QOS MEASUREMENT AND EVALUATION IN PRIVATE NETWORK OF SPP PRIOR TO VOIP IMPLEMENTATION

Peter CHOCHOL
Section of IT Development, Division of Information Technologies,
Slovak Gas Industry, Mlynske nivy 44/a, 825 11 Bratislava, Slovak Republic

ABSTRACT

Voice over Internet Protocol (VoIP) has many advantages for enterprises, chief of which its cost-effectiveness. However, to compete with traditional telecommunications technologies, the quality of speech over a VoIP connection must be comparable to or better than that of the Public Switched Telephone Network (PSTN). Enterprises must perform a full network assessment to discover all network impacts prior to the IP telephony deployment. This will help avoid service problems, unplanned additional costs and deployment delays of IP telephony implementations. Several objective testing methods are available to measure, monitor and analyse the speech quality of VoIP services. These include Perceptual Speech Quality Measure (PSQM). This article explains the technical principles of these methodologies, shows how they are applied in network assessment in private network prior to the IP telephony deployment by the Siemens network engineering and routing tool. This assessment verified that the solution is capable of delivering acceptable voice quality across the network from end points within all important office locations.

Keywords: *VoIP, voice, quality, PSQM, E-model, QoS measurement*

1. INTRODUCTION TO VOIP

VoIP is the name of a new communications technology that changes the meaning of the phrase telephone call. VoIP stands for Voice over Internet Protocol, and it means “voice transmitted over a computer network.” Internet Protocol (IP) networking is supported by any type of network - corporate, private, public, cable, and even wireless networks - not just the Internet. The corporate sector usually prefers private dedicated networks. You can access your account on the VoIP network via a desktop telephone, a wireless IP phone (similar to a cell phone), or the soft screen dial pad of your laptop or desktop computer. With this setup, you can literally pick up your things and move to another location without having to forward your calls to a new telephone. What’s more, you can access the Web from your IP phone, enabling you to get announcements and e-mail on the go. As you can imagine, VoIP is a win-win for everyone. The added flexibility and quicker response times translate into greater customer satisfaction and increased productivity throughout your organization.

2. QUALITY OF SERVICE

Quality of service (QoS) is the collective effect of service performances, which determine the degree of satisfaction of a user of the service [1]. The term quality of service (QoS) is extensively used today, not just in the telecommunications world in which it has its roots, but increasingly regarding broadband, wireless and multimedia services that are IP-based. A typical user is not concerned with how a particular service is implemented. However, the user is interested in comparing the same service offered by different providers in terms of universal, user-oriented performance parameters. This implies that performance should be expressed by parameters that [2]:

- Take into account all aspects of the service from the user's point of view.
- Focus on user-perceivable effects, rather than their causes within the network.
- Are independent of the specific network architecture or technology.
- Can be objectively or subjectively measured at the service access point.
- Can be easily related to network performance parameters.
- Can be assured to a user by the service providers.

3. DEFINITION OF CATEGORIES OF SPEECH TRANSMISSION QUALITY

The erlang (symbol E) as a dimensionless unit is used in telephony as a statistical measure of offered load. It is named after the Danish telephone engineer A. K. Erlang, the originator of traffic engineering and queueing theory. Traffic of one erlang refers to a single resource being in continuous use, or two channels being at fifty percent use each, and so on, pro rata. For example, if an office had two telephone operators who are both busy all the time, that would represent two erlangs (2 E) of traffic, or a radio channel that is occupied for thirty minutes during an hour is said to carry 0.5 E of traffic.

Alternatively, an erlang may be regarded as a "use multiplier" per unit time, so 100% use is 1 E, 200% use is 2 E, and so on. For example, if total cell phone use in a given area per hour is 180 minutes, this represents 180/60 = 3 E. In general, if the mean arrival rate of new calls is λ per unit time and the mean call holding time is h, then the traffic in erlangs E is:

$$E = \lambda h$$

This may be used to determine if a system is over-provisioned or under-provisioned (has too many or too few resources allocated).

Traffic measured in erlangs is used to calculate grade of service (GoS) or quality of service (QoS).

The mean of opinion (MOS) scores, i.e., of the values on a predefined scale that subjects assign to their opinion of the performance of the telephone transmission system used either for conversation or for listening to spoken material. The abbreviation MOS (Mean Opinion Score) is defined in [3]. A MOS can range from 5 down to 1.

