Modeling and Evaluating Feedback-Based Error Control for Video Transfer

by

Yubing Wang

A Dissertation
Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment
of the Requirements for the Degree
Doctor of Philosophy

in

Computer Science

by

____

August 2008

APPROVED:

Prof. Mark Claypool
Advisor

Prof. Robert Kinicki
Co-Advisor

Prof. Dan Dougherty
Committee Member

Prof. Ketan Mayer-Patel
Committee Member

Prof. Michael Gennert
Head of Department

Contents

List of Figures

List of Tables

Abstract

Packet loss can be detrimental to real-time interactive video over lossy networks because one lost video packet can propagate errors to many subsequent video frames due to the encoding dependency between frames. Feedback-based error control techniques use feedback information from the decoder to adjust coding parameters at the encoder or retransmit lost packets to reduce the error propagation due to data loss. Feedback-based error control techniques have been shown to be more effective than trying to conceal the error at the encoder or decoder alone since they allow the encoder and decoder to cooperate in the error control process. However, there has been no systematic exploration of the impact of video content and network conditions on the performance of feedback-based error control techniques. In particular, the impact of packet loss, round-trip delay, network capacity constraint, video motion and reference distance on the quality of videos using feedback-based error control techniques have not been systematically studied.

This thesis presents analytical models for the major feedback-based error control techniques: Retransmission, Reference Picture Selection (both NACK and ACK modes) and Intra Update. These feedback-based error control techniques have been included in H.263/H.264 and MPEG4, the state of the art video in compression standards. Given a round-trip time, packet loss rate, network capacity constraint, our models can predict the quality for a streaming video with retransmission, Intra Update and RPS over a lossy network. In order to exploit our analytical models, a series of studies has been conducted to explore the effect of reference distance, capacity constraint and Intra coding on video quality. The accuracy of our analytical models in predicting the video quality under

different network conditions is validated through simulations. These models are used to examine the behavior of feedback-based error control schemes under a variety of network conditions and video content through a series of analytic experiments.

Analysis shows that the performance of feedback-based error control techniques is affected by a variety of factors including round-trip time, loss rate, video content and the Group of Pictures (GOP) length. In particular: 1) RPS NACK achieves the best performance when loss rate is low while RPS ACK outperforms other repair techniques when loss rate is high. However RPS ACK performs the worst when loss rate is low. Retransmission performs the worst when the loss rate is high; 2) for a given round-trip time, the loss rate where RPS NACK performs worse than RPS ACK is higher for low motion videos than it is for high motion videos; 3) Videos with RPS NACK always perform the same or better than videos without repair. However, when small GOP sizes are used, videos without repair perform better than videos with RPS ACK; 4) RPS NACK outperform Intra Update for low-motion videos. However, the performance gap between RPS NACK and Intra Update drops when the round-trip time or the intensity of video motion increases. 5) Although the above trends hold for both VQM and PSNR, when VQM is the video quality metric the performance results are much more sensitive to network loss. 6) Retransmission is effective only when the round-trip time is low. When the round-trip time is high, Partial Retransmission achieves almost the same performance as Full Retransmission. These insights derived from our models can help determine appropriate choices for feedback-based error control techniques under various network conditions and video content.

Chapter 1

Introduction

1.1 Motivation

The growth in power and display capabilities of today's computers has enabled streaming video with a range of quality to be viewed by end-users. High-end users with modern desktop displays can watch videos in full-quality, wide-screen mode at their desk-tops while low-end users with video-capable mobile phones can watch low resolution video on their mobile phones. The growth in computer technology has been matched by an equal the growth in capacity and connectivity of networks. Users on high-speed corporate and academic networks have had sufficient bandwidth to stream video for some time, but the pervasiveness of broadband networks has also given home users access to high-quality streaming video. Moreover, increasing bandwidth for digital cellular networks has enabled streaming video to mobile laptops, PDAs and even mobile phones. However, despite the increase in network power and connectivity, many network connections still lose data packets. Lost packets are especially detrimental to streaming video because of the dependency between video frames during encoding where one lost video packet can result in error propagation to many other video frames.

Many error recovery techniques have been proposed to repair damaged video due to packet loss. These techniques can be broadly categorized into three groups by whether the encoder or decoder plays the primary role, or both are involved in cooperation with each other [1][2]. Examples of error control techniques at the encoder side include Forward Error Correction (FEC) [3][4], joint source and channel coding (JSCC) [5][6], and layered coding [7][8]. Essentially, they all add redundancy at either the encoding or the transport layer to minimize the effect of transmission errors. While error control techniques at the encoder such as FEC can effectively reduce error propagation, they require additional data to be added to the video stream and encoding and decoding of these techniques can be somewhat complicated. Error control techniques at the decoder side include spatial and temporal smoothing [9], interpolation [10], and filtering [11]. In general, these techniques attempt to recover the damaged videos by estimation and interpolation. While local concealment techniques can visually cover up the loss, the ability to adequately repair video without help from the encoder is limited. The error controls that have interaction between encoder and decoder are called *feedback based error control* [12]. Examples in this category include Retransmission [13][14], Reference Picture Selection (RPS) [15]-[17] and Intra Update [12].

Feedback-based error control [12] techniques use information on the data sent by the decoder to adjust the coding parameters at the encoder or retransmit lost packets to achieve better error repair. The feedback information provided by the decoder indicates the location of damaged parts of the video stream. Based upon the feedbacks, the encoder can identify the affected areas and treat them differently. Generally, since the encoder and decoder cooperate in the error control process, feedback-based error control techniques can achieve better error resilience than error control techniques where only the encoder or decoder play the primary role [1]. This thesis focuses on major feedback-based error control techniques, including Reference Picture Selection (RPS), Retransmission and Intra Update.

A promising repair technique for delay-sensitive video is Reference Picture Selection (RPS)[1] [15]-[17]. Broadly, in RPS, the video encoder uses one of several previous frames that have been successfully decoded as a reference frame for encoding. The reference frame can, by default, be the previous frame (called *RPS NACK*), or the reference frame can be several frames older if the encoder waits for receiver confirmation of successful frame reception (called *RPS ACK*). In the negative acknowledgement (NACK) mode, when a transmission error is observed by the decoder, the decoder sends an NACK message for an erroneous frame, along with the number of a previously received, correctly-decoded frame that can be used as a reference for prediction, to the encoder. Relying on the feedback information provided by the decoder to locate the lost packets, the video quality with RPS NACK degrades for a period of one round-trip time when a transmission error occurs. However, instead of retransmitting the lost video packet, which requires extra bandwidth, the encoder only transmits the encoded frame that uses the previously-received frame for prediction, consuming less bandwidth. In the RPS positive acknowledgement (ACK) mode, all correctly received frames are acknowledged and the encoder only uses acknowledged frames as a reference. Since the encoder usually has to use an older frame for prediction, the coding efficiency degrades as the round-trip delay increases. On the other hand, using RPS ACK mode can entirely eliminate error propagation.

Unlike forward error control techniques (such as FEC), Retransmission can recover the distorted video without incurring much bandwidth overhead because packets are

[1] Chapter 2 provides detailed information about RPS.

retransmitted only when they are determined lost. However, retransmission of lost packets takes at least one additional round-trip time and thus may not be suitable for interactive video applications such as video conferencing that require short end-to-end delays. In some wireless video applications, such as mobile video, where the packet loss rate and the end-to-end delay can be high and capacity is limited, Retransmission alone may not be sufficient for packet loss recovery. Most conventional retransmission schemes delay frame playout times to allow the retransmitted packets to arrive before the display times of their video frames in order to accommodate the added latency. Any packets received after their display times are then discarded. We adopt a retransmission scheme [13] that is different in that packets arriving after their display times are not discarded but instead are used to reduce error propagation.

With Intra Update[2] error control, based upon the feedback information from the decoder, the encoder knows which portions in a frame are damaged and simply encodes those damaged portions in Intra[3] mode. Using Intra Update can stop error propagation in about one round-trip time. However, Intra coding reduces the coding gain and hence degrades the video quality under the same bit-rate constraint.

The choice of Retransmission, Intra Update, RPS NACK or RPS ACK within a video flow with inherent inter-frame encoding dependencies depends upon the network conditions (such as capacity constraints, packet loss rate and round-trip time) between the

[2] The detailed information about Intra Update can be found in Chapter 2.
[3] If a frame is encoded in INTRA mode, it is encoded directly without reference to previously encoded and reconstructed frames

video server and client, application requirements (such as end-to-end delay), and the impact of reference distance[4] on the encoded video quality.

1.2 The Dissertation

Although numerous studies have detailed the benefits of various repair schemes to video quality [1][2][12][66][67], to the best of our knowledge, there has been no systematic exploration of the impact of video and network conditions on the performance of feedback-based error control schemes. This thesis derives a series of analytical models to predict the quality of videos streamed with RPS NACK, RPS ACK, Intra Update or Retransmission. These models are then used to analyze performance of feedback-based error control schemes under various network conditions and video contents through a series of analytic experiments.

In order to validate and then exploit our analytical models to analyze the performance of feedback-based error control techniques, we adopt the following methodology:

1) Determine the input parameters for the analytical models;
2) Measure the impact of reference distance on video quality;
3) Build the analytical models;
4) Validate the analytical models through simulation;
5) Explore the performance of feedback-based error repair techniques using the analytical models

In order to compare the performance of RPS ACK and RPS NACK, we need to determine how the reference distance affects the video quality. The existing studies detailing the benefits to video quality for various repair techniques typically do not vary

[4] The distance between the encoding frame and the reference frame that is used for motion compensation prediction.

the reference distance during encoding. To the best of our knowledge, the effects of encoding distance on video quality have not been quantitatively studied. We conducted systematic measurements of the effects of reference distance on video quality for a range of video coding conditions [81]. High-quality videos with a wide variety of scene complexity and motion characteristics are selected for baseline encoding. The videos are all encoded using H.264 [18]-[22], an increasingly popularly deployed compression standard with support for RPS, with a bandwidth constraint and a range of reference distances. Two objective measures of video quality are used: the popular Peak Signal to Noise Ratio (PSNR), and the reportedly more accurate Video Quality Metric (VQM) [23]. Analysis shows that for both measures of quality, the scene complexity and motion characteristics determine the degradation of video with higher reference distances. In particular, videos with low motion degrade more with higher reference distance since they cannot take advantage of the similarity between adjacent frames. Videos with high motion do not suffer as much with an increase in reference distance since the similarity between frames is already low. The scene complexity determines the overall starting quality with a default, encoding reference distance of one and the bandwidth constraint.

Our analytical models for feedback-based error control techniques captures the relationship between the video quality that can be achieved using these error control techniques and various network characteristics and video contents [82] [83]. The models target H.264 videos since this standard incorporates all these four feedback-based error control techniques, but can generally represent any video encoding technique that uses feedback-based repair.

The accuracy of our analytical models in predicting video quality under different network conditions is validated through simulation. Comparing performance predicted by the analytical models against simulated performance provides an indication of the model accuracy. The simulations modify the input video sequences based on the given loss probability and round-trip delay to mimic the effect of packet loss as well as the change of reference distance on the video quality. The modified input sequences are encoded using H.264 and the average video quality in terms of PSNR and VQM is measured and compared against that predicted by our analytical models.

By employing the analytic models that predict the quality of videos streamed with RPS NACK, RPS ACK, Intra Update or Retransmission, this thesis provides detailed analysis of feedback-based error control schemes over a range of network loss and latency conditions using a variety of videos chosen to represent a diverse range in video scene complexity and motion characteristics. The basis for our video encoding model is H.264. Both PSNR and VQM are used to measure video quality. The models incorporate a bandwidth constraint and a range of reference distances from the network.

1.3 Contributions

The main contributions of this dissertation are the design, validation, simulation, and evaluation of the analytical models for feedback-based error control techniques. The specific contributions of the dissertation include:

1. A systematic study of the effects of reference distance on video quality for a range of video coding conditions [81]. A set of video clips with a variety of motions are selected for study, and the video sequences are shuffled to change

the reference distances. For each reshuffled video sequence, an H.264 encoder encodes the sequence and measures video quality with PSNR and VQM.

2. Two utility functions that characterize the impact of reference distance on video quality based upon the study [81]. While the relationship between PSNR and reference distance can be characterized using a logarithmic function, with VQM as the video quality metric, the same relationship can be characterized using a linear function.

3. Modeling the prediction dependency among GOB[5]s for RPS NACK [82][83] and Intra Update. Based on these two models, the probabilities of correctly decoding a GOB encoded with RPS NACK or Intra Update can be calculated.

4. Study of the impact of bandwidth constraint on video quality in terms of VQM and PSNR. For both video quality metrics, the impact of bandwidth constraints on video quality can be characterized using a logarithmic function.

5. A Partial Retransmission scheme in which only a fraction of lost packets are retransmitted based on their priorities. The analytical model for this retransmission scheme is created and used to analyze its performance.

6. Analytical models for feedback-based error control techniques including Full Retransmission, Partial Retransmission, RPS ACK, RPS NACK and Intra Update. These models characterize the feedback-based error control techniques, incorporating the impact of reference distance, bandwidth constraint, and Intra

[5] GOB (Group of Blocks) contains a fixed number of successive macro-blocks (MB's)

coding on video quality, prediction dependency among GOBs in the reference chain and Group of Picture (GOP) length.

7. Simulations that verify the accuracy of our analytical models. The simulations modify the input video sequences based on the given loss probability and round-trip delay to mimic the effect of packet loss as well as the change of reference distance on the video quality.

8. Analytic experiments over a range of loss rates, round-trip times and video content using our models. The experiments explore a wide range of factors that may impact the performance of feedback-based error control techniques. The analysis based on these experiments is useful for helping select the feedback-based repair techniques to improve video quality.

1.4 Road Map

The remainder of this thesis is organized as follows: Chapter 2 provides background knowledge on coding standards, feedback-based error control techniques; Chapter 3 describes related work; Chapter 4 provides a detailed description of our analytical models; Chapter 5 details the study of impact of reference distance on video quality; Chapter 6 presents the experimental analysis; Chapter 7 validates the accuracy of our analytical models; Chapter 8 summarizes our conclusions and finally Chapter 9 presents possible future work.

Chapter 2

Background

This chapter provides background knowledge for our thesis. Section 2.1 provides an overview of media repair techniques. Section 2.2 introduces feedback-based error control techniques, including Retransmission, Reference Picture Selection (RPS) and Intra Update. Section 2.3 discusses some of the local concealment techniques. Section 2.4 introduces H.264, one of the most popular video compression standards today, and discusses some of the error control techniques embedded in H.264. Section 2.5 describes video buffering techniques. Section 2.6 describes media scaling techniques. Section 2.7 describes the methods of video quality measurement including PSNR and VQM. Section 2.8 summarizes this chapter.

2.1 Error Control Techniques

Many error recovery techniques have been proposed to repair damaged video due to packet loss. These techniques can be broadly categorized into three groups by whether the encoder or decoder plays the primary role, or both are involved in cooperation with each other 0[2]. Examples of error control techniques at the encoder side include Forward Error Correction (FEC) [3][4], joint source and channel coding (JSCC) [5][6], and layered coding [7][8]. Essentially, they all add redundancy in either the source coder or the transport coder to minimize the effect of transmission errors. The error control techniques at the decoder side are called local concealment. Examples of decoder side error control techniques include Motion Compensated Temporal Prediction (MCTP) [2], Spatial Interpolation [24], and Filtering [11]. In general, these techniques attempt to repair the damaged videos by estimation and interpolation. The error controls that have interaction between encoder and decoder are called *feedback based error control* [12]. Examples in this category include Retransmission [13][14], Reference Picture Selection (RPS) [15]-[17] and Intra Update [12].

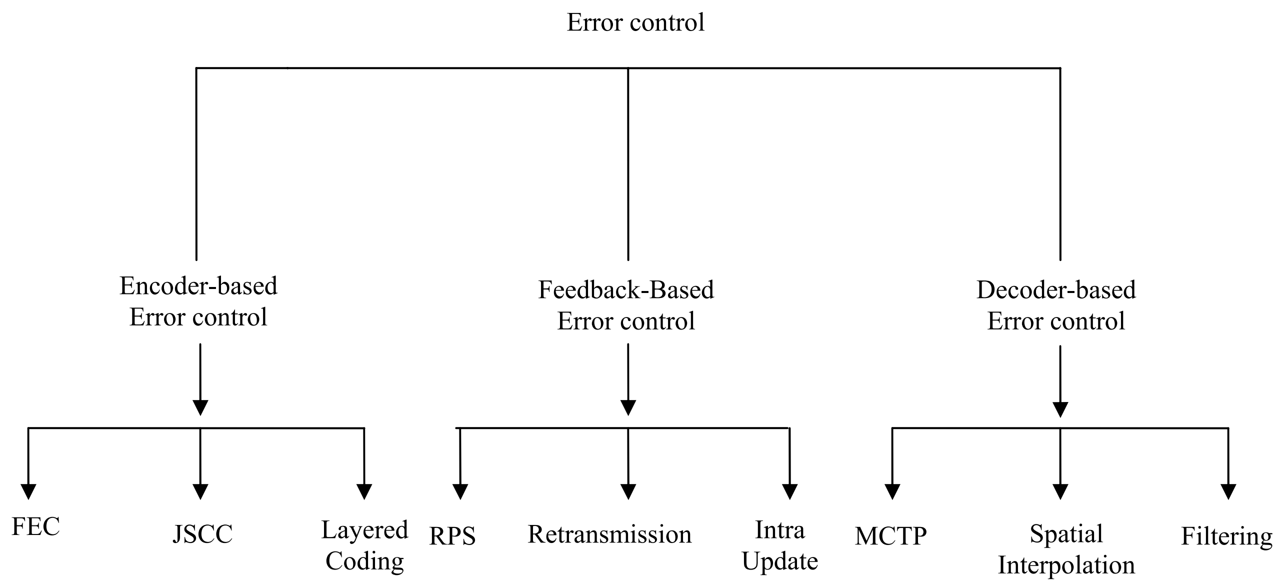

Figure 2.1 Error control techniques

2.2 Feedback-based Error Control Techniques

Feedback-based error control techniques [12] use the acknowledgement from the decoder to adapt the source coder to the channel conditions. The adaptation can be achieved at either the transport level or at the source coding level. At the transport level, the feedback information can be employed to trigger retransmission of lost packets or change the percentage of the total bandwidth used for retransmission. At the source coding level, coding parameters (such as reference frame selection) can be adapted based on the feedback from the decoder. In this section, we first describe retransmission, which is adopted at the transport level, and then Reference Picture Selection (RPS) [15]-[17] and Intra Update, both of which are adopted at the source coding level.

2.2.1 Retransmission-Based Video Error Control

Retransmission [13][14] is the most commonly used error recovery technique for reliable data transport. Since repair packets are retransmitted only when some packets are lost, retransmission incurs very little unnecessary overhead. The conventional retransmission schemes delay frame playout times to allow the retransmitted packets to arrive before the display times of their video frames. These schemes add at least one round-trip time to the display time of a frame after its initial transmission. The retransmission technique we employ is different from conventional ones in that packets arriving after their display time are not discarded but instead used to reduce error propagation [13]. Figure 2.2 illustrates how this retransmission scheme works. Here we assume that each network packet contains one Group of Macro-blocks (GOB). During the transmission, one GOB (GOB 2) in Frame 2 was lost, and at time t1 the receiver detected that GOB 2 was not received. The receiver then sent a negative acknowledgement (NACK) message to the sender, explicitly requesting the retransmission of GOB 2. The sender got the NACK at time t2 and retransmitted GOB 2. The retransmitted GOB 2 arrived at time t3 which is after Frame 2, 3 and 4 were displayed but before Frame 5 was displayed. Due to transmission error and error propagation, Frame 2, 3 and 4 cannot be decoded correctly. However, instead of discarding Frame 2, 3 and 4, the decoder restored them using the retransmitted GOB 2 and then used them to restore Frame 5, which can be decoded and displayed without error.

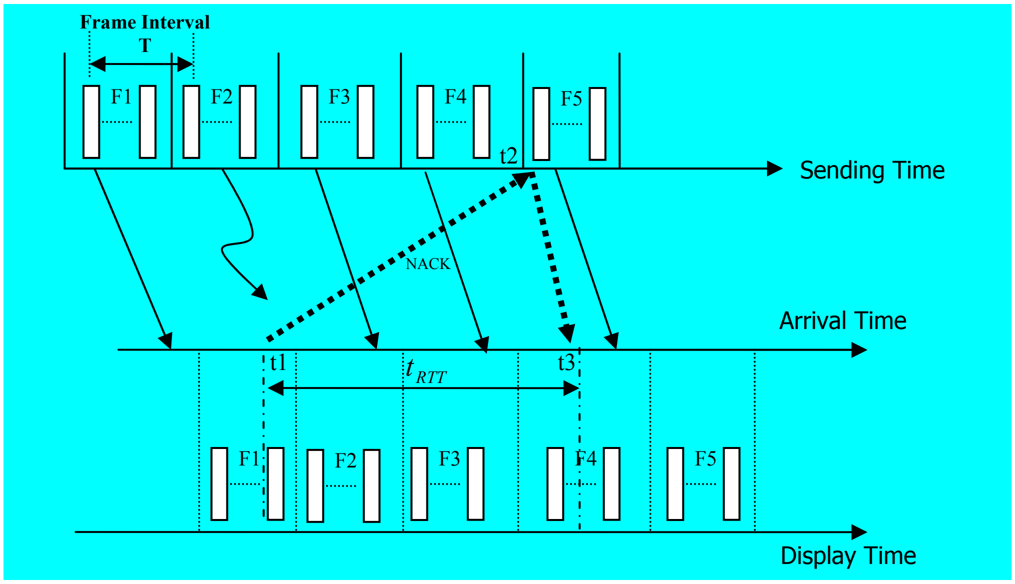

Figure 2.2 Illustration of retransmission scheme

2.2.2 Reference Picture Selection (RPS)

Reference Picture Selection (RPS) [15]-[17] is a feedback-based error control technique that uses information sent by the decoder to adjust the coding parameters at the encoder to achieve better error repair. With RPS, the encoder does not always pick the previous frame, but instead selects a previously-received, correctly-decoded frame as a reference when doing predictive encoding. RPS has two modes. In RPS negative acknowledgement (NACK) mode, when there is a transmission error, the decoder sends the encoder a NACK message with the number of a previously-received, correctly-decoded GOB as a reference for prediction. The encoder, upon receiving the NACK, uses the indicated correctly received GOB as a reference to encode the current GOB. In ACK mode, the decoder acknowledges all correctly received GOBs and the encoder only uses acknowledged GOBs as a reference. In NACK mode, only erroneously received GOBs are signaled by sending NACKs.

