Ref. Ares(2016)3118724 - 30/06/2016

PUBLIC

Contract n. 643072

QoE-Net

innovative Quality Of Experience maNagement in Emerging mulTimedia services

D5.1 QoS/QoE Monitoring Tools

PROJECT INFORMATION

Project Number	643072	**Acronym**	QoE-Net
Full Title	innovative Quality Of Experience maNagement in Emerging mulTimedia services		
Topic	MSCA-ITN-2014-ETN Marie Skłodowska-Curie Innovative Training Networks (ITN-ETN)		
Starting date	01/01/2015		
Duration in months	48		
Coordinator	UniCA – UniversitàdegliStudi di Cagliari		
Project URL	http://www.qoenet-itn.eu		

DOCUMENT INFORMATION

Deliverable Number	5.1	**Title**	QoS/QoE Monitoring Tools
Workpackage Number	5	**Title**	QoE monitoring, control and management
Due date of deliverable	M18		
Submission date of deliverable	21.06.2016		
Responsible partner	DT		
Responsible author/s	Werner Robitza, Saman Zadtootaghaj (editors), Avsar Asan , Alcardo Alex Barakabitze, Elisavet Grigoriou		
Dissemination level	Public		

CHANGE RECORDS

Date	Author	Modifications	Version
March. 11, 2016	Werner Robitza, SamanZadtootaghaj	First version	0.1
March 12, 2016	Werner Robitza	Updates	0.2
March 17, 2016	Werner Robitza	Updates based on Telco	0.3
March 17, 2016	Saman Zadtootaghaj	Updates based on comments	0.4
March 21, 2016	Werner Robitza	Updates on sections 4, 5	0.5
March 21, 2016	Werner Robitza	Updates on VoIP solutions and Citations	0.6

April 05, 2016	Werner Robitza	Updates on VoIP solutions	0.7
April 07, 2016	Saman Zadtootaghaj	Updates on Conferencing solutions	0.8
April 08, 2016	Elisavet Grigoriou	Updates on Virtual probes and section 4.3	0.9
April 08, 2016	Alcardo Alex Barakabitze	Updates on sections 4.1 and 4.2	0.10
April 08, 2016	Saman Zadtootaghaj	Updates on sections 4.1 and 4.2.4	0.11
Aril 13, 2016	Elisavet Grigoriou	Update references in Section 4.3 and Update References Section	0.12
Apr. 13, 2016	Werner Robitza	Accept changes from updates, minor copyediting, comments	0.13
Apr. 13, 2016	Elisavet Grigoriou	Section 4.3 – Update table`s structure	0.14
Apr. 18, 2016	Saman Zadtootaghaj	Update Conferencing and gaming sections based on	0.15
Apr. 18, 2016	Elisavet Grigoriou	Section 4.3 Update	0.16
Apr. 19, 2016	Avsar Asan	Section 8.1.2 and minor comments added	0.17
Apr. 19, 2016	Saman Zadtootaghaj	Updates on section 9	0.18
Apr.19, 2016	Alcardo Alex Barakabitze	Updates on section 4.1 and added some comments	0.19
May 2, 2016	Werner Robitza	Accept changes, minor copyediting, additions to some of the later sections	0.20
May 2, 2016	Elisavet Grigoriou	Updates on section 4.3	0.21
May 3, 2016	Alcardo Alex Barakabitze	Updates on section 4.1	0.22
May 4, 2016	Saman Zadtootaghaj	Updates on Table 4 and comments	0.23
May 4, 2016	Saman Zadtootaghaj	Minor changes based on the meeting	0.24
May 5, 2016	Werner Robitza	Changes based on discussion w.r.t structure	0.25
May 10, 2016	Alcardo Alex Barakabitze	Minor changes based on comments from the meeting	0.26
May 12,2016	Avsar Asan	Additions on Section 4.2.2	0.27
May 13, 2016	Werner Robitza	Copyediting, additions in section 4	1.00
June 13, 2016	Saman Zadtootaghaj	Updates based on comments	1.01
June 13, 2016	Werner Robitza	Updates based on comments	1.02
June 14,2016	Alcardo Alex Barakabitze	Updates on sec.4.1 based on comments	1.03
June 16, 2016	Avsar Asan	Updates and edits based on comments	1.04
June 16, 2016	Saman Zadtootaghaj	Updates on introduction	1.05
June 16, 2016	Werner Robitza	Copyediting, additions in section 4	1.06
June 16, 2016	Avsar Asan	Updates based on the comments	1.07
June 17, 2016	Werner Robitza	Copyediting	2.00
June 21, 2016	Werner Robitza	Final version	3.00

Table of Contents

Table of Figures

Executive summary

This report provides an overview of current approaches for Quality of Service (QoS) and Quality of Experience (QoE) monitoring, focusing on QoE monitoring. It will first explain the different quality measurement approaches, from a subjective and instrumental (objective) perspective, addressing various types of quality models. It describes how the different quality measurement techniques can be classified according to whether they use active or passive probes, whether they are client-centric or network-centric, whether the system is real or virtualized, and what the different points of quality monitoring in the network are. It then focuses on the state of the art of QoS/QoE monitoring techniques by giving an overview of representative approaches for voice and telephony, IPTV and Video on Demand, conferencing, and mobile gaming. Finally, it discusses the future challenges for quality monitoring, based on a rapid growth of the Internet and the increased diversity in the services sent over the network.

Abbreviations

Term	**Meaning**
ARQ	Automatic Repeat request
DPI	Deep Packet Inspection
EADT	Effective Average Downlink Throughput
FEC	Forward Error Correction
FF / RR/ NR	Full Reference / Reduce Reference / No Reference
FPS	First-Person Shooter
FTP	File Transfer Protocol
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
HTTPS	Hypertext Transfer Protocol Secure
ISP	Internet Service Provider
IPTV	Internet Protocol television
KPI	Key Performance Indicator
KQI	Key Quality Indicator
MMS	Multimedia Messaging Service
NFV	Network Function Virtualization
NSA	National Security Agency
OTT	Over the Top

PSTN	Public Switched Telephone Network
QoS	Quality of Service
QoE	Quality of Experience
RNC	Radio Network Controller
RTP	Real-time Transport Protocol
RTT	Round-Trip delay Time
SDN	Software-Defined Networking
SGSN	Serving GPRS Support Node
SIN	Specification Implementation Note
SLA	Service-Level-Agreement
SRTP	Secure Real-time Transport Protocol
SSL	Secure Sockets Layer
TCP	Transmission Control Protocol
TLS	Transport Layer Security
UDP	User Datagram Protocol
VoD	Video on Demand
VoIP	Voice over IP
WAP	Wireless Application Protocol

1 Introduction

Nowadays, most Internet Service Providers (ISPs) are interested to see whether or not customers are delighted with the services being provided to them. It is expected that users of a particular service tend to stick to a provider when their QoE expectations have been met. Therefore, in order to meet customers' expectations, service providers need to permanently measure the current level of service quality, which we call a quality monitoring process. Here, we can distinguish between Quality of Service (QoS) and Quality of Experience (QoE) monitoring. Quality of Service (QoS) is defined as the "totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service" [1]. Quality of Experience (QoE) is the "degree of delight or annoyance of the user of an application or service" [2]. QoE is a multidimensional concept which requires a consideration of several factors (system, human and context factors, see [3]) for different applications, in terms of measuring, managing and controlling QoE from both the end user's and service provider's perspectives.

Basic monitoring methods measure QoS-related network level parameters like packet loss, jitter and throughput. Key Performance Indicators (KPIs) – such as delay, packet loss, and throughput –, and corresponding thresholds for those KPIs are typically defined by network and service operators. Those KPIs are continuously monitored to obtain a current level of quality for a given service. These basic monitoring tools do not explicitly consider the users' perception, so they can be considered technology-centric rather than user-centric. Several factors impact the users' perceived quality which is not fully addressed by QoS concepts. These factors include user characteristics and other context factors such as the location or time a service is being used in. Furthermore, QoS typically focuses on telecommunications services and deals with performance aspects of physical system, whereas QoE covers a much broader domain which deals with a user's assessment of system performance considering context factors and users expectations. Therefore, QoE can never be fully predicted from QoS measurements. Appropriate monitoring tools should not only consider QoS factors but also include QoE factors to obtain better performance (in the sense of correlation with human opinions). Moreover, monitoring tools are often only developed for one specific service, which makes it challenging to develop a "universal" QoE monitoring tool.

With the number of multimedia providers and the diversity of applications and services increasing, endeavors to improve the quality and to provide better services have increased, from the perspective of the network as well as service providers. To meet the increasing user expectations, the concept of Quality of Experience has been increasingly considered in quality monitoring approaches. Accordingly, QoS–QoE mapping is a technique to predict perceived user quality from QoS factors measurable on the network or the client device. In this approach, a mapping between KPIs and key quality indicators (KQIs, for example, service availability, reliability, usability, etc.) would provide an estimate of the perceived quality for the typical service user. Several approaches have been used to find relationships between QoS indicators and human quality ratings obtained from subjective user tests. These include regression, statistical analyses, machine learning or crowdsourcing [4].

In recent approaches, the focus of monitoring tools has shifted: It was understood that including contextual factors would better address QoE concepts and result in more accurate predictions for the user perceived quality. The factors could, for example, include the user's location, their previous experience and expectations, the device types, etc.

In this report we will give an overview of different monitoring tools for diverse application scenarios and content types. In Section 2, we provide an overview of subjective and objective quality assessment methods and their classifications. Section 3 describes specific aspects to be considered for QoS/QoE monitoring and Section 4 provides currently available tools for QoE monitoring. In Section 5, we give an outlook into the future of QoS/QoE monitoring.

2 Multimedia Quality of Experience Assessment Techniques

In general, multimedia quality assessment has to be performed principally subjectively. Given the fact that human visual and auditory systems are the ultimate quality assessors of the multimedia content, subjective quality assessment can be said as the reference way to measure quality of users. So-called objective quality assessment tools try to estimate results of a subjective assessment. The subjective assessment results serve as benchmark for objective assessment. However, subjective assessment is usually expensive and time-consuming. Objective (instrumental) quality assessment instead seeks to replace human judgments by using *models* that can automatically predict quality. Objective assessment is also in compliance with the principle of reproducibility and scalability.

2.1 Subjective Quality Assessment

Typical subjective quality assessment procedures have been presented in the QoE-Net Deliverable D2.1 [5]. The reader is referred to this document for further information.

2.2 Objective/Instrumental Quality Assessment

We can categorize objective quality assessment techniques according to different aspects [6].

The first classification is based on the amount of source signal information that is needed to run the instrumental assessment. In this classification, we define the following three model types:

- **No-reference models (NR)** have no knowledge of the original signal before transmission.
- **Reduced-reference models (RR)** use some extracted features of the source signal and predict the quality by combining this information with measurements of the received signal.
- **Full-reference models (FR)** have full access to the original signal.

This classification can be seen in Figure 1.

