Immersive Services over 5G and Beyond Mobile Systems

Zinelaabidine Nadir[⋆], Tarik Taleb[⋆,⊕], Hannu Flinck[⋄], Ouns Bouachir[ξ], and Miloud Bagaa[⋆]
[⋆]*Aalto University, Espoo, Finland* ; [⋄]*Nokia Bell labs, Espoo, Finland*
[ξ]*Zayed University, Dubai, UAE* ; [⊕]*Sejong University, Seoul, Korea*
[⋆]firstname.lastname@aalto.fi; [⋄]hannu.flinck@nokia-bell-labs.com; [ξ]ouns.bouachir@zu.ac.ae

***Abstract*—5G and beyond mobile systems target a plethora of emerging industrial and entertainment verticals that incur extra overhead to the network. These verticals are characterized by vigorous, continuous, and conflicting requirements that make the desired system's mission strenuous and more challenging. These verticals, such as autonomous driving, will accommodate immersive services, including virtual reality/augmented reality (VR/AR) and Holography services. Immersive services, in particular, have strict requirements for latency, throughput, and positioning. This paper discusses VR-based remote services' potential, which occupies an important place among immersive services such as remote surgery, remote space control, and remote driving of Unmanned Aerial Vehicles (UAVs) or cars. Such services require an ultra-low Glass-to-Glass latency to avoid any failure and accidents when remotely controlling devices. We evaluate an immersive remote control service from the end-to-end communication perspectives using different camera devices that stream real-time 360° videos to a VR Head Mounted Device (HMD). The obtained results demonstrate the challenges of such service and the need for more advanced and optimized techniques, devices, and protocols to achieve less than 20 ms of Glass-to-Glass latency.**

***Index Terms*—Immersive services, XR, VR, AR, Holography, UAV, GTG latency, 5G and beyond, MEC, and mobile network.**

I. Introduction

The limitations of conventional interactive systems relate to simple interaction devices (i.e., 2D display, mouse, keyboard, and voice), leading to weak interactivity and relatively poor user experience. Be it in a real environment or a virtual one, an immersive interactive system offers a rich and immersive experience to end users, also involving the users' full interactions with their respective surroundings. Immersive applications define a new breath of interactive services whereby users interact within a 3D world for achieving a common objective (e.g., collaborative learning and virtual tourism). This world could be virtual, real, or mixed.

It is expected that both Virtual Reality (VR) and Augmented Reality (AR) would be used by millions of users for online shopping and in-store. Both technologies will play a significant role in marketing by enabling various users to visualize products in unprecedented settings. They would also allow teachers and students to remotely attend classes using Head Mounted Devices (HMD) and conduct collaborative learning in an immersive fashion. These technologies would also allow medical doctors to make remote surgeries without the need for their mobility to the venue of surgery. These technologies can be also used to remotely control robots in a smart factory setting, preventing physical contacts between workers and life-threatening machines. However, these technologies require a massive amount of data traffic that could stress the network. To be perceived with an acceptable level of quality of experience, they also need to operate within a strict latency budget. Fortunately, 5G and beyond mobile systems come to fill this gap, promising better network connectivity, higher network bandwidth, and lower network latency.

Along with the ongoing advances in consumer-device technologies, immersive services could facilitate the life of human beings and cope with delicate situations, such as the ongoing Covid-19 and during the yearly Atlantic hurricane and tropical cyclones seasons. They should have been adopted earlier for similar conditions. On the other hand, they could also enable other use cases missing from our daily life, such as tele-surgery and remote-driving. Owing to communication technologies, infrastructure, and software, these services remain unable to reach consumers in masses. For instance, Facebook space, a virtual reality social network, allows only three people to connect in a virtual environment. This is not efficient for many essential use cases, such as virtual classrooms. Fortunately, a ray of optimism arises shining the way for immersive applications along with the ongoing advances in cloud computing, multi-access edge computing (MEC), micro-services, and machine learning techniques. The latest achievements in these technologies and practices offer the network the elasticity to cohabit according to the network state and achieve the desired objectives. This paper discusses the different network-relevant challenges facing a global deployment of immersive services. The paper also surveys various relevant works conducted by the research community and standards developing organizations (SDOs). The deployment of immersive services at a global scale hinges on multiple factors, including the underlying protocols used for encoding, streaming, and decoding, the used devices and their configuration settings, and the network technology. This paper showcases the impacts of these factors considering a VR-based remote control (i.e., of Unmanned Aerial Vehicles - UAVs) service as a use case and focusing on one important metric, namely the Glass-to-Glass (GTG) latency. Based on this study, the paper presents different open research directions that would be explored in the future to support a wide deployment of immersive services.

The rest of this paper is organized as follows. Section II gives a brief introduction to the 5G and beyond mobile systems. Section III introduces various immersive services and

applications. Section IV presents the different research works carried out by the research community and the various standards activities made by different SDOs. Section V presents a study case, namely a VR-based remote control service, and shows the performance evaluation of the envisioned scenarios. Finally, the paper concludes in Section VI.

