Toward Truly Immersive Holographic-Type Communication: Challenges and Solutions

Alexander Clemm, Maria Torres Vega, Hemanth Kumar Ravuri, Tim Wauters, and Filip De Turck

Abstract

With significant advances in holographic display technology, a plethora of interactive applications, such as tele-conferencing and tele-surgery, are well on their way to integrating holographic technologies. However, hologram-based applications will place significant demands on networking infrastructure, which are not supported today. These include support for ultra-low delays, high bandwidth, and the ability to coordinate, synchronize, and dynamically adapt multiple data streams. This article articulates these challenges and points out gaps in existing networks that solutions must address. In addition, it provides an experimental analysis of novel network architectures that address one of these challenges, namely the ability to dynamically set up new flows with very low latency incurred by the first packet.

Introduction

Holographic-type communication (HTC) will enable the ability to transmit and interact with holographic data from remote locations across a network. Far from being just a gimmick, useful scenarios for such applications abound [1]. For example, holographic telepresence will allow remote participants to be projected as holograms into a meeting room to facilitate real-time interactions with local participants. Likewise, training and education applications can provide users the ability to dynamically interact with ultra-realistic holographic objects for teaching purposes.

This raises the question of which capabilities will be required from networks in order to support HTC. Are holograms just another encoding that can be streamed like video, perhaps requiring additional bandwidth, but otherwise handled at the application layer with little impact on existing networking infrastructure, or are there other ramifications?

To appreciate the specifics of HTC, it is important to understand the differences between holographic and 3D content. A 3D image is formed by two 2D static views of the same scene (left and right eyes). The image is the same regardless of the viewer's position. A hologram, on the other hand, adds parallax, meaning that the viewer can interact with the image, which changes depending on the viewer's position from which it is being seen. This changes the role of the user (from passive in 2D and 3D video content to the active and interactive one of holograms), increasing the requirements of HTC massively. Thus, it calls for solutions for capturing, transmission, and interactivity.

This article articulates the technical networking challenges to enable truly immersive HTC. Many of them are rooted in HTC approaches to reduce the networking bandwidth by devising clever compression as well as optimized streaming schemes. However, they in turn introduce new requirements to allow for coordination and synchronization of concurrent streams, with very high precision. In addition to challenges, this article pinpoints possible solution approaches. Along those lines, an experimental analysis of the performance of novel network architectures aiming to address one of the network challenges is provided.

The remainder of this article is structured as follows. The following section describes the state of the art in hologram capture and creation. It also reviews current streaming techniques for holograms. This provides the foundation that is used to highlight the technological challenges to enable truly immersive HTC discussed in the next section. Given the challenges, we then present a number of research tracks and solution approaches. We next show an analysis of the effects of decentralization of networks on the latency in holographic streaming services. The final section offers conclusions.

With significant advances in holographic display technology, a plethora of interactive applications, such as tele-conferencing or tele-surgery, are well on their way to integrate holographic technologies. However, hologram-based applications will place significant demands on networking infrastructure, which are not supported today. These include the support for ultra-low delays, high bandwidth, and the ability to coordinate, synchronize, and dynamically adapt multiple data streams.

Current Technologies for Capturing and Streaming Holograms

Streaming a hologram over the network means to capture, render, and stream a target object. Capturing is performed by a camera array outputting images of the target object from multiple angles and views. The camera feeds need to be merged, rendered into a hologram, and encoded. Finally, the hologram is streamed across the network. On the receiver side, a client receives the stream, decodes it, and renders it for a holographic display or beamer [1]. This section provides a brief review of the current state of the art of technologies for representation and encoding of holograms. Furthermore, it surveys the most common multimedia streaming techniques that will cope with HTC's requirements.

Representation and Encoding

Computer-generated holograms come in two flavors: image- and volumetric-based. In image-based holograms, a computer receives an array

Alexander Clemm is with Futurewei; Maria Torres Vega, Hemanth Kumar Ravuri, Tim Wauters, and Filip De Turck are with Ghent University.

Digital Object Identifier: 10.1109/MCOM.001.1900272

 0163-6804/20/$25.00 © 2020 IEEE

Authorized licensed use limited to: University of Exeter. Downloaded on May 08,2020 at 12:12:29 UTC from IEEE Xplore. Restrictions apply.

Even with compression, holograms will require massive bandwidth. To cope with it, the current tendency is to come up with clever techniques aiming to reduce data that needs to be transmitted by means of eliminating portions of the content that will go unnoticed by a user.

of images capturing an object (or set of objects) at different angles and tilts. One example is light field video (LFV) [2]. Image-based holograms are conceptually simple, but require significant storage and transmission bandwidth due to the massive number of concurrent individual images in the array. For example, if for representing an object, images every 0.3° were required, a hologram accommodating 30° of viewing angle and 10° of tilt would need a 2D array of 3300 separate images. However, as individual images across the image array include only minimal differences, sophisticated compression schemes can be applied to reduce the volume at the cost of computation [1].

