A Stimulus-Response Mechanism for Charging Enhanced Quality-of-User Experience in Next Generation All-IP Networks

Reichl, Peter ; Kurtansky, Pascal ; Fabini, Joachim ; Stiller, Burkhard

A Stimulus-Response Mechanism for Charging Enhanced Quality-of-User Experience in Next Generation All-IP Networks

Peter Reichl •
Telecommunications Research Center Vienna (ftw.), Austria

Joachim Fabini
Institute for Broadband Communication, University of Technology Vienna, Austria

Pascal Kurtansky
Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland

Burkhard Stiller
Department of Informatics (IFI), University of Zurich, Switzerland

Abstract

Traditional Internet pricing focuses on the link between Quality-of-Service (QoS) provided by the network and corresponding charges paid by the user who thus reacts to the current network status. However, the key success factor for future packet-switched networks is not network-centric QoS, but rather the service quality as perceived by the user (Quality-of-User Experience, QoE). With respect to pricing, this implies a fundamental change of perspective. This paper discusses a promising model for QoE-based charging which requires direct user feedback as a joint signal of current perceived quality and willingness-to-pay and explores technical requirements and options for realizing this approach in the framework of Next Generation All-IP network architectures.

Keywords: Internet Economics, pricing and charging, Quality-of-Experience, utility function, IP Multimedia Subsystem (IMS), All-IP architecture

1. INTRODUCTION

The traditional model of the Internet as a packet-based best-effort communication system providing a plethora of services basically for free is expected to undergo substantial changes within the near future. For this transition, Quality-of-Service (QoS) is still one of the central topics under discussion, especially since the current evolution towards convergence of fixed and mobile networks based on all-IP framework architectures like the IP Multimedia Subsystem (IMS) has renewed the research interest in this issue. On the other hand, it has been argued since a long time that there is no use in providing differentiated service quality levels without price discrimination. Therefore, due to this intimate link, also the question of how to charge for QoS-enabled Internet services is experiencing growing interest in the

• Address: Forschungszentrum Telekommunikation Wien (ftw.), Donaucitystr. 1, A-1220 Vienna, Austria.
E-mails: reichl@ftw.at, joachim.fabini@tuwien.ac.at; kurtansky@tik.ee.ethz.ch, stiller@ifi.unizh.ch.

community, which eventually has led to the establishment of "Internet Economics" as a new interdisciplinary research are investigating communication networks from an economic rather than a technical perspective.

In parallel, we observe a paradigm shift concerning the concept of QoS itself. So far, QoS has usually been described in terms of parameters like packet loss rate, link band-width, delay, delay jitter etc., which are observed directly in the network. More recently, however, there has appeared a second notion of "service quality" which focuses on the quality experienced by the end user ("Quality-of-Experience", QoE) rather than on pure engineering parameters. QoE is mainly determined through user trials and expressed in terms of scalar "Mean Opinion Scores" (MOS, usually on a scale between 1 = "bad" and 5 = "excellent"), and a variety of measurement and estimation techniques for QoE have been developed and standardized.

In this paper, we discuss some of the consequences of this paradigm shift and propose a QoE-based charging mechanism inspired by the concept of "stimulus and response" which is one of the cornerstones of modern psychology. The idea is to give the customer the opportunity for direct feedback to the network (stimulus) if two conditions are fulfilled: (a) the perceived (best-effort) quality is becoming insufficient, and (b) the customer is willing to pay for a short time an additional fee for improved quality. The response of the network is of course to deliver satisfying quality.

2. RELATED WORK

In parallel to the intense discussion about introducing QoS-enabled Internet architectures like IntServ or DiffServ, over the last decade a huge variety of interesting concepts and proposals for charging QoS-enabled Internet services has been developed, including Edge Pricing, Expected Capacity Pricing, Congestion Pricing e.g. through auction mechanisms (Smart Market, Progressive Second-Price auctions, the Multilink Dutch Auction Scheme, second-chance auctions and multibid auctions), ECN Pricing, Resource Pricing, Proportionally Fair Pricing, Expected Bandwidth Pricing, the Cumulus Pricing Scheme, the Contract and Balancing Process, Paris Metro Pricing etc., to name but a few. For further details we refer e.g. to [1] – [3] and references therein, whereas [4] provides a comprehensive introduction into Internet Economics in general.

