

Quality of Service and Quality of Experience: Network Design Implications

by William Stallings, Independent Consultant
Florence Agboma, BSKyB

The Internet and enterprise IP-based networks continue to see rapid growth in the volume and variety of data traffic. Cloud computing, big data, the pervasive use of mobile devices on enterprise networks, and the increasing use of video and image traffic all contribute to the increasing difficulty in maintaining satisfactory network performance. Two key tools in measuring the network performance that an enterprise desires to achieve are *Quality of Service* (QoS) and *Quality of Experience* (QoE). QoS and QoE enable the network manager to determine if the network is meeting user needs and to diagnose problem areas that require adjustment to network management and network traffic control. This article provides an overview of QoS and QoE concepts, the relationship between the two, and the design implications of a QoE/QoS architecture.

Background and Overview

Historically, the Internet and other IP-based networks provided a *best-effort* delivery service. The term “best effort” refers to the connectionless, datagram nature of the interconnected set of networks. With best effort, a packet may be lost, duplicated, delayed, or delivered out of order, and the network does not inform sender or receiver. Traditionally, a best-effort network attempts to allocate its resources with equal availability and priority to all traffic flows, with no regard for application priorities, traffic patterns and load, or customer requirements. To protect the network from congestion collapse and to guarantee that some flows do not crowd out other flows, congestion-control mechanisms were introduced; they tended to throttle traffic that consumed excessive resources. As the intensity and variety of traffic increased, various QoS mechanisms were developed, including *Integrated Services Architecture* (ISA) and *Differentiated Services* (DiffServ) (for example, see [1]). *Service-Level Agreements* (SLAs) were also developed so that the service provided to various customers was tunable and somewhat predictable. These mechanisms and services serve two purposes: (1) allocate network resources efficiently so as to maximize effective capacity; and (2) enable networks to offer customers different levels of QoS on the basis of their requirements.

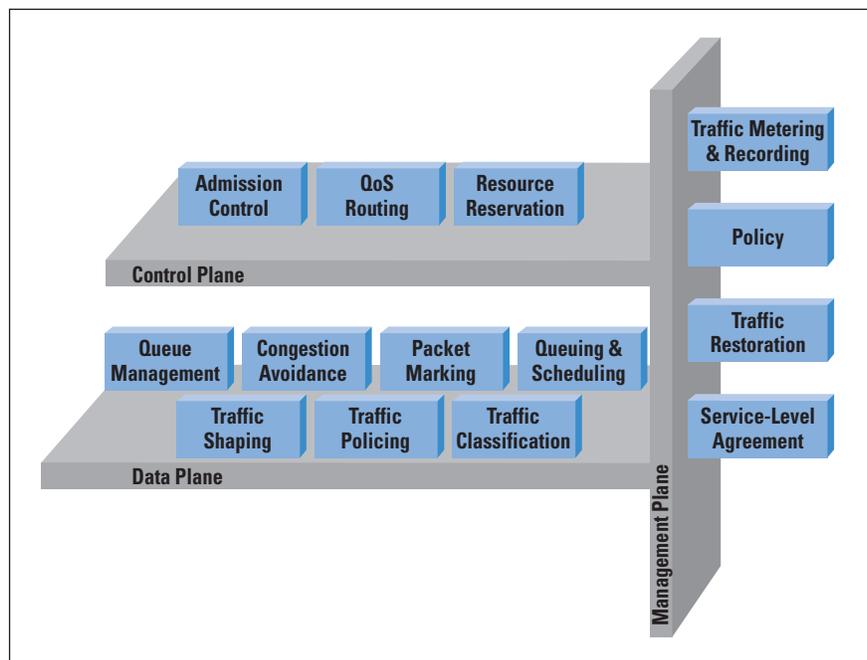
QoS is an important but increasingly insufficient tool for providing network services for many of today’s high-volume applications. To meet the needs of such applications, QoS has recently been augmented with the concept of QoE, which is a subjective measure of performance as reported by the user. Unlike QoS, which can be precisely measured, QoE relies on human opinion. QoE is important particularly when dealing with multimedia applications and multimedia content delivery.

Because QoE extends the concept of QoS to more fully tailor network services and performance to customer and user needs, it is garnering increasing attention by network protocol and system designers. The management of QoE has become a crucial concept in the deployment of future successful applications, services, and products. The greatest challenges in providing QoE are developing effective methods for converting QoE features to quantitative measures and translating QoE measures to QoS measures. Whereas now it is easy to measure, monitor, and control QoS at both the networking and application layers, and at both the end system and network sides, its management is still quite intricate.

QoS Architectural Framework

The *Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T) Recommendation Y.1291* provides an overall architectural framework that relates the various elements that go into QoS provision^[2]. Figure 1 shows the relationship between these elements, which are organized into three planes: data, control, and management. This architectural framework is an excellent overview of QoS functions and their relationships and provides a useful basis for summarizing QoS.

Figure 1: Architectural Framework for QoS Support



The *Data Plane* includes those mechanisms that operate directly on flows of data. The following discussion briefly describes each mechanism in turn.

Queue Management algorithms manage the length of packet queues by dropping packets when necessary or appropriate. Active management of queues is concerned primarily with congestion avoidance. In the early days of the Internet, the queue management discipline was to drop any incoming packets when the queue was full; it was referred to as the *tail-drop* technique. This technique has many drawbacks, including^[3]:

- There is no reaction to congestion until it is necessary to drop packets, whereas a more aggressive congestion-avoidance technique would likely improve overall network performance.
- Queues tend to be close to full, causing an increase in packet delay through a network and possibly resulting in a large batch of dropped packets for bursty traffic, necessitating many packet retransmissions.
- Tail drop may allow a single connection or a few flows to monopolize queue space, preventing other connections from getting room in the queue.

One noteworthy example of queue management is *Random Early Detection* (RED)^[4]. RED drops incoming packets probabilistically based on an estimated average queue size. The probability for dropping increases as the estimated average queue size grows.

Queuing and Scheduling Algorithms, also referred to as “queuing discipline algorithms,” determine which packet to send next; they are used primarily to manage the allocation of transmission capacity among flows.

Congestion Avoidance deals with means for keeping the load of the network sufficiently under its capacity such that the network can operate at an acceptable performance level. The specific objectives are to avoid significant queuing delays and, especially, to avoid congestion collapse.

Packet Marking encompasses two distinct functions. First, packets may be marked by network edge nodes to indicate some form of QoS that the packet should receive. An example is the *DiffServ* (DS) field in the IPv4 and IPv6 packets and the *Traffic Class* field in *Multiprotocol Label Switching* (MPLS) labels^[5]. An edge node can set the values in these fields to indicate a desired QoS. Such markings may be used by intermediate nodes to provide differential treatment to incoming packets. Second, packet marking can be used to mark packets as nonconforming; such packets may be dropped later if congestion occurs.

Traffic Classification refers to the assignment of packets to a traffic class at the edge of the network. Typically, the classification entity looks at multiple fields of a packet, such as source and destination address, application payload, and QoS markings, and determines the aggregate to which the packet belongs. This classification provides network elements a method to weigh the relative importance of one packet over another in a different class. All traffic assigned to a particular flow or other aggregate can be treated similarly. The flow label in the IPv6 header can be used for traffic classification.

