

IP Telephony over Satellite Networks: Control of the Playout Delay to Maximize the Perceived Quality

Fabrizio Boi, Luigi Atzori and Mirko Luca Lobina

MCLab, Department of Electronic Engineering, University of Cagliari

Cagliari 09123 Italy

f.boi, l.atzori, m.lobina @diee.unica.it

ABSTRACT

In streaming applications, playout buffering is operated at the receiver side to compensate network delay variations. This function allows the receiver to reduce packet lateness at the expenses of an increase in the end-to-end delay. IP Telephony is one of the applications where this function is crucial since transmission delays heavily affect the interactivity and significant losses injure the signal quality. Many approaches have been proposed in the last decade to control the playout delay in IP Telephony, where the most recent ones rely on the use of the ITU-T E-Model to evaluate the quality perceived by the end-user. However, none of these specifically deals with the scenario where the service is provided over satellite networks. Indeed, nowadays, this scenario is quite frequent with the reduction of the costs of the satellite channels and equipments. In this case, there are some specific issues that need to be dealt with to improve the effectiveness of the playout control algorithms. These issues are related to the delays that are encountered when satellite links are involved. These aspects are described in this paper, which also presents a new quality-based playout algorithm that provides better results with respect to alternative approached when satellite transmissions are involved.

Keywords

IP Telephony, geo-synchronous (GEO) satellite network, playout mechanism

1. INTRODUCTION

The deployment of reliable and high-quality real-time multimedia services is one of the most challenging activities for Internet providers. This demands for complex management procedures of the service and network, with careful dimensioning of link capacity and servers, appropriate setting of the QoS parameters, traffic engineering operations at both the backbone and the access parts of the network as well as a careful configuration of the servers for signaling, data coding, transcoding and processing.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Mobimedia'07, Month 8, 2007, Nafpaktos, Aitolokarnania, Greece
Copyright 2007 ICST 978-963-06-2670-5

Additionally, the customer premise equipment has to be correctly configured since the quality can be hampered by a mistaken setting of several aspects that may impact the quality of the coded and decoded signal as well as the way the packets are routed and prioritized along the path to the destination. Indeed, higher and higher complexity is introduced in the terminals to provide user-aware, channel-aware and terminal-aware adaptation of the service. The terminal is becoming a part of the network and needs to be able to adapt the content coding and transmission according to the heterogeneous set of communication scenarios.

Within this context, in this paper we focus on the problem of playout buffering for IP Telephony services. This procedure is operated at the receiver side to compensate delay variations at the expenses of an increase in the end-to-end delay; it heavily influences the final perceived quality since the end-to-end delay heavily affects the interactivity and significant losses injure the signal quality. Many approaches have been proposed in the last decade to control the playout delay in IP Telephony, where the most recent ones rely on the use of the ITU-T E-Model to evaluate the quality perceived by the end-user. However, none of these specifically deals with the scenario where the service is provisioned over satellite networks. In this case, there are some specific issues that need to be dealt with to improve the effectiveness of the playout control algorithms. These issues are related to the delays that are encountered when satellite links are involved. These aspects are described in the following and we then propose a new quality-based playout algorithm that specifically takes into account these features.

The paper is organized as follows. Section 2 describes the GEO satellite network, the GEO speech frame model and the delays that are observed in the end-to-end communication. Starting from this analysis, a novel ad-hoc playout mechanism is proposed in Section 3. Comparative results are provided in Section 4 and conclusions are drawn in section 5.

2. ANALYSIS OF THE TRANSMISSION DELAYS OVER SATELITE NETWORKS

In the last years, the satellite system has become a suitable alternative to the conventional fixed/terrestrial backbone, being: i) distance-insensitive both for the QoS (delivery at almost identical quality of service across an endless number of locations) and hardware connection (theoretical capability of reaching all the sites in the network regardless of their distance); ii) flexible, since

many satellite IP networking equipment are available to help in covering long-distance connections with easy to configure procedures. Thanks to these advantages, there have been many efforts towards the integration of satellite channels with terrestrial networks making also use of QoS technologies to assure the required QoS [1, 2, 3].

