Document 5D/ E
21 December 2022
English only

IAFI[1]

FURTHER UPDATES TO WORKING DOCUMENT TOWARDS A PRELIMINARY DRAFT NEW REPORT ITU-R M. [IMT.MULTIMEDIA]

Capabilities of the terrestrial component of IMT-2020 for multimedia communications

1 Introduction

During the 42nd meeting of Working Party 5D, work was continued to further develop the working document addressing the new multimedia related capabilities of IMT-2020

2 Proposal

Further changes are proposed to the working document in the attachment:

- In Section 2, Introduction (Multimedia Applications), three latest applications are added.
- A new section in the table of contents has been added and updated this is regarding various acronyms that have been added to the document
- Additional text has been added in Section 5 regarding trend and demand of application of multimedia, utility of video and video streaming and usefulness of cloud architecture is explained.
- Additional text has been added in Section 6 of the document regarding the stages of development of MBMS by 3GPP are explained.
- In Section 8.11, some text has been redrafted and four latest capabilities have been further updated.
- Section10 (Summary) The document has been summarized for a holistic view of the capabilities of the terrestrial component of IMT 2020 for Multimedia communications

The proposed changes to the working document are in the attachment. These are highlighted in yellow

[1] ITU-APT FOUNDATION OF INDIA (https://itu-apt.org)

Annex 3.3 to Working Party Chairman's Report

WORKING DOCUMENT TOWARD A PRELIMINARY DRAFT NEW REPORT ITU-R M. [IMT.MULTIMEDIA]

Capabilities of the terrestrial component of IMT-2020 for multimedia communications

(Question ITU-R 262/5)

[Editor's note: As this document is developed, WP 5D may consider incorporating any overlapping information into Report ITU-R M.2373, as appropriate.]

(20YY)

TABLE OF CONTENTS

[Editor's note: TBD]

[Editor's note: ToC to be updated once the document has been consolidated.]

Page

1 Scope

This Report addresses the capabilities of IMT-2020 to distribute multimedia content such as video, audio, text and graphics, including support for real-time multimedia interactive applications. This report also addresses the capabilities of IMT-2020 user devices and base stations to support such multimedia communications with low latency and wider transmission bandwidth.

This new report complements Report ITU-R M.2373 on “Audio-visual capabilities and applications supported by terrestrial IMT systems” which addresses the capabilities of IMT systems for delivering audio-visual services to the consumers and also covers some aspects of production of audio-visual content.

2 Introduction

Multimedia applications include network video, digital magazine, digital newspaper, digital radio, social media, mobile TV, digital TV, touch media, etc., that are enabled by IMT-2020 technologies. Beyond the traditional media service, the new media application not only supports accurate delivery of content, but also supports real-time interaction and real-time uploading of user-generated content. The users can be both consumers and producers of new media content.

These applications for multimedia content include but are not limited to:

– audio-visual applications,
– network video applications,
– digital online magazine applications,
– digital online news paper applications,
– internet radio applications,
– social media applications,
– mobile internet TV applications,
– touch media applications,
– online information distribution applications,

– on-demand video applications,
– imaging and audio distribution applications,
– content dissemination applications,
– file delivery application.
_ Real time uploading of self-produced content
_ Electronic classroom presentation technology
_ Full motion video conferencing

This report covers the application of IMT technology to the specific applications mentioned above. For details of applications of the Broadcasting service for multimedia, please refer to the list of ITU-R Recommendations and ITU-R Reports in Section 3 below.

3 Relevant ITU-R Recommendations and Reports

Recommendation ITU-R BT.1833 – *Broadcasting of multimedia and data applications for mobile reception by handheld receivers*

Recommendation ITU-R BT.2016 – *Error-correction, data framing, modulation and emission methods for terrestrial multimedia broadcasting for mobile reception using handheld receivers in VHF/UHF bands*

Recommendation ITU-R M.2083 – *Framework and overall objectives of the future development of IMT for 2020 and beyond*

Report ITU-R BT.2049 – *Broadcasting of multimedia and data applications for mobile reception*

Report ITU-R BT.2295 – *Digital terrestrial broadcasting systems*

Report ITU-R M.2373 – *Audio-visual capabilities and applications supported by terrestrial IMT systems*

4 Acronyms

2K resolution Content having 2000 pixels in horizontal resolution

4K Resolution 3840 wide x 2160 tall pixels = 8.30 megapixels

8K Resolution 7680 wide x 4320 tall pixels = 33.20 megapixels

AMF Access and Mobility Management Function

AR Augmented Reality

CA Carrier Aggregation

CLI Cross-Link Interference

CLI-RSSI CLI Received Signal Strength Indicator

CN Core Network

CSI Channel Status Information

DC Dual Connectivity

DSS Dynamic Spectrum Sharing

E1 Interface between gNB-CU-CP and gNB-CU-UP

eMBB Enhanced Mobile Broadband

eNB Evolved Node B – Radio Base Station used in 4G LTE

F1 Interface Interface is between gNB-CU and gNB-DU

FDD TDD CA Carrier Aggregation between TDD and FDD bands

FR2 Frequency Band above 24.250 GHz

GBR Guaranteed Bit Rate

gNB Next Generation Node B – Radio Base Station used in 5G NR

GPU Graphic Processing Unit

HARQ Hybrid Automatic Repeat Request

HD HD means – 1020 tall x 1920 wide pixels = 2 million pixels

HPHT High Power High Tower

HSDPA High Speed Downlink Packet Access

HSUPA High Speed Uplink Packet Access

IoT Internet of Things

IPTV Internet Protocol Television

L1/L2 Control Signaling

LAN-VN Local Area Network-Virtual Network

LTE Long Term Evolution

MBMS Multimedia Broadcast Multicast Service

MCG Master Call Group

MIMO Multiple Input and Multiple Output

mMTC Massive Machine Type Communications

MPMT Medium Power Medium Tower

MR Mixed Reality – VR + AR

MR Mixed Reality, combination of AR and VR

MR-DC Multi-Radio Dual Connectivity

ACK/NACK Used in Data Transmission

NG-RAN New Generation Radio Access Network

NPN Non-Public Network (Private Network)

NR New Radio O-FDMA Orthogonal Frequency Division Multiple Access

PAPR Peak to Average Power Ratio

PCF Policy Control Function

PDSCH Physical Downlink Shared Channel

PDU Protocol Data Unit

PLMN Public Land Mobile Network

PRACH Preamble Random Access Channel

PTM Point to Multipoint

PTP Point to Point

PUCCH Physical Uplink Control Channel

QFI QoS Flow ID

QR Code A type of Bar Code

RAN Radio Access Network

RIM Remote Interference Management in 5G

RIT Radio Interface Technology

RRC Radio Resource Control

SC-FDMA Single Carrier- Frequency Division Multiple Access

SFN Single Frequency Network

SMF Session Management Function

SRIT Set of Radio Interface Technology

SRS Sounding Reference Signal

SRS-RSRP SRS-Reference Signal Received Power

TRP Transmission and Reception Points

UE User Equipment

UHD 5K Resolution – 5000 pixels wide

UMTS Universal Mobile Telecommunications Service

UPF Use Plane Function

URLLC Ultra-Reliable (UR) Low Latency Communications

URLLC Ultra-Reliable Low Latency Communications

UTRAN UMTS Terrestrial Radio Access Network

V2V Vehicle to Vehicle

V2X Vehicle to Everything

VR Virtual Reality

XRM External Resource Management

5 Trends and demands of applications for multimedia content supported by IMT-2020 technologies

IMT-2020 is expected to revolutionize the mobile experience with much faster, always-on, always connected, and responsive mobile Internet.

By end of 2025, video traffic is estimated to account for 69 percent of all mobile data traffic, a share that is forecast to increase to 79 percent by 2027[2] . as user trend is shifting from text to image to video and now to video streaming. After invention of smart phone and internet, anyone can be a video maker. With increase of video popularity, hundreds of hours of videos are uploaded every minute on YouTube. Multimedia is a very effective method of communicating information because it enriches presentations, retains the audience's attention, and allows multiple and flexible interaction

Many Multimedia applications are already in use and many more such applications are being developed quite rapidly. The New Trend to develop latest multimedia application towards high-definition video, real-time multimedia interaction, AR/VR and real-time uploading of self-produced content.