Traffic Policing determines whether the traffic being presented, is on a hop-by-hop basis compliant, with prenegotiated policies or contracts. Nonconforming packets may be dropped, delayed, or labeled as nonconforming. As an example, ITU-T Recommendation Y.1221^[6] recommends the use of token bucket to characterize traffic for purposes of traffic policing.

Traffic Shaping controls the rate and volume of traffic entering and transiting the network on a per-flow basis. The entity responsible for traffic shaping buffers nonconforming packets until it brings the respective aggregate in compliance with the traffic. The resulted traffic thus is not as bursty as the original and is more predictable. For example, Y.1221 recommends the use of leaky bucket and/or token bucket for traffic shaping.

The *Control Plane* is concerned with creating and managing the pathways through which user data flows. *Admission Control* determines what user traffic may enter the network. This decision may be in part determined by the QoS requirements of a data flow compared to the current resource commitment within the network. But beyond balancing QoS requests with available capacity to determine whether to accept a request, there are other considerations for admission control. Specifically, network managers and service providers must be able to monitor, control, and enforce use of network resources and services based on policies derived from criteria such as the identity of users and applications, traffic/bandwidth requirements, security considerations, and time of day or week. RFC 2753^[7] discusses such policy-related issues.

QoS Routing determines a network path that is likely to accommodate the requested QoS of a flow. This function contrasts with the philosophy of traditional routing protocols, which generally are looking for a least-cost path through the network. RFC 2386^[8] provides an overview of the issues involved in QoS routing, which is an area of ongoing study.

Resource Reservation reserves network resources on demand for delivering desired network performance to a requesting flow. The resource-reservation mechanism that has been implemented for the Internet is the *Resource Reservation Protocol (RSVP)*^[9].

The *Management Plane* contains mechanisms that affect both control- and data-plane mechanisms. It deals with the operation, administration, and management aspects of the network. A *Service-Level Agreement* (SLA) is the agreement between a customer and a provider of a service that specifies the level of availability, serviceability, performance, operation, or other attributes of the service. SLAs are discussed subsequently.

Traffic Metering and Recording concerns monitoring the dynamic properties of a traffic stream using performance metrics such as data rate and packet-loss rate. It involves observing traffic characteristics at a given network point and collecting and storing the traffic information for analysis and further action. Depending on the conformance level, a meter can invoke necessary treatment (for example, dropping or shaping) for the packet stream. A subsequent section discusses the types of metrics that are used in this function.

Traffic Restoration refers to the network response to failures. It encompasses numerous protocol layers and techniques.

Policy refers to a set of rules for administering, managing, and controlling access to network resources. The rules can be specific to the needs of the service provider or reflect the agreement between the customer and service provider, which may include reliability and availability requirements over a period of time and other QoS requirements.

Service-Level Agreements

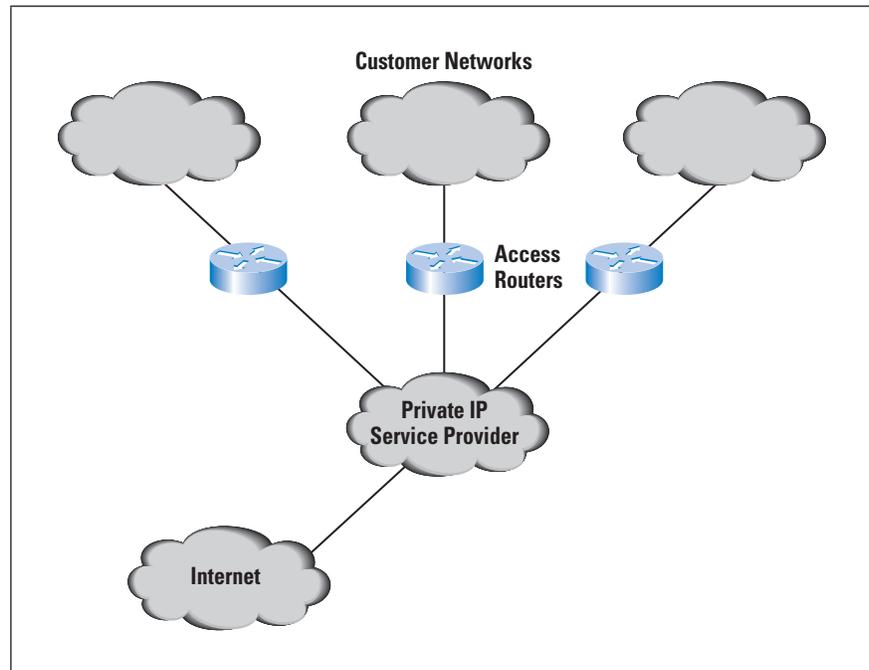
An SLA is a contract between a network provider and a customer that defines specific aspects of the service that is to be provided. The definition is formal and typically defines quantitative thresholds that must be met. An SLA typically includes the following information:

- *A description of the nature of service to be provided:* A basic service would be IP-based network connectivity of enterprise locations plus access to the Internet. The service may include additional functions such as web hosting, maintenance of domain name servers, and operation and maintenance tasks.
- *The expected performance level of the service:* The SLA defines numerous metrics, such as delay, reliability, and availability, with numerical thresholds.
- *The process for monitoring and reporting the service level:* This function describes how performance levels are measured and reported.

The types of service parameters included in an SLA for an IP network are similar to those provided for *Frame Relay* and *Asynchronous Transfer Mode* (ATM) networks. A key difference is that, because of the unreliable datagram nature of an IP network, it is more difficult to realize tightly defined constraints on performance, compared to the connection-oriented Frame Relay and ATM networks.

Figure 2 shows a typical configuration that lends itself to an SLA. In this case, a network service provider maintains an IP-based network. A customer has many private networks (for example, LANs) at various sites. Customer networks are connected to the provider via access routers at the access points. The SLA dictates service and performance levels for traffic between access routers across the provider network. In addition, the provider network links to the Internet and thus provides Internet access for the enterprise.

Figure 2: Typical Framework for Service-Level Agreement



For example, the standard SLA provided by Cogent Communications for its backbone networks includes the following items:

- *Availability*: 100% availability
- *Latency (delay)*: Monthly average Network Latency for packets carried over the COGENT Network between Backbone Hubs for the following regions: Intra-North America: ≤ 45 ms; Intra-Europe: ≤ 35 ms; New York to London: ≤ 85 ms; Los Angeles to Tokyo: ≤ 120 ms.

Latency is defined as the average time taken for an IP packet to make a round trip between Backbone Hubs within a region. COGENT monitors aggregate latency within the COGENT Network by monitoring round-trip times between a sample of Backbone Hubs on an ongoing basis.

- *Network Packet Delivery (reliability)*: Average monthly Packet Loss no greater than 0.1% (or successful delivery of 99.9% of packets). Packet Loss is defined as the percentage of packets that are dropped between Backbone Hubs on the COGENT Network.

An SLA can be defined for the overall network service. In addition, SLAs can be defined for specific end-to-end services available across the carrier's network, such as a virtual private network, or differentiated services.

IP Performance Metrics

The *IP Performance Metrics Working Group* (IPPM) is chartered by the *Internet Engineering Task Force* (IETF) to develop standard metrics that relate to the quality, performance, and reliability of Internet data delivery. Two trends dictate the need for such a standardized measurement scheme:

- The Internet has grown and continues to grow at a dramatic rate. Its topology is increasingly complex. As its capacity has grown, the load on the Internet has grown at an even faster rate. IP-based enterprise networks have exhibited similar growth in complexity, capacity, and load. The sheer scale of these networks makes it difficult to determine quality, performance, and reliability characteristics.
- The Internet serves a large and growing number of commercial and personal users across an expanding spectrum of applications. Similarly, private networks are growing in terms of user base and range of applications. Some of these applications are sensitive to particular QoS parameters, leading users to require accurate and understandable performance metrics.

