END-TO-END VOIP QUALITY MEASUREMENT

Peter ČÁKY, Martin KLIMO, Peter PALÚCH, Ondrej ŠKVAREK
Department of InfoCom Networks, Faculty of Management Science and Informatics,
University of Žilina, Univerzitná 8215/1, 010 26 Žilina, Slovak Republic,
E-mail: {Peter.Caky, Martin.Klimo, Peter.Paluch, Ondrej.Skvarek}@fri.utc.sk

SUMMARY

The VoIP quality of service can be either estimated by analyzing the voice data flow characteristics, or by recording a transmitted voice probe over the whole VoIP transport chain and comparing it to the original. When the quality estimation by the voice data flow characteristics is deployed, the three most prominent degradation factors are observed – the packet loss, the packet delay and the delay variance (also called jitter). Depending on the estimation model, the values of these degradation factors are used to compute the resulting quality rating estimation. The end-to-end quality measurement provides more dependable results, however, it is also more technically and computationally demanding. We describe our experiences with both methods of quality evaluation and measurement. An experimental VPN for VoIP QoS measurements was created and a voice probe was sent through this network while end-to-end quality and degradation factors were observed. For end-to-end measurements we used our own developed equipment that allowed us to synchronize the original and recorded voice probe and obtain the timing characteristics. The obtained results are also discussed in this article.

Keywords: *quality, service, voice, E-model, QoS measurement*

1. INTRODUCTION TO VOIP

Voice over IP (VoIP) is an emerging and promising family of services that provides telephone services over computer and data networks running the IPv4 and IPv6 protocols. There are various benefits of deploying VoIP:

- VoIP is an integral part of the converged network paradigm where data and voice services are provided by the same network, as opposed to the former approach to separate the data and telecommunication networks
- VoIP transports voice packets as ordinary data, thus allowing the use of well-known supported techniques to control their flows in data networks
- Under circumstances, the VoIP may prove more scalable and easier to deploy than the traditional PSTN services
- VoIP may be beneficial to the overall costs for voice calls

VoIP aims to eventually replace the PSTN services; however, there are still many inherent problems that must be solved. Among them, the Quality of Service is one of the most important.

2. QUALITY OF SERVICE

The quality of service (QoS) is defined in the ITU-T recommendation [1] to be „the collective effect of service performance which determines the degree of satisfaction of a user of the service“. Particularly, the quality of VoIP service comprises overall voice quality, user comfort, additional services and so forth.

The End-To-End VoIP QoS (abbrev. E2E) measurement evaluates the complete transmission chain between two subscribers, starting at first subscriber's microphone and ending at other subscriber's speaker. It is the most complete evaluation of the provided service.

There are many approaches to QoS evaluation and measurement. The *subjective* measurements are done by groups of people that evaluate the QoS under given conditions and methodics. The most common subjective method is specified in [2] and it evaluates the quality using a scalar value called Mean Opinion Score (MOS). The *objective* methods assess the QoS using various computational models. During the time, many objective methods were devised, the most prominent being Perceptual Speech Quality Measure (PSQM, [3]) and Perceptual Evaluation of Speech Quality (PESQ, [4]). As the objective methods are usually calibrated and recalculated to the MOS score, both subjective and objective methods remain important for QoS evaluation.

3. DEGRADATION FACTORS

Degradation factors comprise all influences of a transporting network that negatively impair the quality of service. Various degradation factors can be identified in the usual telephone systems, relating mainly to acoustic transporting characteristics of the transmission chain. Moreover, in data networks, these specific factors have been identified as the additional and application-specific source of quality degradation:

- Packet loss
- Delay
- Jitter

3.1. Packet loss

Packet loss occurs when a packet that has been sent from the sender will not be replayed at the receiver. The obvious reasons for packet losses are datagram collisions on shared media networks, congestion in network, traffic shaping and bandwidth limiting, and so on. However, a packet might not be replayed also due to its late arrival. The results of packet loss are drop-outs, interruptions and deteriorations in the replayed sound.

Recent studies show that concealment techniques and voice activity detection routines built into the VoIP phones have a significant influence on the perceived QoS, and that the impact of the packet loss has to be studied dependently on the transported sound. Then the speech has to be modelled by a cyclostationary stochastic process. This topic was discussed in more detail by Bachratá in [8].