Next subsection illustrates the satellite network that is used in our laboratory for the experiments that constitute the basis of the proposed playout buffering algorithm. The following subsection presents the analysis of the impact on the end-to-end data transport performance of the satellite links.

2.1 The CNIT GEO satellite network

In 2000, the CNIT (Italian University Consortium of Telecommunications) has set up its GEO satellite network to create a corporate network connecting its research units and laboratories at the main Italian Universities. This solution allows the members of the consortium to share both common Internet services and audio/video/data, with very high quality and reliability. The core of this satellite network is the Skyplex technology [4]. Skyplex is a digital multiplexer, based on the European standard DVB (Digital Video Broadcasting) and compatible with the DVB/MPEG and the DVB-RCS technologies. The network operates in the Ka-band (20-30 GHz) and provides a guaranteed bandwidth of 2 Mbps shared among the connected terminals. In our typical applications, most of this bandwidth is reserved for audio and video multicasting, so that a significant improvement of their quality is achieved. Furthermore, the bandwidth requirements of each terminals is periodically adjusted by means of a BoD (Bandwidth on Demand) mechanism, following the indications of a Regional Network Control Centre (RNCC).

EUTELSAT [5] is the space operator partner of the network. The satellite (i.e., Hot Bird 6, positioned at 15 degrees East) covers all European countries through 4 uplink areas. Hot Bird 6 has 8 Skyplex units on board, with 18 channels/2 Mbps that operate both in continuous mode (SCPC) and in burst mode (TDMA [6, 7] for each user: 350 kbps - 2 Mbps), depending on the traffic pattern. Our tests have been conducted in a burst mode with timeslots assigned in a best effort mode. In our experiments are mainly related to the two communications between two sites, i.e., Cagliari (UniCA) and Florence (UniFI), by means of the Skyplex technology station with an Ethernet interface. Every site implements an IP Telephony solution based on Asterisk, an open source PABX-IP (Private Automatic Branch eXchange-IP) [8], implementing the IETF SIP protocol [9].

2.2 GEO Frame Structure and Associated Delay “Saw-Tooth” Effect

The Skyplex TDMA structure [7] is organized in two different levels: user timeslots and MPEG2 burst. At a first level of analysis, every TDMA frame is made of N user timeslots for each user connected to the GEO satellite network. The user timeslot contains M MPEG2 transport packets. The detailed description regarding the process of creating the MPEG2 transport stream from the speech codec packets is provided in [7].

In the Skyplex TDMA structure N and M are combined in a way that $N \cdot M = 48$. This is the default configuration, as suggested by the network provider for optimizing the bandwidth usage versus

the number of connected users. Furthermore, the resources of the network are assigned to the users with the already cited BoD mechanism. That is, first of all the RNCC (Regional Network Control Center) monitors the network and then produces a ticket, which is sent to all the satellite gateway terminals to which the users are directly connected every interval of 820 ms. This ticket contains the exact satellite backbone capacity assigned by the network control centre to all the users on the basis of their requirements. As a result of this frame structure, a variable number of MPEG2 packets per TDMA frame are assigned to each user. In fact, the first user timeslot contains, other than the MPEG2 transport stream packets, the ticket produced by the RNCC about the network capacity given to that user. Instead, the next frames, of duration of 820 ms, include several RTP packets, because there is no need for additional traffic information. This structure containing few/many MPEG2 packets is repeated at fixed intervals for the entire duration of transmission.

This analysis is of crucial importance for the study of the end-to-end delays encountered by the RTP packets. From what has been said, the Skyplex proprietary structure of the frame modifies the frame stream: all the RTP packet delays are dependent on the last RTP packet of the TDMA frame. In fact, the transmission over the satellite link can start only when the last RTP packet closes the frame. As a consequence, the associated end-to-end delays of the RTP packets present the behavior showed in Fig.1, which is a typical “saw tooth” trace.

The control of the playout delay needs to take into account this “saw tooth” effect not to introduce too high delays. Accordingly, in the following we present the analysis of the delay components that characterize the considered architecture.

