Content Aware Video Streaming

A Technical Paper prepared for SCTE•ISBE by

Srinath V Ramaswamy
Principal Solution Architect
Comcast
1717 Arch Street, Philadelphia, PA 19103
(215) 286 5444
srinath_ramaswamy@cable.comcast.com

Table of Contents

List of Figures

List of Tables

1. Introduction

Streaming video represents nearly 80% of all traffic carried over the internet. It's growing by 22% each year [1]. Technological evolution in video compression, from MPEG-4/AVC [2] to HEVC [3], paired with techniques like Adaptive Bit Rate (ABR), to "right size" a video segment into multiple bit rates, represent some of the core building blocks that contributed to the huge marketplace success that is streamed video. However, a third building block is becoming necessary, to go a step further in both an improved, adaptive video experience, at the device end, and optimized bandwidth usage, to lessen network strain. Specifically, and considering the remarkable and mushrooming growth of IP video streaming, there exists a growing need for systems to leverage Content Aware Video Streaming solutions, like the one discussed in this paper.

In IP video streaming, video is typically generated at several bitrates, in order to accommodate the varying network bandwidth available between IP Video consumer devices and content delivery networks (CDNs) and varying consumer devices. Consumer devices will typically try to download the highest bitrate video possible that the device can support, with the intent of providing higher video quality to the viewer. Higher bitrate video segments for certain content do not necessarily translate to higher video quality -- if there is any improvement at all, it might be marginal. Acquiring higher bitrate segments when it does not really enhance user video viewing experience results in wasted network bandwidth. This wastage can be avoided by having the devices only download segments that would make a difference in video quality for the viewer. For the devices to do a quality-based download decision, they would need to be provided with info on the content in the segment. In this paper, an algorithm that determines if a segment is going to enhance viewer video experience is presented. This algorithm is based on video analysis without a full video decode of frames in the segment. This algorithm would need to be implemented at the content origination end and the generated info transmitted to consumer devices.

With the content information communicated in each segment, devices can then make a network bandwidth and quality based decision on the next segment to be downloaded. This would minimize network bandwidth usage and at the same time providing the highest video quality. The algorithm presented supports processing of live linear stream segments and would minimally impact the content generation workflow and therefore scalable. Included in this paper are results showing network bandwidth savings realized by this algorithm. This algorithm could also be used to reduce storage needs by only storing segments from a bitrate ladder that enhances video quality on various devices.

2. IP Video Architecture

At a high level, the IP Video architecture is as shown in the Figure 1 below. In this figure, the compressed or uncompressed video stream is transcoded into several streams (variants), each at different fixed bitrate, to accommodate the varying network bandwidth available between IP Video consumer devices and the content delivery network. The transcoded streams might be encoded using one of the MPEG compression codecs (MPEG-2[4], MPEG-4/AVC, HEVC). This is then packaged into one of the many streaming formats, for example MPEG-DASH [5], or HLS [6]. These are then placed on the Origin Server and IP CDN for delivery to IP Video clients. This would apply to both linear and on demand IP video.

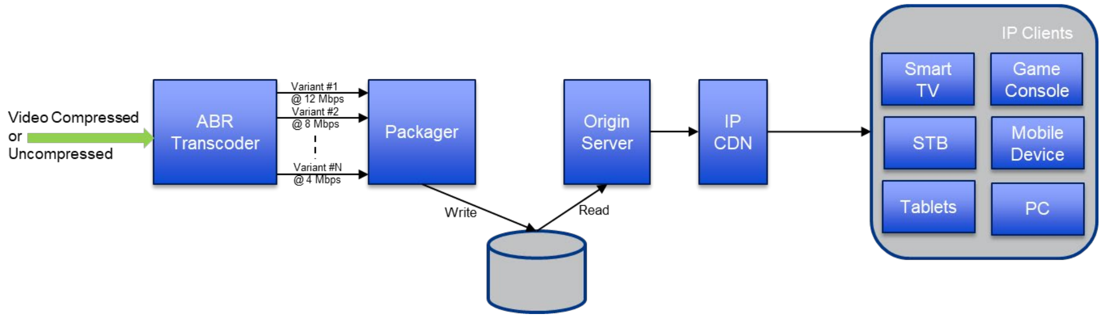

Figure 1 – IP Video Architecture

3. Segment Video Analysis

The IP Video variant streams shown in the Figure 1 are typically split into segments of a few seconds duration, and these segments across the variants are assumed to be time-aligned. The IP Video variant streams are part of a constant bitrate ladder as shown in Table 1, below:

