

Interleaved Source Coding (ISC) for Predictive Video over ERASURE-Channels

Jin Young Lee, *Member, IEEE* and Hayder Radha, *Senior Member, IEEE*

Abstract— Packet losses over unreliable networks have a severe impact on the playback quality of many predictive coded sources such as compressed video. Prior efforts (e.g., [1]-[5] [7]-[9][11]-[13]) have developed a variety of coding methods that are resilient to packet losses. We propose a new packet-loss resilient coding approach, *interleaved source coding* (ISC), which is based on an optimum interleaving of predictive video coded frames transmitted over a *single* erasure channel. We develop a Markov Decision Process (MDP) and a corresponding dynamic programming algorithm for identifying the optimal interleaving pattern for a given channel model. This method improves the overall quality of predictive video coded stream over a lossy channel without complex modifications to standard video coders. ISC provides a viable alternative to (or it could be combined with) path-diversity based approaches, and hence, ISC eliminates (or reduces) the need for content distribution, path diversity routing, and related synchronization issues. Simulations of a wide range of video sequences over practical traces of Markov erasure channels showed significant improvements (up to 4 dB) when compared with traditional predictive video over the same channels.

Index Terms—Dynamic Programming, Interleaving, Markov Decision Process, Packet Losses, Video Coding

I. INTRODUCTION

Streaming video is emerging as one of the most popular on-line realtime Internet applications. It is often used for multimedia content transmission such as video chat, live news, video conferencing, etc. Such realtime streaming video services often lack Quality-of-Service (QoS) guarantees which in turn degrades playback quality due to network impairments, e.g., packet losses. Therefore, for playback quality improvement of realtime streaming video under such condition, special coding techniques resilient to packet losses are required. Techniques such as scalable coding [11][12], multi-hypothesis motion estimation and compensation [7][9], multi state video compression [1], and multiple description coding (MDC) with path diversity [2]-[5] are few examples of methods to be resilient to packet losses.

In this paper, we propose a new packet loss resilient video-coding approach based on *interleaved source coding* (ISC) for predictive video sequences. This method codes a single video sequence into two sub-sequences and transmits

them over a *single* erasure channel. Our proposed ISC interleaving method reduces the frequency and impact of the cascaded effect of packet losses and related propagation of errors resulted from the predictive nature of coded video. Particularly, we target the design of optimum interleaving such that the impact of losses caused by a given erasure channel model (with memory) is limited to a minimum number of video frames. In addition, in case of decoder failed frame replacement, frozen frames, ISC presents smoother video compared to the non-interleaving method.

The proposed ISC video coding differs from previous Multiple-Description-Coding (MDC) based methods (e.g., ones proposed in [2]-[5]) since ISC is primarily designed for transmission of encoded sequences over a *single* channel. This eliminates channel selection, content distribution, and synchronization issues known to present with MDC [2]-[5]. Furthermore, interleaving could reduce the level of coding inefficiency that normally characterizes MDC coding. Nevertheless, we believe that the proposed interleaved coding framework can be generalized for transmission over multiple channels, and hence, it could include some form of MDC. In this paper, however, we focus on interleaved coding for the *single* erasure-channel case. To find an interleaving set, we employ a Markov Decision Process (MDP) and a Dynamic Programming algorithm in association with a realistic packet loss model. We also take into consideration some coarse measure of the temporal correlation among pictures within a given video sequence. This temporal correlation results in interleaving sets that are unique to each video sequence.

The remainder 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 Sub-Section II-A and a mathematical approach to find the optimal interleaving set using a Markov reward process, Markov Decision Process (MDP), and a Dynamic Programming algorithm are described in Sub-Section II-B. In Section III, our proposed method is evaluated using MPEG-4 video simulated over an Internet Markov-based lossy channel model.

II. METHODOLOGY

A. General Interleaving

Traditional predictive video coding partitions a single lengthy sequence into a number of shorter length Group Of Video object planes (GOVs). It is well known that this

J. Y. Lee is with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea, on leave from Michigan State University, East Lansing, MI 48824 USA (Phone: +82-42-860-5383; Fax: +82-42-860-1342 Email: jinlee@etri.re.kr or leejinyo@egr.msu.edu)

H. Radha is with Michigan State University, East Lansing, MI 48824 USA (radha@egr.msu.edu).

partitioning limits the impact of possible errors or losses into individual GOVs.

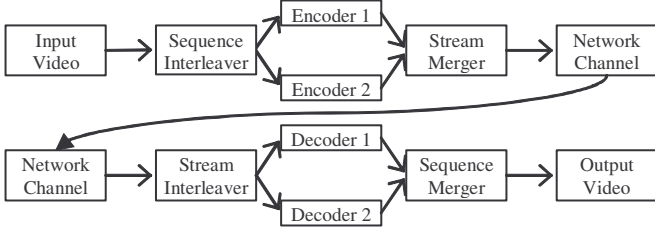


Fig 1. Interleaving of Predictive Video Coding.

The proposed *interleaved source coding* (ISC) is a pre- and post-process of predictive source coders¹ (Fig 1.). ISC reduces the impact of losses within a given GOV and improves the overall quality of predictive video over lossy packet networks.

