GTI 5G Metrics and Test Methods Towards XR White Paper

GTI 5G Metrics and Test Methods Towards XR White Paper

Version:	v_1.0.0
Deliverable Type	□ **Procedural Document** ☑ **Working Document**
Confidential Level	□ **Open to GTI Operator Members** □ **Open to GTI Partners** ☑ **Open to Public**
Program	5G Technology and Product
Working Group	
Project	Technology Evolution
Task	5G-A Technology
Source members	China Mobile, HUAWEI, ZTE
Support members	Qualcomm, ASR, Unisoc, Airtel
Editor	Hongbiao Zhang(CMCC), Luting Kong(CMCC), Lei Cao (CMCC), Tianming Jiang(CMCC), Shijun Zhang(CMCC), Meng Li(CMCC), Ruofei Ma(HUAWEI), Yi Li(HUAWEI),Yi Qin(HUAWEI),Yu Fu(HUAWEI), Na Li(ZTE), Xintai Wang(ZTE), Runbin Liu(ZTE), Kaifeng Lu(ZTE)
Last Edit Date	2023-09-01
Approval Date	

Document History

Date	Meeting #	Version #	Revision Contents
2023.9.6	GTI#38	V1.0	Initial Version

Table of Contents

Executive Summary

Extended Reality (XR), as the primary carrier of the early Metaverse, is changing people's way of life and production. 5G networks with large bandwidth, low latency, and high reliability can help the XR industry thrive. However, the immersive experience of XR have higher requirements for communication networks. This report aims to promote the formation of service quality evaluation metrics and 5G performance metrics for XR in the industry, propose requirements of equipment capability and develop testing methods for typical XR service, provide objective metrics of quality evaluation and a unified testing methods for the key technology to guarantee user experience of XR service in 5G networks.

Abbreviations

Abbreviation	Explanation
XR	Extended Reality
VR	Virtual Reality
AR	Augmented Reality
MTP	Motion to Photons
FPS	Frame Per Second
ATW	Asynchronous Timewrp
FOV	Field of View
KQI	Key Quality Indicators
KPI	Key Performance Indicator
QoS	Quality of Service
HMD	Head Mounted Display
PDCP	Packet Data Convergence Protocol
MAC	Medium Access Control
TB	Transport Block
UE	User Equipment
USB	Universal Serial Bus
SSB	SS/PBCH Block
RSRP	Reference Signal Receiving Power
SINR	Signal to Interference plus Noise Ratio
GoP	Group of picture
MTHR	Motion to High Resolution
5GC	5G Core
ARQ	Automatic Repeat reQust
IMAX	Image MAXimum
FEC	Forward Error Correction
E2E	End to End
IP	Internet Protocol
RAN	Radio Access Network
GPU	graphics processing unit
VOD	Video on Demand
gNodeB	the Next Generation Node B

Introduction

XR, as the primary carrier of the early Metaverse, is bringing revolutionary changes and innovative experiences to people's production and lifestyle in fields such as gaming, social media, shopping, and cultural tourism, becoming an important carrier for stimulating data traffic and driving consumption upgrading. The market is accelerating to embrace the emerging trend of the XR industry, with a good development trend. IDC believes that from a long-term perspective, the global shipment volume of VR/AR displays continues to grow significantly, reaching 76.7 million units by 2024, with a compound annual growth rate of 81.5%[1]. At the same time, 5G networks with large bandwidth, low latency, and high reliability can help the XR industry thrive.

The XR service has the characteristics of frame transmission, such as the 60 FPS XR cloud rendering service, which needs to transmit one frame in every 16.6ms. Although the head display terminal has the ATW algorithm, if the network transmission cannot be completed in a timely manner, the head display can predict and generate the next frame of image based on the previous frame, which can reduced the requirement of the MTP delay to 70ms. It still has high requirements for network delay, throughput, reliability, and other performance, Further research on key technology solutions is needed for the 5G network.

However, before research on key technology solutions, it is necessary to clarify and define the objective metrics of 5G performance and service quality evaluation for user experience. In the field testing, the user experience mainly refers to the subjective feelings and subjective records of the testers, such as stalling duration, artifact, black borders . There are personal factors that lead to biased results and are difficult to compare the result from other field testing. In addition, there are differences in the statistical methods used by various XR service platforms for network performance.It is necessary to unify the definition, data acquisition, measurement reporting, and statistical methods of frame latency, frame throughput, and frame reliability for XR services, and establish mapping relationships with quantified subjective experiences for problem finding.

