Video Ingest for the Cloud: Common Video Subscription Support for Advancement of Multi-Video Service Architectures

By Yasser Syed, Ph.D.
Dan Holden
Comcast

ABSTRACT

This paper discusses how a video model can be developed and evolve from preparing content for a single specific service (e.g. QAM) to an architecture where multiple types of video services (e.g. SDV, Adaptive Streaming, Mobile) can subscribe to a common set of video feeds. Moving towards this architecture can increase scaling efficiencies in transcoding and monitoring, improve content quality consistency across platforms and allow for innovation in services/platform/devices. Legacy services and new multi-streaming technologies will be discussed. Additionally, this paper will cover common encoding approaches for acquisition, transcoding and service wrapping. These types of architectures will be needed to transport content efficiently for existing and new delivery infrastructures.

INTRODUCTION

Video services have always determined how content sources have been acquired. Until recently, cable video services have been a single linear service for delivery to the television. Initially, it was an analog transmission in 6Mhz channels, then digital information (also over an analog transmission), but intended for delivery to a television display. The video was encoded for linear delivery and the costs were burdened on the service.

Since then services have changed into non-linear VOD, switched digital video (SDV) and Internet streaming, but still the burden of the costs for video encoding has been taken on by the individual services.

With cloud-based delivery networks, there is one ingest point for content, but there are many services and platforms being addressed: VOD, linear broadcast, PCs, tablets, TVs, smartphones and personal devices. With the traditional connection between services and content acquisition so strong, we have to examine if there is a better model that could save costs and allow us to design and initiate services in a faster manner.

A NEW APPROACH

In this section, we examine the benefits of designing content acquisition systems to beg more loosely tied to the service. With recent advancements in networking and technology, convergence in content acquisition is possible. Though there are more services today, the end customer experience is nearly the same.

There is a convergence of content in many of the services being planned or offered. It's either the same linear feeds or the same VOD content. What determines which content it is delivered to the service is dependent on physical availability and license determination. Physical availability is how much more effort in design and costs it takes to acquire the content in the service. In addition, the new services are negotiating licensing fees for sources and files. It is easier to negotiate for content that already exists and is provided for customer consumption rather than create new content for the service. There is little new content for any specific service due to the cost to exclusively produce and promote the content. From this, there is a convergence of type content demanded by video services.

Even with this, the licensing negotiation process can determine how, and in what format, source linear feeds and content files are acquired. This is determined by an acceptable content acquisition expense and expected ROI for that service. The negotiated costs can often dictate the final quality of these feeds and files.

The new approach separates the content acquisition expense from the expected ROI of any particular service by creating a distribution network for the content to which services can subscribe and deliver. The quality of the content sources will not be dependent on the negotiated licensing fees, but on existing distribution output delivery channels. It is expected that the content sources will be the same for most services and that standardizing content acquisition and processing for all services will mitigate and distribute the costs across services. It creates a content distribution channel system that can be matched to multiple services including a CDN delivery architecture. Also, quality and customer experience can be consistently maintained across all services.

There have been a number of recent technology improvements and convergences that have made designing standardized content acquisition and processing practical. One is network multicast distribution which allows multiple processing devices to pull from the same source. An early contributing factor was the development of an edge QAM. This also simplifies the physical planning where an in-place GIG-E network does not mean adding an entire separate physical infrastructure for each new candidate service. Additionally, there has been a convergence in the video technology area where standardization in the industry has reduced the number of supported video service codec formats needed. This is also happening at the transport level where standards and specifications have been developed that can support these popular video formats (e.g. SCTE 128, 14496-15).

There are also convergences happening on the consumer device side. There are far fewer video codecs that need to be supported in these devices. Resolutions, bit rates and frames rates are approaching fuller video rates. And delivery bandwidth is approaching rates that are acceptable to carrying a full screen video. Applying these same advantages and convergences with a common content acquisition and distribution structure will result in a reduction of costs for getting the same content (or subset) to services sourcing from these devices. It will also allow newer devices to anticipate and design for these types of signals. Lastly this architecture will be able to maintain content for an expected quality and customer experience for all services delivering to these devices and separate these issues from delivery factors

Legacy Platforms

Applying these principles can have multi-generational advantages with legacy services receiving immediate benefits. We define legacy services as MPEG-2 distribution structures for Broadcast, SDV and VOD for High Definition (HD) and Standard Definition (SD) sources.