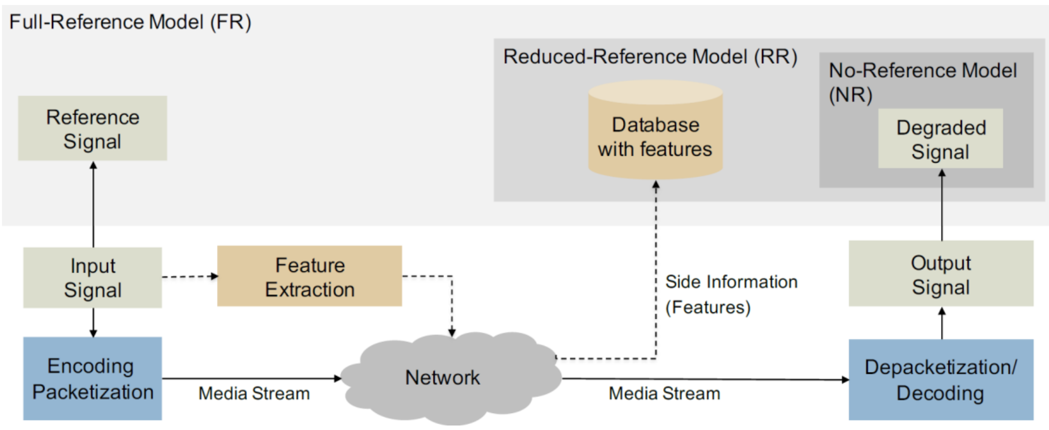

Figure 1 Classification of the objective quality assessment techniques according to information available from source.

In another approach, we can classify the objective quality assessment according to the level of information used. Note that in this classification, all types of models (NR, RR, FR) can be contained:

- **Planning model**: These models take assumed network and client parameters as input and calculate an expected quality score.
- **Bitstream header-based model**: Quality can be predicted by packet-header information (e.g., HTTP or RTP) without accessing the payload information. Therefore, these models do not explicitly consider source and encoding information to evaluate the quality. However, these models try to

predict the quality by considering coding and IP network impairments. Examples of such models include the ITU-T P.1201 recommendation series.

- **Bitstream payload-based model**: Quality prediction is not only achieved by application and transport layer information (e.g. HTTP/TCP or RTP/UDP) but also by extracting and analyzing content features from the coded bitstream. An example of this case is the ITU-T P.1202 family of recommendations.
- **Signal-based model:** Quality is determined by analyzing the decoded media signal.
- **Hybrid model**: In hybrid models, decoded signals are combined with information from the bitstream or the packet layer.

The above classification can be seen in Figure 2.

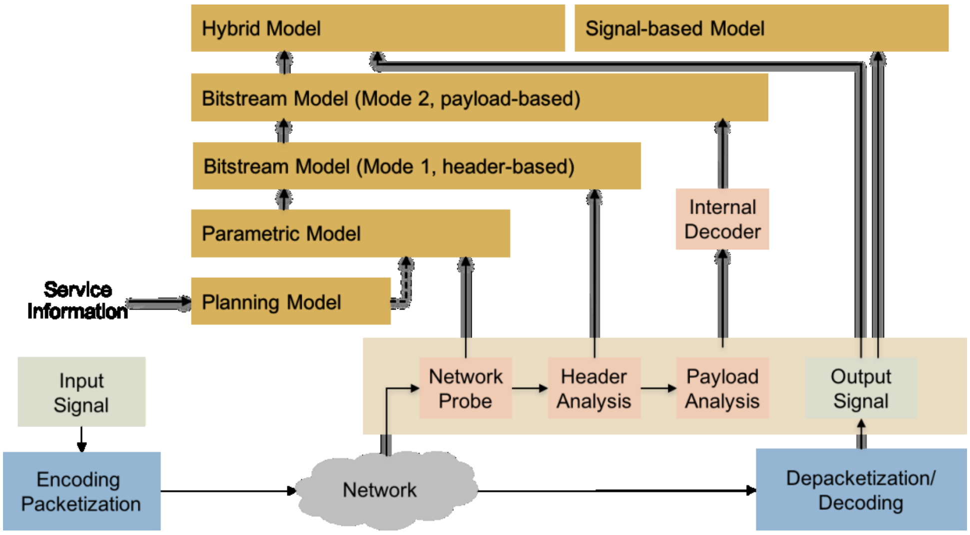

Figure 2 Overview of objective quality assessment models according to different levels of information extracted from the media stream .

These classifications are important for QoS/QoE monitoring since the possibilities for implementation of monitoring systems heavily depend on the amount of information that is available. For example, the use of a full-reference model for large-scale monitoring of an IPTV system would be practically unfeasible, as the amount of information that would have to be transmitted over the network from source to monitoring points would be too large.

3 Main Features of QoS/QoE Monitoring Tools

QoS/QoE measurement can be performed in three distinct stages: First, operators can predict quality before a service is developed. Here, planning models (as described in Section 2.2) are typically used. Strictly speaking, this is not a case of *monitoring*, as there is no real service to continuously monitor yet. However, planning models can still be used in a later phase. Second, after the service is deployed, its performance can be tested before delivering content to real users. Finally, after a service is deployed and content is delivered through it, continuous quality monitoring can take place, based on real user data.

In QoS/QoE monitoring solutions, data from client devices or network elements is collected and can be used in quality assessment procedure. This data is captured by so-called *probes*. Probes are devices that are used to collect data from different points in the delivery chain. When developing a QoS/QoE monitoring system, several aspects have to be considered related to: the time at which quality is measured, whether the probes simulate a connection or just measure traffic (Section 3.1), the placement of probes in the network (Section 3.2), and whether the monitoring takes place on real or virtualized devices (Section 3.3). Also, the choices made with respect to the previously mentioned aspects will influence which data is sent to models as input. This again determines the precision or efficiency of the quality predictions that can be made. The quality calculation point (Section 3.4) also has implications on the architecture of a QoS/QoE monitoring system.

3.1 Active vs. Passive Monitoring

There are two basic categories of probes used in monitoring systems, as described in the following.

3.1.1 Active Probes

Active probes – as the name implies – actively initiate a service connection or send and receive data to measure quality for a certain time frame. For example, an active probe could be implemented to set up a test VoIP phone call to another active probe. Or, it could be a computer using a browser to fetch a website, in order to measure its download time. With active probes, no real users are involved in the measurement process. Instead, they are often built in such a way that they resemble the user scenario as closely as possible.

Active probes are often very complex in the sense that they have full access to the communication data and transmitted signals. Video service probes for example may store the entire video signal transmission for analysis. Due to the amount of information captured by probes, they can make fairly precise predictions about the quality compared to what a user would have experienced in case they had been using the service at that specific time. Furthermore, since the measurement parameters are exactly known, the negative influence of unexpected factors can be limited and even avoided. Beside all these advantages of active probes, due to their complexity, using active probes for monitoring purpose are mostly more difficult than passive probes in terms of set up and maintain. Moreover, active probes cannot be placed at every possible location in the network. Instead, an operator has to choose representative points in their network to place these probes at, in order to have a good overview of the quality for their entire customer base.

3.1.2 Passive Probes

Passive probes passively measure network traffic passing through the network. They are placed all along the delivery chain, at the server side, in core networks, in the access network, in the home network or at the user's terminal. . When they are used in network-centric locations (see also Section 3.2), passive probes can potentially measure a large number of parallel streams. In a simple form, passive probes can measure KPIs like jitter, packet count, packet loss for any given stream. However, they can also make more fine-grained analyses related to higher levels of data, e.g., HTTP contents. This is done using Deep Packet Inspection (DPI).

Passive probes can also be used in combination with other (active) test systems. For example, a passive probe could be placed between a user's modem and an (active) IPTV set top box to calculate quality scores for the TV transmission, without interfering with the traffic. In such a case, the passive probe would perform quality calculations that could not be implemented on the (active) device itself, for example because its software cannot be changed or its software platform does not allow for capturing network data.

Since they often need to monitor more than one stream at the same time, passive probes dos not typically provide the most in-depth analyses. There is a trade-off between accuracy and number of measurements that can be performed, since practically, not every packet stream can be inspected at a certain depth.

3.2 Network-Centric and Client-Side Approaches

With regard to monitoring points, probes can be categorized into two main classes. The first category is network-centric probes. These probes in the form of network appliances are either deployed on network elements such as routers or switches, or come in the form of dedicated hardware attached to other network elements. The second class is the client-side probes. These probes can be software agents deployed on user devices or dedicated devices placed at user locations. What differentiates network-centric from client-side approaches therefore is the location in the delivery chain. Typically, client-side probes describe measurement points at or behind a user's home gateway.

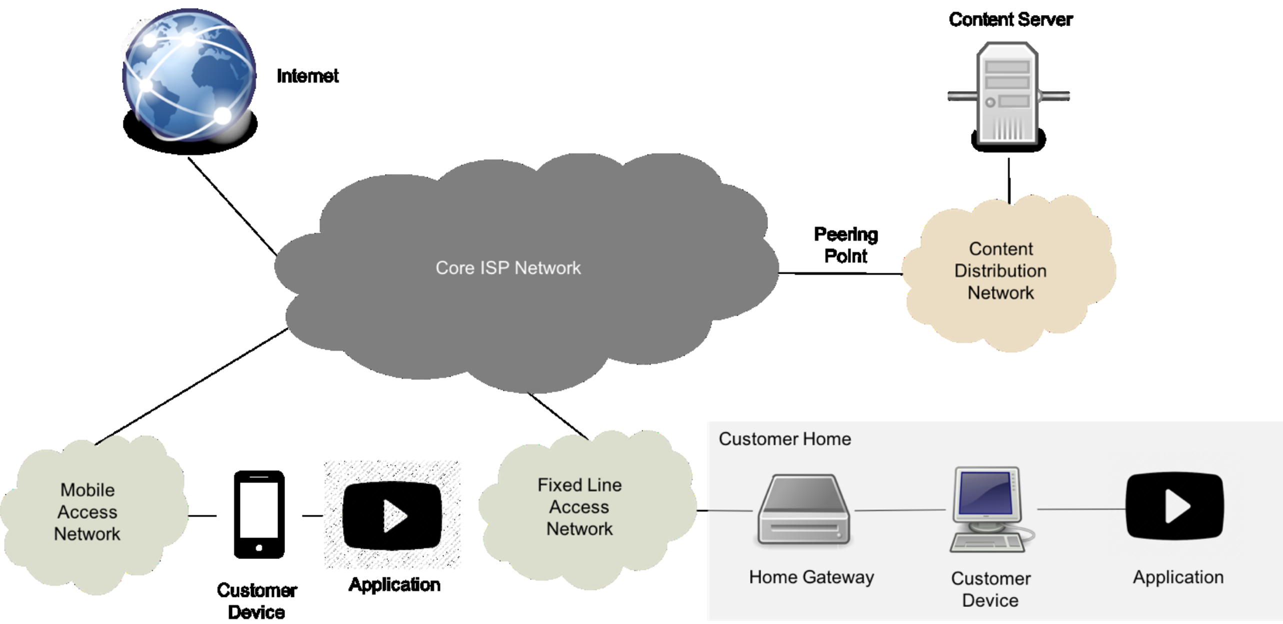

Figure 3 Network schema with possible probe locations.

In Figure 3, a simplified scheme of a network is shown. Here we can see the possible locations for placing probes, such as the CDN, the core ISP network, the access network, or at the customer's home. In the latter case, we would refer to client-side probes.

