Quality of Service in VoIP Communication

Tadeus Uhl

Dedicated to Professor Rudolf G. Schehrer on the occasion of his 65th birthday

Abstract: The focus of this paper lies in the practical aspects of voice over IP communication. VoIP configurations in the H.323 standard will be presented briefly. Following that, the fundamental protocol procedure of H.323 communication will be briefly explained. A further part of the paper will address the subject QoS (quality of service), and present the common measurement methods used in QoS. Results gained from experiments conducted in a VoIP environment will then follow. The investigations concentrate primarily on the load behavior of voice packets in relation to important parameters of this service. The results obtained are presented and evaluated in diagrams. The paper concludes with a summary.

Keywords: Real time communication, Multimedia application, Measurement methods, Traffic engineering

1. Introduction

The underlying technical prerequisites for Voice over IP (abbreviated VoIP) have existed for some time now. Protocols such as H.323 [1] and SIP [2], that were published a number of years ago, are now providing the basis for packet-based multimedia communication. There are several VoIP systems available on the market (e.g. Siemens' HiPath 5500 and Innovaphone IP from TLK Computer PLC, Gateways from Cisco, Lucent, Apolio, Selsius, et al.) that incorporate the standards mentioned above. Today, this form of communication is enjoying a remarkable level of popularity and is the prerequisite for new and innovative services such as teleteaching and tele-learning [3]–[4].

The Internet and most intranet systems are, however, ill equipped to deal with the transfer of time-critical data streams. Runtime discrepancies between packets and packet loss are two of the most notable features of IP networks and present significant problems to voice and video communication. To counter this, processes have been implemented in the upper protocol layers to monitor and support the Quality of Service (QoS) schemes. And it is this problem area that this paper addresses.

2. VoIP architecture

Fig. 1 presents an overview of the architecture and interoperability of VoIP.

Tadeus Uhl, Flensburg University of Applied Sciences, Institute of Communications Technology, Kanzleistr. 91–93, D-24943 Flensburg, Germany. Phone: +49/461/805-1763, Fax: +49/461/805-1474 E-mail: uhl@fh-flensburg.de

Fig. 1 shows clearly the four main constituents of VoIP architecture: terminal, gatekeeper, gateway and multipoint control unit. The terminal comprises the user terminal equipment and supports voice, video and data communication. The gatekeeper provides the terminals and gateways with services for authentication and call control. The gateway is the interface between heterogeneous networks such as the LAN and the ISDN net. The multipoint control unit is needed to support conferencing.

3. H.323 protocol stack

Fig. 2 presents an overview of the H.323 protocol suite.

The ITU recommendation H.323 consists of a large suite of protocols. The various protocols have been overlaid on the transport protocols Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP).
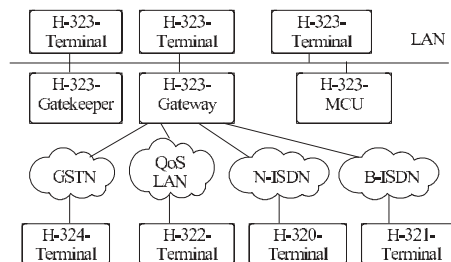

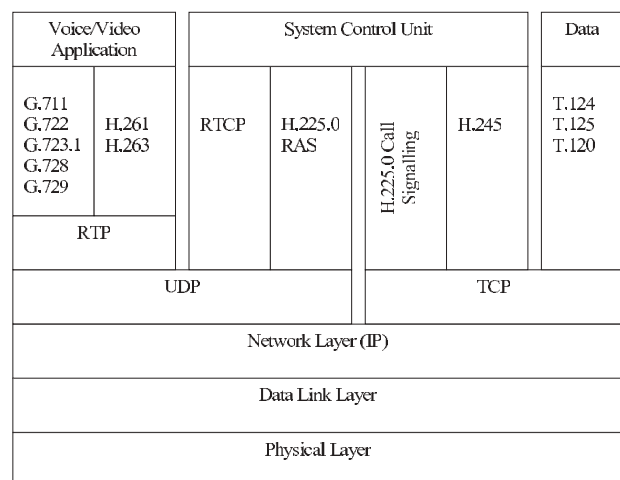**Fig. 1.** Architecture and Interoperability in VoIP.

Fig. 2. Overview of the H.323 Protocol Suite.

The data from the signalling protocols H.255.0 and H.245 are transmitted over the secure TCP. Voice and video data, however, are transmitted over the UDP, because error correcting would only worsen the quality of voice transmission. The registration and admission protocol RAS is also transmitted with the aid of the real-time capabilities of UDP.

