Web QoE Evaluation in Multi-Agent Networks: Validation of ITU-T G.1030

Eva Ibarrola, Fidel liberal, Ianire Taboada
Department of Electronics and Telecommunications
University of the Basque Country, UPV/EHU
Bilbao, Spain
e-mail: {eva.ibarrola, fidel.liberal,ianire.taboada}@ehu.es

Rodrigo Ortega
Department of Industry. System Integration Northline
Atos Origin S.A.E.
Bilbao, Spain
e-mail: rodrigo.ortega@atosorigin.com

***Abstract*— User's requirements have become a key factor for any Quality of Service (QoS) management model to succeed. The advent and rise of new broadband services and network architectures (Triple-Play-Services, NGN…) depends on the ability of providers to achieve user's expectations in these scenarios. For that reason, the overall end user's perception (Quality of Experience – QoE) must be audited, on a regular basis, to address changing user's needs. This paper presents a general system developed to evaluate QoE on IP networks. The system architecture is designed to be capable of emulating multi agent networks and dynamically changing conditions. In addition, the results of a Web browsing QoE experiment, laid out within this emulation system, are described. The experiment was conducted on the basis of ITU-T Recommendation G.1030, and aimed to update the perceptual model, provided in this Recommendation, to today's user requirements and technical improvements.**

Keywords - QoS, QoE, QoE evaluation, Network Emulator

I. INTRODUCTION

Although technical Quality of Service (QoS) has been one of the most important research hot topics in recent years, the idea of a notion of QoS extended to both user's perception and network performance parameters is mostly accepted [1, 2]. Therefore, a lot of discussion has taken place concerning the extension of Quality of Experience (QoE) [3-6]. ITU-T has defined QoE [7] as "*the overall acceptability of an application or service, as perceived subjectively by the end-user*." ITU-T emphasizes that Quality of Experience includes the complete end-to-end system effects (client, terminal, network, services infrastructure, etc.) and that the overall acceptability may be influenced by user expectations and context also.

Estimating Quality of Experience seems not to be a simple task. ITU-T Recommendation G.1030 [8] provides some guidelines to achieve this goal. Nevertheless, this recommendation's proposed model suffers from not being amended to current user's requirements and today's network architectures. In order to update G.1030 we considered feedback from previous research on opinion models and Web service analysis [9-12]. Nevertheless, carrying out real world end-to-end experiments is usually too difficult. Therefore, software tools to emulate the behavior of the network are very useful since they allow researchers to easily handle QoS parameters to study how different conditions may affect to final user's perception.

Many network emulation systems have been traditionally used to explore the behavior of network protocols and to test and evaluate user's applications. However, the complex interactions between different elements in today's networks lead to an end to end (e2e) behavior that is difficult to model and, therefore, emulate. So, only multi-agent approach architectures can provide a realistic emulation for these global networks. Unfortunately, most of widely available emulation tools [13-15] are based on a single network node, not satisfying current networks emulation requirements.

In this paper we present a QoE evaluation system that, by means of a complex virtualized architecture, provides researchers with the capability of emulating multi-agent networks and changing conditions and estimate QoE for different services and applications.

The paper is organized as follows. In Section 2, we present the general architecture of the multi-agent network emulation system and describe its characteristics and functionalities. In Section 3, we give the details of ITU-T Rec. G.1030, a basic framework for QoE evaluation. In Section 4, we relate the experiment conducted to validate G.1030 perceptual model within the emulation system. In Section 5, we present results of the experiment and discuss their implications. Finally, in Section 6, we conclude the paper with final remarks and future work.

II. NETWORK CONDITION EMULATION SOFTWARE ARCHITECTURE

In order to fulfill the need for multi-agent network emulation software to evaluate QoE, we have developed a system that allows emulating both single network elements and several interconnected network equipments and links. Since virtualization techniques allow building complete test environment in a regular PC we use virtual machines to emulate each node or network equipment. In this way, infrastructure and maintenance costs can be reduced, provision of network elements is immediate and execution of emulated applications and services can be simplified. VNUML virtualization system [16] has been chosen since it provides the possibility of multi-agent network emulation based on User Mode Linux (UML) and XML based scenario deployment tools.

In our testbed network conditions are distributed to each element and managed by a central system. Netem GNU/Linux kernel module [14] is used on each one of these virtualised network elements in order to emulate different network conditions and characteristics. We chose Netem because it is simple and easy to integrate in virtual machines. In Fig. 1, we represent the general architecture of the network emulation system.