II. 5G & BEYOND MOBILE SYSTEM

Recent years have witnessed tremendous growth of consumer devices. These devices, which empower our daily routine, have become smaller and powerful. Indeed, they come with diverse capabilities, ranging from sensing and display technologies to local computing resources. Furthermore, the advances in tracking and vision technologies (e.g., depth, event, and 360 cameras) have expanded the concept of interaction, involving the user's body and its surrounding. These capabilities are not only present in smartphones, but also in fitness watches, games consoles, cameras, smart TVs, and autonomous cars. From another side, in contrast to the previous generations of mobile communications systems, 5G envisions single-digit latency, higher data rates (i.e., ten times more than 4G), five-nines (i.e., 99,999%) network reliability and availability, as well as more dense network coverage with 10 to 100 times more connected users and devices [1]. As of this year, 5G smartphones entered the market and are expected to be widespread in a few years along with other devices. It is likely that within a few years, billions of these devices will be connected to the Internet, making people more connected with valuable immersive services such as tele-presence and remote-surgery. On the other hand, these services require high throughput varying from 1 Gbps up to 1 Tbps, and ultra-low latency less than 10 ms [1], making current cellular networks and the underlying transport networks unable to deliver immersive services with guaranteed quality of experience.

The expected digital society of 2030 (Fig. 1) will be mainly driven by immersive services; in addition to other emerging services such as smart cities and factory automation [2]. Furthermore, along with the success of social networks in bringing people together, immersive social networking would push this a step further by considering 3D virtual communities and societies. Many startups and companies have already started working towards realizing this vision (e.g., Oculus room, Microsoft AltspaceVR, Facebook space, and Linden labs Sansar). Furthermore, the visual experience is only one factor that contributes to the conscious mental representation of real/virtual. In this regard, "Ishin-Denshin", which means communicating with the mind through the mind[2], is a project that aims to provide happiness through intelligent communications. In what follows, we summarize the expected digital transformation of immersive services and applications that will significantly affect the daily lives of an important portion of the World population in the near future.

- **Immersive Applications:** These applications, such as VR, AR, or holography, immerse users into a 3D world. These applications will truly change how we live and work in many aspects, such as telepresence, virtual class, and telesurgery.
- **Internet of Everything:** Building an interactive and intelligent environment such as smart homes, situational-awareness, and digital twins, requires the provisioning of connectivity to billions of sensors and devices. Thanks to these technologies, such as digital twins, the virtual world would embrace the real world creating new experiences and services never seen before [3], [4].
- **Internet of Skills:** This would enable the delivery of human skills using haptic, visual, and audio technologies such as remote management of robots, remote teaching, remote surgery, and virtual mechanic.
- **Vehicular Communications:** The objective hereby is to expand connectivity everywhere for high-speed devices (i.e., autonomous cars, UAV, and high-speed trains) to ensure the autonomy, safety, and reliability of these devices.

These services have many strict requirements that may go beyond the requirements of 5G and the typical use cases of 5G, namely enhanced mobile broadband (eMBB), ultra-reliable and low latency communications (URLLC), and massive machine-type communications (mMTC). This would then advocate for another generation of mobile communications systems (i.e., 6G)[1]. In this regard, beyond 5G and 6G are expected to further boost the performance of 5G networks by supporting higher throughput and much lower latency to ensure connectivity for every device everywhere [5], [6]. To achieve the desired objectives and respect the required KPIs, the following concepts, techniques and technology enablers will play a pivotal role: i) micro-service based architecture, ii) Zero-touch network and Service Management (ZSM); iii) technology enablers such as Software Defined Networking (SDN) and Network Function Virtualization (NFV); and, last but not least, iv) Artificial Intelligence (AI) and Machine Learning (ML) techniques.

- **Micro-services based architecture:** The recent adaptation of the micro-services concept in telecom has created a new concept called a cloud-native environment (CNE). CNE offers the network elasticity, softwarization, and customizability that shall ease the deployment of immerse services and applications. Indeed, leveraging this concept, it is possible to horizontally or vertically scale the virtual resources running immersive applications so as to enhance the service responsiveness and ultimately the perceived QoE. Here, it is worth noting that in contrast to the vertical scaling that may require service disturbance due to the changes in the resources (i.e., CPU and RAM), the horizontal scaling happens fast and without any service disruption thanks to the simultaneous launching of multiple replicates. Moreover, micro-services are characterized by their "fine granularity" nature that facilitates their deployment in the vicinity of their respective end-users.
- **Network Analytics and Automation:** The use of the

[1]NTT DOCOMO 6G White Paper. https://www.nttdocomo.co.jp/english/corporate/technology/whitepaper_6g/; accessed on 27-05-2021.

[2]http://cscn2015.ieee-cscn.org/Ishin-Denshin-Intelligent-Communications-Shigeyuki%20Akiba-.pdf