The UDP is in principle not suitable for real-time data transmission and the voice and video data must first be processed by the RTP or RTCP before a real-time transmission can take place and the receiver end can obtain additional time and synchronization information necessary for the reassembly of the packets in the correct original order.

Three different types of connection can be made when an H.323 connection is established. The connection establishment can be done with either with a dynamic gatekeeper, with a static gatekeeper or without a gatekeeper at all. Different courses of actions are taken depending on which type of connection is supported by the application. Fig. 3 shows clearly the phases of an H.323 connection with gatekeeper.

The connection can be divided into five phases:
- The signalling phase performed by H.225.0 (Client A calls Client B),
- the connection control of logical voice channels by H.245,
- the conversation phase during which not only voice packets but also control packets are constantly exchanged,
- the release of the logical voice channels by H.245 (started when the receiver is replaced on the hook), and
- the release of the bandwidth used by the gatekeeper and confirmation to the subscriber that the connection has been terminated.

A detailed description of the protocol procedure and its separate phases can be found in works [5].
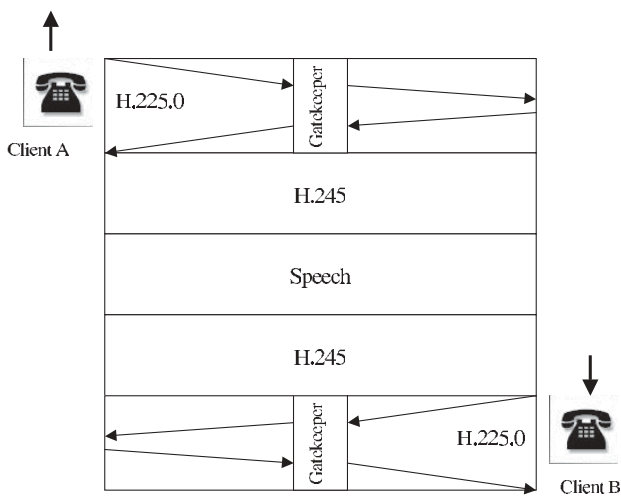
4. Methods to achieve QoS in VoIP

Four methods are used to evaluate the quality of voice transmission. The MOS method [6] is based on the subjective judgment of a group of individuals. The other methods, i.e. PAMS [7], PSQM [8] and PESQ [9] are objective assessment methods that work with a computer-aided evaluation of randomly chosen samples of transferred voice data compared with corresponding references.

The Mean Opinion Score (MOS) method specified in the ITU-T P.800 Recommendation of August 1996 is a procedure for evaluating quality of speech based on the subjective appraisal of a number of subjects. This method was the first method of note for measuring quality of speech. The recommended methods for evaluating quality of speech are the conversation opinion test, the listening opinion test und the interview and survey test. The benchmarks for quality of speech are the so-called MOS values (see Table 1).

The Perceptual Analysis Measurement System (PAMS) was developed by British Telecom in order to evaluate perceived speech quality. It was the first model to use objective methods to assess speech quality in end-to-end networks. PAMS was the forerunner of the international standard Perceptual Speech Quality Measure (PSQM; ITU-T Recommendation P.861) used nowadays. PAMS represents values for quality of speech in two ways: a) listening quality and b) listening effort. Table 2 presents both scales of quality.

The Perceptual Speech Quality Measure (PSQM) method was developed by KPN Research in the Netherlands and was first specified in the ITU-T Recommendation P.861 of February 1998. This method, based on the psychoacoustic properties of the human sense of hearing, defines methods with which to assess the speech quality

Fig. 3. Connection phases of an H.323 connection with gatekeeper.

Table 1. MOS scale of values.

MOS Value	Speech Quality
5	excellent
4	good
3	fair
2	poor
1	bad

Table 2. PAMS scale of values for quality of speech.

PAMS Value	Listening Quality	Listening Effort
5	excellent	noeffort required
4	good	no appreciable effort required
3	fair	moderate effort required
2	poor	considerable effort required
1	bad	no meaning understood without effort

Table 3. PSQM and MOS values in comparison.

PSQM Value	MOS Value	Quality of Speech
6.5	1	bad
6	1.31	
5	1.92	
4	2.54	
3	3.15	
2	3.77	
1	4.38	
0	5	excellent

Table 4. Comparison of PESQ und MOS values.

PSQM Value	MOS Value	Quality of Speech
4.5	5	excellent
4	4	good
3	3	fair
2	2	poor
1	1	bad

of a codec and therefore objectively evaluate the speech quality of a telephone connection. PSQM is primarily employed in networks that use speech compression, digital speech interpolation and packet-based transmission. The measure of quality is given by so-called PSQM values (see Table 3).

