Marek Repcik *

VOIP TRANSMISSION QUALITY INCREASE USING ADAPTIVE PLAY-OUT BUFFER

In this paper we consider some quality aspects of VoIP (Voice over Internet Protocol) transmission. We try to increase the voice transmission quality using the adaptive play-out buffer instead of common buffer at the receiver's side. The hypothesis is that the less packet-loss probability increases the transmission quality more that the additional jitter decreases it. The aim of this paper is to formulate the problem.

1. Introduction

In this paper we consider some quality aspects of voice transmission over the Internet. Voice transmission over the Internet, or IP telephony, has been taking over by the analogue telephony because of its cheaper operation.

1.1 Problem

VoIP transmission consists of several steps described in ITU-T Recommendations. The most important steps are: equidistant voice sampling, coding, compression, equidistant packetization, transmission that violates the inter-packet equidistance, buffering, decompression, and finally playing out, Fig. 1, and Fig. 2 respectively.

Each step that has been performed at the voice processing decreases the quality of the transmission.

The main consequences of decrease of the transmission quality are: packet delay, jitter and packet loss.

Scientists have already investigated each single segment of the information chain in order to decrease the consequences of the
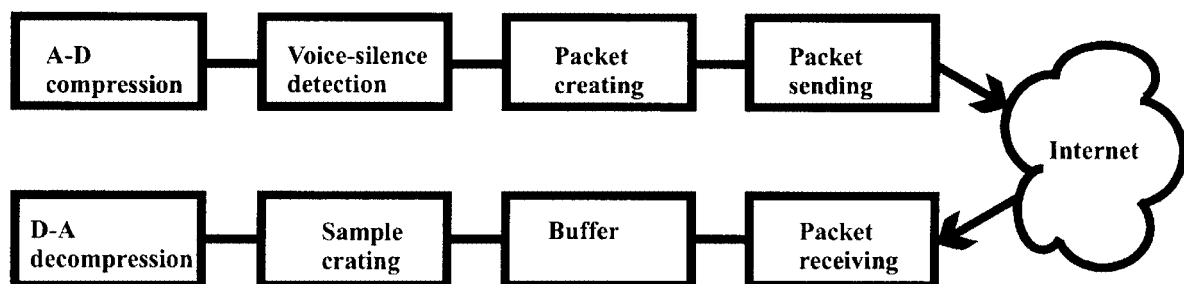

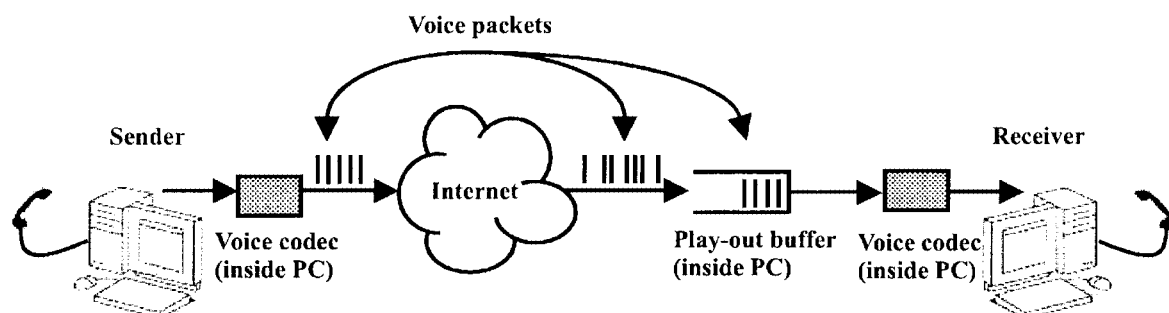*Fig. 1 Information chain of VoIP*

Fig. 2 Packet transmission over Internet.

* Marek Repcik
Department of Information Networks, Faculty of Management Science and Informatics, University of Žilina, Slovakia,
E-mail: pi@frcatel.fri.utc.sk

segment's negative influences. We focused to the buffer segment believing that our idea had not been studied yet.

What is a buffer? A buffer is an output memory, where the incoming packets have been stored until they can be preceded to be played out equidistantly at the receiver's side. What is the main task of a buffer?

1. A buffer helps to eliminate the random deflections (jitter) between the incoming packets that have appeared during the transmission process. Remember that the voice has been sampled in equidistant intervals at the sender's side and therefore it is reasonable to expect the equidistance at the receiver's side too.