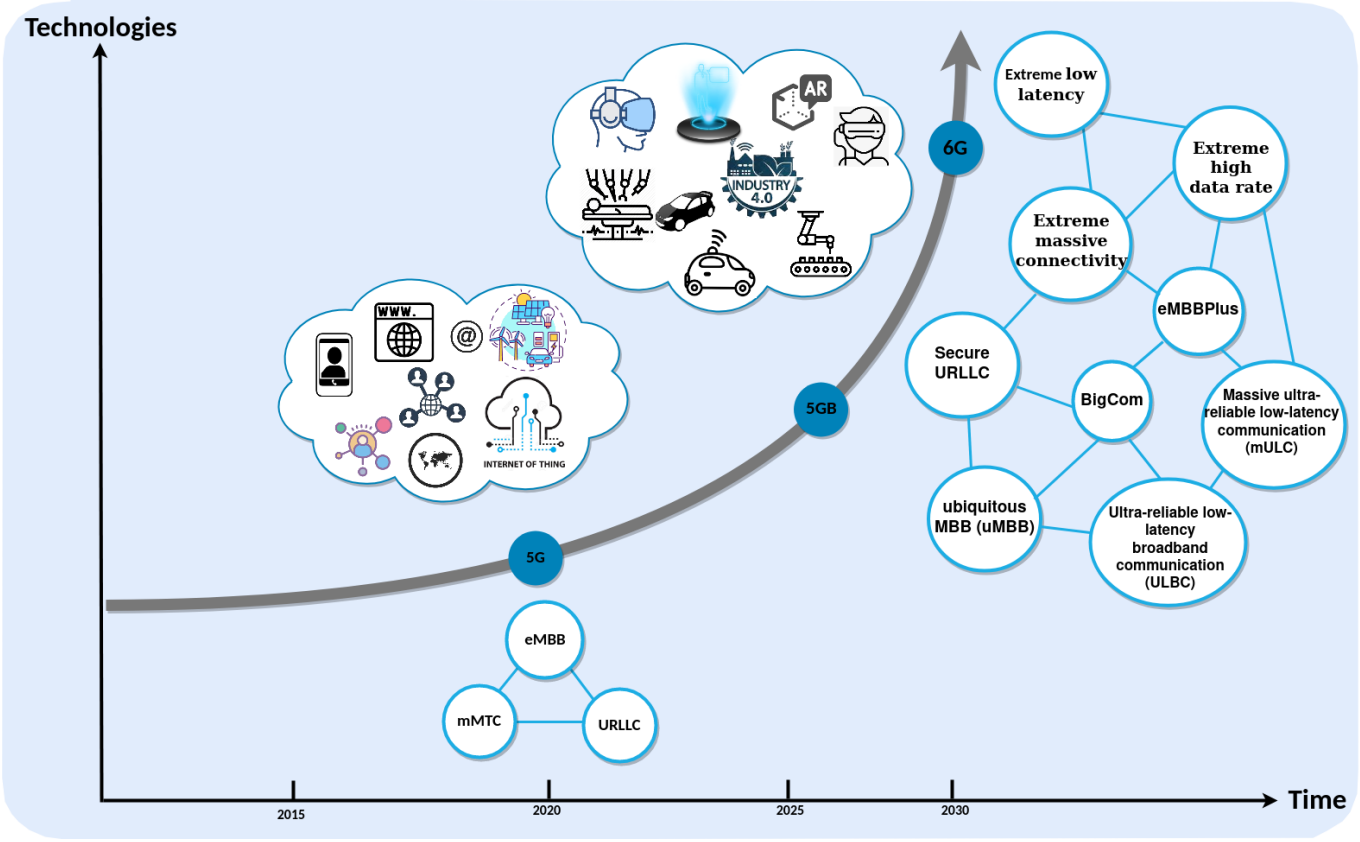


Fig. 1: Evolution towards Network 2030 and the expected services.

ZSM concept enables the management of various immersive services in an autonomous and harmonized way. Similar to the control system, the ZSM concept enables the collection of different monitoring information, and then accordingly adapts both the network and service to meet the desired KPIs with a reduced cost. This information includes computational and network resources, user/device mobility, traffic, QoS, and perceived QoE. To ensure smooth and intelligent closed-loop automation, various AI and machine learning (ML) techniques would be exploited, including supervised, unsupervised ML, deep learning, and deep reinforcement learning.

- **Artificial Intelligence (AI) techniques:** AI techniques (e.g., metaheuristics and ML techniques) have enabled the development of more sophisticated and efficient solutions that have been applied to networking problems, such as service lifecycle management, resource orchestration, and traffic management. Two factors have enabled the success of AI in communications and networking: i) the network (re)programmability leveraging SDN technology; ii) the diverse data generated by various ubiquitous devices from sensors, actuators, IoT devices and smartphones to autonomous cars and smart industry equipment [7].

III. Immersive Services

Immersive systems, with limited capabilities, have been available since a while. However, recent advances in technologies used for tracking, vision sensing, and immersive displays (e.g., head-mounted devices, holographic displays) have come in rescue providing a new whole dimension for immersive applications. In these immersive applications, interactions take place in a 3D-generated world, or a mixed environment where 3D objects are projected on a real-world using HMD devices or smartphones. Immersive technologies are categorized into three types of technologies: Virtual Reality (VR) applications, Holography, and Augmented Reality (AR). In the VR environment, as being separated from the real world, interactions take place within a 3D virtual world. The VR applications require the ability to track the user body movements and then accordingly adapt the projection in the 3D virtual world. On the other hand, in AR applications, interactions take place in a mixed environment whereby 3D objects are projected on a real-world using HMD devices or smartphones, adding more processing delays to merge the two environments (i.e., real and virtual). In contrast to VR, enabling the interactions with the real world in AR is a time-consuming process and requires more advanced tracking and vision sensors. Unlike VR and AR, 3D objects in Holography are projected in the real world. Thus, a user does not have to wear any HMD.

Despite the advances in 3D technologies, immersive services are still not ready to hit the market. These services may have different network requirements in terms of latency and bandwidth from one hand, and computing and storage resources from the other side. From cloud-based VR gaming to telepresence and remote surgery, they are considered to be the killer application for beyond 5G and 6G systems, particularly due to their strict latency/bandwidth requirements. In the remainder of this section, we will first present the

ongoing projects relevant to immersive services as well as some of their use-cases provided in the literature. We then introduce their network and computational requirements.

A. Ongoing Projects and Use cases

The emergence of immersive technologies has opened the door for a wide variety of potential use cases serving many vital verticals. Even though these immersive technologies provide all an immersive and collaborative environment, their applications differ from each other and fall into two categories: i) entertainment and education, such as cloud gaming and medical training; and ii) critical mission services, such as remote surgery, remote command and control of UAVs, and remote driving. A wide deployment of these immersive services would come with different challenges and would necessitate different requirements that are essential to ensure service reliability and continuity. Critical mission services, which may be life-threatening, impose strict requirements in terms of ultra-low latency, high bandwidth, and high reliability. On the other hand, entertainment and education services (e.g., cloud gaming, social tourism and virtual classes) may have less demanding requirements and may tolerate some lag in the latency. In the following, we list up some immersive services.

