Methods for Multimedia Service Adaptation in Next Generation Networks

Blazej Lewcio, Sebastian Möller, Marcel Wältermann, Niklas Kirschnick, and Pablo Vidales,
Deutsche Telekom Laboratories, Berlin, Germany.
Email: {*Blazej.Lewcio, Sebastian.Moeller, Marcel.Waeltermann, Niklas.Kirschnik, Pablo.Vidales*}(at)telekom.de

***Abstract*—In Next Generation Networks (NGNs), vertical handovers between different wireless access technologies provide seamless roaming during active multimedia sessions. The quality of these multimedia sessions, as well as the quality of user experience, depends on the network handover policies, link layer characteristics, and codecs used in the respective networks. Even though extensive research has been carried out on seamless mobility, a thorough analysis of user perception of this aspect is still missing. Knowledge of user perception is however necessary in order to successfully design and further improve mobility management solutions for always-on multimedia services. We present an NGN testbed and an approach for mapping user experience to network conditions, focusing on phenomena caused by a user roaming across diverse wireless technologies. Current experiments address the quality of Voice over IP in Next Generation Networks, as an example of multimedia sessions. These define when a network handover should be scheduled in order to reach optimum quality for the user.**

I. Introduction

Next Generation Networks (NGNs) will provide ubiquitous access to multimedia services available on the Internet. Internet access will be given to nomadic users by diverse wireless transmission technologies that will be connected together to create an All-IP platform. The network heterogeneity and convergence will decouple the service usage from the end-device's point of attachment to the network. Users on the move will be allowed to seamlessly use multimedia services, like voice or video over IP. However, the roaming between diverse wireless networks will have an impact on the transmission quality. Therefore, a tradeoff between network coverage and transmission quality has to be found to improve the experience of multimedia users. It is an open challenge to satisfy user expectations and adapt service performance to variable network conditions as the ones occurring in NGNs. Even though speech quality can be estimated for VoIP services [20], new phenomena resulting from mobility in heterogeneous environments have not been investigated — or even understood.

Roaming among heterogeneous networks during an active multimedia session can rapidly change the streaming quality. Due to changing link layer characteristics and the network handover itself, the quality of real-time multimedia services can drastically degrade or the employed codec must be changed. Varying network bandwidths and transmission delays (e.g. HSDPA vs. WiFi), audio bandwidth changes between wideband (50-7000 Hz) and narrowband (300-3400 Hz) codecs, and the temporal position of variations within a call, are factors that potentially affect user perception. Until now, most of the research efforts have been focused on the reduction of network handover latency and packet loss, while neglecting the user perception of mobility phenomena.

In order to optimize user experience, a relationship has to be established between network characteristics and user perception. A controlled setting is necessary for this purpose, so that the impact on quality of individual parameters like network load, codec type and bandwidth, and the scheduling of network handovers and codec changeovers, can be assessed. Therefore, an NGN testbed has been set up in the frame of the Mobisense project at Deutsche Telekom Laboratories. Using this testbed, subjective experiments have been carried out quantifying the perceptual effects of each network parameter, so that algorithmic models can be derived for quality prediction at a later stage.

The paper is organized as follows. The next section summarizes the most relevant related work. The Mobisense experimental setup is presented in Section III by listing the software and hardware components, their interaction, and the main features of the testbed. Section IV describes the research methodology including the subjective test design, sample processing, evaluation methods, and some preliminary results of the listening-only test. Section V summarizes the main findings and provides a plan for future work.

II. Related work

The work on service adaptation in NGNs is distributed over several research areas: (1) beginning in the networking layer and mobility protocols, (2) going through the application layer and the use of various encoding techniques, and (3) ending by the assessment of user perception.

Existing mobility protocols like Mobile IP [19], SIP [23], and HIP [18], enable end-devices to change the attachment point to the Internet during an ongoing communication. Considerable work has been done to optimize this procedure. In [25], Vidales et al. propose different improvements to the Mobile IP solution, targeting issues such as the reduction of handover latencies or packet loss. The impact of frequent vertical handovers on transmission quality has been investigated in [5]. The authors show that frequent handovers to the occasionally best possible network (here, from GPRS to WiFi) has a major impact for real-time services. Nevertheless, the quantitative results presented in the above papers are not complemented by the analysis of user perception of these mobility phenomena.

