QoS FOR FIXED AND MOBILE ULTRA-BROADBAND

TONI JANEVSKI

IEEE PRESS

WILEY

QoS for Fixed and Mobile Ultra-Broadband

QoS for Fixed and Mobile Ultra-Broadband

Toni Janevski
Ss. Cyril and Methodius University
Skopje, Macedonia

WILEY

IEEE PRESS

This edition first published 2019

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Janevski, Toni, author.
Title: QoS for fixed and mobile ultra-broadband / Toni Janevski, Ss. Cyril and Methodius University, Skopje, Macedonia.
Description: Hoboken, NJ, USA : Wiley IEEE Press, 2019. | Includes bibliographical references and index. |
Identifiers: LCCN 2018060368 (print) | LCCN 2019000192 (ebook) | ISBN 9781119470496 (Adobe PDF) | ISBN 9781119470489 (ePub) | ISBN 9781119470502 (hardcover)
Subjects: LCSH: Mobile communication systems–Quality control. | Wireless communication systems–Quality control. | Broadband communication systems–Quality control.
Classification: LCC TK5102.84 (ebook) | LCC TK5102.84 .J36 2019 (print) | DDC 384.3068/5–dc23
LC record available at https://lccn.loc.gov/2018060368

Cover Design: Wiley
Cover Image: © monsitj/iStock.com

Set in 10/12pt WarnockPro by SPi Global, Chennai, India

Printed in Great Britain by TJ International Ltd, Padstow, Cornwall

10 9 8 7 6 5 4 3 2 1

To my great sons, Dario and Antonio, and to the most precious woman in my life, Jasmina.

Contents

1

Introduction

The telecommunications world has been developing at a rapid pace since the growth of the Internet in the 1990s and 2000s. Nowadays telecommunications is also referred to as information and communication technologies (ICT), as stated by the largest telecommunications agency in the world, the International Telecommunication Union (ITU) [1], a specialized agency of the United Nations. Telecommunications has been around for more than 150 years, starting with telegraphy in the nineteenth century. In fact, the telecommunications world and the ITU have been interrelated since 1865 when the ITU was formed as the International Telegraph Union. Nowadays, telegraphy belongs to history (it has become redundant since the appearance of email and other messaging services available worldwide today). But speaking about the history of telecommunications, after Alexander Graham Bell invented the telephone in 1876 the following century was marked by the development and deployment of telephony, with fixed telephony until the 1980s accompanied by mobile telephony worldwide from the 1990s. Of course, one should not forget to mention television and radio as important telecommunication services during the twentieth century, and they continue to be so in the twenty-first century.

ICT has created a globally connected world, not only giving people the ability to communicate with each other but also opening up access to information and facilitating the exchange of information. The foundation of such an ICT world (the terms ICT and telecommunications will be used interchangeably in this book) lies in the introduction of digital systems and networks in the 1970s and 1980s, which provided the possibility for all information from different sources and of different types (e.g. voice, video, data, multimedia) to be transferred over the global telecommunications infrastructure by using series of bits and zeros (i.e. in a digital form). When information is coded at the source as series of ones and zeros, all such information can be transferred over the same telecommunications networks and accessed via the same devices – if, of course, the networks are created in such a manner. The Internet has provided the required openness to transfer all different types of information over the same network, which is present and working well in the second decade of the twenty-first century and is expected to continue in a similar manner in the future. However, telecommunication networks are interconnected on a local, regional, and global scale to be able to transfer the information between any two or more communication endpoints on the Earth, so the telecommunication services are global. Therefore, the quality of telecommunication services which is applied in a single network or in a single country has influence on the end-to-end quality of that service. So, the quality cannot be considered only at national or regional level, it needs

QoS for Fixed and Mobile Ultra-Broadband, First Edition. Toni Janevski.

to be considered globally. Today, citizens around the world rely on ICT to conduct their everyday activities in personal or business life, and that requires having certain quality of services (QoS). Therefore guaranteeing QoS in the socio-economic environment of users is becoming very important. For that purpose there is a need for technical mechanisms and functions for implementation of the QoS in networks and end-users' devices, and for its regulation in a harmonized and globally accepted approach. That will enable greater quality of services provided to users as customers or consumers or content generators, irrespective of their location or service provider (the entity that provides access to telecommunications services), where the provider can be a telecom operator (on local or national levels) or global over-the-top (OTT) provider (e.g. Google, Amazon, etc.).

1.1 The Telecommunications/ICT Sector in the Twenty-First Century

The telecommunication/ICT world in the twenty-first century is characterized by two important developments:

- It is becoming fully based on Internet technologies, including all networks (with fixed and mobile access) and all services (including all applications working over the Internet).
- It is becoming broadband, which means that there are enough high bitrates in access networks (and also end-to-end) which provide all available applications/services to run smoothly and with satisfactory quality experienced by end-users (we will define what the quality means later in the chapter).