In order to carry out traffic classification and emulation each module is composed by a queue discipline architecture. These queue disciplines, for BW control, together with Netem, provide a reliable and effective network condition emulator system which allows applying different characteristics to different types of traffics. The queue system divides different user's traffic in different queues. It is also possible to divide a user's traffic in different flows.

Once the administrator has configured the emulated scenario, system is booted and system users are able to reserve, modify and free their queues with the conditions established in the configuration files of the request. This remote service is available through the Web Service management system [17]. For each request the native system communicates with each Web Service virtual machine.

III. QoE Evaluation Framework

In the previous sections, we concluded that Quality of Experience estimation is not an easy issue. QoE is specified as end to end and should be agnostic to network deployment architectures and transport protocols or access network technology (xDSL, xPON, wireless) [18], although these technologies do have an impact on e2e network-based QoS (NQoS). Furthermore, QoE is definitely tightly linked to user experience and expectation and each specific experience is related to a role [19]. Therefore, QoE should be analyzed on a service-by-service basis.

We will center our study in the Web services because the number of Web users and Web applications are still rapidly increasing.

Many authors [20-23] have proposed QoE models to correlate performance metrics and user's perception on web service. Still, none of them is general enough to deal with all the QoE extensions.

ITU-T Recommendation G.1030 gives some guidelines to fill this gap. This Recommendation provides a framework to evaluate QoE based on [8]:

- Performance of the IP network of interest: based on relevant measurements or network simulation results.
- Specifications of the application of interest: in terms of its governing protocols with specified options, or a model of the application using network performance and customer appliance performance as input and producing a key metric of application performance, i.e., session time, as a result.
- A perceptual model: intended for the applications of interest to interpret the application performance as an estimation of the quality experienced by a typical population of users.

Fig. 2 illustrates the complete proposed process. In addition, a methodology for validating this framework, applied to Web browsing applications, is specified (see section IV). Nevertheless, all network, application and perceptual suggested models in this recommendation fail to come up to current network capabilities and today's user's needs. As an example, download times used for Web browsing experiment went up to 60 seconds (Fig. 3). Most users of today's Internet services would not wait 60 seconds for a web page download. So, our research attempts to bridge the gap between G.1030 and current services characteristics in order to bring QoE models up-to-date.

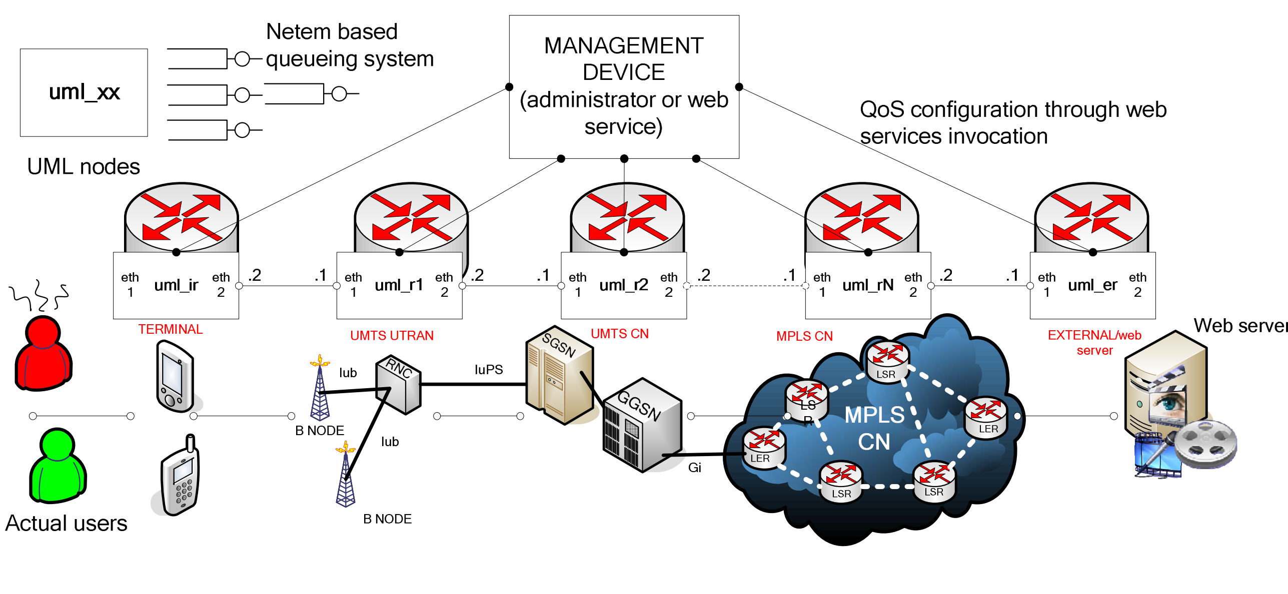

Figure 1. Network emulator system.

