Video Quality Experts Group (VQEG)

VQEG Technical Report 2026/001

(VQEG_TR_2026_001)

VQEG White Paper on Quality of Experience-Aware Management for Collaboration Between Network and Application Providers

March 2026

https://doi.org/10.66537/OLKA7578

Authors

Pablo Pérez [Editor] (Nokia)

François Blouin [Editor] (Meta)

Balu Adsumilli (YouTube)

Sara Baldoni (University of Padova)

Federica Battisti (University of Padova)

Neil Birkbeck (YouTube)

Kjell Brunnström (RISE)

Luis M. Contreras (Telefónica)

Markus Fiedler (Blekinge Institute of Technology)

Jesús Folgueira (Telefónica)

Narciso García (Universidad Politécnica de Madrid)

Emir Halepovic (AT&T)

Tobias Hoßfeld (University of Würzburg)

Theo Karagioules (Audible, work done prior to joining Audible/Amazon)

Ioannis Katsavounidis (Meta)

Kamil Koniuch (AGH)

David Lindero (Ericsson)

Szilveszter Nadas (Ericsson)

Marta Orduna (Nokia)

Pablo Rojo (Nokia)

Alexander Raake (RWTH Aachen)

Rakesh Rao Ramachandra Rao (TU Ilmenau)

Werner Robitza (AVEQ)

Tina Tsou (TikTok)

Michele Zorzi (University of Padova)

Table of Contents

Abbreviations

3GPP: 3rd Generation Partnership Project

5QI: 5G QoS Identifier

6DoF: Six Degrees of Freedom

ABR: Adaptive Bitrate

ACR: Absolute Category Rating

ACR-HR: Absolute Category Rating with Hidden Reference

AR: Augmented Reality

ATM: Asynchronous Transfer Mode

AV1: AOMedia Video 1

BJ: Bandwidth Jitter

BRAS: Broadband Remote Access Server

CAP: Content and Application Provider

CDN: Content Delivery Network

CMCD: Common Media Client Data

CS: Circuit Switched

CSP: Communication Service Provider

CPE: Customer Premises Equipment

DL: Downlink

DMOS: Differential Mean Opinion Score

DNS: Domain Name System

DSCP: Differentiated Services Code Point

DSIS: Double Stimulus Impairment Scale

E2E: End-to-End

ECN: Explicit Congestion Notification

EVS: Enhanced Voice Services

fps: frames per second

FR: Full Reference

FTTH: Fiber-to-the-Home

FWA: Fixed Wireless Access

GDPR: General Data Protection Regulation

GEO: Geostationary Orbit

gNB: next-generation Node B

GoB: Good-or-Better (ratio)

GoP: Group of Pictures

GPON: Gigabit-capable Passive Optical Network

GTP: GPRS Tunneling Protocol

HAS: HTTP-based Adaptive Streaming

HD: High Definition

HDR: High Dynamic Range

HLS: HTTP Live Streaming

HR: High Resolution

HVGA: Half-size Video Graphics Array

IETF: Internet Engineering Task Force

IOAM: In-situ Operations, Administration, and Maintenance

IP: Internet Protocol

IPTV: Internet Protocol Television

ISO: International Organization for Standardization

ITU-T: International Telecommunication Union – Telecommunication Standardization Sector

KPI: Key Performance Indicator

KQI: Key Quality Indicator

L2TP: Layer 2 Tunneling Protocol

L4S: Low Latency, Low Loss, Scalable

throughput

LAN: Local Area Network

LEO: Low Earth Orbit

LR: Low Resolution

MCS: Modulation and Coding Scheme

MEC: Multi-access Edge Computing

MEF: Metro Ethernet Forum

MMS: Multimedia Messaging Service

MOS: Mean Opinion Score

MPLS: Multi-Protocol Label Switching

MPEG-DASH: Moving Picture Experts Group – Dynamic Adaptive Streaming over HTTP

MTU: Maximum Transmission Unit

MUSHRA: MUltiple Stimuli with Hidden Reference and Anchor

NR: No Reference

NRLP: Network Rate-Limit Policies

NWDAF: Network Data Analytics Function

OAM: Operations, Administration, and Maintenance

OLT: Optical Line Terminal

ONT: Optical Network Terminal

OS: Operating System

OSI: Open Systems Interconnection

OTT: Over-The-Top

PaDIS: Provider-aided Distance Information System

P.BBQCG: ITU-T Recommendation P.BBQCG (Parametric Bitstream-based Quality model for Cloud Gaming)

PDU: Protocol Data Unit

PEAQ: Perceptual Evaluation of Audio Quality

POLQA: Perceptual Objective Listening Quality Assessment

PoW: Poor-or-Worse (ratio)

PQoMS: Perceived Quality of Media Signal

PRB: Physical Resource Block

QCI: QoS Class Identifier

QMC: QoE Measurement Collection

QoE: Quality of Experience

QoS: Quality of Service

QP: Quantization Parameter

QUX: Quality of User eXperience

RA: Router Advertisements

RAN: Radio Access Network

RG: Residential Gateway

RISE: Research Institutes of Sweden

RMSE: Root Mean Squared Error

RR: Reduced Reference

RSRP: Reference Signal Received Power

RSRQ: Reference Signal Received Quality

RSSI: Received Signal Strength Indicator

RTC: Real-time Communications

RTT: Round-Trip Time

SAMVIQ: Subjective Assessment of Multimedia Video Quality

SCONE: Standard Communication with Network Elements

SD: Standard Definition

SDH: Synchronous Digital Hierarchy

SG12: Study Group 12

SLA: Service Level Agreement

SMS: Short Message Service

SNI: Server Name Indication

SNR: Signal-to-Noise Ratio

T-CONT: Traffic Container

TCP: Transmission Control Protocol

TRONE: Transparent Rate Optimization for Network Endpoints

TSG-SA5: Technical Specification Group – Service and System Aspects – Working Group 5

UE: User Equipment

UHD: Ultra High Definition

UL: Uplink

UPF: User Plane Function

UVQ: Universal Video Quality, a YouTube video quality metric

UX: User Experience

VMAF: Video Multimethod Assessment Fusion

VoD: Video on Demand

VoLTE: Voice over LTE

VoNR: Voice over New Radio

VoWIFI: Voice over Wi-Fi

VPN: Virtual Private Network

VQA: Video Quality Assessment

VQEG: Video Quality Experts Group

VR: Virtual Reality

XR: Extended Reality

Executive Summary

This VQEG white paper addresses the challenge of improving end-user Quality of Experience (QoE) for Internet services. It begins by outlining the core problem: Content and Application Providers (CAPs) and Communication Service Providers (CSPs) operate largely independently without a common view of their users' experience, given that the default Internet connectivity provided is Best-Effort. This separation makes it difficult to diagnose end-user issues and optimize performance from an end-to-end perspective.

The goal of the white paper is establishing a framework which facilitates a common view of the end-user's experience of service quality. Then, exchanging respective information between CAPs and CSPs allows for managing those services efficiently.

While aiming for broad applicability, it offers a more in-depth analysis of specific services to provide practical insights: short-form video, long-form video, and interactive services such as cloud gaming and video conferencing.

The white paper first establishes a common foundation by reviewing existing QoS and QoE definitions, QoE models and relevant industry standards. It presents a layered model to define key concepts, separating network-level Key Performance Indicators (KPIs), application-specific Key Quality Indicators (KQIs), and the user-centric QoE, proposing clear definitions for some important QoE-related terms, such as user-reported QoE, modeled QoE, or system QoE. This provides a common language for understanding the remainder of the paper and discussions in the research community.

The potential benefits of QoE management are discussed with respect to typical issues and common information gaps, making the case for closer collaboration between CAPs and CSPs. The core proposal is a framework for structured information exchange between those stakeholders. This mechanism, described as a shared state table, allows for the exchange of relevant metrics – either in near real-time or periodically, and with different granularity (e.g. aggregated) – to create a shared view of service and network performance.

This exchange of information enables cooperative optimization. CSPs, on the one hand, can use QoE-related data from CAPs (e.g., video fidelity scores, stalling events) to better understand the impact of network conditions and adjust resource management accordingly. CAPs, on the other hand, can use network status information from CSPs (e.g., congestion levels, available throughput) to make network-aware adaptation decisions, such as selecting an appropriate video quality to avoid stalls.

To demonstrate the framework's real-world value, the white paper also illustrates its application through practical use cases for short and long-form video streaming, as well as interactive services like cloud gaming. These examples show how specific metrics can be used to improve startup times, reduce stalling, and manage latency. Finally, the framework addresses key privacy considerations, proposing a voluntary, opt-in system that uses practices such as temporary, pseudonymized session identifiers.

In conclusion, this VQEG white paper presents a structured approach for improving QoE through enhanced cooperation between CAPs and CSPs. The proposed framework provides a foundation for developing and sharing metrics that can lead to more efficient and effective

service delivery. The recommended next steps include further development and validation of the proposed models through a proof-of-concept, with the long-term goal of contributing the findings to relevant standardization bodies such as ITU-T and IETF.

Status and Disclaimer

This document is provided solely for informational and discussion purposes. It presents hypothetical concepts and potential scenarios intended to explore possibilities and does not constitute an actual business case, recommendation, proposal, commitment, or statement of intent. Any references to strategies, products, services, features, timelines, costs, performance, or outcomes are illustrative only and may not reflect real-world availability or results.

The views and opinions expressed in this document are those of the individual contributor(s) in their personal capacity and do not represent the views, positions, policies, or endorsements of any contributor's employer(s) (including any affiliated entities). No part of this document should be construed as creating any obligation, warranty, or reliance, nor should it be used as the basis for any commercial, legal, financial, or operational decision.

1 Introduction and Scope

1.1 Motivation

As the digital ecosystem continues to evolve, the collaboration between Content and Application Providers (CAPs) and Communication Service Providers (CSPs) becomes increasingly critical to ensure high Quality of Experience (QoE) for end users. However, current industry practices are hindered by persistent structural and informational silos between these two stakeholder groups. These "high walls" lead to fragmented visibility and misaligned optimization strategies, ultimately limiting the effectiveness of network and application-level enhancements (Tomar et al., 2025).

One problem, in a monitoring sense, in today's Internet is the fact that the traffic is typically encrypted. Encryption hides packet content, making it difficult to identify the type of traffic for CSPs, such as video, gaming, or general browsing. This makes it hard to prioritize or optimize network resources based on the type of service to improve the quality. For the same reason, it is hard to analyze usage patterns and forecast future demand accurately, which is crucial for long-term capacity planning and infrastructure investments (Papadogiannaki & Ioannidis, 2021)

Due to encryption, the monitoring of quality is - with reasonable effort - only possible at the application layer. It is by far more difficult otherwise, e.g., within the network, resulting in a challenge for CSPs due to lack of accessibility and thus visibility. Sophisticated approaches based on available and accessible information below the application layer are required, which may result in inaccuracies.

For example, advances in video streaming technology have led to the development of HTTP Adaptive Streaming (HAS), which adapts the video stream to the network characteristics. Also, newer generations of video codecs have decreased bandwidth requirements significantly, while optimizing visual quality, making it hard for CSPs to directly relate bandwidth to QoE. Combined with encryption, these factors obscure content details, complicating CSPs' ability to accurately monitor and ensure consistent quality with respect to the needs of the corresponding application (Robitza et al., 2017). Consequently, algorithms predicting end-user experienced QoE from network-layer data are prone to such inaccuracies and hardly actionable. Contrary, the end-user device and the application layer provide the most insights into when and what actually reaches the user and how the user interacts with the service or device.

A common issue is that Quality of Service (QoS) has traditionally been the primary focus for optimizing end-to-end networks (including end-devices). QoS refers to the technical performance of a service, quantified by easily measurable, objective metrics like bandwidth, throughput, latency, jitter, and packet loss at the network or application layer, indicating how effectively the network delivers the service. However, QoS is mainly implemented and enforced at the networking layers, although the original definition includes conceptual elements and controls at higher layers as well. In practice, the term QoS is often (wrongly) used as a synonym for network performance parameters. To overcome this issue, we introduce proper definitions and notions in Section 2.

In contrast, Quality of Experience (QoE) offers a user-centric perspective on overall service quality. It measures the end user's satisfaction (or frustration) with an application or service, incorporating subjective factors beyond raw network metrics (Le Callet et al., 2012).

This distinction is crucial: QoS is system-focused, dealing with service delivery parameters, whereas QoE is user-focused, reflecting the service quality as perceived by the user. Importantly, high QoS does not automatically guarantee high QoE if other aspects of the user experience are lacking. In collaborations between Communication Access Providers (CAPs) and Communication Service Providers (CSPs), it is essential to align efforts so that robust network QoS translates into a superior QoE for end users.

Due to the problems mentioned above, CSPs face a number of systemic challenges, including limited access to granular QoE metrics and traffic type identification. As a result, they rely heavily on inference-based models that are often inaccurate and difficult to operationalize. Additionally, CSPs must contend with diverse QoE targets and QoS requirements across multiple traffic classes and service types (e.g., mobile vs. fixed), while reconciling best-effort delivery with heterogeneous fairness and cost expectations. The emergence of new traffic patterns and services further compounds these difficulties.

Simultaneously, CAPs face challenges due to a lack of real-time network visibility. This includes insufficient insight into congestion, latency, link utilization, and buffer occupancy. This limited visibility hinders their ability to choose effective adaptation strategies and implement network-aware policies. Additionally, without end-to-end visibility throughout the delivery chain, CAPs struggle to accurately identify the source of Quality of Service (QoS) impairments (e.g., pinpointing the specific network segment or service component experiencing issues) and effectively manage Quality of Experience (QoE).

A key underlying issue is the absence of standardized key performance indicators (KPIs) and interfaces to enable mutual understanding and cooperative optimization. This lack of KPI uniformity not only impedes internal efforts within CAPs and CSPs but also prevents meaningful collaboration between them.

One can argue that standardization offers appropriate means like 5G QoS Identifier (5QI) as defined by 3GPP in TS 23.501, which define the quality requirements for different types of network traffic, like voice calls, video streaming, or browsing. Each 5QI is associated with certain QoS parameters such as latency, packet loss tolerance, and priority level. However, the corresponding tables were designed with static traffic in mind and assume predictable traffic patterns and service types. Static and predictable traffic refers to consistent and regular data flows with behavior that can be anticipated based on the application or service type. However, modern applications often exhibit dynamic behavior, meaning traffic patterns change frequently, and encryption obscures the nature of the traffic, making it harder to classify and manage using traditional QoS frameworks. Hence, the dynamic and unpredictable nature of modern applications represents one of the most important points that we consider to be critical in practice.

Addressing these challenges calls for a coordinated, cross-industry approach to harmonize metrics, improve transparency, and enable intelligent, end-to-end QoE optimization. This topic has been widely discussed in the Video Quality Experts Group (VQEG). Under the umbrella of its 5G-KPI working group, VQEG hosted a workshop on this in its plenary meeting held in Klagenfurt in July 2024, where the vision on the matter was updated. It was also decided to write this white paper to structure the state of the art, identify the main challenges coming from the industry, and propose an initial framework to address them.

1.2 Related Work

Approaches to collaboration between CSPs and CAPs have evolved significantly over the past two decades, driven by the need to optimize QoS and QoE for end users. We can identify two main approaches: resource sharing whereby CAPs offer capacity to CSPs, and an explicit sharing of information to help CAPs/CSPs optimize quality via means of data exchange.

With respect to resource sharing, Netflix launched its Open Connect program[1] in 2012, representing one of the most significant practical implementations of CSP–CAP collaboration. The program provides two main components: embedded Open Connect Appliances (OCAs) and settlement-free interconnection, both architected in partnership with CSPs to maximize benefits for end-users. Netflix currently partners with over a thousand ISPs to localize substantial amounts of traffic through embedded OCA deployments. In a similar vein, Google implemented its Global Cache (GGC) program[2] to allow ISPs to serve Google content from within their own networks. The GGC features transparent operation to users, reducing external traffic for cacheable content. As another example, Akamai developed the Akamai Accelerated Network Partner (AANP) program,[3] designed to benefit network operators, end users, and Akamai's customers while minimizing transit costs through deployment of Akamai servers in operator facilities. Complementary, CSPs can share inbound traffic engineering policies in the form of Border Gateway Protocol (BGP) advertisements, enabling traffic to be delivered into the CSP network from an optimal (localized) location. Programs like these are a successful example of CSP–CAP collaboration today. This collaboration is currently targeted at improving network efficiency, which is a starting point to the wider goal of optimizing QoE.

Regarding explicit information sharing, Fiedler (2009) introduced the concept of "Quality Feedback Flows in Future Networks," proposing a feedback system employing self-organizing overlay technology to provide explicit feedback between users, applications, networks, and service providers. The framework identified network-to-application feedback (moving beyond implicit feedback through packet loss and delays to explicit quality notifications) and application-to-network feedback: enabling applications to communicate their perceived network conditions back to CSPs. The paper also proposed a cross-stakeholder communication, facilitating coordinated quality management across different entities.

Standards bodies like ETSI have formalized the complexity of measuring/improving QoE (ETSI TS 103 294); they reference the ARCU (Application, Resource, Context, User) model from Skorin-Kapov & Varela (2012), which categorizes QoE influence factors into four distinct domains: Application (A): Content type, video resolution, codec. Resource (R): Network QoS, device CPU/memory. Context (C): Location, time of day, subscription cost. User (U): Expectations, mood, demographics. Neither the CSP nor the CAP has access to all four domains. The CSP partially sees the 'R' (Resource) and 'C' (Context) domains. The CAP sees the 'A' (Application) domain and has some insight into 'U' (User) and 'C' (Context) via its application analytics.

---

[1] https://openconnect.netflix.com/en/

[2] https://support.google.com/interconnect/answer/9058809?hl=en

[3] https://www.akamai.com/site/en/documents/corporate/2021/akamai-accelerated-network-partner-aanp-faq.pdf

This shared but incomplete view is one justification for why collaboration is beneficial. Robitza et al. (2017) proposed a "QoE API" for enabling systematic information exchange between CAPs and CSPs. Their research highlighted the quality monitoring issues related to pervasive encryption and identified three possible approaches for information sharing: 1) active CAP to CSP communication: the CAP actively sends quality information to the CSP, 2) active CSP to CAP requests: the CSP actively requests information from CAP servers, 3) passive monitoring integration: Information is embedded in regular transmissions for passive monitoring. For the first two approaches, dedicated servers must be established at either location, requiring negotiation of a formal "QoE API" between CAPs and CSPs. This proposal addressed the challenge that OTT providers possess detailed application-level quality information (such as video stalling events) that ISPs cannot access due to encryption.

Floris et al. (2018) also claim that effective QoE monitoring requires collaboration because no single provider has access to all necessary measurement tools or control over all quality-influencing factors. CAPs typically have access to application-level metrics, user context information, and device capabilities, while CSPs control network resources and can monitor network-level performance indicators.

Frank et al. (2013) present NetPaaS, a protocol that "increases CDN capacity on-demand, enables coordination, reduces download time, and achieves multiple traffic engineering goals leading to a win-win situation for both ISP and CDN." The protocol enables CDNs to provide server specifications, and CSPs to communicate resources and prices. It builds upon research by Poese et al. (2010), which presents the Provider-aided Distance Information System (PaDIS). This system is operated by CSPs to influence user-to-server assignments by utilizing information available only to them, such as network conditions and end-user location. PaDIS exploits server diversity exposed by CDN server selection processes, enabling joint optimization of network and hosting infrastructure. More information about the group can be found on their website.

To summarize, these research-oriented approaches to collaboration between CSPs and CAPs have shown promising ideas, but few have materialized in practice, with some exceptions presented later in Section 3.3.3.

1.3 Vision

The long-term vision of VQEG 5G-KPI working group is to develop QoE models and frameworks for QoE management in current and future network architectures such as 5G-advanced, 6G, GPON or WiFi.

The goal is to achieve more efficient end-to-end multimedia service delivery by improving collaboration and cooperation among ecosystem stakeholders (CAPs, CSPs and users). This will involve identifying and optimizing the most critical influencing factors for user QoE and resource utilization. By searching for an ecosystem win-win-win approach between CAPs, CSPs and users, it should be possible to overcome lack of trust and enhance cooperation among stakeholders.

1.4 Scope

The aim of this white paper is to rationalize and provide a **clear definition** and understanding of QoS and QoE metrics in a way which is practical for the main industry players (CAPs, CSPs), so that they can agree on which metric(s) to use, share and expose, how to measure, and take action on the results.