The current trend is to base holographic representation on volumetric media, "point clouds." Thereby, the target objects are represented as sets of 3D volume pixels, or voxels, in a conceptual 3D box. The actual image is then dynamically rendered from any viewing angle at the local endpoint, placing the point cloud object into a scene or even rendering multiple point cloud objects simultaneously. Thus, it involves complex preprocessing and rendering, using setups with multiple camera angles with RGB as well as depth cameras. At the same time, volumetric media is highly compressed, as each voxel is transmitted only once; thus, the volume is independent of number of angles or tilts. To standardize volumetric-media compression, MPEG has recently selected a reference encoder, V-PCC [3]. This encoder converts point clouds into two separate video sequences, which capture the geometry and texture information, and applies traditional video coding techniques to compress the data.

Independent of the type of holographic media, compression and decompression come at the price of additional computation, which will heavily influence the latency incurred. Therefore, a higher level of compression trades off computation bandwidth and latency vs. required networking bandwidth and latency, and vice versa.

Streaming

Over-the-top (OTT) multimedia streaming provides the best solution to transmit video (2D, 3D, virtual reality — VR, HTC) over the Internet when no dedicated network is available. Video streaming services are categorized as either using the Transmission Control Protocol (TCP) or the User Datagram Protocol (UDP). While streaming solutions prioritizing quality over latency tend to select TCP, solutions that optimize real-timeliness opt for UDP.

Among TCP-based solutions, Real-Time Media Protocol (RTMP) and HTTP Adaptive Streaming (HAS) are the most widely adopted [4]. RTMP splits streams into fragments, dynamically negotiating sizes between client and server. Thus, fragments from different streams can be interleaved, and multiplexed over a single connection. HAS, on the other hand, segments at the application level, encoding the stream at different quality levels and temporarily splitting them into segments of predefined duration. When streaming occurs, the client decides the quality to request next, where requests and transmissions are performed using HTTP. In an attempt to optimize the bandwidth requirements of ultra-high bandwidth applications, HAS also allows splitting the stream not only in time but also in space (tiles). Thus, the HAS client is able to decide which areas of a scene to download at higher qualities. This becomes very useful for viewpoint or viewing-angle-based applications such as VR and HTC. Kara *et al.* [1] have recently demonstrated the possibilities of applying the HAS principles to the streaming of LFV.

TCP-based solutions prioritize quality over latency, inducing additional latency due to the need for continuous packet acknowledgment. This makes them unfit for real-time and interactive transmission. In contrast, solutions based on UDP are better suited for low-latency, non-quality-assuring services. Such is the case of the Real Time Protocol (RTP) combined with the control channel provided by RTCP, widely used for low-bandwidth and packet-loss-resilient 2D video services. Furthermore, approaches like Quick UDP Internet Connections (QUIC) and Web Real-Time Communication (WebRTC) improve the performance of UDP-based solutions, but their application to HTC remains unexplored.

The User Interactivity Challenge and Its Requirements from the Network

Even with compression, holograms will require massive bandwidth. To cope with it, the current tendency is to come up with clever techniques aiming to reduce data that needs to be transmitted by means of eliminating portions of the content that will go unnoticed by a user. For example, some areas may be obstructed from the user's viewpoint, or some angles may not come into view based on the user's position. Schemes that take advantage of these aspects can dynamically adapt which parts of the contents to stream, and at what quality, at any given time. However, the effectiveness depends on the ability to predict the user's movement, thus rapidly adapting the data as needed. This is referred to as the "user interactivity challenge," and will enforce not only ultra-high bandwidth but also ultra-low latency (to ensure interactivity with the content). In addition, perfect synchronization of concurrent flows will be needed. Contrary to other types of multimedia services (e.g., UHD streaming), the interactivity challenge of immersive HTC will require ultra-low latency even if dealing with prerecorded content that does not involve real-time interaction with a remote party, as the user still interacts with the content simply by virtue of changing viewing angle and position. The remainder of this section discusses qualitative and quantitative details of the three network requirements that need to be fulfilled simultaneously to enable the user interactivity challenge for HTC.

