Impact of the codec and various QoS methods on the final quality of the transferred voice in an IP network

Oldřich Slavata[1,2], Jan Holub[1]

[1] Czech Technical University, Faculty of Electrical Engineering, Dept. of Measurement K13138, Technicka 2, 166 27 PRAGUE 6, Czech Republic

[2] Czech Technical University in Prague, University Center for Energy Efficient Buildings, Nam. Sitna 3105, Kladno, Czech Republic

slavao1@fel.cvut.cz

Abstract. This paper deals with an analysis of the relation between the codec that is used, the QoS method, and the final voice transmission quality. The Cisco 2811 router is used for adjusting QoS. VoIP client Linphone is used for adjusting the codec. The criterion for transmission quality is the MOS parameter investigated with the ITU-T P.862 PESQ and P.863 POLQA algorithms.

1.   Introduction

At the beginning of the 21st century, increasing transmission capacity of the network and improved methods for digital processing of video and acoustic signals allowed the Internet to be used for real-time voice and video communication. VoIP (Voice over Internet Protocol) allows the transmission of voice in digital form in UDP / TCP / IP packets. Networks that are used for simultaneous data, voice or video transfer are known as Converged Networks. Before networks converged, protocols were developed to occupy most of the available network capacity and to deliver data as fast as possible. This can lead to packet loss, or to receiving the packet in the wrong order. Lost packets are then resent. In the end point, all packets are re-grouped in the correct order.

For voice transmission, it is important that voice packets are delivered with minimum delay, with little variability of delay, and in the correct order. To some extent, we can accept loss or dropping of some packets in return for low delay and increased fluency. Different transmission requirements have led to the need to handle simultaneous transmission of voice and data (or video) over a single network. The simplest way to solve these problems is to increase the network bandwidth so that all protocols have sufficient capacity. However, this method is expensive and is not always realizable (because of some technological limits). Another option is to introduce QoS classes and to prioritize network traffic according to its importance. Typical classes are:

Voice - Class for voice transmission.

Business-Critical - Class for important data (business applications, access to databases, etc.)
Best-Effort - Class for normal traffic (e-mail, web access, etc.)
Scavenger - Class for unwanted traffic that can also be inhibited

Voice and business-critical traffic should have sufficient bandwidth, and voice packets should additionally be sent as preferred. The following mechanisms are used to provide this capacity:

Classification - manually configured or automatic classification of packets / frames
Marking - specific values are written in the header of the packet / frame
Congestion Management - header tags are used in deciding on the class / queue to which the packet will be assigned
Congestion Avoidance - preventive dropping of packets is used to prevent congestion of the line
Policing and shaping - packets are delayed or dropped in traffic that has reached a defined limit
Link Efficiency Mechanisms - the efficiency of the line is increased increasing by using header compression and packet fragmentation

The MOS scale (mean opinion score) is used to assess the quality of voice transmission (Table 1). The term MOS is defined in Recommendation ITU-T P.800.

Table 1 MOS scale

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

There are several methods for obtaining MOS values. The most accurate is a subjective test, where the MOS value is obtained directly from users. It is time-consuming and expensive to carry out these subjective tests. Objective methods based on computer algorithms are therefore used instead.

Intrusive methods provide results that are near to the results of the subjective tests. They are based on comparing the original sample with the transferred sample. These algorithms use psychoacoustic models of human perception, seeking to provide a mathematical description of human sound perception and to find variables that have a direct impact on the perceived quality of the voice signal. Intrusive methods include PAMS (Perceptual Analysis Measurement System), developed by British Telecommunications, PSQM (Perceptual Speech Quality Measurement), described in Recommendation ITU-T P.861, PESQ (Perceptual Speech Quality Evaluation of), according to ITU-T P.862 (P.862.1) [1], [2], and newly ITU-T P.863 - POLQA (Perceptual Objective Listening Quality Analysis) [3], [4].

Non-intrusive methods are a different type of quality measurement. These methods do not use the reference signal, and the final MOS is calculated using the parameters of the transferred sample only. The disadvantage of these methods is their lower accuracy and reliability. An example of a non-intrusive method is 3SQM, which is defined in recommendation ITU-T P.563.

