sandvine ®

Intelligent Broadband Networks

Contents

Video Quality of Experience: Requirements and Considerations for Meaningful Insight

An Industry Whitepaper

Executive Summary

Video streaming, whether over-the-top (OTT) applications or on-deck services, accounts for the majority of Internet bandwidth.

There are a range of business motivations for communications service providers (CSPs) to monitor subscriber quality of experience (QoE) for streaming video services. Whether for network engineering purposes, business intelligence insight, customer experience management initiatives, or some other reason, the ability to determine the QoE from the subscriber's perspective, for each video stream, is of paramount importance.

In fact, with the right solution in place, every single video playback is a chance to measure and monitor the subscriber experience on the network; in effect, such a solution turns the entire subscriber base into a dynamic collection of real-time probes.

With the wrong solution—one that's ignorant of the end user experience or that relies on sampling approaches—every video playback is a missed opportunity.

By understanding the complexities associated with truly measuring video QoE from the end user perspective—and the requirements that these complexities impose—CSPs will be in a position to make an informed decision and choose the most suitable solution.

This paper explains why, to correctly determine the user's true QoE for video streaming, a solution must understand how user-initiated actions manifest against a backdrop of observed and expected video streaming behavior, unique to the video application being used—in an environment in which a single video might be split across multiple flows, a single household might be watching multiple videos at once, and the video traffic is encrypted.

WHITEPAPER

Introduction to Streaming Video

Video is the dominant driver of Internet traffic; depending on the region, video is either already the majority of bandwidth on communications service provider (CSP) networks or it's destined to become so soon.[1]

While some of this video traffic is "on-deck" (i.e., part of a CSP's brand or company), the majority is characterized as being "over-the-top" (OTT); OTT services exist in two forms:

- Services provided by a third-party: examples include Netflix, HBO, Hulu, Amazon, Vimeo, Vevo, Periscope, Twitch, Qello, BBC (iPlayer), Sky TV and other television networks, etc.

- Services provided by the CSP: examples include Bell's Fibe TV App[2] and Comcast's Xfinity TV App[3]

The popularity of video in general, and OTT video in particular, is responsible for two fundamental shifts in consumer behavior:

1. **Higher peak bandwidth levels, concentrated in a shorter period of time:** Since OTT video is an "on demand" application, it drives traffic when it is viewed; previously, video content was often acquired in bulk and throughout the day via P2P networks and then consumed later. Collectively, subscriber demand for video typically peaks in the range of 8 PM to 11PM.

2. **Heightened subscriber sensitivity to quality:** Video is a sensory experience with rapidly changing sights and sounds, so shifts in quality (e.g., stalls, pixelization, compression artifacts, shifts up or down in resolution, changes in frame-rate) are instantly recognized by the viewer.

From the CSP perspective, video is decreasing network efficiency in the macro sense, in that the peak-to-trough ratio is increasing and there's a large amount of available but unutilized capacity throughout the day. During these high peaks, the network is more prone to congestion; from the viewer's perspective, congestion can very visibly manifest as degradation in video streaming quality.

Accordingly, CSPs worldwide are looking for solutions that measure and monitor the subscriber quality of experience (QoE) for video; and, in particular, for OTT video.[4]

Genuinely measuring the factors that influence or define a subscriber's video QoE is a major technological challenge. As a poor substitute, many vendors pitch solutions that aren't built around providing real insight, but instead are built around what's easy to measure. These solutions, which are often based on sampling technologies like NetFlow or meaningless (in the sense of video QoE) measurements like TCP packet drops, don't get CSPs any closer to understanding the actual customer experience.

To understand how to measure the video quality of experience (QoE) from the viewer's perspective—and to appreciate why so many solutions fall short—it helps to start with a simple example.

Figure 1 shows the instantaneous bandwidth of a 1080p Netflix video stream (y-axis), plotted over time (x-axis). This graph is the result of a mix of user actions and Netflix behavior:

- The user opens Netflix and selects a movie: this action is shown as the small (approximately 500 Kbps) bandwidth spanning either side of the 16:33:00 time label

- When the user selects a movie, Netflix bandwidth bursts to ~40 Mbps to build a sufficient buffer

- A few seconds later, the movie begins to play

- After a minute or so, Netflix has built a sufficient buffer and settles into a steady-state of playback

- At roughly 16:36:40, the user pauses playback: bandwidth drops to zero

- At roughly 16:38:15, the user resumes playback: bandwidth increases again (the movie continued to play through the end of the graph)