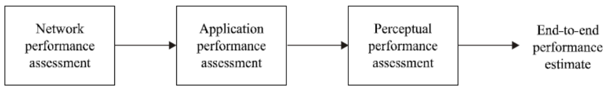

Figure 2. G.1030 Framework

IV. Evaluation of QoE in Web Services

In this section, we describe the QoE experiment carried out over Web browsing service for validating the QoE emulator system and updating ITU-T Rec. G.1030.

Fig. 4 illustrates the different alternatives suggested in G.1030 to solve the QoE evaluation process. The highlighted options was considered the most suitable to be adopted in our QoE evaluation system. The experiment was carried out as close as possible to G.1030's conditions. Some updating of these conditions was made to enhance results to today's user's requirements and technologies.

The experiment was designed to describe the response of users to download time in HTTP transactions. It was important to simulate a real experience and, in order to determine the effect of different environments, three web browse experiments with different time scales were defined. G.1030 used scales around 6, 15, and 60 seconds, representing fast, moderate and slow network contexts, respectively. To be applicable to present networks, we updated these values, based on previous experiments, to 1.8 sec. for fast networks (0.8 standard deviation), 4.3 sec. for moderate networks (3 deviation) and 6.8 for slow networks (5.5 deviation).

G.1030 included three different experiments within the three different contexts (slow, moderate and fast context). A single experiment consisted of 49 sessions and, in each session, the participant retrieved two pages: one search page and a page that results from the requested search. Fig. 5 shows the time line of such a session.

For each of the 49 sessions different combinations of T1 through T4 were configured. This was done using java scripting and recorded movies of the web browser content (using NISTNET). Thus, the sum T1+ T2+ T3+ T4 was ranged from 0 to the maximum time scale in the context of experiment (fast, moderate or slow).

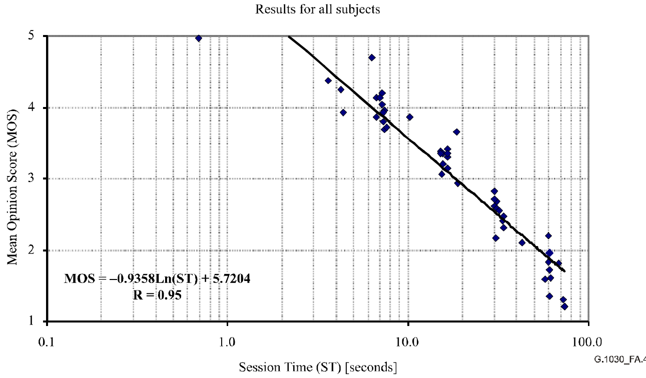

Figure 3. Results with time scale 60 (G.1030)

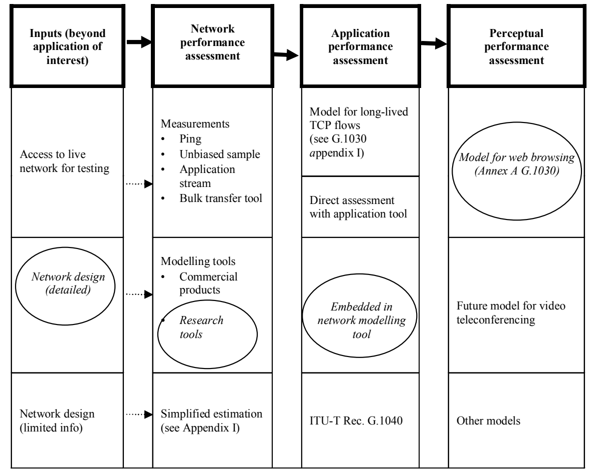

Figure 4. QoE evaluation alternatives (G.1030)

In our experiment, we followed these guidelines introducing some improvements. First, for the network and application performance assessment, we used our Network Emulator Architecture, which granted greater precision and flexibility to the session time simulation. We emulated fast, moderate and slow environments within our system, so it was easier to adjust the required session times of each session. The emulator system gave us the direct relation between network performances (delay) and the application performance metrics (session time) which, in G.1030, is considered the first step for QoE evaluation. Fig. 6 shows some of these correlations drawn up from the system.

