Journal of VLSI Signal Processing 41, 345–354, 2005

DOI: 10.1007/s11265-005-4157-x

Network-Adaptive Video Communication 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, USA

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 network characteristics and video content, to achieve a better trade-off between coding efficiency and forming independent streams to increase error-resilience. The optimization is achieved within a rate-distortion framework, so that the expected end-to-end distortion is minimized under the given rate constraint. The expected distortion is calculated based on an accurate binary tree modeling with the effects of channel loss and error concealment taken into account. With the aid of active probing, packets are sent across multiple available paths according to a transmission policy which takes advantage of path diversity and seeks to minimize the loss rate. Experiments demonstrate that the proposed scheme provides significant diversity gain, as well as gains over video redundancy coding and the NACK mode of conventional reference picture selection.

Keywords: Path diversity, network-adaptive video coding, error-resilient video coding, source-channel coding, reference picture selection, low-latency, rate-distortion optimization, video streaming

1. 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 diversity in media communication, multiple streams can be sent, in a distributed manner, over independent or largely uncorrelated network paths with diversified loss and delay characteristics [1–9]. In this way, the probability of a negative disturbance, such as packet loss, impacting all channels at the same time will be small. Path diversity also alleviates the problem that the default path determined by the routing algorithm is not optimum, which might often be the case according to [10]. Recently, path diversity is also used with optimized scheduling of packet transmission to achieve enhanced performance [11, 12].

In order to maximize the benefits of path diversity we select transmission paths that exhibit largely uncorrelated jitter and loss characteristics. Sending streams along different routes from source to destination naturally leads to path diversity which could include streams traversing different ISPs or even streams being sent in different directions around the globe. With today's Internet protocols, the path a packet takes across

the Internet is a function of its source and destination IP addresses as well as the entries of the routing tables involved. Selecting a specific path for a packet is largely unsupported in today's infrastructure. As discussed in [1], IPv4 source routing is usually turned off within the Internet for security reasons. More promising is to implement path diversity by means of an overlay network that consists of relay nodes [1, 2, 13], 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 such that the packets from different description streams travel along as few common links as possible. In the context of peer-to-peer networking [4, 15], 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) [16–18]. With the next-generation IP protocol IPv6, the source node has a larger amount of control over each packet's route. IPv6's loose source routing (LSR) allows packets to be sent via specified intermediate nodes. This source routing feature of IPv6 will provide more flexibility for future implementation of path diversity.

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 [19]. A Sync frame is encoded by all threads at regular intervals to start a new thread series and stop error propagation. If one thread 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 [20] uses reference picture selection (RPS) to terminate error propagation based on feedback. With RPS (proposed in Annex N of H.263+ [21]), 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 [22]. The multiframe prediction support in Annex N was later subsumed by the more advanced Annex U of H.263++ and is now an integral part of the new H.264 standard [23]. The scheme in [20] employs the RPS NACK-mode [22] by always choosing "the last frame that is believed to be transmitted reliably as the reference frame." When 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 propagation in time. This scheme has not fully taken advantage of path diversity, and the performance largely depends on the feedback delay and channel loss rate. In our earlier work [24], in the scenario of only one transmission path, we extend the RPS concept by allowing the use of a reference frame whose reception status is uncertain but whose reliability can be inferred, for live-encoding.

Most of the past work on path diversity has focused on increasing the communication robustness over error-prone networks. In this work we take advantage of path diversity not only to improve the quality of media communication by reducing the effective packet loss rate, but also to reduce the latency for applications with very stringent delay requirement. We use rate-distortion (R-D) optimized RPS (ORPS) and packet path diversity to increase the robustness of video transmission. Different from the schemes discussed above, the proposed scheme is network-adaptive. Within an R-D optimization 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 less than one second.

This paper is structured as follows: we first introduce the concept of packet dependency management and its implementation in Section 2. Then we describe network-adaptive packet dependency management over multiple paths in Section 3. In Section 4, we describe the selection of the network path for packet transmission. Experimental results are presented in Section 5.

2. Packet Dependency Management and Reference Picture Selection

In [25, 26], long-term memory (LTM) prediction is used for both improved coding efficiency and

error resilience over wireless networks. Different macroblocks in a frame may be predicted from different reference frames, which makes it difficult to put an entire frame into an IP packet and manage the prediction dependency at the packet level during transmission. Throughout this work, we select the reference at the frame level and assume that each predictively coded frame is coded into one IP packet (the proposed scheme can also be extended to the case where a frame is coded into multiple packets). In this way we manage the frame prediction dependency at the packet level.