In [9], the implementation of a testbed to emulate a wireless Internet access provider is presented. The main features of this testbed are multimedia call setup using SIP signaling, seamless mobility features, fast handovers, and QoS features for mobile clients. However, the relation between network characteristics and user perception is not addressed in this work.

Perceived speech quality can be assessed in two different ways: using auditory quality assessment or instrumental quality prediction [20]. While auditory tests provide real quantitative *measurements* of perceived quality, e.g. in terms of an average overall quality judgement (Mean Opinion Score, MOS) [13], instrumental methods are limited to *predict*, i.e. estimate such quality judgements. Since auditory tests are time-consuming and expensive, the usage of instrumental quality prediction models is in some cases preferable. Quality prediction can be based on a speech signal analysis (as in the PESQ model [14]), on network parameters (as in the E-Model [11]), or on protocol information like header data (like presented in [12]). However, none of the mentioned instrumental methods is yet able to handle mobility events, and so to provide reliable quality estimates in heterogeneous NGN environments.

Most models estimate quality for a limited period of time (e.g. an 8-second long sample used by signal-based models), or for an entire call with constant network characteristics. In NGNs, however, quality changes over time. The temporal position of a mobility event (such as a network handover or a codec changeover) within a call has shown to affect perceived quality. For example, degradations occuring at the end of a call have a stronger impact on the final quality rating than degradations occurring at the beginning of a call (so-called "recency effect" [8]). Such effects are not yet taken into account by the mentioned models. In addition, changes between narrowband and wideband codecs are to be encountered in NGNs; such effects are not explored yet, neither in the auditory nor in the instrumental domain.

There are already a few seminal research papers in the direction of quality assessment in NGNs. In [21], for example, Rajendran et al. introduce *OverPhone*, which is a VoIP overlay network evaluating the quality of networks simulated using PlanetLab [7]. The OverPhone network uses the E-model to make routing decisions. The authors conclude that the quality of VoIP calls can be increased by estimating quality to derive routing decisions. As the E-Model is currently limited to single network types, first it has to be extended on the basis of auditory tests in order to derive routing decisions for wireless and heterogeneous networks. Such auditory tests are part of the research described in this paper.

The User Satisfaction Index (USI) is developed in [6] to quantify the VoIP user satisfaction. The model behind the USI is derived from characteristics of Skype [2], such as call duration, bit rate, jitter, and round trip delay. In [6], the USI is validated by an independent set of metrics for speech interactivity derived from user conversation patterns. The impact of the bit rate, jitter, packet loss, and delays is evaluated to provide hints on the importance of these network parameters. Even though the networking analysis is complemented by an analysis of user interaction, the results are not verified by any auditory tests.

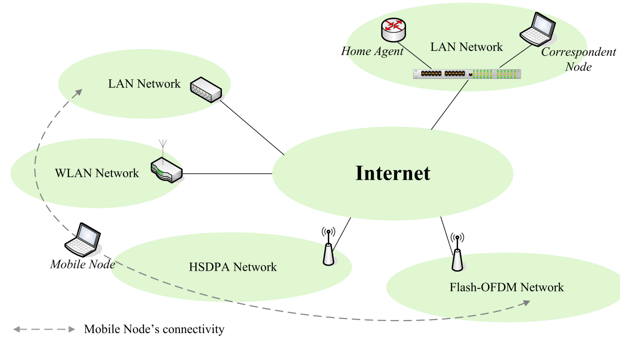

Fig. 1. Overview of the Mobisense testbed.

In [10], Grönvall and Marsh implement and evaluate different WiFi/GSM network handover policies based on signal strength, packet loss, and jitter. The users are asked to follow a predefined walking route while performing a speech call using a handheld device. Although the approach is quite realistic, the experiments do not provide completely controlled network conditions and do not address audio bandwidth changes. In contrast to their experiments, our research provides a reliable evaluation procedure and takes into account changes between narrowband and wideband codecs.

III. NGN TESTBED

Figure 1 illustrates the experimental setup designed to analyze the impact of seamless mobility in heterogeneous networks on running multimedia applications, such as VoIP. The Mobisense testbed supports connectivity to five networks attached to the Internet that are based on four different data transmission technologies: two LAN networks, one WiFi network, one HSDPA network, and one Flash-OFDM network. All of them employ an IPv4 protocol stack, so that Mobile IPv4 provides transparent mobility management to overlaying protocols. The deployed Mobile IPv4 infrastructure consists of a Home Agent (HA), a Correspondent Node (CN), and a Mobile Node (MN) and does not involve additional Foreign Agents. Therefore, the MN individually obtains IP addresses in particular networks and acts directly as one end of the Mobile IPv4 tunnel.

