ETSI TR 103 891 V1.1.1 (2023-10)

TECHNICAL REPORT

Speech and multimedia Transmission Quality (STQ); Parametric non-intrusive QoS evaluation of Cloud Gaming Services over RTP/UDP streaming

Reference
DTR/STQ-00232m

Keywords
4G, 5G, cloud gaming, QoE, QoS, service, video

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Speech and multimedia Transmission Quality (STQ).

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Executive summary

The present document proposes leveraging and improving the existing Recommendations ITU-T P.1201 [i.1], G.1072 [i.4] and ETSI TR 103 702 [i.6] to provide QoS parameters for users who experience Cloud Gaming Services over RTP/UDP streaming. In particular, the present document proposes to extend the Recommendation ITU-T G.1072 [i.4] model, which predicts the expected QoS/MOS for given network conditions, to the single user analysis allowed by non-intrusive passive monitoring and provided by feeding such model with measured or estimated end-user packet loss, round-trip time, video bitrate, framerate, and information derived from the game platform capabilities and streaming data. Moreover, in addition to Recommendation ITU-T H.264 [i.8] characterization provided by Recommendation ITU-T G.1072 [i.4], it proposes also to address Recommendation ITU-T H.265 [i.9] and VP9 video codecs and audio chat.

The identified QoS parameters refer to categories that aim to address a practical Root Cause Analysis.

Introduction

Cloud Gaming is predicted to represent a new opportunity to boost cloud architecture and the 5G network. The success of cloud gaming is closely linked to the ability of gaming and network providers to build a convenient business model for stakeholders while delivering a QoS that is similar to or better than the current levels offered by the game consoles in the market.

Few standards may be applied for Cloud Gaming QoS.

Recommendation ITU-T P.1201 [i.1] provides algorithmic models for non-intrusive audio and video quality monitoring of IP-based video services based on packet-header information; however, the model works for one mode of operation, referred to as "CC-mode". The CC-mode assumes the model feed at the endpoint (at the receiver side), with packets already processed with FEC or retransmission mechanisms and having further information about the media buffer. It requires the analysis of the type of frames impacted by packet loss. All the above conditions can not apply to passive non-intrusive systems located between the transmitter and the receiver, referred to as "NN-mode" in Recommendation ITU-T P.1201 [i.1].

Recommendation ITU-T G.1072 [i.4], derived from Recommendations ITU-T G.1071 [i.3] and P.1201 [i.1], defines a model to predict the gaming quality of experience for cloud gaming services to support network optimization. It is built considering the Recommendation ITU-T H.264 [i.8] codec and does not estimate the audio quality of users in chat. Moreover, being developed for network optimization, its model provides the expected QoS/MOS for given network conditions, but it is not based on single-user analysis. The model proved consistent when tested against the assessed mean opinion score.

ETSI TR 103 702 [i.6] aims to put in place means to assess 5G network capabilities and network readiness to support applications that require enhancement Mobility BroadBand (eMBB), CG among them. The identified test scenarios refer to measurements at the end user device, while the MOS-based approach and real-time analysis of network traffic generated by subscribers are not mentioned.

The present document leverages the above standards to provide a method to assess the user's CG QoS.

1 Scope

The scope of the present document is to provide a parametric non-intrusive evaluation of Cloud Gaming Services based on RTP/UDP streaming through the assessment of the QoS of Cloud Gaming users obtained passively by monitoring the network traffic generated using active test devices or by subscribers. Such evaluation also includes service and user detection together with other indicators to assess the ability of the network to deliver the cloud gaming service.

The present document leverages the existing Recommendations ITU-T P.1201 [i.1], G.1072 [i.4] and ETSI TR 103 702 [i.6] to provide QoS parameters for users who experience Cloud Gaming Services over RTP/UDP streaming. In particular, it extends Recommendation ITU-T G.1072 [i.4] model, which predicts the expected QoS/MOS for given network conditions and is mainly focused to network design and optimization activities, to user analysis based on non-intrusive passive monitoring, feeding such model with measured or estimated end-user packet loss, round-trip time, inter-arrival jitter, video bitrate, framerate, and further information derived from the game platform capabilities and streaming data. Moreover, the present document, in addition to Recommendation ITU-T H.264 [i.8] characterization provided by Recommendation ITU-T G.1072 [i.4], addresses Recommendation ITU-T H.265 [i.9] (and VP9) video codecs and audio chat.

The Recommendation ITU-T G.1072 [i.4]-derived model has been applied to three gaming platforms under several network conditions: Blacknut®, Nvidia GeForce Now®, and Google Stadia®. They delivered objective results in line with the user's subjective. However, the applicability needs to avoid the periods when games are paused or when gamers select game options. In such a situation, the scenes are mostly static, and encoders output a low video bit rate that does not represent the one used to stream the game while in action.

The identified QoS parameters refer to the following categories with the aim also of addressing a practical Root Cause Analysis:

- Service-Centric (Subscriber IP address, Cloud gaming Platform Name).
- Cloud Gaming Session (Platform Start Time, Game Start Time, Streaming Start Time, Play Start Time, Streaming End Time, Streaming Duration).
- Media (MOS, Audio/Video Bitrate, Video Framerate, Audio/Video Packets, Audio/Video Downlink/Uplink bytes).
- Transport (Network Round Trip Time, Packet Loss Rate, Packet retransmission, Server addresses, Server ports, SSRC).

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long-term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] Recommendation ITU-T P.1201: "Parametric non-intrusive assessment of audiovisual media streaming quality".

[i.2] Recommendation ITU-T P.1201.2: " Parametric non-intrusive assessment of audiovisual media streaming quality - Higher resolution application area".

[i.3] Recommendation ITU-T G.1071: "Opinion model for network planning of video and audio streaming applications".

[i.4] Recommendation ITU-T G.1072: "Opinion model predicting gaming quality of experience for cloud gaming services".

[i.5] Recommendation ITU-T G.1032: "Influence factors on gaming quality of experience".

[i.6] ETSI TR 103 702: "Speech and multimedia Transmission Quality (STQ); QoS parameters and test scenarios for assessing network capabilities in 5G performance measurements".

[i.7] IETF RFC 3550: "RTP: A Transport Protocol for Real-Time Applications".

[i.8] Recommendation ITU-T H.264: "Advanced video coding for generic audiovisual services".

[i.9] Recommendation ITU-T H.265: "High efficiency video coding".

[i.10] Teemu Kämäräinen et al.: "A Measurement Study on Achieving Imperceptible Latency in Mobile Cloud Gaming". 2017 ACM. SBN 123-4567-24-567/08/06. .

[i.11] J. Deber et al.: "How Much Faster is Fast Enough? User Perception of Latency & Latency Improvements in Direct and Indirect Touch". ACM 978-1-4503-3145-6/15/04.

[i.12] Vlahovic et al. "The impact of network latency on QoE for an FPS VR game". Proceedings of the 2019 11th International Conference on Quality of Multimedia Experience, Berlin 5-7 June 2019.

[i.13] IETF RFC 5481: "Packet Delay Variation Applicability Statement", A. Morton, B. Claise, March 2009.

[i.14] Shea et al. Cloud Gaming: "Architecture and Performance", IEEE Network July/August 2013.

[i.15] tc Ubuntu Manual.

[i.16] Linux Traffic Control.

[i.17] The tool "clumsy".

[i.18] Recommendation ITU-T P.1401: "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models".

3 Definition of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the following terms apply:

average frames per second: average number of successfully transmitted frames per second during a video stream

NOTE: As defined in Recommendation ITU-T G.1072 [i.4].

cloud gaming: execution of the game logic, rendering of the virtual scene, and video encoding are performed on a server in the cloud, while the client is responsible for video decoding, capturing and sending user input

NOTE: As defined in Recommendation ITU-T G.1032 [i.5].

commands: information sent to the game server to update the game logic resulting from users actioning devices like keyboard, mouse, joystick, or gamepads

DownLink: direction from the game server to the client

First-Person Shooter game: video game in which the player shoots at targets seeing from the viewpoint of the character controlled by the player

freezing artefacts: artefacts introduced when the Packet Loss Concealment (PLC) scheme of the receiver replaces the erroneous frames (either due to packet loss or error propagation) with the previous error-free frame until a decoded picture without errors has been received

NOTE: As defined in Recommendation ITU-T G.1071 [i.3].

game: rule-based system with a variable and quantifiable outcome, where different outcomes are assigned different values, the player exerts effort to influence the outcome, the player feels emotionally attached to the outcome, and the consequences of the activity are optional and negotiable

NOTE: As referenced in Recommendation ITU-T G.1072 [i.4]).

game server: server that streams to the client the encoded audio/video packets of the game, and receives audio and commands from the client

input quality: playability of a game scenario in terms of the component's responsiveness, immediate feedback, and controllability

NOTE: As defined in Recommendation ITU-T G.1072 [i.4].

model, model algorithm: algorithm to estimate the subjective (perceived) quality of a media sequence

NOTE: As defined in Recommendation ITU-T G.1072 [i.4].