This report aims to promote the formation of service quality evaluation metrics for XR service in the industry, define metrics of 5G performance, form a statistical method for the metrics of 5G performance, explore requirements of equipment capability and test methods for typical XR service. In the chapter 1 of this report, an overview and definition of service quality evaluation metrics towards XR are provided. In the chapter 2 of this report, an overview and definition of 5G metrics towards XR are provided. In the chapter 3 of this report, the requirements of equipment capability and testing methods toward XR service are elaborated.

This report is jointly written by China Mobile, Huawei and ZTE.

1 Metrics of Service Quality Evaluation Towards XR

Service quality indicators are formulated to evaluate network quality by considering technical factors that affect user watching experience during typical XR services.

XR services include panoramic/field of view (FOV) videos, IMAX movies, interactive games, and interactive education. This report studies service quality evaluation methods and model based on XR videos and games, and can also apply to services with similar features. For example, the experience model for interactive games is also applicable to interactive education.

1.1 Overview of Service Metrics

Typical XR services are evaluated based the VR MOS model and key quality indicators (KQIs), including presentation quality indicators (such as multi-modal synchronization, stalling, artifacts, and black borders), interaction quality indicators (such as interaction response and MTP latency), and source quality indicators (such as bit rate, frame rate, resolution, and FOV). The XR service experience quality grading standards are formulated based on the quantitative research on experience satisfaction and the test results of experience satisfaction on the 5G live network.

As different network qualities have differentiated impacts on user experience, certain key evaluation indicators are specified. These indicators include different service KQIs and aim to analyze the possible impacts on mobile network quality key performance indicators (KPIs) from the service perspective. The service KQIs consist of three types of user-perceivable, measurable KQIs that can reflect network quality of service (QoS): source quality indicators, presentation quality indicators, and interaction quality indicators. The following figure shows the service quality evaluation framework.

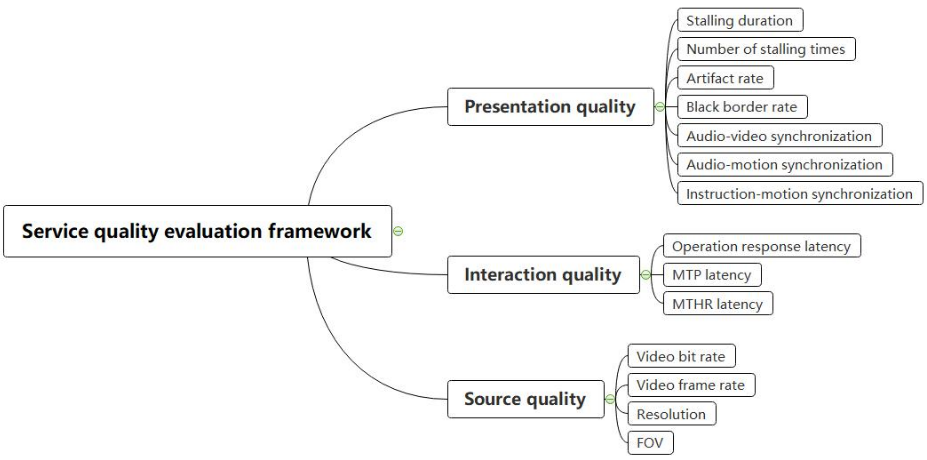

Figure 1 The Framework of Service quality evaluation

1.2 Definition of Service Metrics

1.2.1 Presentation Quality

（1） Stalling duration

This indicator refers to the total inter-frame interval (in the unit of millisecond) of all stalling images during a single test. The terminal records the millisecond timestamp of each video frame arrival, decoding, or rendering. If the interval between two consecutive frames exceeds a specified threshold, a stalling event is recorded. When the inter-frame interval is restored, stalling recovery is recorded. The video frame timestamp when stalling occurs is used as the stalling start time ts_x, and the video frame timestamp when stalling recovers is used as the stalling recovery time te_x. Then, the total stalling duration t_{total} can be calculated as follows:

$$t_{total} = \sum_{x} (ts_x - te_x)$$

In the formula:

ts_x : time when the x(th) image stalling occurs

te_x: time when the image recovers from the x(th) stalling

t_{total} : total stalling duration

（2） Number of stalling times

This indicator refers to the number of image stalling (waiting for buffering) times during a single test.

