AN SDN BASED APPROACH TO MEASURING AND OPTIMIZING ABR VIDEO QUALITY OF EXPERIENCE

Sangeeta Ramakrishnan, Xiaoqing Zhu
Cisco Systems

Abstract

The volume of video traffic is continuing to grow rapidly over cable networks. While a majority of IP video on DOCSIS networks today is over-the-top video from third parties, operators are increasing the amount of programming their subscribers can access on IP video-capable devices. Additionally many operators are pursuing IP video deployments that are at various stages from planning, testing, trialing to deploying. Most OTT deployments use Adaptive Bit Rate (ABR) technology (also known as HTTP Adaptive Streaming). And ABR has also been adopted by cable operators as the technology of choice as they begin their migration to IP video.

ABR video was primarily developed to work as well as possible despite the network. This approach made sense for over-the-top content providers who have no control or influence of the network. Such a server-client method has so far worked well for OTT providers. Some of the challenges with such an ABR delivery method are masked because of end-users' acceptance of a poorer user-experience since OTT providers are perceived to be lower cost options. Also the amount of ABR video on the cable network today, while significant, will still be dwarfed by the amount of ABR video that will come on to the network when cable operators migrate to a ABR-based IP video delivery method. Hence the combination of more ABR traffic along with higher user-experience expectations of a cable subscriber, may pose challenges to cable operators as they deploy ABR-based IP video.

Another concern with ABR video is that different segments of the network can have similar network utilization levels yet there could be a large difference in end user Quality-of-Experience (QoE) between these segments of the network. Currently operators do not have good visibility to the actual QoE of subscribers; instead they primarily monitor network utilization levels to identify those segments that need to be upgraded. Operators need better tools to identify where QoE is below their service objectives so that they can target their network investments accordingly.

In this paper, we present a novel SDN-based solution to solve some of the challenges that we anticipate will occur with heavy ABR usage. Our proposal will help operators to improve their visibility of QoE and optimize their network investments. We also present a technique to help operators improve aggregate QoE of all users on their network. Alternatively this method can be used to pack significantly more streams in the given bandwidth with comparable quality to conventional ABR. Our studies indicate that with this approach bandwidth requirements can be reduced by a third or more, thereby saving DOCSIS channels and HFC spectrum, and reducing the cost of the overall solution.

INTRODUCTION

Over the past few years, video streaming traffic has been growing at a rapid pace, and is anticipated to dominate next generation networks. According to [1] global consumption of Internet video viewed through a TV doubled in 2012; video-on-demand traffic is projected to nearly triple by 2017. Cable operators, are seeking novel solutions to fend off the impending bandwidth crunch introduced by video traffic on their existing network infrastructure.

Additionally, the cable industry is on the cusp of a migration to IP video. Various factors are contributing to the desire to migrate. One of the biggest drivers is the desire to deliver any content to subscribers, anywhere and anytime they want it. This includes the ability to deliver content to consumer owned devices such as tablets, smartphones, PCs, laptops, game consoles and Internet enabled TVs. Delivering video to such consumer owned devices not only helps cable operators to meet subscribers' needs and expectations but has the added benefit that operators do not have to incur capital expenditure to deploy and maintain these devices. STBs deployed in subscribers' homes are a significant portion of the capex budget in offering video services. Reducing that expenditure by either leveraging consumer owned devices or lower cost IP STBs could help operators to improve their bottom-line significantly. Additionally, operators may be able to charge fees for making services available on new outlets such as tablets. So overall IP video provides an opportunity to both improve top-line and bottom-line for operators.

Additionally with IP video, operators may be able to leverage the rapid innovation in the Internet space to offer newer services, and a better user experience. It also enables operators to consolidate the video services infrastructure.