- **Holographic Telepresence:** This is a game-changer for immersive services in many applications, such as meetings and concerts. It consists of capturing, in real-time, real objects, compress them and send them to remote locations where they will be projected using laser beams or HMD.
- **Health care:** Many applications that fall under this use case, could be for i) critical mission services such as remote surgery or for ii) educational purposes, such as VR Medical training (e.g., OramaVR[3]). For instance, OramaVR is a Virtual reality application that helps orthopedic surgery residents to acquire the needed skills to perform safe surgery.
- **Automotive industry**: There are a variety of immersive applications for the automotive industry. For example, there are many VR relevant applications, such as "Audi VR experience" as a virtual showroom and "ZeroLight VR" used by Toyoya to launch their new car models. VR-based remote command and control of UAVs or remote driving is a critical application whereby any failure could result in serious human fatalities or damages.
- **Social tourism:** Using mixed reality, the immersive tourism provides the experience of remotely visiting a location. A digital twin of the real visited location is presented in a VR-based scene whereby locals act as guides using AR. Indeed, a set of cameras and sensors are deployed in the site in order to gather real time data that will be provided to all participants (i.e., remote and local users). This application allows the participants to interact with each other since remote users appear as AR objects on the AR HMD of local users and vise versa. Social tourism has several potential requirements: i) the system should support scenarios with high density (at least 0.2 node per sq.m), ii) the end-to-end latency should be less than 10ms, and iii) the positioning precision of local users should be highly accurate (i.e., less than 0.5meter error and refreshed within less than 100ms).
- **Cloud gaming:** One of the most popular immersive activities is the entertainment applications, such as CloudGaming. These applications require heavy resources to reach a high level of the immersive experience. This level is based on the interactivity between the various players participating in the same game, usually demanding more delay-sensitivity and a high refresh rate.

[3]https://oramavr.com/

B. Network Requirements

Immersive services depend heavily on the processing of 3D data, regardless whether it is a 3D data generated or mapped in real-time from/to the real-world. While the former requires high bandwidth in downlinks (i.e., cloud gaming), the latter requires high bandwidth in both uplinks and downlinks (i.e., in case of Telepresence). Effectively, holography services, such as telepresence, are major bandwidth consumers since they require up to 30 Gbps for both directions. However, the bandwidth requirement of AR/VR applications differs from one use case to another and depends on where the 3D data are processed and rendered (i.e., whether the vision and rendering engines are deployed locally at the device or in the cloud). For such applications, the bandwidth requirements go from 100 Mbps up to a few Gbps [4].

The Motion-to-Photon (MTP) or Glass-to-Glass (GTG) latency is the most crucial factor in making a service experience immersive or not. The GTG latency is defined as the time needed for a user action to be fully projected on the other users' devices. In other words, this delay comprises the scanning delay, the network latency, and the projection delay (i.e., refresh rate and response time of the display technology). For the most immersive service, the required GTG latency varies between $10 - 200ms$, intuitively depending on the target use case [8].

C. Computing Resources

Immersing users into a 3D environment presents a new experience, but comes at a huge price in terms of computation resources. Effectively, such experience largely depends on computation-intensive tasks, represented mainly by rendering and vision engines performed in real-time. Vision engines map data from real-world such as eye and hand tracking, Point Cloud reference, and markers, whereas the rendering engine renders new 3D scenes upon receiving these data. Moreover, encoding is necessary to encode immersive data into a particular streaming format (e.g., Omnidirectional MediA Format (OMAF) and Point Cloud Compression (PCC)). Apart from the eye and hand tracking which relatively require less computing resources, these tasks require massive computing resources and it is worth offloading them to edge cloud

[4]Huawei white paper, PREPARING FOR A CLOUD AR/VR FUTURE. https://www-file.huawei.com/-/media/corporate/pdf/x-lab/cloud_vr_ar_white_paper_en.pdf; accessed on 27.05.2021

to relieve the end devices from the data processing burden and to also reduce the processing delays (i.e., that could be too long on devices with constrained Central Processing Units - CPUs). Furthermore, these tasks will also depend on processing extensive data, i.e., from a few up to tens of Gigabytes. Thus, storage/caching also plays a fundamental role in the overall processing delays.

IV. Enabling Technologies & Standards Activities

In the era of immersive services, the ongoing advances in tracking sensors (position and orientation) and vision sensors, such as depth and event cameras, play an essential role in making the immersive experience more realistic. These sensors are essential to map real environment objects into digital world and vise versa. As mentioned above, such mapping requires significant computing resources for both vision and rendering. This can be currently accommodated by special processors such as Vision Processing Units (VPU) and Graphics Processing Units (GPU). However, despite the fact that VR/AR HMDs come equipped with several of these resources, including many smartphones, there is still a huge gap between the current offerings of immersive systems and the expectations for the end user experience of immersive services. To cope with the different limitations, there is need for improvements at different levels, including communications networks, video coders and consumer-end devices.