2.2.2.1 ACK Mode

In ACK mode, the decoder sends acknowledge messages for all correctly received GOBs and the encoder uses only the acknowledged GOBs as a reference. Due to the delay between decoder and encoder, the encoder has to use those intact GOBs, which are several frames before the current frame, as a reference. Thus, the accuracy of motion compensation prediction is impaired and the coding efficiency decreases, even if no transmission errors occur. Thus ACK mode performs best when the round-trip delay is short. On the other hand, error propagation is avoided entirely since only error-free pictures are used for prediction. Figure 2.3 illustrates the use of RPS with ACK mode. In this example, there are no transmission errors for the first 3 GOBs, allowing the encoder to receive an ACK for GOB 1 while encoding GOB 4. Thus, the encoder uses GOB 1 as a prediction reference to encode GOB 4. Similarly, the encoder uses GOB 2 as a reference for GOB 5, and GOB 3 as a reference for GOB 6. However, since no ACK is received for GOB 4, GOB 7 uses acknowledged GOB 3, instead of GOB 4, as the reference GOB. RPS ACK mode requires additional GOB buffers at the encoder and decoder to store previous GOBs to cover the maximum round-trip delay of ACK's. For instance, after encoding GOB 8, the encoder should store GOB 5, 6, 7 and 8.

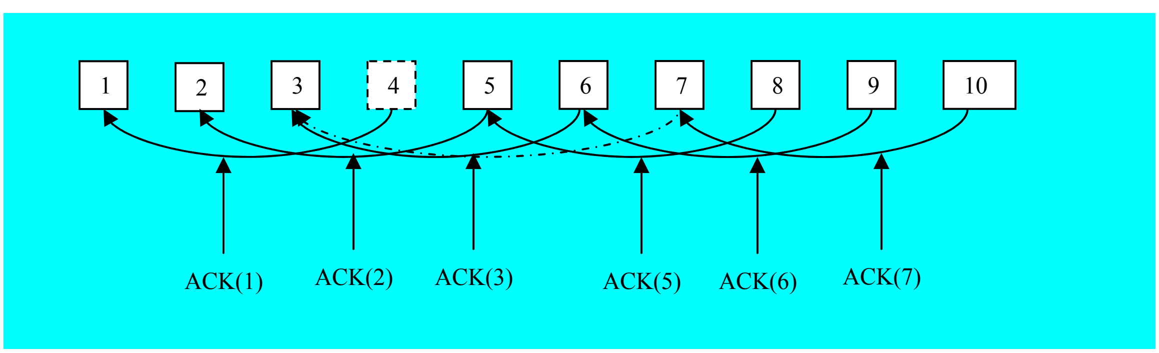

Figure 2.3. Illustration of the encoding of GOBs using RPS with ACK mode, where GOB 4 has a transmission error and the arrows indicate the selected reference pictures.

2.2.2.2 NACK Mode

In NACK mode, one of the GOBs in the previous frame is used as a reference during the error-free transmission. After a transmission error, the decoder sends a NACK for the erroneous GOB with an explicit request to use an older, intact GOB as a reference. As illustrated in Figure 2.4, when GOB 4 is determined to have a transmission error, the decoder sends a NACK to the encoder with an explicit request to use GOB 3, which has been decoded correctly, for prediction. Due to network latency, the NACK arrives back at the encoder only before GOB 7 is encoded. When the NACK arrives, the encoder then uses GOB 3 as the reference to encode GOB 7. Note, in the absence of receiving NACK messages, RPS NACK optimistically uses the most recently transmitted GOB as the reference for encoding. In NACK mode, the storage requirements of the decoder can be reduced to two GOB buffers. Compared to the ACK mode, the NACK mode can maintain better coding performance during error-free transmission. However, if a transmission error occurs, the error propagates for a period of one round-trip delay; that is, the time delay between the NACK being sent and the requested GOB being received.

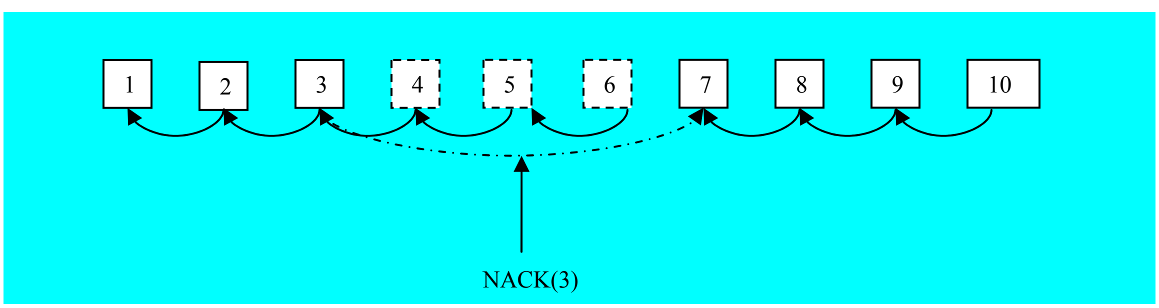

Figure 2.4. Illustration of the encoding of GOBs using RPS with NACK mode, where GOB 4 has a transmission error and the arrows indicate the selected reference pictures.

2.2.3 Intra Update

Similar to RPS with NACK mode, during error-free transmission, Intra Update [12] uses one of the GOBs in the previous frame as a reference. However, when it receives a NACK from the decoder, instead of using older, intact GOBs as a reference, Intra Update simply encodes the current GOB with intra mode. As illustrated in Figure 2.5, when the encoder receives a NACK from the decoder, it codes GOB 7 in intra mode to stop error propagation. But Intra coding reduces the coding efficiency and hence degrades the video quality under the same bit-rate constraint. If the encoder limits the use of Intra coding to macro-blocks that are severely distorted rather than the whole GOB, the coding efficiency can be greatly improved. The *Error Tracking* [12][49][50] approach uses intra mode for some macro-block's to stop inter GOB error propagation but limits its use to severely affected image regions only. Based on the information of an NACK, the encoder reconstructs the resulting error distribution in the current GOB by tracking the error propagation from a few GOBs back to the current GOB using a low complexity algorithm. If a macro-block is determined to be severely damaged, it will be coded in intra mode; otherwise local concealment is used to recover it.

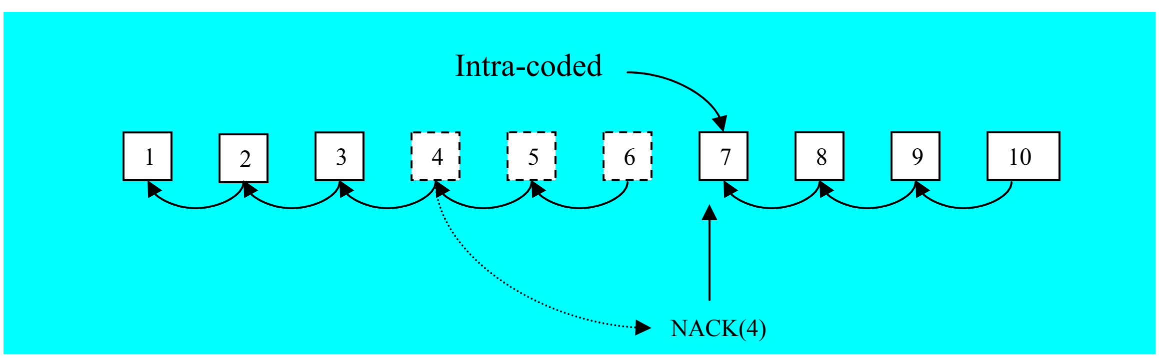

Figure 2.5. Illustration of the encoding GOBs using Intra Update, where GOB (4) is not received correctly and 5 and 6 cannot be decoded correctly.

2.3 Local Concealment

Local concealment is a media repair technique conducted at the decoder aimed at recovery of lost information of a damaged video frame due to transmission errors. The decoder can try to estimate the lost portions of a video frame based on the surrounding received blocks by making use of inherent correlation among spatially or temporally adjacent macro-blocks. There are three types of information that may need to be estimated in a damaged macro-block: the texture information, including the pixel or DCT coefficient values, the motion information, and the coding mode of the macro-block.

2.3.1 Recover Texture Information

The simplest way to recover texture information is by copying the corresponding macro-block in the previously decoded frame based on the motion vector for this damaged macro-block. This approach is referred as *Motion Compensated Temporal Prediction* (MCTP) [2]. The effectiveness of this local concealment technique depends largely on the recovery of the motion vector. Another simple local concealment technique to recover texture information is called *Temporal Interpolation* [24]. Temporal Interpolation interpolates pixels in a damaged block from pixels in adjacent correctly received blocks. Instead of interpolating individual pixels, a simpler approach is to estimate the DC coefficient (i.e. the mean value) of a damaged block and replace the damaged block by a constant equal to the estimated DC value. One way to facilitate such spatial interpolation is by an interleaved packetization mechanism so that the loss of one packet will damage only every other macro-block.

2.3.2 Recover Motion Vector

There are several simple methods to recover the lost motion vectors [26]. (a) assume the lost motion vectors to be zeros, which works well for video sequences with relatively small motion; (b) using the motion vectors of the corresponding block in the previous frame; (c) using the average of the motion vectors from spatially adjacent blocks; (d) using the median of motion vectors from the spatially adjacent blocks; (e) re-estimating the motion vectors. Typically, when a macro-block is damaged, its horizontally adjacent macro-blocks are also damaged, and hence the average or mean is taken over the motion vectors above and below. It has been found that the last two methods produce the best reconstruction results [29].

2.3.3 Recover Coding Mode

One way to estimate the coding mode for a damaged macro-block is by collecting the statistics of the coding mode pattern of adjacent macro-blocks, and finding a most likely mode given the modes of surrounding macro-blocks [25]. A simple and conservative approach is to assume that the macro-block is coded in the INTRA-mode, and use only spatial interpolation for recovering the underlying blocks [27].

2.4 H.264

As the state of the art in video compression standards, H.264 [18]-[22] is used throughout this thesis to encode/decode the video clips. H.264 is a video compression standard developed by ITU-T Video Coding Experts Group (VCEG) together with the ISO/IEC Moving Picture Experts Group (MPEG) [19]. H.264 supports a wide range of applications from low bit-rate Internet streaming to HDTV broadcast. H.264 is designed as a simple and straightforward video coding with enhanced compression performance and "network-friendly" video representation. H.264 has achieved a significant improvement in rate-distortion efficiency, providing a factor of two in bit-rate savings compared with MPEG-2 video, which is the most common standard used for video storage and transmission.

2.4.1 H.264 Data Structure

An H.264 picture is made up of macro-blocks (16x16 luminance samples and two corresponding 8x8 chrominance samples). In each image, macro-blocks are arranged in slices where a slice is a set of macro-blocks in raster scan order. In this thesis, a fixed number of successive macro-blocks in a slice are called a Group of Blocks (GOB). Macro-blocks themselves are classified as one of three types: Intra-coded (I), Predictive-coded (P) and Bidirectional predictive-coded (B). I macro-blocks are encoded independently of other macro-blocks and contain all information required to decode the macro-block. P macro-blocks are encoded using the previous I or P macro-block as a reference, allowing similarities between the successive blocks to be used for better compression. B macro-blocks further exploit motion compensation techniques by using motion information contained in the previous and following I or P macro-blocks. The encoder can select which previous block to use as a reference for motion-compensated prediction. However, as temporal distance for the reference block increases, coding efficiency tends to degrade as similarities between the encoding frame and the reference frame decrease. A P-block can be further divided into partitions, blocks of size 8x8, 16x8, 8x16 or 16x16 luminance blocks. These finer partitions can be used for motion-compensated prediction to achieve better prediction accuracy and, hence, better compression.

H.264 defines five types of slices, and a coded H.264 picture may be composed of different types of slices. I-slices contain only I macro-blocks, P-slices contain P and I macro-blocks, and B-slices contain B and I macro-blocks. SI (Switching I) slices contain SI macro-blocks, a special type of intra coded macro-block. SP (Switching P) slices contain P and I macro-blocks. SP slices are specially-coded slices that enable efficient switching between video streams and efficient random access for video decoders. SP slices are encoded in such a way that one slice in a sequence can be decoded using a motion-compensated reference picture from another sequence. SI slices are encoded without using a reference frame. If one bitstream is corrupted, the encoder can send an SI-frame to the decoder to stop the error propagation and switch to another stream.

2.4.2 H.264 Transport

In order to distinguish between coding specific features and transport-specific features, H.264 makes a distinction between a Video Coding Layer (VCL) and a Network Abstraction Layer (NAL). The output of the encoding process is VCL data which are mapped to NAL units prior to transmission and storage. Each NAL unit contains a Raw Byte Sequence Payload (RBSP), a set of data corresponding to coded video data or header information. In a packet-based network, each NAL unit may be carried in a separate packet and is organized into the correct sequence prior to decoding.

2.4.3 RPS in H.264

RPS can be used on whole pictures, picture segments (slices or GOBs), or on individual macro-blocks. The main difference between these schemes is the signaling in the bit-stream. In case of RPS operation on whole pictures or picture segments, the to-be-used reference picture information needs to be transmitted only once per picture or picture segment. When using macro-block-based RPS, every coded macro-block has to contain reference information, thereby yielding three-dimensional motion vectors (the reference picture time being the third dimension). RPS was first included in H.263 Annex N as an error repair tool [53][54]. By including multiple reference frames in the predictive coding loop, H.263 Annex N was designed to improve error repair as well as coding efficiency [54], but only supported per-picture or per-slice RPS. H.263 Annex U extends Annex N to support not only per-picture or per-slice RPS but also per-macro-block RPS. This enhanced reference picture selection mode was later subsumed into the H.264 video coding standard.

In applications that are based upon multicast or broadcast communication mechanisms, back channels may not be applicable. However, Reference Picture Selection may be used with or without a back channel with H.263 Annex N's sub-mode, known as Video Redundancy Coding (VRC). Since this thesis is focused on feedback-based media repair techniques, details of VRC are not discussed further.

When a back channel is used (as assumed in this thesis), it can be either multiplexed onto the H.263+ data stream in the opposite direction (the *VideoMux* back channel sub-mode), or conveyed out of band (the separate logical channel sub-mode). The VideoMux back channel sub-mode is only applicable for bi-directional video communication, because the back channel messages are conveyed within the video data in the opposite direction. The ITU-T Recommendation H.245 [56] defines dedicated messages to carry H.263+ back channel information and allows the encoder and decoder to build an out-of-band channel on which the decoder can return packet loss information. In particular, the decoder informs the encoder which pictures or parts of pictures have been incorrectly decoded. The H.245 information is convoyed using RTP/RTCP packets to be synchronized with the flow of real-time media. Recently ITU-T finalized Rec. H.271 [57] which defines syntax, semantics, and suggested encoder reaction to a video back channel message for all H.26X (including H.264) codecs. In particular, H.271 provides mechanisms for signalling a reference to a single lost slice of H.264 and signalling a reference to a suggested reference slice. The feedback messages according to H.271 are convoyed using RTP/RTCP or RTP/AVPF.

RPS requires additional frame buffers at the encoder and decoder to store enough previous frames to cover the maximum round-trip delay of NACK's or ACK's. In RPS NACK mode, the storage requirements of the decoder can be reduced to two frame buffers and if only error-free GOB's are displayed, one frame buffer is sufficient. In the RPS ACK mode no such storage reduction is possible. H.264 maintains a multi-picture buffer at both the encoder and decoder to enable multiple reference picture motion compensation for better coding efficiency, but the same buffers can be used for error repair. Two distinct picture buffering schemes with relative indexing are employed for efficient addressing of pictures in the multi-picture buffer. One is a sliding window in which most recent preceding (up to *M*) decoded and reconstructed pictures are stored and the other is adaptive memory control in which the pictures are inserted into and removed from the multi-picture buffer explicitly controlled by the encoder. In order to keep both reference buffers at the encoder and decoder synchronized transmit frame deletion instructions are transmitted from the encoder to the decoder. Such messages are sent using the memory management control operations defined in H.264. The decoder buffer follows the encoder buffer by acting on these instructions as specified by the encoder.

2.4.4 Local Concealment Techniques in H.264

The specific schemes suggested for the H.264/AVC standard in [28][30] involve intra and inter picture interpolations. The intra-frame interpolation scheme uses interpolation based on weighted average of boundary pixels. A lost pixel is deduced from boundary pixels of adjacent blocks. If there are at least two error-free blocks available in the spatial neighborhood, only those blocks are used in interpolation. Otherwise the surrounding "concealed" blocks are used.

For inter-frame interpolation based concealment, the recovery of lost motion vectors is critical. Like in spatial concealment, the motion vector interpolation exploits the close correlation between the lost block and its spatial neighbors. Since the motion of a small area is usually consistent, it is reasonable to predict the motion vector of a block from motion vectors of its neighboring blocks. However, the median or averaging over all neighbors' motion vectors does not necessarily give better results [28]. Therefore, the motion activity of the correctly received slice is first computed. If the average motion is less than a threshold (i.e., ¼ pixel), the lost block will be concealed by directly copying the co-located block from the reference frame; otherwise the motion vector recovery is done using the procedure described in [28]. Note that the selected motion vector should result in the minimum luminance change across the block boundary when the corresponding block of the previous frame replaces the lost block of the current frame.

2.4.5 Other Error Control Techniques in H.264

H.264 includes a number of features to aid the handling of transmission errors. Some of these error control features are incorporated in our analytical models, including: [22]

- The use of random intra macro-block refresh helps stop temporal error propagation and also avoids bit rate variations, such as *Macro-block Line Intra Update* (MLIU), where a group of blocks are intra coded every N frames.
- The use of slices helps improve robustness by stopping spatial error propagation. The macro-blocks belonging to a slice can be decoded independently from other slices since no inter-slice dependencies are allowed.
- Reference picture selection (RPS) allowing the encoder to select one of several previous frames that have been successfully decoded as a reference frame for prediction.

The following error control features are incorporated in our analytical models since they are not feedback-based error control techniques:

- *Flexible Macro-block Ordering* (FMO), wherein the sender can transmit macro-blocks in non-scan order, essentially aims at dealing with packet loss bursts and provides greater flexibility than does simple slice interleaving.
- SP-slices make use of motion-compensated predictive coding to exploit temporal redundancy in the sequences, like P-slices do. Unlike P-slices, however, SP-slice coding allows identical reconstruction of a slice even when different reference pictures are being used.

- Parameter Sets contain information that can be applied to a large number of coded pictures in a sequence, including picture id, the number of reference pictures in list 0 and 1 that may be used for prediction, etc. The parameter sets are separate from the coded slices themselves and may be sent to decoder well ahead of the slices that refer to them. The decoder can use the information contained in the parameter set to recover the lost macro-blocks.
- Data Partitioning allows the coded data that makes up a slice to be placed in three separate Data Partitions (A, B and C), each containing a subset of the coded slice. Each partition can be placed in a separate Network Abstraction Layer (NAL) unit and may therefore be transported separately. Partition B and C can be made to be independently decodable so the decoder may decode A and B only, or A and C only. Note that Partition A contains the slice header information and thus is highly sensitive to transmission errors.
- Redundant coded picture contains a redundant representation of part or all of a coded picture. If a primary coded picture is damaged, the decoder may replace the damaged area with decoded data from a redundant picture if available.

2.5 Video Buffering

The playout time of a video frame is defined as the time the frame is to be displayed at the receiver. Each frame must be delivered before and decoded by its playback time, and a frame that arrives after its decoding and display deadline is discarded. However, different applications may tolerate certain amount of playout delay depending on the characteristics of the applications. In interactive applications the playout delay is limited by the perceptual tolerance of the user, which is around 200 ms [14]. For applications such as Internet video streaming and broadcasting, the playout delay can be relaxed to a few seconds.

The playout buffer at the receiver stores the video packets before they are used for decoding. The use of playout buffer essentially relaxes playout delays and effectively extends the display deadlines for all video frames. The playout buffer provides a number of benefits: [47]

1. The buffer can be used to smooth the video stream and reduce the jitter introduced by changing network delays.
2. The extended display deadlines for the video frames by using playout buffer allow retransmission to take place when packets are lost. Our model shows the playout buffer can greatly improve the effectiveness of retransmission.
3. The use of buffering allows interleaving to transform possible burst loss in the channel into isolated losses, thereby enhancing the concealment of the subsequent losses.

2.6 Quality Scaling

In times of network congestion, some video frames have to be dropped either by routers or by applications to reduce the bandwidth or processing consumption. The dropping of frames by a router may seriously degrade multimedia quality since the encoding mechanisms for multimedia generally bring in numerous dependencies between frames. A multimedia application that is aware of these data dependencies can discard the frames that are the least important much more efficiently than can the router [31]. The adaptation of the data rate of a media stream to the capacity of the network is called *media scaling* [31][32].

Media scaling techniques for video can be broadly categorized as follows [32]:

- *Spatial scaling*: In spatial scaling, the size of the frames is reduced by encoding fewer pixels or by increasing the pixel size, thereby reducing the level of detail in the frame.
- *Temporal scaling*: In temporal scaling, the application drops frames. The order in which the frames are dropped depends upon the relative importance of the different frame types.
- *Quality scaling*: In quality scaling, the quantization levels are changed, chrominance is dropped or compression coefficients are dropped. The level of quantization determines the image quality. A large quantization step size can produce unacceptably large image distortion. Similarly, too fine a step size can lead to lower compression ratios. Thus, the quantization scale is a trade-off between quality and compression.

2.7 Video Quality Measurement

Since providing human subjects in statistically significant numbers for a user study to view and evaluate streamed videos is expensive and often impractical, several algorithms for predicting subjective video quality have been developed, among which the Peak-Signal-to-Noise-Ratio (PSNR) and Video Quality Metric (VQM) [23] are the most commonly used in the research community. Throughout this thesis, both PSNR and VQM are used to measure video quality.

2.7.1 PSNR

PSNR is derived by setting the mean squared error (MSE) in relation to the maximum possible value of the luminance (for a typical 8-bit value this is 2^8-1 = 255) as follows:

$$MSE = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left[(f(i,j) - F(i,j)\right]^2}{M \cdot N}$$

$$PSNR = 20 \cdot \log_{10}\left(\frac{255}{\sqrt{MSE}}\right)$$

Where f(i,j) is the original signal at pixel (i, j), F(i, j) is the reconstructed signal, and M x N is the picture size. The result is a single number in decibels, ranging from 30 to 40 for medium to high quality video.

It is well-known that PSNR does not necessarily accurately model perceptual quality. Despite this drawback, PSNR continues to be the most commonly used video quality metric in literature due to its simplicity. The previous published test results (e.g.[36][37]) showed that the performance of most objective video quality models are statistically equivalent to root mean squared error [36] and PSNR [37]. Therefore, PSNR is adopted as one of the video quality metrics in this thesis.

2.7.2 VQM

A second method to evaluate video quality is VQM [23], developed by the Institute for Telecommunication Science (ITS)[6]. VQM attempts to provide an objective measurement for perceived video quality by separately considering video impairment features that include blurring, uneven motion, global noise, and block and color distortion. VQM combines these component measures into a single metric, *D*, the video degradation (a value between 0 to 1) based on user studies. Results show VQM has a high correlation

[6] http://www.its.bldrdoc.gov/

with subjective video quality, leading to the adoption by ANSI of VQM as an objective video quality standard. Refinements to VQM can be computed using various models based on certain optimization criteria including: (1) television, (2) videoconferencing, (3) general, (4) developer, and (5) PSNR. This study uses the general model to evaluate video quality because the H.264 coding standard studied is increasingly used for a wide range of applications. The VQM general model uses a linear combination of seven parameters to determine video quality:

$$VQM = -0.2097 * si\_loss - 2.3416 * si\_gain + 0.5969 * hv\_loss + 0.2483 * hv\_gain$$
$$+ 0.0192 * chroma\_spread + 0.0076 * chroma\_extreme + 0.0431 * ct\_ati\_gain$$

Among these parameters, four parameters are based on spatial gradient features of the Y luminance component, two parameters are based on the two chrominance components (CB, CR) features, and one parameter is based on features from the contrast and absolute temporal information, both extracted from the Y luminance component.