（3） Artifact rate

Data losses or forward error correction (FEC) failures in VR video data transmission cause image distortion, mosaic, or other artifacts. The artifact rate is calculated by summing up the time-based weighted pPLR(t), which represents the data loss rate in second t in a single test. Weights are determined according to positions in a media session and are attenuated according to the forgetting curve.

（4） Black border rate

This indicator refers to the proportion (%) of black borders that occur in images from secondary rendering when users' heads rotate or move. The direction quadruple OrientQuat_HMD after each HMD movement and the direction quadruple OrientQuat_Graph of the currently rendered image are recorded. The angle difference along the x-axis, which is represented by Degree, between OrientQuat_HMD and OrientQuat_Graph is then obtained. This difference (Degree) indicates the black border angle on the horizontal FOV. It is then divided by the HMD FOV, the result of which equals the single-time black border rate. The second-level or global average black border rate can also be calculated as required.

（5） Audio-video synchronization

This indicator refers to the synchronization latency (in the unit of millisecond) between XR video and audio. In a single test, the terminal records in each frame the millisecond timestamp t_v when the video frame arrives or is rendered and the millisecond timestamp t_a when the corresponding audio frame arrives or is played. The audio-video synchronization latency of this frame is calculated as follows: $t_v - t_a$ (ms). The second-level or global average value can also be calculated as required.

（6） Audio-motion synchronization

This indicator refers to the synchronization latency (in the unit of millisecond) between motion/mouth/expression images and audio of the digital person in the virtual space. In a single test, the terminal records in each frame the millisecond timestamp t_v when the audio frame arrives and the millisecond timestamp t_a when the corresponding motion image arrives or is rendered. The audio-motion synchronization latency of this frame is calculated as follows: $t_v - t_a$ (ms). The second-level or global average value can also be calculated as required.

（7） Instruction-motion synchronization

This indicator refers to the synchronization latency (in the unit of millisecond) between instructions from users and motions of the digital person. In a single test, the terminal records in each frame the millisecond timestamp t_m when an instruction is provided and the millisecond timestamp t_v when the corresponding motion image arrives or is rendered. The instruction-motion synchronization latency of this frame is calculated as follows: $t_v - t_m$ (ms). The second-level or global average value can also be calculated as required.

1.2.2 Interaction Quality

（1） Operation response latency

This indicator refers to the latency (in the unit of millisecond) between the time when a user performs an operation and the time when the user sees the response to the operation. In a single test, the terminal records in each frame the millisecond timestamp t_m when an instruction is provided and the millisecond timestamp t_v when the expected image arrives or is rendered. The operation response latency of this frame is calculated as follows: $t_v - t_m$ (ms). The second-level or global average value can also be calculated as required.

（2） MTP latency

This indicator refers to the lapse between a head rotation and an image refresh. For cloud XR services, there are complex real-time interactions between users and cloud applications. After performing calculation, rendering, compression, and encoding on an interaction instruction, cloud applications send response images as video streams to terminals for decoding and display. The E2E service latency is also called MTP latency (in the unit of millisecond). During a single test, the terminal records the millisecond timestamp (t_m) for the HMD motion collection time after each motion is complete and the millisecond timestamp

(t_v) for the corresponding motion image arrival or rendering time. The difference between a single instruction and a motion synchronization is calculated as follows: $t_v - t_m$ (ms). A second-level or global average value can be calculated as required.

（3） MTHR latency

For multiple region-dependent VR streams, a VR application may use posture data to determine a new viewport, corresponding to the high-quality video stream used in the new viewport region and the low-quality video stream used in other background regions. When the user's head rotates, they may experience a transition from low-quality, low-resolution viewport to high-resolution,high-quality viewport display, and the switching time is defined as the MTHR (Motion to High Resolution, MTHR) . During a single test, the terminal records the millisecond timestamp t_L for a low-quality, low-resolution viewport and the millisecond timestamp t_H for the corresponding high-quality, high-resolution viewport. The MTHR latency is calculated as follows: $t_H - t_L$(ms). A second-level or global average MTHR latency can be calculated as required.

