

Improving Interactive Video in Wireless Networks Using Path Diversity¹

Ahmed Abd El Al^{*}, Chitra Venkatramani[†], Tarek Saadawi^{*}, Myung Lee^{*}

^{*}City College and Graduate Center of City University of New York,
New York, NY 10031
aabdelal@ieee.org, {saadawi, lee}@ccny.cuny.edu

[†]IBM T.J. Watson Research Center
Yorktown Heights, NY 10598
chitrav@us.ibm.com

Abstract. The increase in the bandwidth of wireless channels and the computing power of mobile devices increase the interest in video communications over wireless networks. However, the high error rate and the rapidly changing quality of the radio channels can be devastating for the transport of compressed video. In motion compensated coding, errors due to packet losses are propagated from reference frames to dependant frames causing lasting visual effects. In addition, the bounded playout delay for interactive video limits the effectiveness of retransmission-based error control. In this paper, we propose a mechanism that combines retransmission-based error control with path diversity in wireless networks, to provide different levels of protection to packets according to their importance to the reconstructed video quality. We evaluated the effectiveness of the mechanism under different network conditions. Simulation results show that the mechanism is able to maintain the video quality under different loss rates, with less overhead compared to error control techniques that depend on reference frame updates.

1 Introduction

The increase in the bandwidth of wireless channels and the computing power of mobile devices increase the interest in video communications over mobile wireless networks. However, in such networks there is no end-to-end guaranteed Quality of Service (QoS) and packets may be discarded due to bit errors. Wireless channels provide error rates that are typically around 10^{-2} , which range from single bit errors to burst errors or even intermittent loss of the connection. The high error rates are due to multi-path fading, which characterizes radio channels, while the loss of the connection can be due to the mobility in such networks. In addition, designing the

¹ Prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011.

wireless communication system to mitigate these effects can be complicated by the rapidly changing quality of the radio channel.

The effect of the high error rates in wireless channels can be devastating for the transport of compressed video. Video standards, such as MPEG and H.263, use motion-compensated prediction to exploit the redundancy between successive frames of a video sequence [1]. Although motion-compensated prediction can achieve high compression efficiency, it is not designed for transmission over lossy channels. In this coding scheme the video sequence consists of two types of video frames: *intra-frames* (I-frames) and *inter-frames* (P- or B-frames). I-frame is encoded by only removing spatial redundancy present in the frame. P-frame is encoded through motion estimation using preceding I- or P-frame as a reference frame. B-frame is encoded bidirectionally using the preceding and succeeding reference frames. This poses a severe problem, namely error propagation (or error spread), where errors due to packet loss in a reference frame propagate to all of the dependent frames leading to perceptible visual artifacts that can be long-lasting.

Different approaches have been proposed to tackle the error propagation problem. One approach is to reduce the time between intra-coded frames, in the extreme case to a single frame. Unfortunately, I-frames typically require several times more bits than P- or B-frames. While this is acceptable for high bit-rate applications, or even necessary for broadcasting, where many receivers need to resynchronize at random times, the use of the intra-coding mode should be restricted as much as possible in low bit rate point-to-point transmission, as typical for wireless networks. The widely varying error conditions in wireless channels limit the effectiveness of classic Forward Error Correction (FEC), since a worst-case design would lead to a prohibitive amount of redundancy. Closed-loop error control techniques like retransmission have been shown to be more effective than FEC and successfully applied to wireless video transmission. But for interactive video applications, the playout delay at the receiver is limited, which limits the number of admissible retransmissions [2].

In this paper, we propose a mechanism to provide error resilience to interactive video applications in wireless networks. The mechanism extends retransmission-based error control with redundant retransmissions on diverse paths between the sender and receiver. The mechanism factors in the importance of the packets as well as the end-to-end latency constraints to minimize the overhead and maximize the quality at the receiver. Our simulation results indicate that the proposed mechanism performs significantly better than reference frame update schemes in terms of perceived quality measured at the receiver as well as the transmission overhead.

This paper is organized as follows. Section 2 provides a review for related works. The proposed mechanism is presented in Section 3. Section 4 discusses the mechanism implementation. Section 5 presents experiments that we performed to examine the proposed mechanism and to compare it to reference frame update error control mechanism. Finally, conclusions are outlined in Section 6.