The mobility of the MN is ensured by a commercial version of Mobile IPv4 Secgo client 3.2 [1]. Two configuration features were deliberately activated. The first one is the "make-before breake" handover policy (for smooth handover) that allows the MN to disconnect from the old network after the connection to the new network has been established. The second activated feature is the reverse tunneling option that forces the Home Agent to provide a bidirectional tunnel for the communication between CN and MN, avoiding NAT problems.

The VoIP infrastructure integrated in the testbed relies on SIP and RTP protocols on top of the UDP protocol. The SIP protocol specifies call control methods, such as call initiation, modification, and termination. Because SIP is used for signaling only [23], the speech data is conveyed by the RTP protocol [24]. The VoIP framework is based on the PJSIP 0.5.10.3 software [4] that has been extended in the following way to satisfy the project needs:

- Codec changeover mechanisms have been implemented to enable switching between various codecs (including wideband-narrowband transitions) during an ongoing call.
- Jitter buffer monitoring has been added to observe how the application buffers data.

In order to allow codec changeover, three solutions have been developed and evaluated [15], [27]. These solutions are based on the SIP/SDP parameter renegotiation [22], a parallel media stream establishment, and internal RTP packet routing in the application. Altogether, these functions enable the VoIP application to avoid potential packet loss that arise from the codec changeover process itself. In this way, the application-layer impact on speech quality can be reduced.

TABLE I
HARDWARE AND SOFTWARE ELEMENTS OF THE MOBISENSE TESTBED.

Entity	Hardware	Software
Home Agent	Cisco 2620XM	IOS 12.2(8r)
Mobile Node	IBM T60 HSDPA Data card Flarion Data card IEEE 802.11 a/b/g	OpenSuse 10.2 PJPROJECT 0.5.10.3 SecGo Mobile IP 3.2 NAT-traversal module TCP server and client UDP sender tcpdump netem
Correspondent Node	IBM T60	OpenSuse 10.2 PJPROJECT 0.5.10.3 TCP client tcpdump

Due to a large number of the involved software elements (listed in Table I), the operational complexity of carrying out a test has been reduced by centralizing the testbed control on the Mobile Node [26].

In sum, the testbed provides the following set of features:

- Access to diverse network technologies (WiFi, HSDPA, Flash-OFDM, LAN).
- Performing of vertical and horizontal network handovers.
- Support of a large number of speech codecs (e.g. G.711, G.722, G.722.2, GSM-FR).
- Triggering of wideband-narrowband codec changeovers.
- Manipulation of network conditions such as injection of jitter, packet loss, or delay.
- Recording of speech samples.
- Logging of network traces.
- Monitoring of VoIP application parameters (e.g. jitter buffer).
- Modularity and possibility of extension.

Various mobility events from the above list have been generated at different temporal positions within a call to generate speech samples for an auditory evaluation. In this way, the perceptual effects of mobility events can be studied in detail, and adaptation rules for always-on speech services can be developed. These experiments are described in the following section.

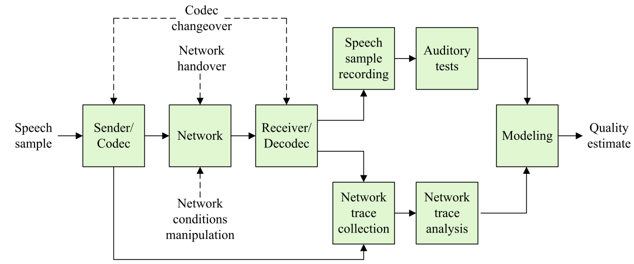

Fig. 2. A twofold evaluation approach of speech quality in NGNs.

IV. EVALUATION OF VOIP QUALITY

The presented study follows a twofold evaluation approach in order to link NGN characteristics to user perception. This approach is presented in Figure 2. It is based on speech samples processed under specified experimental conditions, using the Mobisense testbed described in previous section.

According to the twofold evaluation approach, the process is split in two ortogonal stages – perception and networking analysis. On the perception layer, the audio samples are recorded and judged in an auditory experiment to obtain average quality scores (MOSs). On the networking layer, network traces are collected on the involved network stations and used for a subsequent network trace analysis. Finally, the results of both layers are merged to develop a quality prediction model that provides speech quality estimates, including the impact of mobility events in NGNs. Such a model enables a successfull service adaptation and mobility management in real-time.