One of the major challenges faced by operators as they make this migration to IP video is the bandwidth required to deliver this service and the capex requirements associated with it. As shown in [2] the bandwidth needs for IP video are highly dependent on the type of service offered and can range typically from 20-40 downstream channels to serve a typical Service Group of 250-300 subscribers. The increasing focus on Cloud DVR like technology will in fact lead to higher capacity requirements due to the unicast nature of the delivery method. Hence there is likely to be significant increasing amounts of Adaptive Bit Rate video on the cable network in the future and operators will need better ways to manage it.

SOFTWARE DEFINED NETWORKING

Software Defined Networking (SDN) is a term that began being used widely in the industry around 2009. SDN is defined as the separation of Control and Data planes using an open standard protocol to communicate between them. This differs from a traditional network device (such as a Router, Switch, or CMTS) in which the data and control planes are vertically integrated. SDN promises flexibility and rapid innovation by virtue of the fact that the control software would be removed from the relatively constrained network device to a generic server that can be easily scaled to have more processing and memory capabilities.

SDN is typically implemented with a Controller based architecture as shown in

`Figure 1`, where the Controller interfaces with network devices via standard interfaces (for example Openflow [7], PCMM). Therefore, the controller presents a level of abstraction of the underlying network. Applications communicate with the controller using "controller APIs" and the controller in turn interacts with the network. In other words the controller acts as middleware that provides a higher lever of abstraction to network application developers.

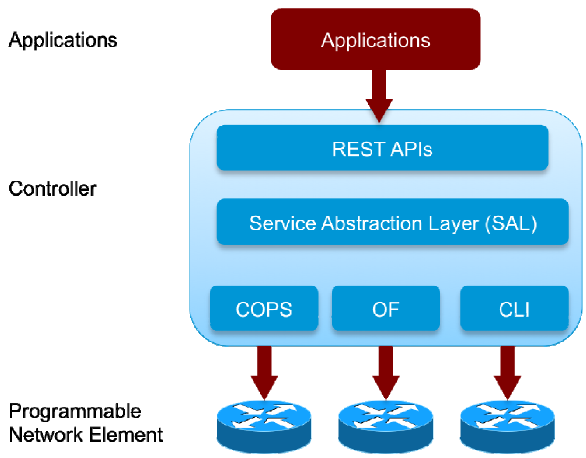

Figure 1 SDN Architecture

Another key attribute of SDN is the two-way communication with the network devices. The network can be thought of as a large distributed database of flows and states. In an SDN architecture, applications can learn from the Controller the state of various devices and can also program the devices via the Controller. An SDN-based architecture enables Applications to be developed in a relatively device-agnostic way.

ADAPTIVE BITRATE STREAMING

Recent years have seen a major technology shift as Internet video delivery solutions are converging towards the adaptive bit rate (ABR) streaming paradigm. Since its inception in 2007 by Move Networks [3], ABR has been quickly adopted by major vendors and service providers, including YouTube, Netflix, Hulu, Akamai, Microsoft Smooth Streaming [4], Apple HTTP Live Streaming (HLS) [5], and Adobe HTTP Dynamic Streaming (HDS).

Figure 1 shows the architecture overview of an adaptive bitrate streaming (ABR) system. The media contents are either pre-stored or captured live at the source. Multiple quality versions of the same video content are generated via transcoding. Moreover, each media file is broken down into many small fragments. The origin HTTP server keeps track of these fragments either as a large collection of separate physical files (e.g., in Apple HLS), or as logical separations via indexing (e.g., in Microsoft Smooth Streaming). Additional content delivery

