OPTIMIZING THE VIRTUAL REALITY EXPERIENCE USING A 360º VIEW OF CLIENT AND NETWORK DATA

S. Mishra[1], E. Widner[1], R. Hadar[2], B. Stevenson[3]

[1]Verizon, USA, [2]CompiraLabs, Israel, [3]Openvu.TV, Canada

ABSTRACT

Virtual Reality (VR), and other forms of Extended Reality (XR), are beginning to drive deep immersive experiences for premium entertainment and sports content. In many cases, data is streamed directly to a headset, requiring the underlying network to accommodate constant throughput of over 25Mbps to support UHD 4K and over 50Mbps for UHD 8K. However, these more demanding video requirements are at odds with networks that are already congested and highly volatile. Granted, networks today have a glut of capacity on both wireline and mobile 5G networks, but available capacity along a delivery path may not be equitably shared across users, and can also drastically change over time and thereby stand in the way of smooth, high-throughput delivery.

INTRODUCTION

Virtual and Mixed Reality are beginning to drive deep immersive experiences for premium entertainment including sports content. These new formats have unlocked additional ways to consume rich 360° and VR content by streaming it directly to a headset, allowing the complete virtual reconstruction of a venue experience. Users may select any seat in the house and control what they want to see, and how they want to see it. Mixed Reality further enhances this experience by overlaying supplementary data, statistics, deep insights, and dynamic computer-generated objects that are contextual to the content stream. MR adds another layer of information, one that augments the action on the field, enhances a player's profile or provides additional information about the venue itself. It also allows content creators to connect the traditional streaming experience with the immersive on-line gaming culture.

COVID-19 also drove massive change as sports and entertainment transitioned from social activities to a virtual singularity. Leagues and venues began experimenting with ways to engage viewers in a more immersive way both at home, and in-person. Although the methods and technology have been evolving the last few years (i.e., minting NFTs during games, real-time stat overlays, multi-viewers), the constant is now that viewers expect an additional interactive dynamic engagement, and an event alone is no longer enough.

The VR CHALLENGE

This change in consumer behavior highlighted a key challenge for Network Operators. Available capacity in the network, particularly within the last mile was not equitably shared across users and could change drastically over time and therefore made a massive impact on the user experience and the delivery of high-bitrate content. Even if ample capacity were available, that alone did not solve packet loss and latency problems. As a result, consumers suffered from volatile throughput and jittery behavior, which adversely impacted their experience. In fact, even “traditional” HTTP-based video on demand frequently suffers from the same buffering, slow video start times, and low resolutions, because of inadequate data delivery paths across the network. These issues are exacerbated for low-lag applications like VR/XR, as the critical motion-to-photon timing required by these applications is severely impacted by packet loss, delay, and jitter, are therefore much more harmful to the consumer experience.

We argue that addressing on-path congestion, packet loss/delays, and throughput jitter, necessitates (1) enhanced visibility into the behavior of VR/XR content delivery, in particular across the last-mile network, and (2) tailoring congestion control to the specific requirements of VR/XR, and the prevailing network conditions with respect to different users.

MOTIVATION

Member companies and individual contributors who participate in Streaming Video Technology Alliance’s (1), Immersive Video Working Group (2) were engaged in several VR and MR deployments during the pandemic and one of the challenges that they quickly identified was that scaling VR content delivery into the home proved difficult.

Even consumers on both Fixed Access and Fixed Wireless Access networks with access to a glut of bandwidth - close to 1 Gigabit/s in some cases – they still experienced slow start, stalling and rebuffering issues that were extremely hard to diagnose, even with the deep client analytics provided by the player. A key challenge is the massive bitrates that are required to deliver VR content. These bitrates contributed significantly to additional packet loss, congestion, and latency in the last mile network.

The Immersive Video Working Group saw an opportunity to help address these issues by completing a Proof of Concept (POC) that would identify potential bottlenecks in the video pipeline between the edge cache and the user device. The deliverables would be comprised of a set of empirical test data, an analysis of that data and a set of best practices and recommendations for Content Owners and Network Operators that are considering launching VR services.

