

An Overview of Quality of Experience Measurement Challenges for Video Applications in IP Networks*

R. Serral-Gracià¹, E. Cerqueira^{2,3}, M. Curado³,
M. Yannuzzi¹, E. Monteiro³, and X. Masip-Bruin¹

¹ Advanced Network Architectures Lab, Technical University of Catalunya (UPC),
Jordi Girona 1-3, 08028 Barcelona - Spain

² Faculty of Computer Engineering, Federal University of Para,
Rua Augusto Corrêa, 01, CEP 66075-110, Belém - Brazil

³ Department of Informatics Engineering, University of Coimbra,
Polo II, Pinhal de Marrocos, 3030-290 Coimbra - Portugal
{rserral,yannuzzi,xmasip}@ac.upc.edu, cerqueira@ufpa.br,
{marilia,edmundos}@dei.uc.pt

Abstract. The increase in multimedia content on the Internet has created a renewed interest in quality assessment. There is however a main difference from the traditional quality assessment approaches, as now, the focus relies on the user perceived quality, opposed to the network centered approach classically proposed. In this paper we overview the most relevant challenges to perform Quality of Experience (QoE) assessment in IP networks and highlight the particular considerations necessary when compared to alternative mechanisms, already deployed, such as Quality of Service (QoS). To assist on the handling of such challenges we first discuss the different approaches to Quality of Experience assessment along with the most relevant QoE metrics, and then we discuss how they are used to provide objective results about user satisfaction.

Keywords: QoE, Measurements, Challenges.

1 Introduction

In the past years Traffic Monitoring and Measurement has been the focus of active research within the networking community. In particular, one important topic in the area has been the network performance analysis [1, 2], which is centered in developing mechanisms for efficient Quality of Service (QoS) and Service Level Agreement (SLA) assessment in the network [3, 4]. QoS assessment addresses the different challenges associated with the accurate computation and estimation (most of the time) of network level metrics, such as, One-Way Delay, Jitter, or Packet Losses. Until recently, most approaches were network oriented and only considered the SLA among network operators and big content providers. This monitoring and measurement methodology has become insufficient as new types of content have emerged on the Internet. As a

* This work was partially funded by Spanish Ministry of Science and Innovation under contract TEC2009-07041, the Catalan Government under contract 2009 SGR1508, and CNPq(476202/2009-4)/FAPESPA(5183.UNI319.4107.07072009)).

consequence, in the recent years the multimedia content consumed by the users has dramatically increased, users are constantly downloading video streams, using video-conferencing applications, or even broadcasting their own video streams to the Internet, which, given the sensitivity of such traffic in terms of delay and packet loss, imposes tighter constraints on the end-user network and applications for a reliable content delivery, opposed to the inter-domain agreements of classic SLA. Therefore, the interest of Content Providers and Network Operators on offering a successful multimedia experience to the users becomes more important than just considering specific network performance parameters. This evolution offers new business models where the relevant factor, and the source of revenue, is not the network or even the content by itself, but rather the degree of satisfaction achieved by the customers who are paying for the service. Therefore, as a side effect, a renewed interest has appeared in operators and content providers to search for mechanisms capable of assessing such user satisfaction through Quality of Experience (QoE) techniques. These techniques can be exploited in order to offer Autonomic Network Management (ANM) capabilities to their networks, e.g., by automating the resource reservation and the traffic engineering through user satisfaction aware metrics.

In the scenario described, QoE assessment requires the definition of a new set of metrics, which must be able to objectively assess the end-user satisfaction. However, in practice, this is a challenging issue because the evaluation involves a number of subjective factors, usually not related to the network performance, e.g., the mood of the user, or the responsiveness of the system, opposed to the classical QoS assessment platforms, which are mostly network centered.

To better understand the new challenges of QoE assessment in multimedia applications, in this paper we overview the different approaches currently available for the QoE evaluation of the delivered multimedia experience, and as an use-case, we highlight the main issues that must be solved in IPTV scenarios.

The rest of the paper is organized as follows, in the next section we overview the different techniques for QoS assessment, addressing specifically the most useful for the QoE environment. In Section 3, we highlight the different approaches for QoE Measurement, which leads to the introduction of the specific metrics used in this environment in Section 4. We continue our dissertation in Section 5 by highlighting the different challenges we must face when designing a complete QoE assessment platform. Next we detail, as an use case, the issues found in the QoE assessment in an IPTV network, and finally in Section 7 we conclude with a summary of our contribution.

