ADAPTIVE MULTIPLE DESCRIPTION VIDEO STREAMING OVER MULTIPLE CHANNELS WITH ACTIVE PROBING

Eric Setton, Yi Liang and Bernd Girod

Information Systems Laboratory, Department of Electrical Engineering
Stanford University, Stanford, CA94305-9510, USA
{esetton, yiliang, bgirod}@stanford.edu

ABSTRACT

We investigate how to transmit video with low latency when multiple burst-loss channels are available. We present a transmission strategy based on feedback for multiple description video. The strategy determines dynamically on which channels video frames and probe packets should be sent. Its goal is to find reliable paths and send independent descriptions on different channels. A multiple description with restart coding scheme is used and ensures a high level of error concealment. When errors are detected and a description is corrupted, the scheme selects a reliable reference frame to restart the stream. We also show that the rate of probe packets can be rate-distortion optimized. Experiments demonstrate significant gains over other schemes such as video redundancy coding.

1. INTRODUCTION

The Internet is a best effort network with no quality of service guarantees. Packets can be lost or can arrive with greatly varying delays. Communication applications requiring low latencies, such as video-on-demand or video conferencing, cannot always wait for packets to arrive or to be retransmitted. Therefore, they need to incorporate a significant amount of error resilience to prevent stalling. Furthermore, in video, errors in decoded frames tend to propagate to subsequent frames. In order to limit this propagation, errors need to be concealed as well as possible.

With multiple description coding, a video transmission can be divided into multiple streams, each of which can be decoded independently at a reduced quality. The independence of the streams offers resilience, as errors will only propagate in one description, at the cost of a lower compression efficiency. The maximum quality is obtained when all the descriptions are correctly received. In video, multiple descriptions can be produced by temporal subsampling as in [1] or by linear combination of motion compensated signals as in [2]. Another approach of multi-stream coding is video redundancy coding[3] where video is compressed into two different streams synchronized periodically by a frame common to both streams.

In the case of bursty channels, losses are temporally correlated; coding can be combined with path diversity to prevent the simultaneous loss of all the descriptions or of all the streams. Multiple paths can be established, for example, by specifying different intermediate routers in sender-driven cases [4]. In the case of wireless transmission, routing protocols can also discover multiple paths as noted in [5].

When timely feedback from the channels is available, dynamic transmission schemes and coding schemes can be used to adapt to the network conditions. They have demonstrated their efficiency compared to other techniques. Some recent schemes use feedback to select reference pictures, e.g. [6] [7], to choose the transmission channels [5] or to do both [8]. Feedback for transmitted packets resides in the reception of negative or positive acknowledgements (NACKs or ACKs) which exist in transport protocols such as TCP.

In this paper, we investigate how to transmit multiple description coded video over multiple channels. The transmission strategy we propose uses feedback to determine how to assign a transmission channel to each description. Its goal is to find reliable paths and to prevent that independent streams be sent on the same channel. The strategy also includes the transmission of header packets to probe channels that are not being used. For the encoding, we separate frames into independent streams by temporal sampling. When a loss is detected in a description, we restart it by choosing a reference in a more reliably transmitted stream.

In the next section, we describe the proposed transmission policy, our multiple description with restart scheme (MDR), and the model used for the channel. In Section 3, we analyze the influence of probes and perform a rate-distortion optimization of their frequency. In Section 4, we show results of comparisons with other schemes.

2. PROPOSED SCHEME

The following transmission scheme can be applied to any kind of multiple description coding. We assume that feed-

back from the channels is available and can be used to assign the state of the channels to good (G) or bad (B). State G, for example, can indicate that packets are always transmitted successfully or could be characterized by a particular delay distribution. Also, we assume that the statistics of the channels are temporally correlated which motivates the use of multiple paths. We consider time to be slotted evenly, each slot representing the transmission time of a video frame encapsulated in a packet.

2.1. Transmission Policy

Here is the transmission policy we propose :

Let N be the number of channels available for transmission and S be the number of descriptions the source is coded into. Select the N_G channels for which the last feedback indicated state G.

If $N_G > 0$, assign each description to one of the N_G channels so that each channel carries either $\lfloor \frac{S}{N_G} \rfloor$ or $\lceil \frac{S}{N_G} \rceil$ descriptions.

If $N_G = 0$, assign each description to one of the N initial channels so that each channel carries either $\lfloor \frac{S}{N} \rfloor$ or $\lceil \frac{S}{N} \rceil$ descriptions.

On the channels that are not being used, send probe packets to ensure feedback information on average every I time slots.

