Performance Analysis of Multi-Source Wireless Multimedia Content Delivery

Anderson Augusto Simiscuka, Mikel Zorrilla, Pablo Cesar, Noel E. O'Connor, and Gabriel-Miro Muntean

***Abstract*—In order to create an improved experience in variable network delivery conditions, immersive multimedia content can be delivered over existing network environments, from multiple sources. These sources are normally servers located in the cloud, in various locations. Storytelling and certain related content, such as the immersive opera multimedia data in the context of the European Horizon2020 project TRACTION, require multimedia players to be able to receive content simultaneously from several locations, and at times, merge the content, creating new content in real-time. For instance, 360° recordings and polygonal 3D content can be delivered from different locations, and the end-user receives the unified content on his or her device. This paper introduces a study of how devices can be analysed, in terms of metrics, when receiving multimedia content from multiple sources, as the network and the devices have constraints regarding performance and video quality.**

***Index Terms*—Multimedia Delivery System, QoS, Wireless, Multi-source, Video.**

I. Introduction

Determining and maintaining high Quality of Service (QoS) levels during multimedia delivery is an important research area, due to device and networks constraints. Compared to single source streaming, multiple-sender based approaches show good performance when dealing with variable network conditions during actual multimedia streaming [1], [2].

The use of multi-source streaming can be useful for projects such as the European H2020 TRACTION project, which aims to develop a collaborative and participatory production toolset, for the co-creation and co-design of operas, with tools that support community dialogue, user-generated rich media capture, immersive audiovisual and 360° content, smart media editing, narrative engines and interactive adaptive media distribution.

In the context of the project TRACTION, which requires the development of collaborative multimedia players, it is important that content from multiple locations can be processed and, when needed, merged, without latency or delays.

Multi-source multimedia players must support, for instance, streaming of multiple pre-recorded recordings and live content from artists playing different instruments, merging the videos into one single experience, even when content is located in various locations. Other video elements that can be played simultaneously with a video stream include user feedback in video, commentators, and sign language interpreters.

A. A. Simiscuka, G.-M. Muntean and N. O'Connor are with Dublin City University, Glasnevin, Dublin 9, Ireland (e-mail: anderson.simiscuka2@mail.dcu.ie, {gabriel.muntean, noel.oconnor}@dcu.ie).

M. Zorrilla is with the Department of Digital Media, Vicomtech Foundation, Basque Research and Technology Alliance (BRTA), Donostia-San Sebastián, Spain (e-mail: mzorrilla@vicomtech.org).

P. Cesar is with the Distribute and Interactive System, Centrum Wiskunde Informatica, the Netherlands (e-mail: p.s.cesar@cwi.nl).

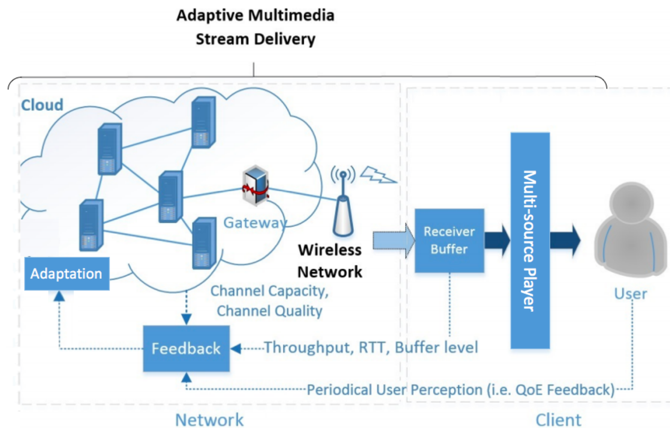

Fig. 1. Adaptive multi-source delivery system

Fig. 1 presents a diagram of an adaptive multi-source delivery system, which considers user feedback among other metrics to adapt the content being delivered from multiple servers in the cloud. A receiver buffer helps the player synchronise the content, and the challenge is to present the content in similar quality (e.g. in a multi-window player, it is not desired to watch one window with poor quality and another one with good quality). Adaptation can also be done at the servers, considering the feedback of the user, type of device consuming the data, and network conditions.

This paper introduces a study of how devices can be analysed, in terms of metrics, when receiving multimedia content from multiple sources, as the network and the devices have constraints regarding performance and video quality. This study is important for the development of novel algorithms and adaptation of video content.

