AN EFFICIENT PACKETIZATION SCHEME FOR VOIP

A. Estepa, R. Estepa and J. Vozmediano
University of Sevilla
Camino de los Descubrimientos s/n - E41092 Seville

Keywords: VoIP, traffic characterization.

Abstract: A number of VoIP audio codecs generate Silence Insertion Descriptor (SID) frames during talk-gaps of conversations to update the comfort noise generator-parameters at the receiver. According to the RFC 3551 packetization scheme, discontinuously-generated SID frames can not be carried in the same IP packet, thus increasing the conversation's bandwidth consumption..

We define a novel packetization scheme in which a set of non-consecutive SID frames may share the same packet, reducing the overhead while keeping the timing between them. We provide analytical expressions and experimental validation for the bandwidth savings obtained with this new scheme, which grows up to a 14% for the G.729B codec.

This work was supported in part by the Spanish *Secretaría de Estado de Universidades y Educación* under the project number TIC2003-04784-C02-02

1 INTRODUCTION

Voice over Internet Protocol (VoIP) is experiencing an exponential growth in recent years due to the implicit cost saving that conveys the usage of the free Internet or corporate networks.

The goal for a VoIP system is the achievement of a high quality of service (QoS) at the minimum cost. The main impairments affecting the VoIP quality are delay, packet loss and the codec's intrinsic quality. Since delay and packet loss are related to the available bandwidth in the network, the reduction of the conversation's bandwidth requirement is a key goal in the design of VoIP systems. The two main factors determining the bandwidth requirement of a voice stream are the codec and the number of voice frames carried in each packet (packetization rate). The chosen codec should balance bandwidth consumption (codec bit-rate) and listening speech quality, while the packetization rate should balance between bandwidth saving (the more codec's frames in each packet, the less overhead ratio) and delay. The packetization rate can be customized by the user in VoIP clients, and could potentially achieve significant bandwidth savings at the cost of increasing the packetization delay.

Codecs periodically generate compressed voice frames. Low bit-rate codecs are usually equipped with a voice activity detection (VAD) feature which pursues bandwidth savings by avoiding the generation of frames during voice inactivity periods. Additionally, some audio codecs such as G.729, G.723.1 or AMR are also equipped with a discontinuous transmission algorithm (DTX) which allows, at the beginning of each voice inactivity period, to send a short-sized type of frames named Silence-Insertion-Descriptor (SID). Reception of a SID frame after a voice frame can be interpreted as an explicit indication of the end of the talk-spurt. In addition, SID frames may be also transmitted at any time during the silence interval to update comfort noise generation parameters. This allows a faithful reproduction of the background noise at the receiver's side, increasing the quality of the conversation at the cost of some additional bandwith (Estepa et al., 2003). But the SID frames, although short-sized, cause the generation of RTP/UDP/IP packets affecting negatively to the bandwidth consumption of the conversation (Estepa et al., 2003).

Voice streams are typically transported using RTP over UDP. The Real Time Protocol (RTP) is aimed to provide the receiver with two main features in addi-

Estepa A., Estepa R. and Vozmediano J. (2006).
AN EFFICIENT PACKETIZATION SCHEME FOR VOIP.
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pages 65-70
DOI: 10.5220/0001569300650070

tion to the packet's sequencing feature:

1. *Payload identification*. The Payload-Type field of the RTP header uses the profile identifiers given at RFC 3551 (Schulzrinne, 2003), as a mean to provide information to identify the codec, the frame type and how frames of a given codec (voiced or SID) are packet. For example, for the G.723.1 three types of frames are distinguised and marked using two bits at the beginning of each frame: Active for 5.3 mode, Active for 6.3 mode and SID. Nevertheless for the G.729 codec, the frame type is indicated by the size of the packet.
2. *Jitter compensation*. A timestamp is inserted in the RTP header which allows to compensate the network jitter. The receiver of a packet containing voice frames can adjust the play-out instant of the first frame of the packet by using this timestamp. Subsequent frames of the packet are time-consecutive, so they can be played-out on the proper instant at the receiver with no additional information other than the audio profile identifier.

Since the RTP header only provides one timing instant, there would be no way to choose the right playout time for SID frames of the same IP packet having any time-slot gap in between them. As a result, the current packetization scheme, RFC 3551 (Schulzrinne, 2003) also imposes that only consecutively-generated SID frames are carried in the same packet (Zopf, 2002). But the large RTP/UDP/IP header size when compared to the reduced size of the SID frames (typically 2 or 4 bytes) makes worth considering to share the same packet for several SID frames.