The following list gives more details about these parameters:

- *si_loss* detects a decrease or loss of spatial information.
- *hv_loss* detects a shift of edges from horizontal and vertical orientation to diagonal orientation.
- *hv_gain* detects a shift of edges from diagonal to horizontal and vertical orientation.
- *si_gain* detects the improvements to quality that result from edge sharpening or enhancements.
- *chroma_spread* detects changes in the spread of the distribution of two-dimensional color samples.
- *chroma_extreme* detects severe localized color impairment.
- *ct_ati_gain* detects the amount of spatial details and the amount of motion.

Note, throughout the remainder of this thesis whenever VQM is discussed or graphed, we have defined VQM as 1–D to provide a direct comparison with other quality metrics, such as PSNR, where higher numbers denote better quality. Thus, using 1-D, a higher VQM value corresponds to a higher quality video image.

2.8 Summary

This chapter provides the background knowledge for building our analytical models. Our analytical models target the feedback-based error control techniques including Retransmission, Reference Picture Selection (RPS) and Intra Update. Since H.264 has been extended to include a variety of feedback-based error control mechanisms, we use H.264 as the video coding standard throughout this thesis. However, as long as we use the same coding standard throughout the entire measurement, the results should hold for different coding standards. Our model also uses "independent segment decoding" (ISD) mode of H.264 standard, and adopts quality scaling to adapt the data rate of a media stream to the capacity of the network. Some advanced error control features contained in H.264 standard, such as Flexible Macro-block Ordering (FMO), SP-slice etc., are not incorporated in our models but are instead possible future work.

Chapter 3

Related Work

This chapter surveys research work that is related to this thesis, seeking to compare existing work with our research work and discussing how the related work has contributed to building our models. Section 3.1 introduces the key error control and concealment techniques for video transmission. In particular this section focuses on feedback-based error control techniques, which is the major topic of this thesis. Section 3.2 presents research work on modeling error control techniques for video transmission. Section 3.3 summarizes this chapter.

3.1 Feedback-based Error Control for Video Transmission

There are various error control and concealment techniques that have been developed for video transmission. Wang et al. 0[2] provided an excellent survey on these techniques. According to their survey, the error control and concealment techniques for video transmission can be classified into three categories based on the roles which the encoder and decoder play: forward error control methods that add redundancy at the source to increase the robustness of the coded video streams; local error concealment techniques that recover the damaged areas based on characteristics of image and video signals at the decoder; and feedback-based error control techniques in which the encoder and decoder cooperate in the process of error concealment. Conceivably, since the encoder and decoder cooperate in the error control process, feedback-based error control techniques can achieve better performance than those error control techniques where only the encoder or decoder play the primary role. This thesis focuses on feedback-based error control techniques.

Feedback-based error control techniques use the feedback information from the decoder to adjust the coding parameters or vary the transport level control. At the transport level, a few Retransmission schemes are introduced. At the source coding level, a few techniques that adjust the coding parameters based on the feedback information from the decoder are described.

3.1.1 Retransmission

Retransmission can provide error repair without incurring much bandwidth overhead because packets are retransmitted only when they are determined lost. However, Retransmission of lost packets takes at least one additional round-trip time and thus may not be suitable for interactive video applications such as video conferencing that require short end-to-end delays [38]-[40]. Nevertheless Retransmission is still effective technique for improving error repair for real-time video applications. The reason is twofold. First, as the speed of network transmissions continues to improve, the additional round-trip delay incurred by Retransmission becomes acceptable for certain applications [41]. For instance, for applications such as Internet video streaming and broadcasting, the delay can be relaxed a few seconds so that several Retransmissions are possible. Second, in the past 10 years, several novel Retransmission schemes have been proposed to tackle the delay problem. These schemes show that careful Retransmission can greatly improve the error repair for video transmission without significantly increasing the delay. Papadopoulos and Parulkar [42] proposed an ARQ that combines selective repeat, Retransmission expiration, and conditional Retransmission. Their experiment in an ATM network showed the effectiveness of their scheme. Marasli et al [43] proposed an error control scheme using Retransmission over an unreliable network to achieve better service quality in terms of delay and loss rate. Smith [44] proposed a cyclical user datagram protocol, which places the base-layer packets of a layered coder in the front of the transmission queue to increase the number of Retransmission trials for the base layer. Instead of trying Retransmission indefinitely to recover a lost packet, as in TCP, the number of Retransmission trials is determined by the desired delay. Zhu [45] proposed a scheme called "Retransmission without waiting". With this scheme, instead of waiting for the retransmitted packet, the receiver recovers the damaged video part by using a chosen local concealment scheme. A trace of the affected pixels and their associated coding information (coding mode and motion vectors) is recorded. Upon the arrival of the retransmitted data, the affected pixels are corrected, so that they are reproduced as if no transmission loss had occurred. Zhu and Yao 0 proposed a prioritized multicopy Retransmission for a very lossy network environment. This scheme provides a flexible tradeoff between delay and reconstructed video quality. Feamster and Balakrishnan [14] leveraged the characteristics of MPEG-4 to selectively retransmit only the most important data in the bit stream. When latency constraints do not permit Retransmission, they used post-processing techniques at the receiver to recover the lost data. Rhee [13] proposed a Retransmission scheme, *called periodic temporal dependency distance* (PTDD) that does not require artificial extension of playout delays. In this scheme, frames are simply displayed at their normal playout times without any delay, as they are decoded. Thus, if a packet arrives after the playout time of its packet, the frame will be displayed with errors.

However, the “late” packet can be used to remove the error propagation. Rather than discarding the late packet, PTDD uses it to restore its frame although the frame has been displayed. Because the frame is used as a reference frame for its succeeding frames, restoring the reference frame stops error propagation. Our Retransmission scheme is similar to Rhee’s scheme which introduces no extra delay and is suitable for interactive media applications. Wei [78] proposes a prioritized retransmission mechanism to protect against the bursty packet losses in wireless LAN environments. Uchida et al [79] propose a proactive Retransmission scheme for hybrid FEC/ARQ. In the proposed scheme, a receiver periodically sends probe packets to a sender in order to check wireless channel state. If the sender does not receive a probe packet during a pre-specified interval, it regards the wireless channel as being in a state of burst loss and proactively retransmits packets expected to be lost during the burst loss period. The buffer management associated with layered video coding is also taken into consideration. Hou et al [80] propose a Differentiated Automatic Repeat Request (DARQ) scheme for MPEG video streaming over wireless links in which the inter-frame dependency and error propagation are jointly considered and a specific retransmission attempt is assigned to each frame in a Group of Pictures (GOP) according to its significance in the reconstruction of the video at the end-user.

3.1.2 Intra Update

After receiving the packet loss information from the decoder, instead of retransmitting the lost packets, the encoder can adapt the source-coding strategy to eliminate or reduce the effect of error propagation. One simple technique is that whenever the decoder detects an error, it sends a request to the encoder so that the next video frame will be encoded in intra mode. However, Intra coding reduces the compression gain and thus degrades the video quality given a fixed bit rate constraint. In order to reduce the bit-rate increase caused by Intra coding, only part of the image needs to be intra-coded due to the limited motion vector range [46][48]. To further improve the coding efficiency, Wada proposed two schemes to perform selective recovery using error concealment [49]. When a packet loss is detected, the decoder sends the identity information of damaged blocks to the encoder. At the same time, error concealment is performed on the damaged blocks. Given the identity information of the damaged blocks, the encoder either avoids using affected areas for prediction or conducts the same local concealment procedure on the damaged blocks as that performed at the decoder. A further refinement of Wada's selective recovery is an error tracking mechanism [12][49][50]. In the error tracking mechanism, the affected picture area is calculated from the point of damaged blocks up to the currently encoded frame through a low-complexity algorithm. Based on the severity of the distortion of an affected macro-block, the encoder can decide whether a macro-block is intra coded or not. Huang et al [75] propose a content-based adaptive intra block update method. It takes different intra block update rate on different image features. The vulnerability of each coded block to channel errors is measured through "error-sensitivity metric", and intra coding mode is carried out for appropriate macro block. Chen et al [76] and Chiou et al [77] propose a two-pass intra-refresh trans-coding scheme for inserting error-resilience features to a compressed video at the media gateway of a three-tier streaming system. The proposed trans-coder can adaptively vary the intra-refresh rate according to the video content and the channel's packet-loss rate to protect the most important macro-blocks against packet loss.

3.1.3 Reference Picture Selection

Rather than switching to intra mode at the encoder to stop inter-frame error propagation at the decoder, the encoder could also encode the current frame with reference to a previous frame that has been successfully decoded. The use of alternative reference frames in the predictive coding loop was first introduced in the standards in an optional annex to H.263 – Annex N called the Reference Picture Selection (RPS) [53][54] mode and in MPEG-4 as NEWPRED [60]. The idea was to make multiple references available at both encoder and decoder so that if feedback from decoder to encoder indicated a reception error both could switch to using a known good reference frame. Inter prediction using an older reference frame would still be more efficient than a complete intra update. Feedback could be in the form of positive or negative acknowledgement – so called ACK or NACK modes. H.263 – Annex U [55] extended this concept to include multiple reference frames in the predictive coding loop as a general coding efficiency tool and the management of the decoder reference store through the use of reference memory management control operation (MMCO) syntax. This enhanced reference picture selection mode was later subsumed into the latest H.264 video coding standard and is one of the reasons H.264 outperforms earlier standards such as MPEG-2. By incorporating H.263+ Annex U, H.264 supports RPS on a per picture, per slice, or per macroblock basis as an error-repair tool in the same way it is used in H.263 [61]. There are several modified versions of RPS proposed to increase coding efficiency in low bit-rate video codecs which use several or even many previous frames for predictions [62]-[64]. To accommodate a high round-trip time that could be many multiples of the frame period, RPS demands a high reference memory storage requirement. Limited reference memory RPS has been studied in [59]. Mulroy and Nilsson [65] proposed a time-windowed approach of reference frames with both encoder and decoder maintaining a subset of frames from specific time periods with respect to the current frame. As frames are coded, this reference subset can be managed so that there are always both very recent references for good compression efficiency and old references suitable for error recovery with high delay feedback. The key benefit of this approach is that the amount of reference picture memory required for reference picture selection is reduced for feedback channels suffering from high round-trip time. Tu and Steinbach [73] propose a framework for error robust, real-time video transmission over wireless networks. In their approach, downlink packet loss is coped with by retransmitting lost packets from the base station (BS) to the receiver for error recovery. Retransmissions are enabled by using fixed-distance reference picture selection during encoding with a prediction distance that corresponds to the round-trip-time between the BS and the receiver. Uplink transmission errors are dealt with by sending acknowledgements and predicting the next frame from the most recent frame that has been positively acknowledged by the BS. Wang et al [74] present three specific feedback based reference picture selection methods using flexible reference frames. In addition, a novel reference frame management method that enables using of flexible reference frame is proposed. The reference frame management method enables much simpler video codec implementations compared to the complex reference frame management methods in H.263 Annex U and H.264/AVC.

3.2 Modeling Error Control for Video Transmission

There are several research works on modeling error control schemes for streaming video transmission. These models aim to identify the optimal error control policy based on network conditions (such as packet loss probability and round-trip time) and application requirements (such as end-to-end delay). Mayer-Patel et al [66] developed an analytical model for predicting the reconstructed frame rate of an MPEG stream. Using this model along with TCP-friendly rate control, they explored the optimal FEC allocation decision as a function of loss rates and proposed an adaptive FEC scheme. Using a similar scheme, Wu et al. [67] derived a similar analytical model for predicting the playable frame rate in a TCP-Friendly MPEG stream with FEC. Based on this model, the variable space is searched to find the MPEG configuration that yields an optimal playable frame rate under the TCP-Friendly throughput constraint. Feamster developed an analytical model to derive the relationship between the packet loss rate and the observed frame rate. They then used this model to evaluate the effectiveness of a selective Retransmission scheme. Marasli et al. [68] used an analytical model to study Retransmission over an unreliable network. Their model showed that better service quality in terms of delay and loss can be achieved by using a limited number of retransmissions, rather than trying Retransmission indefinitely as in TCP. Zhai et al [69] used an analytical model, which is built upon an integrated joint source-channel coding (JSCC) framework, to study the performance of pure FEC, pure Retransmission, and their combination. A hybrid of FEC and Retransmission is shown to outperform each component individually due to its greater flexibility. Stuhlmüller et al [70] derived a theoretical model which covers the complete transmission system including the rate-distortion performance of the video encoder, forward error correction, interleaving, and the effect of error concealment and inter frame error propagation at the decoder. The channel model used is a 2-state Markov model describing burst errors on the symbol level. Liang et al [71][72] derived a model which estimates the expected mean-squared error distortion for different packet loss patterns. The model explicitly considers the effect of different loss patterns, including burst losses and separated (non-consecutive) losses, and accounts for inter-frame error propagation and the correlation between error frames. Based on this model, a packet interleaving scheme to combat the effect of bursty losses is proposed.

3.3 Summary

As described in Chapter 1, error repair techniques for video transfer can be categorized into three groups by whether the encoder plays the primary role, the decoder plays the primary role or both are involved in cooperation. Conceivably, since the encoder and decoder cooperate in the error control process, the feedback-based error control techniques can achieve better performance than those error control techniques where only the encoder or decoder play the primary role. Furthermore, the feedback-based error control schemes have been shown to be effective for a variety of video applications and have become standard features of the major video standards (such as MPEG and H.26x). Therefore, my thesis is focused on the feedback-based error control schemes, including Retransmission, RPS (NACK and ACK), and Intra Update.

Although researchers have developed analytical models to evaluate the performance of Retransmission and FEC, there are no existing analytical models to evaluate the performance of RPS. Furthermore, there are no existing analytical models to help systematically determine the optimal feedback-based error control scheme under different network conditions and application requirements. Our analytical models help fill this gap.

Chapter 4

Modeling of Feedback-Based Error Control Techniques for Video Transmission

This chapter derives analytical models for three major feedback-based error control techniques: Retransmission, Reference Pictures Selection (RPS), and Intra Update. The models aim at capturing the relationship between the video quality that can be achieved using these three error control techniques and various network characteristics including the packet loss rate, round-trip time and a capacity constraint. This chapter is arranged as follows: Section 4.1 derives the analytical model for retransmission; Section 4.2 for RPS in both ACK and NACK modes, and Section 4.3 for Intra Update.

The models target H.264 videos since this standard incorporates RPS and Intra Update, but can generally represent any video encoding technique that uses feedback-based error control techniques.

Our models make the following assumptions:

1. Each frame is encoded in the independent segment decoding (ISD) mode of H.264 where each GOB is encoded as an individual sub-video independently from other GOBs in the same frame, and the reference frame is selected on a per-GOB basis, i.e., for all macro-blocks within one GOB the same reference GOB is used. Since errors inside a GOB do not propagate to other GOBs, the video sequence can be partitioned into independent video sub-sequences. An independent video sub-sequence is referred to as a *reference chain*, illustrated in Figure 4.0. Without

the assumption of ISD, the dependency in our models still hold. However, it is difficult to measure the impact of reference distance on video quality if GOBs are within a frame.

2. Each GOB is carried in a single network packet. This is a reasonable assumption since the number of GOBs per frame can be adjusted so that each GOB can be fit into a network packet without fragmentation. Without this assumption, our models would have to be modified significantly since the loss probability of a GOB would be different than that of a network packet.
3. Transmission of feedback messages are reliable. This assumption is reasonable since feedback is usually not part of the video syntax and is transmitted via a separate network connection where control information is exchanged [12]. The feedback connection is assumed to not suffer from congestion as does the forward link carrying the video, or that the feedback connection uses retransmission or other methods to ensure reliable delivery. Furthermore, since the requirements for the transmission of feedback messages is modest compared to that for video transmission on the forward channel, bit rates for the feedback message can be neglected. Considering the loss of feedback messages can be incorporated into our existing models, but would increase complexity greatly.
4. Packet loss is independent with a random loss distribution. This is an assumption typically made in analytic models of a network and well represents many computer networks. However, in some network situations, packet loss

may be bursty, such as in wireless environments. Incorporating bursty loss requires fundamentally changing our current models.

5. Erroneously-decoded GOBs are repaired by local concealment and make no assumption on specific local concealment techniques.

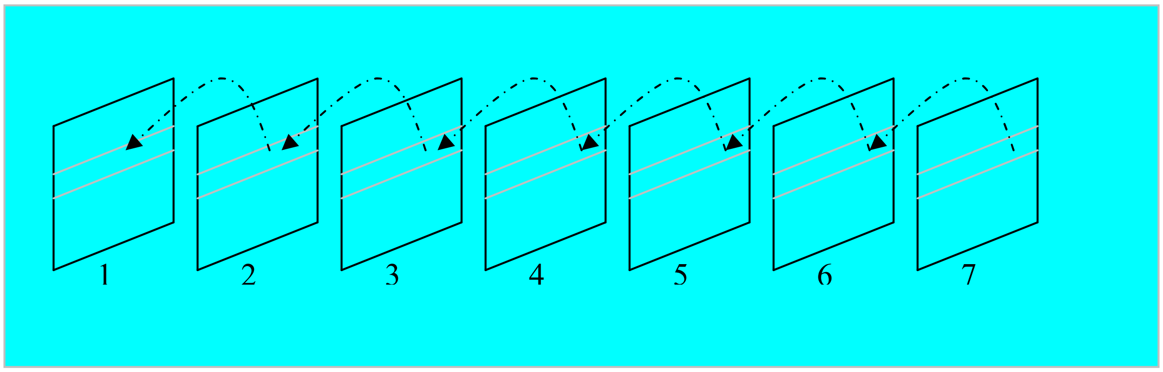

Figure 4.0 Illustration of a reference chain, where each rectangle represents a video frame, the area between two lines in each rectangle represents a group of macroblocks (GOB), and the arrows indicate the selections of reference-GOB.

4.1 Model Parameters

The parameters for our analytical models are categorized into system parameters and derived parameters, and system parameters are further categorized into two layers: encoder layer and transport layer.

Input Parameters (Encoder)	
R_F	Encoded frame rate (in frames per second or fps - typical full-motion video frame rates are 25-30 fps)
N_G	GOP size (in frames)
U_r	Average video quality[7] for a GOB that is encoded using a reference GOB that is *r* GOBs backward in the reference chain
U_0	Average video quality for an intra-coded GOB
U'	Average video quality for a GOB that is repaired using local concealment
Input Parameters (Transport)	
t_{RTT}	Round-trip time (in milliseconds)
p	Packet loss probability (fraction)
Derived Parameters	
t_{INT}	Time-interval between two frames (in milliseconds, so 40 msec. for 25 fps video)
q_n	Probability that the *n*-th GOB in reference chain is decoded successfully
$q_{n,r}$	Probability that the *n*-th GOB in the reference chain is decoded successfully using the *r*-th GOB as a reference
Q_n	Expected video quality value for *n*-th GOB in the reference chain

Table 4.1 Model parameters

Given the network capacity constraint and a specific video clip, the values for U_r and U_0 are obtained from our previous work [6]. The values for U' are obtained using fixed percentages of the best value of U_r. The actual percentage used is varied in the experiments.

In our analytical models, video quality refers to either PSNR or VQM. In other words, our models are independent of video quality metrics adopted.

[7] Video quality in terms of PSNR and VQM

4.2 Retransmission Modeling

This section derives an analytical model for predicting the video quality for a video sequence using retransmission over a lossy network.

4.2.1 Playout Time Constraint and Playout Buffer

A playout time of a GOB (packet) is defined as the time of the frame to which the GOB belongs is displayed at the receiver. A GOB is considered delayed if it does not arrive by its determined playout time. However, different applications may tolerate different amounts of playout delay depending on the characteristics of the applications. In interactive applications the playout delay is limited by the perceptual tolerance of the user, which is around 200 ms [2]. For applications such as Internet video streaming and broadcasting, the playout delay can be relaxed to a few seconds. Since our models are mainly applied to interactive media applications, the maximum playout delay (or playout buffering) is assumed to be around 200 ms. Since the retransmitted GOBs (packets) have to arrive before their designated playout time, the maximum number of retransmissions is determined by the playout delay and round-trip delay.

4.2.2 Full Retransmission

We first assume that every lost packet will be retransmitted and derive the average video quality for a GOB in a video frame.

4.2.2.1 Retransmission Range (RR)

Retransmission Range (*RR*) is defined as the distance between the current GOB and the closest reference GOB in the reference chain which can be retransmitted and arrive before the current GOB is played out. Figure 4.1 illustrates an example of *RR*. As shown in the figure, R1, R2, R3, R4 and R5 are a sequence of GOBs in the reference chain. During the transmission, R2 is lost, and the receiver detects the loss at time t1 and sends a retransmission request (NACK) to the sender. The sender gets the NACK at time t2 and retransmits the lost GOB. The retransmitted GOB arrives at time t3 which is before R5 is displayed but after R3 and R4 have been displayed. Therefore, R2, R3 and R4 cannot be decoded and displayed correctly. However, since R3, R4 and R5 are received successfully, the retransmitted R2 can be used to restore R3 and R4, and thus R5 can be decoded and displayed without error. R2 is the closest reference GOB in the reference chain which can be retransmitted before R5 is played out. Between R2 and R5, R3 and R4 are within the *RR*.