2 Background

This section discusses metrics and techniques historically used as a starting point for objective network quality assessment. In particular, metrics used in the QoS area which are the basis for some aspects of QoE assessment are described.

The efficient assessment of real-time multimedia applications, such as video streaming, mobile IPTV, urban monitoring cameras and other kinds of audio and video applications, is a key requirement to the success of all-IP networks. The quality level control of multimedia services aims to maximize the user's satisfaction and the usage of network

resources, as well as, to keep and attract customers, while increasing the profits of network providers.

Differently from traditional applications, such as Web browsing and File Transfer Protocol (FTP), where the quality of network delivery is not critical, given that these applications are elastic and can tolerate certain amount of network impairments, multimedia services need the content delivery with low impact on the voice and video quality levels perceived by users. However, due to the nature of IP-based networks, various types of network impairments occur along the communication paths, which need to be measured and managed.

Most Quality of Service (QoS) assessment techniques are based on network measurements to control the quality level of applications in wired and wireless systems [5]. Existing QoS metrics, such as packet loss rate, packet delay rate, packet jitter rate and throughput, are typically used to indicate the impact on the video quality level from the network's point of view, but do not reflect the user's perception. Consequently, these QoS parameters fail in capturing subjective aspects associated with human experience. These network layer objective metrics are defined as follows:

- *One Way Delay*: the amount of time from the packet generation until its reception at the destination [6]. Reasonable values for this metric range from milliseconds for Real-Time Interactive (RTI) traffic to minutes in some Video on Demand (VoD) scenarios, such as Peer-to-Peer TV (P2PTV).
- *Packet Loss*: a packet loss is a packet generated in a node, but which does not reach its destination due to network issues [7]. Packet loss is the first cause of QoS and QoE disruption, since if a packet is lost, and not retransmitted on time it will get discarded, consequently producing video or audio disruptions. In real-time scenarios packet losses can be caused by buffer underruns or by network malfunction.
- *Jitter*: after packet loss, jitter is the second cause of service disruption. Jitter is defined as the variation in delay of consecutive packets in the stream [8]. If there is high jitter in the network it can cause buffer underruns, and therefore packet losses in the communication. Jitter is smoothed in the application layer with the de-jitter buffers, which increase delay, but reduce quality degradation.

In order overcome the limitations of current QoS-aware measurement schemes regarding subjective aspects related to human perception and to enable more accurate assessment of the quality experienced by users, Quality of Experience techniques have been introduced [9, 10]. QoE measurement operations in wired and wireless system can be used as an indicator of how a networking environment meets the end-user needs. Existing quality metrics aim to meet wider expectations by providing a good correlation with subjective ratings, cross-content, cross-distortion accuracy, and low computational complexity in order to enable in-service applications.

3 QoE Measurement Approaches

Two main QoE assessment approaches are receiving an important attention from industry research groups, commonly referred as Content Inspection/Artifact-based Measurement and Network-based Measurement. The former operates in the decoded content and

can be configured to calculate the application quality level by using QoE metrics that range from simple pixel-to-pixel comparison schemes to sophisticated Human Visual System (HVS)-based frame-level artifacts analysis [11]. While the latter aims to predict the multimedia quality level based on information gathered from packet and network conditions without accessing the decoded video [12].

Objective and subjective content inspection measurement approaches are used to find distortions that can be introduced at any stage in the end-to-end multimedia delivery system, such as coding/decoding or network congestion. Existing objective video quality assessment models have been developed to estimate/predict the quality level of multimedia services by using QoE metrics that approximate results of subjective quality assessment [13]. QoE subjective measurements assess how audio and/or video streams are perceived by users [14], i.e., what is their opinion on the quality of particular audio/video sequences, as described in ITU-T recommendation BT 500 [12].

Content inspection measurement solutions can be assessed with No-Reference (NR), Reduced-Reference (RR) or Full-Reference (FR) approaches. The FR method accesses the original and received content in the evaluation process. The RR uses an alternative channel between senders and receivers to transmit the parameters of the delivered content in a reduced way or use watermarking schemes, while the NR approach assesses the content quality level without any knowledge of the original material.

