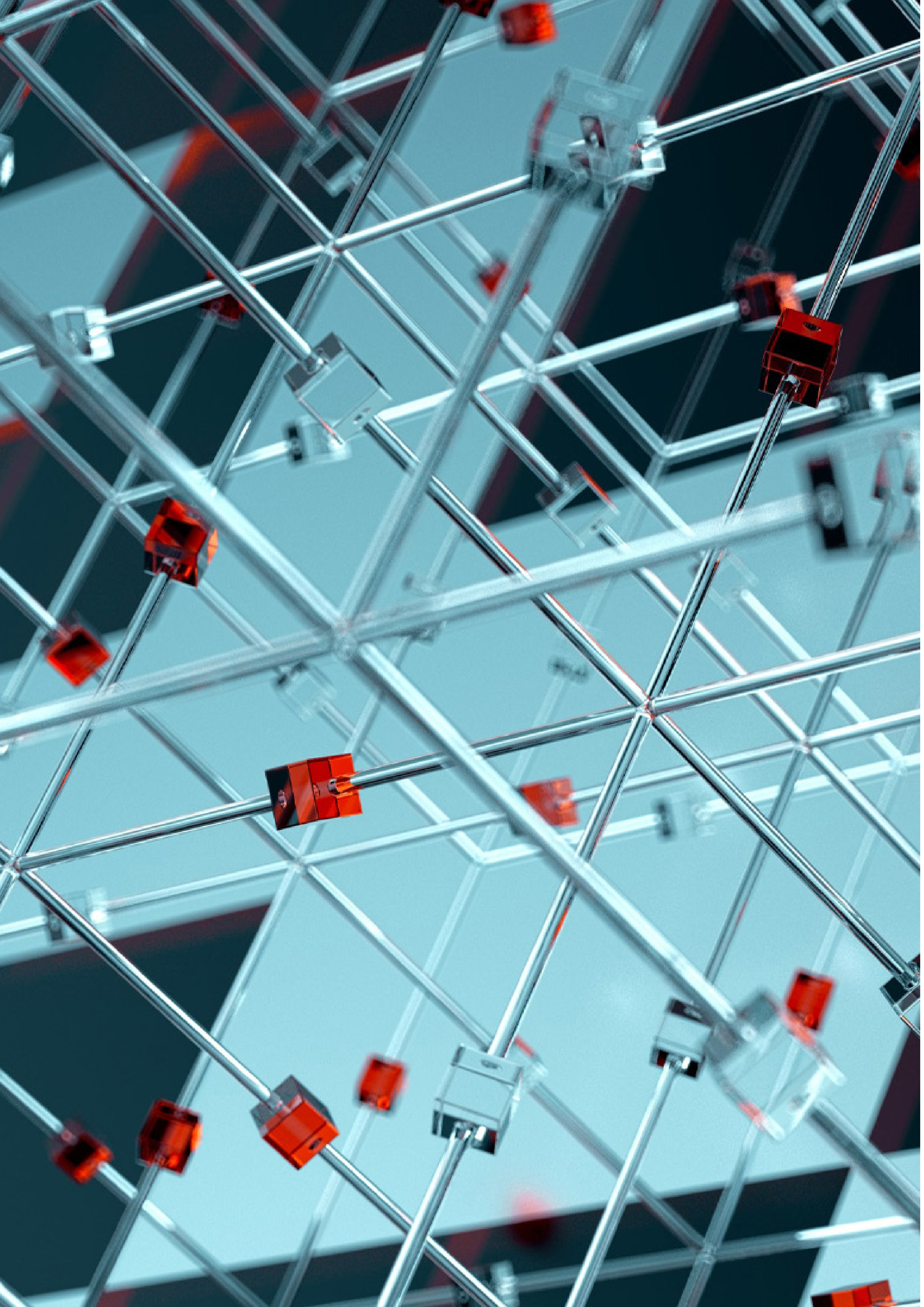


THE BROADCAST BRIDGE
Connecting IT to Broadcast

The Big Guide To OTT

Part 2 - Content Origination

A Themed Content Collection from The Broadcast Bridge

Themed Content Collection

Contents

Introduction

By Paul Martin. The Broadcast Bridge.

The Big Guide To OTT provides deep insights into the technology that is changing and enabling a new media industry.

As OTT delivery grows, driven by both consumer demand and content provider strategy, there are many adjustments to manage, including new production approaches, scaling content distribution, personalising, protecting, and monetising content, and assuring audience QoE.

Streamers are delivering a mix of live, linear, and on-demand content. Business models are blending - subscription with advertising and direct-to-consumer with service aggregation. The internet-enabled OTT delivery model is driving the media industry through a giant transformation.

There are many broadcast disciplines to leverage in OTT – the concept of "broadcast-grade streaming" means streaming should match broadcast's capacity for highly consistent, highly scalable delivery of high-resolution content at low latency. There are also new disciplines for content providers to embrace, like delivering highly personalized content and building new relationships with ISPs.

The OTT technology domain builds on core broadcast distribution disciplines and adapts them to internet-based delivery. New contribution methods, ultra-low latency encoding, and high speed broadband streaming, could mean that 'streaming-grade' will become a new gold standard for content delivery. But the fixed and mobile broadband networks we rely on, and the myriad of devices we use, mean that we need to work differently to manage content accessibility and quality. So while the content may be largely the same, there are significant differences between the worlds of OTT and OTA.

The Big Guide To OTT is a multi-part series. Each part tackles a different theme and there are three or more articles per part.

Available now:

Back To The Beginning

OTT Content Origination

New Parts coming in 2023:

Broadcast Grade Streaming

Customer Experience (CX)

The Business Impact Of QoE

Monetization & ROI

Managing Latency

Internet Infrastructure

Content Delivery Networks (CDN)

Assuring Viewer QoE

Streaming Sustainability

OTT Content Origination

By Paul Martin. The Broadcast Bridge.

Content Origination is in the midst of significant transformation, like all parts of the OTT video ecosystem. As OTT grows and new efficiencies are pursued, Origination must play its part as a fundamental element of the delivery chain. But Origination is not just about smooth and efficient content delivery. It's also about providing key features to the OTT service.

Content Origination is the point at which Live, Linear and VOD content are prepared for final delivery and streamed into the delivery networks. It is where the push systems of broadcast playout and VOD asset publication meet the pull system of streaming to multiple device types according to the required bit-rate and format.

Today, this push-pull line can be blurred. Linear OTT often still treats the Content Origination platform as part of the push system, passing through all formats and bit-rates to the CDNs regardless of whether they were requested or not. In many CDN environments this is required to "warm up the caches" so content is ready for when it is requested to achieve lowest possible latency with minimal requests back to the Origination platform.