---

1. Up-to-date region-specific statistics can always be found at: https://www.sandvine.com/trends/global-internet-phenomena/
2. Learn about it here: https://www.bell.ca/Fibe-TV/Fibe-TV-App
3. Learn about it here: http://corporate.comcast.com/comcast-voices/xfinity-tv-app-2
4. OTT is usually the larger component of traffic; plus, CSPs often have access to server logs or other information to understand how well their own on-deck services are delivered.
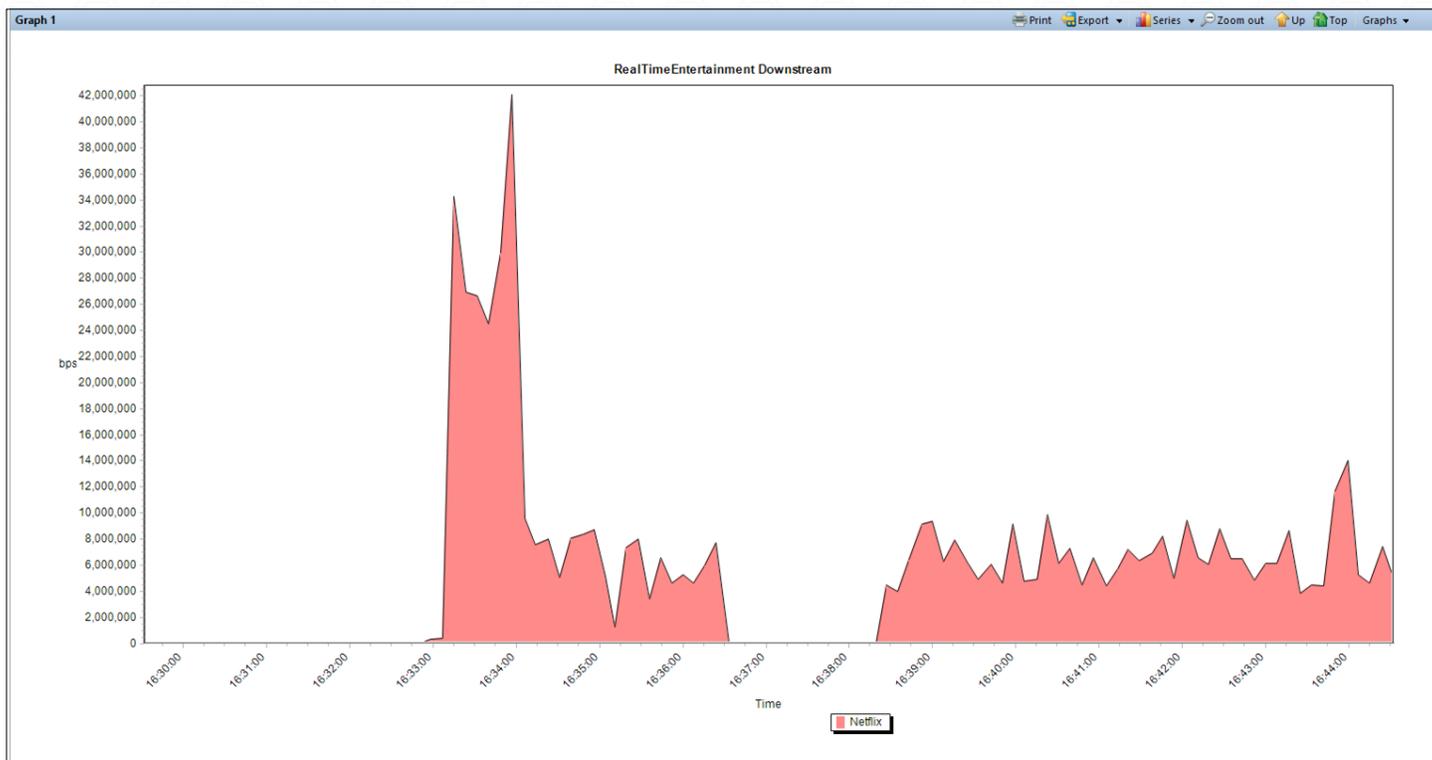
Figure 1: The anatomy of a 1080p Netflix video stream, shown in Sandvine's **Control Center** GUI. Note that this video played perfectly – the bandwidth gap is the result of a user pause.

From the user's perspective, the quality of experience was perfect: smooth 1080p playback with no buffer stalls or bitrate shifts.

Of course, that's not always the case. While some videos will play perfectly, others will suffer from buffer stalls and experience bitrate downshifts and upshifts. Some videos will play from start-to-finish without any additional user commands; others will have pauses, seeks (i.e., jumping forward or backward), and stops.

On a related note: Netflix can play continuously for up to four minutes without an active Internet connection. In this scenario, a solution that only looks at network behavior would indicate a terrible user experience, while the actual user experience would be perfect so long as the connection restored prior to buffer exhaustion.

And now we begin to appreciate the complexity of the goal: to truly measure the subscriber quality of experience for video, a solution must be able to detect and measure user actions and the behavioral characteristics of the video itself; plus, the solution must do so in an environment where most (likely to become all) video streaming traffic is encrypted, content associated with any individual stream can be split across multiple flows, and an individual household (or mobile hotspot, or other analogous point) can have multiple concurrent video streams in play.

The rest of this whitepaper explains and expands upon these requirements, so that communications service providers are able to understand the intricacies of measuring video quality of experience. The ultimate goal of this paper is to give CSPs the knowledge they need to select a vendor who can actually provide meaningful video QoE insight.

Solution Requirements

The following subsections dive deeply into particular aspects of an overall solution.