3.2. Delay

Delay is the time that passes between the moment of recording of a sound sample and its replaying at the receiver (often called as the mouth-to-ear delay). The overall delay consists of smaller delays that occur in various stages of the sound recording and transmission. The most delay-incurring operations are the sound encoding and decoding in the voice codec (algorithmic delay), the cumulation of enough samples to be sent at once in a packet (packetization delay), and the time needed to transmit a packet over a connection with some bandwidth and throughput (serialization delay).

The effects of delay are the echo (with relatively small values of delay – in tens of milliseconds) and the talker overlapping (larger delay – hundreds of milliseconds). Particularly, the talker overlapping is very unpleasant as it destroys the interactive essence of the telephone service.

The impact of delay on the network performance was further studied by Uramová in [11].

3.3. Jitter

Jitter is a measure of delay variance. The delay is not a constant value; rather it varies with time depending primarily on network load. There are more approaches for jitter calculation but the most common is the average of differences in delay between subsequently transmitted packets.

Because of jitter, all VoIP terminals must be equipped with a jitter buffer to alleviate its effects. However, the size of this buffer is a critical parameter. If the buffer is too small then it will not be effective against higher values of jitter resulting in packet losses. On the other hand, if the jitter buffer is too large then also the delay will increase.

4. IMPACT OF DEGRADATION FACTORS

Modelling the impact of degradation factors on the VoIP QoS is dependent on the method of QoS evaluation. As this paper discusses methods using the E-model described in [6], we will discuss the impact of degradation factors as modelled in the E-model.

The E-model is a computational model that is used to estimate the QoS in a network with given transmission parameters. It is intended primarily for transmission planning. Though it is not intended to actually predict the customer QoS opinions, it has shown itself to closely correlate with the MOS scoring. Originally, the E-model took only acoustic transmission parameters into account. Since its proposal, various researches, including ours, are aimed to further improve the accuracy of the E-model for the VoIP applications by including the levels of delay, jitter and packet loss in the computation.

The E-model evaluates the QoS by computing a scalar value called the R-factor. The basic paradigm of the E-model is that the quality is perceived psychologically and thus the psychological factors are additive on the psychological scale.

The basic formula for R-factor computation is:

$$R = Ro - Is - Id - Ie,eff + A \tag{1}$$

where Ro represents in principle the basic signal-to-noise ratio, including noise sources such as circuit noise and room noise. The factor Is is a combination of all impairments which occur more or less simultaneously with the voice signal. Factor Id represents the impairments caused by delay and the effective equipment impairment factor Ie,eff represents impairments caused by low bit rate codecs. It also includes impairment due to packet-losses of random distribution. The advantage factor A allows for compensation of impairment factors when there are other advantages of access to the user. The term Ro and the Is and Id values are subdivided into further specific impairment values. Details about calculating these factors are beyond scope of this paper and are discussed in detail in [6].

5. E2E QoS MEASUREMENT

It is not a goal of this paper to give an overview of methods used for end-to-end VoIP quality evaluation. Reader can find the descriptions of these methods in many references, e.g. [9], [10]. Instead, in this paper we are focusing to an approach that combines the E-model with a classical telecommunication channel quality evaluation given by transmission characteristics.

To obtain a VoIP QoS evaluation from E-model, all of its parameters have to be measured or evaluated. We assume that in the mouth-to-ear VoIP chain, a jitter buffer is also included, which is usually located in a VoIP gateway or VoIP phone. The jitter buffer eliminates all jitter, paying in exchange by increased delay and packet loss. Compared with the digital telephony, delay and packet loss are the main degradation factors that have to be newly measured at the E2E level to evaluate mouth-to-ear speech quality. Other degradation factors are known from digital telephony, and for simplicity we assume they are set to default values. To assess the E2E values of delay and packet loss, we propose an intrusive method of their measurement. The basic configuration of the measurement testbed contains end interfaces which are located between the telephone headset and the telephone body using the headset jack as the interface point. Due to standardised and simple construction of headset, the complete mouth-to-earth chain can be evaluated.