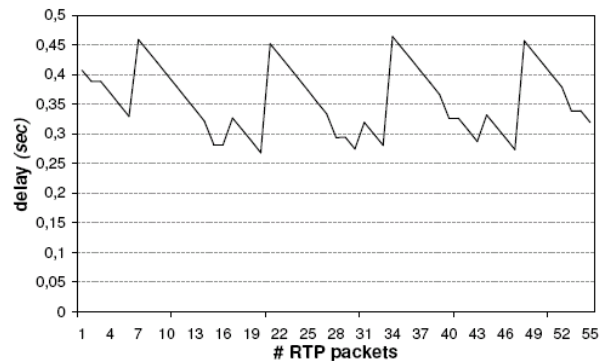


Fig. 1. Delay of RTP packets.

At both sites, the transmission through the segments of IP networks introduces two delay components, which we refer to with D^1 and D^2 , respectively. These are random delays that follow the temporal variability of the traffic loads and congestion occurrences. Two constant delays are then added, which are the transmission delay t_x of a generic RTP packet through the satellite link and the t_{prop} propagation delay of the electromagnetic wave. Both these quantities have to be considered twice for the go and back links. t_{prop} is equal to $(\text{sat distance})/(\text{speed})^1$. The last component is D^5 : it is the delay introduced to each packet transmission that waits at the satellite transmitter buffer till the

¹ $\approx (35790\text{km})/(300\text{km/s})$

satellite frame is full and ready to be transmitted. Based on this analysis, the end-to-end delay for a generic RTP packet is the following:

$$d = D^1 + D^2 + D^s + 2t_x + 2t_{prop}. \quad (1)$$

We now introduce a subscript to index the delay according to the position of each packet in a satellite frame; packet 0 is the last in frame, packet 1 is the second last, and so on, so that the delays can be written as follows:

$$\begin{aligned} d_0 &= D_0^1 + D_0^2 + 2t_x + 2t_{prop} \\ d_1 &= D_0^1 + D_1^2 + 2t_x + 2t_{prop} + D^{int} \end{aligned} \quad (2)$$

...

$$d_i = D_0^1 + D_i^2 + 2t_x + 2t_{prop} + i \cdot D^{int}$$

Note that since the frame is transmitted when the last packet arrives at the satellite transmitter, the delay of every packet includes D_0^1 . In (2), D^{int} is the length of the speech frames transmitted in a RTP packet and depends on the specific speech codec adopted in the VoIP system.

In this way, we have been able to remove the random component D^s in (1) and the delay is then influenced by the random components $x = D^1 + D^2$. Predicting the probability density function (pdf) $f(x)$ of x , we are able to obtain the pdf for all the packets $f_{d_i}(d_i)$ in a frame according to their position within the frame:

$$\begin{aligned} f_{d_0}(d_0) &= f(x - 2t_x - 2t_{prop}) \\ f_{d_1}(d_1) &= f(x - 2t_x - 2t_{prop} - D^{int}) \end{aligned} \quad (3)$$

...

$$f_{d_i}(d_i) = f(x - 2t_x - 2t_{prop} - i \cdot D^{int})$$

3. PLAYOUT BUFFERING

The playout buffering algorithm that we propose relies on the quality-based approach that is adopted by most of the algorithms that appeared in the last years [10], that is, it is designed to adjust the buffer size so as to maximize the expected end-user perceived quality. The aim is to jointly consider the expected end-to-end delay and information loss making use of a perceptually motivated optimality criterion that allows the receiver to automatically balance packet delay versus loss. In this way, no thresholds for the loss and/or delay have to be considered: the playout algorithm automatically adapts the buffer size so as to maximize the expected quality. The following main operations are performed: statistics relevant to loss and delay are predicted by means of the previously sent packets; based on this information, the buffer setting is accomplished so as to maximize the expected conversational quality during future conversational units. To evaluate the quality we make use of the ITU-T E-Model [11, 12] with some simplifications. Accordingly, the overall output index

is a function of several system parameters, including the de-jitter buffer size.