2 Related Work

Analysis for the effects of packet loss on the quality of MPEG-4 video is presented in reference [3], which also proposes a model to explain these effects. The model shows that errors in reference frames are more detrimental than those in dependant frames, due to propagation of errors, and therefore reference frames should be given a higher level of protection.

Forward error correction (FEC) has been proposed to provide error recovery for video packets by adding redundant information to the compressed video bit-stream so that the original video can be reconstructed in presence of packet loss. Reference [4], presents Priority Encoding Transmission (PET) where different segments of video data are protected with redundant information according to their priority, so that information with higher priority can have a higher chance of correct reception. Typical FEC schemes are stationary and must be implemented to guarantee a certain QoS requirement for the worst-case channel characteristics. Due to the fact that wireless channel is non-stationary, and the channel bit error rate varies over time, FEC techniques are associated with unnecessary overhead that reduces the throughput when the channel is relatively error free.

Unlike FEC, which adds redundancy regardless of correct receipt or loss, reference [5] proposes retransmission-based error control schemes, such as Automatic Repeat Request (ARQ), for real time data. Retransmission-based schemes resend only the packets that are lost, thus they are adaptive to varying loss characteristics, resulting in efficient use of network resources. However, retransmission schemes are limited by the receiver's playout delay, as well as the Round Trip Time (*RTT*). Reference [6] presents Time-Lined TCP (TLTCP), which extends the TCP retransmission to support time-lines. Instead of treating all data as a byte stream TLTCP allows the application to associate data with deadlines.

An overview on different error concealment mechanisms proposed to minimize the visible distortion of the video due to packet loss is presented in [7]. Error concealment techniques depend on the smoothness property of the images as well as that the human eye can tolerate distortion in high frequency components than in low frequency components. Reference [2] shows that detectable artifacts can still exist after the error concealment, and that the degree of these artifacts depends on the amount of lost data, the type of the stream and the effectiveness of the concealment algorithm. High-quality concealment algorithms require substantial additional computation complexity, which is acceptable for decoding still images but not tolerable in decoding real-time video. In addition, the effectiveness of concealment depends on the amount and correct interpretation of received data, thus concealment becomes much harder with the bursty losses in wireless channels.

Error-resilient encoding, such as Multiple Description Coding (MDC) and Layered Coding (LC), are proposed to combat channel-induced impairments. MDC generates multiple equally important, and independent substreams, also called descriptions [8]. Each description can be independently decoded and is of equal importance in terms of quality, i.e. there is no decoding dependency between any two of the descriptions. When the decoder receives more descriptions, the quality can be gradually increased no matter which description is received. LC generates one base-layer bitstream and

several enhancement-layer bitstreams [9]. The base-layer can be decoded to provide a basic video quality while the enhancement-layers are mainly used to refine the quality of the video that is reconstructed from the base-layer. If the base-layer is corrupted, the enhancement-layers become useless, even if they are received perfectly.

3 Prioritized Retransmission over Diverse Paths

The ability to successfully decode a compressed bitstream with inter-frame dependencies depends heavily on the receipt of reference frames, and to a lesser degree on dependent frames. Thus, we propose a mechanism to provide adaptive end-to-end unequal error protection for packets belonging to different frames, without sacrificing the timely-delivery requirement for interactive video. We achieve the unequal error protection through redundant retransmissions over diverse paths between the sender and receiver, based on the importance of the packets. There are several ways to set up multiple diverse paths in a wireless network. In single hop wireless network a mobile node would need to establish channels to multiple base stations. In a multi-hop wireless network, routing protocols can utilize the mesh structure of the network to provide multiple loop-free and maximally disjoint paths. Due to the statistical independence of the packet loss events over different paths, by re-transmitting the packets over separate paths, we are maximizing the probability that at least one packet is received error-free, in least number of retransmissions. With a network loss rate l , the error rate can be reduced to

$$Error\ Rate = l^{1+\sum_{i=1}^L M_i} \quad (1)$$

where L is the maximum number of retransmission trials, which is typically determined by the initial playout delay in the receiver as well as the round-trip delay. M_i is the number of retransmission copies during the i^{th} retransmission, which depends on the importance of the retransmitted data to the reconstructed video quality. The maximum number of copies $MAX(M_i)$ is equal to the number of available paths between the sender and receiver.