A standardized and effective set of metrics enables users and service providers to have an accurate common understanding of the performance of the Internet and private internets. Measurement data is useful for a variety of purposes, including:

- Supporting capacity planning and troubleshooting of large complex internets
- Encouraging competition by providing uniform comparison metrics across service providers
- Supporting Internet research in such areas as protocol design, congestion control, and quality of service
- Verification of service-level agreements

The metrics are defined in three stages:

- *Singleton metric*: The most elementary, or atomic, quantity that can be measured for a given performance metric. For example, for a delay metric, a singleton metric is the delay experienced by a single packet.

- *Sample metric*: A collection of singleton measurements taken during a given time period. For example, for a delay metric, a sample metric is the set of delay values for all of the measurements taken during a one-hour period.
- *Statistical metric*: A value derived from a given sample metric by computing some statistic of the values defined by the singleton metric on the sample. For example, the mean of all the one-way delay values on a sample might be defined as a statistical metric.

The measurement technique can be either *active* or *passive*.

Active techniques require injecting packets into the network for the sole purpose of measurement. This approach has several drawbacks. The load on the network is increased, in turn possibly affecting the desired result. For example, on a heavily loaded network, the injection of measurement packets can increase network delay, so that the measured delay is greater than it would be without the measurement traffic. In addition, an active measurement policy can be abused for denial-of-service attacks disguised as legitimate measurement activity.

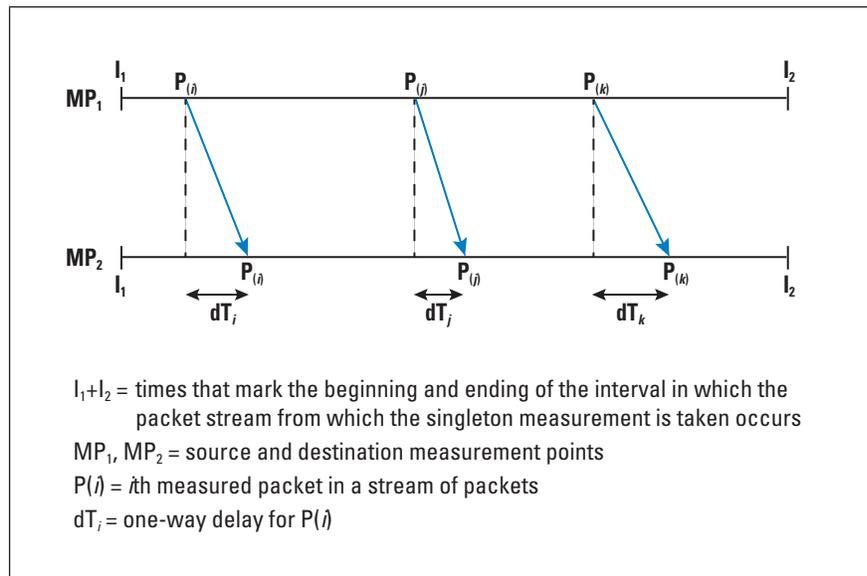
Passive techniques observe and extract metrics from existing traffic. This approach can expose the contents of Internet traffic to unintended recipients, creating security and privacy concerns. So far, the metrics defined by the IPPM working group are all active.

For the sample metrics, the simplest technique is to take measurements at fixed time intervals, known as periodic sampling. There are several problems with this approach. First, if the traffic on the network exhibits periodic behavior, with a period that is an integer multiple of the sampling period (or vice versa), correlation effects may result in inaccurate values.

Also, the act of measurement can perturb what is being measured (for example, injecting measurement traffic into a network alters the congestion level of the network), and repeated periodic perturbations can drive a network into a state of synchronization (for example, [10]), greatly magnifying what might individually be minor effects. Accordingly, RFC 2330^[11] recommends Poisson sampling. This method uses a Poisson distribution to generate random time intervals with the desired mean value.

Figure 3 illustrates the packet-delay variation metric. This metric is used to measure jitter, or variability, in the delay of packets traversing the network. The singleton metric is defined by selecting two packet measurements and measuring the difference in the two delays. The statistical measures use the absolute values of the delays.

Figure 3: Model for Defining Packet-Delay Variation



QoS for Streaming Video

It is worthwhile to comment on the relationship between QoS concerns and streaming video, which perhaps accounts for the greatest volume of traffic on the Internet. First, consider the network transmission requirements for real-time video, such as video teleconferencing. For such applications QoS metrics such as latency and jitter are important, and they impose significant requirements on the networks through which the real-time traffic passes.

For streaming video, in contrast, such QoS requirements can be relaxed by equipping the application to handle a wider range of responses. A motivation for this approach is that QoS is not ubiquitously deployed in networks and thus is not available for all possible streams between content deliverers and content consumers. In the case of streaming video in particular, applications have come to assume that the receiving device is equipped with generous levels of memory, and the application uses this memory as a playback buffer. This system allows the application to use TCP for reliable delivery. The TCP connection can be exposed to far higher levels of network jitter and even packet loss than would be tolerable by a real-time packet stream. Thus, to a significant extent, this approach to video streaming avoids the need to use QoS as a strict precondition for streaming video over the Internet.

From QoS to QoE

The literature contains numerous different, though similar, definitions of QoE. To provide a common working definition, the EU-sponsored *European Network on Quality of Experience in Multimedia Systems and Services* (see [12]) has developed a definition of QoE that reflects broad industry and academic consensus:

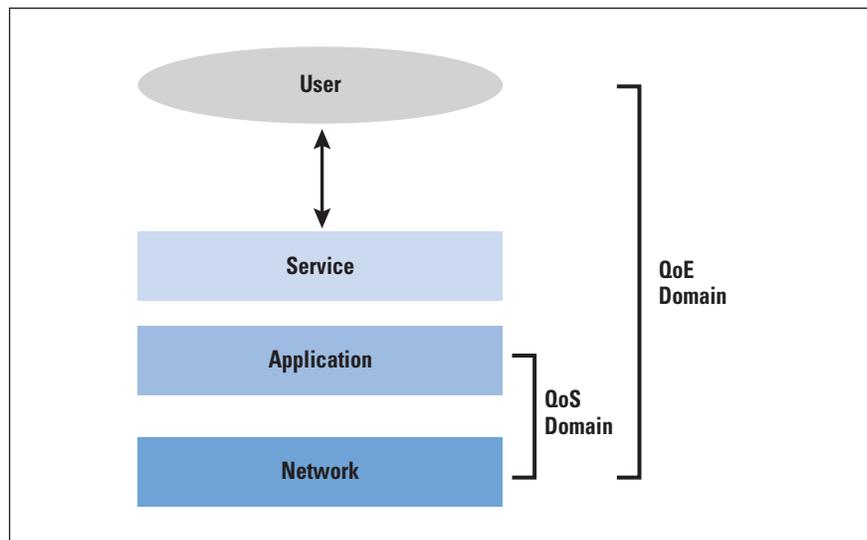
“Quality of Experience (QoE) is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user’s personality and current state.”

