Extremely-interactive and low latency services in 5G and beyond mobile systems

Tarik Taleb*, Zinelaabidine Nadir*,Hannu Flinck † and JaeSeung Song‡

*Aalto University, Espoo, Finland
†Nokia Bell Labs, Espoo, Finland
‡Sejong University, Seoul, South Korea

***Abstract*—This paper provides an overview of extremely-interactive and low latency immersive services as well as on the relevant industry and standardization activities. Immersive services immerse a viewer or the viewed digital objects into an environment that is either real, virtual or a mixture of both. The applications are accordingly named as Virtual Reality, Augmented Reality, Extended Reality, and Holography applications. These applications benefit from the ongoing advances in user interfaces, computing technologies, and networking technologies. Such applications are expected to generate most of the traffic in the next generation networks, particularly 6G networks. In this paper, the main relevant use cases are introduced along with their respective requirements. The paper also provides insights on the relevant architectures and solutions, and highlights some research challenges and directions.**

***Index Terms*—XR, URLLC, Virtual Reality, Augmented Reality, Holography, and 6G networks.**

I. Introduction

Extremely-interactive and low latency services, or immersive systems, have been around alongside other technologies. However, the recent advances in tracking and vision sensors, and immersive displays' technologies have boosted up a new level of immersive experience. The term immersive refers to the interaction of a user with his/her surrounding, which subsequently gives the feeling of being completely inserted and involved within the local surrounding. The surrounding environment could be completely virtual or real. The former is known as Virtual Reality (VR), whereby a user is immersed into a 3D virtual environment that totally hides the real environment. As for the latter, known as Augmented Reality (AR), the real environment coexists alongside with virtual objects. The term extended Reality (XR) refers to the existence of both realities together in one scenario. It is also called Mixed Reality (MR), denoting the blending of real and digital environments. Holography is another emerging service that rather than immerses a user into a new environment, teleports a digital copy of the user or a real object into a remote environment (real or virtual), creating another dimension for immersive services.

Even with the recent advances in the immersive technologies, the numerous envisioned and already experimented immersive applications are not yet ready for the mass markets. The current experimental applications have still very limited usability and functionality, compared to what is expected to come. The current limitations are that the applications and services have very strict latency and bandwidth requirements for the connectivity as well as high demand for computing and storage resources, limiting their applicability mostly to standalone and costly installations. However, from VR Cloud gaming to telepresence and remote surgery, the immersive applications are anticipated to be dominant applications of the next-generation networks in terms of their popularity and the traffic volumes they will generate. It would be safe to anticipate that the XR-applications will be the mainstream applications by 2030 (Fig. 1, under the category of both Very large Volume –VLV – and Tiny Instant Communications – TIC), when both the computing and networking capabilities are expected to reach the envisioned 6G network performance levels along with suitable costs for mass markets.

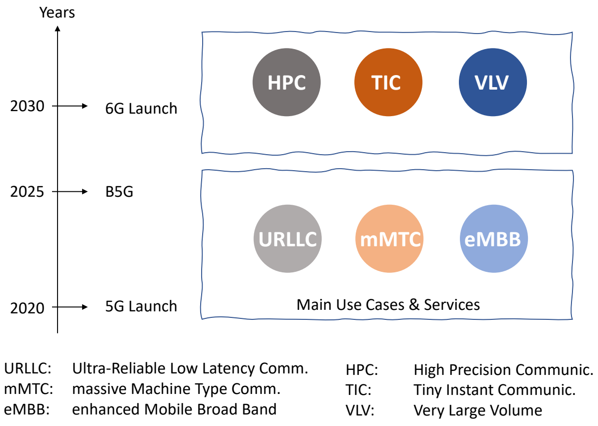
Years
2030
6G Launch
2025
B5G
2020
5G Launch
HPC
TIC
VLV
URLLC
mMTC
eMBB
Main Use Cases & Services
URLLC: Ultra-Reliable Low Latency Comm.
mMTC: massive Machine Type Comm.
eMBB: enhanced Mobile Broad Band
HPC: High Precision Communic.
TIC: Tiny Instant Communic.
VLV: Very Large Volume

Fig. 1: Evolution toward 2030 networks and the expected supported use cases & services.

