# Leveraging IP ABR Video Delivery for Linear QAM TV

**Weidong Mao**
**Comcast Cable**

**David Brouda**
**Imagine Communications**

*Abstract*

*As video service providers increase deployment of managed IP video service to various consumption devices (IPTV set-tops, tablets, smartphones, etc.) there has been an increase in unicast (HTTP) video delivery mechanisms using H.264 Adaptive Bit Rate (ABR) video encoding. At the same time, the MPEG-2-only legacy QAM set-top population is declining, leading to a tipping point in which the population of set-tops that are HD H.264-capable outnumbers the legacy devices.*

*The confluence of HTTP ABR/H.264-enabled devices and H.264 QAM set-tops leads to an interesting decision point: should service providers manage two core video platforms (IP and QAM) as many are doing today, or should these platforms converge at least at some level? The answer is "Yes." That is, the core should converge to HTTP, but QAM delivery to set-tops still need to be supported.*

## INTRODUCTION

Digital linear video distribution over QAM has been increasingly deployed over IP networks since the mid-to-late 2000's when edge QAMs with Gigabit Ethernet interfaces were first deployed for VOD and then broadcast video. While many edge QAMs are capable of performing multiplexing of MPEG-2 Transport Streams there are still dedicated platforms that groom and ad insert MPEG-2 and MPEG-4 video streams to desired bit rates for packaging on QAMs. This packaging might be in the form of Constant Bit Rate (CBR) or Capped Variable Bit Rate (VBR) in order to be delivered as Single Program Transport Streams (SPTS) for Switched Digital Video (SDV) or the packaging involves the multiplexing of more than one video stream into a Multi-Program Transport Stream (MPTS). Typically MPTS are not only multiplexed bit-for-bit but also statistically multiplexed to take advantage of variability in the encoded video across multiple programs such that there is an increase in efficiency without a decrease in video quality.

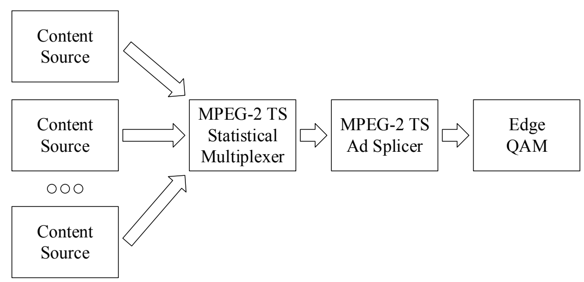


**Figure 1 – Traditional Statistical Multiplexing & Ad Splicing**

In addition to grooming for statistical multiplexing the function of ad splicing also takes place, see `Figure 1`; the platforms that perform these functions tend to be proprietary hardware platforms in which operational densities have not kept pace with technological innovations found in COTS (Common Off-The-Shelf) computing platforms.

QAM set-top boxes were originally deployed in support of Standard Definition (SD) MPEG-2 video. Subsequently High Definitional (HD) capable MPEG-2 set-tops were deployed followed by H.264-capable HD set-tops. While H.264-capable HD set-tops have been deployed for a number of years most operators have not deployed much if any H.264 HD linear and VOD content for QAM video due to incompatibility of the content with the MPEG-2-only set-tops in the majority of customer homes. Driven by the desire to reclaim HFC bandwidth and reduce storage and network cost, the HD MPEG-2-only set-tops are starting to be replaced with H.264 capable HD set-tops. A tipping point will soon be reached at the vast majority of multichannel service providers that all the HD QAM set-tops are H.264 capable. This will allow conversion of existing linear or VOD MPEG-2 HD content to H.264 HD delivery. H.264 offers better quality and lower bit rates, but the traditional methods of statistical multiplexing and ad splicing are more complex in H.264 than MPEG-2; increased complexity directly correlates to increased cost and lower physical densities for processing. Service providers therefor have been challenged by a dilemma in which they want to save bandwidth but the cost and complexities to do so have not yet outweighed the benefits.