The QoE methods to measure this degree of delight or annoyance are discussed later in this article.

The Layers of QoE

For any type of service (for example, IPTV), multiple QoS parameters contribute to the overall user’s perception of quality. The concept of QoE in addition to QoS mechanisms has been proposed in [13], [14], [15], and [16] as a QoE/QoS layered approach by which the requirements of the users drive network-dimensioning strategies. The QoE/QoS layered approach does not substitute the QoS aspect of the network, but instead, user and service-level perspectives are both complementary as shown in Figure 4.

Figure 4: QoE/QoS Layered Model



The levels from the layered approach follow:

- *User*: The user interacts with the service. It is the user’s degree of delight or annoyance from using the service that is to be measured.
- *Service*: This level is the virtual level where the user’s experience of the overall performance of the service can be measured. It is the interface where the user interacts with the service (for example, the visual display to the user).
- *Application-level QoS (AQoS)*: This level deals with the control of application-specific parameters such as content resolution, bitrate, frame rate, colour depth, codec type, layering strategy, sampling rate, and number of channels.

- *Network-level QoS (NQoS)*: This level is concerned with the low-level network parameters such as service coverage, bandwidth, delay, throughput, and packet loss.

Background on the Layers of QoE

Being linked to human perception, QoE is hard to describe quantitatively, and it varies from person to person. The complexities of QoE at the user level stem from the differences between individual user characteristics, of which some might be of a variant nature (that is, changing often with time), whilst others are of a relatively stable nature. Examples could include gender, age, attitudes, prior experience, expectations, socio-economic status, cultural background, educational level, etc. Thus, it becomes a challenge to derive unified QoE metrics for all users and their contexts. Reiter et al.^[17] provide more details on the human factors that may influence QoE. The current practice in any QoE measurement is to identify and control the relatively stable characteristics of a user in a way that is satisfactory to at least a large proportion of the potential user group.

Multiple layers might impact the user QoE for any application. The service level provides a virtual level where the user's tolerance thresholds for a particular service could be measured. As an illustration, the QoE measures from the user perspective for streaming applications could be a) start-up time, b) audiovisual quality, c) channel-change delay, and d) buffering interruptions, whilst the QoE measures for web-browsing applications could be waiting times.

Quite often, the network capacity will dictate the bandwidth that may be used for transmission. At the application-level QoS, there might be a trade-off between quality and size. For audio, a higher sampling rate, for example, 96 kHz, might allow for more information to be perceived compared with 48 kHz, but at the expense of a bigger file size. Traditional telephony speech might be limited to 8 kHz because of the bandwidth capacity. For video, a high resolution might require more bandwidth than low resolution. There are huge varieties of device screens in all kinds of sizes, featuring varied aspect ratios. The one commonality in this array of equipment is that they are all capable of rescaling the video, as is the case for a native player going to the full-screen mode. For a given bitrate, there might be a trade-off between lower resolutions in pixels (images being slightly blurred) with fewer artifacts versus higher resolutions that provide a sharper image but possibly with more artifacts. Most compression standards might use a block-based and motion-compensation coding scheme, and as a result additional compression artifacts are added to the decoded video.

Network-level QoS parameters could impact QoE in a variety of ways.

The network delay could impact QoE, especially for interactive services. For instance, the interactive nature of web browsing that requires multiple retrieval events within a certain window of time might be affected by delay variations of the network. *Voice over IP* (VoIP) services might have the stringent response-time demands, whereas email services might tolerate much longer delays. The different distribution methods of streaming video over the network might affect QoE in different ways. HTTP-based adaptive streaming, which uses TCP, might react to bandwidth constraints and CPU capacity in two ways: either the streaming switches between streaming the different bitrate encodings depending on available resources, or a frame freeze (rebuffering) due to incoming packet starvation in the player buffer. The continuous bitrate switches and rebuffering affect QoE badly. The other distribution method, *User Datagram Protocol* (UDP), might use multicast to replicate the streams throughout the network. Quite often, a resilient coding scheme and a flow-control mechanism might be implemented to maintain the viewing experience despite the effects of poor network conditions.

For reasons that should now be apparent, the background on the layers of QoE suggests that the effect of QoE could be an attribute of only the application layer or a combination of both the application and network layers. Although the trade-offs between quality and network capacity may begin with application-level QoS due to network capacity considerations, an understanding of the user requirements at the service level (that is, in terms of QoE measures) would enable a better choice of application-level QoS parameters to be mapped onto the network-level QoS parameters. A scenario that aims at controlling QoE using QoS parameters as actuators is discussed later in this article. Taking the QoE/QoS approach as a whole entity rather than single entities might aid in providing better QoEs, and potentially could lead to a better-managed delivery infrastructure.

Key Factors That Determine QoE

The nature of QoE, which comprises many layers of interaction between the enabling elements of service delivery and the human user, makes measuring and improving user experiences a challenging task. To understand QoE we must account for both technical and nontechnical factors. Many factors contribute to producing a good QoE. Moller et al.^[18] provide useful perspectives on factors that influence QoE. Here, we discuss the following key factors that might influence QoE:

User Demographics: The context of demographics herein refers to the relatively stable characteristics of a user that might indirectly influence perception, and intimately affects other technical factors to determine QoE. The findings from [19] suggest that user demographics may influence QoE. In the context of adopting HD voice telephony, the different user groups had significantly different quality ratings.

The grouping of users was based on demographic characteristics such as their attitudes towards adoption of new technologies, socio-demographic information, socio-economic status, and prior knowledge. Cultural background is another user demographic factor that might also influence perception due to cultural attitude to quality^[20].

Type of Device: Different device types possess varying characteristics that may affect QoE. An application designed to run on more than one device type, for example on a connected TV device such as Roku and on an iOS device such as an iPhone, may not deliver the same QoE on every device.

Content: The content being distributed via Internet delivery can range from interactive content specifically curated according to personal interests to content that is produced for linear TV transmission. The different characteristics of the video might require different system properties in terms of the quality being produced. According to [21], people tend to watch *Video on-Demand* (VoD) content with a higher level of engagement than its competing alternative, linear TV. This trend may be because users make an active decision to watch a particular VoD content, and as a result, give their full attention to it. One could infer that for VoD, users might be less tolerant of any quality degradations because of their high level of engagement.

Connection Type: The type of connection used to access the service influences users' expectations and their QoEs. Users have been found to have lower expectations when using 3G in contrast to a wireline connection when in fact both connection types were identical in terms of their technical conditions^[22]. Agboma et al.^[23] found users' expectations to be considerably lowered and more tolerant to visual impairments on small devices. Conventional QoS management practices cannot account for these psychological factors.

Media (audiovisual) Quality: This factor may be observed as a significant one affecting QoE, because it is the part of a service that users notice most. The integration of audio and video quality appears to be content-dependent. For less-complex scenes (for example, head and shoulder content), audio quality is slightly more important than video quality. In contrast, for high-motion content, video quality has been found to be significantly more important than audio quality^[24]. Other studies^[25] suggest that the optimum audio/video bitrate allocation depends on scene complexity. For instance, visually complex scenes would benefit from the allocation of higher bitrates, with relatively more bits allocated towards audio, because high-audio bitrates seem to produce the best overall audiovisual quality.

