

# Interleaved Source Coding (ISC) for Predictive Video Coded Frames over the Internet

Jin Young Lee<sup>1,2</sup>

<sup>1</sup>Broadband Convergence Networking Division  
ETRI  
Daejeon, 305-350 Korea  
jinlee@etri.re.kr

Hayder Radha<sup>2</sup>

<sup>2</sup>Dept. of Electrical and Computer Engineering  
Michigan State University  
East Lansing, MI 48823 U.S.A  
radha@egr.msu.edu

**Abstract**— Unreliable network’s packet losses severely impact the playback quality of many predictive coded sources such as compressed video. Prior researches (e.g., [1]-[4] [6]-[9] [11]-[13]) have shown various packet loss resilient coding methods to overcome such deficiency. In this paper, we propose a new packet-loss resilient coding method, *interleaved source coding* (ISC), based on an interleaving of predictive video coded frames transmitted over the internet via a *single* erasure channel that is resilient to packet-losses. To select an interleaving pattern for a given erasure channel model, we employ a Markov Decision Process (MDP) and a corresponding dynamic programming algorithm. ISC improves the overall playback quality of predictive coded video frames over a lossy channel without complex modifications to standard predictive video coders. ISC eliminates the need for content distribution, path diversity routing, and related synchronization issues and hence, it provides a variable alternative to the path-diversity approaches. Simulations of a various video sequences showed that ISC significantly improves the playback quality of the predictive video over practical traces of Markov erasure channel model when compared with traditional non-interleaving predictive coding method.

**Keyword**— Dynamic Programming, Interleaving, Markov Decision Process, Packet Losses, Video Coding

## I. INTRODUCTION

With the expansion of the underlying infrastructure of the IP network, streaming video is emerging as one of the most popular on-line realtime Internet applications. However, despite of the growth of internet infrastructure, due to the unreliable nature of IP network, e.g., packet delay and packet losses, streaming services often lack Quality-of-Service (QoS), which results in degraded playback quality. To improve the playback quality of realtime streaming video without QoS guarantees, special coding techniques resilient to packet losses are required. Multi state video compression [1], Multiple-Description-Coding (MDC) with path diversity [2]-[4], multi-hypothesis motion estimation and compensation [6][9], packet interleaving [8], and Scalable coding [11][12] are examples of techniques that are resilient to packet losses.

In this paper, we propose a new video-coding approach that is resilient to packet losses. Our proposed method is based on *interleaved source coding* (ISC) for predictive video sequences. ISC codes a single video sequence into two sub-sequences and transmits them over a *single* erasure channel, hence reduces the frequency and impact of the cascaded effect of packet losses along with related propagation of errors that are resulted from the predictive nature of coded video. In particular, the design

object of *interleaving* is that the impact of losses caused by a given erasure channel model (with memory) is limited to a minimum number of video frames. Moreover, when the lost frames are replaced with the last successfully decoded frame (frozen frame), ISC produces smoother video than the traditional predictive-coding/transmission methods.

The difference between ISC and the previous interleaving related methods (e.g., ones proposed in [2]-[4][8]) are as following. ISC transmits encoded sequences over a *single* channel where MDC based methods [2]-[4] uses multiple channels for transmission, hence, ISC eliminates content distribution, channel selection, and synchronization issues known to present in MDC. In addition, ISC uses a statistical approach to identify an *optimum frame based* interleaving set whereas most frame-based MDC video approaches employs a rather heuristic method that separates frames into even and odd numbered pictures without taking into consideration the channel characteristics and other important video parameters. To find an ISC set, we employ a Markov Decision Process (MDP) and a Dynamic Programming algorithm with a realistic packet loss model.

The outline of this paper is organized as follows: In Section II, we describe the proposed ISC coding method. A general description on interleaving is given in Section II-A and a mathematical approach finding an ISC set using a Markov reward process, Markov Decision Process, and a Dynamic Programming algorithm are described in Section II-B. In Section III, ISC is evaluated using MPEG-4 video simulated over an Internet Markov-based lossy channel model.