For VOD, the push system ends as content is moved into Central Storage, ready to be streamed on-demand. That said, some CDN environments will replicate an entire VOD library in their own storage, pushing content deeper into the delivery network.

Content Origination Functions

Content Origination combines a set of functions that prepare content for OTT delivery. The diagram below shows the primary functions.

The first step is at encoding and transcoding, to ensure the live streams and VOD files are in the correct bit-rates for OTT delivery. Generally, an ABR group is prepared according to a set of pre-defined profiles, designed to handle the variations in network conditions and the range of devices that can request content. The choice of codec is an important consideration for the whole Origination Platform, given that newer codecs offer improved efficiency but at the cost of higher levels of processing (e.g., HEVC is more intensive than H.264). Recently, the concept of encoding on-demand has grown, as technology leaders have leveraged cloud compute to adjust encoding workloads according to consumer demand, resulting in more efficient use of resources.

Once content is encoded/transcoded there can be multiple workflows depending on whether the content is live or VOD, whether timeshifted viewing is available to the viewer, and whether or not low latency is required.

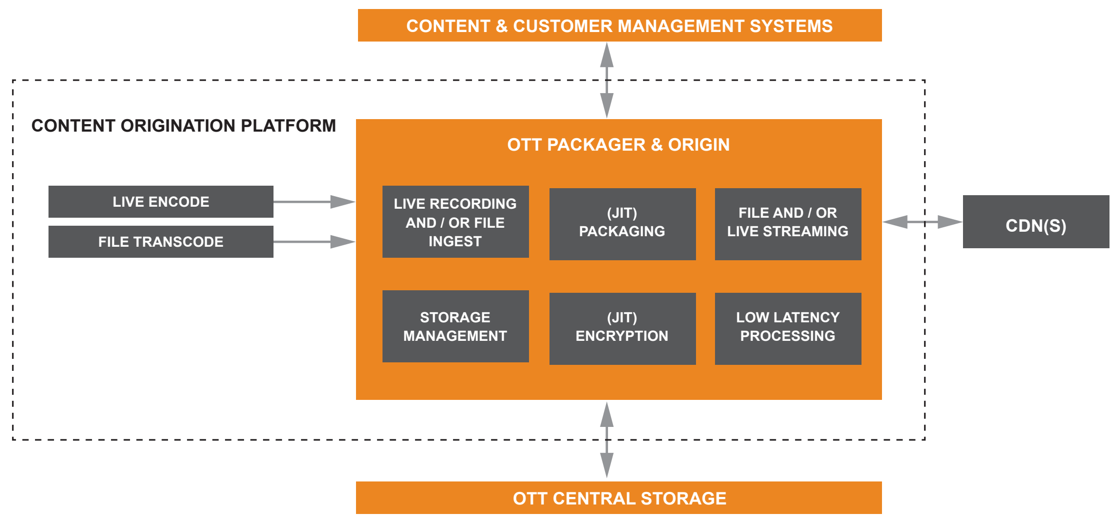


Figure 1 – Primary functions of OTT Content Origination.

The core functions are therefore used in different ways according to the workflow. In general:

- Live Recording and/or File Ingest – to provide Live TV catch-up services, now a basic feature of OTT services, the live stream is recorded to create a back-up for content held in the CDN. This ranges from short-term Live Pause to long-term CloudDVR. For VOD, files need to be ingested into storage. These functions integrate closely with the OTT Central Storage and synchronize with the Content Management Systems in order to confirm their availability for viewer consumption. The "and/or" distinction means that some platforms are unified for both Live and VOD services while some are separate, which is generally decided based on different operational requirements, which include scalability, for Live and VOD services.
- Storage Management – not only is content recorded and ingested into Storage, but the Storage must be managed. As metadata is updated, as content ages, and as content is deleted, something must manage where that content is and synchronize with the Content Management System. This content life-cycle management is often a software function in a module within the Content Origination platform.
- (JIT) Packaging – this function is required in order to deliver content to different device operating systems, like Android (DASH), Apple (HLS) and Microsoft (MSS). Some OTT systems package every piece of content regardless of demand – below this is referred to as Legacy Linear OTT. Just-in-Time (JIT) Packaging is best practice for large VOD libraries where

Figure 2 – 1st generation Broadcast OTT – Linear OTT with Catch-up VOD.

it is inefficient to package every piece of content before storing it.

- (JIT) Encryption – once packaged, content is encrypted for secure delivery over the internet to authorized viewers. Each package type has a respective encryption method (e.g. HLS uses Fairplay), although some package types use multiple encryption methods (e.g. DASH can use Widevine and Playready). Encryption follows packaging, so if content is packaged and then stored, it is generally also encrypted. JIT-Encryption accompanies JIT-Packaging for a more efficient storage model for large VOD libraries.

- Low Latency Processing – this can be isolated as a specific function of the Origination platform. The compute resources can be set up to act on specific streams or specific pieces of content to deliver in Low Latency formats which involves reducing GOP sizes and managing many more connection requests as smaller segments are delivered. The decision to do this is made further upstream in the Content Management Systems, but then executed at Origination.

- File and/or Live Streaming – once one or more of these content processing steps are complete, the content is streamed, pulled by requests from the CDN(s). Typically, the "origin server" has been an internet-facing web server specifically designed for delivering streams, which passes through streams from the Packager. Today these are software functions rather than discrete servers, although depending on workflow scalability and operational models it can make sense to deploy the software in dedicated hardware environments.

Figure 3 – 2nd generation Broadcast OTT – JIT Model for Live and large-scale VOD.

Evolution Of Content Origination

Today there are 2 primary models for Content Origination, with 2 new models on the horizon.

The 1st generation Broadcast OTT model (Figure 2) was introduced when Linear TV channels began to be streamed OTT. It is still widely in use today because it is a simple way to provide OTT content for linear broadcasters. This model takes an output from a broadcast channel which is then encoded, packaged and encrypted for OTT delivery. It is also recorded for catch-up TV. In this model, VOD utilizes a similar workflow to package file-based content before storage, to simplify operations and onward delivery to Streaming.

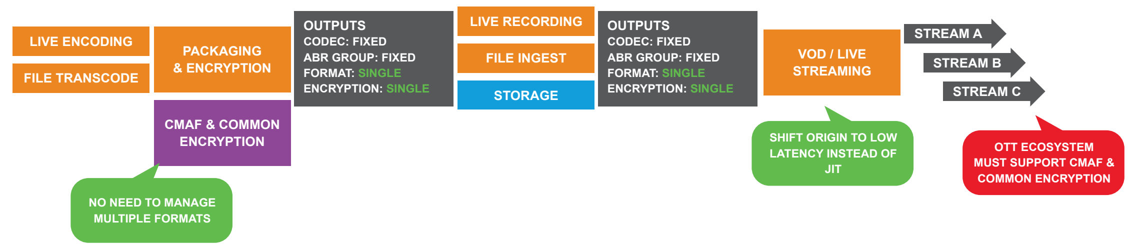


Figure 4 – 3rd generation Broadcast OTT – Common Format & Encryption.