Motivated by the huge potential of VR, AR and XR-applications, many Standards Development Organizations, (SDOs), are addressing their respective parts of the enabling technologies to ensure interoperability between the various components and building blocks that are needed to work together to deliver an immersive experience. For instance, the IEEE P2048 working group is developing 12 standards for VR and AR. ETSI has defined the Augmented Reality Framework, and 3GPP has published studies and specifications on 5G media streaming dealing with AR and VR. As for immersive data representation, the Moving Picture Experts Group The Motion Picture Expert Group (MPEG) working group of ISO/IEC is developing metadata for immersive video together with media formats and video/audio codecs for XR. World Wide Web Consortium (W3C) hosts Immersive Web Working Group that is enabling XR-capability in browser to interact with devices and sensors through open Web APIs and extensions. Internet Engineering Task Force (IETF) is developing streaming protocols that provide low latency and high bandwidth to serve immersive applications from remote servers across the Internet.

This paper provides some insights on the recent standards supporting immersive services in mobile networks. Section II provides an overview on immersive services and their enabling technologies. Section III dives into 5G and beyond mobile systems, XR-use cases, and their potential requirements. Section IV covers the technical challenges relevant to the support of XR services over 5G and beyond systems. It also highlights some solutions and defines some future research directions. Finally, concluding remarks are drawn in Section V.

II. Immersive services: An overview

Immersive services rely on bringing both the real and virtual worlds together. In general, this comprises capturing elements from the real world and projecting them on the virtual world and vice-versa. The advances in sensing and tracking devices have opened the door to numerous ”previously-unthinkable” possibilities, e.g., to even create a world parallel to the real world. However, such merging of scenes and digital objects requires intensive computing resources and a powerful network infrastructure for the most immersive experience. Effectively, the end-devices should be able to capture, in real-time, data and stream it at speeds ranging from few Kbps (e.g., users' movements) up to Tbps (e.g., holograms and digital twins). The advances in display technologies, such as in Head Mounted Device (HMD) and holographic displays, have enabled such vision, whereby real objects could be inserted into a virtual world and, vice-versa, virtual objects could be projected in the real world. For example, instead of creating an avatar for an individual player in a game, the real body of the player could be captured and inserted in the game to represent the real person. Similarly, computer-generated objects could be also projected on the real world as holograms.

There are different levels at which both realities (i.e., actual and virtual) could be brought together:

- The very basic level of immersive services, namely VR, consists of reacting to only the user head and hand motions, and the body movements within a defined area and project them on the virtual world. The user movement is referred to as the Degree of Freedom (DoF). Three levels of DoF are defined: 3DoF, 3DoF+, and 6DoF. 3DoF usually refers to pitch, yaw and roll movements of the head. 3DoF+ adds movement of the head towards six directions along three directional axes (i.e., front, back, left, etc.). 6DoF adds rotational degrees of freedom to these three directional axes.
- The second level of immersive services, AR, consists of projecting virtual objects onto the real world. This requires processing data, called markers, from the real world to extract information to assist the system in generating virtual objects, referred to as computer-generated objects, and project them on the real world.
- The third and last level of immersive services, Holography, consists of capturing an object as it is in the real world and projecting it directly either in the real world or a virtual one. This could only be achieved using point clouds or light field holograms.

It is worth noting that objects, projected onto the real world as holograms, add parallax where the user can interact with them regardless of his/her position [5]. On the other hand, objects projected using VR or AR in HMDs as 2D views (i.e., left and right eyes) require rendering each time a user changes his/her position. Immersive audio is also a subject of many standardization efforts such as IEEE P2048.10 and 3GPP TS 26.254, whereby the objective is to define formats and codecs for immersive audio.

A. Holographic-Type Communication

Holographic-type Communication (HTC) consists of capturing, in realtime, real objects, compress them and send them to remote locations where they will be projected on the real world using holographic displays. A hologram consists of a set of points in space called Points cloud. Points cloud are produced from processing multiple images of the object captured from different angles. Streaming holograms require bandwidth from 2.06 Gbps up to one Tbps [9]. HTC is seen as a key driving technology after AR/VR in the Network 2030 initiative of ITU-T together with multi-sense networks, including haptic communication services [10], [11].

B. Immersive data representation

MPEG is working on multimedia application formats through the specifications of their MPEG-A series. They have also developed mixed reality and augmented reality reference model (MPEG MAR). Omnidirectional Media Application Format (OMAF) is an MPEG format for 360 videos. It covers both equirectangular and cubemap projections. This media type supports only 3-DoF. The standards also specify the file/segments encapsulation (i.e., file format and DASH extensions). OMAF provides additional features, such as overlays and multiple viewpoints, giving additional information or even real-time commercials.