Ultra-High Bandwidth

HTC streaming will lead to skyrocketing bandwidth demands into the terabits-per-second range, an increase by several orders of magnitude over HD or even 3D VR video (Fig. 1). A camera sensor such as the Microsoft Kinect for Windows v2 outputs a 1080p image (1920 × 1080 pixels) with 4 bytes of color data per pixel and a depth image with a resolution of 512 × 424 and 2 bytes of depth data per pixel. This is equivalent to 8.8 MB (70.4 Mb) of raw data per

Authorized licensed use limited to: University of Exeter. Downloaded on May 08,2020 at 12:12:29 UTC from IEEE Xplore. Restrictions apply.

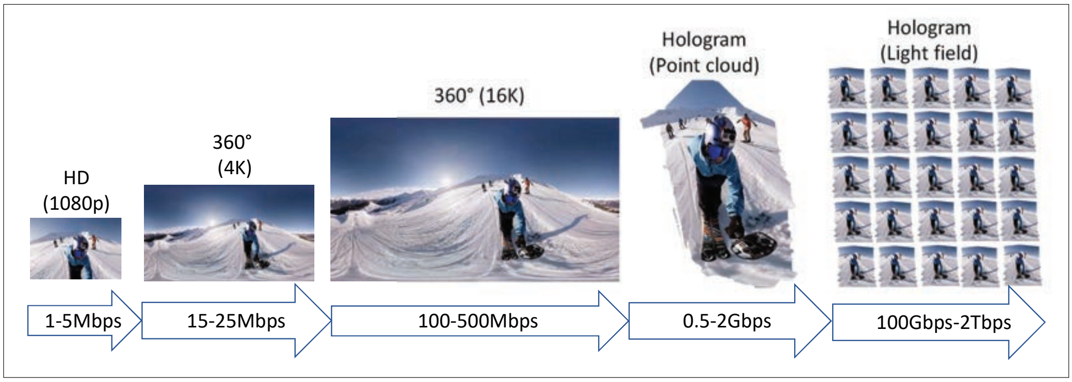

Figure 1. Bandwidth requirement evolution of video streaming services.

frame, or 2.06 gigabits of raw data per second at 30 FPS. This amount increases with more sensors and viewpoints as well as higher resolutions and frame rates. For instance, LFV content is expected to require bandwidth in the range between 100 Gb/s and 1 Tb/s [2].

In this situation, there is no question that very high bandwidth needs to be supported in order to support HTC, even when sophisticated compression and user interactivity prediction schemes mitigate the problem. 5G utilizes higher regions of the electromagnetic spectrum to increase available bandwidths. However, to unlock their full potential, novel management techniques need to be devised [5].

Ultra-Low Delay

In addition to the very high bandwidths with novel wireless infrastructures, the 5G paradigm sets the limit of 1 ms round-trip latency [5]. While achieving this over long distances is nearly impossible due to the physical limitation of the speed of light, ultra-low latency is crucial for true immersiveness to alleviate simulator sickness, especially if head-mounted displays (HMDs) are involved. Furthermore, applications involving real-time communications impose stringent latency requirements on the transmitted stream.

Transport protocols enable reliable and faster delivery of data. However, the well-known tradeoff between latency and reliability of TCP and UDP becomes much more critical for the delivery of HTC. Many applications are highly sensitive to both loss and latency. In addition, the ability to deploy schemes that adjust the quality of fields in the image array based on user interactivity depends on the network's ability to rapidly adjust streams.

Stream Bundle Synchronization

When handling HTC traffic, the network needs to manage a massive number of synchronized streams originating from either different sensors, an object at different angles, or a processed volumetric fusion [6]. Figure 2 shows a case of streams from misaligned sensors resulting in a badly rendered image [6]. Likewise, in the case of moving images, it is unacceptable for fields in the array to be slightly ahead or behind each other. In addition, when streams involve data from multiple sources (video, audio, tactile), stringent synchronization between streams is required (i.e., not just in time but right on time).