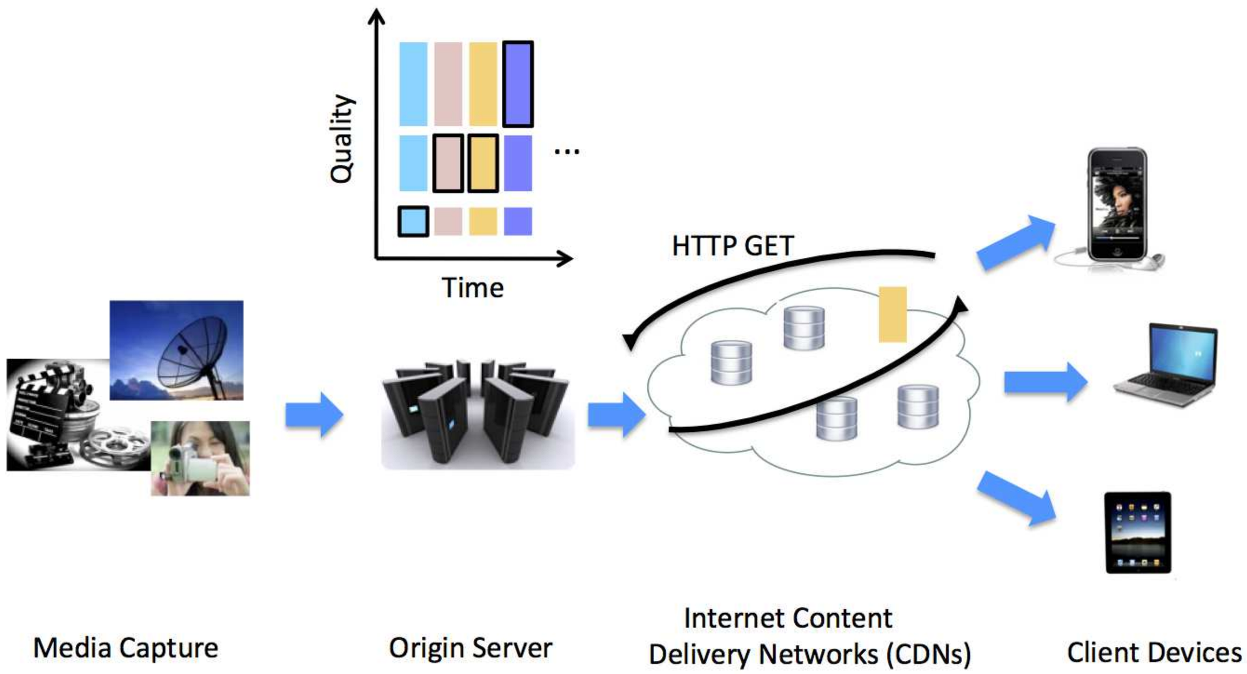

Figure 2 Architecture overview of an adaptive bitrate streaming (ABR) system. Multiple quality versions of the same video content are generated via transcoding or parallel live encoding. Each media file is broken down into many physical or logical fragments. The client can adaptively request different quality versions of the media fragments based on its own estimate of available network bandwidth

networks (CDNs) may also be leveraged at the edge of the network, so as to assist in disseminating video contents to a wide range of end users.

In ABR, the client can dynamically change its video rate and quality on a fragment by fragment basis. In face of temporary network congestion, the client can switch to a lower rate (and hence quality) version of the video to avoid buffer underflow; when connection speed recovers, the client can switch back to higher quality. Such flexibility in rate adaptation can be advantageous in the presence of dynamic network conditions, especially in mobile environments. In 2012, the Motion Picture Experts Group (MPEG) has joined forces with 3GPP (3rd Generation Partnership Project) in defining the recently standardized Dynamic Adaptive Streaming over HTTP (DASH) specifications [6]. The MPEG-DASH standard has intentionally left out of its scope the definition of client behavior for content fetching, rate adaptation heuristics, and video playout, thereby allowing plenty of space for innovation-based competition in industry.

QoE MEASUREMENTS

Today most operators measure the quality-of-experience (QoE) of their subscribers in terms of network utilization levels. Network utilization is typically averaged across over few minutes (typically 5~15 mins). Such average utilization is measured throughout the day. Most operators consider their networks to be congested if utilization exceeds a certain threshold such as 70-80% during peak hours.

If many such measured samples of utilization level exceed their pre-determined threshold, operators typically declare those interfaces to be congested and plan upgrades of their network to address the congestion. The advantage of such a QoE measurement is it is simplicity – easy to measure and easy to monitor.