The MPEG-I standardization project develops standards for volumetric video that enables telepresence for a far better experience than 360-degree video wherein a viewer cannot move in the scene. MPEG introduces Point Cloud Compression (PCC) and defines two representations, namely video-PCC [12] and geometry-PCC. V-PCC decomposes Point Clouds

TABLE I: Requirements of immersive services.

Use case	QoS requirements		
	latency	Data rate	Reliability
HTC telepresence	320 ms	2 Gbps - 2 Tbps	99.999%
Remote services	$<$ 5 ms	15 Mbps - 2 Tbps	99.999%
Social tourism	320 ms	15 Mbps - 2 Gbps	99.999%
Cloud gaming	30 - 150 ms	15 Mbps - 500 Mbps	99.99%

into two separate video sequences which capture the texture information and geometry. Traditional codec, such High Efficiency Video Coding (HEVC) or Versatile Video Coding (VVC), are then applied. However, G-PCC decomposes the 3D space into a hierarchical structure of cubes; each point is encoded as an index of the cube it belongs to.

C. Immersive services architecture

If a use case involves many users, each user should be only immersed either in a real or virtual world. For example, in cloud gaming, all users are immersed in a virtual world. Virtual Social tourism where tourists are being remotely introduced to a new site by local people, could represent a scenario where both realities are blended together. ETSI has published the Augmented Reality Framework [8] that provides a functional reference architecture for AR-components, systems and services with the aim of identifying interfaces for interoperability between various building blocks. The framework architecture is based on three layers, namely hardware, software and data layers. The architecture defines where each component should be (i.e., cloud, edge, or user equipment) and whether the component could be offloaded to an edge/cloud. Both VR and holography may have a similar architecture. VR devices are able to render and project spherical scenes with light tracking sensors to capture user movements in 6DoF. VR rendering engine could be offloaded to the edge for maximum performance. AR-HMD devices require more sensing and tracking to extract and map relevant markers from the real world and map them to virtual objects. In HTC, an object is captured from different angles using multiple cameras. This requires rendering and vision engines to render and encode into a volumetric scene.

D. Open Source and Industry Consortia

The Khronos Group is an industry consortium creating specifications and technology for 3D graphics, AR and VR, computer vision and machine learning. It has previously published, among others, popular OpenGL and Vulkan graphics and compute APIs. In order to unify access to AR/VR platforms and devices, Khronos released the first version of OpenXR to the public in mid 2019. OpenXR has two components: an API and a device plugin interface. The API enables the applications to run on any system that exposes that API. The device plugin interface allows to integrate device drivers from different technology vendors (e.g. Oculus) into OpenXR.

III. USE CASES & SERVICE REQUIREMENTS

Every generation of mobile networks was primarily designed for a specific set of target services. The evolution of the mobile networks has progressed from the first generation that targeted mobile voice calls to the current fifth-generation that is designed to enable, among others, Industrial IoT applications, time-sensitive networking, and enhanced mobile broadband. Beyond 5G and 6G will be driven by immersive services, supported by the future advancements in the user interface and sensor technologies as well as availability of edge cloud resources [3]. Immersive services will change the form factors of the end user devices and the way how services will be consumed. However, this implies that the beyond 5G networks should be ready to support the extreme latency sensitivity, and the compute- and bandwidth-intensity of immersive services whilst still provisioning special offerings to other industrial verticals (e.g., Vehicle to Everything - V2X). Table I provides a summary of the QoS requirements stemming from each use case of immersive services and that is in terms of latency, bandwidth, and reliability. Similarly, Fig. 2 illustrates the requirements of immersive services from the cloud perspectives and as a function of their interactivity – the horizontal line shows how much bandwidth each service requires and the vertical line shows how interactive these services are.

A. Holographic Telepresence

Holographic Telepresence, at the top right corner of Fig. 2, is a game-changer for immersive services. It requires from two Gbps up to two Tbps to achieve the true holographic telepresence experience. According to the technical specifications of 3GPP TR 26.923, the recommended requirements of the one-way delay between two devices, to achieve the desired Quality of Experience in a holographic service, are i) less than 320 ms for the video streaming, ii) less than 280 ms for the audio streaming, and iii) the maintained synchronization between the video and the audio should be between 40-60 ms. The audio streams should be less than 40 ms ahead of video streams, and less than 60 ms behind video streams.