## MULTICAST CORE NETWORK

In order to reduce operational complexities many service providers migrated their digital video network to using IP connectivity over Gigabit Ethernet. With encapsulating MPEG-2 Transport Streams in UDP/IP for delivery the best method for linear video distribution was multicast. IP multicast allowed operators to receive a linear channel at one site and distribute it simultaneously to many sites. Multicast IP networks were built using off-the-shelf networking hardware, which allowed service providers to migrate from their legacy point-to-point video distribution networks built on proprietary hardware platforms.

In support of redundancy, many service providers built two parallel IP network paths carrying duplicate multicast signals in case of network or video equipment failure. At the same time these same service providers also provide IP backbone and regional network for Internet access to their customers. Some providers keep these networks completely segmented from the core into their regional headends while others converge their networks. Segmented networks essentially are more costly to deploy, operate, and maintain while converged networks can be more complicated to operate particularly in the presense of multicast traffic. One operational complication is that multicast does not offer easy-to-implement security mechanisms in order to prevent multicast traffic from traversing networks where it may be administratively forbidden.

Some examples of administrative reasons for preventing the multicast from traversing specific networks might be: the network does not have sufficient bandwidth to carry all available multicast groups; without careful administration the multicast flows could easily make the network segment unusable due to congestion. Some multicast groups may be needed on the network but there are hosts or users on the network that are not permitted to receive traffic for security or business reasons. While Layer 2 and Layer 3 methods exist to solve these problems these methods are not easy to implement and configuration and operational errors can easily cause harm to the network or put business interests at risk.

As was mentioned, many operators will duplicate their video network paths in order to create a level of redundancy that protects from physical network failures. Since multicast implementations are a real time transmission of MPEG Transport Streams over UDP/IP there is no ability for receiving devices to requests retransmission of data that might a have been lost by the network. While both standard and proprietary methods in support of UDP transmission have been developed to support retransmission of lost data or recovery via adding Forward Error Correction, most operators have not implemented these solutions and continue to operate primary and redundant video delivery network paths to maintain the desired level of Quality of Service (QoS). Receiving devices are responsible for determining which multicast signal is the better one to use; methods for making this determine can vary from implementation. In comparison, the duplications of equipments and network paths incur added cost and complexity with custom design that does not exist in general Internet technologies and standards.

In addition to lack of native resiliency in the real-time transmission of MPEG TS in UDP/IP there is also a lack of security. As was mentioned, when multicasting MPEG TS there can be operational difficulties to prevent one or more multicast streams from reaching an administratively forbidden destination; multicast can be disabled on a network segment but if some but not all multicast traffic is required to be present it can lead to operational complexities. Conditional Access and Digital Rights Management (DRM) are valid solutions to this problem but often do not scale well inside an operator core video network and these solutions are often operational complex.

A more modern solution leveraging HTTP unicast adaptive video delivery is ready to solve the challenges of deliverying linear video for QAM H.264 HD set-tops.

## HTTP ADAPTIVE BIT RATE

Multi-screen video delivery has grown substantially in recent years. Rather than leverage any existing processing or delivery platforms in place for QAM, service providers have built parallel delivery platforms. These new platforms have taken advantage of COTS hardware and new methods for ad insertion that do not require proprietary hardware platforms to implement. However, other than leveraging the same content sources as are used for QAM delivery, multi-screen video platforms have been built in parallel to the traditional network (`Figure 1`).

Multi-screen linear video is encoded as H.264 and packaged and delivered over HTTP to consumer devices through a Content Delivery Network (CDN). Adaptive Bit Rate (ABR) is used to ensure a smooth playback experience for the consumer regardless of varying network conditions.

The video formats (or profiles) set as outputs of the ABR Transcoder have traditionally been targeted at small screens, as such the resolutions of the encoded video tend to be less than that of the original content source.