Multicast distribution allows the same source encodes to be used for SPTS and MPTS solutions since the core elementary stream is MPEG-2. Development of a distributed multiplexer allows MPTS streams to be created from a set of SPTS sources. The distributed mux can statistically multiplex media services together in an efficient way either at a cable headend or in a national feed. The MPTS combined streams feeds broadcast to MPEG-2 STBs, while the separate SPTS feeds service SDV and HITS where the channel line-up can be determined at the headend side while the SPTS are distributed by fiber or satellite.

Other generational improvements are to create SD streams from the same corresponding HD sources. This is accomplished through processes such as center-cuts/ pan and scan, and adding in progressive to interlace conversions. With improvements to encoders, more than one output stream can be developed from the same input source stream.

These and other first generational improvements benefit the existing distribution infrastructure while enabling a common content acquisition and distribution infrastructure.

Multi- Streaming Platforms

Two developments occurring now aid in creating the next stage in this approach. One is the advent of adaptive streaming technology and the second is the plethora of personal video devices emerging that can operate in managed and unmanaged networks.

With the emergence of Over The Top (OTT), a video service delivered over the Internet to the PC, TV, handheld pads and mobile devices, content providers like You Tube, Hulu, Netflix, Blockbuster, Vudu, Boxee, Sony, Apple, Samsung, and a long list of others, are providing media content options to consumers in a session based, streaming delivery format. The next step is delivering live content to these same devices via the same adaptive streaming technology in a session-based delivery format. This session-based delivery format is an opportunity to compete with the OTT providers on these stand alone devices, and it opens up more programming opportunities..

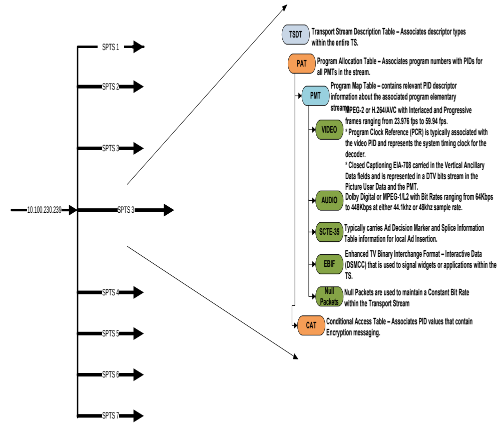
Adaptive streaming develops the idea that a single encoding process can create multiple aligned transcoded outputs. Before this, most encoding processes resulted in a single output, but now alignment between streams needs to be coordinated enough to switch between bit rates through the concept of a fragmentation structure (a grouping concept separate from a GOP (Group of Pictures) structure but can be used to create a session-based delivery format). To achieve this, a unified distribution network for specific content should be in place. Additionally, encoding platforms must be developed to handle multiple output streams that can extend beyond a single hardware box/ software set of processors.

Developing the content sources for this system involve working with the same content providers along with same limitations of present day content source delivery infrastructures (satellite, network) as before. The separate bandwidth generated by this process is quite large. It is not feasible to increase the incoming plant bandwidth just to provide the same content, but for an adaptive streaming service. It is more reasonable and achievable to transcode content that was already there for broadcast purposes. With adaptive streaming this became an early case of using a common mezzanine/contribution format that could be transcoded into a second distribution format.

There are also other advantages of using a common source format. For instance along with the broadcast content source, there is SCTE 35 information that the source stream contains, This can be reutilized by the adaptive streaming service to create its own ad-insertion features in the service.. If a separate content source were used instead, it would be difficult to justify the extra logistics and costs for putting this ad triggering signal in place. In this case, this important revenue feature is added to the service at minimal cost..

One consideration for designing the adaptive streaming service was the number of streams, bit rates and resolutions. In terms of transcoder performance, adding lower bit rate streams and resolutions is not a tremendous burden. There is commonalities between personal devices and television in terms of bit rates and resolutions; many devices can use a single stream or subset of streams for there own needs without having to use the entire adaptive streaming set. The resolutions, bit rates , video format (H.264/ AVC or MPEG-2) and the elementary streams structures are compatible, yet with changes above the ES (elementary stream) layer can make this suitable for different services. For adaptive streaming, this happens at the point of fragment creation where the layer above the ES layer is stripped out and replaced. Separating the transcoder processes from the fragmentation processes (service distribution wrapping) allows for the output of the transcoders to be reused including devices outside of the adaptive streaming or traditional broadcast technology. The transcoded streams could be readapted to use in managed or unmanaged networks through the use of lower-cost reformatting devices.