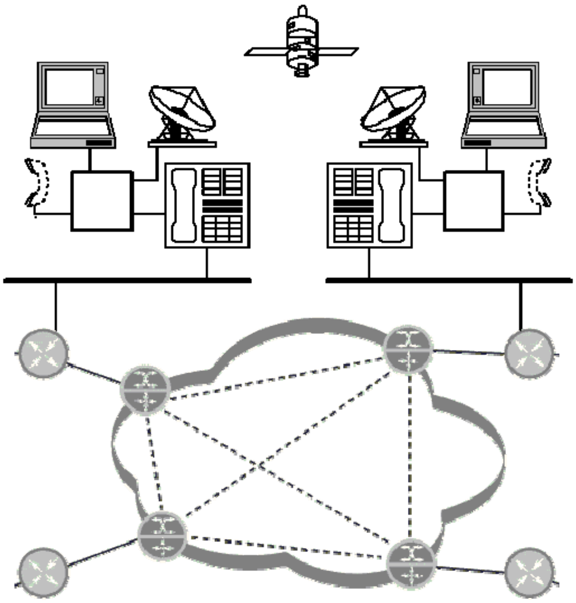

Fig. 1 Testbed configuration

The sample sound file is located in the sending computer, and it is transmitted over a soundcard interface and the measurement device to the headset jack in the telephone. To have a reference file, the sending computer also records this sound. VoIP telephone converts the generated sound into packets that are transmitted over local area network (Ethernet), IP WAN network, and local area network at the receiving side into the VoIP telephone. The receiving VoIP telephone converts obtained packets into sound again. This waveform is routed over measurement device and soundcard interface to the computer, where it is stored as a received sound. E2E QoS evaluation is based on comparing the original sound file with the received one. To allow an exact fitting of these two sound files, a common timing is a must. The GPS system is used as a source of time basis. GPS receiver receives timing information from satellite; time signal is decoded by the measurement device and stored in parallel soundcard channel together with generated or received sound.

5.1. Source file

To obtain relevant results, it is necessary to prepare the source sound file with sufficient care. There are two basic sounds within this file: the speech and sine waveforms are recorded in it. Packet loss and delay are independent on speech content; therefore language, speaker or content is not interesting. However, if the voice activity detection is used in the codec, the packet stream may be modified also by language or content. To avoid this, the G.711 constant bit rate sample stream is modulated by ON/OFF states that are distributed according to [5]. Sine waveforms are pre-recorded on defined levels that are needed for channel characteristics measurement. These levels have to take into account the properties of used soundcard and they can be calibrated in a limited range by the measurement device.

5.2. Delay measurement

According to the used codec, packets carry 20 ms or 30 ms frames of sound (speech or sine waveforms). Individual packet transmission delay is given as a difference between the time instant of packet being received and the time instant of the same packet being sent. By comparing the two records from the sending and receiving side we can obtain this difference. Both these records contain transmitted signal as well as the timing signal obtained from GPS receiver. Then the main task of merging process of sound records is to synchronise the recorded samples according the absolute time axis given by GPS clock.

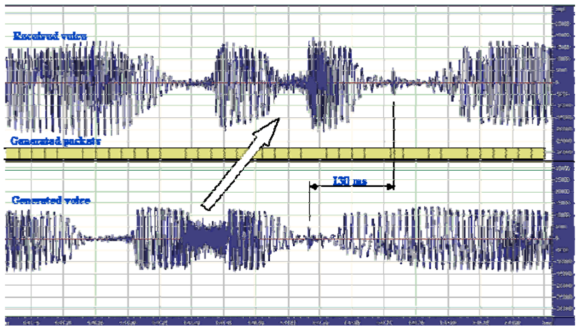

Fig. 2 Voice record sample

After merging sounds into one stereo record, the pairs of corresponding packets in sender record and receiver record are identified by correlation analysis, and the individual packet delay is calculated. Finally the stochastic properties of individual delay can be calculated, e.g. distribution, mean or variance.

5.3. Packet loss measurement

When the individual packets are identified in the recorded sound files as was described in the previous section, it may happen that the sent packet has no corresponding packet at the receiver side. In this case a packet loss is indicated and the number of lost packet is recorded. Record of lost packet numbers is post processed, and probabilistic characteristics are calculated. The basic parameter is a packet loss probability, but also more descriptive characteristics like distribution of interval between consecutive packet losses can be evaluated.

Some difficulties may occur when VoIP telephone uses some loss recovering algorithms. Then simple correlation methods for packets assignment may fail, and additional criteria have been used.