Measure Actual Subscriber QoE

To support CSPs' customer experience management initiatives, the video QoE solution must measure the video quality of experience from the perspective of the end user.

Because of widespread industry confusion and misconception, it is worthwhile to repeat that a foundational assumption of this paper is that CSPs want to measure the actual end user's experience; that is, the subscriber quality of experience for streaming video, or "video QoE" for short.

By this definition, video QoE is distinct from network health metrics, transport layer performance (e.g., TCP packet drops, round-trip time, jitter, etc.), bandwidth averages, etc. Any solution that purports to build a video quality of experience metric from those factors can (and should) be disqualified from consideration.

Be Encryption Agnostic

To be effective for today's and tomorrow's Internet, the video QoE solution must work with encrypted video.

Here are three truths about video streaming:

- Someday, practically all video streaming traffic will be encrypted

- Someday soon, *most* video streaming traffic will be encrypted

- Already, *an enormous amount* of video streaming traffic *is encrypted*

As a consequence, solutions that rely on reading video header fields (e.g., container type, codec, resolution, etc.) are already limited in their efficacy and have a very limited remaining operational life.

In addition to obscuring information about the video stream, encryption makes it quite challenging to correlate multiple flows that correspond to the same video. Weak solutions might produce an incorrect QoE score per flow, mistakenly interpreting the flows as separate streams.

Support Adaptive Bitrate Streaming

The vast majority of video streaming consists of adaptive bitrate streaming. To measure the factors that contribute to quality of experience for this type of streaming, the video QoE solution must be both streaming technique aware and application aware.

Streaming Technique Awareness

Despite the countless providers of Internet video, a small handful of streaming techniques account for the vast majority of traffic[5]:

- Dynamic Adaptive Streaming over HTTP[6] (DASH, or sometimes known as MPEG-DASH)

- HTTP Live Streaming[7] (HLS), by Apple

- HTTP Dynamic Streaming[8] (HDS), by Adobe

- Smooth Streaming[9], by Microsoft

---

5. "Vast majority" in the extreme literal sense; while there will probably always be proprietary techniques here and there, all the major (and most of the minor) services will definitely always use one of a few dominant streaming techniques.
6. Learn about DASH at: https://en.wikipedia.org/wiki/Dynamic_Adaptive_Streaming_over_HTTP
7. Learn about HLS at: https://en.wikipedia.org/wiki/HTTP_Live_Streaming
8. Learn (a little) about HDS at: https://en.wikipedia.org/wiki/Adaptive_bitrate_streaming#Adobe_HTTP_Dynamic_Streaming
9. Learn (a little) about Smooth Streaming at: https://en.wikipedia.org/wiki/Adaptive_bitrate_streaming#Microsoft_Smooth_Streaming

It's worth noting the absence of RTP, RTSP, RTMP, etc. Today's streaming techniques are almost entirely based on HTTP, and are specifically designed to work efficiently for Internet streaming. From a transport perspective, these all rely on TCP (see below for a note about QUIC).

The streaming technique defines the overall behavior of the video stream. Generally, the source content is encoded at multiple bitrates, and the overall files are segmented into chunks. The streaming technique, working with the video client, determines when to begin playback, how to fetch audio and video chunks, how to build the buffer, whether or not to split content across multiple concurrent flows, and other factors.

When a video QoE solution is streaming technique aware, that means it can first determine what streaming technique is being used and as a result understands how to properly interpret the observed behavior. As an added twist, a single application provider (see below) might choose to use a different streaming technique for different client devices, operating systems, or other variables.

Obviously, to truly understand the subscriber's experience, the solution must observe 100% of the video traffic. Furthermore, to achieve the CSP's objectives of having actionable insight, the solution must observe 100% of streams. This requirement disqualifies any sampling-based solutions from consideration: either they sample parts of an extended video flow, and therefore miss many short-term phenomena and actions; or they sample some video flows, missing others, and therefore can't provide detailed insight into any observed phenomena; or both.

In addition to encrypting their content, video providers are adopting other technologies to provide the best possible user experience. Technologies such as Quick UDP Internet Connections[10] (QUIC – initially designed by Google) are designed to optimize TCP traffic and remove head-of-the-line blocking as well as a way to multiplex TCP streams. To be viable, any video QoE solution must provide the same video QoE analysis for video streamed over QUIC as it does for video streamed over standard TCP.

Application Awareness

Each streaming technique itself only defines the high-level behavior of the video stream; it's up to the individual applications to determine the specifics (e.g., buffer length, chunk size, etc.) of each implementation.

In other words, while two application providers (e.g., Netflix and Amazon) might both choose to use DASH, their individual DASH implementations might vary. Plus, as noted in the section above, an application provider might use more than one streaming technique.

To truly understand the subscriber's quality of experience, the video QoE solution must know the specific implementation details of each application.

With a complete understanding of both streaming technique and application provider, for each video stream observed the solution understands the behavioral parameters, including the buffering strategy.

These parameters—and others—are required for a solution to correctly interpret observed, real-time phenomena.

