CCITT SGXV
Working Party XV/1
Experts Group for ATM Video Coding

Document AVC-376
November 1992

SOURCE: Telecom Australia*

TITLE: Theoretical Study on the Use of Priority

PURPOSE: Information


## Abstract

We analyse the performance of predictive coding in a statistically multiplexed situation. In particular we study techniques for improving loss resilence, namely adaptive leaky prediction and prioritisation coupled with layered coding. The results indicate that both techniques, and especially prioritisation, are extremely useful when the utilisation is high.


## 1. Introduction

One way of making meaningful comparisons among different loss resilience coding schemes in a statistically multiplexed environment (e.g. layered coding with prioritised transport versus non-layered coding with error concealment) is by means of the "load-distortion function", where for a given type of source (with a given peak and mean rate etc) the number of such sources that can be supported on a given link with a given capacity is plotted against the reconstruction error variance or SNR achievable using a particular scheme. This idea originates from earlier work by Garrett and Vetterli [1], and takes a joint, closed and combined view of the system including source coding and the network transport as a whole. Instead of taking the cell loss ratio, for example, as given, the way the source coder behaves directly impacts on the network performance. While such a view is arguably simplistic considering a real situation which involves heterogeneous traffic, and does not take into account operational and other costs (e.g. in providing and making use of priority), it is perhaps the most appropriate way to compare the relative performance of different coding arrangements in terms of efficiency. The appropriateness of such an approach also extends to, for example, the study on CBR/VBR.

In what follows, theoretical comparisons on relative performance among various techniques, including the use of basic concealment, adaptive leaky prediction and multiple priorities, will be given on the basis of the load-distortion function. DPCM is used as the underlying predictive coding scheme. For more details please refer to [2] and [3].

## 2. Source and multiplexing models

We assume $N$ i.i.d. exponential on/off sources with activity factor $s$ sharing a link with capacity $C$. The aggregate can be viewed as the result of multiplexing $N$ (a fixed number) VBR sources of a particular type (Bernoulli), each with a peak/mean ratio of $s^{-1}$. Alternatively it can be seen as the result of multiplexing a time-varying number of CBR sources (at most $N$, on average $sN$).

When active a source transmits at a rate of $p$, $\alpha$ of which is high priority in the prioritised cases.

We assume zero-buffer in the multiplexer and use the fluid approximation. For the non-prioritised case the average loss ratio is given by [4]

$$\xi = \frac{1}{Nps} \sum_{n=\left\lceil \frac{C}{p} \right\rceil}^{N} (pn-C)\pi_n \qquad (1)$$

* For more information please contact Ming Chan by email: m.chan@trl.oz.au

where $\pi_n$ is the binomial distribution

$$\pi_n = \binom{N}{n} s^n (1-s)^{N-n}. \tag{2}$$

There are two slightly different ways of providing multiple priorities. In *Selective Discard*, packets are discarded during congestion, with precedence being given to high priority packets as indicated by the CLP bit or the VCI/VPI. The average loss ratios in high and low priorities are given by [2]

$$\xi_h = \frac{1}{N\alpha ps} \sum_{n=\left\lceil \frac{C}{\alpha p} \right\rceil}^{N} (\alpha pn - C)\pi_n \tag{3a}$$

$$\xi_l = \frac{1}{N(1-\alpha)ps} \left[ \sum_{n=\left\lceil \frac{C}{p} \right\rceil}^{\left\lfloor \frac{C}{\alpha p} \right\rfloor} (pn - C)\pi_n + \sum_{n=\left\lfloor \frac{C}{\alpha p} \right\rfloor + 1}^{N} (1-\alpha)pn\pi_n \right]. \tag{3b}$$

Note that high and low priority losses are correlated, and the various joint probabilities of loss are given by

$$\xi_{lh} = \xi_h, \qquad \xi_{0h} = 0, \qquad \xi_{l0} = \xi_l - \xi_h, \qquad \xi_{00} = 1 - \xi_l. \tag{3c}$$

The alternative to selective discard is to divide explicitly the total capacity $C$ among *Independent Channels*, with $\beta C$ allocated to the high priority channel. High and low priority losses in this case are independent and the various probabilities are given by [2]

$$\xi_h = \frac{1}{N\alpha ps} \sum_{n=\left\lceil \frac{\beta C}{\alpha p} \right\rceil}^{N} (\alpha pn - \beta C)\pi_n, \tag{4a}$$

$$\xi_l = \frac{1}{N(1-\alpha)ps} \sum_{n=\left\lceil \frac{(1-\beta)C}{(1-\alpha)p} \right\rceil}^{N} \left[(1-\alpha)pn - (1-\beta)C\right]\pi_n, \tag{4b}$$

$$\xi_{lh} = \xi_l \xi_h, \qquad \xi_{0h} = (1-\xi_l)\xi_h, \qquad \xi_{l0} = \xi_l(1-\xi_h), \qquad \xi_{00} = (1-\xi_l)(1-\xi_h). \tag{4c}$$

Note that selective discard always performs better because of the "cross-multiplexing" between high and low priority traffic, but requires "smarter" switches with the capability to selectively discard.

