Guaranteeing Video Quality in IP Delivery Systems

By Göran Appelquist, Ph.D., Chief Technology Officer, Edgeware AB

This article explores some of the challenges and solutions for operators to guarantee video quality as they struggle to keep pace with the massive growth of IP video traffic.

Göran Appelquist, Ph.D.
Chief Technology Officer,
Edgeware AB

Göran Applequist currently serves as Chief Technology Officer at Edgeware AB. He has more than 15 years of experience from the digital media industry in different management and executive positions. Göran has a Ph.D in Physics from Stockholm University.

The first IP video services emerged over a decade ago, using managed operator networks to reach set-top boxes. Since then, there has been a rapid transition to a much more complex world. Adaptive streaming is used to reach a multitude of devices over different networks, while operators struggle to keep pace with the massive growth of IP video traffic.

This complexity, combined with viewers' expectation for more advanced services at higher quality, can be impossible to manage without the proper solutions in place. This article explores some of the challenges and solutions for operators to guarantee video quality under these new conditions.

What is IP video delivery?

Delivering premium Pay TV services over IP has become increasingly important for network operators competing with cable and satellite broadcast services and for Over-The-Top (OTT) service

> **The need to support a wide range of device types and networks is a strong driver for consolidation.**

providers. IP networks are also used for hybrid delivery, such as IP backbones feeding cable head-ends or satellite/terrestrial services delivering on-demand functionality over IP. A recent trend is to use the IP capabilities of cable networks (DOCSIS) to complement or replace legacy QAM delivery, while wireless networks are widely used to reach mobile devices.

The first IPTV services were launched by telcos in the late 1990s, mainly using UDP technology, limiting them to the operators' managed DSL networks. In 2010, Move Networks introduced the first HTTP-based streaming technology, enabling premium video delivery over existing Content Delivery Networks (CDN). The use of HTTP also made it possible to traverse firewalls and gateways without configuration. New services were launched by content aggregators without network infrastructure, such as Netflix and Hulu. Since then, multiple competing HTTP-based formats have emerged including Microsoft Smooth, Apple HLS and Adobe HDS. The MPEG-DASH standard was recently developed with the hope of eventually consolidating and replacing all of these.

These new formats also proved efficient for reaching smartphones, tablets and connected TVs. Today, the need to support a wide range of device types and networks is a strong driver for consolidation and HTTP-based adaptive streaming is increasingly being used even for the primary video service delivered to the living room.

Quality issues

From an end-user perspective, most Quality of Experience (QoE) issues associated with IP delivery fall within one of four categories. Listed in order of severity, they are:

- Paused playback or buffering.
- Poor interactivity (slow responsiveness) and delays.
- Low quality (low resolution and/or frame rate).
- Glitches in playback (visible or audible).

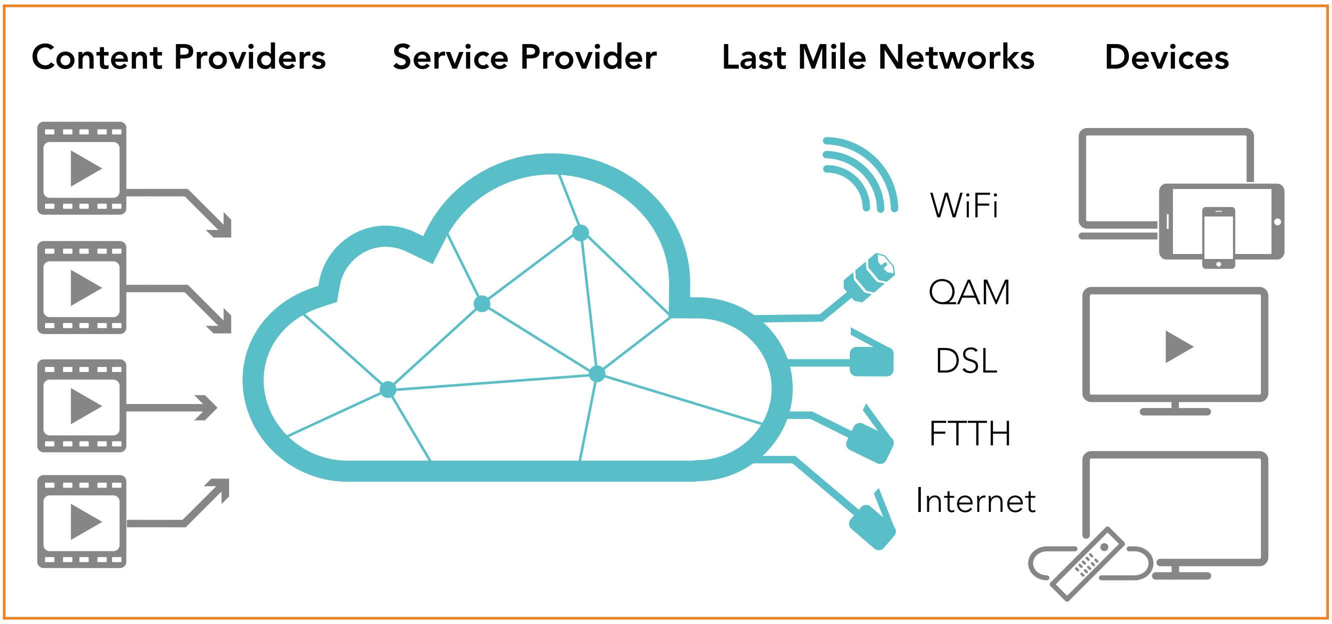
Mapping these issues into the network domain results in the following primary root causes:

- Lack of bandwidth.
- Packet losses.
- Latency.

The root causes and fixes need to be addressed separately, considering both a traditional IPTV and an HTTP-based streaming model.

Traditional IPTV

First, we'll look at how these network issues affect traditional UDP-based IPTV systems. UDP is a unidirectional protocol without client feedback. Video is delivered as a continuous stream at constant bitrate, which is fully controlled by the streaming



Today's premium video services need to address many types of networks and devices

> **HTTP-based adaptive bitrate streaming introduces two new technologies: TCP and multiple quality profiles.**

servers. Live TV streams use multicast to effectively reach many simultaneous users, while on-demand streams use unicast with RTSP for session control. In a traditional IPTV system:

- 'Lack of bandwidth' results in no service, as a new stream cannot be delivered. The system must keep track of all used resources, or overprovision the network to sustain any anticipated traffic.
- 'Packet loss' can cause a video glitch, or an audio pop. Since UDP is a fire-and-forget protocol, lost packages are not recovered. Low levels of packet loss are usually acceptable, or a side channel can be added for re-transmission of packets.
- 'Latency' is typically not an issue, since there is no "handshaking" involved. Latency could have some impact on user-initiated control signalling, but the human responsiveness is normally much slower than any realistic network latency.

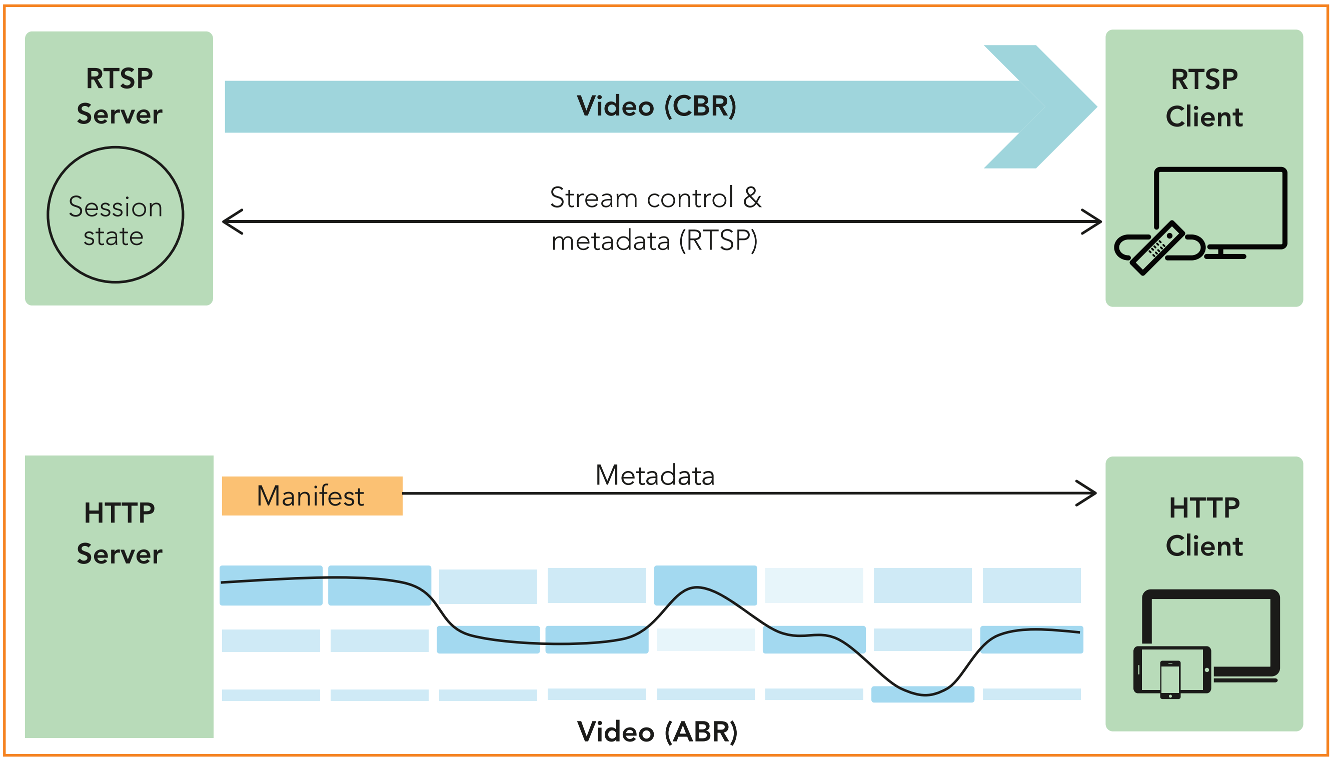
HTTP-based adaptive streaming

HTTP-based adaptive bitrate streaming introduces two new technologies: TCP and multiple quality profiles. TCP is a bi-directional unicast protocol with built-in congestion control and re-transmission. Combined with multiple quality profiles, it allows clients to adapt to changing network conditions by continuously requesting the optimal bitrate.

This enables high quality video delivery over a broader range of IP networks, including wireless and unmanaged Internet. In contrast to traditional IPTV, where a single Constant Bit Rate (CBR) signal is delivered, adaptive streaming video is chopped into segments (typically 2-10 seconds long allowing a smooth change of quality levels). The following illustrates the key differences:

With HTTP-based adaptive streaming:

- 'Lack of bandwidth' is dealt with by delivering content at lower quality, requiring less bandwidth. Even though low quality is an issue, it is better to receive a continuous service at low quality than no service at all. However, bandwidth that is too low or too variable stalls the delivery, due to empty buffers in the client.
- 'Packet loss' usually does not cause glitches, because lost packets are automatically recovered. However, the TCP



Traditional IPTV vs. HTTP adaptive bitrate delivery

It is the operator's responsibility to provide the best network conditions and the adequate use of available resource.

congestion algorithms cause packet loss to severely impact the available bandwidth. To guarantee a certain bandwidth that accommodates packet loss typically requires twice the normal bandwidth, resulting in the need for the network to be over-provisioned.

- Network 'latency' also lowers the available bandwidth, because TCP relies on low-level "handshaking" of packets. Doubling the latency will halve the maximum bandwidth available which may prevent delivery of high-bitrate HD/UHD streams over long-haul networks. End-to-end latency is crucial for live streams, especially sports. When delivering through a CDN, it may add up to one minute or more.

Ensuring quality

In an HTTP delivery system, to ensure optimal video delivery operators need to consider the following techniques to address these challenges.

Resource management

Adaptive streaming theoretically allows 'any' number of clients, as they adapt to the available capacity. However, this would seriously impact the overall video quality

In an HTTP-based adaptive streaming, a viewing session consists of hundreds or thousands of fragments instead of a single stream. This creates a challenge to adequately track and allocate bandwidth to deliver high or continuous quality without over-provisioning the network.

A solution is to use "virtual sessions", created when receiving a client request using URL-naming conventions or cookies. Virtual sessions are used to allocate resources, and to help monitor video delivery. Relevant session-based data is easier to interpret and more valuable to the operator than tracking.

Because the bandwidth varies, it is not evident which bitrate to allocate. Using the highest quality profile is optimal for the end-user, but not for the network. Using an intermediate bandwidth is usually a good compromise. More advanced algorithms for quality selection can include device type and access network.

