BITMOVIN

BITMOVIN WHITE PAPER

Per-Title Encoding and Total Cost of Ownership

How Per-Title Encoding Keeps the Lid on CDN and Storage Costs

Table of Contents

PROFITABILITY AND COST CONTROL

Adaptive bitrate (ABR) encoding is the foundation of all modern online streaming architectures. By encoding a video into multiple bitrates (these renditions are collectively referred to as the "bitrate ladder"), the adaptive streaming video player can select a stream that is appropriate to the device and bandwidth conditions of each user. HLS and MPEG-DASH are the popular modern streaming protocols using adaptive bitrate technology.

The industry has been deploying online streaming technologies as trends towards direct to consumer content delivery and "TV Everywhere" implementation by Pay TV service providers continue unabated.

As consumer adoption and competition heats up, the imperative for profitable growth has set in. Profitable growth, in turn, is driven by monetization and revenue maximization through subscription models, as well as ad-supported models with online video ad technologies. Equally important is containing the underlying costs of processing and delivering content using online streaming.

This paper addresses the cost containment aspect of a profitable online streaming deployment, specifically addressing VOD, either subscription- or ad-based models.

VOD ENCODING TECHNOLOGY REVIEW

Currently most content providers use a fixed bitrate ladder to encode their content. There is a variety of approaches but options generally revolve around the various screen resolutions that are most popular with a particular audience and devices and anticipate a variety of possible bandwidth conditions.

Example of a fixed bitrate ladder

RESOLUTION	TRACK TYPE	BITRATE (BPS)
1280x720	HD	4,763,000
1280x720	HD	3,254,000
960x540	SD	2,584,000
704x396	SD	2,000,000
512x288	SD	1,433,000
384x216	SD	958,000
340x192	SD	626,000
340x192	SD	411,000
	AUDIO	256,000

The example above is a typical fixed bitrate ladder that could be applied to an entire video on demand library or live signal and achieves an entry level adaptive streaming performance. Although this "blanket" approach has the advantage of simplicity in encoding operations, it leaves either video quality or cost or both on the table.

The next level of optimization is to create per genre bitrate ladders. This is an approach that many content providers use today. It involves breaking your library up into different complexity categories. For example, Action Films, Documentaries, Sports, News, and Cartoons. The assumption is that within each of these categories you will find similar levels of video complexity, consisting of detail, motion, and production styles.

A separate bitrate ladder is developed for each category. This is a step in the right direction, but it relies on manually configuring, testing and improving your bitrate ladders and in the end it is very similar to the "blanket" approach.

And the assumption of equivalent content within a category fails when you think of a newscast turning to sports highlights and all of a sudden, the "talking head" of a news anchor cuts to a high complexity video scene exhibiting high motion, and high detail such as the audience in a stadium or the grass on the playing field.

Per-Title Encoding (PTE) is an encoding technology that analyses each and every title (asset) in your library. It then automatically builds a bitrate ladder that is specifically optimized for the content of each and every title - determining not only the individual encoded rendition (or encoded profile), but also the ladder itself (number of renditions, spacing between renditions, etc). This optimization revolves around assigning the minimum amount of bitrate before you hit diminishing returns where additional bits applied to the content would not be perceptually visible to the human eye.

The idea of creating a bitrate ladder to best suit the contents of the video file you are encoding is not a new one. Early research was being done as early as 2012. But implementation was challenging, mostly because it requires each title to be analysed before the encoding can be performed.