## 3. The effect of loss on reconstruction quality

We only deal with the one dimensional case, with an active source being 1st order Gauss-Markov. In the context of video coding this is like studying interframe coding of a single pixel image sequence, or pure interframe coding (with or without MC) without exploiting the remaining spatial correlation. The coding scheme studied is DPCM with 1st order linear prediction. While the setup is crude, it nevertheless captures most of the essence of video coding, especially in terms of its performance in view of channel loss. We assume that loss of a sample can be detected at the decoding end, with a MMSE replacement. We have therefore assumed the use of the most natural concealment throughout - a lost sample will be replaced by its natural prediction (e.g. interframe prediction in video coding, or motion compensated prediction if MC is used).

We further assume

i. 1st order Gauss-Markov input with $R_{xx}(1) = \rho$,

ii. 1st order linear prediction with prediction coefficient $h_1$,

iii. pdf-optimised MMSE quantisation and

iv. fine quantisation and white prediction error†.

### 3.1 Non-prioritised DPCM

For the non-prioritised case the reconstruction error variance conditioned on both low and high priority loss is given by [3]

$$\sigma_w^2 = \frac{1+h_1^2-2h_1\rho}{(1-\varepsilon_q^2 h_1^2)(1-h_1^2)} \left[\xi + (1-h_1^2-\xi)\varepsilon_q^2\right]\sigma_x^2 \qquad (5)$$

where $\varepsilon_q^2$ is the quantiser performance factor [5] and depends on $B$, the number of bits used in quantisation. For Gaussian pdf and $B=8$ $\varepsilon_q^2=4.119E-5$. The special case $h_1=0$ corresponds to PCM, while the special case $h_1=\rho$ corresponds to non-adaptive DPCM which is optimal when there is no loss ($\xi=0$). Note that using $h_1=1$ (as is quite common in video coding) when $\xi\neq0$ would be disastrous if not for other mechanisms such as periodic refresh and I-frames. In all our comparisons we use either optimal source prediction ($h_1=\rho$, and therefore leaky) or adaptive leaky prediction ($h_1<\rho$ for $\xi>0$).

## 3.2 Prioritised DPCM

A simple way to prioritise DPCM is to simply break up each $B$-bit channel codeword into two, with the most significant $\alpha B$ bits in high priority. (In video coding using hybrid DPCM/DCT this is like putting low sequency DCT coefficients in high priority.) The reconstruction error variance is given by [3]

$$\sigma_w^2 = \frac{1+h_1^2-2h_1\rho}{(1-\varepsilon_q^2 h_1^2)(1-h_1^2)} \left[\xi_{lh} + \xi_{0h}\varepsilon_{q''}^2 + \xi_{l0}\varepsilon_{q'}^2 + (\xi_{00}-h_1^2)\varepsilon_q^2\right]\sigma_x^2, \qquad (6)$$

where $\varepsilon_{q'}^2$ and $\varepsilon_{q''}^2$ are the performance factors of equivalent quantisers when low and high priority bits alone are lost respectively [2]. $\varepsilon_{q'}^2$ is close to the performance factor of an optimal $\alpha B$-bit quantiser, while $\varepsilon_{q''}^2$ is very close to 1. See the following table for the case of Gaussian input and $B=8$.

| $\alpha B$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $\varepsilon_{q'}^2$ | 0.3634 | 0.1240 | 0.0361 | 9.760e-3 | 2.538e-3 | 6.475e-4 | 1.637e-4 |
| $\varepsilon_{q''}^2$ | 0.999416 | 0.999799 | 0.999984 | 0.999999 | 1. | 1. | 1. |

The resulting system is *non-embedded* [6], meaning that the effect of losing the low priority information is more severe than what it could have been, but retains 100% efficiency compared to the non-prioritised case when there is no loss. Proposals in MPEG on scalability which simply put DCT coefficients of different sequency into different layers are exactly of this nature, with the so called "drift" problem.

