Comparison of Transmitter-Based Packet-Loss Recovery Techniques for Voice Transmission

Moo Young Kim and W. Bastiaan Kleijn

Sound and Image Processing Lab., Dept. Signals, Sensors and Systems
KTH (Royal Institute of Technology), Sweden
{moo.young.kim,bastiaan.kleijn}@s3.kth.se

Abstract

To facilitate real-time voice communication through the Internet, forward error correction (FEC) and multiple description coding (MDC) can be used as low-delay packet-loss recovery techniques. We use both a Gilbert channel model and data obtained from real IP connections to compare the rate-distortion performance of different variants of FEC and MDC. Using identical overall rates with stringent delay constraints, we find that side-distortion optimized MDC generally performs better than Reed-Solomon based FEC. If the channel condition is known from feedback through the Real-Time Control Protocol (RTCP), then channel-optimized MDC can be used to exploit this information, resulting in significantly improved performance.

1. Introduction

Real-time voice communication over IP networks can offer a lower cost and more flexible service than conventional communication networks. However, due to the unreliable nature of packet delivery, the quality of the received data stream is often adversely affected by packet loss and delay. In real-time communication systems, which have very strict delay constraints, packets delivered after playout time must be treated as lost. Various techniques are being used to counter the quality degradation resulting from packet loss. However, the comparative performance of these methods is currently not well understood. This paper contributes towards the understanding of the relative merits of a number of techniques used to recover packet loss.

Increased knowledge about the effectiveness of methods that address packet loss is of great practical value. Numerous coding standards are in common use in the packet-network environment. For speech transmission, these include G.711, G.723.1, G.729, and Internet low-bit-rate coding (iLBC) [1, 2]. All of these codecs include adaptations to the packet-network environments. The selection of the best method to counter packet loss was constrained by the strict separation between network layers and a lack of knowledge of comparative performance. In our aim to extend knowledge of comparative performance, we will not consider the constraints caused by a lack of communication between network layers.

Packet-loss recovery techniques can be divided into two classes: transmitter-based and receiver-based techniques [3]. When losses occur that cannot be repaired with transmitter-based schemes, receiver-based error concealment schemes, such as insertion, interpolation, and regeneration, produce a replacement for a lost packet. The basic transmitter-based mechanisms available to recover from packet loss are: automatic repeat request (ARQ), interleaving, layered coding (LC), forward error correction (FEC), and multiple description coding (MDC). If the end-to-end delay is of secondary importance, ARQ is the most powerful recovery technique that can be used. Interleaving reduces the number of effective burst errors, but it also requires a large delay to be effective against burstiness. If the delay is of critical importance, low-delay schemes, such as FEC, MDC, and LC, are preferred.

In FEC, lost data are recovered at the receiver without further reference to the transmitter. Both the original data and redundant information are transmitted to the receiver. Two kinds of redundant information exist: those that are either dependent and those that are independent of the medium stream.

In media specific FEC (MS-FEC), redundant data packets, which are dependent on the specific medium, are used to recover the loss if an original data packet is lost. The first transmitted data is referred to as the primary encoding and redundant transmissions as secondary encodings. Media independent FEC (MI-FEC) does not need to know the medium type. In MI-FEC, original data together with parities are transmitted to the receiver. The parities are derived from the original data. To this purpose, we can use, among other techniques, a Reed-Solomon (RS) code. The main advantage of the RS code is that the error pattern is not of importance to recover the lost packets.

MDC divides data into several dependent descriptions such that the decoding quality using any subset is acceptable, and better quality is obtained by using more descriptions [4]. MDC usually considers the case that the probability of losing all the descriptions is small. It does not require additional delay, and requires additional bandwidth only if effective channel coding is required.

LC separates the transmitted information into one base layer and several enhancement layers, while MDC separates the transmitted information in a number of equally important descriptions. If the base layer information of LC is lost, then its enhancement layer is useless. Thus, LC requires the support of prioritized transmission of base layer from networks or applications. The results of [5, 6] showed that MDC is better than LC for the case of a strict delay constraint and no feedback.