## II. METHODOLOGY

### A. General Interleaving

The proposed *interleaved source coding* (ISC) is a pre- and post-process of predictive source coders<sup>1</sup> (Fig. 1) that improves

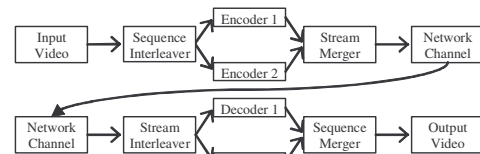


Figure 1. Interleaving of Predictive Video Coding.

<sup>1</sup> The *Interleavers and Mergers* can be integrated into the predictive source coders and use a single encoder and decoder; however, to simplify ISC adaptation to the various predictive video coders, we employ ISC as a pre- and post-process of the coders and leave the coders untouched.

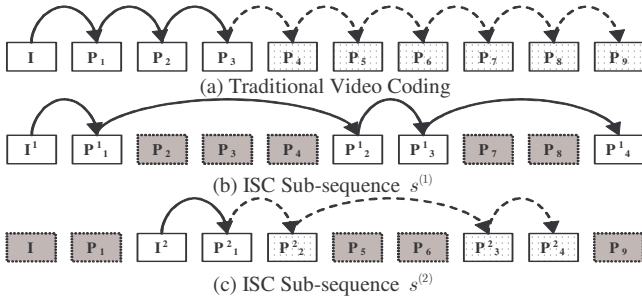


Figure 2. Traditional vs. ISC Video Coding with a packet loss in the frame location of  $P_4$  in (a). The arrowed lines represent the coded frames temporal dependencies in the predictive video coding. The dotted frames are the decoder failed frames due to the loss. The shaded frames are belonged to the other sub-sequence in (b) and (c).

the quality of predictive video over lossy packet networks by limiting the impact of losses within a given Group Of Video object planes (GOVs). ISC separates a single sequence into sub-sequences<sup>2</sup> with the *Sequence Interleaver* and encoded using separate encoder. Then the *Stream Merger* merges encoded frames into a single stream in an original sequence's frame order for transmission. Prior to the transmission, the interleaving information must be transmitted to the recipient for the decoder end's *Stream Interleaver* and *Sequence Merger*.

When separating a single sequence into two sub-sequences,  $s^{(j)}$ ,  $j = \{1, 2\}$ , we adopt the following ISC interleaving constraints where  $2N$  is the number of frames in a GOV in the original non-interleaved sequence  $s$ , and hence, the same interleaving is applied to all GOVs in the sequence or a scene.

$$s = \{0 \ 1 \ \dots \ 2N - 1\} = s^{(1)} \cup s^{(2)} \quad (1)$$

$$s^{(1)} \cap s^{(2)} = \emptyset, \quad \text{size}(s^{(j)}) = N, \quad j = \{1, 2\}$$

$$\sum \{s^{(j)}(2) - s^{(j)}(1), \dots, s^{(j)}(N) - s^{(j)}(N-1)\} > N - 1$$

For a non-interleaved sequence with a GOV size of 10, let  $\mathbb{S} = \{s^{(1)}, s^{(2)}\}$  be an interleaving sub-sequence set (Fig. 2), where the numbers in  $s^{(j)}$  represent the frame locations in the non-interleaved sequence and the coded stream's frame transmission order. Once separated, the sub-sequences are encoded separately as shown in Fig. 2-(b) and (c), and they are transmitted in the same frame transmission order as the non-interleaved traditional video coder (Fig. 2-(a)). It is important that the interleaving information set,  $\mathbb{S}$ , must be transmitted prior to the stream transmission. ISC is expected to decode more frames than traditional coding methods if the decoder encounters problems due to the packet loss or errors within. For example, if a part of the 5<sup>th</sup> frame ( $P_4$ , in Fig. 2-(a)) is lost, non-interleaved coding would not be able to decode 6 frames from  $P_4$  to  $P_9$  where ISC would successfully decode sub-sequence  $s^{(1)}$  and would not be able to decode only three frames,  $P_{22}$ ,  $P_{32}$ , and  $P_{42}$ , from the sub-sequence  $s^{(2)}$ . Hence, ISC improves overall playback quality by limiting errors (due to packet losses) to  $s^{(2)}$ . For a given GOV size  $2N$ , the size of the set of all possible interleaving sets,  $\mathbb{K}$ , is expressed as following:

$$\text{size}(\mathbb{K}) = \frac{1}{2} \binom{2N}{N} - N = \frac{(2N)!}{2 \times (N!)^2} - N \quad (2)$$

<sup>2</sup> ISC framework could support more than two sub-sequences. Here, we only focus on the simple case of two sub-sequences.

