Qinpei Luo *School of Electronics Engineering and Computer Science, Peking University, Beijing, China*
Hongliang Zhang, Boya Di *School of Electronics, Peking University, Beijing, China*
Minrui Xu *School of Computer Science and Engineering, Nanyang Technological University, Singapore*
Anthony Chen, Shiwen Mao *Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA*
Dusit Niyato *School of Computer Science and Engineering, Nanyang Technological University, Singapore*
Zhu Han *Electrical and Computer Engineering Department, Computer Science Department, University of Houston, TX, USA*

Editor: Nirupam Roy

AN OVERVIEW OF 3GPP STANDARDIZATION FOR EXTENDED REALITY (XR) IN 5G AND BEYOND

In recent years, aiming to enhance and extend user experiences beyond the real world, Extended Reality (XR) has emerged to become a new paradigm that enables a plethora of applications [1], e.g., online gaming, online conferencing, social media, etc. XR refers to the human-machine interactions that combine real and virtual environments with the support of computing/communications technologies and wearable devices. The XR content is generated by providers or other users, including audio, video and other metadata. In general, the generated XR content is transmitted to XR devices and rendered into XR scenes (i.e., to generate an image from a 2D or 3D model by means of a computer program), where users can experience a hybrid experience of the real and virtual worlds.

Photo, istockphoto.com

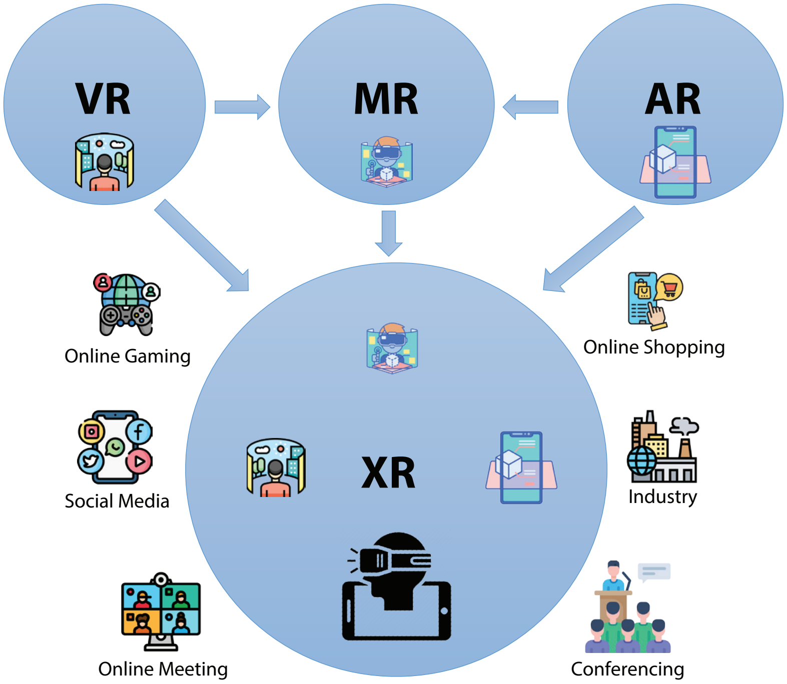

FIGURE 1. VR, AR, MR, and XR with some representative applications.

As shown in Figure 1, XR is a broad concept that encompasses Virtual Reality (VR), Augmented Reality (AR), Mixed Reality (MR), and all the areas interpolated among them. Virtual Reality provides users with a delivered visual and audio scene by mimicking the sensory stimuli from the real world as naturally as possible [2]. The user often wears a head-mounted display (HMD), e.g., Vision Pro just released by Apple Inc., and moves within the limits determined by the application. Augmented Reality is based more on a real environment, upon which additional information or content like artificially generated virtual objects or audio is added [3]. Mixed Reality is a combination of both VR and AR. Digital objects are placed into the physical environment and can interact with the real-world including humans in real time [4], with emerging applications, such as autonomous driving [5].

However, the vivid experience of XR often comes with higher quality of audio, video and other media formats, which pose a new challenge for future cellular networks to support XR services. Moreover, various XR applications, for example, XR gaming and conferencing, require frequent interactions among users and users' motion, and thus future cellular networks should provide ultra-low latency to provide satisfactory user experiences [6].