Real-Time Strategy Game: strategy video game in which each participant builds and maneuvers units to defeat the opponents in real time

Role-Playing Game: video game in which a character can be created and controlled in a fictional world

slicing artefacts: artefacts that are introduced when packet losses are concealed through the use of a Packet Loss Concealment (PLC) scheme to repair erroneous frames

NOTE: As defined in Recommendation ITU-T G.1071 [i.3].

spatial video quality: dimensions unclearness and fragmentation, which can be influenced by artefacts such as blockiness and blur

NOTE: As defined in Recommendation ITU-T G.1072 [i.4].

temporal video quality: dimensions discontinuity which can be influenced by artefacts such as freezing and jerkiness

NOTE: As defined in Recommendation ITU-T G.1072 [i.4].

UpLink: direction from the client to the game server

3.2 Symbols

Void.

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

ABR	Adaptive BitRate
ARQ	Automatic Repeat Request
AvgFPS	Average Frame Per Second
CBR	Constant BitRate
CD	Control Delay
CG	Cloud Gaming
DASH	Dynamic Adaptive Streaming over HTTP
DL	DownLink

e2e, E2E	End-to-End
ELIF	Else If
FEC	Forward Error Correction
FPS	First-Person Shooter game
fps	frame per second
FTTH	Fiber To The Home
HLS	HTTP Live Streaming
HTTP	HyperText Transfer Protocol
HTTPS	HTTP Secure
ID	IDentifier
IP	Internet Protocol
JNT	Just Noticeable Time
KPI	Key Performance Indicator
MOS	Mean Opinion Score
ms	milli-second
MTP	Motion To Photon
NSA	Non Stand-Alone
OD	Playout Delay
PC	Personal Computer
PD	Processing Delay
PL	Packet Loss
PLC	Packet Loss Concealment
PLCC	Pearson Linear Correlation Coefficient
QoE	Quality of Experience
QoS	Quality of Service
QUIC	Quick UDP Internet Connection
RAN	Radio Access Network
RCA	Root Cause Analysis
RMSE	Root Mean Square Error
RPG	Role-Playing Game
RTCP	Real Time Control Protocol
RTP	Real Time Protocol
RTS	Real-Time Strategy Game
RTSP	Real Time Streaming Protocol
RTT	Round Trip Time
SA	Stand-Alone
SSL	Secure Socket Layer
SSRC	Synchronization Source
SVOD	Subscription Video On Demand
TCP	Transmission Control Protocol
TLS	Transport Layer Security
TW	Time Window
UL	UpLink
USB	Universal Serial Bus
VBr	Video Bitrate
VFR	Video Frame Rate
VP9	Video Predictor 9
WebRTC	Web Real Time Communication
Y4M	Yuv4Mpeg2

4 Parametric non-intrusive evaluation of Cloud Gaming Services based on RTP/UDP streaming

4.1 Cloud Gaming and its phases

A Cloud Gaming (CG) service is a platform that, similarly to SVOD services, offers a palimpsest from where users can choose and play the games of their liking.

Users subscribed to a cloud game platform pay a monthly fee for the delivery of the streaming service. The subscription may include games, although in several cases, users need to own the game license in addition to the game platform subscription.

Cloud gaming is based on a technology that integrates web browsing and real-time communication services. WebRTC has often been the best matching technology for such needs, and it has the advantage of being integrated with major browsers.

When a user connects to a CG service, three main logical functions, often delivered by distinct servers, are involved:

- A web server that provides the landing page where user clients access the gaming platform through a web browser or a dedicated application on their portable devices. Besides user authentication and profile management, the web server sends the palimpsest to the client. Users navigate the palimpsest and select the game of their choice.
- A session control function, which may be a specific server, orchestrates the streaming session, managing the exchange of the session description between the game server and the client.
- A game server that manages the client commands/events receives the client audio stream, makes the rendering, encodes, and streams audio and video to the client. The game logic may be run on the game server or on separate servers.

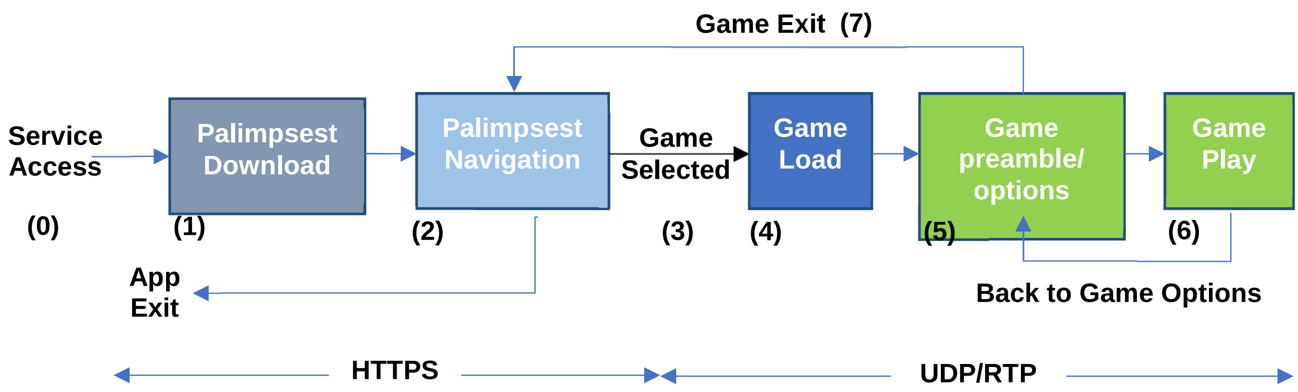


Figure 1: Phases of Cloud Gaming

Many activities flow between the user client and the CG servers; in the following, only the ones impacting this document's subject are described and summarized in figure 1.

Initial phases

- The user logs into the CG site with the browser or the application (0).
- The user client retrieves the HTML palimpsest from the platform web server (1).
- The user can navigate the palimpsest (2) to choose the game, select other options, or exit the application.

These phases are all conducted using a client-server HTTPS (TLS/TCP or QUIC/UDP) approach.

When a game is selected (3)

- The user client receives from the game server on a separate window a "lobby" audio/video streaming for the time necessary for the server to load the game and get it ready to run (4).
- When ready, the game server starts streaming the game preambles, typically an initial narration, which the user may skip to move to the game options (5) and start playing (6).
- The user may eventually exit the game and return to the palimpsest (7).

The game server streams all activities from (3) to (6) using RTP/UDP.

The client-server connection over HTTPS remains active throughout the game session exchanging information about the user capabilities, authentication, and other administrative tasks.

4.2 Monitoring Point

Network events are gathered using a non-intrusive method corresponding to the *Static operation mode (NN) inside the network* defined in Recommendation ITU-T P.1201 [i.1].
The observation point may be placed everywhere between the *Sender Point* and the *End Point*.

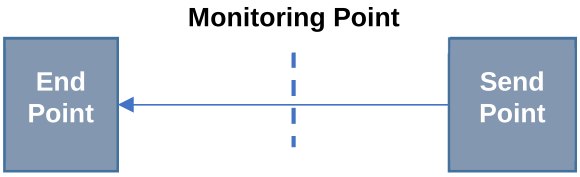


Figure 2: Monitoring Point, Recommendation ITU-T P.1201 [i.1], Static operation mode (NN)

Mobile networks' non-intrusive monitoring interfaces may, as an example, be S1U for 4G and 5G NSA, and N11 for 5G SA.

The term DownLink (DL) will be used to indicate the packets sent by the game server to the client, whereas the term UpLink (UL) will be used in the opposite direction.

4.3 Service Centric Analysis

4.3.1 Client and CG platform identification

CG platform identification involves identifying the connection between the client, the web server, and the platform game server, which streams the game to the client.

CG platform identification is mandatory. All QoS parameters defined in the present document may be categorized per CG platform, so benchmarking among platforms is possible.

Client IP addresses are mandatory as well to address RCA activities efficiently.

4.3.2 CG session identification and analysis

4.3.2.1 Trigger points

CG platform and phases are identified from network events as a non-intrusive setup is considered that corresponds to the "Static operation mode (NN) inside the network" defined in Recommendation ITU-T P.1201 [i.1].