A. Smart Edge

Highly immersive services usually require significant real-time processing power, memory and storage. These capabilities are usually not available at commercially available off-the-shelf (COTS) consumer devices, often characterized by constrained processing and power resources. Partly or fully hosting computation-intensive operations of immersive services (e.g., rendering and vision) at the edge cloud is a viable solution to cope with the limitations of end consumer devices. The success of such offloading operation largely hinges on, among others, the network capability as well as on the latency to the edge cloud nodes. Regarding the latter, with an efficient planning and deployment of edge cloud and extreme-edge cloud supported by an efficient edge cloud resource orchestration system, it becomes possible to allocate, on demand and dynamically, computing resources (e.g., CPU, GPU, Storage) near to the end users to ensure immersive streaming with low latency. Regarding the network, no matter how speedy it is, it is always limited in the rate it can allocate in the uplinks and downlinks [9]. It is therefore vital to ensure an optimal usage of the available network throughput, particularly in the context of real-time streaming of either spherical or volumetric videos whereby the needed throughput for an immersive service ranges from hundreds Mbps to a few Gbps. In this vein, different solutions can be leveraged. For example, by leveraging AI techniques to predict the zone gazed by the fovea of a user, or FoV, the throughput needed for the streaming of a spherical or volumetric video could be reduced using foveated rendering which consists of rendering FoV with high quality and rendering the peripheral zones with less quality [10]. All in all, leveraging edge cloud and relieving end devices from computation-intensive tasks shall yield promising opportunities for both consumer-device manufacturers and immersive application developers to deliver thin and cost-efficient devices capable of supporting highly immersive applications (i.e., 16K service, high refresh rates of 120 and 240 Frame Per Second - FPS).

B. Efficient media encoders

Efficient video encoders are recognized by their high compression rates and low complexity. H265 (HEVC) and H264 (AVC) are the most widely-used encoders. Compared to the latter, H265 achieves up to 50% bitrate saving without compromising the video quality [5]. In terms of resolution, H265 provides a maximum 8k of resolution and 120 FPS whilst H264 provides only a maximum 4k resolution and 60 FPS. Unfortunately, H265 is yet not fully supported by many devices. For immersive applications, these encoders cannot be applied directly to represent spherical videos or volumeteric data [11]. Therefore, MPEG introduces Omnidirectional MediA Format (OMAF) for projecting spherical (360 degrees) videos into 2D videos. It is also planning to develop PCC (Point Cloud Compression) [12] for volumeteric data. Volumetric data are 3D objects that consist of a collection of points called point clouds. OMAF supports only 3-DoF (Degree of Freedom). This standard also specifies the file/segments encapsulation (i.e., File format and DASH extensions). OMAF supports both mono and stereo omnidirectional videos and specifies two types of projection: EquiRectangular Projection (ERP) or CubeMap Projection (CMP). Furthermore, MPEG envisions to develop other formats to support 3DoF+ and 6DoF.

C. Low Latency Scalable Throughput communications

Immersive services are demanding in terms of throughput and end-to-end latency. They require highly-performing communication protocols, such as WebRTC, which is an essential real-time streaming protocol for ultra-low latency immersive video streaming. Such communications should be resilient to any event (e.g., network congestion or path loss) in the network. Recent developments in congestion control mechanisms have shown a shift from loss-based congestion management algorithms (e.g., Reno and CUBIC) to delay-, bandwidth-, and/or loss-based congestion management algorithms (e.g., Google's BBR and Facebook's COPA). Moreover, the Quick UDP Internet Connection (Quic) protocol has been suggested by Google to be an alternative to TCP. Quic provides more secure and low latency mechanisms compared to TCP. In contrast, Quic leverages IDs to identify the connections, which allows the migration between different IPs and traversing different middleboxes. Furthermore, while the header is plaintext, the payload is encrypted to ensure security and prevent middleboxes modification. Another key factor for efficient streaming is the adaptive streaming feature to handle the

[5]https://newsletter.fraunhofer.de/-viewonline2/17386/465/11/14SHcBTt/V44RELLZBp/1; accessed on 27.05.2021.

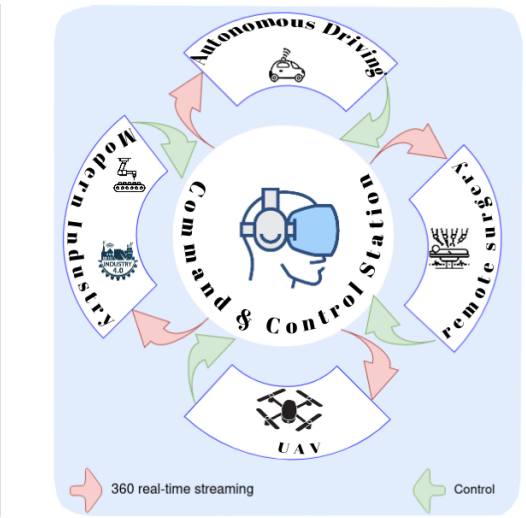


Fig. 2: VR-based remote control services.

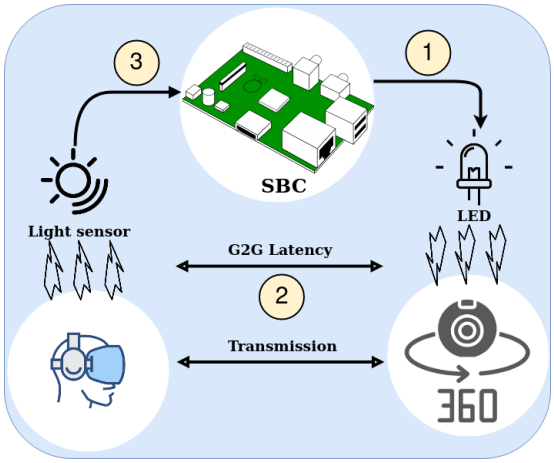


Fig. 3: GTG latency measurement method.

frequent changes in the delay and bandwidth (e.g., foveated rendering and scalable streaming). In addition, semantic routing and qualitative communications as conceptualized in the New IP architecture can also help in coping with the demands of immersive services in terms of throughput and end-to-end latency [13].

