# DISTORTION CHAINS FOR PREDICTING THE VIDEO DISTORTION FOR GENERAL PACKET LOSS PATTERNS

*Jacob Chakareski*[*†], *John Apostolopoulos*[†], *Wai-tian Tan*[†], *Susie Wee*[†] *and Bernd Girod*[*]

[†]Streaming Media Systems Group
Hewlett-Packard Labs, Palo Alto, CA 94304

[*]Information Systems Laboratory
Stanford University, Stanford, CA 94305

## ABSTRACT

When designing a system for video communication over a lossy packet network, it is highly beneficial to have a mechanism for accurately predicting the mean-squared error (MSE) distortion that results from different packet loss patterns. This paper proposes Distortion Chains model for accurately predicting the end-to-end distortion for different general packet loss patterns. The performance is examined using JVT/H.264 encoded video sequences and previous frame error concealment. It is shown that for all tested sequences the proposed model predicts the total distortion due to a packet loss pattern within a 10 % error bound 80 % of the time, as compared to the conventional additive approach which achieves the same accuracy less then 40 % of the time.

## 1. INTRODUCTION

Video communication over lossy packet switched networks is often impaired by packet losses due to congestion, erasures and/or late delivery. This has placed new demands on source coding algorithms and network transport schemes in order to simultaneously account for the channel introduced losses and maximize the reconstructed video quality at the receiver. The challenge of error-resilient video communication has received significant attention in recent years, and a variety of techniques have been proposed, including intra/inter-mode switching [1, 2], dynamic control of prediction dependencies [3], forward error correction [4], multiple description coding [5], and most recently Rate-Distortion Optimized (RaDiO) packet scheduling [6–9]. All these approaches are designed and operated based on models for the effect of losses on the reconstructed video quality. Therefore, their performance crucially depends on the accuracy of the employed distortion models.

Prior work on modelling the effect of losses generally models the distortion as being proportional to the number of losses that occur [2, 10]. For example, in [2] first a model for the total distortion associated with a single (isolated) loss is proposed that accounts for the effects of error propagation, intra refresh, and spatial filtering. Then, using this model the effect of multiple losses is represented as a superposition of multiple independent losses. With this linear or additive model, the expected distortion is proportional to the average packet loss rate. This additive model is accurate as long as the burst loss does not lead to the loss of more than a single frame, where the number of lost frames depends on the number of packets per frame relative to the burst length in packets. For example, this model is accurate for low-bit-rate video, where each coded video frame fits within a single packet, when single losses occur that are spaced sufficiently far apart with respect to the intra-refresh period, e.g. when the loss rate is low and the losses are not bursty. However, in many important applications, for example low-bit-rate video communication (where each coded frame may fit within a single packet) over the Internet or over a wireless link, the losses may be bursty and may result in the loss of multiple frames. In [5], it was recognized that the length of a burst loss has an important effect on the resulting distortion, where longer burst lengths generally led to larger distortions. This was extended in [11] where a simple model was proposed that distinguishes loss events based on the length of the burst loss and explicitly accounts for the different distortions that result for different burst lengths. In [12] a model is proposed that captures the correlation between the error frames associated with single (isolated) packet losses in order to describe more accurately the distortion resulting from a burst loss pattern. The effect of burst losses is particularly pronounced for low bit rate video, and less pronounced for high bit rate video [13].

In this paper, we propose a model, which we refer to as the Distortion Chains model, for predicting the mean-square error (MSE) distortion at the receiver in the event of packet loss. This model provides a simple, causal approach for predicting the distortion in the reconstructed video for general packet loss patterns. The experimental results indicate that even a Distortion Chain model of order 1 can predict the total increase in distortion due to higher order packet losses quite accurately, for packet loss rates of practical interest.