So, the ICT world is becoming all-IP (Internet protocol) and broadband on a global scale (Figure 1.1) [1]). However, it is even more interesting to note that mobile communications are spreading at a faster pace than fixed communications regarding access

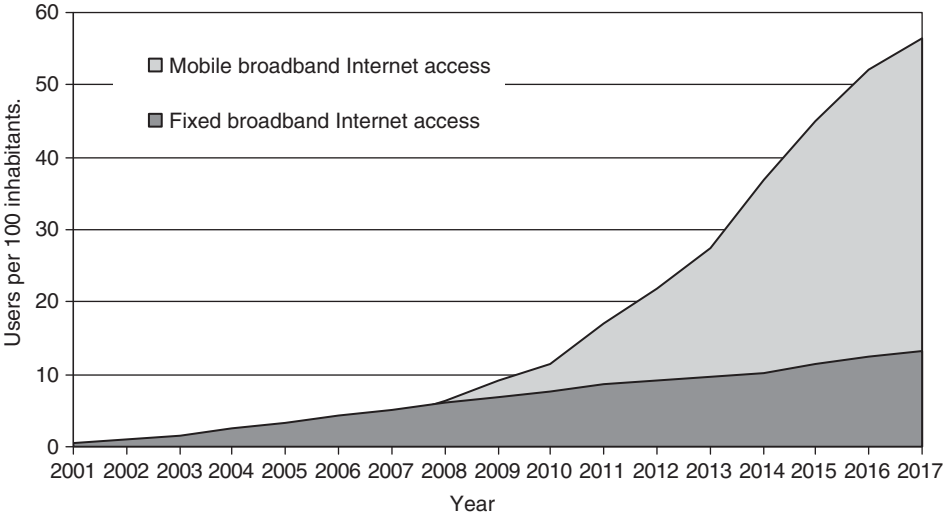

Figure 1.1 Global telecommunication/ICT broadband developments.

networks. One may note that in the 1990s only about 1% of the global population had a mobile cellular subscription and about 11% had a fixed telephone subscription. Nowadays, mobile subscribers' penetration is near saturation, with almost everyone on this planet having a mobile phone/device. Mobile broadband is the most dynamic market segment, surpassing fixed broadband access in the second decade of the twenty-first century. The number of Internet users is also constantly growing over the entire world population (Figure 1.1).

What is broadband? One may define broadband as access network bitrates and end-to-end network bitrates in both directions (downstream and upstream) which support all available types of services with satisfactory quality. For example, in the first decade of the twenty-first century broadband access meant offering hundreds of kbit/s, while a decade later it means access with Mbit/s or tens of Mbit/s (as measurement units for data speeds), enabling, for example, the provision of HD (high definition) video and ultra-HD video streaming (some of the most demanding services).

Broadband can be divided into two main categories (similar to type of access networks):

- *Fixed broadband*. Fixed broadband access technologies are provided by copper (twisted pairs) by reusing local loops deployed for fixed telephony for IP-based access via ADSL (asymmetric digital subscriber line), VDSL (very-high-bit-rate digital subscriber line), or generally x digital subscriber line (xDSL); or by cable access (by reusing coaxial cable networks, primarily developed for TV distribution, and FTTH (fiber to the home) or more general FTTx (fiber to the x), which is a long-term future for fixed broadband access in all regions. On the other side, almost all transport networks nowadays are fiber-based (accompanied by satellite networks, where optical transport is not present), so the differences remain mainly in the last mile.
- *Mobile broadband*. Mobile broadband access technologies appeared with the 3G (third-generation) mobile networks in the late 2000s, and they continue with 4G (fourth-generation) and 5G (fifth-generation) mobile networks in the second and third decades of the twenty-first century. The widespread mobile technologies which belong to the mobile broadband world include UMTS/HSPA (universal mobile telecommunication system/high speed packet access) as part of the 3G standards umbrella, Mobile WiMAX (3G and 4G umbrella), and LTE/LTE-Advanced (LTE stands for long-term evolution) as the most successful 4G technologies. One may note that WiFi is not presented under a separate bullet here, due to the fact that WiFi is more a local wireless extension of fixed or mobile broadband access networks.

According to the ITU [2], by early 2016 total international Internet bandwidth had reached 185,000 Gbit/s, a significant increase from the globally available bandwidth of about 30,000 Gbit/s in 2008. However, there is no equal connectivity of all regions on the global scale. For example, Africa has the lowest international connectivity of all regions, while there is twice as much bandwidth per inhabitant available in Asia and the Pacific, eight times as much in the Americas, and more than 20 times as much in Europe [3]. Also, one may note that always-on access, mobile broadband penetration, as well as the massive adoption of broadband-enabled devices have irreversibly changed the consumers of telecommunication/ICT services, including their social and economic

behaviors as well as QoS expectations, which are constantly increasing over time for certain services (e.g. video).

The more diverse telecommunications/ICT world requires more effort in maintaining the QoS, including provision, measurements, and enforcement. QoS and quality of experience (QoE) are becoming more and more complex because the quality can be impacted by many different factors coming from the telecommunication networks as well as along the value chain, which includes the end-user's device, the available hardware, the network infrastructure, the offered services/applications, etc. As usual, in this process influenced by many factors (which may appear to be different to different users of the same telecommunication service), some differences may arise between perceived and assessed QoS.

QoS is important for all parties involved with the telecommunication services, including both customers (or end-users) and service providers. Therefore there is a need for QoS standards that can form a basis for establishing QoS policies in each country – this is the remit of the appropriate authority in the telecommunication/ICT area (e.g. the national regulatory authority (NRA), ICT ministry, or other ministry or government body). Then the QoS provisioning for the services offered to the customers should be monitored as well as encouraged and/or enforced when necessary.

1.2 Convergence of the Telecom and Internet Worlds and QoS

The telecom and Internet worlds developed in parallel in the 1970s and 1980s. On one side, the telecom world was focused on traditional telecommunication services, telephony (i.e. voice) and television (i.e. TV), which were primary telecommunication services during most of the twentieth century. Traditional telephony was initially fixed based, which refers to the fixed access network via so-called twisted pairs (a pair of two copper wires twisted with the aim of reducing the echo in the opposite direction). A pair of wires also is needed to create a circuit between the telephone (as end-user terminal) and the telecom operator's equipment (e.g. the exchange), therefore such communication was also referred to as circuit-switching (exchanges are network nodes placed on the operator's side, and they provide switching between number of channels on their input and their output).