As an emerging technology, Virtual Reality (VR) being capable to develop immersive and interactive video content, now becoming a key technology to upgrade or even subvert traditional industries. Its application fields include video games, event live streaming, video entertainment, healthcare, real estate, retail, education, engineering and public safety.

The Augmented Reality (AR) (enhanced version of VR) applications are also being developed very rapidly and the industry is in the market start-up period. AR hardware products are constantly evolving, and the ecosystem is gradually maturing. Mixed reality also referred as Hybrid Reality is the merging of AR and VR world to produce new environments and visualizations where physical and digital objects can co-exist and interact in real time.

Extended Reality (XR) is a newly added term and defined as an umbrella for all the three realities namely VR, AR and MR. Extended Reality (XR) refers to all real-and-virtual combined environments and human-machine interactions generated by computer technology and wearables

With the rapid development of the mobile internet and XR (eXtended Reality), B2B2C and B2C based mobile applications have become important exploration areas for AR industries, mainly serving enterprise-level users such as automobiles, machinery manufacturing, real estate, TV stations, publishing houses and exhibition halls. From the perspective of application scenarios, commodity advertising and early childhood education may be the most popular scenarios at present.

Video distribution supported by IMT-2020, whether it is on-demand, live streaming, or video surveillance, etc. will evolve to 4K resolution, thereby improving user experience, information transmission and image recognition experience. The interactive user experience - selection of and loading videos, switching channels and other viewing operations through the terminal during the user's viewing process - is another developing trend. Users expect operations to complete virtually instantaneously. In order to meet user expectations, initial video loading should be completed within 1s and the channel switching within 500 ms. One study[3] predicts that in the next 15 years, users of mobile video streaming will continue to grow steadily in all regions of the world and become the most important application form.

---

[2] *Source*: Ericsson Mobility Report, June 2022. https://www.ericsson.com/49d3a0/assets/local/reports-papers/mobility-report/documents/2022/ericsson-mobility-report-june-2022.pdf.

For additional information refer to the interactive graph at Ericsson Mobility Visualizer - Mobility Report - Ericsson.

[Editor's note: Double check and web-links for the footnotes 2 and 3 are needed.]

[3] Strategy Analytics.

IMT-2020 and cloud technology enable the use of cloud VR. This means moving the most complex GPU rendering processing of VR to the cloud and to deliver video streams via downlink, with action command via uplink, through broadband networks. Compared with the codec function implemented in VR terminal, cloud VR can achieve more than 50[4] times video streaming compression ratio while the codec function is implemented in the VR cloud. Cloud VR utilizes the powerful processing power of the cloud to reduce the requirements for local processing, which helps to reduce the complexity of the terminal, thereby lowering the threshold for consumers to use. The cloud VR application scenarios require a large user experienced data rate of around 100 Mbit/s to ensure a high-definition video experience (> 2K), and a network delay of 5~10 ms (MTP delay of 20 ms) to eliminate user motion sickness effects.

With the continuous development of mobile augmented reality, users have increasingly higher requirements for augmented reality application experience: smooth presentation, real-time interaction, and persistent operation, which pose challenges to the computing power and media processing capabilities of mobile terminal. Efficient invoking the hardware capabilities of mobile terminals, quick identification and capture of augmented reality targets in different business execution environments, and superimposing the smooth display of augmented reality content of various media types in real time, all greatly affect the user experience. Cloud AR can balance the computing power and cost. In the future, with the help of the Cloud AR platform, the great computing power required by rendering videos or pictures can be provided with lower cost. It is predicted that the Cloud AR is future direction of AR development, which provides increased computing power while reducing terminal costs. At the same time, the flexible deployment method of the Cloud AR platform will help to achieve low network delay which is very important for AR applications.

Edge Cloud: It is an architecture is used to decentralise (processing) power to the edges (clients/devices) of the networks. Traditionally the computing power of servers are used to perform tasks such as data minimisation or to create advanced distributed systems. Within the cloud model, such intelligent tasks can be performed by servers, so that a normal mobile with less or almost no computing power can be used.

Co-working with edge cloud services, IMT-2020 can connect the user to a high-definition virtual world on their mobile device. Live events with High Definition (HD) and Ultra-High Definition (UHD) content can be streamed via an IMT-2020 radio network with higher throughput. HD and UHD TV content can be accessed on mobile devices without any interruptions through IMT-2020 higher user experienced data rate. The entertainment industry be will hugely benefited from the IMT-2020 wireless networks, which are expected to provide high resolution, and high dynamic range video streaming without interruption. Cloud AR and Cloud VR with HD or UHD video require higher user experienced data rate and low latency. HD virtual reality games are becoming popular while IMT-2020 network can offer a better real-time interactive gaming experience. It is expected that with the support of IMT-2020 technologies, an amazing virtual experience will bring to the users and the above mentioned multimedia services will become the basic services in the future mobile Internet.

[4] IMT-2020(5G) PG 5G white paper: new multimedia demands and technical research report clause 2.2.3

6 Overview of the technical characteristics of IMT-2020 technologies for multimedia communications

[Editor's note: The texts of this chapter in the square brackets need to be further discussed.]

[Editor's note: information received in input document 5D/1531 needs to be considered for possible inclusion]

Three primary 5G NR use cases defined by 3GPP are:

1. *Enhanced Mobile Broadband (eMBB): data-driven use cases requiring high data rates across a wide coverage area.*
2. *Ultra-reliable Low Latency Communications (URLLC): strict requirements on latency and reliability for mission critical communications, such as remote surgery, autonomous vehicles or the Tactile Internet.*
3. Massive Machine Type Communication (mMTC): need to support a very large number of devices in a small area, which may only send data sporadically, such as Internet of Things (IoT) use cases.

Multicast and Broadcast services (MBS) in IMT-2020 has been one of the most important feature required by IMT system to tackle the increased emergence of video services, ad-hoc multicast/broadcast streams, software delivery over wireless, group communications and broadcast/multicast IoT applications.

It aims to provide a flexible and dynamic allocation of radio resources between unicast, multicast, and broadcast services within the network , to offer multimedia services (including new media, audio, and visual applications) over specific geographic areas (e.g., spanning a limited number of base stations) as well as wide geographic areas respectively.

The IMT-2020 specification supporting Multicast and Broadcast services has been developed step-by-step.

The Multimedia Broadcast/Multicast Service (MBMS) was first introduced in the 3GPP standard released in 2005 (Release 6) to optimize the distribution of broadcast and multicast services via the 3G Cellular systems.

In Dec 2008, 3G PP vide Release-8 first time introduced LTE and finalized standards for High Peak Data Rates up to 300 Mbps in downlink and 75 Mbps in uplink, by using 4×4 MIMO and 20 MHz bandwidth, High Spectral Efficiency, Flexible Bandwidths as 1.4 MHz, 3 MHz, 5 MHz, 10 MHz, 15 MHz and 20 MHz and OFDMA in downlink and SC-FDMA in uplink etc.

In Dec 2009, 3G 3GPP Release-9 added world-wide interoperability between LTE, WiMAX, and Universal Mobile Telecommunications System (UMTS) and also included dual-cell HSDPA with MIMO, LTE Home eNodeB (HeNB) and dual-cell HSUPA.

3GPP since then in various Releases from 10 to 14 further upgraded the standards for enhancement in eMBMS capabilities and many other features of 4G services, released total six 3GPP standards (9-14) in 10 years.

In 2017-18, 3GPP via Release-15, for the first time introduced 5G specification (5G Phase-I) in three phases. The first phase focused primarily on mobile broadband for non-standalone (NSA) 5G architecture. The second phase introduced for standalone (SA) 5G architecture. The third phase introduced an architecture for migration from 4G to 5G. Its primary focus was enabling enhanced

Broadband (eMBB), as offer very high uplink throughput, lower latency and/or higher capacity to support real-time multi-media information/interaction and real-time uploading of self-media content.

In 2019-20, 3GPP vide Released 16 (5G Phase-II) further improved the features of 5G (Release 15) and incorporated Dynamic Spectrum Sharing (DSS), Network Slicing and other features designed for private 5G Network. These developments marked a new focus on enterprise and business centric capabilities. Other capabilities included New Radio (NR)- based access to unlicensed spectrum (NR-U) and satellite access. The completion of Release 16 occurred in two stages, Stage -1in late 2019 for the physical layer aspects and Stage - 2 in late 2020 for the higher layer aspects.