Network: Content delivery via the open Internet is highly susceptible to the effects of delays, jitter, packet loss, and available bandwidth. For users, delay and jitter cause frame freeze and the lack of lip synchronization between what is heard (audio) and what is seen (video). Content delivery is guaranteed using a TCP/IP delivery mechanism. However, under bad network conditions, the frequency of rebuffering, and the implementation of a video player heuristics, might affect QoE. Rebuffering interruptions in IP video playback is considered the worst degradation on user QoE and should be avoided at the cost of startup delay^[26]. On the same note, QoE for a given startup delay strongly depends on the concrete application context and user expectations^[27]. In spite of the different QoE factors that are concerned with the network, reliability and a strong wireless signal is crucial for consuming TV-like services anytime, anywhere, and from any device.

Usability: Another QoE factor is the amount of effort that is required to navigate through the service. The design should render good quality without a great deal of technical input from the user before or after service consumption.

Cost: The long-established practice of judging quality by price implies that expectations are price-dependent. If the tariff for a certain service quality is high, users may be highly sensitive to any quality degradations. A scenario of QoE-based charging is demonstrated in [28] and [29] to analyze a situation where price is used to reflect the value of a service, and at the same time, part of the user-context factor. While [28] offers a trade-off between user expectations and price, its deployment may yield unexpected complexities.

QoE Measurement Methods

QoE measurement techniques evolved through the adaptation and application of psychophysics methods during the early stages of television systems (See [30] for more details). Here, we introduce three QoE measurement methods: subjective assessment methods, objective assessment methods, and end-user device analytics as an alternative to measure QoE. Hereafter, streaming video will be the focus of discussion.

Subjective Assessment

In *Subjective Assessment* of QoE, experiments are carefully designed to a high level of control (such as in a controlled laboratory, field tests, or crowdsourcing environments) so that the validity and reliability of the results can be trusted. It might be useful to consult expert advice during the initial design of the subjective experiment, because the topics of experimental design, experimental execution, and statistical analysis are complex. The different phases discussed in the following paragraphs are an abstract. A methodology to obtain subjective QoE data might consist of the following phases:

Characterize the Service: The task at this stage is to choose the QoE measures that affect the users' experience the most. As an example, consider a multimedia conferencing service; previous studies have shown that the quality of the voice takes precedence over the quality of video^[31]. Also, the video quality required for such applications does not demand a very high frame rate, provided that audio-to-video synchronization is maintained. Thus, the resolution of individual frames can be considerably lower than the case of other video streaming services, especially when the size of the screen is small (such as that of a mobile phone). Therefore, in multimedia conferencing, the QoE measures might be prioritized as voice quality, audio-video synchronization, and image quality.

Design and Define Test Matrix: Once the service has been characterized, the QoS factors that affect the QoE measures can be identified. For instance, the video quality in streaming services might be directly affected by network parameters such as bandwidth, packet loss, and encoding parameters such as frame rate, resolution, and codec. The capability of the rendering device will also play a significant role in terms of screen size and processing power. However, testing such a large combination of parameters may not be feasible. This draft matrix could be reduced to more achievable realistic test conditions by eliminating the combinations that have similar effects on QoE.

Specify Test Equipment and Materials: Subjective tests should be designed to specify test equipment that will allow controlled enforcement of the test matrix. For instance, to assess the correlation between NQoS parameters and the perceived QoE in a streaming application, at least a client device and a streaming server separated by an emulated network are needed. If the objective is to evaluate how different device capabilities impact QoE, then a video content is chosen to produce formats that can run in each of the client devices under scrutiny.

Identify Sample Population: A representative sample population is identified, possibly covering different classes of users categorized by the user demographics that are of interest to the experimenter. Depending on the target environment for the subjective test, at least 24 test subjects (that is, participants) have been suggested for a controlled environment (for example, a laboratory) and at least 35 test subjects for a public environment. Fewer subjects may be used for pilot studies to indicate trending^[32]. The use of *crowdsourcing* in the context of subjective assessment is still nascent, but it has the potential to further increase the size of the sample population and could reduce the completion time of the subjective test.

Subjective Methods: The recommendations include several subjective assessment methodologies, for example [33], [34], and [35]. A recent recommendation^[32] addresses the context of Internet video and distribution of quality television in any environment. These recommendations provide guidelines on topics such as the different subjective test designs, rating scales, and experiment durations.

There has also been some interest in developing alternative subjective methodologies for time-varying system characteristics, see [36] and [37]. Typically, each test subject is presented with the test conditions under scrutiny along with a set of rating scales that allows the correlation of the users' responses with the actual QoS test conditions being tested. There are several rating scales, depending on the design of the experiment. Other scale methods can be found in [32], and acceptance methods in [38].

Analysis of Results: When the test subjects have rated all QoS test conditions, a post-screening process might be applied to the data to remove any erroneous data from a test subject who appears to have voted randomly. Depending on the design of the experiment, a variety of statistical approaches could be used to analyze results. For the sake of brevity, statistical analysis of the results is outside the scope of this article. The most common quantification method is the *Mean Opinion Score* (MOS), which is the average of the opinions collected for a particular QoS test condition. The results from subjective assessment experiments are used to quantify QoE, and to model the impacts of QoS factors. Other authors^[39] have gone beyond deriving QoS MOS functions to QoE management. The rationale and limitations of MOS are discussed later in this article.

Subjective experiments require significant planning and design so as to produce reliable subjective MOS ratings. However, they are time-consuming and expensive to carry out and are not feasible for real-time in-service monitoring. Therefore, the use of objective models is often desirable.

Objective Assessment

In *Objective Assessment* of QoE, computational algorithms provide estimates of audio, video, and audiovisual quality as perceived by the user. Each objective model targets a specific service type. For example, ITU-T P.1201 predicts the impact of IP network impairments for IPTV applications and multimedia streaming applications^[40]. ITU-T J.341 targets video-quality measurement in *High-Definition Television* (HDTV) noninteractive applications^[41]. Other proprietary objective models do exist. For a given scope of quality features, the goal of any objective model is to find the optimum fit that strongly correlates with data obtained from subjective experiments. The following phases presented here should not be considered exhaustive, but they are meant to illustrate a process of obtaining objective QoE data. A methodology to obtain objective QoE data might consist of the following phases:

Database of Subjective Data: A starting point might be the collection of a group of subjective datasets as this list could serve as benchmark for training and verifying the performance of the objective model. A typical example of one of these datasets might be the subjective QoE data generated from well-established subjective testing procedures, as discussed earlier. The selection of the subjective datasets should typically reflect the use cases of the objective model.

Preparation of Objective Data: The data preparation for the objective model might typically include a combination of the same QoS test conditions as found in the subjective datasets, as well as other complex QoS conditions. A variety of pre-processing procedures might be applied to the video data prior to training and refinement of the algorithm.

Objective Methods: Various algorithms in the literature could provide estimates of audio, video, and audiovisual quality as perceived by the user. Some algorithms might be developed, and test subjects trained, for a specific perceived quality artifact, whilst other algorithms and test subjects might be trained for a wider scope of perceived quality artifacts. Examples of the perceived artifacts might include blurring, blockiness, unnatural motion, pausing, skipping, rebuffering, and imperfect error concealment after transmission errors. See [42] for a good overview, and for the development of objective video-quality prediction.