The performance of MS-FEC can be improved by adopting the basic concept of MDC. Instead of transmitting primary and redundant packets, we can send equally important MDC packets. If a single packet is delivered to a receiver, MS-FEC and MDC have the same R-D performance. However, if more than one packet is delivered, the quality of MS-FEC cannot be better than that of a primary encoding, while that of MDC increases as more descriptions arrive. Thus, MDC always gives better performance than MS-FEC, and we will compare only MI-FEC with MDC in the following sections from a rate-distortion (R-D) view point.

In section 2, we will discuss delay constraints for voice over
IP (VoIP). In section 3, the R-D bounds of FEC and MDC based on a Gilbert channel model will be given. In section 4, comparisons between FEC and MDC will be made for both the Gilbert model and real IP connections. Section 5 presents concluding remarks.

2. Delay Constraints for VoIP

VoIP has very strict delay constraints. There are four different causes of delay: packetization, propagation, transport, and jitter-buffer delay [7]. Fig. 1 shows the entire encoding, transmission, and decoding process for real-time voice transmission over IP. Each processing and transmission stage has an associated delay. \(a, b, c, d, e, f, g, h, i\) and \(N\) represent frame size, lookahead, source encoding delay, channel encoding delay, propagation delay, jitter-buffer delay, channel decoding delay, decoding buffer delay, and the number of frames per packet. Here, we assume propagation delay includes transport delay.

In the transport and link layers of IP networks, a packet that is larger than maximum transfer unit (MTU) size can be fragmented. Thus, the packet size should be smaller than MTU size. On the other hand, to reduce header overhead problem, the packet should be as large as possible. It is common to assume MTU sizes of around 1500 bytes and 100 bytes for wired and wireless IP networks, respectively [8]. The header adds redundant information to the payload. For example, the payload of ITU-T G.729 speech coder per frame is 10 bytes. The header size for a single packet of IPv4 is 40 bytes without header compression. To reduce the header redundancy, the payload size should be close to MTU size. However, increasing the payload results in degradation of QoS because of the additional delay.

ITU-T G.114 states the following one-way delay limits for voice communications [9]:

- 0-100ms: toll quality;
- 100-150ms: high quality for most applications;
- 150-400ms: acceptable for international connections;
- > 400ms: very poor quality.

Let us consider voice communication systems where one-way delay should be less than 150ms. For the G.729 speech coder, \(a=10\)ms and \(b=5\)ms in Fig. 1. If we assume \(c + d + g + h + i\leq 5\)ms, propagation delay \(e=60\)ms and jitter \(f=60\)ms, the delay constraint is

\[aN + b + c + d + e + f + g + h + i \leq 150,\]

\[10N + 130 \leq 150,\]  \hspace{1cm} (1)

and only two source frames can be used per packet. In this case, one packet consists of a 40 byte header and a 20 byte payload. Channel coding may require additional delay and hinder the real-time delivery of voice stream. Thus, we allow only a single-packet delay for channel coding in this paper.

3. R-D Bounds of FEC and MDC based on Gilbert Channel Model

Before comparing packet-loss recovery techniques, it is necessary to characterize the packet-loss patterns introduced by the Internet channel. Since it is reported that single and duplicate losses dominate burst patterns over the Internet [10], we use a Gilbert model (i.e., a two-state Markov model) as depicted in Fig. 2. In the Gilbert model, one state represents a packet loss, which is called state ‘0’, and the other state represents the situation when a packet is successfully delivered to the destination which is referred to as state ‘1’. \(p\) is the probability of going from state ‘0’ to state ‘1’, and \(q\) is that of going from state ‘1’ to state ‘0’. For steady-state case, the initial probability distributions of states 0 and 1 are \(\Pi_0 = q/(p + q)\) and \(\Pi_1 = p/(p + q)\) respectively.

For real-time voice communication over IP, we consider the case that only a single-packet additional delay is allowed. Thus, we consider the case that two source packets, \(A\) and \(B\), are used for FEC and MDC. Since most losses are reported to be single or two consecutive losses, we add two redundant packets to the source packets for both methods. We assume that original data are from a memoryless Gaussian source.