The paper is organised as follows. Section 2 introduces the state-of-the-art related works in multi-source multimedia communications and immersive content. Section 3 presents the system architecture and quality metrics considered for analysis. Section 4 concludes the paper with final remarks and future work directions.

II. Related Works

A. Multi-Source Streaming

Substantial research has been performed on the field of multi-source streaming, such as the patented work presented in [3]. This patent focused on the efficient delivery of web content from one or more other content sources, such as images or events synchronised on screen at predetermined time increments in the media presentation. Another patent introduced the reception of digital multimedia data signals from multiple sources. It considers streaming over a fixed bandwidth, and the communications path may be accomplished by subscribing to a base layer of a first source and a base layer of a second source, and subscribing to an enhancement layer of the first source [4].

The paper presented in [5] introduced the design and evaluation of an adaptive streaming mechanism from multiple senders to a single receiver in Peer-to-Peer (P2P) networks. The approach, P2P Adaptive Layered Streaming (PALS), allows a receiver to adapt the streaming of encoded video streams from multiple congestion controlled senders, supporting multiple non-interactive streaming applications.

One of the main video adaptation approaches currently is the ISO/IEC MPEG Dynamic Adaptive Streaming over HTTP (MPEG-DASH), which has two main components: Media Presentation and Media Presentation Description (MPD). Media Presentation is a sequence of one or more segments that incorporate periods, adaptation sets, and representations, which break up the media from start to finish. The MPD is a document defined using the eXtensible Markup Language (XML), and it identifies the various content components and the location of all alternative segments, providing the relationship between them [6], [7].

The authors in [8] employed Content Centric Networking (CCN) for adaptive multimedia streaming in mobile environments. The approach uses the MPEG-DASH standard. CCN is used so devices efficiently switch between multiple links encountering bandwidth variations, in 3G and WiFi networks.

The authors in [9] proposed a multi-video stream bundle framework for interactive video playback with support to dynamic switching among multiple parallel video streams of different viewpoints of the same scene. An optimisation framework is part of the approach and an adaptive protocol improves quality of the parallel streams based on bandwidth conditions and stream switching probabilities.

B. Immersive Multimedia Content

Several tools are used in the process of immersive media authoring. These tools allow creators to produce several types of assets, such as 360° videos in monoscopic and stereoscopic formats, volumetric captures using specific tools and scripts, 3D models in tools such as Google Tilt Brush and Blocks from the Google Poly platform, audio/ambisonics created with Facebook's 360 spatial workstation, etc. [10], [11], [12]

The creation process aims to generate several files and assets that can be combined in the final applications compiled for multiple platforms, and installed in user devices (e.g. smartphones, desktops and VR headsets). Unity can be used in the creation of these immersive applications, as it provides features for importing various audio and visual assets, creation of 3D environments, and compilation into executable applications [13].

WebVR and the newest version of it, WebXR, are open specifications for VR experiences in desktop and mobile web browsers, allow the use of simple VR headsets, such as the Google Cardboard, with smartphones that support a compatible web browser. The main goals of WebVR and WebXR are to render content into two screens (one for each eye) and capture motion control from phones' accelerometers and gyroscopes. They can also be used with advanced headsets (e.g. Oculus Rift) and most recent desktop browsers, such as Chrome, Chromium, Firefox, Microsoft Edge and Safari. WebVR and WebXR have been important technologies for bringing VR to the general public, as they only require a web browser, found in computers and smartphones [14].

Plenty of web-based players and platforms support 360° content, such as YouTube, Facebook and Vimeo. YouTube also supports 180° content, which is more affordable to produce. Other websites usually create their own customised players, using libraries such as Three.js and Video.js. Three.js is a JavaScript 3D library with cross-browser compatibility that contains an API for the creation and display of animated 3D computer graphics in web browsers. Video.js is an open source library for video support on the web and it supports HTML5, Flash video, YouTube and Vimeo content, VR and 360° videos, and other features, through its extensive plugin library. Developers can decide which features can be supported in their custom players through the addition of available plugins. Video.js supports video playback on desktops and mobile devices [15], [16].

III. The TRACTION Player

The TRACTION player is one of the technologies being developed in the TRACTION project to provide support to immersive media from multiple sources in a heterogeneous network with diverse types of devices. These immersive files are mainly polygonal 3D models and scenarios and 360° videos.