While the parameters mentioned above describe the individual factors affecting speech transmission quality, it is the combined effect of all parameters together which leads to the overall level of speech transmission quality as perceived by the user. For transmission planning purposes, the E-model [4] is a useful tool for assessing the combined effect of all parameters and hence differentiating between categories of speech transmission quality.

The primary output of the E-model is the Transmission Rating Factor, R. Table 1 gives the definitions of the categories of speech transmission quality in terms of ranges of Transmission Rating Factor R provided by Recommendation G.107. Also provided are descriptions of "User satisfaction" for each category.

Table 1 ITU-T G.107 – Provisional guide for the relation between R-value and user satisfaction

***R*-value (lower limit)**	**MOS_{CQE} (lower limit)**	**User satisfaction**
90	4.34	Very satisfied
80	4.03	Satisfied
70	3.60	Some users dissatisfied
60	3.10	Many users dissatisfied
50	2.58	Nearly all users dissatisfied

It is very important to fully understand the principle recommended in this Recommendation. The R-value is a measure of a quality perception to be expected by the average user when communicating via the connection under consideration: quality is a subjective judgement such that assignments cannot be made to an exact boundary between different ranges of the whole quality scale.

4. VOIP IMPAIRMENTS

Voice quality is directly affected by three major factors:

- Jitter
- Packet Loss
- Delay

4.1. Jitter

Jitter is the difference between the ideal packet-arrival time and the actual packet-arrival time averaged over multiple packets. Any single measurement between two packets is the interpacket delay and should not be confused with jitter, which must be calculated in accordance with the RTP protocol specification [5].

4.2. Packet loss

Although IP networks are equipped to retransmit dropped packets, a common problem in data transmissions, this practice is inappropriate for RTP packets containing encoded voice information. By the time a retransmitted RTP packet arrives, it is far too late to be of any use in reconstructing the voice signal. For this reason, it is very important that providers configure their networks to prevent them from dropping RTP packets. RFC 1889 identifies RSVP, the Resource Reservation Protocol, as the method for ensuring that RTP packets have guaranteed bandwidth.

4.3. Delay

Delays are more prevalent on packet based networks than on circuit switched networks. And end-to-end delays, caused by the accumulation of delays from every routing and switching device in the network, become even more pronounced with CODEC. When delay becomes excessive, normal conversation becomes difficult.

Table 2 Typical CODEC delays

CODEC	**G.711**	**G.729A**	**G.723.1**
Type	PCM	CS-ACELP	MPC-MLQ&ACLEP
Quality	Toll	Toll	< Toll
Bit Rate	64 Kbits/s	8 Kbits/s	6,3 & 5,3 Kbits/s
Frame Size	0,125 ms	10 ms	30 ms
E-to-E Delay	0,25 ms	25 ms	67,5 ms

5. E2E QOS MEASUREMENT AND EVALUATION

Objective methods refer to those algorithms carried out by machines without the involvement of human listeners. These methods have been developed to substitute for the subjective MOS method. Objective methods can be classified into two categories, intrusive or non-intrusive, based on whether a reference speech signal is employed during the test or not.

5.1. Intrusive methods

During intrusive tests, a reference signal is injected into the network under test, and this is used to compare with the signal that is being tested to produce a MOS rating. The algorithms that are widely used today are perceptual in that they transform speech signals into the psychoacoustic relevant domain such as bark spectrum or loudness, and then incorporate human auditory models to yield a MOS rating. The best-known measures of this sort include Perceptual Speech Quality Measure, Perceptual Analysis/Measurement System, Perceptual Evaluation of Speech Quality.

PSQM - This method was developed by J. Beerends and J. Stemerdink of KPN Research, The Netherlands [6], and standardised by ITU-T as Rec. P.861 in 1998.

PAMS - This was based on the model introduced by M. Hollier of the PsyTechnics group at British Telecommunications [7]. PAMS is the first model designed for robust end-to-end speech quality assessment in telephone-band transmission systems.

PESQ - The strengths of PSQM99 (an updated version of PSQM which was released in 1999) and PAMS were combined into a new measurement algorithm called Perceptual Evaluation of Speech Quality. It was standardised [8] in 2001 and it made PSQM obsolete. PESQ retains the perceptual model of PSQM99 and the variable delay estimation of PAMS and adds new methods for transfer function equalisation and averaging distortion over time.

5.2. Non-intrusive methods

In this category, no reference signal is needed. Instead, nonintrusive methods are based on an analysis of degraded speech signals, voice packet header information or voice and network parameters. Non-intrusive methods are quite challenging because of the absence of a reference signal. Leading non-intrusive methods include:

- Voice payload processing approaches like the ITU-T Rec. P.563 [9] and the Auditory Non-Intrusive Quality Estimation plus (ANIQUE+).
- Parameter models which is also known as the E-model.