Brief description of the overall ISC process is the following: First, ISC separates a single video sequence into two sub-sequences² using a *Sequence Interleaver*, and the resulting sub-sequences are encoded using separate video encoders. Then, a *Stream Merger* merges the encoded frames into a single stream in the original-sequence frame order for transmission. In addition to the ISC merged-stream, information regarding the interleaving pattern employed by the encoder must be transmitted to the decoder prior to the ISC merged-stream transmission. At the decoder side, the interleaving pattern is used by a corresponding pair of *Stream Interleaver* and *Sequence Merger*. Hence, the decoder side's *Stream Interleaver* separates the incoming frames or associated packets into two sub-streams according to the transmitted interleaving pattern information. The separated streams are decoded independent to each other and the *Sequence Merger* finalizes the process by merging the sub-sequences' frames into the proper order for playback.

When separating a single sequence into two sub-sequences, $s^{(j)}$, represented by an index set, $j = \{1, 2\}$, we adopt the following ISC interleaving constraints;

$$s = \{0 \ 1 \ \dots \ 2N - 1\} = \bigcup_{j=1}^2 s^{(j)}$$

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

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

where $(2N - 1)$ is the number of frames in the original non-interleaved sequence s . In practice, $(2N - 1)$ could be the

¹ It is possible to integrate *Interleavers and Mergers* into the predictive source coders and use a single encoder and decoder; however, to simplify ISC adaptation, we employ ISC as a pre- and post-process of the coders and leave the coders untouched.

² The proposed interleaved coding framework could support more than two sub-sequences. Here, we only focus on the simple case of two sub-sequences.

number of frames in a GOV, and hence, the same interleaving is applied to all GOVs in the sequence or a scene.

For example, for a non-interleaved sequence with a GOV size of 10, let $\mathbb{S} = \{s^{(1)}, s^{(2)}\}$ be an interleaving sub-sequence set with $s^{(1)} = \{0 \ 1 \ 5 \ 6 \ 9\}$ and $s^{(2)} = \{2 \ 3 \ 4 \ 7 \ 8\}$ (Fig 2).

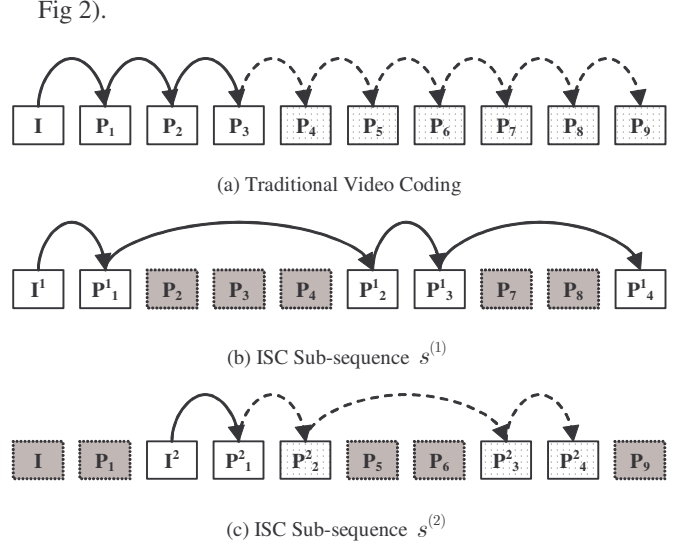


Fig 2. Traditional vs. ISC Video Coding with a packet loss in the frame location of P_4 in (a). The arrowed lines represent the coded frames temporal dependencies in the predictive video coding. The dotted frames are the decoder failed frames due to the loss. The shaded frames are belonged to the other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

Here, the numbers in $s^{(j)}$ represent the frame locations in the non-interleaved sequence and the coded stream's frame transmission order. This interleaving information is required to be transmitted (e.g., as meta data) with the coded sub-streams as stated previously. Once separated, the sub-sequences are encoded as $I^1 P^1_1 P^1_2 P^1_3 P^1_4$ and $I^2 P^2_1 P^2_2 P^2_3 P^2_4$ for $s^{(1)}$ and $s^{(2)}$, respectively, and they are transmitted in the following order: $I^1 P^1_1 I^2 P^2_1 P^2_2 P^1_2 P^1_3 P^2_3 P^2_4 P^1_4$; in other words, the merged coded sequence is transmitted in the same frame transmission order of the non-interleaved traditional video coder. During transmission, if a packet is lost that, for example, is a part of the 5th frame (P_4 , in

Fig 2-(a)), all 6 frames from P_4 to P_9 of the non-interleaved coding are impacted severely and would not be decoded correctly. However, with interleaving, all the frames in sub-sequence $s^{(1)}$ are decoded successfully and only three frames, P^2_2 , P^2_3 , and P^2_4 , from the sub-sequence $s^{(2)}$ are not decoded. Hence interleaving improves overall playback quality by limiting errors (due to packet losses) to $s^{(2)}$.

Since the formation of the optimal interleaving set could vary depending on the channel model and the transmitting sequence, a problem rises here in choosing the optimal set from the set of all possible interleaved sequences. Let \mathbb{K} be the set of all possible interleaving sets for a given GOV size. The size of the set \mathbb{K} can be expressed as follows:

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

TABLE 1. NUMBER OF POSSIBLE INTERLEAVING SET, \mathbb{K}

GOV Size	10	12	14	16	18	20
Size K	121	456	1709	6427	24301	92368

As shown in Table 1, the size of the set \mathbb{K} could be quite large for any reasonable GOV size $(2N - 1)$. Hence, identifying the optimum interleaving set that produces the best quality decoded video transmitted over a lossy network channel could be very computationally expensive task due to the vast size of \mathbb{K} (Table 1). Therefore, an efficient decision-based search algorithm is required to choose and optimal interleaving set that gives the best quality video for a given erasure-channel model and a video sequence.