Table 1 – Constant AVC Video Bitrate Ladder

Resolution	Video Bitrate (Kbps)
1920x1080	6000
1920x1080	4000
1280x720	3000
1280x720	2000
720x480	1600
640x480	1000
512x384	700
384x288	400
384x288	300

An IP Video consumer device traverses up and down this ladder, based on varying network conditions between it and the CDN. During stable network conditions the device might be playing segments from the 6 Mbps video stream but for some of these segments it might be of the same quality as the 4 Mbps segments, therefore the device could downshift to the 4 Mbps segments for those segments and then go back to the 6 Mbps segments when there is an actual gain in the video quality by doing so. By taking this approach network bandwidth usage is reduced. Similarly, in another scenario when the network condition improves after a degradation and the device sees that the network can sustain a higher bitrate, the device which is currently playing a 4 Mbps video stream would try to display a high quality video by moving up the bitrate ladder. In this case, it would try to download and play a 6 Mbps higher bitrate variant segment, even though the gain in the video quality by doing so may be minimal. So in both these scenarios, during the download process, if the device is told that downloading the 6 Mbps segment would result in a minimal video quality gain, then it could avoid the unnecessary download, which saves network bandwidth. Therefore, there is a need for an approach that can detect if a segment would enhance video quality and communicate that to IP Video devices. Also, with this information, the device could look at downloading the next higher bitrate variant segment that really would improve the user experience.

Similarly, when the network condition is worsening and the device is moving down the bitrate ladder, and if it is told that the next lower bitrate stream does not enhance quality, then it could select the one below it in the bitrate ladder, again saving network bandwidth.Video Analysis Algorithm Details

In this section, an encoded video analysis algorithm is detailed that determines if an IP video segment does not enhance video quality and is therefore optional to download. The results of this algorithm are then communicated to the IP video device, which would utilize it to determine the segments to download, in order to save network bandwidth and enhance video quality.

The algorithm analyzes all encoded video frames in a segment to determine if it is optional to download the segment. This is done in the following manner:

1. The compressed bitstream in the segment is first entropy decoded and inverse quantized to access the many frames in the segment and its compression details.
2. The frame type (Intra [I], Predictive [P], Bidirectional [B]) is determined for each frame in the segment. The frame type determines how that frame is going to be analyzed to determine its contribution to the segment relevancy.
3. If the frame is an I type, we determine the frame's compression ratio (defined as the number of bytes to represent the frame / the number of bytes to represent the compressed frame). If the compression ratio is less than a predetermined threshold, then this frame is marked as a frame with high detail and activity. If the compression ratio is low, it would mean that there are a lot of details in this frame and less redundancy.
4. If the frame is a P or B type, we again determine the frame's compression ratio. As with I-frames, if the compression ratio is less than a predetermined threshold then this frame is marked as one with high detail and activity. If the compression ratio is higher, then some of its compression attributes are analyzed to determine if it would still qualify as a frame with high detail and activity. The compression attributes that are analyzed are listed below:

    - *Number of skipped Macroblocks/Coding Unit (CU):* The number of skipped macroblocks in the frame is determined. If it exceeds a certain threshold, then it is counted towards the determination of a high level of detail and activity and is marked as a High Skipped MB/CU frame. A frame with a lot of skipped macroblocks indicates that there isn't much motion between this frame and its previous frames, used for motion estimation. A macroblock is the basic processing unit in AVC/MPEG-2, and the CU, in HEVC, is where the prediction type is decided, as well as which frames are skipped when their motion vector is zero and coefficients are zero.
    - *Partition (Prediction) block size:* A prediction block is the block split from a macroblock or coding unit that is used for motion estimation. This attribute determines the size of the prediction block that is the most widely used in this frame. If it exceeds a certain threshold, then it is marked as a Large Partition Block frame. A comparatively small size indicates a lot of details in the frame.
    - *Inter coded blocks count*: Determines the number of inter-coded blocks in frame. If the number of inter-coded blocks count exceeds a threshold, then it is indicated as High Inter-Coded Block Count frame. A low number of inter-coded blocks implies that there is a scene change, therefore unable to predict from reference frames.
    - *Motion vector component measure*: Determines motion vectors for each inter-prediction partition block in the frame and determines the 90% trimmed standard deviation of the motion vectors' horizontal and vertical components. The 90% trimmed standard deviation of the components is to ignore extreme values to give a better measure and is determined by ignoring the lower 5% and top 5% of the values. These vectors are

encoded differentially, with respect to predicted values from nearby vectors. If greater than a certain threshold, it would be considered a high MV measure frame. Large variations or standard deviations would indicate a lot of activity.

Note that the thresholds used for these compression attributes for P frames are different from those used for B frames, and these thresholds for P and B frames also factor in video frame rates and resolution.

The information gathered by analyzing these compression attributes is then used to determine if this P or B frame is a high detail and activity frame, as shown in Table 2, below:

Table 2 – P/B Frame High Activity and Detail

High Skipped MB/CU	Large Partition Blocks	High Inter Coded Blocks Count	High MV Measure	High Activity/ Detail?
False	False	False	True	True
True	N/A	N/A	N/A	False
False	True	False	True	True
False	True	True	True	False
False	True	True	False	False
False	False	True	True	True
False	False	True	False	False
False	False	False	True	False

Once all the frames in a segment have been analyzed using the approach detailed above, the percentage of high activity/ detail frames in the segment is determined. If this percentage is below a certain threshold value, then it is marked as a segment for optional download. So, when an IP video consumer device is trying to display a higher quality video by moving up the bitrate ladder and encounters this optional download segment, it would not download it, because there is no appreciable gain in video quality by doing so. It would, however, consider downloading a higher bitrate segment that is not marked as an optional download.

Table 3 shows the resultant bitrate ladder after the video for segment "n" has been analyzed, using the algorithm detailed in this paper:

Table 3 – Bitrate Ladder after Video Analysis for Segment n

Resolution	Video Bitrate (Kbps)	Download for Higher Quality?
1920x1080	6000	No
1920x1080	4000	Yes
1280x720	3000	Yes
1280x720	2000	Yes
720x480	1600	No
640x480	1000	Yes
512x384	700	Yes

Resolution	Video Bitrate (Kbps)	Download for Higher Quality?
384x288	400	No
384x288	300	Yes

This video segment analysis does not need to be done for all the video segments across the variants -- only those necessary to determine a high level of activity and detail for each segment time duration. For example, each segment duration could be 4 seconds across the variants. Since the encoded video analysis is done without a full decode, this algorithm does not add latency, in the case of linear video, or increase encoding time, in the case of non-linear video.

3.1. Enhanced IP Video Architecture with Video Analysis

An enhanced architecture that utilizes the algorithm detailed in this paper is shown in the Figure 2 below. The key difference between this and the architecture shown in Figure 1 is the addition of a functional block that processes the IP video segments and generates content-specific information sent to the client.