Challenges with ABR QoE Measurements

As the industry moves toward ABR-based video delivery, this new paradigm poses some challenges for the aforementioned QoE measurement mechanism. This is because with ABR, clients up-shift to higher rate profiles when bandwidth is available. This is in contrast to standard Internet browsing, where higher bandwidth availability simply causes the file download to finish faster. For example, downloading a file when plenty of bandwidth is available does not cause file size to grow! TCP simply takes advantage of the bandwidth available, and file download completes faster.

With ABR, however, clients download larger and larger files along with growing available bandwidth, by upshifting to higher profiles. So bandwidth demand from ABR delivered video is elastic in nature (up to a certain point) and can easily use up all available bandwidth.

As a result, the measured network utilization level can always look high. But in some cases, end users may be quite happy because they are all receiving highest quality streams. Or, in other cases, end users can be quite unhappy because they are all putting up with lowest-profile streams. An operator can no longer tell whether end users are having a good quality-of-experience or not by simply looking at network utilization measurements.

In fact, high utilization on multiple segments of the network may cause operators to assume all such network segments need to be upgraded whereas if they had a better view of the users' QoE they may be able to better determine which segments need to be upgraded immediately and which can be deferred. Having better visibility on actual QoE would enable operators to spend capital on upgrading segments of the network that would yield the most improvement in subscriber QoE.

We ran a simple test to illustrate this point and the results of that test are included in figures that follow. Two interfaces were configured with bandwidth just under 20Mbps, however the number of ABR clients on each interface was varied, 4 in one case and 3 in the other. As seen in `Figure 3` and `Figure 5` below, the utilization levels of the two interfaces are comparable. However as seen in `Figure 4` and `Figure 6` the actual rate profiles selected by the clients in the two cases are quite different.

When 4 clients are sharing an interface we see rate profile selections generally in the range of 3~5 Mbps. However when 3 clients are sharing an interface we see rate profile selections in the range of 4~7Mbps, which typically result in better QoE than when rate profiles selected are in the range of 3~5 Mbps.

The above example is obviously overly simplistic but it does illustrate the point that higher order data such as interface utilization may not give a good indication of QoE due to ABR's behavior of stepping up or down in profiles to match bandwidth available.

Additionally the mix of client devices, content mix, etc., may also be quite different between two network segments. For example the network segment where rates were lower, could in fact be the segment with a larger mix of premium content/subscribers and/or big screen devices. Whereas, the network segment receiving higher rates might have actually a larger mix of handheld devices. So a better measure of QoE should also take into account other characteristics such as device type, content type, and subscriber type