Furthermore, we can distinguish probes according to who owns them, or where the quality is monitored (see also Section 3.4):

- **Content-provider domain:** The supplier of the content, that is, the over-the-top provider (e.g. *YouTube*)
- **Service-provider domain:** The supplier of the service itself, which could be a Content Distribution Network provider (e.g. *Akamai*)
- **Network-provider domain:** Typically the ISP itself, that is, the supplier of the access network

The placement of probes touches two important aspects in quality monitoring scenarios:

- **Privacy vs. Context Data:** Client-side probes, especially when implemented in real user devices, have access to much more information than what is available in transmitted network streams. For example, a mobile device knows about the user's location and surroundings, although it is not necessarily visible on the transmitted network data. Naturally, privacy concerns limit monitoring systems from collecting all user context data, although it would be required for a more accurate prediction of QoE for any given service. But even in network-centric aspects, privacy concerns play a large role.
- **Encryption:** When probes are placed on core network elements, they are typically blind to certain application data that is transmitted over encrypted channels. Encryption (e.g., in the form of HTTPS or SRTP) makes any data sent over the TCP level inaccessible to the network provider, since the secure data channel is set up between the server and client end points. Therefore – also given the increase in the adoption of encryption over the last years – network-centric probing approaches will become increasingly blind to the data being transmitted. For some use cases such as video or audio transmission, missing knowledge about used codecs or lost packets will make it virtually impossible to provide accurate quality predictions.

Given the above concerns, a combination of active and passive as well as client-side and network-centric probes seems the most promising for a quality monitoring system that delivers accurate predictions for both real users of the service and a general overview of service quality from sampling points in the network.

3.3 Real and Virtualized Monitoring

Recently, endeavors to simplify network management have increased, with two main innovations being Software-Defined Networks (SDN) and Network Function Virtualization (NFV).

The main goal of SDNs is to decouple the network control and data planes which brings programmability and centralization control to computer networks. Therefore, it simplifies the network management. Nowadays, SDN has attracted significant attention from both academia and industry. ForCES [7] and OpenFlow [8] are the two main SDN technologies which are implemented and brought SDN from the academia to the actual market.

NFV is a novel procedure to transfer network functions from hardware devices to software applications. Therefore, the goal of this concept is to decouple network function from hardware appliances without an impact on functionality. In the field of performance improvement and network simplification, this concept is highly complementary to SDN. These two topics are independent from each other but are mutually beneficial.

With current network technologies, we can use virtual probes for monitoring and testing of applications and networks. A virtual probe is a test entity that enables addressing various aspects of performance management across the life cycle of business-critical service delivery using NFV, cloud- and SDN-based infrastructure. The virtual probes are able to work as virtualized test functions on any virtualization-capable network element (such as routers and switches). The use of virtual probes is reducing the cost of the hardware and the complexity of the network. Furthermore, they are flexible, thus they can be placed in virtual environments easily unlike the real probes. One important advantage of these is that via a virtual probe, all the interfaces are visible to it. Thus, it is able to extract the appropriate information for customer experience management. The network can be provisioned efficiently and the customer experience can be improved.

3.4 Quality Monitoring Points

In QoS/QoE monitoring systems we can distinguish between the points where quality is monitored and the points where quality is calculated [9]. When developing a QoS/QoE monitoring architecture, it is important to define data acquisition points and calculation points. This decision is based on a trade-off between the amount of information available for calculation and the requirements on the performance of the used systems.

For example, when measuring video quality with a no-reference or reduced-reference model, the quality calculation point should coincide with the measurement point itself, since the video data will only be available at the end device and sending the video stream back to a central monitoring location would practically be consuming too much traffic or take too much time. However, in this case, only the quality scores (e.g., measured in MOS) and a number of KPIs would be sent to the monitoring system. This would make it difficult to gain insight into specific degradations or perform a post-fault analysis, since the underlying data to calculate the quality score is not centrally available.

In [10] there are five specific monitoring points defined. While they have been developed for IPTV and VoD applications, they can also be generally applied to any other service that is sent from a content provider to a user at home (e.g., web browsing, VoIP, teleconferencing). The points are the following:

1. **Monitoring Point 1 (PT1)**: Located at the border between service provider (ISP) and content provider (OTT). The fundamental objective of monitoring at this point depends on the application and service. For example, for IPTV, the ISP wants to assure that the video streams are generated with the correct policy (encoding, QoS parameters, bitrate, codec etc.) and without any errors. For other applications such as VoD content delivered through CDNs, the ISP could check that the peering route is optimally chosen and that there are no congestions. It would be wise for the ISPs to already monitor the quality or try to predict quality problems at this point. This would help the ISPs in preventing problems further down the transmission chain, which would require a more sophisticated troubleshooting or monitoring closer to the user-end, which is much more time-consuming and expensive in practice.
2. **Monitoring Point 2 (PT2):** This point marks the boundary between what is identified in the Figure below as “service provider” and “network provider”. In practice, these are usually the same entities. However, for some services like IPTV and VoIP, the “service” is a dedicated application running under the ISP’s control, which is differentiated from the underlying core network. In the case of IPTV, PT2 supplies information on video parameters before they are sent over the core network. Such parameters then can be used to verify the video stream quality from the head-end. Generally, through PT2 ISPs can gather information on quality and performance parameters of streaming media and check service-related policies.
3. **Monitoring Point 3 (PT3):** Defines the boundary between IP core and access networks. For instance, this is where the multicast should be considered in IPTV. The purpose of monitoring at this point is to measure IP network related parameters such as: mean one-way delay, packet delay variation, transmission codecs, packet loss ratio and profile, path availability and multicast IP

performance parameters (e.g., successful join time, group mean one-way delay, group IP service availability and mean group loss ratio) as specified and detailed in [11] and [12].

4. **Monitoring Point 4 (PT4):** This point is located between the IP access network and end-user domain. Monitoring at this point helps to get information on packet loss ratio, synchronization errors of audiovisual content, IP network parameters and reliable delivery. This is the last point at which service providers (ISP) can monitor data that is as close as possible to what end users receive, unless ISPs have access to end-user equipment.
5. **Monitoring Point 5 (PT5):** The so called “QoE point” is located at the end-user equipment. The data obtained here can be used to evaluate the user quality of experience directly. In this point user-end equipment- or application-specific information can be gathered.

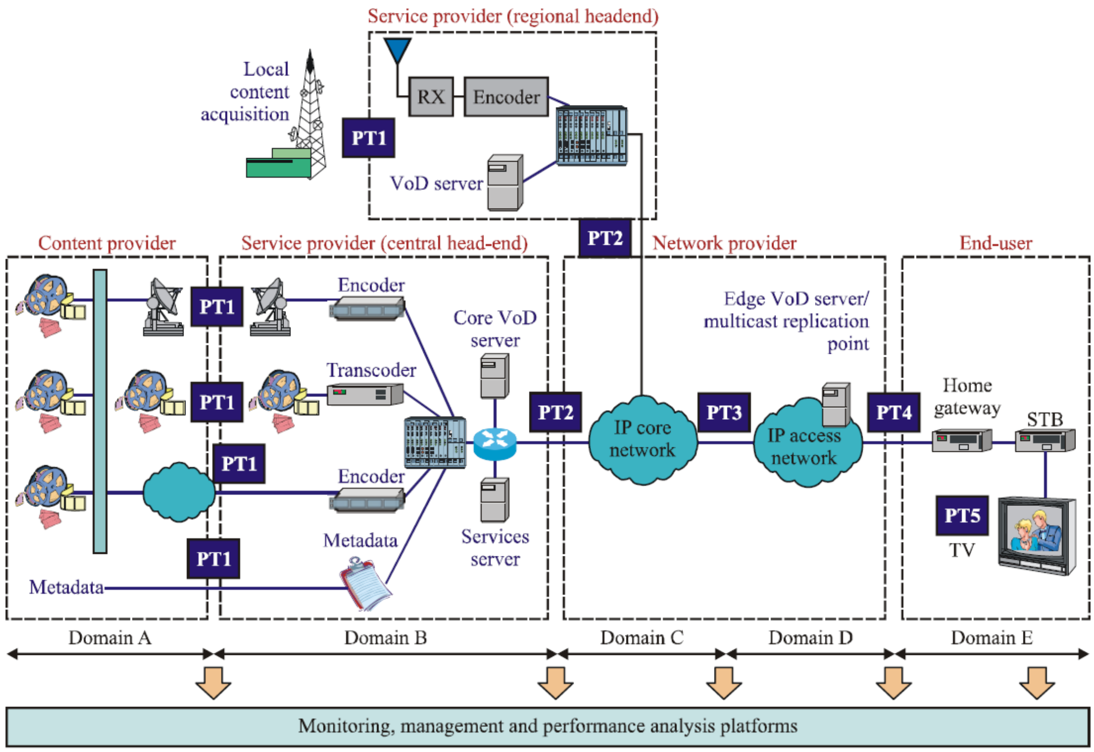

Figure 4 Network Measuring points in path from service provider to end-user [10]

4 Current QoS/QoE Monitoring Solutions

In this section, we give a representative overview of the state of the art of QoS/QoE monitoring tools, with example solutions for quality monitoring tasks that apply to the common performance monitoring scenarios for specific applications, telephony (VoIP), IPTV, Video on Demand), conferencing, and gaming. Those are based on either industry standards or recent publications from academic or industrial research.

We base our classification on applications instead of modalities, since different applications call for different service architectures, which in turn require different kinds of monitoring tools (i.e., a combination of unique choices with respect to the classification given in Section 3). In other words, for example, there are no dedicated monitoring tools for audio, but instead we find solutions for telephony (full-reference approaches) as well as the audio component of video streaming (no-reference approaches).

4.1 Network-Level QoS Tools

This section describes QoS monitoring tools and methods from network level perspective. We provide overview of network/transport monitoring tools and related performance parameters which have an influence on QoS. However, there is still a need for determining those KPIs too, for example in order to find faulty network equipment or global network issues. In fact, network monitoring tools helps in the overall understanding of data flow in the network, avoid network bandwidth bottlenecks and optimize the network performance which results into an improved QoS and QoE for the end users.

4.1.1 Network Parameters that influence QoS

QoS is the measure and the determinant metric of service availability and transmission quality of a network from a technological perspective. The network transmission quality is mainly determined by the following factors: network delay/latency, available bandwidth, jitter, packet loss, error rates and network throughput [12]. Generally, any QoS degradation may result in low QoE for the applications running on the network. We describe in brief several factors/parameters which have an influence on QoS/QoE at different aspects of network/transport and physical level.

Jitter/ Packet Delay Variation describe the time difference in packet inter-arrival time to their destination. It is the variation of the delay of received packets. The delay between transmitted packets can vary due to improper queuing, network congestion or configuration errors. In VoIP technologies, jitter refers to a delay in receiving voice data packets which affects the transmission of voice quality and voice data. In order to ensure a successful transmission of voice/video data, a jitter buffer is used to eliminate the effect of jitter in video/audio signals transmitted over a network.

Packet Loss is mainly caused by congestion as a result of insufficient or non-optimal usage of network resources. For a connection-less transport protocol such as UDP, the loss may directly affect application performance. The effect of packet loss is usually expressed by packet loss rate (PLR). PLR is defined as the ratio of lost packets during transmission to the total number of transmitted packets. Although, measuring the QoE in a video service such as IPTV has some complexity, a better understanding of the QoE perceived by the end-user is normally done through the perceptual investigation of the packet loss effects on the video/audio quality. For example, different studies have investigated the impact of packet loss on user perceived video quality [13] [14].

