Redundancy Control in Real-Time Internet Audio Conferencing

Isidor Kouvelas, Orion Hodson, Vicky Hardman and Jon Crowcroft
Department of Computer Science
University College London
Gower Street, London WC1E 6BT, UK

ABSTRACT *The use of redundant audio encoding has been advocated for lossy networks like the Internet[1, 2] as a way of reducing the impact of loss in audio-conferences. We present a model of loss and determine how the amount of redundancy should be varied with the loss rate. In addition, we make loss measurements and make a preliminary investigation of the position of redundant encodings relative to the original encoding.*

1 INTRODUCTION

IP Multicast allows real-time multiway audio and video conferencing over the Internet, and is now moving from piloting stage to a service in countries such as the UK and the US [3]. Whilst video and shared data are essential to many distributed tasks, sufficient quality audio is a essential for real-time interaction. The effect of packet loss on audio quality is currently the largest obstacle to the realisation of multimedia conferencing over the Mbone.

When packet loss occurs, a fill-in section of audio is provided by the receiver to maintain timing synchronisation. The most common technique used in first generation audio tools is the use of silence as the replacement, but this results in degradation of speech quality, which increases rapidly with loss rate and packet size. Mbone audio packet sizes range from 20 to 160ms, with 40ms packets being the default. For interactive conferencing the coding delay and propagation time should be small ($\leq$ 250ms[4]), however smaller packets require more work by the routers and have a larger relative packet header overhead [5]. The loss of larger packets has a more pronounced effect at the receiver. Our experience shows that a 10% loss rate for 40ms packets is the highest that can be tolerated with silence substitution before a significant drop in speech intelligibility occurs[1].

Receiver based techniques [6, 7, 8] can be applied to mask the loss by utilising data on one or both sides of the loss to fill in with appropriate pitch waveforms for voiced speech and noise for unvoiced speech. Receiver based repair techniques are fundamentally limited by the non-stationary nature of speech, thus the longer the loss the less chance of a receiver has making an appropriate repair. Phonemes are typically 80-100ms in length[9], thus the longer the gap in comparison to the length of a phoneme the less the chance of receiver repair being made successfully.

The problem is compounded by the packet loss on international links, which is commonly 20-25%. This means that audio tools that only use receiver based mechanisms to counteract packet loss cannot provide the minimum quality required to conduct a conversation over international and sometimes national links.

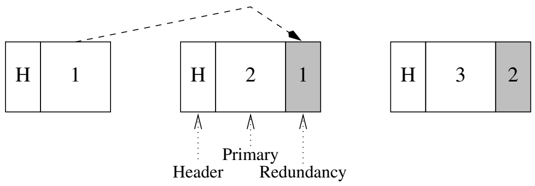

Figure 1: Adding redundancy to audio packets

The Robust-Audio Tool, RAT[10], from UCL uses an audio redundancy technique that was developed in collaboration with researchers at INRIA[11]. Redundant audio information is piggy-backed onto a later packet (figure 1), which is preferable to the transmission of additional packets as this increases the amount of overhead for the transmission in respect to routing decisions and bandwidth. This can easily be accomplished over the Mbone, since IP allows variable length packets. If a packet is lost then the missing information may be reconstructed at the receiver from the redundant data that arrives in the following packet, provided that the average number of consecutively lost packets is small. For the redundant blocks a lower quality encoding with higher compression can be used, such as LPC, to reduce the overhead in packet sizes[1].

Coder	Bandwidth (kb/s)	Complexity (10^x ips)	Quality (MOS)
μ-law	64	4	4.2
ADPCM	32	5	4.1
GSM	13.2	7	3.6
G.723.1	5.3/6.3	7.3	3.5/4
LPC	2.4-5.6	6	2.4-3

Table 1: A comparison of the speech codecs (after Cox and Kroon [12]).

The addition of a piggy-backed redundant encoding for the previous packet can provide fill-in audio for one lost packet. If two or more consecutive packets are dropped by the network then both the primary and redundant information for some period of time will be lost. Theoretically by increasing the number of redundant blocks that we add to each packet we can cover for a loss of any length. In practice there are limitations on the packet size we can use and the maximum bandwidth it is desirable for our audio stream to use. Receivers that are suffering from loss must ensure their buffers are large enough to cater for the amount of redundancy and so incur extra decoding delay, whereas receivers not without the loss incur no extra delay.