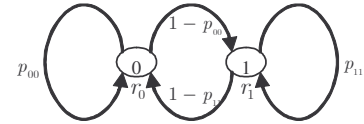


Figure 3. The *Gilbert Model* (Two state Markov model) with rewards  $r_i$ . Since the size of  $\mathbb{K}$  could be quite large for any reasonable GOV size, identifying an interleaving set that produces the best quality decoded video transmitted over a lossy network channel could be very computationally expensive task. Hence, an efficient decision-based search algorithm is required to choose an ISC set from  $\mathbb{K}$ .

## B. Decision Based Interleaving

### 1) Markov Reward Process

Efforts on the analysis and modeling of packet losses over the Internet (e.g., [14][15]) and wireless networks (e.g., [7]) have shown that these losses exhibit Markovian properties. A *Markov Reward Process* (MRP) (e.g., [5][10]) is used to estimate the system's performance by employing the system's transition probabilities and some model for the *rewards* associated with each system state. For an erasure-channel Markov model, MRP can be used to estimate the channel's performance after  $n$  packet transmissions based on the channel's packet transmission probabilities. Hence, this could guide the design of our interleaving coding system (as explained further below). Let  $\{0, 1\}$  be the corresponding state space to good (0) and bad (1) packet transmissions<sup>3</sup> for an erasure-channel model. For each state, instant reward  $r_i$  is assigned and it is awarded to the process whenever it reaches state  $i$  (Fig. 3). For a predictive coded sequence transmission over an erasure-channel with transition matrix  $p$ , the aggregated reward  $v(n-1)$  ([5][10]) is defined as a function of the number of transmitted packets. Hence, after  $n$  packet transmission, the aggregated reward  $v_i(n-1)$  represents the performance of predictive sequence transmission over a lossy channel with a channel's state transition matrix  $p$  with the initial packet transmission at state  $i$ .

$$v(0) = r = \{r_0, r_1\}^T$$

$$v(1) = r + p \cdot v(0), \quad p = \begin{bmatrix} p_{00} & 1 - p_{00} \\ 1 - p_{11} & p_{11} \end{bmatrix} \quad (3)$$

$$v(n-1) = \begin{bmatrix} v_0(n-1) & v_1(n-1) \end{bmatrix}^T = r + p \cdot v(n-2)$$

$$= r + p \cdot r + p^2 \cdot r + \dots + p^{n-2} \cdot r + p^{n-1} \cdot r = \left( I + \sum_{m=1}^{n-1} p^m \right) \cdot r$$

### 2) Markov Decision Process

Associating Markov reward process with a series of actions and decision criteria results a Markov Decision Process (MDP) [5][10]. To find an ISC set that is most suitable for a given decision criteria, maximizing the number of frames (or associated packets), that can be decoded *correctly*, we employ MDP as a guide toward an "optimal" interleaving for a given erasure channel model. The set that achieves our objective is the one with highest MRP aggregated reward. Furthermore, in

<sup>3</sup> Higher order Markov models may be used; however, to reduce computational complexity, we use two state Markov model, *a.k.a. Gilbert Model*, which is proven to replicate an acceptable erasure-channel model as the higher order Markov models ([14][15]).

TABLE I. PROPERTIES OF MDP FOR MULTIMEDIA STREAM INTERLEAVING

Policies {Action, Current State}	Instant Reward		Discount Factor		Transition Probabilities	
	$r_a$	$\gamma_a$	$\gamma_a$	$\gamma_a$	0	1
{C,0}	$r_C$	1	$\gamma_S$	1	$p_C$	$1-p_m$
{C,1}		0		0	0	1
{S,0}	$r_S$	0	$\gamma_S$	1	$p_S$	$1-p_m$
{S,1}		0		1	$1-p_m$	$p_m$