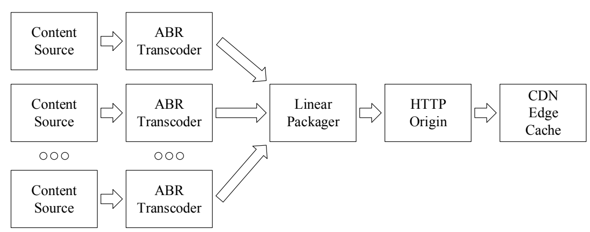


**Figure 2 - Linear HTTP ABR Delivery**

However as service providers consider supporting IP set-top boxes or higher-resolution displays such as computers or tablets it becomes necessary to retain the resolution of the content source but still deliver a seamless user experience during sub-optimal network conditions. As such, ABR is still a necessity.

As service providers have started to deploy cloud DVR (cDVR) services or linear streaming to big screens (connected TVs, game consoles, etc.) the resolution and bit rates have increased beyond what was originally deployed for small mobile screens. As the number of delivery formats (multiple delivery protocols and different DRMs) has increased service providers have turned to using Just In Time Packagers (JITP). JITP ingests a single common format inside the service provider network and converts it to a device-specific format for delivery to a specific consumer platform. A similar approach can be taken with QAM such that a common linear input ABR format can be converted for delivery to a QAM set-top box.

As was discussed previously, UDP transmission of video is not protected against packet loss. Additionally, managing the flow of multicast traffic over specific router-to-router links can be challenging. With HTTP being based on TCP there are inherent retransmission capabilities. Widely deployed methods also exist to manage the traffic flows around the network with HTTP; CDN technology can be utilized in order to deliver content data only when requested. Other widely used IP network techniques for scalability and redundancy such as load-balancing, clustering, and virtualization are applicable for HTTP/TCP where UDP falls short.

A significant business challenge for service providers had been caused by technical limitations; securing multicast content on a core network has proven not only to be difficult but operationally complex. Securing that same content transmitted over HTTP can be as simple as using HTTP with TLS connections but added layers of security using software-based DRM are also feasible. Beyond a blanket protection of the content, the use of DRM and CDNs would allow service providers to easily manage and prevent where specific content is not authorized for distribution due to business constraints.

Even with HTTP based ABR IP video delivery in the core network that can be shared for both IP and QAM H.264 HD linear traffic, edge QAMs still require UDP/IP multicast carrying MPEG-2 Transport Streams as their inputs; separate or multiplexed audio and video segments delivered through ABR IP video at the core still need to be converted to MPEG-2 SPTS or MPTS format before delivering to the edge QAM devices. Linear content requires ads to be inserted before streamed to edge QAMs. It is also desirable to leverage the benefits of statistical multiplexing but without the burden of proprietary hardware platforms.

## HTTP AD INSERTION

Standards-based methods have already been developed and deployed to simplify ad insertion in the HTTP realm as compared to ad splicing in Transport Streams. A common approach for HTTP ad insertion places additional stream processing responsibility on the ABR Transcoder and Linear Packager. The ABR Transcoder must detect SCTE-35 messages in-band and properly align the encoding frame structure such that a complete Group-of-Pictures (GOP) exists before and after the temporal locations of the ad (and not spread between the content and ad). The linear packager must also recognize ad placement locations and the boundaries created by the transcoder and so it can identify replaceable segments in the playlist it creates for the down-stream devices.

Now that the heavy lifting to cleanly mark where an ad starts and stops has been taken care of by the transcoder and packager any down-stream device need only evaluate the playlist and insert advertisements downloaded over HTTP that are similarly formatted with clean frame boundaries.

So, all of this exists, nothing is *new*. How do we get from an HTTP delivered ABR stream sitting on a CDN cache to an edge QAM that does not speak HTTP or understand ABR? The solution is simple: create a multiplexing gateway platform to convert between HTTP ABR and UDP/IP Transport Streams.

