IEEE Fifth Workshop on Multimedia Signal Processing,
MMSP, St. Thomas, Virgin Island, Dec. 2002

Channel-Adaptive Video Streaming Using Packet Path Diversity and Rate-Distortion Optimized Reference Picture Selection

Yi J. Liang, Eric Setton and Bernd Girod
Information Systems Laboratory, Department of Electrical Engineering
Stanford University, Stanford, CA 94305-9510, USA
{yiliang, esetton, bgirod}@stanford.edu

Abstract— **In this paper, we present error-resilient Internet video transmission using path diversity and rate-distortion optimized reference picture selection. Under this scheme, the optimal packet dependency is determined adapting to channel characteristics and video content, to achieve a better trade-off between coding efficiency and forming independent streams to increase error-resilience. Packets are sent over the selected path that minimizes the distortion, while taking advantage of path diversity. Experiments demonstrate that the proposed scheme provides significant gains over video redundancy coding and the NACK mode of conventional reference picture selection.**

I. Introduction

Internet video streaming today is plagued by variability in throughput, packet loss, and delay due to network congestion and the heterogeneous infrastructure. Recently, packet path diversity has been proposed to increase the robustness of multimedia communication over best-effort networks. Using multiple description (MD) coding, the source signal is coded into separate streams, e.g., even and odd video frames, and sent over multiple network paths. The source signal will be reconstructed in full quality if all description streams are received. If at least one description is received, the source signal can still be reconstructed, though possibly at a lower quality. To maximize the benefits of path diversity, multiple streams are sent over independent or largely uncorrelated network paths with diversified loss and delay characteristics [1], [2]. In this way, the probability of a negative disturbance, such as packet loss, impacting all channels at the same time will be small. Multi-path transmission also mitigates the problem that the default path determined by the routing algorithm is not optimum, which might often be the case according to [3].

Path diversity can be implemented by means of an overlay network that consists of relay nodes [1], [2], where packets are sent along different routes as being encapsulated into IP packets that have the addresses of different relay nodes as their destination. At the relay nodes, packets are forwarded to other relay nodes or their destinations. In this way, the packets from different description streams travel along as few common links as possible. In the context of a peer-to-peer framework, every peer could serve as a relay node for media traffic, potentially leading to a number of different paths a stream could take from its source to its destination. Path diversity can also be achieved by content delivery networks (CDN) [4], or source routing [1], [2].

One of the previous approaches of multi-stream coding is video redundancy coding (VRC), where the video sequence is coded into independent threads (streams) in a round-robin fashion [5]. A Sync frame is encoded by all threads at regular intervals to start a new thread series and stop error propagation. If one threads is damaged due to packet loss, the remaining threads can still be used to predict the Sync frame. Another approach is the multiple state coding proposed in [1], in which even and odd frames are coded into independent streams respectively and sent over two paths. With VRC or multiple state coding, independent streams are formed to provide high resilience against non-simultaneous channel errors, but with the penalty of lower coding efficiency due to the wider separation of the frames used for prediction.

A different scheme proposed in [6] for ad-hoc networks uses reference picture selection (RPS) to terminate error propagation based on feedback. With RPS (proposed in Annex N of H.263+ [7]), when the encoder detects that a previous frame is lost, instead of using the most recent frame as a reference, it can code the next P-frame based on an older frame that is known to be correctly received [8]. The scheme in [6] employs the RPS NACK-mode [8] by always choosing "the last frame that is believed to be transmitted reliably as the reference frame." When the transmission channels are in good state, prediction is made using the most recent frame as a reference. Although the coding efficiency is higher than VRC, error-resilience is limited since the coded streams are not independent. Due to the feedback delay, the NACK might be too late to induce a reference selection to stop the error in time. This scheme has not fully taken advantage of path diversity, and the performance heavily depends on the feedback delay and channel loss rate.

In this work we use rate-distortion (RD) optimized RPS (ORPS) and packet path diversity to increase the robustness of video transmission. Different from the schemes discussed above, the proposed scheme is channel-adaptive. With an RD framework, we are able to better trade off coding efficiency and forming independent streams to increase error-resilience. With the increased robustness against channel error, the need for packet retransmission is eliminated and the streaming latency can be reduced to below one second.

This paper is structured as follows: we first describe the management of packet dependency over multiple paths. In Section III, we present the selection of the network path to send a packet. Experimental results are presented in Section IV.