MDP, each *policy*, a mapping from states to actions, is associated with a set of *discount factors*,  $\gamma_a$  [5][10] which decide the amount of aggregated reward to be propagated to the next state. Incorporating (3) with  $\gamma_a$  and  $k$ , *interleaving set indicator*  $k \in \mathbb{K}$ , gives aggregated MDP equation:

$$\begin{aligned}
 v^{(k)}(n-1) &= r_{a^{(k)}(n-1)} + \gamma_{a^{(k)}(n-1)} \times p \bullet v^{(k)}(n-2) \\
 &= r_{a^{(k)}(n-1)} + \gamma_{a^{(k)}(n-1)} \times p \bullet r_{a^{(k)}(n-2)} + \\
 &\quad \dots + \left( \prod_{l=2}^{n-1} \gamma_{a^{(k)}(l)} \right) \times p^{n-2} \bullet r_{a^{(k)}(1)} + \left( \prod_{l=1}^{n-1} \gamma_{a^{(k)}(l)} \right) \times p^{n-1} \bullet r_{a^{(k)}(0)}
 \end{aligned} \quad (4)$$

In ISC, each frame in a GOV is considered as a state iteration in the *Gilbert model*. For each state iteration, based on the policies described in TABLE I, one of the two actions, Coding ( $C$ ), or Skip ( $S$ ), is taken. Here,  $a^{(k)}(n)$  denotes an action taken for the  $n^{\text{th}}$  frame in a GOV.

$$s^{(k,j)}(n) = s^{(k,j)}(n) - s^{(k,j)}(0), \quad \text{for } 0 \leq n \leq N-1 \quad (5)$$

Let the set of ISC sub-sequences in Fig. 2 be the interleaving set  $k$  and  $\mathbb{S}$  is rewritten as  $\mathbb{S}^{(k)} = \{s^{(k,1)}, s^{(k,2)}\}$  with respect to  $k$ .

Hence, the frame numbers in  $s^{(k,j)}$  are rewritten so that the reward computation for each sub-sequence starts from the time instance 0.

$$\begin{aligned}
 m &= \bigcup_{n=0}^{N-1} \{s^{(k,j)}(n) + 1, \dots, s^{(k,j)}(n+1) - 1\} \\
 \forall n \mid & s^{(k,j)}(n+1) - s^{(k,j)}(n) > 1
 \end{aligned} \quad (6)$$

For each sub-sequence, action  $C$  is performed, or in other words, frames are coded at frame locations specified in  $s^{(k,j)}$ . If the difference between two adjacent frame numbers in  $s^{(k,j)}$  exceeds 1, an indication of the presence of skipped frames, action  $S$  is performed for the frames in location  $m$ .

In our proposed MDP model, based on the policies, the instant rewards, discount factors, and transition probabilities are assigned as follows (TABLE I): For all policies with action  $S$ , since the skipped frames do not have any coding dependencies with other frames and do not affect the process regardless of the state, no reward is given to the process and the channel's transition probabilities are left untouched. Hence, the process propagates the rewards to the next state and the discount factors are set to  $I$ . For policies with action  $C$ , when the process is in state  $0$ , the process is awarded with  $I$  for a successfully transmitted and decoded frame and the discount factor is set to  $I$  since the process propagates the aggregated rewards to the next state. Hence, the channel's transition probabilities are left untouched. However, if the process is in state  $1$  where action  $C$  is in effect, the decoder of predictive coding is forced to stop when a lost packet is detected within a GOV and the state is considered as a trapping state for action  $C$ ;  $p_{11} = 1$ . When the decoder is stopped due to a lost packet, to provide smooth video presentation without blank screen, it uses the last successfully decoded frame prior to the failed one from either sub-sequence to replace the missing and effected frames, and then it restarts

when a successfully transmitted  $I$ -frame of a new GOV arrives to the decoder. Therefore, in such case, no rewards are given to the process and the process won't be able to propagate the aggregated rewards.

For the aggregated rewards computation, the instant reward is multiplied by a stationary probability  $\pi$  for the initial state due to the periodic appearance of the new  $I$ -frame which does not have any temporal dependencies to the previously decoded frames and hence, it is assumed that the first packet in  $I$ -frame arrives to the process with the stationary probability. In addition, when ISC coded sequences are packetized for transmission, the number of packets per frame varies with the bitrate, frame rate, and the packet size. Moreover, the number of packets per frame within a sequence varies depending on the coding type, (e.g., Intra-frame coding ( $I$ -frame) and Inter-frame coding ( $P$ -frame)), and the motion of the sequence. Therefore, due to the unpredictability of the variation of the number of packets per each coded frame, an average number of packets per frame  $\eta$  are used for our proposed MDP model to compute the aggregated reward.