To achieve this vision, the 3rd Generation Partnership Project (3GPP) recently has made great efforts on standards and guidelines on support of XR services in the next generation of communication systems. As shown in Figure 2, dating back to 2016, 3GPP Technical Specification (TS) 22.261 took the VR environment into consideration when concerned about the low latency and high-reliability requirements of the 5G system [7]. Later in 2018, 3GPP Technical Specification Groups (TSG) Service and System Aspects Workgroup (SA) 4 specified typical traffic characteristics of 360° VR in Technical Report (TR) 26.925 [8]. Focusing on VR, 3GPP TR 26.918 has attempted to identify potential gaps and relevant interoperability points that may require further work and potential standardization in 3GPP to support VR use cases [9], but the discussion is constrained to three degrees of freedom (DoF) for 360° video and its corresponding audio. Following the report, 3GPP released TS 26.118 in August 2018, which defines the interoperable formats for VR streaming, including operation points, media profiles and presentation profiles [10]. In March 2020, 3GPP released TR 26.928 [11], in which more XR applications are supported by 5G. Most recently, 3GPP Radio Access Network Radio Layer 1 (RAN1) has conducted an evaluation of New Radio (NR) aiming to enhance the XR services upon NR [12]. In March 2022, 3GPP SA released TR 26.998 [13], which examines the glass-type AR/MR devices in the context of 5G radio and network services and provides documentation on the aspect of functional structures of these devices and core use cases.

In this article, we provide an overview of 3GPP standardization for XR in 5G and beyond. The remainder of the article is organized as follows. In the next section, we present the typical scenarios of XR use cases and their specific Quality of Service (QoS) requirements. In the section titled "5G Media Centric Architecture for XR," we introduce four types of key architectures for XR services and their corresponding traffic characteristics. Finally, we summarize the potential standardization issues for XR in 5G and beyond and conclude this article.

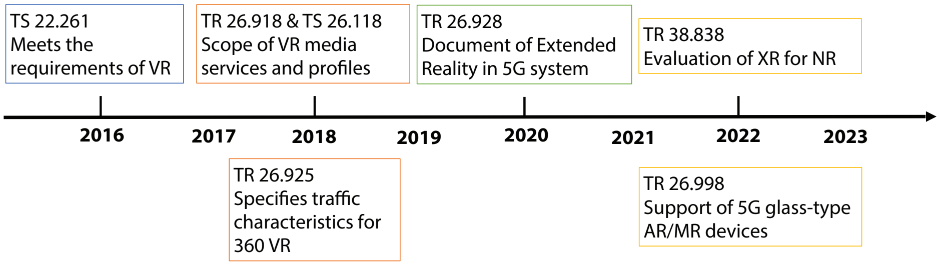
TS 22.261
Meets the requirements of VR
TR 26.918 & TS 26.118
Scope of VR media services and profiles
TR 26.928
Document of Extended Reality in 5G system
TR 38.838
Evaluation of XR for NR
2016
2017
2018
2019
2020
2021
2022
2023
TR 26.925
Specifies traffic characteristics for 360 VR
TR 26.998
Support of 5G glass-type AR/MR devices

FIGURE 2. Timeline of major 3GPP standardization activities for XR services.

TYPICAL USE CASES AND SCENARIOS FOR EXTENDED REALITY

In this section, we introduce the basic types of use cases and scenarios for XR services according to their functions and QoS requirements. According to 3GPP TR 26.928 [11], there are roughly seven types of XR use cases and scenarios as elaborated below.

Offline Sharing of 3D Objects: This type of use case deals mainly with the sharing of 3D objects/models and 3D XR scenes among users without interaction. In this scenario, user A can either download the media contents of 3D objects from the cloud or obtain them from other devices such as 3D cameras. Once the object contents are generated and uploaded onto the cloud, user B can download them and render the 3D object in his/her own environment. The XR scene that user B renders can be captured and fed back to user A. For example, when one of the users is shopping, the user can build a 3D model of an item with the depth camera of the smartphone and send it to other users via the Multimedia Message Service (MMS), and thus the other users can also see the item in sufficient detail and provide comments and suggestions. This kind of service does not require real-time interaction. Therefore, its demand on latency is not stringent. However, depending on the quality of the 3D object, the required uplink/downlink data rates may be high.

Real-time XR Sharing: This type of XR application is an augmentation of the previous use case, i.e., Offline XR Sharing, in which real-time interaction is now required. Users A and B need to conduct real-time XR multimedia streaming directly through a bidirectional/unidirectional Audio/Video (A/V) channel or indirectly through the cloud. The XR experience of user A can be affected by the input from user B and motion signals and effects from user A. Sharing XR experiences among different users can be achieved with a Spatial Computing Server (SCS), which can support collocated users to simultaneously view and interact with XR objects using up-to-date position data. Digital Twin can be a possible specific application of this use case. One user can build the digital twin of a tangible object in the metaverse and share it with other users through XR in real time. In this case, as real-time interaction is required, there is a high demand for latency, while the downlink/uplink data rates still depend on the quality of what is being shared.