Due to the time and processing demands, as well as feasibility issues of content-based assessments, multimedia quality prediction mechanisms can be used. These schemes predict the quality level that a specific content will have after the encoding process, based on the encoding parameters, packet inspections, and network conditions. Further processing of the original data is not required, minimizing this way the associated complexity and resource consumption. Opposed to this, network-based approaches verify all transmitted packets related with the application, the main issue is that such approaches need to perform deep-packet inspection. Additionally, they need to gather information about the current network conditions, such as packet loss rate and packet one way delay, to be used in the predictive process. The final quality level assessment decision can be taken based on previous information together with information about the multimedia characteristics, such as frame-rate, Group of Picture (GoP), frame type and dependence only available at application level.

Hybrid Content Inspection and Network Measurement approaches have also been proposed [15]. The main reason for the development of this kind of scheme is to allow network operators to combine the benefits of the two previous approaches and adjust performance, complexity and feasibility, as well as operational cost issues according to different needs, multimedia content type, networks and equipments.

4 Metrics

A metric is defined as “*a system of related measures that facilitates the quantification of some particular characteristic*”¹, and as such it implies a well defined and measurable component. However, when defining metrics for QoE assessment there are many

¹ <http://www.dictionary.net/metric>

subjective factors which might bias the user satisfaction towards a service, hence increasing the complexity of computing objective metrics. As a consequence, there are notable research efforts in objectifying such factors in order to have a measurable value of the user satisfaction in general.

The Mean Opinion Score (MOS) is metric typically used in QoE scenarios as we will detail later in this section. There is however, a main drawback of using this metric alone, because it only considers video and audio quality, ignoring other important aspects such as application responsiveness or degree of interactivity, which are also relevant for a complete QoE assessment framework.

Many alternatives have been proposed regarding classification schemes of mechanisms for video assessment [16]. Going beyond those approaches, we propose a more generic classification, which considers not only the video quality assessment methods but also other classes of metrics not directly related to perceived video quality. The two different types of metrics are, *i*) direct metrics, and *ii*) indirect metrics. The direct metrics are those which directly affect the perceived quality of the multimedia flows, while indirect metrics refer to other factors such as response time, or degree of interactivity, that impact on the service usability, but do not affect directly to the multimedia traffic. Both direct and indirect metrics have different relative importance depending on the service, and they affect the perceived experience in the delivered service in different ways.

4.1 Direct Metrics

Direct metrics consider factors that directly affect the user perception of the multimedia experience, i.e., received audio and video quality. Direct metrics are obtained from various types of data (potentially at different layers), such as delay variations or packet (frame) losses related with network performance, and codec information as determined by the application.

There is a broad diversity of direct metrics, it is not the goal of this work to provide a full list, but instead we refer to the most relevant metrics in the area of video quality assessment. For a more detailed list the reader is referred to [15, 16].

- *Peak Signal to Noise Ratio* (PSNR): it is a basic, yet important metric that assesses the similarity between two different images. It uses the Full Reference QoE measurement approach described above. PSNR computes the Mean Square Error (MSE) of each pixel between the original and received images, represented in dB. Images with more similarity will result in higher PSNR values.
- *Structural Similarity* (SSIM): the main drawback of PSNR is that it does not consider how human perception works, hence in some cases it cannot detect some human perceptible video disruptions. To address this shortcoming, SSIM combines luminance, contrast, and structural similarity of the images to compare the correlation between the original image and the received one. Similarly to PSNR, SSIM is also based on the Full Reference scheme.
- *Video Quality Metric* (VQM): improving the approaches described above, VQM detects human perceivable artifacts on the images, by considering blurring, global noise, and block and color distortions. This metric also uses the original video, hence using the Full Reference QoE measurement approach.

- *Mean Opinion Score (MOS)*: MOS was originally devised for audio streams, it combines delays, perceived jitter at application layer, codec used for the communication, and packet losses—also at application layer. In the area of video QoE assessment can be considered as a meta-metric given that it considers values from other metrics to generate the final computed user perception. The most used extension of MOS was proposed in [17] where the authors propose a mapping between PSNR and MOS.