Understand User-Initiated Actions

To distinguish between user-initiated phenomena and transport/streaming phenomena, the video QoE solution must be able to recognize and correctly interpret the result of user actions (e.g., play, pause, seeking).

In its simplest form, streaming video playback consists of the user starting a video and then sitting back while the video rapidly builds a buffer by fetching audio and video chunks, begins to play (pulling audio and video chunks from the buffer), reaches steady-state (the rate of chunk fetching and chunk playback is equal), and continues until the video is complete (no more chunks to fetch, buffer is exhausted). There may or may not be adaptive bitrate upshifts and downshifts to prevent buffer exhaustion.

In a more complex form, those same actions are interspersed with user-initiated pauses, plays, and forward- or backward-seeking.

Recall that in an encrypted world (i.e., our world), the solution tasked with measuring video QoE only sees data flying by; that is, it can't simply read user actions from the video control traffic. In this reality, the same user

---

10.   Learn about QUIC at: https://en.wikipedia.org/wiki/QUIC

action (e.g., a pause) can manifest differently in terms of the observed traffic depending upon the 'state' of the video stream. Only by carefully observing everything (i.e., 100% of packets associated with a video stream) and understanding everything (e.g., the individual application's implementation of a streaming technique, and user actions) can a solution make sense of what is being observed.

That is, to correctly determine the user's true quality of experience for video streaming, the solution must be able to understand how user-initiated actions manifest against a backdrop of observed and expected video streaming behavior, unique to the video application being used.

Support Progressive Video Downloads

Progressive video downloads are declining as a percentage of total video traffic, and are primarily used for short videos. Nevertheless, they contribute to subscriber quality of experience, so it's important that a video QoE solution provide detailed insights.

In a progressive video stream, a 'seek' action can occur when a user moves a control (such as a slider) to choose a different time or the client re-buffers on a stall; from a delivery standpoint, this corresponds to selecting and delivering different byte-ranges. Consequently, the duration of a progressive video stream can only be determined by keeping track of these seek actions dynamically for all HTTP transactions (i.e., not just the initial GET on an HTTP flow).

Similarly, tracking these seek actions is the only way in which the delivery quality of a progressive flow can be measured. To understand the display quality, the measurement system must be able to extract information directly from the video container or—in an encrypted download—infer based upon observations.

Figure 2 shows how a buffer stall can be detected by monitoring the actual and required progress of the stream; when the actual progress drops below that which is required for the requested display quality, the stream is forced to buffer.
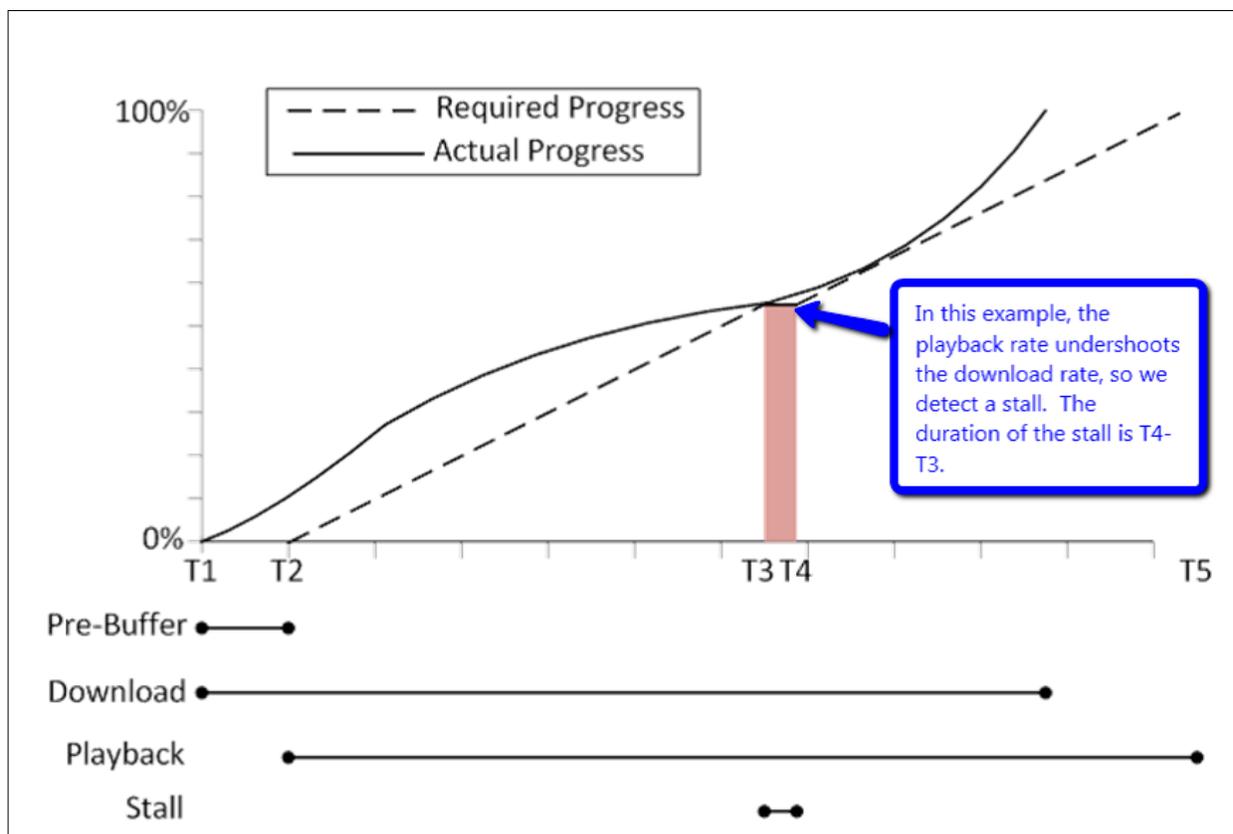