$$\eta = \text{bitrate} / (\text{framerate} \times \text{packet size}) \quad (7)$$

$$v^{(k,j)}(s^{(k,j)}(0)) = r_C(0) \times \left[ (p_{00})^{\eta-1} 1 - (p_{00})^{\eta-1} \right]^T \times \pi \quad (8)$$

$$\begin{aligned}
 v^{(k,j)}(s^{(k,j)}(n)) &= r_C(n-1) + \gamma_C \times \left[ (p_{00})^{\eta-1} \right] \times \left[ p_C \bullet p_S^{(s^{(k,j)}(n) - s^{(k,j)}(n-1) - 1)} \right. \\
 &\quad \left. \bullet v^{(k,j)}(s^{(k,j)}(n-2)) \right] \\
 &\quad \text{for } 1 \leq n \leq N-1
 \end{aligned} \quad (9)$$

Here,  $\left[ (p_{00})^{\eta-1} 1 - (p_{00})^{\eta-1} \right]^T$  is multiplied since our MDP model assumes that a frame is decoded if and only if all the packets in that coded frame are successfully transmitted. For each interleaving set  $k$ , the sum of aggregated rewards corresponds to the expected number of successfully decoded frames and the following equations find an ISC set  $k$  that satisfies our decision criteria, a set with the highest MRP aggregated reward.

$$v^{(k)} = \sum_{j=1}^2 \sum_{n=0}^{N-1} v^{(k,j)}(s^{(k,j)}(n)) \quad (10)$$

$$\arg \max_k [v^{(k)}] \quad (11)$$

Hence, the set of aggregated rewards is expressed as:

$$V^{(k)} = \bigcup_j V^{(k,j)}(s^{(k,j)}) \quad (12)$$

$$\text{where } V^{(k)}(s^{(k,j)}(n)) = v^{(k,j)}(s^{(k,j)}(n)), \quad \text{for } 0 \leq n \leq N-1$$

### 3) Dynamic Programming with MDP

When predictive video coding uses frozen frames for the decoder failed frames, the distances between the replacement and the replaced frames have effects on the smoothness of the sequence flow and the overall quality of the playback sequence since the shorter distance between the replacing frames indicates highly correlated frame replacement in place of decoder failed frames. Hence, ISC is expected to produce smoother and higher quality video over erasure channels when frames are replaced, since the distance from a replacement frame to the replaced ones are expected to be shorter.

To incorporate frame replacement action, ISC adopts a Dynamic Programming to find an ISC set with the highest MDP sum of the aggregated reward with correlation gain  $g^{(k)}$ . To compute the sequence specific correlation gain, following steps

are used: First, temporal correlation of the transmitting sequence is computed with average PSNR between original sequence and temporally shifted sequences (13), where  $d$  is temporal distance between the original and shifted sequence. Second, Minimum Mean Square Estimator (MMSE) is employed to obtain a function that represents temporal correlation of a given sequence (14). Third, a distance matrix  $D^{(k)}$  is generated for each ISC set  $k$  for single-packet-loss per GOV cases since the main purpose of ISC is to isolate decode failure to one sub-sequence.  $D^{(k)}$  is a  $2N$  by  $2N$  upper triangular matrix and its diagonal indices indicate the first frame location in a GOV impacted by a single packet loss. Hence, the non-zeros entries represent the distances from replacement frames to the replaced ones. At last, the correlation weight matrix  $W^{(k)}$  (15) with respect to the distances from replacement frames to the replaced ones, and the correlation computed aggregated reward gain matrix  $G^{(k)}$  (17) is computed to obtain  $g^{(k)}$  (18). Hence, the optimal interleaving set with correlation model is found with (19).