Table 1: Overview of the triggers used for the QoS parameter definition

Trigger ID	Abstract Description	Technical Description / Protocol part
tr-1	Session Access	The client attempts the connection to the CG Web server. It may receive back denial of service or successfully connect. The connection is typically HTTP over SSL or HTTP over QUIC.
tr-2	Game Start	The time a user selects a game. This trigger may be derived by tracing the session control messages between the client and the session server. There is no common standard for the protocol used to manage sessions; it may be HTTP, RTSP, or a proprietary protocol over UDP.
tr-3	Streaming Start	Time of the first detected audio or video packet sent by the game server to the client.
tr-4	Streaming Stop	Time of the last detected audio or video packet sent by the game server to the client.
tr-5	Session Stop	The time the client leaves the CG platform.

4.3.2.2 Session QoS parameters

Service Access Failure Ratio [%]

Sum(tr-1 failed) / Sum(tr-1)

The overall failure rate for the player access.
Passive monitoring lacks visibility of HTTPS failures; therefore, this parameter could not be available for such tools.

Session Duration [s]

Time (tr-5) - Time (tr-1 successful)

Streaming Failure Ratio [%]

Sum (tr-3 failure) / Sum (tr-3)

The overall failure rate of streaming for the requested games.

Streaming Start Delay [s]

Time (tr-3) - Time (tr-2)

Streaming Duration [s]

Time (tr-4) - Time (tr-3)

4.4 Media Components Identification and QoS Analysis

4.4.1 Cloud Gaming media components

The Game server streams encoded audio and video packets to the client; the client may optionally stream to the server audio packets to allow the gamer to chat with peers during gameplay.

In addition, the client sends the user commands to the game server.

Media components are mapped to RTP packets mainly using proprietary P-Type. An additional RTP stream may be used to retransmit lost video packets.

Audio and Video packets are exchanged using RTP methods ruled by IETF RFC 3550 [i.7] and related standards, including telemetry information over RTCP, that may be encrypted.

Telemetry is used by the client to signal the server packet(s) that have been judged lost and the packets' arrival time. The server adjusts the encoding parameters, which may result in changes in video quality, resolution, and frame rates according to the information it receives from the client.

Commands are transported using a reliable connection, i.e. over TCP or a proprietary protocol over UDP, to ensure a reliable connection.

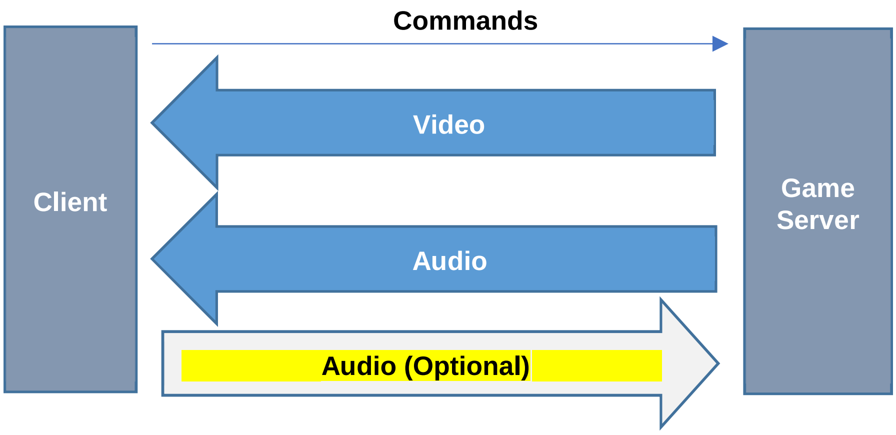


Figure 3: Cloud Gaming media components

4.4.2 Media Components Parameters

Media parameters may be recursively evaluated over a Time Window (TW). TW, as an indication, can be 10 seconds, in line with the hints provided for the MOS evaluation. A short 10-second TW enables more precise troubleshooting.

Table 2 shows the media parameters measured at the monitoring point.

Table 2: Media parameters measured at the monitoring point

Abstract Description	Technical Description / Protocol part
VideoPckDL	The number of RTP packets, classified as video, flowing from the game server to the client (DL). Obtained by counting the RTP packets with the P_Type used to map the encoded video samples.
AudioPckDL	The number of RTP packets, classified as audio, flowing from the game server to the client (DL). Obtained by counting the RTP packets with the P_Type used to map the encoded audio samples.
AudioPckUL	The number of RTP packets, classified as audio, flowing from the client to the game server (UL). Obtained by counting the RTP packets with the P_Type used to map the encoded audio samples.
VideoPckRtxDL	The number of RTP packets, classified as retransmitted video samples, flowing from the game server to the client (DL). Obtained by counting RTP packets with the P_Type used to map video retransmissions.
VideoFrDL	The number of video frames pertaining to VideoPckDL. This parameter can be obtained by counting the number of RTP timestamp variations when the same RTP timestamp is used for all video packets belonging to the same frame. In some cases the RTP 'marker' can be used as it is set at the end of each video frame.
VideoPckMissDL	The number of unseen RTP packets classified as video, flowing from the game server to the client (DL). Calculated by analysing the RTP sequence numbers.
AudioPckMissDL	The number of unseen RTP packets classified as audio, flowing from the game server to the client (DL). Calculated by analysing the RTP sequence numbers.
AudioPckMissUL	The number of unseen RTP packets classified as audio, flowing from the client to the game server (UL). Calculated by analysing the RTP sequence numbers.
VideoPayloadLenDL	DL Video packets payload length in bytes. Obtained stripping out the protocol stack headers from video DL network packets.
AudioPayloadLenDL	DL Audio packets payload length in bytes. Obtained stripping out the protocol stack headers from audio DL network packets.
AudioPayloadLenUL	UL Audio packets payload length in bytes. Obtained stripping out the protocol stack headers from audio UL network packets.

4.4.3 Media QoS analysis at the monitoring point

Average Video Bitrate DL in TW

VideoBitrateDL [Kbps] = (VideoPckDL + VideoPckMissDL) × VideoPayloadLenDL × 8) / (1 000 × TW)

Average Video Framerate DL in TW

VideoFrameRate [fps] = VideoFrDL / TW

Average Audio Bitrate DL in TW

AudioBitrateDL [Kbps] = (AudioPckDL × AudioPayloadLenDL × 8) / (1 000 × TW)

Average Audio Bitrate UL in TW

AudioBitrateUL [Kbps] = (AudioPckUL × AudioPayloadLenDL × 8) / (1 000 × TW)

Video Downloaded Size DL in TW

VideoDwLdSizeDL [KB] = VideoPckDL × VideoPayloadLenDL , TW = tr-4 - tr-3

Audio Downloaded Size DL in TW

AudioDwLdSizeDL [KB] = AudioPckDL × AudioPayloadLenDL, TW = tr-4, tr-3

Audio Downloaded Size UL in TW

AudioDwLdSizeUL [KB] = AudioPckUL × AudioPayloadLenUL, TW = tr-4, tr-3

Video Packet Miss Rate DL in TW

VideoPckMissRateDL [%] = VideoPckMissDL / (VideoPckDL + VideoPckMissDL)

Audio Packet Miss Rate DL in TW

AudioPckMissRateDL [%] = AudioPckMissDL / (AudioPckDL + AudioPckMissDL)

Audio Packet Miss Rate UL in TW

AudioPckMissRateUL [%] = AudioPckMissUL / (AudioPckUL + AudioPckMissUL)

4.5 Media QoS end-to-end analysis

It may be desirable to infer the end-to-end (e2e) QoS between the client and the game server. E2E QoS parameters will support operators in gathering information on users' received service.

Crucial e2e QoS indicators are e2e Packet Loss and Round-Trip Time. These two indicators are of paramount importance for Cloud Gaming. Packet loss may impact image quality; the user's commands responsiveness is affected by both packet loss and round-trip time.

Video Packet Loss Rate DL

a) When the monitoring point is next to the receiver:
VideoPckLossRateDL [%] = VideoPckMissRateDL [%]

b) Video packet loss needs to be inferred when the monitoring point is far from the receiver. The inference method depends on the gaming platform. Two examples:

EXAMPLE 1: When the RTCP reporting is accessible, VideoPckLossRateDL may be inferred by the lost packets and the received packets the receiver reports to the sender. In principle, this might not precisely reflect the lost packets in the network but the ones the receiver considers lost; however, such distinction is tolerated.

EXAMPLE 2: When ARQ techniques are used, video lost packets may be inferred from the retransmitted video packets, VideoPckRtxDL.

Audio Packet Loss Rate DL

Audio loss rate DL may be neglected as it follows the same path as video DL, which is already a good indicator of line losses.
The proposed MOS model does not include audio DL for the time being.