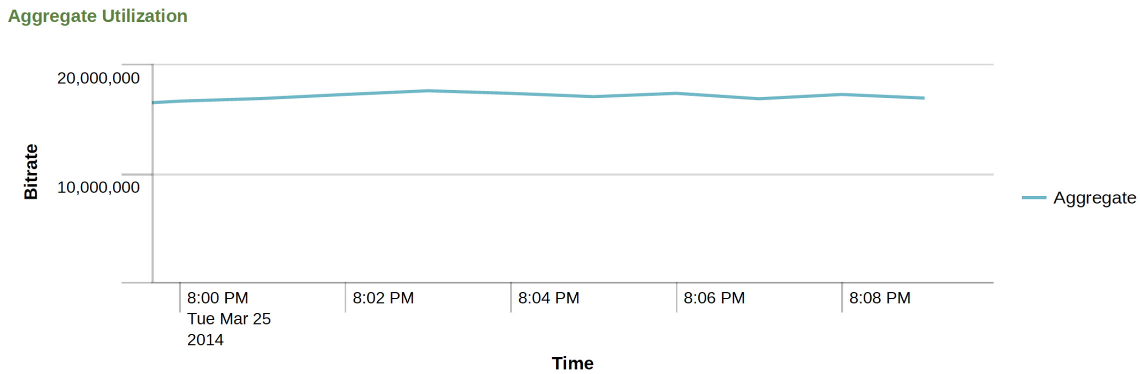

Figure 3 Aggregate utilization of an interface with 4 ABR clients

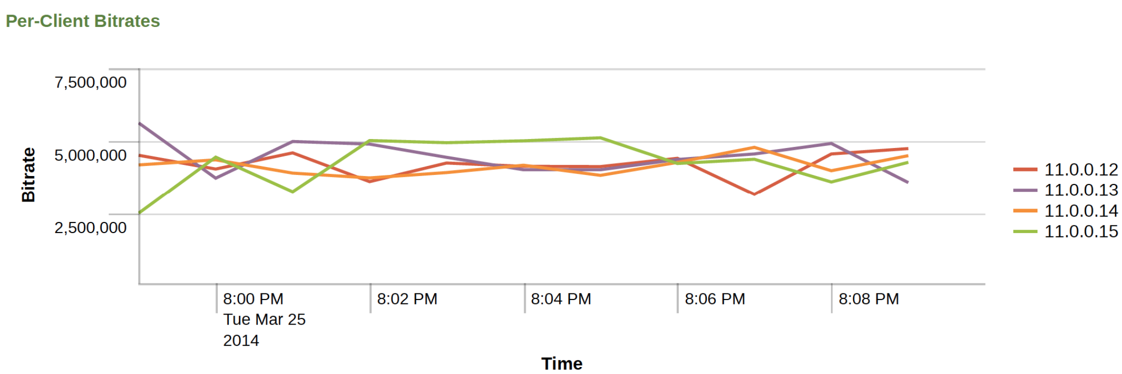

Figure 4 Bitrate per client when 4 ABR clients are sharing an interface

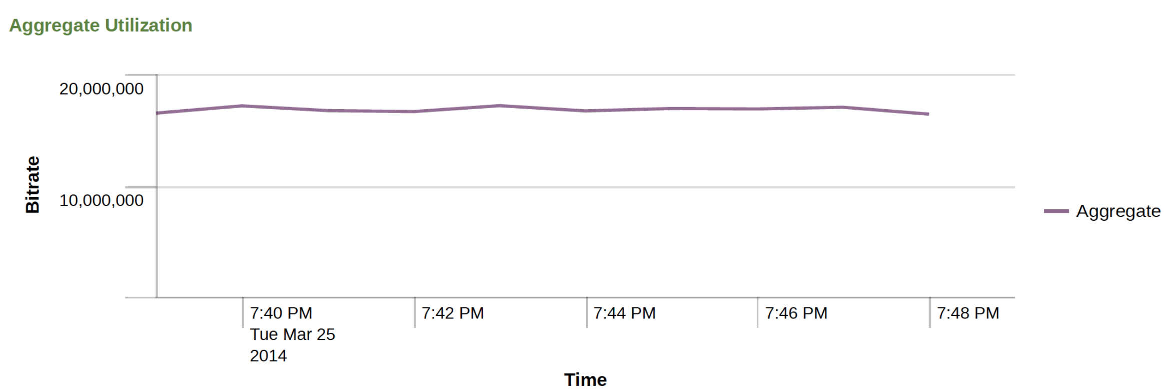

Figure 5 Aggregate utilization of an interface with 3 ABR clients

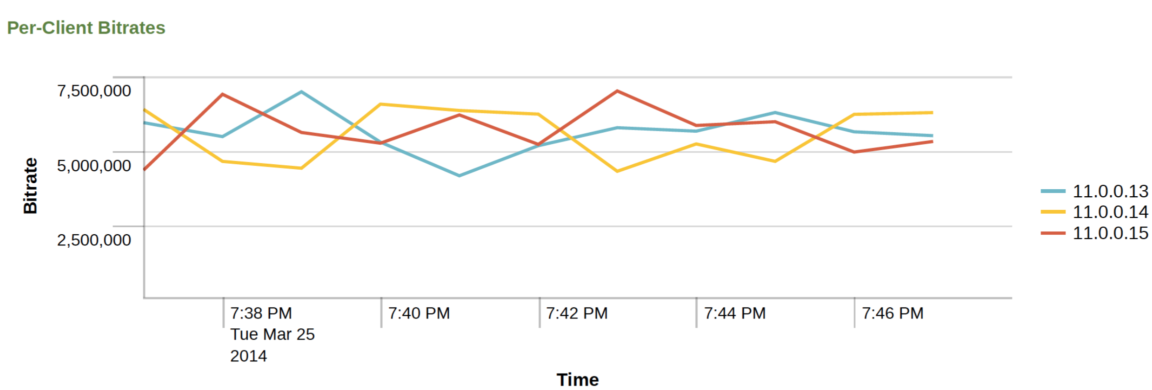

Figure 6 Bitrate per client when 3 clients are sharing an interfa

NETWORK AGNOSTIC ABR DELIVERY

Today Adaptive Bit Rate delivery of video works despite the network. This was fine for over-the-top (OTT) content since OTT providers anyway has no control over the network. Additionally OTT originally started off being a very small portion of the network bandwidth. So in fact each client being greedy and maximizing its own experience was a reasonable approach for a OTT provider solely focused on maximizing their own subscribers' experience when competing for bandwidth with other applications.

However for MSO managed content, the landscape will change in two ways – one, a much larger proportion of the traffic on the network will be ABR video and secondly cable subscribers will be competing with each other for bandwidth. Therefore, operators may want to ensure fairness across a pool of users, to maximize quality of experience across their set of users rather than be satisfied with a greedy approach where each client is operating solely to maximize its own benefit at the cost of other users' experience.

PROPOSED SOLUTION

We hereby propose an alternate architecture ---- built upon SDN --- that is better suited for monitoring and delivery of managed video. In this our SDN-based solution, streamers, networks and clients *work together* to both report on and improve users' quality-of-experience (QoE) in the video network.

The SDN architecture aids in this solution by providing a common framework for collecting information from various network entities. By design, different applications can interface with the SDN Controller to extract the particular information they need for their application-specific purposes. For example, there may be a number of applications interested in topology information. Instead of each application collecting such topology information it is more efficient and scalable to have the Controller be the single entity that collects such information from the CMTS and then have the applications interface with the Controller.