The first condition for the appearance of the Internet on a global scale was the digitalization of the telecommunication networks built for telephony as the primary service. The process of digitalization of initially analogue telecommunication networks (which used to carry analogue voice signals end-to-end) happened in the 1970s and 1980s, supported by the appearance and development of computers. The telecommunication networks were among the first to accept computers in place of the older analogue technology (based on electrical equipment), with the new digital equipment being based on hardware (i.e. electronics) and software. Digitalization changed the design of telecommunication networks so they started to carry digits as signals (instead of analogue audio for telephony or analogue video and audio for TV), where typically the digits were bits (i.e. ones and zeros) as the most appropriate form for representing different types of media (only a single threshold was needed to distinguish between one and zero at the receiving end of the transmission link, and also computers' architectures are based on

storing and processing the data in a digital form, consisting of ones and zeros). With digitalization, the path was traced to transmit all types of information (audio, video, data) over the same network because all signals were in fact converted into ones and zeros before the transmission (at the sender's end) and vice versa (to the original form of the information) at the receiver's end. So, after the digitalization of telecom networks, data services started to increase in importance; however, there was also a need for a new technology which would suit all types of services, including voice, television, and various data. The next technology was packet-switching. Unlike with circuit-switching where a given channel is occupied all the time during the connection between two ends (e.g. between two telephones), regardless of whether or not there is information to transmit (e.g. voice), the new technology was thought to be based on a different approach. That approach was transmitting a chunk of information in a unit called a packet where each packet has a header (which is heading the data) which includes the address information of the sending and/or receiving end and certain control information regarding the given type of information carried in the packet, called a payload. This technology was packet-switching and there were two main competitors for it in the 1990s: the European-based ATM (asynchronous transfer mode) and the US-based Internet. The ATM was mainly based on the philosophy of traditional telecommunications where all intelligence was placed in the network nodes on the side of the telecom operators while the users had simpler equipment (such as telephone devices). The virtual circuits (introduced by the ATM) are established on signaling messages between the end nodes (which is similar to establishing a telephone call in a telephone network). Meanwhile, the Internet was created on several principles that made it a global success, from which the following ones can be considered as the most important:

- There is separation of applications and services from the underlying transport technologies (e.g. mobile or fixed access networks, transport networks).
- All network nodes and user terminals have the main IP stack based on transport layer protocols, which are primarily UDP (user datagram protocol) [4] and TCP (transmission control protocol) [5], over the IP in its two existing versions, IP version 4 [6] and IP version 6 [7].

The traditional telecommunications layering protocol and the IP model are compared in Figure 1.2. Initially, in the early days (the 1970s) the IP model was based on three layers: the Interface layer at the bottom, network control program (NCP) in the middle, and the application layer on the top. In 1981 the NCP split into TCP (or UDP) over IP, so they became four protocol layers as the native Internet model from the 1980s. However, the network interface layer is typically split into the physical layer and data-link layer by all standard development organizations (SDOs), so with such classification the basic IP layering model has five layers.

The network layer of the Internet is the IP – version 4 (IPv4) or version 6 (IPv6) is currently present in every host, router, and gateway in every network. So, the Internet had won the packet-switched networking technologies battle by the end of the 1990s, which further resulted in the telecommunication/ICT world moving toward Internet-based networking and Internet technologies, a process that continued in the 2000s and 2010s.

But was the Internet a separate network from the traditional telecommunication networks? Well, corporations connected to the Internet via the Ethernet-based networks (where Ethernet is the IEEE 802.3 family of standards) or WiFi (which is the IEEE 802.11

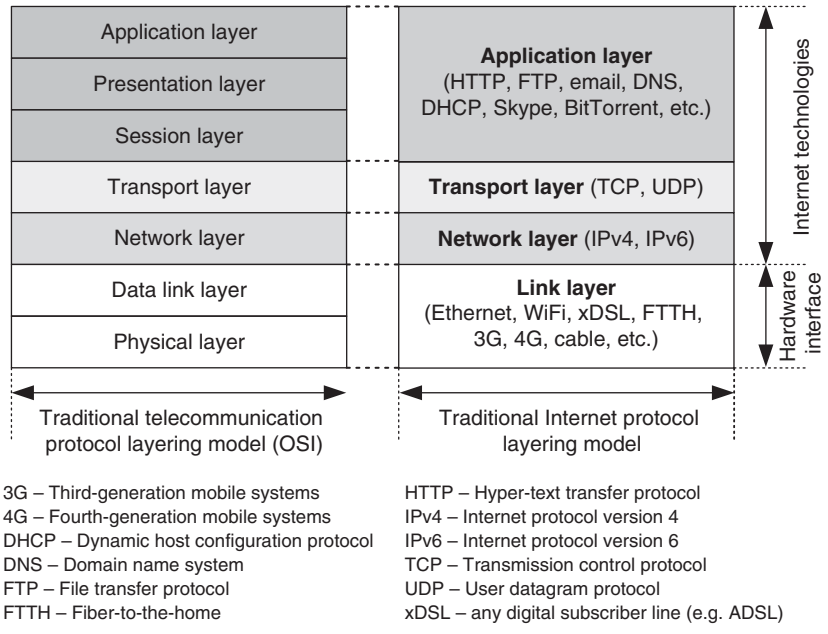

Figure 1.2 Comparison of traditional protocol layering model (Open Systems Interconnection (OSI) model) and Internet protocol layering model.