As shown in the grey Outputs boxes, this model produces a set of outputs which are consistent across the delivery chain. There are two drawbacks of this model: 1) it requires an unnecessary amount of central storage (shown with a value of "X") and 2) it creates a pre-formatted VOD library that is inflexible to changes in format type that routinely occur as new consumer devices come to market.

The "JIT" model (Figure 3) is increasingly used today. Pioneered by large VOD businesses like Cable TV operators, JIT addresses the two main weaknesses of the 1st generation model. First, it provides flexibility to the fast-changing consumer device world by storing content libraries in a mezzanine format and then packaging and encrypting on-demand. This means that if a new format is required, the OTT operator does not need to transcode all VOD assets to the new format, potentially requiring weeks of processing. Secondly, it significantly reduces the size of the central storage. For example, storing in HLS, DASH and MSS increases the number of files stored by 3-times more than necessary when compared with a JIT model. As broadcaster OTT libraries grow into the multi-PB range this makes a big difference, not just for storage costs but also for streaming performance.

The 3rd generation model can be called "Common Format & Encryption" (Figure 4). This is enabled by the CMAF and CENC formats, which are the basis of today's DASH and HLS low latency formats, and are moving us towards a truly common format. From a storage and streaming efficiency perspective, this model no longer requires JIT packaging or encryption of the media segments. Instead, the single format and encryption type means that the stream or file is prepared, then stored and streamed. Not only does this result in a simplification of processing for packaging and encryption, but it also greatly improves cache efficiency through the use of common media segments, it retains the same storage efficiency as the JIT model, and it reduces the complexity of egress for the Streaming component of the platform which previously egressed multiple package types.

profile according to the specific piece of content and the type of device requesting it.

This model creates the most efficient Origination platform based on intelligence being applied before video is processed, and it enables even more efficient storage by leveraging the benefits of newer codecs where possible and creating more optimal ABR profiles. It is another

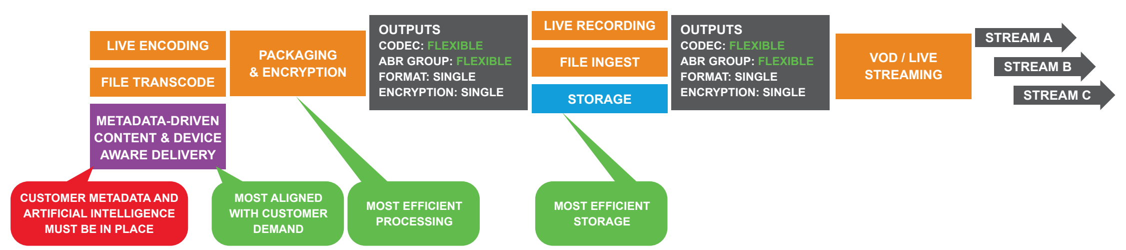


Figure 5 – 4th generation Broadcast OTT – Consumption Aware.

This model is deployable today but is likely to be limited to a low percentage of use cases based on end-to-end adoption of CMAF (including players, devices, etc.). Leaders in this space envisage the generation 3 model to become the standard over the next 2-3 years, but to work alongside models 1 and 2 for many years. In the meantime, benefits of VOD library flexibility and storage efficiency can still be achieved through the JIT model.

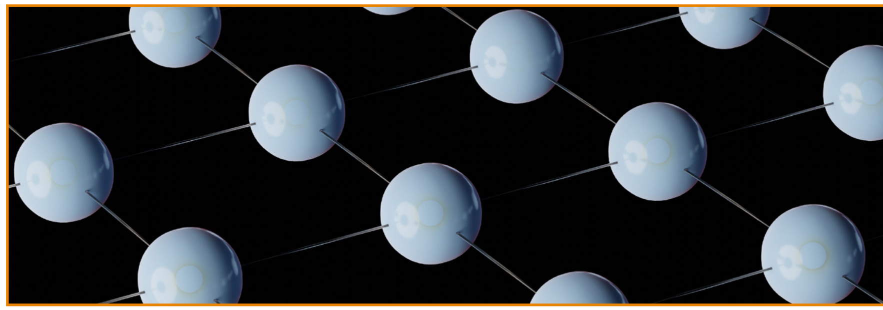
The 4th Generation model is emerging now, more in discussion than in deployment, aiming to leverage artificial intelligence, machine learning and the most modern codecs. We can call this the "Consumption-Aware" model (Figure 5). In this model a new level of consumer personalization is achieved as the system applies the video codec and the bit-rate

step towards the perfect pull-system that seeks to optimize content delivery at scale.

This model is not about individual consumer customization. Decisions must be made about groups of customers and groups of content. This model will use metadata about customers' devices and each piece of content, plus artificial intelligence to analyze consumer behavior and the level of demand for particular content.

An example of this in action on a live stream could be the decision to stream a graphically rich program in a higher-quality codec like HEVC or AV1. This would optimize the infrastructure being used, perhaps on a pay-per-use basis, and allow premium content to have premium viewing experiences. On a VOD library, AI-driven processing could be used to streamline or enrich codecs and bit-rates available in the library based on consumer demand. Codec licensing could become part of the variable factors used to optimize customer satisfaction, OTT throughput and total cost.

Content Origination is the execution point for delivering required content formats to consumers. While it is not a part of the OTT infrastructure that needs to dramatically expand in capacity as audiences grow (unlike storage and edge caching – see other articles in The World of OTT series), it is the first part of the OTT delivery system where the dual objectives of customer personalization and delivery efficiency are simultaneously addressed. Which is why the Content Origination platform directly enables OTT operators to achieve the vision for OTT – the delivery of highly personalized viewing experiences, at scale.



OTT's Unique Storage Requirements

By Paul Martin. *The Broadcast Bridge.*

Central storage systems for OTT are unique in the world of media storage. They combine a set of requirements which no other media use case must handle.

Most OTT services combine live and on-demand content, especially broadcaster OTT services that incorporate linear channels, live programming, and VOD libraries. We start by considering the different storage requirements of Live vs. VOD.

Live-Only

Storage for Live-only is all about small high-performance storage environments for both central storage and CDN storage. Storage in a live-only environment is largely the responsibility of the Edge Cache in the CDN. This is to enable low-latency time-shifted viewing (i.e., live-pause, rewind, restart and lookback). Most live-only OTT services have a small number of channels or streams and therefore a relatively small amount of content to cache. This results in a relatively small storage requirement.