A. Multi-Source Delivery and System Architecture

After the creation process, the assets and applications developed by authors in the various authoring platforms are uploaded to a server, the Media Vault, with multi-source capabilities, which means it supports multiple file formats for any location, as seen in Fig. 2. This server also provides tools for the management of the content and its metadata as well as for download/upload of immersive applications and assets, and streaming of 360° and 2D video (HTTP and FTP). Other important features of the Media Vault include storing accessible content (e.g. subtitles, audio description, sign language videos) and the transcoding of content into different resolutions and into MPEG-DASH, for adaptation and compatibility with the content player and web browsers.

The TRACTION player in conjunction with the media vault can adopt concepts of the delivery system from Fig. 1 and briefly described in the introduction. The architecture of the proposed DASH-based adaptive multimedia delivery system is deployed at both client and server sides. At the client side, a web player renders various segments into video/audio streaming and 3D scenarios, which will be deployed via different devices. At the server side, the platform annotates different information (e.g. QoS, QoE metrics) and encodes all information into an adaptive multimedia stream and associates it with the Media Presentation Description (MPD) files.

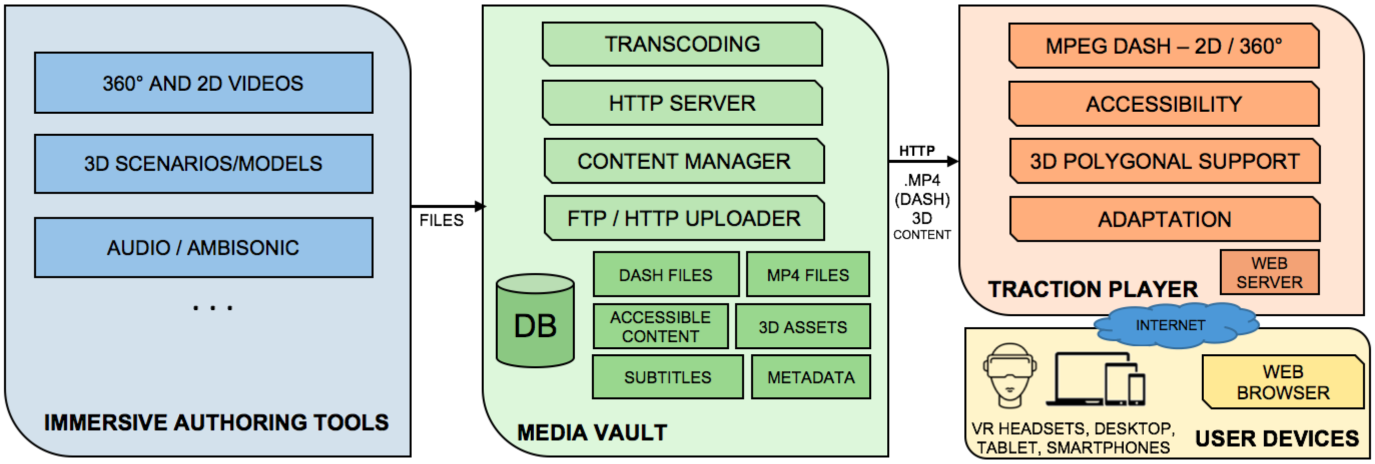

Fig. 2. TRACTION player architecture

After encoding, the sender delivers the adaptive multimedia stream from the sender buffer via the network based on the network conditions and user preferences. The stream is received at the receiver buffer at the client side, enabling the video/audio content playout.

The TRACTION player is web-based in order to run on devices' web browsers, available in smartphones, computers, tablets and VR headsets. The player requires a webserver (e.g. Tomcat), a HTTP server (e.g. Apache), and it has the capabilities of playing 2D and 360° content encoded in MPEG-DASH. The player also contains accessible features developed in the EU H2020 ImAc project [17], such as voice control, large menus, subtitling, audio description and sign language support. The components of the player are illustrated in Fig. 2.

The player is being extended to support immersive 3D models, and the main technology to be considered is WebXR, which allows the delivery of polygonal immersive applications through web browsers, and support for WebXR can be added with JavaScript APIs.

B. Performance and Quality Metrics

As the TRACTION player is being designed for use in a variety of devices, it is important the support for novel adaptive algorithms and schemes based on device, user and network requirements. The development of research in these algorithms will allow multiple concurrent users located in areas with limited Internet bandwidth and with a variety of devices to access and produce content at higher quality, even in constrained environments. The player must generate KPIs for the analysis of the network status, and the algorithms must adapt content in order to maintain higher quality of service and quality of experience.