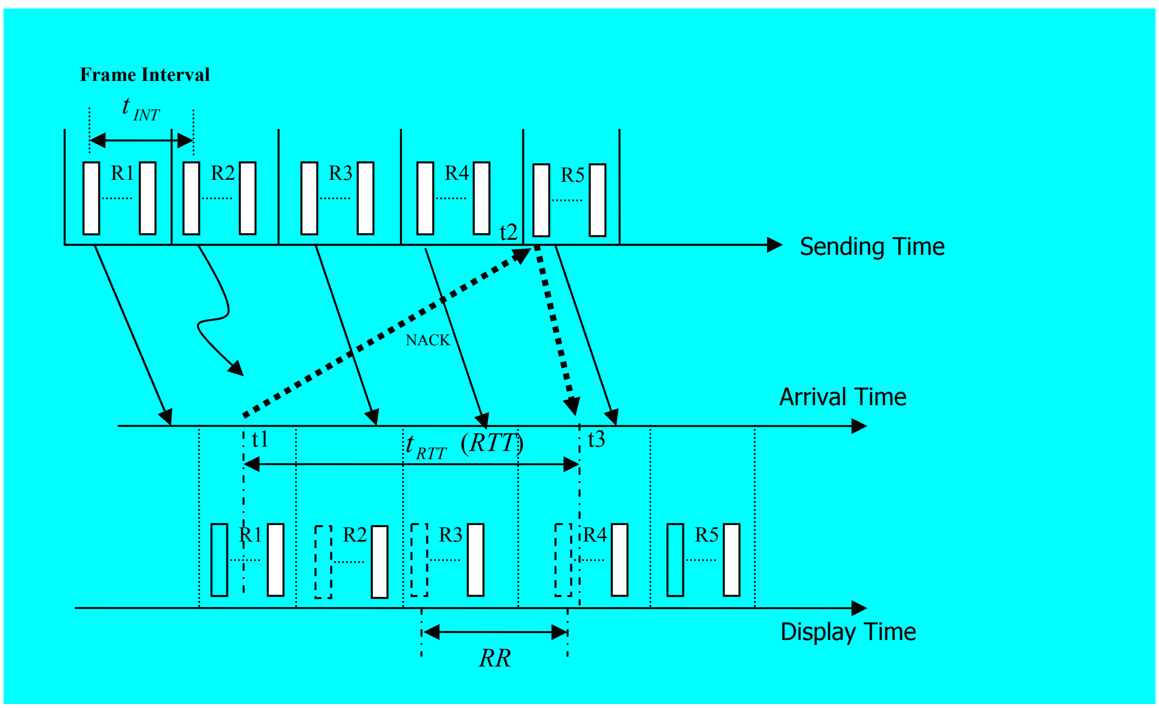

Figure 4.1 Illustration of Retransmission Range (RR), where each rectangle represents a GOB, and the rectangle with dashed-line indicates the GOB is either lost or cannot be decoded correctly due to error propagation.

Note that as long as all GOBs in *RR* (R3 and R4 in this example) and the current GOB (R5 in this example) are successfully received, and the GOB preceding *RR* (R2 in this example) can be retransmitted successfully if it is lost, the current GOB can be correctly decoded. *RR* is computed in number of GOBs as follows:

$$N_{RR} = \left\lfloor \frac{t_{RTT} - t_{buf}}{t_{INT}} \right\rfloor \tag{4.1}$$

Where t_{buf} accounts for the fact that introducing display buffering extends the display delay and thus reduces the effective round-trip delay.

4.2.2.2 Capacity Constraint

Given the bandwidth constraint *C*, we have this constraint:

$$\begin{gathered} C_E * [1 + (p + p^2 + p^3 + \cdots) * (1 - \frac{N_{RR}}{N_G})] = \frac{C_E * (N_G - N_{RR} p)}{(1-p) N_G} \le C, \\ C_E \le \frac{C * N_G (1-p)}{N_G - N_{RR} p} \end{gathered} \tag{4.2}$$

where, C_E is the maximum bandwidth allowed given the bandwidth constraint *C*; $(1 + p + p^2 + p^3 + \cdots)$ accounts for the multiple retransmission scenarios in the case the Retransmissions are lost and $(1 - \frac{N_{RR}}{N_G})$ accounts for the fact that the last (N_{RR} +1) frames of a GOP will not be retransmitted.

4.2.2.3 Achievable Video Quality

In the event that the first GOB in the reference chain is lost, retransmission cannot be used to recover any of the first (N_{RR} +1) GOBs in the reference chain because it is impossible for any of the retransmitted GOBs to arrive before their display times due to the round-trip delay. Since each subsequent GOB in the reference chain depends upon the success of the preceding GOBs, the probability of the n-th GOB in the reference chain being successfully decoded is:

$$q_n = q^n, \quad 1 \le n \le N_{RR} + 1 \qquad (4.3)$$

The erroneously-received GOBs are locally concealed. The average video quality for a GOB that is repaired using local concealment is denoted as $U^{'}$ and the average video quality for a GOB that uses the previous GOB in the reference chain as its reference GOB is denoted as U_1. Thus, the expected video quality for the n-th GOB in the first $(N_{RR} + 1)$ GOBs in the reference chain is:

$$Q_n = U_1 q^n + U^{'}(1 - q^n), \quad 1 \le n \le N_{RR} + 1 \qquad (4.4)$$

For the remaining GOBs in the reference chain, the successful decoding of a GOB depends on the successful receipt of all the GOBs within its *Retransmission Range (RR)* because the other GOBs that precede its *RR* can be retransmitted before its display time. Thus, the probability of the n-th GOB in the reference chain being successfully decoded is:

$$q_n = q^{N_{RR}+1}, \quad n \ge N_{RR} + 2 \qquad (4.5)$$

And the expected video quality for the n-th GOB is:

$$Q_n = U_1 q^{N_{RR}+1} + U^{'}(1 - q^{N_{RR}+1}), \qquad n \ge N_{RR} + 2 \quad (4.6)$$

In summary, the expected PSNR for a GOB(n):

$$Q_n = \begin{cases} U_1 q^n + U^{'}(1 - q^n), & 1 \le n \le N_{RR} + 1 \\ U_1 q^{N_{RR}+1} + U^{'}(1 - q^{N_{RR}+1}), & n \ge N_{RR} + 2 \end{cases}$$

The average video quality over a GOP can be computed as follows:

$$Q = \frac{\sum_{i=1}^{N_{RR}+1} (U_1 q^i + U^{'}(1-q^i)) + \sum_{i=N_{RR}+2}^{N_G} (U_1 q^{N_{RR}+1} + U^{'}(1-q^{N_{RR}+1}))}{N_G}$$
$$= (U_1 - U^{'})(N_G - N_{RR} - \frac{1}{1-q}) q^{N_{RR}+1} + U^{'} N_G + \frac{q}{1-q}(U_1 - U^{'}) \quad (4.7)$$

4.2.3 Partial Retransmission

When the channel capacity is extremely constrained, retransmission of every lost packet may not be feasible. Instead, only a fraction (p_r) of lost packets will be retransmitted, which we define as *Partial Retransmission*. We next derive the average achievable video quality for a GOB in a video frame with partial retransmission.

4.2.3.1 Retransmission Range

Whether a lost GOB will be retransmitted or not is determined on a GOP-by-GOP basis. In other words, the reference chain is limited to within a single GOP. Given the fraction of lost packets that could be retransmitted (p_r) and the fact that the early GOBs in the reference chain should be given higher priority for retransmission over the later ones, the average number of GOBs that are retransmitted is:

$$N_R = \min(\lceil N_G * p_r \rceil + N_{RR} + 1, N_G) \quad (4.8)$$

4.2.3.2 Capacity Constraint

With partial retransmission, the bandwidth constraint becomes:

$$C_E + C_E * p_r * (p + p^2 + p^3 + \cdots) \le C,$$
$$C_E \le \frac{C(1-p)}{1 - p(1-p_r)} \quad (4.9)$$

C_E is the maximum bandwidth allowed given the bandwidth constraint *C*.

4.2.3.3 Achievable Video Quality

For the first $N_{RR}+1$ GOBs in the reference chain, the average video quality is the same as that of full retransmission:

$$Q_n = U_1 q^n + U^{'}(1-q^n), \quad 1 \le n \le N_{RR}+1 \qquad (4.10)$$

For the GOBs in the frames between $N_{RR}+2$ and N_R, they will be retransmitted if they are lost. Thus, the expected video quality can be computed using (4.6):

$$Q_n = U_1 q^{N_{RR}+1} + U^{'}(1-q^{N_{RR}+1}), \qquad N_{RR}+2 \le n \le N_R \qquad (4.11)$$

The remaining GOBs in the reference chain (within a GOP) will not be retransmitted if any of them are lost because there is no more retransmission left in the "budget". Therefore, the probability of the n-th GOB in the remaining portion of the GOP being successfully decoded is:

$$q_n = q^{N_{RR}+1} q^{n-N_R}, \quad N_R+1 \le n \le N_G$$

The expected video quality for the n-th GOB in the last portion of the GOP is:

$$Q_n = U_1 q^{N_{RR}+1} q^{n-N_R} + U^{'}(1-q^{N_{RR}+1} q^{n-N_R}), \qquad N_R+1 \le n \le N_G \qquad (4.12)$$

In summary, the expected video quality for a GOB (n) with partial retransmissions:

$$Q_n = \begin{cases} U_1 q^n + U^{'}(1-q^n), & 1 \le n \le N_{RR}+1 \\ U_1 q^{N_{RR}+1} + U^{'}(1-q^{N_{RR}+1}), & N_{RR}+2 \le n \le N_R \\ U_1 q^{N_{RR}+1} q^{n-N_R} + U^{'}(1-q^{N_{RR}+1} q^{n-N_R}), & N_R+1 \le n \le N_G \end{cases} \qquad (4.13)$$

The average video quality over a GOP can be computed as follows:

$$Q = \frac{1}{N_G}\left(\begin{array}{l}\sum_{i=1}^{N_{RR}+1}(U_1 q^i + U^{'}(1-q^i)) + \sum_{i=N_{RR}+2}^{N_R}(U_1 q^{N_{RR}+1} + U^{'}(1-q^{N_{RR}+1})) \\ + \sum_{i=N_R+1}^{N_G}(U_1 q^{N_{RR}+1} q^{i-N_R} + U^{'}(1-q^{N_{RR}+1} q^{i-N_R}))\end{array}\right) \quad (4.14)$$

4.3 Reference Pictures Selection (RPS) Modeling

This section derives the analytical models for Reference Picture Selection. Section 4.3.1 describes the model for RPS ACK; and Section 4.3.2 for RPS NACK.

4.3.1 Analytical Model for RPS ACK

RPS ACK uses acknowledged GOBs as references. Since it takes at least one round-trip time for the encoder to receive an ACK for a GOB, the current GOB has to use a GOB which was captured at least δ[8] GOBs before it as a reference. The age of the GOB selected as a reference GOB grows linearly with the length of the round-trip time. When the encoder uses an older reference GOB, video quality is inherently lowered given the network capacity constraint. As long as GOB *n* is successfully received, it can be decoded successfully since it can use any previously-acknowledged GOB as a reference. Therefore, the probability of GOB *n* being successfully decoded is:

$$q_n = 1 - p \quad (4.15)$$

Since the encoder selects the last GOB available without errors at the decoder as a reference, the reference GOB for GOB *n* could be chosen from GOB 1 up to GOB (*n*-*δ*). The probability of decoding GOB *n* correctly using GOB (*n*-δ-*i)* as a reference is:

---

[8] $\delta = \left\lceil \frac{t_{RTT}}{t_{INT}} \right\rceil$

$$(1-p)p^i q_{n-\delta-i}, \qquad 0 \le i \le n-\delta-1 \qquad (4.16)$$

where $q_{n-\delta-i}$ is the probability of GOB (n-δ-i) being successfully decoded, p^i is the probability of i consecutive GOBs (proceeding the GOB (n-δ)) having transmission errors and (1-p) is the probability of GOB (n) being successfully received.

The use of older reference frames for prediction degrades the effectiveness of compression for a GOB. In order to maintain a constant frame rate and bit rate, the encoder thus has to use a coarser quantizer and the overall video quality may decrease. To account for the video quality degradation due to the use of older reference GOBs for prediction, U_r denotes the average video quality for a GOB n whose reference GOB is r GOBs back in the reference chain.

The expected video quality for n-th GOB is as follows:

$$Q_n = \begin{cases} (1-p)\sum_{i=0}^{n-\delta-1} U_{\delta+i} p^i q_{n-\delta-i} + p * U', & n > \delta \\ (1-p)U_0 + p * U', & n \le \delta \end{cases} \qquad (4.17)$$

where U' denotes the average video quality for a locally concealed GOB and U_0 the average video quality for an intra-coded GOB. Note that the first δ GOBs of a GOP have to be encoded in intra mode since no ACK messages from the decoder will be received prior to encoding.

Since $q_{n-\delta-i}$ is a constant (1-p), equation (4.17) can be further simplified as follows:

$$Q_n = \begin{cases} (1-p)^2 \sum_{i=0}^{n-\delta-1} U_{\delta+i} p^i + p * U', & n > \delta \\ (1-p)U_0 + p * U', & n \le \delta \end{cases} \qquad (4.18)$$

The average video quality over a GOP can be computed as follows:

$$
\begin{aligned}
Q &= \frac{1}{N_G}\left(\sum_{n=1}^{\delta}((1-p)U_0 + pU^{'}) + \sum_{n=\delta+1}^{N_G}((1-p)^2\sum_{i=0}^{n-\delta-1}U_{\delta+i}p^i + pU^{'})\right) \\
&= \frac{1}{N_G}\left(((1-p)U_0 + pU^{'})\delta + \sum_{n=\delta+1}^{N_G}((1-p)^2\sum_{i=0}^{n-\delta-1}U_{\delta+i}p^i + pU^{'})\right)
\end{aligned}
\tag{4.19}
$$

4.3.2 Analytical Model for RPS NACK

For NACK mode, one of the GOBs in the previous frame is used as a reference during the error-free transmission. After a transmission error, the decoder sends a NACK for the erroneous GOB with an explicit request to use older, intact GOBs as a reference. Therefore, the encoder may use a GOB in the previous frame or one in an older frame as a reference to encode the current GOB (n) depending upon whether it receives a NACK from the decoder or not. If it does not receive a NACK from the decoder, it uses a GOB in the previous frame as a reference. The probability of correctly decoding GOB (n) using a GOB in the previous frame as reference is denoted as $q_{n,1}$, where 1 indicates using the preceding GOB in the reference chain as a reference. If the encoder does receive a NACK, it uses the GOB requested by the encoder as a reference. As in ACK mode, the reference GOB for GOB (n) could be chosen from GOB (1) up to GOB (n-δ) depending upon which GOB is the last correctly decoded GOB at the decoder. $q_{n,\delta+i}$ ($0 \le i \le n-\delta-1$) denotes the probability of decoding GOB (n) correctly using GOB (n- δ -i) as a reference. Since any of the first δ GOBs cannot receive a NACK before being encoded, the successful decoding of each subsequent GOB depends upon the success of the preceding GOBs. Therefore, the probability of GOB (n) being successfully decoded is as follows:

$$
q_n = \begin{cases} q_{n,1} + \sum_{i=0}^{n-\delta-1} q_{n,\delta+i}, & n > \delta \\ (1-p)^n, & n \le \delta \end{cases}
\tag{4.20}
$$

The expected video quality for GOB (n):

$$Q_n = \begin{cases} U_1 q_{n,1} + \sum_{i=0}^{n-\delta-1} U_{\delta+i} q_{n,\delta+i} + (1-q_n)U', & n > \delta \\ (1-p)^n U_1 + (1-(1-p)^n)U', & 1 < n \le \delta \\ \left(1-p\right)*U_0 + p*\big)U', & n = 1 \end{cases} \quad (4.21)$$

The average video quality over a GOP can be computed as follows:

$$Q = \frac{1}{N_G}\left(\sum_{n=1}^{\delta} ((1-p)^n U_1 + (1-(1-p)^n)U') + \sum_{n=\delta+1}^{N_G} (U_1 q_{n,1} + \sum_{i=0}^{n-\delta-1} U_{\delta+i} q_{n,\delta+i} + (1-q_n)U') \right) \quad (4.22)$$

4.3.2.1 GOB Dependency Modeling

To estimate $q_{n,1}$ and $q_{n,\delta+i}$ ($0 \le i \le n-\delta-1$), it is essential to model the prediction dependency between GOBs in the reference chain. A binary tree is used to model GOB dependency for RPS with NACK mode. Two input parameters are required to build the dependency tree: packet loss probability (p) and round-trip time (δ). Figure 4.2 illustrates a binary tree for the possible decoded versions of a GOB and the corresponding reference GOB selections while using RPS with NACK mode[9]. In the illustrated example, there are four GOBs and the round-trip time equals the length of time to capture two GOBs.

[9] The first p branch and its descendants are not shown because there are no correctly decoded GOBs under this branch.

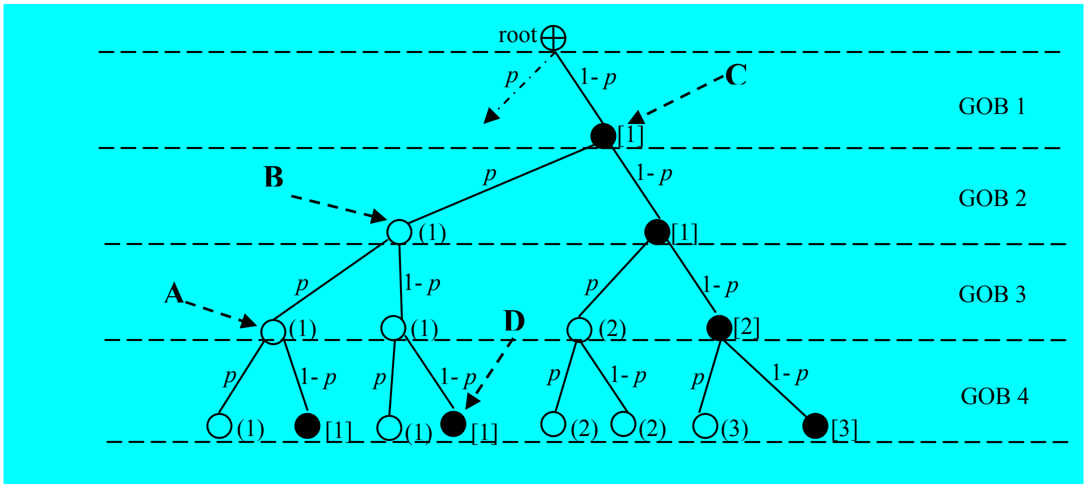

Figure 4.2 Binary tree for the possible decoded versions of a GOB with RPS with NACK mode

A node in the tree represents a decoded version of a GOB in a video frame. The nodes with hollow circles are those decoded erroneously while those with solid circles are those decoded correctly. Branches leaving a node represent the two cases that either a packet[10] is received erroneously with probability p or received correctly with probability (1-p). For a node whose decoded status is "Erroneous", the number (in parenthesis) besides the node represents the last GOB that has been decoded correctly at the decoder when entering this node; for a node with "Correct" decoded status, the number (in brackets) represents its reference GOB number. The GOB number between two dashed lines represents the GOB in transmission. Note that the root node (labeled with a crossed circle) represents an intra-coded GOB. Therefore, each time the encoder intra-codes a GOB, the binary tree is refreshed.

To illustrate how the decoded status of a node is decided, consider node A (GOB 3). Since the ancestor of node A (node C) is received correctly, node A (GOB 3) did not receive a NACK message from the decoder and thus used its parent (node B, GOB 2) as a

[10] As stated earlier, our model assumes each packet contains one GOB.

reference. Since the decoded status of node B is "Erroneous", the decoded status of node A is "Erroneous" as well. Upon entering node A, the last correctly decoded GOB at the decoder is GOB 1; therefore, the number (in parenthesis) besides node A is 1. Next, consider node D (GOB 4). Since D's ancestor 2 frames back (node B, GOB 2) is received erroneously (under the *p* branch), node B receives a NACK message from the decoder, which explicitly requires using GOB 1 as a reference. Thus, the reference GOB for node B is GOB 1. Furthermore, node B is under the (1-*p*) branch to indicate GOB 4 was correctly received. Therefore, the decoded status of node B is "Correct".

4.3.2.2 GOB Dependency Tree Creation

Each node in the GOB dependency tree contains the following information:

- *Decoded Status*: Correct or Erroneous
- *Probability of occurrence of this node* (decoded version of a GOB)
- *The latest GOB that has been decoded correctly* (LDC [11] for short) *at the decoder when entering this node* - this information is recorded only for a node with "Erroneous" decoded status.
- *Reference GOB used for motion compensation prediction* - this information is recorded only for a node with "Correct" decoded status.
- *GOB number in the reference chain*

The creation of the GOB dependency tree begins with a successfully decoded intra GOB. Since no correctly decoded GOBs are under the *p* branch of the root node, that part of the tree can be ignored. For each node, the four parameters described above are determined using the following algorithm:

1. If this node is under the *p* branch:

[11] Latest Decoded Correctly

- Its decoded status is set to “Erroneous”;
- Its probability is set to: $p * parent \rightarrow prob$, where $parent \rightarrow prob$ is its parent’s probability;
- Its LDC is determined based upon its parent’s decoded status: if its parent is decoded correctly, its LDC is set to its parent’s GOB number; otherwise it is set to its parent’s LDC;
- Its GOB number is set to its parent’s GOB number plus one.

2. If this node is under the (1- p) branch:
   - Its decoded status is determined based upon whether its ancestor δ frames back is received correctly and the decoded status of its parent. If its ancestor δ frames back was received correctly (no NACK), then check its parent’s decoded status; if its parent is decoded correctly, its decoded status is set to “Correct”, otherwise it is set to “Erroneous”. If its ancestor δ frames back was NOT received correctly, its decoded status is set to “correct” since it received a NACK message from the decoder and used an older, correctly decoded GOB as a reference.
   - Its probability is set to: $(1-p) * parent \rightarrow prob$, where $parent \rightarrow prob$ is its parent’s probability;
   - If its decoded status is “Correct”, its reference GOB is determined based upon the decoded status of its ancestor δ frames back. If its ancestor δ frames back was received correctly (no NACK), its reference GOB is set to its parent’s GOB number. If its ancestor δ frames back was NOT received correctly, its reference GOB is set to its ancestor’s LDC.

- If its decoded status is “Erroneous”, its LDC is determined based upon the decoded status of its parent. If its parent was decoded correctly, its LDC is set to its parent’s GOB number; otherwise it is set to its parent’s LDC.
- Its GOB number is set to its parent’s GOB number plus one.

4.3.2.3 Estimate of $q_{n,r}$ using the GOB Dependency Tree

After building the GOB dependency tree, $q_{n,r}$ is estimated in two steps. First, the GOB dependency tree is traversed to find all “Correct” nodes with GOB number equal n and reference GOB number (n-r). Then the probabilities of each node from step 1 are added together to produce an estimate for $q_{n,r}$.

4.4 Intra Update Modeling

This section presents the analytical model for Intra Update. Similar to RPS with NACK mode, during error-free transmission, Intra Update uses one of the GOBs in the previous frame as a reference. However, when it receives a NACK from the decoder, instead of using an older, correctly decoded GOB as a reference, Intra Update simply encodes the current GOB with intra mode.

Without receiving a NACK from the decoder, the encoder uses the previous GOB as a reference to encode the current GOB. The probability of decoding GOB n correctly using previous GOB as a reference is denoted as $q_{n,1}$. Upon receiving a NACK, the current

GOB is intra coded. The probability of decoding GOB (n) correctly using Intra coding is denoted as $q_{n,INTRA}$. Therefore, the probability of GOB n being successfully decoded is:

$$q_n = \begin{cases} q_{n,1} + q_{n,INTRA}, & n > \delta \\ (1-p)^n, & n \le \delta \end{cases} \qquad (4.23)$$

The expected video quality for GOB n:

$$Q_n = \begin{cases} q_{n,1}U_1 + q_{n,INTRA}U_0 + (1-q_n)U', & n > \delta \\ (1-p)^n U_1 + (1-q_n)U', & n \le \delta \end{cases} \qquad (4.24)$$

Where, U_0 is the average video quality for an intra-coded GOB (n).

