International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.9976

(12/2021)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

Access networks – In premises networks

Supporting ultra-high-definition video service over G.hn

Recommendation ITU-T G.9976

ITU-T

ITU-T G-SERIES RECOMMENDATIONS

TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200–G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300–G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400–G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450–G.499
TRANSMISSION MEDIA AND OPTICAL SYSTEMS CHARACTERISTICS	G.600–G.699
DIGITAL TERMINAL EQUIPMENTS	G.700–G.799
DIGITAL NETWORKS	G.800–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999
MULTIMEDIA QUALITY OF SERVICE AND PERFORMANCE – GENERIC AND USER-RELATED ASPECTS	G.1000–G.1999
TRANSMISSION MEDIA CHARACTERISTICS	G.6000–G.6999
DATA OVER TRANSPORT – GENERIC ASPECTS	G.7000–G.7999
PACKET OVER TRANSPORT ASPECTS	G.8000–G.8999
ACCESS NETWORKS	G.9000–G.9999
Metallic access networks	G.9700–G.9799
Optical line systems for local and access networks	G.9800–G.9899
In premises networks	**G.9900–G.9999**

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T G.9976

Supporting ultra-high-definition video service over G.hn

Summary

Recommendation ITU-T G.9976 studies the specificities of the transmission of ultra-high-definition (UHD) video service over G.hn. This Recommendation provides an analysis of the typical deployment of UHD video types in home networks, typical scenarios (including network topology, medium usage, support endpoints) and network requirements.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T G.9976	2021-12-17	15	11.1002/1000/14634

Keywords

G.hn, network requirement, UHD video.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Page

1	Scope	1
2	References	1
3	Definitions	1
3.1	Terms defined elsewhere	1
3.2	Terms defined in this Recommendation	2
4	Abbreviations and acronyms	3
5	Conventions	6
6	Video service in home network	6
6.1	General information	6
6.2	Video types in home networks	13
6.3	Application scenario	13
6.4	Network capability for supporting video	24
7	Video deployment in the home network	25
7.1	Distribution of 4K UHD video streams throughout the home	25
7.2	Network topology	26
8	Performance requirements for supporting video in home networks	26
8.1	Target architecture of carrier network for video service	26
8.2	Throughput requirements	27
8.3	Latency requirements	28
8.4	PLR requirements	29
9	Supporting video transmissions over ITU-T G.hn technology	30
9.1	General indoor system based on G.hn node	30
9.2	Requirements for devices in home networks	31
Annex A	Classification method for UHD video services	32
Annex B	Calculation algorithm model and equations for U_MOS	35
B.1	Calculation algorithm model	35
B.2	Calculation algorithm for Qs	36
B.3	Calculation algorithm for Qa	38
B.4	Calculation algorithm for Qi	39
B.5	Calculation algorithm for Qv	39
Appendix I	General requirements for supporting UHD broadcast TV over home networks	41
I.1	Request on network KPI from channel switching	41
I.2	Request on network KPI from no dappled screen	42

Page

Appendix II – General requirements for supporting UHD video on-demand over home networks 43

II.1 The request on networks for fast interaction 43

II.2 The request on the network for smooth view 44

Bibliography 45

Recommendation ITU-T G.9976

Supporting ultra-high-definition video service over G.hn

1 Scope

This Recommendation analyses typical video deployments within homes, including video types, network topologies, usage of the medium and network requirements.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T F.700] Recommendation ITU-T F.700 (2000), *Framework recommendation for multimedia services*.

[ITU-T G.9960] Recommendation ITU-T G.9960 (2018), *Unified high-speed wire-line based home networking transceivers – System architecture and physical layer specification*.

[ITU-T G.9961] Recommendation ITU-T G.9961 (2018), *Unified high-speed wireline-based home networking transceivers – Data link layer specification*.

[ITU-T G.9962] Recommendation ITU-T G.9962 (2018), *Unified high-speed wire-line based home networking transceivers – Management specification*.

[ITU-T G.9963] Recommendation ITU-T G.9963 (2018), *Unified high-speed wireline based home networking transceivers – Multiple input/multiple output specification*.

[ITU-T H.265] Recommendation ITU-T H.265 (2021), *High efficiency video coding*.

[ITU-T P.800] Recommendation ITU-T P.800 (1996), *Methods for subjective determination of transmission quality*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 quality of experience (QoE) [b-ITU-T P.195]: The degree of satisfaction of the user of an application or service. It results from the fulfilment of his or her expectations with respect to the utility or enjoyment of the application or service in the light of the user's personality and current state.

3.1.2 quality of service (QoS) [b-ITU-T F.1399]: The collective effect of service performance which determines the degree of satisfaction of a user of the service.

3.1.3 mean opinion score (MOS) [b-ITU-T E.800]: The mean of opinion scores, i.e., of the values on a predefined scale that subjects assign to their opinion of the performance of the telephone transmission system used either for conversation or for listening to spoken material. ([b-ITU-T P.800.1])

NOTE – There are some further different types of MOS. Their definitions are in [b-ITU-T P.800.1].

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 audio quality (Q_a): A quality indicator that depends on factors such as the clarity, fluency and fidelity of the audio source and can be measured by objective indicators such as the audio sampling rate, number of channels, bit rate, encoding method, encoding parameters and signal layer quality.

3.2.2 burst congestion: There is much bursty traffic, such as variable bit rate (VBR) video, in high-speed communication networks. If a large amount of bursty traffic simultaneously arrives at a node, or if a high-speed link emerges into a slower one, overflow may be buffered or congestion may arise. Unless the protocol software can detect congestion and reduce the sending rate of packets, in severe cases network communication services will even stop or deadlock, and the network will be paralysed due to congestion.

3.2.3 convergence ratio: For network equipment, cove usually refers to the ratio of the capacity of signal output port to the capacity of signal input port in the same direction. The advantage of convergence is that it can save shared equipment. The disadvantage is that some ports of network equipment will be congested when they are busy, and some packets will be discarded.

3.2.4 Interactive quality (Q_i): A quality indicator that measures the user's experience in interactive processes and which mainly depends on the system's response to user interactive operations. It is influenced by the platform, network and terminal. It can be quantified by objective indicators such as response to operations in the electronic programme guide in a single viewing behaviour, video initial loading time, channel switching time, response to fast-forward, response to fast-rewind, etc. Other factors of human–machine interaction (HMI) are also very important.

3.2.5 U_MOS: U_MOS = $f\,(Q_s, Q_a, Q_i, Q_v)$ is the comprehensive score of a subscriber's experience as a function of video quality (Q_s), audio quality (Q_a), interactive quality (Q_i) and viewing quality (Q_v). Each of these four indicators (Q_s, Q_a, Q_i, Q_v) are integers ranging from 1 to 5, representing the related subscriber's experience or service quality from low to high.

NOTE – The relationship among these indicators and functions is shown in the following equations that objectively reflect the subjective QoE of video service of subscribers. The U_MOS model mainly divides specific functional scenarios into session scenario and instant scenario. In session scenario, the user's viewing behaviour is longer than 1 minute at a time. The instant scenario is usually used for real-time quality monitoring.

$$U\_MOS = (Q\text{av} - 1) \cdot (1 - c_1 \cdot (5 - Qi) + c_2 \cdot (5 - Qv)) + 1 \tag{3-1}$$

Q_{av} is the overall quality of video and audio experience:

$$Q\text{av} = \alpha + \beta \cdot Qs + \gamma \cdot Qa \cdot Qs \tag{3-2}$$

Parameters c_1 and c_2 are the dynamic weighting coefficients of interactive experience and viewing experience, respectively. The initial c_1 and c_2 may be obtained from the results of data surveys carried out by the service operator. At the same time, the weighting factors α, β and γ are superimposed. When the score of (Q_s, Q_a, Q_i, Q_v) changes, the weighting factors α, β and γ will be adjusted accordingly. There are no specific requirements or methods for the adjustment of weighting factors α, β and γ; this is a way to divide different situations by adjustment of the weighting factors α, β and γ.

3.2.6 video quality (Q_s): Quality indicator that depends on the definition, smoothness, fidelity (chroma, contrast) of the source video and other factors. It covers the dimensions of the video source's resolution, frame rate, bit rate, content, encoding and terminal (screen size, PPI). It can be measured by objective indicators such as encoding parameters, resolution, frame rate, code rate and signal layer quality.

3.2.7 viewing quality (Q_v): Quality indicator that measures the user's viewing experience, which depends on the quality of the programme signal that appears during the viewing process. The influencing factors include dappled screen, mosaic, screen pause (stalling), and sound and picture asynchrony, which can be quantified by objective indicators such as the transmission performance and quality impairment of video and audio information.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AAC	Advanced Audio Coding
ABR	Adaptive Bit Rate
AI	Artificial Intelligence
AP	Access Point
app	application
AR	Augmented Reality
BER	Bit Error Rate
BTV	Broadcast TV
CATV	Community Antenna Television
CBR	Constant Bit Rate
CDN	Content Delivery Network
CO	Central Office
CPE	Customer Premise Equipment
DLL	Data Link Layer
DNS	Domain Name System
DOCSIS	Data Over Cable Service Interface Specifications
DPI	Dots Per Inch
DRM	Digital Rights Management
DSL	Digital Subscriber Line
DSLAM	Digital Subscriber Line Access Multiplexer
E2E	End-to-End
EPG	Electronic Programme Guide
FCC	Fast Channel Change
FF	Fast-Forward
FHD	Full High Definition
FR	Full Reference
FTP	File Transfer Protocol
FTTH	Fibre to the Home
FTTR	Fibre to the Room
FTTx	Fibre to the *x*

GARP	Generic Attribute Registration Protocol
GE	Gigabit Ethernet
GMRP	GARP Multicast Registration Protocol
GOP	Group of Pictures
HAS	HTTP Adaptive Streaming
HD	High Definition
HDMI	High Definition Multimedia Interface
HEVC	High Efficiency Video Coding
HLS	HTTP Live Streaming
HMI	Human–Machine Interaction
HPD	HTTP Progressive Download
HTML5	Hyper Text Markup Language 5.0
HTTP	Hyper Text Transfer Protocol
HTTP-FLV	Hyper Text Transfer Protocol Flash Video
IGMP	Internet Group Management Protocol
IP	Internet Protocol
IPB	Intracoded frames, Predicted pictures and Bidirectional predictive pictures
IPTV	Internet Protocol TV
KPI	Key Performance Indicator
KQI	Key Quality Indicator
LAN	Local Area Network
LOS	Light of Sight
MAC	Media Access Control
MCU	Multipoint Control Unit
MDU	Multiple Dwelling Unit
MIMO	Multi-Input Multioutput
MOS	Mean Opinion Score
MP2MP	Multipoint-to-Multipoint
MP2P	MultiPoint-to-Point
MPEG	Moving Picture Experts Group
MTU	Multitenant Unit
MV	Motion Vector
NAS	Network Attached Storage
NE	Network Element
NR	No Reference
NT	Network Terminal
OAM	Operations, Administration and Maintenance

OFB	Operational Frequency Band
OLT	Optical Line Terminal
ONU	Optical Network Unit
OTN	Optical Transport Network
OTT	Over the Top
PC	Personal Computer
PCM	Pulse Code Modulation
PE	Provider Edge
PHY	Physical Layer
PIM	Protocol Independent Multicast
PLR	Packet Loss Rate
PON	Passive Optical Network
PPI	Pixels Per Inch
P2MP	Point-to-Multipoint
P2P	Point-to-Point
Q&A	Questions and Answers
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QoE	Quality of Experience
QP	Quantization Parameter
RAN	Radio Access Network
RCM	Robust Communication Mode
RET	Retransmission
REW	Fast-Rewind
RGW	Residential Gateway
RR	Reduced Reference
RTMP	Real-Time Messaging Protocol
RTSP	Real-Time Streaming Protocol
RTP	Real-time Transport Protocol
RTT	Round Trip Time
SD	Standard Definition
SP	Service Provider
STB	Set Top Box
TCO	Total Cost of Ownership
TCP	Transfer Control Protocol
TV	Television
UC	User Centre

UDP	User Datagram Protocol
UHD	Ultra High Definition
VBR	Variable Bit Rate
VDSL2	Very-high-speed Digital Subscriber Line 2
VLAN	Virtual Local Area Network
VoD	Video on-Demand
VR	Virtual Reality
WLAN	Wireless Local Area Network
xDSL	x Digital Subscriber Line
xPON	x Passive Optical Network
YUV	luminance (Y), blue luminance (U), red luminance (V)

5 Conventions

None.

6 Video service in home network

6.1 General information

The pursuit of higher quality of experience (QoE) for subscribers continues to evolve. Figure 6-1 shows the roadmap of video services in the near future, where improvements in terms of higher resolution, more screens, stronger interactions and more ways to view, are driving the development of video technologies and the requirements on communication data traffic.

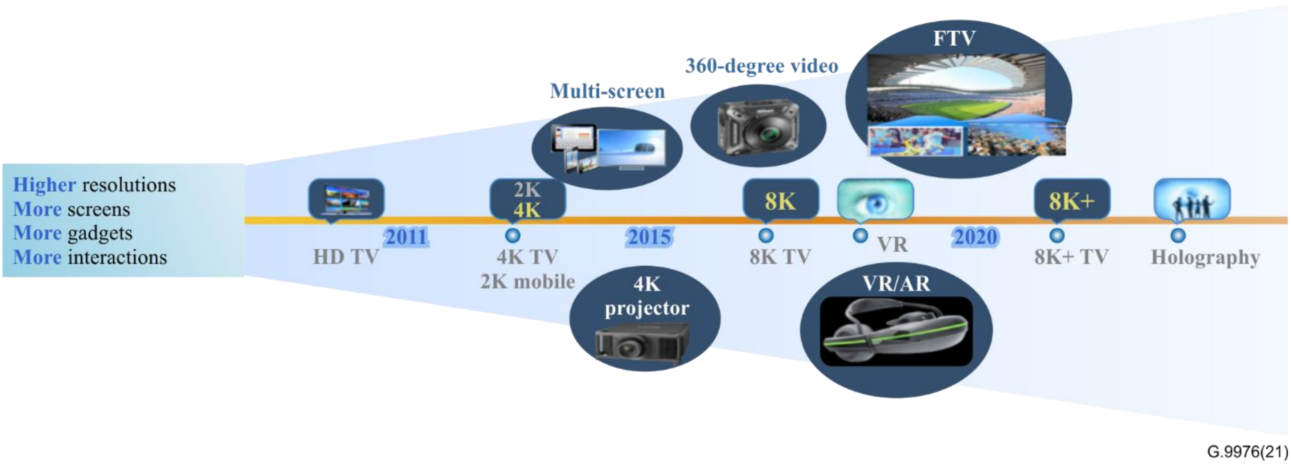

Figure 6-1 – Road map of video services

Video service is the main service requiring a high transmission data rate, which takes up most of the volume of an access network. It is also seen to be one of the most important services triggering the upgrade of an operator's network. Video service is approaching virtual reality (VR) and augmented reality (AR), in the future, and then holography.

With the fast and wide deployment of 4K video and the appearance of 8K video service, stakeholders have noticed the importance of guaranteeing the QoE of subscribers. In practice, user experience can be evaluated by key quality indicators (KQIs) and key performance indicators (KPIs) of the video service.

6.1.1 Evaluation method and information model of the QoE of video service

The audio MOS perception score defined by [ITU-T P.800] divides the subjective perception of QoE into five levels, which is a sequential scale method. At present, the quantitative perception score of each video experience adopts a similar method to the audio MOS perception score, which divides subjective perception into five levels, with integer 5 score being the highest and integer 1 being the lowest.

The objective evaluation method for QoE measures the degree of distortion of the output video and the original video, which can be divided into a full reference (FR) evaluation method, a reduced reference (RR) evaluation method, and a no reference (NR) evaluation method.

The FR evaluation method obtains quality evaluation by comparing the system output video sequence (test video) and input video sequence (reference video).

The RR evaluation method is also called the partial reference evaluation method, which selects partial parameters of input and output video for comparison to obtain quality evaluation. Its calculation accuracy and calculation degree are between FR and NR.

The NR evaluation method only performs quality evaluation based on the output video sequence. It can be divided into bitstream layer-based evaluation and media layer-based quality evaluation according to the analysed characteristic parameters.

The NR method is an important method used by this Recommendation to evaluate the QoE of video services.

The input information of the NR method is divided into four levels, as shown in Figure 6-2.