In a conventional encoding and transmission scheme without any awareness of network losses, an I-frame is typically followed by a series of *P*-frames, which are predicted from their immediate predecessors. This scheme is vulnerable to network errors since each *P*-frame depends on its predecessor and any packet loss will break the prediction chain and affect all subsequent *P*-frames. If each *P*-frame is predicted from the frame preceding the previous frame instead, the scheme is more robust against network errors due to the changed dependency and the higher certainty of the reference frame. Consider, for example, a fixed coding structure where each frame uses the reference that is v frames back for prediction, where v is used to denote the *coding mode*, or *prediction mode*. The n-th frame in the sequence thus depends on $\lceil \frac{n}{v} \rceil$ previous frames, where $\lceil x \rceil$ represents the smallest integer number that is greater than or equal to x. An example for $v = 3$ is illustrated in Fig. 1. Assuming each packet is lost independently with probability p, the probability that the n-th frame in the sequence will be affected by a previous loss is hence

$$p_e = 1 - (1 - p)^{\lceil \frac{n}{v} \rceil}. \quad (1)$$

This probability is plotted in Fig. 2 for $p = 0.10$, $n = 10$, and $v = 1, 2, \ldots 5$, and INTRA coding (we use $v = \infty$ to denote INTRA coding).

As illustrated in Fig. 2, using frames from the long-term memory with $v > 1$ for prediction, instead of using

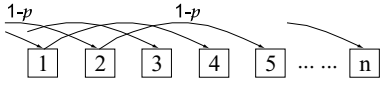

Figure 1. A coding structure where each frame uses the third previous frame as a reference ($v = 3$). Each frame is correctly received at the decoder with probability 1-p. Frame 5 in the sequence depends on $\lceil \frac{5}{3} \rceil = 2$ previous frames, and the probability it will be affected by a previous loss is $p_e = 1 - (1 - p)^2$.

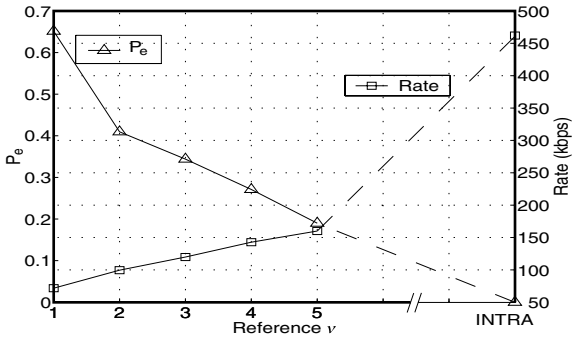

Figure 2. The probability of the 10th frame being affected by a prior loss (left axis) and the sequence-averaged rates (right axis) using different reference frames. Rates are obained by encoding the first 230 frames of *Foreman* sequence (30 frame/sec) using H.264 TML 8.5 at an average PSNR of approximately 33.4 dB. $p = 0.10$.

an immediately previous frame ($v = 1$), reduces prediction efficiency and increases error resilience. The robustness is normally obtained at the expense of a higher bitrate since the correlation between two frames becomes weaker in general as they are more widely separated. A special and extreme case is the I-frame, which is the most robust over lossy networks, but generally requires 5-10 times as many bits as the *P*-frame. In Fig. 2, we also show the average rates of encoding the *Foreman* sequence at close PSNRs using different coding modes v, including INTRA coding.

Fixed reference selection schemes provide different amount of error resilience at different coding costs, as is shown in Fig. 2. In this work, we consider the dependency across packets and dynamically manage this dependency while adapting to the varying network conditions. Due to the trade-off between error resilience and coding efficiency, we apply *Optimized Reference Picture Selection (ORPS)* within an R-D optimization framework, by considering video content, network loss probability and channel feedback (e.g., ACK, NACK, or time-out). The proposed scheme is compatible with the new ITU-T standard H.264 [23].

3. Network-Adaptive Packet Dependency Management 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. We minimize the distortion of the frame to encode by determining the optimal prediction dependency as well

as the path to send the frame. Under this greedy algorithm, path selection and reference selection can be performed sequentially. We discuss reference selection in this section and path selection in Section 4.

3.1. Rate-Distortion Optimized Reference Picture Selection

Due to the trade-off between error-resilience and coding efficiency, we select the reference picture within an R-D optimization 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

$$J_v = \overline{D}_v + \lambda R_v. \tag{2}$$

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.264 TML 8 used to select the optimal prediction mode [27]. Q is the quantization parameter set to trade off rate and distortion.