This work has been supported by a gift from HP Labs, Palo Alto, CA.

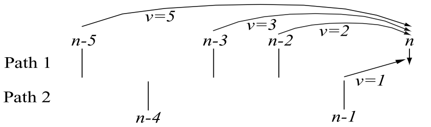

Fig. 1. An example of reference selection over two transmission channels.

II. Packet Dependency Control over Multiple Paths

Assuming the typical scenario where an IP packet contains one video frame, packet dependency can be managed through the selection of the reference frame (or the use of INTRA coding) for the next frame to encode. A conventional coding scheme of predicting P-frames from their immediate predecessors is vulnerable to channel errors, since any packet loss will break the prediction chain and affect all subsequent P-frames. If a frame is predicted from an older frame that is estimated to be more reliable, e.g., v frames back ($v > 1$), the coded frame is more robust against channel errors due to the changed dependency. This is normally obtained at the expense of a higher bit rate since the correlation between two frames becomes weaker in general as they are more widely separated.

We minimize the distortion of the frame to encode by determining the optimal prediction dependency and the path to send the frame. Under this greedy algorithm, reference selection and path selection can be performed independently. We discuss reference selection in this section and path selection in Section III.

A. *Rate-distortion optimized reference picture selection*

Due to the trade-off between error-resilience and coding efficiency, we select the reference picture within an RD framework.

While coding a frame n, assuming V previously decoded frames are available from the long-term memory (V is referred to as the *length of LTM*), we use $v(n)$ to represent the reference frame that Frame n may use and $v(n)$ indicates the prediction dependency. For example $v(n) = 1$ denotes using the previous frame and $v(n) = 2$ denotes using the frame preceding that frame, and so on. For a particular $v(n) = v$, a rate R_v is obtained from encoding and the expected distortion of all decoded outcomes $\overline{D}_v$ is obtained from a binary tree modeling to be described next. With the obtained R_v and $\overline{D}_v$, the Lagrangian cost corresponding to using the reference frame $v(n) = v$ is

$$J_v = \overline{D}_v + \lambda R_v. \tag{1}$$

where λ is a Lagrange multiplier. We use $\lambda = 5e^{0.1Q}(\frac{5+Q}{34-Q})$, which is the same as λ_{mode} in H.26L TML 8 used to select the optimal prediction mode [9]. Q is the quantization parameter used to trade off rate and distortion.

In the case of a single path transmission as described in our previous work [10], to encode a frame n, several trials are made, including using the I-frame as well as INTER coded frames using different reference frames taken from the long-term memory, e.g., $v(n) =$ 1, 2, 3, ... V and ∞ (to denote INTRA coding). The optimal reference frame $v_{opt}(n)$ is selected such that the minimal RD cost J_v is achieved.

In the case of multiple paths, we have to consider not only the RD cost, but also the formation of independent streams to increase error-resilience. Denoting the path Frame n is sent over by $C(n)$ (see Section III), trials are made using $v(n) \in \mathcal{V}$, where the set of candidate references is further restricted by

$$\mathcal{V} = \{v = 1,\ \infty\} \cup \{v = 2, 3, ...V | C(n - v) = C(n)\}\,. \tag{2}$$

In (2) $v = 1$ is the most thrifty in bit rate and $v = \infty$ provides the highest robustness; while for all other candidates we impose the restriction that only frames sent over the same channel as $C(n)$ will be considered as a reference, which keeps the frame to code independent of other streams. In the two-path example in Fig. 1, where the LTM size $V = 5$, if Frame n is to be sent over Path 1, $\mathcal{V} = \{1, 2, 3, 5, \infty\}$; otherwise, $\mathcal{V} = \{1, 4, \infty\}$.

The optimal reference frame $v_{opt}(n)$ for encoding Frame n is the one that results in minimal J_v

$$v_{opt}(n) = \arg\min\nolimits_{v \in \mathcal{V}} J_v(n).$$

The optimal selection is determined within an RD framework, considering video content, channel loss probability and channel feedback (e.g., ACK, NACK, or time-out). For example, if Frame $n - 1$ is estimated to be very reliable, or, in case of loss, if it can still be concealed very well due to the low motion in the video content, it is more likely $v(n) = 1$ will be used to save bits, even if the independence between streams may be broken. Compared to the RPS-NACK scheme proposed in [6], our proposed scheme is able to take more advantage of path diversity by maintaining independent threads when higher error-resilience is desired. Compared to VRC [5] and multiple state encoding [1], the proposed scheme is more RD efficient since the reference selection is adaptive and $v = 1$ is allowed.