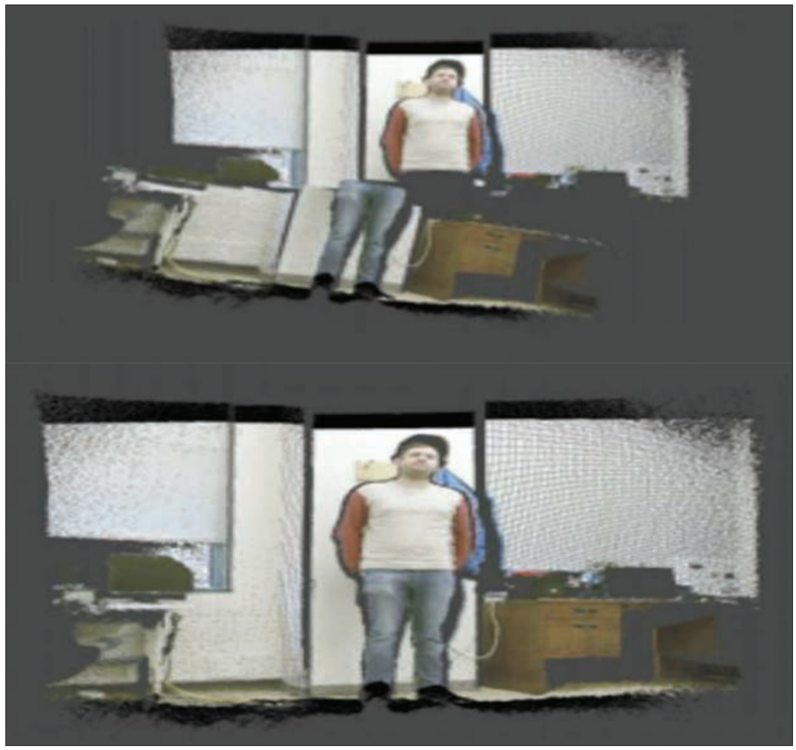

Figure 2. Images showing a misaligned (top) and a correctly aligned dataset [6].

Toward Fully Interactive Holographic-Type Communications: A Cross-Layer Approach

The user interactivity challenge of HTC cannot be addressed by focusing only on a single networking aspect, but needs to be addressed across network layers. Based on the state of the art of networking solutions for immersive media (ranging from 3D to VR to LFV), this section aims to pinpoint challenges, open areas of research, and possible solutions to enable truly immersive HTC. We have focused the analysis on three upper layers (Fig. 3). As approaches at the higher levels impose requirements on the lower ones, the discussion of this section starts at the level of the end user.

End User

The user of a holographic system only looks at two of the potentially thousands of fields composing the hologram at each instant of time. This

Authorized licensed use limited to: University of Exeter. Downloaded on May 08,2020 at 12:12:29 UTC from IEEE Xplore. Restrictions apply.

There are tradeoffs between QoS parameters and perceived QoE. The perceived quality of a hologram depends on many parameters (frame resolution and rate, degree resolution, etc.). Understanding the effects and "sweet spots" of each of the parameters on the user perception is fundamental to help improve the bandwidth consumption while maintaining the user's QoE.

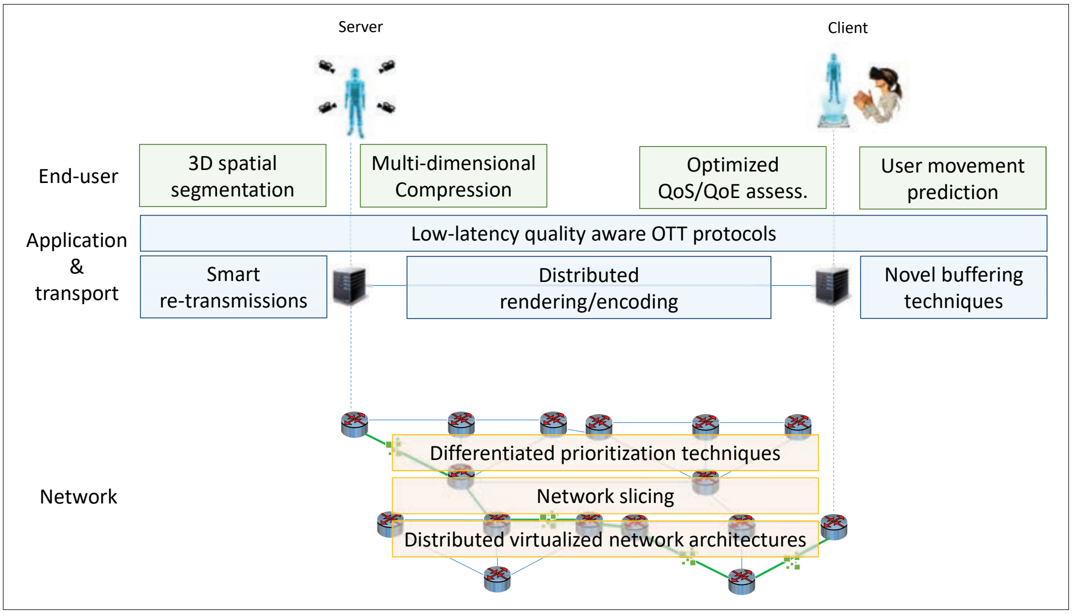

Figure 3. Challenges and solutions for a cross-layer optimization of infrastructure.

translates to bandwidth wasted on contents that are never viewed. Predicting the user's movement allows optimizing the streaming of the holographic views. For example, by determining which parts of the hologram the user is most likely to see, the quality of the image streams can be adapted.

