Measuring Perceptual VoIP Speech Quality over UMTS

Roland Kwitt, Thomas Fichtel, and Thomas Pfeiffenberger

Salzburg Research
Advanced Networking Center
5020 Salzburg, Austria,
{roland.kwitt,thomas.fichtel,thomas.pfeiffenberger}@salzburgresearch.at,
WWW home page: http://www.salzburgresearch.at

Abstract. This paper focuses on VoIP speech quality measurements over a 3G UMTS network, by means of perceptual quality models. We present conducted measurement results and point out that perceptual measurements are a necessity when it comes to evaluating the capability of wireless networks for VoIP transmissions, since they provide measures for the impact of the network on the perceived speech quality.

1 Introduction

Voice over IP (VoIP) is an emerging service in the Internet and VoIP applications are becoming more and more popular in use. However, this imposes new requirements on today's networks, especially regarding Quality of Service (QoS) guarantees. Moreover, the rapid area–wide deployment of VoIP requires the evaluation of potential networks, whether they can provide for the required QoS. Regarding these point, the focus is on the specific demands on a network to provide toll quality service or predefined levels of QoS suitable for VoIP.

However, besides assured delay guarantees and packet losses etc., the most important factor when evaluating networks for VoIP transmissions is the impact of the network on the speech quality perceived by the user. The determination of speech quality is a fundamental aspect when talking about QoS in VoIP environments. It is obvious that speech quality should be measured by humans themselves. Unfortunately, such subjective measurements are quite expensive and extremely time consuming. In order to ease speech quality determination, so called *quality models*, simulating human perception, have been developed. These models can calculate a quality rating based on given network metrics and can thus predict speech quality. Such quality ratings for VoIP transmissions are not only of great importance to customers, but also to providers, since they provide means to compare and audit QoS agreements.

In this paper, we focus on measuring perceptual VoIP speech quality over the Universal Mobile Telecommunications System (UMTS), by means of the Communications Measurement Toolset (CMT) II, which already provides means for perceptual VoIP assessment. The conducted measurements include stationary (non–moving) as well as non–stationary point–to–point scenarios.

The remainder of this paper is organized as follows. In Sect. 2 we briefly review data transmissions over UMTS, followed by an introduction on the CMT II and its measurement capabilities in Sect. 3. Sect. 4 presents the employed perceptual quality model and Sect. 5 discusses the measurement results. Sect. 6 concludes this paper and gives an outlook on further work.

2 Data Transmission in UMTS

Mobile data communication is evolving quickly because of the Internet, Intranet, Laptops, PDAs and increased requirements of workforce mobility. 3G UMTS is supposed to be the commercial convergence of fixed line telephony, mobile, Internet and computer technology. New technologies are required to deliver high speed location and mobile terminal specific content to users.

The UMTS transport network is required to handle high data traffic. A number of factors were considered when selecting a transport protocol: bandwidth efficiency, quality of service, standardisation stability, speech delay sensitivity and the permitted maximum number of concurrent users. In the UMTS network, ATM (Asynchronous Transfer Mode) is defined for the connection between UTRAN and the core network and may also be used within the core network itself [1].

UMTS Architecture

A UMTS network consist of three interacting domains: the Core Network (CN), the UMTS Terrestrial Radio Access Network (UTRAN) and the User Equipment (UE). The main function of the core network is to provide switching, routing and transit for user traffic. The CN is logically divided into the circuit–switched domain (CS) and the packet–switched domain (PS). The CN also contains the databases and network management functions. The basic CN architecture for UMTS is based on the GSM network with GPRS (Release 99). All equipment has to be modified for UMTS operation and services. The UMTS Terrestrial Radio Access Network (UTRAN) provides the air interface access method for User Equipment.

UMTS Services

UMTS offers teleservices (like speech or SMS) and bearer services, which provide the capability for information transfer between access points. It is possible to negotiate and renegotiate the characteristics of a bearer service at session or connection establishment and during an ongoing session or connection. Both, connection–oriented and connection–less services are offered for Point-to-Point and Point-to-Multipoint communication.

Bearer services have different QoS parameters for maximum transfer delay, delay variation and bit error rate. Offered data rate targets are:

– 144 kbits/s satellite and rural outdoor
– 384 kbits/s urban outdoor
– 2048 kbits/s indoor and low range outdoor

UMTS network services have different QoS classes for four types of traffic:

– Conversational class (voice, video telephony, video gaming)
– Streaming class (multimedia, video on demand, webcast)
– Interactive class (web browsing, network gaming, database access)
– Background class (email, SMS, downloading)

UMTS Handover

The following categories of handover (also referred to as handoff) exist: [1]:

Hard Handover means that all the old radio links in the UE are removed before the new radio links are established. Hard handover can be seamless or non–seamless. Seamless hard handover means that the handover is not perceptible to the user. In practice a handover that requires a change of the carrier frequency (inter–frequency handover) is always performed as hard handover.