The goal of this policy is to send independent descriptions on different paths when it is possible while avoiding transmission of packets on paths that are presumably in state B. In this way, as long as error bursts do not occur simultaneously on every channel, successful transmission of at least one of the independent streams is achieved and guarantees a minimal quality. Because we are concerned about low latency, we do not consider the option of retransmission and we always send a video frame at every time slot, even in the case of all channels in bad state.

2.2. Multiple Description Encoding with Restart

To generate multiple descriptions we divide the video sequence into S frame-interleaved streams. Each frame is coded as a P-frame and its reference is the previous frame of the stream. We use the notation P-n for a generalized P-frame using a reference n frames away. Here, frames are encoded as P-S frames.

When feedback indicates a frame has been lost, we need to mitigate the error by restarting the stream. For this purpose, we use the latest frame that we believe will be decoded successfully as reference. Thus, the long term memory buffer at the encoder and at the decoder needs to store the S last frames transmitted. An example is shown in Figure 1 for $S = 2$. At time 7, the encoder learns that Frame 3 has been lost and chooses 6 as the reference frame for 7.

If feedback indicates that all streams have been corrupted, we send an I-frame as soon as we receive an ACK from one of the channels and use it to restart all the streams.

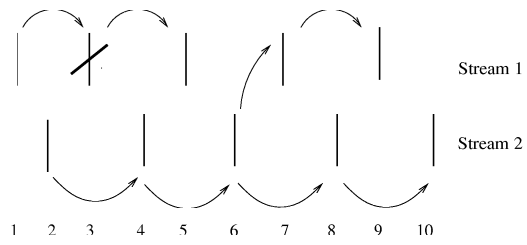

Fig. 1. MDR coding for two streams with $t_{delay} = 4$

2.3. Channel Model

We use the two-state Gilbert model to approximate the bursty behavior of each channel. State G corresponds to the case where packets are received correctly and timely. In state B, packets are lost, either due to network congestion, insufficient channel bandwidth or late arrival. The model is fully determined by the transition probabilities from state G to state B and from state B to state G. These model parameters can in practice be estimated from the accumulated channel statistics [8]: the average probability loss p and the average burst length L_B.

We assume that feedback delay is fixed so that we know if a data packet or a probe packet has arrived t_{delay} after its transmission. This implies that the feedback channel is free of transmission errors. We limit the number of frames transmitted to one per time slot even when $N_G > 1$. This keeps the required bandwidth at every time slot comparable to schemes where only one channel is used. The number of probe packets is not limited and is included in our rate computation.

2.4. Example

Figure 2 shows a simulated transmission trace of a video sequence encoded in two descriptions and streamed over 3 channels. At the beginning of the transmission, Channels 1 and 2 are selected for the transmission of the two streams. At time 10, Channel 2 enters a burst and as soon as the first NACK is received, at time 16, Channel 3 replaces it. Stream 2 is restarted as shown by the encoding of a P-1 frame at that time. At time 28, all the frames are transmitted on Channel 1 as the burst on Channel 3 is detected. The acknowledgement received from the transmission of a probe packet on Channel 2 allows the transmission to continue on two channels at time 35. At time 73, the two descriptions are lost as Channels 2 and 3 enter a burst almost simultaneously. Both streams are restarted by the encoding of an I-frame at time 79 as an ACK from a probe packet from Channel 1 is received.

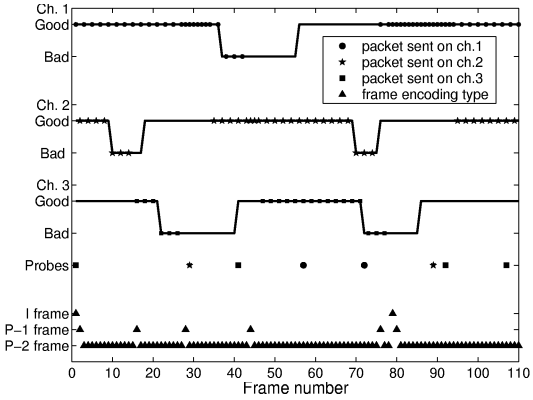

Fig. 2. Transmission over 3 channels with $t_{delay} = 6$

3. INFLUENCE OF PROBES

As probes we use header packets without payload as for example an RTP header encapsulated in an IP header which would have a size of 40 bytes.

Probes are sent to ensure that feedback is received on every channel every I time slots. Figure 3 shows the influence of I on the quality of decoded sequences of 150 frames with $S = 2$ and a quantization parameter $Q = 22$ set in the H.26L codec, for 3 channels with loss rate $p = 15\%$ and average burst length $L_B = 8$ or 12. When I increases, the rate of probe packets decreases and the feedback on idle channels is spaced further. This makes the transmission policy less efficient as the state of the channels is known with less accuracy and results in a linear degradation of the quality as shown in Figure 3. As I approaches the length of the sequence, probes are hardly ever sent and we notice a saturation in the decay.