Content quality in streaming needs to be constantly assessed and analysed by Quality of Service (QoS) and Quality of Experience (QoE) metrics, so content can be delivered at higher quality given network constraints. QoS metrics are related to data transport and network parameters, such as packet loss, delay, jitter, round trip time, etc. QoE metrics focus on the quality perceived by users. Another important metric for video quality assessment is the Peak Signal to Noise Ratio (PSNR), which is the ratio between the maximum power of a signal and the power of the signal's noise, and is used to measure the quality of video reconstruction during video compression [18], [19].

Several multimedia content adaptation techniques have been proposed over the years, reducing buffering times and adapting content to devices with various screen sizes and network capabilities. These techniques can be incorporated into the technologies used in TRACTION, in order to increase perceived user quality even in constrained devices and network conditions. Adaptation can also be based on type of device and content and considering user prioritisation.

Content encoded in the MPEG-DASH standard can be adapted by the use of several algorithms. These algorithms make the use of metrics and schemes such as stateful bitrates, bandwidth estimations, QoE models and Markov Decision Process (MDP) [20].

Image quality, especially in the context of opera content, can be adjusted for improved user experiences. Adaptation based on resolution and region of interest can improve quality of the video. The region of interest-based adaptive scheme [21], for instance, performs adaptation at the level of regions within clip frames, based on user interest obtained from eye-tracking monitoring. The scheme adjusts the quality of those regions from the multimedia frames the viewer is the least interested in, if necessary due to network conditions. Regions in which the viewers are the most interested in, either do not change or involve little adjustment, resulting in high overall end-user perceived quality.

360° VR videos and the underlying 3D geometry can be also divided into spatially partitioned segments/tiles in the 3D space, and be adapted with more or less priority, according to the regions the user are more likely to look. Colour can also be improved. For instance, 360° content recorded in theatres also film the audience, which is a region that is normally dark and the quality can be decreased as users are not interested in seeing it. The overall video size is then decreased, improving performance. On the other hand, colour quality of the stage can be improved with increased brightness and contrast [22], [23].

IV. Conclusion and Future Work

This paper presented one of the components of the EU H2020 project TRACTION, the TRACTION player. The player aims to support content from multiple sources and deliver it in a unified web application that supports immersive content, such as 360° videos and 3D environments.

This web application is intended to be used in several devices with different network and video requirements, therefore, metrics for adaptation of content and performance analysis were presented.

The TRACTION project is working to build a player that supports immersive media, includes algorithms for content adaptation and processes 3D models. Other features to be considered include intelligent annotation of faces, images and audio, as well as inclusion of novel user experiences, such as the integration of olfactory and gestural technologies.

Acknowledgement

This work was supported by the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement no. 870610 for the TRACTION project. The support of the Science Foundation Ireland (SFI) Research Centres Programme Grant Numbers 12/RC/2289_P2 (Insight) and 16/SP/3804 (ENABLE) is gratefully acknowledged.

References

[1] S. B. Lee, A. F. Smeaton, and G. M. Muntean, "Quality-oriented multiple-source multimedia delivery over heterogeneous wireless networks," *IEEE Transactions on Broadcasting*, vol. 57, no. 2 PART 1, pp. 216–230, jun 2011.

[2] B. Ciubotaru, C. H. Muntean, and G. Muntean, "Mobile multi-source high quality multimedia delivery scheme," *IEEE Transactions on Broadcasting*, vol. 63, no. 2, pp. 391–403, 2017.

[3] S. S.-P. Chang, H. Chen, J. S.-C. Chen, and E. Feig, "Efficient streaming of synchronized web content from multiple sources," U.S. Patent 6 715 126, Mar. 30, 2004.

[4] R. Hazra, "Method of streaming video from multiple sources over a network," U.S. Patent 6 510 553, Jan. 21, 2003.

[5] V. Agarwal and R. Rejaie, "Adaptive multisource streaming in heterogeneous peer-to-peer networks," in *Multimedia Computing and Networking 2005*, S. Chandra and N. Venkatasubramanian, Eds., vol. 5680. SPIE, jan 2005, pp. 13–25.