Figure 2: Measuring a buffer stall in a progressive video download.

Be Effective in the Real World

To be a useful tool for CSPs' CEM efforts, the video QoE solution must work in the real world.

That is, it must do all of the things described above, at line rate, in an environment in which:

- many thousands of videos are being streamed concurrently, and

- a single video might be split across multiple flows, and

- a single household might be watching multiple videos at once,

- on multiple devices and clients, and

- 100% of the video traffic is encrypted

In practical terms, for each individual video stream, the system must have the ability to recognize a client-side GET for video and the ability to correlate multiple bitrate streams into a single video playback

At a slightly higher level, the system must be able to distinguish between multiple concurrent video flows associated with a single access device (whether or not they're spread across multiple client devices).

And at the highest level, the system must do so at whatever the scale of the network, and subject to the network's routing environment.

These realities impose some practical system requirements:

- Performing the requisite packet operations demands an enormous amount of CPU capacity

- Maintaining the requisite state demands an enormous amount of RAM

- Fulfilling the requirement for observing 100% of the packets for 100% of the video flows requires that the solution maintain processor core affinity[11]

An Industry Whitepaper

Solution Considerations

So far, this paper has only explored the requirements that must be met for a video quality of experience solution to understand and interpret real-time traffic observations; but these are really just the inputs to a quality of experience system or calculation—what of the outputs?

Measurements vs Metrics

There are two fundamental types of output that can emerge from a video quality of experience system:

- A collection of measurements (e.g., number of upshifts, number of downshifts, number of buffer stalls, duration of buffer stalls, etc.) associated with each and—ideally—every video stream

- A single metric (e.g., a score out of five) associated with each and every video stream; presumably, this metric is calculated using the constituent measurements

There isn't really a right or wrong answer as to which is better, and CSPs should select a solution that can deliver both: a metric that clearly indicates the user experience for a video stream, and access to the foundational measurements that went into the metric's calculation.

With these atomic pieces in hand, the CSP can examine, leverage, and aggregate as needed. For instance:

- Examine in real-time; report on historic trends; leverage predictive analytics

- Contribute to individual subscriber profiles (e.g., for churn prevention); explore in aggregate to monitor the QoE of every piece of the network; compare the performance of different content delivery networks (CDNs), peering, and transit links

Calibration and Interpretation

CSPs should ask their vendors how the system is calibrated, and how to interpret different results.

Raw statistics (whether a calculated metric or foundational measurements) are one thing, but meaning is another. For instance, if a network operator sees that videos in a particular part of the network are playing with an average QoE score of X, then the operator will know:

- If X represents an improvement or degradation to past scores, by comparing to the past

- If X is better or worse than other parts of the network, by comparing to other locations' scores

However, the operator won't be able to tell if X is good or bad, in absolute terms.

To interpret the results in absolute terms, the operator has to rely upon some absolute scale.

How is that scale created? Since the goal of this solution is to understand subscriber quality of experience, then it makes sense that the system should be calibrated by real subscriber experiences.

To reflect regional differences that reflect legitimate differences in subscriber expectations, it should be possible for the CSP to manually calibrate (i.e., regionalize) their scales.

Accessing the Data

CSPs should ask detailed questions about how the information (whether metrics or measurements) is accessed. For instance:

- Are there proprietary GUIs that can be used to examine and explore the data? Figure 3 and Figure 4 show two such examples.
- Can the data be streamed to an existing big data system?
- Can the vendor's database be queried directly by third-party systems?