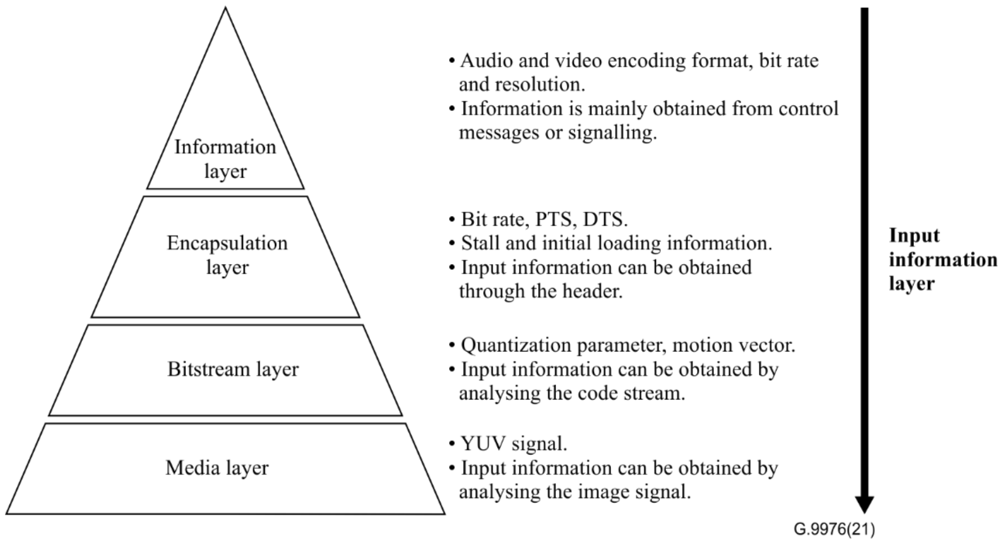

Figure 6-2 – Information model for estimation of video

According to the different layers of the acquired characteristic parameters of the video service, it is necessary to select algorithms of different input information layers, and the accuracy of the algorithm is also different. According to the layer of the acquired characteristic parameters, it can be roughly divided into the following layers:

The evaluation of the video quality through the parameters of the information layer is called the planning model, which is mainly used to estimate the expected value of the service quality during the service deployment period and estimate the average value of the service quality.

The evaluation of the video quality through the parameters of the encapsulation layer is called the parametric model, which mainly evaluates the video quality through the interactive signalling of the video and the analysis of the message header information.

Evaluating the video quality by bitstream layer parameters is called the bitstream model; it needs to obtain the video stream information characteristics for evaluation.

Evaluating the video quality by media layer parameters is called the hybrid model; it needs to restore the video image and analyse the video content.

The complexity of the planning model, parametric model, bitstream model and hybrid model is increasing. When conditions permit, the parameter set of the parametric model should include the parameter set of the planning model, and the parameter set of the bitstream model should include the parameter set of the parametric model. The parameter set of the hybrid model should include the parameter set of the bitstream model, that is, the most advanced hybrid model should include not only the exclusive KPIs for the media layer, but also the exclusive KPIs for the bitstream, encapsulation and information layers.

The parametric model, bitstream model and hybrid model can all be used for real-time quality monitoring, and the accuracy of model evaluation is positively correlated with computational complexity. The hybrid model has the highest accuracy and complexity, while the bitstream model is the second-most accurate and complex. The accuracy and complexity of the parametric model are the lowest.

The indicators are affected by many factors such as video server, core network, access network, indoor network and even the network terminal. The relevant indicators for video service are shown in Figure 6-3.

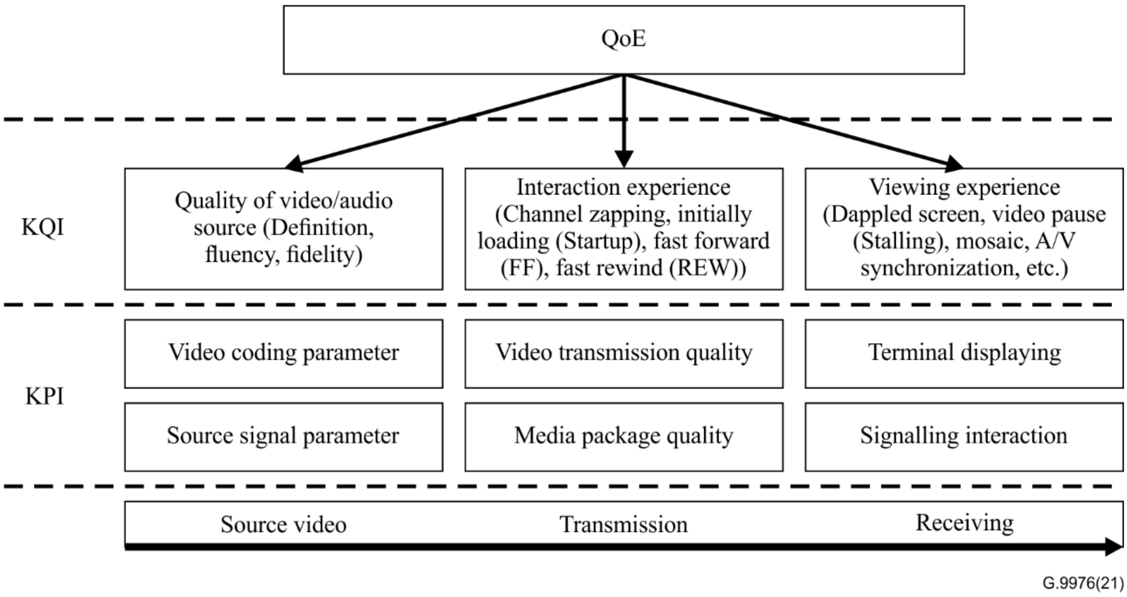

Figure 6-3 – Framework of quality indicators for a video service

6.1.2 Framework of evaluation of the QoE of video service

The QoE of a video service is determined by the quality of the source video, transmission quality and the quality of the receiving end equipment including the monitor. For end-to-end (E2E) video displaying, evaluation of the QoE of video service can be carried out on these three levels.

The quality of the source video is mainly affected by factors such as the source signal parameter and video coding parameter. For example, the quality of the source video is mainly determined by the quality of the encoding of the original video received from the service provider (SP) via the hardware device, or the transcoding on the existing video file. An evaluation of the source video can be used to verify whether the quality of the source video meets the user's requirements. The index system for evaluating the quality of the source video is shown in the left KPI modules in Figure 6-3. Among them:

- The source signal parameter module includes the quality loss caused by the original signal acquisition, such as still frames, over-brightness, over-darkness and blurred images. These parameters are related to the media layer. The module includes parameters within the Q_s parameter set and Q_a parameter set. In the Q_s parameter set, the source signal parameters include bit rate, video frame rate and resolution. In the Q_a parameter set, the source signal parameters include bit rate, sample rate and number of channels.
- The video coding parameter module is mainly used to define the loss of image quality caused by loss of information due to lossy compression. These parameters are related to the bitstream layer. The module includes parameters within the Q_s parameter set and Q_a parameter set. In the Q_s parameter set, the video coding parameters include codec type (e.g., H.264, H.265, MPEG2-TS, Flash (FLV), MP4, WebM and AVS2), coding para. (e.g., frame type, total bytes per frame, QP, MV, skip ratio) and coding method (i.e., CBR/ VBR). In the Q_a parameter set, the video coding parameters include codec type (e.g., PCM, MP3, WMA, AAC and AC3), scale factor, and coding method (i.e., CBR/VBR).

The quality evaluation of source video can only be applied to unencrypted streams. The quality of the video during transmission is shown in the middle KPI modules in Figure 6-3. It is mainly affected by two factors: video transmission and media package.

- The main factors affecting the quality of video transmission include network packet loss, network packet retransmission and network packet jitter.
- The main factors that affect the quality of media packaging include incorrect media packaging, such as asynchronization of audio and video caused by incorrect timestamps.

When evaluating the video quality at transmission side, the quality of the source video needs to be considered. Video transmission quality and video encapsulation quality are encapsulation layer information.

If the video is encrypted during transmission, it is difficult to obtain the video bitstream layer and media layer information.

The quality of the video at receiving end is shown in the KPI modules on the right side of Figure 6-3. It is affected by many factors such as the receiving end, network, platform and source video.

The terminal displaying factor includes the size of screen, whether the terminal hides the error of the received video, i.e., the decoder mode. When using error hide mode, the received packet with error will be directly discarded. When using error non-hide method, the received packet with error will be displayed. The following types exist:

1) Broadcast TV (BTV) / video on demand (VoD) + non-transfer-control protocol (TCP) + intracoded frames, predicted pictures and bidirectional predictive pictures (IPB) coding + no hide mode:
   a) When the whole intracoded frame is lost, which generally is the first frame in the group of pictures (GOP), black screen and video pause (stalling) will occur;
   b) When the whole predicted picture is lost, the screen will be mosaic/blurred;
   c) When the whole bidirectional predictive picture is lost, the screen will be mosaic/blurred.
2) In other VoD scenarios, a possible question is video pause (stalling). Generally, VoD is used with TCP.
3) In BTV + non-IPB coding + no-hide mode, a possible question is whether the screen will be mosaic/blurred;
4) In BTV + hide mode, a possible question is whether video pause (stalling) will occur.

The signalling interaction includes the initial signalling interaction at the receiving end, which affects the initial loading time of the video, the video buffering time at the receiving end and the requirement of the time interval for the receiving end to receive fragments.

6.1.3 Key indicators

The process of a user using the VoD/BTV function includes starting the terminal/player, accessing the VoD/BTV system, selecting content to watch and performing operations such as sending on-demand request and watching. When the user sends out a VoD request, the streaming media service system will retrieve the programme information stored in the source library according to the VoD information and transmit it to the user's terminal through the high-speed transmission network as video and audio stream files. VoD services are generally implemented by unicast mode.

The factors affecting the QoE of the VoD/BTV function come from the terminal's software, hardware, transmission protocol, codec, network, platform, content source quality and other aspects. The three key aspects are quality of video/audio source, interaction experience and viewing experience, as shown in the KQI modules in Figure 6-3.

The quality of the source video depends on the definition, fluency and fidelity (chroma, contrast) of the source video. It covers the six indicators in dimensions of the source video including the resolution, frame rate, bit rate, content, encoding and terminal. They can be measured by objective indicators such as encoding parameters, resolution, frame rate, bit rate and signal layer quality.

The quality of the audio source depends on factors such as the clarity, fluency and fidelity of the audio source, and can be measured by objective indicators such as the audio sampling rate, number of channels, bit rate, encoding method, encoding parameters and the quality of signal layer.

The user's interaction experience depends on the system's response speed to the user's interactive operations. It covers performance indicators of the platform, network and terminal. It can be used for the user's response to electronic programme guide (EPG) operation in one viewing behaviour, and the initial loading time of video, channel zapping time, fast-forward (FF) and fast-rewind (REW) response and other objective indicators. Other human–machine interaction (HMI) factors are also very important.

The viewing experience depends on the quality of the programme signal that appears during the viewing process. The influencing factors include dappled screen, mosaic, video pause (stalling) and A/V synchronization, which can be measured based on objective indicators such as the transmission performance and quality impairment of video and audio information.

In addition, the user environment, the inertia of people's perception, the size of the screen in terminal, the installation and maintenance of equipment, the richness and continuity of services and other indicators also have a certain impact on the QoE.

The environment in which the evaluation of QoE of VoD/BTV is carried out will affect the QoE. There is a significant difference between watching video on a hand-held mobile terminal in a mobile environment and watching video in front of a TV or computer in a static environment. For example, the degree of tolerance to video pause (stalling) will be different.

For terminals with screens of different sizes, users have different requirements for video clarity.

The influence of objective indicators of interaction experience and viewing experience on the QoE will weaken as the viewing behaviour continues. When a dappled screen / video pause (stalling) event occurs, the QoE drops immediately, and when dappled screen / video pause (stalling) ends and normal playback resumes, QoE will gradually and slowly recover. If the follow-up can continue to play normally, real-time QoE will gradually return to a normal value.

In order to scientifically evaluate and quantify the QoE of the VoD/BTV function, the evaluation scenarios are divided into session scenario (characterizing the experience quality of the user's complete viewing behaviour) and instant scenario (characterizing the user's experience quality during real-time viewing). Statistical analysis of the QoE in a certain period (week, month or year) in the session scenario can reflect the user's real QoE when using the VoD/BTV function.

Carrying out evaluation of the QoE of VoD/BTV requires a comprehensive consideration of various factors such as audio and source video quality, interaction experience quality, display experience quality, user environment, time continuity and frequency. The user's score of QoE on video service, which combines the quality of the source video, the quality of the interaction experience and the quality of the viewing experience, is a mean opinion score (MOS) that truly reflects the QoE.

6.1.4 Framework of quality indicators of video service

The quality of a video stream can be evaluated by KPI (network related) and KQI (service related) parameters. KPI generally refers to indicators based on a network device, including terminal performance, network transmission quality, video transmission quality, media encapsulation quality and video coding parameters. KPIs determine the actually available E2E throughput from the subscriber's terminal to the Internet, impacting the transmission latency between the video server and the subscriber's end device, therefore affecting the subscriber's experience (i.e., QoE).

KQI indicators are generally obtained through KPI indicator processing.

Starting from the subscriber level, QoE indicators can be defined as the overall subjective perception of a terminal subscriber of video services. The perceptual experience of the video service is a comprehensive score reflecting the subscriber's experience during watching the video service, which is generally represented by U_MOS (U_MOS = $f\,(Q_s, Q_a, Q_i, Q_v)$).

6.1.5 Collecting parameters and processing indicators

The video service is finally presented to the subscriber through the terminal, which is the closest entity to the subscriber in the video service system. Therefore, the source of the parameters that affect the quality of the video service can be the terminal's network packet captured data, the terminal system information related data, or the audio/video information related data reproduced by the player.

Parameter collection can be achieved in multiple locations of the network, such as set top boxes (STBs) and players, or through network probes to assist in fault location and diagnosis. When the parameters are obtained from the terminal, the calculated service quality accuracy is higher, but support of parameter acquisition by the terminal device is required.

After the acquired data is processed, KPI indicators (network element (NE) related indicators) and KQI indicators (service related indicators) can be obtained. KPI indicators and KQI indicators form the QoS indicators for video service quality analysis. The QoE indicator is based on the subscriber perception level and the KQI indicator can reflect the subscriber's service experience to a certain extent.

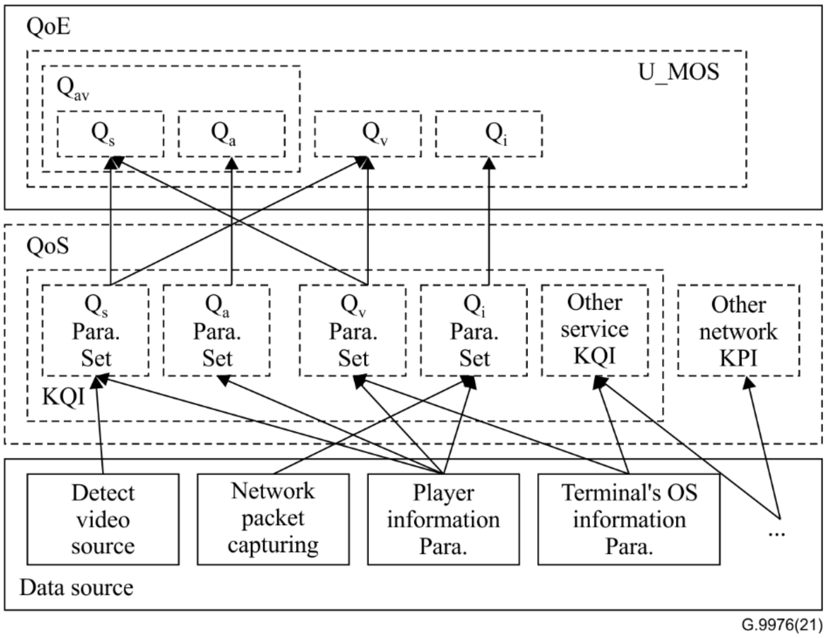

Figure 6-4 – Data processing flow

Figure 6-3 shows the KPI parameters related to KQI at different stages. In practical applications, most of the parameters are directly collected by the data source, as shown in Figure 6-4. Data sources are divided into four main categories. These four data sources are combined with other network management systems, platforms, etc., and then a total of six parameter sets can be obtained as follows:

- Q_s parameter set. This includes parameters of the source video such as bit rate, video frame rate, resolution, codec type, coding para., information layer para., coding method (CBR/VBR).
- Q_a parameter set. This includes parameters of audio source such as bit rate, sample rate, number of channels, codec type, scale factor information layer para., coding method (CBR/VBR), audio loss/volume of sound too high/low.
- Q_i parameter set. This includes parameters during interaction such as number of times of VoD/BTV request, number of times of VoD/BTV request succeed, number of times of EPG request, number of times of EPG request succeed, number of times of index request (HTTP live streaming (HLS)), number of times index request succeeded (HLS), latency of VoD/BTV request, latency of EPG request, number of occurrences of abnormal interrupt, response time of FF/REW/pause, applicability of HMI to the terminal.
- Q_v parameter set. This includes parameters during viewing such as buffer frequency, buffer length, buffer start time, number of occurrences of dappled screen, length of dappled screen (in time), area percentage of dappled screen, screen size and screen dots per inch (DPI).
- Other service KQI. This includes domain name system (DNS) resolution time, index getting time, content delivery network (CDN) address getting time, video segment TCP connection time, segment error times, segment download average rate, segment download peak rate and segment resource distribution. This is also mainly collected by the terminal.
- Other network KPI. This includes committed accessing throughput, round trip time (RTT), packet loss rate (PLR), etc.