This paper proposes a new packetization scheme to allow the transmission of non-consecutively-generated SID frames in the same packet. The new scheme does not modify the existing packetization scheme for active frames, and is backward-compatible with the RFC 3551 packetization profile. This scheme will save bandwidth in conversations using SID-capable codecs, specially in scenarios with a large number of potential users, where the header-compression technique shows scalability problems. This is may be the case of a corporate network branch to central office communication, or the case of the VoIP Service Provider transport service, since the cost of the transport can be considered proportional to the required bandwidth.

The remainder of this paper is organized as follows: the mechanism is described in section 2. The analytical expression of the achieved rate reduction is deduced in section 3. The analytical expressions are experimentally validated in section 4 and, finally, section 5 concludes the paper.

2 PROPOSED PACKETIZATION SCHEME

To reduce the bandwidth usage of the current IETF packetization scheme and at the same time solving the problem of obtaining the right playout time for every SID frame but the first one of each packet, we define a new payload type called *multi-SID*. To maintain backward compatibility, we re-define the format of the existing RTP payload profile identifier number 13 (Schulzrinne, 2003) currently devoted to indicate a generic non-proprietary SID frame. This generic SID frame, whose format is defined in (Zopf, 2002), uses the first byte to indicate the power level of background noise with a number between 0 and 127 (note that the most significant bit is always set to 0.)

We propose to extend the aforementioned format to include also our new *multi-SID* payload type, which will be indicated by setting the first bit of the first byte to one as shown in figure 1. In our format, the remaining bits of the first byte of this *multi-SID* frame identify the codec (and thus the SID frame size being used.) Following bytes carry the first SID frame (e.g. two-bytes size in figure 2.) At the end of the every SID frame but the last one, an additional field called *NF* stands for the number of frame-period gaps to be inserted before playing the next SID frame. The *NF* field has a length of one octet.

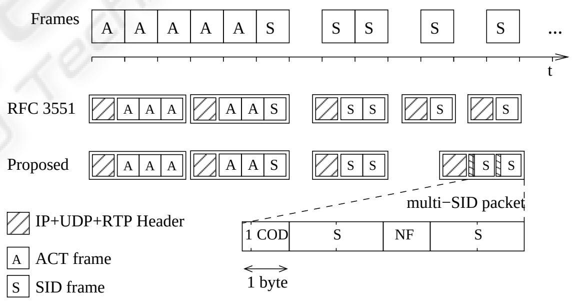

Figure 1: Proposed packetization scheme.

Finally, we propose to use the *multi-SID* payload type only whenever two or more SID frames generated within N_{fpp} frame intervals would be forced to travel in different packets according to the current RFC 3551 scheme. This implies that the bandwidth consumption will always be equal or less than with the RFC 3551 packetization scheme. The bandwidth saving achieved by this procedure will be estimated in next section.

3 BANDWIDTH GAIN ESTIMATION

The aim of this section is to quantize the benefits of our proposed scheme. The reader should keep in mind that the contribution of this paper is the new packetization scheme, since the deduction of the conversation's mean bit-rate analytical expression was first presented in (Estepa et al., 2005).

3.1 Mean Bit-rate for RFC3551 Packetization Scheme

We assume that, during voice activity periods, a new packet transporting N_{fpp} compressed-voice (ACT) frames is generated every $T \cdot N_{fpp}$. During voice inactivity periods, a new packet loaded with SID frames is sent according to the RFC 3551 packetization scheme (in the VoIP case). Since SID frames generation is a random process, we will use a discrete random variable, X, to indicate the inter-arrival time (in number of periods) between SID frames. Moreover, we assume that SID frame generation is a renewal process.

The mean bit-rate of one conversation can be calculated as the sum of the contributions of the traffic generated during voice activity (R_{ON}) and voice inactivity periods (R_{SID}). Let ρ be the conversation activity rate. Then:

$$R = \rho \cdot R_{ON} + (1-\rho) \cdot R_{SID} \tag{1}$$

where ρ is the conversation mean activity rate, p is the peak rate and $R_{SID}$is the mean rate during voice inactivity periods caused by the transmission of SID frames. For those codecs which do not generate SID frames, obviously $r_{SID} = 0$.

The peak rate depends on both the codec characteristics and the number of frames per packet. Thus, it is clearly given by:

$$R_{ON} = \frac{H + N_{fpp} L_{ACT}}{N_{fpp} T} \tag{2}$$

where H is the header size of the encapsulating protocols (i.e. 40 octets for the VoIP case,) L_{ACT} is the voice frame size and T is the frame generation period of a given codec. Table 1 shows the characteristics of some VoIP codecs.

Regarding to the R_{SID} factor of equation 1, an analytical expression for the VoIP transport case was deduced and validated in (Estepa et al., 2004). The deduction was based in the separation of the contribution of the header and the SID frames to the mean bit-rate so $r_{SID} = R_H + R_{fr}$. The contribution of the SID frames can be obtained by application of the Elementary Renewal Theorem (ERT) which states that the SID frames arrival long-term rate is the inverse of the expected inter-arrival time ($E[X] \cdot T$).