The scheme is adaptive in the sense that the retransmission overhead will only be added when there is loss in the stream, and the degree of the overhead is proportional to the importance of the lost packets. To ensure in-time delivery of retransmitted packets, and to prevent retransmitting expired packets, the retransmission is controlled by the packet lifetime, as well as estimate(s) of the path delays.

The priority for each data unit in the stream is determined by the application. Thus in the context of motion compensated coding, the application can assign higher priority for I-frames data, than P- or B- frames data. Also P-frames might be assigned varying priority levels, since P-frames that are closer to the preceding I-frame are more valuable for preserving picture quality than later P-frames in the group of pictures (GOP). The prioritization scheme can also be applied on the macroblock basis in coding schemes which provides the encoder with the flexibility to select the coding mode, i.e. intra or inter coding, on the macroblock level [10].

3 Implementation

We implemented the mechanism as a sub-layer above Real Time Protocol (RTP) [11]. Fig. 1 shows the system architecture. We refer to this sub-layer as Multiple Path-RTP (MP-RTP).

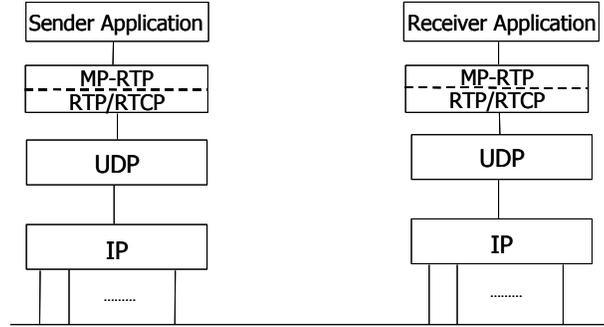


Fig. 1. System architecture

MP-RTP is responsible for:

1. Maintaining the reliability level and the lifetime for each packet, as well as implementing delay constrained retransmission,
2. Monitoring the status of the available paths, and selecting the suitable path(s) for packet retransmission.

For each video frame, the sending application assigns a priority level, which is based on the frame's importance to the reconstructed video quality. I-frames are assigned higher reliability level than P- or B- frames. Also P-frames are assigned varying reliability levels based on their location in the GOP. In addition, the sending application calculates the lifetime for each video frame N , $T_L(N)$, as follows:

$$T_L(N) = T_R(N) + D_S \quad (2)$$

where $T_R(N)$ is an estimate for the rendering time of frame N at the receiver, and D_S is a slack term to compensate the inaccuracies in estimating the One-Way-Delay (*OWD*) from the sender to the receiver, as will be discussed later, as well as the receiver's processing delay. Assuming that there is no compression and/or expansion of total display time at the receiver, the rendering time for frame N , $T_R(N)$, is calculated as follows:

$$T_R(N) = T_0 + T_D + N / R \quad (3)$$

where T_0 is the video session initiation time, T_D is the receiver's playout delay, which determines the rendering time for the first frame in the sequence. Playout delay can be obtained from the receiver during the session initiation. R is the frame rate. As the MP-RTP sub-layer receives a frame it fragments it, if required, into multiple packets, then RTP headers are added and the packets are sent to the receiver. In

addition, a copy of each packet is kept in a retransmission buffer, along with its lifetime and priority. Typically, all the packets within one frame will have the same lifetime and priority. MP-RTP clears packets from the retransmission buffer, as it receives the Real Time Control Protocol-Receiver Reports (RTCP-RR), which are sent regularly from the receiver, indicating the highest sequence number received, as well as other information regarding the quality of the received stream [11]. Initially, packets are sent on a primary path with the receiver, selected by the sender during the session initiation.

The MP-RTP at the receiver is responsible for sending retransmission requests to the sender as soon as it detects a missing packet. The format of the retransmission request, shown in Fig. 2, is similar to RTCP-RR [11], except that it is extended to include the 32 bits sequence number of the missing packet. As the retransmission request is susceptible to losses, the MP-RTP retransmits these reports on different paths to the sender.