The average video quality over a GOP can be computed as follows:

$$Q = \frac{1}{N_G}\left(\sum_{n=1}^{\delta}((1-p)^n U_1 + (1-q_n)U') + \sum_{n=\delta+1}^{N_G}(q_{n,1}U_1 + q_{n,INTRA}U_0 + (1-q_n)U') \right) \quad (4.25)$$

4.4.1 GOB Dependency Tree Creation

In order to estimate $q_{n,1}$ and $q_{n,INTRA}$, an approach similar to estimate $q_{n,1}$ for RPS with NACK mode is adopted. The first step is to create the GOB dependency tree. Figure 4.3 illustrates a binary tree for the possible decoded versions of a GOB using Intra Update. In the illustrated example, there are 4 GOBs and the round-trip time delay equals the length of time to capture 2 GOBs.

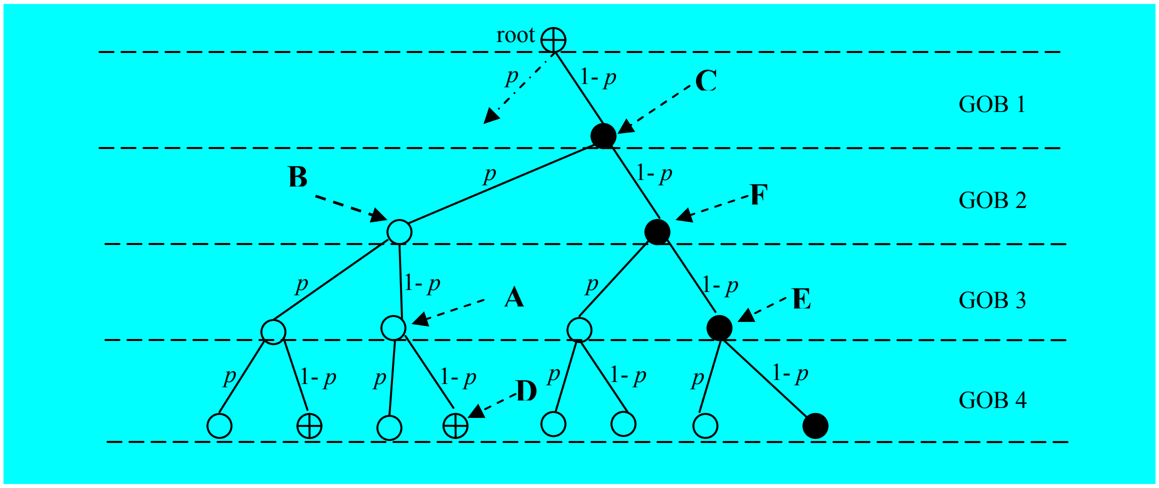

Figure 4.3. Binary tree for the possible decoded versions of a GOB using Intra Update

A node in the tree represents a decoded version of a GOB in a video frame. The nodes with hollow circles are those decoded erroneously, solid circles are those decoded correctly using the previous frame as a reference and crossed circles are those intra-coded. Branches leaving a node represent two cases that either a packet (GOB) is received erroneously with probability p or correctly received with probability (1-p). The GOB number between two dashed lines represents the GOB in transmission. Note that the root node represents an intra-coded GOB. Therefore, each time the encoder intra-codes a GOB, the binary tree is refreshed.

To illustrate how the decoded status of a node is decided, consider node "A" (GOB 3). Since the ancestor of node "A" (node "C", GOB 1) was received correctly, node "A" did not receive any NACK message from the decoder and thus used its parent (node "B", GOB 2) as a reference. Since the decoded status of node "B" is "Erroneous", the decoded status of node "A" is "Erroneous" as well. Next, consider node "D" (GOB 4). Since its ancestor 2 frames back (node "B", GOB 2) was received erroneously (under p branch), node "D" receives a NACK message from the decoder and thus is intra-coded. Also node

"D" is under the (1-*p*) branch, which indicates (GOB 4) is correctly received; therefore, the decoded status of node "D" is "Intra-Coded". Last, consider node "E" (GOB 3). Since its ancestor 2 frames back (node "C", GOB 1) is received correctly, node "E" does not receive any NACK message from the decoder and thus uses its parent (node "F", GOB 2), which was decoded correctly, as a reference. Also, node "E" is under the (1-*p*) branch, therefore its decoded status is "Correct".

Each node in the GOB dependency tree contains the following information:

- *Decoded Status:* Erroneous, Intra-Coded and Correct
- *Probability of occurrence of this node (decoded version of a GOB)*

The creation of the GOB dependency tree begins with the root node. Since no correctly decoded GOBs are under the *p* branch of the root node, we can simply ignore that part of the tree.

For each node, the four parameters described above are determined using the following algorithm:

1. If this node is under the *p* branch:
   - Its decoded status is set to Erroneous;
   - Its probability is set to: $p * parent \rightarrow prob$, where $parent \rightarrow prob$ is its parent's probability;
2. If this node is under the (1- *p*) branch:
   - Its decoded status is determined based upon whether its ancestor δ frames back is received correctly and its parent's decoded status. If its ancestor δ frames back is received correctly or intra coded, then check its parent's decoded status; if its parent's decoded status is either Intra-Coded or Correct, its decoded status is set to

Correct otherwise is set to Erroneous - it did not receive any NACK message from the decoder and uses its parent, whose decoded status is Erroneous, as a reference. If its ancestor δ frames back is received erroneously, its decoded status is set to Intra-Coded since it receives a NACK message from decoder.

- Its probability is set to: $(1-p) * parent \rightarrow prob$, where $parent \rightarrow prob$ is its parent's probability;

4.4.2 Estimate of $q_{n,1}$ and $q_{n,INTRA}$ using the GOB Dependency Tree

After building the GOB dependency tree, $q_{n,1}$ is estimated in two steps. First, the GOB dependency tree is traversed to find all "Correct" nodes with GOB number equal *n*. Then the probabilities of each node from step 1 are added together to produce an estimate for $q_{n,1}$.

Similarly, to estimate $q_{n,INTRA}$, the GOB dependency tree is traversed to find all "Intra-Coded" nodes with GOB number equal *n*. Then the probabilities of each node from step 1 are added together to produce an estimate for $q_{n,INTRA}$.

Chapter 5

Impact of Reference Distance for Motion Compensation Prediction on Video Quality

This chapter provides a systematic study of the effects of reference distance on video quality for a range of video coding conditions. High-quality videos with a wide variety of scene complexity and motion characteristics are selected for baseline encoding. The videos are all encoded using H.264 with a bandwidth constraint and a range of reference distances. Two objective measures of video quality are used, the popular Peak Signal to Noise Ratio (PSNR), and the reportedly more accurate Video Quality Metric (VQM). Section 5.1 provides the hypothesis for this study; Section 5.2 describes the methodology used to test the hypothesis; Section 5.3 analyzes the experimental results; and Section 5.4 draws conclusions.

5.1 Hypothesis

As the reference distance increases, the coding efficiency decreases since the similarities between the current frame and the reference frame decreases. If the network capacity is constrained, the video quality degrades as the coding efficiency drops. The degree of the coding efficiency degradation is affected by the video content. For instance, if a video sequence contains high motion scenes, then the similarities among adjacent frames are low. Thus, there are more macro-blocks within the video that must be intra encoded. On the other hand, if a video sequence contains low motion scenes, it is more likely the macro-blocks within the video can be inter-coded using motion-compensation predictions since the similarities among frames are high. Since the intra-coded macro-blocks are independent of the reference frames, the coding efficiency for pictures containing more intra-coded macro-blocks (high motion) degrades less with an increase in reference distance than those containing more inter-coded macro-blocks (low motion). Figure 5.1 depicts our hypothesis of the relationship between video quality and reference distance for videos with high motion and low motion. As shown in the figure, as the reference distance increases, the video quality degrades as the coding efficiency drops for both high-motion and low-motion videos. However, the video quality for high-motion videos degrades slower than those low-motion videos as the reference distance increases.

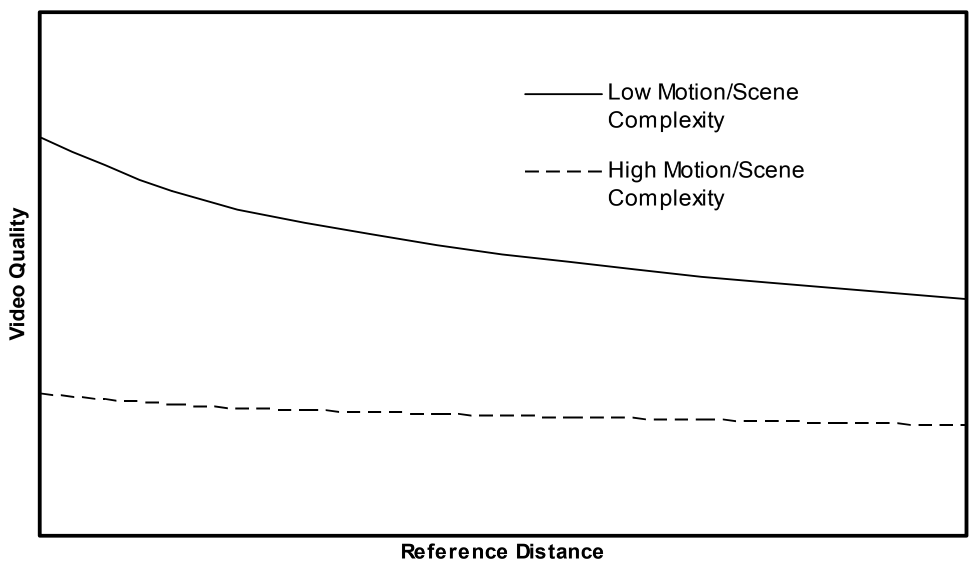

Figure 5.1. Hypothesis of the relationship between video quality and reference distance for videos with high motion and low motion.

5.2 Methodology

In order to explore the relationship between video quality and reference distance, the following methodology was used:

- Select a set of video clips with a variety of motion content (see Section 5.2.1).
- Change reference distances for each selected video sequence (see Section 5.2.2).
- Encode the video clips using H.264 (see Section 5.2.3).
- Measure video quality using PSNR and VQM (see Section 5.2.4).
- Analyze the results (see Section 5.3).

5.2.1 Select Video Clips

A set of video clips with a variety of motion content are selected to determine the effects of reference distance over a wide range of videos. These video clips are all in YUV 4:2:2 formats which are widely used in the video research community. The picture resolutions are all common intermediate format (CIF, 352x288 pixels). Each video sequence contains 300 video frames with a frame rate of 25 frames/second (fps). The content of these video clips can be roughly categorized into three groups: high motion/scene complexity, medium motion/scene complexity and low motion/scene complexity. Table 5.1 provides an approximate content classification of each video clip, with an identifying name and a short description of the video content.

Video Clip	Motion	Description
Akiyo	Low	A news reporter talking
Container	Low	A container ship moving slowly
News	Low	Two news reporters talking
Silent	Medium	A person demonstrating sign language
Mom & Daughter	Medium	A mother and daughter talking
Foreman	High	A foreman talking
Mobile	High	Panning of toy train moving
Coastguard	High	Panning of a coastguard ship moving

Table 5.1. Video clips used in the experiments.

Table 5.2 shows the fraction of P-Blocks in the encoded video clips. For the low-motion videos (Akiyo and Container), the majority of the macro-blocks are inter-coded (P-blocks), whereas for the high-motion videos (Foreman and Coastguard), only around half of the macro-blocks are inter-coded.

Video Clip	Fraction of the Inter Blocks (P-Blocks)
Akiyo	0.9666
Container	0.9246
News	0.8746
Silent	0.8637
Mom & Daughter	0.8423
Foreman	0.5947
Mobile	0.5722
Coastguard	0.5225

Table 5.2. The fraction of the inter blocks for different video clips

5.2.2 Changing Reference Distance

The main purpose of this study is to explore the relationship between video quality and reference distance. Thus, the encoder needs to be able to alter the reference distance in a controllable manner while encoding a video sequence. One way to achieve this is to modify the encoder to select the reference pictures specified by users. This approach is complicated as it involves modifying the encoder and may result in inaccurate measurements if done incorrectly. We take an alternative approach by changing the input video sequences. For instance, to use a reference picture which is two frames before the current frame instead of one frame before, the odd-number frames (1, 3, 5, …) are extracted and then the even–number frames (2, 4, 6, …). Both sequences are fed into the unmodified encoder, resulting in two video quality values. The video quality for the original video sequence is the average of these two quality values. The same approach is applied to other reference distances.

5.2.3 Encode/Decode

H.264 is used for video compression to encode/decode the video clips. The H.264 encoder/decoder used by this study is the Joint Model (JM 10.2)[12] developed by the Joint Video Team (JVT) which consists of experts from ITU-T VCEG and ISO/IEC MPEG. In this study, the following settings are applied to all experiments:

Since our study mainly explores how changing reference distance affects the quality of P-frames, primarily used in videoconferences, all the video frames are encoded as either P-frames or I-frames, and no B-frames are used in the experiments.

Under a bit-rate constraint, changing reference distance affects encoding efficiency and thus video quality. For fair comparison, the same bit-rate constraint is imposed for all experiments.

[12] http://iphome.hhi.de/suehring/tml/download/

H.264 supports multiple reference picture motion compensation which allows the encoder to select among several pictures that have been decoded and stored in the decoder. Since in our study the reference distance is between the encoding frame and a previously encoded frame, only one single reference frame is used in our experiments.

5.2.4 Measure of Video Quality

PSNR and VQM are used as metrics to measure video quality. The purpose of using two different quality metrics is to investigate whether different quality metrics have different relationships between video quality and reference distance. The PSNR measurement is conducted by JM as it reports the resulting PSNR for each video sequence being encoded. VQM is not reported by JM, so we have to use a VQM measurement tool named VQM-PC, downloaded from the VQM web site[13]. This VQM tool takes the original and the processed video clips as input and measures the video quality of the processed video clips relative to the original video clips. The resulting VQM score is in the range of (0, 1), where 0 represents no impairment and 1 represents the maximum impairment.

5.3 Analysis of Impact of Reference Distance on Video Quality

Section 5.1 hypothesizes that the video quality degrades as the reference distance increases, and the video quality for high-motion videos degrades slower than those low-motion videos as the reference distance increase. A series of experiments are conducted as described in Section 5.2.

[13] http://www.its.bldrdoc.gov/n3/video/vqmdownload_US.htm

5.3.1 Impact of Reference Distance on PSNR

The impact of reference distance on PSNR is first examined. Figure 5.2 depicts PSNR versus reference distances for eight video clips with different content. The bit-rate constraint for this experiment is 4.8 Mbps. GOP length is 22. As shown in Figure 5.2, as the reference distance increases, the PSNR for all the video clips degrade. However, the degrees of quality degradations of the eight videos are different. Akiyo shows the steepest quality degradation: as the reference distance is increased from 1 to 8, the PSNR drops from 48.06 db to 43.74 db. Coastguard shows the slowest quality degradation: as the reference distance is increased from 1 to 8, the PSNR drops from 35.47 db to 33.8 db.

Figure 5.3 depicts the trendlines and equations for Akiyo, Mom & Daughter, and Coastguard. As Figure 5.3 shows, the curves can be well described using the logarithmic function: $y = a\ln(x) + b$, where a is the gradient of the logarithmic function, determined by the motion of a video clip, and b is the y-intersection of the logarithmic function, determined by the scene complexity of a video clip. The coefficients of the logarithmic functions for all the curves shown in Figure 5.3 are presented in Table 5.2. Table 5.2 shows that as the amount of motion increases in the video clip, the gradient (a) of the quality degradation decreases, and as the scene complexity increases, the intersect (b) decreases. Note that Mobile has the most complex scene among the eight video clips. The R-Squared values for all the logarithmic functions are also presented in Table 5.2.

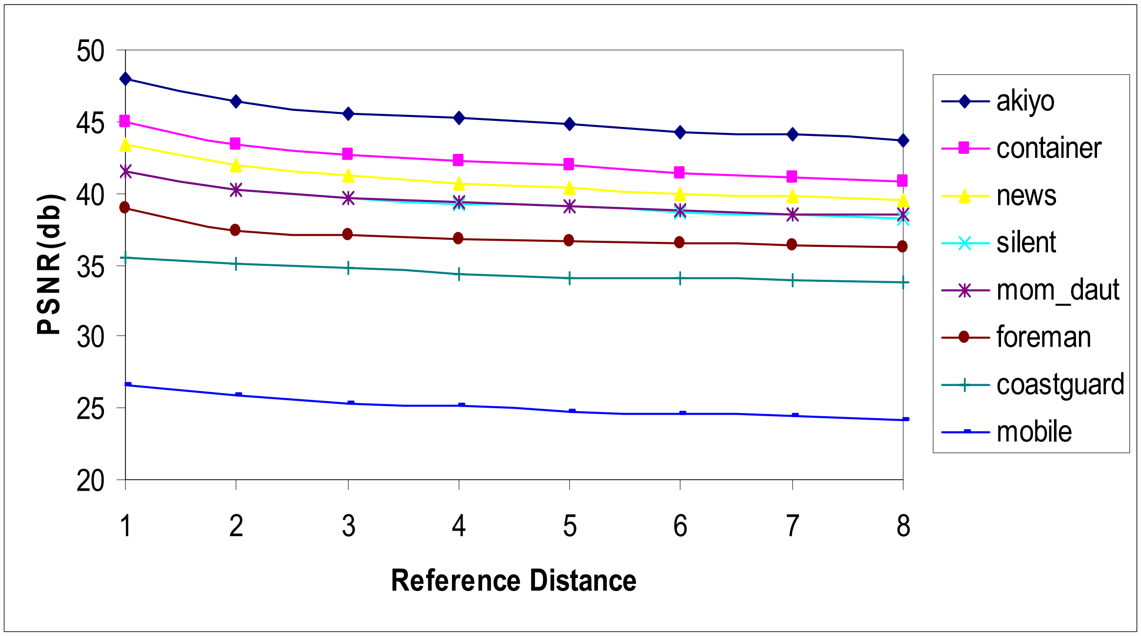

Figure 5.2 PSNR vs. reference distance for video clips with different content characteristics

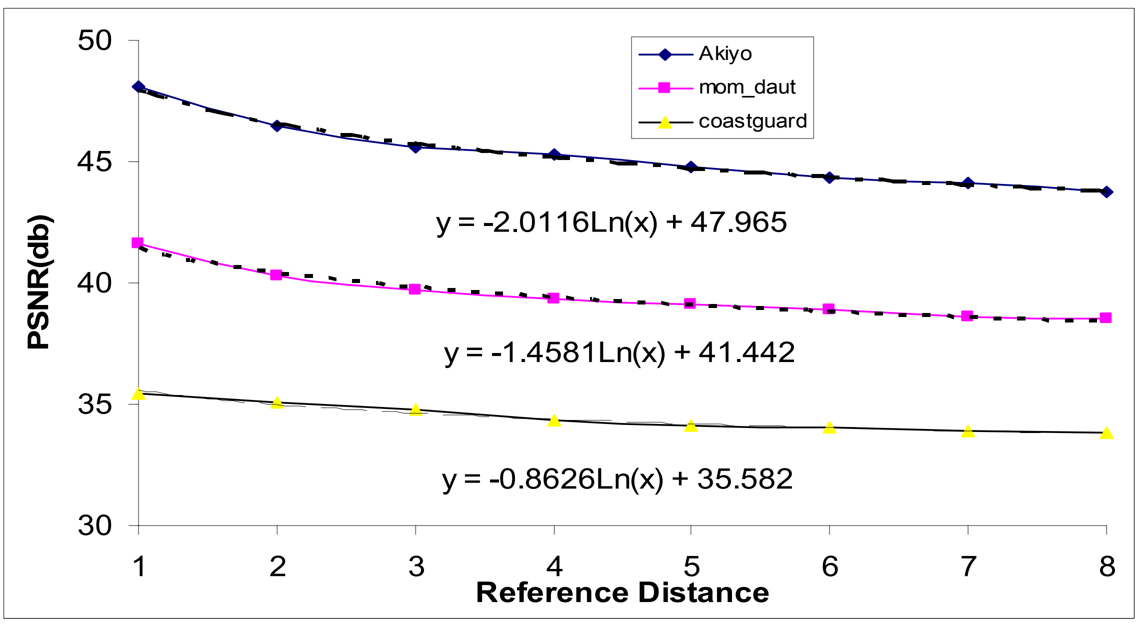

Figure 5.3. Trendlines and equations for Akiyo, Mom & Daughter, and Coastguard

Video Clips	Gradient (a)	Intersect (PSNR) (b)	R-Squared
Akiyo	-2.0116	47.965	0.9953
Container	-1.9023	44.838	0.9948
News	-1.8556	43.295	0.9984
Silent	-1.5283	41.41	0.9929
Mom & Daughter	-1.4581	41.442	0.9904
Foreman	-1.1681	38.511	0.9265
Mobile	-1.1553	26.663	0.9754
Coastguard	-0.8626	35.582	0.9975

Table 5.2 The coefficients that describe the relationship between PSNR versus reference distance

5.3.2 Impact of Reference Distance on VQM

We further test the hypothesis by using VQM to examine whether the results using PSNR hold for VQM. Figure 6 depicts VQM versus reference distance for eight video clips, the same clips used in the PSNR experiment. The bit-rate constraint for this experiment again is 4.8 Mbps. GOP length is 22. We use (1-VQM) as the quality metric for better comparisons with PSNR (i.e. higher values are better). Notice, by adopting this quality metric, 1 represents the best quality and 0 represents the worst. Figure 5.4 shows the same trend as Figure 5.2: as the reference distance increases, the video quality (1-VQM) degrades. Akiyo shows the steepest quality degradation: as the reference distance is increased from 1 to 8, the quality drops from 0.972 to 0.890. Coastguard shows the slowest quality degradations: as the reference distance is increased from 1 to 8, the quality drops from 0.843 to 0.831.

Figure 5.5 depicts the trendlines and equations for Akiyo, Mom & Daughter, and Coastguard using VQM. As Figure 5.5 shows, the lines can be well described using linear functions: $y = ax + b$, where a is the gradient of the linear function, determined by the motion of a video clip, and b is the y-intersection of the linear function, determined by the scene complexity of a video clip. The coefficients of the linear functions for all the lines shown in Figure 6 are presented in Table 5.3. Table 5.3 shows that, as the amount of motion increases in the video clip, the gradient (a) of the quality degradation decreases, and as the scene complexity increases, the intersect (b) decreases. The R-Squared values for all the linear functions are also presented in Table 5.3.