This white paper examines **QoS and QoE information** at the interface between Content Application Providers (CAPs) and Communication Service Providers (CSPs), recognizing them as the most crucial stakeholders for multimedia communication services. CAPs directly serve users, providing optimal access to QoE information, while CSPs manage the communication network, offering the best QoS insights. Although other stakeholders like technology providers and device manufacturers contribute to the end-to-end chain, their interactions are not explicitly detailed here. Similarly, the role of Test and Measurement Providers (T&Ms) as independent validators will not be covered either. The primary focus remains on **CAP and CSP interactions**, as they control the most critical resources impacting end-user QoE and overall service quality.

The target of the white paper is being generic (applicable to as many use cases as possible), but also to present examples of the proposed approach with specific use cases. To this aim, the white paper will provide detailed analysis of long-form and short-form video streaming, as well as interactive services such as cloud gaming of video conferencing.

The white paper intends to be precise and clear in the proper use of terminology and modeling of Quality of Experience. It does not aim to replace existing definitions of QoS and QoE, but selects a working set of precise and clear definitions that are consistent with each other and suitable for its context. Simultaneously, it intends to be realistic in the considerations of what can be practically achieved by CAPs and CSPs within their normal operation constraints. To this end, the analysis builds upon the challenges identified by CAPs and CSPs.

The specific scope of this white paper includes:

- Structuring the state of the art and provide clear definitions and guidelines on how to manage QoE in communication networks.
- Reviewing existing standards for QoE models and metrics.
- Analyzing the challenges identified by CAPs and CSPs and sketch possible solutions.
- Defining a conceptual QoE management framework to exchange QoS and QoE information between CAPs and CSPs which facilitates end-to-end QoE optimization.

The following items are explicitly outside the scope of this white paper:

- Specifying which individual QoE models and tools should be used to monitor QoE in the network.
- Specifying which protocols and technical mechanisms should be used by CAPs and CSPs to exchange QoS and QoE information.
- Specifying legal requirements or regulatory policies. These are subject to local laws and regulations, which may vary by jurisdiction.

1.4.1 Use Cases under Consideration

A communication network is used to handle a large variety of services in parallel. Table 1 shows the most relevant services under consideration.

Category	Type	Examples
Traditional 3GPP Applications	Messaging	SMS and MMS, RCS
	Voice	Circuit-Switched Voice, VoLTE, VoNR and VoWIFI
	Web Browsing	News, eLearning, shopping

Connected Application	Social Media	Facebook, Instagram, TikTok, X
	Internet of Things	Wearables, Connected home and car, industry applications, remote sensors, asset tracking
	Online Gaming	Massively multiplayer online games
Real-time Communications (RTC)	Video Conferencing	WhatsApp, FaceTime, Zoom, Teams, etc.
	Voice	
	Messaging	
Streaming Media	Long-form Video	Netflix, YouTube, etc.
	Short-form Video	TikTok, Instagram Reels, YouTube Shorts
	Live Video	TV, Live Events (Sport, Music, etc.)
	Audio Streaming	Spotify, Tidal, etc
Immersive Applications	Cloud Gaming[4]	Nvidia, Xbox, Luna, PlayStation, etc.
	Augmented and Virtual Reality	Industrial Applications
File Transfer	Backup, Upgrades and Transfers	Dropbox, Windows Update, SpeedTest, etc.

Table 1 – Most relevant services in communication networks.

As previously stated, the work in this white paper is aimed at ultimately covering any possible service. However, for practical reasons, the paper shows a deeper analysis on a few of them, so that the work can be specific enough to be useful. In particular, the following three use cases will be addressed:

- Short-form video
- Long-form video
- Interactive services: Cloud gaming/video conferencing

QoE in media services is influenced by different factors throughout the production and delivery chain, from content generation and production, through content placement in CDNs, end-to-end communication, to end device and client capabilities and behaviors. Different kinds of degradations in QoE can appear and we group them under these general attributes of QoE:

- Media fidelity, e.g., audio/video quality degradations (Resolution, frame rate, color depth, blockiness, blur) caused by issues in source quality, encoding process, adaptive protocols, as well as by the capabilities of the device and by the behaviour of the end-user.
- Media delivery continuity, e.g., stalls and freezes that interrupt smooth playback, often caused by throughput loss, bursty packet loss, adaptation logic issues, etc.

---

[4] Cloud gaming refers to remotely rendered games, where a cloud computer processes and streams video to the end user device for remote play.

- Media delay and interactivity, e.g., long startup delay, delayed interaction, overlapping speech, often caused by transmission latency, video swiping, network congestion, etc.

Each of these attributes may have a different overall effect in the QoE, depending on the use case. Table 2 shows the (high-level) relative importance of such attributes for each of the considered use cases, which are selected to highlight variation in requirements.

Type	Media Fidelity	Media Delivery Continuity	Media Delay and Interactivity
Video Conferencing	X	XXX	XXX
Long-form Video	XX	XXX	X
Short-form Video	XX	XX	XXX

Table 2 – Priority and importance of key attributes for each service. This is an illustrative view to provide an example of how different QoE influencing factors are impacting different use cases.

1.4.2 Network Architecture Landscape

Figure 1 shows a blueprint architecture of the most relevant network scenarios that will be considered in this document. The top part shows a simplified view of a CAP, including an application in the client device (potentially running QoS/QoE estimation models), an optional step deployed in the edge of the network (typically a CDN cache) and a backend running in the cloud. The bottom part shows the main components of the network of a CSP. We are considering two access technologies: 5G mobile access and G-PON FTTH access. Additionally, we may consider the presence of a local area network (LAN) in the end user location.

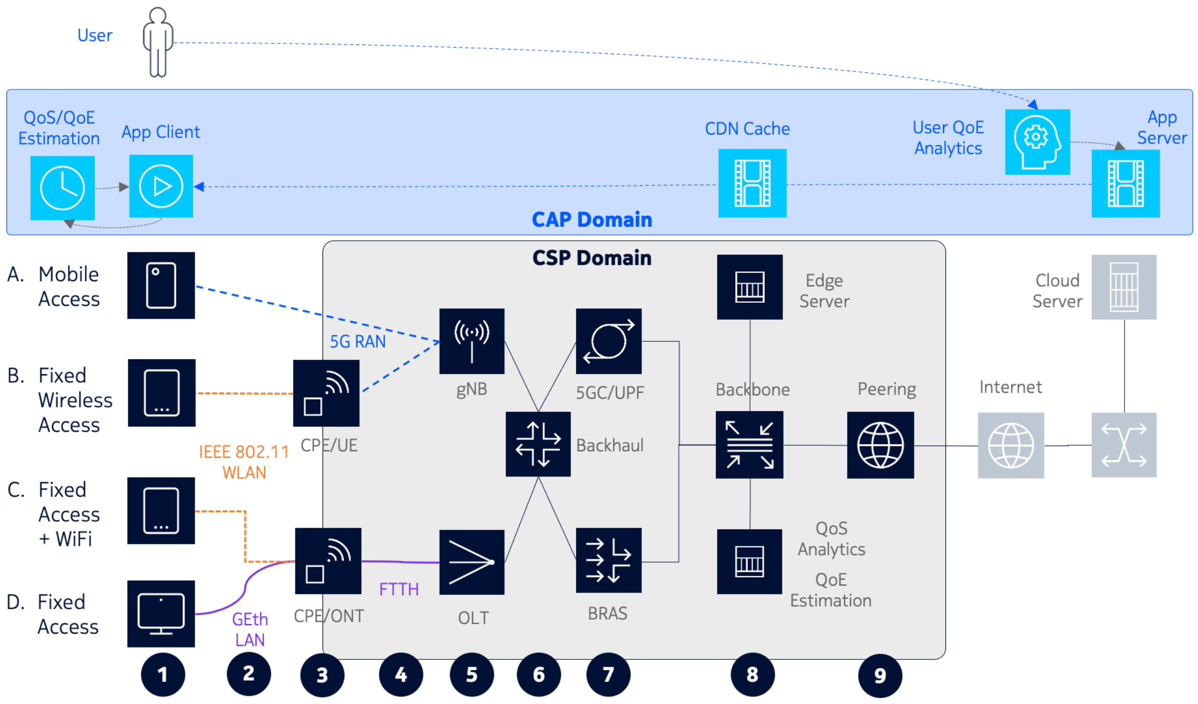

Figure 1 – Blueprint architecture of most relevant network scenarios for this white paper

In the networking part, data flows from the user device toward the broader internet through a series of interconnected components and networks:

1. Starting on the left, we have the **device** where the application runs, such as a smartphone, tablet, or computer. If directly connected to the mobile network (5G) it is referred to as User Equipment (UE).
2. This device can connect to the **Local Area Network** (LAN), which may be a wireless network following the IEEE 802.11 Wi-Fi standard or a wired network using IEEE 802.3 Ethernet.
3. Via LAN, the data travels to a piece of **equipment that connects the user's premises** to the Communication Service Provider (CSP) network, known as Customer Premises Equipment (CPE) or Residential Gateway (RG). This equipment is also referred to as User Equipment (UE) in mobile networks or an Optical Network Terminal (ONT) in a fiber (FTTH) broadband case.
4. The next segment is the **Access Network Link**, which connects the UE or CPE to the CSP's access infrastructure.
5. This infrastructure terminates at an **access network end-point**, e.g., the next-generation Node B (gNB) in 5G mobile networks or the Optical Line Terminal (OLT) in fiber-to-the-home deployments.
6. Following this, the data enters the **backhaul network**. The backhaul provides high-capacity links between access nodes and the core network. Various transport technologies are used in this segment, including Ethernet, Multi-Protocol Label Switching (MPLS), Internet Protocol (IP), microwave links, Synchronous Digital Hierarchy (SDH), or Asynchronous Transfer Mode (ATM). This part of the network typically uses private IP addressing schemes. Data traffic is encapsulated and tunneled using protocols such as GPRS Tunneling Protocol (GTP) in mobile networks or Layer 2 Tunneling Protocol (L2TP) in fixed access networks.
7. The end point of these transport tunnels is a **gateway function**. In 5G networks, this is the User Plane Function (UPF), while in fixed networks it is often the Broadband Remote Access Server (BRAS). These gateways decapsulate the tunneled traffic and forward it to the public segment of the CSP network.
8. The **public network** of the CSP is where public IP addresses are assigned. This segment often hosts edge computing capabilities, such as Content Delivery Network (CDN) caches, which may reside directly on the CSP's infrastructure or be connected through dedicated links to third-party data centers.
9. Finally, the data reaches the Internet **interconnection point** (IXP), commonly known as the peering point. This is where the CSP connects with other networks and global internet infrastructure, enabling end-to-end data communication.

From a network perspective, the bottleneck (resource limitation) is normally at the access link. Two main access technologies are considered: wireless (5G) and wired (FTTH). In addition, there may be an additional Wireless LAN step, where bottlenecks and resource limitations may appear too. The combination of these elements result in four networking scenarios:

A. Mobile Access
B. Fixed Wireless Access (FWA)
C. Fixed Access with Wireless LAN
D. Fixed Access with wired (Gigabit Ethernet) LAN

As with use cases, other network configurations are possible, but they will not be explicitly addressed in this paper. A particularly relevant one is satellite networks, i.e., low-earth orbit (LEO) or geostationary orbit (GEO) satellite-based communication. These networks may be more prone to degradations in terms of latency or throughput than their terrestrial counterparts. Additionally, complex scenarios (e.g., multi-operator networks) complicate end-to-end QoS monitoring and management.

In a well-dimensioned network, there should be no bottlenecks at the backbone or peering segments except under failure conditions, which may be beyond the control of CSPs. There may be bottlenecks in the backhaul, especially with wireless links (e.g., gNB which is connected to the core using radio access link may experience QoS drops on adverse weather conditions). However, for video on demand (VoD) and live streaming, peering arrangements between CSPs may have a significant impact on the end-user QoE during peak hours (e.g., through increased initial loading delay, or even aborted streams), as overflow peering may be used by content providers to handle traffic spikes. Massive file transfers (e.g. game updates) can also cause very significant load on the network, both in terms of direct QoE (download speed) and collateral impact on other applications. For CSPs it is important to have insights into whether peering causes user-facing QoE issues.

Proper QoS management between different users is critical in shared media, especially in the radio access link (5G). Additionally, QoS management between different traffic/application classes can further help optimize network performance (scheduler behaviors) to meet the needs of those classes. In 5G networks, two techniques can be applied for QoS management: marking QoS priorities of the different connections (bearers) using 5G QoS Identifiers (5QI, called QCI in 4G) or splitting the link resources into different network slices. Similarly, in FTTH access links it is possible to use GPON T-CONTs (Gigabit-capable Passive Optical Network Traffic Containers) to differentiate between different traffic types. In DOCSIS (Data Over Cable Service Interface Specification) networks, service groups serve a similar purpose. The different priority levels at the access link are mapped to QoS levels in the backhaul network using DSCP (for IP) or 802.1p (for Ethernet) marking (Contreras, 2024). Therefore, QoS can be managed end-to-end within the mobile or fixed network, although typically no hard Service Level Agreement (SLA) guarantees are provided end-to-end.

This QoS management has, however, several limitations:

- Due to network neutrality, it may not be possible to discriminate against QoS in a per application flow/sub-flow level, only in a per-traffic class.
- Although several traffic classes are defined in the standards, most traffic in public networks is best-effort internet traffic and it is handled without any specific QoS management policy. The most relevant counterexample is voice calls using VoIP (Voice over New Radio, VoNR, for 5G), which is prioritized over best-effort traffic and connected to the telephony network instead of the internet. Linear IPTV traffic in fixed networks follows a similar approach.
- QoS enforcement is limited to the access network, specifically from the user terminal to the UPF/BRAS gateway. A key challenge for Communication Service Providers (CSPs) in implementing uplink QoS enforcement is the "QoS trust" boundary, as user terminals and in-home equipment may not be owned or managed by the CSP.

The LAN can also apply QoS policies. However, they will only (by definition) arbitrate devices from the same CSP subscriber, there is no between-subscriber resource management in the LAN.

2 QoE Definitions and Models

In this chapter we define a common language and terminology as the basis of the white paper. It does not aim to replace existing definitions of QoS and QoE, but selects a working set of precise and clear definitions that are consistent with each other and suitable for its context.

The term QoE and its distinction from related or derived terms (user-reported QoE, modeled QoE, Key Quality Indicators (KQIs), Key Performance Indicators (KPIs)) is highlighted. First, the problem with the existing use of the term "QoE" and its interpretation in different research communities is discussed. Then the text focuses on the guidelines for the usage of the most relevant models to use depending on the specific application (e.g., planning purposes, codec design/testing, live monitoring, enforcing/managing QoE, optimization/prioritization regarding the QoE target, diagnostic/troubleshooting). Understanding and defining the service/application characteristics and its impact on users' expectations is a complex and challenging task.

2.1 Challenges with Current QoE Definitions and Usage

The rationale behind establishing a common language and terminology is to foster mutual understanding and clear communication, especially across different research communities and stakeholders (CSPs, CAPs, vendors). Especially, the terms *QoE* or *QoE metric* are sometimes misused in the literature, and it is not clear what is meant. As a consequence of such sloppy usage, the term QoE becomes meaningless and no longer conveys useful information, which indicates the need for precise definition. For example, QoE is mentioned as subjective user experience, but the reported metrics are then "only" touching application-level metrics like video bitrates or network-level metrics like throughput. The connection between those network performance, QoS metrics and QoE metrics are appropriate QoE models, which are discussed in Section 2.2.

From a telecommunication perspective, the term QoE is not as consistently defined and established as QoS. Existing network monitoring and management frameworks are implemented with respect to network performance metrics or even network-level end-to-end QoS metrics. However, there is a need to understand the quality, which is emerging from the transported traffic and the delivery of services or data to the users. Roughly, QoE measures how well the delivered data meets user expectations for service quality. It reflects the end-user's satisfaction with the quality of transmitted content and the networked services, going beyond technical metrics to include perceived experience of services.

Use case requirements are currently defined in terms of QoS instead of QoE. Current metrics in end-to-end deployments suffer from an *information mismatch between CAPs and CSPs*, with CAPs focusing on application-level user experience metrics and CSPs relying on network-centric measures. This disconnect hinders a unified view of service quality, making it difficult to align on standards and optimize QoE collaboratively. The same is also true for the associated research communities, such as multimedia (CAPs) and telecommunications (CSPs).

One can also argue that QoS optimization will result in QoE optimization and that QoS optimization will be sufficient. But two network-related metric values (KPIs, see definition below) may be acceptable individually (e.g., packet loss below 1% and throughput above 1 Mbps) while their combined impact on the user perceived quality proves unsatisfactory. Further, it remains unclear if all relevant KPIs are considered (e.g., jitter may be missing). A purely QoS-driven approach may fail to accurately reflect the relative influence of one KPI over another on the user-perceived quality. There is a trade-off between the costs and the available resources and the QoE. Thus, suitable QoE metrics are necessary to identify relevant operational points in practice, which may result in different operational points when using QoS metrics, see for example Hobfeld et al. (2017). Operational points refer to specific configurations, decisions, or actions taken within a system or process to achieve desired performance or service outcomes.

The goal of Section 2.2 is to provide definitions of QoE and the corresponding QoE metrics to be usable for CAPs and CSPs. We emphasize that those QoE metrics may include a certain degree of uncertainty to quantify QoE, e.g., arising from subjective user variability, measurement noise, contextual factors, or limitations in the underlying QoE models; but only in such a way that it is still useful in practice for CSPs and CAPs.

2.2 Definition of QoE and the QoE Hierarchy

A commonly accepted definition of QoE is provided in the Qualinet white paper on definitions of Quality of Experience (Le Callet et al., 2012), which is partly integrated in standardization of the ITU-T Recommendation P.10/G.100 on the vocabulary for performance, quality of service and quality of experience.

> **Quality of Experience (QoE)** is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in light of the user's personality and current state.

This definition of QoE indicates a clear overlap with the User Experience (UX) research community, see also Timmerer & Hoßfeld (2025). An approach to merge was done Hammer et al. (2018), by introducing the concept of Quality of User eXperience (QUX). A comprehensive overview of the field can be found in the QoE book by Möller & Raake (2014).

Measuring true QoE, which reflects a user's delight or annoyance, is challenging because it heavily depends on user mood, context, and other intangible factors. However, we assume that the degree of delight or annoyance, i.e., the QoE, that the user experiences when using a specific application and service will lead to changes in their behavior, either in a short term (e.g., continue or stop watching the content) or in the long term (e.g., canceling a subscription), which is the ultimate business motivation for monitoring and managing QoE in multimedia communication services.

Therefore, QoE models should focus exactly on monitorable and manageable parameters. The ITU-T Recommendation P.10/G.100 also discusses QoE assessment that involves measuring or estimating QoE through a structured process. This process takes into account various influencing factors, which may be controlled, measured, or reported. In general, QoE encompasses a large number of influencing factors which can be categorized into three very different, but interconnected, perspectives: Human (e.g., preferences, expectations, previous experience), System (e.g., codec, resolution, bitrate) and Context (e.g., cost, use case, interactivity level) (Reiter et al., 2014; Le Callet et al., 2012). For a comprehensive QoE assessment, relevant and most known factors should be included, whereas a limited assessment considers only a few factors, see Chapter 4 of this white paper.

QoE is, in general, a multi-dimensional and multi-factor unobservable variable in the user's mind. To consider QoE holistically, we should probably measure and quantify all its constituents (see e.g., ITU-T P.1320). In practice, however, we are not interested in considering the overall degree of delight or annoyance of the user, but rather the effect that the manipulation of a few system conditions (e.g., compression, network throughput, latency) has on such delight or annoyance. This effect is usually measured by asking users to rate the QoE on a one-dimensional scale. The most used one is the so-called Absolute Category Rating (ACR) scale, which is a 5-level Likert-type scale with the labels Excellent, Good, Fair, Poor and Bad, normally assigned to numerical values 5 to 1.

To avoid confusion with the general concept of Quality of Experience, we will name this restricted version of the QoE as *user-reported QoE*. The term "user-reported QoE" highlights that we understand the Quality of Experience from the user's perspective, based on their subjective impression (and subjective ratings) of the service or content, and instantiated in an actual rating provided for observation by others.

> We define **user-reported QoE** as the quantification of the impact of a system on user delight or annoyance, through self-report, behavioral, or psychophysiological studies. This impact can be caused by the application, network, or hardware and is moderated by the usage context.

This impact can be positive (enhancement) or negative (impairment). Impairments can degrade overall QoE and influence attitudes toward the application, service, or hardware. The system can compensate for impairments and enhance overall quality. For example, Hoßfeld et al. (2011, December) conducted a measurement study, which showed how VoIP or video-conferencing applications adapt to networking issues like packet losses. The applications were enhanced to have a positive impact on QoE, referred to as QoE provisioning in Hoßfeld et al. (2011, June), by using forward error correction on application layer or codec switches to avoid information loss and a degradation of the service.