As the POC progressed it was quickly identified that while the Nice People at Work (3) player analytics provided rich data that quantified the Quality of Experience (QoE) for each user - e.g., average bitrate, rebuffering ratio, start time, etc. – additional data would be required to determine *why* those metrics were negatively impacted. That data would need to be gathered from the network.

This study puts forth a new approach for augmenting client player analytics with real-time network analytics. Not only can specific issues relating to user experience be identified in a timely manner, but these can also be correlated with the specific network conditions that have inserted the lag or jitter. We also propose mechanisms for resolving these issues by customizing congestion control logic to the VR/XR service and to the network. We present an in-depth empirical investigation of issues uncovered for VR content delivery and of how these were resolved. To this end, we leverage data gathered at Verizon labs.

POC METHODOLOGY

The Immersive Video Working Group then began to define the overall Proof of Concept methodology. The solution involved gathering HTTP session-based data from the Nice People at Work player plugin and blending that with Network data collected by the CompiraCloud (4) monitoring product. The Compira product provides a view into the last mile network and gathers information related to round trip time, throughput, packet loss, etc., on the TCP connection at one second granularity.

	NPAW	CompiraCloud
Source of data collected	Client/Player	Edge Cache Thin Agent
Type of Data	QoE	Network Analytics
Sample Metrics	Average Bitrate, Rebuffering Ratio, Startup time etc.	RTT, Throughput, Packet Loss, etc.
Data Granularity	HTTP Session	TCP Connection at 1 sec granularity
Purpose	QoE Metrics	Understand root cause for QoE issues.

Table 1 - An example of client-side and Network side KPIs captured in the POC

As shown in Table 1 above, The POC objective was to gather an aggregate data set where the sum of the parts provided the ability to determine a root cause for instances where the user session was negatively impacted. A drop-in data rate or increased rebuffering rations could be correlated directly with real-time changes in network conditions allowing for a much deeper understanding of the issues being faced when delivering a VR experience.

POC IMPLEMENTATION

The actual implementation and data collection was straightforward. Network Data was gathered from edge caches deployed in the Verizon Network (5) and uploaded to CompiraCloud instance running in the public cloud. Similarly, a Nice People at Work (Youbora client) was embedded within the Viaccess-Orca (6) player (VO), the VO player was integrated to run on Meta's Oculus Quest 2 Head Mounted Device (HMD) to collect client-side analytics which were then uploaded to the respective NPAW cloud. Both data sets were then merged on the back-end to generate a 360° view of the overall user experience compared to Network Conditions at a given time.

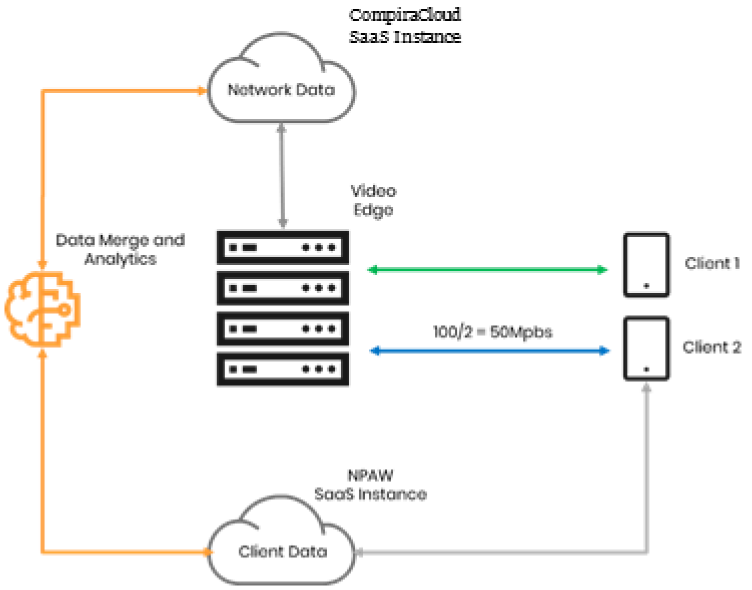

Figure 1 - POC high-level architecture

Deployment consists of deploying CompiraEdge, which consists of a thin agent (TA) in conjunction with the Congestion Control (CC) mechanism. Both the CC and TA modules are installed on the Edge Node. The TA is responsible for gathering performance-related statistics from the CC element and relaying them to CompiraCloud.