In 2019-20, **3GPP Release 17** further improved the features of release – 15 and 16 for the 5G network including DSS and private 5G network capabilities. It is the successor to Release 16 and is one of the most versatile release in the history of 3GPP with significant enhancements in the Next-Generation Radio Access Network (NG-RAN) and 5G Core Network (5GCN).

The NR MBS in 3GPP Release 17 enables such a service over a specific geographic area which further enables a more efficient and effective delivery system for real-time and streaming multicast/broadcast content. NR MBS is able to carry IP multicast as well as Ethernet multicast packets with better QoS support using dynamic delivery mode switching.

The International Telecommunications Union has adopted the radio interface standards for 5G as part of its IMT-2020 recommendation. These standards are the 3GPP 5G-SRIT and 3GPP 5G-RIT from the 3GPP and 5Gi from the Telecommunications Standards Development Society India.

The IMT-2020 technologies consisting of 3GPP 5G-SRIT and 5G RIT MBS (including “LTE-based 5GBroadcast”[5] and NR MBS) which allows the IMT-2020 technologies to efficiently deliver Multicast and broadcast services.

The “LTE-based 5G Broadcast” finished in 3GPP Release 16 is based on legacy LTE eMBMS that can be traced back to 3GPP Release 8&9 which lay down the foundation for cellular broadcast. “LTE-based 5G Broadcast” is able to apply 100% of radio resources of one or more radio carriers for the delivery of broadcast/multicast content [in a stand-alone multicast/broadcast network comprising of multiple cells] with inter-site distances of up to 200 km. [ETSI TS 103 720 V1.1.1 (12/2020) specifies LTE-based 5G terrestrial broadcast system, including the radio and the core parts of the system. The 3GPP-developed LTE based 5G Broadcast technology is the radio part, profiled in ETSI TS 103 720 V1.1.1. ITU-R has described this system in Recommendation ITU-R BT.2016-3 as “System L”].

The NR MBS in 3GPP Release 17 enables such a service over a specific geographic area which further enables a more efficient and effective delivery system for real-time and streaming multicast/broadcast content. NR MBS is able to carry IP multicast as well as Ethernet multicast packets with better QoS support using dynamic delivery mode switching.

“LTE-based 5G Broadcast” is the term utilized by 3GPP during the specification development phase. It has been included one of the three radio interfaces of IMT-2020, 3GPP 5G-SRIT, and, it is adopted in Recommendation ITU-R M.2150-1 (02/2022) – *Detailed specifications of the terrestrial radio interfaces of International Mobile Telecommunications-2020 (IMT-2020)*. [6] The motion-to-photon latency is defined as the latency between the physical movement of a user's head and the updated picture in the VR headset. The motion-to-sound latency is the latency between the physical movement of a user's head and updated sound waves from a head mounted speaker reaching their ears.

Beyond the MBS in IMT-2020, from 3GPP Release 15, it has been developed related IMT-2020 technologies to offer very high uplink throughput, lower latency and/or higher capacity to support real-time multi-media information interaction and real-time uploading of self-media content.

[Interference coordination is one of key characteristics. If non-synchronized frame structures are deployed for public and non-public networks, the interference between gNB and gNB or UE and UE would cause big impacts on the overall system performance. Interference coordination specified in 3GPP Release-16 could mitigate such interference. / Interference between gNB and gNB or UE and UE should be avoided as it would cause big impacts on the overall system performance. 3GPP Release-16 extends interference mitigation techniques to certain operational environments, such as those in non-synchronized TDD networks.]

Carrier Aggregation (CA) and Dual Connectivity (DC) are the straightforward and effective ways to boost uplink throughput by directly increasing the uplink frequency resources and possibly cell capacity. With CA technology, a UE can receive or transmit on one or multiple contiguous/non-contiguous component carriers. The CA was firstly introduced in LTE Release 10. CA in IMT-2020 was specified from 3GPP Release 15 and support maximum 16 carriers with 400 MHz each, thus it can support up to 6.4 GHz bandwidth. It is continuously developed in 3GPP Release 16, Release 17 and so on to support more flexible schedule and interference mitigation. The Multi-Radio Dual connectivity (MR-DC) was introduced in 3GPP Release 15. It is comprised of LTE+NR DC, NR+NR DC, NR+LTE DC etc. The RATs utilized in the first node and second node can be different. It is also continuously enhanced in 3GPP Release 15, Release 17 and so on to improve radio resource efficiency and reduce latency.

UL MIMO is another technology to improve up link data rate. It was specified for IMT-2020 from 3GPP Release 15. It introduces CSI feedback and reference signalling design which are more flexible than LTE. With high resolution codebook and beam forming characteristics, the data throughput and capacity of IMT-2020 system are expected to be met. In 3GPP Release 16, Release 17, it continuously developed CSI feedback, Coordinated Multiple Points (CoMP), beam management, power control, PAPR reference etc. to offer higher radio resource efficiency and communication performance for new media e.g., XR applications.

7 Use cases

[Editor's note: The texts in square brackets need to be further reviewed.]

7.1 Ultra-high-definition [audio-visual for live event]

Live events (such as concerts, stage shows, sports events, etc.) [which is a typical audio-visual application] require more content details. For instance, in the concert, the 4K-based multi-screen and multi-camera live streaming can capture extremely vivid details such as subtle "micro-expression" changes, bringing a strong visual impact to the audience. Meanwhile, the application of HDR and wide colour gamut technology will perfectly present the effect of live lighting and stage design to the audience. In band performances, multi-camera allows the audience to see their favourite players. In addition, ultra-high-fidelity audio transmission is also necessary for stage shows. For large concerts (such as orchestras), multi-channel audio transmission with Hi-Fi music quality with inaudible noise and distortion, and a flat frequency response within the human hearing range, will allow listeners to experience a truly immersive experience.

With the support of UHD streaming by IMT-2020, the above scenarios will be feasible. All of the above features require huge bandwidth as a basis, and the IMT-2020 network can enable high-quality video and audio to be distributed to audiences with minimal delay.

7.2 Virtual reality (VR) panoramic video

[Virtual Reality panoramic video can be utilized for multiple multimedia applications, e.g. network video applications, digital online magazine applications, digital online newspaper applications, social media applications, on-demand video applications, mobile internet TV applications etc.] It is developed and extended on the technology of 720-degree or 360-degree panorama. It converts static panoramic pictures into dynamic video images. Panoramic videos can be viewed at any angle of 360 degrees from left to right, up and down, so that we have a truly immersive feeling. It will not be limited by time, space and region. Panoramic video is not a single static panoramic picture, but has depth of field, dynamic image and sound, etc., and also has sound and picture alignment, sound and picture synchronization. Panoramic videos show surprising effects that are beyond the reach of traditional 720-degree panoramas. Compared with the traditional 720-degree panorama, panoramic video can be said to have a huge leap in quality, quantity, form and content. But it requires very higher user experienced data rate, lower latency and higher availability and multiple information sources synchronization capabilities. With the support of IMT-2020 capable of higher bandwidth and ultra-reliability and low latency, the VR panoramic video is expected to be feasible.

7.3 Augmented reality (AR)

[Augmented reality (AR) can be utilized for multiple multimedia applications, e.g., audio-visual applications, digital online magazine applications, digital online newspaper applications, social media applications, on-demand video applications, online information distribution applications etc.] Augmented reality (AR) refers to the technology that allows the virtual world on the screen to be combined and interacted with the real-world scene through the actuarial calculation of the position and angle of the camera image and the addition of image analysis technology. AR can be utilized in multiple scenarios. For example, AR can help visualize construction projects. Computer-generated structural images can be superimposed onto a real partial view of the property prior to constructing a physical building on top of it. AR systems have been used as collaborative tools for design and planning in a build environment. For example, AR can be used to create augmented reality maps, buildings and data sources projected onto a desktop for collaborative viewing by built environment professionals. In educational settings, AR has been used to supplement standard curriculum. Text, graphics, video and audio can be overlaid into the student's real-time environment. Textbooks, flashcards, and other educational reading materials may contain embedded "markers" or triggers that, when scanned by AR devices, provide supplemental information to students presented in multimedia formats.