B. Remote services

Coupling immersive services with haptic communications will result in new forms of applications in diverse verticals, such as digital health and Industry 4.0. This is highly motivated by the need to save resources (e.g., human lives, transportation, energy, and water). Remote services could be deployed using VR or AR applications or using HTC. An example of this is a remote robotic surgery, whereby a surgeon operates remotely while receiving a live 360 videos of the remote location projected on his VR-HMD. The surgeon uses haptic communications to control the remote robot. Another potential application is remote driving, such as controlling UAVs via VR-HMD to reach areas affected by a natural disaster. These services are highly sensitive to network latency: the latency should be as short as possible with an upper bound of 5 ms [13].

Fig. 3 shows how remote services work using immersive services. They could be deployed either as VR, AR or HTC

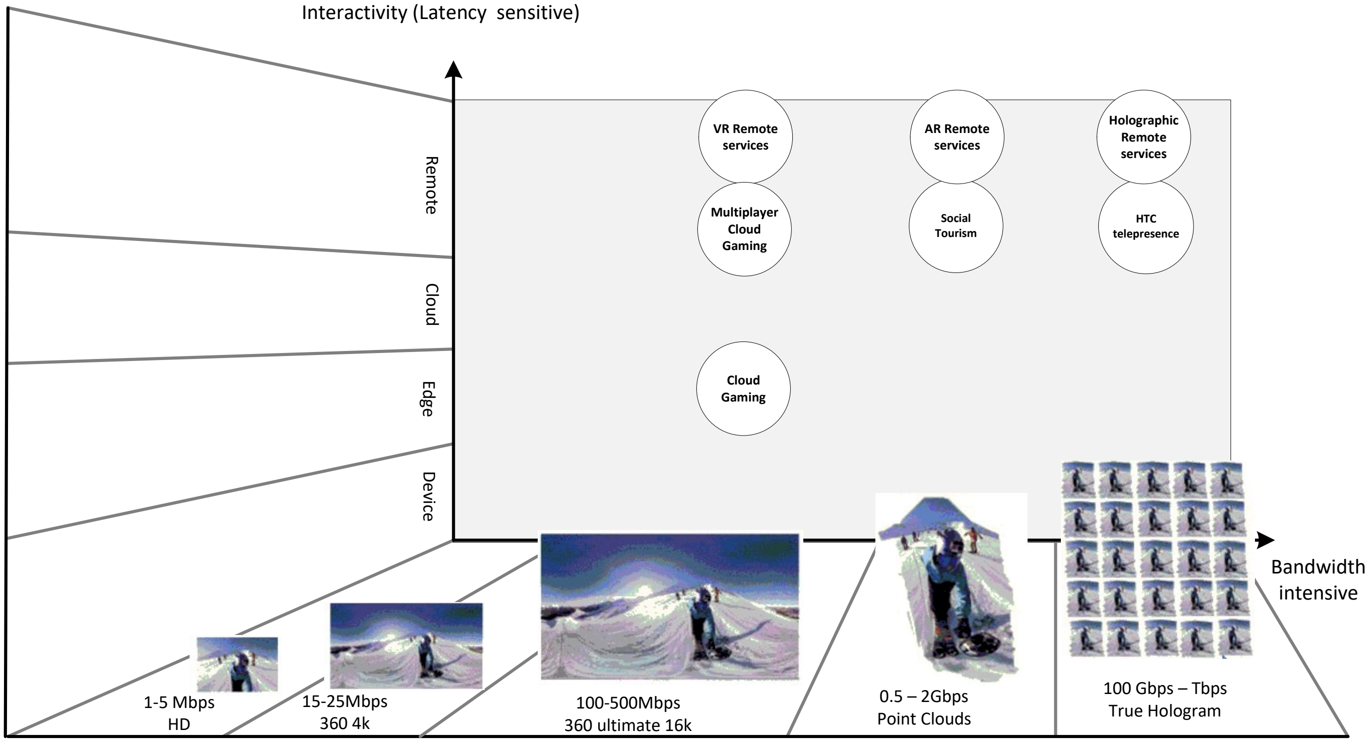

Fig. 2: Immersive services use cases and their requirements.

applications. Overall, a user controls the remote device using haptic communications. However, the difference resides in how the user perceives the remote location. This could be achieved via $i)$ VR where a 360-degree of the remote location is projected on VR-HMD, $ii)$ via AR – volumetric videos of the remote objects being projected on an AR-HMD, or via $iii)$ the remote object being projected in the real world as a hologram. Holograms give the user more flexibility in inspecting and controlling the remote object. However, this comes at the price of very low latency and much higher throughput compared to VR applications.