1.2.3 Source Quality

（1） Video bit rate

This indicator refers to the video data volume sampled per unit time (in the unit of bit/s) during VR video encoding. During a single test, record the size of each frame of an XR video, the total size M (in the unit of Mb or Kb) after the test or video playback is complete, and the test or video playback duration t (in the unit of s). Then, the average video bit rate of a single test is M/t (in the unit of Mbit/s or Kbit/s).

（2） Video frame rate

This indicator refers to the number of frames per second (FPS) during VR video encoding. During a single test, record the number of frames rendered or played by an XR video per second as a second-level frame rate, as well as the total frame rate M and duration t (in the unit of s) of the test or video playback. Then, the average frame rate is M/t (in the unit of FPS).

（3） Resolution

This indicator refers to the number of image pixels during VR video encoding. During a test, the number of pixels (H x W) of the video frame rendered or played by the XR per second is recorded. In the adaptive resolution scenario, the average resolution can be calculated by second.

（4） FOV

This indicator refers to the horizontal monocular FOV of the VR HMD. The value of the horizontal monocular FOV is read through the system or physical interface.

2 Metrics of 5G Performance Towards XR

2.1 Overview of 5G Metrics

Considering the characteristics of frame transmission, it is necessary to redefine the metrics of 5G performance for XR immersion experience. Key metrics can be described in terms of frame delay, frame throughput, frame reliability, etc. This chapter will define the key metrics of 5G performance towards XR and form a calculation method.

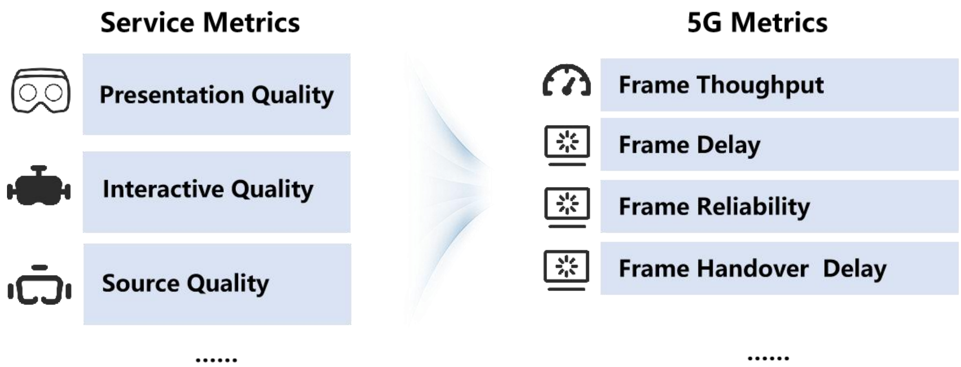

Figure 2 The System Framework of 5G metrics

2.2 Definition of 5G Metrics

2.2.1. Fame Level Metrics

At present, the encoding methods for XR services are mainly based on Group of Picture (GoP) encoding and Slice-based encoding. When the server transmits each frame data to the network, it presents a periodic feature of frame transmission, Taking the 4K resolution 60fps XR e-sports service as an example, the code rate is about 35 to 40 Mbps, and each frame generates several IP packets. All IP packets are transmitted to the base station with pattern of burst in a short period of time. The statistical granularity of traditional 5G performance, such as latency, throughput, and reliability, is mostly at the packet level, making it difficult to fully characterize the guarantee performance of 5G by expanding the statistical granularity of key network bearer performance indicators that affect business experience based on the frame level characteristics of XR services. The statistical granularity of network bearer indicators such as latency, throughput, and reliability is mostly at the packet level, making it difficult to fully characterize the performance which can describe the capacity of network transmission for XR services. Therefore, it is necessary to further establish a frame level network metrics, expanding the statistical granularity of the metrics to all data packets in a frame, changing the packet level statistical granularity to frame, which all data packets in a frame are measured together.