In this paper we investigate the problem of controlling the amount of redundancy that should be used under different network conditions. We first make an assumption about the model of packet loss on the network and show the limits to which audio redundancy can be useful. We then present results from real network measurements and show how the deviations from our assumptions affect the results. The aim being to develop a mechanism that automatically adapts the number of levels of redundancy transmitted based on feedback from receivers. This feedback can be obtained from RTCP reports[5] that are periodically transmitted by each receiver.

Bolot and Vega-Garcìa[2] have also collected and analysed packet losses for real-time audio streams over the Internet for unicast and multicast connections. They found comparable loss variations and suggested a relative reward based approach to covering losses. The reward is calculated from the increase in bandwidth from using redundancy against the decrease in loss in the audio stream. In this paper we consider benefits to the decoded stream rather than attributing scores based on the overhead and quality improvement. Handley has made similar observations of the loss patterns reported in this paper for video data over the Mbone [13].

2 REAL-TIME AUDIO OVER THE INTERNET

The Real-Time Protocol (RTP) is used to transport audio data over the network. The pertinent features for the purposes of our discussion are that every packet includes a sequence number, a timestamp, and a payload indicator for the data it carries[5, 14]. A separate channel carries Real-Time Control Protocol (RTCP) information. One of the purposes of RTCP is to relay from receivers to senders the observed packet loss rates.

Our aim is to use the information contained in RTCP reports and gathered through intelligent probing to calculate an optimal amount of redundancy a transmitter should put in the packets and an optimal spacing of the redundant blocks within the packet. For the purposes of the analysis in this paper, the optimal redundancy is the amount of redundancy that must be added to have a particular loss rate in the decoded audio. We refer to the loss in the decoded audio as the *stream loss*. The optimal spacing is the amount of offset of the redundancy relative to the primary encoding and this can be determined from loss measurements.

3 ANALYSIS

In transporting audio across the Internet we are only concerned with packet loss and delay. Bit errors rates are small ($\leq 10^{-5}$) and cause whole packets to be discarded if detected by the IP header checksum. A first approximation is that the losses occur independently and and can be modelled by a Bernoulli random variable.

Assume a packet loss rate p over the network and n levels of audio redundancy. In order for a unit not to arrive at the receiver, all of the n packets carrying it have to be lost. Hence the perceived unit loss rate l at the receiver will be p^n. This can also be shown as follows. The probability $P(k)$ of the network loosing k consecutive packets and correctly delivering packet $k+1$ is given by:

$$p^k(1-p)$$

So the probability of the network loosing n or more packets is:

$$l = \sum_{k=n}^{\infty} p^k(1-p) = p^n \tag{1}$$

By solving for n we can calculate the number of levels of redundancy that are required to achieve a desired unit loss rate l at the receiver given the network packet loss rate p:

$$n = \left\lceil \frac{\log(l)}{\log(p)} \right\rceil \tag{2}$$

Figure 2 shows the number of levels of redundancy required to achieve different stream loss rates at the receiver against network loss and was plotted using equation 2.

Figure 3 shows the number of levels of redundancy required to achieve 5%, 10% and 20% stream loss rates at the receiver against network loss rate. A 5% loss rate in the audio stream represents a reasonable loss rate that can be patched using receiver techniques like waveform substitution. Even with the highest available compression used for the redundancy (at the moment LPC) it is not reasonable to expect more than 5 or 6 levels since it increases the end-to-end delay. For instance 5 levels of redundancy and 20ms packets, requires an extra delay of 100ms meaning that the end-to-end delay must be less than 150ms to remain interactive.