The user-reported QoE is normally reported on a one-dimensional scale and processed statistically. The simplest processing is just taking the mean of the reported opinion scores (MOS) from several users under the same conditions (i.e., same network and service settings, context factors), but other functions (or even the complete distribution) can be used. An important property of user-reported QoE scores is that they depend on the context: the same conditions observed in two different experimental situations may lead to different score distributions (see e.g., Cavanaugh et al. (1976)). Thus, user-reported QoE can be seen as a subjective measure of expectation fulfillment. Consequently, we define QoE metrics as follows, see also ITU-T P.10/G100. A mathematical definition of those measures is provided, e.g., by Hoßfeld et al. (2016).

> A (user-reported) **QoE metric** is a quantitative measure that assesses the user-reported QoE statistically. Examples are the Mean Opinion Scores (MOS), the ratio of users rating good-or-better (GoB), the ratio of users rating poor-or-worse (PoW).

For the common 5-point absolute category (ACR) rating scale, the ratio of users rating good-or-better considers all users rating the quality as excellent (5) or good (4), and the ratio of users rating poor-or-worse considers all users rating the quality as poor (2) or bad (1).

User-reported QoE, or elements of such user-reported QoE (e.g., the audiovisual quality of a video sequence) can be predicted by **QoE models** which analyze the received media signal at different levels (bitstream vs decoded, full-reference/reduced-reference/no-reference, etc), or which rely on network and service parameters (e.g., codec, resolution, bitrate) to estimate QoE (parametric models). We call the output of these models **Modeled QoE** to make this distinction even clearer. Modeled QoE focuses on measurable aspects of perception, using models that mimic human sensory processing to evaluate how content quality and services might be perceived. This approach typically involves algorithms to estimate perceived quality in a standardized way, often without direct user feedback.

The Modeled QoE is designed to have the highest agreement possible with the user-reported QoE of a given signal, content or communication situation, typically assuming some reference observation context. As an agreement measure, Pearson or Spearman's rank correlations are commonly used. An alternative metric is the Root Mean Squared Error (RSME) as described, for example, in ITU-T Rec. P.1401 (ITU, 2020). Correlation indicates the extent to which subjective and objective scores move together, but systematic errors may still exist in absolute values; the degree of agreement quantifies these errors (e.g., mean absolute error, RMSE). From the perspective of the network, the modeled QoE provides an estimated quantification of the quality being transmitted in the network.

> The **modeled QoE** is the output of a QoE model, which is based on or predicts a user-reported QoE metric. A **QoE model** considers various input signals and parameters to predict (user-reported) QoE in terms of a (user-reported) QoE metric. We assume that modeled QoE can be instrumentally measured using a QoE model, in the absence of subjective ratings.

With an appropriate model, **Modeled QoE** is affected by the output of the application layer. For instance, the bitrate of the video signal, the buffering time before the start-up of the media playing, or the loss of information caused by video packet losses (which, with lossy transmission protocols, may generate audio-visual artifacts or, with lossless transmission protocols, lead to longer transmission times, sometimes causing stalling in the player, with a spinning indicator). Modeled QoE metrics, also called objective quality metrics, predict user-reported QoE metrics like the MOS.

Quantitative measures of these effects, which have obvious influence on the QoE, are sometimes also called QoE metrics or QoE measurements (see e.g., 3GPP QMC). However, we denote them as **Key Quality Indicators (KQIs)**, which is in line with the discussion in 3GPP TSG-SA5. KQIs are derived from measurements at the application layer but may also consider measurements at the network layer. For example, the video bitrate V measured on the application layer and the available throughput B on the network layer may indicate the reception ratio B/V, which is a KQI. A reception ratio less than one indicates that the network is not delivering the contents fast enough. However, video buffers can compensate for this. Thus, the KQI does not directly quantify the user's perceived quality. Another example of a KQI is the frequency of playback interruptions, which can be measured at application level in the video player or approximated from network measurements (Orsolic & Skorin-Kapov, 2020; Seufert et al. 2024); it indicates potentially severe QoE impairment. In general, KQIs can be measured on an application or network layer and are related to a specific service or application (like video streaming in the examples above). The KQIs quantify the performance on the application layer in a way that correlates with QoE but do not fully capture the user perception. The modeled QoE may be a function of KQIs and KPIs.

> A **Key Quality Indicator (KQI)** is a metric that directly or indirectly reflects the overall end-to-end quality of a specific service or application.

Finally, KQIs are influenced by the availability of the resources in the communication network, such as throughput, latency, or amount or distribution of packet losses, as well as on other parts of the system (e.g., computing resources). We will group these technical factors under the overall umbrella of the Quality of Service (QoS). To highlight network performance as key focus, the term Key Performance Indicator (KPI) is used, see for example 3GPP TS 28.554 (3GPP, 2024).

> A **Key Performance Indicator (KPI)** is a specific type of network layer metric used to measure and evaluate the performance of a system or service at network level. KPIs are collected from the network or calculated from network measurements.

It is the goal of CSPs and CAPs to actively analyze and assess the QoE of all users in a system. This is expressed by the term “System QoE” to indicate the assessment of user experience from a provider’s perspective for *all users* of a particular service. Monitoring provides essential information to assess the system QoE forming the foundation for effective actions like resource and network management to ensure high-quality services and enhance user experience. Typically, system QoE (for a group of users) relies on modeled QoE (for individual users) by mapping monitoring data to modeled QoE, although direct user feedback and user-reported QoE may be used for system QoE. As defined by Hoßfeld and Pérez (2024), different system QoE metrics may be relevant in practice.

> **System QoE** is defined as the assessment of the modeled or user-reported QoE of the users of a particular service or system from a provider’s perspective over a dedicated time frame. Typically, system QoE relies on modeled QoE through objectively measurable parameters and appropriate QoE models. The **expected system QoE** is a system QoE metric, quantifying the average QoE rating of an arbitrary user in the system.

Similarly, the system GoB ratio and the system PoW ratio quantify the ratio of users in that system experiencing good-or-better and poor-or-worse quality, respectively. QoE fairness quantifies the variability of the experienced QoE across the users in the system. Typically, system QoE relies on modeled QoE and appropriate QoE models. To be more precise, the modeled QoE for individual users depends on concrete KPIs, KQIs, and context factors; for a group of users with potentially *varying* KPIs, KQIs, and context factors, the resulting modeled QoE values are aggregated with appropriate system QoE metrics.

2.3 QoE Layers: Quality Models and Metrics

The relation between the different terms is summarized in Table 3, with some illustrative examples. The table structures the different terms inspired from the Open Systems Interconnection (OSI) model by the International Organization for Standardization (ISO) in ISO/IEC 7498 (ISO, 1994). This set of definitions provides a hierarchical approach to QoE at different levels of abstraction: the network layers (L1-4) include the physical (L1), link (L2), network (L3), transport layer (L4) from the ISO/OSI model; L7 is the application layer; L8 is the user layer above L7 to indicate that individual users are interacting with the application. Above L8, the system layer[5] focuses on a group of users. Accordingly, the QoE related terms (system QoE, user-reported QoE, modeled QoE, KQIs, KPIs) are structured into this hierarchical layer model.

Quality models and metrics can mainly be divided into two types, namely, subjective and instrumental/objective metrics. At different layers, different types of metrics are available for evaluating a given service. Table 3 includes example metrics for each layer.

Finally, the definitions provided in the previous section are not exhaustive: other QoE-related frameworks may provide related, but different, constructs. However, the provided levels of abstraction should be sufficient to map those frameworks into ours, so that at least it is possible to identify related categories of concepts. Table 3 also shows some of these frameworks in the appropriate layer.

Layer	Information	Example Metrics	Related Concepts and Frameworks
System Layer	**System QoE**, focusing on the QoE of a group of users of a particular service or system (based on modeled or user-reported QoE)	**System QoE Metrics**: Expected system QoE, System GoB, System PoW, QoE Fairness	Long-term behavior: customer churn, e.g., user retention rate.
Layer 8: User Layer	QoE scores based on subjective ratings (**user-reported QoE**). QoE metric = function (subjective ratings)	**QoE Metrics**: MOS, GoB, PoW Rating scales: ACR, DCR (ITU-T P.910) (regarding fidelity, timeliness, or any other aspect of the experience) Short-term behavior: Engagement Long-term behavior: User retention	QoE Questionnaire (Zhu et al., 2015) Engagement, watch time (Balachandran et al., 2012) Acceptability / annoyance (Li et al., 2019) Perceived Quality of Media Signal (PQoMS) (Koniuch et al. 2024) Overall quality, subjective MOS (ITU-T P.10/G.100)
	QoE scores based on QoE models (**modeled QoE**) mapping KQIs and KPIs to user-reported QoE metrics. QoE metric = function(KQIs, KPIs)	**Video QoE Metrics**: VMAF, UVQ, ITU-T P.1204.3/4/5 (Raake et al., 2020), Avqbits (Rao et al., 22) **Speech/Audio QoE Metrics**: ITU-T P.863 (POLQA), ITU-R BS.1387 (PEAQ)	Transmission Rating (Pérez 2023), estimated MOS, objective MOS (ITU-T P.10/G.100)

---

[5] We avoid L9 since there are different common understandings of what Layer 9 means.

Layer 7: Application Layer	Key Quality Indicators (**KQIs**): application-specific QoS parameters measured at the application (client, server), estimated from lower layer KPIs, aggregated into KQIs	Video streaming: rebuffering ratio, video quality, reception ratio, initial loading delay Speech: audio quality, coding bitrate Web browsing: page load times, speed index	QoE measurement collection (QMC) in 3GPPTS 28.404
Layer 1 – 4: Network Layers	Key Performance Indicators (**KPIs**): network-centric QoS parameters measured on physical (L1), link (L2), network (L3), transport layer (L4) and potentially aggregated into KPIs	L4: TCP goodput or throughput, and variability thereof (i.e., jitter) L3: IP Packet Loss Ratio (PLR) L2: Collision Rate, Frame Error Rate L1: Signal-to-Noise Ratio (SNR)	Network-level Quality of Service (QoS), reliability, availability (ratio of time a system or service is operational), fault tolerance, utility maximization (Nádas et al. 2021)

Table 3 – Layered approach relating the different notions of QoE, KQIs, and KPIs with the available information and corresponding metrics. Related concepts are included as reference.

- **System Layer**: This mainly uses subjective metrics such as expected system QoE, system GoB, QoE fairness, etc. A theoretical framework to handle these metrics can be found in Hoßfeld and Pérez (2024). Additionally, platform-wide statistics about long-term user behavior, such as customer churn or retention rate, is also part of the system layer.
- **Layer 8: User laye**r: In this layer both subjective and instrumental metrics are available to measure quality. Some measurement frameworks, such as the QoE questionnaire (Zhu et al., 2015), address the evaluation of the overall QoE (degree of delight or annoyance of the user). However, most of the frameworks focus on measuring the effect of the system in the user delight or annoyance, i.e., the **user-reported QoE**, either by measuring the user short-term behavior (Balachandran et al., 2012) or by user ratings (Hoßfeld et al. 2016; Li et al., 2019; Koniuch et al., 2024). User ratings are typically reported as metrics, the most frequent being the Mean Opinion Score (MOS) or the ratio of users rating good or better (GoB), see Hoßfeld et al. (2016) for an overview on user-reported QoE metrics. **Modeled QoE** metrics, also called *objective quality metrics*, predict user-reported QoE metrics like the MOS based on some measurements of the media signal.
- **Level 7: Application layer**: This layer uses instrumental methods for assessing quality. The instrumental methods focus on the modeled QoE of the different user-reported QoE from layer 8.
- **Layer 1-4: Network layers**: In these layers, quality can be measured through KPIs such as network throughput, packet loss rate, jitter, etc. These all correspond to instrumental ways of measurement. Derived KPIs can be used, e.g., when utility maximization determines the resource sharing between flows, marginal utility is a rich congestion measure (Nadas et al., 2021).

2.4 QoE to QoS Relationship

The objective of this section is to understand the relationship between a given service's impairment (QoS layer/Network Layer) and its impact on the application (KQI layer) and on the end-user's QoE.

The goal is to opportunely bridge QoE and QoS requirements to find a link between metrics and quality models and other aspects such as guidelines and rules for traffic engineering and allowable operating ranges. The QoS requirements are used to engineer the network so that services carried will meet their QoE targets.

It is worth underlining that aspects of a packet switched network are probabilistic (non-deterministic), and the associated QoS requirements are expressed accordingly, for instance in terms of the proportion of time that a requirement will be met. From this understanding, we can derive QoS requirements that correspond to particular desired QoE outcomes, or even trade off QoE with other factors such as cost. This translation is non-trivial, due to the difficulty in developing a "one size fits all" equation or algorithm that will allow one to compute or derive an individual QoS requirement from a corresponding QoE target.

The industry and academic research have been actively pursuing the topic of QoE for multiple decades, with a particular focus in the early 2000s on services such as VoIP, IPTV, and various Over-the-top (OTT) services.

This continuous effort in understanding, measuring and designing the QoE of multimedia technologies is witnessed by the definition of many standardization and implementation guidelines.

A non-exhaustive list of topics covered by standards for real-time applications is reported here:

- **Video**: Objective quality assessment models to predict the impact of audio and video media encodings and observed IP network impairments on QoE in multimedia streaming applications; measurement approaches, diagnostic analysis and KPIs/KQIs for video-based services, including video, audio quality estimation and quality integration.
- **Audio**: Like the video use case, audio-related models take into account a wide range of impairments including coded type, packet loss, delay, echo etc. Useful for transmission planning tools, to assess VoIP audio performance, establish benchmark networks for comparison, and compare design alternatives.
- **Gaming (cloud and terminal based):** Like the video-related models, aspects such as resource allocation and configuration of IP-network transmission settings such as the selection of resolution and bitrates, under the assumption that the network is prone to packet loss, throughput and latency are considered.
- **Telemetry and QoE-QoS Planning:** Measurement approaches, diagnostic analysis and KPIs/KQIs. Proactive analysis of network performance and support for customer service troubleshooting. New on-path per-packet telemetry information (piggybacking metadata on packet) to be collected and extracted from the network, and techniques being developed with real-time notification to complement ping/traceroute.

- **AR/VR/XR (including Metaverse):** The different aspects considered for this use case include human and system factors, metrics, affecting the user perceived experience of virtual reality (VR) and augmented reality (AR) services; service quality monitoring requirements; latency and synchronization aspects including motion-to-photon latency, motion-to-sound latency, A/V synchronization. In addition to this, subjective test methodologies to evaluate aspects of QoE for 360° video viewed in head-mounted display; measurement methods to spatial audio telemeeting systems have been under discussion. More recently, metaverse QoE requirements are under development.

To select the most appropriate QoE models for a given purpose, several factors should be considered to support an informed decision. In this context, although algorithms such as VMAF and UVQ are widely adopted, we will focus on reviewing relevant international standards currently in force (listed in Annex A). We will consider the following factors:

- **Primary purpose**: why the model is needed, what problem it aims to solve, and how it contributes to the understanding, estimation, or improvement of QoE in a communication or multimedia service (e.g., what is evaluated, which is the considered context, which is the specific use case).
- **Technical factors**: analyzed technical parameters that affect perceived quality, depending on model type (e.g., codec type, bitrate, display size and screen resolution, bandwidth, operating system and platform); an important consideration is the availability of these parameters at the specific measurement interfaces.
- **User factors:** considered user-related factors that explicitly or implicitly reflect real-world perception differences (e.g., user experience, cognitive load, perception variability).
- **Evaluation method**: how the accuracy of the model or method is validated (e.g., objective models, subjective testing).

We will also provide an analysis of strengths and weaknesses of the considered standards.

2.4.1 Video

	Primary Purpose	Technical Factors	User Factors	Evaluation Method
ITU-T G.1071	Opinion model for network planning of video and audio streaming applications	Audio-related (bitrate, codec, decoder packet loss concealment, number of channels, packet loss degradation), Video-related (codec, bitrate, framerate, freezing, framerate, packet-loss degradation, decoder packet loss concealment, GoP)	MOS	Subjective Testing
ITU-T G.1070	Opinion model for video-telephony applications	Speech-related (delay, coding distortion, packet-loss robustness, packet-loss rate, talker echo loudness rating), video-related (delay, codec, resolution, key-frame interval, packet-loss rate, framerate, bitrate)	MOS	Subjective Testing
ITU-T P.1203.1	Metadata- or Bitstream-based Prediction model for video quality assessment (3 models of varying complexity depending on the input)	Metadata (bitrate, resolution, framerate, video codec); framesize information (I and non-I frames); Quantization Parameter (QP)	MOS (Derived from overall HAS session MOS)	Subjective Testing
ITU-T P.1203.3	Prediction model for quality assessment of an overall HTTP-based adaptive streaming session	Stalling and quality switching information	MOS	Subjective Testing
ITU-T P.1204.3	Bitstream-based video quality prediction model for short-term video quality assessment	Metadata, QP, motion in x-direction, frame size information	MOS	Subjective Testing

ITU-T P.1204.4	Reduced reference video quality prediction model for short-term video quality assessment	Pixel information	MOS	Subjective Testing
ITU-T P.1204.5	Hybrid no-reference video quality prediction model for short-term video quality assessment	Metadata and pixel information	MOS	Subjective Testing
ITU-T P.940	Bitstream-based prediction model for video-telephony applications	Metadata (audio codec and rate, video codec, resolution…), terminal factors, packet loss, jitter, delay	MOS	Subjective Testing

Table 4 – Main characteristics of the considered video standards.

ITU-T Rec. G.1070 is a model that estimates quality for video-telephony services for videos up to a resolution of full-HD and framerate of 30fps. The model is only applicable for the scenario in which the videos are encoded with H.264. **ITU-T Rec. 1071** is an opinion model for network planning of video and audio streaming applications for services such as IPTV. It consists of two models with one focusing on the low resolution (LR) use case and the other focusing on the high resolution (HR) use case. The LR variant of the model is applicable for videos up to a resolution of HVGA (480×320) and framerate of 30fps encoded with MPEG-4 or H.264. The HR variant of the model is applicable for SD and HD videos of framerates up to 60fps encoded with H.264. The HR model is applicable for both interlaced and progressively sampled videos.

ITU-T Rec. P.1203 was the first series of models standardized for quality assessment for the HTTP-based adaptive streaming (HAS) use case. It consists of **ITU-T P.1203.1** focusing on quality assessment of short-term videos via analyzing codec metadata ("Mode 0") or bitstream parsing ("Mode 3"), **ITU-T Rec. P.1203.2** for audio quality prediction, and **ITU-T Rec. P.1203.3** for overall integral quality of a HAS session. The drawbacks of this recommendation are that it is only applicable for videos of up to a resolution of full-HD and framerate of 24fps for videos encoded with H.264 (libx264), although extensions to the model have been proposed.[6] Furthermore, the short-term video quality models were trained and validated using a "reverse-engineered" estimation of MOS from overall integral quality scores.

The **ITU-T P.1204** series of Recommendations (**P.1204.3**, **P.1204.4**, **P.1204.5**) mainly focuses on short-term video quality prediction. The original standardized models are applicable for videos up to a resolution of 4K/UHD-1 and framerate of 60fps encoded using H.264 (libx264) or H.265 (libx265) or VP9 (libvpx-vp9) codecs. In addition to short-term video quality estimation, these recommendations also provide approaches to evaluate the overall integral quality of a HAS session considering initial loading delay, stalling, and quality switches as additional input parameters. These models are applicable for both the PC/TV and

---

[6] https://github.com/Telecommunication-Telemedia-Assessment/itu-p1203-codecextension

mobile/tablet use case as target display devices. ITU-T Recs. P.1204.4 and P.1204.5 have further been extended to AV1 (libaom-av1) encoded videos with the corresponding AV1 extension of ITU-T Rec. P.1204.3 underway. These models have been shown to outperform state-of-the-art models for their respective use cases in different studies. However, these models are not applicable for quality estimation of HDR videos. A metadata-based (mode 0) variant of P.1204, which should provide a drop-in replacement for P.1203.1, is currently under development by ITU-T members.

ITU-T P.940 is a model that estimates the quality of video-telephony services based on the technical characteristics of the audio signal (codec, rate), video signal (codec, rate, resolution…) and terminal, as well as network-level measures such as packet loss, jitter and delay. It covers H.264/H.265 video coding, opus audio, and up to UHD resolution for large screens and full-HD for mobile devices.