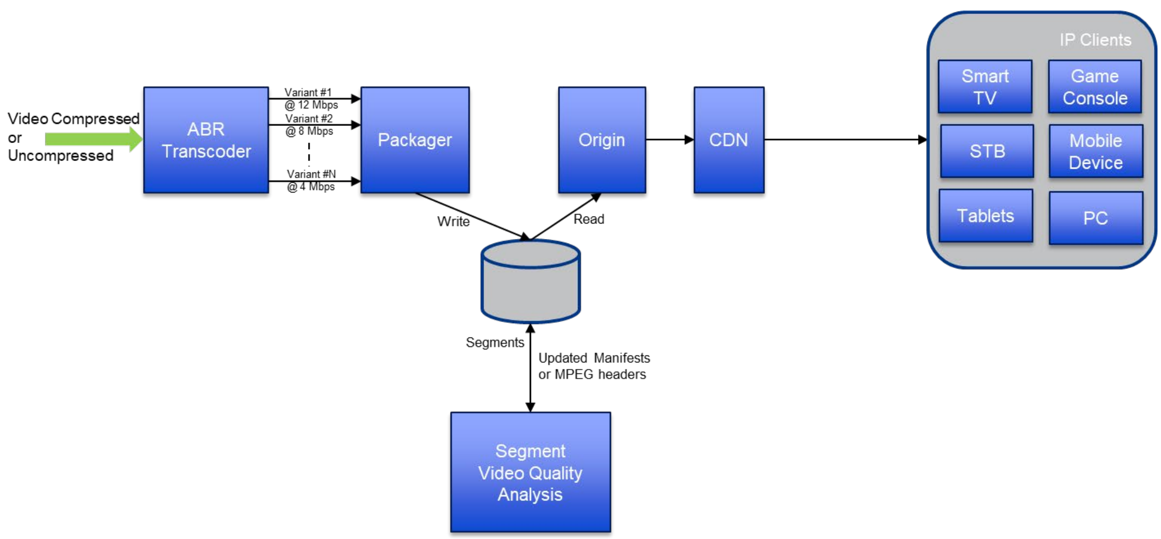

Figure 2 – IP Video Enhanced Architecture

4. Carriage of Optional Download Information

There are multiple approaches to send the optional segment download information to IP Video consumer devices.

One approach would be to signal those segments that are optional to download in the manifest files. For example, in the case of DASH, this would happen in the Representation element; in HLS this would happen in the Media Playlist. In the case of live linear, as the segments are generated, the optional download information is determined and included in the updated manifest files. When signaled in the manifest, the consumer device could base its next download decision on this value.

Another approach would be to remove the segment that is optional to download from a Representation, in a DASH manifest file, or, in the case of HLS, to remove the media file URL from the Media Playlist.

Potentially one other approach would be to indicate the optional segment download information via fields in the segment. For example, in user data fields in the MPEG-2 Transport packet headers, or ISOBMFF m4s file box structures.

Please note that none of these are currently standardized.

5. Benefits

There are many benefits reaped by utilizing the Content Aware streaming solution described in this paper. Primarily by not downloading higher bitrate segments that does not necessarily provide a gain in video quality streaming network bandwidth usage is reduced. Lab studies have shown a reduction of 20%.

Since this algorithm identifies segments that do not contribute to higher video quality, one approach would be to not store those segments and instead point the consumer device to an alternate segment that has the same quality as this one. This would save storage in applications like Cloud DVR, Video On Demand and CDN. Another benefit from this is that since the segments are of smaller size during certain times, the downloads will be faster.

The algorithm detailed in this paper does not do a full decode of the compressed bits in a segment but does partial decode resulting a fast video quality assessment.

The intent of this solution is to not impact the existing transcoders and packagers but execute this algorithm on the content generated by packagers before they are made available to IP video consumer devices.

6. IP Video Network Bandwidth Utilization

Figure 3 shows instantaneous video bitrate when consumer device is downloading segments with and without using the solution detailed in this paper. The black curve plots the video bitrate when segments are downloaded without looking at the segment quality whereas the blue curve shows the bitrate when the consumer device downloads segments based on its quality. So, for example at Time 7 the device downloads a segment from a lower bitrate stream that has the same video quality as the segment downloaded at the higher bitrate indicated on the black curve. It is observed that the video quality is the same with the two type of downloads.