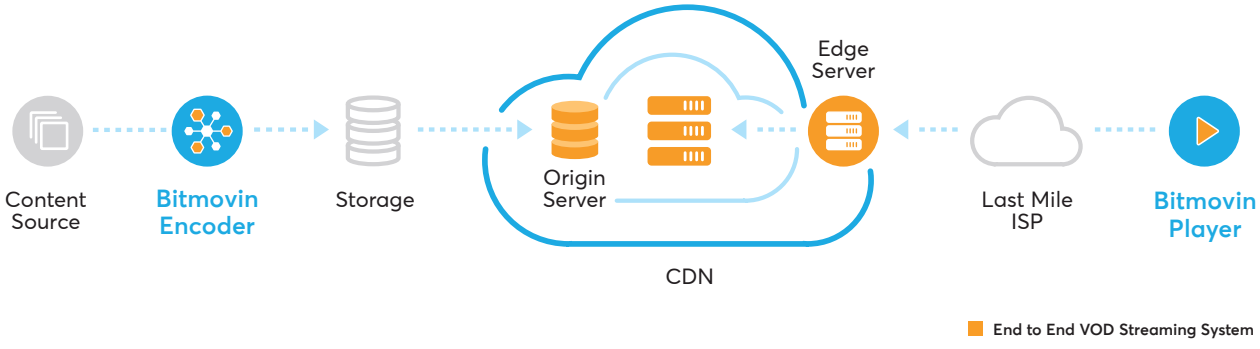
Netflix was the first to overcome this problem and launched the first commercial PTE workflow in 2015. Bitmovin's PTE solution was up and running shortly after that and offers ready access to this type of technology through on-prem or cloud-based encoding.

PTE IN AN END TO END SYSTEM

A typical video on demand end to end system using Per Title Encoding (PTE) consists of content files arriving in a mezzanine format, such as ProRes or mp4. This format typically has a single, high quality encoded version (or "track"), and in addition to the video, contains audio, captions / subtitling, thumbnail images, and other metadata content. The file transcoder, located on-premise, in private cloud, or on the public cloud (or multiple public clouds), then catches this mezzanine file and transcodes it into an adaptive bitrate format like HLS or MPEG-DASH, including encrypting the content through the interfacing of a digital rights management (DRM) system. The resulting file outputs are then stored in an origin server, and ready for distribution through the CDN upon request from multi-device players.

The main ongoing costs in such a VOD online streaming architecture are:

A. Encoding, typically charged by encoded total rendition minutes (duration of content), and varies by output resolution, and value-added features.
B. CDN delivery, charged by total data delivered typically in GB, incorporating number of viewers, length of viewing session, rendition delivered (which in turn is driven by factors like receiving device and last mile and in-home network).
C. Storage associated with origin serving. This includes the carrying costs of the VOD library back catalog.



PTE CASE STUDY

As already discussed, per-title encoding (PTE) is content- ("asset" or "title") dependent, and to get to the cost profile, we need to run a case study with specific content assets using a fixed ladder and PTE-generated ladders. The VOD library will then be modeled as a combination of assets of these different types of content. Three different types of content are examined with the resulting bitrate ladders from per-title encoding shown below.

Cartoon content sample

Standard fixed ladder definition

RESOLUTION	TRACK TYPE	BITRATE (BPS)
1280x720	HD	4,763,000
1280x720	HD	3,254,000
960x540	SD	2,584,000
704x396	SD	2,000,000
512x288	SD	1,433,000
384x216	SD	958,000
340x192	SD	626,000
340x192	SD	411,000
	AUDIO	256,000

Cartoon Content Per Title Encode Ladder

Cartoon (simple) content is modeled here with a PTE operation run and the resulting per-title encode bitrate ladder show below:

RESOLUTION	TRACK TYPE	BITRATE (BPS)
640x360	SD	244,000
1024x576	SD	443,000
1280x720	HD	847,000
1920x1080	HD	1,200,000
1920x1080	HD	2,300,000
N/A	AUDIO	256,000

Concert Content Per Title Encode Ladder

Talking heads (medium complexity) content encode results are shown below:

RESOLUTION	TRACK TYPE	BITRATE (BPS)
680x360	SD	240,000
768x432	SD	380,000
1024x576	SD	720,000
1280x720	HD	1,100,000
1920x1080	HD	2,100,000
1920x1080	HD	2,450,000
N/A	AUDIO	256,000