## 3.3 Embedded DPCM

The embedded DPCM system was invented by Ching [7] and studied in detail by Goodman [6]. By anticipating possible losses in the channel, and accordingly removing information from the prediction loop that are likely to be lost, better error resilience can be achieved compared to prioritised but non-embedded DPCM. The system proposed in [8] is of exactly the same principle.

Considering both low and high priority losses, the error variance is given by [3]

$$\sigma_w^2 = \frac{1+h_1^2-2h_1\rho}{(1-\varepsilon_{q'}^2 h_1^2)(1-h_1^2)} \left[\xi_{lh}+\xi_{0h}h_1^2 + \xi_{0h}(1-h_1^2)\varepsilon_{q''}^2 + (\xi_{l0}-(1-\xi_{00})h_1^2)\varepsilon_{q'}^2 + \xi_{00}(1-h_1^2)\varepsilon_q^2\right]\sigma_x^2. \qquad (7)$$

By setting all loss probabilities to zero and $\xi_{00}$ to 1 in (7) it can be seen that the efficiency of the system at no loss is slightly inferior to the two previous systems.

## 4. Comparing end-to-end performance

(5), (6) or (7) can be combined with (3) (for selective discard) or (4) (for independent channels) so that the $SNR$ of the various systems are expressed in terms of $N$ for a given $C$, $s$, $p$ and $\rho$. Also involved are parameters related to the actual coding, $h_1$ and $B$, and parameters related to source partitioning, $\alpha$. In the case of independent channels the network partitioning, $\beta$, is yet another variable. The values of $h_1$, $\alpha$ and $\beta$ can be optimised for a given $N$ to maximise the eventual $SNR$. Alternatively by using fixed values of $h_1$, $\alpha$ and $\beta$ for all $N$ we can investigate the performance of various non-adaptive systems.

Here we only show the results of using selective discard, which works slightly better in all cases compared to independent channels. All results are for a given $C=100$, $p=2$, $s=0.5$, $\rho=0.95$ and $B=8$.

With reference to Figure 1, at moderate loading (with low loss ratio) PCM (plot 1) works well. As loading increases the performance quickly deteriorates. Using two priorities with an adaptive $\alpha$ (source partitioning) results in significant improvement (plot 2). Using a fixed $\alpha=0.5$ also results in substantial improvement (plot 3). At moderate loading DPCM with fixed prediction ($h_1=\rho$) also works well (plot 4), as in PCM. Again as loading increases the performance quickly deteriorates, to the point where it is no better than PCM. With adaptive leaky prediction the situation is improved (plot 5)†. Using two priorities with adaptive $\alpha$ (plot 6) the improvement is far more significant. Adding adaptive leaky prediction (plot 7) results in further improvement. Compared to simple DPCM (plot 4) and even DPCM with adaptive leaky prediction (plot 5), using two priorities with fixed $\alpha$ and fixed prediction (plot 8) still represents significant improvement. Embedded DPCM with adaptive $\alpha$ (plot 9) works very well indeed, and adding adaptive leaky prediction (plot 10) makes little difference. Finally with both $\alpha$ and the prediction fixed (plot 11), embedded DPCM using two priorities still works very well. While it is not apparent from the figure, its performance at very low loss ratio is indeed slightly inferior compared to a non-embedded system.

Note that in the fixed $\alpha$ cases one can always choose and fix $\alpha$ at values which favours certain regions of loading which are of interests (e.g. $N<100$). Starting with some broad assumptions (e.g. intended utilisation) it is therefore feasible to make use of multiple priorities with substantial gains in efficiency even without dynamically adapting the source coding to the network condition.

## 5. References


[1] M. Garrett and M. Vetterli, "Joint Source/Channel Coding of Statistically Multiplexed Real Time Services on Packet Networks", *to be published in IEEE Trans on Communications* (submitted March 1991).

[2] M.H. Chan and J.P. Princen, "On the use of loss priority in the transport of variable rate real-time information", *submitted to IEEE Trans on Networking* (submitted September 1992).

[3] M.H. Chan, "Prioritised Statistical Multiplexing and Recursive Coding", *submitted to ICC'93*, Geneva (May 1993).

[4] N. Yin and M. Hluchyj, "A Dynamic Rate Control Mechanism for Source Coded Traffic in a Fast Packet Network", *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 7, pp. 1003-1012 (September 1991).

[5] N. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall (1984).

[6] D. Goodman, "Embedded DPCM for Variable Bit Rate Transmission", *IEEE Trans on Communications*, Vol. 28, No. 7, pp. 1040-1046 (July 1980).

[7] Y.C. Ching, "Differential pulse code communications system having dual quantisation schemes", *U.S. Patent 3781685* (Dec 1973).