With lower compression schemes, such as ADPCM, maximum packet size becomes a limiting factor. The curves show that above a network loss rate of about 50% it becomes very hard to consistently cover for the packet loss since the number of required redundancy levels becomes very high. There should be a cutoff point for the

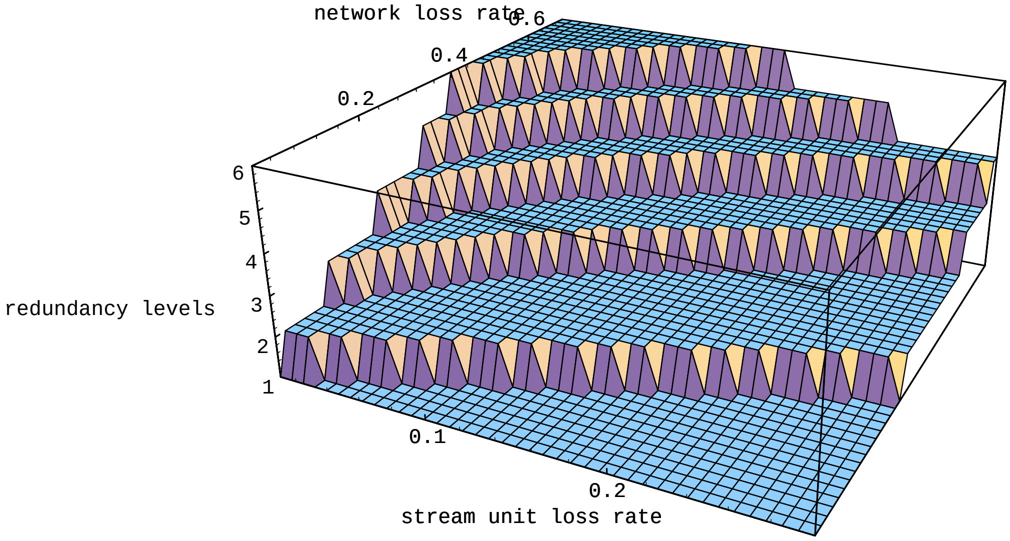

Figure 2: Number of redundancy levels required against loss rate and perceived quality

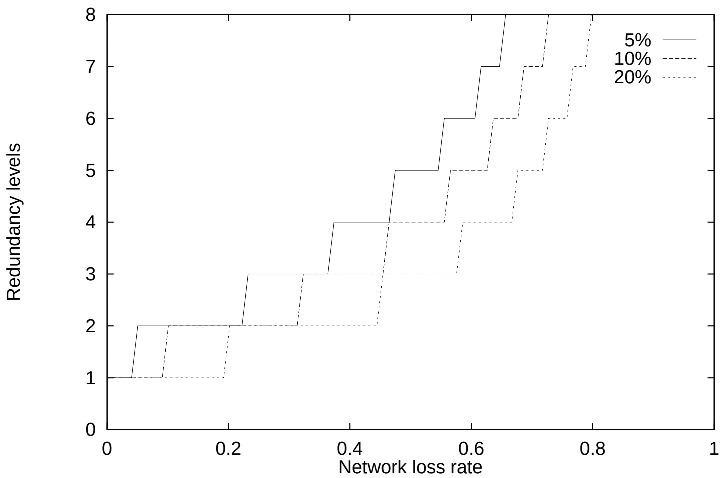

Figure 3: Number of redundancy levels required against network loss rate

network loss rate where the audio application decides, or indicates, that interactive communication is not possible.

$$\begin{aligned} \mathrm{E}\,(\text{gap length}) &= \sum kp^{k-1}(1-p) \\ &= \frac{1}{1-p} \end{aligned} \tag{3}$$

Another potential measure for determining the number of redundant encodings to use per packet can be the average packet gap length of the network. This is calculated in equation 3.

4 MEASUREMENTS

In order to confirm the above supposition, and to obtain data from which a design could be made, loss statistics were collected using a variety of both unicast and multicast links between UCL, France, Greece and the US. The results are not an exhaustive set of measurements, but rather are indicative of the conditions that might be expected. A number of data sets were collected for different loss rates, by looking for links that showed the approximate desired loss rate. Each data set has been constructed using 336 byte packets, generated at 80ms intervals, which is typical of voice traffic generated by current audio tools[1]. Approximately 10,000 packets per data set were collected.