family of standards) were initially denoted as computer networks (in the 1990s) because they were connecting computers (called hosts, either clients or servers). Meanwhile, the telecommunication networks typically connected "dumb" devices to the "smart" network, which included end-user devices such as telephones and TV sets. But residential users started to connect to the Internet in the 1990s using dial-up modems over their telephone lines, that is, over the established global telecommunication infrastructure for telephony. Also, the Ethernet access networks were interconnected via the global telecommunication infrastructure created for transmission primarily of digital telephony, based on 64 kbit/s dedicated bitrate in each direction of the voice communication (i.e. ITU-T G.711 voice codec standard, based on pulse code modulation (PCM)). The "chunk" bitrate of 64 kbit/s was the basis for all digital telecommunication systems in the 1990s, and even in the 2000s. For example, the main transport technology in digital PSTNs (public switched telephone networks), the SDH (synchronous digital hierarchy) was based on the 64 kbit/s chunks made for voice, so the STM-1 (synchronous transport module, level 1) has a bitrate of 155.52 Mbit/s, which equals exactly 2430×64 kbit/s, and higher transport modules, STM-N, have bitrates of $N \times 155.52$ Mbit/s, which again is a multiple of 64 kbit/s as basic voice throughput in one direction (from user A to user B) in digital telephony. However, the appearance of the Internet on the global ICT scene in the 1990s (sped up by the invention of the World Wide Web and its growth in that period) resulted finally in transport of Internet traffic (various data carried with IP packets and transported by using the IP stack end-to-end) over such SDH transport networks created for transport of digital voice signals. With the convergence of ICTs to the Internet networking principle, Internet networking principles started to penetrate from access

networks (e.g. Ethernet) to metro and transport networks, thus making SDH and other technologies from the digital PSTN era obsolete.

What did this convergence of the telecom and Internet worlds mean? Instead of having separate networks for transmission of different services (e.g. a telephone network for telephony, broadcast network for television, and separate network for data transmission) as the main characteristic of the traditional telecom/ICT world, the transition to Internet networking principles (based on separation of underlying transport technologies from the networking protocol above them, as well as separation of networking protocols from the application on the top) provided the possibility for the realization of one network for all telecommunication/ICT services. So, Internet networking and Internet technologies (standardized primarily by the Internet Engineering Task Force (IETF)) have become the main approach in the telecommunication world, and have become the telecommunications from the network layer up to the application layer. Thus the Internet is not something separate from the telecommunication world, as it was considered to be at the beginning (in the 1980s and even the 1990s), nowadays it is the main part of that world.

With this convergence, certain issues have also arisen. One of the main such issues is the QoS. Why? Because the telecom world based on telephony and TV had strict specifications for QoS requirements for those services, and the telephone and TV networks were designed to provide end-to-end QoS (we will define QoS later). For example, in circuit-switching, PSTNs typically are established two channels end-to-end (each 64 kbit/s), one per direction between the two users (talking to each other over the telephone), and such bitrate is dedicated to the given call for its duration regardless of whether there are voice signals (i.e. talk) to transport over the line or not (i.e. silence). The same approach is present in TV broadcast networks; however, the bandwidth (in bit/s) for TV (i.e. for video) is many times higher than the one needed for voice. On the other side, the native Internet was built on best-effort principles, which means that the network will make the best effort to carry each IP packet from its source to its destination address; however, there are no strict guarantees that the packet will be delivered. So, the Internet in its native design does not contain mandatory QoS mechanisms. Overall, the traditional Internet world has no mandatory QoS, while the traditional telecom world has mandatory QoS mechanisms and functions implemented in the network.

Regarding the Internet, one should note that there were efforts to provide certain QoS options in IP standards from the start. IPv4 (IP and IPv4 are used interchangeably in the following text) has a type of service (ToS) field in which it can specify QoS requirements on precedence, delay, throughput, and reliability. In a similar manner, IPv6 has a DSCP (differentiated service code point) field which can provide support for QoS per flow on the network layer. Neither IPv4 nor IPv6 guarantees the actual end-to-end QoS as there is no reservation of network resources, which is something that should be provided by other mechanisms in IP-based networks.

What about the main Internet architecture? The Internet is built on the basis of autonomous systems (ASs). Each AS is in fact an autonomous administrative domain and it is identified by a 16-bit or 32-bit AS number, which is allocated by the IANA (Internet Assigned Numbers Authority), a department of the ICANN (Internet Corporation for Assigned Names and Numbers), which governs the Internet in terms of domain naming and IP addressing, as well as other well-known numbers from various standardized protocols for IP-based networks (e.g. port numbers for different

protocols on the application layer). The autonomous system is called "autonomous" for a reason, that is, it can apply within its administrative domain the traffic management schemes and routing protocols independently from other ASs, which directly impact the QoS. However, one company or operator can have several ASs and can apply similar traffic management and QoS functions in them. In general, the Internet and the global telecommunication networks are based on IP networks consisting of about 50,000–100,000 active ASs (this number is constantly increasing) [8], which are interconnected. Every AS is connected with either one other AS (e.g. so-called stub AS), or with several adjacent ASs, thus creating the global telecommunication network of today (which is completely based on Internet networking technologies). The global Internet infrastructure consisting of interconnected ASs is crucial in understanding why the traditional way of QoS implementation and enforcement (that is, the same approach in all countries, e.g. for digital telephony, i.e. PSTN/Integrated Services Digital Network (ISDN)) is no longer possible in an IP-based environment. Why? Because in IP environments there is heterogeneity of various IP networks, variety of applied network and traffic management techniques, and the plethora of services and applications which are constantly being offered (e.g. huge OTT applications/services ecosystems, with millions of applications). So, one may note that the QoS as an end-to-end characteristic is becoming more complex in the telecommunications/ICT completely based on Internet networking and Internet technologies.