In the case of a single path transmission as described in our previous work [24], 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., $v(n) = 1, 2, 3, \ldots V$ and ∞ (to denote INTRA coding). The optimal reference frame $v_{\text{opt}}(n)$ is selected such that the minimal R-D cost J_v is achieved.

In the case of multiple paths, we have to consider not only the R-D cost, but also the formation of independent streams to increase error-resilience. Denoting the path Frame n is sent over by $C(n)$ (determined by the scheme described in Section 4), trials are made using $v(n) \in \mathcal{V}$, where the set of candidate references is further restricted by

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

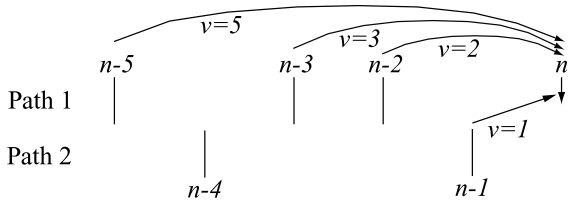

Figure 3. An example of reference selection over two transmission channels.

In (3) $v = 1$ is the most thrifty in bitrate usage and $v = \infty$ provides the highest robustness; while for all other coding modes we impose the restriction that only frames sent over the same channel as $C(n)$ will be considered as candidate references, which keeps the frame to code independent of other streams. In the two-path example in Fig. 3, 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 $v_{\text{opt}}(n)$ for encoding Frame n is the one that results in minimal J_v

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

The optimal selection is determined within an R-D optimization framework, considering video content, network 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 VRC [19] and multiple state encoding [1], the proposed scheme is more R-D efficient since the reference selection is adaptive and $v = 1$ is allowed. Compared to the RPS-NACK scheme proposed in [20], our proposed scheme is able to take more advantage of path diversity by maintaining independent threads when higher error-resilience is desired.

In (2), the expected distortion $\overline{D}_v$ is estimated using a binary tree modeling that describes the prediction dependency between frames, as illustrated in Fig. 4. A *node* in the tree represents a possible decoded outcome (frame) at the decoder. In the example shown in Fig. 4, 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

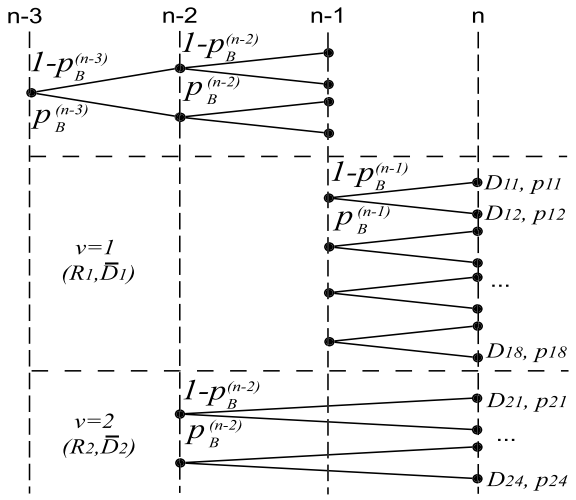

Figure 4. The binary tree structure for the estimate of error propagation and optimal reference selection. $v = 1$ represents using frame $n - 1$ as the reference for prediction. $v = 2$ represents using frame $n - 2$ as the reference for prediction.

the following two cases: 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 network loss 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 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

$$\overline{D}_v = \sum_{l=1}^{L(n)} p_{vl} D_{vl}, \tag{4}$$

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. 4. For example,

$$p_{11} = \left(1 - p_B^{(n-3)}\right)\left(1 - p_B^{(n-2)}\right)\left(1 - p_B^{(n-1)}\right),$$

while

$$p_{12} = \left(1 - p_B^{(n-3)}\right)\left(1 - p_B^{(n-2)}\right)p_B^{(n-1)},$$

and so on. The P_B's for different frames may be obtained from the characteristics of different transmission paths. 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. The complexity of the formulization (e.g. the size of the binary tree in Fig. 4) depends on the length of the LTM and the channel feedback delay [24].

3.2. *The Network Loss 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 network 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 network conditions vary, and could be different for each channel.

If Frame i is sent over the same path as $i - 1$, its loss probability $p_B^{(i)}$ is conditioned on the reception status of Frame $i - 1$:

$$p_B^{(i)} = \left(1 - I_B^{(i-1)}\right)p_{GB} + I_B^{(i-1)}(1 - p_{BG}), \tag{5}$$

where $I_B^{(i-1)} = 0$, if Frame $i-1$ is received and $I_B^{(i-1)} = 1$ if $i - 1$ is lost. If Frame $i - k$ ($k \geq 1$) is the most recent frame that was sent over the same path as i, the loss probability of Frame i is

$$p_B^{(i)} = \left(I_B^{(i-k)} - \overline{P}_B\right)(1 - p_{GB} - p_{BG})^k + \overline{P}_B. \tag{6}$$

The loss probability obtained from (6) is used in the tree model in Fig. 4.