Packet loss burstiness is commonly acknowledged as a key significant factor on the performance of networking protocol in wireless networks. This is so because; packet losses over wireless channels typically have strong correlations and occur in a bursty fashion such that their performance affects directly the effectiveness and efficiency of communication [15]. The design of wireless systems for QoS and QoE provisioning to the end users requires an accurate evaluation of the packet loss burstiness, which can reveals the performance and characteristics of the wireless channels.

End-to-end Delay is the time required for transmitting a data packet along its entire path. Delay has a strong impact on the end user's perception video/audio quality and therefore the QoE. The quality of transmitted audio and video is much affected by delay in the network. This effect in turns lowers the end-user's perception on quality of received videos. A delay can either occur before service consumption or during service consumption. Initial delay is the waiting time before service consumption while delay for a live transmission/service interruption is the waiting time occurring during service consumption for a certain time period [16]. In particular, user-perceived quality suffers from initial delays when applications are launched, as well as from freezes during the delivery of the stream. In fact, delays may results in an increased user waiting time for a particular service/application. Since users are normally dissatisfied with a service due to the delays, it is very important to investigate and develop appropriate approaches for minimizing the waiting time or delay which have an impact on end user's QoE.

Available Bandwidth describes the maximal amount of data per unit time that can be transmitted from one node to the other. The access downlink bandwidth has an impact on the QoE of popular applications such as video on demand services when are accessed through smartphones. Measuring and managing bandwidth usage is more essential than ever to assure better subscriber QoE and increase customer retention.

Network throughput describes the amount of useful data that can be transmitted per unit time. In the networking industry, the downlink throughput is the mostly used and an accepted QoS related feature for network performance monitoring and reporting. Focusing on Mobile ISP quality monitoring and reporting systems, the downlink throughput is the relevant metric for determining the performance of mobile networks. Recently, in the QoE research field, the average downlink throughput has been investigated and translated into a measure of user satisfaction for bandwidth-sensitive services such as video streaming and file sharing [17]. In order to improve the end user's QoE especially for video streaming, we need a better understanding of the throughput predictability, stability and efficient throughput estimation algorithms. Evaluation of various modelling approaches that make use of KPIs such as Effective Average Downlink Throughput (EADT) has proven that some metrics that take into account both the duration and intensity of throughput drops perform better in predicting the accuracy of the QoE than models which are only based on average throughput [18]. Table 1 shows various quality influencing factors on the transport/network and physical layers.

Table 1 Network-level KPIs influencing QoE

Layer	Quality influencing factors
Transport and Network	• Packet loss burstiness, packet loss ratio • Jitter, network delay, congestion period, available bandwidth • Latency, network throughput
Physical	• Bit rate, SNR, Packet/Symbol/Bit Error Probability. • Ergodic capacity, throughput, diversity order/coding gain • Outside capacity and probability • Energy efficiency

The QoS influencing factors such as jitter, delay and packet loss can be monitored through traffic monitoring tools that can monitor a continuous traffic on the network which in turn can enable taking necessary actions for ensuring a good network performance and QoE. Figure 5 illustrates an end-to-end QoS/QoE monitoring tools. The user's terminal may support the following tools:

- Application layer tools such as Ping, FTP, HTTP Browsing, MMS, SIN and WAP
- Protocol analyzer for protocol stack performance analysis at any interfaces and field measurement tools for radio measurements.
- Mobile Quality of Service (QoS) agents for performing measurements and monitoring across ISO/OSI layer 1 to layer 7. For example, passive monitoring methods such as connectivity monitoring, performance monitoring, SLAs/ELAs monitoring or probe monitoring can be done at layer 2 to monitor service availability, frame loss, throughput, frame delay variation, frame disordering etc. NetFlow and sFlow tools can be used at layer 3 for monitoring packet loss ratio, inter arrival time, delay and burstiness etc.

Protocol analyzers perform monitoring of delay and packet loss across the Radio Network Controller (RNC), Serving GPRS Support Node (SGSN), Gateway GPRS Support Node (GGSN) and application servers.

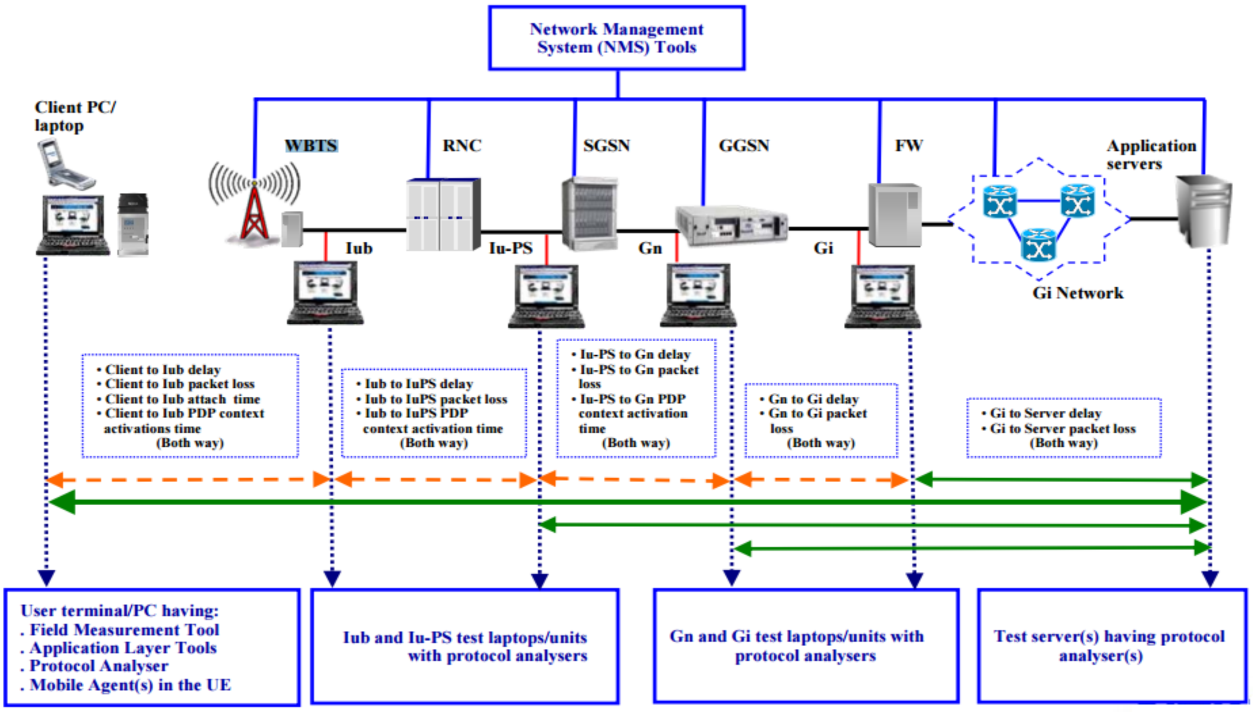

Figure 5 End-to-end QoS/QoE tools [19]

4.2 QoE Monitoring Tools

In contrast to the network-level monitoring tools described in the previous section, this section focuses on QoE monitoring tools instead. Those are typically built for higher-level measurements of QoE, for example using Transport Layer or Application Layer data. Data captured includes, for example, speech signals or video packets. To predict the QoE for the end-user, those monitoring systems typically employ QoE models that use the captured network data.

4.2.1 Voice Telephony

Under the term Voice over IP (VoIP), there exist many technologies to send speech data over existing IP networks rather than classical circuit-switched networks (public switched telephone networks, PSTN). Generally, we can distinguish between

1. services aiming to replace circuit-switched telephony providing reliable QoS
2. over-the-top best effort services that are typically accessed through third-party software (e.g., Skype) and run on a multitude of platforms and devices

A typical VoIP transmission chain consists of a sender and a receiver, connected through a handset, headset or hands-free device, a modem and the transmission network (consisting of access and core networks of one or more ISPs). In a non-fully VoIP-switched network ("full service"), one side of the transmission could still use PSTN equipment, requiring VoIP–PSTN bridges.

VoIP typically builds on UDP-type data transmission (e.g., through RTP) due to its real-time requirements, as TCP transmission would result in data overhead and latencies. UDP – because of its non-reliable transmission mode – naturally incurs transmission problems due to packet loss, delay or jitter, which may manifest in degradations on the voice component. Those problems are usually mitigated by introducing packet buffers (to smooth out jitter) and packet loss concealment methods, although the former introduces a delay component in the transmission chain.

VoIP systems can be monitored both actively and passively, with many providers relying on active measurements for detailed performance and degradation evaluation as well as passive KPI-oriented measurements to assess the overall network state. In the following, we show examples of VoIP monitoring solutions.

A good overview of VoIP monitoring literature can also be found in [20], although only considering background work up to the year 2009. Another paper by Möller et al. gives an overview on speech quality estimation until the year 2011 [21], including different model types. In the following, we will briefly show the most important concepts and tools for VoIP monitoring used in practice.

4.2.1.1 Network Planning with the E-Model (ITU-T G.107)

The E-Model – described in ITU-T Recommendation G.107 [22] – is a planning model that assists network operators to build VoIP systems. An overview of the principles in the E-Model is shown in [23], too. The E-Model outputs a value *R* in the range between 0–100, where 100 is the highest possible quality. The R-score can then be transformed to a MOS value.

According to our classification, it is therefore already used *before* the implementation of a service and therefore not a traditional monitoring model, since it relies on assumptions only. However, it can still be used in practice to estimate the quality of a call, by setting the levels of the individual parameters (as seen in Figure 6) – although this is not the intended application purpose of the model.

An example of using the E-Model for monitoring purposes is given in [24], in which the authors simplified the model and derived coefficients for different speech codecs. They also describe a possible monitoring system, which captures RTP packets at the terminal and processes them to estimate a MOS for a given measurement time. Often, the E-Model is extended to create new models, as shown, for example, in [25].