2.4.2 Audio

The ITU-T documents presented in Table 4b are complementary. Specifically, G.114, G.107, and G.109 are technical, quantitative, and planning-focused, thus being applicable for engineering and network deployment. P.1305 and P.1310 are human-centric, subjective, and qualitative, better suited for research on user experience and advanced communication scenarios (such as virtual meetings). P.1201.2 and P.1203.2 apply to one-way video streaming specifically. We provide a detailed comparison in the following.

	Primary Purpose	Technical Factors	User Factors	Evaluation Method
ITU-T G.109	Categorizing telephone speech quality based on impairments	Network and signal impairments (e.g., echo, loss, noise, delay, clipping)	Perceived speech quality tied to satisfaction	Quality categories (e.g., Excellent, Poor) linked to G.107
ITU-T G.114	Delay planning and impact on conversational speech	One-way delay, jitter, codec buffering, echo	Delay tolerance, impact on interaction	Delay limits (≤150ms optimal, ≤400ms max)
ITU-T P.1201.2	Quality of audio streams in IPTV	Audio frame loss and burstiness (transmission artifacts), compression	–	Subjective tests, R-to-MOS mapping
ITU-T G.107	Predictive model for voice quality (R-value)	Delay, codec, jitter, packet loss, echo	MOS/user satisfaction estimated from R-value	Subjective tests, R-to-MOS mapping
ITU-T G.107.1	Computational model for planning wideband speech (50–7000 Hz) services	Talker echo, absolute delay, wideband speech coding, packet loss	Perceived conversational quality, echo perception, delay tolerance, codec-related distortion impact	Subjective validation; R-factor converted to MOS
ITU-T G.107.2	Computational model for planning fullband	Fullband coding, background noise	Impact of background noise on user comfort.	Subjective validation; R-

	speech (20–20,000 Hz) services	(send/receive), packet loss, delay	delay-related conversational impairments, packet-loss impact on perceived continuity	factor converted to MOS
ITU-T P.565	ML-based models to assess the impact of networks on speech in mobile packet-switched voice services	Jitter, packet loss, codec settings, jitter buffer behavior	Listening quality; impact of jitter, packet loss, and codec resilience on perceived quality; quality fluctuations during calls	Statistical validation; independent dataset tests
ITU-T P.1305	Impact of delay on human interaction	Conversational structure, system delays	Communication flow, turn-taking, perceived effort	Subjective testing and conversation analysis
ITU-T P.1310	Subjective quality in spatial audio meetings	Audio rendering, audio-visual alignment, speaker intelligibility	Cognitive load, task performance, spatial clarity	Subjective testing
ITU-T P.1203.2	Quality of audio streams in HAS	Audio compression artifacts	–	Subjective tests, R-to-MOS mapping

Table 5 – Main characteristics of the considered audio standards.

The E-model described in **ITU-T G.107** performs a multi-factor prediction by accounting for various impairments with a single scalar output (R-value), which can then be mapped to user satisfaction levels. This model has the drawback of being optimized mainly for 3.1 kHz narrowband telephony (although extensions exist) and its applicability depends heavily on the accuracy of the input data. **ITU-T G.107.1** focuses on conversational quality in wideband speech (50–7000 Hz), where factors such as talker echo, absolute delay, and the effects of wideband speech coding and packet loss are included because they strongly influence how users perceive conversations. **ITU-T G.107.2** extends this logic to fullband services (20–20,000 Hz), capturing higher fidelity listening expectations and explicitly considering the annoyance caused by background noise at sender and receiver, alongside the effects of delay, packet loss, and coding on perceived quality. **ITU-T G.109** translates technical performance into user-centric speech quality categories, offering a simplified framework for understanding the results of more complex models like **ITU-T G.107**. The main issue with G.109 is that it lacks its own measurement framework and it depends on ITU-T G.107 for values. Moreover, it does not provide detailed analysis or prediction tools but only a classification scheme.

ITU-T P.565 is centered on packet-switched mobile voice services and models the user's listening experience in terms of MOS-LQO (Mean Opinion Score – Listening Quality Objective), focusing only on impairments caused by the transport network—such as jitter, packet loss, codec settings, and jitter buffer behavior—while excluding device- and environment-related factors like background noise, automatic gain control, or voice enhancement.

ITU-T G.114 provides straightforward, absolute thresholds for one-way delay, making it particularly useful in network design where minimizing latency is essential, though it does not account for other transmission impairments. While these three recommendations focus on technical and predictive aspects, P.1305 and P.1310 emphasize subjective experience in telemeetings. **ITU-T P.1305** examines how delays affect the flow and naturalness of multiparty conversations, capturing subtle human interaction impacts that quantitative models often miss. It is ideal for evaluating communication systems where conversational efficiency and spontaneity are relevant issues. It does not provide quantitative thresholds for delay and its results are contextual and qualitative. These aspects make it less effective for high-level planning or large-scale evaluations.

ITU-T P.1310 goes further by focusing on spatial audio quality in immersive communication settings, assessing factors like audiovisual alignment, cognitive load, and task performance through controlled subjective testing. Although it requires more complex setups, it represents the reference for evaluating advanced telepresence or virtual meeting environments both accounting for media signal quality and communication effectiveness. However, it does not provide predictive modeling (like G.107) and consequently cannot be used for planning or simulation.

ITU-T P.1201.2 was developed for IPTV audiovisual streaming applications in the presence of packet loss. The model predicts quality on the R scale, which is then translated into MOS, and considers compression artifacts as well as transmission artifacts. Parts of the model have been re-used in **ITU-T P.1203.2,** with transmission artifacts excluded due to the model being used only in an HAS context.

In practice, these recommendations should be seen as part of a continuum from technical planning and impairment modeling to real-world user experience evaluation.

2.4.3 Gaming

ITU-T G.1032 and G.1072 have complementary roles in the assessment of gaming QoE. ITU-T G.1032 identifies technical and user-related influence factors, thus giving a relevant contribution for test design and model development. ITU-T G.1072 provides a validated, application-ready model that quantifies QoE for cloud gaming based on measurable network and encoding parameters.

	Primary Purpose	Technical Factors	User Factors	Evaluation Method
ITU-T G.1032	Identification of influence factors (user, system, context) affecting gaming QoE	Resolution, frame rate, delay, jitter, bandwidth, packet loss, device type, display size, compression strategy (e.g., GoP, bitrate)	Gaming experience, emotional state, demographic	Qualitative guide to ensure relevant variables are included in testing and modeling
ITU-T G.1072	Prediction of Mean Opinion Score (MOS) for cloud gaming QoE, based on video encoding and network parameters.	Bitrate, resolution, encoding frame rate, average fps, delay, packet loss, Group of Pictures (GoP), encoding complexity	Non-expert users: highly skilled players or VR scenarios are excluded	Subjective test data (non-expert users); uses RMSE and correlation with MOS

ITU-T P.809	Subjective evaluation methods for gaming quality	Test environment setup, platform design, network parameters	Player engagement, immersion, flow, presence, enjoyment, frustration	Passive tests, interactive gameplay tests with questionnaires

Table 6 – Main characteristics of the considered game standards.

ITU-T G.1032 provides a comprehensive conceptual framework that identifies a wide range of factors influencing QoE in gaming, including both technical and user-related factors. However, ITU-T G.1032 does not provide any computational model or prediction algorithm, nor does it quantify the relative importance of the factors it discusses. It serves more as a foundation for further study than a tool for direct application. In contrast, **ITU-T G.1072** presents a parameterized opinion model designed to predict MOS for cloud gaming services, based on specific encoding and network parameters. The model's main strength is its ability to generate objective, reproducible QoE predictions for planning and configuring cloud gaming networks. However, its limitations include a narrow scope, excluding more complex scenarios like VR gaming. Furthermore, **ITU-T P.809** provides standardized methods for subjective evaluation of gaming QoE, including guidelines for interactive tests, passive viewing studies, and validated questionnaires addressing aspects such as immersion, flow, and engagement and includes an overview of best-practice questionnaires for assessing player experience..

While ITU-T G.1032 offers flexibility and wide applicability in conceptual design, ITU-T G.1072 provides a focused, validated tool for quantitative assessment in specific cloud gaming contexts. ITU-T P.809 complements both by offering standardized user-centric test methods to assess gaming QoE in practice.

2.4.4 Telemetry and QoE-QoS Planning

The choice among the considered standards depends on context. MEF 23.2 and ITU-T Y.1541 are best suited for deterministic, SLA-driven network performance management. ITU-T GSTR-5GQoE and ITU-T P.1211 offer valuable insights for user-centric, service-specific quality evaluation, particularly in 5G and streaming. ITU-T J.1631 addresses the needs of immersive VR over cable networks. IETF IOAM stands out for real-time telemetry in modern, programmable infrastructures. Each standard brings distinct strengths, and their effective application depends on how technical objectives are aligned with the service environment and user expectations.

	Primary Purpose	Technical Factors	User Factors	Evaluation Method
ITU-T Y.1541	Standardized IP-layer performance classes for transport networks and inter-domain SLAs	Delay, delay variation, loss, availability	-	Deterministic, engineering-grade parameters tested against hypothetical paths and use cases
MEF 23.2	Definition of performance objectives for Carrier Ethernet	Bandwidth profiles, delay, jitter, loss, frame	Technical Class of Service levels mapped to application types	Quantitative thresholds per class/tier with

	services across multiple operators	coloring, Class of Service identifiers	(e.g., VoIP, video), but lacks subjective testing	normative definitions
ITU-T J.1631	Definition of network and service delivery requirements for immersive VR content via broadband cable networks	Latency, jitter, bandwidth, packet loss; it also considers edge/cloud rendering performance	-	Qualitative evaluation models
ITU-T GSTR-5GQoE	Establish QoE indicators for vertical 5G use cases (e.g., XR, remote driving)	Network latency, reliability, and 5G QoS model integration	End-user experience across various industries; it includes expected behavior and service quality	Scenario-specific KPIs and reference architectures
IETF IOAM	Real-time, in-path measurement of flow-level performance for automated and SLA-driven networks	Packet delay, jitter, loss, and path information in real time using embedded telemetry	-	Packet-level precision
ITU-T P.1211	Definition of the relationship between specific impairments (i.e., media quality levels and stalling) and perceived quality of adaptive streaming services.	Media encoding quality, resolution, bitrate shifts, stalling events	Final media quality (MOS) scores and direct assessment of perceptual impact using Shapley values	Established ITU QoE models (P.1203/P.1204)

Table 7 – Main characteristics of the considered Telemetry and QoE-QoS Planning standards.

The six documents serve distinct purposes, with strengths and limitations depending on their intended use.

ITU-T GSTR-5GQoE takes a forward-looking, application-centric approach by focusing on QoE in 5G specific applications like tele-operated driving and XR services. Its value lies in emphasizing user-centric indicators beyond just QoS, yet it remains an informative document rather than a formal standard and is applicable mainly to very specific or emerging scenarios. In contrast, **ITU-T Y.1541** is highly effective in IP backbone and interconnection settings, offering deterministic thresholds for delay, jitter, and packet loss, which are critical for transport-layer service guarantees. Still, it overlooks user-level QoE and lacks direct application mapping.

ITU-T P.1211 is suited for analyzing adaptive streaming quality, using cooperative game theory to identify which impairments most significantly degrade user experience. While powerful in streaming diagnostics, its scope is narrow, excluding other media or service types, and it depends on the fidelity of underlying quality models.

ITU-T J.1631 addresses the specific challenge of delivering cloud-rendered VR over cable networks by laying out functional and performance needs for immersive video applications. Though it bridges technical and user considerations well, it is confined to broadband cable infrastructure and early-stage VR ecosystems.

MEF 23.2 offers a robust and structured framework for managing Carrier Ethernet services through defined Classes of Service and performance tiers, making it highly suitable for service providers handling multi-operator SLAs. However, it is tightly bound to Ethernet environments and requires significant coordination across domains, limiting its flexibility outside such contexts.

Finally, the **IETF IOAM** framework shows real-time, programmable telemetry capabilities, offering granular, per-packet insights into live network conditions. It is best suited for advanced, autonomous networks but still requires considerable complexity in deployment.

In conclusion, the best suited standard depends on the specific scenario and requirements. ITU-T GSTR-5GQoE and ITU-T P.1211 are useful in service-specific QoE analysis, ITU-T Y.1541 and MEF 23.2 for structured network performance, ITU-T J.1631 in immersive media delivery, and IETF IOAM in modern, telemetry-driven network operations.

2.4.5 Metaverse AR/VR/XR

The available standards for the Metaverse are aimed at different goals within the XR/VR/AR landscape. Specifically, 3GPP reports are focused on system and media delivery design and, similarly, Y.3109 addresses network-level QoS frameworks. In contrast, the ITU-T G-series is more concentrated on user experience and perceptual quality. In addition, P-series standards target audio-based evaluations of telemeetings. We provide a detailed comparison in the following.

	Primary Purpose	Technical Factors	User Factors	Evaluation Method
ITU-T P.1310	Test of spatial audio in telepresence and videoconferencing	Microphone capture, beamforming, spatial codecs	Speaker localization, audibility in multiparty calls	Subjective tests (e.g., conversation tasks, multiscale ratings)
ITU-T G.1035	Analysis of VR user perception, simulator sickness, and immersion	Device latency, rendering delays, simulator-sickness sources (e.g., vergence conflict)	Demographics, immersion, fatigue, motion sickness	Descriptive, qualitative QoE evaluation
ITU-T Y.3109	Analysis of network-level QoS assurance for VR	MEC functions, 5G slicing, QoS flows, latency/ bandwidth profiles	–	Framework-based, QoS model mapping

ITU-T G.1036	Definition of QoE metrics for AR applications	Tracking accuracy, display optics, networking delays	Visual comfort, task context, environmental lighting	Analytical model proposal
ITU-T P.1320 ITU-T P.1321	Evaluation of collaborative XR meetings (enterprise, tele-ops)	Rendering pipeline, network and compression stack	Presence, co-presence, body-language cues, cognitive load	Mixed-mode: lab vs. crowdsourcing considering several protocols (e.g., ACR, DSIS)
ITU-T P.812	Evaluation of interactive VR	Test design, system setup, interaction tasks	Presence, comfort, fatigue, usability	Subjective methods (qualitative and quantitative)

Table 8: Main characteristics of the considered Metaverse standards.

The documents collectively address different layers of the XR/VR/AR multimedia systems, each suited to specific roles and priorities.

ITU-T G.1035 is valuable for its exploration of psychophysical impacts (e.g., cybersickness, presence, eye strain) and user diversity (e.g., age, experience, emotional state). This document is suggested for design decisions focused on usability and comfort. The performed analysis is mainly qualitative, and it does not contain built-in test methods or direct measurement guidance. It also lacks system/network layer integration.

Similarly, **ITU-T G.1036** extends this user-focused approach into the AR domain, outlining how hardware, context, and perception shape quality of experience, though it leans more toward conceptual modeling than direct measurement. It suggests a QoE framework for emerging AR use cases, but it lacks established metrics and it does not include practical test methodologies.

For telepresence systems emphasizing audio quality and localization, **ITU-T P.1310** provides detailed methods for spatial audio testing, though it does not address broader XR interaction. **ITU-T P.1320** is the most appropriate when user experience is the priority as it considers social presence, system context, and cognitive load, along with test setups. It is highly meeting-specific and thus it is not suited for solo or non-telepresence XR scenarios. Moreover, it may require controlled environments or high-end setups to implement. **ITU-T P.1321** describes how to design, perform and analyze subjective experiments to assess the QoE constituents defined in P.1320.

In addition, **ITU-T P.812** provides principles for subjective test methods tailored to interactive VR applications. It offers guidance on designing and conducting controlled experiments, including minimal requirements for test environments, equipment, and participant profiles. Importantly, it emphasizes QoE aspects such as presence, immersion, cybersickness, motivation, engagement, and social acceptability, and outlines best-practice questionnaires and measurement approaches.

ITU-T Y.3109 is targeted at network architects and telecom engineers, focusing on network-level QoS parameters for VR/MEC, offering a technical framework. It is not user-focused, so it ignores perceptual and human experience aspects.

3 Industry Alignment on Common QoE Metrics, Models & QoE Management

QoE can only be delivered and optimized from an end-to-end perspective. As we are looking to continue to deliver media the most effectively, an end-to-end QoE management approach can prove to be a win-win for the ecosystem. This chapter will first introduce QoE management concepts, then identify current issues and gaps, and describe the possible benefits of a collaboration. Finally, it will highlight some recent initiatives for enabling CSP and CAP data exchange.

3.1 Introduction to QoE Management

Figure 2 illustrates how each segment of an end-to-end delivery chain contributes to specific QoS impairments, with control over these impairments depending on segment ownership.

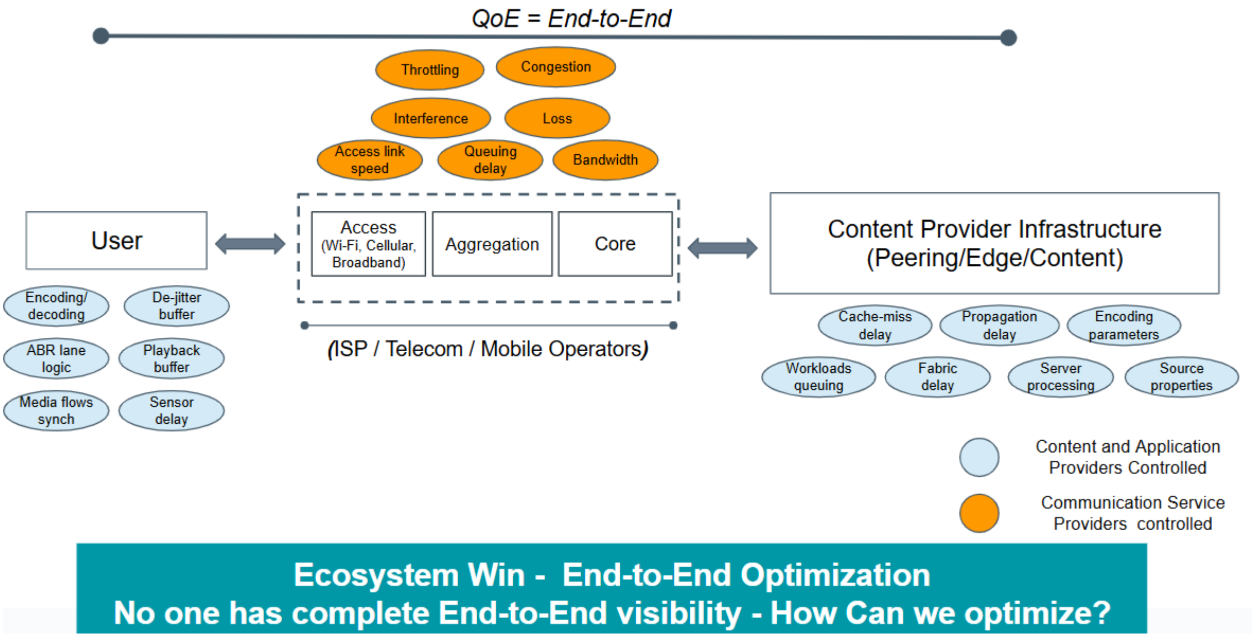

Figure 2 QoS metrics and other factors that are impacting QoE.

- **Left Section (Multimedia Client):** Blue bubbles represent factors typically managed by Content Application Providers (CAPs), such as application and user software stack. Some factors can also be attributed to the devices or operating system.
- **Middle Section (ISP/Telecom/Mobile Operators):** This section signifies a complex network (e.g., home, Wi-Fi, cellular, broadband). Orange bubbles indicate factors typically controlled by Communication Service Providers (CSPs).
- **Right Section (Content Provider Infrastructure):** Factors attributed to service and streaming content from content data networks are listed in blue bubbles, which are also controlled by CAPs.

In practice, the situation could even be more complex, e.g. involving several CSPs in different segments of the chain. However, even in this simplified scenario, no one has complete end-to-end (E2E) visibility. So the question is, how can one optimize for QoE, understanding QoE requires an end-to-end approach?

3.2 QoE Management – Issues and Gaps

A number of issues and gaps have been identified from the CAP's perspective as well as CSP's perspective in Tables 9 and 10 below.[7]

CAPs Challenges	End-User Impact	Opportunity
Difficulty to determine CSPs link capacity and network congestion	Increases probability of stalls due to inadequate network link capacity detection	Improve bandwidth estimator convergence time and precision
ABR (Adaptive Bitrate) track bitrate/resolution selection Codec selection	Non-optimal video quality	Improve ABR track selection
Difficulty to predict network dynamics and network capacity variability	Increases probability of stalls due to buffer underflow in client devices app Increase data consumption due to retransmissions	Reduce mismatch in link speed and CAPs bandwidth estimates with more accurate tracking of network link capacity Improve de-jitter/playback buffer adjustment
Understanding subscriber data plan/budget	Increase data subscriber's plan limit increase of buffering/prefetching – danger of wastage	Optimize video delivery efficiency overall Control egress data volume

Table 9 – CAPs challenges.