4. Path Selection

Packets are sent across multiple available paths according to a transmission policy which takes advantage of

path diversity and seeks to minimize the loss rate. This policy adapts to the network conditions by analyzing both the passive feedback originating from media data packet transmission and the active feedback created by probe packets sent over idle channels at a reasonable R-D cost [28].

4.1. Transmission policy

In order to fully benefit from path diversity, the transmission policy should distribute packets optimally across all available paths, in order to minimize the packet loss rate. Transmission should be scheduled according to the state of each channel. In this way, channels with comparable statistics should hold the same rate of packet transmitted, and priority should be given to channels with better condition. In our test model, we assume a simplified scenario in which all the transmission channels follow a Markov chain and the delay over each path is equal and constant. Therefore, the better channel is the one from which the most recent ACK has been received. Thus, the optimal transmission policy, in this simplified case, is to send the next packet over the better path, i.e. the one from which the latest ACK is generated. In this way, if all the channels are in good state, packets are sent alternately across the different paths.

When a channel is not used or experiences burst losses it is possible that no packet will ever be sent over that path due to channel inactivity and the absence of ACKs. To avoid keeping using only one particular channel, we send probe packets over idle channels to induce "active" feedback. This guarantees a minimal flow of information indicating the state of each channel periodically.

In the extreme case when all the paths fail at the same time, we simply send packets in a round robin fashion to detect the next state transition.

The proposed path selection scheme is different from what is used in [20], 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 for 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.

4.2. Influence of the probe

Probe packets are transmitted to induce additional feedback and test the state of the channels, so that the optimized transmission policy can be determined. For the probe, we suggest using an RTP header encapsulated in a UDP-IP header. The routing along different paths is specified in the IP header in the same way as for media data packets. RTP provides packet identification as well as time stamps needed to determine the transmission policy. The minimal size of such a probe is 40 bytes for IPv4. Although it does not compare to the size of a media data packet (ranging from a few hundred to a few thousand bytes for video), the rate associated with probes should be included in the total data rate budget, especially when transmission over a large number of paths is considered.

As the rate of probe packets increases, path selection plays a more active and important role, as the scheme is more responsive to the variation of the network condition. This contributes to reducing the packet loss rate as well as the distortion of the decoded video. In this way, the cost in terms of data rate is traded for an enhanced quality. The influence of the probe can be analyzed in terms of rate and distortion to determine the optimal transmission rate. Figure 5 shows the influence on the video quality when probes are transmitted at a period I (in ms). We simulate sending the first 150 frames of the *Foreman* sequence, encoded at a fixed quantization parameter using the H.264 codec. Here, we limit the reference picture selection to the two previous frames

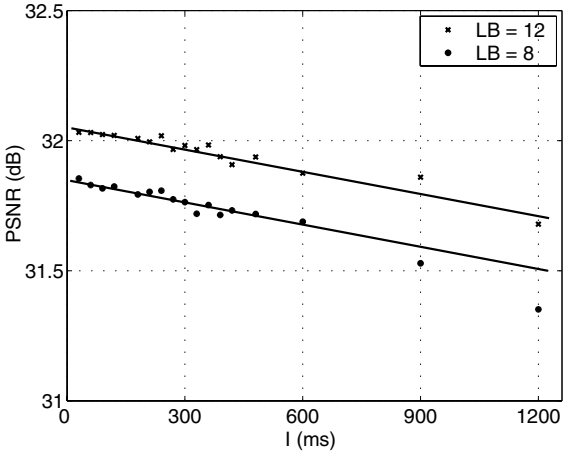

Figure 5. Influence of probing period I on the decoded video quality. *Foreman* sequence transmitted at 30 fps. Three channels are employed; the loss rate on each channel is $\overline{P}_B = 15\%$ and the average burst loss length $\overline{L}_B$ is given in frames.