3.1. R-D Bounds of FEC and MDC

We first consider the application of RS-FEC to the case that two memoryless Gaussian sources are encoded separately and transmitted over a single channel. The mean distortion bound of FEC at a given rate \(R\) is:

\[D_{FEC} = \nu 2^{-2R} + \omega.\]  \hspace{1cm} (2)

FEC can fully correct a given number of missing packets, but it cannot recover anything if the number of missing packets exceeds its recovery capacity. For the Gilbert model, the constants \(\nu\) and \(\omega\) of (2) are:

\[\nu = \frac{q}{2(p + q)} (2 + 3p - 3p^2 - pq^2 + 3p^2q),\]

\[\omega = \frac{p}{2(p + q)} (2 - 3q + 3pq + q^3 - 3pq^2).\]  \hspace{1cm} (3)

We consider a balanced MDC where the same rates are assigned to both encoders, and the same distortion bounds are achieved by both side decoders. We apply interleaving to MDC, which requires the single-packet delay that we have allowed for all FEC schemes. This interleaving improves robustness against burst packet-loss while maintaining the network load. The mean distortion bound of MDC is

\[D_{MDC} = \alpha D_0 + \beta D_1 + \gamma,\]  \hspace{1cm} (4)
where $D_0$ and $D_1$ are the central and side distortions, respectively, and

$$\alpha = \frac{q}{p+q} (1 - 2p + p^2 + pq),$$

$$\beta = \frac{2pq}{p+q} (2 - p - q),$$

$$\gamma = \frac{p}{p+q} (1 - 2q + q^2 + pq).$$

(5)

### 3.2. Informed Coding Case: Channel Optimized MDC

From the Ozarow’s bound [4], the achievable R-D bound of MDC for a one-dimensional memoryless Gaussian source with unit variance is illustrated in Fig. 3. If no detailed information about the packet-loss characteristics is available to the transmitter, then central-distortion optimized MDC (CD-MDC) is a good choice for relatively high-quality channels and side-distortion optimized MDC (SD-MDC) can be used for situations where better protection against packet loss is required. However, if the transmitter can be informed about channel conditions, it is advantageous to change the operating point of MDC between CD-MDC and SD-MDC based on the given information. MDC has inherent characteristics that facilitate adaptation to changing channel conditions.

It is possible to express the mean distortion bound of MDC, at a given rate $R$, as a function of the side distortion:

$$D_{MDC,R}(D_{1,R}) = \alpha D_{0,R}(D_{1,R}) + \beta D_{1,R} + \gamma,$$

where $\alpha \geq 0$, $\beta \geq 0$, and $\gamma \geq 0$ from (5). We note that $D_{MDC,R}(D_{1,R})$ is a convex and non-increasing function of $D_{1,R}$. This implies that, at a given fixed rate $R = \hat{R}$, the function $D_{MDC,R}(D_{1,R})$ is convex for $2^{-2R} \leq D_{1,R} \leq \frac{1}{2}(1 + 2^{-4\hat{R}})$.

Furthermore, the side distortion is never selected to be more than $\frac{1}{2}(1 + 2^{-4\hat{R}})$, since $D_{0,R}(D_{1,R})$ has a constant value if $D_{1,R} \geq \frac{1}{2}(1 + 2^{-4\hat{R}})$.

The best operating point of MDC can now be determined by finding the corresponding stationary point of equation (6). By setting the derivative to be zero, we find that the best operating point on $2^{-2R} \leq D_{1,R} \leq \frac{1}{2}(1 + 2^{-4\hat{R}})$ is

$$\frac{\partial D_{0,R}(D_{1,R})}{\partial D_{1,R}} = -\frac{\beta}{\alpha}.$$  

(7)

Thus, finding the best operating point corresponds to finding the correct slope of the curve in Fig. 3. The $\alpha$ and $\beta$ in (7) can be found from the Gilbert model parameters from (5). The described procedure implements an MDC optimization that depends on the channel conditions, which results in the minimum distortion of MDC. We call this new method the channel optimized MDC (optMDC).