Figure 3: Screenshot of a Netflix Dashboard (Video QoE joined with other usage statistics) built in Sandvine's Network Analytics business intelligence product
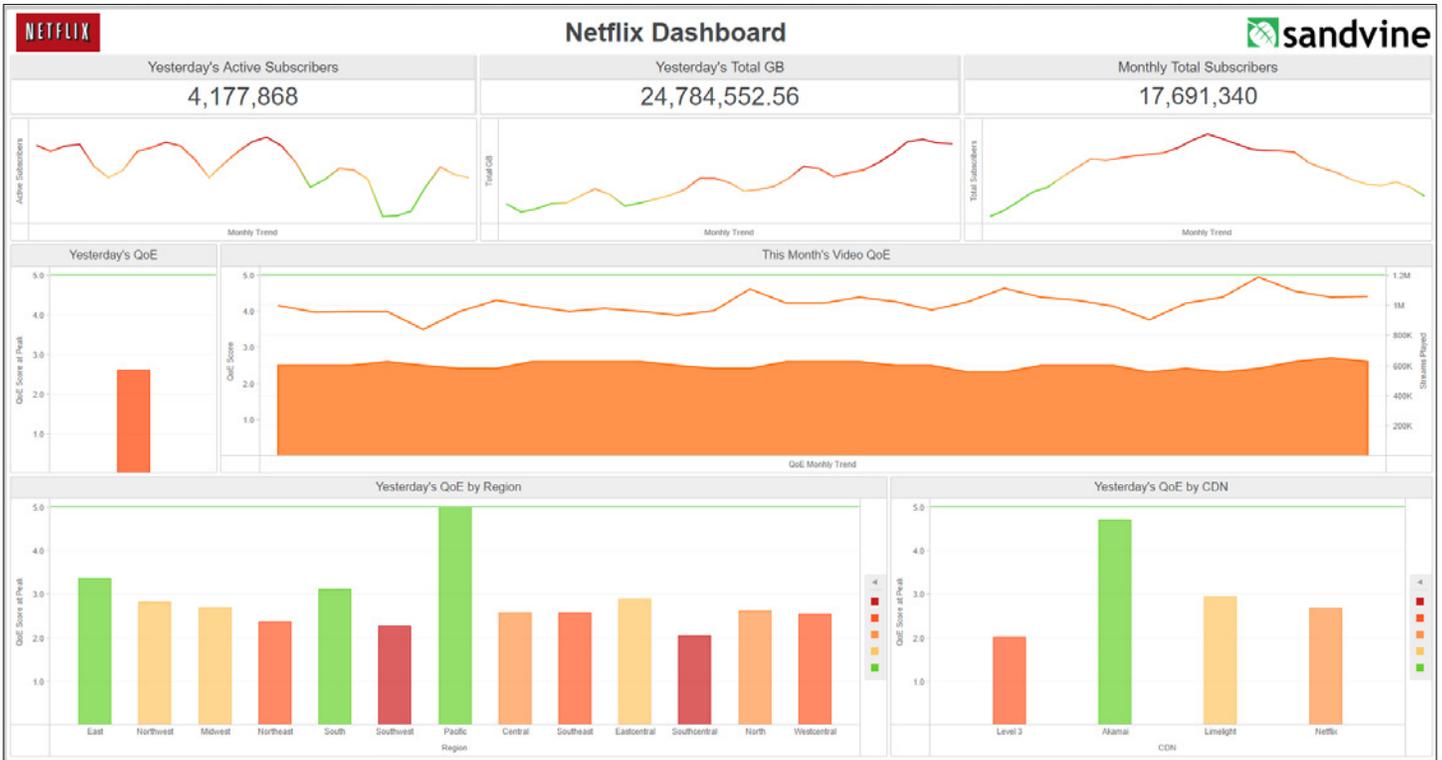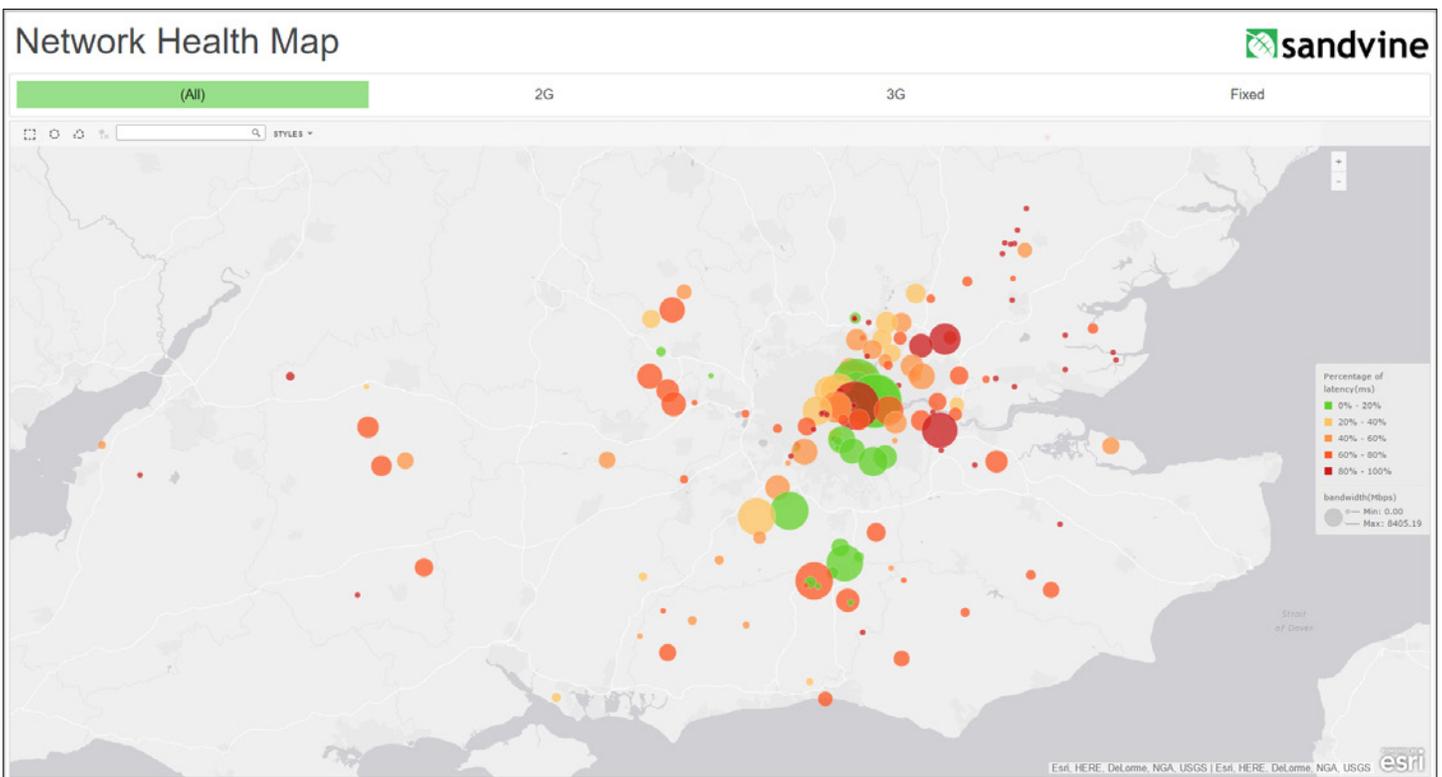