B. Decision Based Interleaving

1) Markov Reward Process

Previous efforts for the analysis and modeling of packet losses over the Internet (e.g., [14][15]) and wireless networks (e.g., [8]) have shown that these losses exhibit Markovian properties. For a Markov channel model, a *Markov Reward Process* (MRP) (e.g., [6][10]) can estimate the system's performance using (a) the Markov channel's transition probabilities based on the packet transmission and (b) some model for the *rewards* that are associated with each system state. This reward-based MRP could be used to measure the system's performance after n packet transmissions, and this, in turn, could guide the design of our ISC coding system (as explained further below).

To establish a MRP for an erasure-channel model, let $\{0,1\}$ be the corresponding state space to good (0) and bad (1) packet transmissions.³ The instant rewards r_i are assigned for each state and they are awarded to the process whenever it reaches state i (Fig 3).

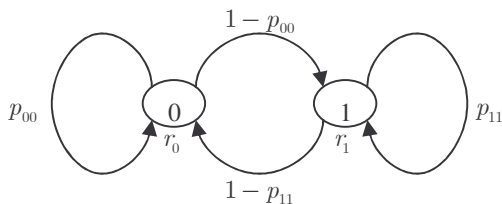


Fig 3. Two state Markov model with rewards r_i

For the transmission of a predictive coded (and packetized) sequence over a lossy Markov channel with a channel's state

transition matrix p (Table 2), we define the aggregated reward $v(n-1)$ ([6][10]) as a function of the number of transmitted packets.

TABLE 2. TWO STATE MARKOV TRANSITION MATRIX, p

(a) General Representation			(b) Actual values from [14][15]		
Current \ Future	0	1	Current \ Future	0	1
0	p_{00}	$1 - p_{00}$	0	0.9734	0.0266
1	$1 - p_{11}$	p_{11}	1	0.2948	0.7052

After n packet transmissions, the aggregated reward $v(n-1)$ represents the performance of predictive sequence transmission over a lossy channel with a channel's state transition matrix p .

$$v(0) = r = \{r_0, r_1\}^T \quad (3)$$

$$v(1) = r + p \cdot v(0) \quad (4)$$

$$\begin{aligned} v(n-1) &= \begin{bmatrix} v_0(n-1) & v_1(n-1) \end{bmatrix}^T \\ &= r + p \cdot v(n-2) \\ &= r + p \cdot r + p^2 \cdot r + \dots + p^{n-2} \cdot r + p^{n-1} \cdot r \quad (5) \\ &= \left(I + \sum_{m=1}^{n-1} p^m \right) \cdot r \end{aligned}$$

For example, in a two state Markov channel model, if the instant rewards are $\{r_0, r_1\} = \{1, 0\}$, the reward process is awarded with I for a successful packet and 0 for a lost packet during the transmission. In this case, after n packet transmissions, the aggregated rewards, $v_i(n-1)$, represent the expected number of good packet transmissions with the initial packet transmission at state i .

2) Markov Decision Process

A Markov Decision Process (MDP) associates a Markov reward process with a series of actions and decision criteria [6][10]. In our case, we employ MDP to find an interleaving set that is most suitable for a given decision criteria. In general, our objective is to maximize the number of frames (or associated packets) that can be decoded *correctly*. Hence, an MDP could guide us toward an "optimal" interleaving for a given erasure channel model that achieves our objective; the interleaving set that provides the highest sum of MRP aggregated reward. Since there are many possible interleaving sets, we use an *interleaving set indicator* k , $k \in \mathbb{K}$. Further, in MDP, a set of *policies*, mappings from states to actions, are associated with a set of *discount factors*, γ_a [6][10]. The discount factors decide the amount of aggregated reward to be propagated to the next state. Incorporating equation (5) with the discount factors and the interleaving set indicator k gives an aggregated MDP equation:

³ It is possible to use higher order Markov models; 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]).

$$\begin{aligned}
v^{(k)}(n-1) &= r_{a^{(k)}(n-1)} + \gamma_{a^{(k)}(n-1)} \times p \bullet v^{(k)}(n-2) \\
&= r_{a^{(k)}(n-1)} + \gamma_{a^{(k)}(n-1)} \times p \bullet r_{a^{(k)}(n-2)} + \\
&\dots + \left(\prod_{t=2}^{n-1} \gamma_{a^{(k)}(t)} \right) \times p^{n-2} \bullet r_{a^{(k)}(1)} \\
&\dots + \left(\prod_{t=1}^{n-1} \gamma_{a^{(k)}(t)} \right) \times p^{n-1} \bullet r_{a^{(k)}(0)}
\end{aligned} \tag{6}$$

In the proposed ISC interleaving method, we consider each frame in a GOV as a state iteration in the Markov model⁴. Based on the policies described in Table 3, one of the two actions, Coding (C), or Skip (S), is taken for each state iteration. $a^{(k)}(n)$ denotes an action taken for the n^{th} frame in a GOV.