Quality-of-Experience (QoE) has been defined first in [5] as "a measure of the overall acceptability of an application or service, as perceived subjectively by the end-user". As traditional subjective tests via extensive user trials are time-consuming and expensive, significant research has been performed with respect to instrumental measurement algorithms. In our context, single-sided speech quality assessment methods like "3SQM" [6] or the non-intrusive E-Model (NIEM, [7]) are especially relevant, but also ideas from related approaches like PESQ [8] or PSQA [9] deserve attention.

The bridge between QoE and charging has been illuminated first from the perspective of charging interface design [10]. In [11], a more general discussion in the context of perceptual speech quality is provided and a distinction of QoE-based charging into *instrumental* (i.e. an instrumental algorithm is used to evaluate QoE, e.g. in the form of utility functions) and *reactive* (i.e. feedback about the actual QoE is delivered directly and in real time by the end user) mechanisms is proposed. In this sense, [12] analyzes an instrumental charging model using neural networks for deriving appropriate user utility functions, whereas the present paper deals with a first example of a reactive scheme and describes its prototype implementation in the context of an IMS (IP Multimedia Subsystem) testbed.

3. A STIMULUS-RESPONSE CHARGING MECHANISM

The concept of “stimulus-response” is a very popular psychological model which has been developed in the first half of last century by leading representatives of the so-called “behaviorism”, most notably J. B. Watson and B. F. Skinner. Essentially, behaviorism claims that all types of behavior may be decomposed into pairs of stimulus and response, and the only way for an organism to learn about and interact with its environment is to receive such stimuli and respond to them accordingly and in real-time.

Translating these ideas into the context of QoE-based charging we start with assuming an IP-based network which offers two classes of service quality, i.e. traditional best-effort, and some sort of enhanced QoS, e.g. in the sense of priority service, additional bandwidth or the like. Additionally, a real-time feedback channel is required which allows end users to express their dissatisfaction with the service quality they are currently experiencing (as a straight-forward example for mobile networks, one could imagine some sort of a “red button” at the mobile handset which can be pressed if the perceived quality becomes poor).

Of course, the quality differentiation between the two QoS classes has to be reflected in the pricing scheme: here, we assume that best effort service is for free (or for some flat rate), whereas enhanced service quality is related to some additional charge. In order to keep this transparent, we propose that enhanced service quality is provided only for a limited time period (e.g. one or two minutes) for which the user is charged a fixed additional amount. At the end of this period, the user can choose between returning to the best effort class or re-pressing the red button in order to extend the enhanced service quality for another period.

In [11], we have modeled this situation as an interplay between user’s perceived quality (QoE) and willingness-to-pay (WTP), see Fig. 1. If the perceived quality is sufficiently high, existing WTP is irrelevant, as e.g. during t_1-t_2 and t_5-t_6. On the other hand, unsatisfying QoE without WTP does neither lead to a reaction of the network, see t_3-t_4 and t_7-t_8. Only after t_9, low QoE and high WTP together trigger a stimulus signal, and as immediate response, the network enhances service quality (dotted line) above the threshold. As the user is happy with the result, she sends another stimulus at t_{10}, because otherwise the network would automatically drop back to the best-effort class.

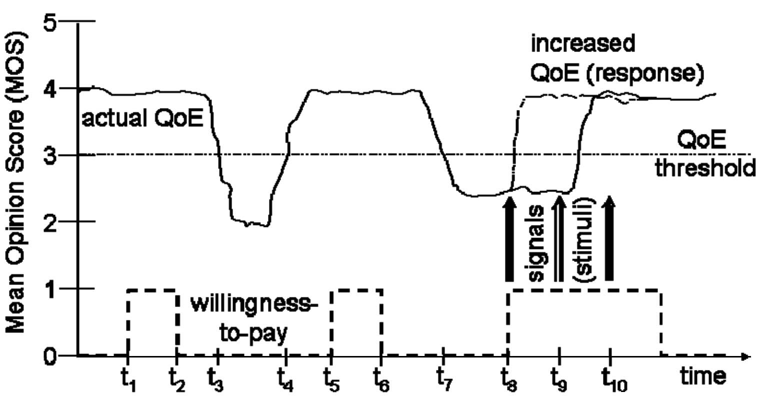

Fig. 1: Quality-of-Experience, Willingness-to-Pay and Stimulus-Response Signals

In the next section, we discuss a prototypical implementation of this approach which is performed in the framework of a 3GPP IP Multimedia Subsystem (IMS) testbed as developed at the Telecommunications Research Center Vienna (ftw.) within the projects CAMPARI/CAIPIRINA.