We find that packet loss patterns are not truly random since the measurements deviate from the hypothesised results. Figure 4(left) compares the measured average packet loss length with the expected loss length

[1]The audio packet size is coded for 80ms of 8000kHz 4-bit monoaural ADPCM.

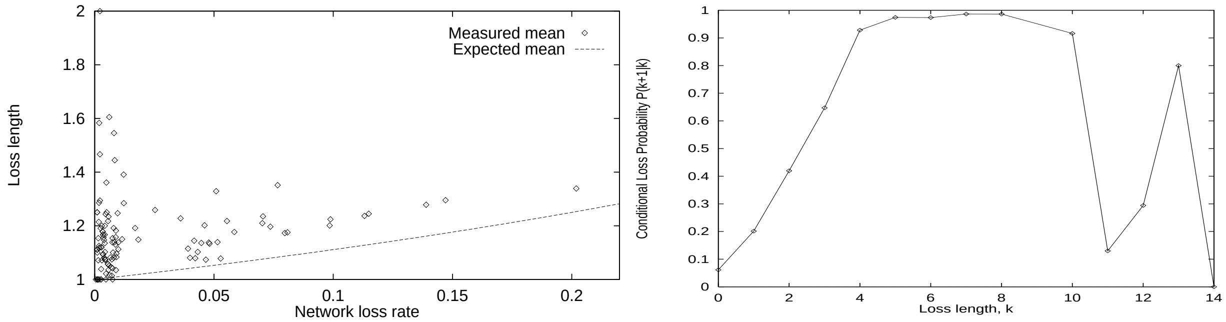

Figure 4: Expected and observed mean packet gap length against network loss rate (left). Conditional loss probability of losing the next packet given loss of k packets (right).

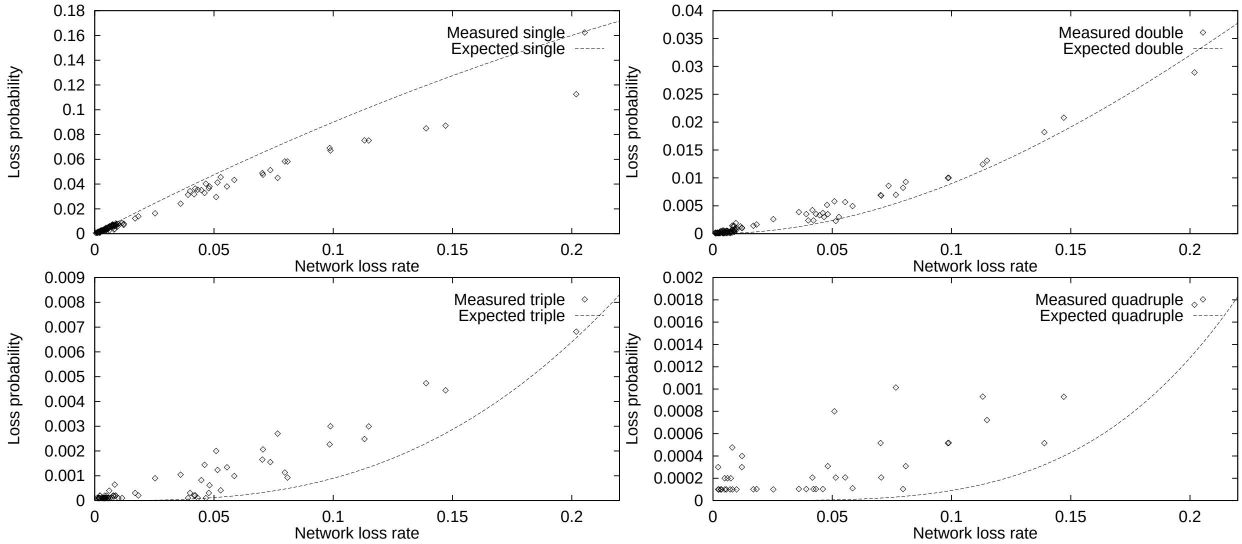

Figure 5: Expected and observed loss length probabilities against network loss rate. (Top row: single packet loss events, double packet loss events. Bottom row: triple packet loss events, quadruple packet loss events).

given that at least one packet has been lost. For low network loss rates ($< 10\%$) we observe high variation in the mean loss length. Closer examination of the datasets indicates that the deviations are caused by a few long duration losses (i.e. 10's of packets). This is attributable to events in the network causing transient burst losses as illustrated by the conditional loss probability distribution shown in figure 4(right).