For ABR video streaming, we envision a Video QoE Application (VQA) that collects information from various points in the network and analyzes it to provide a more accurate estimate of end users' QoE. For example, as shown in Figure 7, the VQA may collect information about the content from the video steamer/CDN, client metadata (such as device type) from clients and network information from the CMTS. By combining information about the content, client, and network, the VQA can generate analytics to aid the operator in better understanding end users' QoE in different segments of the network.

The Controller can be used to collect network topology information from the network devices (such as CMTS), and end device type (smartphone, STBs etc.). The VQA can then query the Controller for required information. Video QoE metrics can be generated per Fiber Node, per CMTS, or per region. QoE could initially be measured as simply the bitrate achieved by video flows. It could be further enhanced to reflect the actual quality of the flows, where quality is measured by some objective metric such as PSNR (Peak Signal to Noise Ratio), VQM (Video Quality Metric), or SSIM. There are proposals in DASH industry forum on enhancing the Media Presentation Description (MPD) files to include quality information of video fragments. By collecting such rate and quality information from MPD files, HFC topology information from CMTS, and client stream and device information from either the client or the server, the VQA can pool together meaningful QoE analytics for the Operator.

Cable operators can use the QoE analytics for a number of purposes. For instance, they can more clearly identify which segments of their network are most congested and need to be upgraded. This will allow them to spend their upgrade budget more intelligently, rather than simply relying on the interface utilization as a poor measure for QoE.