The combination display of virtual and real objects, motion tracking, network communication, fusion rendering and human-computer interaction are the key elements of AR which require very high use experienced data rate, higher accuracy positioning and tracing, and real-time interaction between human and computer. With the capabilities of IMT-2020 especially for uplink higher throughput capability, the AR is expected to be developed step by step.

7.4 Live streaming

Live streaming has become an important information distribution application, which has gradually become a very popular form of entertainment. The popularity of live streaming is no less than the media that traditional entertainment platforms rely on (such as TV or Radio). For instance, top tier live streamers can attract millions of viewers when they are streaming. However, unlike those produced programs, there is no fixed schedule for live streaming, which means the live streamers could start their streaming anytime, anywhere. This characteristic will trigger an explosive surge in online viewing within a geographic area or within a period of time, which brings great challenges to the network transmission of the streaming platform. Therefore, the streaming platform has to pay high operating costs to expand bandwidth and IDC capacity to cope with the peak traffic demand.

However, the massive number of equipment will cause a huge squandering of resources when the traffic drops. Moreover, the way of live streaming is evolving as well. For instance, the streaming video is evolving to ultra-high definition and immersive, which requires the higher video bit rate and stricter latency tolerance.

IMT-2020 network empowered streaming will be benefit to solve above issues. By integrating communication and information distribution application by supported IMT-2020, the super large access capacity can effectively address the concentration of viewing time and geography caused by super-influencer across the country or region, to solve the explosive traffic peak caused by influential contents. Within the ultra-high bandwidth and excellent latency performance of IMT-2020 networks, the realization difficulty of higher definition and real-time streaming will also be greatly reduced.

7.5 Live commerce

The term “Live commerce” refers to the use of live webcast technology to carry out new sales methods such as online display of merchandise, customer Q&A, and shopping guide sales through the Internet platform. Manufacturers usually use professional platforms to build their live streaming booth to sell a variety of products by streamers. Unlike traditional home shopping channel, Live commerce could interact well with customers. Furthermore, live commerce bypasses the traditional intermediate channels such as dealers so it could offer more favourable prices. It achieves seamless connectivity between merchandise and consumers, and attracts consumers' attention through the promotion of reliable quality and cost-effective products. This new way of shopping effectively reduces the cost of trust in the consumer choice process, and promotes purchase behaviour.

Considering the characteristics of live commerce, the following interaction requirements should be met:

– The live streamers can receive the feedback from the viewer and response further with supplemental product introduction, which help enhance the viewer's confidence in the quality of the product.

– The live streamers can use the virtual reality (VR) / augmented reality (AR) to provide product trials and related functions, allowing customers to “experience” the product.

– The viewer can order and pay for the products online in real-time through the payment methods of live streamers (Payment link/QR code, etc.).

– The live streamers can conduct online prize draws or mini-games, to improve viewer stickiness.

Therefore, in a live commerce, distributing low-latency and high-quality live video to a large number of viewers, and the interaction with the viewer through the network will be a strong demand. Through the integration of empowered streaming, as well as the traditional wired networks will be a good practice to meet the above requirement.

7.6 Enhanced venues for live events

With the construction of IMT-2020 networks, the concept of “Smart Venue” that empower large venues with IMT-2020 networks and provide a new viewing experience to live events or activities has emerged. In top events, new viewing methods such as multi-view point and free-view point have been introduced, bringing a different experience to the website viewer and spectator. Taking "multi-view point" as an example, by deploying multiple cameras on stadium, content service providers can allow the viewer watch the game with different viewing angles from the regular streaming.

In addition to providing a diverse experience for viewers, another core concept of "Smart Venue" is to provide spectator with live services. For instance, during a game or event, the spectator can choose a different perspective from the current seat to watch the game via the client. During the intermission, the spectator could watch the replay of highlights or behind-the-scenes.

Generally, a large event often has tens or even hundreds of thousands of spectators, and millions of viewers. Such a dense population leads to huge construction and O&M costs for communication network coverage. When only using IMT-2020 networks to provide live services to users, service providers will face the great challenges of ultra-low latency and ultra-dense connectivity.

5G empowered streaming could provide service to users in or out of the venue, which will effectively get rid of the network transmission bandwidth limitation and greatly reduce the cost of venue network construction. Meanwhile, the venue owners or event organizers will be able to leverage 5G empowered streaming to bring viewers (or spectators) an unprecedented experience.

7.7 Events Live streaming

The production of traditional large sports events is mainly completed by sportscast production systems and equipment such as sportscast vans. However, ultra-high-definition sports video has gradually become the mainstream, and the traditional sportscast system is becoming more and more inadequate in terms of transmission bit rate and delay. At the same time, the content of live events is becoming more and more complex, with a large number of cameras, microphones and sensors working at the same time, and a large amount of data need to be transmitted back to the control center. In addition, as events become more globalized, sportscast resources need to be quickly deployed to all corners of the world in a very short period of time, and traditional methods (such as sportscast van) are difficult to achieve. In fact, the event broadcast production is gradually developing towards centralized, remote and lightweight. The IMT-2020 network empowered streaming has the characteristics of low latency, high bandwidth and wide coverage, which could improve the sportscast production process, optimize resource allocation, and have the ability to quickly distribute content to viewers.

In particular, the high bandwidth and low latency transmission of 5G enables the signals of some independent camera (such as Video Adjudication System) to be backhauled through the 5G network, which is conducive to the lightweight and rapid deployment of live streaming equipment. For cameras with special mobility (such as Spidercam), the use of 5G signal transmission will expand the moving range of the cameras in the venue. In addition, the use of high-speed streaming to quickly distribute the camera signals from multi-viewing point, AR/VR signals, leaderboard and other content generated on the event.

Specifically, the following aspects could be enhanced:

– Short-distance transmission from multi-camera to live streaming system: multiple cameras (such as Steadicam, panoramic VR, AR hardware, etc.) are simultaneously transmitted to the sportscast van through IMT-2020 for live streaming production. The Single-channel high-definition video (mainly 1 920 × 1 080i) is encoded with a bit rate of 15 Mbit/s, and the uplink bit rate of the transmission channel will not be lower than this value. The relative delay between multiple cameras (same signal types) is less than 40 ms to meet synchronization requirements.

– Lightweight deployment of live streaming equipment: IMT-2020 has extremely low delay for both DL&UL transmission, multiple signals (such as video/audio/data) can be transmitted to the remote production center through wireless transmission. Therefore, the transmission of control, tally, and voice signals can ensure extremely low delay in both uplink and downlink. These devices only need to be connected to the IMT-2020 network, whereas in the past they required thousands of cables to connect to each other, which avoids the complicated deployment process.

– Multi-content streaming services: Through high-speed streaming, various content (such as multi-camera, VR, Box score, etc.) can be distributed to a large number of spectators in real time, allowing them to fully grasp the details of the event and provide a better viewing experience.

8 Capabilities of multimedia communications supported by IMT-2020 technologies

[Editor's note: The text in this chapter needs to be improved further.]

8.1 Required capabilities of multimedia communications

8.1.1 The following capabilities are

needed from IMT-2020 for supporting broadcast/multicast communications:

1. to support operation of downlink only broadcast/multicast over a specific geographic area (e.g., a cell sector, a cell or a group of cells).
2. to support operation of a downlink only broadcast/multicast system over a wide geographic area in a spectrally efficient manner for stationary and mobile UEs.
3. to enable the operator to reserve 0% to 100% of radio resources of one or more radio carriers for the delivery of broadcast/multicast content.
4. to allow the UE to receive content via a broadcast/multicast radio carrier while a concurrent data session is ongoing over another radio carrier.
5. to support broadcast/multicast of UHD streaming video (e.g., 4K/8K UHD).
6. to allow the operator to configure and broadcast multiple quality levels (i.e., video resolutions) of broadcast/multicast content for the same user service in a stand-alone IMT-2020 based broadcast/multicast system.
7. to support parallel transfer of multiple quality levels (i.e., video resolutions) of broadcast/multicast content for the same user service to the same UE taking into account e.g., UE capability, radio characteristics, application information.
8. to support parallel transfer of multiple multicast/broadcast user services to a UE.
9. to support a stand-alone multicast/broadcast network comprising of multiple cells with inter-site distances of up to 200 km.
10. to support multicast/broadcast via a 5G satellite access network, or via a combination of a 5G satellite access network and other 5G access networks.