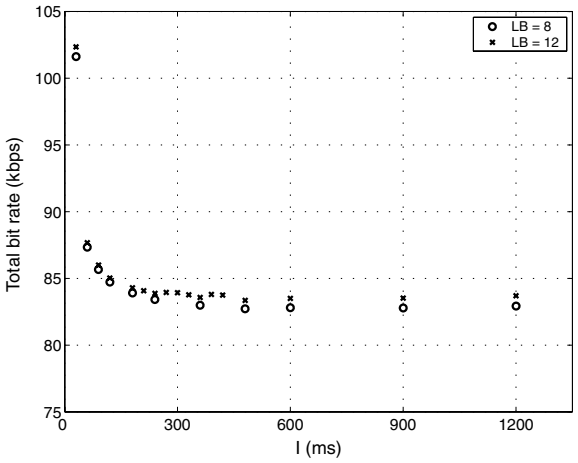

Figure 6. Influence of probing period I on the data rate. *Foreman* sequence transmitted at 30 fps. The average burst loss length $\overline{L}_B$ is given in frames.

in order to focus on the study of probes. As illustrated in Fig. 5, the PSNR of the decoded video is observed to decrease linearly with the probing period I. Fig. 6 shows the influence of I on the total data rate using the same policy in which both media packets and probe packets are counted. The rate of the probes decreases inversely with the probing period, and is less than 4 Kbps when $I \geq 120$ms.

An optimal probing period should minimize the Lagrangian cost $J = D + \lambda R$ similar to (2), except that both probe packets and media data packets are counted in estimating the data rate. In later simulations we use I=120 ms, which is close to optimal across the bitrates for our set of experiments.

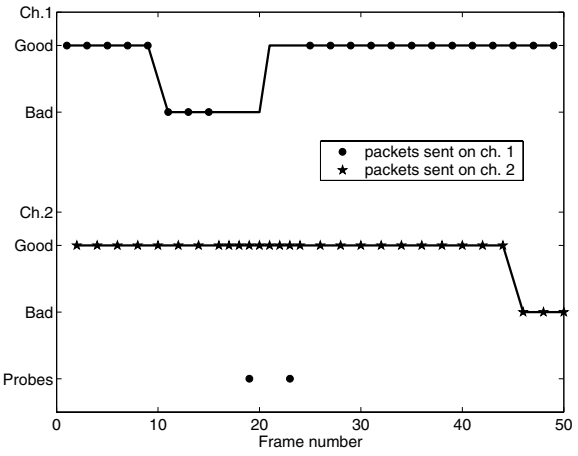

Figure 7. An example of path selection. Two transmission channels.

Figure 7 shows an example of transmitting video over two paths according to the transmission policy. The sequence is streamed at 30 fps and the feedback delay is fixed at 180 ms. During the first half second (up to Frame 17), packets are alternately sent across both channels which are initially in a good state. Due to the feedback delay, Packets 11, 13 and 15 are lost over Channel 1. When these losses are detected, all the packets are directed to Channel 2 and probes are transmitted periodically over the first channel to obtain the state information. As the second probe is acknowledged, alternating transmission resumes over both channels.

5. Simulation Results

We compare the performance of four schemes in transmitting video over two network paths: (1) the proposed ORPS scheme with path diversity; (2) RPS-NACK scheme in [20]; (3) VRC in 2-13 mode [19], where two threads are used and a Sync frame is coded for every 13 frames; (4) the ORPS scheme using only one transmission path [24].

We have implemented the four schemes by modifying the H.264 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 network conditions with a range of loss probabilities. It is assumed that the averaged long-term network characteristics are updated accurately at the sender side, and the instantaneous feedback reaches the sender after a certain delay. The PSNR of the decoded sequences is averaged over 30 random network loss patterns. The first 30 frames of a sequence are not included in the statistics to exclude the influence of the transient period.

Figure 8 shows the R-D performance of sending the *Foreman* sequence over the network with an average loss rate of 15%, and an average burst loss length of 8 frames. Here the losses include packets dropped over the network, as well as late packets that have missed the delivery deadline and cannot be used. 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 $v>1$

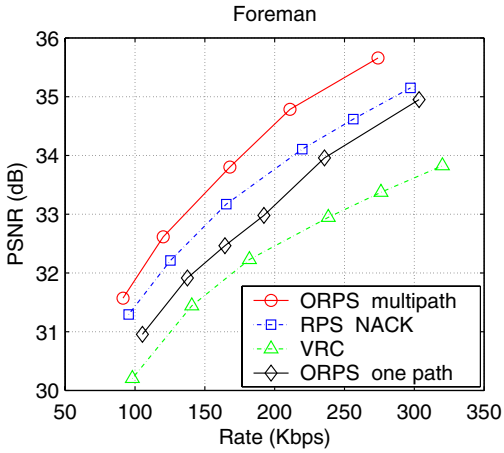

Figure 8. R-D performance of *Foreman* sequence. $\overline{P}_B = 0.15$, $\overline{L}_B = 8$.