In (1), the expected distortion $\overline{D}_v$ is estimated using a binary tree modeling that describes the prediction dependency between frames, as illustrated in Fig. 2. A *node* in the tree represents a possible decoded outcome (frame) at the decoder. In the example shown in Fig. 2, Frame $n - 3$ has only one node with probability 1 (e.g., due to the reception status confirmed by feedback). Frames $n - 2$ and $n - 1$ both, for instance, use their immediately preceding frames as references. Two *branches* leave the node of Frame $n - 3$ representing two cases that either reference frame $n - 3$ is properly received (and decoded) with probability $1 - p_B^{(n-3)}$ or lost with probability $p_B^{(n-3)}$, where $p_B^{(i)}$ is the loss probability of a corresponding node of frame i, which is estimated using the channel model discussed in the next subsection. These two cases lead to two different decoded outcomes of Frame $n-2$, provided that Frame $n-2$ is available at the decoder. The upper node of Frame $n - 2$ is obtained by normal decoding process using the correct reference (decoded $n - 3$); and the lower node corresponds to the case when Frame $n - 3$ is lost. In the latter case, a simple concealment is done

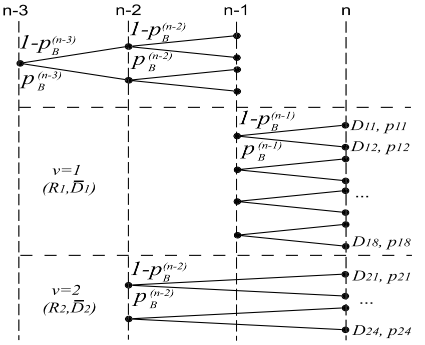

Fig. 2. The binary tree structure for the estimate of error propagation and optimal reference selection.

by copying $n-4$ to $n-3$, and Frame $n-2$ hence has to be decoded using the concealed reference. This leads to the mismatch error that might propagate at the decoder, depending on the prediction dependency of the following frames. The distortion associated with these two cases is evaluated by decoding $n-2$ at the *encoder* side.

In encoding Frame n, the expected distortion of all decoded outcomes for a particular trial v is

$$\overline{D}_v = \sum_{l=1}^{L(n)} p_{vl} D_{vl}, \qquad (3)$$

where $L(n)$ is the number of nodes for Frame n, and p_{vl} is the probability of outcome (node) l, which can be calculated from the model in Fig. 2. For example, $p_{11} = (1-p_B^{(n-3)})(1-p_B^{(n-2)})(1-p_B^{(n-1)})$, while $p_{12} = (1-p_B^{(n-3)})(1-p_B^{(n-2)})p_B^{(n-1)}$ and so on. D_{vl} is the distortion associated with the decoded outcome l. Note that D_{vl} includes both the quantization error and possible decoding mismatch error, which is calculated accurately at the encoder.

B. Channel Model

We use the two-state Gilbert model to approximate the bursty behavior of each channel. The two states are state G (good), where the packets are received correctly and timely, and state B (bad), where the packets are lost, either due to network congestion or late arrival of packets. The model is fully determined by the transition probabilities p_{GB} from state G to B, and p_{BG} from state B to G. These model parameters in practice are estimated from the accumulated channel statistics, i.e., the measurable average loss probability $\overline{P}_B = p_{GB}/(p_{GB}+p_{BG})$, and the average burst loss length $\overline{L}_B = 1/p_{BG}$. These parameters are updated as the channel conditions vary, and could be different for each channel.