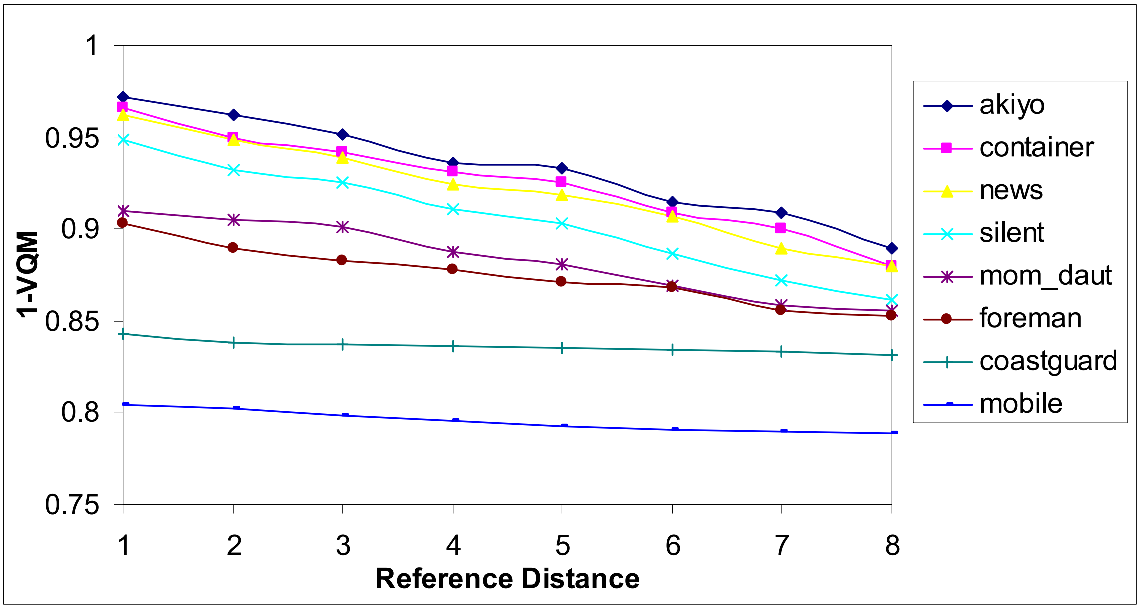

Figure 5.4. VQM vs. reference distance for video clips with different content

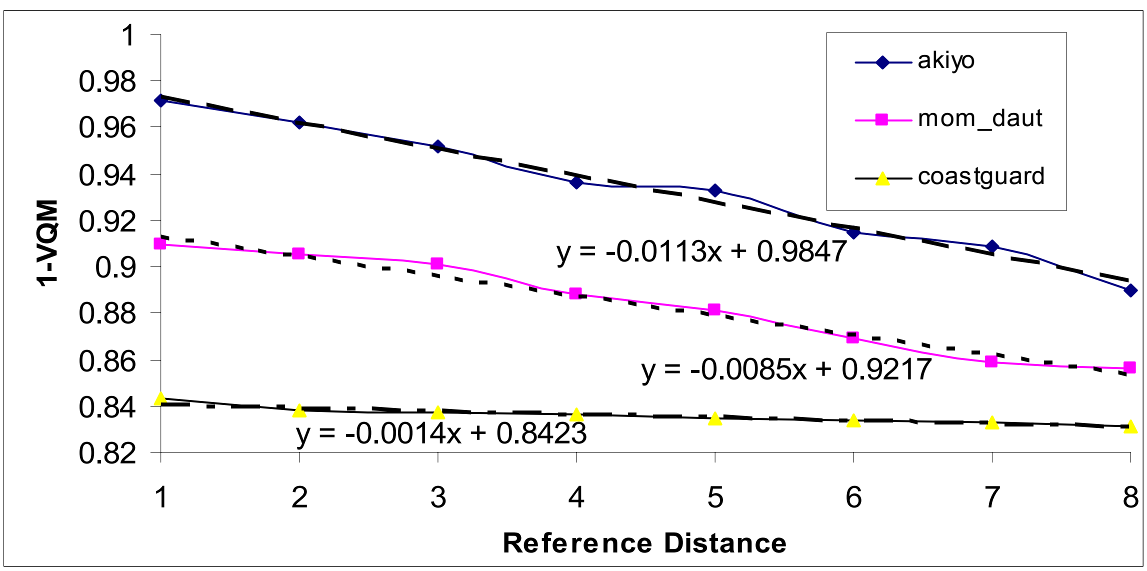

Figure 5.5. Trendlines and equations for Akiyo, Mom & Daughter, and Coastguard

Video Clips	Gradient (a)	Intersect (1-VQM)	R-Squared
Akiyo	-0.0113	0.9847	0.9869
Container	-0.0114	0.9766	0.9848
News	-0.0115	0.9732	0.9931
Silent	-0.0124	0.9606	0.9937
Mom & Daughter	-0.0085	0.9217	0.9821
Foreman	-0.0068	0.9059	0.9779
Mobile	-0.0022	0.8055	0.9076
Coastguard	-0.0014	0.8423	0.9671

Table 5.3. The Coefficients that Describe the Relationship between (1-VQM) vs. Reference Distance

5.4 Conclusion

In this study, a series of experiments are conducted to reveal how the change of reference distance affects video quality. A set of video clips with a variety of motions are selected for study, and the video sequences are shuffled to change the reference distances. For each reshuffled video sequence, an H.264 encoder encodes the sequence and measures video quality with PSNR and VQM.

From analysis of the experimental results, the relationship between video quality and reference distance can be determined. Both PSNR and VQM video quality degrade as reference distance increases. The degree of the video quality degradation is affected by the video content. The video quality for videos with high motions tends to degrade slower than that for those videos with low motion. This is largely because high-motion videos have a much larger number of inter-coded macro-blocks (P-blocks) and are thus less sensitive to the change of reference distance than the low-motion videos. Although these findings hold for both PSNR and VQM, the characterizations of the relationship between video quality and reference distance are different. While the relationship between PSNR and reference distance can be characterized using a logarithmic function, with VQM as the video quality metric, the same relationship can be characterized using a linear function.

Chapter 6

Model Validation

Chapter 4 describes the analytical models for feedback based error repair techniques. This chapter validates the accuracy of these analytical models in predicting the video quality under different network conditions through exploring one specific cases in details: RPS NACK and Intra Update. Section 6.1 describes the methodologies used for model verifications. Section 6.2 provides the verification results and analysis.

6.1 Methodology

Simulation experiments are designed to verify the accuracy of using these analytical models in predicting video quality under various network conditions. The video quality is measured by PSNR and VQM. Comparing performance predicted by the analytical models against simulated performance can provide an indication of the accuracy of our analytical models.

The simulations modify the input video sequences based on the given loss probability and round-trip time to mimic the effect of packet loss as well as the change of reference distance imposed by RPS on the video quality. The modified input sequences are encoded using H.264 and the average video quality in terms of PSNR and VQM are measured and compared against those predicted by our analytical models.

The simulation experiments make the same assumptions as those for analytical experiments:

- Reliable transmission of feedback messages through the feedback channel.
- An accurate estimate of the packet loss probability and the round-trip time from the network protocol.
- Independent network packet losses.
- Fixed round-trip times for the life of the flow.
- The independent segment decoding (ISD) mode of H.264 where each GOB is encoded as an individual sub-video independently from other GOBs in the same frame, and the reference frame is selected on a per-GOB basis, i.e., for all macro-blocks within one GOB the same reference frame is used.

For each experiment, the video quality predicted by the analytical models is compared to the actual quality achieved through the simulations. These comparisons evaluate the accuracies of our models.

In order to simulate the action of the H.264 encoder in response to frame loss, the encoder needs to be able to alter the reference distance in a controllable manner while encoding a video sequence. One way to achieve this is to modify the encoder to select the reference pictures specified by users. However, this approach is complicated as it involves modifying the encoder and may result in inaccurate measurements if done incorrectly. We take an alternative approach by changing the input video sequences as described later in this chapter, resulting in a decreased chance of human-induced error with comparable fidelity.

For all the analytical models, the simulation experiments adopt the following methodology:

1. Randomly drop a controllable number of frames in the input sequence based on the given loss probability.
2. Based on a given round-trip time and the randomly selected lost frames, regenerate the video sequence with RPS NACK.
3. Encode the video sequence generated in step 2 using H.264.
4. Measure the average PSNR and VQM for the encoded H.264 video sequence.
5. Calculate the average PSNR and VQM based upon the video quality measured in step 4 as well as the video quality of the locally concealed frames.

The simulation experiment for all the analytical models follow the same procedures as described above. The only difference is in step 2 where the input sequence is modified to simulate the different feedback-based error control techniques.

The following sections provide the details on the methodology of modifying the input sequence to simulate RPS NACK, RPS ACK and Intra Update. In the graphs presented in the following sections, each rectangle represents a video frame, the number represents the frame number and the letter indicates the frame type (I for intra-coded and P for predictive-coded).

6.1.1 RPS NACK

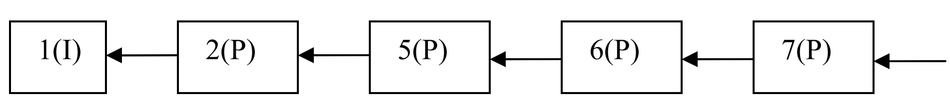

Figure 6.1 RPS NACK, round-trip time = 2 frames, frame 3 is lost

Figure 6.1 illustrates an example of modifying the input sequence for RSP NACK where the round-trip time is equal to two frames and frame 3 is lost. Since frame 3 is lost and the round-trip time is two frames, frame 4 cannot be decoded correctly, and frame 5 is encoded using frame 2 as a reference frame.

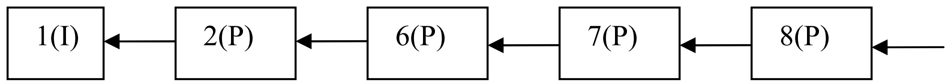

Figure 6.2 RPS NACK, round-trip time = 3 frames, frame 3 and 4 are lost

Figure 6.2 illustrates another example for RPS NACK where the round-trip time is equal to three frames and frame 3 and 4 are lost. Since frame 3 and 4 are lost and the round-trip time is three frames, frame 5 cannot be decoded correctly, and frame 6 is encoded using frame 2 as a reference frame.

6.1.2 Intra Update

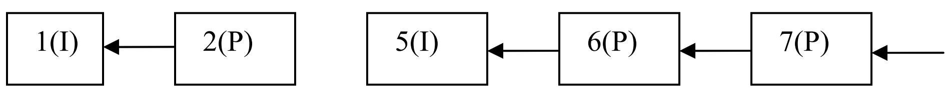

Figure 6.3 Intra Update, round-trip time = 2 frames, frame 3 is lost

Figure 6.3 illustrates an example of modifying the input sequence for Intra Update where the round-trip time is equal to two frames and frame 3 is lost. Since frame 3 is lost and the round-trip time is two frames, frame 4 cannot be decoded correctly, and frame 5 is intra-coded.

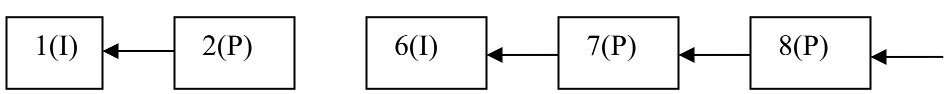

Figure 6.4 Intra Update, round-trip time = 3 frames, frame 3 and 4 are lost

Figure 6.4 illustrates another example for Intra Update where the round-trip time is three frames and frame 3 and 4 are lost. Since frame 3 and 4 are lost and the round-trip time is three frames, frame 5 cannot be decoded correctly, and frame 6 is intra-coded.

6.1.3 RPS ACK

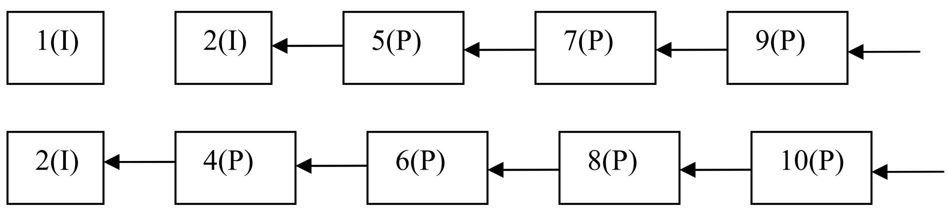

Figure 6.5 RPS ACK, round-trip time = 2 frames, frame 3 is lost

Figure 6.5 illustrates an example of modifying the input sequence for RPS ACK where the round-trip time is two frames and frame 3 is lost. Since the round-trip time is two frames, frames are normally encoded with reference to a frame that is two frames before the current frame. For instance, frame 4 is encoded with reference to frame 2, and frame 6 is encoded using frame 4 as reference etc. Two video sequences are generated: one is the odd-number frames (1, 3, 5, …) and other is the even–number frames (2, 4, 6, …). However, since frame 3 is lost, frame 5 cannot be encoded using frame 3 as a reference and instead frame 2, which is decoded correctly, is used as a reference frame.

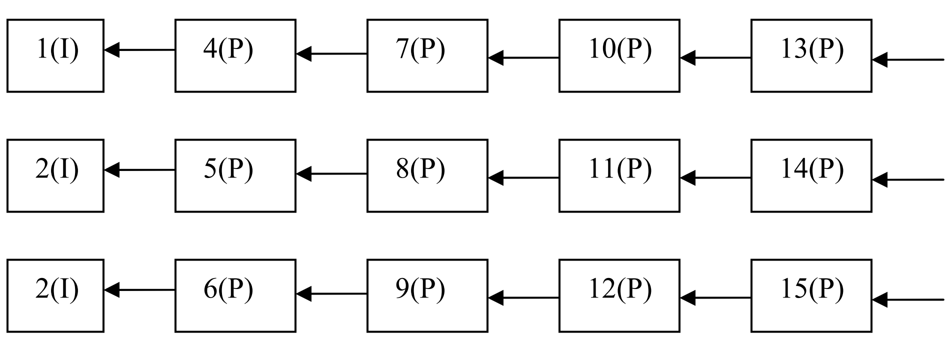

Figure 6.6 RPS ACK, round-trip time = 3 frames, frame 3 and 4 are lost

Figure 6.6 illustrates another example of modifying the input sequence for RPS ACK where the round-trip time is three frames and frame 3 is lost. In this case, since round-trip time is three frames, three video sequences are extracted: (1, 4, 7, …), (2, 5, 8, …) and (3, 6, 9, …). Since frame 3 is lost, frame 6 cannot be encoded using frame 3 as reference and instead frame 2, which is decoded correctly, is used as reference frame to encode frame 6.

6.2 Results and Analysis

The simulations vary round-trip time, loss probability as well as video contents. For a specific video clip, four input sequence with different loss patterns are generated based on the given loss probability and round-trip delay. Each modified input sequence is encoded using H.264 and the video quality in terms of PSNR and VQM for each encoded video sequence is measured and compared against the video quality value predicted by the analytical model. A set of video clips with a variety of motion content are selected for our simulations. These video clips are all in YUV 4:2:2 formats which are widely used in the video research community. The picture resolutions are all common intermediate format (CIF, 352x288 pixels). Each video sequence contains 300 video frames with a frame rate of 25 frames/second (fps). The content of these video clips can be roughly categorized into three groups: high motion/scene complexity, medium motion/scene complexity and low motion/scene complexity. An approximate content classification of each video clip, with an identifying name and a short description of the video content, is shown in Table 5.1.

To conserve space, only the simulations for RPS NACK are presented but the results of RPS ACK and Intra Update are similar. Section 6.2.1 presents the result with PSNR and Section 6.2.2 is the result with VQM.

6.2.1 PSNR

Figure 6.7 shows the average PSNR values for the simulations along with the PSNR values predicted by the analytical model for *Akiyo* which is a low motion video clip. The loss probability is varied from one percent up to thirty percent and the round-trip time is varied from 80 ms to 240 ms. For the given loss probability and round-trip time, the average PSNR over four simulation experiments is presented along with its 95 percent confidence interval shown with an error bar. The two curves illustrate the PSNR values predicted by the analytical model under varying loss probability and round-trip time. Both curves are within the 95 percent confidence intervals of all simulation samples, indicating the PNSR values predicted by the analytical model are consistent with the simulations results. As the loss probability increases, the variance is also increased from 0.4 db to 0.77 db for 80 ms round-trip time and 0.49 db to 0.83 db for 240 ms round-trip time. The increment of round-trip time increases the variance, which is expected since increased round-trip time produces longer sequences of error propagation.

Figure 6.8 and Figure 6.9 show the simulation results and model values for two videos, *News* and *Coastguard,* respectively. The simulation results are also consistent with the values predicted by the analytical model for both videos. It can also be observed that as motion contained in the video increases, the variance decreases due to the fact that high motion videos (such as Coastguard) contain more intra-coded macro-blocks and thus are less sensitive to error propagation.

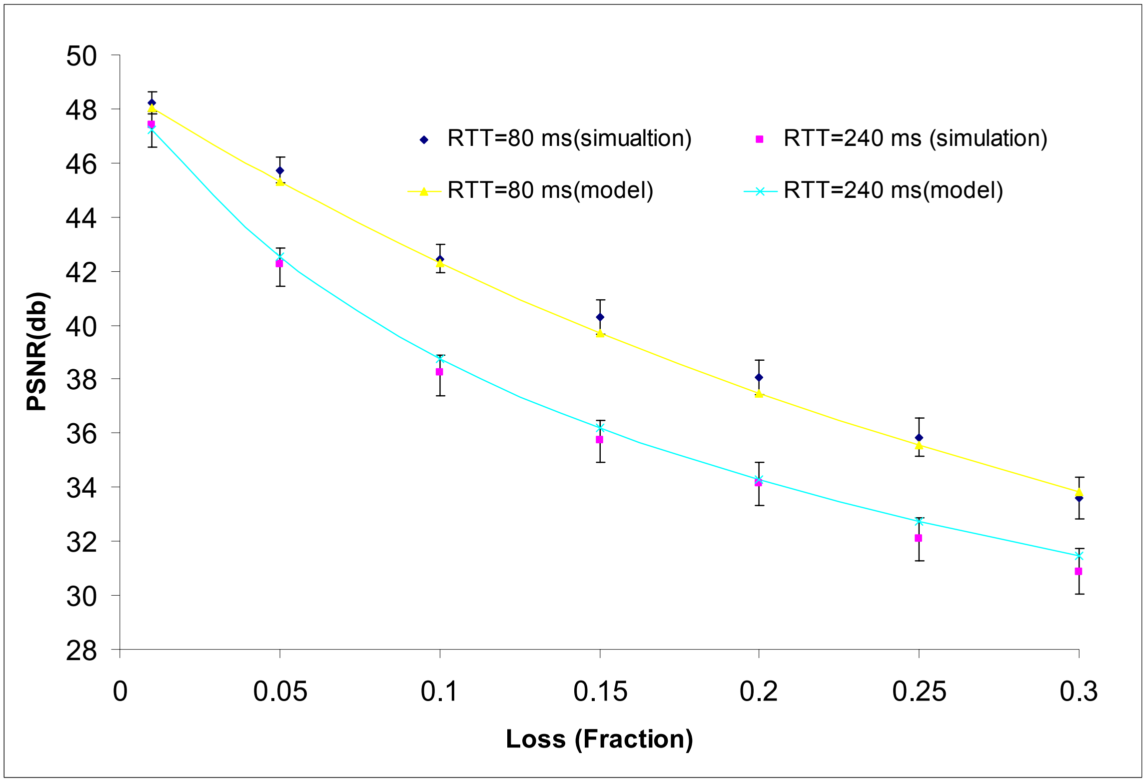

Figure 6.7 PSNR vs. loss with RPS NACK (video clip: Akiyo)

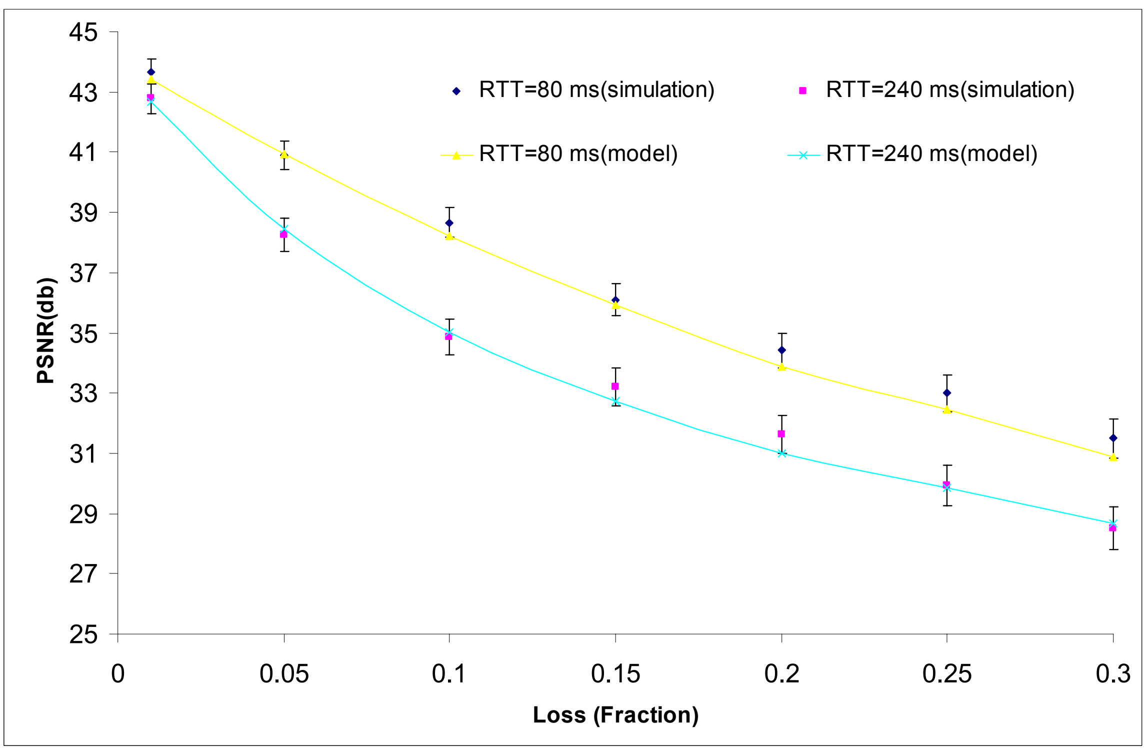

Figure 6.8 PSNR vs. loss with RPS NACK (video clip: News)

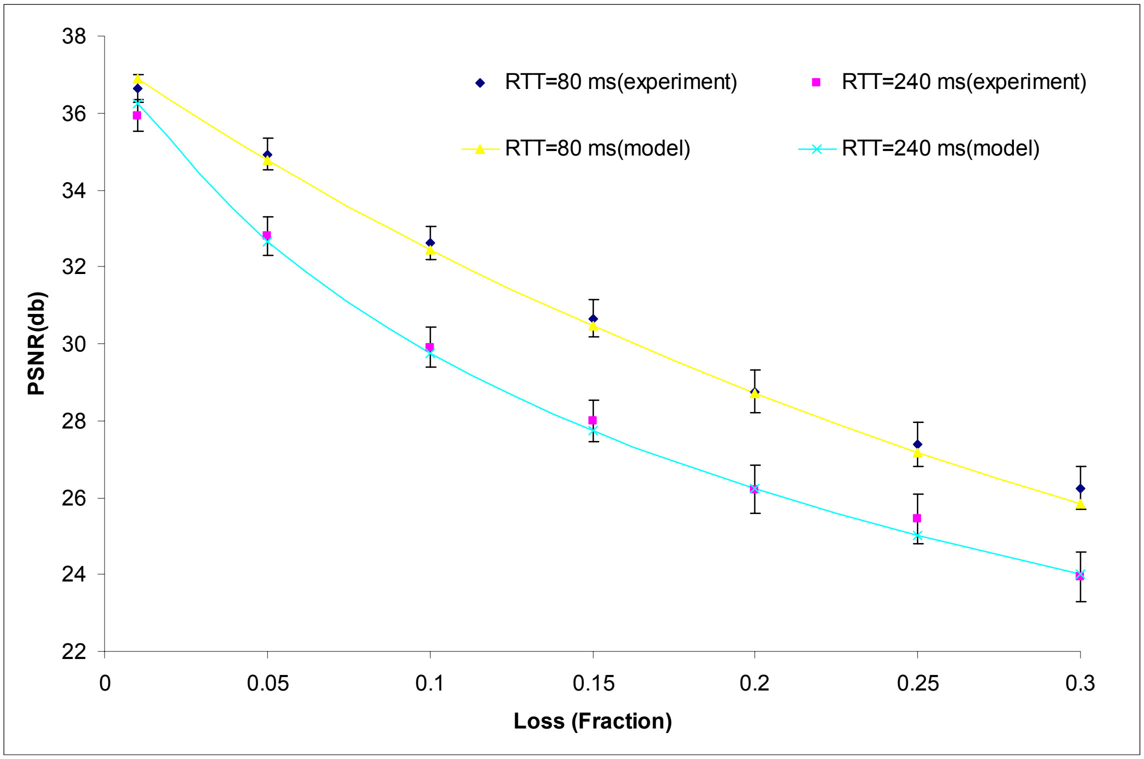

Figure 6.9 PSNR vs. loss with RPS NACK (video clip: Coastguard)

6.2.2 VQM

We next present the simulation results for RPS NACK using VQM as the video quality metric. The simulation experiments using VQM adopt the similar procedures as those using PSNR. However, VQM is not built into the H.264 encoder, so a VQM measurement tool named *VQM-PC* is used.[14]. This VQM tool takes the original and the processed video clips as input and measures the video quality of the processed video clips relative to the original video clips. The resulting VQM score is in the range of (0, 1), where 0 represents no impairment and 1 represents the maximum impairment. For better comparisons with PSNR (i.e. higher values are better), (1-VQM) is used as the quality metric so that a 1 represents the best quality and 0 represents the worst.