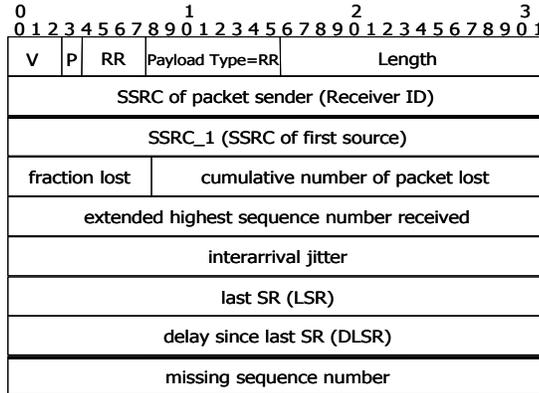


Fig. 2. Extended RTCP-RR to include the missing sequence number

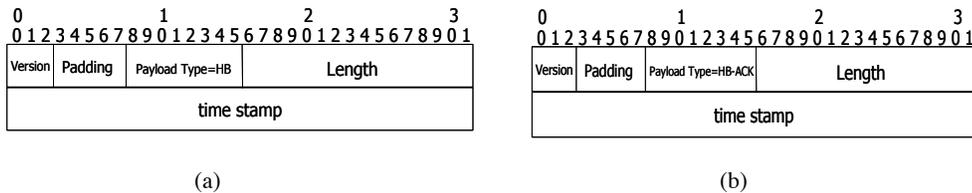


Fig. 3. (a) Heartbeat packet (b) Heartbeat acknowledgement packet

MP-RTP uses Heartbeat packets, shown in Fig. 3.a, to maintain an estimate for the *RTT* of the available paths. The *RTT* estimate is an exponential average of current and past *RTT* measurements. Each heartbeat packet includes a time stamp indicating the transmission time. The MP-RTP at the receiver responds to the heartbeat packet by

sending a Heartbeat-Acknowledgment packet, shown in Fig. 3.b, on the same path from which the heartbeat was received. The heartbeat-acknowledgement includes a copy of the timestamp in the corresponding heartbeat packet. The RTT estimates are used to obtain an approximation for the paths OWD , i.e., $OWD \approx RTT / 2$. The application can compensate the inaccuracies in the OWD approximation as it assigns the frames lifetime, as shown in equation 2. In addition, MP-RTP uses the RTT estimates to switch the primary path, which can break due to the mobility in the wireless network. To minimize the interruption for the interactive video session, as the primary path RTT increases beyond a certain threshold, MP-RTP sets the alternative path with the shortest RTT to be the primary path. The switching threshold can be based on the maximum delay allowed for the interactive video application. Currently, we are using a fixed value for the switching threshold. In future work, we are planning to investigate techniques to dynamically adapt the value of the switching threshold.

As soon as the sender receives a retransmission request, it performs the following algorithm:

1. If the lost packet has a low priority, go to step 2, otherwise go to step 3
2. Check the round trip time estimate RTT_i for all the available paths, maintained using heartbeat packets. Select the retransmission path i with the minimum OWD_i , such that the following condition holds:

$$T_c + OWD_i < T_L(j) \quad (4)$$

where T_c is the current time at the sender and $T_L(j)$ is the lifetime for frame j , to which the retransmitted packet belongs.

3. For high priority packets, the sender selects all the available path(s) that satisfies condition 4, and retransmits the packet on these paths simultaneously.

By controlling the retransmission through the frames lifetime, as well as estimate(s) of the path(s) delay, MP-RTP prevents retransmission of expired packets while trying to meet the frames lifetime constraint. If no path(s) is suitable in step 2 or 3, the retransmission is discarded, as the packet will not be received before the rendering time for the frame to which it belongs. At the same time the upper layer application is notified about the dropped packet to allow the encoder to utilize schemes, such as error tracking, to limit the error propagation [2].

4 Performance Analysis

In order to examine the performance of the proposed mechanism, we implemented the mechanism in OPNET simulation and modeling tool [12]. We simulated a Multi Path Transport (MPT) system, with configurable number of single hop paths between the sender and receiver. For simplicity we assumed that the paths are identical in terms of available bandwidth, equal 2.0 Mbps. A two-state model Markov model, shown in Fig. 4, is used to simulate the bursty packet loss behavior in wireless channels [13].