All the above direct metrics have the main issue that they require the presence of the original video frames in order to compare and assess the user perceived QoE. On the other hand, when there are no reference points of analysis, the solutions require low level performance information about frame losses, delays, and jitter, using the NR and RR models above, as we found in the Moving Pictures Quality Metric (MPQM) model—with its particular implementation V-Factor—in the context of IPTV traffic [18]. Whilst MPQM results are less accurate than the ones obtained with the previous approaches, it is usable in more scenarios, in particular when performing real-time assessment or in environments where the source video is not available.

4.2 Indirect Metrics

Indirect metrics consider properties that affect the multimedia experience, but that are not directly related to the quality of the multimedia content, such as artifacts that might cause undesirable side effects that affect the final user perception of the service. It is worth noticing that indirect metrics have not been studied in detail in the networking area, where the focus has been clearly centered in the network performance and in direct metrics assessment. Nevertheless, as the multimedia applications continue to expand, this research area will become more appealing in the near future.

Given their nature, indirect metrics are closely related with the type of delivered service, and in general they require access to application or run-time information since they depend on user actions such as pressing the “play button”. In this work, we describe the most relevant indirect metrics for the assessment of video applications.

- *Start-up time*: this metric defines the time span since the user queries the system about a specific content until he/she receives it. A classic example of this is the waiting time experienced by an user from the time he/she clicks on a video link until the video is actually reproduced on the user’s screen.
- *Response time*: this metric is an extension of the start-up time. It is defined as the time span since the user issues a command or performs an action until it is acknowledged by the system and an answer arrives at the user, e.g., when the user pauses a video. In video-conferencing this metric is interpreted as interactivity degree, which determines the fluency of a conversation among the conference participants.
- *Delivery Synchronization*: when several users are using a service, the content should be received at the same time by all the participants. This is critical in on-line gaming where the users need to react fast to the actions of others.
- *Freshness*: this metric specifies the time span from content generation until its reception by the user. It is specially important in live video streaming, where the users

want the content as fast as possible (e.g., to celebrate a goal in a football match). This metric greatly affects applications performing live streaming in P2P networks (P2PTV).

- *Blocking*: this metric is closely related the direct metrics because it models the jerkiness of the video, normally caused by empty buffers on the receiver.

As it can be observed, all the above indirect metrics are related in one way or another with the computation of time intervals. The issue of time interval computation has been broadly studied in the QoS and performance analysis fields [1], which involves high precision time synchronization techniques between the end-points. Luckily, in the area of QoE, such sub-millisecond accuracy is not required, since in most scenarios, human perception has resolutions which range from tens to hundreds of milliseconds, which is fairly easy to accomplish with state-of-the-art hardware.

5 Challenges

When designing a QoE assessment platform many factors must be considered. In this paper we focus our the discussion on highlighting the main concerns found in such designs in the area of multimedia applications.

It is not our goal to propose an integrated solution for QoE assessment, but rather to identify the main issues from two different viewpoints, namely, issues related with data acquisition, and issues related with results analysis and interpretation.

5.1 Data Acquisition

Any measurement platform, whether it is aimed at QoS or QoE assessment needs to collect data. Classically in QoS environments such traffic collection involved computationally intensive operations coping with huge amounts of data in high speed links which forced sampling [4] or data aggregation schemes to provide some means of scalability. Contrary to this, in the case of QoE environments, since we are coping with user perception, we cannot use backbone equipment to measure network performance, because the traffic collection must be issued as close to the end-user as possible to have an accurate estimate of the user's perception. This has the advantage that the amount of data to process is greatly reduced at the user end-point. But, despite of this, if the application does not have any integrated measurement techniques, it greatly complicates the traffic analysis, given that application dependent data is necessary. On top of that, when using handheld mobile devices, in general the computational constraints and battery life can impose important issues to perform such assessment.

When acquiring the information in order to assess the QoE level, we have to answer three different questions, *i) Which* information do we have to acquire?, *ii) Where* do we have to collect it?, and *iii) How* do we perform the acquisition?. The *Which* is determined by the metric selection, while the *Where* and the *How* determine the accuracy we will obtain. These issues are further detailed in the next sub-sections.

Metric selection: The context in which the collection is performed will determine the metrics that can be computed and the QoE measurement approach that can be used, i.e., FR, NR, RR. Moreover, this selection will have critical impact on the assessment accuracy.

Normally, when working on real-time QoE assessment platforms the original transmitted data is not available. This implies that Reduced Reference or Non Reference techniques must be used, and thus accuracy will be bounded, specially when Non Reference is used. On the contrary, Full Reference techniques guaranty more accurate results, but with the limitation of not being usable in real-time environments.