To illustrate:

CDN Storage: If an OTT Operator gives its customers a 4-hour time-shifted viewing window on a single ABR stream (let's assume 7 bit-rate profiles, totalling 22Mbps), then it will require 40GB of storage (22Mbps / 8 bits in a byte * 60 seconds * 60 minutes * 4 hours) in a single Cache location. This grows to 120GB when considering unique files required for the 3 primary package types of HLS, DASH, and MSS. That is not a lot of storage for the important time-shifted user experience. If we extend the lookback window to a typical 7 days, then this grows to 1.7TB. Still small, given a single HDD is now up to 16TB. Even a month of lookback is only 7TB of storage. The CDN needs to distribute this content across all Edge Cache locations. So, if you have 20 Edge Caches which would make you a large OTT Operator your business model should justify the 140TB of distributed storage. And for every additional channel at the same bit-rates and time-shifted offering you would add the same amount of Edge Cache storage.

Central Storage: The central storage records each live stream so it can deliver any video segment to any CDN Edge Cache when required. Today, many OTT streams are encoded and packaged before being recorded and stored. This means that content is multiplied by the number of package types. Some solutions offer the ability to encode, then record and store, before packaging and streaming. This can significantly reduce the amount of central storage required depending on the number of package types offered. In the 22Mbps per channel scenario, the central storage would either be 7TB (storing packaged content) or 2.33 TB (storing un-packaged content) per channel. This choice makes a difference for larger multi-channel / multi-stream operators.

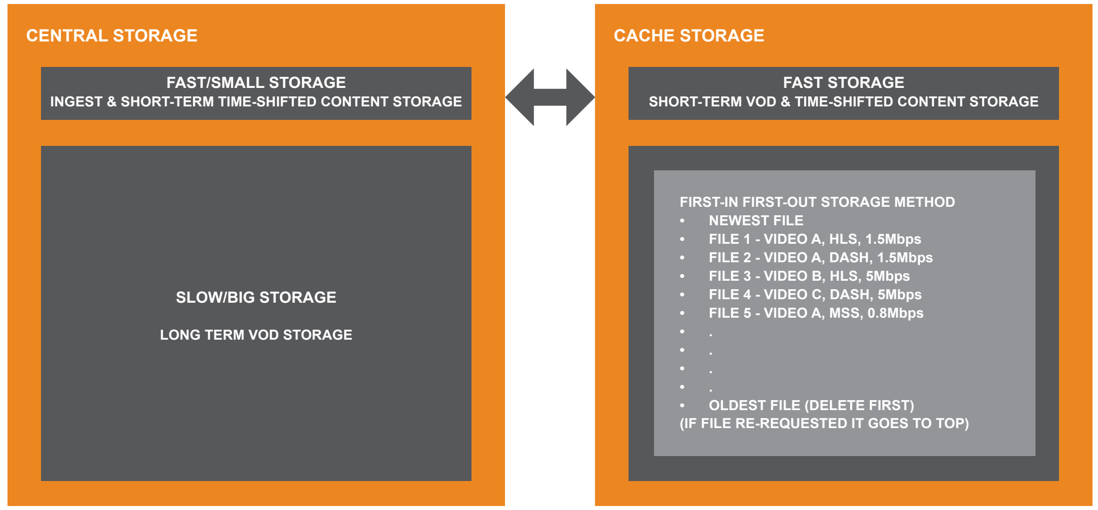


Figure 1 – Central Storage & Cache Storage work together in OTT service delivery.

VOD-Only

Storage for VOD-only is all about large, scalable, secure, and cost-effective central storage plus intelligent caching of the most popular content within the CDN. The central storage stores a large content library. Leading OTT Operators store multiple petabytes depending on resolutions, formats, and the storage of packaged or un-packaged files. Storing un-packaged files is a big benefit for larger libraries. The CDN then stores the content as it is streamed on-demand to the consumer, so the next consumer request for exactly the same content can be delivered from the CDN, instead of calling back to the central storage. In VOD use cases the storage must ingest a large amount of transcoded files (JIT-Transcoding is not recommended due to the excessively high cost of processing power to meet customer latency expectations vs. the relatively lower cost of storage), store them securely for the long-term, and stream hundreds or thousands of unique files simultaneously to the audience. Central Storage ingress and egress is therefore subjected to a very different workload compared to the Live use case.

To illustrate:

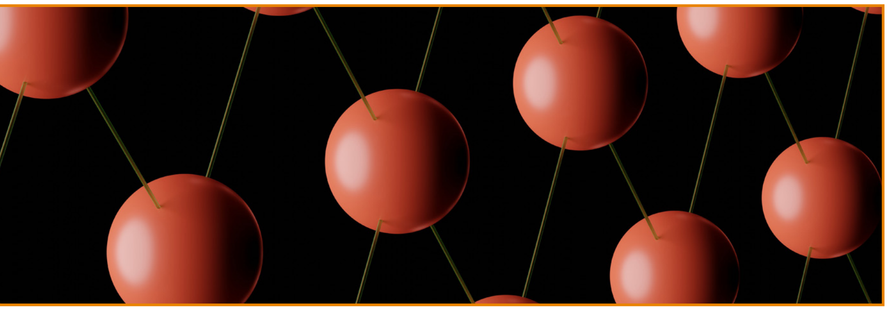
Small VOD library: If a VOD library has 5,000 titles with an average length of 60 minutes, then the ABR-transcoded content (let's assume 22Mbps again) will total 50TB. If packaged into 3 formats before storing, this becomes 150TB. If there are 20 Edge Caches and 10% of the titles are cached for subsequent low latency delivery, then each Edge Cache needs 15TB, which means 300TB (15TB * 20 Caches) of total cache storage.

Larger VOD libraries can reach many times greater than 5,000 titles. A library with 100,000 titles (i.e., more typical of broadcaster VOD) and an average asset length of 60 minutes would require 3PB. Across 20 Edge Caches, using the 10% figure, this would mean 6PBs of cache storage (300TB * 20 Caches). To manage costs, a leaner method is typically used which either requires extensive intermediate caching or more utilization of bandwidth between the central storage and edge caches. But the magnitude of the investment is clear. It is worth highlighting that these figures do not take into account storing multiple copies for redundancy or the typical 20% capacity overhead required by the storage software.

Storage For The OTT Use Case

Most OTT services with combined Live and VOD content observe about 10-20% of their consumption on live content and 80-90% on VOD content. As live streaming grows, some OTT services are reporting 30% Live and 70% VOD. These types of OTT services therefore focus on a strong VOD library (which generally means a large library to offer compelling consumer choice) and a strong live content delivery capability (which generally means sufficient Cache storage to provide customer-satisfying time-shifted viewing options).

The combination of fast file ingest, live recording, time-shifted viewing features, simultaneous streaming of live streams and potentially thousands of unique VOD streams, plus long-term archiving is a unique workload for OTT storage to manage. Traditional editing, playout and archive operations do not combine these requirements. In essence, an OTT central storage system must be very large, very fast, and cost-effective. It's a tough ask.