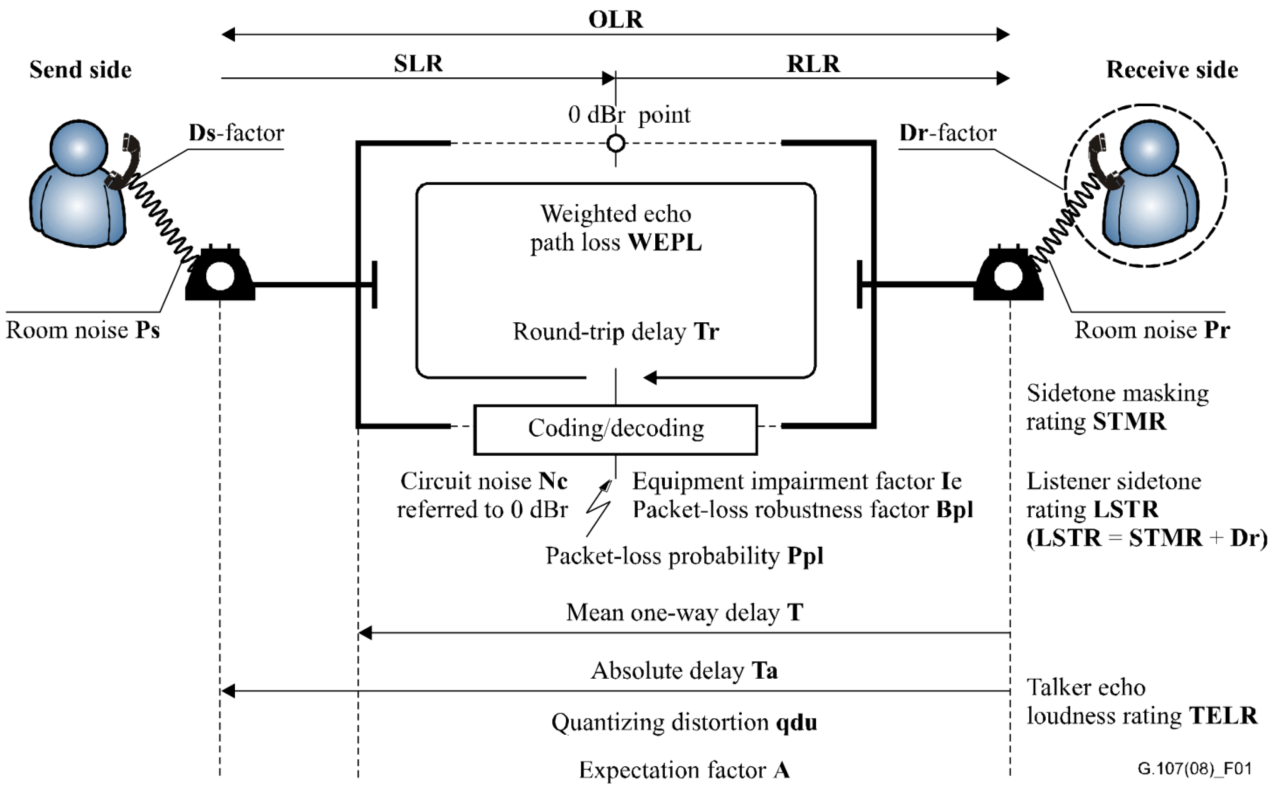

Figure 6 ITU-T Rec. G.107 schema

4.2.1.2 PESQ and POLQA for Active Full-Reference Monitoring

PESQ [26] (Perceptual Evaluation of Speech Quality) and POLQA [27] (Perceptual Objective Listening Quality Assessment) are full-reference models standardized in ITU-T Recommendations P.862 and P.863, respectively. POLQA superseded PESQ, but the latter is still in use today, despite having been standardized more than fifteen years ago and only being valid for narrowband speech transmission (300–3400 Hz). P.862.2 amended PESQ for wideband transmission (50–7000 Hz). POLQA is valid for super wideband transmissions (50–14000 Hz). PESQ and POLQA analyze the quality of a speech transmission on a 5-point Mean Opinion Score scale using an algorithm that compares an original with a degraded signal.

Since both models are FR-based, they require access to the source and received (degraded) signal in order to assess the quality of the transmission. This has practical implications. Since it would be wasteful to always send the source or received signal to the receiver or sender, respectively, or send both signals to another measurement point, those models use predefined source samples that are known to the receiver as well. These samples contain short speech sequences. Therefore, the models are used for *active* speech quality monitoring. They are not used to passively monitor connections of real service users, as it would require not only intercepting signals from the network but also recording the original signal at the user's location.

PESQ and POLQA are available on the market in various forms.[1] Implemented as software, the program could simply calculate the MOS of two recorded audio samples (original and degraded). This may be a

[1] A list of licensees and vendors can be found on: http://www.polqa.info/licensees.html

useful scenario for the development of services, simulation and lab testing, where access to the samples is easy. For use in real VoIP services, the software however should be implemented on probes that actively initiate connections to other probes, reporting the quality (and KPIs) to a central monitoring location.

These probes typically sit behind customer-equivalent modems or telephone connections to give a more customer-centric view. From the perspective of the operator, they should be placed in strategically important locations and run on a regular schedule (e.g., every half hour) so as to identify possible issues with the network.

The organization of probes and measurements is part of monitoring systems, which themselves are not standardized. Numerous monitoring systems based on PESQ and POLQA are used in the industry today.

4.2.1.3 No-Reference VoIP Quality Assessment

For NR / single-ended quality monitoring, a standardized model only exists for the narrowband case, as ITU-T P.563 [28]. P.563 works by dividing the quality analysis into three functional blocks: 1) vocal tract analysis and unnaturalness of speech, 2) analysis of additional noise, 3) interruptions, mutes and time clipping. Internally, a number of parameters is defined, which are then used for classifying which distortions are present in the signal. After that, an overall speech quality measure is calculated.

4.2.2 IPTV and Video on Demand

As multimedia services consumption is becoming more popular, solid standardization efforts have been made to ensure the service quality through proper measurement techniques. In this section, IPTV and VoD services, their differences, standardized measurement and monitoring methods in ISP and OTT perspectives will be explained.

4.2.2.1 The Difference between IPTV and VoD, the Role of Network Protocols

IPTV and VoD differ in terms of the underlying transport protocols, which influence the way in which the services are built as well as the perceivable problems on the user side in case of network degradation. Since IPTV typically transmits live video and is sent simultaneously to a large number of users, it uses the UDP based multicast protocol, which is not reliable in the sense that lost packets will not be automatically retransmitted. VoD services on the other hand nowadays typically use the reliable TCP communication protocols. In the case of lost or delayed packets, degradations in UDP-based services are typically in the domain of visual artifacts (slicing errors, mosaicking), whereas with TCP-based services – which usually employ a longer buffer at the client – users are more likely to see coding artefacts and stalling events (stalling of freezing shows up when the buffer is running empty).

Therefore, the perceptual distortions visible to the user are quite different in the two cases. Considering this distinction, it is evident that the employed QoE models will also differ in the way they consider visual artifacts; they will also use different application-specific parameters and define different internal quality indicators. For example, in IPTV, packet loss leads to video slices being lost, which results in degradations being carried over multiple frames until the next complete intra-coded frame is sent. In VoD, packet loss as an indicator is used completely differently: packet loss over time only results in longer buffering periods. Because of all the above-mentioned differences, a monitoring solution for a VoD system is not necessarily usable for IPTV and vice-versa, both in terms of monitored parameters as well as internal models.

4.2.2.2 Standardization in ITU-T: Towards Monitoring Models for IPTV and VoD Services

With the ever-increasing competition on the market for both VoD and IPTV services, operators tend to become more and more publicly compared each other according to their service quality. In order to enable a fair comparison of two services, however, it is vital that the performance measurements done use the same underlying metrics. Standardization is an important factor that contributes to a common understanding of "quality", and enables operators to use methods and models that have been extensively validated.

On the quality measurement and estimation of IPTV and VoD services, standardized models are developed by ITU-T study groups 9 and 12. For this reason the following subsections include an overview of the relevant ITU-T standards, models and their usage areas on such services. ITU-T models cover TCP and UDP-based multimedia streaming, and current ongoing work focuses on HTTP adaptive streaming (e.g. in the ITU-T "P.NATS" work item).

The approaches that are used in standardized models are often directed by the nature of the services themselves, which have a strong influence on how monitoring systems are built around these standardized models (which themselves do not prescribe a specific monitoring architecture). Accordingly, for IPTV and VoD services, typically, a no-reference method is used, as for such time-critical applications – in which a reference video is not available – the quality prediction is obtained only based on the transmitted sequences.

4.2.2.3 Video on Demand Monitoring Techniques – Perspectives from OTTs and ISPs

There are two different perspectives in VoD monitoring: the OTT and the ISP perspective. The reason is found in the way Video on Demand services work: the clients of the users (apps or players in browsers) establish direct, dedicated connections to the OTT's content servers, only using the ISP's network for transport. ISPs typically do not have access to the original video signals that OTTs send to their customers. For VoD, the distinction *who* is monitoring the quality therefore plays a larger role:

4.2.2.3.1 The OTT Perspective

An Over the Top (OTT) video service provider can typically easily gather KPIs from its clients (such as loading times and quality switches) and has access to the original video as well as all of its representations (in the case of HTTP adaptive streaming). This makes monitoring the quality for individual users very easy, with the only limitation being computing power (e.g. for calculating full-reference quality model scores) or user privacy concerns. However, OTTs have little to no influence on how ISPs route their packets. In fact, most OTTs may choose CDNs to distribute their files, so they hand off caching and file distribution to another entity. Therefore, quality degradations observed at the user end may not even be the OTT's fault, but rather a problem of the CDN, the ISP, or the peering between them, which makes a diagnosis (and correction) of the underlying problems hard.

4.2.2.3.2 The ISP Perspective

ISPs are limited to gathering network traces from which they filter out VoD streams traveling over their network, typically over HTTP. Assuming that those streams are unencrypted (sent via HTTP), ISPs could employ buffer models and adaptive quality switching models to predict how a certain network transmission would translate into perceived user issues (e.g. in the form of stalling events). This method is not 100% accurate, but can give insights into the current level of quality for a given customer and VoD service provider. However, with the recent trend of encrypting their streams (through HTTPS) to increase user privacy, VoD providers make it impossible for ISPs to access to the previously mentioned information. Instead, ISPs can only monitor throughput and traffic patterns.

Currently, no standardized monitoring solutions exist that would allow to make any judgment on the experienced quality by the user when the OTT video streams are fully encrypted. Therefore, ISPs have to resort to using user-centric, active probes, which enable them to access all the information a user would have. This approach requires a large number of probes to enable a reliable view on the current network status. It also calls for a correlation of network-centric parameters with the results obtained from the active monitoring systems.

4.2.2.3.3 VoD Monitoring with ITU Standards

Currently, no ITU-T standard is dedicated to VoD monitoring itself, apart from P.1201 Amendment 2, App. III [29]. It is an extension to the IPTV-based standard P.1201. It allows to calculate a quality score for audiovisual transmissions as well as a degradation score for initial loading and rebuffering events, which are the most important quality factors for progressive download video services.

The ongoing ITU-T Study Group 12, Question 14 work item “P.NATS” (Parametric non-intrusive assessment of TCP-based multimedia streaming quality, considering adaptive streaming) will standardize a first model for the assessment of HTTP adaptive streaming services in 2016.

4.2.2.3.4 Commercial VoD Monitoring Solutions

Several commercial solutions for monitoring VoD exist on the market, not specific to any VoD provider. Those are typically aimed at ISPs wanting to assess the quality for several VoD providers in their network. They are also used for benchmarking tests, in which different ISPs are compared against each other.

Commercial solutions include products marketed by Avvasi, Opticom, Rohde & Schwarz (SwissQual), and others.

4.2.2.4 IPTV Monitoring with ITU-T P.1201 and P.1202

The main ITU recommendations related to IPTV monitoring are ITU-T P.1201 [30] and P.1202 [31]. P.1201 includes models for non-intrusive monitoring of audio, video and audiovisual services based on packet header information analysis, whereas P.1202 includes bit stream-based models only for video services. The gain of exploiting bit stream information of the sequences such as coding-related information is that the quality estimations are more accurate nevertheless more complex and require more computational power. Table 2 gives a comparative overview of the recommendations and correspondent models.