The speech samples are processed in the Mobisense testbed described in the previous section. Using the testbed, all system characteristics, i.e the applied codec, the MN attachment point to the network, and the other network conditions, can be controlled in real-time to emulate different mobility scenarios.

An example of a mobility pattern could look like this:

> *A mobile user equipped with a dual-mode mobile device initiates a VoIP call at home. He starts using his private WiFi network and enjoys wideband speech quality. However, during the call he decides to leave the house. He moves away from the WiFi access point and loses connectivity. As a result, the dual-mode device is forced to switch to another available wireless network, HSDPA. Along with the network handover, the mobile device switches the speech codec to a narrowband codec and so adapts to the new network conditions.*

The mobility situation can be emulated in the Mobisense testbed in the following way:

> *The Mobile Node starts a VoIP call using the WiFi network interface and employing a wideband codec (ITU-T G.722.2 at 23 kbit/s). In order to reproduce*

the MN leaving a particular WiFi coverage area, 5% packet loss is injected after some time to emulate the signal degradation as the MN moves away from the WiFi access point. Consequently, a network handover to the HSDPA network is triggered and such event is immediately followed by a codec changeover to the narrowband ITU-T G.711 codec. The packet loss injection is deactivated in the new network after MN re-establishes the connection.

During the sample processing, network traces and the generated speech samples are recorded for the auditory and networking analysis.

The perceptual speech quality evaluation focuses on how users perceive the overall call quality, as well as the quality of particular parts of the call. Linking the overall call quality to the quality of individual parts of the call, we can investigate how users remember the quality, i.e. the recency effect.

In order to assess overall sample quality, an auditory test following ITU-T Rec. P.800 has been carried out [13]. A balanced set of speech samples (approximately equal number of good and bad quality samples) has been designed, generated, and rated by a group of representative test users, providing an auditory quality rating (MOS) for each network condition.

The recency effect and the temporal position of a mobility event within a call is addressed in a second – independent – auditory test. According to the specification described in [3], such a test consists of a set of 5 samples (each approximately 6 s in duration) which are presented to the listeners in a meaningful sequence, with pauses after each sample. To produce a telephone call feeling, the listeners are asked to answer a content-related question during the pause after every sample. At the end of the simulated call, the listeners rate the speech quality using a 5-point overall quality scale. In this way, the overall call quality and the impact of a bad sample in a particular moment of a call can be assessed.

Figure 3 and Table II present example results of the call quality test. It is obvious that using a wideband speech codec in a WiFi network throughout the call provides the best speech quality (#1). If the wideband conditions are affected by a certain amount of packet loss, switching to a loss-free narrowband channel does not improve the quality (#10, #5, #6, #7). Similarly, switching from narrowband to wideband is only of benefit if it occurs early in a call (#2, #3, #4). Furthermore, 5% packet loss in the middle of a wideband call is perceptually equal to the case of switching to narrowband for the same period (#8, #10). A detailed evaluation of the listening-only test results can be found in [17].

On the basis of the obtained results (cf. [17]), the following importance order of mobility aspects could be determined:

- Packet loss is the factor degrading quality most.
- Wideband-narrowband codec changeover strongly affects user perception, compared to constantly using either narrowband or wideband channels.
- Network handover (smooth handover) only marginally affects speech quality.

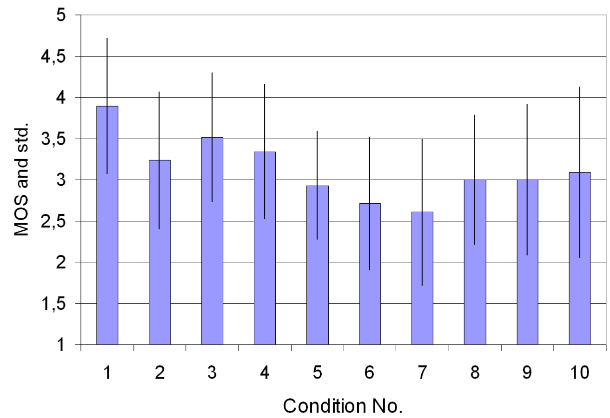

Fig. 3. Example results of the overall call quality test.