These six parameter sets, especially the first four, will be calculated and combined to obtain a set of variables that reflect QoE in four aspects (Q_s, Q_a, Q_i and Q_v).

After the relevant variables are calculated by the equations (see Annex B), a total of four scores (i.e., for Q_s, Q_a, Q_i, Q_v, ranging from 1 to 5) and a total U_MOS score (ranging from 1 to 5) will also be obtained.

According to the U_MOS algorithm model, subscribers can obtain good QoE on an E2E network, such as $Q_i \geq 4$ (i.e., loading delay for VoD ≤ 1 second, channel zapping delay for BTV ≤ 500 ms) and $Q_v = 5$ (i.e., "0 screen pause (stalling)" in VoD service and "0 dappled screen" in BTV service).

6.2 Video types in home networks

Video services can be characterized into different types, e.g., SD, HD, FHD and UHD (4K/8K) based on resolution. The parameters, compression techniques and throughput requirements of various videos are summarized in Table 6-1 (for the calculation of the throughput requirements, please refer to Appendices I and II).

Table 6-1 – Throughput requirements for different video format [b-ETSI WP]

Type	SD	HD	FHD	UHD					
				Basic 4K	True 4K	Excellent 4K	Basic 8K	True 8K	Excellent 8K
Resolution (pixels)	640 × 480	960 × 720	1920 × 1080	3840 × 2160			7680 × 4320		
Frame rate (fps)	25/30	25/30	25/30	25/30	50/60	100/120	25/30	50/60	100/120
Sampling bits/colour depth (bits)	8	8	8	8	10	12	10	12	14
Sampling/compression	YUV 4:2:0 & H.264			YUV 4:2:0 & H.265/HEVC					
Min. bit rate (Mb/s)	2	4	8	15	30	50	50	100	220
NOTE – Other frame rates may be used for dedicated services, such as 24 for movies.									

Moreover, carrier-grade video services have two different modes: VoD and BTV.

According to Table A.3, distribution video services can be divided into two categories: live BTV and VoD. The former requires a communication delay of "real-time", and the "time continuity" is usually "isochronous". The subscriber cannot control the temporal order of the video content. The latter requires a "specified time" for "communication delay", and the "time continuity" is usually "non-isochronous".

UHD VoD service is generally downloaded based on a TCP connection. Because TCP is a reliable transmission protocol, it can check the received packets and retransmit the wrong or lost packets. Therefore, there is no dappled screen caused by packet loss in VoD mode. The main QoE indicator is fast video response (such as no pause) when video is displayed.

BTV services are generally broadcast and multicast. The interactive experience in the multicast mode is impacted mainly by the speed of channel zapping, and the viewing experience requires that the entire process does not have mosaic, black screen, etc.

6.3 Application scenario

6.3.1 Application scenario of BTV

BTV video is an important service for entertainment video. Subscribers can watch sport events, major events and news online through the network. At this time, subscribers cannot control the temporary order of the video content.

BTV can be conveyed by radio, television networks and IP networks. BTV conveyed by radio and television networks are generally provided by SPs. BTV by IP networks can be divided into multicast and unicast according to the configured number of subscribers and types of terminals.

In multicast mode, the network establishes a multicast tree through a multicast routing protocol (such as protocol independent multicast (PIM)). After the head-end system encapsulates the streaming media encoding data, the multicast tree is used to distribute the video stream data to the multicast P2MP node at the edge of the network. When a piece of subscriber equipment (such as a set top box (STB)) requests to watch a channel, it sends an Internet group management protocol (IGMP) join message through a residential gateway (RGW) to the multicast P2MP node. After the subscriber joins the multicast, the multicast P2MP node immediately copies the streaming media data to the subscriber device. The multicast P2MP node can be a provider edge (PE) router / an optical line terminal (OLT) / a multiple dwelling unit (MDU) optical network unit (ONU) / a multitenant unit (MTU) ONU / a digital subscriber line access multiplexer (DSLAM).

P2MP multicast can be realized through Layer 3 multicast, namely IP multicast. It usually uses the IGMP protocol. The multicast VLAN function solves this problem. After the multicast VLAN is configured on the Layer 2 device, the Layer 3 device only needs to copy the multicast data in the multicast VLAN and send it to the Layer 2 device instead of copying it in each user VLAN, thus the network throughput is saved, and the load of the Layer 3 equipment is also reduced. There are two implementation and configuration methods for multicast VLAN: sub-VLAN-based multicast VLAN and port-based multicast VLAN.

Layer 2 multicast refers to the processing flow of IP multicast packets on Layer 2 devices. The Layer 2 switch performs table entry learning according to the source MAC address of the message. Since the multicast MAC address is the destination address, it is necessary to establish the multicast table entry in two ways: static configuration and dynamic learning. The main learning method is generic attribute registration protocol (GARP) multicast registration protocol (GMRP) and IGMP snooping protocol.

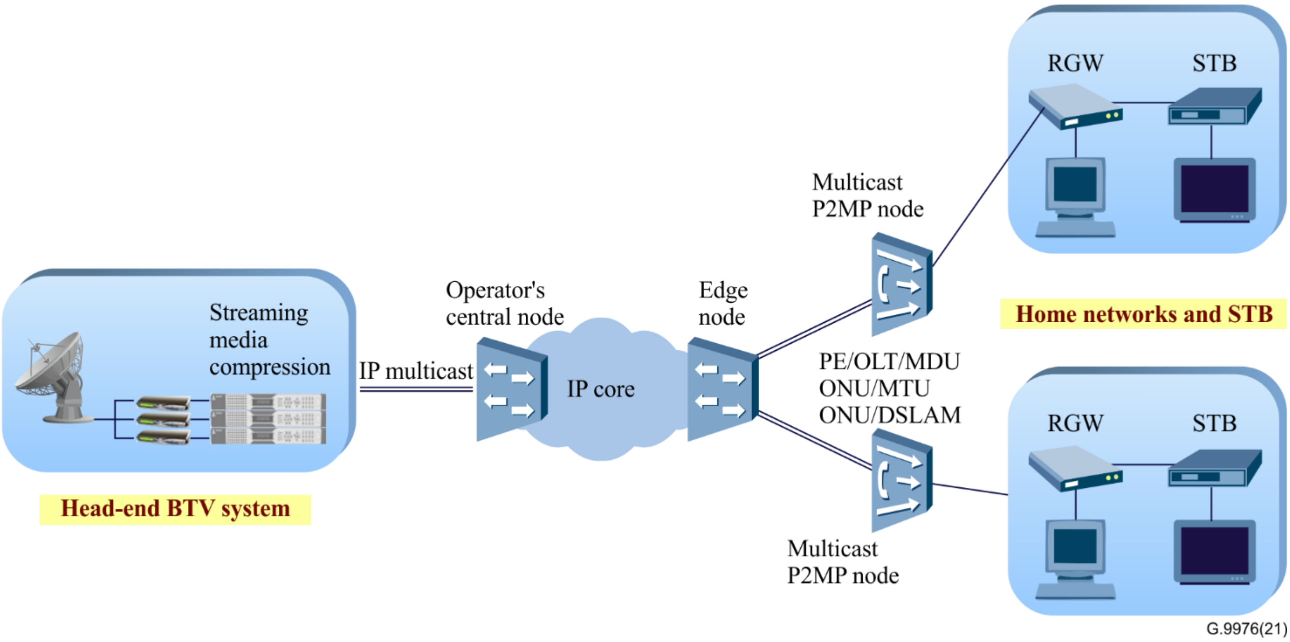

Figure 6-5 – Schematic architecture diagram of BTV multicast

BTV videos are generally presented to subscribers in the form of channels. During video transmission, different access encapsulation formats and carrier methods can be adopted according to different network structures and the total number of subscribers.

The video encapsulation format is a way of combining information such as video tracks, audio tracks and subtitle tracks. The suffixes of common video files such as .mp4, .avi, and .rmvb are all encapsulation formats of video files. The encapsulation format has a certain corresponding relationship with the video encoding method. For example, the .mov encapsulation format file can be encoded with MPEG-2, MPEG4-ASP (XviD), H.264 and other encoding methods. Video content can be carried in carrier methods such as BTV/VoD, multicast/unicast, TCP / user datagram protocol (UDP), etc.

Most operators' self-built IPTV BTV uses multicast schemes and UDP transport layer protocol. It is a real-time streaming medium and has no service layer interaction control. Packets will not be retransmitted when error or packet loss occur during transmission. Between the media server and the edge node of the network, multicasting is generally used as shown in Figure 6-5, which can save the bandwidth resources of the metropolitan area network. In addition, the edge node to the subscriber may adopt unicast or multicast mode.

In multicast mode, UDP is basically used to push video streams. In unicast mode, TCP or UDP may be used. The live video scenario mainly considers the most common live broadcast scenario of multicast UDP. The scenario where TCP transmits live video is similar to the on-demand scenario.

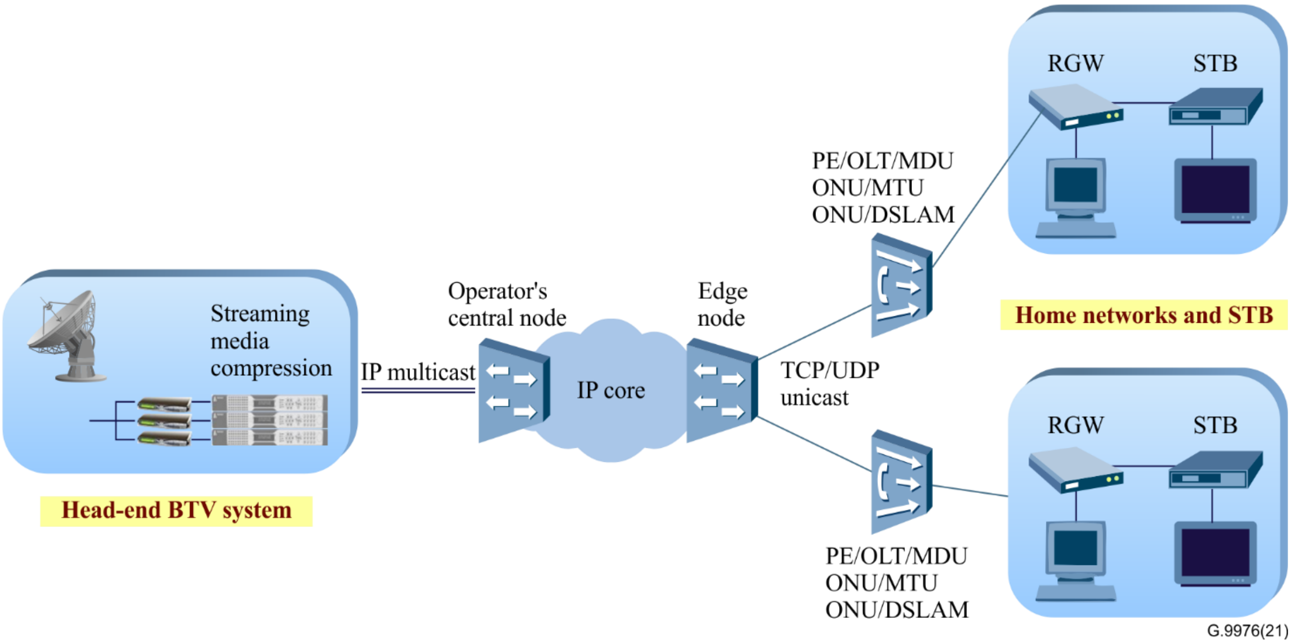

Figure 6-6 – Schematic architecture diagram of BTV unicast

As shown in Figure 6-6, some IPTV live channels currently provide services to subscribers through unicast. In such a scenario, TCP or UDP may be used for transmission. Prior to unicast, the content provider pushes the live stream to the operator's central node through multicast. The central node to the area or edge node may be multicast or unicast relay.

In unicast mode, the head-end system first pushes the live stream to the network edge nodes through multicast. Subscriber's equipment (such as an STB) requests content from edge nodes via TCP or UDP unicast.

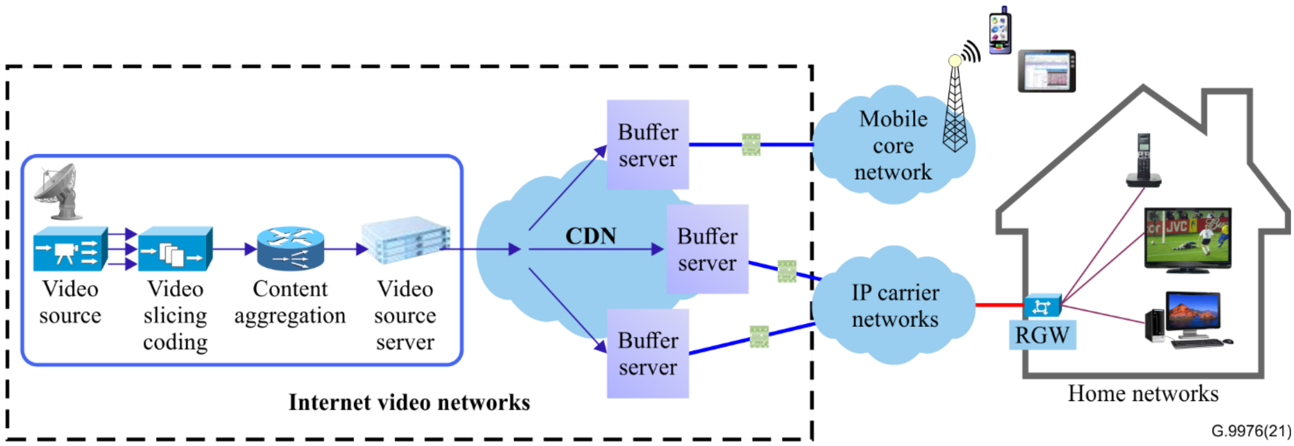

Figure 6-7 – Schematic architecture diagram of BTV by Internet

The popular Internet live broadcast now belongs to an application of IP network live broadcast. The system structure diagram is shown in Figure 6-7. It provides the service of integrating IP video and Internet applications for TV transmission through the public Internet. The terminal is an integrated Internet television or an STB connected to a television or a mobile terminal.

There are two main ways to implement Internet video: HTTP progressive download (HPD) and HTTP adaptive streaming (HAS). Traditional over the top (OTT) video generally uses HPD technology. The HPD-based client only needs to wait for a short period of time before downloading and buffering the first part of the data of the media file before it can be played while downloading. HPD OTT video has many limitations, for example:

- It is not suitable for the transmission of live programmes with high real-time requirements;
- The waiting time for initial play is generally long;
- When the network throughput is unstable, it is prone to video pause (i.e., stalling);
- When the subscriber abandons the programme in the middle, some segments of the video files continue to download, resulting in wasted throughput.

In order to overcome the limitations of HPD OTT video technology, HAS OTT uses video fragmentation and adaptive bit rate (ABR) technology, which can provide video files with suitable resolutions to terminals according to different screen sizes, and achieve smooth video display under different network throughput conditions.

Mobile video service is also an application scenario of BTV by IP network. Its system structure is shown in Figure 6-8. It mainly involves networks including Internet core network, IP carrier network, IP metropolitan area network, wireless network and wired access network. The difference in service processing is mainly reflected in the carrier network. As shown in Figure 6-8, there are four types of video service processes:

- The first is to provide video content by the local CDN;
- The second is to provide video content by DNS resolution and the local Web cache (web page snapshot);
- The third is to provide video content by CDN or Web cache in another city;
- The fourth is to provide video content by the external network CDN or the source website.