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 network. The gain of diversity is also significant by comparing Schemes 1 and 4, when ORPS is applied in both scenarios.

Figure 9 shows the R-D 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. Performance over unbalanced channels of 10% and 20% loss respectively, with average burst loss lengths of 8, is

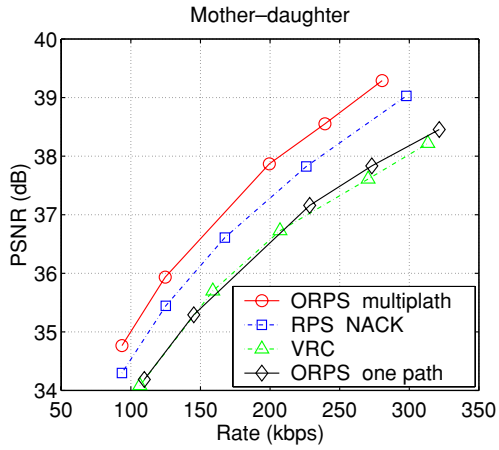

Figure 9. R-D performance of *Mother-Daughter* sequence. $\overline{P}_B = 0.15$, $\overline{L}_B = 8$.

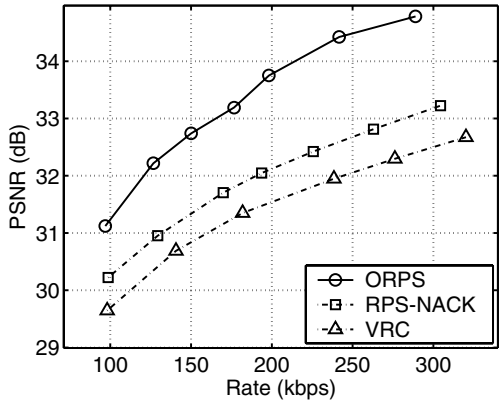

Figure 10. Performance over unbalanced paths. $\overline{P}_{B1} = 0.10$; $\overline{P}_{B2} = 0.20$. *Foreman* sequence.

shown in Fig. 10. 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.

6. Conclusions

We propose an adaptive video transmission scheme using path diversity and rate-distortion optimized reference picture selection, to achieve an improved trade-off between coding efficiency and error-resilience. With the aid of active probing, packets are sent across multiple available paths according to a transmission policy which takes advantage of path diversity and seeks to minimize the loss rate. Experiments demonstrate that the proposed scheme provides significant diversity gain of typical 1 dB in PSNR compared to one-path transmission, when advanced optimized reference picture selection is employed in both scenarios. When compared with video redundancy coding and the NACK mode of conventional reference picture selection, the gain provided by the proposed scheme typically ranges from 0.4 to 1.5 dB.

Acknowledgments

The authors would like to thank Hewlett Packard Laboratories and the Stanford Network Research Center (SNRC) for their sponsorship for this work.