The P.1201 models are very lightweight and can be implemented on set top boxes or dedicated passive monitoring devices. However, the move to “Cloud IPTV services”, in which there is no need for a dedicated set top box anymore, the quality models could be built into the IPTV software that runs on general-purpose hardware (e.g. on a Smart TV). In the latter case, a network-centric monitoring is also possible, e.g. with the deployment of probes in the access network to capture multiple IPTV streams at the same time. Here, scores could be gathered continuously over a long period of time, with a central monitoring system evaluating the MOS output and additional network-level KPIs such as packet loss.

Reference	Modules	Supported codecs	Application	Usage	Method	Parameters	Calculations / Models
P. 1201	P. 1201.1	H.264 MPEG-I Layer II MPEG-2 AAC-LC MPEG-4 HE-AACv2 MPEG-4 HE-AACv3	Lower Resolution (Mobile TV)	audio	Packet-based	audio bitrate (A_BR) received RTP audio packets (A_RP) lost RTP audio packets (J)	$A\_BR = \frac{8 \; x \; 10^{-3}}{A\_MT}\left(\sum_{i=1}^{A\_RP} A\_receivedBtye_i + \sum_{j=1}^{J} A\_lostBytes_j\right)$
				video		video bitrate (V_BR) video frame type (V _TVF) video measurement time (V_MT) total number of bytes (TBpF)	$V\_BR = \frac{\sum_{i=1}^{V\_TVF} V\_TBpF_i}{V\_MT}$
				audiovisual		Audiovisual quality (AV _MOSC) Visual quality (V _MOSC) Audio quality (A _MOSC)	$AV\_MOSC = av1 \; x \; V\_MOSC + av2x \; A\_MOSC + av3 \; x \; V\_MOSC \; x \; A\_MOSC + av4$
	P. 1201.2		Higher Resolution (IPTV)	audio		Frameloss (FRL) Burstiness (BS) Overall est. audio quality (QcodA) Quality of Audio (QA)	$QA = 100 - QcodA - ((b1A - QcodA)x \frac{FRL}{FRL + b2A \; x \; BS + b3A})$
				video		Overall est. video quality (QcodV) Quality of Video (QV)	$QV = 100 - QcodV - (c1V \; x \; log\left(c2V \; x \; \frac{Lossmagnitude}{QcodVn}\right) + 1)$
				audiovisual		Est. audiovisual quality (QQAV) Quality of Video (QV) Quality of Video (QA)	$QQAV = \alpha + \beta \; x \; QV + \Upsilon \; x \; QA \; x \; QV$
P. 1202	P. 1202.1	H.264 MPEG-4 Part2 AMR-WB AC3 AAC-LC MPEG-4 HE-AACv1 MPEG-4 HE-AACv2	Lower Resolution (Mobile TV)	video	Bit-Stream based	Combined quality value (d_cv) d[0], compression_artifact_value d[1], d_slicing_artifact_value d[2], d_freezing_artifact_value d[3] d_rebuffering_artifact_value	$d\_cv = \alpha[0] * dpp[0] + \alpha[1] * dpp[1] + \alpha[2]$
	P. 1202.2		Higher Resolution (IPTV)			Combined quality value (d_cv) d[0], compression_artifact_value d[1], d_slicing_artifact_value d[2], d_freezing_artifact_value	$d\_cv = \alpha[0] * dpp[0] + \alpha[1] * dpp[1] + \alpha[2]$

Table 2 - IPTV/VoD Standardized Models from P.1201 and P.1202

4.2.2.5 Commercial IPTV Monitoring Systems

Despite the standardization efforts shown above, standardized IPTV tools are not often used in the market. In many cases, the IPTV service supplier itself offers a built-in monitoring solution, which is based on proprietary models. The service operator, who buys the IPTV equipment as an off-the-shelf solution or has it customized for its implementation, usually relies on those built-in monitoring tools for evaluating the current service quality (e.g., for IPTV solutions offered by Huawei, Ericsson, Cisco and others). The reasons for choosing such an approach over the separate implementation of standardized models are manifold, with the main one being a simple cost factor.

Information supplied by the monitoring tools could include basic network KPIs, such as packet loss rates or jitter, but also higher-level KPIs, for example relating to channel switching time, which is an important quality factor for users. However, the set top boxes used for decoding the transmitted stream could also include packet-header- or bitstream-based quality models that output "MOS" values for every customer, which then can be collected and evaluated centrally. The validity of such MOS values can be questioned if the underlying models have been developed and evaluated proprietarily.

It will be a challenge for standardization to make its way into the commercial solutions available on the market, with the goal to have vendors provide standardized monitoring tools built into their products.

4.2.3 Conferencing Services

Over the last decade, the usage of Internet conferencing services have extended tremendously for remote meetings, distance learning, tele-medicine, etc. At the same time, service providers have faced issues in deploying and managing large-scale services on the internet, which are related mostly to low-quality services due to change in network condition which causes users to experience impairments such as video frame freezing and voice dropouts.

As mentioned in section 2.2, objective reference models can be categorized into three groups, no-reference, full-reference, and reduced-reference. For monitoring real-time services – such as conferencing – using FR models would be an inappropriate choice, since accessing to the source signal would not always be possible in this scenario. However, for other non-monitoring purposes (i.e., lab testing), a proper FR model for conferencing services should consider both video and audio quality during the assessment. Therefore, it is possible to use FR metrics by combining current video and audio metrics, or using audiovisual quality metrics. It should be noted that although for measuring quality in conferencing services consideration should be taken to the video and audio quality, at the same time there exist some other features such as perceptual aspects and sensitivity to network condition which differ from other applications. Consequently, it is not enough to consider only the video and audio aspects in order to have an appropriate quality monitoring in conferencing service. Many aspects of multiparty[2] quality are still unknown [32].

In the following sections, we show current solutions for quality assessment in video conferencing.

4.2.3.1 Opinion Model for Video-telephony Applications – Network Planning (ITU.T G.1070)

Recommendation G.1070 has been proposed by the ITU-T. Before proposing the recommendation, the concept of non-intrusive parametric and bitstream quality modeling was introduced in [33]. G.1070 can be used for QoE/QoS planning to guarantee the end-to-end user quality while the service is being provided.

---

[2] Conferencing between more than two persons.

Therefore, the application of this model is limited to planning only. However, it can be used to estimate video conferencing quality, by setting the levels of the individual parameters – similar to the E-model.

In G.1070, the multimedia quality is calculated by using the network, application, and terminal equipment parameters. In this model, there are some assumptions for terminals, environments, and evaluation context. For the terminal, it is assumed that a handset is used as an interface for the voice path, and also a monitor with specifications listed in Table 3 in [34] is required. The model assumed the ambient noise is Hoth noise at 35 dB(A), and video content to be a so-called "bust shot" with a gray background. Finally, the only task that has been considered is "free conversation", in which only the video sequence of a user is displayed on the screen.

The framework of this model is illustrated in Figure 7. The inputs consist the important QoE/QoS planning parameters from video and speech data, which are listed in Table 3. The main functions of this model are video and speech quality estimation, and quality integration. For speech quality estimation, it uses the same function as in the E-model, but it is simplified to a smaller number of input parameters. The video quality and multimedia quality integration functions can be found in section 11 of [34].

Table 3 Model inputs

<table>
<tr><th>Audio parameters</th><th colspan="2">Video parameters</th></tr>
<tr><td>Speech coding distortion</td><td colspan="2">Video delay</td></tr>
<tr><td rowspan="3">Speech packet-loss robustness</td><td rowspan="3">Video codec specifications</td><td>Codec type and implementation</td></tr>
<tr><td>Spatial resolution</td></tr>
<tr><td>Key-frame interval</td></tr>
<tr><td>Speech packet-loss rate</td><td colspan="2">Video packet-loss rate</td></tr>
<tr><td>Talker echo loudness rating</td><td colspan="2">Video frame rate</td></tr>
<tr><td></td><td colspan="2">Video bit rate</td></tr>
</table>

There are some other research endeavors which extended the ITU-T G.1070 model to improve the perceptual quality and assess the real-time videos quality. [35] used the Sum of Absolute Differences (SAD) as an estimation of the video spatial-temporal activity to improve the G.1070 model. [36] proposed a video quality monitoring application based on the G.1070 model. [37] utilized this model to propose ROI-based video quality assessment and regulation for mobile videoconferencing. Finally, [38] proposed a no-reference objective quality metric for stereoscopic 3D video by extending G.1070.

4.2.3.2 PSQA: "Pseudo Subjective Quality Assessment" for No-Reference Monitoring

PSQA [39] is a no-reference model for real-time audiovisual streaming application such as telephony conferencing. It is a parametric approach which is network-dependent and application-dependent. PSQA estimates the quality by mapping QoS parameters and source-based parameters into perceptual quality. It is based on statistical learning tools. The main idea behind this algorithm is to use machine learning to build the application-based quality model. The first step is to find the main parameters for the desired application. PSQA doesn't recommend any set of parameters since they could be changed from application to application. Then, in the learning phase, the model is built by MOS value and selected parameters. In a sentence, PSQA could be defined as "a technique based on merging subjective assessment with a statistical learning tool which allows to produce subjective-like quality estimations based on measurable network and application parameters" [40]. There is a preprocessing for the MOS values to remove outliers. In the paper, the RNN method was used as a machine learning algorithm, but another method could be used to obtain better performance. Based on the parameters that are selected in the model, PSQA can be used as a monitoring tool in different points of the network [41].