$$\rho = \bigcup_{d=1} \frac{\text{avg.PSNR}(s, s+d)}{\text{avg.PSNR}(s, s)} \quad (13)$$

$$\arg \min_{\{\rho, a, c\}} \left[ \text{MSE} \left\{ \rho, a \times \exp(-d^b) + c, \forall d \right\} \right] \quad (14)$$

$$W_{x,y}^{(k)} = \begin{cases} a \times \exp\left(-\left(D_{x,y}^{(k)}\right)^b\right) + c, & \forall D_{x,y}^{(k)} \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

$$V^{(k)*} = \left[ V^{(k)}(2N-1) \quad V^{(k)}(0) \quad \dots \quad V^{(k)}(2N-2) \right] \quad (16)$$

$$G_{x,y}^{(k)} = \begin{cases} W_{x,y}^{(k)} \times V^{(k)*}(y - D_{x,y}^{(k)}), & \forall D_{x,y}^{(k)} \neq 0 \\ V^{(k)}(y), & \text{otherwise} \end{cases} \quad (17)$$

$$g^{(k)} = \sum_{x,y=1}^{\text{GOV SIZE}} G_{x,y}^{(k)}, \quad \forall x, y \leq \text{GOV SIZE} \quad (18)$$

$$\arg \max_k \left[ v^{(k)} + g^{(k)} \right] \quad (19)$$

However, due to delay, complexity, and memory constraints, measuring the temporal correlation among video frames within a complete GOV may not be always feasible for realtime applications. Therefore, a generic correlation model may be required if the actual correlation cannot be computed.

$$VI^{(k)} = \bigcup_j VI^{(k)}(s^{(k,j)}) \text{ where } VI^{(k)}(s^{(k,j)}(0)) = v^{(k,j)}(s^{(k,j)}(0)), \\ VI^{(k)}(s^{(k,j)}(n)) = v^{(k,j)}(s^{(k,j)}(n)) - v^{(k,j)}(s^{(k,j)}(n-1)), \quad (20) \\ \text{for } 1 \leq n \leq N-1$$

$$W^{(k)*} = \left( \left( D^{(k)*} \times VI^{(k)*T} \right) \div \sum_{y=1}^{\text{GOV SIZE}} D_{x,y}^{(k)*}, \forall x \right) \wedge D^{(k)}, \quad D^{(k)*} = \begin{cases} 0, & \forall D_{x,y}^{(k)} \neq 0 \\ 1, & \text{otherwise} \end{cases} \quad (21)$$

$$G_{x,y}^{(k)*} = \begin{cases} W_{x,y}^{(k)*} \times V^{(k)*}(y - D_{x,y}^{(k)}), & \forall D_{x,y}^{(k)} \neq 0 \\ V^{(k)}(y), & \text{otherwise} \end{cases} \quad (22)$$

$$g^{(k)*} = \sum_{x,y=1}^{\text{GOV SIZE}} G_{x,y}^{(k)*}, \quad \forall x, y \leq \text{GOV SIZE} \quad (23)$$

Here,  $VI^{(k)}$  is the set of the reward increments at each sub-sequences' reward calculation iteration. In (21), the term  $\left( \left( D^{(k)*} \times VI^{(k)*T} \right) \div \sum_{y=1}^{\text{GOV SIZE}} D_{x,y}^{(k)*}, \forall x \right)$  is the average reward increment of the successfully decoded frames in case of a single error in a GOV and the decrement of the weight matrix is assumed to be exponential with respect to temporal distances from the replacement frames to the replaced ones. Hence, the optimal

interleaving set using the above generic correlation model can be found using the following equation.

$$\arg \max_k \left[ v^{(k)} + g^{(k)*} \right] \quad (24)$$

### III. SIMULATIONS AND RESULTS

We have used CIF sequences of *Akiyo*, *Foreman*, *Coastguard*, and *Mobile* for simulation, and they were coded into *Intra-* (I) and *Inter-* (P) frames using an MPEG-4 encoder. When coding the sequences, un-interleaved GOV sizes of 10, 12, 14, 16, 18, and 20 were used with frame rate of *15fps*, bitrate of *250kbps* and *500kbps*, and packet size of *512Byte*. When coded frames are packetized, to limit the impact of a single packet loss to a single frame, no packets are shared among two consecutive coded frames, in other words, each packet contains data that belongs to only one video frame. Furthermore, partial decoding is not employed for the frames with errors and they are replaced with the last successfully decoded frames (frozen frames). Three ISC scenarios are simulated: (a) non-correlation model (11), (b) correlation gain model (19), and (c) generic correlation gain model (24) referred as ISC-NC, ISC-C, and ISC-GC, respectively. The ISC-NC generates an interleaved pattern that is independent of the video sequence, and hence, it generates ISC pattern depending only on the erasure-channel Markov model. It is important to note that the ISC-GC case captures the correlation among frames in a generic sense, and it does not measure correlation based on actual computation of the correlation among the video frames. Hence, the ISC-GC scenario is mainly dependent on the original GOV size of the video sequence being coded. To simulate a statistically viable experiments and to capture a realistic network loss patterns, ten error traces were generated using the packet-loss Markov transition probabilities from [14][15] ( $p_{00} = 0.9734$   $p_{11} = 0.7052$ ). Each ISC pattern is fitted into these error traces and the PSNR values are averaged to provide statistically satisfying results for analysis.