Building Object Storage For OTT

In recent years, there has been a visible move by media businesses towards using disk-based software-defined object storage for video use cases. Object storage is built for unstructured data, and particularly for larger datasets. At its core, object storage provides a highly scalable, highly resilient platform for long-term storage. In the Media industry it is becoming more and more common to see new Archive storage environments selecting object storage for its strong data-protection capabilities, scale-out and cloud-integration abilities – AWS S3, Ceph and proprietary object storage technologies are expanding their presence. But object storage is not typically associated with high-performance storage.

However, there is a part of the Media industry that has dealt with this "large + fast" challenge for years already. MVPDs (Multichannel Video Programming Distributors) have been streaming thousands of VOD files and hundreds of live channels to millions of consumers for almost 2 decades since the introduction of VOD in the cable TV and IPTV industry. MVPDs built their own private CDNs and deployed managed set-top boxes with on-board storage. During the first half of the 2010s, MVPD requirements for cloudDVR or networkPVR started to grow. Their aim was to remove expensive hard-drives from set-top-boxes to achieve a significant cost-saving and concurrently remove a leading cause of customer dissatisfaction (as anyone who has ever lost or replaced a set-top box with tens of hours of recorded content can attest to). Instead, they wanted to leverage their network connectivity and content caching to centralize content storage and streamline their customer premise equipment infrastructure.

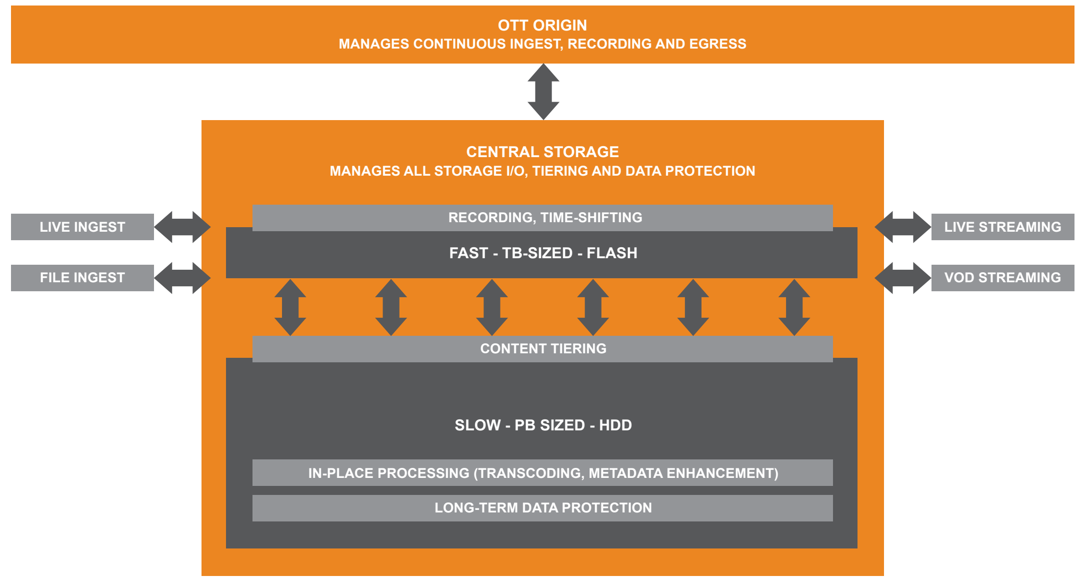


Figure 2 – OTT Central Storage performs many storage tasks simultaneously and at scale.

MVPDs asked for something new: multi-petabyte archive-capable storage, to include cloudDVR capacity, that could record every live stream, ingest thousands of new files, and deliver time-shifted TV and thousands of VOD files simultaneously to millions of individual consumers. In some countries, like the USA and Germany, the MVPDs needed to have enough cloudDVR capacity to store unique copies of every consumer recording to comply with copyright law. While this law requires the addition of many petabytes of storage capacity to a central storage system, the cost saving versus having SSDs in millions of set-top boxes still makes this the right investment.

Leading storage businesses responded with software-defined hybrid object storage. It was the only storage technology that could have a chance of meeting the requirements for size and cost. The biggest challenge was to make it fast, but huge steps have been made with video-centric protocols and algorithms, plus tight integration with Origin platforms for managing ingesting, recording and streaming. The leading solutions have focused on automatic tiering between high-performance flash storage and HDD storage so that a cost-effective multi-petabyte system can stream any file on-demand. That tiering is becoming more finely grained as storage hardware technology evolves and the latest innovations involve usage-based data placement and geographically dispersed data tiers. The leading vendors remain focused on the storage software that manages the data across the drive types, storing it safely and enabling low-latency on-demand streaming.

OTT Operators have very similar requirements. While most are not replacing SSDs inside set-top boxes with a cloudDVR offering, nor streaming hundreds of live channels, nor delivering VOD assets from almost every single content provider in a geographical market, they are placing the same pressures on the central storage. They are ingesting, recording, streaming and protecting their growing content libraries. And as Broadcasters deploy and expand their OTT services, the same mentality they have thrived on for decades persists – as household-name rightsholders delivering live content, they cannot fail. Black to air or rebuffering isn't an option.

Central storage is a key contributor to achieve this goal. It is the protector of the content library and of the network-connected Live and VOD viewing experience. If content is not in the central storage and a CDN needs it, then the customer won't see it. At that point the OTT pull system has failed. As this is not an option for OTT Operators, the central storage needs to handle the pressure and do it cost-effectively. It's what hybrid object storage is built for.

The Importance Of CDN Selection To Broadcast-Grade Streaming

By Paul Martin. The Broadcast Bridge.

Multi-CDN is a standard model for today's OTT Operators, which automatically requires a CDN selection solution. As streaming aims to be broadcast-grade and cost-effective, how are CDN Selection solutions evolving to support these objectives?

The Rise Of Multi-CDN & CDN Selection. For a long time, CDN selection and multi-CDN were in a chicken-and-egg situation – how could one emerge and grow without the other? To many OTT Operators, multi-CDN was not a priority because price and quality were good enough from their single supplier. Plus, a multi-CDN set-up brought new complexities such as managing multiple suppliers, inserting an extra layer of technology into the chain, and supporting multiple delivery points from the Origin.

But things changed as content delivery volumes increased. The CDN market began to see significant price competition. CDNs started to have failures and "good-enough performance" was no longer good enough for OTT Operators seeking strong business results from their growing investments.