C. Automotive industry

A variety of AR and VR applications has been demonstrated for the automotive industry, such as ”Audi VR experience” as a virtual showroom, and ”ZeroLight VR” used by Toyota to launch their new cars. VR-based remote car driving is a critical application where less than 5ms of network latency is required to prevent any failure that could result in human fatalities or damages.

D. Social tourism:

Immersive social tourism is a mixed reality use case where both VR and AR are combined to provide an environment for tourism where locals act as guides using AR and remote users act as tourists using VR. The virtual scene on the VR-devices of remote users is a digital twin of the real site of interest. A set of sensors and cameras are deployed in the real world to get real time data; this data is provided to both remote and local users. All users can interact with each other, and remote users appear as AR objects on the AR-HMD of the local users and vice versa. The potential requirements of this use case are as follows: $i)$ the system should support scenarios with high density, at least 0.2 node per sq.m, $ii)$ the latency should be less than 10ms, and $iii)$ the positioning precision of local users should be less than 0.5 meter error and less than 100ms positioning latency [1].

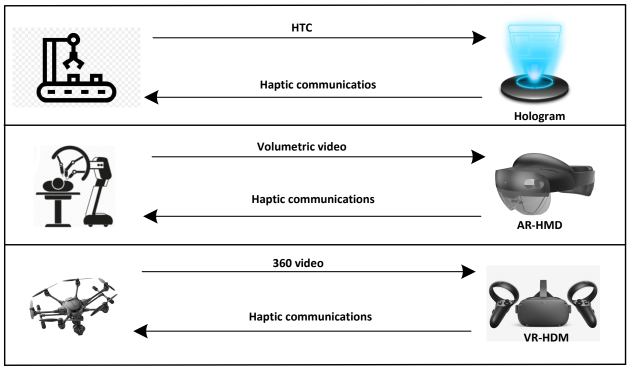

Fig. 3: Examples of remote services – VR-, AR-, and HTC-based variants.

E. Cloud gaming:

Entertainment applications, such as Cloud Gaming, require heavy computing resources for an immersive experience. Overall, multiplayer cloud gaming applications are more delay-sensitive and require a high refresh rate due to their high-interactivity feature compared to other applications. Cloud/Edge/Split Rendering is used for these applications to offload heavy computation from the end user devices and to ultimately save power. However, this requires a reliable and low latency communication (i.e., video streaming) between the end-devices and the edge cloud host, as the rendering of the scene is in real-time and is dependent on all the users' actions. This challenges the scalability of both the network and the edge cloud. In such applications, there is no buffering at the receiver, and packet loss may have a great impact on the streaming quality. Studies show that 0.1% packet loss will reduce the Mean Opinion Score by 10%.

IV. Technical Challenges & Potential Solutions

A. Technical Challenges

Bringing immersion to end users often requires significant real-time processing (i.e., rendering, vision, and physics engines) at the end-user devices, a fact that may strain all the local resources (i.e., computing, storage, and memory). Devices without sufficient resources for an immersive experience, such as thin-client devices, may lean towards offloading the relevant computing tasks to the cloud. In the ETSI reference architecture [8], both engines, rendering and vision, could be offloaded to the edge. Enabling the computing to be performed at the edge has a great impact on the service reliability and continuity, and may provide performance improvements to end consumer devices. Indeed, consumer devices may generate a large amount of data that might be inefficiently processed locally for several reasons (i.e., limited computation power and limited energy). In contrast, leveraging edge cloud resources will clearly reduce power consumption at the end consumer devices and increase the service reliability. Furthermore, relieving such computing from end consumer devices will open the door to service and consumer-device providers to offer cost-efficient devices, capable of supporting most immersive applications (i.e., 16K service, high refresh rates between 120 FPS and 240 FPS), to end users. However, such a processing split is feasible only when suitable edge cloud resources are available nearby end-user devices and they are accessible through a suitable allocation of radio and network resources.