The rest of the paper is structured as follows. Section 2 introduces our notation and reviews the distortion produced by packet loss. Section 3 describes the proposed Distortion Chains approach for modelling the effect of packet loss and how this model is employed to predict the distortion that results from different packet loss patterns. Section 4 evaluates the prediction accuracy of the proposed model for JVT/H.264 coded video. Finally, concluding remarks are provided in Section 5.

## 2. DISTORTION PRODUCED BY PACKET LOSS

We first introduce some necessary notation and background. We follow closely the notation of [12]. We analyze the case where a sequence starts with an I-frame, followed by P-frames that have a certain number of macroblocks periodically Intra updated for increased error-resilience. For simplicity, we assume that each P-frame is coded into a single packet, so that the loss of a packet corresponds to the loss of an entire frame. This corresponds to

This work was performed when Jacob Chakareski was a summer researcher at HP Labs, Palo Alto.

the practically important case of low bit rate video communication over lossy packet networks, e.g. QCIF video at less than 150 kb/s. However, the results in this paper can also be extended to the case when each frame is coded into multiple packets.

The original video signal is a discrete space-time signal denoted by $s[x, y, k]$, where $k \in Z$ is the frame index. To simplify notation, the 2-D array of $M = M_1 \times M_2$ pixels in each frame $k$ are sorted in the 1-D vector $f[k]$ (of length $M$) in line-scan order. We use the 1-D vector $f[k]$ to represent an original video frame, $\hat{f}[k]$ to denote the loss-free reconstruction of the frame, and $g[k]$ to denote the reconstruction at the decoder after loss concealment. The error frame at frame $k$ introduced by one or more packet losses that occurred earlier is defined as

$$e[k] = g[k] - \hat{f}[k]$$

which is also a 1-D vector. We assume previous frame loss concealment. Therefore, if frame $k$ is the first occurrence of packet loss then $g[k] = \hat{f}[k-1]$. Since our primary concern is the effect of channel loss, quantization error is not included in our study. Finally, the Mean Square Error (MSE) associated with error frame $e[k]$ is given by

$$(e^T[k] \cdot e[k])/M = \sigma^2[k].$$

The above MSE quantifies the error power introduced in a single frame due to previous packet losses. Now, let $L$ be the length of a video sequence in frames and let $\boldsymbol{k} = (k_1, k_2, \ldots, k_N)$ denote a loss pattern of length $N$, i.e., $N$ frames are lost during transmission where $k_i < k_j$, for $i < j$. Then, the total distortion, denoted by $D$, due to the loss pattern is the sum of the MSEs over all the frames affected by the loss pattern $\boldsymbol{k}$, i.e.,

$$D(\boldsymbol{k}) = \sum_{l=1}^{L} \sigma^2[l] = \sum_{l=k_1}^{L} \sigma^2[l]. \quad (1)$$

## 3. DISTORTION CHAIN MODEL FOR PREDICTING DISTORTION

We define $D(k_{N+1}|\boldsymbol{k})$ to be the additional increase in distortion due to losing frame $k_{N+1} > k_N$ given that frames $k_1, \ldots, k_N$ are already lost, i.e.,

$$D(k_{N+1}|\boldsymbol{k}) = D(k_1, \ldots, k_{N+1}) - D(k_1, \ldots, k_N). \quad (2)$$

A Distortion Chain model of order $N$ is comprised then of the distortion quantities $D(\boldsymbol{k})$ for every loss pattern $\boldsymbol{k}$ of length $N$ satisfying $k_i < k_j$, for $i < j$, and of $D(k_{N+1}|\boldsymbol{k})$ for every loss pattern $(\boldsymbol{k}, k_{N+1})$ of length $N+1$ satisfying $k_N < k_{N+1}$. These quantities can be generated at the encoder by simulating the corresponding loss events, decoding the video sequence, and then computing the resulting distortions. We next examine how $D(\boldsymbol{k})$ and $D(k_{N+1}|\boldsymbol{k})$ can be used to predict the total distortion for loss patterns of lengths greater than $N$.