XR Multimedia Streaming: This category includes both live and offline multimedia XR streaming, which is delivered from the content provider to users. The most common use of this service is volumetric video streams applied in AR to support 3DoF and 3DoF+ (referring to 3DoF with additional constrained movements, specifically, head movements, along the axes of X, Y, and Z) immersive experiences, which are already implemented by Google [14] to provide content services for other companies in production promotion, virtual touring, online shopping, etc. In this scenario, the user is able to receive XR streaming from the content server and render it with local graphic devices. The user is also able to control the playback with information from various sources, including HMD, body gestures and positions, biometric information, etc. A typical use case of this scenario is *emotional streaming*. When the user is watching an AI-generated film, his/her emotional reactions are collected and transmitted back to the content provider to create a personalized plot and style of the film. Multimedia streaming often requires low latency to provide a seamless experience, and the required downlink bitrate is usually high due to the high quality of audio and video content.

Online XR Gaming: This category covers various types of online gaming with XR devices. The game server is capable of simultaneously providing service for multiple online gamers. Each User End (UE) of the player receives the XR game streaming from the server and feeds back control signals, which are supposed to determine how the game is being played. Other users may join the online game as spectators. They can also receive the XR game streaming from the game server. The streaming may be exactly the same as one player receives if the spectator wants an immersive perspective as the specific player, or it could be from a perspective independent of all the players. As online gaming often requires real-time feedback and interaction among users, its latency demands can be low or ultra-low, and the required bandwidth depends on the resolution of the game scene.

XR Mission Critical: This special-use case describes the scenario of a team of users, each equipped with multiple mission gears connected to a centralized control center. The control center provides XR graphics such as maps, indicative texts and pointers of other team members, hostiles, or just objects in the surroundings to support the mission. It also allows control drones equipped with cameras to collect and extract important information from the field with its capabilities of A/V processing. For instance, in a police mission with AR, each team member is equipped with a helmet deployed with AR glasses, stereo headphones/microphones, 360° VR cameras, and 5G connectivity with very accurate 5G localization. One or several

TABLE 1. Types of Use Cases and Their QoS Requirements

Types of Use Cases and Scenarios	QoS Requirements					
	Latency		Bitrates/Bandwidth		Packet Error Rate (PER) Packet Loss Rate (PLR)	Reliable Delivery
	Downlink (DL)	Uplink (UL)	DL	UL		
Offline Sharing of 3D Objects	Non-critical[1]	Non-critical	Depending on the quality of the 3D Object representation		Low	Required
Real-time XR Sharing	Low		Depending on the quality of the 3D Object representation		Non-critical	Not Required
XR Multimedia Streaming	Low[2]	Moderate	High[2]	Depending on the type of controlling information	Non-critical	Not Required
Online XR Gaming	Depending on the type of game, can be low or ultra-low		Depending on the quality of the game scene, mostly 2Mbps [11] is enough	Moderate	Non-critical	Not Required
XR Mission Critical	Ultra Low		High		Non-critical	Not Required
XR Conference	Low or Ultra-low		Depending on the type of capture/ user representation transmitted [11]	Moderate and almost constant	Non-critical	Not Required
Spatial Audio Multiparty Call	Low		Depending on the quality of the video call		Non-critical	Not Required

Note 1: The metric of latency is according to the threshold of Roundtrip Time (RTT). Ultra-low: 50ms; Low: 100ms; Moderate: 200ms; Non-critical: >200ms.
Note 2: With viewport adaptive streaming and split rendering, the requirements on bitrates can be lower, meanwhile the latency requirement is higher.

drones are also deployed, which can be controlled by the team or command center to enhance their AR experience. The latency of the communications between the team and command center and among the team members must be extremely low to ensure the commands and feedback are up to date.

XR Conferencing: The conference space shared in the form of XR in this scenario can be divided into three types: First, a physical space shared with remote participants of an immersive stream; second, a virtual space that simulates the physical space sent to both local and remote conference participants; third, a completely virtual space provided by an application server. Each user end needs to capture the user's media information such as actions, motions, and expressions, meanwhile rendering the received multi-media stream from the conference server. The conference server is concentrated on processing all the session signals and setting up communication channels for the exchange of media data and metadata. A typical use case of this scenario is a virtual convention/poster session. In this setting, the poster session includes both real and remote attendees. Each remote attendee wears an HMD to obtain a VR conference hall experience, in which he/she can walk from poster to poster and present his/her own poster with a VR controller. By using AR glasses, a real participant can see the remote attendees as if they were physically present in the scene. They can listen to the presentation and interact with a remote presenter. This kind of application requires real-time feedback as two conference attendees communicate; thus, the required latency shall be low or ultra-low.