2. A buffer decreases the packet loss.

How can we increase the voice transmission quality paying attention to the buffer segment? The way is – an adaptive play-out buffer.[1]

What is an adaptive play-out buffer? It is a buffer, in which the incoming packets have been stored until they can be preceded to be played out at the receiver's side not equidistantly but adaptively.[2] It means that the packet processing at the receiver's side has been performed once faster and once slower depending on the buffer load. When the buffer is fuller, the service intervals are shorter tending to empty the buffer and prevent the buffer overload and vice versa, when the buffer is emptier, the service intervals are longer tending to load the buffer and prevent the buffer underload.

What is a difference between a common buffer and an adaptive play-out buffer?

1. An adaptive play-out buffer generally creates a jitter because the adaptive play-out violates the inter-packet equidistance. The jitter causes a noise. This is a disadvantage of the adaptive play-out buffer.

2. There exists such an adaptive play-out buffer that produces less packet-loss than a common buffer. This is an advantage of the adaptive play-out buffer.

Because the common buffer is a special case of an adaptive play-out buffer, for sure there is an adaptive play-out buffer that causes at least the same voice transmission quality as the common buffer!

We should say something about other researchers dealing with adaptive play-out buffers. In [1] an adaptive play-out of packet voice is considered. The authors try to find out how much of a voice spurt should be buffered. They do not consider their adaptive play-out buffer the same way as we do. They are constricted only to a question how much the beginning of a voice spurt should be buffered (delayed) otherwise their buffer works as a common buffer. In [2] the approach is the same. Adaptive techniques perform

continuous estimation of the network delays and dynamically adjust the play out delay at the beginning of each talk spurt. The play-out adjustment is performed during silent periods between talk spurts. The adjustment is done on the first packet of the talk spurt; all packets in the same talk spurt are scheduled to play out at fixed intervals following the play-out of the first packet. Similar approach is also in [3] and [4]. So these are definitely different approaches from ours. We continue and develop an approach mentioned in [5].

We can formulate the problem:

• to find the best set-up of an adaptive play-out buffer parameters such that the less packet-loss increases the transmission quality more than the additional jitter decreases it. Then the overall effect is a quality increase. What does it mean 'set-up of an adaptive play-out buffer'? It means a determination of service time lengths for particular buffer loads.

• to find out how much such an innovation (using an adaptive buffer) increases the voice transmission quality.

2. Way of Advancing

In order to find out the answers we may divide the problem into four steps:

1. to define the transmission quality function,
2. to make a mathematical model of voice transmission using the adaptive buffer,
3. to incorporate the model results into the quality function,
4. to optimize the buffer set up through maximizing the quality function through its parameters.

Let's say something more about each of these steps.

3. Voice Transmission Quality

To define and measure the quality of VoIP service we follow the ITU-T Recommendations. To determine the quality it is necessary to determine the relation amongst the QoS (*Quality-of-Service*) parameters and NP (*Network-Performance*) parameters.

The last recommendations G.175, G.107 and G.109 prefer expressing the NP parameters influence with the Transmission Rating Factor R, which is increasing with increasing quality. For its evaluation so called E-model is used. The E-model is based on the assumption that the all impairment influences can be transformed into the psychological factors whose influences are additive.

3.1 E-model

According to the recommendations ITU-T G.175 and ITU-T G.109 Definitions of Grades of Quality all influences at the voice

---

[1] Adaptive play-out buffer in our meaning is significantly different from until now used adaptive play-out buffers but with its nature it still can be considered to be an adaptive play-out buffer.

[2] We said that it is reasonable to expect the equidistance at the receiver's side but as we'll see later it's not exactly like that. Sorry for that:-).

transmission are expressed aggregately by the transmission rating factor R, which reaches the values in the interval $\langle 0, 100 \rangle$, where $R = 100$ means a very high quality and $R = 0$ means a very bad quality. Transmission with the parameter less than 50 is not recommended; the rest of the values is uniformly distributed in 5 classes.

3.1.1 Pre-Defined Values of Parameters and Acceptable Ranges

According to the recommendation ITU-T G.107 there are the default values of the input parameters of the E-model. There is a strict recommendation to accept these values for those parameters, which are not going to be changed during the planning process. If all of the default values are applied, the results lead to a very high value of the transmission rating factor $R = 94.1$.

See illustration of the general reference connection of the E-model in Fig. 3.