Let $DC^N$ denote our distortion chain of order $N$ and let $\boldsymbol{k} = (k_1, \ldots, k_P)$ be an arbitrary loss pattern of length $P$ such that $N < P \leq L$, where again $L$ is the length of the video sequence in frames. Then, let $\widetilde{D}(\boldsymbol{k})$ denote the estimate of the total distortion due to the loss pattern $\boldsymbol{k}$ obtained from $DC^N$ as follows,

$$\widetilde{D}(\boldsymbol{k}) = D(k_1, \ldots, k_N) + \sum_{i=N}^{P-1} D(k_{i+1}|(k_{i-N+1}, \ldots, k_i) \quad (3)$$

This general formulation suggests that we need the $N$ previous losses (loss pattern of length $N$) in order to predict the distortion for a length $P$ loss pattern that begins with the pattern of length $N$. While this may be impractical for large $N$, in our work we have found that even small values of $N$, such as 1, still provide good prediction results. In addition, when losses are spaced far apart (larger than the intra refresh interval) the losses become decoupled (their effects are independent) and the prior losses have very limited effect on the subsequent losses. In the next section, we validate the accuracy of Equation (3) against the actual measured distortion $D(\boldsymbol{k})$.

## 4. EXPERIMENTAL RESULTS

This section examines the performance of the proposed Distortion Chains model for $N = 1$ using simulation experiments. We simulate different loss patterns on standard test video sequences, and compare the measured distortion with that predicted by $DC^1$. In addition, we also examine the performance of an Additive model which treats the individual losses as independent; this is equivalent to a zeroth-order Distortion Chain model ($DC^0$). The video sequences are coded using JM 2.1 of the JVT/H.264 video compression standard [14]. Two standard test video sequences in QCIF format are used, Foreman and Carphone. Each has at least 300 frames at 30 fps, and is coded with a constant quantization level at an average PSNR of about 36 dB. The first frame of each sequence is intra-coded, followed by P-frames. Every 4 frames a slice is intra updated to improve error-resilience by reducing error propagation (as recommended in JM 2.1), corresponding to an intra-frame update period of $4 \times 9 = 36$ frames.

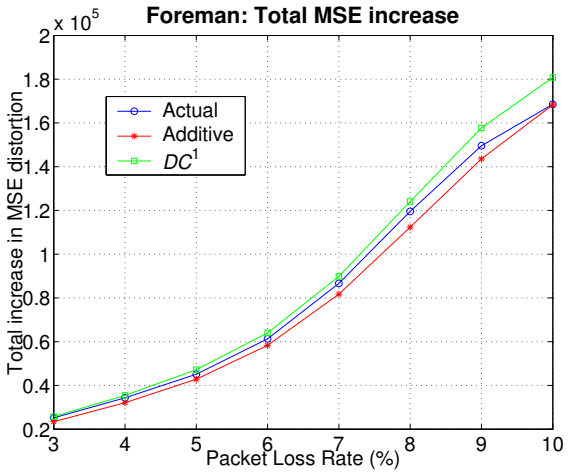


**Fig. 1**. Total increase in MSE distortion for *Foreman*.

In the first set of experiments, the performance is examined across a range of packet loss rates (PLR) of 3-10 %. For each loss rate we generate a set of 50,000 random loss patterns. For each loss pattern we decode the video and record the total resulting distortion. This actual measured distortion is then compared against the corresponding predicted distortions obtained from the Additive model and from the proposed $DC^1$ model. Then, for each packet loss rate the mean distortion value is computed over all of the 50,000 patterns for each of the 3 cases: Actual, Additive and $DC^1$. These quantities are shown in Figure 1. There are

a few conclusions that follow from the figure. On average, $DC^1$ overestimates the actual distortion, while the Additive approach underestimates it. A more revealing graph related to the same set of experiments is shown in Figure 2. Here, we examine the prediction gain in dB of the $DC^1$ model over the Additive model, which is defined as the absolute error of the prediction distortion from the actual distortion, for each packet loss rate averaged over all loss patterns. It can be seen that $DC^1$ provides a significant gain for low loss rates, which decreases as we move towards higher loss rates. This is due to the fact that at very high PLR both models fail to accurately predict the distortion for a given packet loss pattern. For example, at PLR of 3 % the gain is 3.2 dB, while at PLR of 8% the gain is around 1.2 dB.