4. PROTOTYPICAL IMPLEMENTATION OF STIMULUS-RESPONSE CHARGING IN THE CONTEXT OF AN IP MULTIMEDIA SUBSYSTEM TESTBED

The 3GPP IP Multimedia Subsystem (IMS) is currently the most promising candidate for providing the basic architecture framework for Next Generation Networks which will bridge the traditional gap between circuit-switched and packet-switched networks and consolidate both sides into one single network for all services [13]. Within the projects N3-CAMPARI and N9-CAIPIRINA, an open-source based IMS testbed [14] has been realized at the Telecommunications Research Center Vienna as sketched in Figure 2. Here, the non-IMS components WE1 and WE2 are two WAN emulators which have been configured as bridges and – at least from an external view – are fully transparent to the traffic above and including OSI layer 3 (IP). These two nodes can – either deterministically or based on statistical distributions – generate QoS impairments like delay, jitter, loss, reordering, etc., and selectively tamper bypassing traffic based on these parameters.

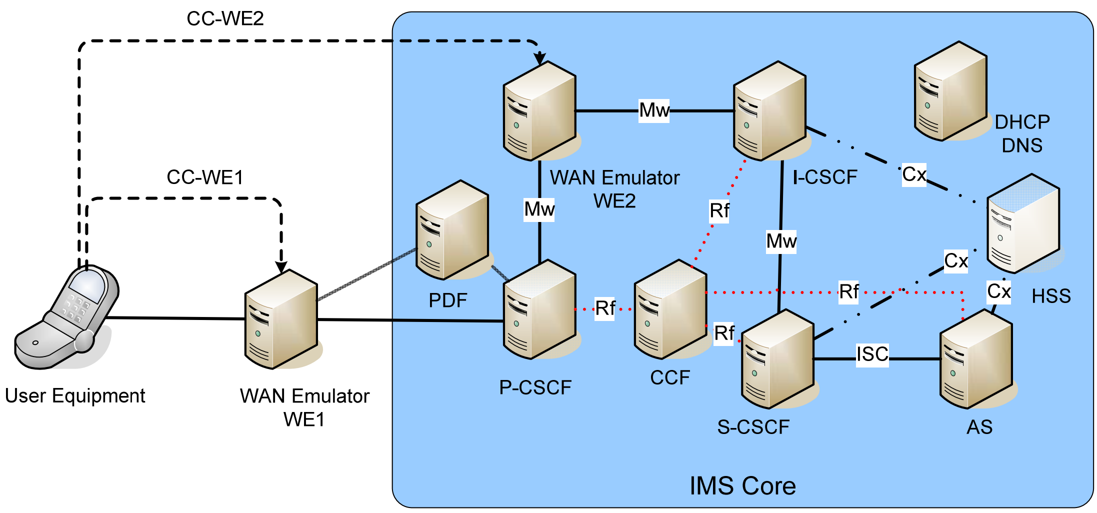

Figure 2: IMS Testbed Components and Setup

The primary use of WE1 and WE2 concerns the simulation of the two main IMS topologies, i.e., the Visited GGSN topology and the Home GGSN topology just by means of software-level WAN emulator reconfiguration, without any change in testbed wiring. For the prototypical realization of the stimulus-response pricing, two control connections, CC-WE1 between the User Equipment UE and the WAN emulator WE1, and CC-WE2 between the UE and WE2, are added. The block diagram in Figure 3 shows the layering of the QoE control. Proprietary SSH- and non-SSH control connections are displayed using dashed lines, IMS signaling connections use solid lines.

WE1 and WE2 have been assigned IP addresses and both can be contacted using SSH-based connections. Within the User Equipment, which, for development purposes, is a laptop computer running SuSE Linux 10.1, a QoS client application with an IMS-enabled KPhone implementation is integrated. This QoS client communicates using SSH with QoS server applications running on the WAN emulators WE1 and WE2 which, on their turn, control the NetEm Linux kernel module responsible for tampering the bypassing traffic $F1(x)$ and $F2(x)$, respectively. Note that the tampering function $F(x)$ is composed of two distinct functions, $F_u(x)$ and $F_d(x)$, which are separately configurable and are able to impact the traffic in the uplink and in the downlink direction, respectively. Thus, the testbed can impair the QoS parameters in the uplink, in the downlink, or in both directions.