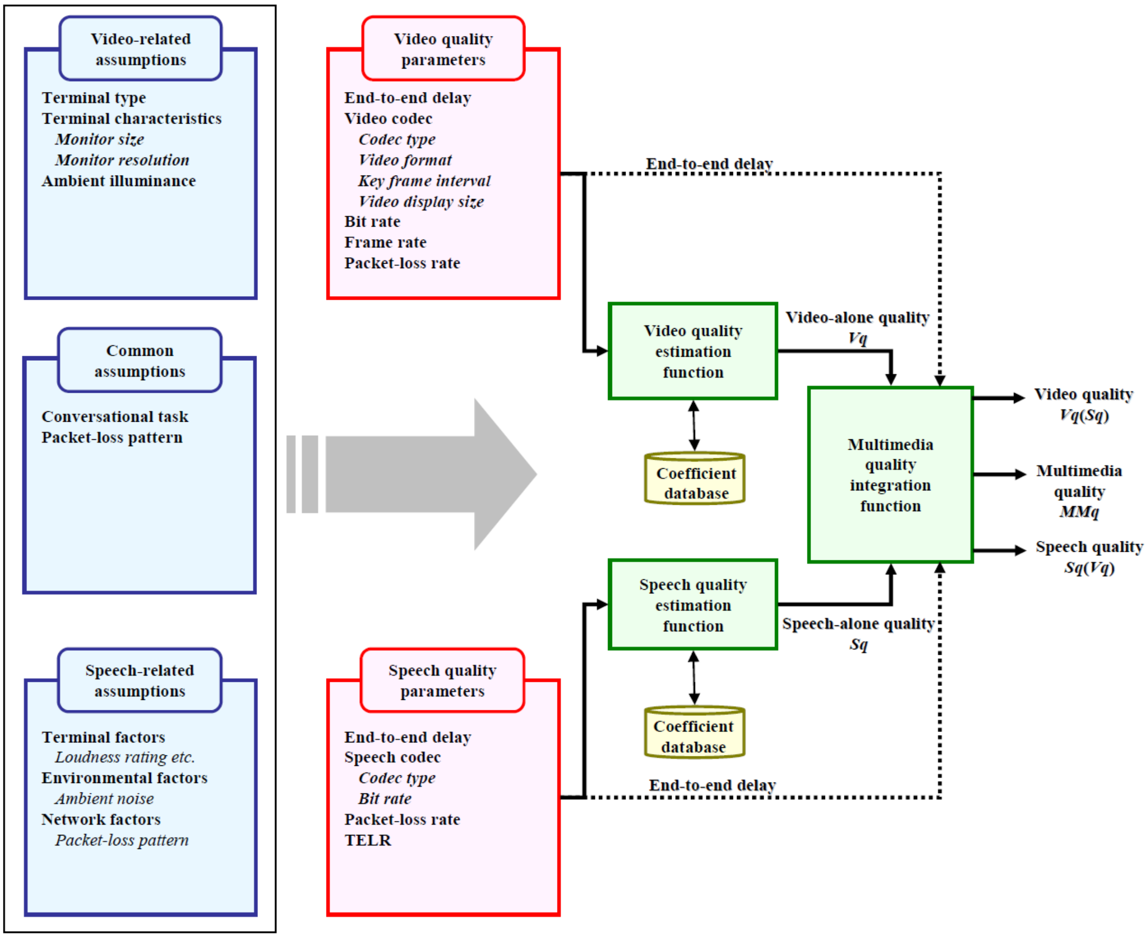

Figure 7 G.1070 Framework

4.2.4 Mobile Gaming

For gaming scenarios, currently, there are no monitoring tools the authors are aware of. However, there exist some quality models which predict the QoE for these applications. It should be noted that there are different gaming scenarios which change the quality assessment procedure. Different mobile gaming scenarios have been presented in the QoE-Net Deliverable D3.1 [42]. The main focus of current research of mobile gaming is devoted to a cloud gaming scenario. Although it seems that quality of a cloud gaming service could be estimated by audiovisual models such as ITU-T G.1071 [43], this service has several factors which influence the quality but are not considered in current audiovisual models [44].

In the following section, two quality models are presented, including the G-Model and the GRF model. Both models are applied to first person shooter (FPS) games, since this genre is highly sensitive to varying network conditions.

4.2.4.1 Gaming Quality Models

There exist several papers which showed the impact of network conditions – such as delay, jitter, packet loss – on perceived quality in FPS gaming. Therefore, current game quality models focus on network performance metrics – especially delay – to estimate the quality. The first model for gaming is proposed in [45], which is inspired by the E-model [22]. The so-called "G-model" predicts the QoE of an FPS game. In this paper, the experiment has been performed for Quake III. The model assumed that FPS games are insensitive to packet loss, therefore, they only emphasize jitter and RTT – RTT called "ping" in this paper.

According to [45] the correlation between the subjective test and this model for Quake III is very high (Correlation coefficient p=0.98). Since the only parameters that are used in this model are RTT and jitter which can be captured by sending packets to the client and vice versa, passive probes can gather this information on the network.

The second model for online gaming has been proposed to overcome the weaknesses of the G-model, which include low number of parameters, and the fact that it was trained on only one game [46]. Extended network parameters such as packet loss are used, moving from a game-specific model to a wide range of existing online games. New parameters such as user-based ones have been included. In the Table 4 parameters that are used in this research paper are listed.

Table 4 Parameters that were considered in [46]

Network parameters	I/O	Game Client
Delay	Mouse speed impairment	Graphic card impairment
Jitter	Display impairment	The memory of the machine
Packet loss		Processor speed impairment

This paper showed that delay and jitter are the most important factors for online gaming that impact the MOS significantly. Since this model needs user device information, it should be placed on the user devices.

4.3 Guidelines for the Use of Monitoring Tools

Objective QoE assessment allows for an automated monitoring of QoE levels. QoE monitoring is considered to become more and more critical due to its relationship with increasingly performing QoE metrics, customer satisfaction and business objectives of providers.

Based on the classification of models in Section 2.2, we now identify typical use cases for the different model types:

1. **Full-Reference models:** Quality verification of delivered content in lab environment, quality assessment of codec/server performance.
2. **Reduced-Reference models:** Quality benchmarks between different remote locations to assess network and application quality, continuous quality monitoring.
3. **No-Reference models:** Continuous quality monitoring and the quality assessment at the client side

In real-life scenarios, it is difficult to add extra information to the transmitted data stream to implement RR metrics, since it would require additional bandwidth as well as a modification of default service implementations. Thus, for a service-wide monitoring, typically NR metrics are considered. Depending on the point in the network where measurements are to be taken and on the processing capabilities of the used devices, the appropriate NR metric type needs to be chosen. In case of an intermediate point in the network – that is a network-centric approach – bitstream-based / parametric metrics for quality estimation may be efficient approach for quality assessment, especially if multiple streams are to be monitored. In case of client-side monitoring, signal-based metrics can also be employed for quality estimation because they are able to access the decoded video at the receiver output. However, in many cases (e.g., when energy or CPU consumption is relevant, and when the monitoring itself cannot influence the performance of the end device), running a signal-based metric is not possible. Here, bitstream-based / parametric metrics are preferred.

As QoE monitoring tools become more widespread, the accuracy and validity of the gathered data (e.g., from probes and other devices) needs to be ensured. For example, using video quality metrics that have been developed for IPTV systems in a VoD system may not yield exact predictions of the perceived quality. In other words: A tool may output a MOS value, but the applicability of that value for a given service has to be checked by comparing the underlying model's scope with the real application scope.

To conclude this section, In Table 5, there is a comparison between existing QoE monitoring approaches based on their quality monitoring points, estimation point, their measurement environment (real life or laboratory), the service type and the used parameters. This has to be intended as a general guideline for selecting the appropriate model on the basis of the high-level requirements.

Table 5 Overview of presented QoE/QoS monitoring approaches

Model	QoS/ QoE Monit-oring Point	QoS/ QoE Estimat-ion Point	Measurem-ent environm-ent	Service type	Network Planning / Monitoring	Parameters
E-model [22]	*n/a*	*n/a*	Laboratory	VoIP	Network Planning	Network, Equipment impairment (+noise and loudness)
G-model [45]	Passive	Network	Laboratory	Gaming (FPS)	Network Planning	Delay, Jitter
PESQ [26]	Client	Network	Laboratory, Real Life	VoIP	Monitoring	Various signal characteristics (FR approach)
POLQA [27]	Client	Network	Laboratory, Real Life	VoIP	Monitoring	Various signal characteristics (FR approach)
P.563 [28]	Client	Client	Laboratory, Real Life	VoIP	Monitoring	Various signal characteristics (naturalness, noise, interruptions)
ITU.T G.1070 [34]	*n/a*	Network	Laboratory	Video Telephony	Network Planning	Has been described in Table 3
PSQA [39]	*n/a*	Network	Laboratory (Could be done in Real life)	Generic Model	Network Planning / Monitoring	Depends on the application
ITU-T P.1201 Amd. 2, App. III ("P.NAMS-PD")	Client	Network, Client	Real Life	VoD	Monitoring	Video, audio codec parameters, stalling times
ITU-T P.1201 [30] ITU-T P.1202 [31]	Client, Network	Client, Network	Real Life	IPTV	Monitoring	Video, audio codec parameters, packet loss information

ITU-T J.247	Client	Client , Network	Laboratory, Real Life	IPTV, VoD	Monitoring	Video signals
ITU-T J.341 [47]	Client	Client, Network	Laboratory, Real Life	IPTV, VOD	Monitoring	Video signals
ITU-T J.343	Client	Client, Network	Laboratory, Real Life	IPTV, VOD	Monitoring	Video signals

5 Conclusions and Future Challenges of QoE/QoE Monitoring Tools

Wherever we look at today's world of rapidly evolving multimedia services, importance of considering novel multimedia services quality become apparent. This includes services that have previously not even been studied extensively in academia, for example gaming and virtual reality. Here, it is still hard to quantify what Quality of Experience even means for users.

At the same time, user's expectations for existing and well-known services (like TV and speech) are changing. Monitoring tools become more complex not only to meet users' expectations, but also to include new application factors.

Recent advances in communication systems have provided for better networks for users, in terms of increased bandwidth, reduced delay, and improved reliability. Therefore, high-quality and highly complex content and applications sent via the Internet are becoming more accessible. This not only includes traditional services which now use more bandwidth (e.g. 4K video) but also novel service types (e.g. Virtual Reality). This fact threatens current QoS monitoring tools, as they become less useful, since they just use network parameters for monitoring purposes. Those have been shown to correlate badly with user perception.

The next generation of multimedia services will include new concepts such as context-aware coding, production and deployment in order to meet user-centric requirements. However, introducing context-aware approaches results in new requirements for multimedia quality monitoring too. Otherwise, the predicted quality may differ from what the user perceived.

5.1 Large Parameter Spaces and New Services

QoE monitoring may involve a large context parameter space including several user factors such as physical environment factors (home, office, mobile, or public usage; space, acoustic, and lighting conditions; transmission channels involved; potential parallel activities of the user; privacy and security issues), user static characteristics (e.g., age, gender, native language), and user dynamic factors (e.g., motivation, emotional status). The influence of each factor still is unknown for several services, especially in areas where there is no QoS/QoE model (e.g., gaming or virtual reality services). Although some researchers introduce different conceptual QoE models, they just classify the parameters and their relationships. For instance – even though for gaming there exist several papers regarding the impact of diverse factors –there is no quality model for gaming [44]. In fact, there is no unified and general mechanism for monitoring or predicting quality. Therefore, the first step to develop new forms of QoE monitoring tools is to search for new QoE factors in each application and measure the impact of QoE factors using dedicated subjective experiments.

New applications and services have been introduced over the last years, which call for more complex QoE monitoring solutions. For instance, QoE in interactive applications such as gaming is influenced by several factors that are not present in classic services like IPTV. The QoE is significantly impacted by the user's interactivity. Therefore, the advent of novel services brings new challenges specific to that service or application. Especially interactive services such as gaming services, social TV and virtual reality need more consideration to extract the influence factors of QoE. Therefore, QoE prediction and monitoring may become more complex in these services in comparison to other services that are mostly passive and rely on technologies that have been around for a long time.

5.2 Privacy Aspects

Privacy issues in context-aware monitoring could be discussed in two views. First, privacy has become an important issue for the users and internet providers in the recent years. After three major information leaks including the leak of classified documents of the NSA [48], the iCloud leakage of celebrity photos [49] and the Panama Papers leaks [50], privacy concerns of users become more and more vital. Therefore, users prefer not to let providers obtain their personal information, which in turn makes the QoE monitoring difficult. However, users who expect higher quality may have to let providers access some personal information. The second view is linked to current assessment in order to abate privacy concerns. Nowadays, most OTT providers switch to application-level encryption such as SSL/TLS (e.g., through HTTPS). Therefore, most user context factors cannot be measured by network-centric probes.

It seems that there are few ways to overcome these issues such as putting probes on the user terminal or cooperation between OTTs and ISPs to achieve QoE monitoring tools. Both of these solutions impact user concerns of privacy which may become more important in the future.