Table 3. Properties of MDP for Multimedia Stream Interleaving

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

Let the set of ISC sub-sequences in

Fig 2 be the interleaving set k . With respect to k , an ISC set is written as $\mathbb{S}^{(k)} = \{s^{(k,1)}, s^{(k,2)}\}$. In our interleaved predictive video model, each sub-sequence has its own Intra-coded I -frame⁵. Consequently, the frame numbers are rewritten so that each sub-sequence's reward computation starts from the time instance 0.

$$s^{*(k,j)}(n) = s^{(k,j)}(n) - s^{(k,j)}(0), \quad \text{for } 0 \leq n \leq N-1 \tag{7}$$

Associating the above equation with $\mathbb{S}^{(k)}$ from

Fig 2 gives

$$\begin{aligned}
\{s^{*(k,1)}(0) \dots s^{*(k,1)}(N-1)\} &= \{0 \ 1 \ 5 \ 6 \ 9\} \quad \text{and} \\
\{s^{*(k,2)}(0) \dots s^{*(k,2)}(N-1)\} &= \{0 \ 1 \ 2 \ 5 \ 6\}.
\end{aligned}$$

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

⁴ Here, we make the simplifying assumption that each video frame is transmitted within a single packet. As our simulations show, this assumption still leads to significant gains in quality even when each video frame is transmitted using multiple packets.

⁵ It is possible to have a single I -frame shared among the interleaved sub-sequences though. This I -frame could be also protected and transmitted in a highly reliable way. In this case, the main design issue will be the interleaving of the predictive frames within the sequence GOVs.

$$\begin{aligned}
m &= \bigcup_{n=0}^{N-1} \{s^{*(k,j)}(n) + 1, \dots, s^{*(k,j)}(n+1) - 1\} \\
\forall n \quad &| \quad s^{*(k,j)}(n+1) - s^{*(k,j)}(n) > 1
\end{aligned} \tag{8}$$

This gives the action sets $\alpha^{(k,j)}$ for the interleaving set $\mathbb{S}^{(k)}$ from

$$\begin{aligned}
\text{Fig 2 as } \alpha^{(k,1)} &= [C \ C \ S \ S \ S \ C \ C \ S \ S \ C] \\
\text{and } \alpha^{(k,2)} &= [C \ C \ C \ S \ S \ C \ C].
\end{aligned}$$

In addition, our MDP model requires modification of the channel's transition matrix p in association with actions. For the policy $\{C, 1\}$, since the decoder of predictive coding is forced to stop when a lost packet is detected, the state I is considered as a trapping state for action C . In our MDP model, once the decoder is stopped due to a lost packet, it uses the last successfully decoded picture 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. For all other policies, the channel's transition probabilities are used since the frame with successfully transmitted packets or lost packets in skipped frames do not affect the decoder. Further, the discount factors for our MDP model for the policy $\{C, 1\}$ is set to 0, since the policy forces the decoder to stop and no further decoding is possible, hence aggregated reward is not propagated unless the decoder is restarted. For all other policies, the process propagates the rewards to the next state and the discount factors are set to 1.

When computing the aggregated rewards, for the initial state $v^{(k,j)}(0)$, the instant reward is multiplied by a stationary probability π . This is due to the periodic appearance of the new I -frame which does not have any temporal dependencies to the previously decoded frames. Hence, it is assumed that the first packet in I -frame arrives to the process with the stationary probability. Therefore, the proposed MDP model's aggregated reward equations for single-packet-per-frame are:

$$v^{(k,j)}(s^{*(k,j)}(0)) = r_C(0) \times \pi \tag{9}$$

$$\begin{aligned}
v^{(k,j)}(s^{*(k,j)}(n-1)) &= r_C(n-1) + \\
\gamma_C \times &\left(p_C \bullet p_S^{(s^{*(k,j)}(n) - s^{*(k,j)}(n-1) - 1)} \bullet v^{(k,j)}(s^{*(k,j)}(n-2)) \right) \\
&, \forall n \in s^{*(k,j)}
\end{aligned} \tag{10}$$

This is valid since the aggregated reward for a skipped frame is:

$$\begin{aligned}
v^{(k,j)}(m) &= r_S(m) + \gamma_S \times p_S \bullet v^{(k,j)}(m-1) \\
&= \begin{bmatrix} 0 & 0 \end{bmatrix}^T + \begin{bmatrix} 1 & 1 \end{bmatrix}^T \times p_S \bullet v^{(k,j)}(m-1) \\
&= p_S \bullet v^{(k,j)}(m-1)
\end{aligned} \tag{11}$$

When coded sequences are packetized, the number of packets per frame varies with the bitrate and frame rate of the encoder, and the packet size. In addition, within a sequence, the number of packets per frame 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, our proposed MDP model uses an average number of packets per frame η and the aggregated reward equations are as follows.

$$\eta = \left\lceil \frac{\text{bitrate}}{\text{framerate} \times \text{packetsize}} \right\rceil \quad (12)$$

$$v^{(k,j)}(s^{*(k,j)}(0)) = r_C(0) \times \left[(p_{00})^{\eta-1} \quad 1 - (p_{00})^{\eta-1} \right]^T \times \pi \quad (13)$$

$$v^{(k,j)}(s^{*(k,j)}(n)) = r_C(n-1) + \gamma_C \times \begin{bmatrix} (p_{00})^{\eta-1} \\ 1 - (p_{00})^{\eta-1} \end{bmatrix} \times \begin{bmatrix} p_C \bullet p_S^{(s^{*(k,j)}(n) - s^{*(k,j)}(n-1))} \\ \bullet v^{(k,j)}(s^{*(k,j)}(n-2)) \end{bmatrix} \quad (14)$$

for $1 \leq n \leq N-1$

The term $\left[(p_{00})^{\eta-1} \quad 1 - (p_{00})^{\eta-1} \right]^T$ is multiplied to the aggregated reward since a frame is decoded if and only if all the packets in the coded frames are successfully transmitted. For each interleaving set k , the sum of aggregated rewards gives corresponding expected number of successfully decoded frames.

