♦ Delivering Quality of Experience in Multimedia Networks

Harold Batteram, Gerard Damm, Amit Mukhopadhyay, Laurent Philippart, Rhodo Odysseos, and Carlos Urrutia-Valdés

Next-generation multimedia networks need to deliver applications with a high quality of experience (QoE) for users. Many network elements provide the building blocks for service delivery, and element managers provide performance data for specific network elements. However, this discrete measurement data is not sufficient to assess the overall end user experience with particular applications. In today's competitive world of multimedia applications, it is imperative for service providers to differentiate themselves in delivering service level agreements with certainty; otherwise they run the risk of customer churn. While QoE for well-established services like voice and Internet access is well understood, the same cannot be said about newer multimedia services. In this paper, we propose parameters for measuring the QoE for newer services. We propose and justify parameter values for satisfactory end user experience and show how standard measurement data can be collected from various network elements and processed to derive end user QoE. © 2010 Alcatel-Lucent.

Introduction

Service providers are looking at new applications to compensate for declining revenues from traditional voice and flat-rate high-speed data services. These are often multimedia applications involving text, voice, pictures, video clips, hyperlinks, and mixed media and can be provided by either the service provider (SP) or an Internet-based application provider (as shown in **Figure 1**). In today's highly competitive environment, users have the option of choosing from a plethora of service providers such as wireless, wireline, and cable operators. Therefore, it is not enough to simply make the services available to users; service providers must deliver those services in such a way that users fully enjoy a rich experience at a reasonable price.

In addition, transition of network traffic from primarily voice to primarily data is a well-known phenomenon in the telecommunication industry. What is more interesting is that data traffic is no longer just a best-effort type of traffic. Many applications which are running as data traffic require a high quality of service to meet user needs and expectations. Consequently, many standards bodies and industry organizations such as the International Telecommuni-cation Union (ITU), 3rd Generation Partnership Project (3GPP*), and Broadband Forum (formerly the Digital Subscriber Line [DSL] Forum) have come up with various classifications of services and associated quality of service (QoS) parameters. However, none of these

Bell Labs Technical Journal 15(1), 175–194 (2010) © 2010 Alcatel-Lucent. Published by Wiley Periodicals, Inc.
Published online in Wiley InterScience (www.interscience.wiley.com) • DOI: 10.1002/bltj.20431

Panel 1. Abbreviations, Acronyms, and Terms

3GPP—3rd Generation Partnership Project
CDMA—Code division multiple access
CLI—Command line interface
DSL—Digital subscriber line
DSLAM—Digital subscriber line access multiplexer
EMS—Element management system
EPG—Electronic program guide
GSM—Global System for Mobile Communications
HD—High definition
IETF—Internet Engineering Task Force
IM—Instant messaging
IMS—IP Multimedia Subsystem
IP—Internet Protocol
IPTV—Internet Protocol television
ITU—International Telecommunication Union
ITU-T—ITU Telecommunication Standardization Sector
KPI—Key performance indicator
KQI—Key quality indicator
MDI—Media delivery index
MOS—Mean opinion score
NE—Network element
NGN—Next-generation networks
NMS—Network management system
NOC—Network operations center
OTT—Over-the-top
PoC—Push-to-talk over cellular
QoE—Quality of experience
QoS—Quality of service
RCS—Rich communications suite
RTP—Real Time Transport Protocol
SAP—Service access point
SBC—Session border controller
SD—Standard definition
SLA—Service level agreement
SMS—Short message service
SP—Service provider
SQM—Service quality management
STB—Set-top box
TM Forum—TeleManagement Forum
TV—Television
VoD—Video on demand
VoIP—Voice over Internet Protocol

matrices alone clearly captures end user quality of experience (QoE) as the measurements deal only with network elements.

In this paper, QoE is defined as the measure of how well a system or an application meets the user's expectations [5, 12]. This concept is different from quality of service, which focuses on measuring performance from a network perspective. For instance, QoE focuses on user-perceived effects, such as degradation in voice or video quality, whereas QoS focuses on network effects such as end-to-end delays or jitter. Of course, QoE is directly related to QoS, but the challenge for a service provider is to have the right set of tools and processes to map the QoS at the network level to the QoE at the user and session levels and have the ability to control it.

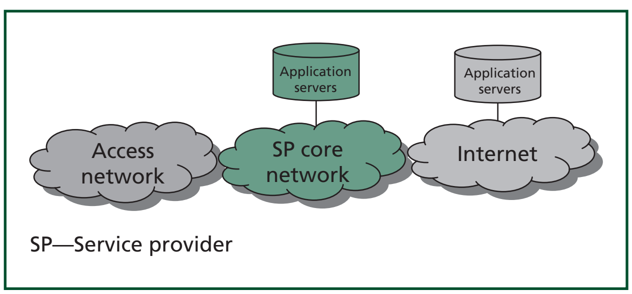

Figure 1.
End-to-end architecture.

Another important point to note is that measurements in individual nodes may indicate acceptable QoS, but end users may still be experiencing unacceptable QoE. Assume, for example, that there is an edge router with an inadequately engineered voice over Internet Protocol (VoIP) buffer size; also assume that packet overflow is monitored at the aggregate queue level. The VoIP buffer may be overflowing quite frequently, but because the VoIP traffic volume is relatively low compared to other types of traffic and because other buffers do not overflow, aggregate buffer statistics may still appear to be satisfactory (acceptable QoS). This, however, is no consolation for the VoIP subscriber, who is consistently experiencing service interruptions (poor QoE because of dropped packets).

Service providers have traditionally focused on determining and managing QoS, not QoE. The most common and time-tested means for measuring QoS is the use of a performance management system that extracts measurement data from network elements or element management systems (EMS) to assess the performance of various network elements across the network. However, as we noted in the previous paragraph, this method does not guarantee acceptable QoE estimation for individual applications, sessions, or users. Several approaches have emerged over the past few years to measure application performance. The focus of this paper is to go a step further and explore the end user experience. The key is not only to measure QoE but also to manage it effectively for the greatest impact on the operator's balance sheet. Ideally speaking, QoE issues should be prioritized based on their relative impact on potential revenue, as it is often impractical to address all the problems at one time.

We begin by providing details on application performance measurement techniques; this section lays the foundation for the rest of the discussions in the paper. The next section provides an overview of standards, and of the standards gaps that exist for current methodologies. We follow that with details on a generic approach we propose to cope with the ever-increasing complexity of new applications. This discussion is followed by a section which provides some examples of key quality indicators/key performance indicators (KQI/KPIs) which lead to a way to measure QoE. In the final section, we discuss further work needed in this area.

Application Performance Measurement Techniques

There are primarily three techniques prevalent in the market today for measuring application performance: 1) using test packets, 2) using probes in network elements and user equipment, and 3) correlating measurements from several network elements. This section provides a brief discussion of these techniques, with the greatest focus on the third technique, since it is a very complex method but has very few drawbacks otherwise. It may be noted, however, that any combination of these techniques may be used in a particular performance measurement tool.

Performance Measurement Using Test Packets

As a general method, test packets are sent from management systems, and performance metrics such as delay, jitter, and packet loss are measured along the way. The results of these measurements are used as a proxy for the performance of real traffic. This method is also often used for troubleshooting.

While this is a very simple technique, care needs to be taken when interpreting the results. It is not desirable to send test packets during busy hours since this will unnecessarily load the network with management traffic. On the other hand, unless testing is performed during peak usage hours, the measurements will not truly reflect user experience at the most important time.

Performance Measurement Using Probes

In this method, probes in the form of software agents or network appliances are deployed on network elements and user devices (for the software agent case). Measurements based on these probes provide a very accurate status of the devices at any time. Furthermore, in the case of software agents, true user experience can be measured unobtrusively since measurements are obtained directly from user devices.

The main drawback of this technique is that it doesn't scale for large networks. While it is very useful for fault isolation and root cause analysis, this technique cannot be used for monitoring large networks with millions of user devices and network elements.