On the other hand, as mentioned above, we changed the time values of the sessions of G.1030. We also decided to set up the experiment in two different phases. In the first one, all the participants went through the three contexts experiments (of 49 randomized sessions) from the slow one to the fast one, and, on the second one, they went from the fastest one to the slowest one. In this way, we could analyze how previous experiences had influenced the perception of current user's perception. At the moment, it is widely accepted that user's expectation influences their QoE [24] so we introduced a simple survey, previous to our experiment, to collect each participant's expectations about Web browsing and other interesting data for statistics (age, male/female, time of use of web browsing, upload/download preferences).

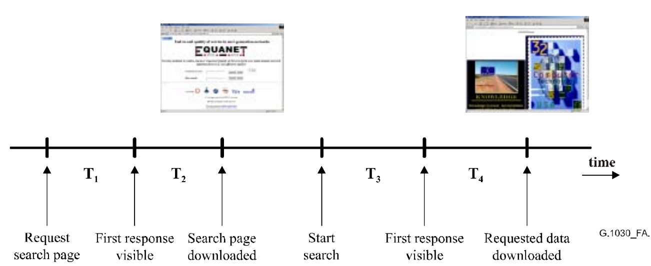

Figure 5. G.1030 timers

All the experiments were carried out through our web survey platform developed along with the network emulator to work together for the QoE evaluation process.

Aware of the differences in behavior between trained and untrained users, and in order to extend the validation of the experiment for a large population of users, two different groups of expert and unskilled users was distinguished. In our experiment a set of 11 professors, lecturers and researchers from the School of Engineering in Bilbao compose the expert group and 25 Bilbao's High school and School of Engineering students the unskilled group.

In what concerns the participant, the experiment was organized according to the following steps: previous to the experiment, we configure network conditions and surveys, for different experiments and sessions, in the network emulator system and the survey server. Instructions for the experiment (about phases and performance of the experience) were e-mailed to participants the day before the experiment, so they become familiar with them. Once everything was ready, the experiment started. The browser showed the welcome page and next participant was required to answer some questions related to personal data and Web browsing experience. After this, a general demo was shown requiring searching a particular search web page. Then the participant was also requested to search for a given page and, after result was displayed, their opinion score was collected according to Absolute Category Rating (1-5) [25].

Once the demo was over, we provided them with some training, where the best and worst session times were shown and associated to 1 through 5 score. From this point on, participants followed the experiment by themselves. When network context changed, new training was provided to update user's expectations. To wrap up this section we show in table I some of the improvements of our experiment with respect to G.1030.

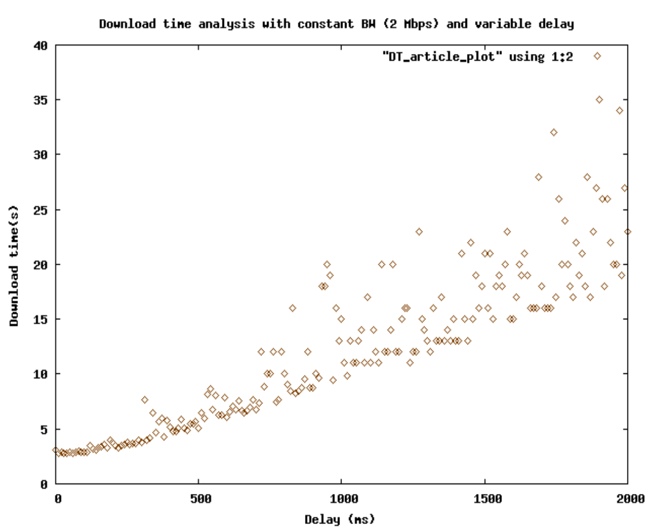

Figure 6. Application performance

TABLE I. EXPERIMENT'S COMPARATIVE

Function	**G.1030**	**Experiment**
Network conditions control	Web browse recorded using NISTNET	Network emulator system control
Environments (Session Time)	Fast: 6 s. Moderate: 15 s. Slow: 60 s.	Fast: 1,8 s. Moderate: 4,3 s. Slow: 6,8 s.
User expectation	Not considered	Considered
User Feedback	Not considered	Considered
Experiment	3 (147 sessions)	6 (294 sessions)
Participants	Expert: 12 N. expert: 17 (media)	Expert: 11 Non expert: 25 (in all cases)

V. RESULTS

In Table II we show the results of the experiment. We can observe, for each context, time limit that leads to maximum MOS 5 (maximum tolerance) and a time limit below which MOS is 1 and, therefore, user cannot appreciate any improvement. The opinion model for each experiment is also presented. In fig. 7, we represent graphically one of the results (first row in table II).