2. Description of the experiment

2.1. Test-bed
The test bed (Figure 1) consisted of two small LAN networks connected with a serial line simulating the WAN connection with bandwidth limited to 384 kbit/s. We used the following network elements: Cisco 7942G IP phones, Switches: Cisco Catalyst 3560, Routers: Cisco 2811, Cisco 1841, PC and servers: Dell.
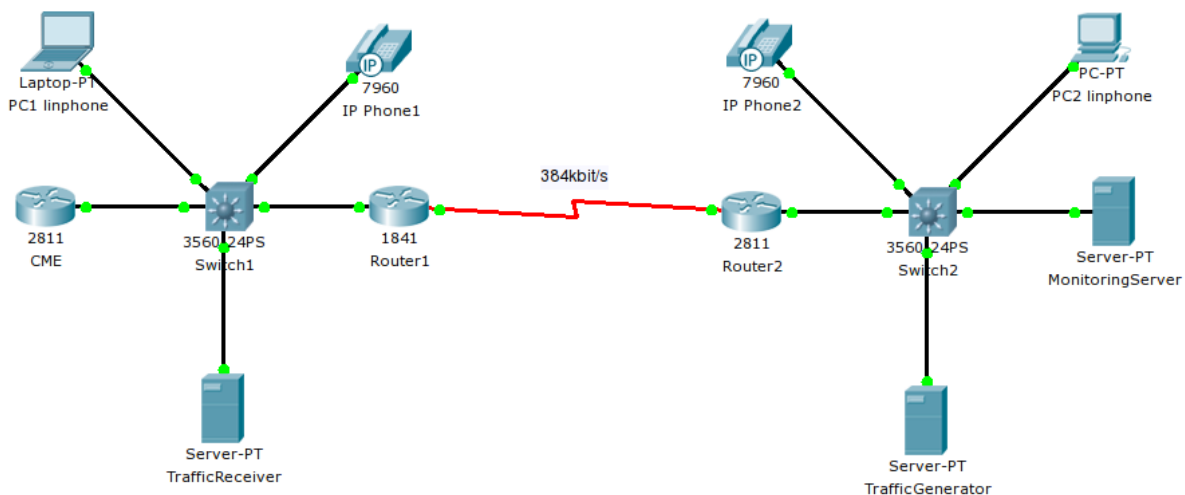
Figure 1. The test-bed

A concatenated speech file in WAV format (8kSa/S, 16bit), 16.75 s in length, was used (Figure 2). The file contains 4 short sentences spoken by 4 different speakers (two men, two women) and adequately covers the entire human speech spectra. The concatenated file that is used therefore effectively replaces tests using multiple speech samples.

Figure 2. Test signal

A VoIP telephone call was made between PC1 and PC2 using Linphone. On the side of the caller, the sample was directly played using Linphone. On the receiver side, it was recorded using Audacity. Various QoS methods were adjusted on Router2. One of the purposes of the experiment was to verify the reliability of the MOS estimates that are provided in Cisco IP phones. Test calls for each QoS setting were also made using these phones. To accelerate the experiment, the dialing and recording was controlled by a bash script. Softphone Linphone was chosen because it allows playback of a voice sample and script control. In previous experiments, it was evaluated as the best in terms of quality of transmission [5]. The following codecs were used in Linphone: G.711, G.722, G729, iLBC. The iperf traffic generator was used to simulate the real traffic in the network.

The PESQ and POLQA algorithms were used to evaluate the voice transmission quality. The output of the PESQ algorithm was recalculated to the value of MOS-LQO (Listening Quality Objective) according to a mathematical prescription defined in ITU-T P.862.1. As follows from the official wording of P.862, it can be used for testing the effect of packet loss on CELP coded transmission. It has not been tested for PCM transmissions affected by packet loss. However, the recommendation itself does not prevent any user from doing this, and it was successfully used in previous experiments [6,7,8,9]. The results can also be affected by Packet Loss Concealment (PLC), implemented by Linphone. Comparing the results of our previous experiment [6,7] with [10], it is obvious that our version of Linphone had no PLC implemented.

2.2. Methods for controlling and preventing congestion
FIFO – First In First Out
Handles packets in the same order as they come. For larger loads, the queue fills up quickly, and this causes a delay and packet losses. This is the simplest type of queue.

WFQ – Weighted Fair Queuing
Forms 16-256 queues according to the bandwidth. Used for classifying packets to queues using automatic classification.

CBWFQ – Class-based Weighted Fair Queuing
Extends WFQ to support traffic classes. For each class, a queue is formed in which all the traffic of this class is routed. Bandwidth is allocated to classes by priority.

LLQ – Low-latency Queuing
Adds a priority queue for real-time traffic (voice, video) to CBWFQ. Real time data is cleared first.

CB-WRED – Class-based Weighted Random Early Detection
Provides a RED mechanism that randomly drops TCP packets before a queue is full, TCP traffic slows down and there is no unsteady flooding of lines. Each class can be assigned a different RED profile.

ECN – Explicit Congestion Notification
Extends WRED by adding information about the network congestion to the packet header.

LFI – Link Fragmentation and Interleaving
Fragments long data packets in order to reduce the delay of voice packets.