Audio Packet Loss Rate UL

a) When the monitoring point is next to the receiver:
AudioPckLossRateUL [%] = AudioPckMissRateUL [%]

b) Audio packet loss needs to be inferred when the monitoring point is far from the receiver. When the RTCP reporting is accessible, AudioPckLossRateUL may be inferred by the lost packets and the received packets the receiver reports to the sender. In principle, this might reflect not precisely the network's lost packets but the ones the receiver considers lost. This is tolerated.
Gaming platforms so far analysed do not retransmit audio packets.

Network Round Trip Time

Figure 4 shows UL and DL network delays (red lines) and the processing delay of a packet moving from the client to the game server back to the client. This path represents a packet transporting a user command from out of the client, and the command results that are included in a video frame returning to the client.

Note that the user experiences additional delays beyond those reported in figure 4 that cannot detected from the network, such as the time from when the gamepad button is pressed to its translation to a network packet, and the time from when the frame is received to when the frame is decoded and shown to the screen.

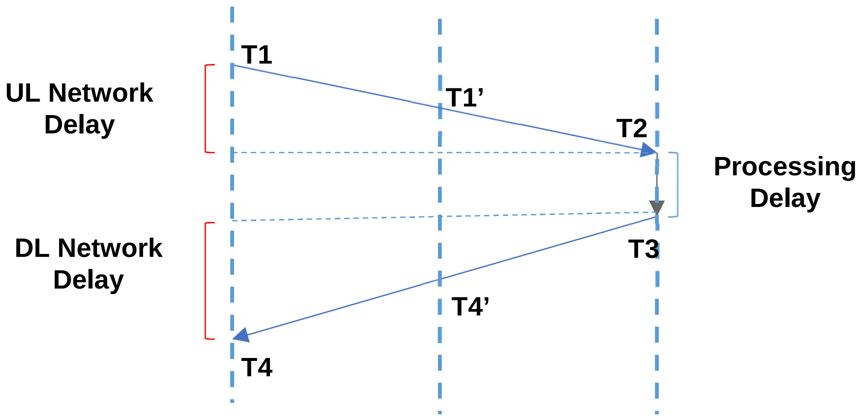


Figure 4: Round Trip Time of a packet from client output, back to client input

NetRTT [ms] = (T4 - T3) + (T2 - T1)

When the monitoring point is next to the receiver, NetRTT may be inferred using a suitable protocol for which it is possible to associate the client requests and the game server responses with short acknowledge packets to minimize the influence of (server) processing delays.

When the monitoring point is next to the sender, and receiver RTCP reports are available, the NetRTT may be obtained from the sender and receiver timestamps included in the RTCP reports, according to IETF RFC 3550 [i.7].

When the monitoring point is in between the sender and the receiver, the RTT can still be estimated when the client and the game server use a protocol for which on both sides exist, and it is possible to associate requests and responses.

Figure 5 shows an example.

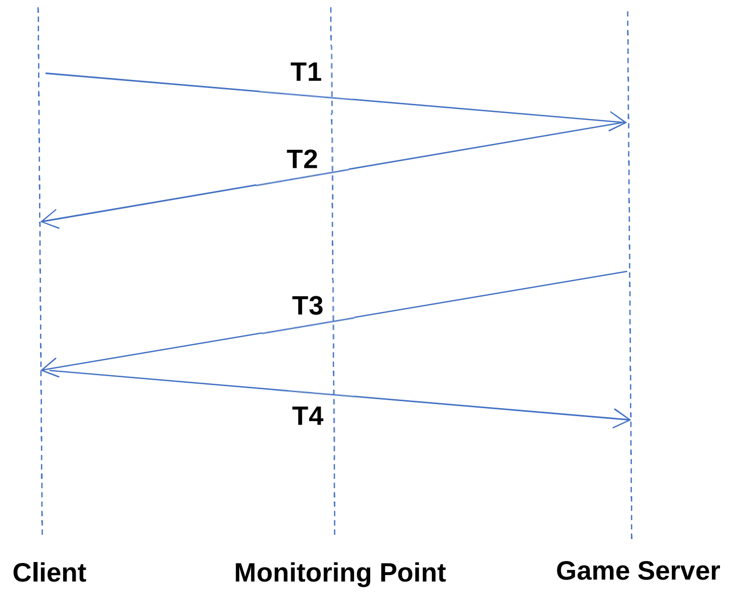


Figure 5: Round Trip Time inferred when the monitoring point is in between Client and Game Server

It is possible to measure the two separate delays:

Round Trip Server Side

RTTSrv = T2 - T1

Round Trip Client Side

RTTCli = T4 - T3.

Two distinct protocols can also be handy as long as they are a good proxy of the round trip from the monitoring point to the game server, back to the monitoring point (RTTSrv), and the round trip from the monitoring point to the client, back to the monitoring point (RTTCli).

When the monitoring point uses the S1U interface for 4G and N3 for 5G, RTTCli is a valuable estimation of RAN delays and RTTSrv of the core delays.

Network Inferred Round Trip Time

NetInfRTT [ms] = average (RTTSrv) + average (RTTCli), with RTTSrv and RTTCli measured in the same TW.

4.6 Transport Analysis

Network operators have the need and the interest to assess whether the network is delivering the service. In case it is not, they need to identify and resolve the problem quickly.

CG monitoring provides valuable information on network performances, being a demanding real-time service. The network indicators listed in table 3 extracted from CG monitoring have a twofold meaning: on one side, they are needed to verify the CG service itself; on the other, such indicators are measures of network performances.

Table 3: Transport parameters

Abstract Description	Technical Description / Protocol part
Game Server IP Addresses and Ports	Extracted from media streaming. Useful to localize server location and benchmark the involved servers.
Media RTP SSRC	Synchronization Sources of the media components are the logical channel used by media components. In most cases, the same SSRC is used for media even when the game server is forced to change address.
VideoPckMissDL, VideoPckMissRateDL	See table 2 for the description. It delivers the network packet loss of the segment "monitoring point - game server". When the monitoring is S1U for 4G networks and N3 for 5G networks, this segment represents the mobile core and the internet part of the connection.
AudioPckMissDL, AudioPckMissRateDL	See table 2 for the description. Similar information than VideoPckMissDL, VideoPckMissRateDL.
AudioPckMissUL, AudioPckMissRateUL	See table 2 for the description. It delivers the network packet loss of the segment "monitoring point - client". When the monitoring is S1U for 4G networks and N3 for 5G networks, this segment represents the Radio Access Network (RAN) part of the connection.
RTTSrv	See clause 4.5. When the monitoring is S1U for 4G networks and N3 for 5G networks, this indicator measures the round-trip contribution of the core network + the connection to the game server.
RTTCli	See clause 4.5. When the monitoring is S1U for 4G networks and N3 for 5G networks, this indicator measures the round-trip contribution of Radio Access Network (RAN).
NetRTT, NetInfRTT	See clause 4.5.
VideoPckLossRateDL	See clause 4.5.
AudioPckLossRateUL	See clause 4.5.
VideoJitterDL	Calculated for VideoPckDL according to IETF RFC 3550 [i.7] formula suggested for inter-arrival jitter.
AudioJitterDL	Calculated for AudioPckDL according to IETF RFC 3550 [i.7] formula suggested for inter-arrival jitter.
AudioJitterUL	Calculated for AudioPckUL according to IETF RFC 3550 [i.7] formula suggested for inter-arrival jitter.

Network RTT and inter-arrival jitter may adversely affect the user experience:

- They affect video quality as packets may not arrive in time to complete a frame, which may cause the client to drop/conceal the frame.
- They impact user commands, especially in highly interactive games, because users do not immediately see reflected in the screen the changes they input.

While FEC/ARQ techniques can compensate for packet drops, delays and packet inter-arrival jitter are more troublesome to recover. Applications can compensate for packet inter-arrival jitter by increasing the receiver buffer but at the expense of additional delays that CG users might not tolerate.

Clause 4.9 overviews the RTT budget allocation for different game categories.

4.7 User Scores Analysis

4.7.1 Model introduction

The present document proposes to apply a Recommendation ITU-T G.1072 [i.4]-derived model for non-intrusive monitoring.

Recommendation ITU-T G.1072 [i.4] was developed to predict cloud gaming services' quality of experience for given network conditions, and is geared at network design and optimization. While not thought for real-time monitoring, it has the advantage of being specific to Cloud Gaming, includes KPIs for command quality, and shows a good mapping of experienced MOS against the model predicted QoE.