Table 1: Codec's characteristics.

Codec	Mode	L_{ACT}	L_{SID}	T (ms)	$E[X]$	P_1
G.729	-	10	2	10	7.33	0
G.723.1	6.3	24	4	30	13.05	0.27
	5.3	20	4	30		
AMR	4.75	12	5	20	7.47	0
	12.2	31	5	29	7.47	0

$$R_{fr} = \frac{L_{SID}}{T \cdot E[X]} \tag{3}$$

where L_{SID} is the size of a SID frame.

In VoIP, the contribution of the packet header generated during inactive periods follows the packet generation pattern imposed by the RFC 3551, where one packet header is sent every non-consecutive SID frame ($i > 1$). For consecutive SID frames, one packet header is sent every N_{fpp} frames, so both cases must be considered. Since the mean time between SID frames is given by ($E[X] \cdot T$), the header contribution (R_H) can be expressed as:

$$R_H = P_1 \cdot \frac{H}{N_{fpp} \cdot T \cdot E[X]} + (1-P_1)\frac{H}{T \cdot E[X]} \tag{4}$$

where P_1 stands for the probability of having two time-consecutive SID frames. Thus, for the VoIP case we have an overall mean bit-rate of:

$$R^{RFC3551} = \rho \cdot \left(\frac{L_{ACT}}{T} + \frac{H}{N_{fpp} \cdot T} \right) + \frac{(1-\rho)}{E[X] \cdot T} \cdot \left(L_{SID} + H \cdot \left(1 + \frac{P_1(1-N_{fpp})}{N_{fpp}} \right) \right) \tag{5}$$

The right side sum of equation 5 represents R_{SID}, which was calculated adding separately the contribution to the mean bit-rate of the packet headers (R^H_{SID}) and the SID frames (R^{fr}_{SID}).

3.2 Mean Bit-rate for *multi-SID* Packetization Scheme

The next step is to deduce the mean bit-rate $R^{multiSID}$ for the new packetization scheme. Clearly, R_{ON} does not change. Following the same approach from (Estepa et al., 2005) for R_{SID}, we separate the rate contribution by the packet headers and the by the SID frames. The novelty now is the packet generation schema and, thus, the header contribution (factor R^H_{SID}.) As the size of the packet header depends on the number of conveyed SID frames, an overhead of an extra byte per SID frame must be accounted whenever the *multi-SID* packet format is used, as shown in figure 2.

The contribution of the header in the binary rate can be readily deduced noting that for an inter-arrival time $X = i < N_{fpp}$, the number of SID frames carried on the packet is $\lceil \frac{N_{fpp}}{i} \rceil$, otherwise packets will transport only one SID frame. Also note that for time-consecutive SID frames (i.e. $X = i = 1$), the *multi-SID* payload is not applied. Thus, the resulting equation for R^H_{SID} is:

$$R^H_{SID} = \frac{1}{T \cdot E[X]} \cdot \left(\frac{H}{N_{fpp}} \cdot P_1 + \sum_{i=2}^{N_{fpp}-1} \frac{H+1}{\lceil \frac{N_{fpp}}{i} \rceil} \cdot P_i + \sum_{i=N_{fpp}}^{\infty} H \cdot P_i \right) \quad (6)$$

Adding equations 6,3 and 2 and reordering, we obtain the overall mean bit-rate using our proposal packetization scheme, which is:

$$R^{multiSID} = \rho \cdot \left(\frac{L_{ACT}}{T} + \frac{H}{N_{fpp} \cdot T} \right) + \frac{1-\rho}{E[X] \cdot T} \cdot \left[L_{SID} + H \cdot \left(1 + P_1 \cdot \left(\frac{1}{N_{fpp}} - 1 \right) \right) + \sum_{i=2}^{N_{fpp}-1} P_i \cdot \frac{1 + H \cdot \left(1 - \lceil \frac{N_{fpp}}{i} \rceil \right)}{\lceil \frac{N_{fpp}}{i} \rceil} \right] \quad (7)$$

From equation 7 and equation 1 the bandwidth gain can be directly computed.

4 VALIDATION AND NUMERICAL RESULTS

This section presents the results of a comparative study of the mean bit-rate achieved with both packetization schemes (RFC 3551 and *multi-SID*) which allow us to quantify the benefits of the *multi-SID* one, and validate the equations presented in previous sections.