If Frame i is predicted from $i-1$ and sent over the same path as $i-1$, its loss probability $p_B^{(i)}$ is conditioned on the receipt status of the corresponding nodes of Frame $i-1$ in the tree model in Fig. 2:

$$p_B^{(i)} = (1-I_B^{(i-1)})p_{GB} + I_B^{(i-1)}(1-p_{BG}), \qquad (4)$$

where $I_B^{(i-1)} = 0$, if Frame $i-1$ is received (upper node) and $I_B^{(i-1)} = 1$ if $i-1$ is lost (lower node). If Frame i is predicted from $i-k$ ($k \geq 1$) and sent over the same path as $i-k$, its loss probability is

$$p_B^{(i)} = (I_B^{(i-k)} - \overline{P}_B)(1-p_{GB}-p_{BG})^k + \overline{P}_B. \qquad (5)$$

In the case that i is predicted from a frame sent over a different path, index $i-k$ in (5) denotes the most recent frame that is sent over the same path as Frame i, and whose receipt status is confirmed by feedback. The loss probability obtained from (5) is used in the tree model in Fig. 2.

III. Path Selection

To select the path over which to send the next packet, we consider minimizing the distortion of the next frame and taking advantage of path diversity: the next packet is always sent over the path from which the most recent ACK is received, unless all paths are in bad state. In the example of two channels, packets are distributed evenly over each channel in an alternating way if both of them are in good state. If one channel turns into bad state, i.e., a NACK is received via feedback, or the packet times out, packets are sent over the other channel if it is believed in good state, until the bad channel is known to have returned to good state. When a channel experiences burst losses, it is possible that no packets will ever be sent over that channel due to channel inactivity and the lack of any ACKs sent. To avoid using only one channel, we start sending small probe packets over the channel once we learn that it enters the bad state. This allows us to resume using that channel as soon as an ACK is received. If all the paths are in bad state, we simply send packets in a round-robin fashion.

The proposed path selection scheme is different from what is used in [6], where packets are always delivered over the paths alternately. Our proposed scheme prohibits the use of a bad channel that experiences burst losses when other channels are good, which decreases the overall packet loss probability. The gain is even higher with unbalanced channels, e.g., channels with different loss probabilities. Packets are distributed properly according to the ACKs received from respective channels with different characteristics. However, the efficiency of this feedback-based path selection depends on the feedback delay.

IV. Simulation Results

We compare the performance of three schemes in transmitting video over two network paths: 1) the proposed ORPS scheme; 2) RPS-NACK scheme in [6]; 3) VRC in 2-13 mode [5]. We have implemented the three schemes by modifying the H.26L TML 8.5. The testing video sequences are *Foreman* and *Mother-Daughter*, representing high and moderate motion, respectively. 230 frames are coded, and the frame rate is 30 fps. Coded frames are dropped according to Gilbert model-simulated channel conditions with a range of loss probabilities.

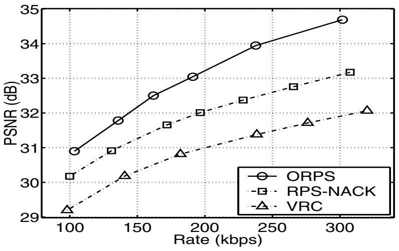

Fig. 3. RD performance of *Foreman* sequence. $P_B = 0.15$, $L_B = 3$.

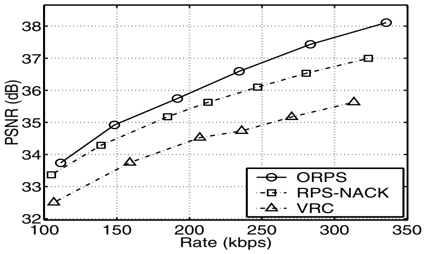

Fig. 4. RD performance of *Mother-Daughter*. $P_B = 0.15$, $L_B = 3$.

The PSNR of the decoded sequences is averaged over 30 random channel loss patterns, and the rates of probe packets are not counted. The first 30 frames of a sequence are not included in the statistics to exclude the influence of the transient period.

Fig. 3 shows the RD performance of sending the *Foreman* sequence over the channel with an average loss rate of 15%. Feedback delay is 8 frames, and the length of LTM is $V = 12$. The distortion at different rates is obtained by varying the Q value and hence the Lagrange multiplier λ. Comparing Schemes 1 and 2, a gain of 1.2 dB is observed at 200 Kbps and 1.5 dB at 300 Kbps by using the proposed scheme, which corresponds to a bit rate saving of 35% at 33 dB. The gain is typically higher at higher rates since at lower rates LTM prediction with $v > 1$ is less efficient and the advantage of ORPS decreases. Note that although no retransmission is used, the video quality is still good over the lossy channel. Fig. 4 shows the RD performance of *Mother-Daughter* under the same experimental conditions. A gain of 0.4 dB is observed at 200 Kbps and 1.0 dB at 300 Kbps. The gain of the proposed scheme is lower compared to *Foreman* since the effect of packet loss is smaller due to lower motion in the sequence.