[14] Downloaded from the VQM web site at http://www.its.bldrdoc.gov/n3/video/vqmdownload_US.htm

Figure 6.10 shows the average VQM for the simulations along with the VQM predicted by the analytical model for *Akiyo*. It can be seen that the VQM predicted by the analytical model are consistent with the simulations results. It is also noticed that as the loss probability and round-trip time increase, the variance is also increased as expected.

Figure 6.11 and Figure 6.12 shows the simulation results and model values for *News* and *Coastguard* respectively. The simulation results are also consistent with the values predicted by the analytical model for both videos. It can also be observed that as motion contained in the video increases, the variance decreases due to the fact that high-motion videos contain more intra-coded macro-blocks and thus are less sensitive to error propagation.

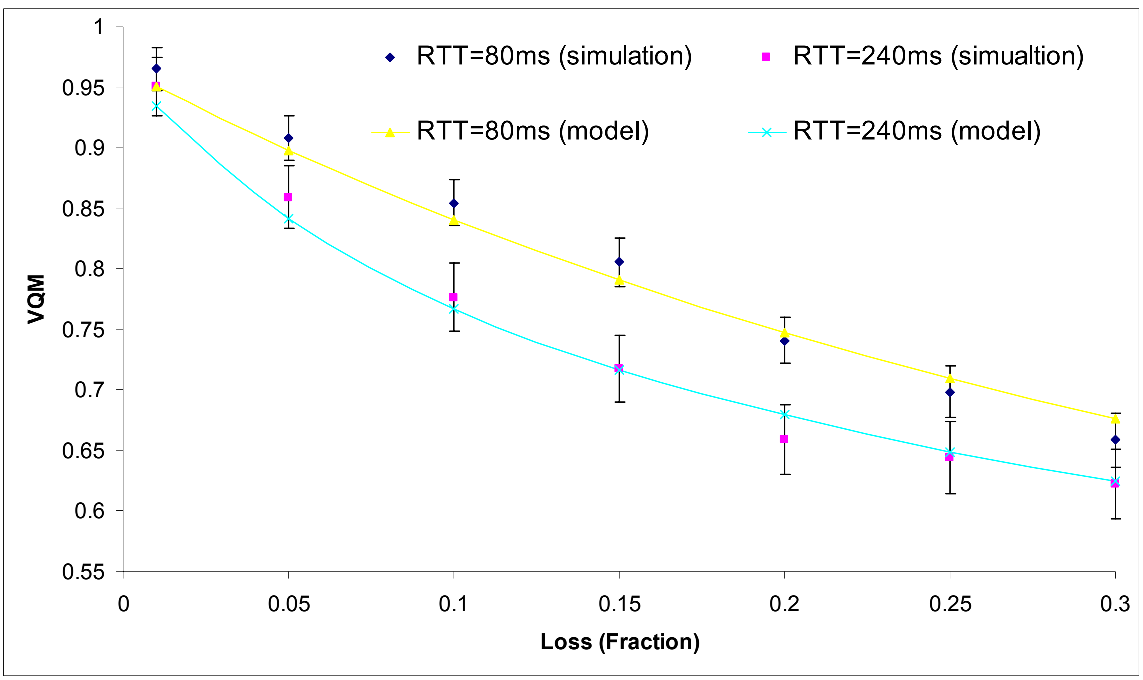

Figure 6.10 VQM vs. loss with RPS NACK (video clip: Akiyo)

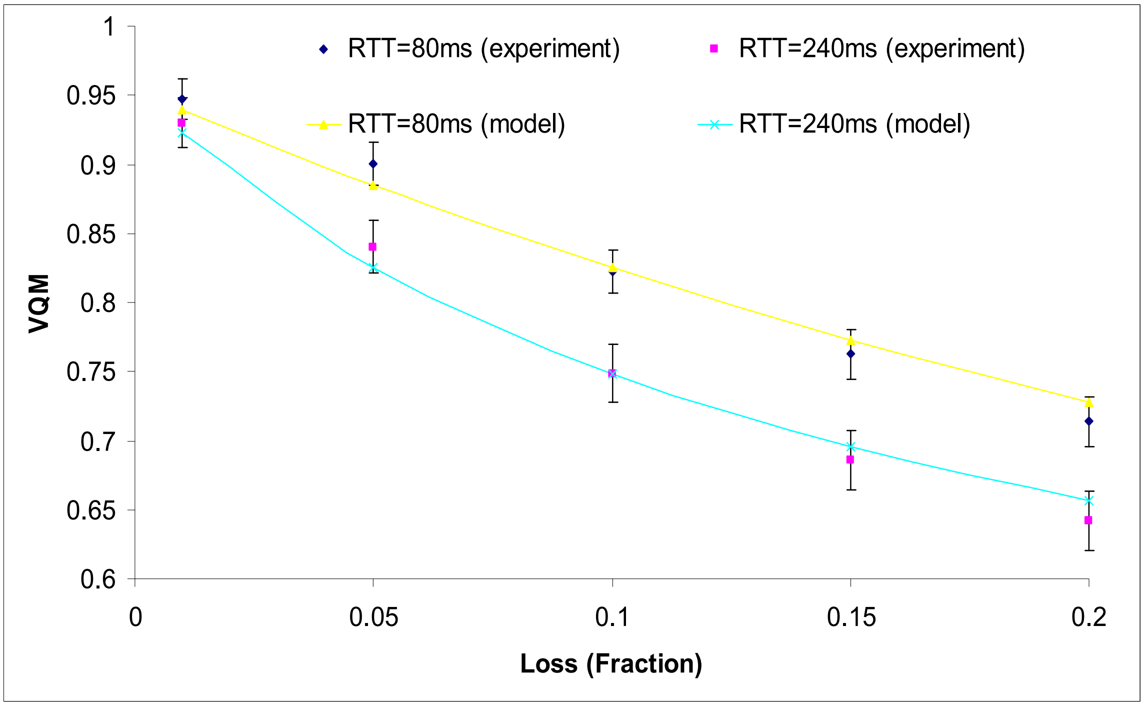

Figure 6.11 VQM vs. loss with RPS NACK (video clip: News)

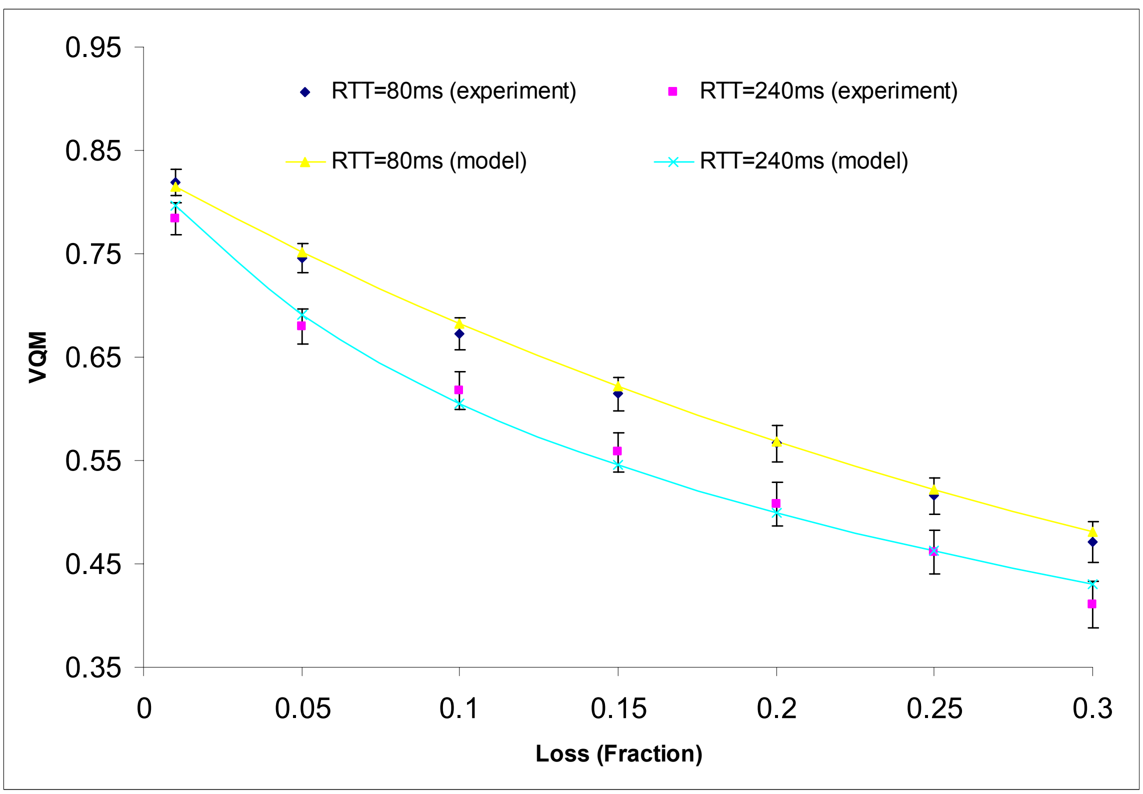

Figure 6.12 VQM vs. loss with RPS NACK (video clip: Coastguard)

Chapter 7

Analysis

This section uses the analytic models presented in the previous section to analyze feedback-based repair performance over a range of network conditions. The analytic experiments select a set of video clips with a variety of motion content. Each video sequence contains 300 video frames captured at a rate of 25 frames per second (fps). The content of these video clips can be categorized into one of three approximate groups based on motion/scene complexity: high, medium and low. The approximate content classification for each video clip, with an identifying name and a brief description of the video content, can be found in Table 5.1.

7.1 Retransmission

This section analyzes the performance of Retransmission under a variety of network conditions. For videos using Retransmission for repair, there are two major factors that affect the video quality: error propagation and retransmission overhead. First, when a video transfer error occurs, the damage due to packet loss propagates at least a period of one round-trip time until the retransmitted packet arrives. The longer round-trip time induces longer error propagation and thus more damage to video quality. Secondly, the increase of packet loss triggers more retransmissions and hence consumes extra bit-rate. Thus to maintain a constant frame rate and bit rate, the encoder uses a coarser quantization and the overall video quality decreases. Figure 7.1.1 shows the impact of bit-rate constraints on the PNSR of video *News* encoded using H.264. Roughly a bit-rate reduction of 60 kbits/s causes 1db degradation of PSNR for *News*. Similar behavior can be observed in Figure 7.1.2 where VQM is used to measure the video quality of *News*.

The impact of loss rate on video quality using Retransmission is first investigated. Figure 7.1.3 provides PSNR versus loss rate curves for one video (*News*) video using Retransmission for repair and GOP size 22, for four round-trip times ranging from 80 milliseconds to 320 milliseconds. The quality for a locally concealed GOB is 50% of the best quality of a GOB. As the loss rate increases, the video quality (PSNR) using Retransmission degrades for all round-trip times. However, with Retransmission, video quality under higher round-trip times degrades faster than does the video quality under lower round-trip times. The reason is twofold. First, with Retransmission larger round-trip time implies longer error propagation periods causing video quality to degrade more rapidly. Second, a successful repair using Retransmission requires that all the GOBs in the Retransmission Range (RR)[15] are received correctly; larger round-trip time implies a larger Retransmission Range and thus a greater probability that a transmission error may occur for a GOB within the Retransmission Range. A similar trend can be observed when VQM is adopted as video quality metric as shown in Figure 7.1.4.

The impact of round-trip time on video quality is examined next. Figure 7.1.5 depicts PSNR versus round-trip time for one video (*News*) using Retransmission for repair under four loss rates ranging from 1% to 20%. As round-trip time increases, in Figure 7.1.3, the video quality (PSNR) degrades for all loss rates. However, the amount of quality degradations is not uniform. With Retransmission, video quality under the higher packet

[15] Please refer to Chapter 4 for the definition of Retransmission Range.

loss rates degrades faster with an increase in round-trip time than under lower packet loss rates. This is because higher packet loss rates induce more frequent GOB error propagation and thus video quality degrades more quickly. Moreover, higher packet loss rates induce more frequent retransmissions and thus consume more capacity. Thus to maintain a constant frame rate and bit rate, the encoder uses a coarser quantization and the overall video quality decreases. A similar trend can be observed when VQM is adopted as video quality metric as shown in Figure 7.1.6.

In the above analysis, we assume that every lost packet is retransmitted. However, when the channel capacity is extremely constrained, retransmission of every lost packet may not be feasible. We then explore the achievable video quality using partial retransmission, where only a portion of lost GOBs are retransmitted and the early GOBs in the reference chain are given higher priority for retransmission over the later ones. The objective is to find out whether retransmitting fewer lost packets can improve video quality or not. Figure 7.1.7 shows PSNR versus fraction of retransmission for four round-trip times with 10% packet loss. For lower round-trip time, PSNR improves as more lost packets get retransmitted while for higher round-trip times, the performance gain by retransmitting more lost packets gradually diminishes. This clearly suggests that retransmission is effective only when the round-trip time is low. It is also noticed that partial retransmission may perform the same as full retransmission. For example, for a round-trip time of 320ms, the retransmission of 50% of the lost packets achieves nearly the same PSNR as does that of 100% of the lost packets. A similar trend can be observed when VQM is adopted as video quality metric as shown in Figure 7.1.8. We further investigate the effectiveness of Retransmission by measuring the ratio of the video quality improvement over the extra bit-rates consumed by Retransmission. Figure 7.1.9 shows this ratio versus fraction of retransmission under four round-trip times. In Figure 7.1.9 *Retransmission Gain* is defined as follows:

$$Rteransmission\,Gain = \frac{The\,video\,Quality\,Gain\,over\,Base\,Line}{Extra\,BitRate\,Needed}$$

The base line is a retransmission rate of 10% of the lost packets and the loss rate of this experiment is 10%. For each fixed fraction of Retransmission, Retransmission is clearly more effective when round-trip time is low than when round-trip time is high. For all four round-trip times, retransmission of 50% of loss packets achieves the best effectiveness in terms of the ratio of performance gain over bit-rate cost.

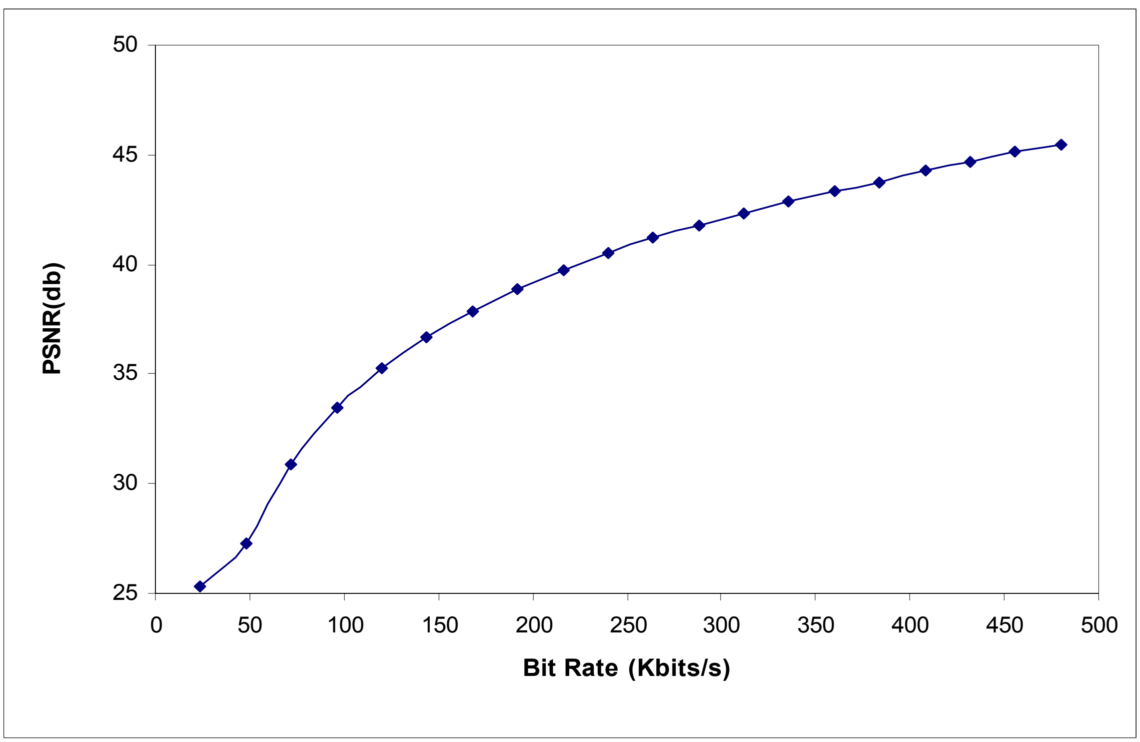

Figure 7.1.1 PSNR vs. bit-rate for video News

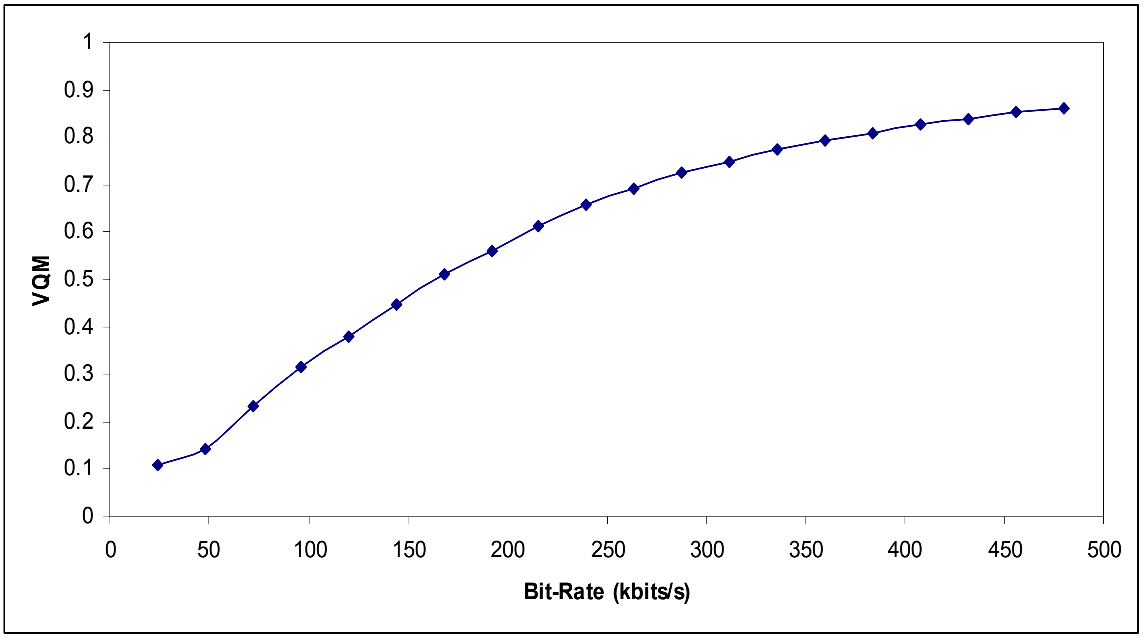

Figure 7.1.2 VQM vs. bit-rate for video News

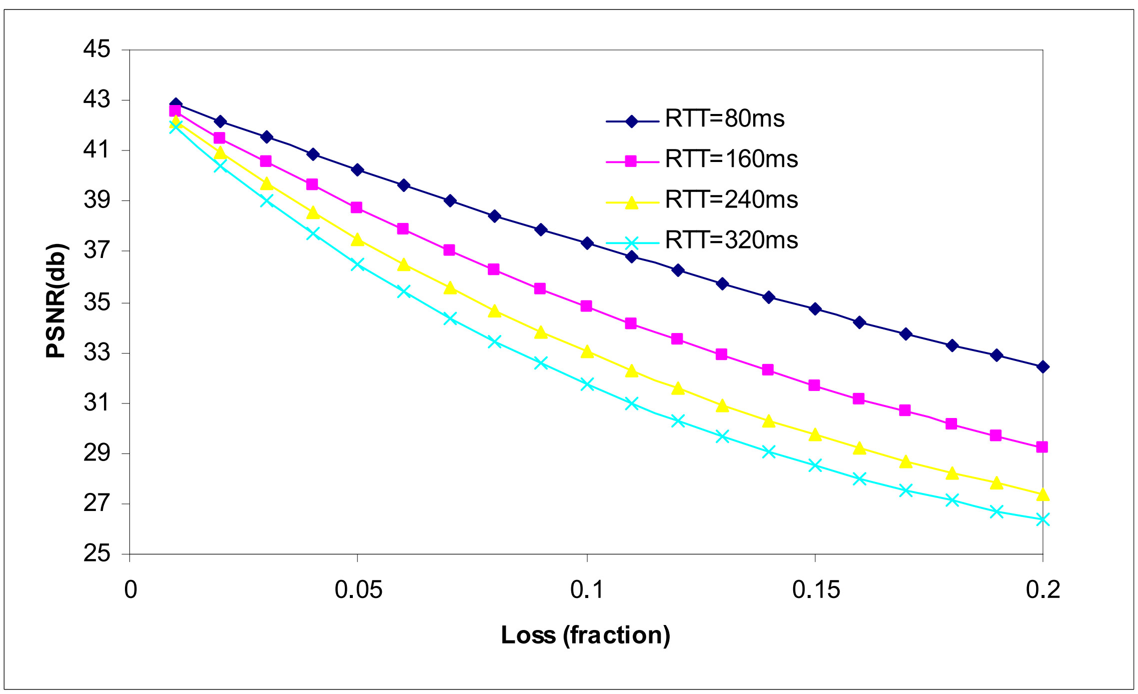

Figure 7.1.3 PSNR vs. loss with Full Retransmission under different round-trip Times for video News