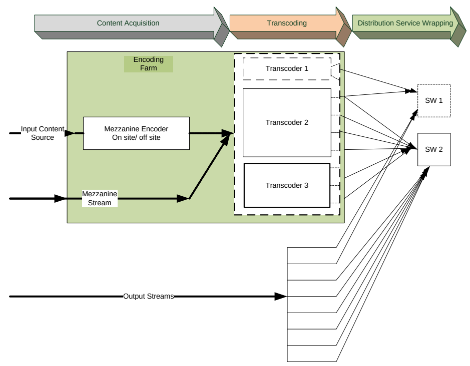
With the right selection of stream resolutions, bit rates, and codecs in a standardized multicast stream format, a number of services could ingest commonly demanded content by subscribing to these specific streams and reformatting them for their own purposes.



Universal Encoding Platform

Some of the basic groundwork for the Universal Encoding Platform has been laid with the development of multi-stream encoding platform. The idea of mezzanine/contribution sources, multiple transcoded H.264/AVC output streams and repurposing of elementary streams are established and proven that it can co-exist with an MPEG-2 distribution system using the same content sources in a multicast environment.

Expanding on this groundwork is facilitated by a set of convergence points occurring in acquisition, transcoding and distribution spaces. In each case where this does not happen for specific content channels, the content processes need to be individualized with extra costs in tools and formatting.



Acquisition

The acquisition convergence point standardizes the ingestion of content and affects the overall quality of the content as it is processed downstream. It can be the place to standardize feature for common customer experience across all services for that content channel. It also provides a natural demarcation point to distinguish between content source issues from downstream factors. Below are expected convergences:

1. Creating common specifications for Mezzanine/Contribution feeds

- Encoders can be placed at MSO or content provider sites
- Direct inputs into a transcoder system that can multicast into multiple devices.
- Formats are limited to standardized inputs such as AVC/H.264, MPEG-2 at bit rates up to 50-60 Mbps and common mezzanine audio format.
- Video would support resolution formats up to 720P or 1080P with the top tier HD resolution supporting frame rates at 60 (59.94), 24 (23.98), and 30 (29.97).

2. Set of Video & Audio pre-processing and filtering operations

- Convert all sources to native frame rates/picture formats to allow downstream service transcoders to adjust frame rates where needed.
- Normalize audio levels before passing down to service transcoders.

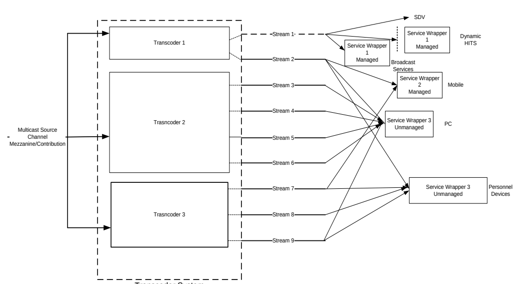
3. Video and Audio encoding processes to optimize quality rather than bit rate

- Video encoders output near-constant quality feeds rather than constant bit rate.
- Intelligent encoding that can signal less expensive downstream transcoders how to process/recover from errors.
- Can also be a single source point to correct content issues that can be sourced to all services.

4. Insert or pass-through auxiliary data information or trigger points

- Also a single source point to update data information to all downstream services.
- Downstream transcoders or service distribution wrappers only pass through information needed for that service.

Transcoders



The transcoding system creates the set of output elementary streams that the service wrappers subscribe to for distribution. The trick here is to establish a stable set of elementary streams based on resolution, frame rate, temporal compression structure and bit rate that are acceptable for multiple services in both a managed (QOS) and unmanaged (OTT) service environments. Some of the lower bit rate streams can also be used for auxiliary applications such as guides.

The adaptive streaming technology for live linear services is a driver because it establishes the concept of aligned multiple transcoded output streams that feed into a service wrapper (a multi-channel dense fragmentation/ encapsulation/ DRM process).
This transcoding point is also useful to retain higher output video quality by tailoring video filtering according to output bit rate and resolution. For example, reducing film grain noise through MCTF filters at lower bit rates allows the transcoder to spend fewer bits on processing noise (which would be lost anyways) and more on the actual video quality. Additionally, perceptual-based video filtering could be completed at lower video rates to help maintain video quality..

An important convergence change here is to treat this as a transcoder system, not a device, outputting a set of multiple bit streams, but can be created using more than one device. This will allow a subset of bit streams to be created by a set of cost effective optimized transcoders rather than limited by one device. It also breaks apart the performance limitation of restricting the number of output streams based upon a single device platform, which can then further reduce the costs of the encoding system through competitive activity. Additionally costs over time can be reduced by putting more functionality in the mezzanine/contribution feeds such that transcoders can operate more simply through cues in the source streams.