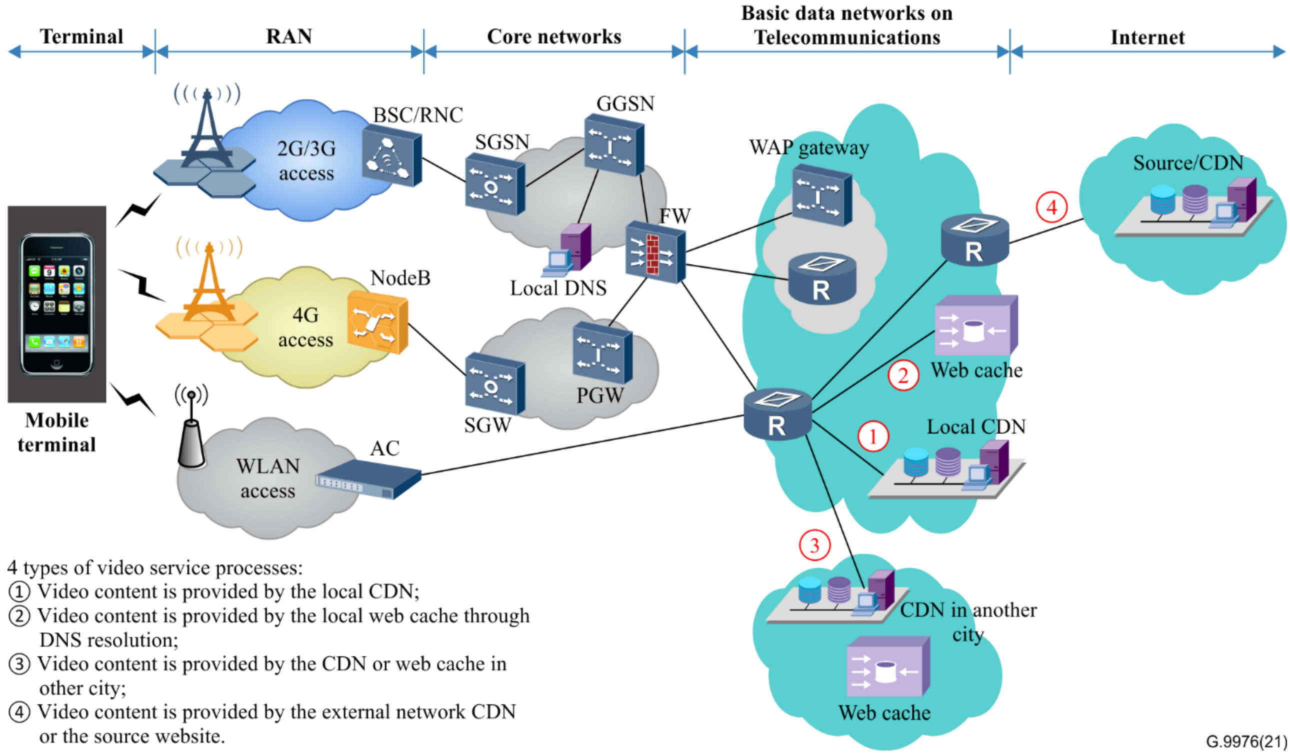

Figure 6-8 – Schematic architecture diagram of mobile BTV video broadcast

6.3.2 Application scenario of VoD

For general home subscribers, entertainment video services are often referred to as audiovisual on-demand services. A content provider and an SP provide video programme services to multiple subscribers, and subscribers can choose to watch any channel of those provided. There is an asymmetric interactive multimedia communication between subscribers and machines (audiovisual resources), providing audiovisual information to subscribers at any time through a telecommunication network. Subscribers can find the information they need with the assistance of the navigation subsystem. After the subscriber chooses the service mode, the audiovisual on-demand service platform sends content information with adequate QoS levels.

The content provided by the audiovisual on-demand service platform can be text, audio, graphic, picture or video information. The media resource can be stored anywhere on the network. The audiovisual on-demand service platform integrates the media resource, performs digital rights management (DRM) and provides navigation service and content delivery through the telecommunication network.

Subscribers can use audiovisual on-demand service in fixed or mobile location such as homes, offices or moving trains. Subscribers should be able to use a PC, STB + TV, mobile phone, tablets or any other device to access the broadband network to obtain service through streaming media communication capabilities.

The VoD mode allows subscribers to select videos to watch according to their needs, instead of passively watching a specific broadcast video.

The VoD implementation process is as follows: When a subscriber issues an on-demand request, the streaming media service system will retrieve the programme information stored in the source video library according to the on-demand information and send the video and audio stream files to the subscriber terminal. On-demand services are generally implemented using unicast. The VoD architecture is shown in Figure 6-9.

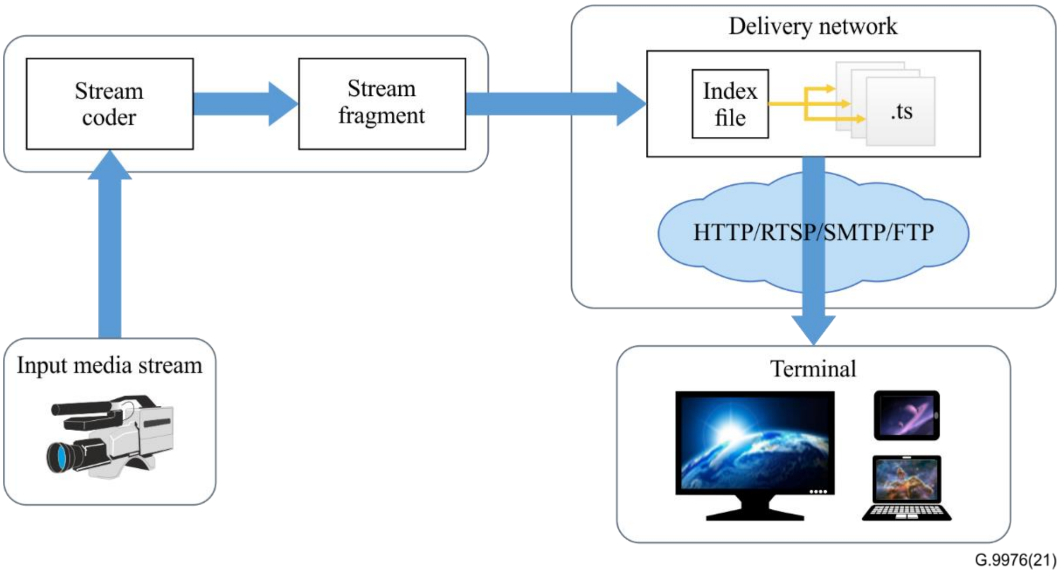

Figure 6-9 – Example of architecture on VoD mode

The network structure consists of a VoD front-end processing system, a server, a delivery network and a client (STB / mobile terminal). From the perspective of the service flow, the subscriber initiates a play request on the client-end. This request is sent through the network, arrives and is received by the network adapter of the VoD server, and is transmitted to the VoD server. After request verification, the server prepares the accessible programme names in the storage subsystem so that subscribers can browse the favourite programme menu. After the subscriber selects a programme, the server fetches the programme content from the storage subsystem and transmits it to the client for playing.

VoD can be an IPTV service provided by an operator or a video service provided by a content provider.

6.3.3 Conferencing application scenario

According to Tables A.2 and A.3, the conferencing service has the "multipoint-to-multipoint (MP2MP)" communication configuration, "real-time" communication delay, "isochronous" time continuity and bidirectional (i.e., source and sink) transmission control entity.

A traditional conference video service provides real-time communication between multiple users in different locations, has good audio facilities and combines moving images of participants and/or the transmission of multimedia information. This type of service is applicable to the company's meeting rooms and public access meeting rooms for hire. It is suitable for various types of multimedia conference terminals.

Conferencing is a bidirectional service based on a telecommunications network, which provides the interconnection of two or more multimedia conference terminals. Other types of terminals can join the conference, such as smartphones or even traditional phones. Although the performance of various types of information in multimedia conferences is usually limited in sending and receiving, at least they can exchange voice to allow conference users to join discussions.

When a conference includes more than two terminals, a multipoint control unit (MCU) is usually required. All locations are connected to the MCU, which selects or appropriately combines signals from various locations, and manages signalling and optional channels.

The conference service allows users to use their own devices to make person-to-person video calls or participate in multiperson video conferences. The video conferencing architecture is shown in Figure 6-10 below, which can be divided into three layers: the subscriber access layer, media processing layer and service platform.

The subscriber access layer includes various types of terminals and meeting rooms. The media processing layer includes MCU, enterprise user centre (UC) and media gateway, etc. It is responsible for media adaptation and forwarding. The service platform is responsible for service scheduling, resource allocation and other functions. The flow of video communication is as follows: the user sends a video communication request, the peer chooses to accept, the two ends establish a communication connection and receive data such as video through a high-speed transmission network to perform video communication.

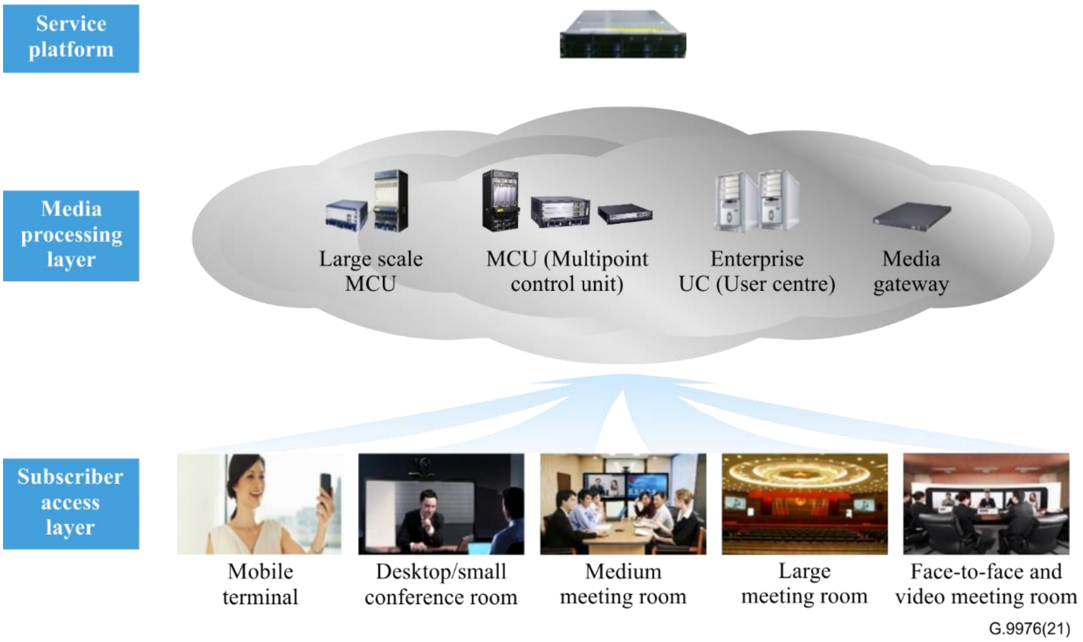

Figure 6-10 – Schematic diagram of video conferencing architecture

A typical diagram of video conferencing for commercial and enterprise users is shown in Figure 6-11. Various types of terminals access the video communication system through the MCU, and use the video communication service under the management of the service platform. Administrators can log in to the service platform through the web portal for configuration and management.

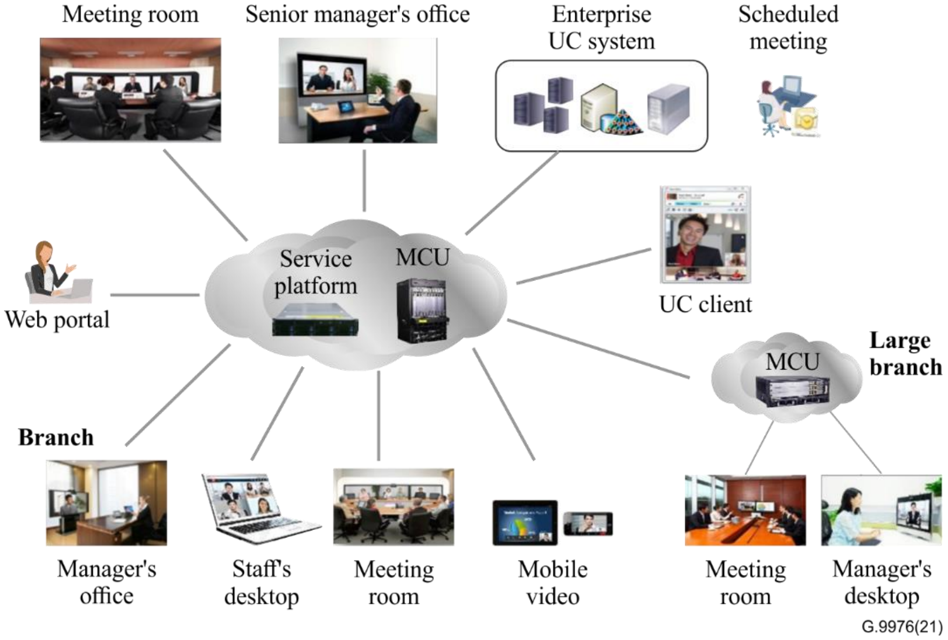

Figure 6-11 – Schematic diagram of video conferencing for commercial and enterprise users

For ordinary home users, the video communication service can be provided by the operator's network, or it can be undertaken by an application, such as various network communication tools, provided by the SP via OTT, so can be used in the home network environment.

In addition to the quality of audio and video, the main factors affecting video communication also include the delay of call establishment and experience during the call. Delay of call establishment is the waiting time from click to connect call to successful connection and screen display. The call experience includes:

- Video pause (stalling), which means that the video is paused due to the jitter of the network, and it can continue to play after a period of buffering;
- The dappled screen, that is, the loss of data packets will cause video distortion;
- E2E delay and synchronization between audio and video.

When performing video communication, the environment of the terminal also affects the user's QoE; for example, in a mobile environment, holding a mobile terminal for video communication compared with video communication via a TV or PC under fixed-line environment, the user will show different levels of patience with the QoS regarding issues such as video pause (stalling). In addition, the size of the terminal is different. For example, compared with a large-screen TV, mobile phones and tablets have different requirements for resolution and therefore the above factors need to be considered.

6.3.4 Conversation application scenario

A multimedia conversation service provides real-time bidirectional communication between users in two different locations through the telecommunication network. It usually transmits the user's audio, images or multimedia information at the same time.

In general, video services require audio. Thus, in order to achieve the basic functions of multimedia conversation services, a terminal should include the following necessary audio communication components: a microphone, amplifier, audio codec and some optional audio related controls. The terminal should also include a network interface unit.

Home users mainly conduct peer-to-peer video calls as shown in Figure 6-12. Home users in different locations conduct video communication through two terminals connected via the Internet.

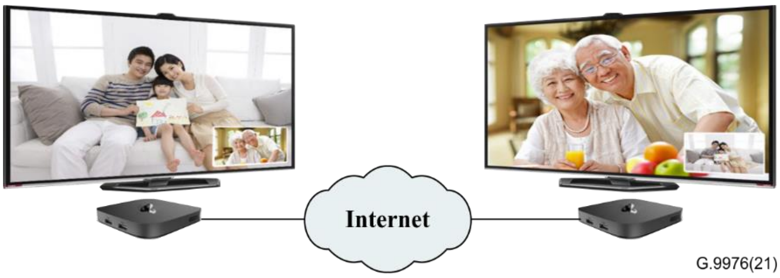

Figure 6-12 – Schematic diagram of video conversation service between home users

The conversation service is similar to the conferencing service, and both belong to the video communication service. The number of participants in the two services and the "communication configuration" characteristics of the communication tasks are, however, different.

6.3.5 New services application scenario

With the rapid development of video services, many new applications are emerging. The typical new applications include online education, real-time telemedicine and live broadcasting (such as webcast). Typically, these services may be subscribed concurrently by multiple family members through various terminals.

With the continuous development of network technology, codec technology and storage technology, the application of video collection services is becoming more and more extensive. Surveillance service in public places is one of the most common forms of video collecting services. At each remote monitoring point, the video encoder compresses and encodes the audio and video signals collected by the camera equipment and then transmits them to the monitoring centre through the network. Both the monitoring host and Internet users can obtain the audio and video signals of various remote monitoring points from the monitoring centre through the network. These signals can be decoded by a video codec and played, or saved by a hard disk video recorder. The collection video service is mainly a point-to-point uplink service. It is mainly manifested as the simultaneous upload of video streams from multiple locations to one platform.

With the development of computer technology and network technology, as well as the popularization and application of multimedia, many new forms of services and information communication methods such as digital libraries and digital museums have emerged.

In traditional database systems, information retrieval is generally based on numerical and character types, while multimedia databases integrate image, video, audio and other types of information, which have the characteristics of large data volume, variable length of information and complex structure. Video retrieval needs to find the required video clips from a large amount of video data. According to the given examples or feature descriptions, the system can automatically find the required video segments. Depending on the content of the submitted video, video retrieval is generally divided into shot retrieval and segment retrieval. This type of service also involves users from multiple locations interacting with the same platform at the same time.

6.3.5.1 Online education

Online education is the interactive/non-interactive multimedia communication between learners and education contents located in different locations. Learners utilize online education services to obtain certification, participate in business trainings or independently learn new knowledge. Online education services can also include teaching centres, which are similar to traditional face-to-face learning in classrooms, individual tutoring, self-study, multirole learning, team learning, etc.

In online education, information may come from a remote database containing education resources or live lectures. The material can be textual, auditory and graphic or video, or it can be stored in a multimedia format. This information can be transmitted in communication configurations such as point-to-point, point-to-multipoint or multipoint-to-multipoint.

Learners may be located in classrooms equipped with related facilities, or in other places such as offices or homes, where they can access the online education service platform. The learner may perform real-time interactions with others during the lessons, he or she may learn according to the class schedule, or may learn using on-demand non-real-time courses. The learner may use a PC, tablet, mobile phone or even a TV with an STB. Learners can change equipment with the assistance of online education service platforms without interrupting learning.

The online education service system is composed of educational resources, an online education service platform and learners. Materials can be obtained from education resources by learners through the online education service platform.

Education resources include live courses broadcast by teachers through audiovisual classrooms, digital video recordings of live courses stored on the server, special software production based on course lectures given by teachers, digital materials (visual library, visual laboratory) and digital news.

In VoD multimedia courseware teaching, the teaching content is pre-made, and the user selects it through the on-demand form.

In online classroom teaching, users can select the content and participate in classroom Q&A interaction through their camera, microphone and so on.

Online education is closely related to video conferencing, VoD, IPTV and web-based applications. The difference is that the goal of online education is teaching and learning, the information used is related to educational resources and learning records, and the participants are teachers and students.