The cloud gaming video is encoded at a Constant Bit Rate (CBR) with no B frames because of its stringent real-time needs. The encoded video bit rate of each game title has a monotonic relationship with the provided resolution or quality, which implies that each encoded frame produces more bytes for higher resolutions or quality and less when encoded at lower resolution or quality. Video resolution or quality is derated based on feedback the receiver sends to the server, which modifies the encoding parameters to adapt the video bitrate. The process is straightforward; no Adaptive Bitrate (ABR) technology (DASH, HLS) is used. For these reasons, the Recommendation ITU-T G.1072 [i.4]-derived model has been tested for non-intrusive monitoring, see figure 6, by feeding it with actual network measurements. Under the conditions described below, it has been found that the model provided good results.

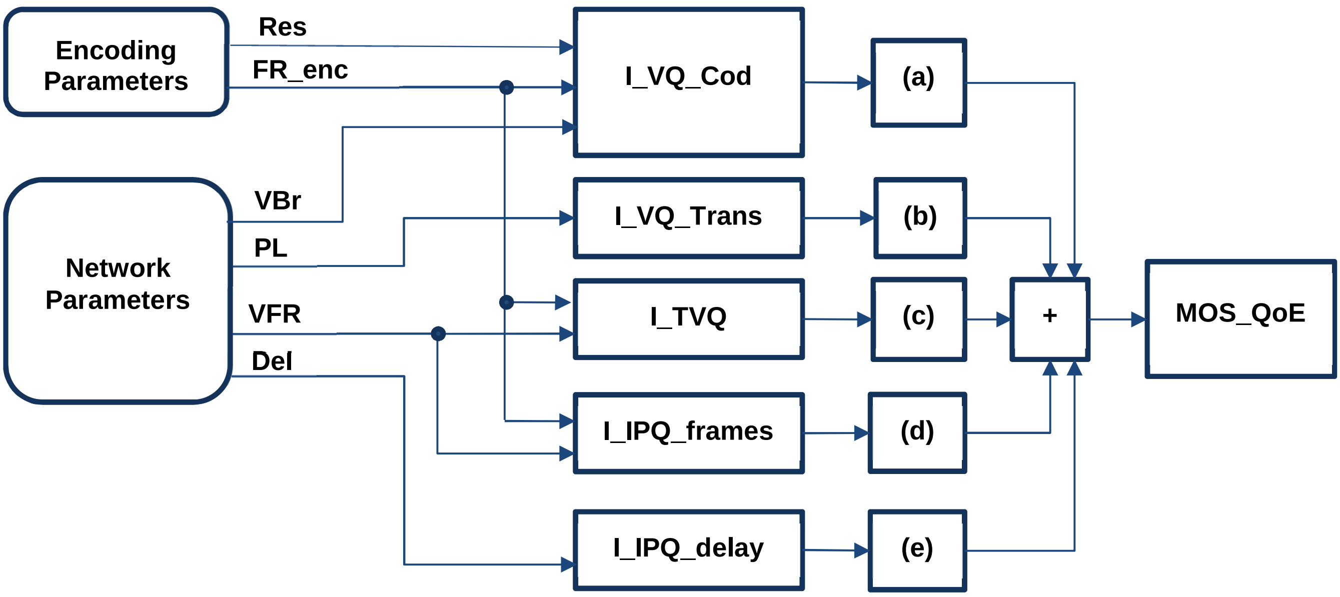
Models specified by Recommendation ITU-T P.1201 [i.1] for parametric non-intrusive assessment of audiovisual media streaming quality are not applicable to monitoring practical cases as they support CC mode - embedded operation mode, assuming the monitoring point located at the terminal side, which is unsuitable for Passive, Non-Intrusive Network Monitoring. In addition, they require the analysis of Video frames (I, P, B) to infer what frame was impacted by packet loss, which is not always possible for encrypted content.

The Recommendation ITU-T G.1072 [i.4]-derived model produces an aggregated QoE indicator (0 to 100, the higher the value, the better the quality) by combining video and input (user commands) quality impairment factors.

A MOS indicator (1 to 5, the higher the value, the better the quality) is derived from the QoE.

The model does not consider the audio component from the game server to the client due to the greater importance of the video in terms of user-perceived quality and network impacts.

The audio from the client to the game server is used as a chat tool for users; it is typically disabled; users need to enable it to join a group. This audio component is not strictly functional to the game logic; it is an "ancillary" function, and is not included in the CG QoE and MOS. Optionally, a separated QoE, MOS may be used for such component, see clause 4.10.



NOTE: The coefficients (a) and (b) are used to weigh video encoding complexity, (c) and (d) frame loss sensitivity, and (e) delay sensitivity.

Figure 6: Cloud Gaming QoE Model (derived from Recommendation ITU-T G.1072 [i.4])

4.7.2 Application range of the model

The model was tested under the conditions listed in table 4.

Three CG platforms, Blacknut®, Nvidia GeForce Now®, Google Stadia®, have been tested in different conditions, both on the fixed network with FTTH technology and on mobile networks of various service providers.

Packet drops, network delays, and packet inter-arrival jitter were injected at different points of the network using the Linux 'tc' command in different flavours [i.15], [i.16] and [i.17].

CG platforms conceal impairment by introducing slicing artefacts. When network conditions become significantly severe, with losses and RTT outside the application ranges of table 4, the game often becomes unresponsive; under such occurrences, CG platforms inform users about the situation and stop streaming.

Table 4: Factors and application ranges of the model

Application Information	Interactive
Sequence Duration	Min 90 seconds of game streaming. Average interactive sequence duration 300 seconds
Screen size	24", 6" (mobile terminal)
Input devices	Mouse, Keyboard, Gamepad
Video Codec	Recommendation ITU-T H.264 [i.8], Recommendation ITU-T H.265 [i.9] (VP9). For the models discussed, Recommendation ITU-T H.265 [i.9] has been considered equivalent to VP9 (see note). The codec used by the CG game was assessed through metadata or test tools provided by each platform
Resolution	1080p, 720p
Codec Video bitrate (Mbps)	1 to 50
Framerate (fps)	30, 60
Encoding mode	CBR, no B-frames
Network packet loss rate	Uniform packets drop 0 % to 10 % (Burst packet drops require further study)
Network delay	Tests were conducted both with static delays up to 400 ms, and with delays composed of a fixed delay plus a variable - gaussian-distributed - delay, to simulate packet inter-arrival jitter, for a total RTT of 400 ms
Indicators' evaluation	10 seconds Time Window
NOTE: At high bitrates (> 5, 10 Mbps) it is difficult discern differences between the two codecs; at lower bit rates (< 1 Mbps) there may be noticeable differences, but it is outside the scope of the present document.	

4.7.3 Modes of operations

The impact of network degradation and encoding distortions on a player's perception of video quality can vary depending on a game's sensitivity to these factors.

To optimize the quality model for video content complexity, Recommendation ITU-T G.1072 [i.4] clause "*6.2 Modes of operation*" introduces three classes of video encoding complexity: low (Class 1), medium (Class 2), and high (Class 3).

To better predict the impact of packet loss and delays on different game categories, two classes of sensitivity are defined: low and high. For example, a gamer playing a First-Person Shooter game is likely to be more sensitive to poor round-trip time than if they were playing a Role-Playing Game.

The present document considers the highest level of encoding complexity (Class 3) and high sensitivity as for non-intrusive monitoring methods, content complexity, and sensitivity information can be challenging to obtain.

To properly evaluate the Quality of Service (QoS) for the model, it is essential to assess it only during active gameplay phases. The video bitrate is typically low during initial streaming, game loading, user-paused phases, and option selections, as these phases contain semi-static, low-complexity content. Including such phases may lead to inaccurate I_VQ_Cod scores that do not reflect the QoS for that dimension.

To ensure accurate troubleshooting and estimation, Time Windows (TW) for measurement and inference should strike a balance between being short enough for precise troubleshooting and long enough for a good estimation. Our experiments indicate that a 10-second TW (equivalent to 300 or 600 video frames) is a suitable compromise.

4.7.4 Model inputs

4.7.4.1 Encoding Parameters

Resolution (Res)

The resolution of the streamed video reflects, in good essence, the user screen resolution. One advantage of cloud gaming resides in managing multiple screen formats.

The screen size information might not always be available to non-intrusive systems; however, it is often possible to recognize the network type.
When user terminal information is not available, but the network type is known, the screen sizes suggested with the current technologies are indicated in table 5.

Table 5: Recommended resolutions when the terminal screen resolution is not available

Network Type	Screen Size	Resolution	Reason
Mobile Network	720p	921600	Current high and medium-end mobile devices have 720p resolution or higher.
Fixed and Wireless Access	1080p	2073600	22 inches or larger screens are typically used.