We briefly summarize the expression of the R-factor which represents the output index of the E-Model that measures the user-perceived quality:

$$R = 100 - I_s - I_d - I_{e,eff} + A \quad (4)$$

I_s includes all the impairments introduced in the circuit-switched part of the end-to-end communication path. I_d mostly comprises the effects due to delay and echo. $I_{e,eff}$ takes mainly into account the impairments caused by low bit-rate codecs and information loss. The expectation factor A raises the conversational quality when the end-user may accept some decreases in quality for access advantage (e.g., mobility). Despite its apparent simplicity, (4) represents a non operative form of the E-Model since the four factors depend on several configuration parameters. However, under some common assumptions these parameters can be simplified as follows [13]:

$$\begin{cases} I_s = 6.8 \\ I_d = 0.024d_{e2e} + 0.11(P_D - 177.3)H(P_D - 177.3) \\ I_{e,eff} = 11 + 40 \ln(1 - 10E) \\ A = 0 \end{cases} \quad (5)$$

where P_D is the total end-to-end delay (in ms), E is the total packet loss rate, and $H(x)$ is the step function ($H(x)=0$ if $x<0$ and $H(x)=1$ otherwise). The loss and end-to-end delay can be expressed in terms of components relevant to the codec, network and playout buffer in the end-to-end path as follows:

$$\begin{cases} P_D = d_{codec} + d_{dejitter} + d_{net} \\ E = e_{net} + (1 - e_{net})e_{dejitter} \end{cases} \quad (6)$$

The total network delay d_{net} is the same delay d used in the previous section. The codec delay is a constant value which is often neglected. The R-factor is then a function of the total packet loss rate and the end-to-end delay, named also playout delay. Note that the amount of losses depend on the playout delay too: the higher the delay is, the lower the loss rate is.

Finding the optimal playout delay for a subsequent talkspurt during the silence period according to the E-Model is the aim of the proposed strategy, whose main steps are shown in Fig. 2. At the reception of each packet, the observed delay d is used to update the pdf $f(x)$ defined in the previous section. Additionally, the length of each received TDMA frame in terms of number of RTP packets conveyed is used to build the probability function $p(s_i)$: it is a function representing the probability of transmitting a packet in position i within a frame. $f(x)$ and $p(s_i)$ during each silence period are then used to represent the relationship between late packet loss probability and playout delay:

$$\begin{aligned}
e_{de jitter} &= L(P_D) = \sum_{i=0}^{\max} p(s_i) \cdot \int_{P_D}^{\infty} f_{d_i}(d_i) dx = \\
&= \sum_{i=0}^{\max} p(s_i) \cdot \int_{P_D - 2t_x - 2t_{prop} - i \cdot D}^{\infty} f(x) dx
\end{aligned} \tag{7}$$

where max represents the maximum number of RTP packets that can be conveyed in a TDMA frame. The following two steps in the proposed algorithm are quite easy to follow. Function $L(P_D)$ is introduced in the expression of E together with the average network packet loss ratio (e_{net}). In this way the final expression of R as a function of only P_D is available. The following operation consists in finding the optimal P_D^* value that will be used during next talkspurt, that is, the one that maximize R.

To increase the effectiveness of the proposed playout control scheme, an essential condition is to define a simple statistical method for estimating $f(x)$ on the basis of observed measurements. Indeed, network characteristics vary with time and, consequently, the delay statistical trends do; a continuous update of this information is then required to take into account these evolutions. Several approaches can be used to predict statistical trends of network delay relying on the analysis of historical and current information. The full aggregation method accumulates data into the probability distribution curve, giving the same weight to old and recent samples; this approach makes the system unable to quickly react to network traffic variations. Differently, the flush and refresh method builds the probability distribution curve on the basis of the last observed M packets and leaves out historic information, generating high overhead. An intermediate approach that we have decided to adopt is the store and track method. This approach does not discard the old data entirely, but gradually reduces its effect on the histogram to approximate the statistical distribution. Each value in the histogram counts the number of occurrences of packets sent within a certain interval of delays (bin). Note that the shorter the bin is, the higher the accuracy in the estimation of $f(x)$ is. On the other hand, short bins increase the algorithm complexity when maximizing the R-factor. To reduce the weight of older samples, each bin of the histogram is periodically scaled down by an aging factor F .

As to the maximization of the R-factor, it is a procedure conducted numerically: the packet loss rate is evaluated function at several P_D values according to the temporal resolution; this are than used in the linear expression of the R-factor; in the current implementation the optimal P_D value is found looking at all the computed R-factor values.