Concern content sample

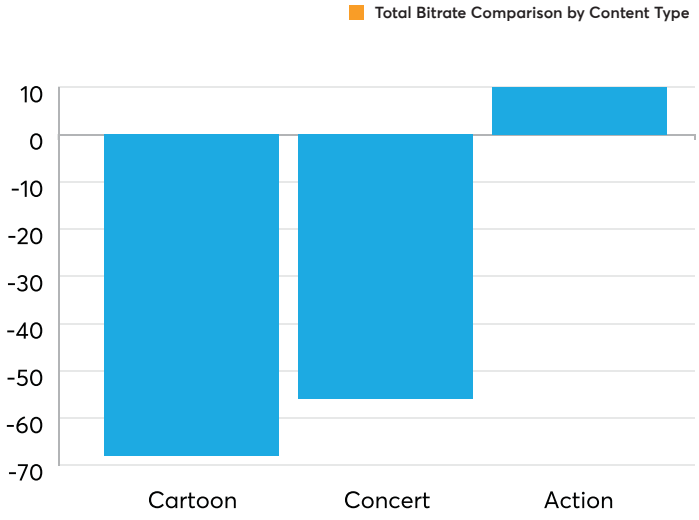
Action Content Per Title Encode Ladder

And finally, the action content (high complexity) and the resulting per title bitrate ladder is shown below:

RESOLUTION	TRACK TYPE	BITRATE (BPS)
512x288	SD	244,000
640x360	SD	381,000
1024x576	SD	726,000
1280x720	HD	1,380,000
1280x720	HD	2,600,000
1920x1080	HD	4,940,000
1920x1080	HD	7,400,000
N/A	AUDIO	256,000

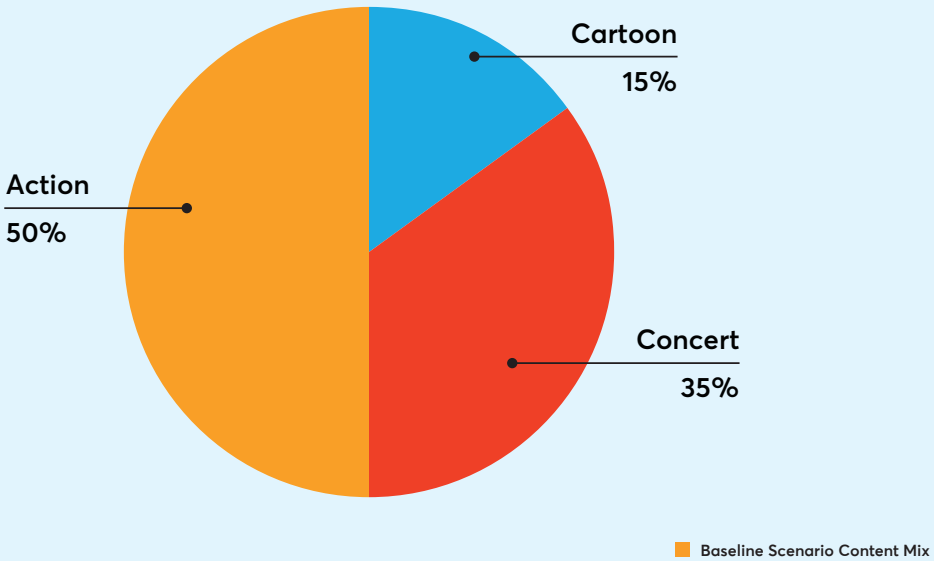
Action content sample

This comparison chart shows the aggregate bitrate difference between the standard ladder and each of three sample assets. As shown, the bitrate savings for the low complexity content is significant - over 50% - whereas the standard bitrate ladder sacrifices too much video quality in the case of the Action content (most complex), and correspondingly, the PTE operation allocated a higher aggregate bitrate to bring the video quality level up.