Viewpoint prediction has recently been explored for VR applications. One prominent example is omnidirectional adaptive tile-based video streaming [7]. However, current prediction models are simple, with error probability up to 40 percent, leaving room for improvement [7]. Extending these VR viewpoint prediction approaches to the HTC case can be a good starting point, but HTC imposes two additional challenges. First, approaches need to go from the three degrees of freedom (DoFs) of VR HMDs (yaw, pitch, and roll) to the full six DoFs (up/down, left/right, forward/backward, yaw, pitch, and roll) supported by holograms. Second, currently developed viewpoint prediction solutions assume the presence of an HMD, which allows for tracking the user's trajectory. However, the actual aim of HTC is to provide a truly immersive experience without wearables. As such, user movement prediction without HMDs is required. Current light-field displays are capable of providing 3D images with continuous motion parallax in a wide viewing zone without having to wear glasses [1]. However, these displays lack user movement tracking mechanisms.

One further bandwidth optimization is 3D spatial and object segmentation. One such technique is 3D tiling [7], an extension of 2D tiling for spherical videos. Holograms are often represented as a sequence of voxelized point cloud frames, grouped into groups of frames (GoFs). Each GoF is spatially divided into 3D tiles. Depending on user viewpoint and throughput, 3D tiles are streamed at different resolutions. 3D tiles are a much greater challenge than 2D tiles. First, there are polynomially more 3D tiles than 2D tiles, which requires efficient indexing and addressing mechanisms. Second, a 3D tile may not be occupied by content and can be empty for a significant time. Third, they can be occluded based on the user's viewpoint , which requires special consideration to deal with sparsely populated spaces efficiently.

Multi-dimensional compression is a well-known technique to reduce the amount of streaming data. As introduced earlier, V-PCC is the standard encoder for point clouds due to its simplicity in acquisition, computation, and storage [3]. However, one fundamental issue still requires more attention. This is inter-frame redundancy, which is considerable due to the high frame rates with which point cloud sequences are captured. In this direction Mekuria *et al.* [8] propose a lossy real-time color-encoding method exploiting interframe redundancy in unorganized point clouds. Regarding light-field approaches, given that an LFV sequence can be seen as a 2D array of single views, the most straightforward procedure of compression is to do so per individual view, as for a 2D video stream [9]. However, while this method considers intra-view relationships, it ignores valuable strong relations between views. A possible manner to improve the prediction of the inter-view relations is to apply multi view coding (MVC) to LFV, reported to achieve 23 percent better compression rate than standard MVC [9]. Independent of the type of hologram and compression technique, the challenge of the trade-off between compression, computational, and time costs still remains. However, with the increasing requirements that HTC imposes, it will not be enough to just concentrate the full computation on the servers. The need to distribute the encoding computation to the network arises, as shown in the next subsection.

Finally, there are trade-offs between quality of service (QoS) parameters and perceived quality of experience (QoE). The perceived quality of a hologram depends on many parameters (frame resolution and rate, degree resolution, etc.). Understanding the effects and "sweet spots" of each of the parameters on the user perception is

Authorized licensed use limited to: University of Exeter. Downloaded on May 08,2020 at 12:12:29 UTC from IEEE Xplore. Restrictions apply.

fundamental to help improve the bandwidth consumption while maintaining the user's QoE.

Analogous to video, where quality of individual images may be traded for frame rates, with holograms, the spatial resolution provides an additional dimension for trade-offs (Fig. 4) that can be leveraged to maximize QoE. However, this requires the ability to dynamically adapt parameters.

Transport and Application

In order to deal with the limitations of the transport layer, developers have been adopting application-level strategies (i.e., OTT streaming protocols). One example is MPEGDASH, optimized for on-demand content [7]. In order to provide real-time communication, there is a tendency to turn to UDP-based solutions. One example is WebRTC. However, its application to HTC and how to deal with quality requirements is still in a nascent stage. Other studies have shown the potential of adopting QUIC over HTTP/3. This could provide a quality-managed low-delay streaming option.

Another possible way to speed up TCP-based solutions is to develop smarter retransmission schemes. Current TCP-based solutions will retransmit any missing packet independent of its importance within the stream. For HTC, the retransmission of every lost packet will increase the latency exponentially. At times, retransmission is required, such as when transmitting/updating the manifest file or the anchors used for rendering virtual views [10]. In other cases, the loss may be acceptable, for instance, in the case of incremental data of less important frames. The transport layer needs to be capable of handling these different scenarios and differentiating what to do for which packet/stream/flow.