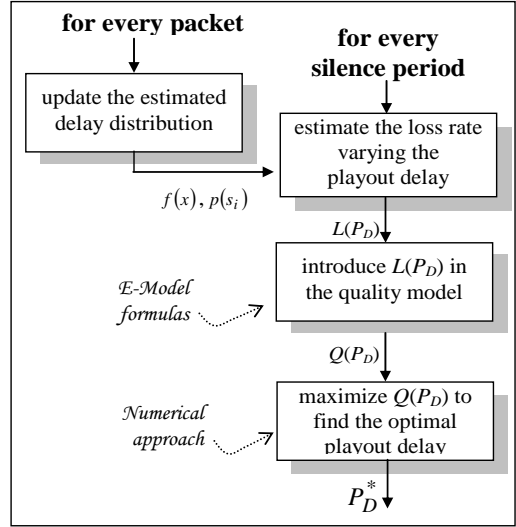


Fig. 2. Sketch of the proposed playout control strategy.

4. EXPERIMENTS

We have recorded and evaluated several traces at different traffic loads of the GEO satellite network. An overview of the main features of these traces is provided in Table I. All the tests have been performed synchronizing both sides clocks by means of the NTP (Network Time Protocol).

Table I. Main features of the recorded traces

Codec	GSM, iLBC, G.711 μ
Call duration (min)	5-10
Average network delay (sec)	0.354-0.400
Network delay variance (sec)	0.0034-0.0040

The “saw tooth” effect is clearly independent from the used codec and a single tooth can be seen as a network spike. For this reason, we have compared the performance of our strategy (named *Ad-Hoc Sat*) with the class of spike detection playout control algorithms. Specifically, the *Adaptive Linear Filter Buffer* with spike detection [14] and the *Enhanced-Normalized Least Mean Square* (E-NLMS) [15] have been chosen. Among the adaptive playout techniques, the first strategy is one of the most used because it offers good performance and a straightforward implementation. The E-NLSM offers results comparable to the Linear Filter. The strategies have been compared in terms of the R-factor over a number of random talkspurts. In this paper we present the results of such comparison for one significant and representative trace: a 5 minutes conversation with the GSM codec. The RTP packet delays for the considered trace is presented in Fig.3, while the resulting R-factor for the initial talkspurts is shown in Fig.4.

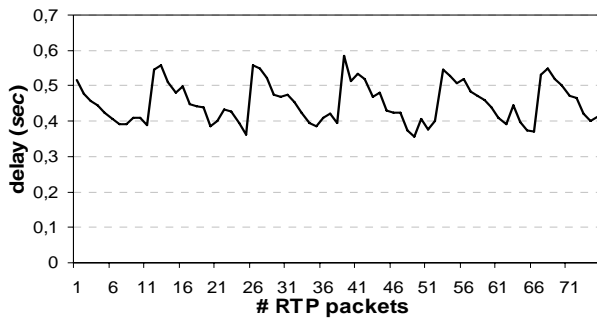


Fig 3. Saw Tooth of GSM trace (zoom)

Fig 3 shows the typical “saw tooth” delay trace, as it was presented and commented in section 2.2. Fig. 4 provides a comparison of the three techniques for the first 30 talkspurts. The initial value is 0 because the initial playout delay is set to 100 ms for any algorithms. We can observe as the performance of *Ad-Hoc Sat* are constant and the value of R-factor remains around 83, that is a high-quality level of voice rating. The reason is that the optimal value of the playout delay P_D^* is calculated for each talkspurt considering both the R-factor maximization and the inner structure of the TDMA frame, that is, the position of the RTP packets with their probability. The *Linear Filter* results are affected by continuous R-factor changes from packet to packet. This highlights that this technique is not able to take into account high spikes of delays between two close RTP packets, resulting occasionally in burst of losses and audible quality degradation. This feature is improved by the E-NLMS because of an additional parameter that allows adjusting the buffer better when high delay variations occur. Table II show the average R-factor values excluding the first initial five talkspurts. It highlights that the proposed algorithm performs more than 4 points better than the *Linear Filter* and almost 1 point than the E-NLMS in terms of R-factor. If we take a look at the delay, we see that the E-NLMS is able to obtain good results at the expense of delays that are higher than the *Ad-Hoc Sat* ones due to the oversetting of the delay when spikes are detected.