Frame Encoding (FR_enc) [fps]

Frame encoding may be obtained from parameters game server and client exchange during streaming.
Frame encoding needs to be constantly evaluated during streaming as the game server may decide to reduce it, for example, from 60 fps to 30 fps, to cope with poor network conditions.

4.7.4.2 Network Parameters

Cloud gaming platforms may use Forward Error Correction (FEC) and Automatic Repeat Request (ARQ) techniques to minimize or eliminate network packet losses. To accurately measure the impact of packet losses on video quality, it is requested to apply a *loss-related parameter mapping* as outlined in Recommendation ITU-T P.1201 [i.1], figure 3-b.

Knowing the receiver buffer length can also be helpful, as the receiver considers packets delayed longer than the buffer length to be lost. Each cloud gaming platform's behaviour should be studied to determine the residual packet loss rate based on the platform's ability to compensate for network loss.

Some cloud gaming platforms that use FEC/ARQ techniques can recover from packet drops of up to 5 % when distributed evenly without causing noticeable video artifacts for the user.

Video Bitrate (VBr) [Kbps]

The "Video Average Bitrate DL in TW" (VideoBitrateDL [Kbps]) is used as VBr.

Once identified the appropriate phases for the QoS evaluation, the "Video Average Bitrate DL in TW" value provides a reliable input for the video quality component of the QoS.

Passive monitoring tools should identify and exclude from the QoS calculation initial streaming phases or paused games, where video complexity is typically low, and VBr is consequently low. The resulted I_VQ_Cod would not likely represent the actual video quality due to compression factors in such phases.

Packet Loss (PL) [%]

Packet Loss (PL) [%] refers to the residual network loss rate after the receiver corrections.

Delay (Del) [ms]

Delay (Del) [ms] is the time a (command) packet takes from when a device outputs it to the network to when the device receives back the frame with the command-related result. Input/output transducer delays, detailed in clause 4.9, are not considered by the model; such delays are not specific to the Cloud Gaming technology and cannot be measured or inferred from network monitoring.

$$\text{Del} = \text{Function}(\text{NetInfRTT} + \text{Processing Delay} + \text{VideoJitterDL})$$

Results in good alignment with user experience were found using:

$$\text{Del} = \textit{Average}(\text{NetInfRTT}) + \textit{StandardDeviation}(\text{NetInfRTT}) + \text{Processing Delay} + \textit{Average}(\text{VideoJitterDL}) + \textit{StandardDeviation}(\text{VideoJitterDL})$$

When Del is measured recording the timestamp difference between the joystick command and the screen results timestamps, Command Delay and Playout Delay should be deducted for homogeneity.

The receiver's buffer size acts as a de-jitter (IETF RFC 5481 [i.13]). The buffer size should be determined for each cloud platform. The formula (10) in clause 4.7.5.4 considers a buffer size of 16 ms, the minimum required for a 60-fps frame to be received, which is a fair assumption given the strict real-time needs of CG and our observation on the three CG platforms.

Video Frame Rate (VFR) [fps]

The Recommendation ITU-T G.1072 [i.4] model is designed to estimate video frame rate based on Packet Loss (PL [%]) and Delay (Del). However, the formula only considers the delay when PL [%] is present, neglecting it when PL [%] is null. This approach overlooks the impact of temporary delays caused by packet inter-arrival jitter.

To address this limitation, the proposed model in figure 6 uses the measured video frame rate:

$$\text{VFR} = \text{VideoFrameRate (see clause 4.4.3).}$$

By considering the user's actual viewing experience, it provides accurate results.

4.7.5 Model Indicators

4.7.5.1 Indicators introduction

Five Indicators are calculated by the model illustrated in figure 6, from Recommendation ITU-T G.1072 [i.4]; they are here briefly described.

4.7.5.2 I_VQ_Cod: video quality impairment factor due to video compression artifacts

It is the estimated spatial video quality impairment for video compression artefacts on the R-scale.

This impairment factor considers the basic video parameters to estimate the user-perceived video quality: Video Bitrate (VBr), encoding frame rate (FR_enc), and resolution (Res).
The model also accounts for the screen resolution in terms of number of pixel per frame:

$$\text{I\_VQ\_Cod} = \text{a1v} \times \exp(\text{ a2v} \times \text{BitPerPixel }) + \text{a3v} \times \text{ContentComplexity} + \text{a4v} \tag{1}$$

$$\text{ContentComplexity} = \text{a31} \times \exp(\text{a32} \times \text{BitPerPixel}) + \text{a33} \tag{2}$$

$$\text{BitPerPixel} = (\text{ VBr} \times 10^3) / (\text{ NumPixelPerFrame} \times \text{FR\_enc }) \tag{3}$$

NumPixelPerFrame derives from screen resolution and format.

EXAMPLES:

- 1080p, typical screen size 1920p × 1080p => BitPerPixel = 2073600
- 720p, typical screen size 1280p × 720p => BitPerPixel = 921600
- a1v, a2v, a3v, a4v, a31, a32, a33 are constant coefficients (table 6).

The model provided good results, as shown in clause 4.8, when fed with the measured VBr in the appropriate TWs, following the Modes of operations in clause 4.7.3.

4.7.5.3 I_VQ_trans: video quality impairment factor due to video transmission errors

It is the estimated spatial video quality impairment for video transmission errors on the R-scale.

This quality impairment represents the contributions of PL and Del to video quality degradation. The Recommendation ITU-T G.1072 [i.4] model is based on models proposed in Recommendations ITU-T G.1071 [i.3] and P.1201.2 [i.2].

$$\text{I\_VQ\_Trans} = \text{c1v} \times \log (\text{ c2v} \times \text{LossMagnitudeE} + 1) \tag{4}$$

$$\text{LossMagnitudeE} = \text{q1} \times \exp(\text{ q2} \times \text{LossMagnitudeNP }) - \text{q1} \tag{5}$$

$$\text{LossMagnitudeNP} = ((\text{ c21} - \text{I\_codn }) \times \text{PL }) / (\text{ c23} \times \text{I\_codn} + \text{PL }) \tag{6}$$

$$\text{IF I\_VQ\_Cod} <= 65\text{: I\_codn} = \text{I\_VQ\_Cod, ELSE: I\_codn} = 65 \quad (7)$$

The math symbol log is the logarithm to the base e.

c1v, c2v, q1, q2, c21, and c23 are constant coefficients derived based on the training dataset (table 6).

The coefficients of the column *Recommendation ITU-T H.264 [*i.8*] Codec, Class 3* of table 6 are the same used in the column *Class 3 (default mode)* of table 3 of Recommendation ITU-T G.1072 [i.4].

Table 6: Coefficients of retrained Recommendation ITU-T G.1071 [i.3] impairments for high complexity class 3 for Recommendation ITU-T H.264 [i.8] and Recommendation ITU-T H.265 [i.9]/VP9 codecs

Coefficient	Recommendation ITU-T H.264 [i.8] Codec, Class 3	Recommendation ITU-T H.265 [i.9] /VP9 Codec, Class 3
a1v	47.7463	46
a2v	-12.07	-15
a3v	9.05168	6
a4v	3.41919	3.3
a31	7.62306	5.336142
a32	-167.838	-117.487
a33	0.0760333	0.053223
c1v	1.57176	17.73
c2v	3.68596	123.08
c21	74.0571	80.61
c23	0.00406	0.00147
q1	0.0008685	0.005175
q2	0.868407	0.04

4.7.5.4 I_TVQ: Temporal video quality impairment factor due to frame rate reduction

It is the estimated temporal video quality impairment for the frame rate reduction on the R-scale.

The most common Cloud Gaming platforms known today conceal packet loss causing slicing artifacts, and it is supposed it will be the one used method, as freezing will even worsen user perception.

$$\text{I\_TVQ} = d1 + d2 \times (\text{ FR\_enc })^2 + d3 \times \text{FR\_enc} + d4 \times \log(\text{FrameLossRate} + 1) \quad (8)$$

$$\text{FrameLossRate} = 100 \times (\text{FR\_enc} - \text{AvgFPS}) / (\text{FR\_enc}) \quad (9)$$

IF (Del < 16 ms OR VideoFrameRate [fps] > FR_enc): AvgFPS = FR_enc, ELSE

$$\text{AvgFPS} = \text{VideoFrameRate [fps]} \quad (10)$$

VideoFrameRate [fps] is the measured framerate in TW using RTP packets (table 3); the nearest to receiver it is assessed, the more accurate results are.

4.7.5.5 I_IPQ_frames: Input quality impairment factor due to frame rate reduction(s)

It is the estimated temporal input quality impairment from the frame rate reduction on the R-scale:

$$\text{I\_IPQ\_frames} = e1 + e2 \times (\text{FR\_enc})^2 + e3 \times \text{FR\_enc} + e4 \times \log(\text{FrameLossRate} + 1) \quad (11)$$

Table 7 lists the constant coefficient values e1, e2, e3, and e4 for the High-Sensitive class.