D. In-proximity communications

The idea of collaborative users is a promising solution to enhance the streaming performance. Indeed, users in proximity could download the same 3D content by catching various parts from the cloud and getting the missing ones from their neighbors' caches, similar in spirit to the Neighbors Buffering Based VoD approach [14]. This solution helps to reduce the individual bandwidth required for each user and optimize the otherwise-affected quality.

Unfortunately, this approach of collaborative users-based streaming and the split process may not be suitable for VR applications. In such applications, each user may receive a totally different stream even if the users are involved in the same virtual environment. Effectively, these streams depend on the users' positions in the virtual environment. As for AR applications, the users may share the same real environment when they are in the same location. The collaboration between the respective AR devices becomes then possible on many aspects such as sensing the real world, requesting the same 3D objects either from each others or collaboratively from the cloud.

E. Other activities

The first version of OpenXR was released to the public in mid 2019. It allows unifying access to the various VR/AR platforms and devices using two components: an API , and a device plugin interface. The API enables the applications to run on any system that exposes that API while the device plugin Interface allows the manufacturers of the devices (e.g., Oculus) to integrate their drivers into OpenXR.

V. Case Study: VR-based Remote Control Services

A. Architecture

In this section, we consider, as an example, an immersive service whereby VR HMDs are used to remotely control devices (e.g., robots and UAVs as in Fig. 2). Thanks to VR/AR and Holography services, end-users are able to remotely manage different industrial verticals in a smooth and more user-friendly manner. The remote devices are equipped with a 360 camera. The devices are controlled by a user using hand motions (i.e., using depth cameras) and HMD controllers. The user controls remote devices based on receiving a 360-degree video of the remote location. Both the streaming of 360-degree video and the signaling to control remote devices require a very low latency and high reliability. While signaling to control the remote devices requires low latency communication, streaming the 360-degree video requires both high bandwidth and low latency. In this paper, only the latter is analyzed.

End devices, namely VR HMD and remote devices, communicate using real time protocols such as RTP or WebRTC. Both protocols provide very low latency. RTP protocols were mainly designed for broadcast scenarios such as live concerts (i.e., VR- or holographic-base) where the live concert is streamed to thousands or millions of users. To relieve the source from streaming the content to all users, an RTP server is used at the cloud to serve many users with the same content. Effectively, the remote device streams the 360 content to the RTP server. The RTP server, in return, re-streams the content to all interested users. A scenario such as VR-based remote controlling is more challenging than live concerts scenarios. Indeed, it requires a direct communication between the HMD and the remote devices in order to have a low end-to-end latency.

B. Results

Fig. 4 shows the performance of the setup in terms of GTG latency, measured between Oculus Quest and the 360-degree camera. Fig. 3 illustrates how the GTG latency is measured in practice. A LED, installed next to the 360 camera, is turned on at a time instant t_1. The light is then detected by a light sensor,

linked to HMD, at a time instant t_2. Both time instants are recorded at a Single Board Computer (SBC). The difference between the two time instants, $(t_2 - t_1)$, represents the GTG latency. In the conducted experiments, two cameras were used: i) Vuze XR and ii) Insta 360 Pro. Both cameras stream at the same frame rate, 30 FPS using RTMP protocol. As illustrated in Fig. 4, Insta Pro outperforms Vuze XR by almost 700ms. This is due to the following two reasons. First, for live streaming, Vuze XR should be connected to a smartphone through WiFi direct which adds more delays. Second, Insta Pro has its own built-in RTMP server, which significantly reduces the GTG latency.

Fig. 5.a plots the performance of the RTMP protocol of Insta 360 Pro in terms of GTG as a function of the Display Refresh Rate (DRR). We observe that the GTG decreases when DDR is increased. This change becomes smaller every time the DDR is increased. This is due to the fact that for a display with a refresh rate of 50Hz, a decoded frame will be displayed with a maximum delay of 20ms. Similarly, with 144Hz and 240Hz refresh rates, a decoded frame will be displayed with a maximum delay of 7ms and 4ms, respectively. This means that increasing the DRR from 144 to 244 will only reduce the GTG by approximately 3ms whilst, on the other hand, it will consume more energy. Overall, current display technologies are characterized by their refresh rates and their response time, and current HMD displays come with maximum 120Hz refresh rate.

Current 360-degree cameras support streaming only at 30 FPS. However, to understand how the streaming frame rate (SFR) affects the GTG latency, a web camera is used instead of the 360-degree camera. This camera is able to stream up to 120 FPS. We also used a display with 244 refresh rate connected to a PC with a CPU i5-9400f and a GPU GTX 1080. The performance of the streaming protocol RTMP is evaluated by varying i) the Display Refresh Rate (DRR) and ii) the Streaming Frame Rate (SFR). To avoid any potential impact of the network on the G2G latency, the web camera is set to stream at a low resolution, 640×480.

Fig. 5.b shows the performance of RTMP in terms of GTG latency as a function of DRR, whereby DRR is varied between 50Hz and 244Hz. This figure shows that an increase in DRR has a slight effect on the GTG latency. As depicted in this figure, for any streaming rate, increasing the DRR from 50 to 244Hz will only reduce the GTG latency by approximately 40ms. This could be explained as follows. At 50 FPS, a decoded frame is displayed on the display with a maximum GTG delay of 460ms. Similarly, at 244Hz, a decoded frame is displayed with a maximum GTG delay of 420ms. That is, increasing DDR will only reduce the maximum GTG delay for displaying a decoded frame by merely 40ms. This intuitively comes at the cost of much higher energy consumption.