## HTTP ABR TO UDP GATEWAY

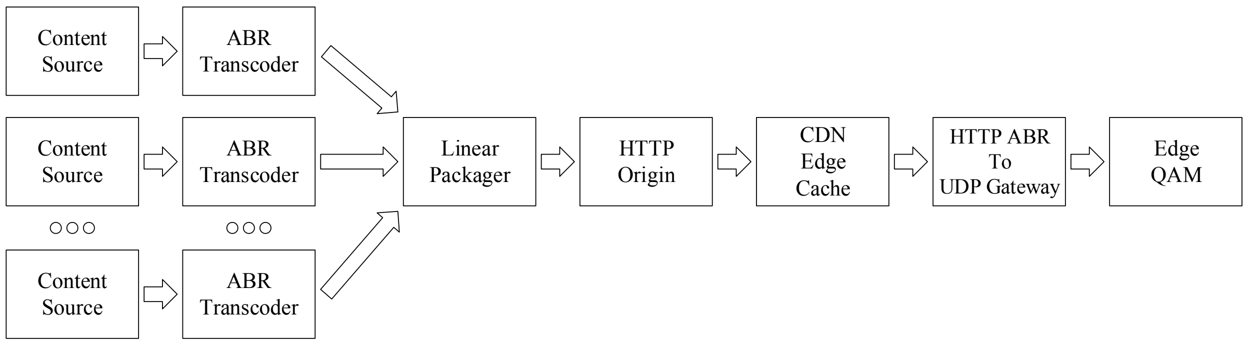


**Figure 3 - HTTP ABR To QAM Gateway**

The new HTTP ABR To UDP Gateway must act like an HTTP ABR client to ingest content including acting like an HTTP ABR ad insertion client. In the case of SPTS UDP/IP output the gateway simply needs to retrieve the designated ABR profile(s) from the CDN Edge Cache, perform the necessary HTTP ad insertion using the ad content profile that is less than or equal to the bit rate of the linear content ABR profile and then stream it to the edge QAM over UDP/IP as a Transport Stream. In the process any PID and Program Number remapping of the ad content should also be performed by the gateway.

MPTS output is a bit more complicated. While ad insertion is taken care of in the HTTP domain as previously described there is still a desire to take advantage of statistical multiplexing. However, doing so in proprietary hardware is not a scalable or desireable solution for many service providers. Luckily, the ABR transcoder and packager have already done the heavy lifting by creating a variety of bit rates for the same resolution video and marking the location of ads. Having the Gateway evaluate the ABR playlist by selecting an appropriate input bit rate profile more eloquently performs statistical multiplexing decisions compared to those previously performed in proprietary hardware.

Example

10 HD H.264 streams are configured to be multiplexed by the Gateway and transmitted as a 38.8 Mbps UDP/IP multicast stream which will be received by downstream edge QAMs. Simplistically as CBR encoded streams the average bit rate would be 3880 kbps. In this basic example (`Figure 4`) the highest encoded profile of each HD stream is actually 5500 kbps with additional profiles of 3750 kbps, 2750 kbps, and 2250 kbps. ABR content is packaged and placed on the CDN and delivered to the edge cache. The statistical multiplexing gateway retrieves the content from the edge cache as files over HTTP. With content encoded at fixed segment intervals (i.e. 2 seconds) the gateway is able to add up the segments from each of the 10 streams and ensure that they are always less than 38.8Mbps as shown in `Figure 4`.