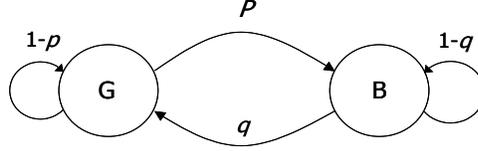


Fig. 4. A two-state Markov model to simulate burst packet losses

The two state model, which is often referred to as Gilbert channel model, has been shown to be able to effectively capture the bursty packet loss behavior of the wireless channels. The two states of this model are denoted as *Good* (G) and *Bad* (B). In state G , packets are received correctly whereas, in state B , packets are assumed to be lost. This model can be described by the transition probabilities p from state G to B and q from state B to G . The average packet loss rate (PLR) is:

$$\text{Average } PLR = \frac{p}{p + q} \quad (5)$$

We vary the error characteristics for channel i by appropriately controlling the channel *Good* and *Bad* durations, according to an exponential distributions with averages p_i and q_i , respectively. Delay for channel i is modeled by an exponential distribution with the mean delay $D_i = 30$ msec. We set the path maximum transfer unit (MTU) of 400 bytes for all the paths. The heartbeat interval is set to 150 msec.

To generate the video sequence used in our simulation, we used open source XviD MPEG-4 compliant video codec [14]. Sixty seconds of a high motion video sequence (football match) are encoded at 15 frames per second (fps), which results in a sequence of 900 frames. The frame resolution is quarter common intermediate format (QCIF, 176 x 144 pixels), which is the most common format at low bit rates, and the coding rate is 200 Kbps. We repeated our experiments with limited motion video sequence (TV news) and we get similar results to that shown here. We limited the playout delay at the receiver to 100 msec., to represent an interactive video application. We set the switching threshold, discussed in Section 4, to 200 msec. We selected this value because given the channel delays and the playout delay at the receiver, having the RTT of the primary path higher than this threshold will result in all frames arriving later than their rendering time at the receiver and will be discarded.

The average Peak Signal to Noise Ratio (PSNR) is used as a distortion measure of objective quality. PSNR is an indicator of picture quality that is derived from the root mean squared error (RMSE). Without transmission losses, the average PSNR of the decoded frames for the video sequence used in our performance study is 27 dB.

After obtaining a transmission trace of a video sequence, we run the decoder on the trace to measure the image distortion due to packet losses, using the PSNR. In order to generate statistically meaningful quality measures, for each simulation scenario we repeated the experiment ten times with different seeds. The presented PSNR values are the average of the ten experiments.

In our performance study we set the application to choose I-frames and half of the P-frames starting from the I-frame in a GOP to be high priority frames, while other frames are set to low priority frames.

4.1 Effect of Packet Loss Rate on Video Quality

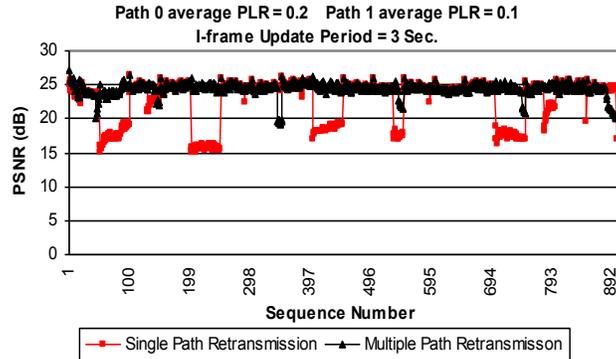


Fig. 5. PSNR versus frame number

We tested MP-RTP using two diverse paths, namely path 0 and path 1, between the sender and the receiver. Path 0 was selected as the primary path during the video session initiation. The channel average packet loss rates for path 0 and path 1 were set to 0.2 and 0.1 respectively. We set the encoder so that the I-frame update period, i.e. interval between two consecutive I-frames, equal 3 seconds. Fig. 5 shows the PSNR for each frame in the video sequence. For comparison we repeated the experiment using the retransmission scheme with single path retransmissions, where missing packets are retransmitted on a single path selected randomly from the paths between the sender and receiver. As can be shown from the figure that the redundant retransmission scheme is able to maintain the video quality, at high packet loss rates. On the other hand, with the single path retransmission scheme, the video quality can be dropped for long durations due to loss of packets in reference frames, and under the high loss rate retransmitted packets can also be lost, leading to error propagation in the following dependent frames up to the next I-frame. Although the sender can keep retransmitting the packet, the receiver will discard these retransmissions, as they arrive after the frame rendering time.