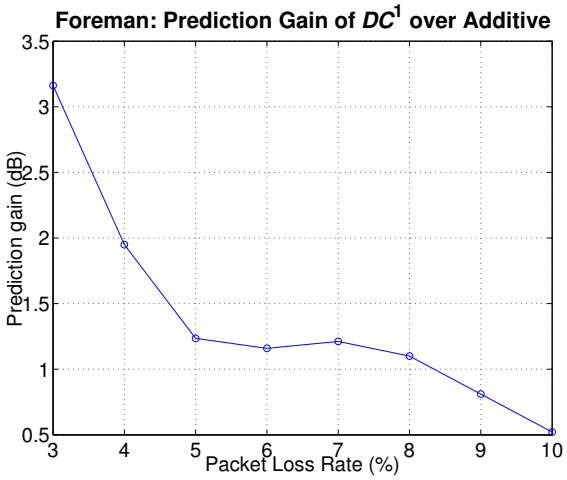


**Fig. 2**. Prediction gain (dB) of $DC^1$ over Additive.

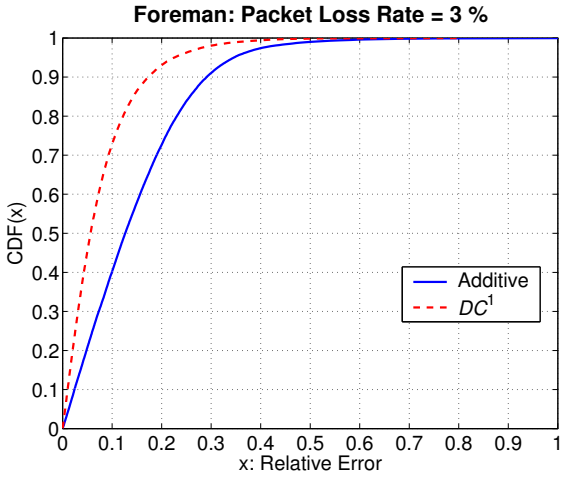


**Fig. 3**. CDF of the Relative Prediction Error for PLR = 3 %.

In the next two figures, we examine the distribution of the relative error for the two models for PLR = 3 % and 8 %. In Figures 3 and 4 we show the Cumulative Density Functions (CDFs) of the relative error for the two approaches, which is defined as the ratio of the absolute prediction error and the actual distortion. It can be seen from Figure 3 that $DC^1$ provides an estimate that is within a 10 % error bound 75 % of the time, while the Additive model does that only 40 % of the time. Similarly, $DC^1$ provides an estimate that is within a 20 % error bound 93 % of the time, while the Additive model does that only 73 % of the time.

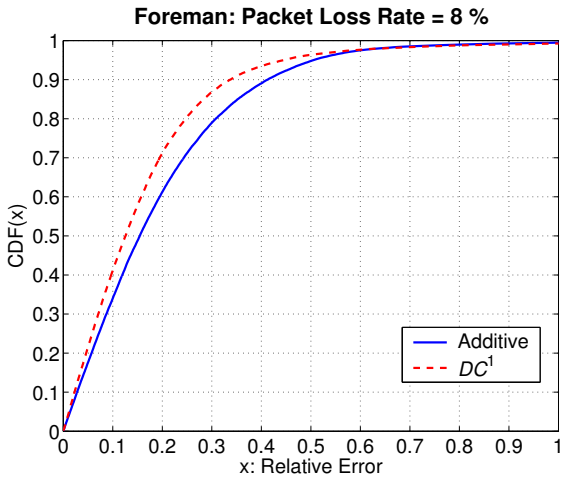


**Fig. 4**. CDF of the Relative Prediction Error for PLR = 8 %.

Figure 4 shows similar improvement due to the Distortion Chains approach, even though now notably due to the higher packet loss rate both of the approaches perform proportionally worse.