Spatial Audio Multiparty Call: This type of use case is often implemented with AR functions on a mobile phone. Captured by the front-facing camera, each party can see other parties displayed in an AR manner on the phone with 2D video streaming. If the user equipment is replaced by AR glasses, then with the spatial information collected, the user can receive more accurate stereo voice from other parties, thus he/she is able to enjoy a more immersive experience. This case demands low latency as it is a type of real-time communication.

Different types of use cases for XR services and their corresponding QoS requirements are summarized in TABLE 1, which is according to 3GPP TR 26.928 [11].

5G MEDIA CENTRIC ARCHITECTURE FOR XR

In this section, based on the core use cases introduced in the previous section, we now discuss how to incorporate the core technologies in 5G media centric architectures to meet the demands of XR applications.

Before presenting the architectures, we first introduce the general tasks of XR services:

- Displaying and viewport (referring to an area within the XR devices from which the user can view information) rendering to generate an XR scene with the content delivered;
- Capturing real-world content including tracking and pose generation;
- Media coding/decoding and media content delivery;
- Fixing media formats, meta data and other data;
- Utilizing 5G communication system;
- Spatial location estimation.

According to the execution locations of the above tasks, we can have roughly four types of architectures to support XR applications. Generally, all the use cases of XR can be attributed to mainly two types, conversational and non-conversational, depending on whether there exist interactions among multiple users. For non-conversational use cases, according to where the rendering takes place, it can be divided into unified rendering and split rendering, and the former can be further divided into viewport independent delivery and viewport dependent delivery contingent on whether there is a request for adaptive media. There is a general architecture for the conversational use case. Figures of the above four architectures are illustrated in Figure 3a to d.

Viewport-independent Delivery

In this architecture, the tracking information of XR sensors is only processed in XR devices without being fed back to the XR server [10]. That is to say, the entire XR scene is encoded by the server isolated from the XR device and delivered unidirectionally to XR devices. The entire delivery process in the viewport-independent architecture is illustrated in Figure 3a.

As shown in Figure 3a, the XR server managed by the XR application provider will take responsibility for processing the media including generation, encoding, and delivery. Then, the media content will be transmitted through a cellular system (e.g., 5G) in the form of downloading or passive streaming. Upon receiving the XR media content, the XR device will decode the content and render it along with the tracking information from the sensors, and then display the rendered media. The delivery and decoding of multiple video and audio streams are required to be in parallel to provide a seamless XR experience. The QoS characteristics and possible use cases of the above architecture are shown as follows.

1) QoS Characteristics As the rendering of XR media is processed locally on the XR device with this architecture, the delivery latency is not coupled with motion-to-photon latency. As the latter can be handled with new technologies on video content and compression, more attention has been paid to the streaming protocol. According to [16], to achieve an 8K resolution for 3D VR, we need at least 2.35 Gbps bandwidth and 10ms latency. From publicly announced demos at the 3GPP workshop "Immersive Media Meets 5G" in 2019 [17], based on devices nowadays and 2-3 years in the future, roughly 100 Mbps bandwidth can be realized to support High Definition (HD) 6DoF VR applications.

2) Possible Use Cases This architecture can be partially applied to *XR Multimedia Streaming* described in this section. As it only considers unidirectional delivery of XR media, it can support the use case, which does not support the interaction between the user and the generated XR scene. For example, with this architecture, a user can experience a football match as if he/she is sitting in the stadium with other audiences, but his/her actions have no effect on the ongoing game.

Viewport-dependent Streaming

Compared to the architecture of viewport independent delivery, the clear difference from the above structure is that the processing of media at the XR server is no longer independent from the XR device. Although the tracking information is still mainly processed within the XR device, the real-time viewport information is now fed back by the Adaptive Media Request to the XR delivery engine so that it can provide personalized information relevant to current viewports. As shown in Figure 3b, the tracking data from XR sensors are rendered to the viewport, and the media is adapted concerning the XR viewport, then transmitted