Fig. 6, shows the average PSNR over the whole sequence versus different channel average packet loss rates for the primary path, i.e. path 0. The channel average packet loss rate for path 1 is set to 0.1. We repeated the same experiment with different I-frame update periods. For our mechanism we used an I-frame update period equal 3 seconds. As can be seen in Fig. 6, the single path retransmission scheme achieves a similar performance to MP-RTP only when the I-frame frequency is increased more than three times to one every 15 frames. As the I-frames have larger sizes than P- and B-frames, increasing the I-frame frequency for the same bit rate translates to reduced video quality since bits are now wasted to code I-frames. If the I-frame frequency is set to one in 45 frames for the single path case, it can be seen that the quality deteriorates rapidly.

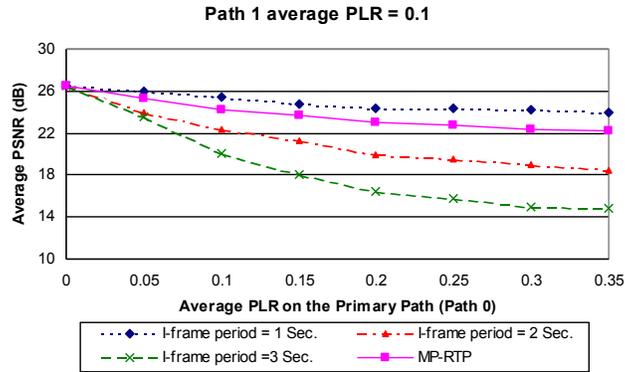


Fig. 6. Average PSNR versus average packet loss rate

Again this is mostly due to losses in reference frames, as a result of the high packet loss rate and the bounded delay for interactive video. The errors are propagated from reference frames to the following frames up to the next I-frame. On the other hand, redundant retransmissions over diverse paths ensures that in the single retransmission allowed at least one copy of the packet will be received, preventing the error propagation.

4.2 Effect of Changing the Number of Paths

We tested the redundant retransmission mechanism with different number of paths between the sender and receiver. In all experiments the I-frame update period is equal 3 seconds.

We varied the channel average packet loss rate on the primary path, i.e. path 0, from 0.05 to 0.3. We represented the independent packet losses for the other paths, i.e. paths 1-3, by choosing different channel average packet loss rates 0.01, 0.1 and 0.2 respectively. As can be seen from Fig. 7, with a single path the quality deteriorates at high packet loss rates, due to error propagation. But, with MP-RTP, increasing the number of paths between the sender and the receiver, improves the quality due to the independent loss characteristics of the paths, which increases the probability that the retransmitted packets will be received before their deadline.

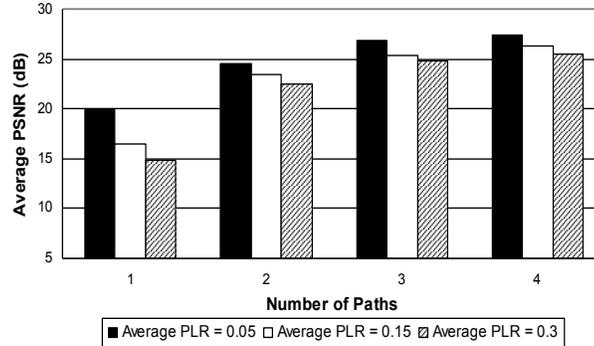


Fig. 7. Average PSNR versus number of paths

5.3 Redundant Retransmission Overhead

In this experiment, we compared the overhead of MP-RTP, due to the redundant retransmissions and heartbeats, to the overhead of error control mechanisms that depend on increasing the I-frame frequency to limit the error propagation.

We define the overhead ratio to be the total number of bytes sent in I-frame update scheme to the total number of bytes sent in MP-RTP, to attain a given video quality represented by the average PSNR. In order to calculate the maximum overhead for MP-RTP, we used 3 paths. We varied the channel average packet loss rate for the primary path, path 0, while the channel average packet loss rates for the other paths, path 1 and path 2, were set to 0.1 and 0.2 respectively.

Fig. 8 shows the overhead ratio for average PSNR equal 23 dB. As was shown before, the single path retransmission case required an I-frame frequency of almost 1 per second, while the MP-RTP required 1 per 3 seconds, for a video quality of around 23 dB. It can be seen from the figure that the overhead of our mechanism is less than that for the I-frames update scheme. The reason behind this is that the redundant retransmission mechanism implemented in MP-RTP is adaptive, in the sense that it only adds the retransmission overhead when there is loss in the video stream. In addition, the degree of the overhead is proportional to the importance of the lost packets. Although heartbeat packets are periodically sent, they have less contribution to the overhead, as they are small in size compared to the size of video frames.

