Interaction Mode Aware Playout Delay Adjustment Algorithm for Packetized Audio

Author H. Levonyan

PhD student of State Engineering University of Armenia,

(phone +374 91208577, fax +374 10442155),

hrant.levonyan@epygilab.am

Abstract

In packet audio applications, packets are buffered at a receiving side and their playout is delayed in order to compensate the effect of network jitter. In such applications the jitter buffer adjustment algorithms are used to estimate the playout delay of the packets. Conventional algorithms estimate playout delay considering the network behaviour only – they don't consider other criterias which can affect the perceptual voice quality. In this paper a new algorithm is proposed aimed to improve the voice quality considering not only network behavior but also the voice interaction mode. The results indicate that the proposed algorithm significantly improves the perceptual voice quality and can be used for extending the existing algorithms.

Keywords

VoIP, network jitter, playout buffer, playout delay, adaptive playout algorithm, perceptual voice quality

1 Introduction

Perceptual voice quality is the main important characteristic in VoIP. As far as IP networks were not designed to support real-time applications, there are many factors which significantly affect the perceived voice quality. From the user's perspective network delay, jitter and the packet loss are the most important factors in VoIP. The impact of jitter is typically reduced by using a delaying mechanism which queues irregularly arrived packets in the playout buffer and thus allows the slowest packets to arrive in time to be played out. The appropriate playout buffer organization is a very important task, which directly affects the perceptual voice quality. The buffer can be neither too big nor too small, because the first case can lead to the long playout delay and the second one may cause packet losses due to their late arrivals. Since packet receiving time is unknown and receiver doesn't know how to select appropriate playout times, playout buffer size can not be fixed. Thus an adaptive playout algorithm should be applied, which considering the network behavior, makes it possible to balance the length of the buffer with the possibility of packet loss. Generally, a good playout algorithm should be able to achieve the best possible trade-off between loss and delay.

Most of the adaptive playout algorithms described in the literature perform continuous estimation of the network delay and its variation to dynamically adjust the packets playout time. The basic adaptive playout algorithm estimates the average packet delay and its variance and uses them to calculate the playout time. The packet delay, its variation and the playout time for playing out the next packet are estimated as:

$$d_i = \alpha * d_{i-1} + (1- \alpha) * n_i \qquad (1)$$

$$v_i = \alpha * v_{i-1} + (1- \alpha) * | d_i-n_i | \qquad (2)$$

$$p_i = d_i + \beta * v_i \qquad (3)$$

where d_i and v_i are the i-th estimates of delay and its variance respectively, n_i is the delay suffered by the i-th packet in the network, α is a weighting(smoothing) factor and β is a safety factor, used to ensure that the predicted delay is greater than the actual delay with a high probability.

As far as parameter α has a critical impact on the estimated playout delay, there are many algorithms which estimate playout delay using different modifications of (1). Some of them set α equal to some optimal fixed value, others change that value in some manner considering network conditions (e.g. spike detection algorithm). But all of them have one thing in common: all of them estimate playout delay considering only network behavior Fig. 1a. But in addition to this VoIP session can be characterized also by the interaction mode, which means the following:

- Interactive mode: both sides are sending information
- Non-interactive mode: one side is only sending the information, the other is only receiving

As an example of an interactive mode, the situation when two people are talking to each other can be brought, and as a non-interactive – the situation of audio broadcasting or when somebody is just listening hold music, some voice message or something else.

There are a lot of VoIP systems supporting many services which can be considered as services of interactive mode, of non-interactive mode and services with dynamic switches between the modes. It is obvious, that if algorithm, besides network behavior, also considers the interaction mode, it can significantly improve perceptual voice quality. In the next section a new interaction mode aware playout delay adjustment algorithm is presented.

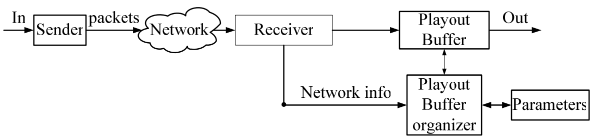

Fig. 1a Conventional approach

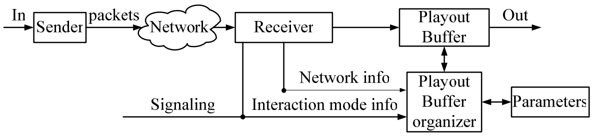

Fig. 1b Proposed approach

2 Proposed Algorithm

Fig. 1b is a practical representation of the proposed algorithm. Playout delay estimation of interactive mode differs from the estimation of the non-interactive one, because in the non-interactive mode the playout delay can be much longer. Besides this interaction mode aware algorithm should be ready to interaction mode dynamic switches (IDS), i.e. the interactive mode switches to non-interactive and vice versa.