The described procedure for evaluating the packet loss is based on analysis of the recorded waveforms. However, for validation and comparison purposes, we have also monitored the packet loss by analyzing the packet stream itself. The VoIP data are transported using the RTP protocol specified in [7]. The headers of individual RTP packets contain the Sequence Number field which is 16 bits wide and is incremented by 1 in subsequently transmitted packets. The packet loss from the sender up to the measurement point can thus be evaluated by analyzing the sequence numbers. We have used the Agilent Distributed Network Analyzer hardware with the supplemental software and the Ethereal packet analyzer software for the actual packet loss analysis. The Network Analyzer was directly attached between the VoIP phone and the data network.

It should be noted that in this configuration, only packet losses due to packet discards in the network can be evaluated. The packets which were discarded by the VoIP phone itself due to their late arrival are not recognizable in this configuration, as they appear to have arrived to the Network Analyzer. Originally we hoped that we could use the RTCP Sender Reports messages (as defined by [7]) to examine the exact packet loss statistics. The RTCP SR reports are reports that the VoIP endpoints should generate and exchange; they contain the statistical information to calculate the packet loss, round-trip delay and jitter. However, we have found out that none of VoIP phones we have used in our experiments implemented the RTCP protocol correctly: some phones were sending malformed RTCP SR reports while others did not send them at all.

5.4. Amplitude distortion measurement

The main advantage of end-to-end measurement is the fact that the quality evaluation is based on comparing of generated sound and received sound independently of all transformations that are applied to the transmitted signal. This approach allows for the abstraction from the type of voice transmission chain (analogue, digital or packetized) and uses only waveforms of original and transmitted sounds. This is similar to classical telephone channel measurements, a topic where rich knowledge and experiences have been gained so far. This approach gives a complex evaluation of the whole transmission chain and can very quickly indicate the linear distortion of the sound. The measurement is organised in a classical way when sine waveforms at given frequencies and level are generated, and level of received waveform is measured at the receiving side (phase characteristic is omitted due to its unimportance for voice quality evaluation). Results are plotted into an attenuation characteristic (see Fig. 3). Nonlinear distortion can be evaluated from an output/input characteristic that can be obtained from sine waveforms generated at standardised frequencies and given levels while the levels of output waveforms are measured (see Fig. 4).

6. MEASUREMENT RESULTS

Authors have developed a measurement device which allows generating and recording test waveforms over the VoIP telephone headset jack, as well as decoding and recording of GPS time signal. As an example of E2E VoIP quality evaluation, we describe results of the measurement that was running on the private IP/MPLS network with mixed voice and data traffic. According to Fig. 1, two measurement sets were installed in Banska Bystrica and Zilina (100 km distance) and the channel quality was measured as it was described in the previous section. The Polycom VoIP telephones were used as the VoIP endpoints.

The following results were obtained:

- mean delay 120 ms
- packet loss 0%
- R factor 82,8 (according to E-model)
- attenuation characteristic – see Fig. 3
- output/input characteristic – see Fig. 4

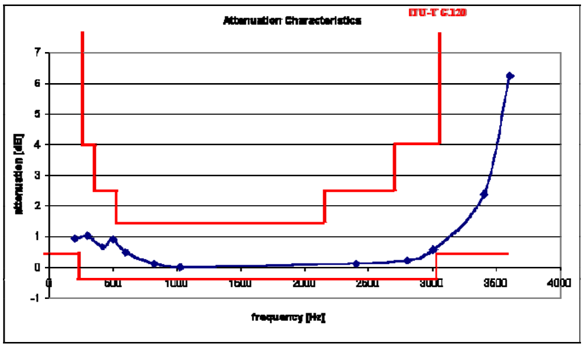

Fig. 3 Attenuation channel characteristic

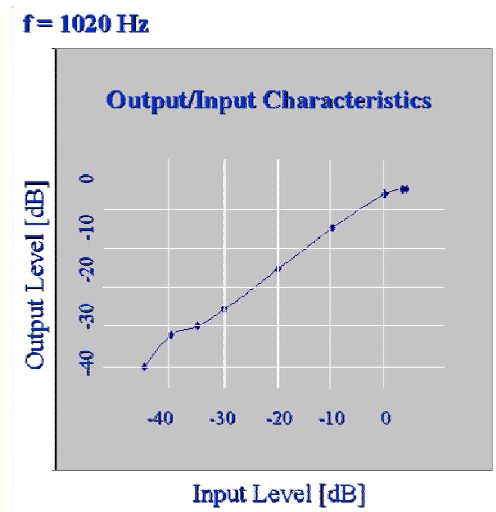

Fig. 4 Output/Input characteristic

Results obtained by this sample measurement show acceptable VoIP quality. The delay appears to be critical, obviously caused by a low speed (256K) connection for separated voice traffic (given by contract as a class of service). The attenuation characteristics satisfied ITU-T G.120 Recommendation and the output/input characteristic is linear in a broad range of signal level.