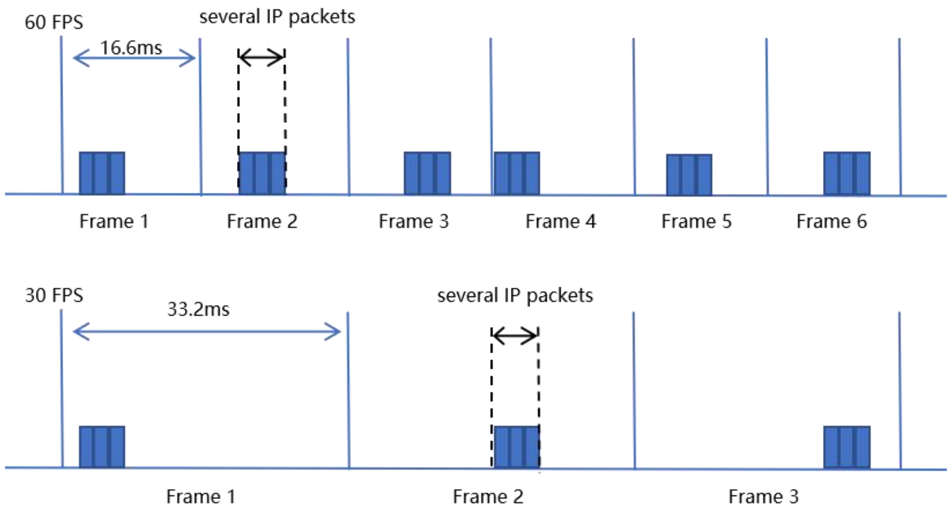

Figure 3 diagram of XR service frame characteristics

2.2.2. I frame and P frame

The GOP group contains one I frame and multiple P and B frames, and the Slice-base encoding contains one I frame and multiple P frames. I-stream belongs to the basic stream, which is encoded and decoded independently, The image content of I-stream can be decoded without relying on other image streams. P-stream is a predictive stream or an enhanced stream, and the encoding of the stream depends on the before and after image streams in the I-stream or P-stream. The data size of I-stream is larger than the P-stream because it is encoded separately. Usually, in the process of network transmission, the importance of I frames is higher than that of P frames. The network should have the ability to distinguish between I frames and P frames for service recognition, and separately calculate the network performance of I frames and P frames.

2.2.3. Frame Delay

Delay is one of the important indicators for measuring network transmission capacity and also a key indicator affecting XR user experience. The current statistical method for delay does not take into account the frame characteristics of XR service, and there are differences in the definition and measurement methods of delay statistics among different XR platforms, and terminals, which cannot be objectively compared the performance with each other. Therefore, it is necessary to establish a unified definition and statistical method for frame delay.

2.2.3.1. Frame Delay of RAN

Frame delay of RAN downlink refers to the time difference between the base station receiving the first downlink data packet of each frame and the terminal receiving the last downlink data packet. Frame delay of RAN uplink refers to the time difference between the terminal sending the first uplink packet of the frame and the base station transmitting the last uplink packet of the frame. The indicators can be calculated and obtained through the following key measurement values, which can be globally averaged as needed.

During downlink transmission, the time when the base station receives the first downlink packet of the frame at the PDCP layer is recorded as T1, and the time when the terminal receives the last downlink packet of the frame is recorded as T2; During uplink transmission, the time when the terminal sends the first instruction packet of the frame is recorded as T3, and the time when the base station completes the transmission of the last uplink packet of the frame is recorded as T4.

$$\text{Frame delay of RAN downlink} = T_2 - T_1$$

$$\text{Frame delay of RAN uplink} = T_4 - T_3$$

2.2.3.2. Frame Delay of Network Transmission

The frame delay of network downlink transmission refers to the time difference between the time when the server transmits the first downlink packet of each frame and the terminal receives the last downlink packet of the frame. The frame delay of network uplink transmission refers to the time difference between the terminal sending the first uplink instruction packet of the frame and the service server receiving the last uplink packet of the frame. The indicators can be calculated and obtained through the following key measurement values, which can be globally averaged as needed.

During downlink transmission, when the server transmits the first downlink packet of the frame to the 5GC is recorded as T5; During uplink transmission, the time when the server receives the last uplink packet of the frame is recorded as T6.

$$\text{frame delay of network downlink transmission} = T_2 - T_5$$

$$\text{frame delay of network uplink transmission} = T_6 - T_3$$

2.2.3.3. Frame Delay Jitter

The network transmission process is often affected by the channel environment, which leads to network transmission delay jitter within a range, resulting in inconsistent XR user experience. Therefore, it is necessary to clearly obtain the frame delay jitter to ensure that users can obtain a deterministic experience. Frame uplink and downlink delay jitter refers to the average delay difference between the current frame and the previous frame within a certain period of time.

$$\overline{D} = \frac{\sum_{i=1}^{N} |D_N - D_{N-1}|}{N}$$

2.2.3.4. Frame Handover Delay

During the cell handover process, the interruption delay of the user interface is generally 40-60ms, which can cause 3-4 frames of XR service to be lost. The traditional handover delay of user interface is referred to the time different between the terminal receiving switching commands and resuming transmission of data packets. However, it should be considered that XR service needs to complete the transmission of all data packets of a frame before displaying a complete frame of the image. Therefore, it is necessary to define a frame handover delay for XR service.