Today, the standards expected from a streaming service are high, both from consumers and from media executives. Quality must be as close to broadcast-grade as possible, to deliver an excellent Quality of Experience (QoE) which leads to better viewer engagement and lower customer churn. Delivery cost must enable a streaming service to economically scale as the volume of content delivered grows exponentially and peak audiences expand towards the size of traditional prime-time linear TV audiences.

With these market-level changes, multi-CDN has taken off and solidified the demand for CDN selection solutions.

Initially, CDN selection was simple. The primary function was to allocate enough traffic to each CDN to ensure that committed traffic levels and expected price per unit of delivery were achieved. But as multi-CDN has matured, and the business importance of streaming has grown, CDN selection has become much more sophisticated.

Now, OTT Operators require CDN Selection solutions to deliver multiple business-critical benefits such as avoiding service outages and reducing churn, eliminating capacity bottlenecks, enabling a global streaming product strategy with consistent distribution platform implementation, and tailoring distribution methods to meet various budgets and performance needs.

CDN TOPIC	CONSIDERATIONS
Performance Variability	How does the CDN work under different types of demand (e.g., Live vs. VOD) and at different times of the day (e.g., evening vs. morning)?
Scalability	How does the CDN perform when audiences scale up quickly, for example due to a large live event or a must-see VOD release?
Location	Where is the CDN capacity physically deployed and how is it connected to the ISP infrastructure (e.g., transit only, direct interconnection, multiple ISP locations)?
Type	Is the CDN a private CDN (i.e., only for use by one D2C Streamer) or is it a public CDN (i.e., multi-tenanted)?
Capacity	Is the capacity required to reach the audience confirmed as available in the right locations at the right times? Is the required capacity guaranteed to be available to my business, or only provided on a best efforts basis? Is the capacity of the CDN aligned with the downstream capacity of the ISPs or other connectivity providers?
Variable Costs	How do costs adjust with changes in the volume of content delivered (i.e., Gigabytes/GB) and the capacity consumed (i.e., Gigabits per second/Gbps)? Is the CDN pricing model capacity-based or consumption-based? How much cost will be variable vs. fixed? How can maximum ROI be achieved considering CDN contracts?
SLAs	Do the SLAs allow for missing a minimum commitment if the CDN performance required a redirection of traffic to other CDNs? Can overages on one CDN be offset by under-usage on another?

Figure 1 – Considerations involved in multi-CDN set-up and CDN selection.

Today's Best Practice

From a technology perspective, there are three areas of focus to assure a high quality of streaming content delivery. First, the actual delivery platform itself – the encoder, packager, origin and CDN must all be set up for successful delivery. Second, the player – ensuring it works well with the wide range of devices. Third, the choice of profiles – encoding profiles must be handled well by the whole delivery chain from encoder to player.

CDN is the primary item in these focus areas where second-by-second dynamism exists. It is important to consider subjects like network performance variability, network scalability, network location, network type, network capacity, and potentially highly variable costs (figure 1). In other words, the way the whole network infrastructure performs after the Origin and before the Player can be impacted by many factors that are outside the control of the technical decisions made by the OTT Operator.

Therefore, today's CDN Selection best practice is based on sophisticated and real-time decision-making logic to adapt quickly to changes in the network environment that affect stream performance.

Because different CDNs will have different answers to the questions in Figure 1, OTT Operators need to closely manage their relationships with each CDN provider. The focus on multi-CDN has led to

OTT Operators working to obtain cost parity between CDNs, so that the CDN selection process can predominantly focus on performance, which is where the impactful day-to-day variability is observed.

Within the focus on performance, the subjects of CDN capacity location, capacity transparency, and capacity guarantees have become the latest hot topics. Simply “delivering 50PB in a month” is not enough. The delivery of those 50PB must meet exacting performance standards. It is like asking a courier to not just drop the package at the right address, but also ensure the customer receiving it is impressed with the condition of the package and the level of customer care and attention provided.

Leading OTT Operators today engage in dynamic CDN measurement and selection. This enables real-time decision-making at every moment of the day. CDN capacity utilisation (i.e., consumer demand) can vary easily depending on consumer behaviour and the content publication activities driven by all the businesses simultaneously utilising the CDN capacity. Capacity supply can also vary due to outages and planned maintenance, not only at the CDN level but also at the ISP level which can force large-scale stream re-routing. Regardless, the streaming show must go on and the streams must somehow be delivered well.

Dynamic CDN Measurement & Selection

Leading CDN Selection solutions enable OTT Operators to operate dynamically. They offer powerful configuration abilities that can apply, in real-time, all defined business rules at a very granular level and optimise stream performance across a range of KPIs that can be defined and prioritised by the OTT Operator.

The theoretical design of these solutions is relatively straightforward. The acid test is how they practically perform under real-life conditions, handling all required real-time computational processing. Although it may be possible in the CDN Selection system to drill down to device type and operating system type by country, ISP and ASN, and make delivery choices depending on live vs. VOD and the type of bitrate profile for different content, the reality is that making large-scale, real-time CDN selection choices at this level might be impractical. But as the world of streaming continues to evolve, these granular capabilities will become critical to success.

As shown in Figure 2, CDN selection is an initial task for each stream request, which becomes a continuous task as the stream is delivered to the consumer over time. Some CDN Selection solutions start with the premise that multi-CDN selection will first be driven by commercial rules that are fixed, such as minimum delivery commitments in a contract. This will set an initial business rule to deliver a certain quantity to one CDN which will primarily be managed by allocating a percentage of traffic to that CDN. This percentage-based solution, which depends on accurately forecasting total content delivery to allocate delivery correctly between CDNs, can lead to under- or over-utilisation of a particular CDN. Therefore some solutions have moved away from having percentage-allocation as a primary rule and instead they allow CDN Selection decisions to initiate with a wide range of factors, like capacity availability (e.g., I have 1 Tbps of capacity on my private CDN that must always be fully utilised), region (e.g., I deliver all my content in Germany, and I should deliver