### 4. Experimental Results

In this section, using both the Gilbert channel model and data obtained from real IP connections, we compare the R-D bounds of RS-FEC and MDC in terms of signal-to-noise ratio (SNR).

Liu and Marsh recorded transmission characteristics of the Internet during 2003 [11]. The nine global sites were interconnected to each other with a full-mesh topology, for a total of 72 interconnections. A pre-recorded conversation was coded with ITU-T G.711 and sent with 20ms packetization intervals between the cooperating sites on an hourly basis. From the measurement repository, we selected the files where packets were transmitted without silence suppression, and estimated the loss percentage and Gilbert model parameters ($p$ and $1-q$). In real-time voice communication systems, if a packet is delivered after its playout time, it must be discarded. Thus, we estimated the loss rate and the Gilbert model parameters for the case that such late packets are considered lost (we used a 60ms jitter buffer).

Based on the Gilbert parameter estimates, we compare the SNR performances of FEC, SD-MDC and optMDC for single-channel transmission case. In Fig. 4 (a), we show the SNR difference between SD-MDC and FEC at 1 bit/sample. In these figures, ‘o’ represents positive SNR-difference corresponding to the $p$ and $1-q$ value of a certain connection. A positive SNR-difference means SD-MDC is more robust against packet loss than FEC. Fig. 4 (a) shows that SD-MDC always has better SNR performance than FEC for 1 bit/sample. At this low rate, the better performance of SD-MDC results from the central decoding when all descriptions arrive. However, for high-rate (9 bits/sample), the SNR’s of both methods approach a similar value. The SNR difference between SD-MDC and optMDC is shown in Fig. 4 (b). The optMDC has a 2.57 dB higher average SNR than SD-MDC for the low-rate case (1 bit/sample).

We also evaluated the performance of FEC and MDC directly for the measured data, without assuming the Gilbert model. We directly applied FEC and MDC to the loss patterns of the measurement repository and calculated the SNR. In Table 1, the direct calculation is compared with the Gilbert model-based estimation. The simulation results with and without Gilbert model are similar. Most of the mismatches are found at the connections to and from the Turkish site, since it contains long consecutive packet losses that are difficult to model with the two-state Gilbert model (the Turkish research network is connected via a satellite link to Belgium).
Table 1: SNR difference (dB) between Gilbert-model based estimation and direct measurements (rate=1 bit/sample).

<table>
<thead>
<tr>
<th>SNR</th>
<th>Gilbert</th>
<th>Direct</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEC</td>
<td>5.81</td>
<td>5.77</td>
</tr>
<tr>
<td>MDC</td>
<td>7.96</td>
<td>7.88</td>
</tr>
<tr>
<td>optMDC</td>
<td>10.53</td>
<td>10.47</td>
</tr>
</tbody>
</table>

While the difference in performance of FEC, SD-MDC and optMDC is significant at low rates, this difference becomes asymptotically negligible with increasing rate. The decrease in the difference can be understood based on the three transmission scenarios for source information: all transmitted information is received, part of the transmitted information is received, or no information is received. At high bit rates, the performance of the coding systems is dominated by the performance for the case that no information is received, which is similar for the three methods.

5. Conclusions

To facilitate real-time voice communication through the Internet, both forward error correction (FEC) and multiple description coding (MDC) are promising packet-loss recovery techniques. Using memoryless Gaussian sources, we compared the performance of Reed-Solomon based FEC (RS-FEC), side-distortion optimized MDC (SD-MDC) and channel optimized MDC (optMDC).

SD-MDC is more robust than RS-FEC for VoIP. If the channel condition is known from a feedback channel, optMDC can be used, which has an optimized ratio between the central and side distortions of MDC. As expected, optMDC always performs at least as well as SD-MDC. Particularly at low source-coding rates this difference can be very large. Thus we conclude that, in most circumstances, SD-MDC is the best packet-loss recovery technique for real-time voice communication over the Internet. If we know the long-term average or short-term variations of channel condition, optMDC is the best choice.

6. Acknowledgment

Work performed with partial support from The Swedish Agency for Innovation Systems.

7. References