CSPs Challenges	Impact	Opportunity
Diversity in networking requirements per traffic class	Network policy design becomes complex. QoS fairness hard to map with QoE fairness	QoS and/or QoE targets exchange Service adaptivity strategy exchange
Accurate and deterministic identification of applications (due to encryption, VPN, mixed traffic classes)	Application of appropriate traffic management policy depends on accurate traffic classification	Traffic class information exchange (in-band or out-of-band signaling)
Absence of QoE information	3GPP QCI/5QI-QoS tables provide priority level, not QoE targets. Reliance on coarse QoE estimation when monitoring effect of traffic management policy	Parametric model to convert QoS to QoE
Absence of real-time QoE feedback	Detecting actions to improve user experience not captured adequately by network performance KPIs	Real-time QoE information exchange

---

[7] References VQEG 5G KPI workshop Austria, August 2024

Lack of visibility of traffic type impacting scheduling and buffer optimization	Application of appropriate traffic management policy depends on QoS parameters of each traffic flow	Real-time information of flow properties

Table 10 – CSP challenges.

3.3 QoE Management Benefits (Two Layers of Benefits)

From a systems design perspective, CAPs have control across most modules of the QoE optimization pipeline, spanning from content acquisition, hosting, processing and distribution, to content-delivery protocols and services, and the clients to render the media. Nonetheless, the E2E (User-CAP-User) path includes the networking service which provides the interconnection between the CDN/backend infrastructure and the multimedia client/end-user – typically managed by CSPs and network operators. While most of the networking, communication and multimedia-delivery mechanisms are standardized and industry-converged, there still exists a significant variety in the available types of multimedia content – each with their own, often competing, networking requirements. Concurrently, QoE objectives or prioritization of the different factors (video quality, stalls, initial start-up) may differ across CAPs, while network management strategies and policies across CSPs come in multiple forms, further increasing the complexity of the QoE optimization space.

In parallel, in the absence of a framework to facilitate content-type signaling and QoE-information availability, network operations are lacking the required context and feedback that could enable QoE-aware network management. The first step towards QoE-awareness is the identification of meaningful QoE metrics that can serve as input to network operations, troubleshooting, planning and policy design, along with QoE-QoS correlation to inform parametric models.

QoE-awareness has the potential to yield a series of benefits for both CAPs and CSPs, and ultimately the end-user, who is concurrently the focus of both network-service and multimedia optimization. Therefore, CAPs and CSPs have a joint interest in serving their common user-base. To further motivate QoE-aware management, in the following section we expand on the two main benefits, namely visibility and efficiency.

3.3.1 Visibility and Increase of Transparency

One of the gaps identified that would bring value to QoE aware management is having visibility of what is happening at the different layers (network layers 1-4 up to user layer 8) and what could be happening if changes are implemented if controllable network parameters are changed. The key benefits of QoE-aware management is visibility and increase transparency:

- Visibility via means of QoE information exchange **from CAPs to CSPs** allows CSPs to enhance network resource allocation and network performance monitoring.
- Visibility via means of QoS information exchange **from CSPs to CAPs** allows CAPs to make adjustments towards optimal QoE.
- Transparency via **traffic type signaling from CAPs** allows CSPs to apply appropriate management policies, free of traffic classification errors.

Potential QoE management actions include:

- Visibility via means of **QoE information exchange from CAPs to CSPs** allows CSPs to monitor the health of the network and predict the impact of network configuration on both short- and long-term timeframes. On short-term timeframes, actions can be taken by CSPs to adjust resource availability and help increase QoE, while long-term QoE trending can be used to inform refinement of management policies and network planning.

- CAPs **understand the effect of service performance decisions** on network usage, balance optimal codec/encoding rates together with resulting delay metrics, and to help understand why users have issues. Also, with better understanding of current and historical network health, it is easier to predict possible issues and remedy them with appropriate action even before issues occur.

3.3.2 Delivery Efficiency, Fairness and Optimization (Take some Actions)

CSPs aim to deliver traffic efficiently, while providing fair use of network resources to all users. Efficiency comes from the optimal balance between delivering all useful traffic with the least use of resources. Resources include network equipment, physical and radio links, spectrum, and operational costs. Useful traffic includes all data that users and applications consume without waste and unnecessary re-transmissions. CAPs delivery efficiency can be measured in terms of Content Distribution Network usage, compute power, encoding resources and volume of data transmitted.

To achieve efficiency, multiple stakeholders' success metrics and actionable measures can be summarized as follows:

- Users should merely request data that they will actually consume (e.g., download movies they will indeed watch).
- Applications should fetch data that is reasonably expected to be consumed (e.g., avoid loading too many advertisements that will not be shown, or buffering and pre-fetching too much of video content).
- Reliable transport protocols should be tuned to the characteristics of the path to avoid unnecessary retransmissions while fully utilizing available bandwidth.

Next, fairness to all users is of paramount importance. Traditionally, fairness was considered to be a very simple concept, namely bandwidth fairness, where all competing network flows converge to equal share of bandwidth. This problem has been solved within transport protocols via congestion control algorithms. However, the network traffic mix has become far too rich for bandwidth fairness to fully satisfy the needs. This is because bandwidth usage (or data rate) does not translate directly into QoE. The two main reasons are that application requirements for bandwidth are no longer equal and latency requirements vary significantly across applications. Additionally, different applications may have different tolerances to the variability of both bandwidth and latency over time. QoE fairness is what needs to be considered for today's traffic mix. Achieving the fair share of QoE should supersede achieving the equal share of bandwidth.

Take, as an example, the video stream and web page downloads of the news mobile app on the 10 Mbps cellular network bottleneck link. Currently the two applications' flows will keep trying to converge on 5 Mbps each but likely never really achieve it, if they are the only users and their radio characteristics are the same. However for optimal QoE for each, a video stream likely needs less than 2 Mbps of relatively stable throughput for a great video quality on a phone screen, with initial and later possibly occasional bursts of more speed for buffering and prefetching, effectively making a trade-off between data rate and time. The news app will benefit significantly from higher speed bursts (8-10 Mbps) periodically as the user browses.

Then, if a video conferencing application joins the network and attempts to establish a constant bit rate traffic of about 500 kbps, it may occasionally run into congestion leading to QoE issues, and the network cannot help no matter how relatively small the requirements are.

How could this QoE fairness be accomplished by the network? The only way is for the network to be aware of the traffic mix and its needs, so it can anticipate the correct traffic volume and profile over time. This can not only enable QoE fairness, but also better network planning, which would help some types of networks to avoid over-engineering and over-provisioning for (inappropriate) peak usage.

While network equipment vendors claim that they can detect and classify almost all traffic inside the network using packet header inspection, there are problems with this approach. First, the sheer scale of traffic makes this an expensive proposition. Sampling is sufficient for statistical analysis and long-term planning, and also high-level management such as reaction to diurnal changes and planned events. However, sampling cannot support pro-active traffic management or QoE fairness. Second, this type of detection is inference or estimation, which is difficult to maintain over time as new apps and traffic types emerge frequently. Third, a further complicating factor is that today's applications have mixed content, and while inference can often be very accurate in detecting the app, it may have difficulty distinguishing multimedia traffic from a web page download.

Hence, a principle emerges that under current conditions, the network cannot assign or impact the priority given to the network data flow because it does not fully understand the traffic requirements within that flow. So how can a network become aware of the traffic and its requirements accurately? And can that be achieved while preserving user's privacy needs and content owner's security requirements, which are both facilitated today via increasing layers of encryption and use of VPNs? Encryption of URLs, Server Name Identification (SNI), HTTP headers, and DNS traffic makes it even more difficult to implement QoE-based traffic management fairness, resources allocation and delivery efficiency.

The truly accurate approach is for an information exchange to occur between CAPs and CSPs. This can be some form of signaling or notifications, with the following aspects and questions to consider:

1. Who is signaling and what? An app or a CAP could signal requirements and QoE, while a network can signal a form of health status, congestion or available bandwidth and latency. L4S (De Schepper et al., 2023) is a successful, and simple, example of the latter[8,9,10]. Additionally, IETF has recently established SCONE as an active effort to develop a protocol for CSPs and CAPs to exchange information. SCONE will be covered in detail in the next section.

2. Frequency of signaling. CAPs may provide periodic or event-driven signals throughout the user session, while applications may do the same. In some cases it may be more important to signal transient events so that overreacting is avoided. Different elements can be signaled with different frequencies as well. QoS-to-QoE description parameters

---

[8] https://www.ericsson.com/en/reports-and-papers/white-papers/enabling-time-critical-applications-over-5g-with-rate-adaptation

[9] https://onestore.nokia.com/asset/213410

[10] https://datatracker.ietf.org/meeting/123/materials/slides-123-tsvwg-l4s-with-apple-devices-00

can be transmitted less frequently and still be relevant, compared to direct QoE estimates that have to be updated frequently to be useful for optimization algorithms.

3. Gradual deployment. It should not be expected that all CAPs or all networks have information exchange enabled, so impact on all flows should be considered.
4. Role of operating systems (OS). The most effective approach may be that the OS and network exchange information due to the vast numbers of CSPs. An OS-based approach is more scalable, and the OS could enforce “honesty” in advertising network requirements. OS can also provide a homogenous interface to hardware accelerators for media encoding/decoding, that in turn can provide the best assessment of media quality. One way of realizing this could be to add a trusted middleware layer, e.g. by extending Pauly et al. (2025), to harmonize between OS and applications. Sometimes, however, OS implementation roadmaps may block or delay some features. Preference for OS involvement may vary depending on content and direction of info exchange.

There are implications of establishing information exchange as described above, which cross multiple spheres, which are out of scope of this white paper, such as:

1. Business: Protecting each stakeholder's business from harm via either revealing proprietary information, abuse of revealed information, or negative customer experience impact.
2. Incentives: The system must align incentives so that participants cannot benefit from misreporting or cheating, ensuring cooperation is honest and verifiable. Overall, the incentive structure should reward efficiency and fairness, while discouraging practices that degrade network performance.
3. Regulatory: Should some aspects of information exchange be mandatory (QoE KPIs or others) or network architecture be more open?
4. Privacy: Mainly thought of user privacy, how much information does not harm?
5. Fair treatment of users: No system-level action or optimization may result in unfair treatment of some customers with respect to others (e.g. by profiling their sensitivity to some QoE degradations).

3.3.3 Existing CAP–CSP Information Exchange Protocols

There are several recent or in-progress initiatives to exchange QoS or QoE-related information between CAPs and CSPs. The scope of this whitepaper explicitly excludes considering which protocol should be used between CAPs and CSPs to exchange QoE metrics. However, it is important to mention the most significant ones, as the QoE metrics discussed in this white paper may have future influence on them.

The IETF has recently chartered the **Standard Communication with Network Elements (SCONE)**[11] working group to define a protocol-based mechanism that could allow applications to receive notifications containing throughput advice for both upstream and downstream directions. The aim is to enable on-path network elements to advise QUIC endpoints about maximum sustainable throughput or rate limits, allowing applications like video streaming to adapt their send rate based on network provider advice over longer

---

[11] https://datatracker.ietf.org/group/scone/about/

timescales. Its goal is to enhance user experience without impacting network efficiency, privacy and net neutrality at the time of addressing the increasing demands of video traffic.

The throughput advice as provided by the network can enable applications, such as video streaming services, to self-adapt their data rates and behaviors in alignment with current network conditions and policies, thereby enhancing user experience and optimizing network utilization. The protocol solution in SCONE (Tomar et al., 2025) is work in progress as working group adopted document[12] with the expectation of being complete during the first half of 2026.

L4S (Low Latency, Low Loss, and Scalable Throughput) is an IETF-standardized approach that enables the network to signal incipient congestion early and precisely, so applications can adapt their sending behavior before queues build and loss occurs. It builds on Explicit Congestion Notification (ECN) with finer-grained feedback (Accurate ECN/AccECN) and introduces Dual Queue Coupled Active Queue Management (AQM), allowing flows that prioritize low latency to coexist fairly with throughput-seeking traffic. Explicit Congestion Notification for L4S (De Schepper, 2023) allows CSP network nodes to explicitly signal congestion before a large queue builds up. More Accurate Explicit Congestion Notification (AccECN, Briscoe et al. (2025)) enables a CAP server to adapt its sending behaviour prior to any packet loss. Dual Queue Coupled AQM for L4S (De Schepper & White, 2023) adds sharing of available bandwidth between applications optimising throughput and those preferring to avoid congestion and packet loss. By shifting from reactive, loss-driven congestion control to proactive, mark-driven signaling, L4S reduces queueing delay, stabilizes throughput, and minimizes packet loss—improving the transport conditions that application-layer adaptation relies on—while preserving privacy because no payload inspection is required.

As a QoE management strategy, L4S provides a continuous, in-band feedback loop between network and applications: frequent ECN marks communicate congestion levels, senders adjust rates accordingly, and dual-queue AQM maintains fair, low-latency service across heterogeneous flows. This early, granular signaling directly improves KPIs such as latency, jitter, and packet loss, which translates into better KQIs (e.g., fewer stalls for streaming, lower interaction delay for real-time services) and ultimately higher modeled and user-reported QoE. L4S can operate alongside broader QoS–QoE frameworks by supplying high-frequency transport-level indicators, complementing lower-frequency QoS-to-QoE descriptors and direct QoE estimates from applications, and supporting gradual deployment while maintaining fair sharing with non-L4S traffic.

An additional standardized framework is provided by the **Application-Layer Traffic Optimization (ALTO) protocol** which exposes abstracted network information to applications, enabling CAPs and CSPs to exchange guidance that can improve traffic distribution and efficiency without revealing sensitive internal topology. Defined in IETF RFC 7285 (Kiesel et al., 2014) and extended by RFC 9439 (Wu et al., 2023) for reporting performance metrics, ALTO allows a network operator to publish cost maps, endpoint properties, and path preferences that reflect factors such as routing distances, capacity, or dynamic load. This information can help applications to select more efficient server endpoints or content replicas, leading to lower latency and improved user experience.

---

[12] https://datatracker.ietf.org/doc/draft-ietf-scone-protocol/

As coordination mechanism, ALTO enables a structured, off-path feedback loop between CAPs and CSPs: the network provides cost information and real-time updates through incremental changesets, applications adapt their content placement or request routing, and operators can tune ALTO services to reflect evolving performance conditions. The fine-grained but privacy-preserving updates introduced by ALTO allow near-real-time responsiveness without imposing high operational overhead, making ALTO suitable for dynamic environments such as CDNs, cloud gaming platforms, or multi-access edge deployments. By aligning content distribution strategies with network conditions, ALTO can complement QoS–QoE frameworks by supplying application-layer guidance that translates into better traffic locality, more stable throughput, and ultimately improved end-user QoE across heterogeneous service environments.

Common Media Client Data (CMCD)[13] is a standardized protocol (from CTA WAVE) that enables media players to share playback data with CDNs for optimizing content delivery and improving user experience. By providing the transmission of this detailed data and information, CMCD-enabled video streaming services can facilitate better troubleshooting, optimization, and dynamic delivery adjustments by CDNs. With CMCD, media clients can send key-value pairs of data to CDNs, providing valuable insights into the streaming session. This data includes information such as encoded bitrate, buffer length, content ID, measured throughput, session ID, playback rate, and more. Version 1 of the standard has been published in 2020.[14] CMCD has been adopted by many CAPs in the streaming ecosystem, and in June 2024, Apple added opt-in CMCD support to AVPlayer, thereby increasing the possible adoption of the standard.[15] However, for privacy reasons, Apple is not supporting all keys, including a “content ID.” Currently, a version 2 of the standard is under development. It focuses on enabling different reporting modes (e.g., a mode where only video playback state changes are reported) and transmission modes (e.g., sending data to a third party out-of-band), as well as adding more fine-grained QoE-relevant data, such as information on stalling events. Note that strictly speaking, CMCD is not aimed at CSPs and thus not a CAP-CSP information sharing protocol. However, its purpose for use by CDNs to understand and improve streaming performance is still relevant in the context of this white paper.

The **3GPP Quality of Experience (QoE) Measurement Collection (QMC)** function enables network operators to gather and manage QoE data from user equipment applications, facilitating the assessment and enhancement of service quality across various network segments. Moreover, 3GPP has introduced the notion of Protocol Data Unit (PDU) Set to differentiate a group of PDUs (belonging to the same or different sessions) collectively for certain purposes such as packet discarding. This proposition was motivated by media use cases such as XR, with the idea of signaling in advance which PDUs are more relevant for keeping users’ QoE expectation.

GSMA TS.43 Service Entitlement Configuration is a technical standard that enables mobile devices to automatically verify if they are eligible for specific network services, e.g. VoLTE, VoWiFi, 5G voice, eSIM activation, and SMS over IP, using secure authentication. It facilitates direct communication between the device and the operator's network, streamlining

---

[13] https://github.com/cta-wave/common-media-client-data

[14] https://cdn.cta.tech/cta/media/media/resources/standards/pdfs/cta-5004-final.pdf

[15] https://developer.apple.com/streaming/Whats-new-HLS.pdf

activation for various features. Its Service Entitlement framework could provide a coarse-grained indicator of an entitlement for certain video quality (e.g., Standard Definition or SD) or bitrate limit. It already provides the network metering status, which can be used by CAPs to engage a conservative ABR logic, limit the top ABR track level, etc.

TS.43 is essential for ensuring that devices are properly configured for modern network services without requiring manual user input, providing a seamless user experience, and may present a low-effort vehicle for deployment of CSP to CAP sharing of certain types of information. Access to this framework is done through the mobile OS API.

Custom APIs are another possibility to directly exchange information between CSPs and CAPs. These out-of-band APIs may be quicker to implement but face a challenge of being too specific to any CSP or CAP.

An example is the Google Mobile Data Plan Sharing API, which allows mobile operators to share user data plan status and offer information with Google applications via the Google Traffic Application Function (GTAF). It facilitates data-aware, optimized streaming by allowing CAPs streaming systems to adapt behavior, such as deferring downloads when a user is low on data. The key aspect of this API is a service that interacts with CSPs to use the data plan information updates to manage streaming. It requires CSPs to authenticate with Google Cloud OAuth2, supporting CPID (Carrier Plan Identifier) for user identification.

4 Proposed CSPs-CAPs Metrics, Models and QoE Management Tools

This chapter represents both the main content of this white paper and the foundation for the next phase of work in aligning industry standards for metrics, KPIs, and QoE management tools. First, we give a high-level overview of how QoE-aware networking may improve end-user experience, and what actionable controls exist. We then list requirements for the QoE metrics to be shared. We then concretely specify how a QoE metrics framework based on a shared state table may work. Next, the chapter presents a description of three use cases we have identified – short-form video, long-form video, and interactive applications. Finally, privacy aspects are discussed.

4.1 QoE-Aware Networking and Actionable Controls

With collaboration between CAPs and CSPs, the network can plan, take action, and give suggestions based on each specific service and the actual service mix. For clarity, the term service here represents an end-user or consumer-based services such as video streaming or video conference. QoE estimates from the user equipment alone can be used to fine-tune scheduling and send recommendations back to the content server to tune media delivery settings.

The end goal here is to balance the QoE levels for the users in a network segment/cell, using both network state (e.g., congestion information, subscriber entitlement) already known by the network in all involved nodes in conjunction with the instantaneous QoE estimates from all devices and services. With a holistic view of the group of users loading each cell, or network node, the network operator (CSP) can give informed guidance to the services to minimize latency and data loss, according to what the services need. Another option is that, in addition to modeled QoE scores, the services provide the network with simplified QoS-to-QoE model coefficients that can be used to better predict the outcome of any suggestion or change the network will have. With this information, the network can balance data rate and latency per user on a per-service basis, prioritizing the more latency sensitive services until the point where it starts hurting the data rate limited services. We will discuss this resource–quality tradeoff in the next section.

The network also needs to know the adaptability strategies of the service, so that they both don't start working against each other. For instance, in-band signal strategies such as L4S already provide feedback to locally adapt the application throughput to the available channel capacity. QoE-aware networking could, on top of that, provide guidance on longer-term resource allocation decisions or cross-service QoE fairness assessment.

Potential actionable controls exist to:

- Assign appropriate bandwidth to achieve the best fairness among competing services, reduce wastage (network, CDN)
- Fine-tune network scheduling and select media quality levels for pre-encoded material based on QoE estimates from client apps
- Optimize ABR track selection based on QoE congestion information already known by the network
- Perform resource planning with simplified QoS-to-QoE model coefficients provided by services to the network.