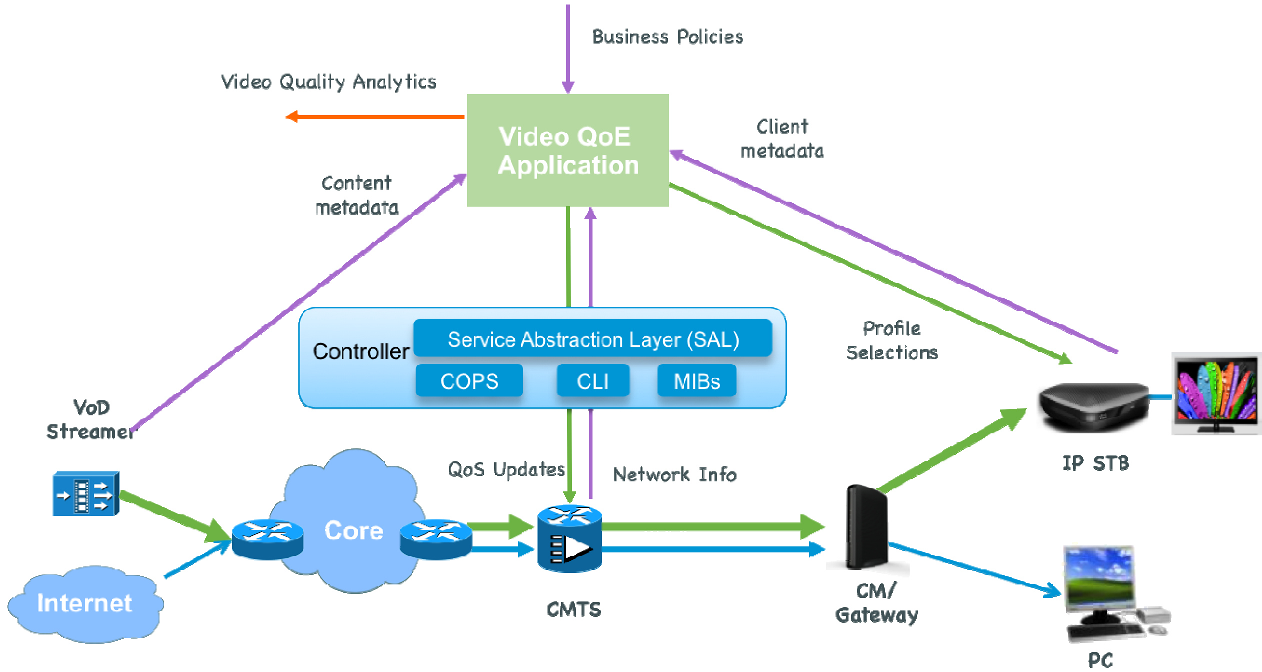

Figure 7 SDN based architecture for QoE Analytics and QoE Optimization

In addition, such analytics can be used to debug and prevent issues in the field. Availability of the QoE analytics makes it much easier for operators to identify where QoE is poor, so they can address the problems in a proactive manner rather than waiting for subscribers to complain about problems.

The proposed solution not only presents a platform for aggregating and visualizing QoE analytics, it further enables the Operator to optimize its network resource for improved end-user quality-of-experience (QoE). For this second use case, the VQA, as shown in Fig. 8, collects a similar set of data as in the Analytics use case. In addition, it can also modify the QoS on the CMTS to optimize users' QoE as needed. One example of how this can be done is via a SDN Controller that has a COPS plug-in that is able to leverage the PCMM support built into CMTSs to apply QoS changes to flows that need it.

The VQA may decide on the bandwidth allocated to each flow based on device type (e.g., STB vs. tablets vs. smartphones), on codec in use (HEVC vs. MPEG4), or on video content complexity (e.g., sports vs. talk shows). Business policies such as premium subscriber or premium content may also influence the QoS that the flows receive. The VQA may in fact use any other information --- e.g., bitrate and video quality scores carried within the Media Presentation Description (MPD) or as companion metadata --- to optimize the QoE for individual flows, so as to achieve certain performance objectives (e.g., maximizing the aggregate quality across all flows, equalizing the quality across all flows in a given user group).