Achieving full user interactivity in HTC also requires novel buffering schemes, possibly extending from the client to the edge of the network. Window-based buffering is a good example in that direction. It allows for fast reactions to unforeseen user behavior by inserting updated content just in time before it needs to be played, rather than at the end of the queue [11]. Accordingly, buffering schemes not only need to be supported by clients that render playback at the host, but also to apply to packets and frames of streams, as they are queued and buffered across the network. Network function virtualization (NFV) and edge computing are being investigated toward improving the end-to-end buffering [12]. However, while these caching mechanisms are promising, their application to HTC is still unknown.

Finally, as mentioned earlier, the trade-off between storage cost, computational cost, quality (encoding rate), and start-up latency could highly benefit from distributed rendering and encoding. One approach to save storage and reduce start-up delay is to save only the master representation at the highest quality and render all others on the fly at the expense of computing [1]. Another option is to offload the encoding task to the network. Toward that, several early works have proposed using NFV to place network encoders or transcoders in such way that the raw content is streamed through the network and is encoded at the required quality along the network path or at the edge [12].

Artefact	Parameters
Still image	Resolution Color depth
Video	Frame rate
VR/AR	Tiles
HTC	Views per 6DoF: Angle, Tilt, Distance

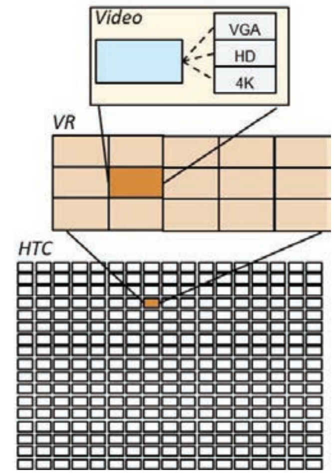

Figure 4. Trade-off between quality and bandwidth requirements for different video streaming services.

Network

Traditional networking infrastructure is limited in its ability to adapt to new requirements. Software defined networking (SDN), which separates the functionality of the data from the control plane, promises flexibility, reconfiguration, and on-demand resource allocation [13]. This makes SDN a natural candidate platform for HTC. However, continuous control communication required as a result of centralized implementations impacts the end-to-end latency considerably and impairs scalability.

This suggests distributed control solutions. Various architectures with different levels of distribution of control intelligence across the network have already been proposed. For example, flat SDN control architectures partition the network horizontally into subsets of switches, each governed by a local controller [13]. However, while the advantages of intelligence distribution in network architectures have been proven in theory, its real impact on network latency has until now only been poorly understood. The next section provides an assessment of the impact of SDN control distribution on latency.

In addition, network slicing can help cope with guaranteeing the QoS requirements of all the possible HTC-based services sharing the same physical infrastructures [12]. In it, a subset of the underlying physical network (i.e., a network slice) is dedicated such that a set of requirements (e.g., latency, throughput, and reliability) are guaranteed. Network slicing is considered one of the key points of the new fifth generation (5G) network infrastructures, such as for the 5G service-based architectures (SBAs).

Finally, HTC-enabled networks will also be required to provide sophisticated differentiated prioritization techniques. Streams belonging to the same scene or object need to be prioritized based on their impact on the user's visual perception. One example concerns the field of view (FOV) (40°–60° nasally, directly in front of the eyes). It can be exploited to prioritize streams, where frame rates, resolution, and color depth of various streams are dynamically changed based on location within the FOV. Prioritization can

Authorized licensed use limited to: University of Exeter. Downloaded on May 08,2020 at 12:12:29 UTC from IEEE Xplore. Restrictions apply.

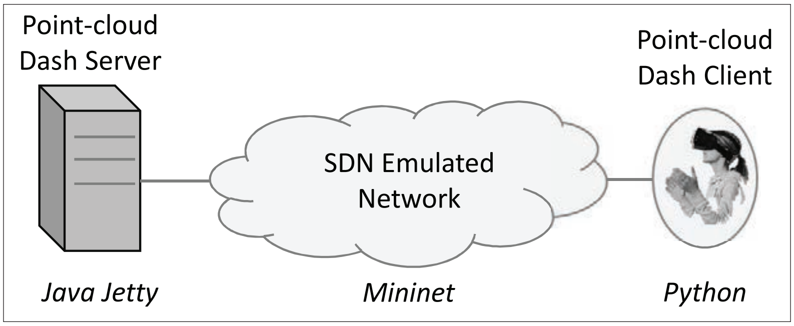

Figure 5. Point cloud streaming over the SDN emulated network.