3.   Results
Ten samples were measured and processed for each codec setting and QoS method. Ten repetitions are enough to achieve satisfactorily low dispersion results and uncertainties. The confidence intervals (CI95) were computed using the following formulae (1):

$$u_{Ax} = \tilde{\sigma}(\bar{x}) = \sqrt{\frac{1}{n \times (n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (1)$$

The graphs show a significant correlation between the QoS method and the final voice quality. Generally speaking, the more advanced the method is, the better the transmission quality will be. Some methods have the same results in transmission quality, but they differ in the length of the delay, represented by the ping response (figure 3.). The length of the delay is important for the quality of the conversation.
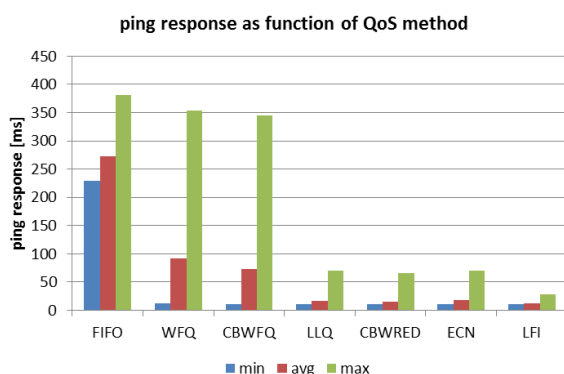

Figure 3. Ping response as a function of the QoS method

The following set of graphs (Figure 4 – 7.) shows the results for various codecs. The PESQ and POLQA algorithms produce very similar results. An estimate of MOS using an IP phone frequently differs from the results of specialized algorithms, but it is adequate for showing whether or not the call is possible.
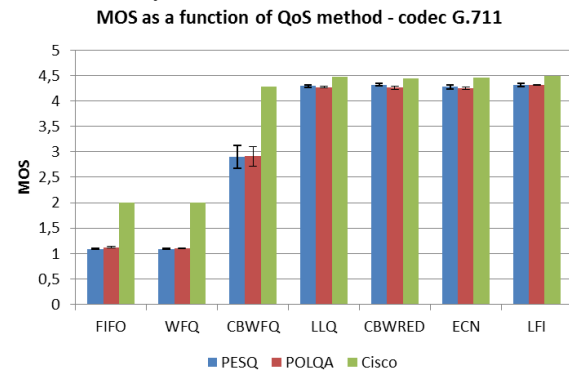
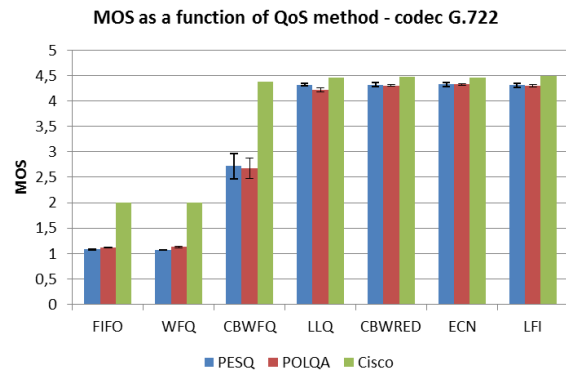Figure 4. MOS as a function of the QoS method - codec G.711

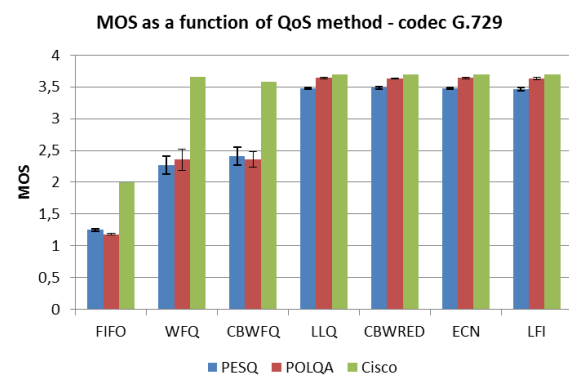Figure 5. MOS as a function of the QoS method - codec G.722

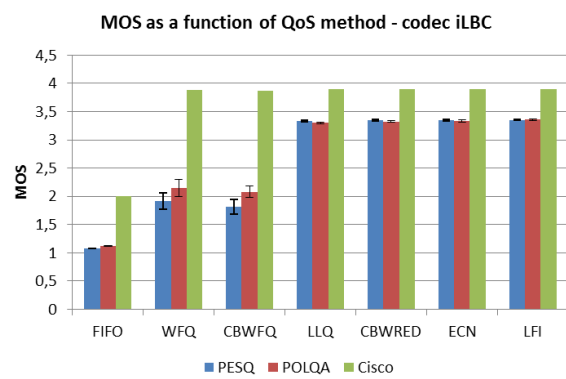Figure 6. MOS as a function of the QoS method - codec G.729