7. CONCLUSIONS

Introducing Voice over IP for telephone services brings new challenges in quality evaluation because of new degradation factors, namely delay and packet loss. To fulfil the customer needs, impact of these factors to VoIP quality of service has to be studied and new measurements methods have to be developed. The paper explains the approach to end-to-end quality evaluation that has been developed at the Department of InfoCom Networks, University of Zilina. This approach is based on original research results in understanding of packet loss impact to the voice quality, as well as on developing measurement tool for E2E VoIP quality evaluation. Usage of this tool and research results were documented on the example of VoIP QoS measurement on the real network. Results of the measurement provide a deep view to the transmission channel properties and give not only the actual transmission quality assessment, but also allow finding the potential limits or week points. This can avoid the quality degradation in the case of network load changes or reconfigurations.

REFERENCES

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

[2] ITU-T: Recommendation P.800: Methods for subjective determination of transmission quality, 1996

[3] ITU-T: Recommendation P.861: Objective quality measurement of telephone-band (300-3400 Hz) speech codecs, 1996

[4] ITU-T: Recommendation P.862: Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs

[5] ITU-T: Recommendation P.59: Artificial Conversational Speech, 1993

[6] ITU-T Recommendation G.107: The E-model, a computational model for use in transmission planning, 2005

[7] IETF: Request For Comment 3550: RTP: A Transport Protocol for Real-Time Applications, 2003

[8] Bachratá K.: Vplyv straty paketov na prenos reči, INFOTRANS 2005, Mezinárodní konference informačních technologií v dopravě, Pardubice 21.-22.9.2005, Univerzita Pardubice, str. 15 – 23, 55-768-05, ISBN 80-719480-7194-792-X

[9] Jakab František: Metódy optimalizácie hodnotenia a riadenia v počítačových sieťach. Meranie a vyhodnocovanie prevádzkových parametrov., Dizertačná práca na získanie vedecko-akademickej hodnosti philosphiae doctor „PhD.“, Košice, TU Košice, 2005, 133 str.

[10] Hronský Martin, Jakab František, Potocký Miroslav, Jakab Róbert, Giertl Juraj: Sampling Algorithms for Nonintrusive Measurement in Network Oriented Educational Systems, Conference Proceedings of the 4th International Conference on Emerging e-learning Technologies and Applications ICETA 2005, Information and Communications Technologies in Education, Košice, Slovak Republic, 13 - 14 September 2005, Košice, elfa, s.r.o., 2005, 4, pp. 165-171, ISBN 80-8086-016-6

[11] Uramová J.: Analytické a simulačné modely priepustnosti informačných sietí, INFOTRANS 2005, Mezinárodní konference informačních technologií v dopravě, Pardubice 21.-22.9.2005, Univerzita Pardubice, str. 304 – 312, 55-768-05, ISBN 80-719480-7194-792-X

BIOGRAPHY

Peter Čáky was born in 1978. In 2001 he graduated (Ing.) at the Faculty of Management Science and Informatics at University of Žilina. Since 2001 he is working as an internal or external (since 2005) PhD student at the Department of InfoCom Networks.

Martin Klimo was born in 1949, received the Ing. and PhD. degrees from University of Žilina in 1973 and 1985 respectively. Since 1990 he is working as Associated Professor and in 1993 he was appointed head of the Department of InfoCom Networks. His professional interests include communication theory, queueing theory and text-to-speech systems as well as e-learning.

Peter Palúch was born in 1981. After graduating in 2004 (Ing.) at the Faculty of Management Science and Informatics at the University of Žilina he continues to study as an internal PhD student at the Department of InfoCom Networks at the same faculty.

Ondrej Škvarek was born in 1963. In 1986 he graduated (Ing.) at the Faculty of Electrical Engineering, University of Žilina. Since 1991 he is working as professor assistant with Department of InfoCom Networks. His professional interests include networking, text-to-speech systems.