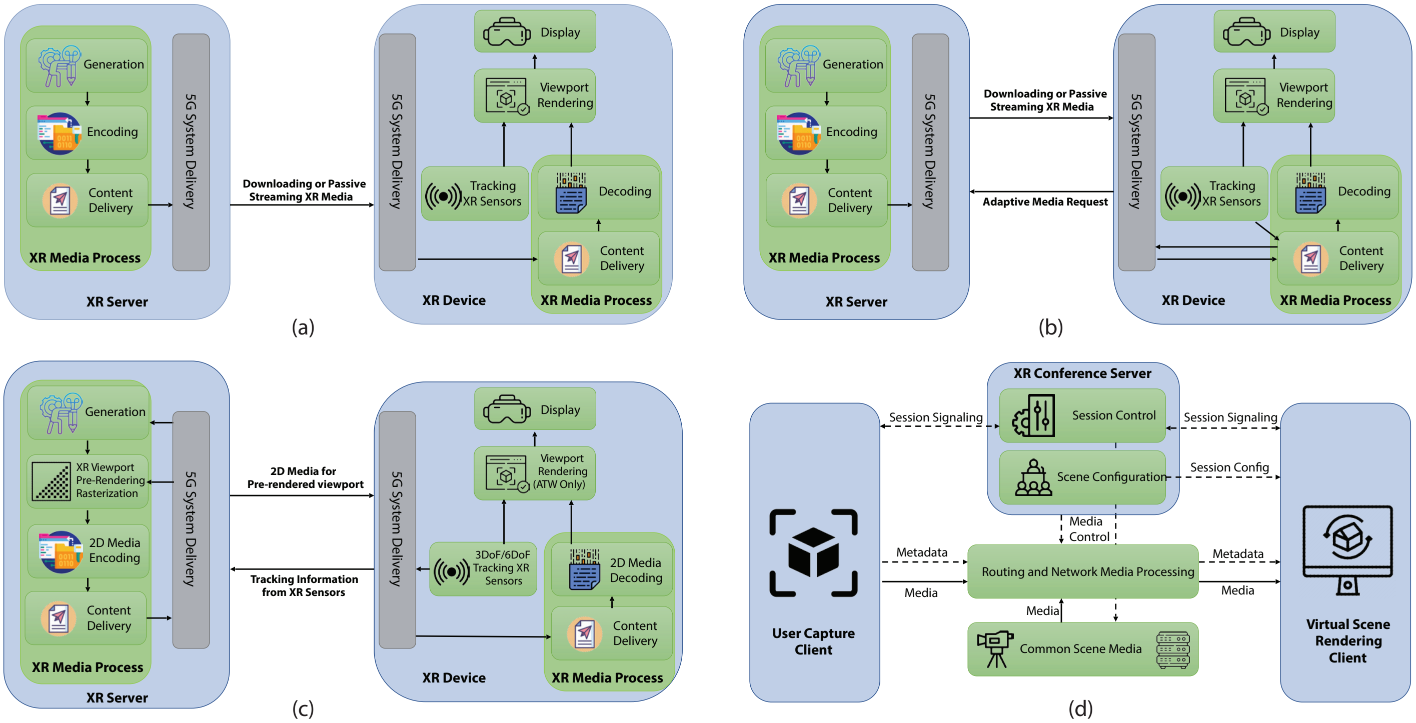

FIGURE 3. Four Architectures for 5G-XR applications: (a) The viewport independent delivery architecture. (b) The viewport dependent delivery architecture. (c) The raster-based split rendering architecture. (d) The general architecture for XR conversational and conferencing use cases.

back to the XR server. After that, a reduced or viewport specified scene is delivered and processed by the device. For example, if the user walks a few steps ahead, the XR server will not deliver the information of objects that are now behind the user, or just deliver them with a low quality. Due to its features, this architecture has the following three QoS characteristics and corresponding use cases.

1) QoS Characteristics With Adaptive Media Request introduced, updated tracking and sensor information can affect the network interactivity. Because of updated viewport information, HTTP/TCP level requests and responses are usually exchanged every 100-200ms [11] in such streaming architecture.

As the viewport information feedback can help the XR server to reduce the XR scene, according to 3GPP TR 26.918 [9] and other studies, such architecture is able to lower the required bitrate compared to viewport independent streaming at the same level of A/V quality.

Another thing worth mentioning about this architecture is that it also allows adjusting the quality of the XR scene to the available bitrate. Thus, such a system may be flexibly designed considering a tradeoff among bitrates, latencies, complexity and quality.

2) Possible Use Cases With the feedback from viewports, this architecture now can support those applications in *XR Multimedia Streaming* that include interactions and motions of users. For instance, a group of tourists who are visiting an XR museum can share the experience together. Each of them will be able to pause, rewind or fast forward the exhibited content, while other members can be synchronized to have the same experience.

Raster-based Split Rendering

In the Raster-based Split Rendering architecture, XR servers take the responsibility of generating the XR scene based on the tracking information from XR sensors in the XR devices.

As shown in Figure 3c, the XR scene is mainly rendered through the XR viewport pre-rendering rasterization process in the XR server. However, the XR device is still able to conduct real-time pose correction (referring to the process of XR devices re-estimate the pose of the human body with the tracking information from the sensors) by asynchronous time-warping (ATW), which is a widely used technique to reduce the motion-to-photon delay [18]. Other XR pose corrections may also be included in applications.

Therefore, we can see that producing the XR scene can be divided into two parts. One is the raster-based XR rendering deployed on an XR server, which can be set in the edge to reduce the network delay. The other is pose correction within XR devices. Because of its special structure, it has the QoS characteristics and use cases listed as follows.