11. to be able to setup or modify a broadcast/multicast service area within [1s].12. be able to apply QoS, priority and pre-emption to a broadcast/multicast service area. 13. to support downlink parallel transfer of the same content, via broadcast/multicast and/or unicast, such that all receiver group members in a given area receive the media at the same time according to user perception. 14. to support a mechanism to inform a media source of relevant changes in conditions in the system (e.g., capacity, failures).15. to provide means for a media source to provide QoS requirement requests to the broadcast/multicast service.16. to provide means for the broadcast/multicast service to inform the media source of the available QoS, including modification of available QoS characteristics and availability of the broadcast/multicast service.17. to support broadcast/multicast of voice, data and video group communication, allowing at least 800 concurrently operating groups per geographic area. 18. to support existing Multicast/Broadcast services (e.g., download,

streaming, group communication, TV, etc.).19. to support dynamic adjustment of the Multicast/Broadcast area based on e.g. the user distribution or service requirements.

20. to support concurrent delivery of both unicast and Multicast/Broadcast services to the users.

21. to support efficient multiplexing with unicast transmissions in at least frequency domain and time domain.

22. to support static and dynamic resource allocation between Multicast/Broadcast and unicast; the new RAT is needed to in particular allow support of up to 100% of DL resources for Multicast/Broadcast (100% meaning a dedicated MBMS carrier).

23. to cover large geographical areas up to the size of an entire country in SFN mode with network synchronization and is needed to allow cell radii of up to 100 km if required to facilitate that objective. It is needed to also support local, regional and national broadcast areas.

24. to support Multicast/Broadcast services for fixed, portable and mobile UEs. Mobility up to 250 km/h is needed to be supported.

25. to leverage usage of RAN equipment (hard- and software) including e.g., multi-antenna capabilities (e.g., MIMO) to improve Multicast/Broadcast capacity and reliability.

26. t*o support on-demand establishment of UE to UE, multicast and broadcast private communication between member UEs of the same network.*

27. to allow member UEs of local area virtual network to join an authorized multicast session.

28. to support for real time and non-real time multimedia services and applications with advanced Quality of Experience (QoE)

29. to support delivery of cached content from a content caching application via the Broadcast/Multicast service.

8.1.2 Required capabilities for real-time multimedia interaction and media content uploading

The audio-visual interaction is characterised by a human being interacting with the environment or people, or controlling a UE, and relying on audio-visual feedback which is also the characteristics of the multimedia interaction and uploading. Interactive multimedia allows the user to control, combine and manipulate a variety of media types, such as text, computer graphics audio and video materials, animation and virtual reality.

To support VR environments with low motion-to-photon capabilities, the IMT-2020 is needed to support:

– motion-to-photon[6] latency in the range of 7 ms to 15 ms while maintaining the required resolution of up to 8k giving user data rate of up to [1 Gbit/s] and

– motion-to-sound delay of [< 20 ms].

To support interactive task completion during voice conversation, the IMT-2020 is needed to support low-delay speech coding for interactive conversational services (100 ms, one-way mouth-to-ear).

---

[6] The motion-to-photon latency is defined as the latency between the physical movement of a user's head and the updated picture in the VR headset. The motion-to-sound latency is the latency between the physical movement of a user's head and updated sound waves from a head mounted speaker reaching their ears.

Due to the separate handling of the audio and video component, the IMT-2020 will have to cater for the VR audio-video synchronisation in order to avoid having a negative impact on the user experience (i.e., viewers detecting lack of synchronization). To support VR environments the IMT-2020 is needed to support audio-video synchronisation thresholds:

– in the range of [125 ms to 5 ms] for audio delayed and

– in the range of [45 ms to 5 ms] for audio advanced.

Table 7.1-1[7] gives an example on such dependences for a VR application in IMT-2020.

TABLE 7.1-1

KPI for high data rate and traffic density scenarios.

	Scenario	Experienced data rate (DL)	Experienced data rate (UL)	Area traffic capacity (DL)	Area traffic capacity (UL)	Overall user density	Activity factor	UE speed	Coverage
3	Indoor hotspot	1 Gbit/s	500 Mbit/s	15 Tbit/s/km^2	2 Tbit/s/km^2	250 000/km^2	Note 2	Pedestrians	Office and residential (Note 2) (Note 3)

NOTE 1: For users in vehicles, the UE can be connected to the network directly, or via an on-board moving base station.

NOTE 2: A certain traffic mix is assumed; only some users use services that require the highest data rates [2].

NOTE 3: For interactive audio and video services, for example, virtual meetings, the required two-way end-to-end latency (UL and DL) is 2-4 ms while the corresponding experienced data rate needs to be up to 8K 3D video [300 Mbit/s] in uplink and downlink.

NOTE 4: These values are derived based on overall user density. Detailed information can be found in [10].

NOTE 5: All the values in this table are targeted values and not strict requirements.

Table 7.1-2[8] provides required capability of AR/VR usages supported by the IMT-2020.

7 3GPP TS22.261 Table 7.1-1

8 3GPP TS22.261 Table 7.6.1-1

TABLE 7.1-2

KPI Table for AR/VR high data rate and low latency service

Use Cases	Characteristic parameter (KPI)			Influence quantity		
	Max allowed end-to-end latency	**Service bit rate: user-experienced data rate**	**Reliability**	**# of UEs**	**UE Speed**	**Service Area (Note 2)**
Cloud/Edge/Split Rendering (Note 1)	5 ms (i.e. UL+DL between UE and the interface to data network) (Note 4)	0,1 to [1] Gbit/s supporting visual content (e.g. VR based or high definition video) with 4K, 8K resolution and up to120 frames per second content.	99,99 % in uplink and 99,9 % in downlink (Note 4)	-	Stationary or Pedestrian	Countrywide
Gaming or interactive data exchanging (Note 3)	10 ms (Note 4)	0,1 to [1] Gbit/s supporting visual content (e.g. VR based or high definition video) with 4K, 8K resolution and up to 120 frames per second content.	99.99% (Note 4)	≤ [10]	Stationary or Pedestrian	20 m x 10 m; in one vehicle (up to 120 km/h) and in one train (up to 500 km/h)
Consumption of VR content via tethered VR headset (Note 6)	[5 to 10] ms (Note 5)	0.1 to [10] Gbit/s (Note 5)	[99.99%]	-	Stationary or Pedestrian	–

NOTE 1: Unless otherwise specified, all communication via wireless link is between UEs and network node (UE to network node and/or network node to UE) rather than direct wireless links (UE to UE).

NOTE 2: Length x width (x height).

NOTE 3: Communication includes direct wireless links (UE to UE).

NOTE 4: Latency and reliability KPIs can vary based on specific use case/architecture, e.g. for cloud/edge/split rendering, and can be represented by a range of values.

NOTE 5: The decoding capability in the VR headset and the encoding/decoding complexity/time of the stream will set the required bit rate and latency over the direct wireless link between the tethered VR headset and its connected UE, bit rate from 100 Mbit/s to [10] Gbit/s and latency from 5 ms to 10 ms.

NOTE 6: The performance requirement is valid for the direct wireless link between the tethered VR headset and its connected UE.

The required capabilities for Video and audio production applications supported by IMT-2020 is listed in Table 7.1-3[9].

