can be delivered to any device, on-net or off-net, managed or unmanaged. Adaptive streaming is becoming a new tool in the service providers' toolkit to deliver advanced video services. To overcome some of the technical challenges of scaled ABR streaming deployments in managed networks, a variety of optimization techniques can be applied to upstream and downstream DOCSIS networks. From upstream optimization of simple expansion of bandwidth, TCP ACK prioritization and TCP ACK suppression, to downstream optimization of VBR coding, stream protection and stream prioritization, these techniques enable cable operators to unleash the full potential of adaptive streaming and pave the way for future video services.

ACKNOWLEDGEMENTS

The author appreciates Jonathan Muslow's great proof of concept work. The author also thanks Li-shen Chin for his help on ABR stream analysis tools and his team's earlier work in the area of ACK suppression.

REFERENCES

[1] Cisco Visual Networking Index, June 2010 http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf

[2] Sandvine 2010 Fall Global Internet Phenomena report, http://www.sandvine.com/news/global_broadband_trends.asp

[3] A. Begen, T. Akgul and M. Baugher, "Watching video over the Web, part 1& 2: streaming protocols," IEEE Internet Comput., vol. 15/2 & 15/3, 2011

[4] J. T. Chapman, "Taking the DOCSIS Upstream to a Gigabit per Second", NCTA Spring Technical Forum 2010

[5] S. J. Huang, "Principle, Applications of Variable Bit Rate Coding for Digital Video Broadcasting, w. Statistical Multiplexing Extension", NAB 1999

[6] X. Liu, A. Bernstein, "Variable Bit Rate Video Services in DOCSIS 3.0 Networks", NCTA 2008

[7] Dynamic Adaptive Streaming over HTTP, http://mpeg.chiariglione.org/working_documents/mpeg-b/dash/dash-dis.zip

[8] S. Abhshabi, A. Begen, C. Dovrolis, "An Experimental Evaluation of Rate-Adaptation Algorithms in Adaptive Streaming over HTTP", Multimedia System, Feb, 2011