TABLE II
CONDITIONS OF THE OVERALL CALL QUALITY TEST. H: HSDPA, W: WIFI; PPL: PACKET LOSS IN %; SWITCHING AT THE BEGINNING (BEG.), MIDDLE (MID.) OR END A CALL.

No.	Network(s)	Codec(s)	Ppl per segment
1	W	G.722.2	0
2	H	G.711	0
3	H→W beg.	G.711→G.722.2	0
4	H→W end	G.711→G.722.2	0
5	W→H mid.	G.722.2→G.711	0,3,3,0,0
6	W→H mid.	G.722.2→G.711	0,5,5,0,0
7	W→H mid.	G.722.2→G.711	0,10,10,0,0
8	W→H→W	G.722.2↔G.711	0
9	H→W→H→W	G.711↔G.722.2	0
10	W	G.722.2	0,0/5,5,5/5,0

The order of importance may change when one of the factors dominates the overall perception; in this case, it may mask the other factors completely.

Currently, the network traces collected during the speech sample generation are analysed. This evaluation is focused on the measurement and positioning of packet transmission anomalies, like variations of Inter Packed Delay (IPD) together with an analysis of IPD density plots, packet loss, and jitter. Furthermore, it is also evaluated how these particularities in the networking layer are handled by the application, especially in the jitter buffer. Finally, it is analyzed whether particular mobility events are located in talkspurts or pauses.

The twofold evaluation approach enables to create a cross-layer quality prediction model that helps to obtain a quality estimate for NGNs. Integration of this speech quality estimates in service adaptation routines are foreseen to improve the Quality of Experience (QoE), rather than relying on the network layer parameters only.

V. CONCLUSIONS AND OUTLOOK

Considering user perception of service quality, the multimedia service adaptation can be improved in terms of its QoE. A detailed quality assessment in NGNs requires a test environment, in which the performance of multimedia services and service adaptation protocols can be measured. The presented NGN testbed fulfils these requirements. It gives a possibility of network and codec switching in real-time, and of simultaneous data recording for subjective experiments. This set of features, as well as the possibility of extension, make the testbed a good tool to evaluate user perception of multimedia service quality in NGNs.

In order to determinate quality-impairing factors, as well as their interaction, a twofold evaluation approach was presented. This approach links network characteristics to user perception, and allows to derive quality prediction models which are useful for defining codec and network switching policies. The generation of arbitrary switching profiles also enables us to measure the impact of the temporal position of a mobility event within a call. The user perception depends on the occurrence and on characteristics of such events, of which importance we were able to sort.

In the future research, a conversational test according to the specifications presented in [16] will be carried out. In such a test, two test participants are asked to perform a telephone call over an NGN connection emulated with our testbed. This nearly real-life situation allows to assess user perception of quality changes, while users concentrate on the information exchange necessary to complete a task, and not just on the surface form of the exchanged speech signals.

The present study was focused on VoIP service quality in NGNs, addressing speech transmission quality as an influencing factor for overall service quality. Transmission quality, in turn, can not yet be fully predicted by existing models. Therefore, we plan to extend quality prediction models like the E-model to include handling of mobility events like network handovers and wideband-narrowband codec changeovers. We also plan to extend our reseach towards real-time video streaming, e.g. in video conferences and mobile TV. For this purpose, the experimental set-up will be extended with video streaming functionality. Similarly, video codec changeovers and network handover are also considered as a possibility of video service adaptation. A future joint speech, audio and video evaluation of quality is necessary to provide high-quality always-on multimedia services and can be used to develop a cross-layer multimedia adaptation protocol for NGNs.

ACKNOWLEDGMENTS

Pablo Vidales is member of of the National System of Researchers (SNI) in Mexico since January 2007. This study was partly supported by the Deutsche Forschungsgemeinschaft (DFG), grant MO1038/5-2.

REFERENCES

[1] Secgo Mobile IP Toolkit. http://www.secgo.com/.

[2] Website of Skype. http://www.skype.com/.

[3] Speech Processing, Transmission and Quality Aspects (STQ); Estimating Speech Quality per Call. ETSI TR 102 506 v.1.1.1, Oct. 2006.

[4] B.Prijono. Open Source SIP Stack and Media Stack for Presence, Instant Messaging, and Multimedia Communication. http://pjsip.org/.

[5] A. Calvagna and G. D. Modica. A user-centric analysis of vertical handovers. pages 137–146, New York, NY, USA, 2004. ACM Press.