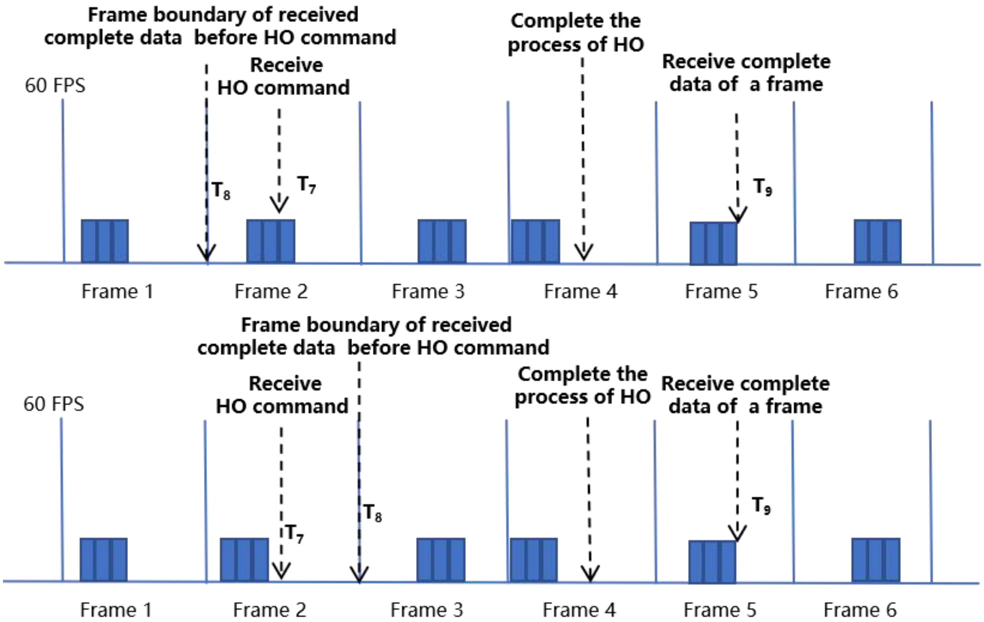

Figure 4 diagram of handover for XR terminal

Frame handover delay refers to the time difference between the frame boundary of received complete data before HO command and the terminal receiving the all data packet of a frame after handover to the target cell. The indicators can be calculated and obtained through the following key measurement values, which can be globally averaged as needed.

The time when terminal receives the handover command is recorded as T7, and before the handover, the time of frame boundary when the terminal fully receives the all data packet of a frame is recorded as T8. After the handover is completed, the time when the terminal fully receives the all data packet of a frame is recorded as T9.

$$\text{Frame handover delay} = \mathrm{T}_9 - \mathrm{T}_8$$

2.2.4. Frame Throughput

The transmission throughput is also one of the important indicators to measure the network transmission capacity. Due to the increase of packet header information by each network element during the data transmission process, the size of data packets varies at each transmission stage. It is also necessary to establish an objective and unified metrics definition and statistical method considering the frame characteristics of XR services.

2.2.4.1. Frame Throughput of RAN

Frame throughput of RAN downlink refers to the ratio of the total size of all downlink data packets transmitted by the base station for each frame to the frame delay of RAN downlink. Frame throughput of RAN uplink refers to the ratio of the total size of all uplink data packets transmitted by the terminal for each frame to the frame delay of RAN uplink. The indicators can be calculated and obtained through the following key measurement values, which can be globally averaged as needed.

The sum of all packet sizes of the frame sent by the base station is recorded as M1, and the sum of all packet sizes sent by the terminal to the base station is recorded as M2.

$$\text{Frame throughput of RAN downlink} = \frac{M_1}{T_2 - T_1}$$

$$\text{Frame throughput of RAN uplink} = \frac{M_2}{T_4 - T_3}$$

2.2.4.2. Frame Throughput of Network Transmission

Frame throughput of network downlink transmission refers to the ratio of the total size of all downlink data packets sent by the server to the core network and the delay of the 5GC receiving the all downlink data of a frame. Frame throughput of network uplink transmission refers to the ratio of the total packet size of the uplink data sent by the 5GC to the service server and the delay of the server receiving the all uplink data of a frame. The indicators can be calculated and obtained through the following key measurement values, which can be globally averaged as needed.