Another aspect regarding the accuracy is related on whether we have access to application dependent information. The critical issue in this case is *Where* to perform the assessment, e.g., in the end-user application, at IP layer in the end-user host, or in the egress router. Choosing the location *Where* to perform the analysis will determine *How* the data acquisition must be performed. As an example, QoS solutions classically use the egress router of the network to perform the traffic analysis. At this point, gathering application dependent data is challenging, specially regarding the computational demands of deep packet inspection, and the broad range of applications and configurations existing nowadays. When the acquisition is done locally at the end user's premises (e.g., in the set-top box in IPTV environments), more application dependent information is available, such as used codec, which greatly improve the computed QoE accuracy. Finally, if the gathering is performed by the end-user's application, we have access to user's important data, such as, user perceived delays and losses, and status of the de-jitter buffers, which help further increasing the accuracy of the user perceived QoE.

5.2 Results Analysis

Selecting the proper metrics, and performing an accurate metric computation is only the first step to a successful QoE assessment. The next step is to decide whether the user is satisfied or not. In general, it is broadly accepted that the lower threshold for a successful experience of a video transmission is a MOS value of 3. Nevertheless, as we have highlighted in Section 4, MOS is derived from other metrics, which in some cases do not consider all the necessary factors for a complete assessment. On top of that, depending on the type of the offered service some metrics—specially indirect metrics—have radically different constraints.

To illustrate this point let's consider three different types of service, namely, Real-Time Interactive (RTI), Video on Demand (VoD), and Live Streaming (LS). On-line gaming platforms and video-conferencing are examples of RTI applications, while generic video streaming applications belong to the VoD category, where the streamed content is not live, but stored in a particular location and then it is reproduced. Finally, streaming applications fall in the LS category, where live content, such as sport events, is streamed to the end user.

Table 1 details the effects of direct and indirect metrics on the different categories of applications identified above. It can be observed in the table that each metric affects each application category in different ways. We grouped the direct metrics together

Table 1. Metric effect on the different application categories

Metric	RTI	VoD	LS
Direct Metrics	High	Medium	High
Start-up	High	Medium	High
Response time	High	Medium	Low
Delivery Synchronization	High	Low	High
Freshness	High	Low	High
Blocking	High	High	High

because all assess the image quality, differing only on the methodology and accuracy but not on the effects depending on the service category.

In the table we can notice that RTI applications are very sensible to all the different metrics, both direct and indirect, since interactivity forces the use of small de-jitter buffers and, due to required small lags in the communication, using retransmissions is not feasible. Another important point to consider is the Delivery Synchronization, when several (more than two) users are sharing a video-conference. If the reception of the audio and video is not synchronized among the interlocutors, it will add unnecessary noise to the conversation, e.g., when one participant is responding to a statement, while some of the others are still listening to the previous sentence.

In the case of VoD applications, the constraints are more relaxed, specially for the indirect metrics. Since Delivery Synchronization and Freshness have little or no interest in this case, as the user selects the content to view individually and is not interested on its age or if anyone else is also receiving it. Regarding the start-up time, it determines the waiting time until the video starts at the destination. High values in this metric can be annoying, but users can cope with lags up to several seconds without much concern. The effects are similar for the Response Time, since in this case, when a user, e.g., wants to fast-forward a video, normally a small lag in the action is acceptable. As expected, the direct metrics have a relevant impact in VoD. But depending on the transmission protocol, the frame losses can be greatly avoided by retransmissions given the large de-jitter buffers used by these applications.

Finally regarding LS, this type of service has tighter constraints than VoD, but looser than RTI. In such environments, the content should arrive as fast as possible to its destination, because the user wants the contents in “real-time”, and if the start-up takes too long, or the content is not “fresh” compared to other sources, the user can decide to switch to other means of broadcasting, e.g., digital TV, satellite, and thus the revenue of the provider will decrease. Another relevant metric in this type of service is the Delivery Synchronization. For instance, if someone is watching a sport event, and is receiving the contents delayed relatively to other sources, the user is most likely to switch to this other delivery system to have more up-to-date content. On the contrary, the Response Time has low relevance in this case, because, in general, in such live events the actions the user can perform are very limited (i.e., stop the playback).