Table 7: Coefficients of I_IPQ for High-Sensitive class (Recommendation ITU-T G.1072 [i.4])

Coefficient	e1	e2	e3	e4
Value (High-Sensitive)	54.71	0.02589	-2.485	9.306

4.7.5.6 I_IPQ_delay: Input quality impairment factory due to network delay degradations

It is the estimated input quality impairment for network delay degradations on the R-scale.

Equations are from Recommendation ITU-T G.1072 [i.4].

$$I\_IPQ\_delay = f1 / (1 + \exp(f2 - f3 \times Del)) + f4 \quad (12)$$

The constant coefficient values f1, f2, f3, and f4 for the High-Sensitive class are shown in table 8.

Table 8: Coefficients of I_IPQ for High-Sensitive class (Recommendation ITU-T G.1072 [i.4])

Coefficient	f1	f2	f3	f4
Value (High-Sensitive)	90	1.191	0.009775	-18.73

4.7.6 QoE and MOS prediction

QoE and MOS are calculated by combining the five indicators I_VQ_Cod, I_VQ_Trans, I_TVQ, I_PQ_frames, and I_IPQ_delay:

$$R\_QoE = R\_Max - (a \times I\_VQ\_Cod + b \times I\_VQ\_Trans + c \times I\_TVQ + d \times I\_PQ\_frames + e \times I\_IPQ\_delay) \quad (13)$$

Table 9 illustrates values for the weighting factors a, b, c, d, and e.

R_QoE is the overall estimated gaming QoE expressed on the R-scale, where zero (0) is the worst quality and one hundred (100) the best.

R_Max is the reference value indicating the best possible gaming QoE (= 100) on the R-scale.

Table 9: Weighting factors to calculate QoS (Recommendation ITU-T G.1072 [i.4])

Coefficient	a	b	c	d	e
Value	0.788	0.896	0.227	0.625	0.848

The MOS estimation derives from R_QoE

```
mosMax = 4.64
mosMin = 1.3

IF 0 < R_QoE < 100:
   MOS = mosMin + R_QoE × (mosMax - mosMin) / 100 +
   R_QoE × (R_QoE - 60) × (100 - R_QoE) × 7.0 × math.pow(10, -6)


ELIF rQoE > 100:
   MOS = mosMax


ELSE:
   MOS = mosMin
```

4.8 Model assessment of video compression artefacts

Validation results of the model for the Recommendation ITU-T H.264 [i.8] codec are shown in Recommendation ITU-T G.1072 [i.4], clause "*9 Performance of the model*".
The model has been used for Recommendation ITU-T H.265 [i.9]/VP9 codecs, for which the new parameters of column *Recommendation ITU-T H.265 [*i.9*]/VP9 Codec, Class 3* of table 6 have been identified and used to assess the validity.

Both Recommendations ITU-T H.264 [i.8] and H.265 [i.9] have been validated using a reference video (Y4M format) with high content complexity which has been encoded with lossless quality and proposed to a panel of score voters as the reference video assuming a MOS of 4,5.

The reference video was encoded Recommendation ITU-T H.264 [i.8] and Recommendation ITU-T H.265 [i.9] at 1 Mbps, 2 Mbps, 3 Mbps, 6 Mbps, 9 Mbps, 12 Mbps, and 15 Mbps CBR with no B frames to be as next as possible to Cloud Gaming encoding methods. The obtained video clips were independently assessed by 25 people of different ages, 60 % male, 40 % female, using the same screen. Each person had to provide a score from 1 to 4,5. The reference video was not part of the set to be scored but was available for reference to the single individual during the assessment, communicating it had a MOS = 4,5.

Table 10 Reports the performance of the model for the two codecs in terms of Root-Mean-Square Error (RMSE) and Pearson Linear Correlation Coefficient (PLCC), as described in Recommendation ITU-T P.1401 [i.18].

Figure 7 and figure 8 plot the model predicted scores (X-axis) versus the testers' scores (Y-axis). It is noticeable the models tend to be optimistic.

Table 10: Final prediction performance of the model on Recommendations ITU-T H.264 [i.8] and H.265 [i.9] test sets

	Recommendation ITU-T H.264 [i.8] Codec MOS-scale	Recommendation ITU-T H.265 [i.9] MOS-Scale
RMSE	0.55	0.53
PLCC	0.92	0.89

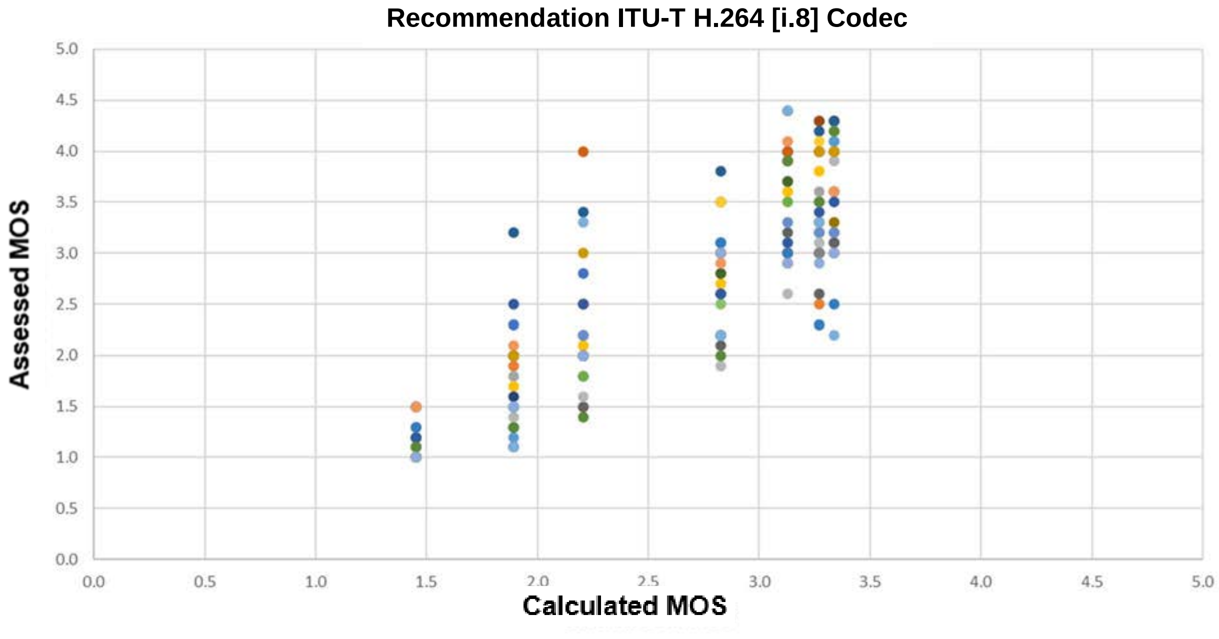


Figure 7: Predicted VS Assessed scores for Recommendation ITU-T H.264 [i.8] codec CBR No B-Frames

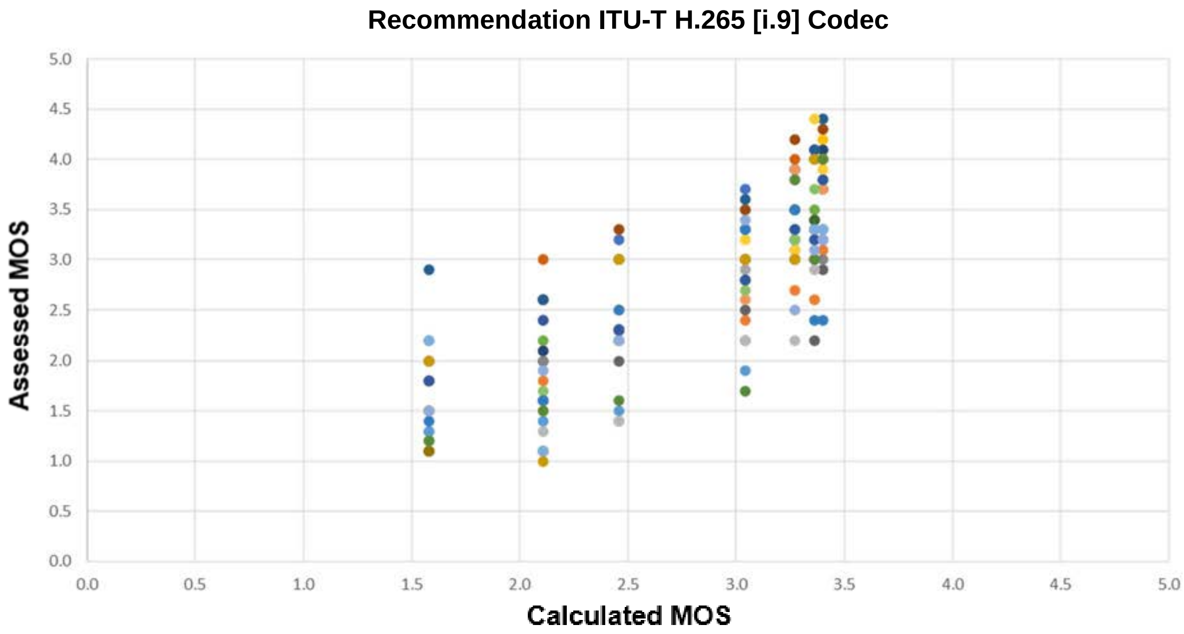


Figure 8: Predicted VS Assessed scores for Recommendation ITU-T H.265 [i.9] codec CBR No B-Frames

4.9 Motion To Photon (MTP)

The total time it takes for a user-command from the gamepad input to the relative result to the screen is called Motion To Photon (MTP).