Performance Management Using Network Measurements

In this method, measurements from various network elements and their EMSs are collected and processed at a central repository, as illustrated in **Figure 2**. The repository also collects data from various network management systems (NMS), e.g., configuration, inventory, subscriber, or fault management systems. Intelligent software in the central repository correlates and analyzes these different sets of data. It also can apply a rules engine to monitor certain events or provide diagnostics. The rules engine may, alternatively, trigger some further data collection to probe deeper into potential problems. Managing QoE starts with well-chosen ongoing and ad hoc measurements

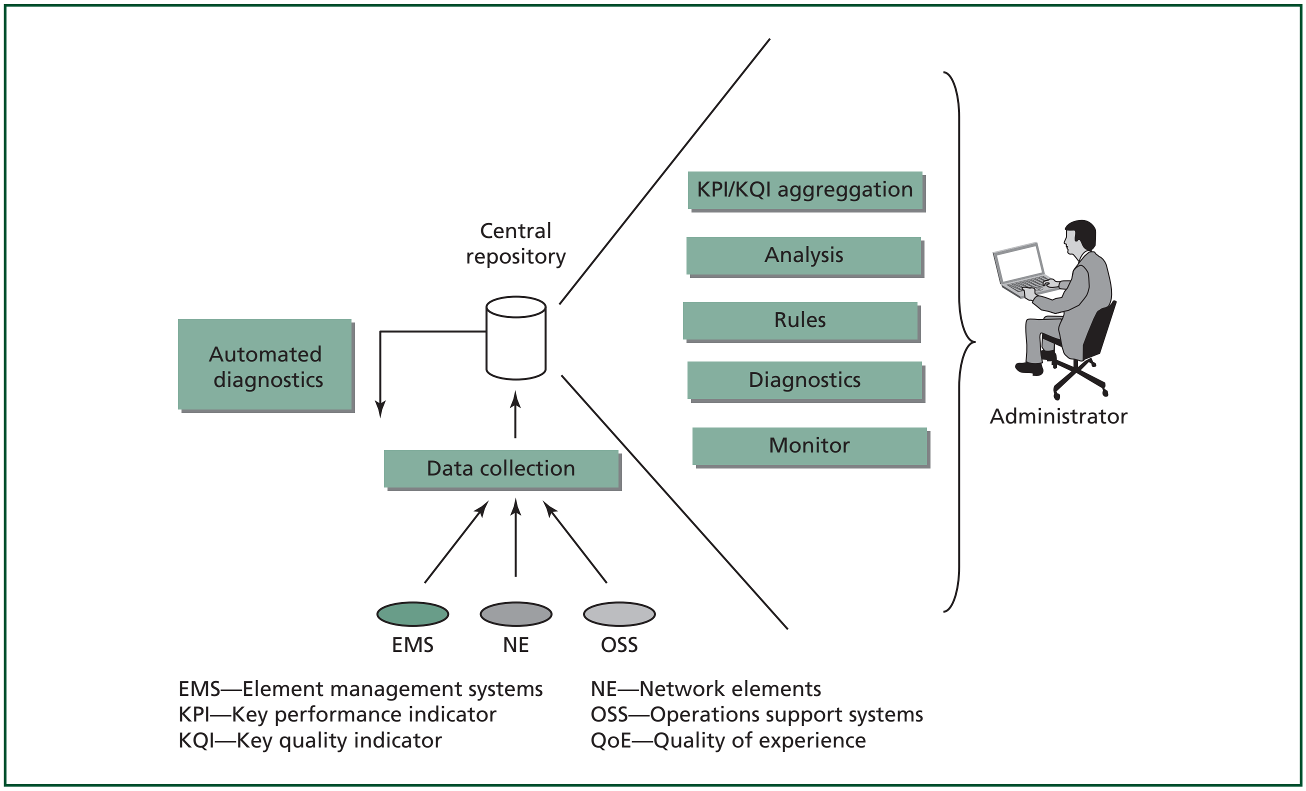

Figure 2.
Use of standard measurement for QoE determination.

since they form the basis for all the analyses needed to determine the level of QoE.

To get a better understanding of the process, let us look at a simplified example application for Internet Protocol television (IPTV) where a human error caused a degradation of QoE. **Figure 3** depicts the assumed architecture.

Here is the sequence of events around the problem and its resolution:

1. On a permanent basis, the following KPIs are collected: set-top box (STB) retries per channel, digital subscriber line access multiplexer (DSLAM) uplink and downlink port loss and bandwidth, edge and core network router and switch port loss and bandwidth, and headend-level monitoring of each channel.
2. Over time these KPIs are aggregated and archived.
3. An operator makes a mistake using a command line interface (CLI) command and misconfigures a bandwidth profile on a router service access point (SAP). This restricts the bandwidth allowed on that SAP but keeps it at a high enough value to allow a significant amount of traffic through.
4. STBs downstream of that router port identify missing data and begin sending retry requests.
5. The KPI threshold for STB retry requests is crossed and alarms are generated (A1).
6. The KPI threshold for DSLAMs and switches is not crossed and no alarms are triggered (since no traffic is dropped).
7. The KPI threshold-crossing rules for the misconfigured SAP may trigger intermittent alarms (A2), based on port loss.
8. The KPI threshold-crossing rules for headend-level monitoring do not raise an alarm.
9. The alarms will appear on the administrator dashboards and in the reports.
10. The logical correlation engine will recognize that A1 and A2 alarms are related by causality

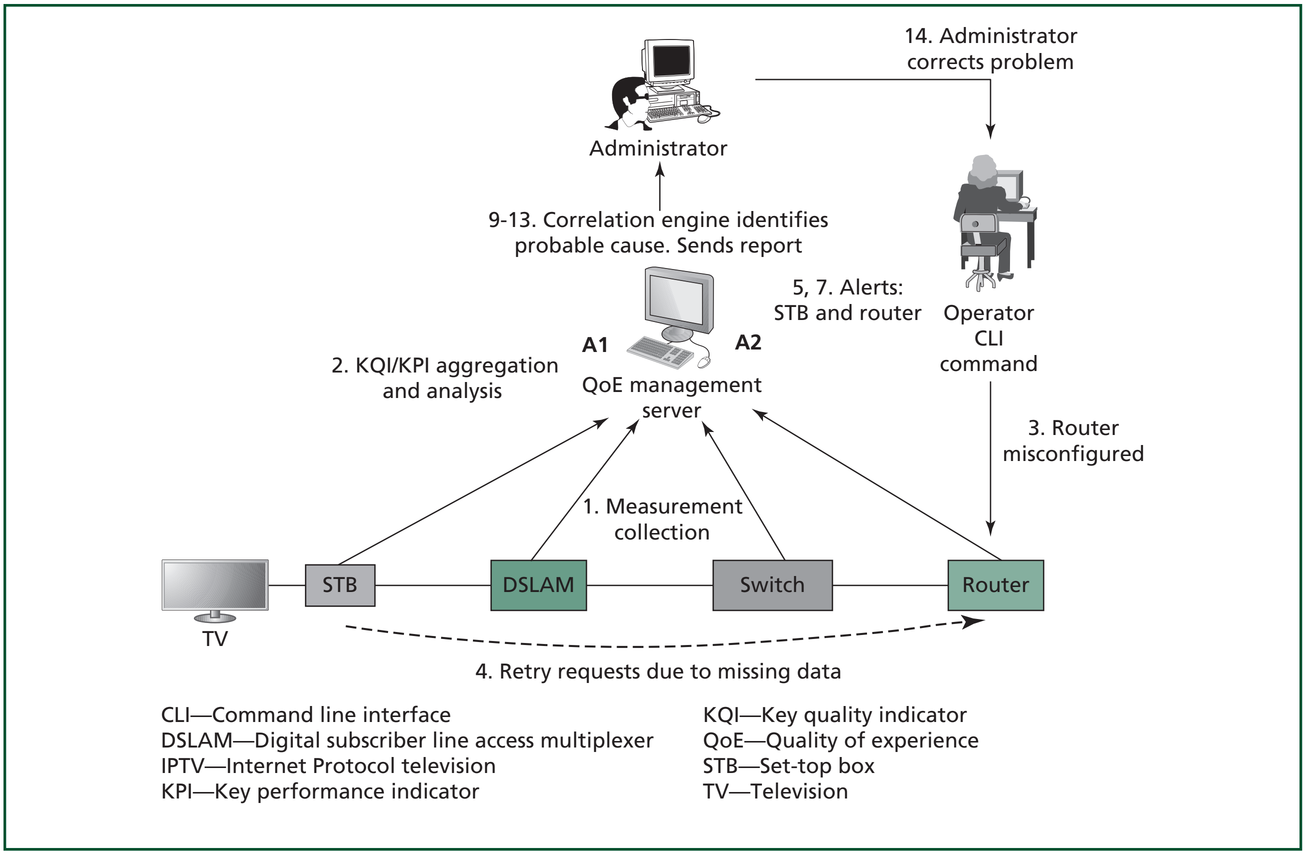

Figure 3.
Simplified IPTV architecture.