Unlike enhanced Mobile Broad Band (eMBB) and Ultra-Reliable Low-Latency Communication (URLLC), current 5G networks do not have dedicated service definitions for VR and AR. Furthermore, XR does not exactly match any of the two 5G use cases, namely eMBB and URLLC. Indeed, VR and AR would require scaling to a massive number of simultaneous users – and that could be a characteristic of eMBB, yet, with a considerable support for low latency and high reliability – that are rather attributable to URLLC. 3GPP is considering to solve this problem by introducing a special quality indicator, 5G QoS Identifier (5QI), to be used for interactive media services. Moreover, having a XR-dedicated slice cloud be also a viable option, especially if the XR applications require end-to-end isolation, involving the radio access as well.

5G supports edge computing by selecting the closest breakout point (i.e., User Plane Function - UPF) relative to the closest edge media server [3]. However, tight coordination between the servicing media servers of XR and 5G core network needs to be ensured. This could be achieved through interactions between the XR service provider as an Application Function (AF) and the core network, more specifically the Network Exposure Function (NEF). Such tight coordination between XR service provider and 5G network becomes important under different events. For instance, in the case of a handover event, particularly when setting up a new immersive surrounding upon the handover, exposure to up-to-date radio network information becomes vital. This can occur only through NEF. In addition to the location of the media servers at the edge, and available radio and core network resources, content caching and positioning need enhancements. Furthermore, rendering of the content requires alignment with the viewer's physical posture relative to the immersive environment. The 5G network could assist the media servers with such position and orientation information. This information can help in fetching for the right cached content.

B. Solutions and Research directions

Immersive services require communication protocols that ensure both high throughput and low latency. Real-time streaming protocols, such as WebRTC, are essential for ultra-low latency immersive video streaming [7]. However, such streaming should be resilient to any event in the network, such as congestion, path and packet loss. Congestion control mechanisms play an important role for reliable communications in addition to embedded redundancy in the coding of the content. Recent developments in congestion control mechanisms showed a shift from loss-based algorithms such as Reno and CUBIC to congestion control algorithms that jointly look into loss, delay and bandwidth (e.g., BBR [4] and COPA [2]). Moreover, the QUIC protocol is gaining attention as a replacement to TCP. Adaptive streaming is another key factor for efficient streaming when the delay and the bandwidth are changing frequently. A potential adaptive streaming solution could be based on Foveated rendering and scalable streaming. In 5G networks, high bandwidth at the radio access can be achieved through carrier aggregation, leveraging the different bands of 5G. Deterministic Networking (DetNet) and Time Sensitive Networking (TSN) are also explored in the context of 5G and XR services may benefit from the respective developments.

For media streaming, high compression rate and low complexity still represent the main feature of efficient video encoders. HEVC and Advanced Video Coding (AVC) are the most used encoders. In comparison to the latter, the former achieves up to 50% bitrate saving while keeping the same quality. HEVC, which is still not fully supported by most devices, provides a maximum 8K of resolution and 120 FPS whereas AVC provides only a maximum 4K resolution and 60 FPS. Versatile Video Coding (VVC), the successor of HEVC, provides up to 50% of bitrate reduction with an increase in the computational complexity [6].

Given the huge data involved in an individual XR session, content caching becomes an important challenge [14]. In this vein, it is expected that collaborative content caching and streaming may improve the service experience as all users involved in an XR session are more likely to be in the same neighborhood/proximity. Overall, collaboratively, users could download the different parts of the common content (or 3D objects) from the cloud and share their dedicated and processed parts through caches at the local network. This would reduce the cumulative throughput for each user across the wide area network while keeping the same quality. However, this would not bring much benefit to VR applications when split-processing is in place. For instance, when two in-proximity VR users are involved in the same virtual environment, each user will receive a totally different stream according to his/her position in the virtual environment. However, for AR applications, users share the same real environment when they are in proximity, and their respective AR devices may collaborate on many aspects such as sensing the real world, requesting the same 3D objects either from each other or collaboratively from the cloud.