How is QoS transitioning from the traditional telecom world (made for telephony and TV) to the all-IP world? It is based on strict standardization of certain functions and certain protocols. For example, the traditional telecommunication approach by default includes end-to-end QoS support in the network. Also, signalization in a standardized manner is required for establishing calls/sessions between any two ends on any two devices connected to any two networks regardless of the types of application, device, or network. That led to implementation of certain functions and approaches from the traditional telecom world in the Internet networks, which were initially best-effort based. The main standardization for such convergence was carried by the ITU in its umbrella of specifications called next generation networks (NGNs), which initially started as an idea around 2003. One of the main purposes of the NGN standardization framework was (and still is) to standardize the end-to-end QoS support in all-IP networks (including all needed functions in transport and service stratums) that is essential for real-time services, such as VoIP (voice over Internet protocol) and IPTV (Internet protocol television). Such services have strict requirements regarding QoS (guaranteed bitrates, losses, delay, delay variation, jitter, etc.). So, NGN provides a standardized implementation of QoS (instead of proprietary case-by-case implementations) [9], which is mandatory for the transition from PSTN and public land mobile network (PLMN) to all-IP networks.

So, the telecommunication world is transiting from circuit-switched networks (e.g. PSTN/ISDN) to all-IP networks, including fixed and mobile access networks, as well as core and transit networks. First, the transition was completed in transit and core networks, then in fixed access networks, and lastly in mobile access networks (due to mobility of users, which makes the continuous QoS provisioning more complex). When the transition to all-IP-based networks/services is completed in the telecommunications/ICT world, the Internet technologies are also used for traffic management and for QoS standardization, monitoring, and enforcement. The transition from separate

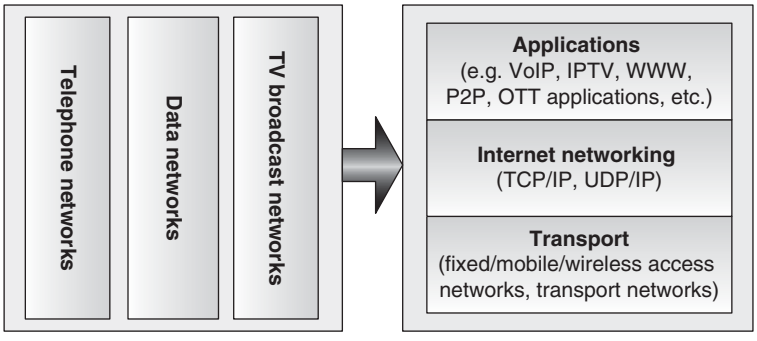

Figure 1.3 Transition from separate networks for different services (the traditional telecom approach) to horizontal separation of services/applications provided via broadband IP networks (the new way).

networks for different services (traditional telecom approach) to a horizontal layered approach in an all-IP environment with broadband access is illustrated in Figure 1.3.

Once the path to packet-switching in the telecom world had been well established by the start of the 2000s, the ITU defined an architectural framework for the support of QoS in packet-switched networks. Nowadays, although there are unified packet-switched networks, the IP-based networks, other different packet-switched networks exist which are also standardized, including the SS7 signaling (as usual, every new technology enters first into the signaling segment of the telecom networks) as well as the already mentioned ATM.

1.3 Introduction to QoS, QoE, and Network Performance

With the convergence of the telecom and Internet worlds, the QoS functions and requirements apply not only to traditional telecommunication and broadcast services but also to broadband Internet-based services.

Overall, telecommunications/ICT services in the twenty-first century are increasingly being delivered using IP based networks, including:

- IP-based networks (access networks, core/backbone networks, and transit networks), and
- IP-based services, which include two main types from the QoS viewpoint:
  - QoS-enabled, i.e. managed services, such as voice, TV, and any other service provided by telecom operators with QoS guarantees end-to-end based on a signed agreement between the telecom operator and the customer (in such case the end-user becomes a customer for the operator);
  - OTT services, which are provided in a best-effort manner over Internet access (either fixed or mobile), without end-to-end QoS, and based on the network neutrality principle being implemented in the Internet (we will refer to network neutrality in more detail in Chapter 8).

So, QoS is clearly moving from its initial definitions targeted at traditional telecommunication networks (e.g. PSTN/ISDN, broadcast networks) to QoS in IP networks and services.

Networks and systems are gradually being designed in consideration of the end-to-end performance required by user applications. In the following subsection we define the QoS, QoE, and network performance (NP).

1.3.1 Quality of Service (QoS) Definition

Traditionally, QoS was mainly addressed from the perspective of the end-user being a person (e.g. telephony), with the ability to hear and see and be tolerant of some degradations of services (e.g. low packet loss ratio is acceptable for voice, while end-to-end delay for voice should be less than 400 ms). But with the advent of new types of communications where services may not require real-time delivery and where the sender or the end-user may not be a person but could be a machine (e.g. Internet of Things), it is important to keep in mind that not all services are the same, and even similar services can be treated in different ways depending on whether they are used by machines or by humans on one or both ends of a given communication session or connection.

> Quality of service (QoS), as defined by the ITU [10], is totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service.