The presented algorithm doesn't modify playout delay estimation mechanism of the interactive mode (M_{inter}) considering that it can be any network behavior based algorithm described in the literature. This algorithm just describes the methods which can be applied in case of non-interactive mode.

Let's set two sub modes of a non-interactive mode in case of possible IDS: NOM – a non-interactive obligatory mode and NNM – a non-interactive non-obligatory mode. Since the playout delay value in case of a non-interactive mode can be much longer than the maximum value allowed in the case of interactive mode, it is obvious that the delay accepted in case of the non-interactive mode cannot be accepted in the interactive mode; thus either playout delay should be decreased while switching to interactive mode or playout delay should not be increased during non-interactive mode. Accordingly to this, sub modes NOM and NNM were set. Prior means that the playout delay cannot be prolonged too much, because all the data are obligatory and have to be played out; the latter means that the playout delay can be increased as much as needed, because in case of switching to the interactive mode playing out the packets received before this moment is not necessary anymore and the playout delay can be reduced if needed. As an example of NOM the situation when before starting the conversation the receiver has to listen to something can be brought (e.g. License Agreement) and as an example of NNM – when the receiver is just listening to hold music or voice message.

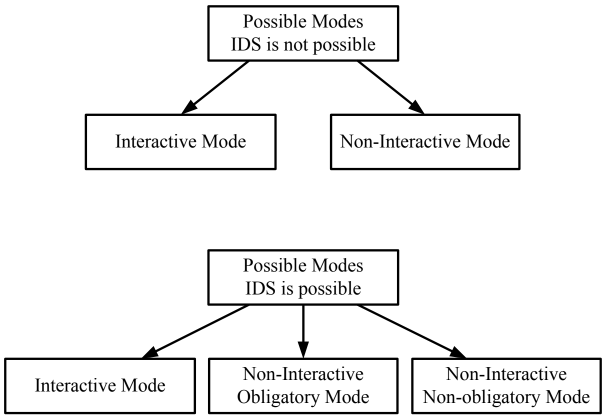

Fig. 2 Interaction modes

As shown on Fig. 2 there are only 4 modes: an interactive mode, a non-interactive mode, a non-interactive obligatory mode and a non-interactive non-obligatory mode. The playout delay of interactive mode will be estimated using M_{inter}, because as it was mentioned above this paper doesn't propose anything new for the interactive mode. As for a non-interactive mode, the description of the three methods which should be applied in case of a non-interactive mode is given further.

2.1 Non-interactive mode playout delay estimation mechanism ($M_{non\text{-}inter}$)

$M_{non\text{-}inter}$ is the mechanism which should be applied in case of non-interactive mode if the IDS is not possible. This mechanism quickly reacting to the network deterioration allows constantly increase the playout delay till some maximum value d_{max}.

if($n_i > n_{i-1}$ && $d_i < d_{max}$) then

$d_i = \alpha * d_{i-1} + (1- \alpha) * n_i$

2.2 Non-interactive mode obligatory playout delay estimation mechanism ($M_{non\text{-}inter\_ob}$)

$M_{non\text{-}inter\_ob}$ is the mechanism which should be applied in case of non-interactive obligatory mode. This mechanism allows quickly react to the network deterioration and slowly react to the network improvement: i.e. quickly increase the playout delay and slowly reduce it.

if($n_i > n_{i-1}$) then

$d_i = \alpha * d_{i-1} + (1- \alpha) * n_i$

else

$d_i = \beta * d_{i-1} + (1- \beta) * n_i$

where $\alpha < \beta$

2.3 Non-interactive mode non-obligatory playout delay estimation mechanism ($M_{non\text{-}inter\_non\text{-}ob}$)

$M_{non\text{-}inter\_non\text{-}ob}$ is the mechanism which should be applied in case of non-interactive non-obligatory mode. This method estimates 2 values playout delay using $M_{non\text{-}inter}$ and switch delay d_{sw} using M_{inter}. In case of switching to the interactive mode the playout delay should be set equal to d_{sw}, since the non-interactive non-obligatory mode allows it.