The remainder of this chapter will provide more detail on those use-cases and actionable controls.

4.2 Characterization and Requirements for QoE Metrics to be accepted and understood by CAPs and CSPs

For QoE metrics to be effectively shared between CAPs and CSPs, several key requirements must be met. These are outlined in this section.

First, **both parties must agree on common metrics using well-established protocols.** These metrics should be easily understood – ideally standardized or non-proprietary –, capturing QoE for users/viewers of a given service at the session level or other units of time (e.g., video segments), using a standardized numerical scale (such as 0-100 or 1-5).

Importantly, **these metrics must demonstrate correlation with subjective opinions**, showing little bias and well-documented accuracy through measures like RMSE, Pearson correlation, and estimates of standard deviation.

The metrics can be implemented in two primary ways: either **calculated independently by the CSP** using CAP-provided data or **supplied directly by CAPs** through agreed-upon interfaces, whether in-band or out-of-band. One example for such data is the CMCD standard. There currently is a discussion about adding video quality assessment (VQA) metrics to the reported data[16], including proposals on how to map them to a common

---

[16] https://github.com/cta-wave/common-media-client-data/issues/131

scale (0-100) – but it is also conceivable that QoE-relevant metrics can be derived from CMCD data by the receiving party.

The metrics may operate across multiple layers, potentially combining user-reported QoE metrics, modeled QoE metrics, and various KQIs and KPIs from lower layers. The goal is to approximate the end-user QoE experience as closely as possible in real-time. Different metrics may measure various QoE attributes, with their relevance varying by application. For instance, while video quality in terms of fidelity is important (as measured by metrics like VMAF[17]), stalling (e.g., as modeled in ITU-T Rec. P.1203, developed for long-form video) may be more critical to monitor and expose in the context of network management. In particular, solutions like P.1203 provide a framework for understanding video streaming quality, as they incorporate aspects of shorter-term video quality, stalling, and initial loading delay. While VMAF provides valuable insights about video encoding quality, it may not be sufficient from a CSP's perspective, and can be costly for smaller CAPs to implement. Alternative metrics such as YouTube UVQ[18], ITU-T Rec. P1204.3, and P.1204.4 may be considered in this context, along with the QoE timeliness features mentioned before (buffering, loading time, and stall counts).

Context plays a critical role in metric interpretation. This encompasses the user context – including the users' expectations, which may differ depending on device type and connection (Krishnan & Sitaraman, 2012) – and the system context, where network attributes like traffic classes and service hierarchies are considered, distinguishing between content types like reels/feeds, VoD, or live streaming.

The metric exchange should ultimately optimize QoE/QoS trade-offs, with CSPs able to evaluate potential quality improvements for specific assets or services or making it possible to mitigate quality degradations. Fairness should be another important principle in these optimization goals. Underlying this exchange is the concept of a resource–quality tradeoff, whereby it can be generally said that improved resources (e.g., higher bitrate, lower latency) result in higher quality (e.g., overall QoE, video fidelity, or responsiveness), as illustrated in Figure 3.

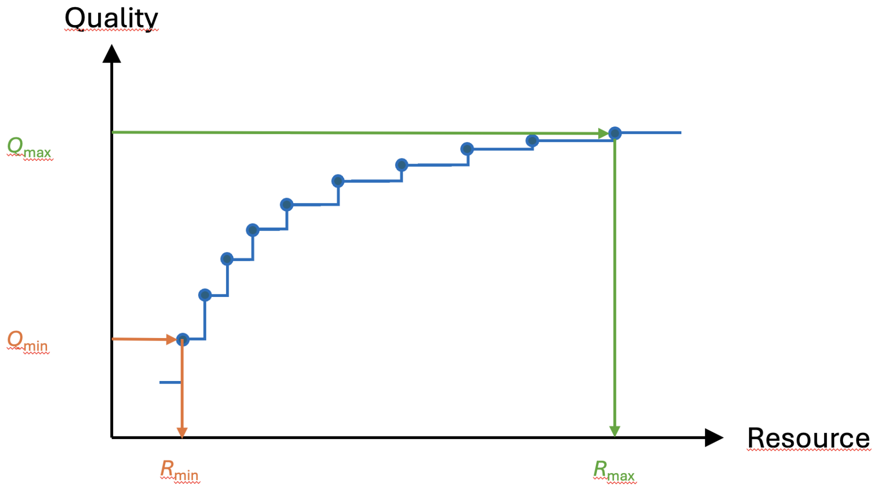

Figure 3 – Typical resource–quality tradeoff curve.

---

[17] https://github.com/Netflix/vmaf
[18] https://research.google/blog/uvq-measuring-youtubes-perceptual-video-quality/

This curve may be then instantiated on a per-service basis, where the tradeoffs depend on the specifics of the application. For instance, for short-form video, increased throughput will lead to higher quality in terms of lower startup delays. For long-form video, assuming static network conditions, a certain sustained bitrate should result in a given quality. By sharing the specifics of this resource–quality curve, a CAP can enable a CSP to provide the required resources $R_{\min}$ for achieving the minimum required quality for the end-users $Q_{\min}$, but also inform about the maximum attainable quality $Q_{\max}$ for a given set of resources $R_{\max}$ without overprovisioning. Exceeding the sustainable amount of resources $R_{\max}$ implies two sustainability risks: 1) spending (or rather wasting) too much additional resources on marginal, hardly perceivable quality improvements, and 2) the risk of provoking quality degradations and even customer pushouts in restricted capacity environments such as mobile cells, due to overconsumption of network resources and subsequent bottleneck effects (Fiedler et al., 2016).

In addition, the shared metrics should enable fair comparison between different service types and providers, promoting transparency while acknowledging the complexity of calculating QoE scores for millions of streams. This requires using standardized or publicly available metrics, as mentioned above, but also sharing of the respective quality-resource aspects.

Metrics may be provided at various aggregation levels and time scales, from real-time reaction to congestion to long-term capacity planning. For instance, real-time metrics on events like rebuffering may inform CSPs about ongoing issues during live events, whereas aggregated weekly or monthly metrics may help CSPs identify underserved regions, or highlight the need to improve peering infrastructure.

Lastly, the type of **measurement and the context** in which it is done must be considered. We assume that most metrics will be collected (passively, in the background) in actual production environments stemming from the actual use of services by end-users. However, synthetic monitoring using active probes may also be possible to provide a measurement free of possible end-user biases (e.g., degraded WiFi resulting in poor QoE, even though the CSP is not at fault). A laboratory context can be used for modeling, e.g., the resource–quality curve.

4.3 Towards a QoE Metric Framework

To create an effective and practical framework, we must establish certain scope limitations. Our primary focus will be on real-time monitoring, with an emphasis on short-form and long-form video streaming as the main target, and video conferencing as a secondary consideration. The framework will primarily address 5G networks, with fixed-access networks as a secondary target.

The proposed QoE management framework is built around a **shared state table** concept, implementing a logical shared view of QoS and QoE status between CAP and CSP. This approach creates virtual shared information where both parties maintain a real-time view of QoS and QoE from their respective perspectives.

Rather than relying on traditional signal and acknowledgment patterns, this system maintains shared state information *continuously*, including potentially divergent views of the same parameters (such as bandwidth estimates) from both parties. This functionality could be implemented through a side-channel or metadata description during handshake, potentially utilizing any mechanisms from Section 3.3.3 or similar ones[19].

The foundation for the reported information builds upon the ITU-T P.1203 model – see Figure 4, which provides modeled QoE metrics pertaining to different aspects of a single video streaming session. These include segment time-frame video and audio coding qualities (O.21 and O.22 respectively – a per-second MOS), long-term stalling effects (O.23), and an integration module for short-term audiovisual quality (O.34), long-term audiovisual quality (O.35), and long-term overall MOS (O.46). While this model specifically targets video streaming, its framework can be adapted using other quality metrics for the video module, such as ITU-T Rec. P.1204.3, P.1204.4, or VMAF. Hence, in the following, the P.1203 framework is just used to define the semantics of the different metrics to be shared. For other use cases, such as gaming (ITU-T G.1702), the model can incorporate additional factors like interactivity.

This approach is also applicable to other services (see Raake et al. (2014)). More generally, analyzing various QoE models reveals common structural elements across different services/application types. These components generally fall into three main categories:

1. (Spatio-temporal) media fidelity
2. Media delivery continuity
3. Media delay and interactivity

These components may be supplemented by aggregation modules that combine sub-components or integrate across different time frames.

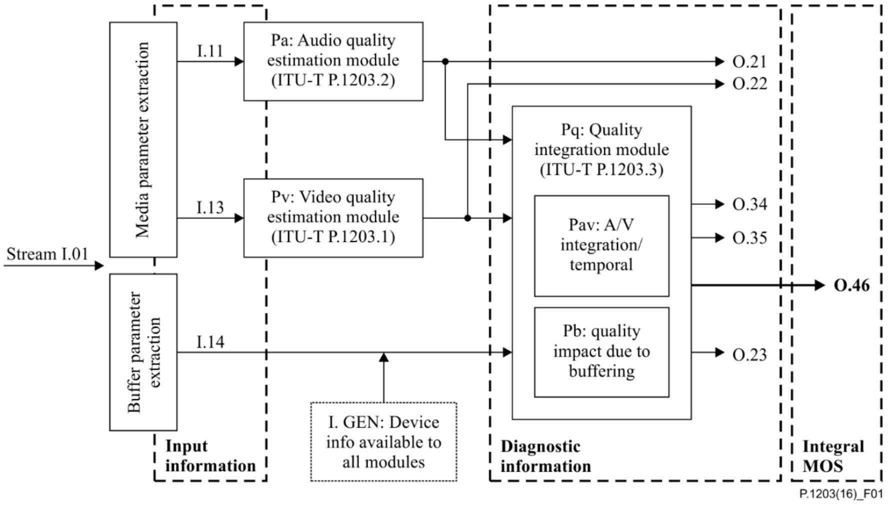

Figure 4 – Modeled QoE metrics pertaining to different aspects of a single video streaming session, per ITU-T Rec. P.1203.

The next table shows the main components for the use cases under consideration. As seen in the table, different use cases may share similar components, which opens the possibility to **share modeled or user-reported QoE metrics across use cases**. At the bottom, metrics that can be obtained at different layers. As seen in the table, application-layer KQIs

---

[19] The specific implementation is out of the scope of this white paper.

are easily related to these QoE components, while network-related KPIs may have impact across several of them.

Type	Media Fidelity	Media Delivery Continuity	Media Delay and Interactivity
Video Conferencing	Resolution & coding quality	Image freezing, “macroblocking” effect Audio continuity	E.g., Overlap Rate, Repeat Rate, Echo (He et al., 2024) End-to-end round-trip delay
Long-form Video	Resolution & coding quality	Rebuffering	Start-up delay, scrubbing behavior
Short-form Video	Resolution & coding quality	Session drop rate	Swiping interactivity Loading delay, stalls/re-buffering
Live Video	Resolution & coding quality	Rebuffering	“Channel change time” Fast FWD interactivity End-to-end single-way delay
Cloud Gaming	Resolution, rendered frame rate, and coding quality	Image freezing, and/or slicing effects due to jitter and packet losses	Motion-to-photon delay, uplink/downlink delay

Table 11a – Main QoE-related components for different use cases.

Layer	Media Fidelity	Media Delivery Continuity	Media Delay and Interactivity
System Layer System QoE	System QoE, System GoB		
Layer 8b: User Layer User-Reported QoE	Overall MOS Engagement / user retention rate / user watch time		
	Fidelity MOS	Continuity MOS	Interaction MOS
Layer 8a: User Layer Modeled QoE	VMAF, UVQ, P.1204.3, P.1204.4 (video) POLQA (speech) PEAQ (audio)	P.1203 O.23	Included in G.1072/G.107 (Cortés et al., 2023)
	P.1203 O.34/O.35 (audiovisual)		
	P.1203 O.46 (overall MOS, Streaming)		
	G.1072 (Cloud Gaming)		
Layer 7: Application Layer KQIs	Resolution, coding bitrate, codec, etc.	Buffering time, number of freezes	Loading delay
	3GPP TS 28.406 (QoE Measurement Collection)		
Layer 1-4: network & Transport Layer KPIs	Average throughput, throughput variation, packet delay / RTT, Jitter, packet loss rate and distribution, signal to noise ratio, MCS, RSSI, RSRP, RSRQ, directionality of traffic (uplink/downlink), congestion level (e.g., RAN PRB utilization) 3GPP TS 32.314 (Layer 2 Measurements) 3GPP TS 32.425 (Performance Management)		

Table 11b – Main QoE-related components for different layers.

4.4 High level architecture on how the ecosystem can work

In this section we will show examples of the benefits of QoE awareness in the management procedures that can be delivered to the ecosystem. We are describing the potential benefits of metrics exchange which in turn are leveraged in real-time operation to perform action through a control loop and exploit the collaboration among different tenants in the delivery of the services. As examples, short-form video, long-form video, and interactive services will be covered.

All use cases will use a common set of metrics to be shared between CAPs and CSPs in the shared state table (Table 12). To build Table 12, a bottom-up analysis has been done, stating, for each use case, which are the relevant metrics that need to be considered. This results in the three use cases sharing most of the metrics, but also showing some variations due to two factors: on the one hand, some QoE metrics do not apply to all cases (e.g. interaction quality for long-form video); on the other hand, some modeled QoE metrics have only been standardized for some use cases, but not for others (e.g. integral MOS in ITU-T P.1203 for long-form video streaming).

Layer	Metric	Definition	Algorithm (example)	Time scale (*)	Provider		(Video) Use case		
					CAP	CSP	Short Form	Long Form	Inter-active
8	Integral MOS	Blended metric of video / audio fidelity metrics, delivery continuity and interactivity	ITU-T P.1203 (O.46)	Segment / Session	X			X	
8	Video Quality	(Instant) video quality score (fidelity) : predicting video MOS	VMAF, UVQ, ITU-T P.1204.3/4/5	Segment / Session	X		X	X	X
8	Audio Quality	(Instant) audio quality score (fidelity) : predicting audio MOS	ITU-T P.1203 (O.22), PEAQ	Segment / Session	X			X	X
8	Interaction Quality	(Instant) interaction quality score: predictor of MOS for interactivity	P.BBQCG: Impact of latency, RTT and stuttering / frame freezes	Segment / Session	X				X
8	Stalls / buffering	Stall duration (ms)	ITU-T P.1203 O.23	Segment / Session	X		X	X	
8	Start-up delay	Start-up or swapping delay (ms)		Segment / Session	X		X		
8	ABR rate-quality function	Rate distortion curve, Complexity coefficients describing QoE to bitrate translations, with resolution and codec taken into account	Bitrate ladder (video quality for each bitrate) (Li et al. 2018)	Segment / Session	X		X	X	X
8	Interaction cost function	Interaction quality impact estimate of delay and jitter	Delay-QoE curve (Cortés et al., 2023)	Segment / Session	X				X
7	Application KQIs (metadata)	Video/audio codec, video/audio bitrate, framerate	DASH metadata (3GPP TS 26.247)	Segment / Session / Daily	X		X	X	X
		Frame freezes / actual frame rate, round trip time	P.BBQCG	Segment / Session	X				X
7	Rate guidance	Available throughput for each stream, if one is to achieve the optimal trade off between compression and delay for the current network state (Fiedler et al., 2016)	R_{max} (Fiedler et al., 2016)	Segment / Session		X		X	X

4	Subscriber network entitlement	Video Policy Rate Limit, Entitlements GSMA TS.43	Platform Operating System and vendor or CSP specific	Session / Daily+		X	X	X	X
1-4	Network state	Indicated as a quantitative or qualitative value, e.g., ECN bit, Low/Normal/High, packet loss, throughput, packet round-trip-time, packet jitter, marginal utility (e.g. Congestion Threshold Level in Nádas et al. 2021)...	Standard or vendor-specific	Segment / Session / Daily+		X	X	X	X

*Time scales: Segment (<10s), Session, Daily+ (days to months)

Table 12 – General overview of metrics exchanged, per CAP and per CSP, for different use cases (*shared state table* concept).

4.4.1 Use Case on Short-Form Video

Short video contents are generally pre-loaded according to the recommendation and ranking engine algorithm and also vary depending on the network types (e.g., cellular vs Wi-Fi). An optimized pre-loading strategy should be applied to satisfy the user experience with reasonable loading time, meanwhile not over-preloading the contents that causes network congestion, and conversely ensuring sufficient capacity to make the prefetching/preloading transparent and invisible to the end-user and not impacting QoE. In addition, the ABR (Adaptive Bit Rate) algorithm that is built-in short-form video applications allows for selection of video quality lanes (with varying levels of video fidelity and resolution) depending on network link capacity and bottlenecks which are generally estimated on the client side with some bandwidth estimation algorithm.

4.4.1.1CAPs → CSPs Actionable Control

CAP can expose metric(s) from the proposed list in Table 12 such as video QoE fidelity score (QoE metric/model unified scale and definition, e.g., VMAF, UVQ, P.1203 ….) or encoding bit rate of the first 5 segments being prefetched in the client video player buffer when the application is launched (note this process can be repeated every x minutes as new segments are being prefetched and videos are consumed by the player). CAP can expose the minimal required bitrate $R_{\min}$ to the CSP, which is related to minimal QoE $Q_{\min}$ to be met in order to allow for a sufficient burst volume (number and time) of prefetched video material to avoid churn. In addition, CAP can potentially expose QoE timeliness metrics such as the current in-play stalls duration, or the typical behaviour of a given (or a typical) user, such as the average number of video clips streamed and visioned per minute, or the average time spent on each video clip.

The CAP provided information on video related information, normally not visible in encrypted streams, can be leveraged by CSP to optimize network resources, such as changing resource block allocation to a given traffic flow that video fidelity is under a certain acceptable threshold. The increased visibility of video session requirements to be delivered is assisting CSPs in the management of their networks on a real-time basis. Some initial simulation and modeling (Wehner et al., 2025) are demonstrating that client video player buffer depletion can be eliminated or avoided thus resulting in stalls reduction and or elimination if such visibility is provided. A Markov model is used to investigate what has the highest impact on the QoE regarding preloading strategies, available network bandwidth, and swiping behavior for realistic assumptions on the video bitrates, network throughput, swipe rates and real-world preloading strategies.

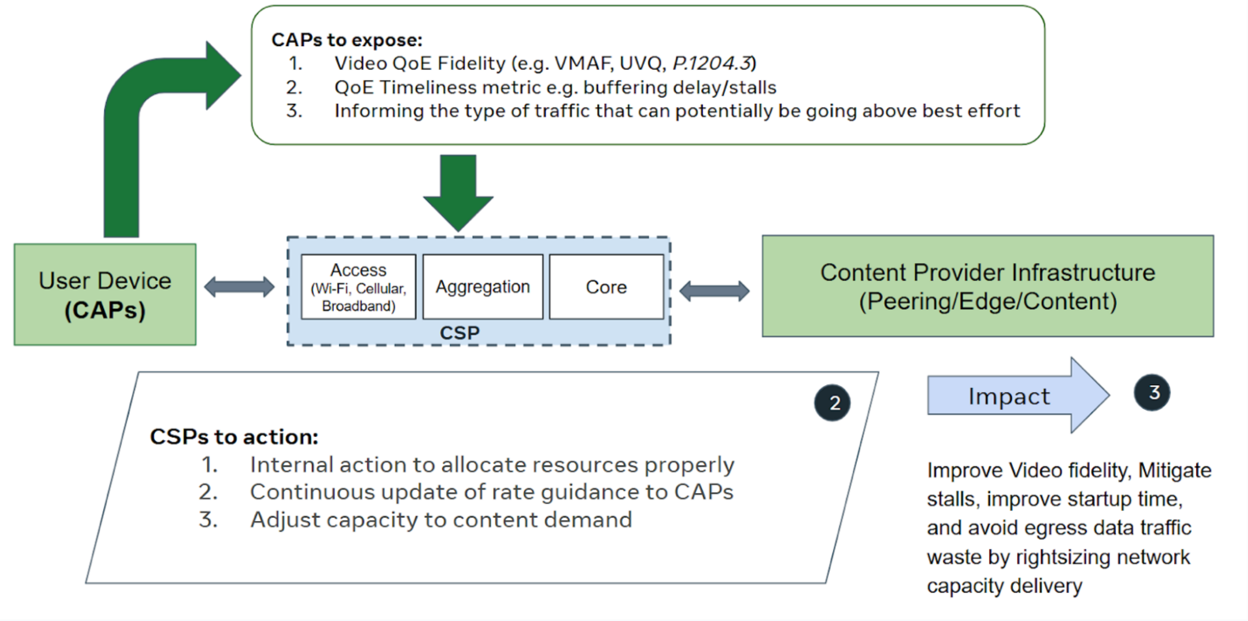

Figure 5 – Ecosystem Collaboration between CAPs and CSPs, with CAPs sharing/exposing metrics for Short-Form Videos

4.4.1.2 *Ecosystem Impact*

The benefits of having such a control loop in the real-time operation allows CSPs to better optimize network resources, during a session, dedicating more resources for some short period of time. Priority for more time-sensitive applications results in improved video fidelity, mitigation of stalls, improved startup time, and avoidance of egress data traffic waste by rightsizing network capacity delivery to deliver acceptable level of service quality and the visibility to track it.