Soft Handover means that the radio links are added and removed in a way that the UE always keeps at least one radio link to the UTRAN. Soft handover is performed by means of macro diversity, which refers to the condition that several radio links are active at the same time. Normally soft handover can be used when cells operated on the same frequency are changed.

Softer handover is a special case of soft handover where the radio links that are added and removed belong to the same Node B (i.e. the site of co–located base stations from which several sector–cells are served. In softer handover, macro diversity with maximum ratio combining can be performed in the Node B, whereas generally in soft handover on the downlink, macro diversity with selection combining is applied.

Generally we can distinguish between intra–cell handover and inter–cell handover. The most obvious cause for performing a handover is that due to its movement a user can be served in another cell more efficiently (like less power emission, less interference). It may however also be performed for other reasons such as system load control.

3 Communication Measurement Toolset II

The Communication Measurement Toolset II (CMT II) [2] is a tool for testing and measuring the quality of end-to-end IP communication channels. It is developed in Java and its architecture is based on the Java Agent Development Framework (JADE) [3], supporting agent communication, administration and mobility. JADE realizes communication through Foundation for Intelligent

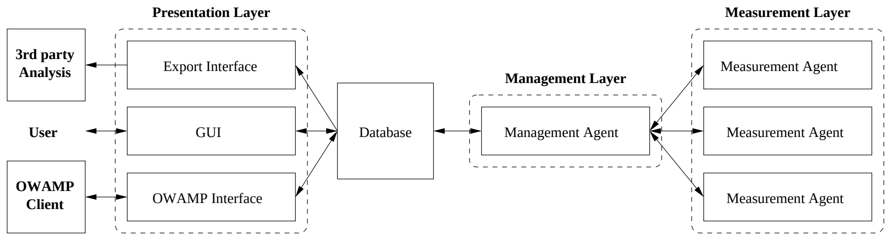


Fig. 1. CMT II framework architecture

Physical Agents (FIPA) compliant interaction protocols, allowing standardized communication to other agent platforms as well.

CMT II provides a management platform to handle the measurement scenarios and to measure the IP performance parameters for different transport protocols including their multiplexing. The different parameters (e.g. packet size) of the load generators can be adjusted and for traceability measurement parameters and results are stored into a database. The impact of different network configurations, protocol parameters and QoS parameters can be analyzed. CMT II is intended to generate traffic, which emulates real applications (in this case VoIP), therefore different traffic models are implemented. CMT II is a highly distributed system, which can be divided into four layers, illustrated in Fig. 1.

The *management layer* is responsible for centralized administration of measurement tasks. It distributes measurement requests to relating agents and handles measurement results and failures.

The *measurement layer* executes measurement tasks and consists of different measurement agents. There are single–point agents (e.g.: for passive monitoring) and multi–point agents (e.g.: for sender–receiver scenario). When a measurement agent is started, it registers itself on the management agent, concerning its measurement capabilities and host specific parameters (e.g.: network interfaces). The measurement agent monitors its own status and utilization and reports those values and events to the management layer, in order to ensure the traceability of results. The user can specify measurement scenarios, which consists mainly of the sending and the receiving host (specified through their IP addresses) and the load, which should be generated by the sender (specified by different parameters like protocol, packet size, packet inter-departure time, etc). The measurement results are throughput, packet loss, one-way delay and the instantaneous packet delay variation conforming to the IETF IPPM standards.

The *presentation layer* is used to make measurement results available for the community. With the CMT II GUI the user on the one hand can create and manage measurement tasks and on the other hand visualize measurement results. Regarding this point simple statistical evaluations and further analysis methods together with user definable graphs are provided. The second possibility to allow access to measurement results is via the export interface. The third alternative for accessing measurement results is the one way active measurement protocol (OWAMP) [4] interface,

4 Perceptual VoIP Quality

The CMT II framework uses the ITU *E-Model* [5–7] for predicting the perceptual quality of VoIP transmissions. The E–model is a very simple computational model, given by

$$R = R_0 - I_s - I_d - I_e^{eff} + A \ . \tag{1}$$

It is calculated by the use of so called impairment factors. These impairment factors can be individually calculated by the use of default values (e.g. echo cancellation) as well as measured values (e.g. packet loss). The E-model results in the so called R-factor which should be converted into a Mean Opinion Score (MOS). Here I_d denotes the impairment factor, taking delay into account. I_d is thoroughly given in [5] and approximated in [8]. I_e^{eff}, takes the audio codec, the packet loss concealment method and the packet loss probability into account (see [5]). The advantage factor A allows for compensation of impairment factors when there are other advantages of access to the user.