A different situation is examined in Figures 5 and 6, where we examine the statistical performance of the proposed model by considering its performance over all possible loss patterns, given a fixed loss rate for the entire sequence. In this case we examine the performance for all relevant possible packet loss patterns (where coupling can occur from the lost packets) where 3 packets are lost in each window of 120 packets, and the windows are sliding across the video sequence. Unlike in the prior figures where we only examined 50,000 packet loss patterns at each packet loss rate, in this case we examine all relevant possible packet loss patterns where 3 packets are lost in each sliding window of length 120, corresponding to about 480,000 packet loss patterns which required 6 weeks of processing on a dual-processor P4 2GHz.

The CDFs in Figures 5 and 6 clearly illustrate that $DC^1$ can provide improved prediction accuracy as compared to the Additive model. Specifically, $DC^1$ is accurate to within a 10 % error bound 80 % of the time for both Foreman and Carphone, while the Additive model achieves that accuracy only 37 % and 28 % of the time for Foreman and Carphone, respectively. Similarly, $DC^1$ provides an estimate that is within a 20 % error bound 95 % of the time for both sequences, while the Additive model achieves this only 64 % and 54 % of the time. This statistical study gives us an indication of the robustness of the proposed model across the wide range of possible loss patterns that correspond to a given average loss rate.

## 5. CONCLUSIONS

This paper proposed Distortion Chains for predicting the distortion for video communication afflicted by general packet loss patterns. We have shown through experiments that the proposed model, even for a minimal amount of memory ($N = 1$), provides a significant

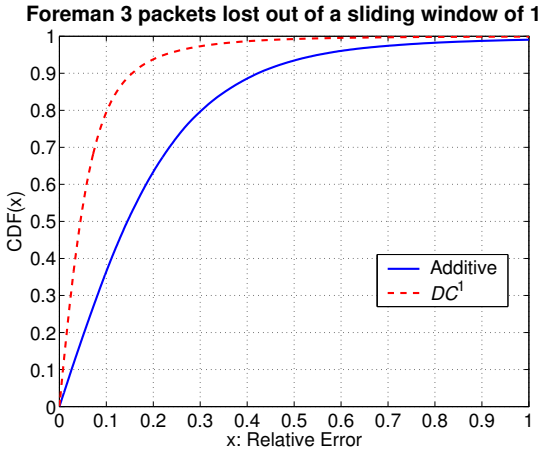


**Fig. 5**. CDF of the Relative Prediction Error for *Foreman*.

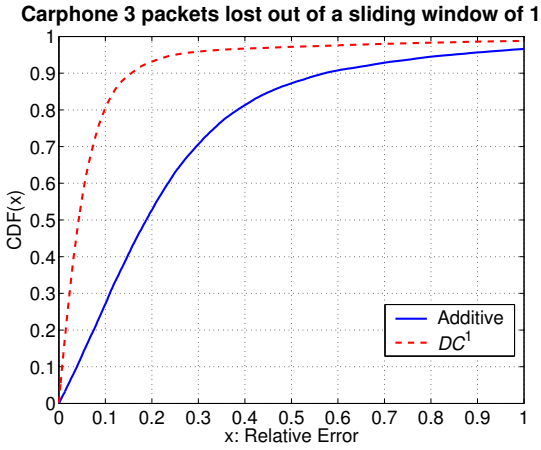


**Fig. 6**. CDF of the Relative Prediction Error for *Carphone*.

improvement in performance over a model that treats the individual packet losses as independent. Moreover, it is observed that for all video sequences under consideration, and for loss of 3 packets in a sliding window of 120 packets, the Distortion Chains model predicts the total distortion due to a packet loss pattern within a 10 % error bound in 80 % of the cases. In contrast, the Additive model achieves this accuracy less then 40 % of the time. This is encouraging, since current video coding algorithms and network transport schemes can leverage the improved accuracy of the proposed model to improve their own performance. In an additional related work [15], we have designed a packet scheduling scheme for video streaming over the Internet that employs the distortion model presented here.