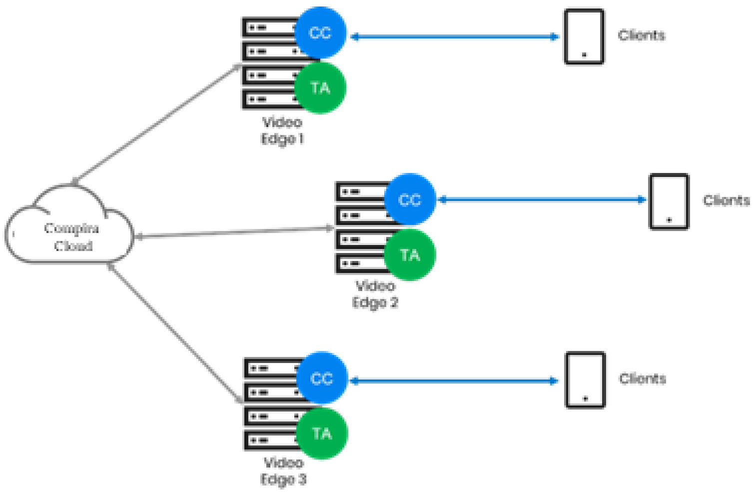

Figure 2 - Example of network-side data collection

PROOF OF CONCEPT – PHASE 1 TESTING (CALIBRATION)

The first step entailed installing CompiraEdge (CE) in the Verizon Network. An initial set of tests were run to validate that the CE was functional, and that the relevant data was being acquired from the network. This was completed using UHD 4K VOD content.

To ensure CE was performing as expected, the data gathered was compared with metrics obtained from a device client. A key consideration is that metrics were obtained at the HTTP/session level on the device while the CE results were measured in the TCP/connection level. As would be expected, the underlying metrics were distinct for each methodology however, the bitrates received and response times measured on the client could be correlated with the throughput and average RTT from CE to validate the test environment. This served as a rudimentary end-to-end sanity test prior to formal phase 1 analysis.

Source of data collected	Client/Player	Edge Cache Thin Agent
Type of Data	QoE	Network Analytics
Metrics	Bitrate=35Mbps Response time=7.2msec	Througput=32Mbps Average RTT=8.1msec
Data Granularity	HTTP Session	TCP Connection at 1 sec granularity

Table 2 - POC Phase 1 data calibration

A further analysis of the data collected by CE commenced based on 180 second windows during the streaming session in order to establish a baseline for all subsequent test cases. This yielded some interesting results.

The network itself was assumed to have low congestion and RTT conditions, however it was apparent that some packets exhibited high round trip times leading to an overall throughput average of 8.1msec.

The initial few seconds of each session displayed aggressive network utilization as the client attempted to fill the player buffer. Throughput peaked at 261Mbps while downloading around 600Mb of data for the first 20 sec of video. Even with a single streaming session, the network became congested resulting in packet loss of up to 2.2%. Following this initial burst fragments were then requested every 4-5 seconds and requiring 144Mbps which alleviated most of the packet loss for that session.

Based on these findings it was determined that even single streaming sessions over 300Mbps on optimal networks can cause congestion, mostly as the initial player buffer is filled, resulting in packet loss and increased round trip times. This provided an initial set of assumptions for testing VR content in an optimal network scenario.