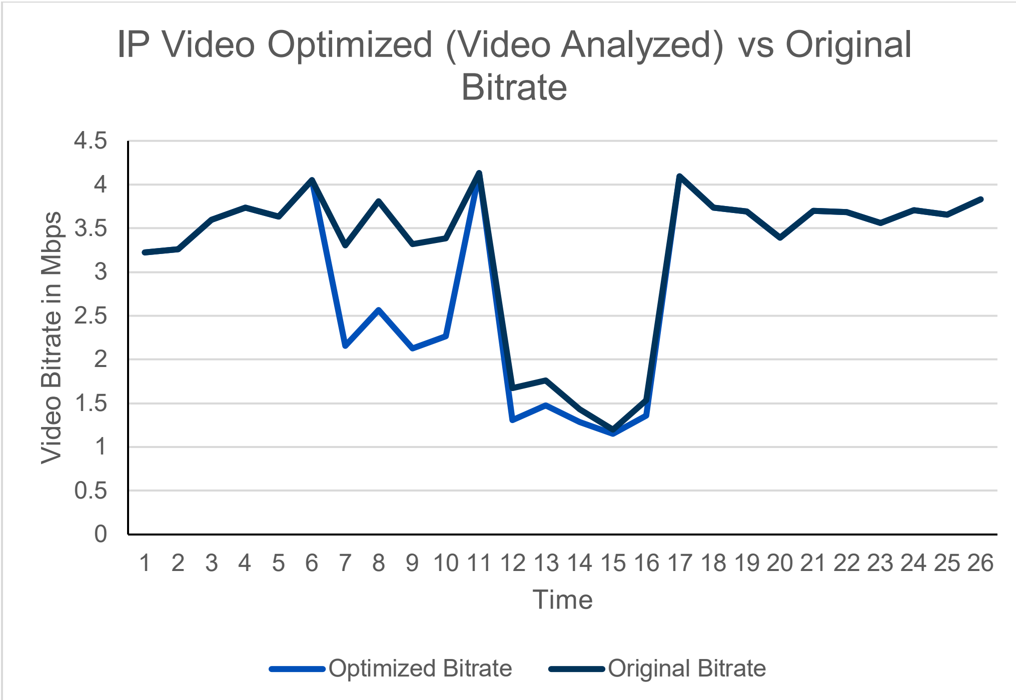

Figure 3 – IP Video Network Bandwidth Utilization

7. Conclusion

IP Video Streaming is widespread and consuming valuable network bandwidth – as in, almost 80% of all internet traffic. This traffic is expected to grow by 22% yearly. Considering this growth, it is important for systems to leverage Content Aware Video Streaming solutions such as this to optimize network bandwidth usage. In this paper an algorithm is presented that determines if an IP video segment is going to enhance viewer video experience. This algorithm, which analyzes encoded video without a full decode, in linear IP video. Also included are approaches to communicate an IP video segment's relevance to enhance viewer video quality to IP video devices. In addition to optimizing network bandwidth usage and adapting to changing network conditions, it also reduces storage requirements and enhances consumer video viewing quality.

Abbreviations

ABR	Adaptive Bitrate
AVC	Advanced Video Coding
CDN	Content Delivery Network
DASH	Dynamic Adaptive Streaming over HTTP
HEVC	High Efficiency Video Coding
HLS	HTTP Live Streaming
IP	Internet Protocol
ISO-BMFF	International Standard Organization – Base Media File Format

MPEG	Motion Picture Experts Group
URL	Uniform Resource Locator

Bibliography & References

[1] https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/United_States_2022_Forecast_Highlights.pdf

[2] H.264, Coding of Audio-Visual Objects — Part 10: Advanced Video Coding (AVC), ISO/IEC 14496-10:2010.

[3] H.265, High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding, ISO/IEC 23008-2:2013

[4] H.262, Generic Coding of Moving Pictures and Associated Audio Information (H.262), ISO/IEC 13818-2:2013.

[5] Dynamic Adaptive Streaming over HTTP (DASH), ISO/IEC 23009-1:2014

[6] R. P. Pantos, editor, HTTP Live Streaming, https://tools.ietf.org/html/rfc8216