This ecosystem collaboration can also be applied in non-real-time time scale (day, week) in certain areas (spectrum and coverage at planning level), backhaul connectivity to the core to support network planning. Further, bandwidth changes for video can also be reflected in the metrics exchange, e.g., if users are changing resolution. Furthermore, feedback collection from surveys of the services can be provided.

4.4.1.3 *CSPs → CAPs Actionable Control*

In this second example, we look into the interaction in the other direction whereby CSPs are exposing and or sharing relevant metrics to CAPs.

The CSP exposes metrics from the proposed list in Table 12 such as subscriber's network entitlements including video policy rate limit, e.g., standard definition video support for the plan (implemented in a number of USA operators). Allowing more resources at the beginning accelerates prefetching and reduces the first video initial loading delay. However, many synchronous big bursts might exhaust server and network capacities. Such a mismatch between resources available (CSPs) and demand from content application providers (CAPs) manifests itself in delays exacerbated by packet loss and retransmissions, yielding starting delays as well as freezes. Exceeding the maximal bitrate $R_{\max}$ implies overuse of available resources and subsequent drops of QoE due to impairments caused by startup delays due to resource shortage.

CSP may expose subscriber's network entitlement (e.g. allowed video policy rate), or network health utilization/congestion state to provide guidance on the allowable $R_{\max}$, CSP may send signaling messages to indicate congestion, enabling CAPs to adapt transport layer flows within a few Round Trip Times, thereby avoiding significant packet loss or increased latency. In situations of prolonged congestion, indicated by consistently high numbers of congestion-marked packets, potentially exacerbated by additional buffer latency and packet drops at a CSP bottleneck, the CSP may suggest an available $R_{\max}$ to other CAPs. This $R_{\max}$ recommendation is contingent on the resources available for the observed number of flows.A CAP can use this information to either select the optimal ABR track or tune the delivery rate. A CAP can even self-regulate the entire delivery to a given rate by understanding the current network limitations.

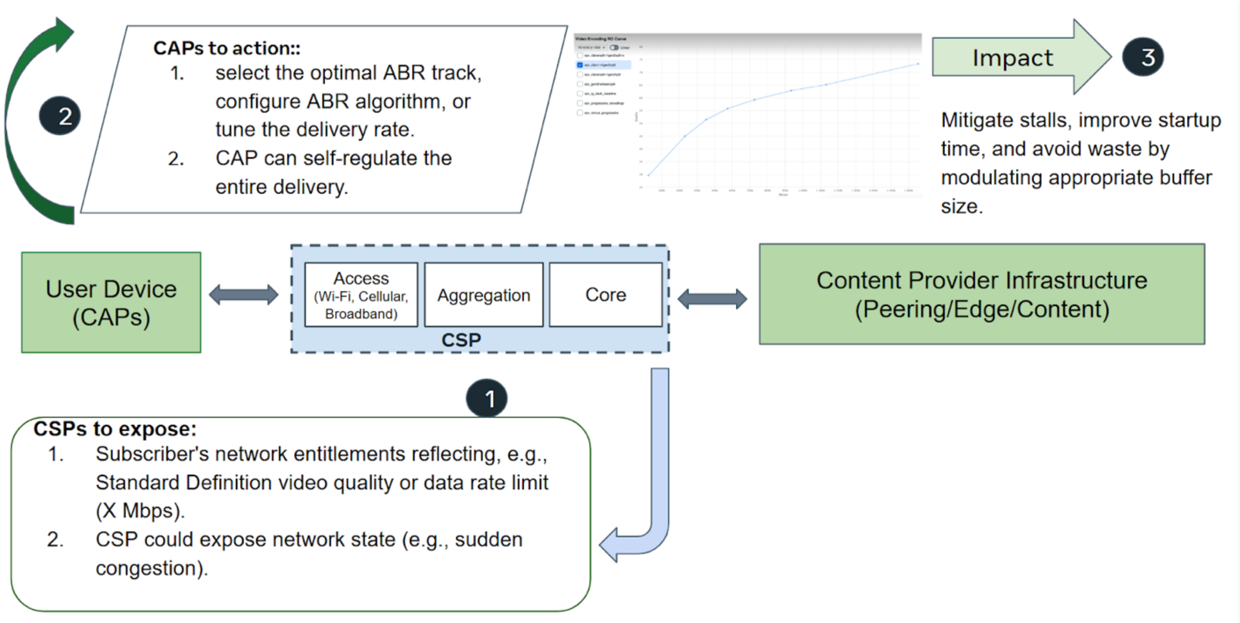

Figure 6 – Ecosystem Collaboration between CAPs and CSPs, with CSPs sharing/exposing metrics for Short-Form Videos

4.4.1.4 *Ecosystem Impact*

Implementing a real-time control loop allows Content/Application Providers (CAPs) to optimize video streaming resources more effectively. This optimization can lead to several benefits, including reduced stalls, improved startup times, and the prevention of resource waste. These improvements are achieved by appropriately adjusting buffer sizes and selecting video encoding lanes. It is crucial to consider the different time scales at which transport protocol control loops and CSP/CAP signaling control loops influence flow behavior, ensuring these control loops operate independently.

With a better understanding of current and historical network health, CAPs can predict possible issues and remedy them with appropriate action even before issues occur. For instance, they can reduce mismatch in link speed and CAPs' bandwidth estimates with more accurate tracking of network link capacity and improve de-jitter/playback buffer adjustment.

Understanding the effect of service performance decisions on network usage provides insights into balancing optimal codec/encoding rates together with resulting delay metrics, and to help understand why users have issues.

4.4.2 Use Case on Long-Form Video

Long-form videos represent a major share of network traffic and have been in focus of many QoE studies and standards (Garcia et al., 2014, Barakabitze et al., 2019). During the past years, the streaming performance has been optimized based on HTTP Adaptive Streaming (HAS, e.g., through MPEG-DASH or HLS) with appropriate bitrate ladders and advanced video segment/chunk selection strategies, allowing for maximizing fidelity and minimizing rate and amplitude of (perceivable) quality switches and/or stallings. As HAS reacts to consequences of network performance degradation by “climbing down the bitrate ladder”, i.e., reducing the bitrate, the maximal bitrate $R_{\max}$ can be deduced from observing the DASH segment selection activities.

As opposed to short-form videos, initial waiting times are considered less disturbing than the appearance of freezes, and users are more likely to tolerate larger loading times (Krishnan et al. 2013). The risk for the latter grows with the amount of additional throughput variability induced by the network due to a shared bottleneck (Fiedler et al., 2000), also denoted as bandwidth jitter (BJ). The knowledge of the latter helps to dimension the buffer size and to optimize the segment selection strategy. The effect of BJ can be mitigated by increasing the client player buffer size, which may affect QoE through initial waiting times. The BJ (and thus the initial waiting time) may be reduced by decreasing the maximal bitrate $R_{\max}$. Due to the complexity and implicitness of the bitrate control process, exchange of metrics between CAPs and CSPs are expected to be beneficial for the overall QoE optimization.

Both CAPs and CSPs may refer to the same underlying parametric model as an effective tool to conduct tradeoff analysis and determine impact (e.g., assign resources) for a given threshold (e.g., the bitrate required to achieve MOS between 3.5 and 4.5), for instance using P.1203, VMAF, or UVQ as commonly accepted metric.[20] This way, both parties do not need to expose “internal” information, but might deduce their “to do’s” based on that.

The relation between the CSP-oriented resource domain (e.g., bitrate R) and the CAP-oriented quality domain (Q) is shown in Figure 3. Agreeing upon a minimal acceptable quality $Q_{\min}$, the CSP can deduce the minimal required bitrate $R_{\min}$. As this will be content- and encoding-dependent, the CAP must at least share information about the current quality, and *should* share a mapping as indicated in Figure 3. Similarly, the maximal “meaningful” $Q_{\max}$ helps to avoid overallocation and thus waste of resources by limiting the bitrate to $R_{\max}$, a value that can be deduced by the CSP through the parametric model, while the CAP should respect $Q_{\max}$ to avoid overuse of resources.

The maximal bitrate $R_{\max}$ should be chosen such that quality disturbances due to resource overallocation and bottleneck effects are avoided. This is usually handled by HAS. However, if that control was not effective enough, the CAP might benefit from receiving a signaling message to indicate congestion from the CSP, thus being able to reduce its bitrate.

---

[20] Note that VMAF does not output a score on a MOS range, and both VMAF and UVQ do not consider stalling in their quality scores. For the purpose of QoE optimization, a range of VMAF scores could be targeted. When metrics are used that do not reflect stalling in their quality calculations, explicit stalling indications by the CAP would be required to assess overall QoE.

It has to be noted that the CSP faces the challenge of using the parametric model "backwards" – basically an inversion of the bitrate ladder as seen from Figure 7 – in order to derive the amount of network resources and other parameters to achieve a given QoE, see ITU-T P.1211 (P.DiAQoSe).

The essential metrics to be exchanged are then, based on Table 12, the current quality (video, audio), or, if unavailable, at least the metadata required to infer the current quality from a CSP side (Rao et al. 22), as well as the resource–quality curve for the given service.

Other metrics, which are useful but most likely not available/not sharable, are user behavior (watch time), user engagement (audience retention rate), or evaluations regarding user abandonment. Rate distortion curve user acceptance reference paper can provide additional background see Hoßfeld et al. (2023).

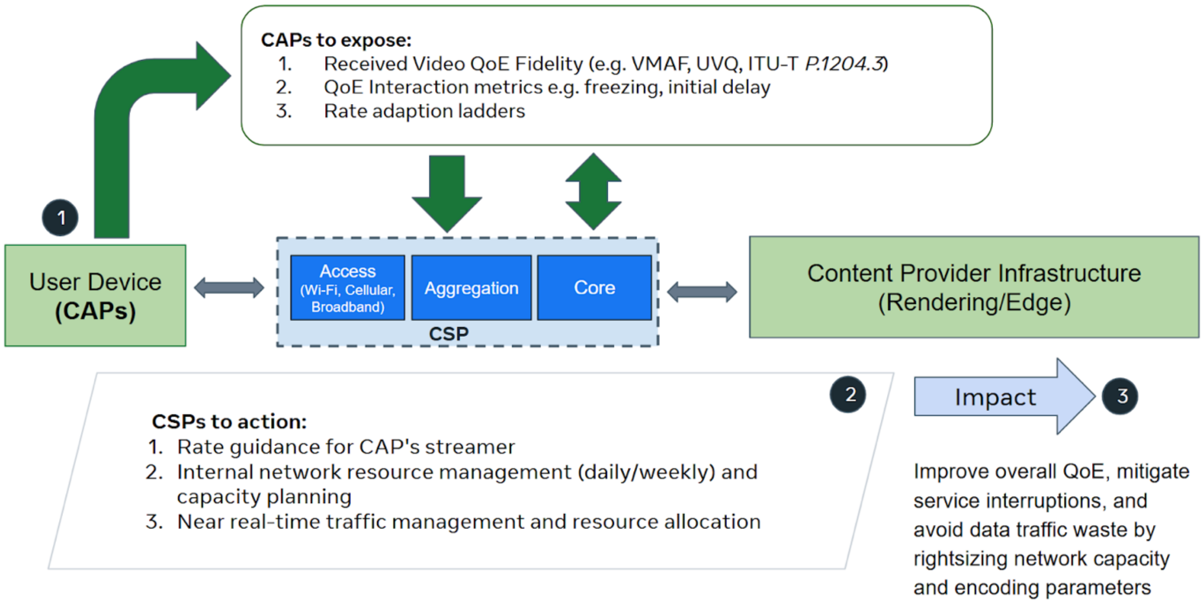

Figure 7 – Ecosystem Collaboration between CAPs and CSPs, with both entities sharing/exposing metrics for Long-Form Videos

4.4.2.1 Ecosystem impact

Sharing a joint view on the "quality window" to be provided (between e.g., Q_{min} and Q_{max}) implies a set of advantages for both parties: (1) A CSP can contribute to the avoidance of unacceptable QoE conditions, which is important for users and both CAP and CSPS user churn, by providing sufficient amounts of resources as needed; (2) Unsustainable overexploitation of CSP's resources is reduced, as the bit rate ladder is only used as far up as a positive (and perceivable) impact on the fidelity can be motivated. Furthermore, the CSP has means to warn the CAP of unfortunate network conditions, such as exceeding the sustainable bitrate. To guide the allowable R_{max}, the CSP can expose information about the subscriber's network entitlement (e.g., permitted video policy rate) or the network's health/utilization/congestion status. In addition, through congestion signaling from CSPs, CAPs can react promptly. This allows for transport layer flows to be adjusted within a few Round Trip Times, effectively mitigating significant packet loss or increased latency.

The combination of shared view on a jointly agreed-upon (parametric) quality model and an explicit notification channel between CSP and CAP facilitates, both a smoother and more sustainable QoE experience of long videos can be reached.

4.4.3 Use Case on Interactive Services (Gaming or Video Conferencing)

Remote rendered services such as cloud gaming work with very small jitter buffers compared to other services, to add as little delay as possible. This makes it sensitive to disturbances and congestion, but it also gives an opportunity to make changes in real time that can have a significant effect on the service overall.

In this example, the CSPs expose congestion levels and recommendations to each service "flow" on rates that optimize the overall QoE levels in the network location (e.g. mobile network cell or, more generically, access network segment). This could mean sending metrics from lower layers in Table 12, such as congestion state, packet round trip time, or throughput.

The CAPs on the other hand expose current video fidelity scores with accompanying parameters and KQIs (codec, resolution, video frame rate, video bitrate), estimated bitrate adaptation trade-offs (fidelity score vs bitrate, as illustrated in Figure 3), minimal required resources like $Q_{\min}$, as well as jitter and delay sensitivity for the service, the latter two based on interaction and continuity components for service appropriate QoE estimation models. Encoding bitrate adaptation ladder parameters must give the CSPs information about what the CAP knows or estimates will happen with the QoE when the encoding bitrate has to change. Similarly, the jitter and delay sensitivity for the service must relay the QoE impact that adjusted delay and jitter will have. Using this generalization, this use case could cover both the remote rendering case but also something with less strict interactivity demands such as video conferencing.

With this information available for all services in an "evaluation area" (such as a mobile network cell, which will be used from hereon), the CSP can make holistic decisions and provide guidance to each service flow to optimize the overall service quality in the cell/system using a QoE fair process.

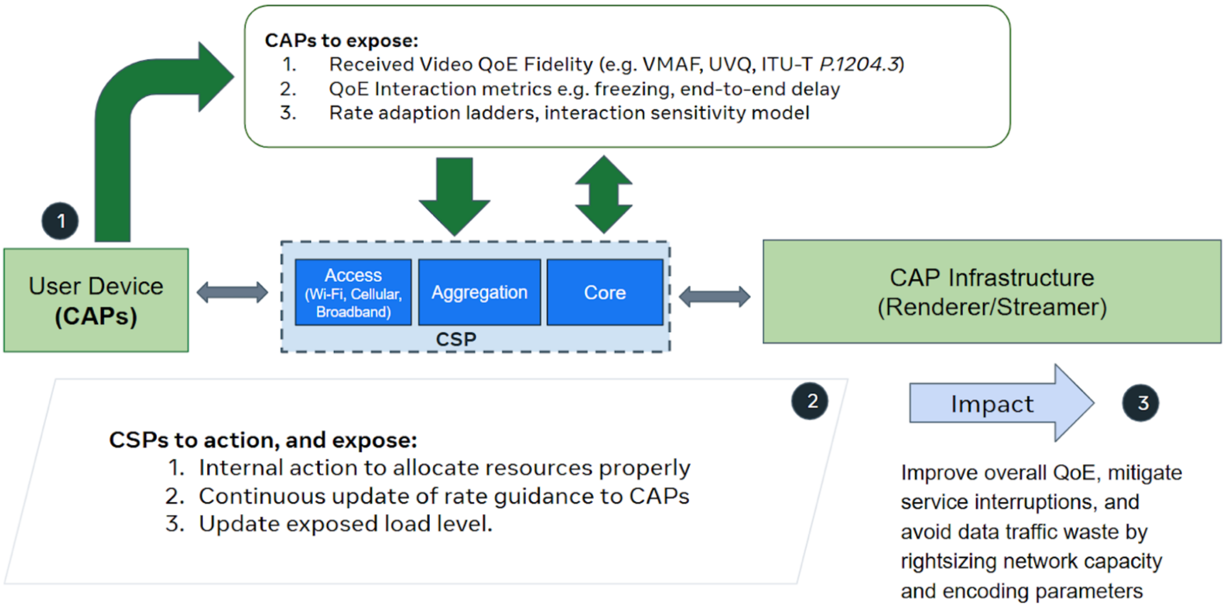

Figure 8 – Ecosystem Collaboration between CAPs and CSPs, with both entities sharing/exposing metrics for interactive services

4.4.3.1 *Ecosystem impact*

By collecting current QoE states of each service flow in a cell, and by having an understanding of what will happen to each service if something is adjusted, the CSP can provide holistic guidance based on a QoE fair resource allocation. Doing this while taking both the bandwidth and delay implications for each service flow, the CSP can prioritize and predict the outcome of those prioritizations to make decisions that are increasing the overall QoE and not over-provision some flows or users unnecessarily. This will make it possible to fit more satisfied users in the same network (Nádas et al. 2024). When those QoE states are updated continuously, the overall QoE can be further optimized by taking into account the current states. Moreover, QoE preferences can be formalized and met using utility functions (Nádas et al. 2024).

Since these services are quickly affected by both delay and bandwidth restrictions, it is important that the controlling entity knows the trade-offs to be able to give good guidance to the streamer/source. It is also important to know the network state to understand what will happen if you allow the streamer to increase the bit rate; will the delay then increase or not, and what will be the final QoE adjustment? In an extreme case, this fully replaces transport layer congestion control, the CSP's controller detects congestion in its network and sends guidance messages to the application, which will reduce congestion. Depending on the potential bottlenecks, congestion control is still needed to control bottlenecks not under the control of the CSP or act as an emergency switch. Congestion control loops are found on CAP and CSP levels and depending on their structure (outer/inner loop), parameterization and in particular reaction speeds (time constants), they might cooperate well or disturb each other, causing instability. Further research in this area is necessary to analyze control loops and propose the most applicable solutions.

4.5 Privacy Considerations

The potential of the CAP–CSP information exchange proposed in this white paper depends on the ability to provide actionable insights, which often require a fine-grained view of performance and QoE, potentially at a per-session or even per-media-segment level, as described in the previous sections.

This level of detail is ultimately required in order to assess quality in a valid manner. For instance, it allows for the precise, real-time diagnosis of issues like sudden video stalling events indicated by CAPs, without resorting to assumptions by the CSP about the playback behavior. Furthermore, QoE management of interactive services such as cloud gaming is only possible when immediate feedback on network conditions for a specific stream in a given network location are available.

Understandably, the implementation of the proposed framework using data tied to actual users' media consumption may raise questions about user privacy. The approach we present is therefore purely opt-in; it is not a protocol that must be used by all stakeholders in the ecosystem. Instead, we highlight the mutual benefits for all involved parties if they choose to implement the described framework.

The practical implementation of the data exchange models must respect strict privacy and business confidentiality boundaries and must not violate the fundamental right to privacy granted to end-users in various jurisdictions. In practice, this means certain data is off-limits or needs sanitization. For example, personally identifying user information (persistent user IDs, subscriber accounts) and detailed session-level logs should typically not be shared outright, due to privacy laws (e.g. EU GDPR) and proprietary concerns. Likewise, enabling the tracking of end-users (e.g., via specific user agents or persistent IDs, as mentioned above) should be prevented by both sides (CAPs and CSPs). The goal is therefore to create a system where enhanced QoE for end-users (characterized by stalling-free streaming, faster load times, and more reliable interactive sessions) is achieved through a transparent, privacy-preserving, and collaborative effort between CAPs and CSPs.