ITU-T Rec. P.563 was standardised in 2004 and is based on three algorithms: Non-intrusive speech Quality Assessment (NiQA), which was developed by Psytechnics Ltd. of the UK, the Non-intrusive Network Assessment (NiNA) developed by SwissQual Inc, Switzerland, and Perceptual Single Sided Speech Quality Measure (P3SQM) from Opticom GmbH, Germany. ITU-T Rec. P.563 is the first ITU-T ingleended method for predicting the subjective quality in arrowband telephone networks.

The E-model. The complexity of modern networks requires that for transmission planning the many transmission parameters are not only considered individually but also that their combination effects are taken into account. This can be done by "expert, informed guessing," but a more systematic approach is desirable, such as by using a computational model. The output from the model described here is a scalar quality rating value, R, which varies directly with the overall conversational quality.

The E-model is based on the equipment impairment factor method, following previous transmission rating models. It was developed by an ETSI ad hoc group called "Voice Transmission Quality from Mouth to Ear".

The reference connection, as shown in Figure 1, is split into a send side and in a receive side. The model estimates the conversational quality from mouth to ear as perceived by the user at the receive side, both as listener and talker.

Speech transmission quality is an important aspect of quality-of-service for many user applications of many telecommunications services. Recommendation ITU-T P.11 [10] identifies the key speech quality parameters and gives the subjective effects of variations in the parameters. Examples of speech quality parameters are speech level, attenuation distortion, transmission delay, echo path loss and delay, circuit noise, background noise, nonlinear distortion (such as the effects of low bit-rate speech codecs, packet loss, etc) and terminal characteristics.

The transmission parameters used as an input to the computation model are shown in Figure 1. Values for room noise and for the D-factors are handled separately in the algorithm for send side and receive side and may be of different amounts. The parameters SLR, RLR and circuit noise Nc are referred to a defined 0 dBr point. All other input parameters are either considered as values for the overall connection such as OLR (in any case the sum of SLR and RLR), number of qdu, equipment impairment factors Ie and advantage factor A, or referred only to the receive side, such as STMR, LSTR, WEPL (for calculation of Listener Echo) and TELR.

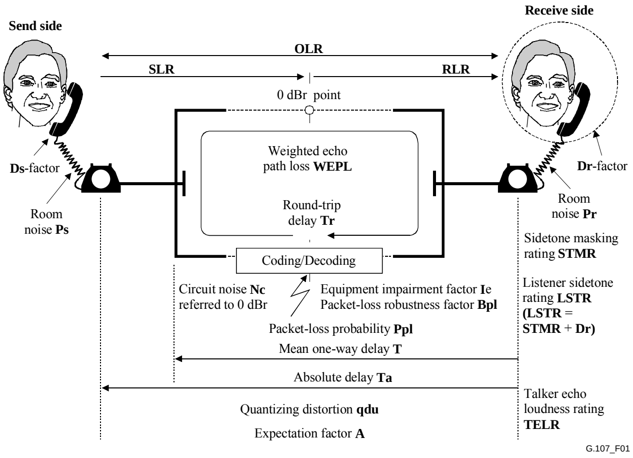

Fig. 1 ITU-T G.107 – Reference connection of the E-model

There are three different parameters associated with transmission time. The absolute delay Ta represents the total one-way delay between send side and receive side and is used to estimate the impairment due to too-long delay. The parameter mean one-way delay T represents the delay between the receive side (in talking state) and the point in a connection where a signal coupling occurs as a source of echo. The round-trip delay Tr only represents the delay in a 4-wire loop, where the "double reflected" signal will cause impairments due to Listener Echo.

5.3. Calculation of the transmission rating factor, R

The result of any calculation with the E-model in a first step is a transmission rating factor R, which combines all transmission parameters relevant for the considered connection. This rating factor R is composed of:

$$R = Ro - Is - Id - Ie\text{-}eff + A$$

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.

The E-model is a conversational model that covers the effects of delay and echo, but it is a parametric, computational model rather than a measurement tool. The MOS derived from the E-model can be regarded as the average voice quality under a certain condition.

6. MEASUREMENT

In general terms, auditing and testing of IP telephony should be done with two objectives in mind: pre-deployment auditing, which needs to establish all impacts on the existing corporate environment from the designed IP telephony solution, and post-deployment, which needs to ensure that the deployed IP telephony solution has acceptable performance in terms of reliability and voice quality (this should be a key component of general acceptance testing).