1) QoS characteristics As ATW is introduced, the motion-photon latency can be largely reduced, so what determines the latency with this architecture for 5G delivery is mainly the roundtrip interaction delay. To meet the requirements of high bitrates and low error rates corresponding to the use cases, it is designed to support the range of 10-20ms latency and 50-100 Mbps guaranteed bitrate.

The uplink transmission predominantly consists of the tracking information from XR sensors. The data rates are at the level of 100kbps, and the latency is set small enough to avoid increasing the overall latency.

2) Possible Use Cases This architecture can be applied to those applications that include a lot of interactions and pose adjustments, for example, *Online XR Gaming*.

XR Conversational

Figure 3d shows the general architecture for XR conversational and conference services. Five steps are required for the XR content delivery as listed below:

1. The first client initiates the call setup.
2. The session control of the XR server triggers network media processing to reserve resources in the network.
3. Session control transmits call setup to the following clients.
4. After the call setup is finished, all the clients are connected to the network processor.
5. Session control works as a router to instruct the network processor to forward each stream.

This architecture has the following QoS characteristics and possible use cases.

THIS PAPER PROVIDES AN OVERVIEW OF THE ONGOING 3GPP STANDARDIZATION OF 5G NR IN SUPPORT OF XR APPLICATIONS

1) QoS characteristics For XR Conversational services, the designed delay is contingent on communication purposes. The ideal minimum delay is expected to be under 150ms. As for bandwidth requirements, the uplink is almost constant, while the downlink largely depends on the type of capture and representation of users transmitted, which is given below.

- 2D+/RGB Depth Camera: >2.7Mbps (single camera), >5.4Mbps (double cameras)
- 3D Mesh: roughly 30Mbps
- 3D VPCC / GPCC: 5-50Mbps.

2) Possible Use Cases This architecture is valid for all the XR applications that require conversational service, including *XR conference and Spatial Audio Multiparty Call.*

According to 3GPP TS 26.928 [11], the traffic characteristics of the above four architectures are shown in Table 2.

FUTURE WORK OF STANDARDIZATION FOR XR IN 5G

In the previous sections, we have discussed the advances in 3GPP standardization of 5G NR support for XR services. In the following, we will focus on the potential standardization issues of XR in future cellular systems.

1) Interface Between XR Devices and 5G Functionalities: XR centric devices are the key components for XR services. Thus, future standardization efforts should be made in a framework that synthesizes interfacing device centric XR functionalities and 5G System (5GS) Radio functionalities.

In 3GPP TR 26.998 [13], 3GPP has already introduced the 5G Media Access function of AR glasses. However, there are still mainly two aspects that need further study. First, more XR devices other than AR glasses should be taken into account. Second, a unified definition or protocol of interface for all XR devices and 5G functionalities is needed.

2) Traffic Characteristics and Models: Up to now for many XR services, their traffic characteristics have not been specifically defined. It should be a prime work to survey and collect data from typical XR applications. The main interest in traffic characteristics may include the following aspects:

- DL/UL data rate ranges;
- Maximum DL/UL Packet Delay Budget;
- Maximum Packet Error Rate;
- Maximum Roundtrip Time;
- Traffic Characteristics at the IP level including packet sizes in both DL/UL.

With these characteristics identified, we shall be able to establish 5G QoS flows optimized for specific types of XR applications.

3) Edge/Cloud Rendering and Processing: In future XR services, Edge/Cloud processing and rendering will be a promising key technology to simultaneously reduce power consumption and latency, which is critical for some XR use cases such as online XR gaming. Some work has been done aiming to reduce localization error and achieve fairness among users [19]. However, the following aspects of standardization still need to be considered:

- Generalized XR cloud and split rendering application framework.
- Formats and protocols that support content delivery and XR tracking information delivery at sufficiently high frequency.
- Distribution of computing resources in the 5G system network.
- 5G QoS Identifiers and other 5G system capabilities defined for edge/cloud rendering and processing of XR services.
- AI-enabled and AI-oriented edge computing for XR applications in the metaverse [20].

4) XR Awareness in RAN: Another intriguing aspect for further standardization work is to improve the performance of XR services by introducing and enhancing the ability of XR awareness in 5G RAN. If the RAN is able to identify which flow belongs to the XR service, one can optimize the transmission by adapting its priority, radio resource allocation, packet discarding and forwarding, with other network features to improve the experience of a specific XR application.

5) Support of Multi-party and Conversational: With the 5G Media Streaming architecture well established in 3GPP TS 26.501 [15], we can adapt it to fit XR applications. However, in conversational XR use cases with multi-party, such architecture may not fit very well, as it includes lots of metadata, spatial information, and audio/video from multiple resources. Thus, there is a need for standardization on the 5G network designed for multi-party and conversational XR cases, including the transport of spatial information, content delivery protocol, and other important issues.