(A2's probable cause is a parent of A1's probable cause in the causality graph root cause analysis).

11. The spatial correlation engine, using its knowledge of the topology, will recognize that A2 alarms are issued by a node which sits on all the network paths of all A1 alarms.
12. These engines in tandem will establish that the A2 alarms are a probable root cause of the A1 alarms, simplifying the administrator's view (since correlated A1 notifications are collapsed under A2) and directing the troubleshooting operations towards the suspect SAP by putting the events in context (e.g., which services have a problem, the severity of the problem, which alarms are raised, which rules are triggered, and which customers [STBs] are impacted).
13. Bandwidth misconfiguration may be listed as a possible cause of the A2 alarms, thus providing even more assistance to the troubleshooting process.
14. The network operations center (NOC), when treating these alarms, will rapidly inspect the suspected SAP, should discover quickly that the bandwidth profile was misconfigured, and will rectify the problem easily. (The whole process from detection to resolution could happen within minutes, before customers even start to complain and call their customer care centers or open tickets on their portal.)
15. If the bandwidth error had been barely detectable (i.e., causing only rare problems), the problem would still be discovered through the historical correlation engine.
16. Rather than STB retries (a network-level KPI), an end user QoE estimation could also be used as input for the correlation engines.

Understanding network services is necessary, but measuring complete end-to-end performance is impractical. The trade-off here involves having the

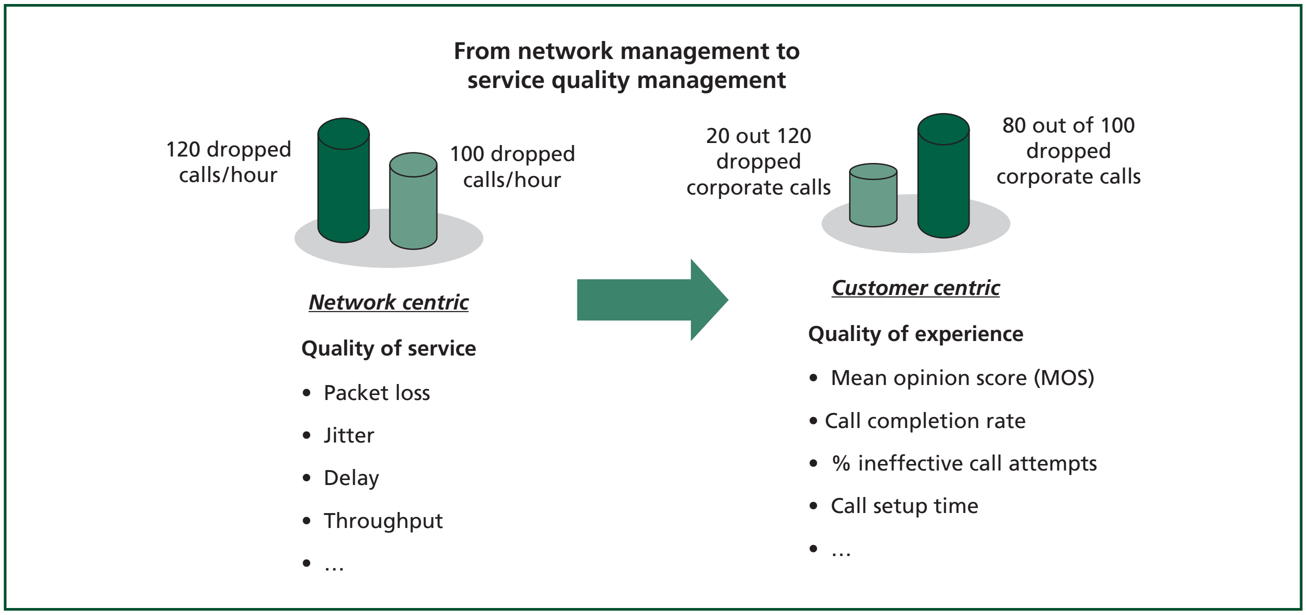

Figure 4.
Network-centric versus customer-centric view.

topology intelligence to reconstruct end-to-end performance data from local performance data and a sampling of the actual measurements.

Performance Measurement to QoE

It is not enough to measure performance and derive KPI/KQIs. One has to take a step beyond to determine QoE and manage it effectively. This requires a paradigm shift from a network centric to a customer centric view, as shown in **Figure 4**.

Managing QoE is essential. It is one thing to monitor and measure it, but another thing to maximize it (to increase profit). Maximizing QoE requires early detection of potential QoE-affecting problems and solving them as soon as possible, *before* they reach the customer complaint stage. It also means being able to prioritize problems, so as to first tackle those with the highest impact in terms of potential revenue: A problem affecting 1,000 customers is more important than a problem affecting only 10; a problem affecting five gold customers may or may not be deemed more important than a problem affecting 10 bronze customers, depending on the weights and the decision rules.

Standards Activities

The concept of QoE has been the study of various standards organizations. In this section we present a survey of the relevant standards. While they provide guidelines on the measurement data, defining the KPIs/KQIs based on the measurements is generally left to the operators.

ITU-T

Within the ITU Telecommunication Standardization Sector (ITU-T), Study Group 12 is the lead group responsible for QoS and QoE. Some of their main recommendations include the following:

- *ITU-T G.1010* provides guidelines with regard to key QoS factors impacting the user. It focuses on delay, delay variation, and information loss and gives performance targets for various applications (e.g., conversational voice, audio streaming, Web browsing, and others) that would meet user expectations. For instance, short message service (SMS) is classified as a low priority transaction service, and therefore it is argued that tens of seconds in delivery delay are acceptable. Again, the purpose of this recommendation is to serve as a guide and not to set forth definitive requirements, since actual target values would be left to the operator to decide.
- *ITU-T G.1030* provides a model for estimating the performance of data applications over Internet

Protocol (IP) networks. This model consists of three steps: 1) network performance assessment, 2) application performance assessment, and 3) perceptual performance assessment. The last step is the one which introduces the idea of user experience (perception). This can be viewed as an "opinion model" similar to the e-model defined in [8], which maps end user experience from the network layer up to the application layer. The recommendation includes a model for Web browsing, but other applications are left for further study.

- *ITU-T G.1070* provides an opinion model for computing video telephony QoE based on a series of speech and video parameters. The model consists of three functions: 1) video quality estimation, 2) speech quality estimation, and 3) multimedia quality integration functions. The outputs are multimedia quality, video quality influenced by speech quality, and speech quality influenced by video quality; however, the model is based on very specific terminals and environments. An extension to accommodate other conditions is a topic for further study.

Other recommendations such as [9] define a series of Real Time Transport Protocol (RTP) statistics that can be collected from the network element (NE) to compute performance metrics (e.g., packet loss, delay, jitter, failures, etc.) Yet others such as [2] and [3] recommend test methods for assessing audio and video qualities.

Broadband Forum

Broadband Forum has also taken on the task of defining QoE and its relationship to QoS [12], although their target is triple play applications (i.e., video [both broadcast and on-demand], audio, and best-effort Internet data). Other applications are left for future work. Their definition of QoE is consistent with that of ITU-T in that it is viewed as a measure of overall performance from the user's perspective, whereas QoS is a measure of network performance. Their goal is to provide a clear relationship between the two so that given a set of QoS measurements, one could estimate the QoE for a user, and likewise, given a target QoE, one could calculate the required network performance. In general, this is a good first step since the Broadband Forum provides a complete methodology and some specific requirements in terms of delay, bandwidth, and other such network parameters. However, it is not clear how QoE would be affected if such requirements are not met, so a complete mapping is the subject for future research.