[6] L. Zou, T. Bi, and G. Muntean, "A dash-based adaptive multiple sensorial content delivery solution for improved user quality of experience," *IEEE Access*, vol. 7, pp. 89 172–89 187, 2019.

[7] A. Polakovic, R. Vargic, and G. Rozinaj, "Adaptive Multimedia Content Delivery in 5G Networks using DASH and Saliency Information," in *International Conference on Systems, Signals, and Image Processing*, vol. 2018-June, aug 2018.

[8] S. Lederer, C. Mueller, B. Rainer, C. Timmerer, and H. Hellwagner, "Adaptive streaming over Content Centric Networks in mobile networks using multiple links," in *2013 IEEE International Conference on Communications Workshops, ICC 2013*, 2013, pp. 677–681.

[9] N. Carlsson, D. Eager, V. Krishnamoorthi, and T. Polishchuk, "Optimized adaptive streaming of multi-video stream bundles," *IEEE Transactions on Multimedia*, vol. 19, no. 7, pp. 1637–1653, jul 2017.

[10] M. Kljun, V. Geroimenko, and K. Čopič Pucihar, "Augmented Reality in Education: Current Status and Advancement of the Field," in *Springer Series on Cultural Computing*. Springer, 2020, pp. 3–21. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-42156-4_1

[11] A. Çamcı, "Some considerations on creativity support for vr audio," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2019, pp. 1500–1502.

[12] A. Yaqoob, T. Bi, and G.-M. Muntean, "A survey on adaptive 360° video streaming: Solutions, challenges, and opportunities," *IEEE Communications Surveys and Tutorials*, 2020.

[13] Y. Kuang and X. Bai, "The research of virtual reality scene modeling based on unity 3d," in *2018 13th International Conference on Computer Science Education (ICCSE)*, 2018, pp. 1–3.

[14] B. MacIntyre and T. F. Smith, "Thoughts on the future of webxr and the immersive web," in *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 2018, pp. 338–342.

[15] S. A. Sarker, M. Rahman, N. Muslim, and S. Islam, "Performance analysis of video streaming at the edge and core cloud," in *2018 5th International Conference on Networking, Systems and Security (NSysS)*, 2018, pp. 1–7.

[16] M. Stanton, T. Hartley, F. Loizides, and A. Worrallo, "Dual-mode user interfaces for web based interactive 3d virtual environments using three.js," in *Human-Computer Interaction – INTERACT 2017*, R. Bernhaupt, G. Dalvi, A. Joshi, D. K. Balkrishan, J. O'Neill, and M. Winckler, Eds. Cham: Springer International Publishing, 2017, pp. 441–444.

[17] M. Montagud, I. Fraile, J. A. Nuñez, and S. Fernández, "Imac: Enabling immersive, accessible and personalized media experiences," in *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video*, ser. TVX '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 245–250. [Online]. Available: https://doi.org/10.1145/3210825.3213570

[18] A. A. Simiscuka, M. Bezbradica, and G. Muntean, "Performance analysis of the quality of service-aware networking scheme for smart internet of things gateways," in *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2017, pp. 1370–1374.

[19] A. A. Simiscuka, T. M. Markande, and G. Muntean, "Real-virtual world device synchronization in a cloud-enabled social virtual reality iot network," *IEEE Access*, vol. 7, pp. 106 588–106 599, 2019.

[20] D. Talon, L. Attanasio, F. Chiariotti, M. Gadaleta, A. Zanella, and M. Rossi, "Comparing dash adaptation algorithms in a real network environment," in *European Wireless 2019; 25th European Wireless Conference*, 2019, pp. 1–6.

[21] B. Ciubotaru, G. Ghinea, and G. Muntean, "Subjective assessment of region of interest-aware adaptive multimedia streaming quality," *IEEE Transactions on Broadcasting*, vol. 60, no. 1, pp. 50–60, 2014.

[22] A. Omidfar Sawyer and K. Chamilothori, "Influence of subjective impressions of a space on brightness satisfaction: an experimental study in virtual reality," *Proceedings of Symposium on Simulation for Architecture and Urban Design 2019*, 2019. [Online]. Available: http://infoscience.epfl.ch/record/273300

[23] A. Zare, A. Aminlou, and M. M. Hannuksela, "Virtual reality content streaming: Viewport-dependent projection and tile-based techniques," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 1432–1436.