TABLE 2. Traffic Characteristics for the Four Typical Architectures

Architecture	DL Rate Range	UL Rate Range	DL Packet Delay Budget (PDB)	UL PDB	RTT	DL PER Range	UL PER Range	Traffic Periodicity Range	Traffic file size distribution
Viewport independent streaming	100Mbps	HTTP Requests every second; TCP handshake.	300ms	300ms	Referred to Adaptive Streaming and TCP connection.	10e-6	10e-6	Almost Constant	Almost Constant
Viewport dependent streaming	25Mbps	More Frequent HTTP requests every 100ms; TCP Handshake.	300ms	300ms	Referred to Adaptive Streaming and TCP connection.	10e-6	10e-6	Almost Constant	Almost Constant
Raster-based Split Rendering with Pose Correction	100Mbps	500Kbps	20ms	10ms	50ms	FFS	FFS	Almost Constant	FFS
XR Conference	3~50Mbps per user	3~50Mbps	Meet the requirements of real-time communication, ideally less than 150ms			FFS	FFS	Almost Constant	>50Mb during start-up Then depending on media consumption or almost constant

NOTE 1: DL PDB and UL PDB are according to suitable 5QIs value defined in 3GPP TS 26.501 [15] for adaptive streaming over HTTP.
NOTE 2: FFS stands for "For further study."
NOTE 3: RTT, UL and DL PDB can't apply simultaneously.

CONCLUSIONS

This paper provided an overview of the ongoing 3GPP standardization of 5G NR in support of XR applications. In this survey, we first discussed why cellular systems were of great importance to support XR services and reviewed the history of 3GPP standardization on this topic. Then, we introduced the typical use cases of XR applications and their corresponding QoS requirements. To meet the requirements of these cases, four architectures were presented. We concluded this article with a discussion of future work. ■

Acknowledgement

This paper was supported in part by the National Key R&D Project of China under Grant No. 2020YFB1807100, in part by the National Research Foundation, Singapore, and Infocomm Media Development Authority under its Future Communications Research & Development Programme, DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-RP-2020-019), and MOE Tier 1 (RG87/22), and in part by U.S. National Science Foundation CNS-2148382, CNS-2107216, CNS-2128368, CMMI-2222810, ECCS-2302469, US Department of Transportation, Toyota, and Amazon.

Qinpei Luo is an undergraduate student of the School of Electronics Engineering and Computer Science, Peking University, where he is currently pursuing his BS degree. His research interests mainly focus on reconfigurable intelligent surfaces, machine learning for communication and reinforcement learning.

Hongliang Zhang received his PhD at the School of Electrical Engineering and Computer Science at Peking University in 2019. He is an assistant professor at the School of Electronics at Peking University. His interests include reconfigurable intelligent surfaces, aerial access networks, optimization theory, and game theory. He is a recipient of the 2021 IEEE Comsoc Heinrich Hertz Award for Best Communications Letters and the 2021 IEEE ComSoc Asia-Pacific Outstanding Paper Award.

Boya Di is an assistant professor at the School of Electronics, Peking University, Beijing 100871, China. She obtained her PhD from the Department of Electronics, Peking University, Beijing, China in 2019. Her interests include reconfigurable intelligent surfaces, edge computing, vehicular networks, and aerial access networks. She received the Best Doctoral Thesis Award from the China Education Society of Electronics in 2019 and is the recipient of the 2021 IEEE ComSoc Asia-Pacific Outstanding Paper Award.

Minrui Xu received his BS degree from Sun Yat-Sen University, Guangzhou, China, in 2021. He is currently working towards his PhD in the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests mainly focus on metaverse, mobile edge computing, deep reinforcement learning, and mechanism design.

Anthony Chen received his MS in electrical and computer engineering from Clemson University, Clemson, S.C., in 2022. Currently, he is pursuing his PhD in the Department of Electrical and Computer Engineering at Auburn University, Auburn, AL. His research interests include wireless communications, deep learning, and optimization.

Shiwen Mao received his PhD in electrical and computer engineering from Polytechnic University, Brooklyn, N.Y., in 2004. Currently, he is a professor and Earle C. Williams Eminent Scholar at Auburn University, Auburn, AL. His research interests include wireless networks and multimedia communications. He is the Editor-in-Chief of *IEEE Transactions on Cognitive Communications and Networking*, a Fellow of the IEEE, and a Life Member of the ACM.