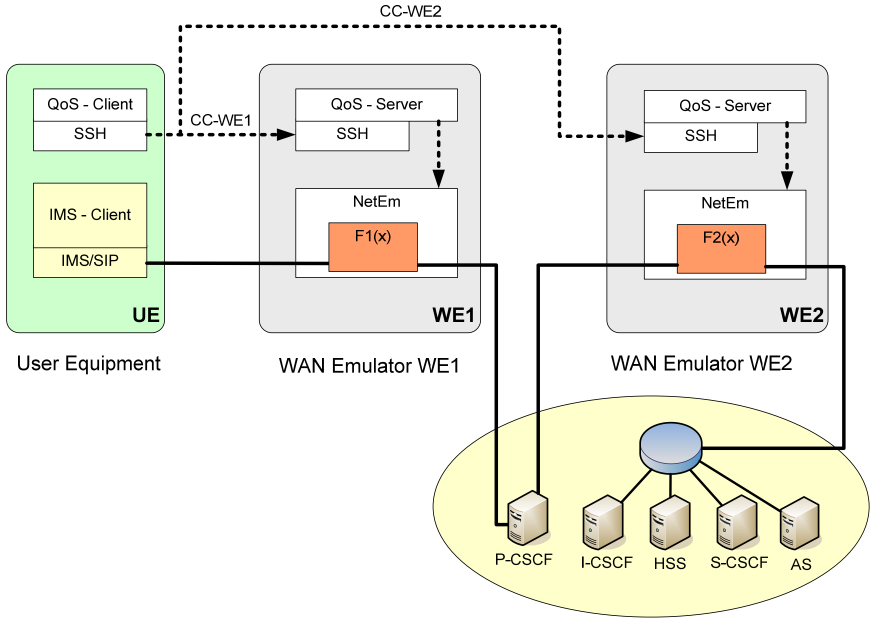

Figure 3: IMS Prototype QoE Control Layering

By default, the QoS servers in WE1 and WE2 generate statistically distributed, periodical QoS impairments. According to the stimulus-response mechanism described in section 3, the end user can react to this impairment by pressing a button in his Kphone VoIP application which is executed on the User Equipment UE, and the testbed then is able to provide enhanced QoS for a limited period of time as required.

The next section discusses signaling issues relevant for QoE-based charging from the complementary perspective of another QoS-enabled all-IP architecture which has been developed in the European project DAIDALOS [15].

5. STIMULUS-RESPONSE CHARGING IN THE DAIDALOS ARCHITECTURE

Figure 4 depicts the overall DAIDALOS architecture, presenting two administrative domains (light and dark grey network clouds) interconnected through edge routers. Within an administrative domain, access networks (AN) are connected through core routers to a core network; in ANs, mobile terminals (MT) are connected by access routers (AR). Providing QoS in ANs requires additional and advanced functions summarized as the Advanced Router Mechanisms (ARM). On every MT, a QoS client is installed and allows the MT to request QoS resources for delivering QoS-enabled services. On the network side, a novel element, the so-called QoS Broker, is present at least once in every AN (several QoS Brokers can be used for load balancing), and performs admission control, manages network resources, and controls the AR according to the active sessions and their QoS requirements.

Advanced IP-based services are provided through the Service Provision Platform (SPP), in the core network, and its proxy instance – the Multi Media Service Platform Proxy (MMSP Proxy) – in the AN. The service QoS definitions are provided by the Policy Based Network Management System (PBNMS), which forwards them to the QoS Brokers, which in turn proxy them to the AR. Another important network element within in the SPP is the so-called A4C server responsible for Authentication, Authorization, Accounting, Auditing and Charging (A4C). Every administrative domain contains at least one A4C server, whereas several A4C servers can be used for load balancing. The QoS Broker within the SPP handles aggregated flows traversing the core network to other administrative domains, for instance for roaming customers. Providing QoS also depends on the actual load of an AN and/or the free resources. Therefore, DAIDALOS uses a real time network monitoring system, which consists of Network Monitoring Entities (NME) located in several points of the network, and a Central Monitoring System (CMS) located in the core network. The NMEs can perform passive and active probing of the network by measuring the QoS characteristics of the data packets belonging to specific flows. The CMS controls the monitoring process, processes the measurements, and propagates the measurements to the QoS Brokers in the network and other entities. These measurement results are used by the QoS Brokers to manage the flows and resources in the access networks as well as for the management of aggregates and resources in the core networks.