Fig. 4 shows the obtained (averaged) PSNR as a function of the GOV size for different bitrates. Here, the non-ISC cases show linear downward trend with respect to the GOV size and bitrate. This implies that such variations have negative impacts on the quality since they increase the average number of packets per frame, which in turn causes an increase in the number of GOVs impacted by lost packets, the average number of replaced frames, and the distance between the replacement frames. For the ISC cases, with the GOV size increment, the average PSNR shows linear trends similar to the non-ISC cases. However, the slope is rather flat when compared to the non-ISC cases, which implies that the GOV size variation has less negative impact on ISC method compared to the traditional non-ISC method. When the sequences are coded using the same coding method at the same GOV size, but with the different bitrates, e.g., *250kbps* and *500kbps*, Fig. 4 shows that variation of bitrate has less impact on the PSNR values for the ISC cases than the non-ISC cases; hence this shows that ISC reduces the negative impact of increased  $\eta$ , the average number of packets per frame, as stated previously. In addition, since the average PSNR gain of ISC cases over non-ISC cases are higher, this implies that the ISC method performs better when coded at higher bitrate. The correlation-based models, both ISC-C and ISC-GC, provide

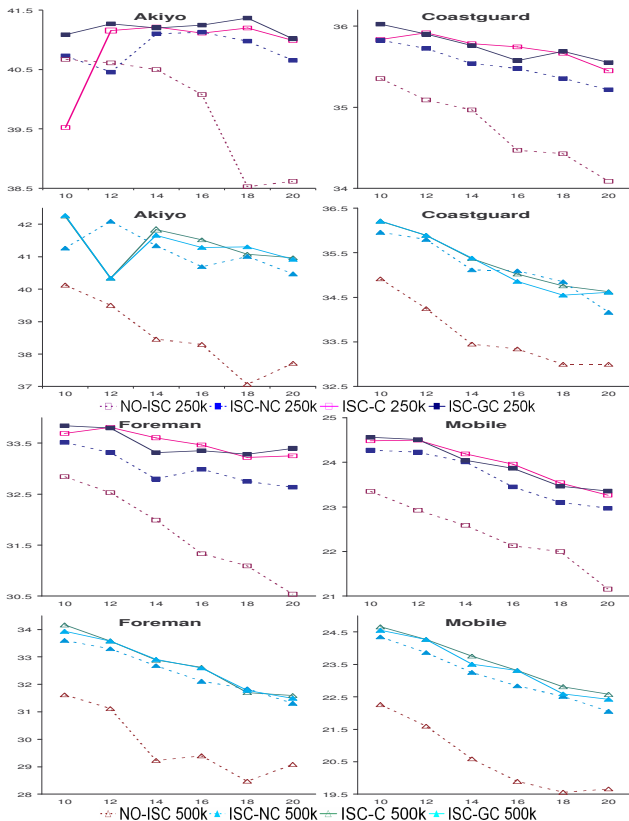


Figure 4. Average PSNR (GOV Size vs. PSNR(dB))

improvements over the non-correlation (ISC-NC) based scenario, and hence demonstrate the advantages of the correlation gain computation. When comparing the two different correlation model sets, the generic correlation model shows competitive results, and it is plausible to use the generic model in cases when the actual temporal correlation for a given sequence is not feasible to compute.