$$v^{(k)} = \sum_{j=1}^2 \sum_{n=0}^{N-1} v^{(k,j)}(s^{*(k,j)}(n)) \quad (15)$$

Hence, the set of aggregated rewards is expressed as:

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

$$\text{where } V^{(k)}(s^{*(k,j)}(n)) = v^{(k,j)}(s^{*(k,j)}(n)), \quad (16)$$

for $0 \leq n \leq N-1$

With the following equation, Markov Decision Process finds an interleaving set k that satisfies our decision criteria, a set with the highest MRP aggregated reward.

$$\arg \max_k [v^{(k)}] \quad (17)$$

3) Dynamic Programming with MDP

In predictive video coding, when the decoder encounters a packet loss (or errors in a transmitted packet), to continue the smooth video presentation (without blank screen or distorted frames), a playback application often replaces the decoder failed frames with the last successfully decoded frame until a successfully decoded frame arrives to restart the decoding process. Here, we refer to this last successfully decoded frame as the *replacement* frame. When the decoder failed frames are replaced, the *distances* (in terms of number of pictures) between the replacement frame and the replaced frames have effects on the smoothness of the sequence flow and the overall quality of the playback sequence. This is due to the fact that the shorter distance between the replacing frames indicates highly correlated frame replacement in place of decoder failed frames. Fig 4 illustrates the frame replacement actions in case of decoder failure.

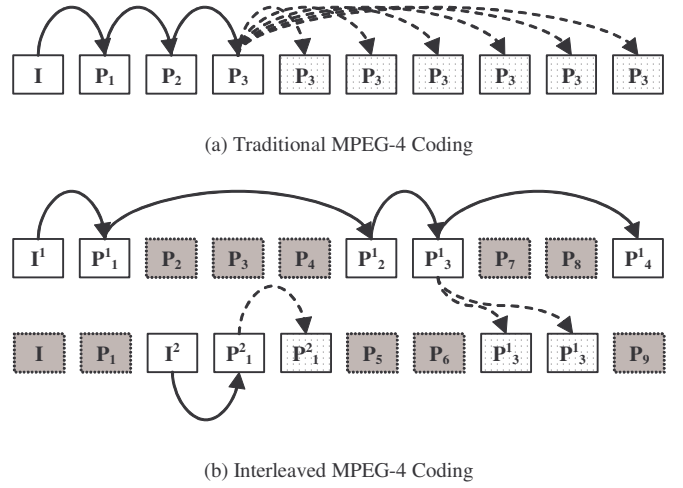


Fig 4. Frame Replacement Illustrations with a packet loss in the frame location of P_4 in (a). The dotted arrowed lines represent the frame replacement relationship for the decoder failed frames (dotted frames).

As shown in Table 4, the average frame replacement distances due to a single lost packet in a GOV is shorter for ISC than the traditional transmission method. Hence it is expected that ISC produces smoother and higher quality video over erasure channels with decoder failed-frame replacements.

TABLE 4. AVERAGE FRAME REPLACEMENT DISTANCE WITH A SINGLE LOST PACKET IN A GOV

GOV SIZE	10	12	14	16	18	20
Non-ISC	4.0000	4.6667	5.3333	6.0000	6.6667	7.3333
ISC	2.8265	2.9686	3.0740	3.1561	3.2230	3.2793

To incorporate the quality improvement from frame replacements, correlation gain $g^{(k)}$ is added to equation (15) and a Dynamic Programming is used to find an interleaving set that produces the highest MDP sum of the aggregated reward with the correlation gain $g^{(k)}$.

$$\arg \max_k [v^{(k)} + g^{(k)}] \quad (18)$$

The correlation gain $g^{(k)}$ is computed with the following steps. First, temporal correlations are computed with average PSNR between original sequence and temporally shifted sequences.

$$\rho(d) = \frac{\text{avg.PSNR}(s, s+d)}{\text{avg.PSNR}(s, s)} \quad (19)$$

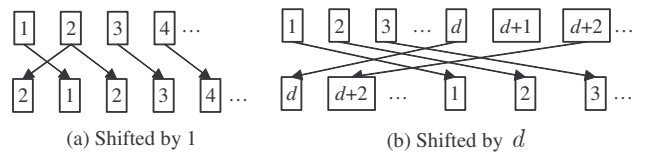


Fig 5. Sequence Shifting for Temporal Correlation Measurement

Fig 5 shows illustration on sequence shifting for the temporal

correlation measurement and the correlations are computed with equation(19) . Second, a curve fitting method with the Minimum Mean Square Estimator (MMSE) is used to obtain a function that represents temporal correlation of a given sequence.

$$\arg \min_{\{a,b,c\}} \left[MSE \left\{ \rho(d), a \times \exp(-d^b) + c, \forall d \right\} \right] \quad (20)$$

TABLE 5. DISTANCE MATRIX $D^{(k)}$ FOR $\mathbb{S}^{(k)}$ SHOWN IN FIG 4-(B)

	1	2	3	4	5	6	7	8	9	10
1	1	2	0	0	0	1	2	0	0	1
2	0	1	0	0	0	1	2	0	0	1
3	0	0	1	2	3	0	0	1	2	0
4	0	0	0	1	2	0	0	1	2	0
5	0	0	0	0	1	0	0	1	2	0
6	0	0	0	0	0	1	2	0	0	1
7	0	0	0	0	0	0	1	0	0	1
8	0	0	0	0	0	0	0	1	2	0
9	0	0	0	0	0	0	0	0	1	0
10	0	0	0	0	0	0	0	0	0	1

Third, a $2N$ by $2N$ upper triangular distance matrix $D^{(k)}$ (Table 5) is generated for each ISC set k for single-packet-loss per GOV cases, since the main purpose of interleaving method is to isolate decode failure to one sub-sequence. The distance matrices' diagonal indices indicate the first frame location in a GOV impacted by a single packet loss. Hence, the non-zeros entries of the distance matrix represent the distances from replacement frames to the replaced ones.