Perceptual Evaluation of Speech Quality (PESQ) is a method to evaluate the quality of speech in telephone networks and the speech quality of speech codecs. The method was developed jointly by British Telecom and KPN Research. PESQ combines the time alignment of PAMS and the perceptual model of PSQM. It can be used not only on speech codecs but is also designed to quantify end-to-end connections. Defined in the ITU-T Recommendation P.862 of February 2001, PESQ has become the recognized standard for measuring speech quality in VoIP networks. Fig. 4 shows the principle underlying of the PESQ method.

In this model, the original signal, or a reference signal, is transmitted over a telephone line. The lossy incoming signal (the test signal) is received at the other end. Any interference acting on the speech signal during transmission over the connection being tested will cause the test signal to differ from the reference signal. So that the signals can be compared with each other, a perceptive hearing model is used. This model is a replica of the human hearing apparatus and is actually very good. The signals must be in the form of spoken language since so many systems have been designed to transport spoken language. After a number of transformations and calculations the discrepancy between the hearing-emulating internal representations of the reference signal and the test signal can

be constructed. The result of this contrasting of the two signals is an indication of the audibility of the interference originating in the line being tested. This can then be used to calculate a value for the quality of speech. Table 4 shows the comparison between PESQ values and MOS values.

Another way to achieve a quantitative analysis of quality of speech in VoIP is to use classic parameters such as end-to-end delay, jitter delay, loss probability and communication patter [10]. These parameters can be measured with protocol analysers, processed and then represented graphically. To be sure, this method fails to include some of the criteria considered in the complex methods to measure QoS described above but it is simple to use. And that is of practical importance. It would be stimulating to draw a comparison between values gained from complex measurement methods on the one hand and simple methods on the other. The next chapter will address that very issue.

5. Practical measurements of QoS

Fig. 5 shows the analysis environment used in the analysis. It consists of the real-time communications system HiPath 5000 [11], the protocol analyser DA-360 (incorporating a load generator) [12], and the analysis software application TraceView VoIP (implementing a PESQ algorithm) [13].

In order to conduct a quantitative and qualitative analysis of quality of speech, the DA-360 measurement sys-

Fig. 4. Overview of the basic structure of PESQ.
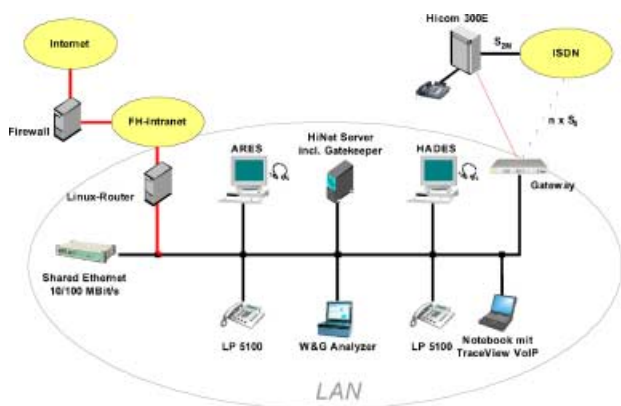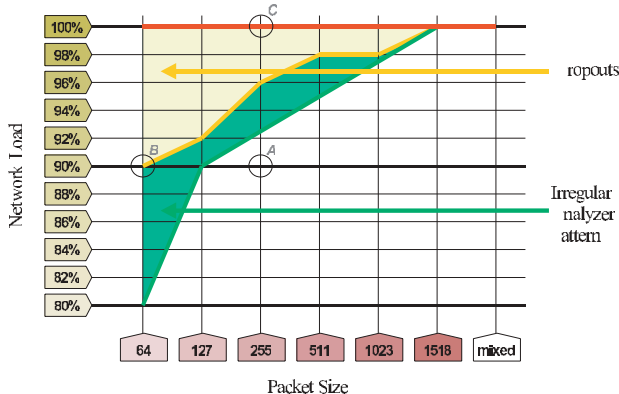

Fig. 5. Analysis environment.

tem was added to the net to propagate 511-byte packets and simulate a varying network load. A series of speech tests was conducted for each of the varying net loads. Ten tests were considered to be an appropriate number for each series. The reference file used for the series of tests was the file Or105.wav in ITU-T [9]. The size of the user data packets transmitted was 480 byte and the jitter buffer was adjusted to 100 ms. The objective values gained for quality of speech using the analysis software TraceView VoIP were then compared with subjective assessments. Table 5 shows the results of the analysis.

With respect to the results of the analysis, it must be said that speech quality remains excellent until net load reaches 85 per cent. It is also noticeable that as the net load increases, the average net delay also increases. The increase in jitter delay (not shown in Table 5) was countered by the size of the jitter buffer. It becomes clear that the objective measurement of speech quality coincides to a large extent with the subjective assessment of speech quality.