Certainly, the same expected behavior than the one provided in G.1030 is found: the user perceived quality goes down linearly with the logarithm of the session time. As we suspected, even though G.1030 is still a reference point for evaluation of QoE, the opinion model needed to be updated. A comparison of the results with other standards and scientific works [26, 27] reveals that, predictably, thresholds of acceptability change over time.

Consequently, it is important to anticipate to these the changes, studying QoE over emulating systems, like the one presented in section II, in order be able to fulfil users' expectations.

A second key finding of the research is that even though the relations between objective QoS and perceived quality have become stronger - the opinion model shows this fairly - the overall acceptability of the application (QoE) is closely linked to other contextual parameters, like the user's previous experiences or his/her expectations. This conclusion can be easily reached by looking at table II and comparing the expert and non expert results. Expert users became earlier frustrated with delays occurred during the experiment because of their higher demanding expectations. It is also interesting to observe how unskilled users became more critical in the second experiment.

TABLE II. RESULTS

Experiment	User	Context	Maximum (5)	Minimun (1)	Opinion model
From slow to fast network	Expert	fast network	3,9	0,6	MOS=4.07-2.25 ln (Session Time)
		medium network	9,1	0,6	MOS=4.36-1.52 ln (Session Time)
		slow network	13,1	0,7	MOS=4.40-1.06 ln (Session Time)
	Non Expert	fast network	9,3	0,6	MOS=4.35-1,50 ln (Session Time)
		medium network	13,1	0,7	MOS=4.53-1.37 ln (Session Time)
		slow network	35,6	0,7	MOS=4.68-1.03 ln (Session Time)
From fast to slow network	Expert	fast network	4,9	0,7	MOS=4.39-2.50 ln (Session Time)
		medium network	9,5	0,7	MOS=4.57-1.58 ln (Session Time)
		slow network	14,5	0,8	MOS=4.75-1.02 ln (Session Time)
	Non Expert	fast network	8	0,9	MOS=5.01-1.09 ln (Session Time)
		medium network	9,8	1	MOS=5.11-1.79 ln (Session Time)
		slow network	16,2	1,4	MOS=5.57-1.64 ln (Session Time)

Influence of previous experiences can be observed looking at the different results deriving from unskilled users when coming from a slow to a fast network or vice versa. Under equal conditions, a user's tolerance decreases when previous experience was better. However, this influence has not affected to expert users who are more aware of actual network conditions than of previous experiences.

Further study should be done to study this contextual factor.

Finally, we must remark that correlation obtained between session time and subjective quality was in most of the experiments over 0,95 so the proposed opinion models can certainly be adopted for mapping QoE with NQoS.

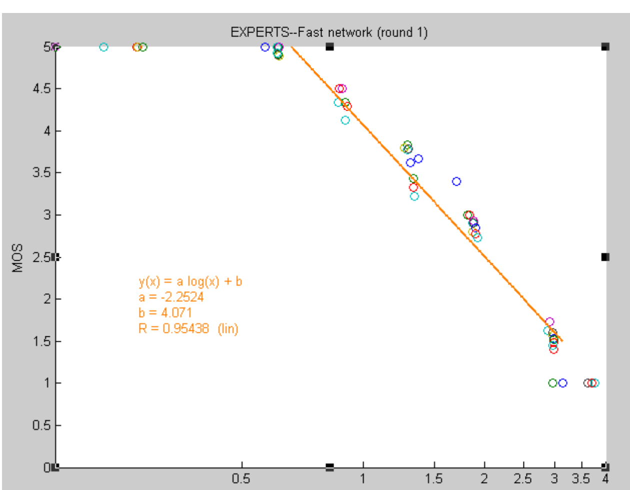

Figure 7. MOS vs. Session Time

VI. CONCLUSIONS

This experiment was designed to evaluate the suitability of our network emulator system for the QoE estimation. This objective has been broadly achieved since it has been used to update ITU-T Rec. G.1030 to today's context. Likewise, the high correlation of the results of the experiment has proved to be really effective.

Further work is needed to integrate new capabilities to the emulation system, like studying QoE in wireless access and NGN networks, where link characteristics are not completely described by bandwidth, delay and losses parameters only.

Nowadays, more QoE experiments are being carried out over voice-over-IP (VoIP) services and new network sceneries.