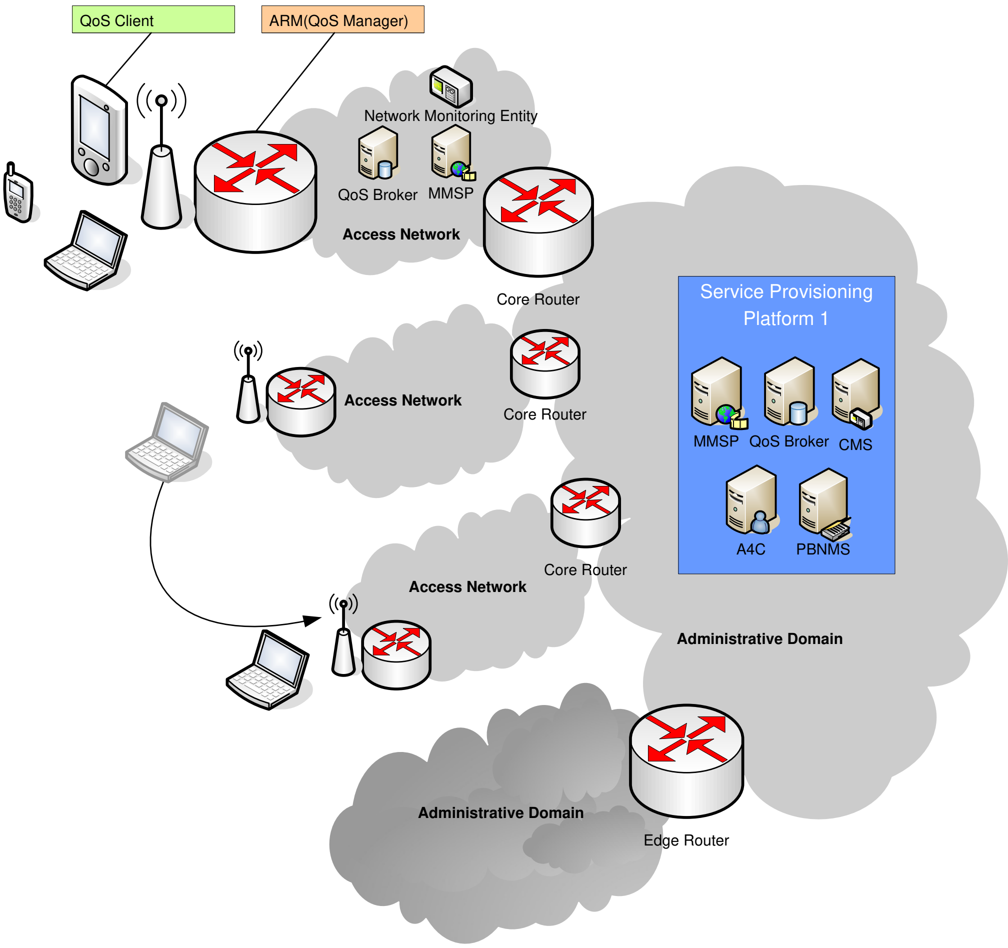

Figure 4: DAIDALOS Architecture

Realizing QoE-based charging with the above described DAIDALOS all-IP architecture is achieved by using the QoS-related network elements (QoS client, AR, QoS Broker and MMSP). In the remainder of this section, we discuss the signaling related to a customer's request for more QoS (see Figure 5).