TeleManagement Forum

The TeleManagement Forum (TM Forum) looks at QoE from the service level agreement (SLA) management perspective, as would be expected. The TM Forum defines KQIs and KPIs as measurements of perceived quality rather than network performance [13–15], which is consistent with ITU-T and Broadband Forum views on QoE. KQIs are constructed from KPIs, and KPIs are derived from network performance measurements. For instance, an example KQI is the "percentage of sessions that experience delay of X or above," and a corresponding KPI is the session start-up delay, which is derived from network performance measurements.

In summary, the goal of all these standards organizations is to provide clear definitions of QoE and QoS and to establish the relationship between the two so that service providers can measure and manage QoE. In many respects this has been achieved for triple play applications but other applications are still left for future work.

The Challenge With NGN Applications

While QoE estimation algorithms for voice quality are well understood and video quality estimation methods are becoming increasingly more practical, methods for other applications are far less mature. QoE estimations for new NGN applications will require a per-application analysis in which the main properties that can impact user-perceived quality are determined. Next-generation services may be provided either by an SP or by an over-the-top (OTT) content provider or application provider. While an SP is in full control of the access network and thus may have the ability to provide desired QoE, OTT players are in a much different position. In today's world, OTT players often allow the user to download client software which helps optimize user experience. However, as bandwidth demand grows for applications and various applications need to interact with each other to provide a rich user experience, a simple client-based approach is unlikely to be successful. Service providers are not likely to be satisfied with the role of bit-pipe provider and certainly would like to get a share of the OTT revenue. However, it is impractical for them to mirror the thousands of applications available on the Internet today—or the exponential expansion of applications we can envision in the future—within their own infrastructure.

In direct competition, service providers have a theoretical ability to provide poor QoE to users accessing OTT applications (even though there are legal debates currently raging on this subject) while providing better QoE for the same applications provided through their own network infrastructure. However, there exists a happy middle ground: OTT players can negotiate business arrangements with service providers for specific QoE for their applications; this can be a win-win situation for both players. Service providers can leverage their QoE infrastructure in several ways: 1) provide QoE to differentiate the high quality content they deliver, 2) provide QoE for content/application providers that are business partners, 3) act as mediator for different application providers to develop higher value services, 4) personalize content on behalf of partner application/content providers.

Service providers have a number of fundamental advantages over OTT players in providing QoE for applications. For example, they have control over the broadband access network and maintain basic demographic information about their subscribers. In addition, wireless operators have full knowledge of user equipment capabilities. OTT players have the basic advantage that they can spread the cost of developing applications or content over a much larger user base, can draw advertisements and sponsorships from a much larger set of interested parties, and spend very little on physical infrastructure. A marriage of these two forces can provide a very powerful business solution.

While IP Multimedia Subsystem (IMS) infrastructure is slowly gaining momentum, and applications over this infrastructure are now taking off, it is also true that numerous "point solutions" are available on the Internet, which do not necessarily depend on an IMS infrastructure. Finding the best way to complement the capabilities of service providers and OTT players is a topic that is likely to be debated over the next several years. However, one trend that has clearly emerged in the industry over the past year is support for the rich communications suite (RCS), which has been championed by the Global System for Mobile Communications (GSM) Association and equally supported by code division multiple access (CDMA) operators in the wireless world. Wireline operators are also embracing the concept with apparent eagerness. A likely scenario to evolve is one in which service providers develop the applications for supporting RCS while OTT players take advantage of the underlying infrastructure to provide all other services.

To illustrate this point, the following three services form the cornerstones of RCS:

- *Enhanced address book*. Provides presence and user equipment capability indications.
- *Rich call*. Allows users to exchange different types of content (e.g., pictures and video) during a call.
- *Rich message*. Multimedia messaging with enhanced user interface.

In addition, service providers often have location information for their subscribers. Suppose an OTT player wants to develop a powerful marketing and sales application. As soon as a user is within a certain distance of a particular store (leveraging presence and location information), a rich message may be sent. The message is customized, based on the subscriber's demographic data, and is tailored for the appropriate device being used at that time (leveraging information from the enhanced address book). If the subscriber is interested, he may click a link to call a salesperson and view pictures/video clips of merchandise during the call for a rich user experience. The important point to note is that the OTT player does not need to build the infrastructure needed for the service and can focus on application development and revenue collection, as long as there is a business relationship of revenue sharing or other arrangement with the service provider. The service provider, on the other hand, must guarantee adequate QoE to make the OTT player's business a success.

One of the main challenges with next-generation network (NGN) services such as the one just described is modeling the service in such a way that meaningful QoE measurements can be extracted. After all, QoE is very subjective but must somehow quantitatively reflect the end user's satisfaction with the overall service in a meaningful way, such that it allows the operator to take appropriate actions based on those values. Service QoE typically includes multiple aspects of the service; for example, service availability, usability, and media quality factors all contribute to the overall service QoE. While calculation models to estimate voice quality using a mean opinion score (MOS) are well studied and validated with numerous human test subjects, the same level of confidence will be difficult to obtain for most other services. Deriving a single MOS-like score for a particular service will be quite complicated when a service consists of multiple contributing factors that influence the overall user-perceived quality. Unless the service QoE is carefully studied for each contributing factor with feedback from sufficiently large groups of actual users, the absolute value of a QoE score has limited meaning. Furthermore, a service that is composed of multiple applications is quite complicated to analyze.

Absolute Versus Relative QoE Measurements

Analyzing a service and understanding the various contributing factors that influence overall QoE can, however, be a very effective way to measure *changes* in QoE. In other words, QoE measurements can be meaningful when variations over time can be correlated to changes in the contributing factors [16]. These changes may be caused by changes in the network, for example, by bandwidth allocation, equipment capacity, or other changes in the service delivery components. It is therefore not only important to understand the key service QoE contributing factors from an end user perspective, but also the relationship between these factors and the individual service components in the end-to-end service delivery chain. This will allow degrading QoE values to be traced back to their root cause and, equally important, allow verification if optimizations in the service architecture component actually result in measurable QoE improvements.

Traditionally, service quality management (SQM) systems focus on measuring the performance characteristics of individual service components. **Figure 5** illustrates traditional metrics collection. Metrics such as voice MOS scores are measured, statistically aggregated, and compared with predefined threshold values. Operators can define target KQI values that indicate the overall performance level of voice quality. For example, a KQI target may state that the average voice MOS score during a measurement period of one week must be above 3.5 for 99.5 percent of all calls. Traditional service quality would be expressed in terms of the number of calls that meet a designated service quality level. A small percentage would be allowed to fall below this level. An example might be 99.5 percent of calls must achieve a setup delay less than 1.0 seconds. When threshold values are exceeded, an alarm can be generated.

While such measurements are clearly valuable, for example, in verifying SLAs, the per-session relationship between QoE contributing factors, such as call setup delay and voice quality within the same voice service session, is not preserved. For an end user perception of the service quality, both factors are important. Either excessive call setup delay or poor voice quality will result in a poor overall QoE evaluation from the end user. If both occur within the same session, the total effect can even be amplified since the end user is likely to be extremely dissatisfied if he experiences both an excessive call setup delay *and* poor voice quality within the same call session. Consequently, performance and perceived service quality require different measurement approaches where the relationship between the various factors that influence the overall service QoE is combined and preserved as contributing factors per service session, as illustrated in **Figure 6**.

Now consider the same QoE analysis process for a "rich call." On top of the traditional MOS for the basic voice service, we need to evaluate additional attributes related to picture and video quality, not just as standalone services, but also as a mixed-media service. Finding the main factors that influence user-perceived service quality, their interdependencies, and their relationship with service delivery components is a challenging task. This process will involve a number of steps, which are discussed in the sections following.