To ensure privacy, we highlight the principle of pseudonymization for any operational exchange of granular data. When a session begins (e.g., a video stream or a gaming session), a temporary, non-persistent, and cryptographically generated session pseudonym could be used. This transient identifier, known only to the directly involved CAP and CSP for the active session's duration, could serve as an operational token for exchanging relevant performance metrics or events, like stalling events. Critically, this pseudonymization should be designed to be completely unlinked from any persistent user or device identifiers, ensuring that the analysis remains focused on the operational characteristics of the data flow, not the individual.

Furthermore, the exchange of such pseudonymized, granular data should be governed by purpose limitation and data minimization. The information shared should always be confined to the essential metrics required for immediate monitoring, operational adjustments, and service optimization – which is in the end-users' best interests. For example, a CAP might signal that a session is experiencing a QoE drop, allowing a CSP to investigate network conditions in a given cell relevant to that specific, pseudonymized flow – without needing to know the specific content or user affected. Also, the data exchange could be limited for immediate use and not for long-term logging or cross-session analysis, which may lead to user profiling. Once its operational purpose is served, data must be discarded (as per GDPR) or further anonymized, e.g., into an aggregated dataset for broader, long-term system performance analysis and planning. Or, if granular data is not needed, pre-aggregated data can be shared by CAPs or CSPs.

Finally, it must be reiterated that the entire framework is built on the premise of voluntary participation (opt-in) by both CAPs and CSPs, driven by clear mutual benefits. This careful approach to granular data exchange can be an enabler for our broader vision.

5 Conclusions and Next Steps

5.1 Conclusions

As described in this white paper, there is a clear need for aligned and practical QoE definitions and metrics for collaboration between CAPs and CSPs. The white paper addresses this need in two steps.

First, it structures the existing state of the art, including both theory and implementations, in a layered approach inspired by the ISO/OSI model, providing precise definitions for the different conceptualizations of quality that appear at each level.

Second, it proposes a flexible framework for exchanging quality information between CAPs and CSPs, based on a shared state table and semantic building blocks, which offers a viable path to address this challenge. Importantly, this framework does not prescribe a single universal QoE metric; instead, it accommodates different QoE measures depending on the specific use case and implementation context. As also described in this document, this information exchange enables tangible benefits for both players in terms of efficiency, fairness, and improved QoE.

In Chapter 4, we applied the proposed framework and demonstrated the potential impact of metrics-KPIs exchange between CSPs and CAPs by establishing an ecosystem collaboration, which in turn can result in an overall improvement of video service delivery performance and efficiency, for example, by reducing the duration of stalling, or reducing waste of resources, all through increasing visibility between the two parties.

5.2 Next steps

This white paper is an initial step in the long-term of VQEG of helping the industry to develop new techniques to improve end-to-end QoE in multimedia services delivery. The next steps have been identified as follows.

1. Further develop a QoE model based on the framework described in Chapter 4, considering at least one use case. From the list of metrics that have been identified, select which ones can be deployed in a first implementation.
2. Perform a short term validation on the idea: proof of concept and test or simulation of basic conditions of the use case.
3. Analyze the complexity of implementing and sharing these metrics in the ecosystem and describe advantages, disadvantages, technical limitations and potential mitigations. Clarify how CAPs and CSPs will reliably identify and correlate traffic flows to enable consistent metric sharing across the ecosystem. Define the required QoS monitoring coverage—favoring end-to-end visibility and safeguards against local optimization—to ensure issues are detected where they occur. Provide recommendations to move forward that CAPs and CSPs can utilize, depending on technical feasibility, business cases, regulatory or other constraints.
4. Contribute the results to the appropriate standardization bodies to promote industry-wide standardization of the framework. On the one hand, contribute to the standardization of the framework itself (e.g., under ITU-T Study Group 12). On the other hand, contribute to the standardization of the necessary protocols and systems to implement the shared table mechanism (e.g., in IETF or 3GPP).

References

Aggarwal, V., Halepovic, E., Pang, J., Venkataraman, S., and Yan, H. (2014). Prometheus: Toward quality-of-experience estimation for mobile apps from passive network measurements. In Proceedings of the 15th Workshop on Mobile Computing Systems and Applications (pp. 1-6).

Balachandran, A., Sekar, V., Akella, A., Seshan, S., Stoica, I., and Zhang, H. (2012, October). A quest for an internet video quality-of-experience metric. In Proceedings of the 11th ACM Workshop on Hot Topics in Networks (pp. 97-102).

Barakabitze, A. A., Barman, N., Ahmad, A., Zadtootaghaj, S., Sun, L., Martini, M. G., & Atzori, L. (2019). QoE management of multimedia streaming services in future networks: A tutorial and survey. IEEE Communications Surveys & Tutorials, 22(1), 526-565.

Briscoe, B., Kühlewind, M., & Scheffenegger, R. (2023). More accurate explicit congestion notification (ECN) feedback in TCP. *Internet Engineering Task Force, Internet-Draft draft-ietf-tcpm-accurate-ecn-28.*

Cavanaugh, J. R., Hatch, R. W., and Sullivan, J. L. (1976). Models for the subjective effects of loss, noise, and talker echo on telephone connections. Bell System Technical Journal, 55(9), 1319-1371.

Contreras, L. M. (ed.). (2024). 5QI to DiffServ DSCP Mapping Example for Enforcement of 5G End-to-End Network Slice QoS, IETF draft-cbs-teas-5qi-to-dscp-mapping-03. Available at: https://www.ietf.org/archive/id/draft-cbs-teas-5qi-to-dscp-mapping-03.txt.

Cortés, C., Pérez, P., and García, N. (2023). Understanding latency and QoE in social XR. IEEE Consumer Electronics Magazine.

De Schepper, K. (2023). RFC 9331: The Explicit Congestion Notification (ECN) Protocol for Low Latency, Low Loss, and Scalable Throughput (L4S).

De Schepper, K., & White, G. (2023). RFC 9332: Dual-Queue Coupled Active Queue Management (AQM) for Low Latency, Low Loss, and Scalable Throughput (L4S).

De Schepper, K., Bagnulo, M., & White, G. (2023). RFC 9330: Low Latency, Low Loss, and Scalable Throughput (L4S) Internet Service: Architecture.

Fiedler, M. (2009). Quality feedback flows in future networks (Flüsse von Qualitäts-Feedback in zukünftigen Netzen). e & i Elektrotechnik und Informationstechnik. Volume 126, pages 269–273.

Fiedler, M., Hoßfeld, T., and Tran-Gia, P. (2010). A generic quantitative relationship between quality of experience and quality of service. IEEE Network, 24(2), 36-41. DOI: 10.1109/MNET.2010.5430142.

Fiedler, M., Popescu, A., and Yao, Y. (2016). QoE-aware sustainable throughput for energy-efficient video streaming. In 2016 IEEE International Conferences on Big Data and Cloud Computing, BDCloud 2016, Social Computing and Networking, SocialCom 2016 and Sustainable Computing and Communications, SustainCom 2016 (pp. 493-500). DOI: 10.1109/BDCloud-SocialCom-SustainCom.2016.78.

Fiedler, M., Tutschku, K., Carlsson P., and Nilsson, A. (2003). Identification of performance degradation in IP networks using throughput statistics. . Teletraffic Science and Engineering, Elsevier, Volume 5, pages 399–408..

Floris, A., Ahmad, A., Atzori, L. (2018). QoE-aware OTT-ISP collaboration in service management: Architecture and approaches. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 14(2s), 1-24.

Frank, B., Poese, I., Lin, Y., Smaragdakis, G., Feldmann, A., Maggs, B. ..., Weber, R. (2013). Pushing CDN-ISP collaboration to the limit. ACM SIGCOMM Computer Communication Review, 43(3), 34-44.

Garcia, M. N., and Raake, A. (2011, September). Frame-layer packet-based parametric video quality model for encrypted video in IPTV services. In 2011 Third International Workshop on Quality of Multimedia Experience (pp. 102-106). IEEE.

Garcia, M. N., De Simone, F., Tavakoli, S., Staelens, N., Egger, S., Brunnström, K., & Raake, A. (2014, September). Quality of experience and HTTP adaptive streaming: A review of subjective studies. In 2014 Sixth International Workshop on Quality of Multimedia Experience (pp. 141-146). IEEE.

Hammer, F., Egger-Lampl, S., Möller, S., and Experience, U. (2018). Quality-of-user-experience: a position paper. Quality and User Experience, 3(1), 9. DOI: 10.1007/s41233-018-0022-0.

He, J., Ammar, M., Zegura, E., Halepovic, E., and Karagioules, T. (2024). QoE Metrics for Interactivity in Video Conferencing Applications: Definition and Evaluation Methodology. In Proceedings of the 15th ACM Multimedia Systems Conference (MMSys '24) (pp. 178-189). Association for Computing Machinery, New York, NY, USA. DOI: 10.1145/3625468.3647622.

Hoßfeld, T, , Seufert, A, Loth F, Wunderer, S, (2023). Industrial User Experience Index vs. Quality of Experience Models. IEEE Communication magazine.

Hoßfeld, T., and Pérez, P. (2024). A theoretical framework for provider's QoE assessment using individual and objective QoE monitoring. In 2024 16th International Conference on Quality of Multimedia Experience (QoMEX) (pp. 235-241). IEEE.

Hoßfeld, T., Fiedler, M., and Zinner, T. (2011, June). The QoE provisioning-delivery-hysteresis and its importance for service provisioning in the future internet. In 2011 7th EURO-NGI Conference on Next Generation Internet Networks. IEEE.

Hoßfeld, T., Heegaard, P. E., Skorin-Kapov, L., and Varela, M. (2017, May). No silver bullet: QoE metrics, QoE fairness, and user diversity in the context of QoE management. In 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX) (pp. 1-6). IEEE.

Hoßfeld, T., Heegaard, P. E., Varela, M., and Möller, S. (2016). QoE beyond the MOS: an in-depth look at QoE via better metrics and their relation to MOS. Quality and User Experience, 1, 1-23.

Hoßfeld, T., Seufert, M., Hirth, M., Zinner, T., Tran-Gia, P., and Schatz, R. (2011, December). Quantification of YouTube QoE via crowdsourcing. In 2011 IEEE International Symposium on Multimedia (pp. 494-499). IEEE.

International Organization for Standardization (ISO). (1994). Information technology – Open Systems Interconnection – Basic Reference Model: The Basic Model. ISO/IEC 7498-1:1994.

Kiesel, S., Previdi, S., Roome, W., Shalunov, S., & Woundy, R. (2014). RFC 7285: Application-Layer Traffic Optimization (ALTO) Protocol.

Koniuch, K., Baraković, S., Husić, J. B., Subramanian, S., De Moor, K., Janowski, L., and Wierzchoń, M. (2024). Top-down and bottom-up approaches to video quality of experience studies; overview and proposal of a new model. Frontiers in Computer Science, 6, 1305670.

Krishnan, S. S., and Sitaraman, R. K. (2012). Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs. In Proceedings of the 2012 Internet Measurement Conference.

Le Callet, P., Möller, S., and Perkis, A. (2012). Qualinet White Paper on Definitions of Quality of Experience (Version 1.2). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland. Available at: http://www.qualinet.eu/images/stories/QoE_whitepaper_v1.2.pdf.

Li, J., Krasula, L., Baveye, Y., Li, Z., and Le Callet, P. (2019). AccAnn: A new subjective assessment methodology for measuring acceptability and annoyance of quality of experience. IEEE Transactions on Multimedia, 21(10), 2589-2602.

Li, Z., Bampis, C., Novak, J., Aaron, A., Swanson, K., Moorthy, A., & Cock, J. D. (2018). VMAF: The journey continues. Netflix Technology Blog, 25(1).

Möller, S., and Raake, A. (2014). Quality of Experience – Advanced Concepts, Applications and Methods. T-Labs Series in Telecommunication Services. Springer International Publishing, Switzerland.

Nádas, S., Ernström, L., Lindero, D., and Lynam, J. (2025) On QoE-Aware Traffic Management for Real-time, Interactive Video with Timevariant Spatial Complexity, arXiv:2507.11798

Nádas, S., Ernström, L., Szilágyi, L., Patra, G., Krylov, D., and Lynam, J. (2024). To QoE or not to QoE. In Proceedings of the 2024 Applied Networking Research Workshop (ANRW '24) (pp. 38-44). Association for Computing Machinery, New York, NY, USA. DOI: 10.1145/3673422.3674892.

Nádas, S., Varga, B., Contreras, L. M., & Laki, S. (2021). Incentive-Based Traffic Management and QoS Measurements. Measuring Network Quality for End-Users, An Internet Architecture Board virtual workshop.

Orsolic, I., and Skorin-Kapov, L. (2020). A framework for in-network QoE monitoring of encrypted video streaming. IEEE Access, 8, 74691-74706.

Papadogiannaki, Eva, and Sotiris Ioannidis. "A survey on encrypted network traffic analysis applications, techniques, and countermeasures." ACM Computing Surveys (CSUR) 54.6 (2021): 1-35.

Pérez, P. (2023). The transmission rating scale and its relation to subjective scores. In 2023 15th International Conference on Quality of Multimedia Experience (QoMEX) (pp. 31-36). IEEE. DOI: 10.1109/QoMEX58391.2023.10178659

Pérez, P., Ruiz, J., Benito, I., and López, R. (2021). A parametric quality model to evaluate the performance of tele-operated driving services over 5G networks. Multimedia Tools and Applications. DOI: 10.1007/s11042-021-11251-x.

Perkis, A., Timmerer, C., et al. (2020). QUALINET White Paper on Definitions of Immersive Media Experience (IMEx). DOI: 10.48550/arXiv.2007.07032.

Poese, I., Frank, B., Ager, B., Smaragdakis, G., Feldmann, A. (2010). Improving content delivery using provider-aided distance information. In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement (pp. 22-34).

Raake, A., and Egger, S. (2014). Quality and quality of experience. In Quality of Experience: Advanced Concepts, Applications and Methods (pp. 11-33). Springer International Publishing, Cham.

Raake, A., Borer, S., Satti, S. M., Gustafsson, J., Rao, R. R. R., Medagli, S., et al. (2020). Multi-model standard for bitstream-, pixel-based and hybrid video quality assessment of UHD/4K: ITU-T P. 1204. IEEE Access, 8, 193020-193049.

Raake, A., Garcia, M. N., Robitza, W., List, P., Göring, S., and Feiten, B. (2017, May). A bitstream-based, scalable video-quality model for HTTP adaptive streaming: ITU-T P. 1203.1. In 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX) (pp. 1-6). IEEE.

Rao, R. R. R., Göring, S., and Raake, A. (2022). AVQBits — adaptive video quality model based on bitstream information for various video applications. IEEE Access, 10, 80321-80351. DOI: 10.1109/ACCESS.2022.3195527.

Reichl, P., Tuffin, B., and Schatz, R. (2013). Logarithmic laws in service quality perception: where microeconomics meets psychophysics and quality of experience. Telecommunication Systems, 52, 587-600.

Reiter, U., Brunnström, K., De Moor, K., Mohamed-Chaker, L., Pereira, M., Pinheiro, A., You, J., and Zgank, A. (2014). Factors Influencing Quality of Experience. In Quality of Experience: Advanced Concepts, Applications and Methods, S. Möller and A. Raake (Eds.) (pp. 45-60). Springer. DOI: 10.1007/978-3-319-02681-7_4.

Robitza, W., Ahmad, A., Kara, P. A., Atzori, L., Martini, M. G., Raake, A., & Sun, L. (2017). Challenges of future multimedia QoE monitoring for internet service providers, Multimedia Tools and Applications, Volume 76, pages 22243–22266.

Seufert, M., et al. (2024). Marina: Realizing ml-driven real-time network traffic monitoring at terabit scale. IEEE Transactions on Network and Service Management.

Skorin-Kapov, L., Varela, M. (2012). A multi-dimensional view of QoE: the ARCU model. In 2012 Proceedings of the 35th International Convention MIPRO (pp. 662-666). IEEE.

T. Pauly et al, Architecture and Requirements for Transport Services, RFC9621, IETF, 2025

Timmerer, C., and Hoßfeld, T. (2025, June). O QoE, Where Art Thou? ACM SIGMM Records, QoE Column. Available at: https://records.sigmm.org/2025/05/02/o-qoe-where-art-thou/

Tomar, A., Ihlar, M., Eddy, W., Swett, I., Tiwari, A., & Joras, M. (2025). SCONE Need for Defining a New On-Path Signaling Mechanism (Internet-Draft draft-tomar-scone-ecn-01)

[Work in progress]. Internet Engineering Task Force. https://datatracker.ietf.org/doc/draft-tomar-scone-ecn/01/

Wehner, N., Blouin, F., Ouellette, M., Pérez, P., & Hoßfeld, T. (2025). Modeling Key Quality Indicators of Short-Form Video Preloading Strategies. In 2025 17th International Conference on Quality of Multimedia Experience (QoMEX). IEEE.

Wu, Q., Yang, Y., Lee, Y., Dhody, D., Randriamasy, S., & Contreras, L. (2023). RFC 9439: Application-Layer Traffic Optimization (ALTO) Performance Cost Metrics.

Zhu, Y., Heynderickx, I., and Redi, J. A. (2015). Understanding the role of social context and user factors in video quality of experience. Computers in Human Behavior, 49, 412-426.

Copyright & License Notice

Acknowledgements

The authors would like to thank Wolfgang Balzer, Vincent Barriac, Andy Furnell, Rüdiger Geib, Tanja Kojic, Craig Sparks, Anoop Tomar, Maurizio Vergari, and Kazuhisa Yamagishi for their thorough reviews and the valuable feedback they provided on this white paper.

Any use of generative AI in this paper adheres to ethical guidelines for use and acknowledgment of generative AI in academic research. Each author has made a substantial contribution to the work, which has been thoroughly vetted for accuracy, and assumes responsibility for the integrity of their contributions.[21]

In particular, LLMs have been used for:

- Grammar improvements and rephrasing
- Checking the document for missing references and abbreviations
- Content summarization

---

[21] https://zenodo.org/records/14845058

Appendix A

Video

1. ITU-T G.1070 (2018) – Opinion model for video-telephony applications
2. ITU-T G.1071 (2016) – Opinion model for network planning of video and audio streaming applications
3. ITU-T P.1203.1/3 (2019) – Parametric bitstream-based quality assessment
4. ITU-T P.1204.3/4/5 (2020) – Video quality assessment of streaming services up to 4K

Table A1 – Video QoE Industry standards.

Audio

1. ITU-T G.109 (1999) – Definition of categories of speech transmission quality
2. ITU-T G.114 (2003) – General Recommendations on the transmission quality for an entire international telephone connection
3. ITU-T P.1201.2
4. ITU-T G.107 (2016) – The E-model: a computational model for use in transmission planning
5. ITU-T P.1305 (2016) – Effect of delays on telemeeting quality
6. ITU-T P.1310 (2017) – Spatial audio meetings quality evaluation
7. ITU-T P.1203.2

Table A2 – Audio QoE Industry standards.

Gaming (Cloud and terminal based)

1. ITU-T G.1032 (2017) – Influence factors on gaming quality of experience
2. ITU-T G.1072 (2020) – Opinion model predicting gaming quality of experience for cloud gaming services

Table A3 – Gaming QoE Industry standards.

Telemetry and QoE-QoS Planning

1. ITU-T Y.1541 (2011) Network performance objectives for IP-based services
2. MEF 23.2 (2016) – Carrier Ethernet Class of Service
3. ITU-T J.1631 -Functional requirements of E2E network platforms to enhance the delivery of cloud-VR services
4. ITU-T GSTR-5G QoE (2022) – Quality of experience (QoE) requirements for real-time multimedia services over 5G networks
5. IETF IOAM – In-Situ flow and on-path telemetry
6. TU-T P.1211 (2023) – Diagnostic assessment of QoS and QoE for adaptive video streaming sessions

Table A4 – Telemetry & QoE-QoS Planning Industry standards.

Metaverse AR/VR/XR

1. ITU-T P.1310 (2017) – Spacial audio meetings quality evaluation
2. 3GPP 26.918 (2020) – Virtual Reality (VR) media services over 3GPP
3. 3GPP 26.928 (2021) – Extended Reality (XR) in 5G
4. ITU-T G.1035 (2021) – Influencing factors of QoE for VR services
5. ITU-T Y.3109 (2021) – QoS assurance related requirements for VR
6. ITU-T G.1036 (2022) – QoE influencing factors for AR services
7. ITU-T P.1320 (2022) – QoE assessment of XR meetings

Table A5 – Metaverse AR/VR/XR Industry standards.