Lastly, the transcoding point provides an opportunity to standardize the carriage of these output streams such that service wrappers can subscribe and ingest streams and data in a common manner. A multicast distribution using an MPEG-2 transport layer is a suitable common interface for service wrappers.

Service Wrappers

Service wrappers adapt the elementary stream to the intended service and can be a point to add DRM/conditional access or XML messaging specific to that service. More than one service wrapper can subscribe to an elementary stream or set of elementary streams.

The advantages are each service does not singularly bear the costs for transcoding the stream. The devices designed for each service have lower costs per a stream because they are mostly wrapping the elementary stream (with some optional light transcoding functions) and more likely to be designed as a dense stream product. The reuse of the ES streams to multiple services is possible by focusing end-devices to support standardized video codecs like H.264/AVC. With the ability to output multiple aligned output transcoded streams, a small subset of streams can source multiple services.

Some examples of service wrapper functionalities are:

1. Groom multiple SPTS to set of MPTS streams
2. Convert a VBR stream to a CBR stream
3. Convert an SCTE 35 trigger to a suitable service equivalent

4. Match specific audio/data streams(s) to the service ES stream
5. Fragment the incoming conditioned ES stream
6. Apply DRM to the stream
7. Add service specific XML data

Additionally, the service wrapper developers do not have to be experts in video encoding, nor will companies that have video expertise have to learn a new set of domain skills for the new service.

IMPLEMENTATION

This architecture solution standardizes the content acquisition and transcoding processes. The output of the system is a set of multicast streams to which service wrappers can subscribe.

Transitioning to this concept is an expansion of existing infrastructure, operations and monitoring that already takes place at some MSOs for operationalizing video systems. Content sources are already ingested for today's broadcast services. Expanding this does not mean increasing the content selection, it means standardizing mezzanine contribution specifications to support the quality and data needs for multiple services. It also means developing transformaters to go from MPEG-2 to AVC/H.264. This will allow content providers to transition over instead of immediately cutting over. It will also allow for the transcoders to convert to and from AVC/H.264 and MPEG-2. Operations and monitoring are already taking place in a centralized system but need to be expanded to create a demarcation point between content source issues, transcoding issues and service issues.

Operation and monitoring should be viewed from three perspectives on both a content and service level:

1) Alerting to loss of content/service or content/service degradation;
2) Troubleshooting loss of content/service or content/service degradation; and,
3) Data and metrics gathering related to quality of content/service.

These three perspectives share common system touch points that provide existing content/service channel condition data. With the separation of content and services, the service operations and monitoring simplifies to handle the wrapping and adaptation modifications to fit the ES stream to the service. Some operations and monitoring features covered are:

1. Alerting to loss or degradation
- System will monitor content loss and degradation for content to all services with each service only needing to monitor modifications
- Existing MPEG tools can be reutilized to monitor loss and degradation in the system for content streams
2. Troubleshooting loss or degradation
- Devices within the system will be capable of at least rudimentary measurements of media service conditions.
3. Metrics Gathering of Media Service Quality
- Devices are able to calculate a Level of Service (LOS) that can be sent to collection points
4. Data Collection Points
- All data (a normal state, an off-normal state, degradation, failure and quality level) will be gathered at a central point for processing and archiving
- The Data Collection Point (DCP) is a system that will provide a user interface to system Operation Engineers (OE)

The benefits of common implementation is the costs of operations and monitoring will be centralized across services with each service only responsible for operations and monitoring for modifications completed at the service wrapper point and beyond.

TRANSITIONAL STRATEGIES

The number of output resolutions and rates supported by this video model should support playable video on devices intended for service. The initial set of streams should support devices for broadcast, VOD, SVC, adaptive streaming, mobile and personal devices for managed and unmanaged networks.

Future services need to justify additional stream formats based on ROI, or need to readapt the existing streams. Adding an additional stream would imply adding far more than the targeted service, which should available to all services and across all content selections.

In regard to existing streams, lower bit rates and resolutions will be more fluid over time since it is believed the device lifetime in these areas is much shorter than hardware desktop devices. Changes in the lower bit rate streams may lean toward increasing frame rates to make the set of output streams more homogenous and available across the platform. Changes in the number of streams may increase to accommodate new service needs or consolidate due to end of legacy services or incorporating new scalable video technologies.

SUMMARY

With new delivery and platform structures, creating a common set of video feeds for content that can span across multiple services has many benefits including a reduction in cost to develop and monitor new services. Without this type of architecture it will be harder to bring up new types of video services with a common set of service features because of the high ROI cost to support\common customer experience across services. With a new set of video streams for content, it will be easier to design and initiate new services that can already subscribe to one or more subsets of these streams