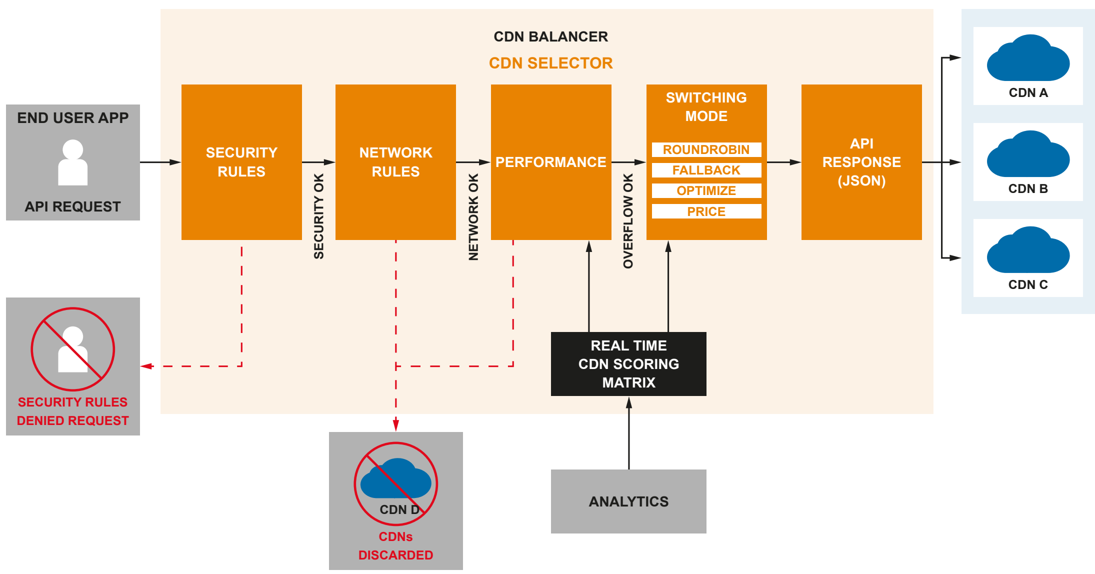


Figure 2 – High-level CDN selection process.

SWITCHING MODE	DESCRIPTION
Round Robin	A method based on percentage allocations to each CDN. The goal is to reach the end of each billing cycle with the best financial result, with optimized performance levels. In some systems, rules can be set up to over-allocate traffic when performance exceeds expectations in order to under-allocate when performance dips, yet still arrive at the end of the period with best financial performance.
Fallback	A method based on static preference-based allocation of traffic to a single CDN, as long as it meets performance and commercial requirements. If it does not, then there is a fallback to a second, third, or further CDN in a prescribed order. This model can be combined with commercial rules that can allow a CDN to be capped, and no more traffic allocated, once a monthly commit level has been reached or if a CDN has a bandwidth limit in place.
Optimize	A method from leading CDN selection providers using proprietary algorithms that select the best scored CDN using multiple KPIs like errors, buffer ratios, bitrate, measured end-user QoE, and viewing time. Every stream is individually allocated based on this real-time measurement. This approach heavily focuses on high-performance streaming, where QoE is top priority and cost is secondary.
Price	A method based on static price-based allocation of traffic to the single cheapest CDN available if it meets performance and commercial requirements. Like the Fallback method, it can be combined with commercial information like contract commitments. Price must be entered into the model, but for confidentiality reasons a relative price percentage to rank CDNs can be used if required.

Figure 3 – Descriptions of Switching Modes in leading CDN Selection solutions.

100% of my content in the Frankfurt region to ISP "1" using CDN "A"), and content type (e.g., deliver 4K live content on CDN A, deliver all other content on any CDN).

The granularity of these business rules depends on how the content, network, and consumer are understood by the system. If the understanding is high-level - "the consumer is on ISP 1 in Country 1" (i.e., a simple DNS approach) - then CDN Selection simply passes responsibility for good delivery to the CDN. But if it is possible to understand the consumer's location more precisely (e.g., through eDNS0 implementations) and profile (i.e., with client-side data), and it is possible to understand how the CDN is configured in terms of server locations, capacity availability, connectivity types with ISPs, etc., then CDN selection can be highly sophisticated, and the OTT Operator can have more control of the delivery process.

Once a CDN has been initially selected by these business rules, then the next step is to complete the selection based on performance. Using performance checks is a win-win for the OTT Operator and the CDN, because content will only be allocated for delivery on a CDN if it is performing well enough at that time.

Performance decisions need to be based on QoE and QoS, as measured by the client-side tools and the server-side tools. QoE takes priority in most solutions, which includes measures like rebuffering ratio, start-up time, start-up errors, and sustained bitrate. The CDN Selection tools must have granular, real-time data for all these metrics to support good decisions.

"In the NPAW Suite there are over 70 different quality KPIs and more than 90 different metadata elements associated with each playback," according to NPAW's CDN Balancer Business Unit Head, Luis Lopez Chousa. "Analysing this multi-dimensional data matrix in real-time can involve millions of calculations, which means OTT Operators can make pinpoint automated decisions about how to deliver their content given the environmental conditions at the time."

Jose Jesus, Director of Product for Experience Insights Suite at Conviva, states "OTT Operators normally set up business policies to meet their CDN commitments while at the same time using each CDN when and where it performs best. It is important that algorithms optimise performance within business policy constraints, but if a CDN's performance deviates too far from the baseline then it needs to be temporarily avoided. Typical metrics OTT Operators want to improve are connection induced rebuffering ratio, bitrate, technical video start-up failures, and technical video playback failures".

Once a CDN has been selected Switching Modes ensure that the usage of each CDN is maximised for ongoing cost and performance. There are multiple switching modes as described in Figure 3.

Performance and cost results both depend on how the CDN Selector performs. Leading Technology providers are working on optimisations required by the leading OTT Operators, to push the performance of CDN selection to the next level.

Managing Public And Private CDNs

OTT Operators are increasingly investing in their own private CDN platforms, dedicated to their own content delivery. CDN Selection solutions may measure each platform in the same way but they should be able to allocate content to each CDN environment using different methods. For example, a private CDN should bring cost and performance benefits that need to be maximised. So, it should have the majority of traffic allocated to it unless there are problems. CDN Selection solutions that initially select using the percentage-based business rule are not ready to support this requirement.

Per Segment Delivery

A recent development in leading CDN Selection solutions is to enable CDN selection on a per video segment basis. A plug-in component at the player determines which CDN to retrieve each segment from based on a range of real-time measurements provided by the client-side analytics. Policies can be set that limit the ability to retrieve from a CDN that is not desirable for the OTT Operator, if there are performance or business constraints. This low level of granularity will provide OTT Operators even more methods in the future for optimising their content delivery.

NPAW's Luis Lopez Chousa states "The per segment approach takes CDN selection capabilities to a new level that could transform streaming performance for the largest audiences. OTT Operators will be able to know that they are doing everything possible to extract the best performance from their delivery platforms, which is particularly important during the biggest events when the pressure to perform is at its highest."