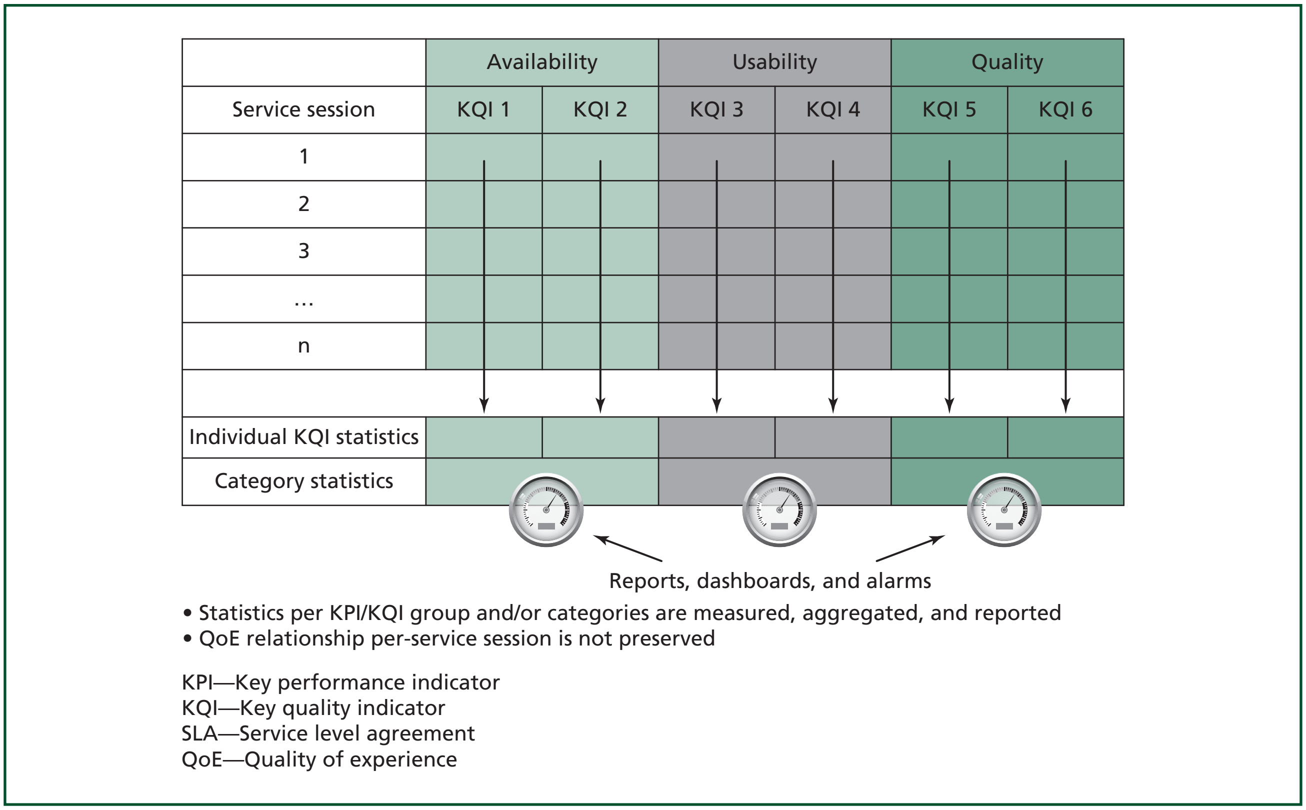

Figure 5.
Traditional performance measurements focus on SLA values.

Service Session Demarcation Points

Service QoE measurement requires that the usage of a service is measurable between clear points of demarcation, i.e., the start and end times of a service session. QoE metrics can only be related to user experience within the context of session boundaries.

A VoIP service has clear session demarcation points. The session starts when the user picks up the phone and ends when the phone is put down. For other services, these boundaries are not always clear. For example, consider live broadcast TV as part of an IPTV service. A user can be watching live TV for many hours per day and will be zapping through various channels during that period. What are the session boundaries in this case? The media quality may be different per channel. One channel may be delivered in high definition (HD) quality while others are standard definition (SD) quality. Clearly the user expectation when watching an HD channel is different from watching an SD channel. Hence, it is reasonable to partition the sessions into channel viewing time. But the channel change time is also a factor that needs to be taken into account. Excessive channel change time is a well-known complaint from IPTV subscribers. The channel change could be considered part of the session, or, alternatively, channel changing can be considered a separate service where each session consists of one or more channel changes within a reasonable time interval. As NGN services become more complex, defining practical session boundaries will not be a trivial task. However, the reliability of any service QoE measurements also depends on the precise definition of the service session demarcation points.

Service Session Decomposition

The next challenge is to decompose the service session into separate, measurable service elements that can contribute to the service QoE. Each of the elements must

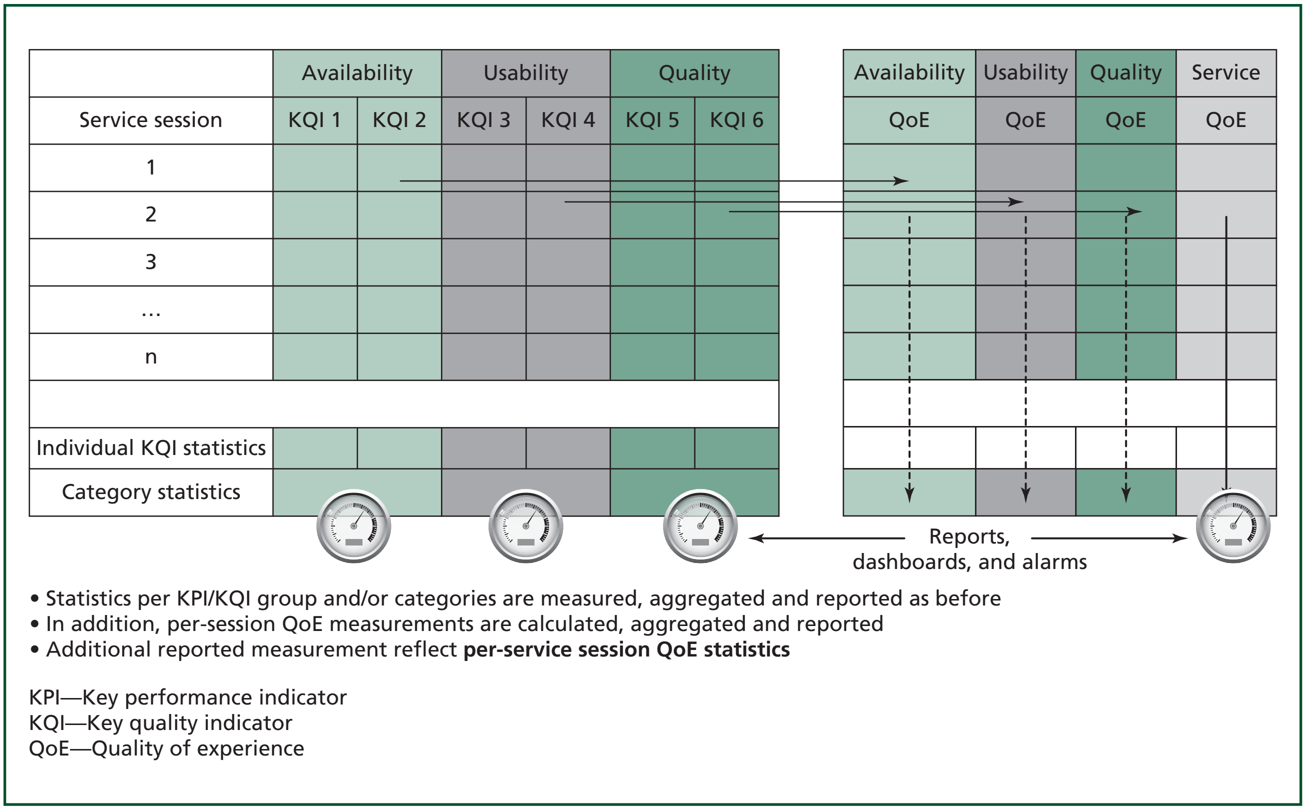

Figure 6.
Service oriented QoE measurements.

be directly perceivable by the end user. These elements may be sequential steps of the service, or they may occur concurrently. For example, a simple VoIP call can be decomposed into the following sequential elements:

- *Service start*. The user picks up the phone and waits for dial tone.
- *Connection setup time*. The user dials the number and waits for ringing notification.
- *Voice quality*. The MOS estimate of the conversation.
- *Call disconnect time*. The user ends the call and waits for confirmation (the network has to free up resources so the user needs to wait until another call can be placed).