Finally, the correlation gain is computed with the following equations. $W^{(k)}$ is the correlation weight matrix with respect to the distances from replacement frames to the replaced ones. In case of replacements, the weight is multiplied by the aggregated reward of the replacement frame and the discounted reward is given to the replaced frame. $G^{(k)}$ is the correlation computed aggregated reward gain matrix.

$$W_{x,y}^{(k)} = \begin{cases} a \times \exp\left(-\left(D_{x,y}^{(k)}\right)^b\right) + c, & \forall D_{x,y}^{(k)} \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

$$V^{(k)*} = \left[V^{(k)}(2N-1) \quad V^{(k)}(0) \quad \dots \quad V^{(k)}(2N-2) \right] \quad (22)$$

$$G_{x,y}^{(k)} = \begin{cases} W_{x,y}^{(k)} \times V^{(k)*}(y - D_{x,y}^{(k)}), & \forall D_{x,y}^{(k)} \neq 0 \\ V^{(k)}(y), & \text{otherwise} \end{cases} \quad (23)$$

$$g^{(k)} = \sum_{x,y=1}^{\text{GOV SIZE}} G_{x,y}^{(k)}, \quad \forall x, y \leq \text{GOV SIZE} \quad (24)$$

Measuring the temporal correlation among video frames within a complete GOV may not be always feasible for realtime applications due to delay, complexity, and memory constraints. Therefore, a more generic correlation model may be required for the cases when the actual correlation cannot be computed. Below, we present such a generic model.

$$VI^{(k)} = \bigcup_j VI^{(k)}(s^{(k,j)})$$

$$\text{where } VI^{(k)}(s^{(k,j)}(0)) = v^{(k,j)}(s^{*(k,j)}(0)), \quad (25)$$

$$VI^{(k)}(s^{(k,j)}(n)) = v^{(k,j)}(s^{*(k,j)}(n)) - v^{(k,j)}(s^{*(k,j)}(n-1)),$$

for $1 \leq n \leq N-1$

$VI^{(k)}$ is the set of the reward increments at each sub-sequences' reward calculation iteration.

With respect to $D^{(k)}$ and $VI^{(k)}$, the weight matrix $W^{(k)*}$ is calculated with the following equation. Here,

$$\left(\left(D^{(k)*} \times VI^{(k)*T} \right) \div \sum_{y=1}^{\text{GOV SIZE}} D_{x,y}^{(k)*}, \forall x \right)$$

is the average reward increment of the successfully decoded frames in case of a single error in a GOV. Since the decoder failed frames are copied by the last successfully decoded frames, multiplying this value by the replacement frame's aggregated reward estimates the correlation-based aggregated reward of the replaced frame. Hence, the decrement is assumed to be exponential with respect to temporal distances from the replacement frames to the replaced ones.

$$W^{(k)*} = \left(\left(D^{(k)*} \times VI^{(k)*T} \right) \div \sum_{y=1}^{\text{GOV SIZE}} D_{x,y}^{(k)*}, \forall x \right) \wedge D^{(k)}$$

$$\text{where } D^{(k)*} = \begin{cases} 0, & \forall D_{x,y}^{(k)} \neq 0 \\ 1, & \text{otherwise} \end{cases} \quad (26)$$

$$G_{x,y}^{(k)*} = \begin{cases} W_{x,y}^{(k)*} \times V^{(k)*}(y - D_{x,y}^{(k)}), & \forall D_{x,y}^{(k)} \neq 0 \\ V^{(k)}(y), & \text{otherwise} \end{cases} \quad (27)$$

$$g^{(k)*} = \sum_{x,y=1}^{\text{GOV SIZE}} G_{x,y}^{(k)*}, \quad \forall x, y \leq \text{GOV SIZE} \quad (28)$$

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

$$\arg \max_k \left[v^{(k)} + g^{(k)*} \right] \quad (29)$$

III. SIMULATIONS AND RESULTS

A. Simulation Setup

For evaluation, CIF sequences of *Akiyo*, *Foreman*, *Coastguard*, and *Mobile* were coded into an *IPPP...* GOV structure using an MPEG-4 encoder. GOV sizes (un-interleaved size) of 10, 12, 14, 16, 18, and 20 were used to partition the evaluation sequences. Frame rate of 15 frames per second, bitrate of 250 kbps and 500 kbps, and packet size of 512 Byte are used to represent emerging Internet-access technologies (e.g., DSL/Cable and LAN connections).