References

1. J.G. Apostolopoulos, "Reliable Video Communication Over Lossy Packet Networks using Multiple State Encoding and Path

Diversity," in *Proceedings Visual Communication and Image Processing*, Jan. 2001, pp. 392–409.
2. Y.J. Liang, E.G. Steinbach, and B. Girod, "Real-Time Voice Communication over the Internet Using Packet Path Diversity," in *Proceedings ACM Multimedia 2001*, Ottawa, Canada, Oct. 2001, pp. 431–440.
3. N. Gogate, D.-M. Chung, S.S. Panwar, and Y. Wang, "Supporting Image and Video Applications in a Multihop Radio Environment using Path Diversity and Multiple Description Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 9, Sept. 2002, pp. 777 –792.
4. D. Comas, R. Singh, and A. Ortega, "Rate-Distortion Optimization in a Robust Video Transmission based on Unbalanced Multiple Description Coding," in *Proceedings of IEEE 4th Workshop on Multimedia Signal Processing*, Cannes, France, Oct. 2001, pp. 581 –586.
5. T. Nguyen and A. Zakhor, "Distributed Video Streaming Over the Internet," in *Proceedings of SPIE Conference on Multimedia Computing and Networking*, San Jose, CA, Jan. 2002.
6. T. Nguyen and A. Zakhor, "Distributed Video Streaming with Forward Error Correction," in *Proceedings of Packet Video Workshop*, Pittsburg, PA, April 2002.
7. T. Nguyen and A. Zakhor, "Protocols for Distributed Video Streaming," in *Proceedings IEEE International Conference on Image Processing*, Rochester, NY, Sept. 2002.
8. A. Majumdar, R. Puri, and K. Ramchandran, "Distributed Multimedia Transmission from Multiple Servers," in *Proc. of the IEEE International Conference on Image Processing*, Rochester, NY, vol. 3, Sept. 2002, pp. 177–180.
9. Y. Wang, S.S. Panwar, S. Lin, and S. Mao, "Wireless Video Transport Using Path Diversity: Multiple Description vs. Layered Coding," in *Proc. of the IEEE International Conference on Image Processing*, Rochester, NY, Sept. 2002.
10. 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, 1999, pp. 289–299.
11. J. Chakareski and B. Girod, "Rate-Distortion Optimized Packet Scheduling and Routing for Media Streaming with Path Diversity," in *Proc. IEEE Data Compression Conference (DCC)*, Snowbird, UT, April 2003.
12. J. Chakareski and B. Girod, "Server Diversity in Rate-Distortion Optimized Media Streaming," in *Proc. IEEE International Conference on Image Processing, ICIP-2003*, Barcelona, Spain, Sept. 2003.
13. D.G. Andersen, H. Balakrishnan, M. Frans Kaashoek, and R. Morris, "The Case for Resilient Overlay Networks," in *Proceedings of the 8th Annual Workshop on Hot Topics in Operating Systems (HotOS-VIII)*, Online at: http://nms.lcs.mit.edu/projects/ron/, May 2001.
14. V.N. Padmanabhan and K. Sripanidkulchai, "The Case for Cooperative Networking," in *Proceedings of the First International Workshop on Peer-to-Peer Systems (IPTPS)*, Cambridge, MA, March 2002.
15. V.N. Padmanabhan, H.J. Wang, and P.A. Chou, "Resilient Peer-to-Peer Streaming," Tech. Rep. MSR-TR-2003-11, Microsoft Research, Redmond, WA, March 2003.
16. 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.
17. J.G. Apostolopoulos, W. Tan, and S.J. Wee, "Performance of a Multiple Description Streaming Media Content Delivery Network," in *Proc. of the IEEE International Conference on Image Processing*, Barcelona, Spain, Sept. 2003.
18. L. Qiu, V.N. Padmanabhan, and G.M. Voelker, "On the Placement of Web Server Replicas," in *Proc. IEEE Infocom*, Anchorage, AK, vol. 3, April 2001, pp. 1587–1596.
19. S. Wenger, G.D. Knorr, J. Ott, and F. Kossentini, "Error Resilience Support in H.263+," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 7, 1998, pp. 867–877.
20. 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. of the IEEE International Conference on Multimedia and Expo (ICME)*, Aug. 2001.
21. ITU-T Recommendation H.263 Version 2 (H.263+), *Video Coding for Low Bitrate Communication*, Jan. 1998.
22. B. Girod and N. Färber, "Feedback-Based Error Control for Mobile Video Transmission," *Proceedings of the IEEE*, vol. 87, no. 10, 1999, pp. 1707–1723.
23. Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, *Joint Final Committee Draft (JFCD) of Joint Video Specification (ITU-T Recommendation H.264, Advanced video coding (AVC) for generic audiovisual services, May 2003 — ISO/IEC 14496-10 AVC)*, Aug. 2002.
24. Y.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)*, Rochester, NY, vol. 2, Sept. 2002, pp. 181–184.
25. M. Budagavi and J.D Gibson, "Multiframe Video Coding for Improved Performance Over Wireless Channels," *IEEE Transactions on Image Processing*, vol. 10, no. 2, 2001, pp. 252–265.
26. T. Wiegand, N. Färber, and B. Girod, "Error-Resilient Video Transmission Using Long-Term Memory Motion-Compensated Prediction," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, 2000, pp. 1050–1062.
27. 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.
28. E. Setton, Y.J. Liang, and B. Girod, "Multiple Description Video Streaming Over Multiple Channels with Active Probing," in *Proceedings of IEEE International Conference on Multimedia and Expo*, Baltimore, MD, July 2003.