The sum of all downlink packet sizes in a frame sent by the server is recorded as M3, while the sum of all uplink packet sizes sent by the 5GC is recorded as M4. The time when the service sends the first downlink data packet of the frame is recorded as T10, the time when the 5GC receives the last downlink data packet of the frame is recorded as T11, the time when the 5GC sends the first uplink data packet of the frame is recorded as T12, and the time when the server receives the last data uplink packet of the frame is recorded as T13

$$\text{Frame throughput of network downlink transmission} = \frac{M_3}{T_{11} - T_{10}}$$

$$\text{Frame throughput of network uplink transmission} = \frac{M_4}{T_{13} - T_{12}}$$

2.2.5. Frame Reliability

In traditional metrics of 5G performance, reliability is usually only counted at the granularity of packet error rate and packet loss rate. Due to the requirement of XR service that all data packets of a frame must complete correct transmission in order to obtain the data of the frame, reliability for XR service need to be counted at the frame level granularity.

2.2.5.1. Frame Reliability of Downlink and Uplink

Frame reliability of downlink and uplink refers to the success rate of transmitting the downlink and uplink data in a frame within a certain period of time. The indicators can be calculated and obtained through the following key measurement values, which can be globally averaged as needed.

If there is a packet error or packet loss in a frame within a certain period of time and the ARQ is not completed within the specified time, it can be recorded as a frame loss. During downlink transmission, the total number of downlink frame transmitted lost is recorded as E1, and the total number of downlink frames is recorded as E3; During uplink transmission, the total number of uplink frame transmitted lost is recorded as is recorded as E2, and the total number of uplink frames is recorded as E4.

$$\text{Frame reliability of downlink} = 1 - \frac{E_1}{E_3}$$

$$\text{Frame reliability of uplink} = 1 - \frac{E_2}{E_4}$$

3 Equipment Capability & Test Methods Towards XR

3.1 Equipment Capability Requirements

3.1.1. Test System Architecture

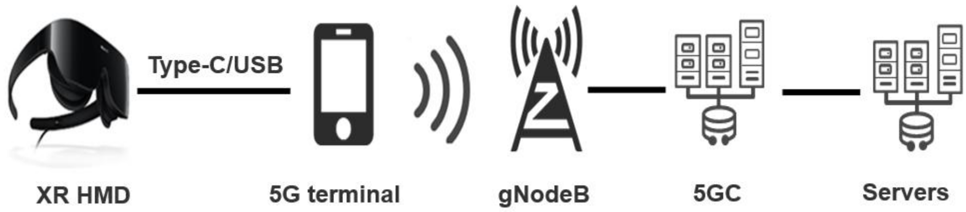

Figure 5 Structure of the test tool and its relationship with the tested network

The test system consists of the terminal-side dialing test and monitoring tool and the platform-side service system. The terminal-side tool is installed in the HMD (all-in-one device) or in the terminal (separate device) connected to the HMD. The platform-side system is installed on the server behind the core network.

During the test, the terminal-side tool client sends an XR service request to the platform side. The platform-side service system receives the instruction and streams the video and game images to the client in streaming media mode. The terminal-side tool collects the indicators of the XR services.

3.1.2. Requirements for Server

In terms of hardware, server needs to have mainstream processors, GPUs, and storage devices, operating systems, and high-bandwidth communication capabilities. In terms of software, XR VOD push service capability and XR video sources with different mainstream resolutions such as 4K and 8K. Mainstream-format-encoding and transmission capabilities for running XR games and pushing video streams to terminals. According to the metrics of 5G performance defined in Chapter 2, server also needs to have the following capabilities:

1. The ability to record the receiving and sending timestamps of data packet in each frame, including the timestamps of receiving uplink data packet in each frame and the timestamps of sending downlink data packet in each frame.

2. The ability to record the size of data packet in each frame, including the size of the uplink data packet in each frame, and the size of the downlink data packet in each frame.