## 6. REFERENCES


[1] R. Zhang, S.L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Selected Areas in Communications*, vol. 18, no. 6, pp. 966–976, June 2000.

[2] K. Stuhlmüller, N. Färber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Selected Areas in Communications*, vol. 18, no. 6, pp. 1012–1032, June 2000.

[3] Y.J. Liang and B. Girod, "Low-latency streaming of pre-encoded video using channel-adaptive bitstream assembly," in *Proc. Int'l Conf. Multimedia and Exhibition*, Lausanne, Switzerland, Aug. 2002, IEEE, vol. 1, pp. 873–876.

[4] W.-T. Tan and A. Zakhor, "Video multicast using layered FEC and scalable compression," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 373–387, Mar. 2001.

[5] J. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *Proc. Visual Communications and Image Processing*, San Jose, CA, Jan. 2001, SPIE, vol. 4310, pp. 329–409.

[6] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Trans. Multimedia*, 2001, submitted.

[7] J. Chakareski, P.A. Chou, and B. Aazhang, "Computing rate-distortion optimized policies for streaming media to wireless clients," in *Proc. Data Compression Conference*, Snowbird, UT, Apr. 2002, IEEE Computer Society, pp. 53–62.

[8] J. Chakareski, P.A. Chou, and B. Girod, "Rate-distortion optimized streaming from the edge of the network," in *Proc. Workshop on Multimedia Signal Processing*, St. Thomas, US Virgin Islands, Dec. 2002, IEEE, pp. 49–52.

[9] J. Chakareski and B. Girod, "Rate-distortion optimized packet scheduling and routing for media streaming with path diversity," in *Proc. Data Compression Conference*, Snowbird, UT, Mar. 2003, IEEE Computer Society, pp. 203–212.

[10] I.-M. Kim and H.-M. Kim, "A new resource allocation scheme based on a PSNR criterion for wireless video transmission to stationary receivers over gaussian channels," *IEEE Trans. Wireless Communications*, vol. 1, no. 3, pp. 393–401, July 2002.

[11] J. Apostolopoulos, W.-T. Tan, S. Wee, and G.W. Wornell, "Modeling path diversity for multiple description video communication," in *Proc. Int'l Conf. Acoustics, Speech, and Signal Processing*, Orlando, FL, May 2002, IEEE, vol. 3, pp. 2161–2164.

[12] Y.J. Liang, J. Apostolopoulos, and B. Girod, "Analysis of packet loss for compressed video: Does burst-length matter," in *Proc. Int'l Conf. Acoustics, Speech, and Signal Processing*, Hong Kong, China, Apr. 2003, IEEE, vol. 5, pp. 684–687.

[13] A. Reibman and V. Vaishampayan, "Quality monitoring for compressed video subjected to packet loss," in *Proc. Int'l Conf. Multimedia and Exhibition*, Baltimore, MD, USA, July 2003, IEEE, vol. 1, pp. 17–20.

[14] Telecom. Standardization Sector of ITU, "Video coding for low bitrate communication," *Draft ITU-T Recommendation H.264*, Mar. 2003.

[15] J. Chakareski, J. Apostolopoulos, W.-T. Tan, S. Wee, and B. Girod, "R-D hint tracks for low-complexity R-D optimized video streaming," in *Proc. Int'l Conf. Multimedia and Exhibition*, Taipei, Taiwan, June 2004, IEEE, submitted.