For the Bernoulli model the conditional loss probability is constant irrespective of the number of packets. Figure 4(right) contradicts this. For this particular set of measurements the probability that the loss continues after 4 losses is high up until 10 losses. After 10 losses the readings may be attributable to noise as less than 0.3% of the losses were in this range. We observe similar patterns in other datasets. We believe the burst losses to be caused by a range of network phenomena including short lived TCP-connections, synchronization between TCP stream, and routing updates.

Each of the graphs in figure 5 shows the deviation of the measurements from the expected results for a different gap length. We observe that because of an existing degree of correlation in packet loss patterns, the measured values deviate from the theoretical ones. The single packet loss graph shows lower actual losses than expected whereas the longer gap length graphs show that the actual values are larger.

5 REDUNDANCY SPACING

Because the measured loss patterns display some degree of correlation, if packet n is lost then there is a higher than average probability that packet $n + 1$ will also be lost. This indicates that an advantage in perceived audio quality at the receiver can be achieved by offsetting the redundancy in the packets. Figure 6 shows how this can be achieved. The change is that packet n carries the redundancy for packet $n-2$ instead of $n-1$. Increasing the distance of the redundancy from the primary encoding, forces the receiver to wait longer before playing out audio in received packets to give a chance for the redundant block to arrive. The further the redundancy offset is, the longer the playout delay at the receiver has to be.

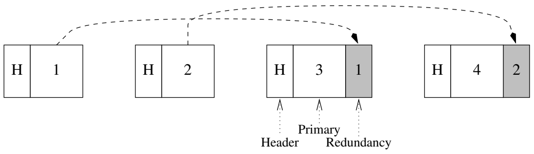

Figure 6: Offsetting redundancy

Figure 7 shows the change in perceived loss rate at the receiver when the redundancy is offset by one packet (as shown in figure 6). Positive values indicate an improvement whereas negative values indicate a degradation. It is clear that for the majority of measurements the perceived loss rate is reduced by offsetting the redundancy. There is also a visible trend indicating that for higher loss rates the improvement is more significant.

6 CONCLUSIONS

The deviations from the Bernoulli model imply that senders will require additional information from receivers to control the number of levels and offset of the redundancy. This information could be in the form of packet gap length histograms. Since RTCP is extensible with application specific message components, this can be readily accommodated and is currently being implemented. The Bernoulli model described in section 3 is still useful as it represents a lower bound on the higher loss length graphs presented in figure 5.

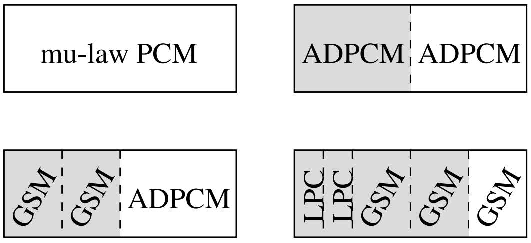

Figure 8: Constant bandwidth equivalents with variable amounts of redundant audio (shaded grey).

The aim of any adaption scheme should be to achieve optimal audio quality and fairly share available bandwidth in the network. A first step in achieving fairness is to avoid increasing the amount of data being transmitted when packet loss is observed. A control algorithm can be tuned to maximise perceived audio quality at the receivers for a given desired network bandwidth usage. This can be done by increasing the compression of the primary encoding and trading it for an increase in redundancy. An example is shown in figure 8.

The measurements presented here and elsewhere [2, 13] there is a strong argument for offseting the redundant encoding from the primary encoding, particularly in non-interactive scenarios. The appropriate amount of offset can be determined by reception reports. For interactive applications the amount of offset that maybe applied is limited by the end-to-end network delay.