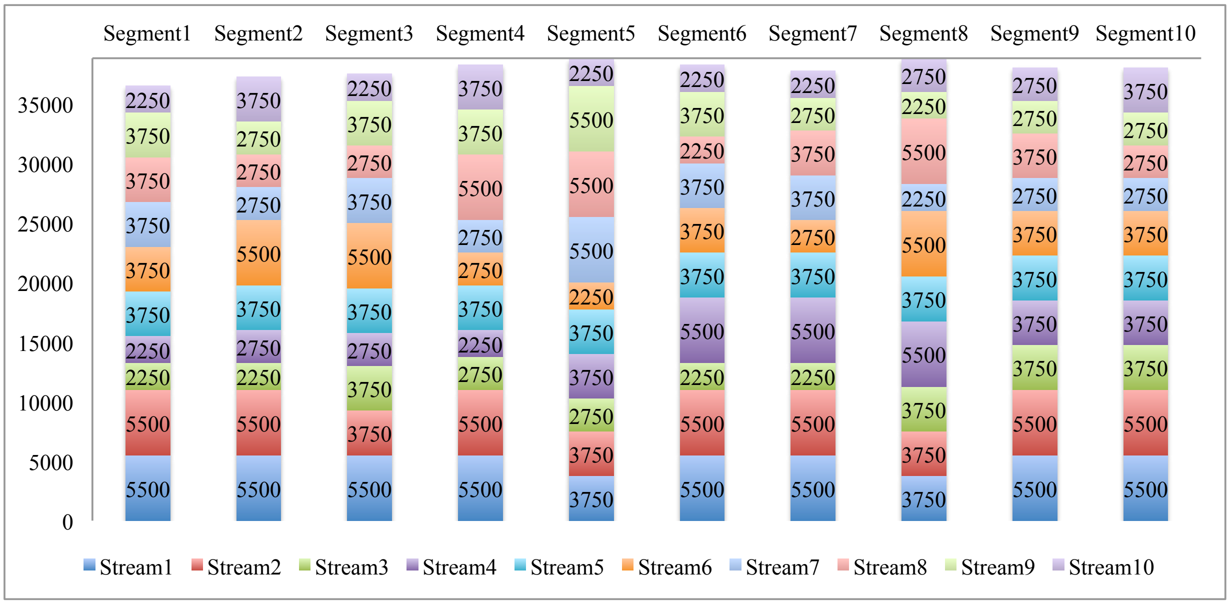


**Figure 4 – Example ABR Stat-Mux**

Increasing the number of profiles (bit rates increments) will allow for higher picture quality permitting the statistical multiplexing decisions to be more granular. Additionally, metadata regarding picture quality can be provided from the ABR transcoder downstream to the Gateway that will allow more intelligent multiplexing decisions about upcoming segments and which profiles to incorporate into the output.

Ad boundaries will add to the variability of the statistical multiplex since they will create scene changes (fade to black) and variable length segments. As ad content is transcoded offline in non-real-time it can be prepared with additional profiles that allow for more granular rate shifting during ad breaks which could potentially occur concurrently across multiple channels in a multiplex.

## Implementation

The HTTP ABR to UDP Gateway acts like an ABR client and transmits the output as an MPEG-2 Transport Stream over UDP/IP. The profile decision process is different than a consumer-device ABR client as decisions are not made on available network bandwidth at the input but rather available bandwidth on the multiplexed output and is done in conjunction with multiple ABR streams simultaneously. These processes are not computationally complex and do not involve manipulation of the video payload like traditional statistical multiplex or ad splicing. While there is some basic manipulation of MPEG-2 Transport Stream data (Program Number remapping and PID remapping), and while concern must be given to the timing of multiplexing packets it is the expectation that all of these functions can be implemented on COTS computer server hardware at scale and can potentially be virtualized.

## CONCLUSION

With the confluence of the increasing deployment of high resolution and bit rate HTTP ABR/H.264 and decline of MPEG-2 HD set-tops service providers can converge their core video distribution network to HTTP ABR, increasing bandwidth efficiency by eliminating redundant multicast core video networks for linear video delivery to H.264 HD QAM set-tops. Content security is increased and QoS is natively provided. Not only does using HTTP delivery reduce the overall network management but also it unifies ad insertion into a common platform and simplifies statistical multiplexing at the same time. HTTP ABR delivery is the next evolution of the multichannel service provider's linear broadcast core video network.