TABLE 7.1-3

KPI of professional low-latency periodic deterministic audio transport service

Profile	# of active UEs	UE Speed	Service Area	E2E latency (Note 1)	Transfer interval (Note 1)	Packet error rate (Note 2, Note 3)	Data rate UL	Data rate DL
Music Festival	200	10 km/h	500 m × 500 m	750 µs	250 µs	10^{-6}	500 kbit/s	-

[9] 3GPP TS22.263 Table 6.2.1-1

	100	10 km/h	500 m × 500 m	750 µs	250 µs	10^{-6}	-	1 Mbit/s
Musical	30	50 km/h	50 m × 50 m	750 µs	250 µs	10^{-6}	500 kbit/s	-
	20	50 km/h	50 m × 50 m	750 µs	250 µs	10^{-6}	-	1 Mbit/s
	10	-	50 m × 50 m	750 µs	250 µs	10^{-6}	-	500 kbit/s
Semi-professional	10	5 km/h	5 m × 5 m	750 µs	250 µs	10^{-6}	100 kbit/s	-
	10	5 km/h	5 m × 5 m	750 µs	250 µs	10^{-6}	-	200 kbit/s
	2	-	5 m × 5 m	750 µs	250 µs	10^{-6}	-	100 kbit/s
AV production	20	5 km/h	30 m × 30 m	750 µs	250 µs	10^{-6}	1.5 Mbit/s	-
	10	5 km/h	30 m × 30 m	750 µs	250 µs	10^{-6}	-	3 Mbit/s
Audio Studio	30	-	10 m × 10 m	750 µs	250 µs	10^{-6}	5 Mbit/s	-
	10	5 km/h	10 m × 10 m	750 µs	250 µs	10^{-6}	-	1 Mbit/s

NOTE 1: Transfer interval refers to periodicity of the packet transfers. It has to be constant during the whole operation. The value given in the table is a typical one, however other transfer intervals are possible as long as the end-to-end latency is ≤ (1 ms – Transfer interval).

NOTE 2: Packet error rate is related to a packet size of (transfer interval × data rate). Packets that do not conform with the end-to-end latency are also accounted as error.

NOTE 3: The given requirement for a packet error rate assumes a uniform error distribution. The requirement for packet error rate is stricter if packet errors occur in bursts.

The required capabilities for low latency Video applications supported by IMT-2020 is listed in Table 7.1-4[10].

TABLE 7.1-4

KPI for low latency video

Profile	**# of active UEs**	**UE Speed**	**Service Area**	**E2E latency**	**Packet error rate (Note 1)**	**Data rate UL**	**Data rate DL**
Uncompressed UHD video	1	0 km/h	1 km^2	400 ms	10^{-10} UL 10^{-7} DL	12 Gbit/s	20 Mbit/s
Uncompressed HD video	1	0 km/h	1 km^2	400 ms	10^{-9} UL 10^{-7} DL	3 .2 Gbit/s	20 Mbit/s
Mezzanine compression UHD video	5	0 km/h	1 000 m^2	1 s	10^{-9} UL 10^{-7} DL	3 Gbit/s	20 Mbit/s
Mezzanine compression HD video	5	0 km/h	1 000 m^2	1 s	10^{-9} UL 10^{-7} DL	1 Gbit/s	20 Mbit/s

[10] 3GPP TS22.263 Table 6.2.1-3.

Tier one events UHD	5	0 km/h	1 000 m^2	1 s	10^{-9} UL 10^{-7} DL	500 Mbit/s	20 Mbit/s
Tier one events HD	5	0 km/h	1 000 m^2	1 s	10^{-8} UL 10^{-7} DL	200 Mbit/s	20 Mbit/s
Tier two events UHD	5	7 km/h	1 000 m^2	1 s	10^{-8} UL 10^{-7} DL	100 Mbit/s	20 Mbit/s
Tier two events HD	5	7 km/h	1 000 m^2	1 s	10^{-8} UL 10^{-7} DL	80 Mbit/s	20 Mbit/s
Tier three events UHD (Note 2)	5	200 km/h	1 000 m^2	1 s	10^{-7} UL 10^{-7} DL	20 Mbit/s	10 Mbit/s
Tier three events HD (Note 2)	5	200 km/h	1 000 m^2	1 s	10^{-7} UL 10^{-7} DL	10 Mbit/s	10 Mbit/s
Remote OB	5	7 km/h	1 000 m^2	6 ms	10^{-8} UL 10^{-7} DL	200 Mbit/s	20 Mbit/s
NOTE 1: Packets that do not conform with the end-to-end latency are also accounted as error. The packet error rate requirement is calculated considering 1500 B packets, and 1 packet error per hour is $10^{-5}/(3*x)$, where x is the data rate in Mbit/s. NOTE 2: Could use either professional equipment or mobile phone equipped with dedicated newsgathering app							

The required capabilities for airborne base stations to support UHD applications is listed in Table 7.1-5[11].

TABLE 7.1-5

KPI for airborne base stations support UHD.

Profile	# of active UEs	UE Speed	Service Area	E2E latency	Packet error rate (Note 1)	Data rate UL	Data rate DL
NPN ground to air UHD up Link	10	500 km/h	700 km^2 x 6 000 m (Note 2)	40 ms	10^{-8} UL 10^{-7} DL	100 Mbit/s	20 Mbit/s
NPN ground to air HD up link	10	500 km/h	700 km^2 x 6 000 m (Note 2)	40 ms	10^{-8} UL 10^{-7} DL	80 Mbit/s	20 Mbit/s
NPN air to ground UHD down Link	2	500 km/h	700 km^2 x 6 000 m (Note 2)	40 ms	10^{-7} UL 10^{-8} DL	20 Mbit/s	100 Mbit/s
NPN air to ground HD down link	2	500 km/h	700 km^2 x 6 000 m (Note 2)	40 ms	10^{-7} UL 10^{-8} DL	20 Mbit/s	80 Mbit/s
NPN radio Camera UHD	10	200 km/h	1 km^2	3 ms	10^{-8} UL 10^{-7} DL	100 Mbit/s	20 Mbit/s
NPN radio camera HD	10	200 km/h	1 km^2	3 ms	10^{-8} UL 10^{-7} DL	80 Mbit/s	20 Mbit/s

11 TS22.261 Table 6.2.1-4.

NOTE 1: Packets that do not conform with the end-to-end latency are also accounted as error. The packet error rate requirement is calculated considering 1 500 B packets, and 1 packet error per hour is $10^{-5}/(3*x)$, where x is the data rate in Mbit/s.

NOTE 2: 6 000 m = height but in a cone formation (i.e. ground coverage with a circle of diameter 30 km).

Further, with development of immersive multiple modal integrated with AR/VR, required capabilities supported by IMT-2020 is studied and listed in Table 7.1-6[12], which will greatly enrich and enhance user experience of the immersive multimedia service.

[12] 3GPP TS22.261 Table 7.11-1.

TABLE 7.1-6

KPI for Multi-modal communication service

Use Cases	Characteristic parameter (KPI)			Influence quantity			Remarks
	Max allowed end-to-end latency	**Service bit rate: user-experienced data rate**	**Reliability**	**Message size (byte)**	**UE Speed**	**Service Area**	
Immersive multi-modal VR (UL: device → application sever)	5 ms (note 2)	16 kbit/s -2 Mbit/s (without haptic compression encoding); 0.8 - 200 kbit/s (with haptic compression encoding)	99.9% (without haptic compression encoding) 99.999% (with haptic compression encoding) [40]	1 DoF: 2-8 3 DoFs: 6-24 6 DoFs: 12-48 More DoFs can be supported by the haptic device	Stationary or Pedestrian	typically < 100 km² (note 5)	Haptic feedback
	5 ms	< 1 Mbit/s	99.99% [40]	1 500	Stationary or Pedestrian	typically < 100 km² (note 5)	Sensing information e.g. position and view information generated by the VR glasses
Immersive multi-modal VR (DL: application sever → device)	10 ms (note 1)	1-100 Mbit/s	99.9% [40]	1 500	Stationary or Pedestrian	typically < 100 km² (note 5)	Video
	10 ms	5-512 kbit/s	99.9% [40]	50	Stationary or Pedestrian	typically < 100 km² (note 5)	Audio
	5 ms (note 2)	16 kbit/s - 2 Mbit/s (without haptic compression encoding); 0.8 - 200 kbit/s (with haptic compression encoding)	99.9% (without haptic compression encoding) 99.999% (with haptic compression encoding) [40]	1 DoF: 2-8 3 DoFs: 6-24 6 DoFs: 12-48	Stationary or Pedestrian	typically < 100 km² (note 5)	Haptic feedback