Assume that a multimedia session with standard QoS has already been established between MT 1 and MT 2, and consider the scenario of Figure 1 where the (network) QoS is decreasing and the corresponding QoE falls below the acceptance threshold. As soon as the user with MT 1 requests more QoS (i.e. at time t_8), the corresponding application contacts the AR 1 (*QoS Info Request*) in order to ask QoS Broker 1 if the network has more available resources. If this is the case, QoS Broker 1 answers with a confirmation message, containing a.o. the supported services (*QoS Info Dec & Profile*). Then, the application on MT 1 issues an *App_Sig Initiation* message containing (1) the DSCP, (2) type of service, (3) possible supported QoS configurations of the MT, and (4) the mapping of applications to network services and QoS parameters. The SPP proxy (MMSP 1) performs service level authorization and filters the QoS configurations accordingly and contacts the home SPP proxy (HoMMSP 2) of MT 2's AN. Then, the *App_Sig Initiation* message is forwarded by the HoMMSP 2 to the MT 2. The callee receives the message, matches the set of QoS configurations supported by the sender with its own and maps them to network services and QoS parameters. Based on the mapping, the MT 2 issues a *QoS Req* message to the local QoS Broker 2. As only the caller is charged for enhanced QoS, the callee can easily accept the QoS change. Thus, if enough QoS resources are available, a successful *QoS Dec* message is sent to the MT 2. The MT 2 sends then an *App_Sig Reply* message with, (1) DSCP, (2) type of service, (3) possible supported QoS configurations in both terminals and the callee network, (4) its care-of address, and, (5) the mapping of applications to network services and QoS. Note that the MMSP 2 also performs service authorization and configuration filtering when receiving this message. The MMSP 2 forwards the *App_Sig Reply* message to the MMSP 1, which forwards it to the MT 1. Upon reception of the *App_Sig Reply* message, the MT 1 issues a *QoS Req* message to the local QoS Broker 1, taking into account the set of common configurations. At this point it is assumed that the QoS resources are still available, thus QoS Broker 1 sends a successful *QoS Dec* message back to the MT 1. Finally, the MT 1 sends an acknowledgement of the session initiation (*App_Sig Ack*) to the MT 2. If this configuration differs from the preferred one in the *App_Sig Reply*, the MT 2 sends a report to the local QoS Broker 2 via the AR 2. Finally, data with the enhanced QoS can flow between the two end-points.