6.1. Pre-deployment auditing –Network Voice Quality Assessment

A network voice quality assessment is an assessment of how well the network can transport voice traffic with acceptable voice quality. Voice quality assessments must be performed before deployment as part of the general network assessment, but it should, in particular, be done as part of the final acceptance testing to verify the performance of the IP telephony solution after it has been implemented.

Although predeployment tests should be used to complement the network topology assessment and the network equipment assessment, the final acceptance testing should be comprehensive and should focus on documenting that the completed solution can support the enterprise telephony needs under normal- and busy-hour situations, as well as a stress test behavior during worst-case situations.

Prior to deployment, enterprises should also consider performing network and traffic simulations to predict voice quality and to ensure that the additional voice traffic and network configuration will not negatively impact the performance of existing applications.

In SPP have been used network engineering and routing tool by Siemens which can simulate voice traffic on the network. This product generates calls from selected points throughout the network, and is received by a server placed at a central point in the network. Main features of the tool are:

- This tool generates synthetic voice calls using an appropriate voice codec.
- Calculates voice quality in a mean opinion score (MOS) per call, and offer detailed statistics of minimum and maximum MOS for each tested network path related to the number of calls.
- Monitors network latency, jitter and packet loss per call, and offer detailed statistics of minimum and maximum MOS for each tested network path related to the number of calls.
- Monitors network metrics within devices in each call path to assist with trouble resolution of any voice quality issues.
- Visualisation of results in network view, routing paths, node load, bandwidth demands, channel sizes.
- Creates customized reports with visual presentations of performance metrics and use levels.

7. RESULTS

Network and call metrics were measured by month in private network during real operations. The 5353 IP phones were simulated in 65 local area networks with codec G.711. Number of concurrent calls has been generated on base of the real experience in everyday operations.

The following average results were obtained:

- Average delay 17,12 ms
- Packet loss 0,22%
- R factor 85
- MOS 4,2

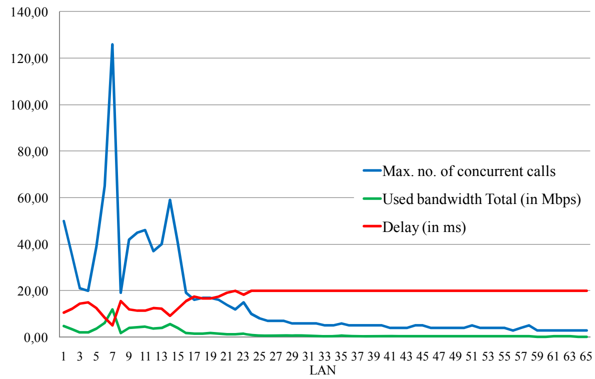

Fig. 2 Relation measured among number of concurrent calls, used bandwidth and delay

Table 1 Measured maximal values of bandwidth and delay

LAN	A	Max. no. of concurrent calls	Used bandwidth Total (in Mbps)	Delay (in ms)
1	ipNetwork	50,00	4,8327	10,53
2	ipNetwork	35,00	3,39828	12,15
3	ipNetwork	21,00	2,01773	14,39
4	ipNetwork	20,00	1,90916	14,94
5	ipNetwork	39,00	3,77601	12,39
6	ipNetwork	65,00	6,17921	8,27
7	ipNetwork	126,00	12,0807	5,09
8	ipNetwork	19,00	1,84395	15,49
9	ipNetwork	42,00	4,02873	11,91
10	ipNetwork	45,00	4,32685	11,46
11	ipNetwork	46,00	4,42167	11,27