The E-Model is a generic model for perceptual speech quality prediction. Thus, not all impairment factors are relevant for VoIP. Here, $I_s = 0$ (e.g.: side noise). Eq. (1) can thus be reduced to

$$R = 93,2 - I_d - I_e^{eff} \tag{2}$$

with the equipment impairment factor given through

$$I_e^{eff} = I_e + (95 - I_e)\frac{P_{pl}}{P_{pl} + B_{pl}} \tag{3}$$

where I_e refers to the equipment impairment factor without packet loss, P_{pl} denotes the packet-loss probability (in our case measured with CMT II) and B_{pl} denotes the packet–loss robustness factor of the used audio codec.

I_d is approximated in [8] through

$$I_d = 0,024T_a + 0,11(T_a - 177,3)H(T_a - 177,3) \tag{4}$$

where T_a indicates the measured one-way delay and H denotes the Heaviside function.

Finally the R-factor can be converted according to [5] into a corresponding MOS through

$$MOS = 1 < 1 + (0,0035R) + R(R - 60)(100 - R)7 * 10^{-6} < 4,5 \tag{5}$$

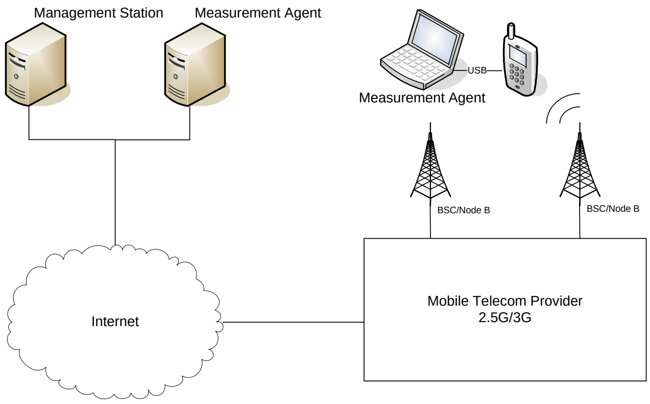


Fig. 2. Measurement Set–up

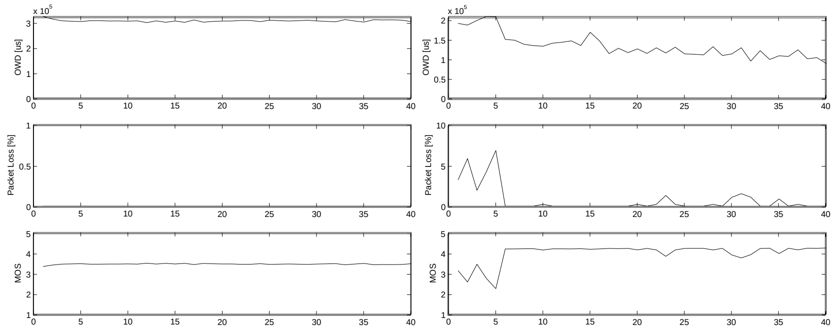
5 Measurements

This section presents the measurement set–up (see Fig. 2) and provides the corresponding measurement results. As illustrated in the measurement set–up diagram, we used a laptop and a Nokia 6630 mobile phone, connected via USB cable, to establish a PPP over UMTS connection to an Austrian mobile telecom provider.

To measure the one way delay between the nodes we use the Network Time Protocol (NTP) for time synchronisation. Furthermore, to measure the perceptual quality of a VoIP stream over the UMTS network, we had to emulate a frequently used audio codec for VoIP transmissions. In this context, the term *emulate* means, that we use the same packet size and sending interval as a real VoIP application, but without real (audio-) content in the payload. We decided to take the GSM–FR (Full Rate) audio codec, described in the *ETSI Recommendation GSM 06.10*, which uses *Adaptive Pulse Code Modulation (APCM)* to encode analog voice signals into digital signals.

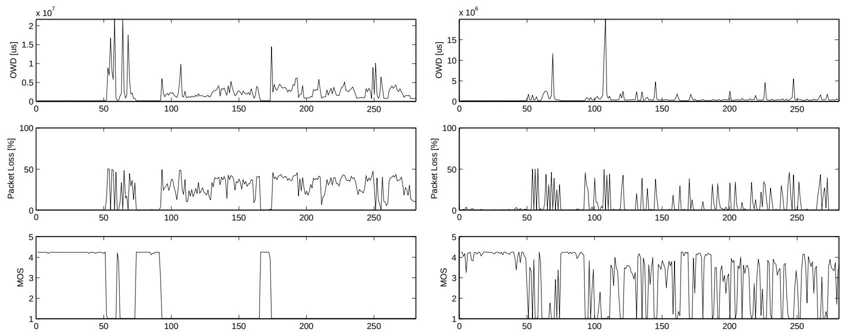
As mentioned in Sect. 1, we conducted stationary as well as non–stationary measurements, with a sending interval of 22 ms and a data rate of 13 kbit/s. Fig. 3 shows the results for the stationary measurement for both directions (upstream and downstream). This is important because here we have an asynchronous data transmission with different transmission rates per direction. The offered data transmission rate by the mobile telecom provider was 384kbit/s for the downstream (to the UE) and 64kbit/s for the upstream (from the UE). The plots at the bottom of Fig. 3(a) and Fig. 3(b) show the resulting MOS for the VoIP transmission. The MOS is calculated according to the equations given in Sect. 4 with $I_e = 26$ and $B_{pl} = 43$ for the GSM–FR audio codec [9]. The MOS values ranges from 1 to 4.5, with a higher value denoting better speech quality.