3 Design

Using "Ethereal – Network Protocol Analizer", fifty seconds of the real voice sent from US to Armenia has been captured (see Fig. 3).

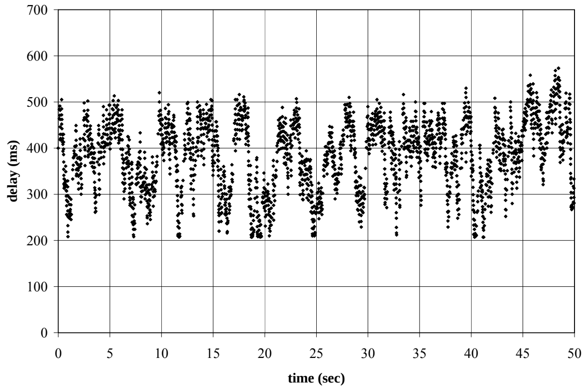

Fig. 3 Delay experienced by the packets

The simulator developed by the author performs the following:

- obtains the real arriving time of the packet from the captured file
- estimates playout delay for the packet based on given parameters (playout delay cannot be less than 40ms)
- calculates the lost packets count and packets average delay.

Thus it allows us to compare the performance of the algorithms under identical network conditions.

The performance of the proposed algorithm is estimated for all interaction modes. The parameters used during the experiments are presented in Table 1 and obtained results – in Table 2.

Modes	Parameters	
Interactive	α = 0,998002	1 Segment 50 sec of interactive mode (conversation)
Non-interactive	α = 0,5	1 Segment 50 sec of non-interactive mode (time, forecast, etc)
Non-interactive Obligatory	α = 0,7 β = 0,9	3 Segments 10 sec of interactive mode (conversation) 20 sec of non-interactive obligatory mode (license agreement) 20 sec of interactive mode (continue conversation)
Non-interactive Non-obligatory	α = 0,5	3 Segments 15 sec of non-interactive non-obligatory mode (1st voice mail) 15 sec of non-interactive non-obligatory mode (2nd voice mail) 20 sec of non-interactive non-obligatory mode (3rd voice mail)

Table 1 Parameters used during the experiments

Modes	Segment 1	
	loss (%)	av. delay (ms)
Interactive	26.3	41
Non-interactive	0.9	116

Modes	Segment 1		Segment 2		Segment 3	
	loss (%)	av. delay (ms)	loss (%)	av. delay (ms)	loss (%)	av. delay (ms)
Interactive	3.9	40	9.2	41	12.7	42
Non-interactive Obligatory	3.9	40	5.2	48	8.5	60

Modes	Segment 1		Segment 2		Segment 3	
	loss (%)	av. delay (ms)	loss (%)	av. delay (ms)	loss (%)	av. delay (ms)
Interactive	7.5	40	5.6	41	12.6	42
Non-interactive Non-obligatory	4.2	48	3	47	6.6	56

Table 2 Result obtained during the experiments

4 Conclusion

In this paper are explored the issues of estimating the playout delay of the packets in the VoIP systems. A new algorithm is proposed that improves the perceptual voice quality by characterizing the VoIP session by interaction mode. Implementation and obtained results show up that the proposed algorithm is very effective in cases when the network behaviour is variable, and it can be used for extending the existing algorithms used in the VoIP systems supporting different services like call hold, voice mail, etc.

5 References

[1] R Ramjee, Jim Kurose, Don Towsley, and Henning Schulzerine, “Adaptive playout mechanisms for packetized audio applications in wide area networks,” Proc. IEEE INFOCOMM, pp. 680–686, June 1994.

[2] Aman Kansal and Abhay Karandikar, “Adaptive delay estimation for low jitter audio over internet,” in Proccedings of IEEE Globecom, San Antonio, USA, November 2001.

[3] S. B. Moon, J. Kurose, and D. Towsley, “Packet Audio Playout Delay Adjustment: Performance Bounds and Algorithms,” Multimedia Systems, vol.6, pp. 17–28, 1998.