In [i.10], the MTP is dissected into the major contributors, **MTP = CD + RTT + PD + OD**, shown in figure 9.

Control Delay (CD)
CD is inside the mobile device, to be a delay between the user initiating a control input and the control sent through the radio interface of the mobile device.

Uplink plus Downlink Network Delay (RTT)
Round Trip Time (RTT) is the total time for a packet to go back and forth in the network (UL Network Delay + DL Network Delay in figure 9).

Processing Delay (PD)
It is the time necessary for the game server to process the command, render and capture the scene, encode the frame.

Playout Delay (OD)
Time for the Client to receive, decode, and display the scene.

Experiments [i.11] show that a person starts perceiving a delay, named **Just Noticeable Time (JNT)**, for MTP values next to 100 ms when tapping the screen, or 60 ms for dragging.

Each MTP addendum is now considered.

Control Delay (CD) for the latest mobile devices is lower than 30 ms for screen tapping, while it is in the range of 17 ms for a USB-connected joystick (see [i.10], section 3.1.1). Cloud gaming providers advertise mobile phones and gamepads as a combined package with the advantage of the lowest CD, besides better usability. To further reduce CD, some CG platforms provide joysticks with Wi-Fi® capabilities that directly connect to the game server through the home modem.

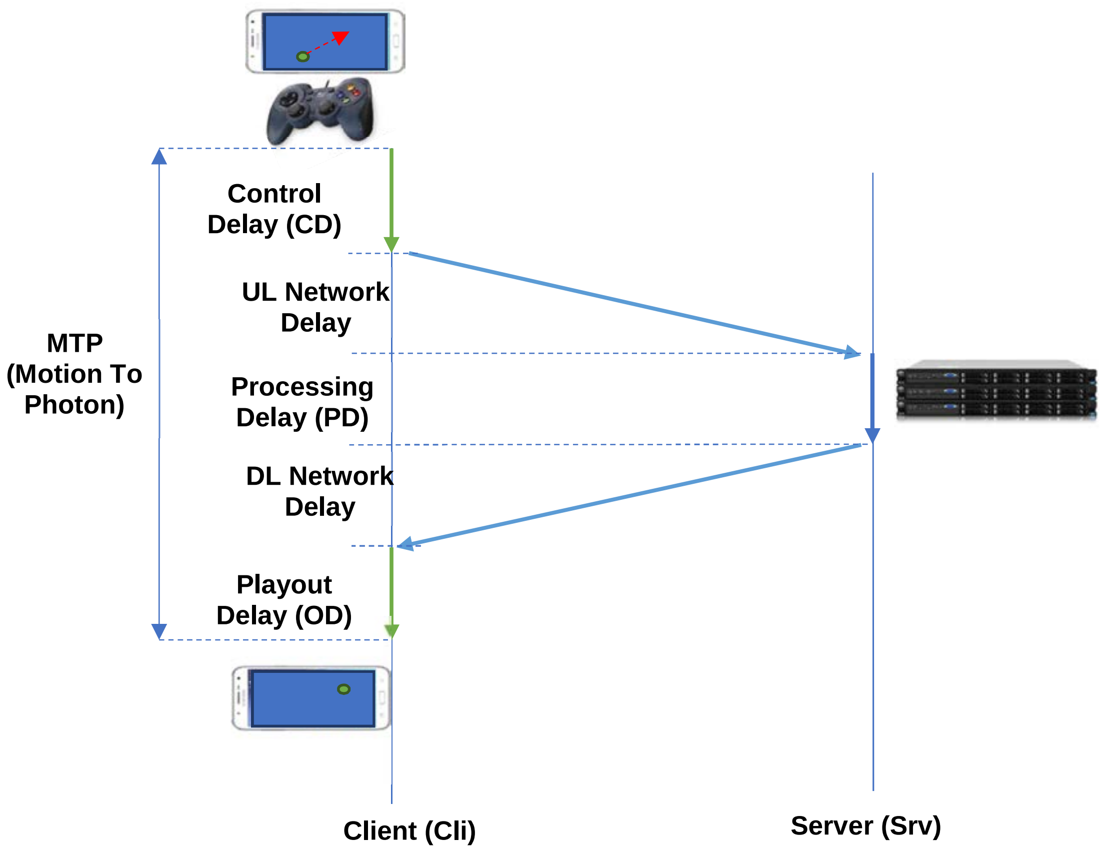


Figure 9: Motion To Photon (MTP) components

Processing Delays (PD) is where CG excels. Cloud games servers use top-notch graphics cards, basically encapsulating high-end console technologies, with a clear advantage over the average capabilities of home PCs. In [i.10], section 3.3, 17 ms PD is considered a reasonable average value.

Playout Delay (OD) includes frame reception, decoding, and display. According to [i.10], section 3.1.2, frame reception is related to the bit rate (resolution) and frame rate used; it does not vary much with the phone model and has average values in the range of 10 ms. With the latest mobile phones, frame decoding may be less than 8 ms, also because low complexity codecs are preferred, and no B-frames are used for encoding to privilege speed. Frame display mostly depends on the screen refresh rate that, in the latest mobile phones, is 60 Hz, which translates to 17 ms. To avoid screen tearing, "double buffering" is used, which brings the average display delay to 25,5 ms (17 x 1,5). Summing up all the components, OD is in the range of 44 (10 + 8 + 25,3) ms.

According to [i.12], MTP values in the range of 100 ms are noticeable by FPS game players; higher MTP values are tolerated; with MTP values greater than 200 ms, users may decide to leave the game due to experience degradation.

The degree of MTP tolerance strongly depends on the game type. FPS games are the most demanding, whereas RPG games may tolerate up to 500 ms MTP, ending with RTS games which allow the higher MTP values of up to one second (see [i.14], table 1).

Table 11 shows the maximum RTT (Max NetRTT) permitted to the network considering the ideal case JNT and the tolerated Total Delay Budget for FPS, RPG, RTS games, assuming all other delays are highly optimized.

Table 11: Maximum Network RTT allowed for JNT and maximum allowed FPS, RPG, RTS games, assuming optimistically that all other delays are highly optimized

		JNT	Max FPS	Max RPG	Max RTS
Total Delay Budget (ms)	**(a)**	**100**	**200**	**500**	**1 000**
Command Delay (see note 1)	(b)	17	17	17	17
Processing Delay (see note 2)	(c)	17	17	17	17
Playout Delay (see note 3)	(d)	45	45	45	45
Max NetRTT	(a)-(b)-(c)-(d)	21	121	421	921
NOTE 1: Lowest, best-in-class delay for Gamepad with USB (see [i.10], section 3.1.1). NOTE 2: Average best-in-class server processing (see [i.10], section 3.3). NOTE 3: 60 Hz screen refresh rate, double buffering of the latest mobile terminals (see [i.10], section 3.1.2).					

For FPS games, when all ingredients making up for the MTP have the values indicated above, Max NetRTT in the range of 20 ms is required to provide a good/excellent experience; values greater than 120 ms may cause gamers to abandon the game.

4.10 Uplink Audio QoE and MOS prediction

The audio from the client to the game server is not strictly functional for the game as it can be used as a chat by users. Users need to enable it to join a chat group.

It is possible to provide a QoE and MOS prediction using the model from Recommendation ITU-T P.1201.2 [i.2] (Higher Resolution, audio part) for the codecs supported by such standard.

History

Document history		
V1.1.1	October 2023	Publication