PROOF OF CONCEPT PHASE 2 TESTING (WITH UHD 4K/8K SBR & ABR CONTENT)

Phase 2 involved gathering detailed metrics from VR content. 4K and 8K Single Bitrate and Adaptive Bitrate were tested and data was gathered and compared from both the network using CE, and from the player using client-side analytics. The results are displayed below.

PHASE 2 VR Testing - local cache

20.6.2022 runs	Youbora						Compira Cloud					
Test Type	Happiness Score	%Buffer Ratio	Join Time [sec]	Buffer Time [sec]	Bitrate [Mbps]	Buffer	Ave TpT [Mbps[	% packet loss	% idle sec	Ave RTT [msec]	Med RTT [msec]	Max RTT [msec]
DtoN 4KP25 - SBR	6.85	0	0.137	0	16.6	0	17.0	1.05	64%	11.5	5.0	119
DtoN 4KP25 - ABR	6.87	0	0.99	0	15.9	0	15.9	0.56	62%	17.0	6.0	101
DtoN 4KP25 - ABR	6.87	0	0.102	0	15.9	0	15.9	0.57	66%	20.1	5.3	116
DtoN 8KP25 - SBR	6.8	0	0.232	0	28.6	0	29.1	0.65	48%	11.5	6.4	150
DtoN 8KP25 - ABR	6.85	0	0.131	0	27.1	0	26.9	0.59	46%	21.4	5.9	152

Table 3 - POC Phase 2 test set

Table 3 above shows test type with test including a client playing content directly from the cache (DtoN) with 4K and adaptive bit rate (ABR) and single bit rate (SBR).

At a very high level, the results were generally as expected and consistent with the baseline results obtained in Phase 1. The average throughput was similar when measured on both CE and the Client and no rebuffing occurred during the session, although some packet loss did occur, mainly for SBR content streams. However, the data did display some anomalies, and a deeper dive into the telemetry was required.

Percentage of idle time shows the percentage of seconds in which a fragment was sent from the network. In the case of 4K, a segment was delivered ~33% of the time (62%-66% idle) while for 8K fragments were sent every second segment (46%-48% idle). This is indicative that the cache took more than 1 second to send an 8K fragment. Another unexpected outcome was that the Average RTT was higher for ABR but overall, there were certain peaks of max RTT on certain packet delivery which exceeded 100 Ms.

When reviewing the data, it was clear that when an ABR session was instantiated there was a few seconds of aggressive network utilization required to fill the player buffer and after that a fragment was then requested every 2-3 seconds. This was similar the base line SBR tests, however periodically the client needed to fetch more than 1 segment as illustrated below:

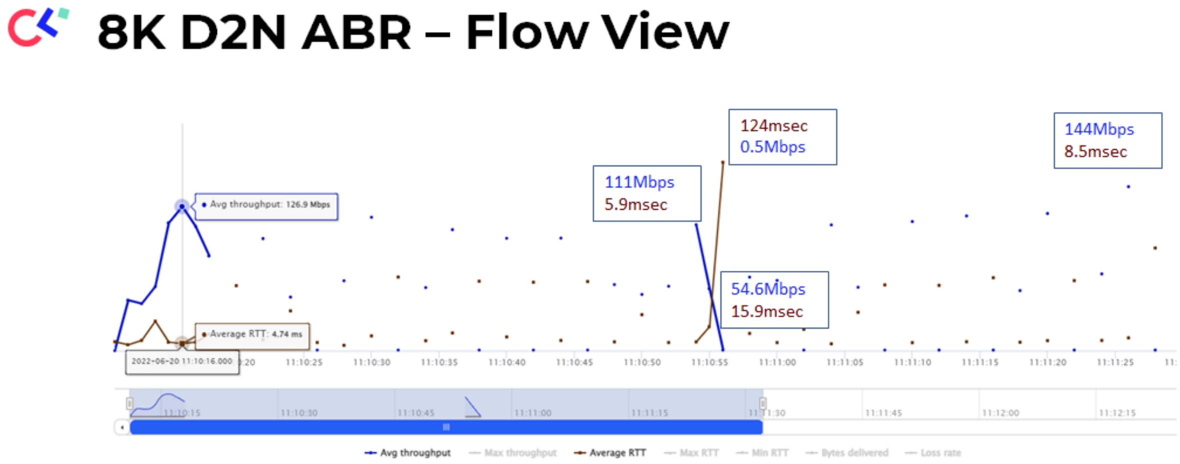

Figure 3 - 8K ABR data capture over time

However, when comparing the behavior of 4K to 8K SBR, it was obvious that network was sufficient to typically fetch a segment of 4K within 1 sec but 8K required multiple seconds to fetch a single fragment.

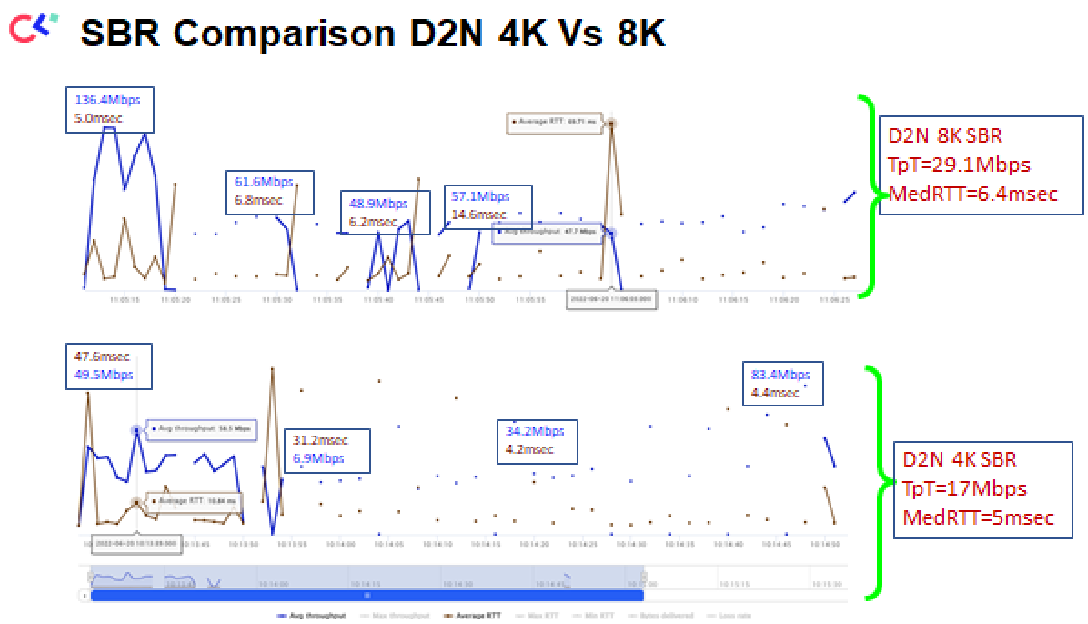

Figure 4 - SBR comparison for 4K vs 8K

The behavior of 8K SBR and 8K ABR was also compared and the SBR average throughput was slightly higher than the ABR. In the case of ABR there also more instances where more than a segment was fetched during the session. An interesting result was that with ABR the average RTT was higher than in the case of SBR as could be seen in the figure below:

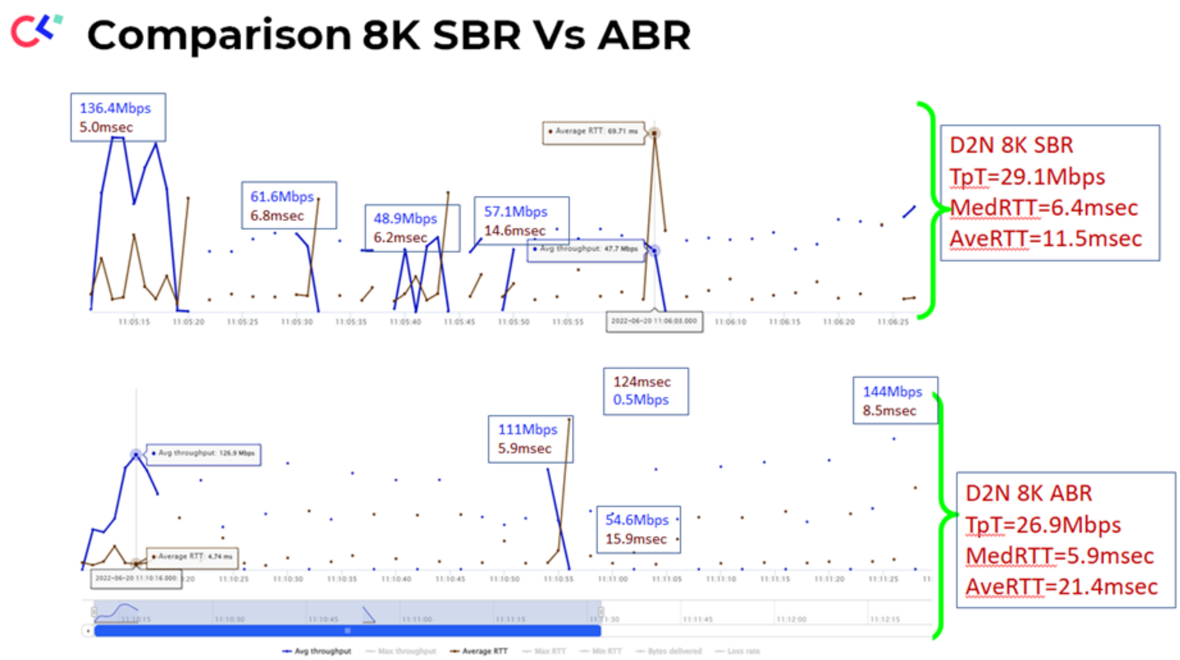

Figure 5 - Comparison of 8K SBR against 8K SABR

Overall, CE was able to provide enhanced telemetry that provided increased visibility into the network behavior and conditions causing loss and latency in the last mile network. This was true for both SBR and ABR streams for all video streams.

When correlated and viewed in conjunction with the client-side data, helped both the networking and caching planning teams identify those issues where the bottlenecks occurred and improved their ability to troubleshoot performance challenges.

CONCLUSION

The Proof of Concept leveraged TCP CUBIC as the default congestion control mechanism, however the POC infrastructure and approach is congestion-control-agnostic. A promising next step is investigating the impact of other choices of congestion control mechanisms and, moreover, of customizing congestion control to the VR context and to the prevailing network conditions. Another relevant use case would be to investigate VR experience exclusively over a 5G powered network.

The POC focused on ability to blend and reference client and network data as a single source and that method proved invaluable when drawing conclusions during the Proof of Concept. Not only could specific issues related to a consumers Quality of Experience be identified, but they could also be correlated with specific network conditions that caused the issue in the first place, and while this provides empirical data for the overall PoC report, it also forms a basis for the follow up best practices and recommendations document.

Additionally, the reference architecture implemented as part of the POC is generic in nature. It could be deployed very simply as an enhanced analytics infrastructure for normal HTTP streaming services that are implemented on commercial or open-source caching infrastructures as both agents are completely unintrusive in the scope of the overall streaming workload.

REFERENCES

1. https://www.svta.org
2. https://www.svta.org/working-group/immersive-video/
3. https://npaw.com/
4. https://www.compiralabs.com/
5. https://www.verizon.com
6. https://viaccess-orca.com

ACKNOWLEDGEMENTS

The authors would like to thank members of the Immersive Video Work Group within the Streaming Video Technology Alliance for their tremendous help in providing test content, tools and their assistance in executing these tests. We specifically would like to thank Vincent Lepec of Viaccess-Orca and Roc Toll of Nice People at Work for all their assistance and without their help this work could not have been done.