It can be noted that regardless the application category, all metrics are affected equally by the *Blocking*, this is caused because, even with large buffers, a multimedia stream must be played smoothly on the end-user premises for a proper QoE.

6 Use-Case: IPTV

The use of IPTV services has been growing up in the recent years, and the user perceived quality is the most important feedback parameter for providers. Recent studies have shown that 84 percent of IPTV providers think video quality monitoring is critical or very important for the success of the business.

Unlike QoS assessment operations that are performed along the communication paths, QoE measurements in IPTV scenarios involve coding/decoding processes, analysis of the transmission links and user perspective, as presented in Fig. 1.

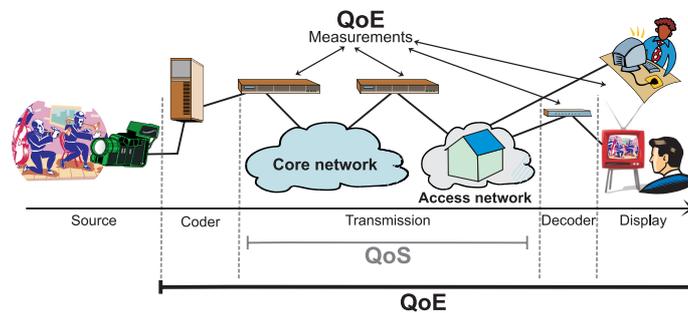


Fig. 1. Applications of Objectives Quality Assessment Model in IPTV

In fixed and mobile IPTV systems, most of the technical challenges are related with terminal capabilities, wireless/wired links, business, mobility issues and QoS and QoE control. In order to simplify the explanation of this use case, only QoS and QoE measurements will be explored. It is assumed the delivery of a MPEG-4 AVC/H.264 High Definition video over Digital Subscriber Link (DSL) IPTV access networks, commonly used in North American and other countries. The GOP length is 32 frames with IBBBP structure to allow a high quality, and the packets are transported over RTP/UDP/IP.

In the first example, Network-based Measurement is configured in the system to assess the quality level of the IPTV service. A centralized assessment mechanism (distributed schemes could also be used) is placed in the access network to predict the video quality level based on information about packet header, packet payload information and network conditions, without decoding the ongoing video content.

The mechanism verifies all packets being delivered associated with the IPTV service, by accessing the packet beyond layer 2 information and collecting data about the frame type, frame dependency, GOP, transport protocol, content characteristics, packet loss rate, packet delay rate and packet jitter. Together with utility function models [19] for IPTV applications, this mechanism maps network-content based information into user utility function and presents a final score about the quality level of the delivered application. For instance, if an *I* frame was lost and due to the application frame-dependency, it will result in 14 impaired frames and low quality video sequence will be received.

Unlike the previous approach, the Content Inspection/Artifact-based Measurement scheme studies the decoded video by using objective metrics. The assessment can be done in network edges, in the set-top box, or even in user's terminals. To provide real-time measurements, Non Reference metrics are used to define the IPTV quality level, such as blockiness, blur, flickering and noise. In this example, the blockiness is configured as an objective metric and each blocking artifact has at least one visible corner. After decoding the video, the mechanism measures the absolute difference of pixels luminance (intra-pairs and inter-pairs) and outputs a ratio between the sum of pixel intra- and inter-pairs difference as a quality score. Based on utility function and blockiness values, a final result is contained to inform the quality level of the IPTV service.

7 Conclusions

In this paper we overviewed the different challenges found in the design of QoE assessment platforms. The study focused on the case of video delivery over the network. As we observed, the first step to have a full-fledged QoE assessment mechanism is to decide which are the most suitable metrics for the quality assessment. To ease the selection the metrics were separated into two different categories, direct and indirect. Direct metrics have been the most studied, these type of metric require specific information about the delay variation and the packet/frame losses in order to assess the quality of the received image from the network point of view. As we pointed out, depending on the amount of available information, three different models can be used, namely, Full-Reference, Reduced-Reference and No-Reference, each one with different requirements and levels of accuracy.

The main challenges concerning data acquisition and results analysis were identified. Regarding data acquisition, we have discussed that three questions must be answered prior to the design of the assessment framework, namely, *Which*, *Where* and *How*. The answer to these questions will determine the different capabilities that will be deployed in the assessment system.