Similar definitions of QoS are used by other SDOs in their standards. However, from the telecommunications/ICT point of view the QoS is always an end-to-end characteristic, which however can be split into different network segments between the ends (e.g. between two hosts on the Internet, or two telephones). However, the end-user's perception of a given telecommunication/ICT service is also influenced by different factors which may include (but are not limited to) social trends, advertising, tariffs, and costs, which are interrelated with the customer expectation of QoS. For example, social trends may be in terms of popular devices (e.g. smartphones), services (e.g. some services are more popular than other similar services over the Internet), applications (e.g. there is different popularity of different applications in their ecosystems), etc. Further, end-user perception of the QoS is not limited only to objective characteristics at the man-machine interface. For end-users, the QoS refers to the quality that they personally experience during their use of a given telecommunication service. The end-user may be satisfied with the QoS of a given service at a certain time and the same user may not be satisfied with the same service 10 years in the future. For example, when a user accommodates to higher resolution of TV or video streaming due to higher bitrates in the access networks, then the same user will increase the expectations of such a service (i.e. TV or video streaming).

Figure 1.4 illustrates how the QoS depends on the end-to-end technical aspects, which include network performance and terminal performance, and non-technical aspects (those that are not directly related to the equipment), which include customer care and point of sale.

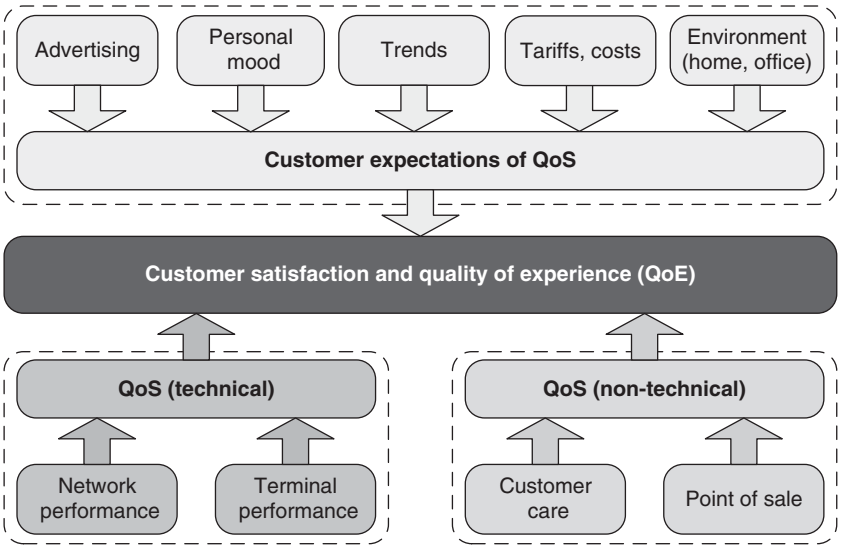

Figure 1.4 Technical and non-technical points of view for quality of service.

1.3.2 Quality of Experience (QoE)

Initially QoE was defined as the overall acceptability of an application or service, as perceived subjectively by the end-user (according to ITU-T Recommendation P.10/G.100, [11]).

ITU-T has replaced the QoE definition developed in 2007 with a new definition adopted in 2016, which is currently the actual QoE definition given as follows [11]:

> Quality of experience (QoE) is the degree of delight or annoyance of the user of an application or service.

However, one should note that there is continuous research on the QoE topic, so the definition is expected to evolve further.

Regarding the QoE there are two topics that need to be addressed together with the definition:

- *QoE influencing factors*. They include the type and characteristics of the application or service, context of use, the user's expectations with respect to the application or service and their fulfillment, the user's cultural background, socio-economic issues, psychological profiles, emotional state of the user, and other similar factors.
- *QoE assessment*. This is the process of measuring or estimating the QoE for a given number of end-users of a given application or a service. The assessment is typically based on an established procedure, and taking into account all important influencing factors. The output of the QoE assessment may result in a scalar value, various multi-dimensional representations of the results, as well as verbal descriptors (e.g. good, bad). In theory, all assessments of QoE should be accompanied by the description of the influencing factors that are included.

Overall, the QoE includes complete end-to-end system effects (end-user equipment and application, various influencing factors on the user for the given services, as well as network and service infrastructure). So, the QoE may be influenced by user expectations and the context (e.g. for the same obtained bitrate for a given service, a user who had lower expectations will enjoy higher QoE than a user who had higher expectations for the same service on the same equipment offered via the same network). QoE takes into consideration certain additional parameters, such as:

- user expectations;
- user context (what is in trend, user's personal mood, environment where the service is being used such as work/home/outside environments, etc.);
- the potential difference between the service being offered to the user and the individual user awareness about the service and additional features (if any) for that service.

Regarding the QoE assessment, the most used measure is the mean opinion score (MOS). Initially, the MOS scale referred to voice service only (ITU-T P.800), but nowadays it is also used for other services such as video (e.g. for IPTV). MOS is expressed as a single number in the range between 1 and 5, where MOS with a value of 1 denotes the worst and a value of 5 denotes the best quality experienced by the user (Table 1.1).

1.3.3 Network Performance (NP)

Network performance differs from QoS because it relates only to the network part of the service provisioning, without taking into account different user influencing factors. On one side, the QoS is the outcome of the user's experience of using a given service and the user's perception of it, while on the other side the NP is determined by the performances of network elements one by one, or by the performance of the network as a whole. So, the NP has an influence on the QoS, and it represents a part of it. However, the QoS is not influenced only by the NP but also by non-network performance parameters. Simply said, QoS consists of network performance and non-network performance, as shown in Figure 1.5.

The NP concept is applied for purely technical purposes, i.e. assessment and analysis of technical functions. NP is the ability of a network portion or the whole network to provide the QoS functions related to communications between the users. Network performance is determined by the performance of network elements one-by-one. The

Table 1.1 Mean opinion score for the quality of experience.