Part of this work has been carried out in the scope of the European FP7 ICT Adamantium R&D project.

REFERENCES

[1] A. Bouch, M. A. Sasse, and H. DeMeer, "Of packets and people: a user-centered approach to quality of service," IWQOS, Eight International Workshop on Quality of Service, Pittsburgh, PA, USA, 2000.

[2] ITU-T, "E.800: Terms and definitions related to quality of service and network performance including dependability," 1994.

[3] A. v. Moorsel, "Metrics for the Internet Age: Quality of Experience and Quality of Business," 5th Performability Workshop, Erlangen, Germany. , 2001.

[4] M. Siller, "Improving Quality of Experience for Multimedia Services by QoS Arbitration on a QoE Framework," presented at 13th Packed Video Workshop Nantes, 2003.

[5] N. Muhammad, D. Chiavelli, S. David, L. Man, "QoS and QoE Management in UMTS Cellular Systems", John Wiley & Sons, 2006.

[6] P. Reichl, "From 'Quality-of-Service'and 'Quality-of-Design'to 'Quality-of-Experience': A Holistic View on Future Interactive

Telecommunication Services," SoftCOM 2007, 15th International Conference on, 2007.

[7] ITU-T, "P.10/G.100 (2006) Amendment 1 (01/07): New Appendix I - Definition of Quality of Experience (QoE) ", 2007.

[8] ITU-T, "G.1030: Estimating end-to-end performance in IP networks for data applications," 2005.

[9] A. Bouch, A. Kuchinsky, and N. Bhatti, "Quality is in the eye of the beholder: meeting users' requirements for Internet quality," presented at Proceedings of the SIGCHI conference on Human factors in computing systems The Hague, The Netherlands 2000.

[10] N. Bhatti, A. Bouch, and A. Kuchinsky, "Integrating user-perceived quality into Web server design," Computer Networks, vol. 33, pp. 1-16, 2000.

[11] Y. Liu, A. H. Ngu, and L. Z. Zeng, "QoS computation and policing in dynamic web service selection," in Proceedings of the 13th international World Wide Web conference on Alternate track papers, NY, ACM, 2004.

[12] M. A. Sasse, A. Watson, J. Mullin, M. Jackson, A. H. Anderson, L. Smallwood, and G. Wilson, "The ETNA Taxonomy," 2004.

[13] Cnet, http://www.csse.uwa.edu.au/cnet/.

[14] Netem, http://www.linuxfoundation.org/en/Net:Netem.

[15] NISnet, http://snad.ncsl.nist.gov/nistnet/.

[16] VNUML:http://www.dit.upm.es/vnumlwiki/.

[17] W3C, , Web Services Home: http://www.w3.org/.

[18] N. Tim Rahrer, F. Riccardo Fiandra, and B. Steven Wright, "Triple-play Services Quality of Experience (QoE) Requirements," DSL Forum TR-126, 2006 2006.

[19] K. Kilkki, "Quality of Experience in Communications Ecosystem",º Journal of Universal Computer Science, vol. 14, pp. 615-624, 2008.

[20] C. H. Muntean and J. McManis, "The Value of QoE-Based Adaptation Approach in Educational Hypermedia: Empirical Evaluation," Proceedings of the 4th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH2006), pp. 121–130, 2006.

[21] A. S. Patrick, J. Singer, B. Corrie, S. Noel, K. El Khatib, B. Emond, T. Zimmerman, and S. Marsh, "A QoE sensitive architecture for advanced collaborative environments," First International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks, QSHINE , 2004.

[22] C. H. Muntean and J. McManis, "A QoS-aware adaptive Web-based system," 2004 IEEE International Conference on, 2004.

[23] P. Varga, G. Kún, G. Sey, I. Moldován, and P. Gelencsér, "Correlating User Perception and Measurable Network Properties: Experimenting with QoE," in Autonomic Principles of IP Operations and Management, 2006, pp. 218-221.

[24] A. Parasuraman, V. A. Zeithaml, and L. L. Berry, "Reassessment of Expectations as a Comparison Standard in Measuring Service Quality: Implications for Further Research," Journal of Marketing, vol. 58, pp. 111-124, 1994.

[25] ITU-T, "P.800: Methods for subjective determination of transmision quality," 1996.

[26] ITU-T, "G.1010: End-user multimedia QoS categories," 2001.

[27] Y. Chen, T. Farley, and N. Ye, "QoS Requirements of Network Applications on the Internet " Inf. Knowl. Syst. Manag., vol. 4, pp. 55-76, 2004..