12	ipNetwork	37,00	3,56895	12,45
13	ipNetwork	40,00	3,86028	12,28
14	ipNetwork	59,00	5,68543	9,07
15	ipNetwork	41,00	3,93257	12,15
16	ipNetwork	19,00	1,77271	15,49
17	ipNetwork	16,00	1,47868	17,50
18	ipNetwork	17,00	1,57785	16,67
19	ipNetwork	17,00	1,65144	16,67
20	ipNetwork	16,00	1,53713	17,50
21	ipNetwork	14,00	1,30373	19,06
22	ipNetwork	12,00	1,14173	20,00
23	ipNetwork	15,00	1,4059	18,34
24	ipNetwork	10,00	0,956742	20,00
25	ipNetwork	8,00	0,700439	20,00
26	ipNetwork	7,00	0,672508	20,00
27	ipNetwork	7,00	0,672508	20,00
28	ipNetwork	7,00	0,626474	20,00
29	ipNetwork	6,00	0,554178	20,00
30	ipNetwork	6,00	0,554178	20,00
31	ipNetwork	6,00	0,529737	20,00
32	ipNetwork	6,00	0,509544	20,00
33	ipNetwork	5,00	0,480722	20,00
34	ipNetwork	5,00	0,480722	20,00
35	ipNetwork	6,00	0,529737	20,00
36	ipNetwork	5,00	0,388807	20,00
37	ipNetwork	5,00	0,388807	20,00
38	ipNetwork	5,00	0,388807	20,00
39	ipNetwork	5,00	0,448722	20,00
40	ipNetwork	5,00	0,402511	20,00
41	ipNetwork	4,00	0,356672	20,00
42	ipNetwork	4,00	0,356672	20,00
43	ipNetwork	4,00	0,356672	20,00
44	ipNetwork	5,00	0,402511	20,00
45	ipNetwork	5,00	0,402511	20,00
46	ipNetwork	4,00	0,323539	20,00
47	ipNetwork	4,00	0,323539	20,00
48	ipNetwork	4,00	0,323539	20,00
49	ipNetwork	4,00	0,323539	20,00
50	ipNetwork	4,00	0,356672	20,00
51	ipNetwork	5,00	0,388807	20,00
52	ipNetwork	4,00	0,301194	20,00
53	ipNetwork	4,00	0,323539	20,00
54	ipNetwork	4,00	0,323539	20,00
55	ipNetwork	4,00	0,323539	20,00
56	ipNetwork	3,00	0,28669	20,00
57	ipNetwork	4,00	0,323539	20,00
58	ipNetwork	5,00	0,402511	20,00
59	ipNetwork	3,00	0,232088	20,00
60	ipNetwork	3,00	0,232088	20,00
61	ipNetwork	3,00	0,28669	20,00
62	ipNetwork	3,00	0,28669	20,00
63	ipNetwork	3,00	0,28669	20,00
64	ipNetwork	3,00	0,232088	20,00
65	ipNetwork	3,00	0,232088	20,00

Results obtained by this sample measurement show acceptable VoIP quality.

8. CONCLUSIONS

The introduction of Internet Protocol (IP) telephony impacts the enterprise data network. Despite numerous discussions of this, some enterprises still do not perform proper predeployment assessments of the enterprise network. This often leads to unplanned additional costs or deployment delays. Properly performed network assessments are mandatory to reduce this risk, and this research outlines how to do these assessments.

Testing the deployed solutions ability to transport voice with an acceptable voice quality as experienced by the user should verify network characteristics in terms of latency, jitter and loss, and should also specifically measure voice quality in an MOS or R-factor as it best relates to end-user perception of voice quality. Generally, an MOS value above four is acceptable. In this case we measured average MOS 4.2. This test verified that the solution is capable of delivering acceptable voice quality across the network from end points within all important office locations.

REFERENCES

[1] ITU-T: Recommendation E.800: Terms and definitions related to quality of service and network performance including dependability, 2008

[2] ITU-T: Recommendation G.1010: End-user multimedia QoS categories, 2001

[3] ITU-T: Recommendation P.800.1: Mean Opinion Score (MOS) terminology, 2006

[4] ITU-T: Recommendation G.107: The E-model, a computational model for use in transmission planning, 2008

[5] IETF: RTP: A Transport Protocol for Real-Time Applications, 1996

[6] J. Beerends and J. Stemerdink, “A perceptual speech-quality measure based on psychoacoustic sound presentation,” J. Audio Eng. Soc.,vol. 42, no. 3, pp.115-123, March 1994.

[7] A. Rix and M. Hollier, “The perceptual analysis measurement system for robust end-to-end speech quality assessment,” in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, June 2000, vol. 3, pp. 1515-1518

[8] ITU-T: Recomendation 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, 2001

[9] ITU-T: Recommendation P.563: Single-ended method for objective speech quality assessment in narrow-band telephony applications, 2004

[10] ITU-T: Recommendation P.11: Effect of transmission impairments, 1993

Received November 26, 2009, accepted January 16, 2010

BIOGRAPHY

Peter Chochol was born on 3.5.1962. In 1987 he graduated (MSc) with distinction at the department of Electronics and Multimedia Communications of the Faculty of Electrical Engineering and Informatics at Technical University in Košice. Since 2007 he is working as Director of IT Development in Slovak Gas Industry. His scientific research is focusing on QoS VoIP.