Fig. 6 shows the results gained from qualitative and quantitative analysis using classical performance values (communication patter and loss probability). In this experiment packet size, too, was varied during the production of network load. When it came to evaluating the quality of speech, three areas were picked out for individual consideration: area A (best quality possible), Area B (disturbed communication) and area C (communication impossible). It is to be noted that the quality of speech is influenced only when net load reaches

90 per cent (using the smallest packets admissible in Ethernet). This area increases when load packet size is increased. In the grey area, it was possible to detect by metrological means (i.e. objectively) irregular communication patters. Yet these had no noticeable effect on the subjective assessment of the quality of speech. The light grey area contains dropouts that were reflected in a bad quality of speech and could also be ascertained, i.e. proved metrologically, through packet loss [10].

When the measurement results from Figs. 5 and 6 are compared it becomes clear that the quantitative analysis of quality of speech using classical performance values (that are easy to measure in practice) on the one hand and using complex measuring methods such as PESQ algorithm (using expensive measuring equipment) lead to similar conclusions.

6. Summary

In the course of this study, the subject of quality of speech in VoIP was covered in detail from theoretical and practical standpoints. After a discussion of the latest standards for QoS and VoIP, practical measurements were conducted. The results obtained were presented in tables and diagrams, and then interpreted.

The measurements have shown that quality of speech in VoIP is very much dependent on the basic network load. Small packets in the environment have a larger influence on VoIP than larger packets. From this behavior it can be concluded that it is not the net load as such that causes poor quality of speech but rather the number of packets sent within a time span that must be considered as the overriding factor. If more than a certain number of packets are sent within a certain time span, jitter will increase exponentially bringing with it all the negative consequences for transmission quality and delay.

An analysis of speech quality in VoIP using complex QoS algorithms confirms the properties described above. It must be stressed that measuring equipment that implements the standards for QoS are expensive. They are, however, more flexible when it comes, for instance, to testing speech codes, end equipment, etc. On the other hand, the investigations showed that is possible, in principle, to evaluate quality of speech in VoIP using simple methods (and less expensive equipment).

The network loads used in this study were relatively simple and could be treated in much more complexity in future work. Of primary interest would be an investigation using multimedia load generators in the network environment. This would have simulated load scenarios more similar to those in real IP networks. It would be preferable if investigations into the quality of speech parameter could be conducted when several VoIP connections are active simultaneously. Such investigations would have complemented well the results presented in this paper. Further studies should point in this direction.

Table 5. Analysis results alongside PESQ values.

Network Load	PESQ Value (mean)	Subjective Evaluation	Average Delay [ms]
0	4.402	excellent	160.4
50	4.402	excellent	167.2
75	4.402	excellent	167.6
85	4.402	excellent	168.4
95	2.746	poor	220.5

Fig. 6. Measurement results using classic parameters.

References

[1] http://www.openh323.org/standards.html.

[2] http://www.sipcenter.com/files/sip2.pdf.

[3] http://www.fernuni-hagen.de/LVU.

[4] http://www.vfh.de.

[5] Roedig, U.; Ackermann, R.; Steinmetz, R.: IP-Telefonie und Firewalls, Probleme und Lösungen. PIK **(2001)** (24), 32–40.

[6] ITU-T Recommendation P.800: Methods for subjective determination of transmission quality. http://www.itu.org.

[7] PAMS: Measuring speech quality over networks . . . as the customers hear it (White paper).
http://www.psytechnics.com/papers/PAMS_1.1.pdf.

[8] ITU-T Recommendation P.861: Objective quality measurement of telephone-band (300–3400 hz) speech codecs. http://www.itu.org.

[9] ITU-T Recommendation 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. http://www.itu.org.

[10] Uhl, T.; Zäncker, O.: Voice over IP exemplifed by Siemens HiNet RC 3000. Proc. of the ICC-2001, St. Petersburg/Russland, Juni 2001.

[11] http://www.siemens.de.

[12] http://www.acterna.com.

[13] http://trafficlyser.de.

Tadeus Uhl received a M. Sc. degree in Electrical Engineering from the Technical Academy of Bromberg (Poland) in 1975 and a Ph. D. from the Technical University of Danzig in 1981. He got his postdoctoral lecturing qualification at the University of Dortmund in 1990. From 1975 till 1982 he worked as an scientific assistant at the Institute of Telecommunications and Electrical Engineering in the Technical Academy of Bromberg and from 1982 till February 1992 as a member of the scientific staff at the Institute of Electronic Systems and Switching in the University of Dortmund. Since March 1992 he is a professor at the Institute of Communications Technology in the Flensburg University of Applied Sciences.

His study subject includes performance evaluation of WAN and LAN protocols and switching techniques in the ISDN.