Dusit Niyato is a professor in the School of Computer Science and Engineering, at Nanyang Technological University, Singapore. He received his BEng from King Mongkuts Institute of Technology Ladkrabang (KMITL), Thailand in 1999 and PhD in Electrical and Computer Engineering from the University of Manitoba, Canada, in 2008. His research interests are in the areas of sustainability, edge intelligence, decentralized machine learning, and incentive mechanism design.

Zhu Han received his PhD in electrical and computer engineering from the University of Maryland, College Park, in 2003. Currently, he is a John and Rebecca Moores Professor in the Electrical and Computer Engineering Department as well as in the Computer Science Department at the University of Houston, TX. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid. He has received many awards, including the IEEE Leonard G. Abraham Prize in the field of Communications Systems in 2016. He is an IEEE fellow, an AAAS fellow, and an ACM Distinguished Member.

REFERENCES

[1] C. Ziker, B. Truman, and H. Dodds. 2021. Cross Reality (XR): Challenges and opportunities across the spectrum, *Innovative learning Environments in STEM Higher Education: Opportunities, Challenges, and Looking Forward*, Springer, New York, NY, 55–77.

[2] G.C. Burdea and P. Coiffet, Virtual Reality Technology (2003), John Wiley & Sons, Hoboken, N.J.

[3] J. Carmigniani and B. Furht. 2011. Augmented reality: An overview, *Handbook of Augmented Reality*, Springer, New York, NY, 3–46.

[4] M. Speicher, B.D. Hall, and M. Nebeling. May 2019. What Is Mixed Reality?, in *Proc. CHI*, New York, NY, No. 537, 1–15.

[5] M. Xu, D. Niyato, J. Chen, H. Zhang, J. Kang, Z. Xiong, S. Mao, and Z. Han. Generative AI-empowered simulation for autonomous driving in vehicular mixed reality metaverses, *arXiv preprint arXiv:2302.08418*.

[6] S. Zeng, H. Zhang, B. Di, Z. Han and L. Song. Jan. 2021. Reconfigurable Intelligent Surface (RIS) assisted wireless coverage extension: RIS orientation and location optimization, *IEEE Commun. Lett.*, vol. 25, no. 1, 269-273.

[7] 3GPP, "5G; Service Requirements for the 5G System," TS 22.261, V16.14.0, Apr. 2021.

[8] GPP, "5G; Typical Traffic Characteristics of Media Services on 3GPP Networks," TS 26.925, V16.0.0, Mar. 2020.

[9] 3GPP, "Universal Mobile Telecommunications System (UMTS); LTE; Virtual Reality (VR) Media Services over 3GPP," TS 26.918, V15.2.0, Mar. 2018.

[10] 3GPP, "5G; Virtual Reality (VR) Profiles for Streaming," TS 26.118, V16.2.1, May 2021.

[11] 3GPP, "5G; Extended Reality (XR) in 5G," TS 26.928, V16.0.0, Mar. 2020.

[12] V. Petrov, M. Gapeyenko, S. Paris, A. Marcano, and K.I. Pedersen. Extended Reality (XR) over 5G and 5G-advanced new radio: Standardization, applications, and trends," *IEEE Network* (to be published), *arXiv preprint arXiv:2203.02242*.

[13] 3GPP, "LTE; 5G; Support of 5G Glass-type Augmented Reality / Mixed Reality (AR/MR) devices," TS 26.998, V17.0.0, Mar. 2022.

[14] "Immersive Stream for XR." Google Cloud. [Online]. Available: https://cloud.google.com/immersive-stream/xr.

[15] 3GPP, "5G; 5G Media Streaming (5GMS); General Description and Architecture," TS 26.501, V16.5.0, Sept. 2020.

[16] E.S. Wong, N.H.A. Wahab, F. Saeed, and N. Alharbi. July 2022. 360-degree video bandwidth reduction: Technique and approaches comprehensive review. *Applied Sciences*, vol. 12, no. 7581, 1-25.

[17] "2nd VR Ecosystems & Standards Workshop" [Online]. Available: https://www.vr-if.org/events/3gpp-vrif-ais-workshop/.

[18] J.M.P. van Waveren. Nov. 2016. The asynchronous time warp for virtual reality on consumer hardware. *Proc. ACM VRST*, New York, NY, 37–46.

[19] H. Zhang, S. Mao, D. Niyato and Z. Han. Jan. 2023. Location-dependent augmented reality services in wireless edge-enabled metaverse systems, *IEEE Open J. Commun. Soc.*, vol. 4, 171-183.

[20] M. Xu, D. Niyato, H. Zhang, J. Kang, Z. Xiong, S. Mao, and Z. Han. Sparks of GPTs in Edge Intelligence for Metaverse: Caching and Inference for Mobile AIGC Services. *arXiv preprint arXiv:2304.08782.*