Use Cases	Characteristic parameter (KPI)			Influence quantity			Remarks
	Max allowed end-to-end latency	**Service bit rate: user-experienced data rate**	**Reliability**	**Message size (byte)**	**UE Speed**	**Service Area**	
Immersive multi-modal navigation applications Remote Site → Local Site (DL)	50 ms [39]	16 kbit/s -2 Mbit/s (without haptic compression encoding); 0.8 - 200 kbit/s (with haptic compression encoding)	99.999% [40]	1 DoF: 2 to 8 10 DoF: 20 to 80 100 DoF: 200 to 800	Stationary or Pedestrian	$\leq$ 100 km^2 (note 5)	Haptic feedback
	<400 ms [39]	1-100 Mbit/s	99.999% [40]	1 500	Stationary/ or Pedestrian,	$\leq$ 100 km^2 (note 5)	Video
	<150 ms [39]	5-512 kbit/s	99.9% [40]	50	Stationary or Pedestrian	$\leq$ 100 km^2 (note 5)	Audio
	<300 ms	600 Mbit/s	99.9% [40]	1 500	Stationary or Pedestrian	$\leq$ 100 km^2 (note 5)	VR
Immersive multi-modal navigation applications Local Site → Remote Site (UL)	<300 ms	12 kbit/s [26]	99.999% [40]	1 500	Stationary or Pedestrian	$\leq$ 100 km^2 (note 5)	Biometric / Affective
	<400 ms [39]	1-100 Mbit/s	99.999% [40]	1 500	Workers: Stationary/ or Pedestrian, UAV: [30-300mph]	$\leq$ 100 km^2 (note 5)	Video
	<150 ms [39]	5-512 kbit/s	99.9 % [40]	50	Stationary or Pedestrian	$\leq$ 100 km^2 (note 5)	Audio
	<300 ms	600 Mbit/s	99.9 % [40]	1 500	Stationary or Pedestrian	$\leq$ 100 km^2 (note 5)	VR

NOTE 1: Motion-to-photon delay (the time difference between the user's motion and corresponding change of the video image on display) is less than 20 ms, and the communication latency for transferring the packets of one audio-visual media is less than 10 ms, e.g. the packets corresponding to one video/audio frame are transferred to the devices within 10 ms.

NOTE 2: According to IEEE 1918.1 [40] as for haptic feedback, the latency is less than 25 ms for accurately completing haptic operations. As rendering and hardware introduce some delay, the communication delay for haptic modality can be reasonably less than 5 ms, i.e. the packets related to one haptic feedback are transferred to the devices within 10 ms.

NOTE 3: Haptic feedback is typically haptic signal, such as force level, torque level, vibration and texture.

NOTE 4: The latency requirements are expected to be satisfied even when multimodal communication for skillset sharing is via indirect network connection (i.e., relayed by one UE to network relay).

NOTE 5: In practice, the service area depends on the actual deployment. In some cases a local approach (e.g. the application servers are hosted at the network edge) is preferred in order to satisfy the requirements of low latency and high reliability.

Beyond these above required capabilities, following service capabilities are required to support multimedia real-time interaction and media content uploading.

1. The IMT-2020 enables an NPN for video, imaging and audio for professional applications.
2. The IMT-2020 network is needed to provide a time reference information to a 3rd party application acting as a master clock with an accuracy of 1 microsecond.
3. The IMT-2020 is needed to securely reconnect within a short period of time (<1s) from UE starting first network reconnection attempt after the UE has detected a UE network connection loss.
4. The IMT-2020 is needed to support uplink and downlink service continuity maintaining acceptable performance requirements while switching between co-located PLMN and NPN (e.g., due to mobility).
5. The IMT-2020 is needed to support service continuity maintaining acceptable performance requirements: for an uplink stream while performing traffic steering, switching, and splitting among co-located PLMN(s) and NPN(s); for downlink while switching between co-located PLMN and NPN.

8.2 Technical aspect of multimedia communications supported by IMT-2020

The necessary capabilities listed in above session are met by different IMT capabilities and in different technical components (radio interface, architectures enhancements, and application layer support.

8.2.1 Radio Access Network

8.2.1.1 LTE-based 5G Broadcast

3GPP technology defined in Rel-16 aims to deliver audio-visual services (including free-to-air services) in single frequency network (SFN) on stand-alone infrastructure, usually in High-Power High-Tower Single Frequency Networks to support larger inter-site distance (e.g., to allow cell radii of up to 100 km).

To allow cell radii of up to 100 km, numerology enhancement with a new 0.37 kHz subcarrier spacing and CP duration ~300 µs was introduced to support broadcast in medium power medium tower (MPMT) & HPHT. To allow reception with UE mobility up to 250 km/h, numerology enhancement with a new 2.5 kHz wider sub-carrier spacing, CP duration ~100 µs with better Doppler resiliency. There are other related enhancements to support broader coverage for the control information like dedicated reference signals (RS) accompany each numerology, enhanced subframe structure, control channel, and less dense RS pattern with reducing overheads.

8.2.1.2 Flexible and dynamic resources allocation

3GPP Rel-17 aims to support flexible and dynamic resources allocation to enable both Multicast and Broadcast support in use cases of public safety and mission critical, V2X applications, transparent IPv4/IPv6 multicast delivery, IPTV, software delivery over wireless, group communications and IoT applications, instead of broader area broadcast like services.

RAN basic functions for broadcast/multicast for UEs in RRC_CONNECTED state to be supported with better reliability and better service continuity.

For multicast communication service, the same service and the same specific content data are provided simultaneously to a dedicated set of UEs (i.e., not all UEs in the Multicast service area are authorized to receive the data). A multicast communication service is delivered to the UEs using a multicast session.

A UE can receive a multicast communication service in RRC_CONNECTED state with mechanisms such as point to point (PTP) and/or point to multi-point (PTM) delivery, with a balance between network efficiency. HARQ feedback/retransmission can be applied to both PTP and PTM transmissions.

For PTM transmission there can be two HARQ feedback schemes: UE specific feedback and NACK only feedback, depending on network decision based on level of reliability requirement or network resources. Network is able to dynamically change Multicast service delivery between PTM and PTP with service continuity for a given UE.

Basic mobility with service continuity. Unicast like mobility mechanism is designed to offer the basic mobility, e.g., the Multicast session resources are to be established along with UE's mobility in the target RAN node which supports the MBS service.

For broadcast communication service, the same service and the same specific content data are provided simultaneously to all UEs in a geographical area (i.e., all UEs in the Broadcast service area are authorized to receive the data). A broadcast communication service is delivered to the UEs using a broadcast session. A UE can receive a broadcast communication service in RRC_IDLE, RRC_INACTIVE and RRC_CONNECTED state.

There are other enhancement including, UE are able to receive MBS with simultaneous operation in unicast reception; enhancement on RAN network interface like Xn, F1, and E1 interfaces; support on dynamic broadcast area.

8.2.1.3 UL enhancement

A) UL MIMO

While UL MIMO offers the capability for reduction in overhead and/or latency, high-speed vehicular scenarios (e.g., a UE traveling at high speed on highways) at FR2 require more aggressive reduction in latency and overhead – not only for intra-cell, but also for L1/L2 centric inter-cell mobility. This also includes reducing the occurrence of beam failure events. Besides of it, enhancements for enabling panel-specific UL beam selection were investigated and specified continuously. This offers some potential for increasing UL coverage including. Then, channels other than PDSCH can benefit from multi-TRP transmission (as well as multi-panel reception) which also includes multi-TRP for inter-cell operations. This includes some new use cases such as UL dense deployment within a macro-cell and/or heterogeneous-network-type deployment scenarios. And, the SRS can be further enhanced for capacity and coverage.

B) UL Carrier Aggregation

Carrier aggregation was developed the unaligned frame boundary capability to provide more flexible beginning TX frame structure configurations among different carriers to offer higher uplink throughput and lower latency. Tx switching is to specify the dynamic switching mechanisms among two uplink carriers. Unaligned frame boundary and Tx switching can be implemented together to achieve larger uplink throughput for TDD CA operation. In one cell group, it supports PUCCH carrier switching semi-statically or dynamically. This could reduce the latency of PUCCH transmissions significantly for CA operation.

C) UL Dual connectivity (DC)

Dual connectivity is capable of Uplink power control i.e., to limit UE's transmission power to assure edge user's communication QoS, migration interference among the users, and to reduce energy consumption of the UE. Uplink DC is also capable of earlier measurement and fast recovery from MCG failure to reduce latency. Beyond these, it is capable of to indicate UE entering the third state to maintain context of the UE which will reduce configuration overhead.