Mean opinion score (MOS)	Quality classification
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

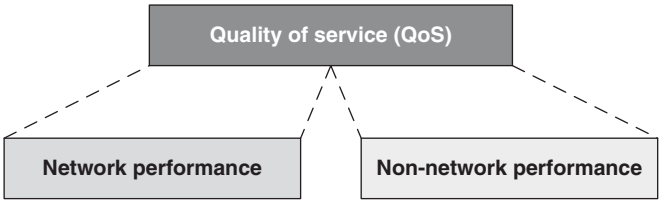

Figure 1.5 Network performance (NP) and quality of service (QoS).

performance of the network as a whole (end-to-end) is determined by the combination of the performance of all single elements along with their interconnections on the communication path between the end-user devices. NP is specified in terms of objective performance parameters, which can be measured, and then the performance value is assigned quantitatively [10].

1.3.4 QoS, QoE, and NP Relations

All three terms – QoS, QoE, and NP – are related one to another. QoE is different from QoS and NP as it has a subjective nature by definition, and because it depends on the end-user's perception. Clearly QoE is impacted by QoS and NP. For example, if NP is lower, it will result in lower quality experienced by the end-user.

Further, NP applies to various aspects of the network provider's functioning, including planning, development, operations, and maintenance of network elements and the whole network. Also, there can be several interconnected networks (along the path between the endpoints of the established communication session), which may be operated by different network providers (i.e. telecom operators). As shown in Figure 1.6,

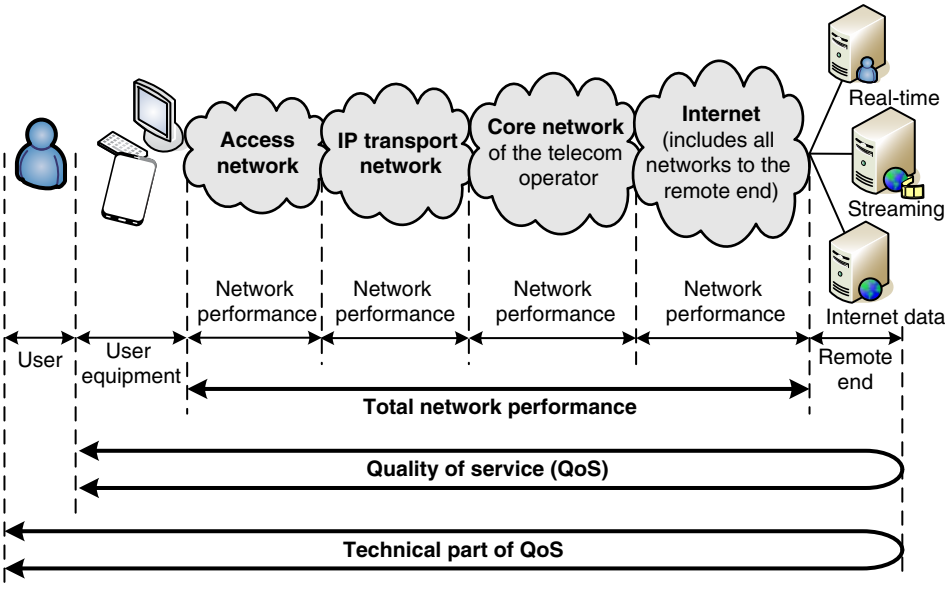

Figure 1.6 Network performance, QoS, and QoE.

NP is the detailed technical part of the offered QoS. Also, NP contributes to QoS as experienced by the user [12]. The functions of a service depend on the performance of the network elements and the performance of the user's terminal equipment. QoS is always end-to-end, which can be user-to-user or user-to-content (one may also add here QoS for machine-to-machine communications, which may be directly not related to a human end-user). Hence, QoS is always an end-to-end characteristic, where it depends on the contributions from different components (Figure 1.6), including the end-user, its equipment (smartphone, computer, etc.), access network (fixed or mobile), IP transport network, core network, and the rest of the path end-to-end (e.g. through the Internet). QoE has a broader scope because it is impacted by QoS as well as by user expectations and the context. Meanwhile, QoS has broader scope than NP.

1.4 ITU's QoS Framework

ITU has developed a QoS framework to suit different networks and services, initially defined in ITU-T G.1000 [12] and then in ITU-T 802 [13]. The framework is continuously evolving. To provide QoS support for a given service, QoS criteria and then QoS parameters based on the criteria need to be defined. Such definitions of QoS criteria were initially given in ITU-T recommendation G.1000, which provides the general QoS framework (by the ITU). Before defining QoS parameters, QoS criteria relevant to the user are required. The main goal is to establish a list of all aspects that could influence the QoS. There are three models for this purpose [13]:

- *Universal model*. This is a generic and a conceptual model.
- *Performance model*. This model is more suited for determining the performance criteria of a telecommunication service.
- *Four-market model*. This model is particularly suited for multimedia services.

1.4.1 Universal Model

In this model all QoS criteria are grouped under four categories:

- *Performance criteria*. These cover the technical parts of the service, and can be qualitative or quantitative (or both). They are further defined within the performance model (next section).
- *Esthetic criteria*. These refer to ease of interaction between the user and the telecommunication service (ergonomic aspects, ease of use of the service, style and look, design of functionalities of the service, etc.).
- *Presentational criteria*. These reflect the presentational aspects of the service to the customer, including the packaging of the service, tariff models for that service, billing options for the service, etc.).
- *Ethical criteria*. These are related to the ethical aspects, such as conditions specified for cutting off the given service, services for disabled users, etc.