Verification of Results: After the objective algorithm has processed all QoS test conditions, the predicted values might benefit from a post-screening process to remove any outliers, the same concept as applied to the subjective datasets. The predicted values from the objective algorithm might be in a different dimension compared to the subjective QoE datasets. The predicted values might be transformed to the same scale as obtained in the subjective experiments (for example, into the MOSs) to allow for a linear comparison, and so that the optimum fit between the predicted QoE values and subjective QoE data can be obtained. This transformation might be an integral module of the objective model. The statistical analysis that might be applied to calibrate the scale of an objective model is outside the scope of this article.

Validation of Objective Model: The objective data analysis might be evaluated with respect to its prediction accuracy, consistency, and linearity by using different subjective datasets. It may also be worth noting that the performance of the model might depend on the training datasets and the verification procedures. See [43] for more details on calibrating and validating objective models. The *Video Quality Experts Group* (VQEG) validates the performance of objective perceptual models that result in ITU recommendations and standards for objective quality models for both television and multimedia applications^[44]. The practical deployment of such objective models is discussed in the background of QoE measurement methods.

End-User Device Analytics

End-user device analytics could provide an alternative to measuring QoE as experienced by the end user. Real-time data such as the connection time, bytes sent, and average playback rate are collected by the video player application for each video viewing session and fed back to a server module where the data is pre-aggregated and then turned into actionable QoE measures.

Some of the metrics reported for per-user and aggregate viewing sessions include startup delay, rebuffering delays, average bitrates, and frequency of bitrates switches. Operators might be more inclined to associate viewers' engagement to QoE because better QoEs might make viewers less likely to abandon a viewing session, leading to increased monetization of assets and low subscriber turnover.

The definition of viewer engagement may have different meanings for different operators and contexts. First, operators might like to know which viewer engagement metrics affect QoE the most in order to guide the design of the delivery infrastructures. Secondly, they might also like to quickly identify and resolve service outages and other quality issues. A minute of encoder glitch could replicate throughout the *Internet Service Providers* (ISPs) and the various delivery infrastructures, and affect all their customers. Operators might like to know the scale of this impact, and how it affects users' engagement. The cost of getting the viewer experience wrong often makes the news headlines^[45]. Finally, they would like to understand their customers' demographics (connection methods, type of device, and bitrates of the consumed asset) within a demographic region so that they can fully monetize their assets, and use their other resources strategically.

QoE enthusiasts advocate QoE measurement to be a multidisciplinary approach that seeks to explain its findings, building on general laws of perception, sociology, and user psychology^[39]. If the end-user device analytics is used as a means of QoE measurement, many unexplained variables may not be accounted for (for example, why a user exits a service). For instance, the viewer's tolerance thresholds of QoS parameters for live and VoD might have different QoE patterns, and another dimension of complexity to include might be premium vs. free content. Also, a lack of interest in watching the content, not necessarily an effect of poor QoE, might make a user exit a service. Viewer engagement metrics is measured objectively bypassing subjective studies and surveys.

Some attempts at addressing these unexplained variables can be found in the literature. For example, [46] used the fraction of video viewed as a measure of engagement because this factor can be measured objectively. The data that appeared to belong to early quitters were systematically removed from their analysis to provide a clearer understanding of how the QoE measures affected viewer engagement. A slightly different approach to measuring viewer engagement can be found in [47]. However, with a huge database of aggregated data in the range of millions of viewing session logs, some profile classes of how the QoE measures affect viewer engagement might emerge.

Background on QoE Measurement Methods

The MOS appears to be the *de facto* standard metric for QoE. The possible reasons could be its long-term establishment in telephony networks, perhaps its widespread acceptance on the merits that it can be easily understood, and a metric for benchmarks. There are different types of MOS values, and different test methodologies to produce them, (see [32] for more details). Figure 5 shows the five-point absolute category rating MOS scale that is commonly used.

Figure 5: Five-Point Rating Scale

Score	Label
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

The MOS value is the average opinion for a given QoS tests condition, not necessarily for the individual users because different users have different opinions. Additional information such as statistical uncertainty in terms of confidence intervals is usually encouraged. The MOS is considered to be characteristic of only the experiment and the group of test subjects from which it was derived.

While there is a reference methodology to produce MOS, it has to be interpreted within context. First, the MOS value obtained for a particular QoS test condition in a subjective experiment may depend on the range of the QoS test conditions used in the experiment. This dependence might be due to test subjects who re-calibrate their use of the rating scale to the conditions in the experiment. An appropriately designed experiment that has a practice period at the start of the experiment, and the test conditions include the best and worst conditions, minimizes the effects of the aforementioned behavior.

Secondly, direct comparisons of MOS scores obtained from separate experiments are generally not meaningful. They are meaningful only if the experiments have been specially designed to enable such comparisons. Data from such specially configured experiments must be studied and shown that their MOS comparisons are statistically valid. Biases in the rating scale interpretation might exist such as differences in interpretation and use of rating scales across cultures; test subject profile, for example, age and technology exposure; test environment; and the presentation order of the test conditions^[48].

Thirdly, it is possible that different objective models that have been trained and optimized using different subjective contexts will predict non-identical MOS values for the same QoS conditions. Objective models are usually developed and optimized for a specific scope of quality features. As a consequence, comparisons between MOS predictions and thresholds can be reliably made only if the thresholds are chosen in the context of the MOS model. See [46] for MOS interpretation and reporting.

The ITU standardization activities^[49] classify objective quality assessment methods into five main categories: media-layer models, parametric packet-layer models, parametric planning models, bitstream-layer models, and hybrid models. Depending on the input parameters to the model and the specific service type (for example, speech, audio, video, and multimedia), each category is aimed at predicting QoE.

The practical deployment of such objective models might benefit in-service monitoring positioned at any one node, or at all nodes, within the content-delivery ecosystem. The node(s) could be at the headend incoming feeds, distribution networks, and at locations of endpoints that are customer equipment. For example, the media-layer model (also known as the perceptual model) might be best suited for quality assurance at the headend, and for benchmark purposes. The media-layer model uses both the source video and the impaired video to predict QoE. The parametric-packet layer model uses only packet-header information such as IP packets, for example, UDP, and extracts the required information to predict QoE. It does not consider encoding distortions, but owing to its light-weight implementation it might be appropriate for in-service monitoring distribution networks. On the other hand, the bitstream-layer models might be appropriate for locations of endpoints that are customer equipment, because they analyze up to the bitstream level. This model in turn does not consider the decoded (impaired) output. The perceptual hybrid model, which combines the media layer and bitstream layer, might use both the bitstream information and the decoded output to predict QoE. While objective assessment seems to offer real-time QoE measurements, some categories might not meet industry demands. Hence, end-user device analytics as a method to QoE measurement appears to be an alternative approach.

Currently, there is the lack of a reference methodology for end-user device analytics as a method of QoE measurement, analogous to MOSs found in subjective assessments^[32] and objective assessments^[42]. A limiting factor to this development might be the restricted rights governing service providers, or the likes, on the usage of such databases. This situation makes it challenging for researchers, service providers, and delivery infrastructures to focus their efforts on developing better delivery infrastructures.

Subjective experiments could still be the most accurate way to measure perceived QoE, and the only way to obtain reliable ground-truth data used in benchmarking objective QoE models.

Linking QoS to QoE

A first glance indicates a considerable mismatch between the concepts of QoS and QoE. This mismatch can be seen in the ITU-T definitions of the two terms. QoS is defined in ITU-T E.800^[50] 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.” These characteristics are objective measures consisting of quantitative variables and attributes that may be present or absent. Thus, QoS characteristics are objective and can be objectively measured. By contrast, ITU-T P.10^[51] defines QoE as “the overall acceptability of an application or service, as perceived subjectively by the end-user.” Thus, QoE is called out specifically as consisting of subjective measures.