6.3.5.2 Telemedicine

Telemedicine can take many forms, and the most common is remote diagnosis and communication between doctor and patient. In addition to online chat functions on websites, remote telemedicine consultations can be carried out through mobile applications (apps). Currently, this mainly focuses on text and pictures. Combined with the current development of social software, there should be no technical difficulties in expanding to video communication in the future. Considering the home environment, this will be a new type of video service that incorporates dedicated software.

Combined with the development of wireless technology such as 5G, there are also operations for controlling medical robots. This requires the cooperation of hospital related equipment and is not suitable for home environments.

Telemedicine and consultation services involve interactive multimedia communication between medical experts located in multiple locations. This communication is usually initiated by a doctor who needs to discuss a specific patient and communicate with other relevant experts. This may take place between the doctor and a consultant, or it may require the doctor to arrange an interactive meeting with several consultants at the same time.

During the consultation process, the required information may be extracted from a remote database containing the patient's medical files: for example, X-rays, ultrasound scans, electrocardiograms or similar medical images are retrieved from several diagnostic testing centres; or technology information, medical imaging instruction or other assistance material are extracted from the reference library. These materials can be text, audio, graphic or image, or they can be stored in a multimedia format.

The partners involved in the consultation may be located in an office or medical facility, which has a full range of broadband multimedia communication facilities. They also may be located somewhere with limited communication access. In order to cope with all possibilities, the requirement to support dynamic resource allocation provisions, both in the "call" initiation and "call" in progress, needs to ensure that more important interaction needs can also be fully met.

6.3.5.3 Online live broadcasting

Online live broadcasting is mostly webcast. It refers to the use of PC, Internet STB, smart mobile phone, tablet and other terminal as carriers, relying on webpage or client streaming technology, to show real-time spatial scenes and activities such as teaching, art creation and art of acting. Hosts and subscribers can interact through multiple online methods. This combines various elements such as images, text and sound and has become a mainstream expression form on the Internet. It currently mainly includes game broadcasts, commodity show broadcasts, live answers, media live broadcasts and host live broadcasts.

A complete streaming media platform should include the following parts:

- Coding tool. This is used to create, capture and edit multimedia data to form a streaming media format, which can be completed by a computer with audio and video hardware interface and the related software.
- Streaming media data. This is convenient to encapsulate the audio and video data for IP network transmission, mostly adopted in some encoding mode such as H.264.
- Server. This is used to store and control the data of streaming media.
- Network. The Internet is suitable, with multimedia transmission protocol or real-time transmission protocol.

•	Video player. Flash players and webpages such as HTML5 are generally used for clients to browse streaming media files.

The process of live webcasting includes video collection, video processing, video encoding and encapsulation, streaming push, streaming distribution and streaming playing (see Figure 6-13).

Video collection is the first step in the entire video push streaming process. It captures the original video data from the system's capture device and outputs it to the next step. Video collection involves two aspects of data collection: image collection and audio collection, which correspond to two completely different input sources and data formats.

After the video or audio is collected, the original data is obtained. In order to enhance some live effects or add some additional effects, it is generally processed before encoding and compression, such as time stamping or watermarking of the company logo, actor or host beautification, and sound processing. Video processing includes the use of artificial intelligence (AI) algorithms to capture specific behaviours in the picture, such as some people not following dress requirements in particular places.

Encoding is to convert the original data into more concise data via a specific method. The principle of encoding is to reduce the amount of video data transmission by removing redundant information in the original data without affecting the subscriber's QoE.

Encapsulation is used to mix multimedia information encoded with specific encoding standards, including video stream, audio stream, subtitle, chapter information, into a video file. It can provide an index for multimedia content.

Streaming push refers to the process of transmitting packaged content to the server.

Streaming pull refers to the process of pulling live content on the server and using the specified address to pull.

There are several types of push-pull streaming protocols: real-time streaming protocol (RTSP), real-time messaging protocol (RTMP), hypertext transfer protocol flash video (HTTP-FLV), HLS and private protocols of various manufacturers.

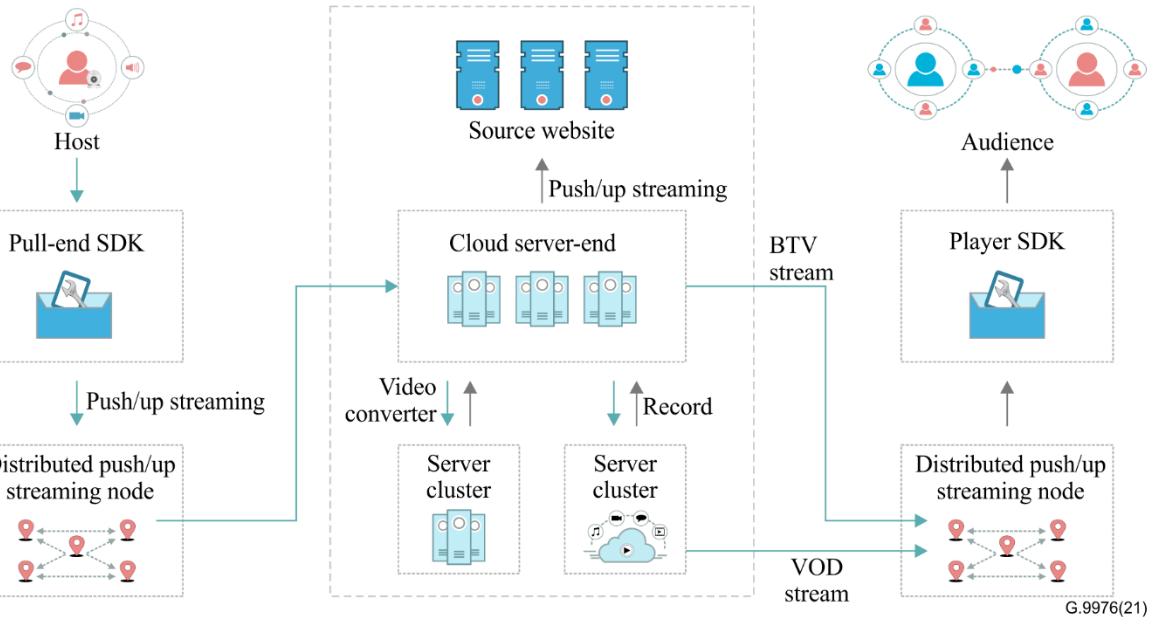

Figure 6-13 – Schematic diagram of online live broadcasting

6.3.5.4 Multitype of video services and concurrent scenario within the home

With the development of network technology, multiple family members may simultaneously use multiple services in a home environment during a certain period, including online working mainly based on video conferencing, online education mainly based on live webcast and video distribution (VoD and BTV) for entertainment. This brings additional requirements for the home network. Measured by "communication task characteristics", the following most stringent requirements for video streaming must be met, such as "multipoint-to-multipoint" (MP2MP) communication configuration, "bidirectional-symmetric" (BD) for video streaming, "source and sink" bidirectional transmission control entity, "real-time" communication delay, video and various media components and "isochronous" time continuity. Home networking also needs to meet the requirements of the coexistence of multiple media streams and multiple types of terminals for displaying and operating. All the above places more complex and stricter requirements on the home network than before.

6.4 Network capability for supporting video

UHD video services are characterized by high throughput, high concurrency, high QoE and low latency. As the proportion of video traffic increases, it brings many challenges to the traditional carrier network. The network KPI indicators that affect QoE mainly consider the following aspects.

The first challenge is network architecture. The features of traditional telecom network (e.g., high convergence ratio and multilevel model) need to be simplified. The home network matches video traffic with a low convergence ratio.

The second challenge is throughput. The operators provide a promised access throughput to fixed broadband subscribers. The home network needs to be checked for whether its access throughput with various physical media can meet the requirement on throughput for 4K video.

The third challenge is latency. The latency is the RTT between the subscriber's terminal and the server where the video fragment file is located. This indicator can be applied to both fixed broadband Internet users and mobile Internet users. From the view of the E2E link, the metro backbone and control layer are prone to congestion, which leads to increased latency and latency jitter. The requirements of transmission latency for network physical media, etc., should be specified in the home network.

The fourth challenge is PLR, i.e., the two-way PLR of the link between the terminal and the server where the video fragment file is located. This indicator applies to both fixed broadband Internet users and mobile Internet users. Among the entire E2E link, there is also the risk of burst packet loss in the metro core and aggregation nodes. The PLR of home WLAN needs to be studied. Devices such as RGW and STBs may only have a small buffer and therefore they are prone to encounter burst packet loss.

At the same time, attention needs to be focused on the challenges of cost and OAM. The video service drives traffic to increase exponentially. If it is expanded in a traditional way, the cost is high. The video service experience is sensitive to network latency and packet loss and has high real-time performance. The carrier network should be able to quickly sense the user's experience and quickly handle faults.

The recommended values for the carrier network KPI are shown in Table 6-2.

Table 6-2 – Network requirements for excellent experience with 4K video service

		VoD		BTV	
		Without cloud optimization	**With cloud optimization**	**Without retransmission (RET)**	**With RET**
Average bit rate of 4KVideo		$\geq$ 25 Mb/s	$\geq$ 25 Mb/s	$\geq$ 30 Mb/s	$\geq$ 30 Mb/s
Network indicator	E2E throughput	$\geq$ 100 Mb/s	$\geq$ 37.5 Mb/s	$\geq$ 56 Mb/s	$\geq$ 56 Mb/s
	RTT	$\leq$ 20 ms	$\leq$ 20 ms	/	/
	PLR	$\leq 3.4 \times 10^{-5}$	$\leq 2.4 \times 10^{-4}$	$\leq 1 \times 10^{-6}$	$\leq 1 \times 10^{-4}$
NOTE – Cloud optimization refers to the optimization technology between the video server and the STB, which mainly includes improving the transmission efficiency through TCP optimization and reducing the demand for throughput bursts by stream smoothing, thereby reducing the requirements on throughput and PLR.					

Home networks need to ensure efficient coverage of high-speed networks to support 4K/8K services through a combination of wired (coaxial, phone line, power line, etc.) and wireless links. At the same time, they also need to meet the following requirements:

- Support different terminals such as TV, mobile phone and tablet to seamlessly share content of 4K/8K BTV and VoD services in the home network;
- Realize secure transmission of 4K/8K video content in the home network;
- The home network needs to have some abilities related to QoS such as self-test, self-analysis and transmission adaptive. Troubles during high-speed transmission shall be automatically and timely discovered.

7 Video deployment in the home network

7.1 Distribution of 4K UHD video streams throughout the home

As discussed in clause 6.2, the data rate of one 4K video stream is in the range of 15 to 100 Mb/s, depending on the video quality. This brings significant implications for the home network technology and places additional requirements on the RGW.

One key factor is how many streams are simultaneously transmitted and where they are sent within the home. For only a single video stream sent to a device, e.g., STB/TV, there may not be a problem since in most of the cases, the STB/TV is located in the same room as the RGW and can be connected using an Ethernet cable from the RGW with an available Gigabit Ethernet (GE) port.

The problem appears when more than one UHD-capable device needs to be supported and/or the RGW is not in the same room as to the video end device. In this case, wireless connection may be one of the possible choices to build up the link. However, wireless coverage and performance to a specific end device within a given home mainly depend on the relative position between the access point (e.g., RGW) and the end device. The supported profile (such as frequency band, antenna number, transmission power) and vendor design of both the access point (AP) and end device also affect the actual performance.

To overcome this drawback, one or more additional APs can be used to extend the coverage, but each of the APs needs robust backhauling links, either wired (e.g., Ethernet, G.hn, FTTR) or wireless (e.g., Wi-Fi) backbone, to the network terminal (e.g., RGW). Different technologies suffer from dedicated challenges for conveying video over the backhaul links. Wi-Fi makes use of the shared unlicensed spectrum, which may generate sudden jitter when packets drop in collision. Phone line, coax and fibre are ideal media for video transmission in the short distances of the home scenario but need to be deployed in advance. Power line is available anywhere and is good for a video end device to connect to the network if a stable data rate can be established.

7.2 Network topology

The practical applications of video create significant difference in the requirements on the network topology. Table A.2 provides a complete analysis of topology requirements based on the service content:

- In private online conversation (such as video surveillance systems), the video connection is established through point-to-point architecture;
- To distribute the same video resources to multiple video end devices, point-to-multipoint (such as broadcast from the RGW) is needed to boost the transmission efficiency;
- Simultaneous recording/playback video in multiple rooms to a network attached storage (NAS) system requires the network to convey the video between multiple locations within the home and one site, forming a multipoint-to-point scenario;
- In-premises conference actually shares video and audio information.

8 Performance requirements for supporting video in home networks

The family of ITU-T G.996x specifies the system architecture, physical layer (PHY) and data link layer (DLL), etc. They focus on wire line-based home networking transceivers which are capable of operating over telephone, coaxial and power-line wiring.

Since the cost of Ethernet lines and fibre is declining, deployments using these technologies are becoming more popular. In addition, transmission technologies based on Ethernet and fibre are also under continuous development.

Finally, UHD video services technologies are also continuing to develop.

8.1 Target architecture of carrier network for video service

Based on the requirements present in previous clauses, the target network architecture of a video carrier network is shown in Figure 8-1, which includes five components.

The first is the home network. A home with ≥ 100 Mb/s coverage solves the seamless ultra-broadband coverage in the home.

The second is ultra-broadband access. This includes ultra-broadband home and ultra-broadband access solutions, providing at least 100 Mb/s throughput capability and solving the problem so that subscribers can start their 4K service.

The third is the simple metropolitan area. It includes IP simplification and optical transport network (OTN) to central office (CO) solutions, with the best total cost of ownership (TCO) to carry tens of millions of 4K video subscribers.

The fourth is optimized bearer network. Under the existing network conditions, the deployment of targeted solutions can help to improve a subscriber's experience and reduce costs. The solutions maybe include QoS deployment and OTN multicast.

The fifth is QoE management. After the service is started, it is necessary to actively manage the experience, quickly discover problems with a subscriber's video experience and solve the problems at the front-end to avoid subscribers complaining and leaving the network.

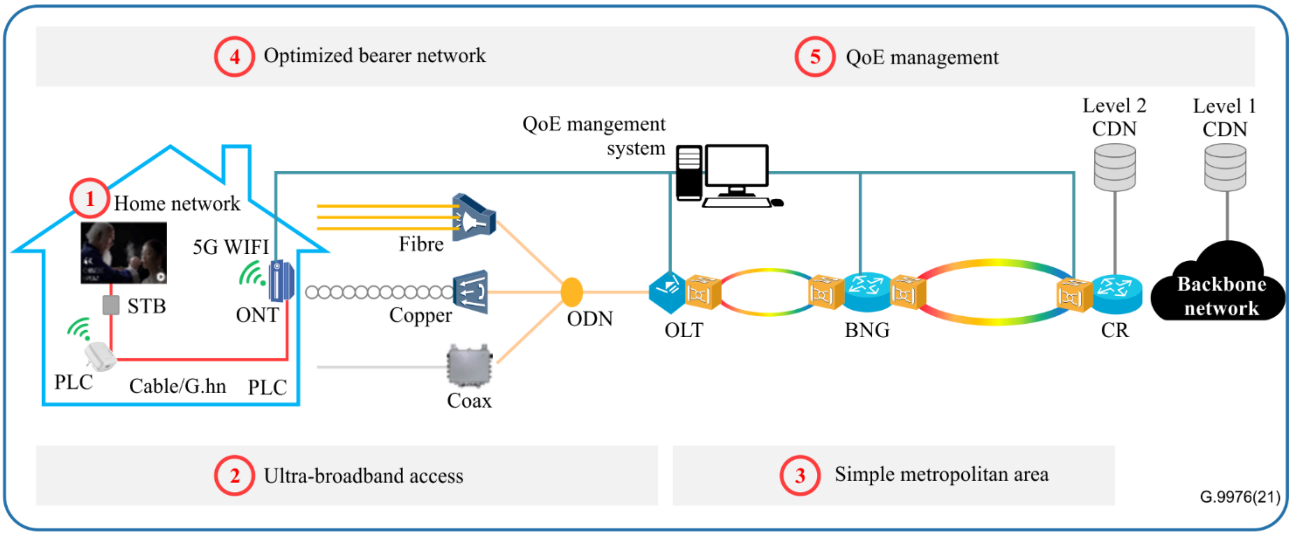

Figure 8-1 – The target architecture of the video carrier network with the best experience

A video carrier network with optimal experience does not need to be deployed in one step. Operators can gradually evolve to the target network based on their actual service development. For example, in the initial stage of a video service, the focus is on the rapid development of subscribers, so the first concern is the rapid coverage of the access network and rapid throughput improvement; in the stage of scale development, a simple metro network is required to relieve traffic pressure, reduce the probability of congestion and consider deployment carrier network optimization and QoE management to reduce subscribers' complaints and customer churn rate.

8.2 Throughput requirements

There are two types of TV signals: analogue and digital. They can be transmitted through large-capacity, wide-band analogue channels or digital channels. Digital channels can reduce noise accumulation and improve transmission quality. TV signals can be transmitted in three forms: terrestrial broadcasting system, satellite TV broadcasting system and cable television system (e.g., community antenna television (CATV)).