Such an approach will ensure that the bandwidth is available for clients that need it the most. While any approach that strives for aggregate fairness across a group of users will inevitably allocate less bandwidth to some clients than what they would have obtained in the absence of such a mechanism, the intent is that the loss of bandwidth doesn't lead to a noticeable reduction in QoE for that user, whereas the increased bandwidth for some other users contribute to their significant QoE improvement.

Benefits to the Operators are two-fold with this approach: improved aggregate QoE for subscribers and the ability to support more video streams that meet a certain QoE criteria at a given network bandwidth. Indeed, increased packing efficiency can provide significant bandwidth savings for Operators, thereby providing capex savings.

To evaluate the proposed approach we ran simulations with varying number of clients sharing a 100Mbps link. Figure 8 shows the average quality across all the flows (measured in terms of PSNR) for each test.

As can be observed from the figure, the proposed quality-optimized approach yields a significant improvement in average quality across clients. Another way to view the results is that the same average PSNR (of ~42 dB in this particular example) can be attained with our quality-optimized approach while supporting 25 streams at the same bandwidth as opposed to only 20 clients via existing methods. This improvement translates to a 25% increase in packing efficiency.

In fact, the packing efficiency improvement would be even greater if the quality metric in use is the minimum quality across clients. This is because the proposed quality-optimized approach tends to reduce the variation in quality across clients and over time, thereby boosting the minimum quality across all video segments. Using minimum quality as a metric is analogous to a Service Level Agreement (SLA) where an operator would like for the video to stay above a certain threshold most of the time.

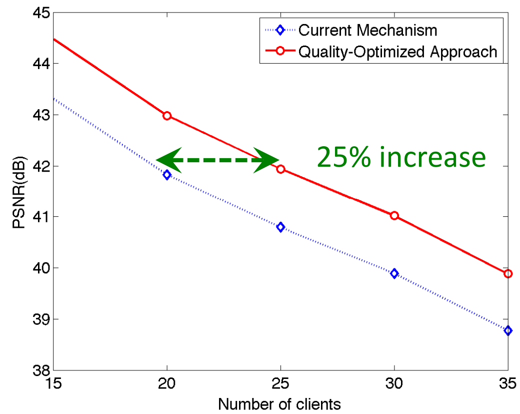

Figure 8 Comparison of the proposed quality-optimized approach against current QoE-oblivious approach. Performance results are plotted in terms of average video quality (PSNR). In this experiment, the number of competing clients vary from 15 to 35 over a 100 Mbps bottleneck link.

SUMMARY

In this paper, we have identified the challenges posed by ABR video for operators in measuring and managing their Access networks. We have also proposed a unique solution to the problem of assessing subscriber QoE and optimizing it. The proposed QoE optimization techniques can be used by operators to either improve QoE of groups of users or to pack more users in a given bandwidth while maintaining QoE. Simulation results have shown that significant improvements in packing efficiency can be achieved thereby providing significant CAPEX savings to cable operators. The proposed solution is based on SDN, and hence can leverage the flexibility of an SDN infrastructure to build a more efficient and intelligent video delivery network.

REFERENCES

1. Cisco White Paper, "Cisco visual networking index - forecast and methodology, 2011-2016," http://newsroom.cisco.com/dlls/index.html.
2. "HFC capacity planning for IP Video", Sangeeta Ramakrishnan, Cisco Systems, SCTE 2011.
3. History of Move Networks, Available Online: http://www.movenetworks.com/history.html
4. A. Zambelli, "IIS Smooth Streaming Technical Overview," http://www.microsoft.com.
5. Apple Inc., "HTTP Live Streaming Overview," https://developer.apple.com/library/mac/documentation/networkinginternet/conceptual/streamingmediaguide/Introduction/Introduction.html
6. MPEG, "ISO/IEC 23009-1:2012 Information technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats," 2012.
7. https://www.opennetworking.org/sdn-resources/onf-specifications/openflow