The service elements can be grouped into categories (e.g., service availability, service usability, and media quality). In the example above, the connection setup and call disconnect time could be grouped into the service usability categories. For example, ITU-T G.1000 [4] defines a category matrix to facilitate identification of communications QoS criteria. The TM Forum also defines several useful service categories in [13]. Depending on the service aspects the operator wishes to measure and monitor, a selection of most applicable categories can be made.

Now consider the same analysis process for a "rich call":

- *Service start*. Similar to VoIP above but depends upon the starting media.
- *Connection setup time*. Not quite similar; we have to consider new media setup time every time the user uses a different media.
- *Voice quality*. This metric needs to be replaced by a combined measure of voice, picture, and video quality. Each media session may have its own measurements, and the overall "rich call" QoE has to be defined as an aggregation of the individual voice, picture, and video session related QoE.
- *Call disconnect time*. Again similar to VoIP but may depend upon the ending media.

Each session QoE assessment element will have attributes such as measurement units and maximum, minimum, and average values. The contribution of each element to the overall QoE will be different and needs to be normalized. Operators may also assign a different weight or importance to a particular factor. We recommend that both raw measured values and weight or normalization factors be registered so that these factors can be modified without losing the original data. **Figure 7** shows a generic approach for modeling service QoE measurements. The KQIs of each service element can be weighted according to operator-defined criteria, to emphasize the relative importance of the measurement, then normalized and grouped into a category. Categories can be combined into an overall QoE indicator, which can be used for high-level system monitoring, reporting, and trend analysis. Exceeding threshold limits can trigger an alert to the operator and diagnostic or root cause analysis processes similar to traditional performance monitoring systems.

Service Architecture Decomposition

The service QoE model has a session-oriented, end user perspective. Service usage is decomposed into measurable service elements that contribute to the overall service QoE. Now the relationship between the functional service elements and the architectural components of the service should be analyzed. For example, in an IMS VoIP application, call setup delay can be measured at various points in the service architecture—in the end user device, at the session border controller (SBC), or at other IMS network elements. A "rich call" will have many more such components. Each of these elements can also be the root cause for an excessive call setup delay value due to congestion, equipment failure, or other factors. When a poor QoE value is measured, the contributing factor(s) must be traced back to the probable cause. Figure 7 illustrates the relationship between service-specific, user-perceivable KQI elements and root cause, performance related KPIs as measured in the network and related service equipment. Note that this relationship does not necessarily mean that the service-specific KQIs can be derived or calculated from the underlying KPIs, rather that the network and equipment KPIs represent the sources of probable cause. Hence, the relationship must be understood between service elements noticeable by the end user

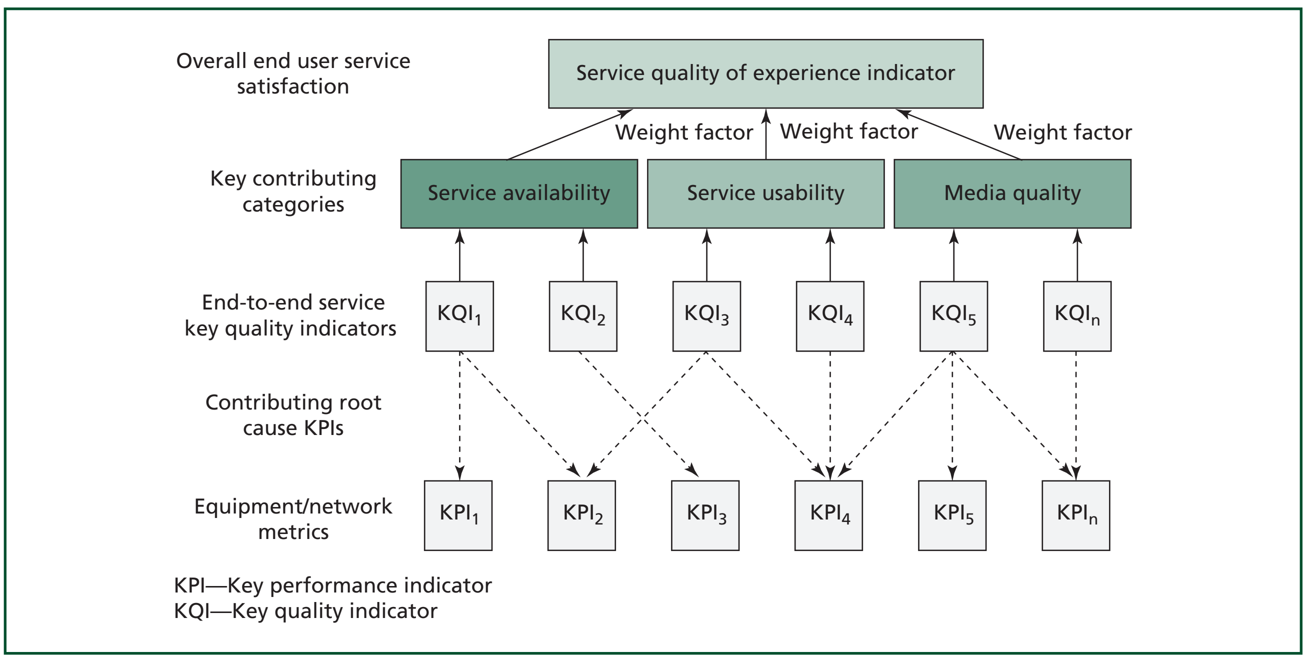

Figure 7.
Service quality of experience model.

and the service architecture components responsible for the service function. This relationship can be used to design a QoE measurement architecture.

QoE Measurement Architecture

The QoE measurement architecture defines the measurement points and required measurement equipment or interfaces in the service architecture. The QoE architecture not only defines what information needs to be extracted and at which points, but also the set of related measurement parameters that need to be obtained and grouped in such a way that the relationship with the service session is preserved. For example, if the service architecture decomposition as explained in the previous section identified the set of network elements that could potentially contribute to significant call setup delays, performance measurements on those elements can be included and grouped as a set of measurement parameters that are collected and stored during the service session. Maintaining these related performance indicators across multiple network elements as a group of measurement parameters associated with a particular service session will help to identify the root cause when the service QoE value is below a certain threshold. The challenge is to find a practical balance with the often large amount of available performance data and the minimum information required for effective root cause analysis.

Session Context and Environment Variables

User expectations can significantly depend on the context or environment in which the service is used. For example, a user may be watching an HD quality movie during a video-on-demand (VoD) service session on a television (TV) which does not support the full resolution quality of the movie. If the user is aware of the limitations of his television set, his expectations of the video quality will be modest. However, if the same user decides to invest in an expensive high resolution TV set to take full advantage of the HD VoD service, his expectations will be quite different. If the delivered video quality does not match the new expectations, the same service will result in a much lower QoE opinion from that user. The service delivery has not changed, but the environment in which the service is used has. It is, therefore, important to understand which environment variables can impact the service QoE and, when possible, to obtain information about these variables per user. Knowledge of user equipment (say from the enhanced address book) will allow the service provider to track and analyze QoE with greater accuracy.

Another complicating factor is that a user will change expectations over time. For example, voice quality in early mobile phone services was not at the same level as it is today, yet early adopters were quite satisfied with the service. However, as technology evolves, user expectations will quickly adapt accordingly.

Example Applications and Corresponding KQI/KPIs

In this section, we present example KQIs and KPIs for the following applications: VoIP, IPTV, Web browsing, video streaming, push-to-talk over cellular (PoC), and instant messaging (IM) with presence. The intention is not to give an exhaustive but rather a representative list.

As stated in the previous section, KQIs and KPIs need to be defined from a user's perceived quality perspective. Furthermore, they can also be defined from a "session" level perspective as well as from an aggregate "network" perspective. With that in mind, for the purposes of this paper we categorize KQIs into the following three classes:

1. *Service availability*. A measure of whether the user can utilize the desired application.
2. *Service responsiveness* (also called service accessibility [6]). A measure of time taken from a specific user's request to its response, or the time it takes for packets to arrive at a destination.
3. *Service quality* (also called service integrity [6]). A measure of information loss, which can either be caused by packet losses, by bit errors, or by degradations introduced due to media coding and decoding.