3.1.2 Calculation of the Transmission Rating Factor R

According to the equipment impairment factor method, the fundamental principle of the E-Model is based on a concept that all psychological factors on the psychological scale are additive. The result of any calculation with the E-Model in the first step is a transmission rating factor R, which combines all transmission parameters relevant for the considered connection, according to the recommendations ITU-T G.107 and ITU-T G.175. This rating factor R is composed of:

$$R = Ro - Is - Id - Ie + A \qquad (1)$$

Ro represents 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 equipment impairment factor Ie represents impairments caused by low bit rate codecs. The advantage factor A allows compensation of impairment factors when there are other advantages of access to the user.

Using the adaptive play-out buffer just the parameters Is and Id will be touched. Why? As said before the factor Is is a combination of all impairments, which occur more or less simultaneously with the voice signal. The adaptive play-out buffer causes an additional jitter that causes a noise. The factor Id will be touched too, as it represents the impairments caused by delay. It is clear that for different adaptive buffer set-ups different delays of packets in buffer are expected.

3.1.3 Case of more impairment sources

The problem may be how to incorporate the adaptive play-out's influence to the Is impairment. We suggest to incorporate it to the Q parameter of the Is impairment as another impairment source, which is a result of noise created by the adaptive play-out.

Let $s(t)$, $t \in R$ be the original signal with the mean power σ_s^2, let $n_1(t)$, ..., $n_M(t)$ are independent noises with the mean powers $\sigma_{n_1}^2$, ..., $\sigma_{n_M}^2$. Then the overall signal-to-noise ratio Q is:
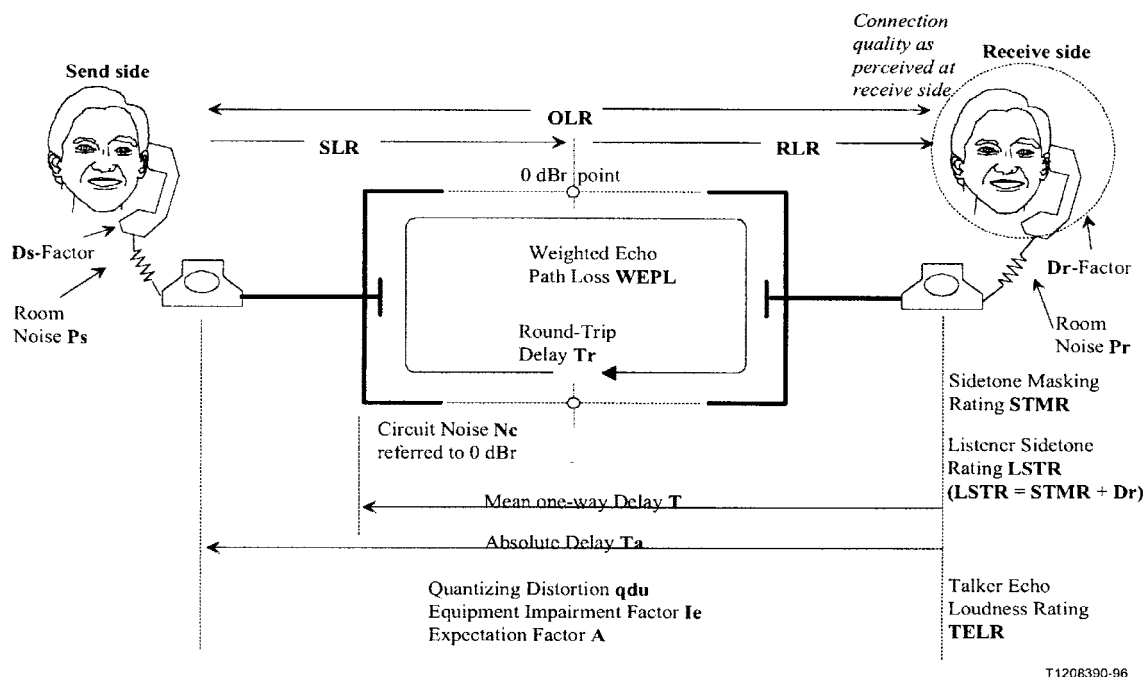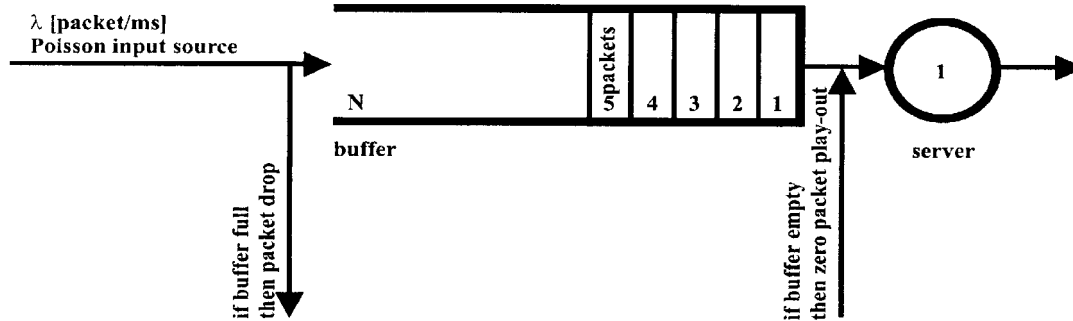