Figure 4: Screenshot of a Nework Health Map built in Sandvine's Network Analytics business intelligence product

Conclusions

There are a range of business motivations for communications service providers to monitor subscriber quality of experience for streaming video services. Whether for network engineering purposes, business intelligence insight, customer experience management initiatives, or some other reason, the ability to determine the quality of experience from the subscriber's perspective, for each video stream, is of paramount importance.

However, doing so is no easy task—especially with encrypted video traffic.

To correctly determine the user's true quality of experience for video streaming, a solution must be able to understand how user-initiated actions manifest against a backdrop of observed and expected video streaming behavior, unique to the video application being used—in an environment in which a single video might be split across multiple flows, a single household might be watching multiple videos at once, and 100% of the video traffic is encrypted.

Furthermore, to be viable and reliable in the real world, the solution must satisfy additional system and deployment requirements.

A number of these requirements are summarized in Table 1.

Table 1: Summary of Solution Requirements

Requirement	Description/Explanation
Measure actual subscriber QoE	The video QoE solution must measure the video quality of experience from the perspective of the end user. By this definition, video QoE is distinct from network health metrics, transport layer performance (e.g., TCP packet drops, round-trip time, jitter, etc.), bandwidth averages, etc. Any solution that purports to build a video quality of experience metric from those factors can (and should) be disqualified from consideration.
Be encryption agnostic	To be effective for today's and tomorrow's Internet, the video QoE solution must work with encrypted video.
Be streaming technique aware	The streaming technique defines the overall behavior of the video stream. The streaming technique, working with the video client, determines when to begin playback, how to fetch audio and video chunks, how to build the buffer, whether or not to split content across multiple concurrent flows, and other factors. When a video QoE solution is streaming technique aware, that means it can first determine what streaming technique is being used, and as a result understands how to properly interpret the observed behavior.
Be application aware	Each streaming technique itself only defines the high-level behavior of the video stream; it's up to the individual applications to determine the specifics (e.g., buffer length, chunk size, etc.) of each implementation. While two application providers (e.g., Netflix and Amazon) might both choose to use DASH, their individual DASH implementations might vary. Additionally, an application provider might use more than one streaming technique, differing the streaming technique depending upon the client device, operating system, or other factors. Additionally, an application provider might use more than one streaming technique, differing the streaming technique depending upon the client device, operating system, or other factors.

Understand user-initiated actions	To distinguish between user-initiated phenomena and transport phenomena, the video QoE solution must be able to recognize and correctly interpret the result of user actions (e.g., play, pause, seeking). Only with this intelligence can a video QoE solution understand how user-initiated actions manifest against a backdrop of observed and expected video transport behavior, unique to the video application being used.
Be effective in the real world	The system must monitor 100% of video streams on the network; for each individual video stream the system must have: • The ability to correlate multiple bitrate streams into a single video playback • The ability to recognize video from audio streams • The ability to track the current size of the buffer • The ability to recognize when a chunk has been downloaded • The ability to recognize a client-side GET for video At a higher level, the system must be able to distinguish between multiple concurrent video flows associated with a single access device (whether or not they're spread across multiple client devices). At the highest level, the system must do so at the scale of the network, and subject to the network's routing environment. These realities impose some practical system requirements: • Performing the requisite packet operations demands an enormous amount of CPU capacity • Maintaining the requisite state demands an enormous amount of RAM • Fulfilling the requirement for observing 100% of the packets for 100% of the video flows requires that the solution maintain processor core affinity

Beyond the hard requirements, there are additional solution considerations; these are summarized in Table 2.