5.3 Measuring Quality over Time and for Certain Users

Proper quality monitoring and user engagement prediction cannot be achieved by measuring the quality in a single location, at a single point in time. In fact, quality should be measured and predicted over time, for a given user of a service. Looking at the repeated use of a service or application, a user may be satisfied or not [51], ultimately leading to churn if the user’s expectations are not met. As it mentioned in [51] there exists a need for developing novel monitoring tools which measure quality over time – rather than based on a single session – for any given user. This enables a better view on how engaged or likely to churn a certain customer is. However, it is not always trivial to collect quality measurements and connect them to a specific user – let alone privacy concerns. Even if the data can be mapped, it is a challenge to train and validate models (using, e.g., machine learning approaches) to predict user engagement or churn. This requires a closer collaboration between different departments or working groups in larger operators, e.g. between the (typically technology-centered) network analysts and the (user-centered) customer-relationship departments.

5.4 Conclusion

Over the last decade, the use and diversity of multimedia applications have extended tremendously. Traditional approaches to QoS/QoE monitoring may at some point not be applicable anymore, for example due to architectural constraints. Also, it has been understood that a real prediction of QoE is only possible with access to detailed indicators from the applications sent over the network, which are not always available.

In this document we first described how QoE is typically assessed – both subjectively and instrumentally. We presented the state of the art of QoS/QoE monitoring tools, as well as the influence factors and parameters considered in these tools. After that, a selection of current quality monitoring tools in different application areas (such as VoIP, IPTV, and VoD) were explained. Finally, we gave an outlook into the future of QoS/QoE monitoring tools, with the most important challenges that need to be considered.

6 Bibliography

[1] ITU-T, "Rec. E.800: Definitions of terms related to quality of service," 2008.

[2] S. Möller, P. Le Callet and A. Perkis, "Qualinet White Paper on Definitions of Quality of Experience," European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), 2013.

[3] U. Reiter, K. Brunnström, K. De Moor, M.-C. Larabi, M. Pereira, A. Pinheiro, J. You and A. Zgank, "Factors influencing Quality of Experience," in *Quality of Experience*, Springer, 2014.

[4] M. Alreshoodi and J. Woods, "Survey on QoE/QoS correlation models for multimedia services," *arXiv,* 2013.

[5] "Deliverable 2.1 – Test Method Compendium," QoE-Net, 2015.

[6] A. Raake, J. Gustafsson, S. Argyropoulos, M.-N. Garcia, D. Lindegren, G. Heikkilä, M. Pettersson, P. List and B. Feiten, "IP-based mobile and fixed network audiovisual media services.," *IEEE Signal Processing Magazine,* vol. 28, no. 6, pp. 68-79, 2011.

[7] "SDN & NFV OpenFlow and ForCES IETF--93," [Online]. Available: https://www.ietf.org/edu/tutorials/sdn-nfv-openflow-forces.pdf. [Accessed 18 4 2016].

[8] " OpenFlow," [Online]. Available: https://www.opennetworking.org/sdn-resources/openflow. [Accessed 18 4 2016].

[9] S. Baraković and L. Skorin-Kapov, "Survey and challenges of QoE management issues in wireless networks," *Journal of Computer Networks and Communications,* 2013.

[10] ITU-T, "ITU-T Recommendation G.1081 - Performance Monitoring points for IPTV," ITU-T, 2008.

[11] ITU-T, "ITU-T Recommendation Y.1540: Internet protocol data communication service and Availability Performance parameters," 2011.

[12] ITU-T, "ITU-T Recommendation Y.1544: Multicast IP Network Performance Parameters," ITU-T, 2008.

[13] M. Mu, R. Gostner, A. Mauthe, G. Tyson and F. Garcia, "Visibility of individual packet loss on H. 264 encoded video stream: a user study on the impact of packet loss on perceived video quality," *IS&T/SPIE Electronic Imaging,* 2009.

[14] Y. Alahari and P. Buddhiraja, "Analysis of packet loss and delay variation on QoE for H. 264 and WebM/VP8 Codecs," 2011.

[15] F. Liu, T. H. Luan, X. S. Shen, and C. Lin,, "“Dimensioning the packet loss burstiness over wireless channels: a novel metric, its analysis and applications," *Wireless Communications and Mobile Computing,* vol. 14, no. 12, p. 1160–1175, 2012.

[16] T. Hoßfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial Delay vs. Interruptions:

Between the Devil and the Deep Blue Sea," in *Fourth International Workshop on Quality of Multimedia Experience (QoMEX)* , Yarra Valley,Australia, 2012.

[17] P. Casas, et al., "On the quest for new KPIs in mobile networks: The impact of throughput-fluctuations on QoE," in *IEEE ICC 2015 - Workshop on Quality of Experience-based Management for Future Internet Applications and Services (QoE-FI)*, London, 2015.

[18] A. Sackl, P. Casas, S. Raimund, L. Janowski and R. Irmer, "Quantifying the impact of network bandwidth fluctuations and outages on Web QoE," in *Quality of Multimedia Experience (QoMEX)*, 2015 .

[19] D. Soldan, "QoS and QoE Monitoring and Data analyses," 2006. [Online]. Available: http://www.netlab.tkk.fi/opetus/s383215/2006/QoE8.pdf. [Accessed 2016].

[20] R. Birke, M. Mellia, M. Petracca and D. Rossi, "Experiences of VoIP traffic monitoring in a commercial ISP," *International Journal of Network Management,* vol. 20, no. 5, pp. 339-359, 2010.

[21] S. Möller, W.-Y. Chan, N. Côté, T. H. Falk, A. Raake and M. Wältermann, "Speech quality estimation: Models and trends," *Signal Processing Magazine, IEEE,* vol. 28, no. 6, pp. 18-28, 2011.

[22] ITU-T, *ITU-T Recommendation G.107: The E-model: a computational model for use in transmission planning,* 2015.

[23] R. G. Cole and J. H. Rosenbluth, "Voice over IP performance monitoring," *ACM SIGCOMM Computer Communication Review,* vol. 31, no. 2, pp. 9-24, 2001.

[24] H. Assem, D. Malone, J. Dunne and P. O'Sullivan, "Monitoring VoIP call quality using improved simplified E-model," in *2013 IEEE International Conference on Computing, Networking and Communications (ICNC)*, 2013.

[25] C. Simão, F. Neves, S. Soares, F. Tavares and P. Assunção, "ArQoS®: System to monitor QoS/QoE in VoIP," in *EUROCON-International Conference on Computer as a Tool*, 2011.

[26] ITU-T, *ITU-T Recommendation P.862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,* 2001.

[27] ITU-T, *ITU-T Recommendation P.863: Perceptual objective listening quality assessment,* 2014.

[28] ITU-T, "ITU-T Recommendation P.563: Single-ended method for objective speech quality assessment in narrow-band telephony applications," 2004.

[29] ITU-T, "ITU-T Recommendation P.1201 Amd. 2, App. III: Use of P.1201 for non-adaptive, progressive download type media streaming," 2013.

[30] ITU-T, "ITU-T Recommendation P.1201: Parametric non-intrusive assessment of audiovisual media streaming quality," 2012.

[31] ITU-T, "ITU-T Recommendation P.1202: Parametric non-intrusive bitstream assessment of video media streaming quality," 2012.

[32] J. Skowronek, K. Schoenenberg and A. Raake, Experience with and insights about the new ITU-T standard on quality assessment of conferencing systems, Fortschritte der Akus-tik (AIA/DAGA'13), 39. Jahrestagung für Akustik, IT-Meran, 18-21 March, Deutsche Ge-sellschaft für Akustik e.V. (DEGA), D-Berlin.

[33] T. Akira, K. Yamagishi and G. Kawaguti, "Recent activities of QoS/QoE standardization in ITU-T SG12," NTT Technical Review, 2008.

[34] ITU-T, *ITU-T Recommendation G.1070: Opinion model for video-telephony applications,* 2007.

[35] J. Jose, C. L. Ardao and R. Sotelo, "Quantitative modeling of the impact of video content in the ITU-T G. 1070 video quality estimation function," in *Informática na educação: teoria & prática* , 2011.

[36] L. Tao, N. Narvekar, B. Wang, R. Ding, D. Zou, G. Cash, S. Bhagavathy and J. Bloom, "Real-time video quality monitoring," *EURASIP Journal on Advances in Signal Processing ,* vol. 1, pp. 1-18, 2011.

[37] Y. Han, Z. Yuan and G.-M. Muntean, "ROI-based Video Quality Assessment and Regulation for Mobile Videoconferencing," in *Proceedings of Workshop on Mobile Video Delivery*, 2014.

[38] H. Yi, Z. Yuan and G.-M. Muntean, "Extended no reference objective Quality Metric for stereoscopic 3D video," in *In Communication Workshop (ICCW)*, 2015 .

[39] M. Varela, Pseudo–subjective Quality Assessment of Multimedia Streams and its Applications in Control, Rennes, France: PhD thesis, INRIA/IRISA University.

[40] A. P. C. da Silva, M. Varela, E. d. S. e. Silva, R. M. Leão and G. Rubino, "Quality assessment of interactive voice applications," *Computer Networks ,* vol. 52, no. 6, pp. 1179-1192, 2008.

[41] S. Mohamed, G. Rubino and M. Varela, "Performance evaluation of real-time speech through a packet network: a random neural networks-based approach," *Performance evaluation 57,* pp. 141-161, 2004.

[42] "Deliverable 3.1 – Review of content and context-ware coding strategies in the envisaged application scenarios," QoE-Net, 2016.

[43] ITU-T, *ITU-T Recommendation G.1071: Opinion model for network planning of video and audio streaming applications,* 2015.

[44] S. Moller, S. Schmidt and J. Beyer, "Gaming taxonomy: An overview of concepts and evaluation methods for computer gaming qoe," in *Quality of Multimedia Experience (QoMEX)*, 2013.

[45] A. F. Wattimena, R. E. l. V. J. M. Kooij and O. K. Ahmed, "Predicting the perceived quality of a first person shooter: the Quake IV G-model," in *5th ACM SIGCOMM workshop on Network and system support for games*, 2006.

[46] L. Venturelli, D. Picovici and J. Kehoe, "Online game quality assessment," in *In Irish Signals & Systems Conference (ISSC)*, 2014.

[47] ITU-T, "ITU-T Recommendation J.341: Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference," 2016.

[48] G. Greenwald, *No place to hide: Edward Snowden, the NSA, and the US surveillance state,* Macmillan, 2014.

[49] C. Arthur, "Naked celebrity hack: security experts focus on iCloud backup theory," 1 September 2014. [Online]. Available: https://www.theguardian.com/technology/2014/sep/01/naked-celebrity-hack-icloud-backup-jennifer-lawrence.

[50] J. Garside, H. Watt and D. Pegg, " The Panama Papers: how the world's rich and famous hide their money offshore," [Online]. Available: http://www.theguardian.com/news/2016/apr/03/the-panama-papers-how-the-worlds-rich-and-famous-hide-their-money-offshore. [Accessed 3 April 2016].

[51] K. Mitra, A. Zaslavsky and C. Åhlund, *QoE modelling, measurement and prediction: A review,* arXiv preprint arXiv, 2014.