These three asset types are then grouped into a complete VOD library to act as the basis of the cost comparison model. The model takes into account deployment experience from Bitmovin's solutions architecture team working with various customers and partners.

Model assumptions:
Content mix in baseline scenario, based on some recent project experience - 15% simple (cartoon style) content, 35% talking heads content, and 50% action content:



The baseline scenario also consists of the following assumptions, all of which represent variables in the model that can be modified to represent different deployments:

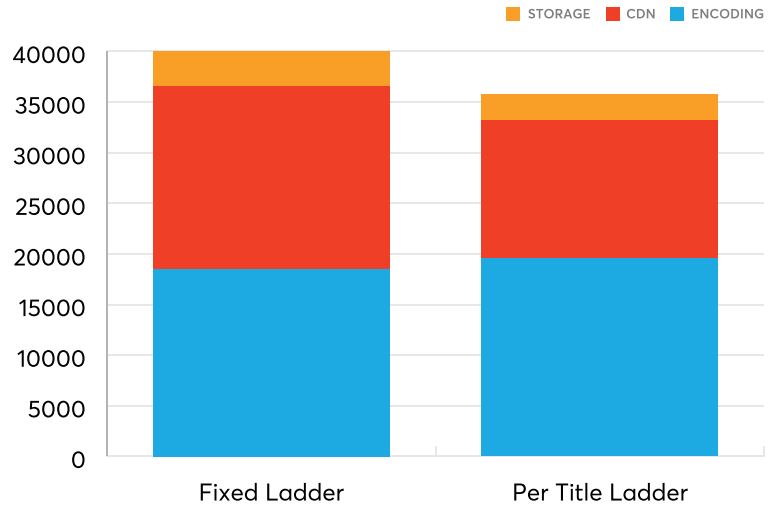
CONTENT ASSUMPTIONS	
Catalog size	10,000 assets
Average asset duration	60 min
Catalog refresh rate	25%
SUBSCRIBER ASSUMPTIONS	
Average hours viewed	2,400,000 - Based on average monthly subscriber base of 200,000 globally, viewing at an average of 3 hours per week and 4 weeks per month
COST ASSUMPTIONS	
CDN	1 cent / GB
Storage	2.3 cents / GB
NETWORK ASSUMPTIONS	
HD viewership (vs SD)	66% - Corresponds to the delivered renditions based on last mile network conditions and playback device type
Video codec	H.264 - Note that a more efficient codec like VP9, HEVC, or AV1 would yield different encoding and storage and streaming results. The model has the ability to model H.264, HEVC, VP9, or a combination
Packaging formats	HLS and MPEG-DASH

TOTAL COST OF OWNERSHIP RESULTS

The baseline scenario total monthly cost results are shown in the chart below, consisting of the main cost elements of CDN, Storage, and Encoding. These costs are opex-based, and represent the ongoing monthly costs.

Total Monthly Costs, Baseline Scenario

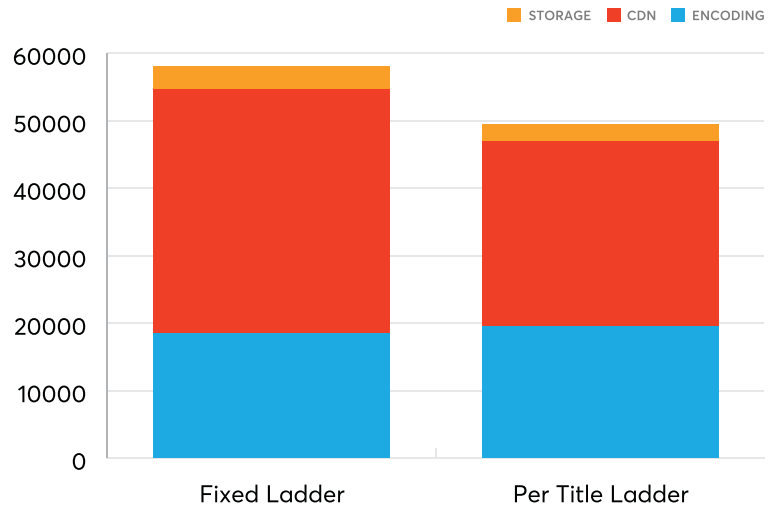
The PTE ladder option shows monthly ongoing costs, including storage carrying costs, of around $40,200, with the fixed ladder coming in at $36,000, a savings of over **11%** monthly. The encoding costs increase with a corresponding decrease in CDN and storage costs, as expected.