The service is split into functional elements, where each of the is cross-checked against the given four quality components and criteria. However, the definitions and measurement methods of the quality parameters are not a part of the universal model, but they are defined within the performance model.

1.4.2 Performance Model

The main aim in the performance model is to determine performance criteria which are further used for defining QoS parameters important to both users and providers [13]. In total, there are seven specified QoS criteria [13], as shown in Table 1.2. They are identified to provide easy translation into QoS parameters.

The given quality criteria (Table 1.2) are mapped on a set of service functions, which include service management (sales and pre-contract activities, service provision, alteration, service support, repair, and cessation), connection quality (connection establishment, information transfer, connection release), billing, and network or service management by the customer. The mapping between the service functions and service quality criteria is referred to as a performance model (in ITU-T E.802).

The QoE is influenced by all seven QoS criteria given in the performance model. For example, speed impacts the available bitrates (in downlink and uplink) and latencies (i.e. delays), which is crucial for the end-user's experience of the service. Upgrading to higher access bitrates (including fixed and mobile networks) improves the overall QoE. Further, availability and reliability are also very important and are directly related to planning and dimensioning of the network (for a given number of users and for a given service) as well as to its operation and maintenance functions and procedures. For example, one typical quality metric for network availability which is used in the traditional telecom world (at the end of the twentieth century) is so-called "five nines," that is, 99.999% of the time service to be available to the end-users, and that poses certain requirements regarding the survivability mechanisms which need to be implemented in the network (e.g. automatic traffic redirections over alternative paths in the network in a case of link or path failures). Also, one may note that security aspects – accuracy (e.g. billing accuracy), simplicity of use of the service (the user should not be required to read a manual in order to use an offered telecommunication service), and flexibility regarding the service use (e.g. ease of change of tariff model or billing method, or even changing the operator in a case of QoS degradations) – influence the QoE.

Table 1.2 QoS criteria.

No.	QoS criteria	Applying QoS criteria to different service functions
1	Speed	Service supply time, call setup time, one-way delay, release time, billing frequency, etc.
2	Accuracy	Unsuccessful call ratio, speech quality, bill correctness, etc.
3	Availability	Coverage, availability of call center, service availability, etc.
4	Reliability	Dropped calls ratio, number of billing complaints within a specific time period, etc.
5	Security	Fraud protection and prevention, etc.
6	Simplicity	Professionalism of help line (i.e. customer care), ease of software updates, ease of contract cessation procedure, etc.
7	Flexibility	Ease of change in contract, availability of different billing methods including online and offline billing, etc.

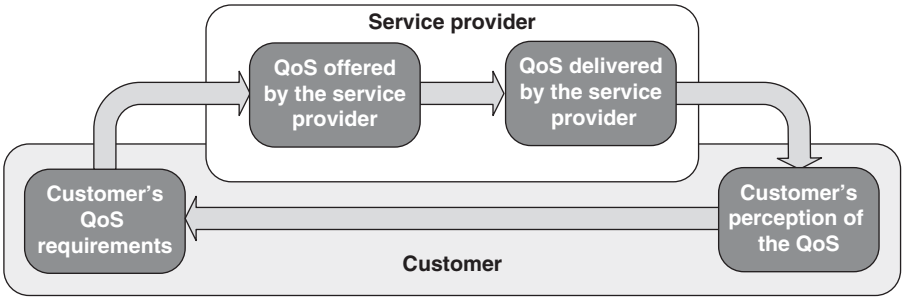

Figure 1.7 QoS viewpoints.

The ITU's performance model is based on four viewpoints of QoS initially defined in ITU-T G.1000 [12]. These fours viewpoints cover QoS from both customer and service aspects, as given in Figure 1.7:

- *Customer's QoS requirements*. This is the QoS level required by the subscriber, which may be specified in non-technical language also (e.g. good service is required), because the customer is interested not in how a given service is provided (e.g. by the telecom operator's network) but primarily in the obtained end-to-end QoS (expressed in terms they can understand, such as bitrates, data volume, etc.).
- *QoS offered by the provider (or planned/targeted QoS)*. This is a statement made by the service provider (e.g. telecom operator) to customers about the QoS offered for a given service. This viewpoint is primarily used for a service level agreement (SLA), which serves as a bilateral agreement signed between the customer and the service provider. This QoS can be specified in terms understandable to the customer on one side and with technical terms for the purposes of implementation of such QoS on the side of the operator. This can also serve as a merit for subscribers to make the best choice from the given service provider's offerings.
- *QoS achieved (i.e. delivered) by the provider*. This viewpoint refers to the actual level of QoS achieved or delivered by the service provider, and for purposes of comparison it should be expressed through the same QoS parameters as the QoS offered to the customer (e.g. specified in the SLA). This QoS viewpoint can be used by the regulator for the purposes of QoS regulation, including publication of the results from QoS audit in the telecom operators' networks and then QoS encouragement or enforcement (when and where needed).
- *Customer QoS perception*. This is the QoS level experienced by the customer, typically obtained from user ratings of the QoS provided by the service operator. This is also a customer viewpoint, so it is not expressed in technical terms but in terms of degrees of satisfaction (e.g. from "not satisfied" up to "very satisfied"). For example, a customer may rate the service on a 5-point scale, with grade 1 being the worst and grade 5 being the best experience with the service. The perceived QoS can be used by service providers or regulators to determine the customer's satisfaction, which may further lead to corrective actions by providers or regulators (e.g. in situations when there is a significant mismatch between the perceived QoS and the QoS offered by the provider).