7 ACKNOWLEDGEMENTS

We would like to acknowledge the contributions of Lorenzo Vicisano, Anna Watson, Colin Perkins, and the networked multimedia research group at UCL.

This work was partially supported by the EPSRC(Award 95011276), British Telecommunications plc(Agreement ML722854), and the MERCI project (Telematics for Research Project RE:1007).

BIBLIOGRAPHY

[1] Vicky Hardman, Martina Angela Sasse, Mark Handley, and Anna Watson. Reliable audio for use over the Internet. In *International Networking Conference (INET)*, September 1995.

[2] J.-C. Bolot and A. Vega-García. The case for FEC-based error control for packet audio in the internet. *ACMMultimediaSystems*, 1997.

[3] M.R. Macedonia and D.P. Brutzman. Mbone provides audio and video over the internet. *IEEE computer*, pages 30–36, April 1994.

[4] Paul T. Brady. Effects of transmission delay on conversational behavior on echo-free telephone circuits. *Bell System Technical Journal*, 50:115–134, January 1971.

[5] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: a transport protocol for real-time applications. Request for comments RFC 1889, Internet Engineering Task Force, January 1996.

[6] D.J. Goodman, G.B. Lockhart, O.J. Wasem, and W.-C. Wong. Waveform substitution techniques for recovering missing speech segments in packet voice communications. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-34(6):1440–1448, December 1986.

[7] O.J. Wasem, D.J. Goodman, C.A. Dvorak, and H.G. Page. The effect of waveform substitution on the quality of pcm packet communications. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(3):342–348, March 1988.

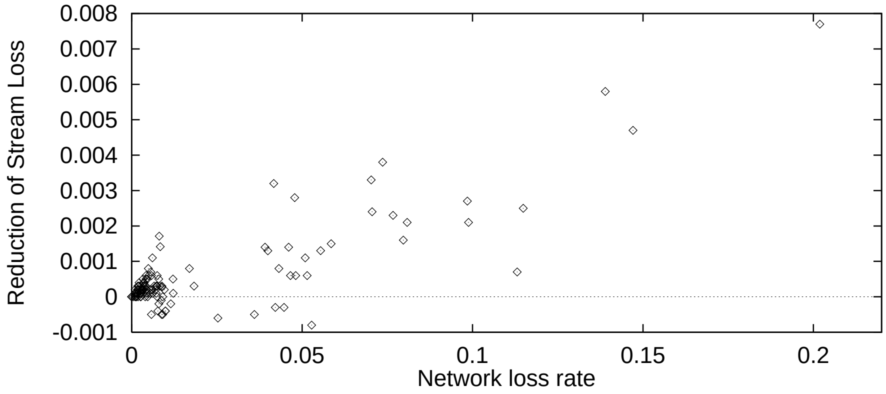

Figure 7: Perceived loss rate difference between normal and offset redundancy against network loss rate

[8] H. Sanneck, A. Stenger, K. Bem Younes, and B.Girod. A new technique for audio packet loss concealment. In *IEEE Global Internet 1996*, pages 48–52. IEEE, December 1996.

[9] R.M. Warren. *Auditory Perception: A new synthesis*. Pergamon Press, 1982.

[10] V. Hardman, M.A. Sasse, and I. Kouvelas. Successful multi-party audio communication over the internet. To appear Communications of the ACM.

[11] J.-C. Bolot, S. Fosse-Parisis, M. Handley, V. Hardman, O. Hodson, I. Kouvelas, C. Perkins, and A. Vega-García. RTP payload for redundant audio data. Internet draft (work-in-progress) *draft-ietf-avt-redundancy-00.txt*, March 1997.

[12] R.V. Cox and P. Kroon. Low bit-rate speech coders for multimedia communication. *IEEE Communications Magazine*, pages 34–41, December 1996.

[13] M.J. Handley. An examination of mbone performance. USC/ISI Research Report: ISI/RR-97-450 (available from http://buttle.lcs.mit.edu/ mjh/mbone.ps).

[14] H. Schulzrinne. RTP profile for audio and video conferences with minimal control. Request for comments RFC 1890, Internet Engineering Task Force, January 1996.