Figure 7. MOS as a function of the QoS method - codec iLBC

The next set of graphs (Figure 8 – 9.) shows the results sorted according to the algorithm that is used. The results show that high bitrate codecs G.711 and G.722 usually provide better transmission quality than low bitrate G.729 and iLBC. However, a low bitrate codec can provide better results when there is low bandwidth or a worse QoS method.
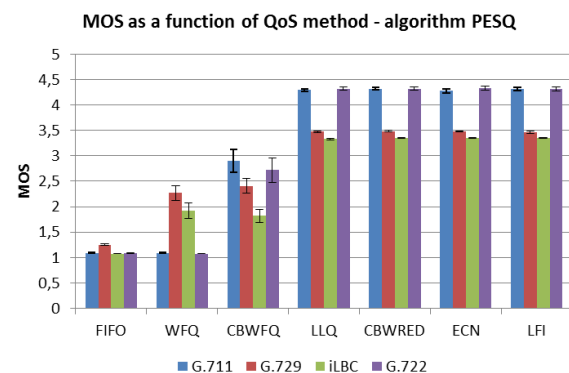

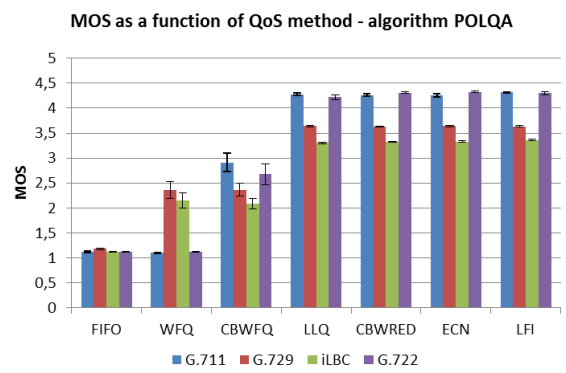Figure 8.   MOS as a function of the QoS method - PESQ algorithm

Figure 9. MOS as a function of the QoS method - POLQA algorithm

The results confirm the assumption that the more advanced the QoS method is, the better the transmission quality will be. A simple FIFO queue is not suitable for the purpose of traffic management in converged networks. The WFQ method offers a slight improvement for low bitrate codecs, but the resulting quality is still very bad. In the CBWFQ method, there is a noticeable

improvement even in high bitrate codecs. The LLQ method and others already provide transmission quality at the maximum limit for the current codec. However, they differ in the length of the response, and therefore in the delay and the conversation quality.

4.   Conclusion

This paper has checked the impact of the codec that is used and the QoS method on the quality of voice transmission in IP networks.

The main objective was to identify the relation between the QoS method and the MOS values as delivered by different objective algorithms. Generally speaking, the more advanced the method is, the better the quality of the transmission will be. Some advanced methods differ only in the delay.

The second objective was to explore differences between widely-used codecs. Low bitrate codecs provide worse results than high bitrate codecs, but they may be useful for low bandwidth.

The results of this experiment can be used in planning and implementing other measurements.

5.   Acknowledgment

References

[1]   ITU-T 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, February 2001

[2]   ITU-T P.862.1 Mapping function for transforming P.862 raw result scores to MOS-LQO, November 2003

[3]   ITU-T P.863 Perceptual objective listening quality assessment, January 2011

[4]   POLQA Technical White Paper, July 2010, Opticom

[5]   Slavata, O.:Měření kvality přenosu hlasu pro sítě typu VoIP, Bachelor project, FEE CTU, August 2007

[6]   Slavata, O.: Neintruzivní měření kvality přenosu hlasu v telekomunikačních sítích, Diplomová práce, Master's thesis, CTU FEE, January 2010

[7]   Slavata, O., Holub, J.: Evaluation of objective speech transmission quality measurements in packet-based networks, Computer Standards & Interfaces 36 (2014)

[8]   Počta, P., Holub, J.: Predicting the Quality of Synthesized and Natural Speech Impaired by Packet Loss and Coding Using PESQ and P.563 Models

[9]   Stein, Y., Druker I.: The Effect of Packet Loss on Voice Quality for TDM over Pseudowires, RAD Data Communications, October 2003

[10]   Gündüzhan E., Momtahan K.: A Linear Prediction Based Packet Loss Concealment Algorithm for PCM Coded Speech, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, November 2001

[11]   Ditech networks: Limitations of PESQ for Measuring Voice Quality in Mobile and VoIP Networks, December 2007

[12]   http://www.voip-info.org

[13]   http://www.itu.ch