- Terrestrial broadcasting system: Transmits electromagnetic wave signals to the surrounding area through the TV transmitting antenna. The advantages are low cost, wide coverage and the places where a TV set can be used are not restricted. The disadvantage is that the TV signal is easily blocked and reflected by ground obstacles (such as high buildings), resulting in multipath interference, and the image often has ghosting. In addition, the TV signal intensity is inversely proportional to the square of the distance. In areas far away from the TV station, the received TV signal intensity is very weak, and it is susceptible to interference from surrounding signals (such as cars, electrical equipment, household appliances), weather and background noise, thus snowflake-like interference bars often appear on the screen. Since the coverage area of a TV transmitting station is the light of sight (LOS) range of the TV transmitting antenna and the receiving antenna on a TV set, the coverage range is from only tens of kilometres to around a hundred kilometres. If one city's TV programmes are to be transmitted to another, distant city, the signal needs to be transmitted over long distances via satellite channel or OTN.
- Satellite TV broadcasting system: Its main advantages are larger coverage, high quality of TV broadcast and strong adaptability. Its disadvantage is high cost, and it requires the user to purchase a satellite receiver STB and install a parabolic microwave antenna.

- Cable TV system: Transmission through coaxial cable, free from external interference and interference with others, so that the spectrum can be fully utilized, and the image quality is the best among all transmission systems.

Early cable TV users can choose any TV channel to watch because all the programme contents are pushed to them. In the same cable network, each TV user has the same throughput of the programme on the coaxial line. The key requirement for UHD transmission capacity is how many UHD video streams need to be carried simultaneously, and the locations of corresponding terminals (i.e., screens) within the home. For a single stream to an STB/TV in the same room as the RGW, as long as the RGW has a GE interface, a direct Ethernet connection can be used. However, problems may appear if it is necessary to support more than one UHD video-capable device, and/or when the RGW and the terminal display device are not in the same room.

With the development of video technologies, the following factors may affect the throughput requirements of home networks:

- The resolution of various video programmes has been determined, but the video service resolution requirements and video compression coding algorithms and standards will continue to develop (e.g., further development to 8K and VR/AR), causing changes in throughput requirements.
- Different terminals can receive different video streams. Even in cable TV, STBs located in different rooms within the same home network may play different programmes, so total throughput demand will see a substantial increase.
- How to meet the different E2E physical throughput requirements. For example, if the target is to have no video pause taking place during VoD mode, it is necessary to appropriately increase the throughput requirements of home networking.
- How to evaluate the throughput requirement model to carry 4K video for the home network. The equipment's capacity needs to be designed taking into consideration support for 8K in the future, scalability on protection (i.e., actual throughput lower than 50% of the capacity of the interface) and maintaining sufficient margin.

8.3 Latency requirements

As shown in Figure 8-2, the E2E transmission architecture of a UHD video service consists of a home network, an access network, a metro network, IP core network and IPTV source e.g., a CDN/video server. The latency affects whether dappled screen occurs. The UHD video service requires an E2E latency of no more than 20 ms. To meet this requirement, there are currently four major challenges on the E2E link:

- The first is that Wi-Fi in home network connections utilizes unlicensed spectrum and contention-based mechanisms, which may introduce relatively larger latency;
- The second is that some access layers have a large latency;
- The third is that the metro network coverage is large, so the impact of fibre transmission latency cannot be ignored;
- The fourth is that the latency within IP equipment in the metro network layer is uncertain.

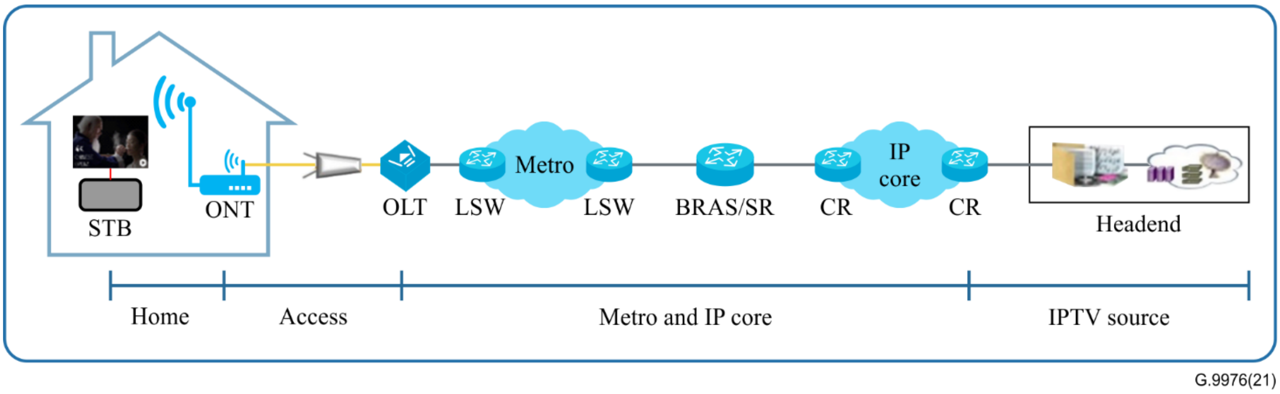

Figure 8-2 – Decomposition diagram of bearer network for 4K

As shown in Table 8-1, current copper access technologies (such as VDSL2, Vectoring, Super Vector, G.fast, etc.) introduce a latency of 10~20 ms. It is recommended to optimize the latency of the access network through optical fibre access and minimize the latency of the metro network. Considering that the transformation of the access network is characterized by high complexity, long cycle and high cost, xDSL will still coexist with FTTx for some time. The latency of FTTH is low, about 3~5 ms. The latency of copper line access is larger than optical fibre.

Table 8-1 – Typical latency for various copper wire access technologies

Access technology	Throughput (downstream)	RTT(ms)
VDSL2	50 M @ < 1000 m	10~20
Vectoring	50 ~ 120 M @ < 800 m	10~20
Super Vector	100 ~ 300 M @ 300 ~ 500 m	10~20
G.fast	200 M ~ 1.2 G @ 100 ~ 500 m	2~6

In the home network, it needs to be clarified how to meet the latency requirements of UHD video services in multiple media environments.

8.4 PLR requirements

A 4K carrier network also needs to meet a requirement of 10^{-5} for PLR. The reasons for packet loss are poor line quality and congestions on the network node.

The main risk of line quality depends on the connection within the home network. Copper wire access will also cause error packet loss due to the line. As shown in Table 8-2, each of the current copper access technologies introduces a PLR of 10^{-4} ~ 10^{-5}. Because of signal interference, the bit error rate (BER) of WLAN is as high as 10^{-3}, it is recommended that the STB access the RGW through a wired connection.

Table 8-2 – Typical PLR for various copper access technologies

Access technology	Throughput (downstream)	PLR
VDSL2	50 M @ < 1000 m	$10^{-4} \sim 10^{-5}$
Vectoring	50 ~ 120 M @ < 800 m	$10^{-4} \sim 10^{-5}$
Super Vector	100 ~ 300 M @ 300 ~ 500 m	$10^{-4} \sim 10^{-5}$
G.fast	200 M ~ 1.2 G @ 100 ~ 500 m	$10^{-4} \sim 10^{-5}$

Avoiding node congestion is an important goal in the design of a network architecture. Since the OLT downlink to the user side is non-blocking, congestion may occur in the metro carrier layer between the OLT and the CDN. The metro carrier network needs to deploy QoS and guarantee light load and low convergence ratio. At the same time, because burst congestion is difficult to avoid, so the buffering capability of equipment and customer premise equipment (CPE) is necessary to absorb instantaneous bursts of traffic and avoid or reduce the impact on QoE.

In home and access networks, because of the diversity of technologies and complex environments, the PLR is difficult to guarantee. Therefore, in order to improve the QoE and cope with the complex access side environment, the PLR should be reduced as much as possible. The goal when constructing a carrier network with the best QoE should be to achieve a situation with no packet loss.

9 Supporting video transmissions over ITU-T G.hn technology

This clause discusses the necessary capability of the G.hn node to support UHD video in a home network.

9.1 General indoor system based on G.hn node

Figure 9-1 describes the general architecture of a home network using ITU-T G.hn technology. Wireline mediums are considered as the backhaul link for wireless access points. In this architecture, power line and coax cable are the sharing medium while phone line belong to point-to-point (P2P) topology.

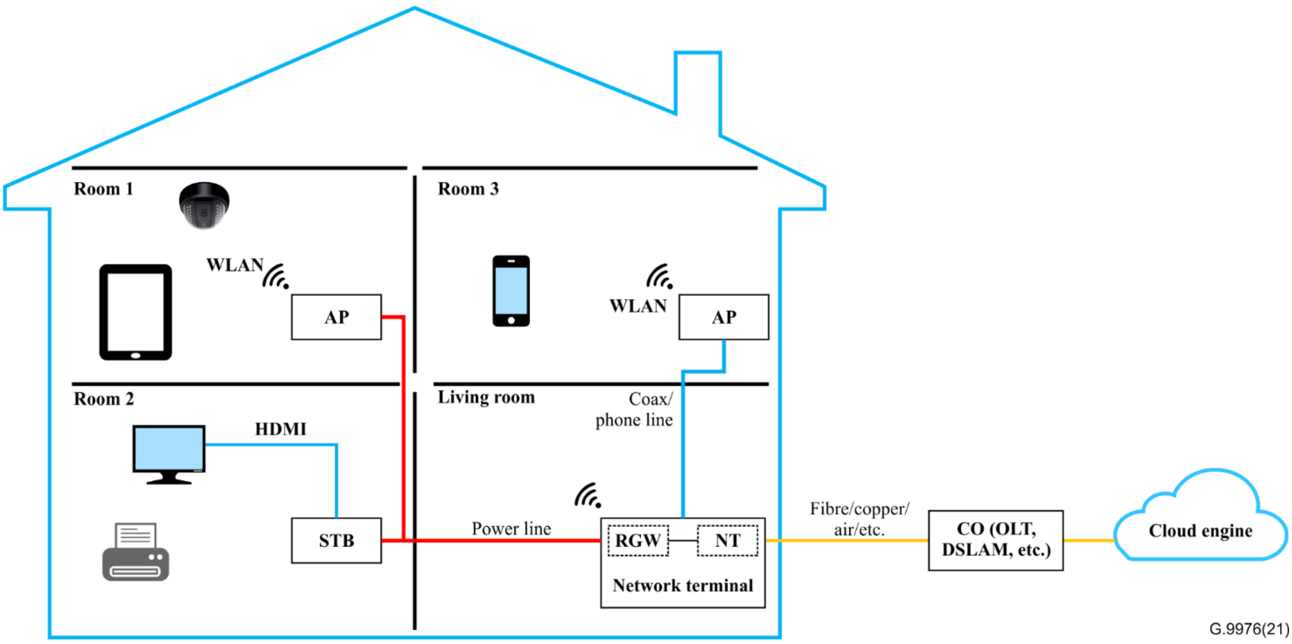

Figure 9-1 – General architecture of home networking using ITU-T G.hn technology as wireless backhaul

The network terminal in Figure 9-1 connects to the broadband access network through northbound interfaces by access technologies, such as DSL, DOCSIS and PON. Extension of the LAN by ITU-T G.hn technology can be achieved by making use of available wireline medium. WLAN or direct video connection such as HDMI convey the "last metre" video data to the end device (e.g., television, projector, HD monitor). In order to support the service requirements of UHD video transmissions, the following clauses describe the requirements that are recommended for home network devices and G.hn interfaces/links in particular.

9.2 Requirements for devices in home networks

9.2.1 Network terminal

9.2.1.1 The northbound interface of a network terminal (NT) shall support one or more of these access interfaces: xPON, 1000Base-T and DSL (e.g., VDSL2, G.fast).

9.2.1.2 The southbound interface of the NT shall support at least a G.hn interface and 1000Base-T.

9.2.1.3 The G.hn interface of the NT shall be able to support centre control function, as the role of controller under EasyMesh protocol and domain master of G.hn interface (responsible for maintaining the home network, e.g., network topology, device registration).

9.2.1.4 The NT shall be able to support management and coordination of multiple APs.

9.2.2 Access point

9.2.2.1 The northbound interface of an AP shall support one or more of interfaces: G.hn interface and 1000Base-T.

9.2.2.2 The southbound interface of the NT shall support one or more of interfaces: Wi-Fi 5 (dual-band 2 × 2 profiles or above), Wi-Fi 6 (dual-band 2 × 2 profiles or above) and 1000Base-T interface.

9.2.3 G.hn node

9.2.3.1 The G.hn interface shall support 100 MHz OFB on power line/twisted pair/coaxial line medium to support physical layer rate up to 1 Gb/s.

9.2.3.2 The southbound interface of the NT shall support one or more of interfaces: 1000 Base-T interface.

9.2.3.3 MIMO should be supported on the power line/CAT5/CAT6 medium to achieve aggregated data rate by two data streams.

9.2.3.4 The G.hn interface shall support 4096 QAM modulation of 12 bits per subcarrier.

9.2.3.5 The G.hn interface shall support RCM mode for data stream to build up robust communication channels.

9.2.3.6 The G.hn network shall support the topology maintenance function, real-time monitoring of the link status of the home networking topology.

9.2.3.7 The G.hn network shall be able to detect and coordinate with neighbour networks.

9.2.3.8 The G.hn network should support mesh function to form the G.hn network.

9.2.3.9 The G.hn network shall be able to maintain at least four nodes for data transmission.

Annex A

Classification method for UHD video services

(This annex forms an integral part of this Recommendation.)

The classification method of video services can be divided into three steps as shown in Figure A.1.

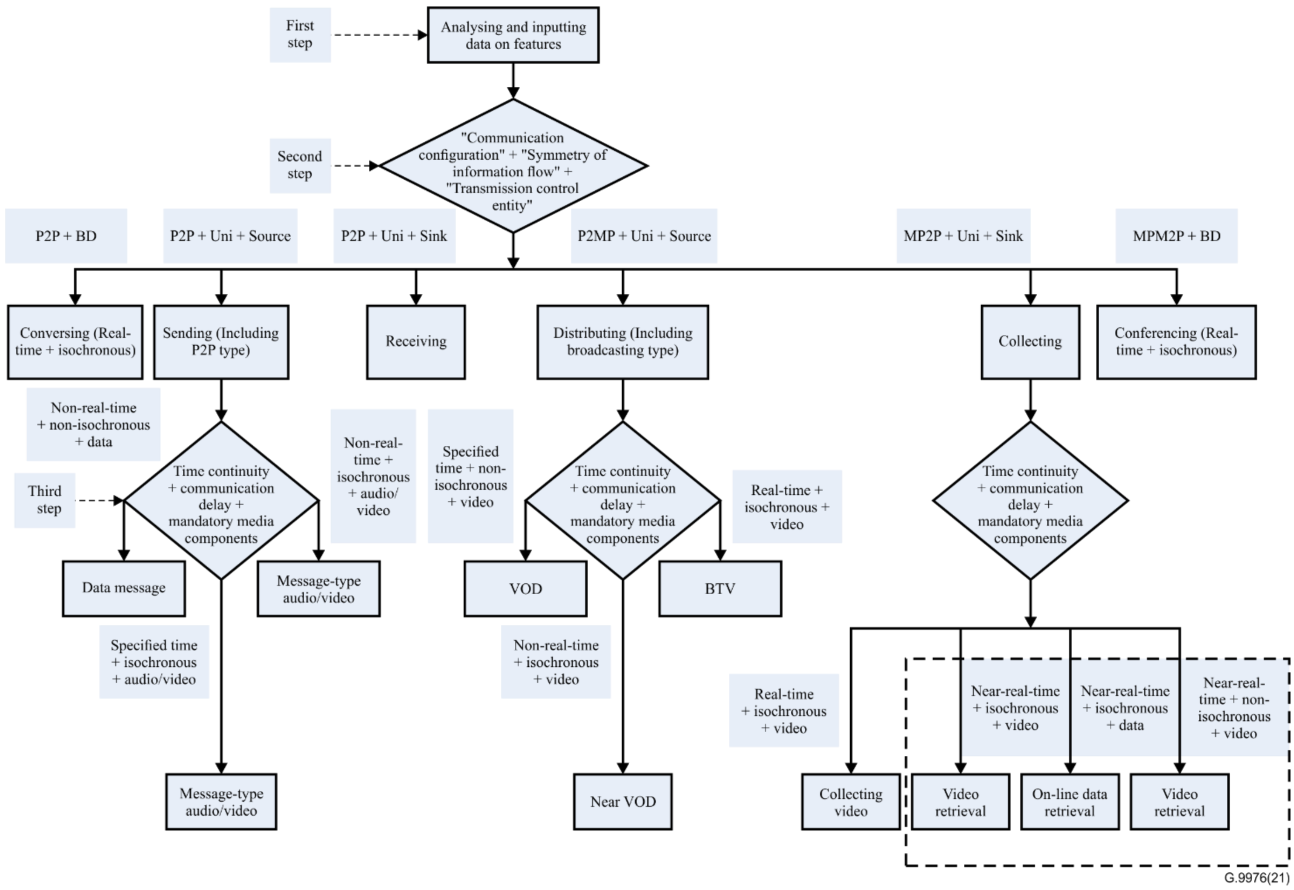

Figure A.1 – Classification method for video services

The first step is to analyse the eight characteristics of various services according to Table A.1 and analyse which characteristics of the video services to be classified are different.