When the coded sequences 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.) In addition, partial decoding is not employed for the frames with

errors and frozen frames for both ISC and traditional (non-ISC) cases. Three ISC scenarios are simulated: (a) correlation gain

computation model (equation (18)), and (b) generic correlation gain computation model (equation (29)), (c) the non-correlation computation model (equation (17)). We refer to these scenarios as ISC-C (correlation model), ISC-GC (generic correlation model), and ISC-NC (non-correlation model), respectively. The ISC-NC scenario generates an optimal interleaved pattern that is independent of the video sequence, and hence, it generates ISC pattern depending on the erasure-channel Markov model only. 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] (Table 2-(b)). Each evaluation case is fitted into these error traces and the PSNR values are averaged to provide statistically satisfying results for analysis.

B. Simulation Results and Analysis

1) Bitrate and GOV size variation effects

Fig 6 shows the obtained (averaged) PSNR as a function of the GOV size for different bitrates. In Fig 6, 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 such changes increase the average number of packets per frame η , which in turn causes an increase in (a) the number of GOVs impacted by lost packets, (b) the average number of replaced frames, and (c) the distance between the replacement frames.

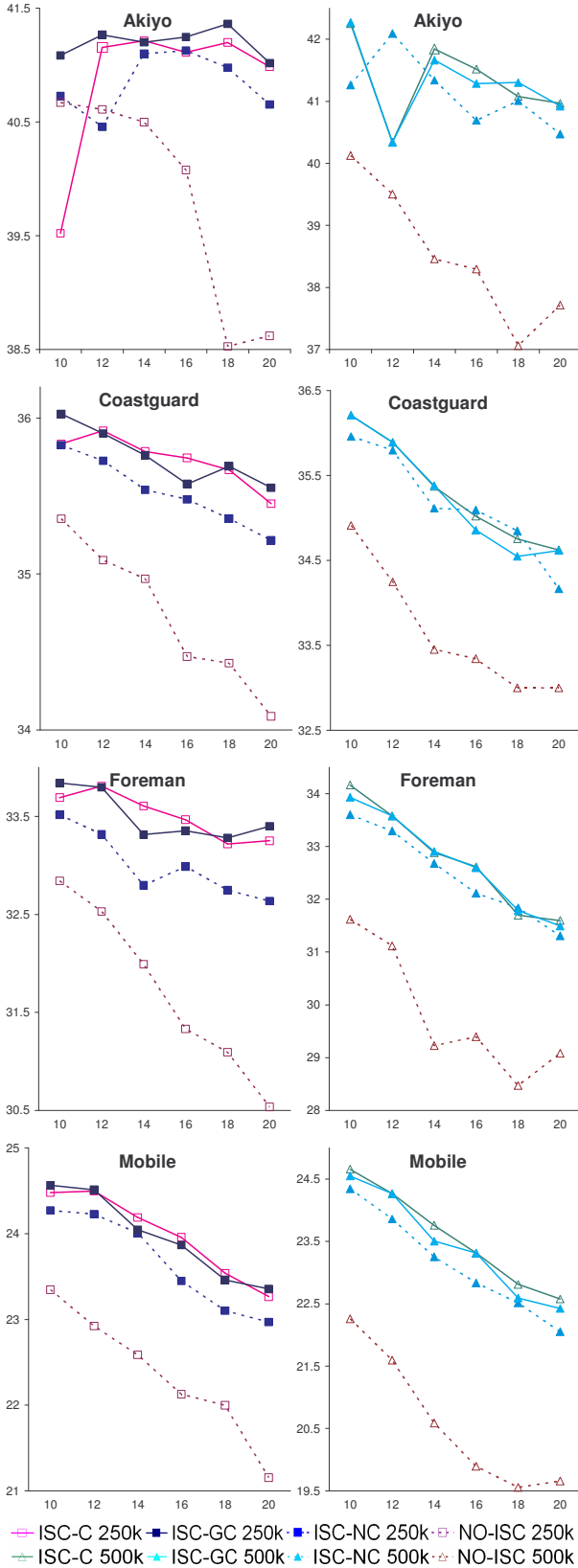


Fig 6. Average PSNR (GOV Size vs. PSNR(dB))

TABLE 6. PSNR DIFFERENCES (dB) @ 500kbps – 250kbps

		10	12	14	16	18	20
Akiyo	ISC-C	2.7269	-0.8161	0.6278	0.4076	-0.1192	-0.0185
	ISC-GC	1.1866	-0.9270	0.4635	0.0376	-0.0599	-0.0950
	ISC-NC	0.5363	1.6376	0.2389	-0.4356	0.0310	-0.1814
	NO-ISC	-0.5449	-1.1055	-2.0402	-1.7833	-1.4637	-0.9060
Coastguard	ISC-C	0.3728	-0.0300	-0.4178	-0.7267	-0.9147	-0.8284
	ISC-GC	0.1818	-0.0129	-0.3814	-0.7227	-1.1446	-0.9390
	ISC-NC	0.1311	0.0701	-0.4260	-0.3865	-0.5083	-1.0518
	NO-ISC	-0.4437	-0.8419	-1.5190	-1.1299	-1.4293	-1.0905
Foreman	ISC-C	0.4749	-0.2295	-0.7225	-0.8510	-1.5208	-1.6541
	ISC-GC	0.0937	-0.2174	-0.4077	-0.7553	-1.4945	-1.9041
	ISC-NC	0.0852	-0.0239	-0.1164	-0.8747	-0.9077	-1.3292
	NO-ISC	-1.2210	-1.4121	-2.7649	-1.9366	-2.6234	-1.4544
Mobile	ISC-C	0.1777	-0.2298	-0.4303	-0.6446	-0.7273	-0.6882
	ISC-GC	-0.0128	-0.2458	-0.5420	-0.5544	-0.8675	-0.9304
	ISC-NC	0.0752	-0.3674	-0.7562	-0.6121	-0.5850	-0.9178
	NO-ISC	-1.0877	-1.3254	-1.9982	-2.2344	-2.4418	-1.4970

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. This 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 , Table 6 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 η , the average number of packets per frame, as stated previously.