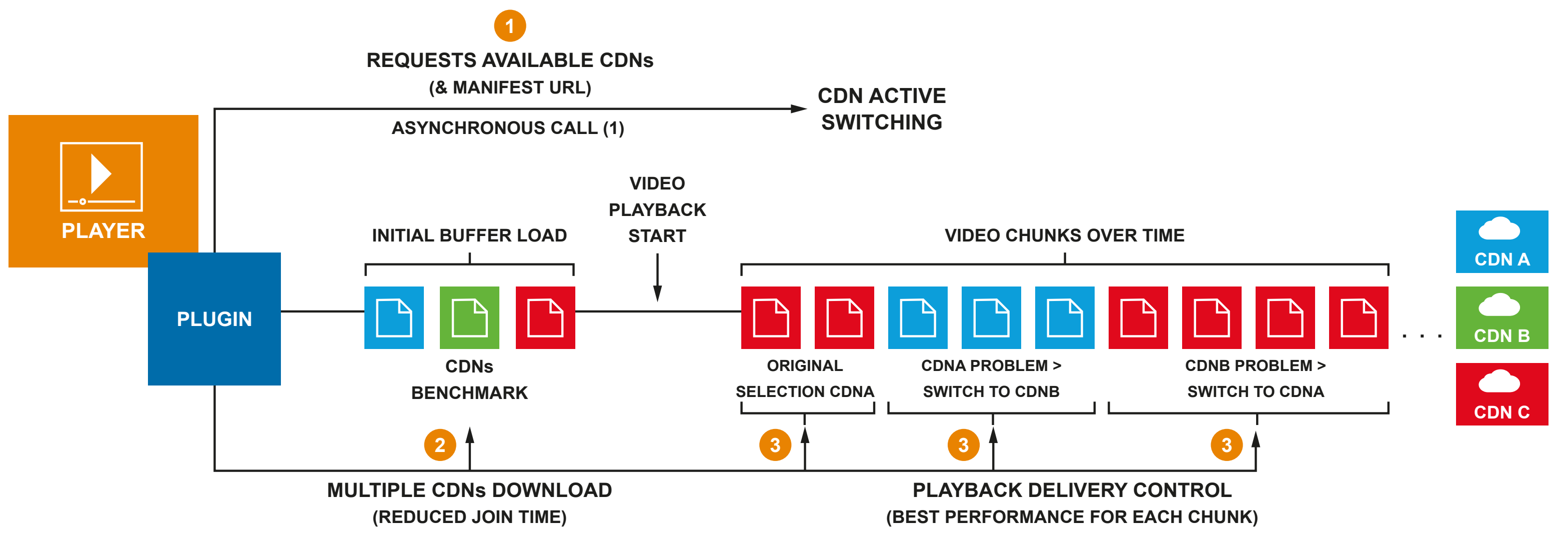


Figure 4 – Active Switching using Per Segment CDN Selection.

Reporting

Completing the CDN selection process is the reporting on decisions made and their results. Leading solutions allow for deep drill-down into session-level data which enables analysts to dissect performance and establish appropriate improvement plans. Real-time alerts of under-performing CDNs or content types can be provided and are connected back to the analytics reports. This feedback loop ensures continuous improvement is embedded in day-to-day CDN selection decisions.

Results

The OTT Operators taking this subject most seriously are seeing important performance improvements. Conviva's Jose Jesus states "Our published reports over the last 7 or 8 years show that the industry has moved from looking at average rebuffering ratio, or the percentage of sessions with less than 2% rebuffering ratio, to measure how many sessions have zero rebuffering, as this represents a proper TV experience. Benchmark rebuffering ratio in Europe and North America, for example, has dropped from 1% to less than 0.1% over the last 7 years because best-in-class OTT Operators are continuously measuring performance and making improvements."

Technology Developments

The general technology drive in CDN Selection is for more data, more real-time analytics, and more detail in decision-making. Additionally, some very specific areas are being researched and developed to support OTT Operators' future needs.

- Internet control planes – to prevent risking an overload of a CDN, leading to performance degradation, customer dissatisfaction, and CDN switching, the CDN Selector must know when a CDN is full. This requires some form of control plane for CDN capacity decision-making. The following Streamer Requirements section explains this further.

- Deep Edge deployments – improving performance most often involves tuning how the players manage the video and working with CDNs to fine-tune capacity availability and locations. The placement of Edge Caching capacity nearer to consumers is a major trend that will ultimately help all OTT Operators to scale their businesses.

- Multiple delivery sources – the Per Segment Delivery approach previously mentioned can be extended to incorporate other delivery sources, like WebRTC or Peer to Peer. This could mean that specific segments could be downloaded from a non-CDN source, like a peer, potentially reducing costs and latency for the OTT Operator.

Streamer Requirements

Advanced OTT Operators, pushing the boundaries and requiring the very best solutions, describe three specific areas of requirement for their future CDN selection solutions.

First, business rule configuration is a must-have. Solutions that focus on technical selection criteria and do not include the ability to use finely tuned business rules are too simplistic. Fine tuning involves, for example, delivering lower value content on a lower cost / lower performance CDN, while higher value content should use the highest performing CDN. Given that CDNs can differ in each country in terms of features, performance, and price, it is important to use their services intelligently to drive best cost.

Second, the CDN Selector is the initial point of the stream delivery before streams spread out inside the CDNs. The CDN Selector is aware of exactly what type of device/consumer a stream is to be delivered to – for example, mobile vs. SmartTV, or a standard vs. premium customer. Therefore, it should be possible to manipulate the manifest at this point. With this functionality, the right number of users can be sent to specific CDNs with capabilities that best match the delivery requirement.

Third, data integration from 3rd party systems is critical. A prime example is the private CDN platform that OTT Operators can have very deep insight into, such as capacity availability by server, and the ability to control the amount of available capacity at each PoP. The requirement is to bring that data into the CDN Selection tools. Today, a lot of CDN Selection functionality requires installing an App plug-in which requires opening the App and inserting the data, and time and cost to integrate. It would be more efficient to bring data from the CDN into the CDN Selector that can be used to drive more granular capacity allocation. The basic rule of streaming performance at a CDN level is that the best performance is achieved when the right capacity is available. Pinpointing the right capacity in real-time would be a major benefit to the OTT Operators.

As Vincenzo Roggio, Head of Distribution at DAZN, states "In our large-scale live event streaming delivery, traffic is spiky. It starts, reaches a high peak, and then stops, many times per day. We need to know exactly how our CDN capacity is performing, and we need to know that during a 2-hour game we are getting optimal decision-making from our tools that help our CDNs to perform their best. We know this leads to the best QoE for our customers, which is business-critical for us. We have a multi-CDN environment and look for every way to continuously improve the performance and efficiency of our content delivery operations."

Conclusion

CDN Selection has a big role to play in delivering broadcast-grade experiences for streaming viewers. It supports OTT Operators to optimise their CDN performance and cost, both of which are critical to their success as streaming grows.

The leading solutions are laser-focused on improving streaming performance through excellent use of data that delivers clear insights, actionable information, and a continuous improvement capability.

We can see that the streaming world is improving in every respect. Devices are better, encoding standards are sharper, networks are better. But, there is always the chance of encountering network congestion. To protect against this impacting stream performance or cost, analytics must be granular, and CDN selection must be dynamic.

In the end, whether a viewer receives their broadcast-grade streaming experience or not relies heavily on which CDN is selected.