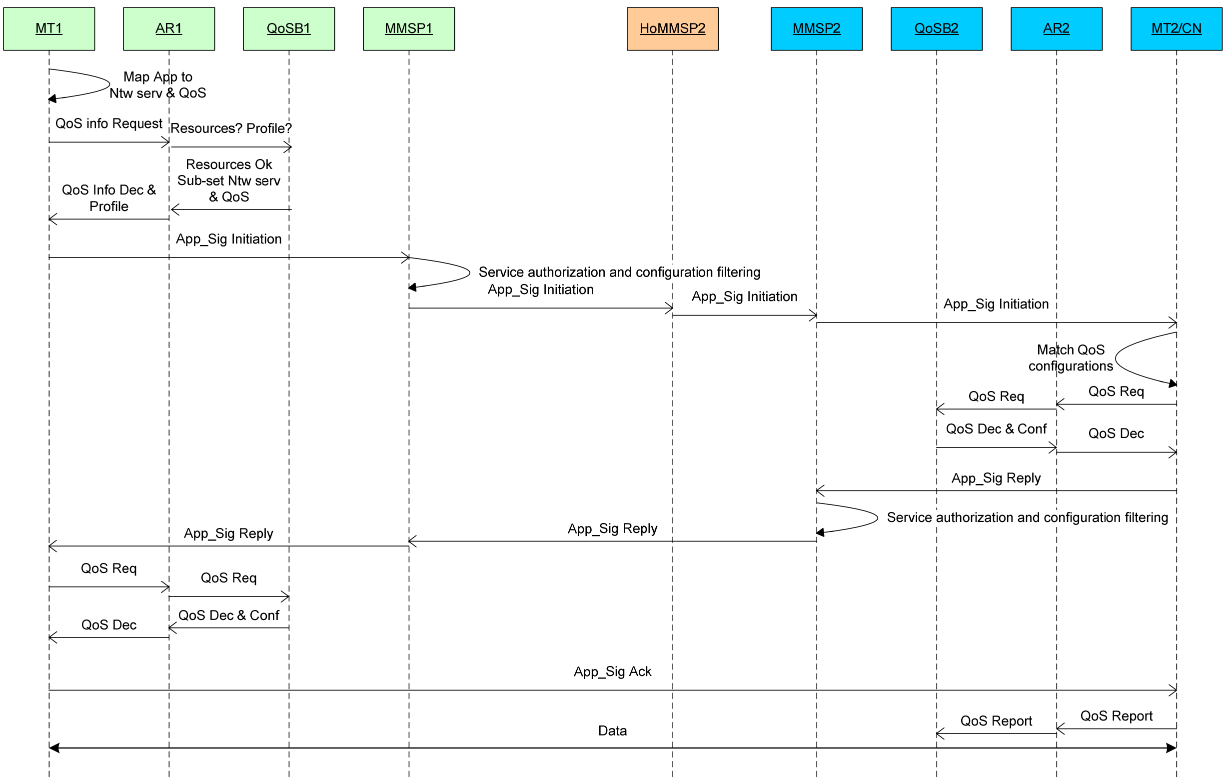

Figure 5: Signaling Flow of a Mobile Terminal requesting QoS in the DAIDALOS Architecture

6. SUMMARY

This paper has discussed the newly arising paradigm of QoE-based charging from a practical perspective. After a brief review of related work, the stimulus-response mechanism has been introduced. Then we have presented a prototypical implementation performed in the context of an IMS testbed, before related signaling issues in the generic All-IP architecture of the DAIDALOS project have been discussed in some detail.

7. ACKNOWLEDGEMENTS

This work has been partially performed in the framework of the Austrian government's Kplus competence center program, and partially in the framework of the EU IST project DAIDALOS II "Designing Advanced Interfaces for the Delivery and Administration of Location-independent Optimized Personal Services" (FP6-2004-IST-4-026943) with the University of Zurich as a Third Party Expert.

8. REFERENCES

[1] M. Falkner, M. Devetsikiotis, I. Lambadaris, *An Overview of Pricing Concepts for Broadband IP Networks*. IEEE Communications Survey, 2^{nd} Q 2000, pp. 2-13.

[2] B. Stiller, P. Reichl, S. Leinen, *Pricing and Cost Recovery for Internet Services: Practical Review, Classification and Application of Relevant Models*, NETNOMICS, Baltzer, Vol. 3, No. 1, March 2001.

[3] B. Tuffin, *Charging the Internet without bandwidth reservation: an overview and bibliography of mathematical approaches*, Journal of Information Science and Engineering, 19(5):765-786, 2003.

[4] C. Courcoubetis and R. Weber, *Pricing Communication Networks: Economics, Technology, and Modelling*. Wiley, March 2003.

[5] International Telecommunication Union, *Definition of Quality of Experience*, ITU-T Delayed Contribution D.197, Source: Nortel Networks, Canada (P. Coverdale), 2004.

[6] International Telecommunication Union, *Single-ended method for objective speech quality assessment in narrow-band telephony applications*, ITU-T Recommendation P.563, May 2004.

[7] International Telecommunication Union, *The E-model, a computational model for use in transmission planning*, ITU-T Recommendation G.107, March 2005.

[8] International Telecommunication Union, *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, ITU-T Rec. P.862, Feb. 2001.

[9] G. Rubino, Q*uantifying the Quality of Audio and Video Transmissions over the Internet: the PSQA Approach*, in: J. Barria (ed.), Design and Operations of Communication Networks: A Review of Wired and Wireless Modelling and Management Challenges, Imperial College Press, 2005.

[10] A. Bouch, M. A. Sasse, *It ain't what you charge, it's the way that you do it: A user perspective of network QoS and pricing*, Proceedings IM'99, Boston, MA, May 1999, pp 639-655.

[11] P. Reichl, F. Hammer, *Charging for Quality-of-Experience: A New Paradigm for Pricing IP-based Services*, Proceedings of the Second ISCA Workshop, Berlin, Germany, September 2006.

[12] Y. Hayel, G. Rubino, B. Tuffin and M. Varela, *A New Way of Thinking Utility in Pricing Mechanisms: A Neural Network Approach*, Proceedings of the 13th CLAIO (Congreso Latino-Iberoamericano de Investigacion Operativa), Montevideo, Uruguay, 2006.

[13] G. Camarillo and M. Garcia-Martin, *The 3G IP Multimedia Subsystem (IMS)*, Wiley 2006.

[14] P. Reichl, S. Bessler, J. Fabini et al., *Practical Experiences with an IMS-Aware Location Service Enabler on Top of an Experimental Open Source IMS Core Implementation*, Journal of Mobile Multimedia, vol. 2 no. 3, pp. 189 – 224, 2006.

[15] DAIDALOS, an EU Framework Programme 6 Integrated Project, http://www.ist-daidalos.org, September 2006.