Distortion at different channel loss rates is shown in Fig. 5 for *Foreman* encoded at approximately the same 200 Kbps using the three schemes. The gain is observed ranging from 0.6 dB to 1.3 dB, depending on the channel loss rate. The advantage of using error-resilient ORPS is more obvious at higher loss rate, while RPS-NACK is known efficient at low loss rate. Performance over unbalanced channels of 10% and 20% loss respectively, is shown in Fig. 6. The gain of Scheme 1 over 2 is even higher than that in the case of balanced channels of 15% loss, which is due to the adaptive reference picture selection and path selection used in the proposed scheme.

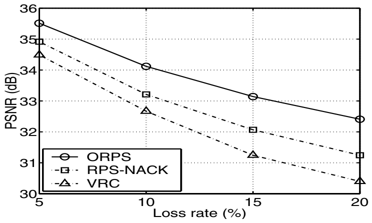

Fig. 5. Distortion at different channel loss rates: $P_B = 0.05$, $L_B = 2$; $P_B = 0.10$, $L_B = 3$; $P_B = 0.15$, $L_B = 3$; $P_B = 0.20$, $L_B = 4$. *Foreman*.

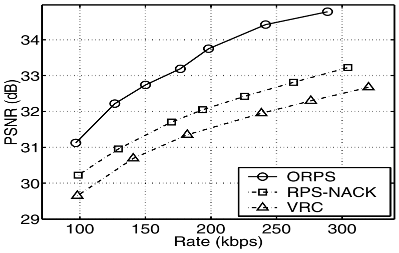

Fig. 6. Performance over unbalanced paths. $P_{B1} = 0.10$, $L_{B1} = 3$; $P_{B2} = 0.20$, $L_{B2} = 4$. *Foreman*.

V. Conclusions

We propose an adaptive video transmission scheme using path diversity and rate-distortion optimized reference picture selection, to achieve a better trade-off between coding efficiency and error-resilience. Experiments demonstrate significant gains over schemes including VRC and RPS NACK.

References

[1] John G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *Proc. Visual Commun. and Image Processing*, Jan. 2001, pp. 392–409.

[2] Yi J. Liang, Eckehard G. Steinbach, and Bernd Girod, "Real-time voice communication over the Internet using packet path diversity," in *Proceedings ACM Multimedia 2001*, Oct. 2001, pp. 431–440, Ottawa, Canada.

[3] S. Savage, A. Collins, E. Hoffman, J. Snell, and T. Anderson, "The end-to-end effects of Internet path selection," *Computer Communication Review, ACM SIGCOMM '99*, vol. 29, no. 4, pp. 289–99, Oct. 1999.

[4] J.G. Apostolopoulos, T. Wong, W. Tan, and S.J. Wee, "On multiple description streaming with content delivery networks," in *Proceedings IEEE INFOCOM*, June 2002.

[5] S. Wenger, G.D. Knorr, J. Ott, and F. Kossentini, "Error resilience support in H.263+," *IEEE Journal on Circuits and Systems for Video Technology*, vol. 8, no. 7, pp. 867–877, Nov. 1998.

[6] S. Lin, S. Mao, Y. Wang, and S. Panwar, "A reference picture selection scheme for video transmission over ad-hoc networks using multiple paths," in *Proc. IEEE Int. Conf. on Multimedia and Expo*, Aug. 2001.

[7] ITU-T Recommendation H.263 Version 2 (H.263+), *Video coding for low bitrate communication*, Jan. 1998.

[8] B. Girod and N. Färber, "Feedback-based error control for mobile video transmission," *Proc. IEEE*, vol. 87, no. 10, pp. 1707 – 1723, Oct. 1999.

[9] ITU-T Video Coding Expert Group, *H.26L Test Model Long Term Number 8*, July 2001, online available at: ftp://standard.pictel.com/video-site/h26L/tml8.doc.

[10] Yi J. Liang, M. Flierl, and B. Girod, "Low-latency video transmission over lossy packet networks using rate-distortion optimized reference picture selection," in *Proc. of the IEEE International Conference on Image Processing (ICIP-2002)*, Sept. 2002, Rochester, NY.