V. Conclusion

This article presented an overview of immersive services which are expected to be one of the killer services of beyond 5G and 6G networks. First, the article introduced some enabling technologies, followed by an overview of relevant use cases under discussions in various Standard Development Organization (SDO)s. These use cases pertain to different sectors, ranging from Health, Industry, to Entertainment and Education. Relevant industry initiatives to standardize the immersive service delivery are also presented. The article also discussed technical challenges hindering a wide deployment of cellular-based XR services, and briefly introduced some potential solutions.

References

[1] 3GPP. Study on Network Controlled Interactive Services, TR22.842, (Release 17), December 2019.

[2] Venkat Arun and Hari Balakrishnan. COPA: Practical Delay-Based Congestion Control for the Internet. In *Proceedings of the Applied Networking Research Workshop*, ANRW '18, page 19, New York, NY, USA, 2018. Association for Computing Machinery.

[3] V. Balasubramanian, M. Aloqaily, F. Zaman, and Y. Jararweh. Exploring Computing at the Edge: A Multi-Interface System Architecture Enabled Mobile Device Cloud. In *2018 IEEE 7th International Conference on Cloud Networking (CloudNet)*, pages 1–4, 2018.

[4] Neal Cardwell, Yuchung Cheng, C. Stephen Gunn, Soheil Hassas Yeganeh, and Van Jacobson. BBR: Congestion-Based Congestion Control: Measuring Bottleneck Bandwidth and Round-Trip Propagation Time. *Queue*, 14(5):20–53, October 2016.

[5] A. Clemm, M. T. Vega, H. K. Ravuri, T. Wauters, and F. D. Turck. Toward Truly Immersive Holographic-Type Communication: Challenges and Solutions. *IEEE Communications Magazine*, 58(1):93–99, 2020.

[6] A. Wieckowski et al. Towards A Live Software Decoder Implementation For The Upcoming Versatile Video Coding (VVC) Codec. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 3124–3128, 2020.

[7] ETSI ISG ARF Group Spec. 001. Augmented Reality Framework (ARF); AR standards landscape. Technical report, March 2020.

[8] ETSI ISG ARF Group Spec. 003. AR framework architecture. March 2020.

[9] Jon Karafin and Brendan Bevensee. 25-2: On the Support of Light Field and Holographic Video Display Technology. *SID Symposium Digest of Technical Papers*, 49(1):318–321, 2018.

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

[11] R. Li et al. New IP: An Extensible Framework to Evolve the Internet. *in Proc. IEEE HPRS 2020 Workshop on New IP, May 2020*, May 2020.

[12] S. Schwarz et al. Emerging MPEG Standards for Point Cloud Compression. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 9(1):133–148, 2019.

[13] ITU-T FG NET-2030 Sub-G2. New Services and Capabilities for Network 2030: Description, Technical Gap and Performance Target Analysis. October 2019.

[14] Sukhmani Sukhmani, Mohammad Sadeghi, Melike Erol-Kantarci, and Abdulmotaleb El Saddik. Edge Caching and Computing in 5G for Mobile AR/VR and Tactile Internet. *IEEE MultiMedia*, 26(1):21–30, January 2019. Conference Name: IEEE MultiMedia.

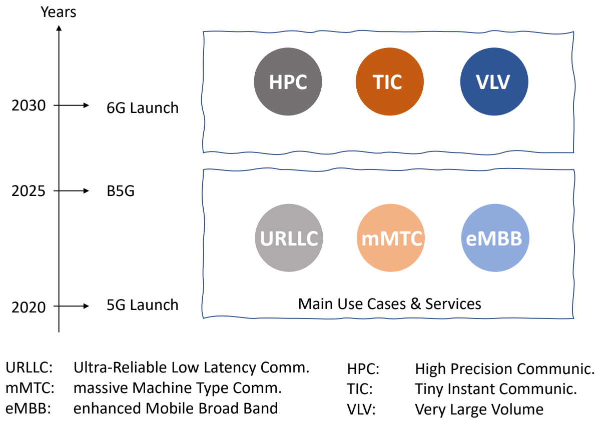
Years
2030
6G Launch
2025
B5G
2020
5G Launch
HPC
TIC
VLV
URLLC
mMTC
eMBB
Main Use Cases & Services
URLLC: Ultra-Reliable Low Latency Comm.
mMTC: massive Machine Type Comm.
eMBB: enhanced Mobile Broad Band
HPC: High Precision Communic.
TIC: Tiny Instant Communic.
VLV: Very Large Volume

Fig. 1: Evolution toward 2030 networks and the expected supported use cases & services.