VoIP

Table I shows some common KQIs and KPIs defined for VoIP. The KQIs can be computed on a session level such as the "percentage of service downtime experienced by sessions from customer X" or at a network-aggregate level such as the "percentage of service downtime experienced for all PSTN calls."

Table I. VoIP KQI/KPI examples.

Category	KQI	KPI
Service availability	% of service downtime	Call setup success ratio
Service responsiveness	Call setup delay	Setup delay
Service quality	MOS score	Delay, jitter, packet loss

KPI—Key performance indicator
KQI—Key quality indicator
MOS—Mean opinion score
VoIP—Voice over Internet Protocol

Table II. IPTV KQI/KPI examples.

Category	KQI	KPI
Service availability	% of service downtime	Registration success ratio, session setup success ratio
Service responsiveness	Session start delay	Setup delay
	TV channel change delay	Control signal delay
	TV control delay	Control signal delay
	Digital video recorder control delay	Control signal delay
Service quality	Video MOS score (MOSv)	Delay, jitter, image element loss
	Audio MOS score (MOSa)	Delay, jitter, packet loss
	Combined A/V quality (MOSav) audio+video synchronization (lip synch)	Delay, jitter, packet loss

A/V—Audio/visual
IPTV—Internet Protocol television
KPI—Key performance indicator
KQI—Key quality indicator
MOS—Mean opinion score
TV—Television

Appropriate target values can then be assigned depending on the operator's business needs. For example, <1% service downtime for enterprise VoIP sessions or 3 seconds average delay for intra-network calls.

IPTV

IP video service quality metrics with a primary focus on the need for correlation of viewer perception with network performance and operation has been analyzed in [1]. IPTV QoE is not only determined by video quality, but also depends on other factors such as zapping (or channel change) time, electronic program guide (EPG) accuracy, and responsiveness to pause, resume, fast forward, fast rewind, record, and stop commands issued by the user. The TM Forum video over IP application note [16] published recommendations for service responsiveness related KPIs.

An example set of KQIs and KPIs for IPTV is shown in **Table II**. As with the VoIP example, these KQIs can be defined at a session level or at a network-aggregate level. Typically delays should be in the milliseconds for all the control signals.

With respect to the video, audio, and A/V MOS score KQIs in this table, the ITU-T presents an objective model in [7] that can be used to compare video quality, but a complete MOS model based on performance measurements is still not defined. Other KQIs based on the Internet Engineering Task Force (IETF) media delivery index (MDI) [17] have been proposed as well.

Web Browsing

Estimating the user perceived QoE of a Web browsing session can be very challenging due to the variety of Web applications as well as the complexity of the systems and components involved. An important

Table III. Web browsing KQI/KPIs examples.

Category	KQI	KPI
Service availability	% of service downtime	Session setup success ratio
Service responsiveness	Response time between request and response	End-to-end delay
Service quality	N/A (see note)	N/A
Note: TCP will attempt to correct all errors—if BER or packet loss is high, it will cause added delays in the transmission or the connection will fail. Thus both effects are included in the other two KQIs.		

BER—Bit error rate
KPI—Key performance indicator
KQI—Key quality indicator
TCP—Transmission Control Protocol

Table IV. Video streaming KQI/KPI examples.

Category	KQI	KPI
Service availability	% of service downtime	Session setup success ratio
Service responsiveness	Session start delay	Setup delay
	Pause delay	Control signal delay
	Fast forward/rewind delay	Control signal delay
Service quality	Video MOS score (MOSv)	Blockiness, jerkiness, bluriness
	Audio MOS score (MOSa)	Delay, jitter, packet loss
	Combined A/V quality (MOSav) audio+video synchronization (lip synch)	Delay, jitter, packet loss

A/V—Audio/visual
KPI—Key performance indicator
KQI—Key quality indicator
MOS—Mean opinion score

attribute of Web browsing in general is the responsiveness to a user action. Users easily tend to become impatient if they cannot access a Web site quickly enough or if the Web page is not responding fast enough to user actions such as pressing a button or entering a search query.

Table III presents example KQIs/KPIs for Web browsing. Example target values for the response time are based on [10]. A response time of <2 seconds is preferred, <4 seconds is acceptable, and 10 seconds is the maximum, as 10 seconds is about the limit for keeping the user's attention focused on the dialog.

Video Streaming

Objective models have been proposed to measure video quality, but like those for IPTV above, a complete MOS model is still not defined. We again refer the reader to [7]. In some cases (e.g., a paid live Webcast) delays could be in the milliseconds, but in other cases like in free user-generated videos expectations are lower so delays could range in the seconds. **Table IV** shows examples of video streaming KQIs and KPIs.

Push-to-Talk Over Cellular

Table V details examples of push-to-talk over cellular (PoC) KQIs and KPIs. Within the table, "talk burst confirm delay" refers to the time required for the signaling messages to flow back and forth in the network from the moment the PoC button is pushed to the playing of the chirp by the user device. "Volley latency" refers to the time it takes to gain floor control.

Open Mobile Alliance PoC requirements [11] state that the talk burst confirm delay should typically be less than 2 seconds. Volley latency from the end user's perspective should be imperceptible so a few hundreds of milliseconds are usually acceptable.

Table V. Push to talk over cellular KQI/KPI examples.

Category	KQI	KPI
Service availability	% of service downtime	Session setup success ratio
Service responsiveness	Talk-burst confirm delay	Setup delay
	Volley latency	Control signal delay
Service quality	MOS score	Delay, jitter, packet loss

KPI—Key performance indicator
KQI—Key QoS indicator
MOS—Mean opinion score

Table VI. IM with presence KQI/KPI examples.

Category	KQI	KPI
Service availability	% of service downtime	Session setup success ratio
Service responsiveness	Message delay	Session setup delay, transfer delay
	Status change delay	Control signal delay
Service quality	See note	See note
Note: IM most likely should have similar values for the response times as Web browsing so a response time of <2 seconds is preferred. Presence updates delays could be less stringent and take a few minutes.		

IM—Instant messaging
KPI—Key performance indicator
KQI—Key QoS indicator

Instant Messaging With Presence

Table VI presents example KQIs and KPIs for instant messaging with presence. IM most likely should have similar values for the response times to Web browsing so a response time of <2 seconds is preferred. Presence update delays could be less stringent and take a few minutes.

Conclusion

In this paper, we reviewed various techniques that can lead to an estimation of QoE. We believe that the use of network measurements is a technique that, although more complex than others, can lead to a better estimation and resolution of QoE issues. We also discussed standard definitions of various measurements and presented proposed values of KPIs/KQIs. Finally, we investigated some of the challenges of next-generation applications and provided a framework for addressing them.

The proposed framework is a starting point to deal with the ever-increasing complexity of QoE issues. For example, even well-established applications such as peer-to-applications like Skype* or BitTorrent* use very sophisticated mechanisms to deliver their services. As long as the service provider is satisfied with the role of bit-pipe provider and there are no guarantees around quality of service, there is no problem. However, if there is indeed an expectation of QoE triggered by “net neutrality” or other similar regulatory issues, the problem cannot simply be wished away. In addition, the demand for higher QoE is expected to increase when amateur videos in YouTube* are things of the past, and the professional quality videos in Hulu* are the norm. Let us also be aware that user equipment (particularly in the wireless industry) is moving from a pure media consumption device to a media creation and consumption device (picture/video editing, multimedia messaging), further highlighting the need for QoS for a wide range of applications. All these issues require further study.

Another item for further study, and also a matter of increasing importance, is the burgeoning world of cloud computing. The user interface, processing, and storage all may be at different physical sites connected to each other via ultra-high-speed connections. There are at least two aspects of QoE in this environment. First, the end users pay the computing application provider based on usage and/or subscription (so that they don't have to build and maintain their own computing infrastructure). Consequently, there is an expectation of QoE for the service provided by the computing resource provider. The cloud computing provider, in turn, has to depend upon the backbone connection provider to deliver end user service with the right QoE. Since the cloud computing service provider tries to minimize idle resources, the highest degree of QoE must be provided by the interconnection provider to facilitate inter-server communication. For economic reasons, it is not practical to provide highly reliable point-to-point links among servers located around the globe. A well-defined framework and methodology are going to be necessary in the near future to find the perfect balance between a high degree of QoE and reasonable economics.