[8] G. Morrison and D. Beaumont, "Two-layer video coding for ATM networks", *Signal Processing: IMAGE COMMUNICATION*, Vol. 3, No. 2-3, pp. 179-195 (June 1991).


† Strictly speaking the assumption on white prediction error is valid only when source prediction is optimal ($h_1=\rho$). Equations (5), (6) and (7) are nevertheless accurate if the loss process is memoryless, otherwise the results given by (5), (6) and (7) for $h_1\neq\rho$ would be over-estimates of the actual SNR.

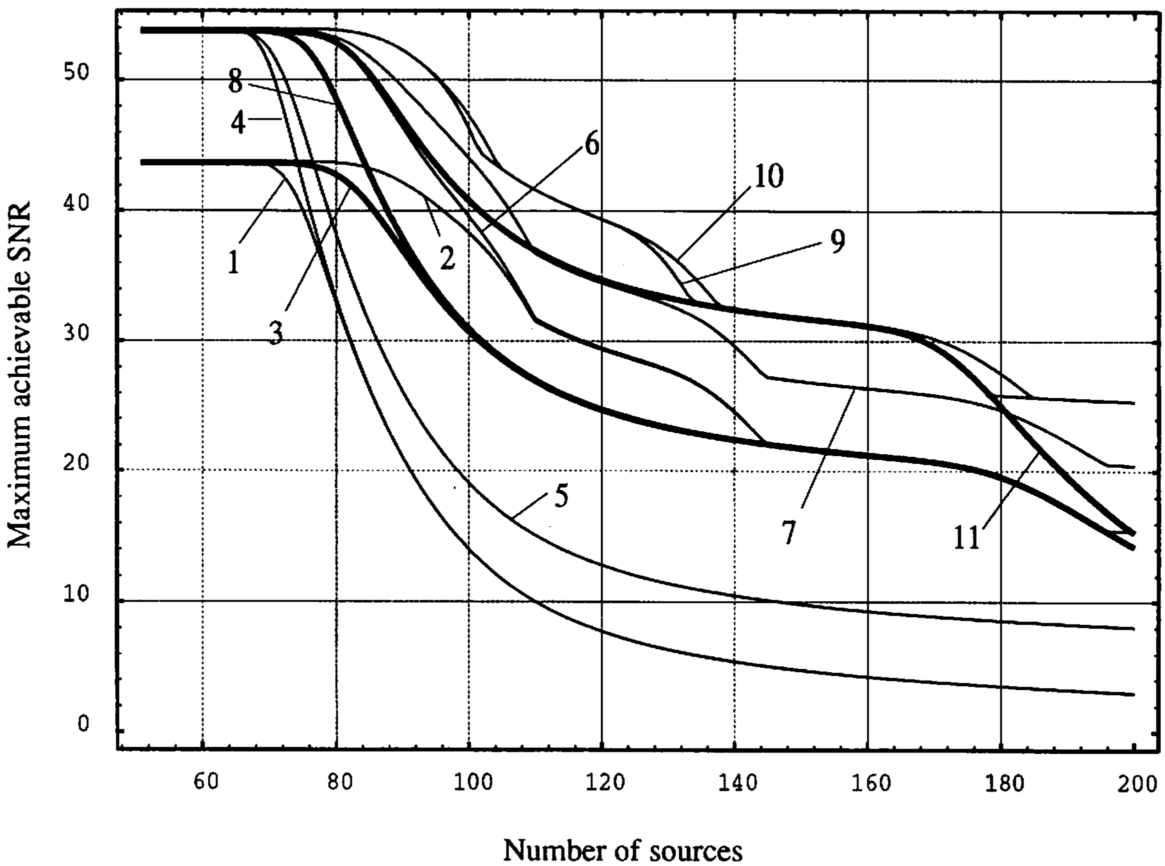


Figure 1. Maximum achievable SNR.

1. PCM
2. Prioritised/embedded PCM
3. Prioritised/embedded PCM with fixed $\alpha = 0.5$
4. DPCM
5. DPCM with adaptive prediction
6. Prioritised but non-embedded DPCM
7. Prioritised but non-embedded DPCM with adaptive prediction
8. Prioritised but non-embedded DPCM with fixed $\alpha = 0.5$
9. Prioritised and embedded DPCM
10. Prioritised and embedded DPCM with adaptive prediction
11. Prioritised and embedded DPCM with fixed $\alpha = 0.5$

1st order Gauss-Markov input with correlation coefficient of 0.95.
8 bit PCM/DPCM.
Exponential on/off sources with peak-to-mean ratio of 2.
Total capacity is 100 times the mean.
Multiple priorities implemented using selective discard.