Table A.1 – Communication task attributes [ITU-T F.700]

Attributes	Possible values
Communication configuration	Point-to-point (P2P)/point-to-multipoint (P2MP)/multipoint-to-point (MP2P)/ multipoint-to-multipoint (MP2MP)
Symmetry of information flow	Unidirectional (Uni)/bidirectional-symmetric (BD)/bidirectional-asymmetric
Transmission control entity	Source/sink/source and sink/third party
Communication delay	Real-time Near-real-time Non-real-time Specified time
Mandatory media components	Audio/video/text/still picture/graphics/data

Table A.1 – Communication task attributes [ITU-T F.700]

Attributes	Possible values
Optional media components	Audio/video/text/still picture/graphics/data/none
Media component interrelations	1) Synchronization between: a) Audio and video (lip synchronism, location related synchronism); b) Audio and text (voice synthesis); c) Text and video/still picture/graphics (subtitles synchronized with images); d) Graphics and audio. 2) Symmetry between media components of the same type to allow for bidirectionality. 3) Conversion between information types (or media components).
Time continuity	Isochronous/non-isochronous

The second step is to classify according to the characteristics of the actual configuration of the video service in the communication network, using the combination of the three parameters i.e., "communication configuration", "symmetry of information flow "and "transmission control entity" in Table A.2. This step of classification is based on the different requirements for the communication network and the terminal equipment, so it is the basis of various classification methods.

Table A.2 – Provisional list of communication tasks

Communication configuration	Communication tasks for three types of information flow		
	Bidirectional	Unidirectional source controlled	Unidirectional sink controlled
Point-to-point (P2P)	Conversing	Sending	Receiving
Point-to-multipoint (P2MP)	The video of the chairman or main venue of the conference scene is watched by the rest of the staff or branch venues	Distributing (VoD and BTV)	When watching a live TV show, the viewer decides whether to watch it.
Multipoint-to-point (MP2P)	Each party of the conference scene sends videos to the main site or chairperson	Upload video content in different locations	Collecting
Multipoint-to-multipoint (MP2MP)	Conferencing		
NOTE – Multipoint-multipoint usually means that each of these points can send to multiple points, and each two points should belong to a bidirectional channel, so there is no such unidirectional situation. This table is based on Table 3 of [ITU-T F.700], adding some other types of services.			

The third step is that existing communication services are becoming more and more sensitive to time characteristics. Therefore, the two time related parameters i.e., "communication delay" and "time continuity" in Table A.3, will be used in combination with "mandatory media components" to construct a classification method. This step of classification is based on the fact that various types of communication services should pay close attention to a user's QoE, which is closely related to latency and continuity.

Table A.3 – Classification method combining of communication delay and time continuity

Communication delay	Service examples	
	Isochronous	**Non-isochronous**
Real-time	Conversation, BTV, collecting	N/A (Note 1)
Near-real-time	On-line audiovisual retrieval (Note 2)	On-line data retrieval (Note 2), video retrieval
Non-real-time	Near VoD	Data messaging (P2P and P2MP)
Specified time	Audiovisual messaging (Note 2)	Audiovisual messaging (Note 2), VoD distributing

NOTE 1 – Real-time services do not require storage and presentation, and data video information that lags in transmission will be discarded directly.

NOTE 2 – Isochronism depends on the terminal device storage and capture capacities.

This table is based on Table 2 of [ITU-T F.700] and add some contents.

What needs to be clear is that "communication delay" has a quantifiable standard that can be clearly defined. Other characteristic classifications are based on characteristics similar to the overall connection topology, and the type parameter selection can be set directly, without real-time data collection and quantitative analysis.

The two time related parameters of "communication delay" and "time continuity" directly affect a user's QoE, and the effect is similar to service failures such as video pause (i.e., stalling) and black screen caused by the unsmooth video stream in "viewing" in KQI. Based on the "communication task attributes" in Table A.1, combining the two time related parameters of "communication delay" and "time continuity" that have a greater impact on a user's QoE, and taking into account the classification of media components, common video services can be classified. Figure A.1 shows the main scenarios and further examples for common services.

Annex B

Calculation algorithm model and equations for U_MOS

(This annex forms an integral part of this Recommendation.)

B.1 Calculation algorithm model

The U_MOS model mainly divides specific functional scenarios into session scenario and instant scenario. In session scenario, the user's viewing behaviour is longer than 1 minute at a time. The instant scenario is usually used for real-time quality monitoring.

$$U\_MOS = (Qav - 1) \cdot (1 - c_1 \cdot (5 - Qi) + c_2 \cdot (5 - Qv)) + 1 \quad \text{(B-1)}$$

Q_{av} is the overall quality of video and audio experience:

$$Qav = \alpha + \beta \cdot Qs + \gamma \cdot Qa \cdot Qs \quad \text{(B-2)}$$

Parameters c_1 and c_2 are the dynamic weighting coefficients of interactive experience, i.e., Q_i, and viewing experience, i.e., Q_v. The initial c_1 and c_2 may be obtained from the results of data surveys done by the service operator. At the same time, the weighting factors α, β and γ are superimposed. When the score of (Q_s, Q_a, Q_i, Q_v) changes, the weighting factors α, β and γ will be adjusted accordingly.

Table B.1 – Variables for calculation U_MOS in instant scenario and session scenario

<table>
<tr><th colspan="3"></th><th>Instant scenario</th><th>Session scenario</th></tr>
<tr><td rowspan="3">BTV</td><td rowspan="2">RTSP/RTP</td><td>Error hide mode</td><td>Qs, Qa,
Qi=Qi_ZapInstant
Qv=Qv_STInstant</td><td>Qs, Qa,
Qi=Qi_ZapSession
Qv=Qv_STSession</td></tr>
<tr><td>Error no hide mode</td><td>Qs, Qa,
Qi=Qi_ZapInstant
Qv=Qv_STInstant +Qv_BLInstant</td><td>Qs, Qa,
Qi=Qi_ZapSession
Qv=Qv_STSession +Qv_BLSession</td></tr>
<tr><td colspan="2">HTTP</td><td>Qs, Qa,
Qi=Qi_LoadInstant
Qv=Qv_STInstant</td><td>Qs, Qa,
Qi=Qi_LoadSession
Qv=Qv_STSession</td></tr>
<tr><td rowspan="3">VoD</td><td rowspan="2">RTSP/RTP</td><td>Error hide mode</td><td>Qs, Qa,
Qi=Qi_LoadInstant
Qv=Qv_STInstant</td><td>Qs, Qa,
Qi=Qi_LoadSession
Qv=Qv_STSession</td></tr>
<tr><td>Error no hide mode</td><td>Qs, Qa,
Qi=Qi_LoadInstant
Qv=Qv_STInstant +Qv_BLInstant</td><td>Qs, Qa,
Qi=Qi_LoadSession
Qv=Qv_STSession +Qv_BLSession</td></tr>
<tr><td colspan="2">HTTP</td><td>Qs, Qa,
Qi=Qi_LoadInstant
Qv=Qv_STInstant</td><td>Qs, Qa,
Qi=Qi_LoadSession
Qv=Qv_STSession</td></tr>
</table>

B.2 Calculation algorithm for Q_s

$$Qs = f(ScreenSize, VideoComplexity, Resolution, BitRate, CodecType, VideoFrameRate, Reserved\ldots) \quad \text{(B-3)}$$

According to its processing signal level, the Q_s module mainly includes three-layer models such as Model0, Model1 and Model2, which correspond to the parametric model, bitstream model and hybrid model, as described in clause 6.1.

B.2.1 Parametric model (Model 0)

$$Qs' = f(PPI, ScreenSize) = c_7 \cdot (1 - \frac{1}{1+(\frac{PPI}{c_8 \cdot ScreenSize^{c_9}})^{c_{10}}}) \quad \text{(B-4)}$$

$$Qs = Qs' - \frac{Qs' - c_{11}}{1+\left(\frac{BitRate}{c_{12}}\right)^{c_{13}}} \quad \text{(B-5)}$$

Q_s' represents the display quality, that is, the highest quality score that can be obtained under the conditions of fixed pixels per inch (PPI) and screen size (such as 720P video playing on 42" TV); C_{11} represents the corresponding lowest score, and 1.0 can be used.

According to different codec types (e.g., H.265, H.264, MPEG-2, AVS2, etc.), there are different corresponding values of C_{12} and C_{13}.

B.2.2 Bitstream model (Model 1)

$$Qs' = f(PPI, ScreenSize) = c_7 \cdot (1 - \frac{1}{1+(\frac{PPI}{c_8 \cdot ScreenSize^{c_9}})^{c_{10}}}) \quad \text{(B-6)}$$

Equation B-6 is the same as Equation B-4.

$$Qs = Qs' - (5 - Qcod) * \frac{(Qs' - 1)}{4} \quad \text{(B-7)}$$

Q_{cod} is the quality analysis result of the video sequence, the value range is [1, 5], and it is linearly mapped to the value range [1, Qs'].

$$Qcod = f\left(FrameType,\ TotalBytesPerFrame,\ Qp,\ MV,\ SkipRatio\right)$$
$$= f\begin{pmatrix} FrameType,\ TotalBytesPerFrame, \\ \left[Qp\_max,\ Qp\_avg,\ Qp\_min\right], \\ \left[MV\_max,\ MV\_avg,\ MV\_min\right], \\ SkipRatio \end{pmatrix} \quad \text{(B-8)}$$

$$Qcod = kfr_{\text{impact}} \cdot exp\left(n_1 \cdot \left(QP\_\ fr_{\text{impact}} + cpx_{\text{video}} + motion_{\text{impact}}\right)\right) \quad \text{(B-9)}$$

Where "*FrameType*" means I/P/B frame in GOP. "*QP*" means quantization parameter.

$$kfr_{\text{impact}} = n_2 \cdot kfr + n_3 \quad \text{(B-10)}$$

Where *kfr* represents the ratio of the video frame rate to the number of frames between two adjacent I-frames.

$$Qp\_fr_{\text{impact}} = n_4 + n_5 \cdot \left(\frac{avgQP_{pic}}{51}\right)^{n_6} + n_7 \cdot \frac{1}{fr} + n_8 \cdot \text{intraflicker} + n_9 \cdot (\text{maxQP}_{\text{pic}} - \text{minQP}_{\text{pic}}) \quad \text{(B-11)}$$

Among them, *intraflicker* is a Boolean variable, which indicates whether there is I-frame flicker in the received frame. The so-called I-frame flicker means that the *QP* value of a received I-frame is obviously suddenly higher than that of the previous and subsequent I-frames.

$$cpx_{video} = \min\left(\sqrt{\frac{br_v}{AvgByte_{I-frame}}} + n_{10} \cdot skipRatio, 1.0\right) \quad \text{(B-12)}$$

$$motion_{\text{impact}} = n_{11} \cdot avgMV_{pic} \cdot \left(1 - \frac{fr}{30}\right) \quad \text{(B-13)}$$

B.2.3 Hybrid model (Model 2)

$$Qs = f(Blockiness, Blurriness, Contrast, Noise) \quad \text{(B-14)}$$

Blockiness refers to the discontinuity of block boundaries caused in the encoding process. When the *blockiness* is severe, the video will have obvious blocky defects, which will affect the visual effect and reduce the QoE during viewing.

$$blockiness = (-10.38 + 17.86 * \text{blockiness\_temp}) / 2 \quad \text{(B-15)}$$

$$\text{blockiness\_temp} = \frac{\text{globalInnerSum}}{\text{globalOuterSum}} \quad \text{(B-16)}$$

Blurriness refers to the measurement of the degree of image blur. *Blurriness* refers to the change in the gradient amplitude based on the grayscale level of the image pixel. This change can characterize the edge information of the image. When the gradient magnitude is too small, the edge will not be clear enough, which then affects the visual effect and reduces the QoE during viewing.

$$blurriness = 5 \cdot \left(\frac{\exp(-1.5 \cdot blur\_temp + 2.87)}{1 + \exp(-1.5 \cdot blur\_temp + 2.87)}\right)^{0.14} \quad \text{(B-17)}$$

$$blur\_emp = \frac{\text{ed}gewidth}{edgeNo} \quad \text{(B-18)}$$

Reasonable contrast can display vivid and rich colours, showing more details, better definition and grayscale levels. However, if the contrast is too high, the video will have a sense of distortion, and if the contrast is too low, the picture will appear grey. Unreasonable contrast will affect the subjective perception of the human eye.

$$C_n = \sqrt{\frac{1}{\text{MN}-1}\left(\sum f(i,j)^2 - \frac{1}{\text{MN}}\left(\sum f(i,j)\right)^2\right)} \quad \text{(B-19)}$$

where M is the number of horizontal pixels in the n^{th} frame, N is the number of vertical pixels in the n^{th} frame, C_n is the image contrast of the n^{th} frame, and $f(i,j)$ is the grayscale level of a pixel (i,j).

Noise defines the measurement of the fluctuation of the pixel chromaticity value. This kind of fluctuation has no positive effect on the overall quality of the picture and has no fixed law.

$$N_n = \frac{\sum_i \sum_j (G_{i,j} - \{w_n\}_{i,j})}{G.height * G.width} \quad \text{(B-20)}$$

where $\{w_n\}$ is the set of extracted image blocks that meet the requirements, $G_{x,y}$ is the calculated value of the pixel point (x, y) of the n^{th} frame after convolution with the operator and N_n is the noise degree of the n^{th} frame of the image.

According to the training data, the corresponding relationship between the above four measurements and MOS is obtained i.e., *Qs_i*, and then the weighted average is used to obtain the total Q_s.

$$Qs = f(Blockiness, Blurriness, Contrast, Noise) = \frac{1}{4}\sum_i \alpha_i Qs\_i$$
$$i = \{Blockiness, Blurriness, Contrast, Noise\} \quad \text{(B-21)}$$

Where α_i is a weighting factor.

B.3 Calculation algorithm for Q_a

Table B.2 – Variables for calculation algorithm for Q_a

Variable	Physical meaning
Bit rate	The average number of bits per second required for audio data
Sample rate	Audio data sampling points per second
Number of channels	The number of sound sources during sound recording or the corresponding number of speakers during display
Codec type	Audio encoding method
Scale factor	Scale factor in advanced audio coding (AAC) coding
Loudness	
Dynamic range	
Skewing	Phase difference between left and right channels
Sonic boom	
Mute length	

$$Qa = f(Bitrate, SampleRate, NumberofChannels, CodecType, Scalefactor, Loudness, DynamicRange \ \ldots) \quad \text{(B-22)}$$

According to its processing signal level, it mainly includes Model0, Model1 and Model2 three-layer models. Among them Model0 contains *BitRate*, *CodecType* and other variables, Model1 contains *BitRate*, *SampleRate*, *NumberofChannels*, *CodecType*, *ScaleFactor* variables. Model2 needs to continuously collect audio PCM from the original audio or the player to obtain key quality information such as *Loudness, Dynamic Range and Skewing*.

B.3.1 Basic model (Model 0)

$$\text{Qcoda} = a_1 \cdot exp(\, a_2 \cdot \text{bit}rate) + a_3 \quad \text{(B-23)}$$

Among them, there are different corresponding values of a_1, a_2 and a_3 according to different *CodecTypes* (e.g., E-AC3, AAC, MPEG1 Layer2 and Layer3). The overall audio quality is expressed as follows:

$$Qa = 100 - Qcoda \quad \text{(B-24)}$$

B.3.2 Parametric model (Model 1)

$$Qa = f(BitRate, SampleRate, NumberofChannels, CodecType, Scalefactor) \quad \text{(B-25)}$$

B.3.3 Bitstream model (Model 2)

$$Qa = f(Loudness, Dynamic\ Range, Skewing, Sonicboom, Mutelength) \quad \text{(B-26)}$$

B.4 Calculation algorithm for Q_i

Table B.3 – Variables for calculation algorithm for Q_i

Variable	Physical meaning
QZapping	Channel switching/zapping time during BTV
QLoading	Initial loading time for VoD
t_{zapping}	Channel switching/zapping delay during BTV
t_{loading}	Initial buffer delay during VoD
t	The duration of the currently played video; System will automatically record it
T	Maximum forgetting time; this is a system parameter.

The interactive experience quality first considers the influence of channel switching/zapping and initial display delay factors on the experience, and other factors can be further supplemented in the future.