Yi Liang received the Ph.D. degree in Electrical Engineering from Stanford University in 2003. His expertise is in the areas of networked multimedia systems, real-time voice and video communication, and low-latency media streaming over the wire-line

and wireless networks. Currently holding positions at Qualcomm CDMA Technologies, San Diego, CA, he is responsible for video and multimedia system design and development for Qualcomm's mobile station modem (MSM) chipsets. From 2000 to 2001, he conducted research with Netergy Networks, Inc., Santa Clara, CA, on voice over IP systems that provide improved quality over best-effort networks. From 2001 to 2003, he had been the lead of the Stanford - Hewlett-Packard Labs low-latency video streaming project, in which he and his colleagues developed error-resilience techniques for rich media communication over IP networks at low latency. In the summer of 2002 at Hewlett-Packard Labs, Palo Alto, CA, he developed an accurate loss-distortion model for compressed video and contributed in the development of the mobile streaming media content delivery network (MSM - CDN) that delivers rich media over 3G wireless. Yi Liang received the B. Eng. degree from Tsinghua University, Beijing, China.

yiliang@stanfordalumni.org

Eric Setton received the B.S. degree from Ecole Polytechnique, Palaiseau, France in 2001 and the M.S. degree, in Electrical Engineering from Stanford University in 2003. He is currently a Ph.D. candidate in the department of Electrical Engineering of Stanford University and is part of the Image, Video and Multimedia Systems group. Multimedia communication over wired and wireless networks, video compession and image processing are his main research interests. In 2001, he received the Carnot fellowship and the SAP Stanford Graduate fellowship. In 2003, he received the Sony SNRC fellowship. He has spent time in industry in France at SAGEM and in the United States at HP labs and at Sony Electronics. He has 4 patents pending.

esetton@stanford.edu

Bernd Girod is Professor of Electrical Engineering in the Information Systems Laboratory of Stanford University, California. He also holds a courtesy appointment with the Stanford Department of Computer Science and he serves as Director of the Image Systems Engineering Program at Stanford. His research interests include networked media systems, video signal compression and coding, and 3-d image analysis and synthesis.

He received his M.S. degree in Electrical Engineering from Georgia Institute of Technology, in 1980 and his Doctoral degree "with highest honours" from University of Hannover, Germany, in 1987. Until 1987 he was a member of the research staff at the Institut fur Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover, working on moving image coding, human visual perception, and information theory. In 1988, he joined Massachusetts Institute of Technology, Cambridge, MA, USA, first as a Visiting Scientist with the Research Laboratory of Electronics, then as an Assistant Professor of Media Technology at the Media Laboratory. From 1990 to 1993, he was Professor of Computer Graphics and Technical Director of the Academy of Media Arts in Cologne, Germany, jointly appointed with the Computer Science Section of Cologne University. He was a Visiting Adjunct Professor with the Digital Signal Processing Group at Georgia Institute of Technology, Atlanta, GA, USA, in 1993. From 1993 until 1999, he was Chaired Professor of Electrical Engineering/Telecommunications at University of Erlangen-Nuremberg, Germany, and the Head of the Telecommunications Institute I, co-directing the Telecommunications Laboratory. He has served as the Chairman of the Electrical Engineering Department from 1995 to 1997, and as Director of the Center of Excellence "3-D Image Analysis and Synthesis" from 1995-1999. He has been a Visiting Professor with the Information Systems Laboratory of Stanford University, Stanford, CA, during the 1997/98 academic year.

As an entrepreneur, Prof. Girod has worked successfully with several start-up ventures as founder, investor, director, or advisor. Most notably, he has been a co-founder and Chief Scientist of Vivo Software, Inc., Waltham, MA (1993–98); after Vivo's aquisition, 1998-2002, Chief Scientist of RealNetworks, Inc. (Nasdaq: RNWK); and, from 1996–2004, an outside Director of 8×8, Inc. (Nasdaq: EGHT). Prof. Girod has authored or co-authored one major text-book, two monographs, and over 250 book chapters, journal articles and conference papers in his field, and he holds about 20 international patents. He has served as on the Editorial Boards or as Associate Editor for several journals in his field, and is currently Area Editor for Speech, Image, Video and Signal Processing of the "IEEE Transactions on Communications." He has served on numerous conference committees, e.g., as Tutorial Chair of ICASSP-97 in Munich and ICIP-2000 in Vancouver, as General Chair of the 1998 IEEE Image and Multidimensional Signal Processing Workshop in Alpbach, Austria, and as General Chair of the Visual Communication and Image Processing Conference (VCIP) in San Jose, CA, in 2001.

Prof. Girod has been a member of the IEEE Image and Multidimensional Signal Processing Committee from 1989 to 1997 and was elected Fellow of the IEEE in 1998 'for his contributions to the theory and practice of video communications.' He has been named 'Distinguished Lecturer' for the year 2002 by the IEEE Signal Processing Society. Together with J. Eggers, he is recipient of the 2002 EURASIP Best Paper Award.

bgirod@stanford.edu