4.1 Experiment Setup

The result above has been validated by using the testbed described in (Estepa et al., 2003), where 5 hours of conversations were recorded from a ISDN line in an low-noise office environment (i.e. $SNR > 20dB$). The raw audio files were encoded using the G.729B codec. This codec, highly available in any VoIP environment, holds the capability of generating SID frames and is widely referenced in the literature, so it will let us to compare our results with previous studies. The resulting sequences of frame types generated by the codec (i.e. ACT, SID or NULL) were processed to empirically determine the values of P_i and the conversation activity factor ρ. These sequences were also used to feed a packetization program that builds up the IP packets according to both packetization schemes presented in this paper: RFC 3551 and the *multi-SID* scheme. The output of the program shows the mean bit-rate for each conversation. We will use these values to validate the analytical expressions deduced here and to probe the bandwidth savings that the proposed packetization scheme provides over the traditional RFC 3551. A similar validation procedure can be followed to extend the results to other codecs.

4.2 Numerical Results

Figure 2 shows the mean-bit rate measured with the aforementioned simulations for both packetization schemes and the plot of the analytical models given by equations 7 and 5.

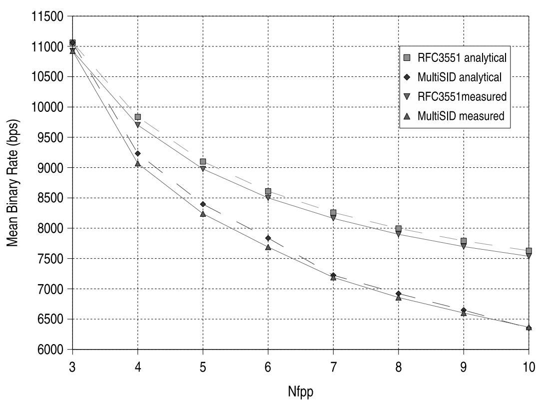

Figure 2: Mean Binary Rate for both *multi-SID* and RFC 3551 packetization schemes.

Figure 3 shows the difference of both packetization scheme s in terms of bandwidth. With the G.729 codec and N_{fpp}=9 (which implies a packetization delay of 90ms), the bandwidth reduction reaches a 14%.

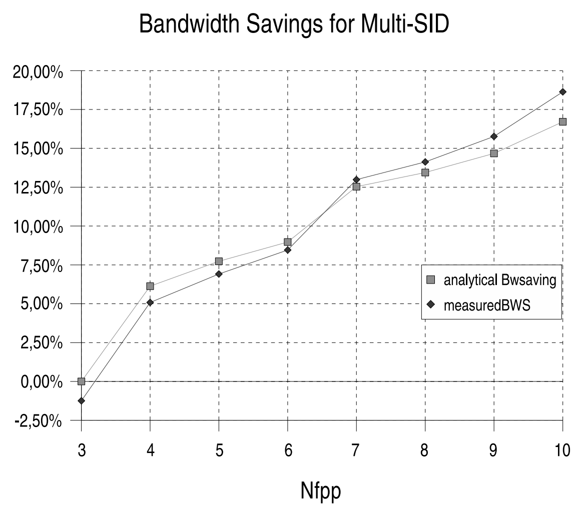

Figure 3: Bandwitdh saving for *multi-SID* respect to RFC 3551: analytical and measurement results.

In the Internet, the network delay and associated jitter are clearly dominant factors in the overall delay leaving no room for large packetization values (typically N_{fpp}=2 or 4 for G.729). However in corporate networks, where the network and jitter delay are low, the admissible packetization delay could potentially allow higher N_{fpp} values, yielding significant bandwidth savings with the *multi-SID* packetization scheme.

5 CONCLUSIONS

This paper presents a new packetization scheme that can be used in VoIP. This backward-compatible scheme permits sending packets loaded with multiple non-consecutively-generated SID frames, reducing the conversation's bandwidth requirement.

An analytical expression for bandwidth saving when new packetization scheme is used has been deduced. Experimental validation confirms that the new scheme improves the existing one up to 14% for admissible packetization values. The *multi-SID* packetization scheme is specially meaningful in the backbone of corporate networks where many voice sources are multiplexed.

REFERENCES

Estepa, A., Estepa, R., and Vozmediano, J. (2003). Packetization and Silence Influence on VoIP Traffic Profiles. *Lecture Notes in Computer Science*, 2899(1):331–339.

Estepa, A., Estepa, R., and Vozmediano, J. (2004). A New Approach for VoIP Traffic Characterization. *IEEE Communications Letters*, 8(10):644–647.

Estepa, A., Estepa, R., and Vozmediano, J. (2005). Accurate prediction of voip traffic mean bit rate. *IEE Electronic Letters*, 8(10):644–647.

Schulzrinne, H. (2003). RTP Profile for Audio and Video Conferences with minimal control. RFC 3551, Internet Engineering Task Force.

Zopf, R. (2002). Real-time Transport Protocol (RTP) payload for comfort noise (cn). RFC 3389, Internet Engineering Task Force.