Total Monthly Costs, Doubled Streaming Hours

We then perform some sensitivity analysis with the model to see the varying impacts of inputs and assumptions made in the baseline model. If the monthly average streaming hours are doubled, corresponding to increased average hours per week or more subscribers accessing the content.

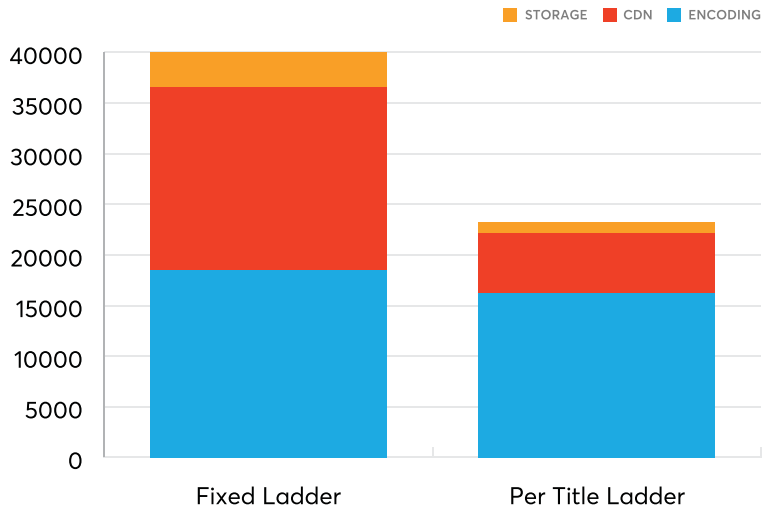
The savings here increases to over **15%** with CDN costs accounting for a larger percentage of the total cost.



Total Monthly Costs, All Simple Content

To see the impact of how content types impact the total cost of operation, we set the two extreme content types to be 100% of the library, which is unrealistic, but illustrates the corner cases. All simple (cartoon) content costs are shown on the right.

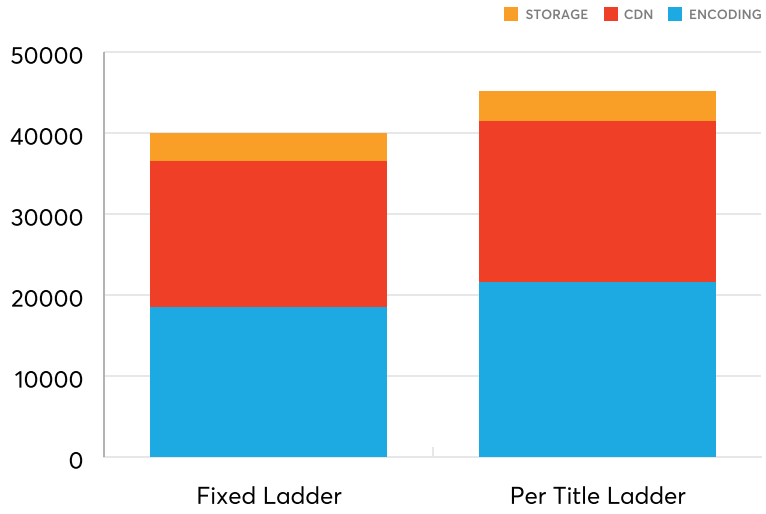
The monthly cost savings here are over **42%** per month, due to the drastically reduced CDN costs.