3.1.3. Requirements for Test Terminals

The terminal needs to have the capabilities of supporting XR video playback and game image rendering, recording the attributes of the current video and game image, such as the resolution, frame rate, and bit rate, by second, recording the uplink instruction sending timestamp and downlink audio and video receiving timestamp of each frame, recording the quadruplet information in the direction of the XR HMD, recording basic information about XR HMDs, such as the screen resolution and FOV. According to the metrics of 5G performance defined in Chapter 2, test terminals also needs to have the following capabilities:

1. The ability to record the receiving and sending timestamps of data packet in each frame, including the timestamps of sending uplink data packet in each frame and the timestamps of receiving downlink data packet in each frame.

2. The ability to record the size of data packet in each frame, including the size of the uplink data packet in each frame, and the size of the downlink data packet in each frame.

3.1.4 Requirements for Network Equipment

For the requirements of immersive experience in XR service, the network should have higher performance such as large bandwidth, low latency, and high reliability. According to the metrics of 5G performance defined in Chapter 2, network equipment also needs to have the following capabilities:

1. The ability to record the receiving and sending timestamps of data packet in each frame, including the timestamp of the base station receiving and sending data of downlink and uplink in each frame, the timestamp of the 5GC receiving and sending data of downlink and uplink in each frame.

2. The ability to record the size of data packet in each frame, including the size of the base station receiving and sending data of downlink and uplink in each frame, the size of the 5GC receiving and sending data of downlink and uplink in each frame.

3. The ability to record frame loss during the transmission.

4. The ability to recognize and perceive XR services, measure the transmitted data packets according to the XR frame characteristics based on the service recognition ability.

3.2 Test Requirement

3.2.1. Scenario

The mobile network quality test shall be performed in different network scenarios such as urban areas, counties, towns, and suburban areas. These network scenarios shall include dense buildings, office areas, streets, subways, scenic spots and parks. Stable sites in the environment shall be selected for the test.

3.2.2. Service

The test should cover different types of XR service, such as Strong interactive VR: VR cloud game. Weak interactive VR: VR videos and movies; Strong interactive AR: AR multi-person assistance. Weak-interaction AR: AR movie. The service configuration should support mainstream transmission protocols, audio and video compression algorithms, and media encapsulation formats which can be presented in the appendix.

3.2.3. Environment

1. Test terminal: The test terminal supports all mobile communication systems and has been connected to the tested mobile communication network.

2. Test HMD: The VR HMD is connected to test terminal through hot spots or USBs, and then connected to gNodeBs through air interfaces. The gNodeBs are connected to XR Server through the core network.

3. Test content: Standardize user behaviors during the test, including the operation frequency and amplitude. Perform cloud games and other services, and record the subjective feelings of XR services under different signal strength, such as stalling, artifact, and black border.

4. Test duration: Standardize the specific duration from when users start XR services during a single test to the end of the test. Single XR service test takes about 5 minutes.

5. Test environment: Before initiating the test, close other applications that are connected to the network to ensure the stability and accuracy of XR service operation during the test.

6. Test scenario: The evaluation sample of the overall network quality of the operator shall cover different geographical regions such as cities, counties, towns, and villages. The test shall cover different scenarios such as indoor buildings, walking, driving vehicles and subways. Perform indoor and outdoor fixed-point tests with different signal intensities, and outdoor vehicle mobility tests.

3.2.4. Data Processing

Number of tests must ensure that the number of samples is sufficient for statistical purposes. Samples can be classified by operator, region, and bandwidth type. The number of samples of each type in each test is generally not less than 50, and the number of samples in a special period (for example, idle network hours) is not less than 20.

The processing of test data should be based on a large number of user test results. Through scientific statistical analysis of these test results, the overall test results can be obtained to reflect the Internet service experience in the country, provinces, municipalities, and autonomous regions.

After obtaining a sufficient number of test data samples, the processing principles for test data are as follows:

1. To avoid sample deviation, too large samples (top 2.5%) or too small samples (last 2.5%) should be discarded during data collection. Therefore, a certain margin shall be reserved for the original quantity of collected data.

2. Within a statistical period, an average value can be obtained by performing weighted averaging on the samples in the country or regions, and the average value is used as the user experience data of the country or regions. The weighting is based on the actual market structure of operators, regions, and bandwidth types.

4 References

[1] IDC,“全球 VR/AR 市场季度跟踪报告全球 VR/AR 市场季度跟踪报告”,2020