Case	Flow setup (ms)	1st segment (ms)	Av. segment (ms)	% HQ
Cent. (floodlight)	53.8	258.54	925.11	98.3
Flat distr. (ONOS)	58.3	263.3	903.24	95.8
Hier. distr. (kandoo)	27.4	236.43	919	98.3
Fully distr. (DAIM)1	5.1	224.8	918.28	98.3

Table 1. Latency comparison for point cloud streaming over different SDN-based architectures.

also extend to video objects, where each video object can be prioritized based on its contribution to the QoE. Similar considerations apply for differentiated retransmission, selectively choosing which data needs to be transmitted in case loss occurs.

Latency Analysis of Holographic Streaming in SDN Arhictectures

As already mentioned, minimizing network latency is critical for HTC. One specific challenge concerns the latency for the first packet in a flow. This generally incurs greater delay than later packets, as flow rules and entries may not yet have been installed. While the problem is not unique to HTC, addressing it becomes of crucial importance in the presence of solutions that involve multiple data streams and packet flows which can be dynamically started and stopped as part of adapting the contents being streamed to user interactivity. To this end, this section presents an analysis of the effects of decentralization on the latency of the flow setup and the transmission for the case of HTC.

In particular, we focus on the case of DASH streaming of point clouds over a Mininet-based emulated client-server network setup (Fig. 5). The client [14] provides support for different camera traces and rate adaptation heuristics. It is connected to an HTTP/1.1-enabled Jetty point cloud server. Between client and server, a network setup with different possible control plane architectures is established. For this analysis four different SDN architectures are put to test: centralized (Floodlight), hierarchically distributed (Kandoo), flat distributed (ONOS), and fully distributed (DAIM) [15]. For the streaming, the dynamic point cloud object ”Loot” (120 s, 30 fps) is used, compressed at two quality levels, LQ (2.4 Mb/s) and HQ (18 Mb/s) [14], and 1 s segments. A buffer of length 4 s is provided on the client, and the network bandwidth was set to 20 Mb/s.

Three latency measurements are performed: flow setup, first segment download, and the average per segment download (Table I). Additionally, the table shows the percentage of segments downloaded at the highest quality. As can be seen, the fully distributed case (DAIM) outperforms the others in terms of flow setup latency, which further results in better performance while downloading the first segment. It also results in good average download time per segment while keeping a very high percentage of HQ segments downloaded. These results encourage research in this direction. However, it has to be noted that the initial bit rate of the dynamic point cloud without the "lossy" compression can be as high as 4 Gb/s, and also, the increased buffer size results in lower interactivity and prediction accuracy. In addition, the current MPEG's reference encoder and decoder cannot be run in real time on contemporary hardware [14]. Thus, there is still room for research on network distribution and integration of HTC.

Conclusion

Holograms are emerging as the media of the future, with many associated future networked applications. Digitized holograms are, in essence, composed of large arrays of views, each containing closely interrelated moving images. Many holographic applications are highly interactive in nature and involve ultra-fast feedback loops. Furthermore, the optimization of holograms itself involves ultra-fast control loops to take into account user interactivity with the hologram. All this imposes large demands and significant challenges to future networking technology. These include not only the need for unprecedented bandwidth and ultra-low delay, but also high-precision synchronization and the ability to support coordinated service levels across massive bundles of interrelated streams. There is a need for intelligent network and end-user schemes focused on providing the greatest holographic quality for those areas with which the user is most likely to interact vs. other parts whose transmission and rendering may in effect be "wasted." All this provides a rich treasure trove of research opportunities for years to come.

Acknowledgments

This research is sponsored by Huawei Technologies through cooperation with Network Laboratory, 2012Labs. Maria Torres Vega is funded by the Research Foundation-Flanders.

References

[1] P. A. Kara *et al.*, "Evaluation of the Concept of Dynamic Adaptive Streaming of Light Field Video," *IEEE TBC*, vol. 64, no. 2, 2018, pp. 407–21.
[2] J. Karafin and B. Bevensee, "25-2: On the Support of Light Field and Holographic Video Display Technology," *SID Symp. Dig. Tech. Pap.*, vol. 49, no. 1, 2018, p. 318–21.
[3] S. Schwarz *et al.*, "Emerging MPEG Standards for Point Cloud Compression," *IEEE JETCAS*, vol. 9, no. 1, Mar. 2019, pp. 133–48.
[4] V. Swaminathan, "Are We in the Middle of a Video Streaming Revolution?" *ACM TOMM*, vol. 9, no. 1, Oct. 2013, pp. 40:1–40:6.
[5] A. Aijaz *et al.*, "Realizing the Tactile Internet: Haptic Communications over Next Generation 5G Cellular Networks," *IEEE Wireless Commun.*, vol. 24, no. 2, Apr. 2017, pp. 82–89.
[6] K. Lesniak and C. S. Tucker, "Dynamic Rendering of Remote Indoor Environments Using Real-Time Point Cloud Data," *J. Comp. Info. Sci. Eng.*, vol. 18, no. 3, June 2018.
[7] M. Hosseini, "View-Aware Tile-Based Adaptations in 360 Virtual Reality Video Streaming," *2017 IEEE VR*, 2017, pp. 423–24.