Table II. Average R-factor and playout delay.

	Linear Filter	E-NLMS	Ad-Hoc Sat
Average R-factor	78.7	82,3	83.2
Average Playout Delay (sec)	0.485	0.5155	0.498

5. CONCLUSIONS

A playout control algorithm for communications over satellite networks, named *Ad-Hoc Sat*, has been presented for IP Telephony services. The major contributions of this paper are the study of the GEO frame model and the proposal of a strategy that improves the end-user perceived quality by taking into account the delay statistics encountered by the RTP packet along the satellite channels. The performed tests show encouraging results when comparing the proposed technique with two alternative algorithms that include spike detection strategies. Future work is

aimed at investigating new algorithms for the prediction of the delay pdf.

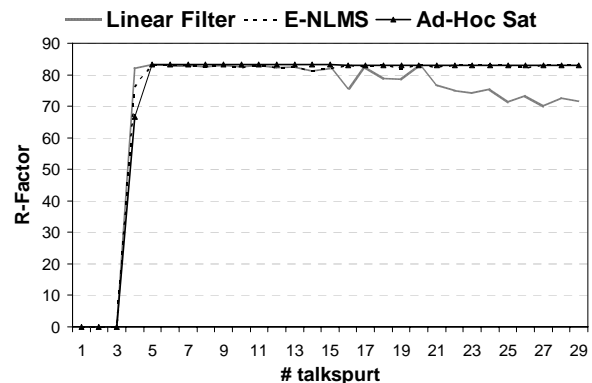


Fig. 4. Comparison of the alternative techniques in terms of R-Factor.

6. ACKNOWLEDGMENTS

The research activities described in this paper have been conducted within the IKNOS project funded by the MIUR (Italian Ministry of Research and Education).

7. REFERENCES

- [1] A. Iera, A. Molinaro, and S. Marano, “IP with QoS Guarantees via geo Satellite Channels: Performance Issues,” IEEE Personal Communication, Vol. 8, Issue 3, pp. 14-19, June 2001.
- [2] R. Toegl, U. Birnbacher, and O. Koudelka, “Deploying IP Telephony over Satellite Links,” IEEE Wireless Communication Systems Conf., pp. 624-628, September 2005.
- [3] A. Iera, A. Molinaro, and S. Marano, “Call Admission Control and Resource Management Issues for Real-Time VBR Traffic in ATM-Satellite Networks,” IEEE Journal on Selected Areas in Communications, Vol. 18, Issue 11, pp. 2393-2403, November 2000.
- [4] CNIT, Skyplex :CNIT satellite network, 2005.
- [5] www.eutelsat.com
- [6] E. Feltrin, E. Weller, E. Martin and K.Zamani, “Design, Implementation and Performances Analysis of an On Board Processor – Based Satellite Network,” IEEE Communication Society, 2004.
- [7] EUTELSAT,” Technical guide: annex D, Skyplex,” 1999.
- [8] www.asterisk.org
- [9] RFC 3261,” SIP:Session Initiation Protocol.,”
- [10] L. Atzori, M.L. Lobina, , “Playout Buffering in IP Telephony: a Survey Discussing Problems and Approaches,” pp. 36-46, 3th quarter 2006.
- [11] L. Atzori, M.L. Lobina, and M. Corona, “Playout Buffering of Speech Packets based on a Quality Maximization Approach,” IEEE Trans. on Multimedia, vol. 8, no. 2, April 2006.

- [12] ITU-T Recommendation G.107, "The E-Model, a computational model for use in transmission planning," 03/2003.
- [13] Cole, R. and Rosenbluth, J. Voice "Over IP performance monitoring," ACM Computer Communication Review, vol. 31, no. 2, Apr. 2001
- [14] R. Ramjee, J. Kurose and D. Towsley, "Adaptive Playout Mechanisms for packetized audio applications in Wide-Area Networks," in Proc. of IEEE INFOCOM '94.
- [15] A. Shallwani and P. Kabal, "An adaptive playout algorithm with delay spike detection for real time VoIP," in Proc. of IEEE Canadian Conf Electrical 2003.