Total Monthly Costs, All Complex Content

With content set to 100% complex (action) content, the costs end up as shown on the right.

In this case, PTE ladder costs resulted in a **12%** increase in costs, showing that the fixed ladder configuration saved monthly costs, but at the expense of incurred video quality degradations. This would result in reduced user experience and possible churn risks.



PUTTING PER TITLE ENCODING INTO OPERATION

Per Title Encoding was shown to achieve the desired goal of video quality maximization while containing total costs for the specific video asset.

What are the other factors to consider when putting per title encoding into operation? This section looks at targeted ad insertion, cost predictability, and video quality measurement.

TARGETED AD INSERTION

Ad insertion, both server-side (SSAI) and client-side (CSAI) work by getting the encoded ad asset delivered for playback at the right time, either network-side or client-side. In the case of VOD, that mostly corresponding to pre-roll ads. Ideally, the ad file is encoded identically to the VOD asset, allowing the viewer to seamlessly cross from the ad to the program, or from program to the ad in the case of mid-roll or post-roll ads. Identical encoding between the VOD asset (e.g. movie or TV episode) and the ad asset means that the screen resolution and frame rate and other encoding parameters are identical during the transition, allowing for a broadcast TV-like experience that maximizes the effectiveness of the ad.

Identical encoding of the VOD asset to the ad asset, however, is operationally challenging with PTE since per-title encoding, by definition, changes based on the content itself. And the VOD asset and ad asset encoding operations are typically done at different times, and typically, by different service providers. Not to mention the fact that the ad asset is likely to be inserted around a variety of VOD assets, so perfect matching between one set of VOD and ad assets means no matching between that same ad and other VOD assets.

Two approaches to handle this issue are:

1. Over-encoding the ad asset. This results in many renditions, pre-encoded and packaged and ready to get inserted into a VOD asset with renditions most closely matching the ladder created by PTE of the VOD asset.
2. Just in time transcode (fast transcode) of the ad asset into the specific rendition or bitrate ladder dictated by the PTE of the VOD asset. This would ensure the ad asset is encoded identically, but ads some operational challenges around a just in time transcoder that would need to scale up with ad insertion needs.

PREDICTING COSTS / PROFITS

The fixed ladder bitrate approach, despite its inferiority in video quality and total cost of ownership results, has the advantage of predictable costs. Per title encoding by definition varies by content type, by asset, and as library content becomes more action-oriented, encoding costs (and corresponding video quality) would increase.

MEASURING VIDEO QUALITY

We discussed video quality in context of optimizing cost. What's the "source of truth" for video quality? There is no one industry standard on the topic, with methods today in use including Structural Similarity Index (SSIM), Peak-signal-to-noise-ratio (PSNR) , Video Multimethod Assessment Fusion (VMAF). Bitmovin uses the "Bitmovin Quality Index" as a tailored measurement system that is seen to correlate most closely to the human visual system, and draws upon elements of the methods in use today.

FIND OUT MORE

The total cost of operation tends to vary greatly with each deployment based on a number of key factors as we saw here. Equipped with modeling capabilities and experience not only at thee encoding level, but also the end to end system architecture and costs, Bitmovin's Solutions Architecture team is ready to help.

In the paper, we mentioned how video quality and an acceptable level resulted in the costs increasing in the case of complex video content. A tool developed by Bitmovin, the Per Title Ladder Benchmark Tool, was built to allow testing of your own content in a Per-Title encoding workflow and benchmarking it against a manually-defined fixed ladder.

Bitmovin, Inc.
301 Howard Street, Suite 1800
San Francisco, CA 94105, USA
+1-833-248-6686

Schleppe Platz 7
9020 Klagenfurt, Austria
+43-463-203-014

www.bitmovin.com

BITMOVIN

Software to Solve Complex Video Problems