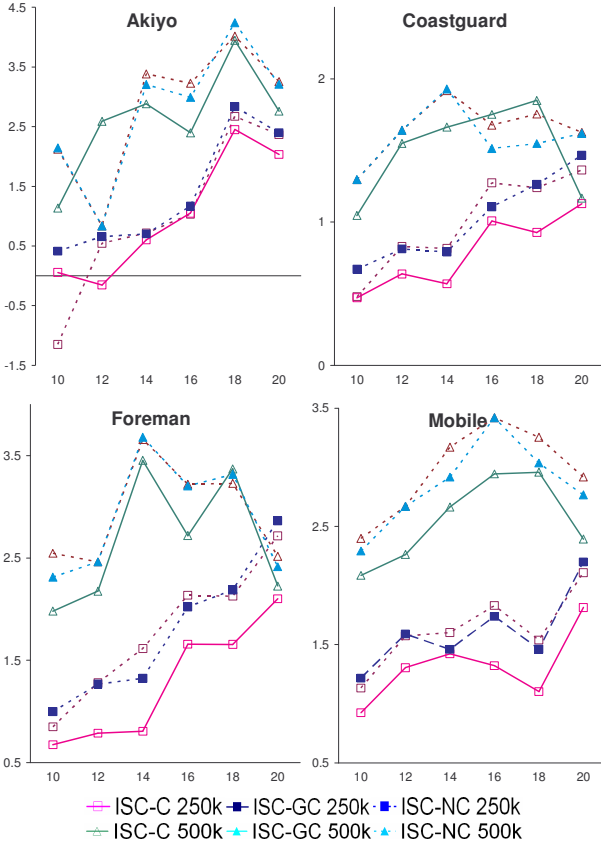


Fig 7. Average PSNR Gain over Non-ISC methods. (GOV Size vs. PSNR(dB))

In addition, as shown in Fig 7, 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.

2) Correlation Gain Improvements

The correlation-based models, both ISC-C and ISC-GC, provide improvements over the non-correlation (ISC-NC) based scenario. In Fig 7, the latter sets show improvements in PSNR gain for most of the evaluation cases, 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.

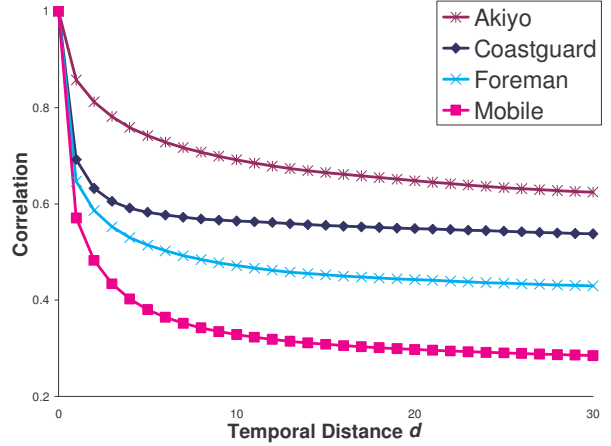


Fig 8. Temporal Correlation of the Evaluation Sequences

3) Evaluation Summary

Overall observation shows that the proposed ISC method improves over the traditional approach on most of the cases, especially for the sequences with high motion or low temporal correlation (Fig 8). Up to 4 dB in average PSNR improvements is observed.

This represents a very significant improvement in quality for compressed video applications. 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, changes in bitrate or GOV size have less impact on ISC coded sequences. Furthermore, when the non-correlation gain computed ISC (ISC-NC) sets are compared to the correlation computed sets (ISC-C and ISC-GC), 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 optimum interleaving set. Once the optimum interleaving is identified for a given channel model, this interleaving can be applied to any video sequence (i.e., without taking into consideration the particular statistical properties of the video sequence).

IV. CONCLUSION

In this paper, we proposed an *interleaved source coding* (ISC) method of predictive coded video sequence for Internet streaming applications. When the coded frames are transmitted over the Internet, this new method provides clear resilience against packet losses when compared with the traditional (without interleaving) approach. This advantage is achieved since ISC limits the errors from packet losses to one of the two sub-sequences (generated by ISC) and minimizes the 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]-[5] [7]-[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] Apostolopoulos, J. G. and Wee, S. J., "Unbalanced Multiple Description Video Communication Using Path Diversity," *IEEE Proc. ICIP*, October 2001.
- [3] Barrenchea, 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.
- [4] Begen, A., Altunbasak, and Y., Ergun, O., "Multi-path selection for multiple description encoded video streaming," *IEEE Proc. ICC*, May 2003.
- [5] Franchi, N., Fumagalli, M., Lancini, R., and Tubaro, S., "Multiple description video coding for scalable and robust transmission over IP," *Packet Video*, April 2003.
- [6] Gallager, R., *Discrete Stochastic Processes*, Kluwer Academic Publishers, 1996.
- [7] 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.
- [8] 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
- [9] Lin, S. and Wang, Y., "Error resilience property of multihypothesis motion-compensated prediction," *IEEE Proc. ICIP*, Rochester, New York, 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 Scharr, 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*, London, 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*, 19