Deploying an agile software-defined network management solution makes it possible to allocate resources, provide load balancing and monitor functions that are scaled on demand.

Getting close to the users

With both packet loss and latency reducing the available bandwidth between a server and a client, it is important to originate the streams as close to the users as possible to maximize video quality. This can be achieved by caching. A tiered architecture is ideal, where the most popular content is cached closest to end-users. By reducing upstream bandwidth requirements, caching can also significantly save costs.

Optimising use of available resources

Deciding that the client should have the final quality decision is not completely accurate. It is the operator's responsibility to provide the best network conditions and the adequate use of available resource.

First, video should be prioritized over non-critical heavy traffic, such as file transfers. This can be achieved using standard network tools, such as the DiffServ features in the IP protocol, traffic classification in mobile networks or DOCSIS service flows in cable networks. However, these methods are usually limited to single network domains or managed networks. End-to-end priority cannot be guaranteed over multiple domains like the open Internet.

The second option is for operators to limit the options available to clients. These options are presented through a separate metadata file, called a manifest or a playlist, which is retrieved when starting playback. Based upon available information, such as type of device, network and current network status, the manifest can be customised to limit the options. For example, it is unnecessary to offer UHD to a smartphone or even to a connected 4K TV if the network is congested. Although this technique is limited to start of playback and cannot adapt to all real-time changes, it can significantly optimise resources and end-user quality.

The main perceived advantage of leaving the decisions to the clients is that the delivery infrastructure can be agnostic to video. The clients can simply request arting playback. Based upon available information, such as type of device, network and assumption, is over-simplified. Premium video delivery requires careful design of the entire infrastructure, in addition to implementing video awareness in several locations.

Despite the growing popularity of on-demand services, live TV is still a major traffic driver.

Using multicast for HTTP live streams

Despite the growing popularity of on-demand services, live TV is still a major traffic driver. This is especially true for live sports streaming, which can cause huge peaks in traffic. Traditional IPTV systems use multicast to efficiently distribute live streams to large audiences. HTTP streams are unicast, a much less efficient method. One solution is to use distributed caches to "fan-out" this stream closer to the end- users. However, in large networks with multiple caches, this can still be inefficient due to the inter-cache traffic.

It is also possible to leverage multicast for live HTTP streams, at least to a nearby cache. The cache could be in the home (residential gateway) or access/aggregation network. Either the actual file fragments are delivered over multicast, using techniques such as PGM, or the video is delivered as synchronized MPEG-2 transport streams, one per bitrate, which are segmented, encrypted and re-packaged in the cache. The first option requires fewer intelligent caches but needs more bandwidth for multiple end-user formats. The latter option enables only one format to be delivered through the network, but requires intelligent caches to process content 'on the fly'.

Pushing live streams to edge caches via multicast reduces the impact of network latency on available bandwidth, facilitating delivery of high-bitrate streams. End-to-end latency is also typically reduced. However, it should be noted that just-in-time processing of the transport stream can add significant latency if not implemented correctly.

Measure and analyse

Since quality decisions are made by the clients, the best way to verify the quality experienced by end users and to evaluate the impact of any improvements is to measure and analyse the traffic that was actually delivered. This can be challenging if the delivery of every little chunk needs to be monitored. A video-aware analytics tool that aggregates and abstracts this data at a meaningful level is therefore strongly recommended.

Broadcasters and OTT content providers who use third-party CDNs usually hand over content long before it reaches the viewer. There is often no visibility of quality at the point of delivery into the CDN or into the broadband access networks.

To take control of quality, and impose SLAs on delivery partners, a much greater level of visibility is needed. This is possible through the establishment of a local origin at the broadcaster or content provider. Local origins perform functions such as time-shifting live to on-demand programmes and delivery into CDNs including all necessary measurement and analysis of performance.

Summary

Ensuring video quality while transitioning from a purpose-built, managed network to multiple IP-based networks has a number of challenges. Different formats exist to alleviate these challenges such as HTTP-based adaptive streaming. However, issues with packet loss, latency, resource management and measurement must still be addressed via careful technology and vendor selection.

Edgeware powers some of the largest IP video services in the world and has been helping broadcasters and operators to successfully deliver high quality services over the last decade.