In summary, observation shows that the proposed ISC method improves over the traditional approach on most of the cases, up to 4 dB in average PSNR, which represents a significant improvement in quality for compressed video applications, especially for the sequences with high motion or low temporal correlation. In particular, this demonstrates that ISC improves the quality of predictive coded sequences over an erasure channel by limiting errors to one of the two sub-sequences, hence minimizing the cascaded effects of lost packets, and/or decreasing the average frame replacement distance. In addition, bitrate or GOV size variations have less impact on ISC coded sequences. Furthermore, when the non-correlation gain computed ISC sets are compared to the correlation computed sets, the latter sets show some modest improvement in PSNR for most of the evaluation cases. Consequently, it is feasible that significant improvements can be gained by taking into consideration the channel model *only*, and hence, reducing the complexity for identifying the interleaving set. Once the interleaving is identified for a given channel model, it can be applied to any video sequence (i.e., without taking into consideration the statistical properties of the video sequence).

#### IV. CONCLUSION

In this paper, we proposed an *interleaved source coding* (ISC)

method for predictive coded video sequences over the internet. This new method is resilient to packet losses since it limits the errors due to packet losses to one of two ISC generated sub-sequences and minimizes cascaded effects of packet losses over a *single* erasure-channel model. Hence, ISC increases the number of successfully decoded frames and overall playback quality of the decoded video sequence. The optimal ISC sets are found using a Dynamic Programming and a Markov Decision Process with respect to the packet loss rate, temporal correlation of the sequences and the bit rate for the coder. Unlike other methods (e.g., [1]-[4] [6]-[9][11]-[13]), ISC does not require complex modification of the coding standards and eliminates the need for content distribution, channel selection and synchronization issues. It is clearly shown that ISC advances traditional predictive coded sequence transmission method; however, improvements on finding the *true* optimal interleaving sets are required and they are left for future work. Some of our future extension includes ISC over wireless, ISC with forward error correction (FEC), and multi-channel ISC.

#### REFERENCES

- [1] Apostolopoulos, J. G., "Error-resilient video compression through the use of multiple States," IEEE Proc. ICIP, September 2000.
- [2] Barrencea, G., Beferull-Lozano, B., Verma, A., Dragotti, P., and Vetterli, M., "Multiple description source coding and diversity routing: A joint source channel coding approach to real-time services over dense networks," Packet Video, April 2003.
- [3] Begen, A., Altunbasak, and Y., Ergun, O., "Multi-path selection for multiple description encoded video streaming," IEEE Proc. ICC, May 2003.
- [4] Franchi, N., Fumagalli, M., Lancini, R., and Tubaro, S., "Multiple description video coding for scalable and robust transmission over IP," Packet Video, April 2003.
- [5] Gallager, R., Discrete Stochastic Processes, Kluwer Academic Publishers, 1996.
- [6] Girod, B., "Efficiency analysis of multihypothesis motion-compensated prediction for video coding," IEEE Trans. Image Processing, vol. 9, no. 2, pp. 173-183, February 2000.
- [7] Khayam, S. and Radha, H., "Markov-based modeling of wireless Local area networks," ACM Mobicom Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM), September 2003
- [8] Liang, Y., Apostolopoulos, J., and Girod, B., "Model-based delay-distortion optimization for video streaming using packet interleaving," Proc. Asilomar Conference on Signals, Systems, and Computers, November 2002
- [9] Lin, S. and Wang, Y., "Error resilience property of multihypothesis motion-compensated prediction," IEEE Proc. ICIP, September, 2002.
- [10] Puterman, M., Markov Decision Processes: Discrete Stochastic Dynamic Programming, John Wiley & Sons, Inc., New York, NY, 1994.
- [11] Radha, H., Chen, Y., Parthasarathy, K., and Cohen, R., "Scalable internet video using MPEG-4," Signal Processing: Image Communication, vol. 15, pp. 95-126, 1999.
- [12] Radha, H., van der Schar, M., and Chen, Y., "The MPEG-4 FGS video coding method for multimedia streaming over IP," IEEE Trans. Multimedia, vol. 3, issue 1, pp. 53-68, March 2001.
- [13] Reibman, A. R., Jafarkhani, H., Wang, Y., Orchard, M. T., and Puri, R., "Multiple description coding for video using motion compensated prediction," IEEE Proc. ICIP, October 1999.
- [14] Yajnik, M., Kurose, J., and Towsley, D., "Packet loss correlation in the Mbone multicast network", IEEE Global Internet Miniconference, part of GLOBECOMM, November 1996.
- [15] Yajnik, M., Moon, S., Kurose, J., and Towsley, D., "Measurement and modeling of the temporal dependence in packet loss", IEEE Proc. INFOCOM, 1999