Fig. 5.c shows the evaluation of DRR in terms of the GTG latency as a function of SFR whereby SFR is varied between 20 FPS and 120 FPS. From the figure, we observe that an increase in SFR reduces significantly the GTG latency. For example, increasing SFR from 20 FPS to 30 FPS, and from 30 FPS to 60 FPS decreases significantly the GTG latency by approximately 100 ms. This change becomes smaller for higher values of SFRs. Indeed, the GTG latency decreases by 30ms and 15ms when SFR is increased from 60 FPS to 90 FPS, and from 90 FPS to 120 FPS, respectively. This is principally due to the fact that any processing of the video from the source (i.e., device, public/edge cloud) to the destination could result in a series of real time actions (i.e., decoding, processing, encoding and streaming) whereby buffering is needed for decoding streams. If SFR is too small (e.g., 20 FPS or 30 FPS), this could result in higher delays each time a video is processed. However, if SFR is high (e.g., 120 FPS), this could result in smaller delays compared to lower SFR scenarios. However, immersive videos (i.e., 360 degree and volumetric) require much higher bandwidth and intensive computing for streaming at higher frames rate. Increasing SFR to 120 FPS, or 240 FPS, will therefore cause bottlenecks in both the network and the computation resources.

C. Challenges

Fig. 6 demonstrates a detailed analysis of the GTG latency[6]. The processing time of immersive data depends heavily on both the cameras' refresh rate and SFR. If the camera refresh rate is 30 FPS, the processing time at the sender counts at least for 66ms without any processing delays (e.g., stitching and encoding). Similarly, any processing at the edge could cost the service at least 33ms. On the other hand, any delayed arrival of packets would certainly affect the decoding time of the relevant frame at both edge cloud and end device. Increasing SFR reduces significantly the GTG latency. However, this requires intensive computation to cope with the increased throughput. Moreover, in the case of holography, in particular, increasing SFR could worsen the streaming experience: significant network resources, beyond the capacity of current networks, would be needed, and in the absence of such network resources, several packets will be lost and delayed arrivals of successfully-transmitted ones may be experienced, ultimately impacting the GTG latency.

VI. Conclusion and future works

Highly interactive immersive services impose strict challenges on 5G and beyond mobile systems. For the most immersive services, such as telepresence and tele-surgery, ultra low Glass-to-Glass latency is a necessity to achieve the desired service quality. Despite the fact that users are only able to capture changes above 13 ms [15], pushing immersive data, such as volumetric videos, every millisecond, or a frame per millisecond, takes the service into the ideal scenario. However, this comes at a price of becoming compute- and bandwidth-intensive. Unfortunately, current mobile systems are still not able to handle such increasing demand.

In this paper, we attempted shedding lights on challenges, particularly relevant to networking, that hamper a wide deployment of immersive services. The paper also introduced

[6]The size of each box does not reflect the processing time of the respective task.

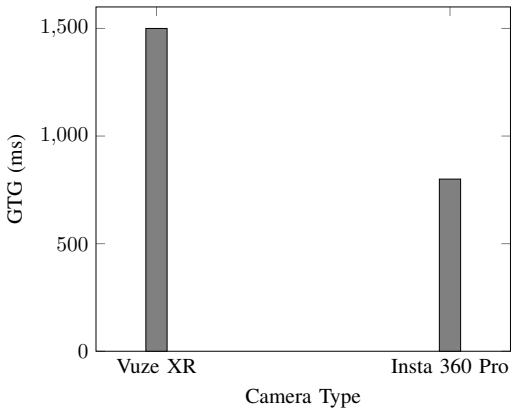
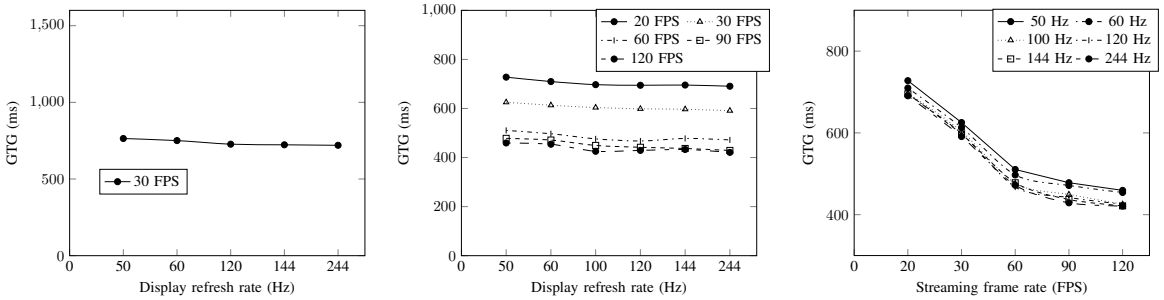


Fig. 4: Performance of the setup using Oculus Quest.



(a) Performance evaluation of RTMP in case of Insta 360 PRO.

(b) Performance evaluation of RTMP for varying DRR values.

(c) Performance evaluation of RTMP for varying FPS values.

Fig. 5: Performance analysis of RTMP.