Authorized licensed use limited to: University of Exeter. Downloaded on May 08,2020 at 12:12:29 UTC from IEEE Xplore. Restrictions apply.

[8] R. Mekuria, K. Blom, and P. César, "Design, Implementation, and Evaluation of a Point Cloud Codec for Tele-Immersive Video," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 27, no. 4, 2017, pp. 828–42.
[9] W. Xiang et al., "Big Video Data for Light-Feld-Based 3D Telemedicine," *IEEE Network*, vol. 30, no. 3, May 2016, pp. 30–38.
[10] N. K. Karn, H. Zhang, and F. Jiang, "User-Perceived Quality Aware Adaptive Streaming of 3D Multi-View Video plus Depth over the Internet," *Multimedia Tools Appl.*, vol. 77, no. 17, Sept. 2018, pp. 22,965–83.
[11] J. Park, P. A. Chou, and J.-N. Hwang, "Rate-Utility Optimized Streaming of Volumetric Media for Augmented Reality," *IEEE JETCAS*, vol. 9, 2018, pp. 149–62.
[12] D. You *et al.*, "Fog Computing as an Enabler for Immersive Media: Service Scenarios and Research Opportunities," *IEEE Access*, vol. 7, 2019, pp. 65,797–65,810.
[13] F. Bannour, S. Souihi, and A. Mellouk, "Distributed SDN Control: Survey, Taxonomy, and Challenges," *IEEE Commun. Surveys & Tutorials*, vol. 20, no. 1, 2018, pp. 333–54.
[14] J. van der Hooft *et al.*, "Towards 6DoF HTTP Adaptive Streaming Through Point Cloud Compression," *ACM Multimedia 2019*, Nice, France, 2019.
[15] H. K. Ravuri *et al.*, "An Experimental Evaluation of Flow Setup Latency in Distributed Software Defined Networks," *5th IEEE NetSoft*, Paris, France, June 2019.

Biographies

Alexander Clemm is a Distinguished Engineer in Futurewei's Future Networks and Innovation Group in Santa Clara, California. He has been involved in networking software and management technology throughout his career, most recently in the areas of high-precision networks and future networking services. He has served on the Organizing Committees of many management and network softwarization conferences. He has around 50 publications, 50 issued patents, and several books and RFCs. He holds an M.S. in computer science from Stanford University and a Ph.D. from the University of Munich, Germany.

Maria Torres Vega is a postdoctoral researcher for immersive multimedia delivery at the Department of Information Technology, Ghent University-imec, Belgium. She received her M.Sc. degree in telecom.munication engineering from the Polytechnic University of Madrid, Spain, in 2009 and her Ph.D. from Eindhoven University of Technology, The Netherlands, in 2017. Her research interests include quality of service and QoE in immersive multimedia systems and autonomous networks management.

Hemanth Kumar Ravuri received his B.Sc. degree in electronics and communication engineering from Jawaharlal Nehru Technological University, India, in 2014 and his M.Sc. degree in electrical engineer.ing with emphasis on telecommunication systems from BTH, Karlskrona, Sweden, in 2016. He is currently pursuing his Ph.D. degree in the Department of Information Technology, Ghent University-imec. His research interests include network and service architectures, next generation multimedia delivery, QoS, and QoE.

Tim Wauters received his M.Sc. and Ph.D. degrees in electrotechnical engineering from Ghent University, Belgium, in 2001 and 2007, respectively. He is a postdoctoral fellow with the Department of Information Technology, Ghent University-imec. His work has been published in over 100 scientific publications in international journals and in the proceedings of international conferences. His research interests include network and service architectures, and management solutions for multimedia delivery.

Filip De Turck received his M.Sc. and Ph.D. degrees in electronic engineering from Ghent University in 1997 and 2002, respectively. He is a full professor with Ghent University-imec, where he leads the Network and Service Management Research Group in the Department of Information Technology. His research interests include network and service management, and design of efficient virtualized networks. He is the Chair of the IEEE Technical Committee on Network Operations and Management and Editor-in-Chief of *IEEE TNMS*.

Authorized licensed use limited to: University of Exeter. Downloaded on May 08,2020 at 12:12:29 UTC from IEEE Xplore. Restrictions apply.