Fig. 3 General reference connection of the E-model.

Fig. 4 Stochastic model $M\,|\,D_k\,|\,1\,|\,N$ for adaptive packet play-out process.

$$Q = 10\log \frac{\sigma_s^2}{\sum\limits_{i=1}^{M} \sigma_{n_i}^2} = 10\log \sum\limits_{i=1}^{M} 10^{-\frac{Q_i}{10}} \qquad (2)$$

where $Q_i \overset{def}{=} 10\log \dfrac{\sigma_s^2}{\sigma_{n_1}^2}$, $i = 1, 2, ..., M$.

According to ITU-T P.11, Annex E, for the overall signal-to-noise ratio it is recommended to use the 'correlated signal-to-noise ratio' Q, which considers the subjective factors and recommends combining the values Q_i on the $15\log_{10}$ base, i.e.:

$$Q_i \overset{def}{=} -15\log \sum\limits_{i=1}^{M} 10^{-\frac{Q_i}{15}} \qquad (3)$$

In our case, if Qjl is the signal-to-noise ratio, where the noise is caused by the jitter of the adaptive buffer play-out and the packet-loss, i.e. $Qjl \overset{def}{=} 10\log \dfrac{\sigma_s^2}{\sigma_{njl}^2}$

where $njl(t)$ is the noise caused by the adaptive buffer play-out and packet-loss, then:

$$Q = -15\log \left[10^{-\frac{37}{15}} qdu + 10^{-\frac{Qjl}{15}} \right] \qquad (4)$$

4. Transmission Model

We need to model the voice transmission over the Internet using a stochastic model of bulk service.

Using the state probabilities of the model we should be able to determine:

- the noise caused by the jitter being caused by the network and by the adaptive play-out buffer,
- the delay of a packet in the adaptive buffer.

The model should describe:
- playing the packets out adaptively depending on the buffer load,
- refusing packets if the buffer is full,
- playing so called zero packets in a case that packets are late (if they haven't arrived to the time of their supposed playing-out). Playing a zero packet means that we play an artificial packet that contains samples of zero signal – silence, as it is preformed in reality.

We suggest a stochastic model $M\,|\,D_k\,|\,1\,|\,N$ where the input source is the Poisson process, the buffer capacity is equal to N and there is one server working deterministically and adaptively depending on the buffer load, see Fig. 4. The request of modelling the zero packet playing-out could be modelled by the negative states of the system. We realize that the Poisson process is not a good approximation of an input packet source but it is quite easy to analyse and it represents the worst case of an input source.

5. Conclusion

We have proposed an adjustment of the voice transmission over the Internet using an adaptive play-out buffer. We have suggested a way to model and determine whether and how much such adjustment is good and relevant.

References

[1] PINTO, J., CHRISTENSEN, K. J.: *An Algorithm for Playout of Packet Voice based on Adaptive Adjustment of Talkspurt Silence Periods,* Department of Computer Science and Engineering, University of South Florida, Tampa, Florida, 1999.

[2] NARBUTT, M., MURPHY, L.: *Adaptive Play-out Buffering for Audio/Video Transmission over the Internet,* Department of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland.

[3] ROSENBERG, J., QIU, L., SCHULZRINNE, H.: *Integrating Packet FEC into Adaptive Voice Play-out Buffer Algorithms on the Internet,* Cornell University, Columbia University, USA, INFOCOM 2000.

[4] FUJIMOTO, K., ATA, S., MURATA, M.: *Adaptive Play-out Buffer Algorithm for Enhancing Perceived Quality of Streaming Applications,* Osaka University, Osaka City University.

[5] KLIMO, M.: *Buffer Management for CBR Traffic,* University of Žilina, Communications – Scientific Letters of the University of Žilina, ISSN 1335-4205.