D) Interference coordination

Interference coordination includes two aspects, i.e., cross-link interference (CLI) and remote interference management (RIM). For CLI, it is left to network implementation for gNB-gNB interference. The UE-UE interference coordination is capable of such as SRS-RSRP/CLI-RSSI based layer-3 CLI measurement and reporting, and network coordination mechanism for CLI with inter-gNB exchange of intended UL/DL configuration. RIM targets to migrate the interference occurring in specific weather conditions with the distance between aggressor gNB and victim gNB hundreds of kilometres. The RIM reference signal (RS) based on PRACH preamble-like RS is introduced for better interference measurement while the detailed remote interference mitigation mechanisms are left to implementation.

8.2.2 Architecture Enhancements

The MBS architecture follows the IMT-2020 architectural principles as defined in 3GPP TS 23.501, enabling distribution of the MBS data from the 5GS ingress to NG-RAN node(s) and then to the UE. The MBS architecture provides:

– Efficient usage of RAN and CN resources, with an emphasis on radio interface efficiency;

– Efficient transport for a variety of multicast and broadcast services.

MBS traffic is delivered from a single data source (e.g., Application Service Provider) to multiple UEs. Depending on many factors, there are several delivery methods which may be used to deliver the MBS traffic in the 5GS.

The 5G MBS also provides functionalities such as local MBS service, authorization of multicast MBS and QoS differentiation.

Between 5GC and NG-RAN, there are two possible delivery methods to transmit the MBS data:

– 5GC Individual MBS traffic delivery method: This method is only applied for multicast MBS session. 5GC receives a single copy of MBS data packets and delivers separate copies of those MBS data packets to individual UEs via per-UE PDU sessions, hence for each such UE one PDU session is required to be associated with a multicast session.

– 5GC Shared MBS traffic delivery method: This method is applied for both broadcast and multicast MBS session. 5GC receives a single copy of MBS data packets and delivers a single copy of those MBS packets to an NG-RAN node, which then delivers the packets to one or multiple UEs.

The 5GC Shared MBS traffic delivery method is required in all 5G MBS deployments. The 5GC Individual MBS traffic delivery method is required to enable mobility when there is an NG-RAN deployment with non-homogeneous support of 5G MBS.

For the multicast session, a single copy of MBS data packets received by the CN may be delivered via 5GC Individual MBS traffic delivery method for some UE(s) and via 5GC Shared MBS traffic delivery method for other UEs.

New network functionality are introduced to support MBS in 5GS:

– MB-SMF, Supporting MBS session management (including QoS control).

– MB-UPF, user plane function for MBS with Packet filtering of incoming downlink packets for multicast and broadcast flows, and QoS enforcement.

Other network functions are enhanced, e.g., SMF to support UE join/leave operation of Multicast session, AMF to support group paging, and signaling routing, PCF to provide policy information and QoS handling for MBS session.

The integration of XR applications within the IMT-2020 System is approached following the model of 3GPP 5G Media Streaming as defined in 3GPP TS 26.501 [x]. Assume a 3GPP 5G-XR Application Provider being an XR Application provider that makes use of 5G System functionalities for its services. For this purpose, it provides a 5G-XR Aware Application on the UE to make use of a 3GPP 5G-XR client and network functions using network interfaces.

The QoS model is described in clause 5.7 of 3GPP TS 23.501 [x]. The 5G QoS model supports both:

– QoS Flows that require guaranteed flow bit rate (GBR QoS Flows)

– and QoS Flows that do not require guaranteed flow bit rate (Non-GBR QoS Flows).

The QoS model also supports Reflective QoS (see clause 5.7.5 of 3GPP TS 23.501 [8]).

A QoS Flow ID (QFI) is used to identify a QoS Flow in the IMT-2020. User Plane traffic assigned to the same QoS Flow within a Protocol Data Unit (PDU) Session receives the same traffic forwarding treatment (e.g., scheduling, admission threshold).

The QFI may be dynamically assigned or may be equal to the 3GPP 5G QoS Identifier (5QI). A QoS Flow may either be 'GBR', 'Non-GBR' or “Delay Tolerant GBR” depending on its QoS profile and it contains QoS parameters as follows:

– For each QoS Flow, the QoS profile includes the QoS parameters:
  - 3GPP 5G QoS Identifier (5QI); and
  - Allocation and Retention Priority (ARP).

– For each Non-GBR QoS Flow only, the QoS profile can also include the QoS parameter:
  - Reflective QoS Attribute (RQA).

– For each GBR QoS Flow only, the QoS profile also include the QoS parameters:
  - Guaranteed Flow Bit Rate (GFBR) - uplink (UL) and downlink (DL); and
  - Maximum Flow Bit Rate (MFBR) - UL and DL; and

– In the case of a GBR QoS Flow only, the QoS profile can also include one or more of the QoS parameters:
  - Notification control;
  - Maximum Packet Loss Rate - UL and DL

In Release 18, PDU Set concept is introduced in FS_XRM study to optimize the delivery of XRM service in 5GS. A PDU Set is composed of one or more PDUs carrying the payload of one unit of information generated at the application level (e.g., a frame or video slice for XRM Services), which are of same importance at application layer. All PDUs in a PDU Set are needed by the application layer to use the corresponding unit of information. In some cases, the application layer can still recover parts of the information unit, when some PDUs are missing.

The following key issues are currently under investigating:

WT#1: Enhancements for supporting multi-modality service: Study whether and how to enable delivery of related tactile and multi-modal data (e.g., audio, video and haptic data related to a specific time) with an application to the user at a similar time, focusing on the need for policy control enhancements (e.g., QoS policy coordination).

WT#2: Enhancements of network exposure to support interaction between 5GS and application:

– WT#2.1: Study whether and how interaction between AF and 5GS is needed for application synchronization and QoS policy coordination among multiple UEs or between multiple QoS flows per UE.

– WT#2.2: Study exposure of 5GS QoS information (e.g., QoS capabilities) and network conditions to the Application to enable quick codec/rate adaptation help to provide desired QoE (e.g., such as assist in alleviating 5GS congestion).

WT#3: Study whether and how the following QoS and policy enhancements for XR service and media service transmission are performed:

– WT#3.1: Study the traffic characteristics of media service enabling improved network resources usage and QoE.
– WT#3.2: Enhance QoS framework to support media units granularity (e.g., video/audio frame/tile, Application Data Unit, control information), where media units consist of PDUs that have the same QoS requirements.
– WT#3.3: Support differentiated QoS handling considering different importance of media units. e.g., eligible drop packets belong to less important media units to reduce the resource wasting.
– WT#3.4: Whether and how to support uplink-downlink transmission coordination to meet RTT (Round-Trip Time) latency requirements between UE and N6 termination point at the UPF.
– WT#3.5: Potential policy enhancements to minimize the jitter, focusing on i.e. requirement provisioning from AF, extension of PCC rule.

WT#4: Study potential enhancements of Mobility and power management considering traffic pattern of media services:

– WT 4.1: void
– WT 4.2: Power saving enhancement e.g. support trade-off of throughput/latency/reliability considering device battery life, whether and how to enhance CDRX, considering XR/media traffic pattern.

9 Case study

[Editor's note: This section provides case studies from various countries associated with applications for multimedia content *in various usage scenarios supported by eMBB, mMTC and URLLC of terrestrial IMT-2020 systems]*

10 Summary

[Editor's note: To be described.]

This contribution document is aimed at summarising the various capabilities of terrestrial component of IMT-2020 for Multimedia.

Multimedia is an immersive technological way of presenting information that combines audio, video, images, and animations with textual data. Multimedia applications include network video, digital magazine, digital newspaper, digital radio, social media, touch media, etc., and can be easily enabled using IMT-2020.

New emerging technologies such as Virtual Reality (VR) and Augmented Reality (AR) are becoming key technologies to upgrade or even subvert traditional industries. Since there is an endless field of applications which include video games, event live streaming, video entertainment, healthcare, real estate, retail and public safety. It is through this contribution the vast capabilities of IMT-2020 regarding multimedia communication have been summarised.