Concerning the stationary measurement, our main focus was on the verification of the measurement set–up. It can be seen from Fig. 3 that because we



(a) Upstream (b) Downstream

Fig. 3. Perceptual VoIP quality evaluation for a stationary VoIP over UMTS measurement. Each unit on the x–axis represents a 10s interval in which the *Mean Opinion Score (MOS)*, mean packet loss and mean OWD have been calculated.



(a) Upstream (b) Downstream

Fig. 4. Perceptual VoIP quality evaluation for a non–stationary VoIP over UMTS measurement. Each unit on the x–axis represents a 10s interval in which the *Mean Opinion Score (MOS)*, mean packet loss and mean OWD have been calculated.

physically stayed in the UMTS network during the whole measurement, there is absolutely no problem regarding the speech quality of the VoIP transmission.

Significant changes of the speech quality can be seen from Fig.4. This plot shows the measurement results for the non–stationary measurement scenario during a period of approximately 3000 seconds. We chose the test track for this measurement because we knew that a switch from UMTS to EDGE[1] would happen along the route. This background information was provided by the Telco provider.

At the beginning of the measurement the mobile node is physically stationary and is connected to a 3G UMTS network. After about 200 seconds the mobile node travels on a highway at a speed of about 100km/h. In the first period of the measurement, which lasted approximately 500 seconds, there is no degradation of the speech quality. Then, the significant degradation results from a hard handover, as it is described in Sect. 2, from the 3G UMTS network into a 2.5G EDGE network. The hard handover is a result of the unavailability of the UMTS network. If the 3G UMTS network becomes available again, then the UE re–selects it. The further significant changes in speech quality also result from hard handovers between 3G and 2.5G networks. Hence, the very long periods of low speech quality are indicators of the unavailability of the 3G UMTS network in this area. Low speech quality is a result from staying in a 2.5G EDGE network. The measurement results give a good overview of the speech quality on an all IP base wireless network. Based on the presented results, we can see that it is no problem to deliver VoIP over an 3G UMTS network, but it is impossible to use VoIP over an 2.5G EDGE network.

6 Conclusion

Concerning an all IP based 3G UMTS network, this paper shows that it is possible to use VoIP in a 3G network. To verify problems during a soft or softer handover, as described in Sect 2, it is important to compare signalling information of the 3G network and IP performance parameters. It is one further step in our work to integrate signalling information from the 3G network in the measurement framework CMT II. For a better understanding of measurement results it is highly important to collect information from different layers. If it is possible to collect, to analyse and to correlate measurement results from different layers during a measurement, the interpretation of the measurement results are more convincing. Another further step in our work should be the analysis of the impact of other audio codecs with different data rates, as well as to not only emulate VoIP traffic but analyse real audio data with different methods (e.g. PESQ).

[1] Enhanced Data rates for GSM Evolution

References

1. Taferner, M., Bonek, E.: Wireless Internet Access over GSM and UMTS. Springer, Berlin (2002)
2. SalzburgResearch Forschungsges. m. b. H.: Communication and measurement toolset. http://cmtoolset.salzburgresearch.at (2004)
3. Telecom Italia Lab: Java Agent Development Framework. http://jade.tilab.com (2005)
4. Shalunov, S., Teitelbaum, B., Karp, A., Boote, J., Zekauskas, M.: A One–way Active Measurement Protocol (OWAMP). http://www.ietf.org/internet-drafts/draft-ietf-ippm-owdp-14.txt (2004) Internet Draft.
5. ITU-T: Recommendation G.107 - The E-Model, a Computational Model for use in Transmission Planning (2003) ITU–T Standard.
6. ITU-T: Recommendation G.108 - Application of the e-model: A planning guide. (1999)
7. ITU-T: Recommendation G.113 Appendix I - Provisional planning values for the equipment impairment factor Ie and packet-loss robustness factor Bpl. (2002)
8. Lingfen, S.: Speech Quality Prediction for Voice over Internet Protocol Networks. PhD thesis, University of Plymouth (2004)
9. ITU-T Study Group 12: Contribution 106: Estimates of Ie and Bpl parameters for a range of Codec types. http://www.telchemy.com/technicalreference.html (2004)