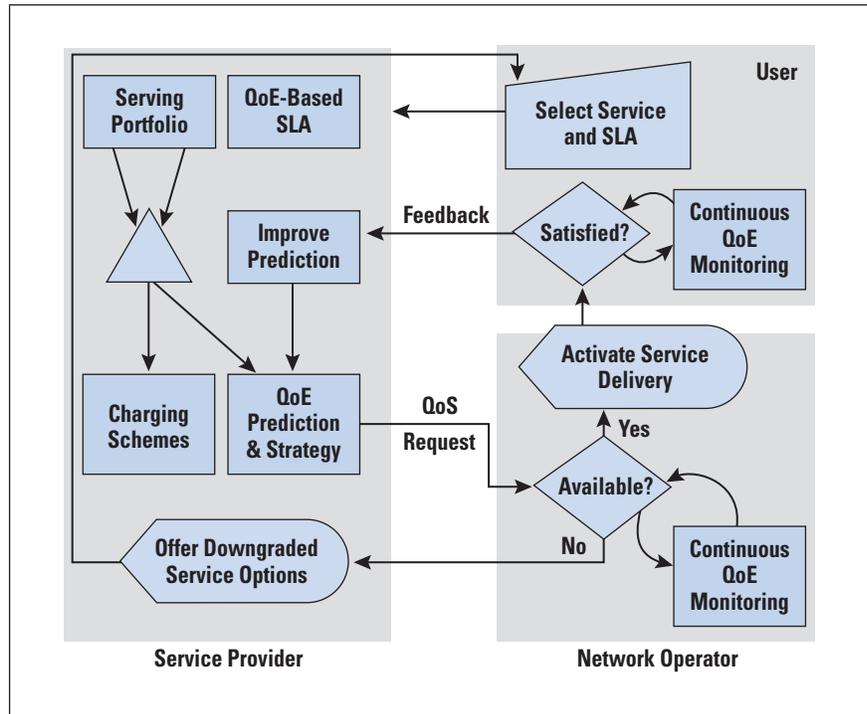
Both network service providers and customers have become accustomed to developing SLAs based on QoS measures. Recent years have seen a growing awareness on the part of both providers and customers that QoE is the more important concept, and that ways must be found to tie network performance parameters, as committed in an SLA, to the desired user QoE. For a typical network and service environment for a particular customer, numerous QoS metrics will impact overall QoE. The focus of ongoing research and product development is on developing reliable techniques, acceptable to both service provider and customer, for correlating QoS performance metrics with QoE as measured by MOS. ITU-T G.1080 identifies two ways in which such correlations can be exploited:

- Given a QoS measurement, predict the expected QoE for a user, with appropriate assumptions.
- Given a target QoE for a user, deduce the required network service performance, with appropriate assumptions.

Service providers can take the first approach and provide a range of QoS offerings with an outline of the QoE that their customers might reasonably expect. Customers can take the second approach by defining the required QoE and then determining what level of service will meet their needs.

Figure 6 illustrates a scenario for the second approach, where the user can make a selection from a range of services, including the required *level of service* (SLA). By contrast to the purely QoS-based management, the SLA here is not expressed in terms of raw network parameters. Instead, the user indicates a QoE target; it is the service provider that maps this QoE target together with the type of service selected, onto QoS demands.

Figure 6: User-Centric Service Delivery



For instance, in the case of multimedia streaming service, the user may simply choose between two QoE levels (high or low). The service provider selects the appropriate quality-prediction model and management strategy (for example, minimize network resource consumption) and forwards a QoS request to the operator. It is possible that the network cannot sustain the required level of QoS, making it impossible to deliver the requested QoE. This situation leads to a signal back to the user, prompting a reduced set of services/QoE values.

Assuming that the network can support the service, delivery can be activated. During service operation, two monitoring and control loops run concurrently: one at network level and the other at service level. The latter allows the user to switch to a different level of QoE (for example, to get a cheaper service or to request higher quality). If the user generates no explicit feedback, it means that the user is satisfied, confirming that the quality-prediction model is working. In this way, the quality-prediction model continues to be redefined during service delivery, allowing it to evolve as user needs and devices change over time.

There are so many varied components with this approach, which extends from the management complexity of a *Content-Delivery Network* (CDN) service based on QoE SLAs to service billing. However, focusing on integrating QoE as part of the management methodology during the design and development of services ensures a user-centric perspective, and helps to move beyond the current network design principles.

References

- [1] Geoff Huston, “QoS—Fact or Fiction?” *The Internet Protocol Journal*, Volume 3, No. 1, March 2000.
- [2] ITU-T, “An Architectural Framework for Support of Quality of Service in Packet Networks,” Recommendation Y.1291, May 2004.
- [3] Sally Floyd and Van Jacobson, “Random Early Detection Gateways for Congestion Avoidance,” *IEEE/ACM Transactions on Networking*, August 1993.
- [4] Jon Crowcroft, Bob Braden, Steve Deering, Sally Floyd, Bruce Davie, Van Jacobson, and Deborah Estrin, “Recommendations on Queue Management and Congestion Avoidance in the Internet,” RFC 2309, April 1998.
- [5] William Stallings, “MPLS,” *The Internet Protocol Journal*, Volume 4, No. 3, September 2001.
- [6] ITU-T, “Traffic Control and Congestion Control in IP-based Networks,” Recommendation Y.1221, June 2010.
- [7] Dimitrios Pendarakis, Raj Yavatkar, and Roch Guerin, “A Framework for Policy-based Admission Control,” RFC 2753, January 2000.
- [8] Eric S. Crawley, Raj Nair, Bala Rajagopalan, and Hal Sandick, “A Framework for QoS-based Routing in the Internet,” RFC 2386, August 1998.
- [9] Lixia Zhang, Steve Berson, Shai Herzog, and Sugih Jamin, “Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification,” RFC 2205, September 1997.
- [10] Sally Floyd, “TCP and Explicit Congestion Notification,” *ACM Computer Communication Review*, October 1994.
- [11] Vern Paxson, Guy Almes, Jamshid Mahdavi, and Matt Mathis, “Framework for IP Performance Metrics,” RFC 2330, May 1998.
- [12] Sebastian Möller, Patrick Le Callet, and Andrew Perkis, “Qualinet White Paper on Definitions on Quality of Experience – Output Version of the Dagstuhl Seminar 12181,” *European Network on Quality of Experience in Multimedia Systems and Services*, COST Action IC 1003, 2012.
- [13] Florence Agboma and Antonio Liotta, “QoE for Mobile TV Services,” In *Multimedia Transcoding in Mobile and Wireless Networks*, IGI Global, 2009.