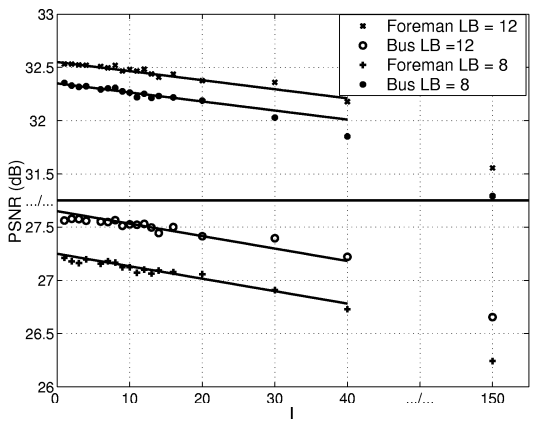

Fig. 3. Influence of I on the PSNR

The influence of I on the bit rate with the same settings is shown in Figure 4 for the Foreman sequence. When I=1 two probe packets (one on every channel not used for transmission) are sent at every time slot which results in a bit rate of more than 19 kbps for probes alone. For I=2, on average, only one probe packet is sent every two time slots which results in a rate of 5 kbps. Note that the difference in bit rate for $I > 4$ is smaller than 4 kbps. The most efficient value of I should minimize the Lagrangian cost function $D + \lambda R$ where R is the rate, D is the distortion and λ is a Lagrange multiplier set to trade off R and D [8]. Here, $I = 16$ is optimal for both sequences. However, the performance of the scheme is not very sensitive to I for $4 \leq I \leq 20$ and in the rest of the experiments we choose $I = 4$ which is close to optimal across bit rates.

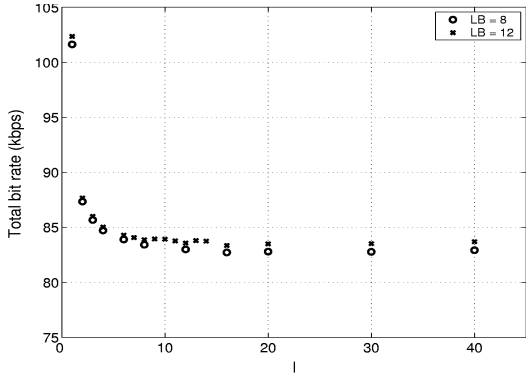

Fig. 4. Influence of I on the bitrate

4. SIMULATION AND RESULTS

In the simulations considered we use the AMR scheme with $S = 2$ which keeps the compression efficiency high while increasing the error resilience.

The first competing strategy we use is the VRC scheme [3] over two channels. A Synch-frame is encoded every 12 frames to restart periodically the streams. If a loss occurs, the Synch frame is predicted by the error-free stream. If both streams are lost, the Synch frame is intra-coded.

The second competing strategy (denoted MD) is very close to the scheme described in [1]. In this scheme, the frames are encoded into two independent streams and sent over two different channels. If a NACK is received on one channel, after the next ACK, an I frame is sent to conceal the error.

For all three schemes, when a frame on one stream is lost while the other stream is correctly received, we recover the lost frame by motion compensated interpolation using the previous and the next frame. Our method is derived from [9]. The simulations were obtained using the H.26L codec, for QCIF sequences of 150 frames ; results are averaged over 40 channel realizations.

Figure 5 and 6 show the performance of our scheme compared to the the other strategies for the Foreman and the Bus