[6] K. Chen, C. Huang, P. Huang, and C. Lei. Quantifying Skype User Satisfaction. In *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications(ACM SIGCOMM 2006)*, pages 399–410. ACM Press, Sept. 2006.

[7] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, and M. Bowman. PlanetLab: an Overlay Testbed for Broad-Coverage Services. *ACM SIGCOMM Computer Communication Review*, 33(3):3–12, July 2003.

[8] A. Clark. Modeling the Effects of Burst Packet Loss and Recency on Subjective Voice Quality. In *Proceedings of Internet Telephony Workshop*, New York, Apr. 2001.

[9] A. Dutta, J. Chen, S. Das, M. Elaoud, D. Famolari, S. Madhani, S. McAuley, M. Tauil, S. Baba, T. Maeda, N. Nakajima, Y. Ohba, and H. Schulzrinne. Implementing a Testbed for Mobile Multimedia. In *Proceedings of the IEEE Global Telecommunications Conference 2001 (GLOBECOM 2001)*, volume 3, pages 1944–1949. IEEE Press, Nov. 2001.

[10] B. Grönvall and I. Marsh. Performance evaluation of voice handovers in real 802.11 networks. In *Proceedings of the Second Workshop on Wireless Network Measurements (WiNMee 2006)*, Boston, MA, USA, Apr. 2006.

[11] ITU-T Rec. G.107. *The E-model, a Computational Model for Use in Transmission Planning.* International Telecommunication Union, Mar. 2005.

[12] ITU-T Rec. P.564. *Conformance Testing for narrowband voice over IP transmission quality assessment models.* International Telecommunication Union, Nov. 2007.

[13] ITU-T Rec. P.800. *Methods of Subjective Determination of Transmission Quality.* International Telecommunication Union, Aug. 1996.

[14] ITU-T Rec. P.862. *Perceptual Evaluation of Speech Quality(PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs.* International Telecommunication Union, Feb. 2001.

[15] B. Lewcio. Assessment and Adaptation of Voice Quality in All-IP Heterogeneous Networks. Diploma thesis (unpublished), Aug. 2007.

[16] S. Möller. *Assessment and Prediction of Speech Quality in Telecommunications*. Springer Verlag GmbH, Jun. 2007.

[17] S. Möller, M. Wältermann, B. Lewcio, N. Kirschnick, and P. Vidales. Speech Quality during Handovers between Narrowband and Wideband in Next Generation Networks. (Submitted for publication – Interspeech 2008).

[18] R. Moskowitz and P. Nikander. Host Identity Protocol (HIP) Architecture. RFC 4423 (Informational), May 2006.

[19] C. Perkins. IP Mobility Support for IPv4. RFC 3344 (Proposed Standard), Aug. 2002. Updated by RFC 4721.

[20] A. Raake. *Speech Quality of VoIP: Assessment and Prediction*. John Wiley & Sons, Sep. 2006.

[21] R. K. Rajendran, S. Ganguly, R. Izmailov, and D. Rubenstein. Performance Optimization of VoIP using an Overlay Network. Technical report, NEC America, Sep 2005.

[22] J. Rosenberg and H. Schulzrinne. An Offer/Answer Model with Session Description Protocol (SDP). RFC 3264 (Proposed Standard), Jun. 2002.

[23] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. SIP: Session Initiation Protocol. RFC 3261 (Proposed Standard), Jun. 2002. Updated by RFCs 3265, 3853, 4320, 4916.

[24] H. Schulzrinne and S. Casner. RTP Profile for Audio and Video Conferences with Minimal Control. RFC 3551 (Standard), Jul. 2003.

[25] P. Vidales, C. Bernardos, I. Soto, D. Cottingham, J. Baliosian, and J. Crowcroft. MIPv6 Experimental Evaluation Using Overlay Networks. *Computer Networks*, 51(10):2892–2915, Jul 2007.

[26] P. Vidales, N. Kirschnick, B. Lewcio, F. Steuer, M. Wältermann, and S. Möller. Mobisense testbed: Merging user perception and network performance. In *4th International Conference on Testbeds and Research Infrastructures for the Development of Networks & Communities (TRIDENTCOM 2008)*, Innsbruck, Austria, Mar. 2008.

[27] M. Wältermann, B. Lewcio, P. Vidales, and S. Möller. A Technique for Seamless VoIP-Codec Switching in Next Generation Networks. In *Proceedings of the IEEE International Conference on Communications 2008 (ICC 2008)*, Beijing, China, May 2008.