- [14] Mario Siller, “An Agent-Based Platform to Map Quality of Service to Experience in Active and Conventional Networks,” University of Essex, 2006.
- [15] Milos Ljubojević et al., “Influence of Resolution and Frame Rate on the Linear in-Stream Video Ad QoE,” *Journal of Information Technology and Applications*, Volume 4, Issue 1, June 2014.
- [16] Jingjing Zhang and Nirwan Ansari, “On Assuring End-to-End QoE in Next Generation Networks: Challenges and a Possible Solution,” *IEEE Communications Magazine*, Volume 49, Issue 7, July 2011.
- [17] Ulrich Reiter et al., “Factors Influencing Quality of Experience,” In *Quality of Experience: Advanced Concepts, Applications and Methods*, 2014.
- [18] Sebastian Möller, M. Waltermann, and M. Garcia, “Features of Quality of Experience,” In *Quality of Experience: Advanced Concepts, Applications and Methods*, 2014.
- [19] Miguel Ríos Quintero and Alexander Raake, “Is taking into account the subjects’ degree of knowledge and expertise enough when rating quality?” In *Fourth International Workshop on Quality of Multimedia Experience (QoMEX)*, July 2012.
- [20] Telchemy, “Voice Quality Measurement,” In Voice over IP Performance Management, 2014. <http://www.telchemy.com/appnotes/TelchemyVoiceQualityMeasurement.pdf>
- [21] Ofcom, *On-Demand Services: Understanding Consumer Choices*, October 2012.
- [22] Andreas Sackl et al., “Wireless Vs. Wireline Shootout: How User Expectations Influence Quality of Experience,” In *Fourth International Workshop on Quality of Multimedia Experience (QoMEX)*, July 2012.
- [23] Florence Agboma and Antonio Liotta, “Addressing User Expectations in Mobile Content Delivery,” *Mobile Information Systems*, Volume 3, Issue 3–4, December 2007.
- [24] David S. Hands, “A Basic Multimedia Quality Model,” *IEEE Transactions on Multimedia*, Volume 6, No. 6, December 2004.
- [25] Stefan Winkler and Christof Faller, “Maximizing Audiovisual Quality at Low Bitrates,” *Proceedings of International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, January 2005.

- [26] Tobias Hossfeld et al., “Initial Delay Vs. Interruptions: Between the Devil and the Deep Blue Sea,” In *Fourth International Workshop on Quality of Multimedia Experience (QoMEX)*, July 2012.
- [27] Sebastian Egger et al., “Waiting Times in Quality of Experience for Web Based Services,” In *Fourth International Workshop on Quality of Multimedia Experience (QoMEX)*, July 2012.
- [28] Peter Reichl, Patrick Maillé, Patrick Zwickl, and Andreas Sackl, “A Fixed-Point Model for QoE-Based Charging,” In *Proceedings of the 2013 ACM SIGCOMM Workshop on Future Human-Centric Multimedia Networking*, 2013.
- [29] Andrew Perkis, Peter Reichl, and Sergio Beker, “Business Perspectives on Quality of Experience,” In *Quality of Experience*, 97–108, 2014.
- [30] Florence Agboma, “Quality of Experience Management in Mobile Content Delivery Systems,” University of Essex, Ph.D. Thesis, 2009.
- [31] Martina Angela Sasse et al., “Remote Seminars through Multimedia Conferencing: Experiences from the Mice Project,” In *INET '94*, 1994.
- [32] ITU-T, “Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in Any Environment,” Recommendation P.913, 2014.
- [33] ITU-R, “Methodology for the Subjective Assessment of the Quality of Television Pictures,” Recommendation BT.500, 2012.
- [34] ITU-T, “Subjective Video Quality Assessment Methods for Multimedia Applications,” Recommendation P.910, 2008.
- [35] ITU-R, “General Methods for the Subjective Assessment of Sound Quality,” Recommendation BS.1284-1, 2003.
- [36] Jesus Gutierrez et al., “Subjective Evaluation of Transmission Errors in IPTV and 3DTV,” In *IEEE Visual Communications and Image Processing (VCIP)*, 2011.
- [37] Nicolas Staelens et al., “Assessing Quality of Experience of IPTV and Video on Demand Services in Real-Life Environments,” *IEEE Transactions on Broadcasting*, Volume 56, Issue 4, December 2010.

- [38] Hendrik Knoche, John D. McCarthy, and Martina Angela Sasse, “Can Small Be Beautiful?: Assessing Image Resolution Requirements for Mobile TV,” In *ACM International Conference on Multimedia*, 2005.
- [39] Tobias Hossfeld et al., “Quality of Experience Management for YouTube: Clouds, Fog and the Aquareyoum,” *PIK – Praxis der Informationsverarbeitung und Kommunikation*, Volume 35, No. 3 2012.
- [40] ITU-R, “Parametric Non-Intrusive Assessment of Audiovisual Media Streaming Quality,” Recommendation P.1201, 2013.
- [41] ITU-T, “Objective Perceptual Multimedia Video Quality Measurement of HDTV for Digital Cable Television in the Presence of a Full Reference,” Recommendation J.341, 2011.
- [42] Marcus Barkowsky et al., “Hybrid video quality prediction: reviewing video quality measurement for widening application scope,” In *Multimedia Tools and Applications*, Volume 74, No. 2, Springer, 2015.
- [43] ITU-T, “Methods, Metrics and Procedures for Statistical Evaluation, Qualification and Comparison of Objective Quality Prediction Models,” Recommendation P.1401, 2012.
- [44] Kjell Brunnström, David Hands, Filippo Speranza, and Arthur Webster, “VQEG Validation and ITU Standardization of Objective Perceptual Video Quality Metrics [Standards in a Nutshell],” *IEEE Signal Processing Magazine*, Volume 26, No. 3, May 2009.
- [45] BBC News, “Xbox and PlayStation resuming service after attack,” <http://www.bbc.co.uk/news/uk-30602609>
- [46] Athula Balachandran, Vyas Sekar, Aditya Akella, Srinivasan Seshan, Ion Stoica, and Hui Zhang “Developing a Predictive Model of Quality of Experience for Internet Video,” In *ACM SIGCOMM Computer Communication Review*, Volume 43, No. 4, October 2013.
- [47] Florin Dobrian, Asad Awan, Dilip Joseph, Aditya Ganjam, Jibin Zhan, Vyas Sekar, Ion Stoica, and Hui Zhang, “Understanding the Impact of Video Quality on User Engagement,” *ACM SIGCOMM Computer Communication Review*, Volume 41, No. 4, August 2011.

- [48] ITU-T, “Mean Opinion Score Interpretation and Reporting,” Recommendation P.800.2, 2013.
- [49] Akira Takahashi et al., “Standardization activities in the ITU for a QoE assessment of IPTV,” *IEEE Communications Magazine*, Volume 46, No. 2, February 2008.
- [50] ITU-T, “Definitions of terms related to quality of service,” Recommendation E.800, September 2008.
- [51] ITU-T, “Vocabulary for performance and quality of service,” Recommendation P.10, July 2006.

WILLIAM STALLINGS is an independent consultant and author of numerous books on security, computer networking, and computer architecture. His latest book is *Foundations of Modern Networking: SDN, NFV, QoE, IoT, and Cloud* (Pearson, 2016). He maintains a resource site for computer science students and professionals at ComputerScienceStudent.com and is on the editorial board of *Cryptologia*. He has a Ph.D. in computer science from M.I.T. He can be reached at ws@shore.net

FLORENCE AGBOMA holds an M.Sc. and a Ph.D. in Electronic Systems Engineering from the University of Essex, United Kingdom, in 2005 and 2009, respectively. Her Ph.D. research focused on Quality of Experience for Mobile Content Delivery Systems. She currently works as a Technology Analyst at British Sky Broadcasting (BSkyB), London. Her interests include subjective video quality assessment, psychophysical methods, and quality of experience management across devices and platforms. E-mail: fagbom@googlemail.com