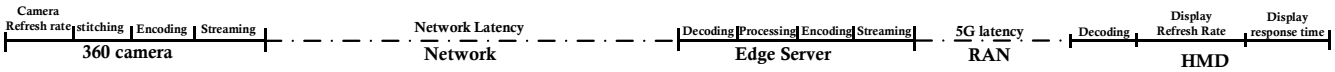


Fig. 6: Detailed analysis of the GTG.

different open research directions that are yet to be explored to facilitate the provisioning of immersive services at a global scale. The paper took a VR-based remote control (i.e., of robots) service as a use case and demonstrated how different factors (e.g., adopted technology, display refresh rate, an streaming frame rate) impact an important KPI of immersive services, namely the Glass-to-Glass latency. Based on real-life implementation, different interesting research observations were made. The provisioning of immersive services, leveraging 5G and beyond, cloud edge, and AI, is a highly promising area of research. The work presented herein represents an initial seed of the authors' contributions to this field, and more research work will be conducted on this area in the future.

ACKNOWLEDGMENT

This research work is partially supported by the European Union's Horizon 2020 research and innovation program under the CHARITY project with grant agreement No. 101016509. It is also partially funded by the Academy of Finland Project 6Genesis under grant agreements No. 318927.

REFERENCES

[1] R. Li. Towards a new internet for the year 2030 and beyond. In *Proc.3rd Annu. ITU IMT-2020/5G Workshop Demo Day.*, 2018.

[2] FG-NET-2030 (ITU-T). Network 2030, a blueprint of technology, applications and market drivers towards the year 2030 and beyond. https://www.itu.int/en/ITU-T/focusgroups/net2030/Documents/White_Paper.pdf.

[3] H. Elayan, M. Aloqaily, and M. Guizani. Digital twin for intelligent context-aware iot healthcare systems. *IEEE Internet of Things Journal*, pages 1–1, 2021.

[4] O. E. Marai, T. Taleb, and J. Song. Roads infrastructure digital twin: A step toward smarter cities realization. *IEEE Network*, pages 1–8, 2020.

[5] K. Samdanis and T. Taleb. The road beyond 5g: A vision and insight of the key technologies. *IEEE Network*, 34(2):135–141, 2020.

[6] T. Taleb et al. *White paper on 6G networking*. 6G Research Visions. University of Oulu, Finland, June 2020.

[7] O. Wahab, A. Mourad, H Otrok, and T. Taleb. Federated machine learning: Survey, multi-level classification, desirable criteria and future directions in communication and networking systems. *IEEE Communications Surveys Tutorials*, 02 2021.

[8] 3GPP. Study on Network Controlled Interactive Services (Release 17). Technical Report TR 22.842, December 2019. Version 17.2.0.

[9] Hernani D. Chantre and Nelson Luis Saldanha da Fonseca. The location problem for the provisioning of protected slices in nfv-based mec infrastructure. *IEEE Journal on Selected Areas in Communications*, 38(7):1505–1514, 2020.

[10] T. Kämäräinen, M. Siekkinen, J. Eerikäinen, and A. Ylä-Jääski. CloudVR: Cloud Accelerated Interactive Mobile Virtual Reality. In *Proceedings of the 26th ACM international conference on Multimedia*, MM '18, pages 1181–1189, Seoul, Republic of Korea, October 2018. Association for Computing Machinery.

[11] 3GPP. Virtual Reality (VR) profiles for streaming applications. Technical Specification Group Services and System Aspects 26.118, 3rd Generation Partnership Project (3GPP), 03 2020. Version 16.0.2.

[12] S. Schwar et al. Emerging MPEG Standards for Point Cloud Compres-

sion. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 9(1):133–148, 2019.
[13] R. Li et al. New ip: A data packet framework to evolve the internet. *in Proc. IEEE HPRS 2020, May 2020*.
[14] T. Taleb, N. Kato, and Y. Nemoto. Neighbors-buffering-based video-on-demand architecture. *Signal Processing: Image Communication*, 18(7):515–526.
[15] Mary C. Potter, Brad Wyble, Carl Erick Hagmann, and Emily S. McCourt. Detecting meaning in RSVP at 13 ms per picture. 76(2):270–279.

Zinelaabidine Nadir received the M.Sc degree in computer science from Laghouat University in 2012. He is a doctoral student at the MOSA!C Lab, Aalto University, Finland. His current research focuses on Immersive Services, 5G and B5G.

Tarik Taleb [S'05, M'05, SM'10] received the B.E. degree (with distinction) in information engineering in 2001, and the M.Sc. and Ph.D. degrees in information sciences from Tohoku University, Sendai, Japan, in 2003, and 2005, respectively. He is currently a Professor with the School of Electrical Engineering, Aalto University, Espoo, Finland. He is the founder and the director of the MOSA!C Lab (www.mosaic-lab.org).

Hannu Flinck is a Senior Research Project Manager with Nokia Standards, Espoo, Finland. He holds an M.Sc. and Lic. of Tech. degree in Computer Science and Communication Systems from the Helsinki University of Technology (TKK, current Aalto Uni). He joined Nokia Telecommunications in 1987 and worked since then in various research including Nokia Research Center and recently in Nokia Bell Labs related to networking architecture and communication protocols.

Ouns Bouachir (M'18) is an assistant professor of computer engineering in the college of Technological Innovation at Zayed University, UAE. She has a PhD degree in computer engineering from the University Paul Sabatier, France. She also holds an engineering degree in Telecommunications from the Higher School of Communications, Tunisia. Prior to joining Zayed University, she worked as an assistant professor and post-doc researcher at Canadian University Dubai, UAE.

Dr. Bagaa Miloud received his Engineer's, Master's, and Ph.D. degrees from the University of Science and Technology Houari Boumediene, Algeria, in 2005, 2008, and 2014, respectively. From 2015 to 2016, he was a postdoctoral researcher at the Norwegian University of Science and Technology, Norway. He was a postdoc researcher at Aalto University from 2016 to 2019, then a senior researcher from 2019 to October 2020. Currently, he is a senior cloud specialist at IT Center For Science LTD.