B.4.1 *Qi_ZapInstant* and *Qi_ZapSession*

$$QZapping = c_{15} \cdot exp\left(c_{16} \cdot t_{zapping}\right) + c_{17} \cdot exp\left(c_{18} \cdot t_{zapping}\right) \tag{B-27}$$

$$Qi\_ZapInstant = QZapping \tag{B-28}$$

Session scoring is based on real-time scoring and decays according to the forgetting curve.

$$Qi\_ZapSession = (\alpha_1 \cdot exp(\beta_1 \cdot \frac{t}{T}) + \alpha_2 \cdot exp(\beta_2 \cdot \frac{t}{T})) \cdot QZapping \tag{B-29}$$

B.4.2 *Qi_LoadInstant* and *Qi_LoadSession*

$$QLoading = c_{19} \cdot t^3{}_{loading} + c_{20} \cdot t^2{}_{loading} + c_{21} \cdot t^1{}_{loading} + c_{22} \tag{B-29.1}$$

$$Qi\_LoadInstant = QLoading \tag{B-29.2}$$

$$Qi\_LoadSession = (\alpha_3 \cdot exp(\beta_3 \cdot \frac{t}{T}) + \alpha_4 \cdot exp(\beta_4 \cdot \frac{t}{T})) \cdot QLoading \tag{B-29.3}$$

B.5 Calculation algorithm for Q_v

Table B.4 – Variables for calculation algorithm for Q_v

Variable	Physical meaning
QBlocking	Scoring based on the proportion of dappled screens
V_AIRF	Average damage ratio of video frames (area percentage)
V_IRpF	Damage ratio of the video frame, the proportion of the affected area in the video frame with damage/degradation.
V_NDF	Number of damaged/degraded video frames in the video stream
V_TNF	The total number of video frames in the video stream
V_IR	Block area ratio, video stream damage ratio (that is, the proportion of the dappled screen duration)
V_PLEF	Number of video damage/degradation occurrences
V_CCF	Video complexity, obtained by analysing the video stream; when the content complexity cannot be extracted, it can be replaced by empirical values
Frequency	Number of times buffering occurred during video viewing

Table B.4 – Variables for calculation algorithm for Q_v

Variable	Physical meaning
Duration	The average of the lengths of all buffer events, the sum of all buffer durations divided by the number of buffers
$BufferLength_i$	The duration of the i^{th} freeze/stalling
Interval	In the case of multiple buffering, the average value of the buffering interval; (only used when more than one re-buffering event occurs (Frequency > 1))
$BufferStartTime_i$	The time when the i^{th} freeze/stalling occurs

The quality of viewing experience first considers the influence of block and stalling factors on the experience, and other factors can be further supplemented in the future.

B.5.1 Qv_BLInstant and Qv_BLSession

$$Qv\_BLInstant = 5 - QBlocking \tag{B-30}$$

$$QBlocking = f(BlockingTimeRatio, BlockingAreaRatio, Frequency) = \frac{\left(\frac{V\_AIRF \cdot V\_IR}{c_{23} \cdot V\_CCF + c_{24}}\right)^{c_{25}} \cdot \left(\frac{V\_PLEF}{c_{26} \cdot V\_CCF + c_{27}}\right)^{c_{28}}}{1 + \left(\frac{V\_AIRF \cdot V\_IR}{c_{23} \cdot V\_CCF + c_{24}}\right)^{c_{25}} \cdot \left(\frac{V\_PLEF}{c_{26} \cdot V\_CCF + c_{27}}\right)^{c_{28}}} \tag{B-31}$$

$$V\_AIRF = \frac{\sum_{i=1}^{V\_NDF} V\_IRpF_i}{V\_NDF}, \quad V\_IR = \frac{V\_NDF}{V\_TNF} \tag{B-32}$$

The session score of the viewing QoE is based on the current real-time score and the previous session score, and the current session score is calculated through the α filter function. The α filter coefficient is affected by the duration of the real-time sampled segment and the played time. The specific equations are shown as follows:

$$Qv\_BLSession_n = \alpha \cdot Qv\_BLSession_{n-1} + (1-\alpha) \cdot Qv\_BLInstant \tag{B-33}$$

$$Qv\_BLSession = \alpha_5 \cdot exp(\beta_5 \cdot BlockingRatio) + \alpha_6 \cdot exp(\beta_6 \cdot BlockingRatio) \tag{B-34}$$

$$BlockingRatio = BlockingTimeRatio \cdot BlockingAreaRatio \tag{B-35}$$

B.5.2 *Qv_STInstant* and *Qv_STSession*

$$Qv\_STInstant = 5 - QStalling \tag{B-36}$$

$$QStalling = f(Duration, Interval, Frequency) = \frac{\left(\frac{Duration}{c_{29}}\right)^{c_{30}} \cdot \left(\frac{Interval}{c_{31}}\right)^{c_{32}} \cdot \left(\frac{Frequency}{c_{33}}\right)^{c_{34}}}{1 + \left(\frac{Duration}{c_{29}}\right)^{c_{30}} \cdot \left(\frac{Interval}{c_{31}}\right)^{c_{32}} \cdot \left(\frac{Frequency}{c_{33}}\right)^{c_{34}}} \tag{B-37}$$

$$Duration = \frac{\sum_{i=1}^{NRE} bufferLength_i}{Frequency} \tag{B-38}$$

$$Interval = \frac{\sum_{i=1}^{Frequency-1} [bufferStartTime_i - (bufferStartTime_{i-1} + bufferLength_{i-1})]}{Frequency - 1} \tag{B-39}$$

$$Qv\_STSession_n = \alpha \cdot Qv\_STSession_{n-1} + (1-\alpha) \cdot Qv\_STInstant \tag{B-40}$$

$$Qv\_STSession = \alpha_7 \cdot exp(\beta_7 \cdot StallingRatio) + \alpha_8 \cdot exp(\beta_8 \cdot StallingRatio) \tag{B-41}$$

$$StallingRatio = StallingDuration / TotalPlayTime \tag{B-42}$$

Appendix I

General requirements for supporting UHD broadcast TV over home networks

(This appendix does not form an integral part of this Recommendation.)

Live video transmission usually adopts broadcast and multicast, while broadcast can be regarded as a kind of multicast with an unspecified multicast number. The interaction experience of multicast is mainly the speed of channel zapping, while the viewing experience requests no dappled screen including mosaic and blank screen, etc.

I.1 Request on network KPI from channel switching

To ensure the experience of channel zapping, the switching time should be no more than 500 ms.

[ITU-T H.265] is a mainstream compression algorithm of 4K video, which adopts an IPB compression algorithm. I-frame adopting interframe compression can decompress itself; P frame is forward predictive frame, B frame is bidirectional predictive frame. Decompression of P frame and B frame depends on I-frame. Frame types during video encoding are shown in Figure I.1.

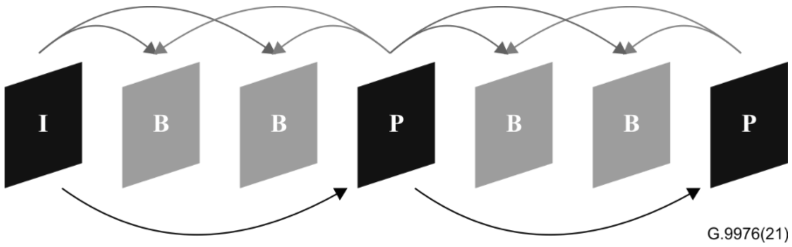

Figure I.1 – Frame types during video encoding

For the mainstream UDP live broadcast system, channel switching can be divided into: signal interacting phase (X: 1 × RTT), one complete I-frame download phase (Y) and player loading preparation phase (Z: generally needs 10–200ms). To ensure that the channel switching time is ≤ 500 ms, it is necessary to ensure that X + Y + Z ≤ 500 ms. See also Figure I.2.

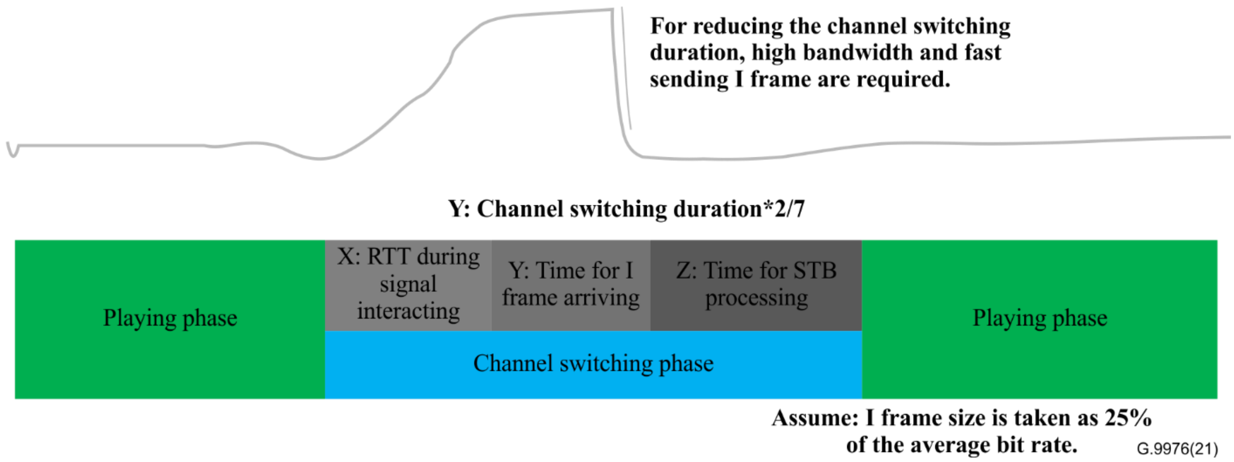

Figure I.2 – UDP initially buffering duration

During T = 500 ms − RTT − 200 ms = 300 ms − RTT, the STB needs to download 30 Mb/s × 2s*25% = 15 Mb medium data (typical GOP of live streaming is 2 seconds, I-frame usually accounts for about 25% of FTP). In Table 6-2 the E2E RTT limit is 20 ms, for UDP live video the system is insensitive to time delay, so E2E RTT usually can be evaluated to 30 ms. Through this condition, the requirement of a subscriber's throughput is more than 56 Mb/s (i.e., 15 Mb / (0.3 − 0.03) s). Thus, the interactive experience requirement of UDP live video to network is E2E throughput ≥ 56 Mb/s.

I.2 Request on network KPI from no dappled screen

Video service adopts moving picture coding, so during decompression each frame has referenced relations between one another after compression. A lost message can make subsequent multiple frames decompress incorrectly. Incorrect or lost message is a primary reason for mosaics in video. The higher the compression ratio, the greater the influence on incorrect or lost message packets will have. The experimental data indicates that 4K live streaming is very sensitive to packet loss. One lost packet will lead to dappled screen or video pause. Viewing experience becomes worse with higher PLR.

According to [b-DSL Forum TR-126], to avoid screen dappling for 4K live video requires a PLR of less than 10^{-6}. According to application layer technology such as RET and fast channel change (FCC), which need RET and FCC servers, the requirement of PLR could be reduced to around 10^{-4}.

In conclusion, the requirements for the network for a good QoE in 4K live video are the following:

- E2E throughput ≥ 56 Mb/s;
- PLR ≤ 10^{-6} (10^{-4}, when considering RET).

Appendix II

General requirements for supporting UHD video on-demand over home networks

(This appendix does not form an integral part of this Recommendation.)

Downloading UHD VoD service are often based on TCP connectivity. TCP is a reliable protocol, since it checks the received packets and retransmits the lost packets. Therefore, package loss and dappled screen will not appear in VoD. Primary indicators influencing QoE are faster interaction and lack of video pause (stalling) during viewing.

II.1 The request on networks for fast interaction

For VoD, initial loading time must be less than 1 second so rapid interaction can be guaranteed.

At present, the mainstream technology for video stream is HLS. Initially, buffering includes: signalling interacting phase (X1: about 9 RTTs), downloading the media data for minimum decoding buffering (generally using video in 2 seconds) phase (Y) and player loading phase (Z: generally requires 10–200 ms). To ensure the initial loading time ≤ 1 second, means ensuring that X1 + Y + Z ≤ 1 second; see Figure II.1.

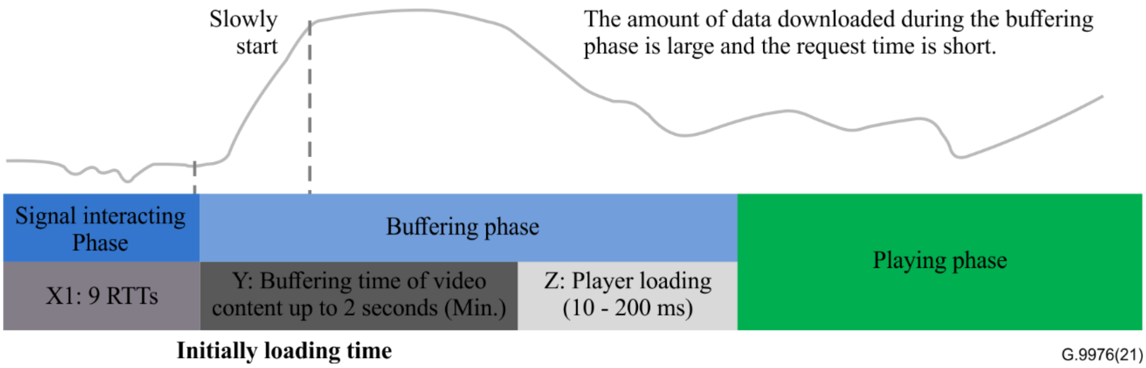

Figure II.1 – HLS initial buffering duration

Slowly starting needs 6 RTTs, which leads to low downloading speed, so the whole download needs at least 7 RTTs. The length of initial buffering time = 9 × RTT + (6 + 1) × RTT + 200 ms ≤ 1000 ms, thus RTT should be equal to or less than 50 ms.

In addition, during T = 1000 ms − 9 × RTT − 6 × RTT − 200 ms = 800 ms − 15 × RTT, the single TCP thread needs to download 25 Mbits (the average bitrates of 4K video) × 2 s = 50 Mbit video data. According to the variable values of RTT chosen, the requirement of TCP throughput can be obtained as shown in Table II.1.

Table II.1 – Different demands on TCP throughput based on E2E RTT

E2E RTT	TCP throughput
10 ms	77 Mb/s
20 ms	100 Mb/s
30 ms	143 Mb/s
40 ms	250 Mb/s
50 ms	1 Gb/s

When E2E RTT = 10 ms, TCP Throughput = 50 Mb/T = 50 Mb/(800 ms − 15 × 10 ms) = 50 Mb/0.65 s = 77 Mb/s.

According to the formula of TCP throughput:

$$\mathbf{Throughput} \leq \mathbf{min}(\mathbf{Max}(\mathbf{BW}), \frac{\mathbf{WSS}}{\mathbf{RTT}}, \frac{\mathbf{MSS}}{\mathbf{RTT}} \times \frac{\mathbf{1}}{\sqrt{\mathbf{p}}}) \tag{II-1}$$

In the Equation (II-1),

BW is physical bandwidth

WSS is window sliding size (usually set as 256K bytes in server)

RTT is round trip time

P is the probability of packet loss i.e., PLR

MSS is the maximum segment size (generally set as 1460 bytes).

According to the requirement of TCP throughput and this formula, the requirement of PLR under different conditions can be obtained.

- When RTT is 10 ms, and TCP throughput ≥ 77 Mb/s, it requires PLR ≤ 2.3×10^{-4};
- When RTT is 20 ms, and TCP throughput ≥ 100 Mb/s, it requires PLR ≤ 3.4×10^{-5}.

Considering that the 10 ms E2E delay requirement is too strict, 20 ms is generally taken as the delay requirement.

II.2 The request on the network for smooth view

Smooth viewing of video is the most important user experience i.e., QoE. If single TCP throughput is greater than 1.5 times average bitrates, more than 95% 4K video could be played smoothly.

For true 4K (see Table 6-2) with cloud optimization, average bitrates of 4K @ H.265 ≥ 25 Mb/s, corresponding TCP throughput should be greater than 37.5 Mb/s.

According to the above formula of TCP throughput, the requirement for the carrier network is throughput ≥ 100 Mb/s and PLR ≤ 3.4×10^{-5} when RTT is 20 ms.

In conclusion, the requirements on the network for excellent QoE in 4K video are the following:

- E2E throughput ≥ 100 Mb/s (after optimizing cloud and end, the requirement generally can be reduced to 37.5 Mb/s, regardless of the requirement on the network for the initial loading);
- RTT ≤ 20 ms;
- PLR ≤ 3.4×10^{-5} (after optimizing cloud and end, and using the 37.5 Mb/s as the throughput in above formula, after calculation the requirement of PLR can be reduced to 2.4×10^{-4}).

Bibliography

[b-ITU-T E.800] Recommendation ITU-T E.800 (2008), *Definitions of terms related to quality of service.*

[b-ITU-T F.1399-1] Recommendation ITU-T F.1399-1 (2001), *Vocabulary of terms for wireless access.*

[b-ITU-T P.800.1] Recommendation ITU-T P.800.1 (2016), *Mean opinion score (MOS) terminology*.

[b-ITU-T P.915] Recommendation ITU-T P.915 (2016), *Subjective assessment methods for 3D video quality.*

[b-ETSI WP] ETSI White paper No. 17, *Next Generation Protocols – Market Drivers and Key Scenarios. 2016.*
http://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp17_Next_Generation_Protocols_v01.pdf

[b-DSL Forum TR-126] DSL Forum Technical Report TR-126 (2006), *Triple-play services quality of experience (QoE) requirements*.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	**Transmission systems and media, digital systems and networks**
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2022