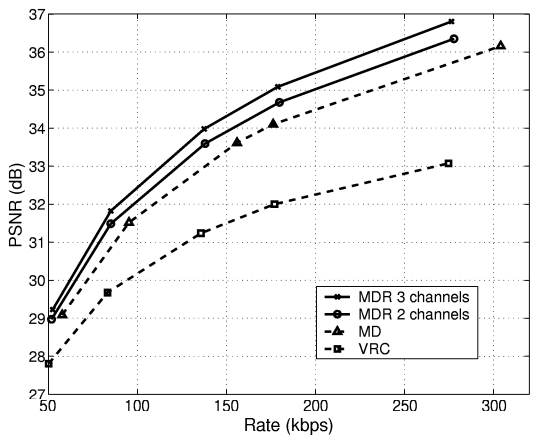

Fig. 5. Foreman sequence, $p = 15\%$, $L_B = 8$

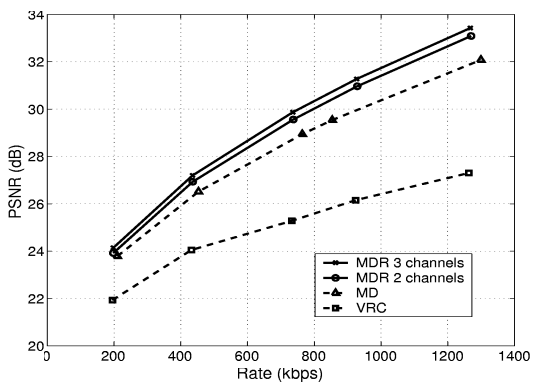

Fig. 6. Bus sequence, $p = 15\%$, $L_B = 8$

sequences. For the Foreman sequence, at 200 kbps MDR over three channels performs better than VRC and MD respectively by 3.1dB and 0.8dB. The increase due to the use of 3 channels is 0.4dB. For the Bus sequence, at 600 kbps MDR over three channels performs better than VRC and MD respectively by 3.9dB and 1dB, and the increase compared to MDR over 2 channels is 0.3dB. For the Foreman sequence at a constant quality of 32dB the bit rate saving is 21% compared to MD and over 50% compared to VRC. Figure 7 shows the performance of our scheme for different channel conditions. The gains against VRC range from 3.1dB to 1.9dB, and from 1.6dB to 0.6dB against MD. They are greater when the burst lengths are higher.

5. CONCLUSION

We demonstrate the efficiency of a transmission strategy based on feedback for multiple descriptions. A multiple description coding scheme with restart is used with this strategy and achieves a high level of error concealment. We show that the rate of header packets used to probe idle channels may be optimized for better results. Significant gains as high as 3.1 dB are achieved against other schemes like VRC for the Foreman sequence at 200 kbps.

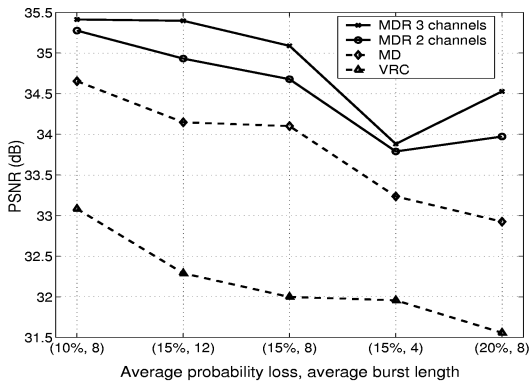

Fig. 7. Foreman sequence at 180 kbps

6. REFERENCES

[1] John G. Apostolopoulos, "Error resilient video compression via multiple state streams," *Proc. Int. Workshop on Very Low Bitrate Video Coding (VLBV'99), Kyoto, Japan*, pp. 168–171, Oct. 1999.

[2] Shunan Lin and Y. Wang, "Analysis and improvement of multiple description motion compensation video coding for lossy packet networks," *Proc. ICIP, Rochester, USA*, pp. 185–188, Sept. 2002.

[3] Stephan Wenger, G. Knorr, and F. Kossentini, "Error resilience support in H.263+," *IEEE Trans. on Circuits Systems for Video Technology*, vol. 8, no. 6, pp. 867–877, Nov. 1998.

[4] John G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," *Proc. VCIP, San Jose, USA*, pp. 392–409, Jan. 2001.

[5] Yao Wang and S. Lin, "Error resilient video coding using multiple description motion compensation," *Proceedings MMSP, Cannes, France*, pp. 441–446, Oct. 2001.

[6] Shunan Lin, S. Mao, and Y. Wang, "A reference picture selection scheme for video transmission over ad-hoc networks using multiple paths," *Proc. ICME, Tokyo, Japan*, Aug. 2001.

[7] Yi J. Liang, M. Flierl, and B. Girod, "Low latency video transmission over lossy packet networks using rate-distortion optimized reference picture selection," *Proceedings ICIP, Rochester, USA*, pp. 181–184, Sept. 2002.

[8] Yi J. Liang, E. Setton, and B. Girod, "Channel-adaptive video streaming using packet path diversity and rate-distortion optimized reference picture selection," *Proceedings MMSP, St. Thomas, US Virgin Islands*, Dec. 2002.

[9] Jonathan K. Su and R. Mersereau, "Motion-compensated interpolation of untransmitted frames in compressed video," *Proc. 30th Asilomar Conf. on Signals Systems and Computers, Asilomar, USA*, pp. 100–104, Nov. 1996.