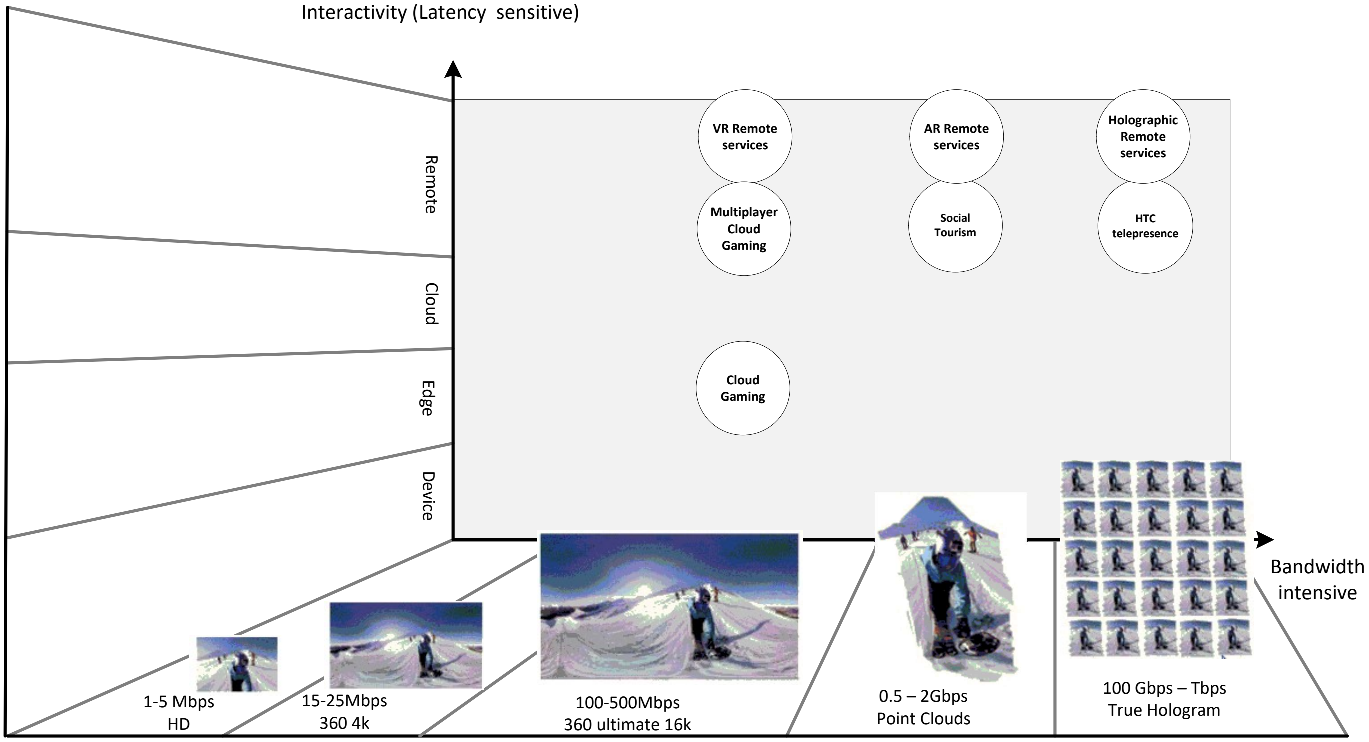

Fig. 2: Immersive services use cases and their requirements.

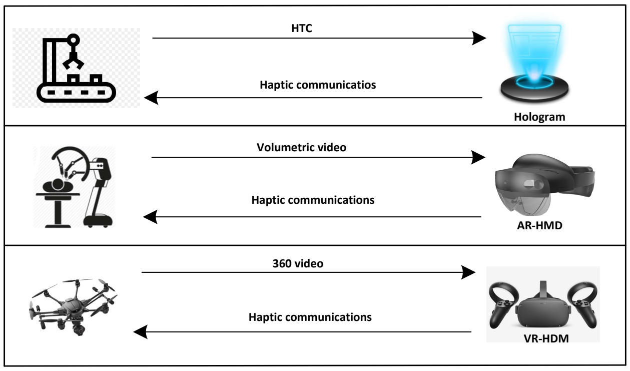

Fig. 3: Examples of remote services – VR-, AR-, and HTC-based variants.