Finally, and as a use case we discussed about a broadly used service, i.e., IPTV, where we detailed the different steps and requirements that must be met in order to deploy a sound QoE assessment platform for such a service.

References

1. Crovella, M., Krishnamurthy, B.: Internet Measurement. Infrastructure, Traffic, and Applications. John Wiley & Sons, Ltd., Chichester (2006)
2. Serral-Gracià, R., Barlet-Ros, P., Domingo-Pascual, J.: Coping with Distributed Monitoring of QoS-enabled Heterogeneous Networks. In: 4th International Telecommunication Networking Workshop on QoS in Multiservice IP Networks, Venice, Italy, February 2008, pp. 142–147 (2008)
3. Serral-Gracià, R., Labit, Y., Domingo-Pascual, J., Owezarski, P.: Towards End-to-End SLA Assessment. In: INFOCOM 2009. The 28th Conference on Computer Communications, pp. 2581–2585. IEEE, Los Alamitos (2009)

4. Sommers, J., Barford, P., Duffeld, N.G., Ron, A.: Accurate and Efficient SLA Compliance Monitoring. In: Proceedings of ACM SIGCOMM, Kyoto, Japan, August 2007, pp. 109–120 (2007)
5. Jian-ren, Y., Rui-min, H., Jun, C., Jian-bo, Z.: A Service-Oriented Framework of Distributed QoS Measurement Based on Multi-Agent for Overlay Network, pp. 158–162 (February 2009)
6. Almes, G., Kalidindi, S., Zekauskas, M.: A One-way Delay Metric for IPPM. RFC 2679 (September 1999)
7. Almes, G., Kalidindi, S., Zekauskas, M.: A One-way Packet Loss Metric for IPPM. RFC 2680 (September 1999)
8. Demichelis, C., Chimento, P.: IP Packet Delay Variation Metric for IP Performance Metrics (IPPM). RFC 3393 (November 2002)
9. Piamrat, K., Viho, C., Bonnin, J.M., Ksentini, A.: Quality of Experience Measurements for Video Streaming over Wireless Networks, pp. 1184–1189 (April 2009)
10. Wang, B., Wen, X., Yong, S., Wei, Z.: A New Approach Measuring Users' QoE in the IPTV, pp. 453–456 (May 2009)
11. Romaniak, P.: Towards Realization of a Framework for Integrated Video Quality of Experience Assessment, pp. 1–2 (April 2009)
12. Lin, T.L., Kanumuri, S., Zhi, Y., Poole, D., Cosman, P.C., Reibman, A.R.: A Versatile Model for Packet Loss Visibility and its Application to Packet Prioritization. *IEEE Transactions on Image Processing* 19(3), 722–735 (2010)
13. He, Y., Huang, T.: Objective quality definition of scalable video coding and its application for optimal streaming of FGS-coded videos. *Computer Communications* 32(1), 34–40 (2009)
14. Janowski, L., Papir, Z.: Modeling subjective tests of quality of experience with a Generalized Linear Model, pp. 35–40 (July 2009)
15. Romaniak, P., Mu, M., Mauthe, A., D'Antonio, S., Leszczuk, M.: Framework for the Integrated Video Quality Assessment. In: 18th ITC Specialist Seminar on Quality of Experience, Blekinge Institute of Technology, Karlskrona, Sweden (May 2008)
16. Cerqueira, E., Janowski, L., Leszczuk, M., Papir, Z., Romaniak, P.: Video Artifacts Assessment for Live Mobile Streaming Applications. In: Mauthe, A., Zeadally, S., Cerqueira, E., Curado, M. (eds.) FMN 2009. LNCS, vol. 5630, pp. 242–247. Springer, Heidelberg (2009)
17. Klaue, J., Rathke, B., Wolisz, A.: EvalVid - A Framework for Video Transmission and Quality Evaluation. In: Proceedings of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation, pp. 255–272 (2003)
18. Cognet, Y.: Measuring IPTV QoS performance at the box (2006), <http://www.videsignline.com/>
19. Mu, M., Mauthe, A., Garcia, F.: A Utility-Based QoS Model for Emerging Multimedia Applications. In: The Second International Conference on Next Generation Mobile Applications, Services and Technologies NGMAST, Cardiff, UK, pp. 521–528. IEEE Computer Society, Los Alamitos (2008)