Table 2: Summary of Solution Considerations

Consideration	Description/Explanation
Measurements or Metrics	For each video stream examined, the video QoE solution will produce a number of measurements (e.g., the number and duration of buffer stalls, the number of upshifts and downshifts, etc.). These measurements can be examined as they are, or can be rolled up into a single video QoE metric or score. There isn't really a right or wrong answer as to which is better, and CSPs should select a solution that can deliver both: a metric that clearly indicates the user experience for a video stream, and access to the foundational measurements that went into the metric's calculation.
Calibration and Interpretation	To interpret the results, whether in relative terms (i.e., better or worse than) or absolute terms (i.e., against an immutable scale), a CSP must rely on a calibrated system. To provide the most insight and meaning, the solution should be calibrated against real-world subscriber experiences.
How is the data accessed?	CSPs should ask detailed questions about how the information (whether metrics or measurements) is accessed (e.g., proprietary GUI, stream of data records, direct database queries, etc.)

By understanding the complexities associated with truly measuring video quality of experience from the end user perspective—and the requirements that these complexities impose—CSPs will be in a position to make an informed decision and choose the most suitable solution.

Additional Resources

Thank you for taking the time to read this whitepaper. We hope that you found it useful, and that it contributed to a greater understanding of how to meaningfully measure subscriber video quality of experience.

In addition to the resources cited in this document, please consider reading these documents related to traffic optimization, all of which are available on www.sandvine.com:

• Maximizing Performance with Core and Processor Affinity[12]

If you have any feedback at all, then please get in touch with us at **whitepapers@sandvine.com**.

---

12. https://www.sandvine.com/downloads/general/sandvine-technology-showcases/maximizing-performance-with-core-and-processor-affinity.pdf
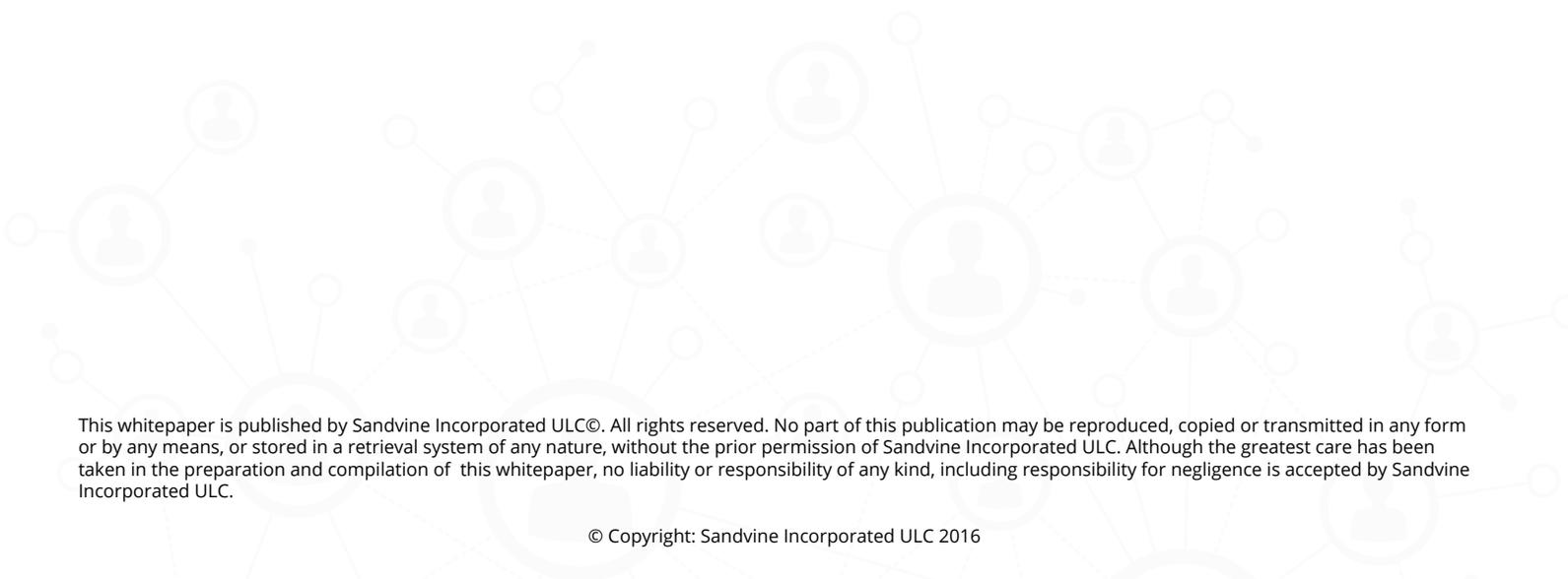
sandvine ®

Intelligent Broadband Networks

408 Albert St. Waterloo, Ontario Canada N2L 3V3

www.sandvine.com