Acknowledgements

The authors would like to thank Jayant Desphande from Bell Labs for his recommendations on speech quality estimation systems and Bilgehan Erman from Bell Labs for his recommendations on video quality.

*Trademarks

3GPP is a trademark of the European Telecommunications Standards Institute.
BitTorrent is a registered trademark of BitTorrent, Inc.
Hulu is a trademark of Hulu, LLC.
Skype is a trademark of Skype Technologies, S.A.
Wi-Fi is a registered trademark of the Wi-Fi Alliance Corporation.
WiMAX is a registered trademark of the WiMAX Forum.
YouTube is a registered trademark of Google, Inc.

References

[1] B. Erman and E. P. Matthews, "Analysis and Realization of IPTV Service Quality," Bell Labs Tech. J., 12:4 (2008), 195–212.

[2] International Telecommunication Union, Telecommunication Standardization Sector, "Subjective Audiovisual Quality Assessment Methods for Multimedia Applications," ITU-T P.911, Dec. 1998, <http://www.itu.int>.

[3] International Telecommunication Union, Telecommunication Standardization Sector, "Interactive Test Methods for Audiovisual Communications," ITU-T P.920, May 2000, <http://www.itu.int>.

[4] International Telecommunication Union, Telecommunication Standardization Sector, "Communications Quality of Service: A Framework and Definitions," ITU-T G.1000, Nov. 2001, <http://www.itu.int>.

[5] International Telecommunication Union, Telecommunication Standardization Sector, "End-User Multimedia QoS Categories," ITU-T G.1010, Nov. 2001, <http://www.itu.int>.

[6] International Telecommunication Union, Telecommunication Standardization Sector, "Quality of Telecommunication Services: Concepts, Models, Objectives and Dependability Planning—Use of Quality of Service Objectives for Planning of Telecommunication Networks, Framework of a Service Level Agreement," ITU-T E.860, June 2002, <http://www.itu.int>.

[7] International Telecommunication Union, Telecommunication Standardization Sector, "Objective Perceptual Video Quality Measurement Techniques for Digital Cable Television in the Presence of a Full Reference," ITU-T J.144, Mar. 2004, <http://www.itu.int>.

[8] International Telecommunication Union, Telecommunication Standardization Sector, "The E-Model, a Computational Model for Use in Transmission Planning," ITU-T G.107, Mar. 2005, <http://www.itu.int>.

[9] International Telecommunication Union, Telecommunication Standardization Sector, "Gateway Control Protocol: Version 3," ITU-T H.248.1, Sept. 2005, <http://www.itu.int>.

[10] J. Nielsen, "Response Times: The Three Important Limits," 1994, <http://www.useit.com/papers/responsetime.html>.

[11] Open Mobile Alliance, "Push to Talk Over Cellular Requirements," Approved Version 1.0, OMA-RD-PoC-V1_0-20060609-A, June 9, 2006, <http://www.openmobilealliance.org>.

[12] T. Rahrer, R. Fiandra, and S. Wright (eds.), "Triple-Play Services Quality of Experience (QoE) Requirements," DSL Forum TR-126, Dec. 13, 2006.

[13] TeleManagement Forum, SLA Management Handbook: Volume 2, Concepts and Principles, Rel. 2.5, GB 917-2, TM Forum, Morristown, NJ, July 2005.

[14] TeleManagement Forum, "Best Practice: Voice Over IP SLA Management—Application Note to the SLA Management Handbook," TM Forum GB 934, Rel. 2.0, v1.10, Apr. 2008.
[15] TeleManagement Forum, "Best Practice: Video Over IP SLA Management—Application Note to the SLA Management Handbook," TM Forum GB 938, Rel. 2.0, v1.10, Oct. 2008.
[16] TeleManagement Forum, "Technical Report: Holistic e2e SQM Framework," TM Forum TR 149, v0.5c, Jan. 2009.
[17] J. Welch and J. Clark, "A Proposed Media Delivery Index (MDI)," IETF RFC 4445, Apr. 2006, <http://www.ietf.org/rfc/rfc4445.txt>.

(Manuscript approved November 2009)

HAROLD BATTERAM is a distinguished member of technical staff with the Alcatel-Lucent Bell Labs Network Modeling and Optimization group in Hilversum, the Netherlands. He has participated in several Dutch national and European collaborative research projects including the European 5th and 6th framework projects. His current work focuses on quality of experience for multimedia applications, IPTV modeling, and real-time sensitive network applications. He has several patents pending in the areas of context aware applications, network latency equalization, IPTV, and Session Initiation Protocol (SIP) signaling. His current research interests are next-generation network applications and multi-party, multimedia real time applications. Mr. Batteram holds a B.S. degree in electrical and computer engineering from the Hogere Technische School (HTS) in Hilversum.

GERARD DAMM is product manager for the 8920 Service Quality Manager (SQM) at Alcatel-Lucent in Cascais, Portugal. His current focus is on end-to-end service quality assurance. He was previously with Bell Labs and the Alcatel Research and Innovation department. He has experience in modeling, simulation, network and service management, IP, carrier-grade Ethernet, multicast, IPTV, virtual private networks (VPNs), schedulers, optical burst switching, and network processors, with several papers and patents in these areas. He is the editor of the TM Forum SLA Management Handbook (GB917 v3). He holds a Ph.D. in computer science from Paris University, France, and is member of the Alcatel-Lucent Technical Academy.

AMIT MUKHOPADHYAY is a distinguished member of technical staff in the Alcatel-Lucent Bell Labs Network Planning, Performance and Economic Analysis Center in Murray Hill, New Jersey. He holds a B.Tech. from the Indian Institute of Technology, Kharagpur, and a Ph.D. in operations research from the University of Texas, Dallas. His current work focuses on 3G and 4G wireless technologies; the convergence of wireless, wireline, and cable services; as well as the evolution of next-generation technologies. Dr. Mukhopadhyay's research interests include architecture, performance, and cost optimization for voice, data, and video as well as converged networks for various access technologies including wireless, cable, DSL, wireless fidelity (Wi-Fi), and WiMAX*. He is a senior member of IEEE and a member of Alcatel-Lucent Technical Academy and has several publications in refereed journals.*

RHODO ODYSSEOS is a product marketing manager and business development associate in Alcatel-Lucent SAI in Cascais, Portugal. Her area of expertise is operations support systems (OSS) and business support systems (BSS). For over 15 years, she has been instrumental in consultation and development for OSS/BSS for service providers worldwide. Most recently she has been involved in consulting with customers to support the design, implementation, integration, and management of OSS/BSS solutions for next-generation networks and IP-based services. Ms. Odysseos holds a master of science degree in computer science and engineering from Czech Technical University in Prague.

LAURENT PHILIPPART is a product line manager at Alcatel-Lucent in Cascais, Portugal and has over 15 years experience in OSS. He has a telecommunications engineering background and holds an M.Sc. in satellite communications and spacecraft technology. He is the leader of the TeleManagement Forum service level agreement management team and served as editor of its IPTV Application Note. He has been involved in a number of large projects, including consulting and implementation for performance and service quality management systems for service providers worldwide and has specialized in voice and audio/video quality assessment for IP-based services.

CARLOS URRUTIA-VALDÉS is a research engineer in the Network Modeling and Optimization group at Alcatel-Lucent Bell Labs in Murray Hill, New Jersey. His current work focuses on the modeling and analysis of multimedia applications, IMS, wireless technologies, and home networks. He has published several papers on 3G networks and on IMS. He has been awarded a patent in the area of wireless backhaul optimization and has patents pending in the areas of next-generation applications and IMS optimization. His current research interests are home networking, application and traffic modeling, and the end-to-end design of wireless and wireline networks. Mr. Urrutia-Valdés holds a B.S. degree in electrical engineering from Florida International University, Miami, and an M.S. degree in computer engineering from the University of Southern California, Los Angeles. ◆