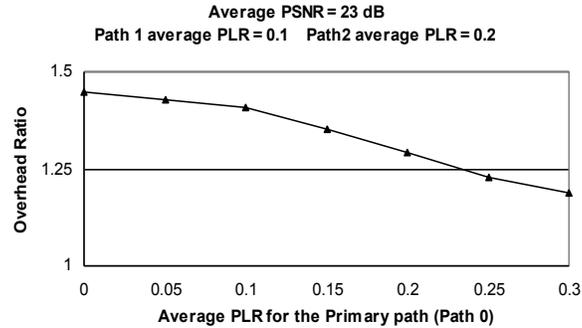


Fig. 8. Overhead ratio versus average packet loss rate on the primary path

6 Conclusion

The nature of video encoded using motion compensation requires higher protection for reference frames than dependent frames, otherwise errors due to packet losses in reference frames propagate to dependent frames. Interactive video complicates the problem by bounding the time available for the error control. To tackle these problems, we propose a mechanism to provide unequal error protection to data within the video stream according to their importance to the reconstructed video quality. The unequal error protection is realized through extending the classic retransmission based error control, with redundant retransmissions on diverse paths, in order to increase the probability that at least one of the retransmitted packets arrive at the receiver in less number of retransmissions. The degree of redundant retransmission depends on the reliability level required for the data within the retransmitted packet. A delayed constrained retransmission, based on the packet lifetime and estimate of the delay from the sender to receiver, is used to prevent re-transmitting expired packets. We implemented the proposed mechanism as an extension to RTP, referred to as Multi Path - RTP (MP-RTP). Performance results show that the mechanism is able to provide a good quality for interactive video under different packet loss rates. In addition, comparing the transmission overhead of the mechanism to the overhead of reference frame updates error control mechanism, it is shown that for a given video reconstruction quality MP-RTP has less overhead, which is an important feature required in wireless networks.

Disclaimer

The views and conclusions in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research laboratory or the U.S. Government.

References

1. International Organization for Standardization, Overview of the MPEG-4 Standard. (1999)
2. Girod, B., Farber, N.: Feedback-Based Error Control for Mobile Video Transmission. Proceedings of the IEEE, special issue on video for mobile multimedia, Vol. 97, No. 10, (1999) 1707-1723
3. Feamster, N., Balakrishnan, H.: Packet Loss Recovery for Streaming Video. International Packet Video Workshop (2002)
4. Albanese, A., Blomer, J., Edmonds, J., Luby, M., Sudan, M.: Priority Encoding Transmission. IEEE Transactions on Information Theory, Vol. 42, No. 6 (1996)
5. Dempsey, B.: Retransmission-Based Error Control for Continuous Media in Packet Switched Networks. Ph.D. thesis, University of Virginia (1994)
6. Mukherjee, B., Brecht, T.: Time-lined TCP for the TCP-friendly Delivery of Streaming Media. International Conference on Network Protocols (ICNP) (2000)
7. Wang, Y., Zhu, Q.: Error Control and Concealment for Video Communications: A Review. Proceedings of IEEE, Vol.86, No. 5 (1998)
8. Goyal, V.: Multiple Description Coding: Compression Meets the Network. IEEE Signal Processing Magazine, Vol. 18 (2001) 74-93
9. Lee, Y., Kim, J., Altunbasak, Y., Mersereau, R.: Layered Coded vs. Multiple Description Coded Video over Error Prone Networks. Signal Processing: Image Communication, Vol. 18 (2003) 337-356
10. ITU-T Recommendation, H.263, in Video Coding for Low Bitrate Communication (1997)
11. Schulzrinne H., Casner S., Frederick, R., Jacobson, V.: RTP: A Transport Protocol for Real-Time Applications. RFC 3550 (2003)
12. Opnet Simulation and Modeling Tool, www.opnet.com.
13. Gilbert, E.: Capacity of a burst-noise channel, Bell System Technical Journal, Vol. 39, No. 9 (1960) 1253-1265
14. XviD MPEG-4 video codec, www.xvid.org