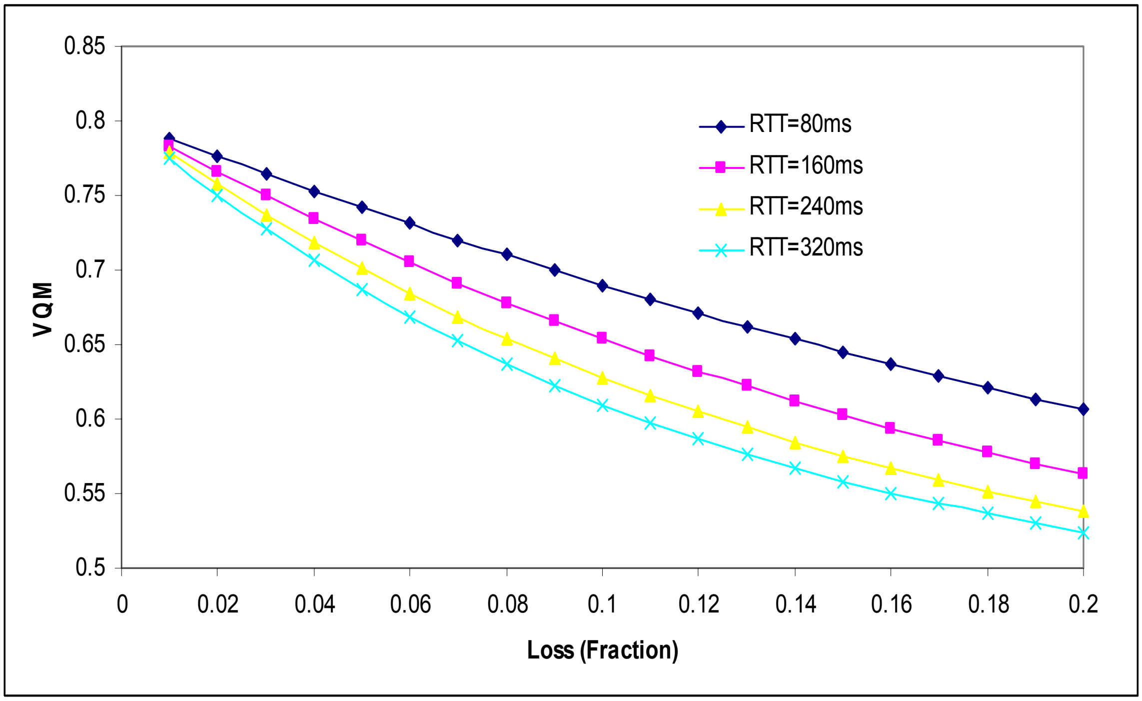

Figure 7.1.4 VQM vs. loss with Full Retransmission under different round-Trip Times for video News

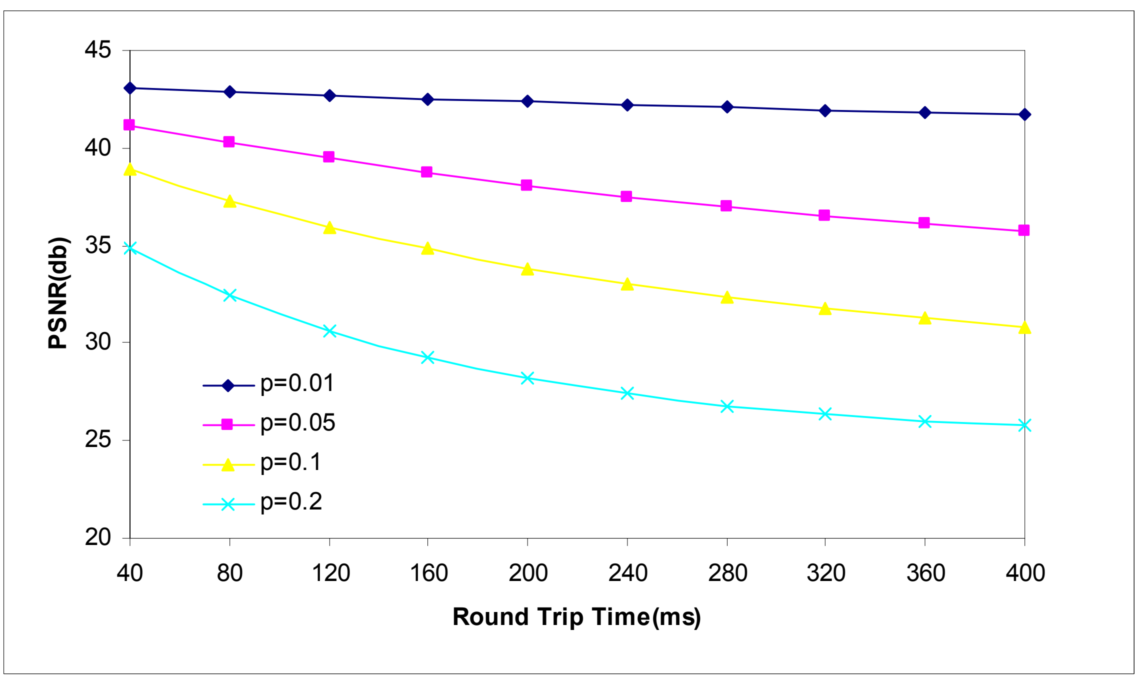

Figure 7.1.5 PSNR vs. round-trip time with Full Retransmission under different loss rates for video News

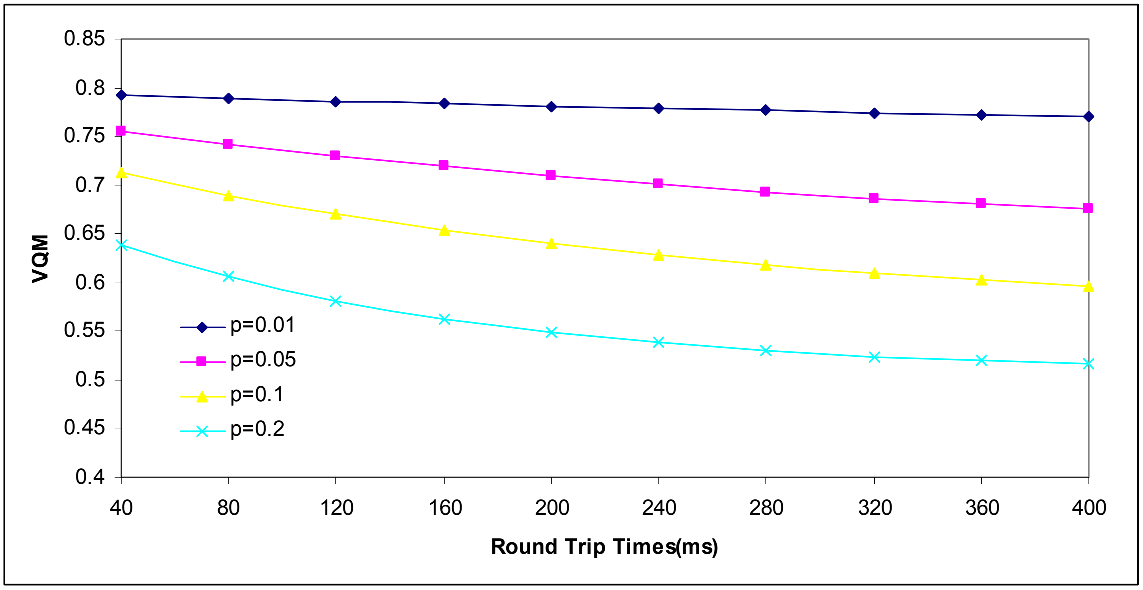

Figure 7.1.6 VQM vs. round-trip time with Full Retransmission under different loss rates for video News

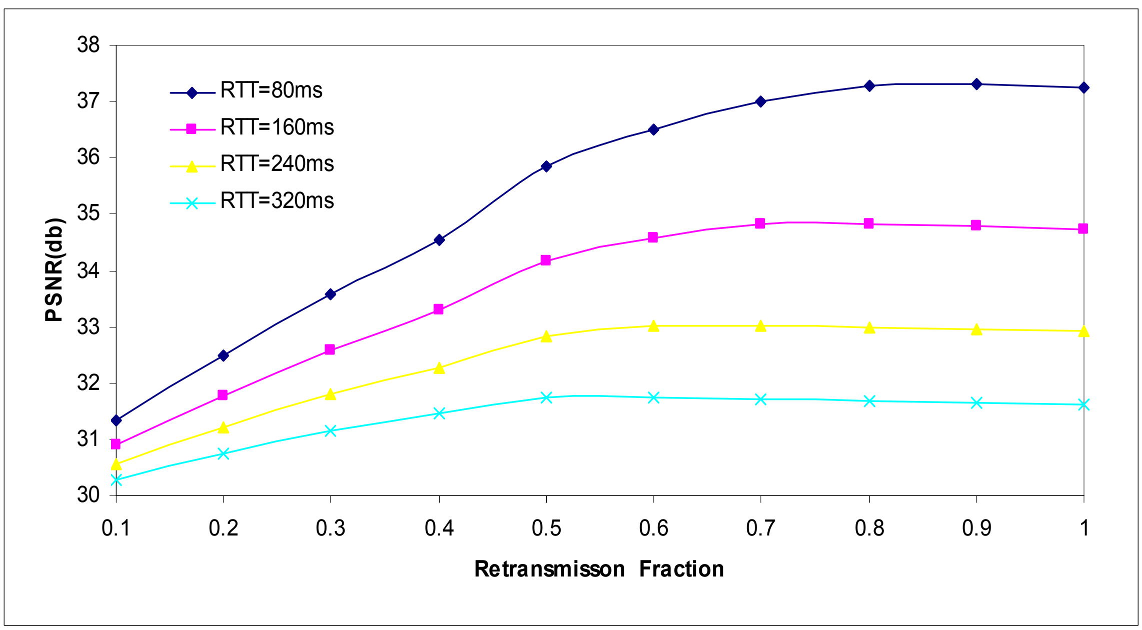

Figure 7.1.7 PSNR vs. retransmission fraction with Partial Retransmission under different round-trip times for video News (loss rate 10%)

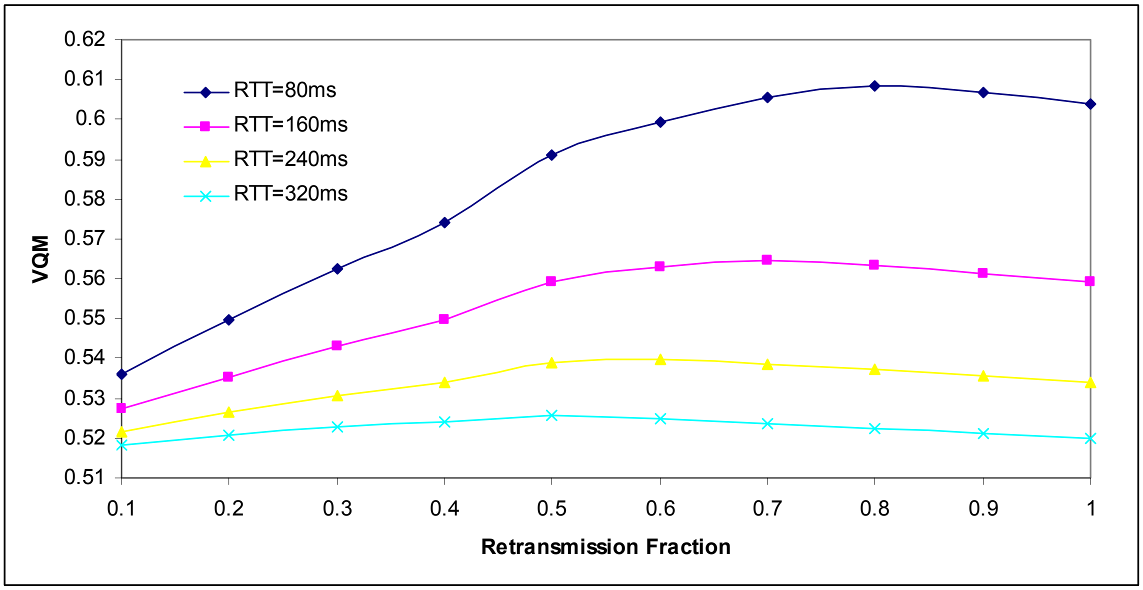

Figure 7.1.8 VQM vs. retransmission fraction with Partial Retransmission under different round-trip times for video News (loss rate 10%)

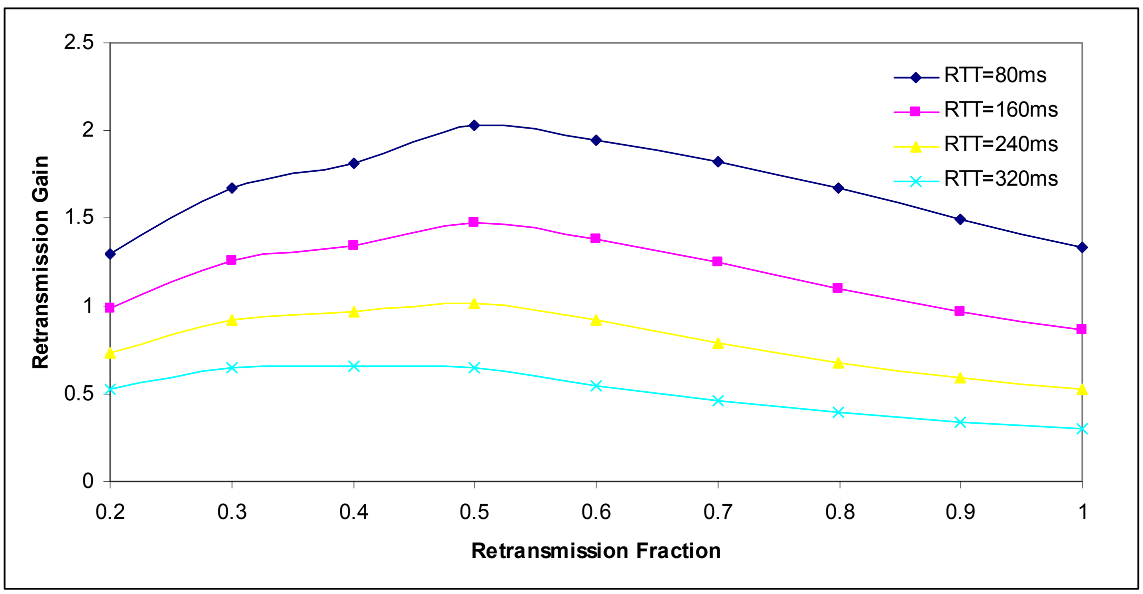

Figure 7.1.9 Retransmission gain vs. retransmission fraction under different round-trip times for video News (loss rate 10%)

7.2 RPS NACK

This section analyzes the performance of RPS NACK under a variety of network conditions. For videos using RPS NACK for repair, there are two major factors that affect video quality: error propagation and reference distance. Error propagation is mainly determined by both loss rate and round-trip time. Higher loss rate induces more frequent error propagation and longer round-trip time causes longer duration of error propagation. When an error is detected, the encoder using RPS NACK uses an older GOB as a reference for prediction. The longer reference distance reduces the coding efficiency and thus lowers video quality. The reference distance is primarily determined by round-trip time.

The impact of round-trip time on RPS NACK video quality is examined first. Figure 7.2.1 depicts PSNR versus round-trip time for one video (*News*) encoded with RPS NACK and GOP size 22, under four loss rates ranging from 1% to 20%. The quality for a locally concealed GOB is 50% of the best quality of a GOB. As round-trip time increases, in Figure 7.2.1, the video quality (PSNR) degrades for all loss rates. However, the amount of quality degradations is not uniform. With RPS NACK, video quality under the higher packet loss rates degrades faster with an increase in round-trip time than under lower packet loss rates. For RPS NACK, each transmission error propagates to subsequent frames for a period of one round-trip time. Thus, higher packet loss rates induce more frequent GOB error propagation and video quality degrades more quickly with higher round-trip times. A similar trend can be observed when VQM is adopted as video quality metric as shown in Figure 7.2.2.

The impact of loss rate on RPS video quality is now investigated. Figure 7.2.3 provides PSNR versus loss rate curves for the *News* video encoded with RPS NACK for four round-trip times ranging from 80 milliseconds to 320 milliseconds. As the loss rate increases, the video quality (PSNR) using RPS NACK degrades for all round-trip times.

However, with RPS NACK, video quality under higher round-trip times degrades faster than does the video quality under lower round-trip times. With RPS NACK, larger round-trip time implies longer error propagation periods that causes video quality to degrade more rapidly. A similar trend can be observed when VQM is adopted as video quality metric as shown in Figure 7.2.4.

Finally, the RPS NACK models are used to investigate the impact of the GOP length on video quality. Figure 7.2.5 depicts PSNR versus GOP length for video News encoded with RPS NACK for round-trip times ranging from 80 ms to 320 ms, as well as a video with no repair. The loss rate for this experiment is 0.05. Below a GOP length of 5, video quality increases in all cases. After the GOP length reaches 5, quality for the video without RPS degrades due to error propagation. With RPS NACK, when the round-trip time is 80 ms, quality increases and becomes asymptotically steady. When round-trip times are 160 ms, 240 ms and 320 ms, quality slightly decreases and becomes asymptotically steady. For all GOP lengths, videos with RPS NACK perform no worse than videos without RPS, and RPS NACK performs better for lower round-trip times than for higher round-trip times since higher round-trip times introduce longer periods of error propagation. The result shown in Figure 7.2.5 can be compared with those in [21], which studied the impact of the choice of GOP on video quality for MPEG video and Forward Error Correction (FEC) repair. The results in Figure 7.2.5, that GOP lengths larger than 5 have diminishing returns, are similar to the result in [21] despite different video encoding (H.264 versus MPEG) and different repair methods (RPS versus FEC). A similar trend can be observed when VQM is adopted as a video quality metric as shown in Figure 7.2.6.

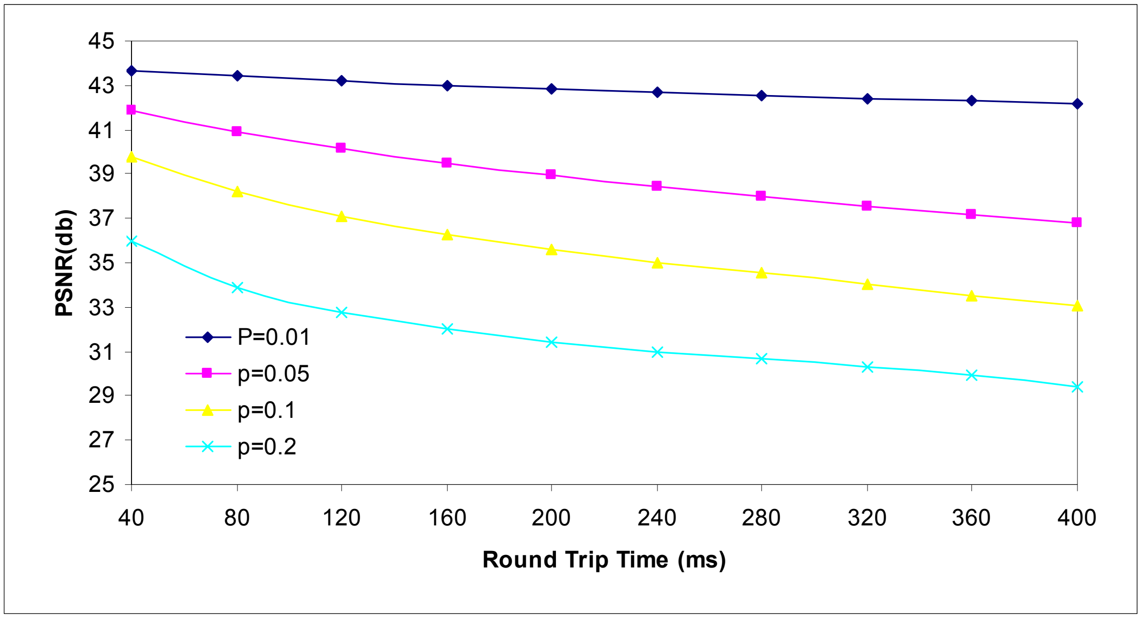

Figure 7.2.1 PSNR vs. round-trip time with RPS NACK under different loss rates (video News)

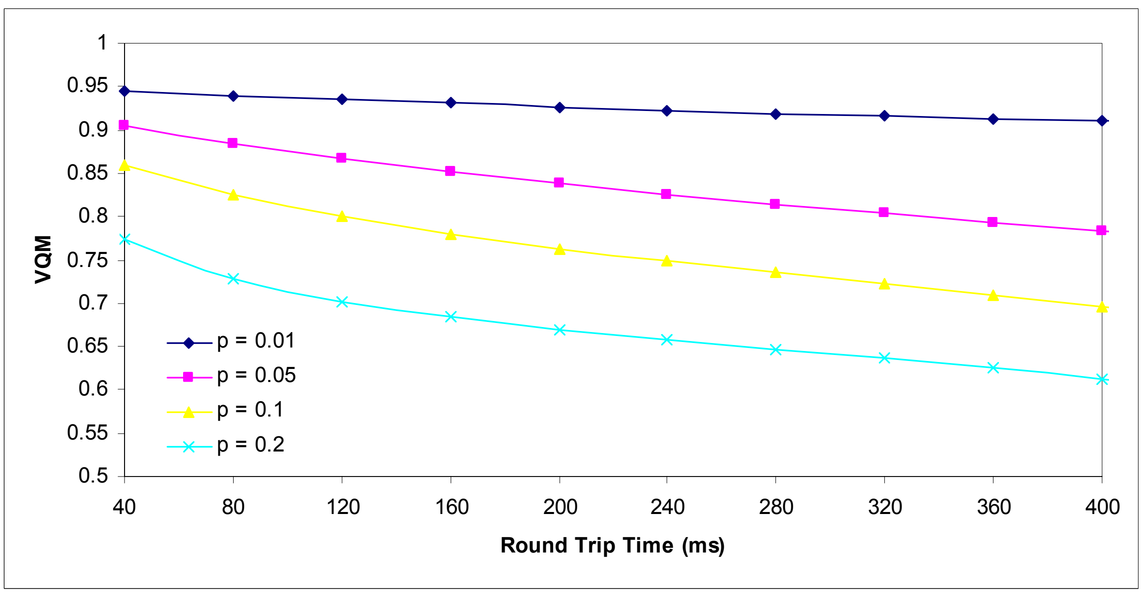

Figure 7.2.2 VQM vs. round-trip time with RPS NACK under different loss rates (video News)

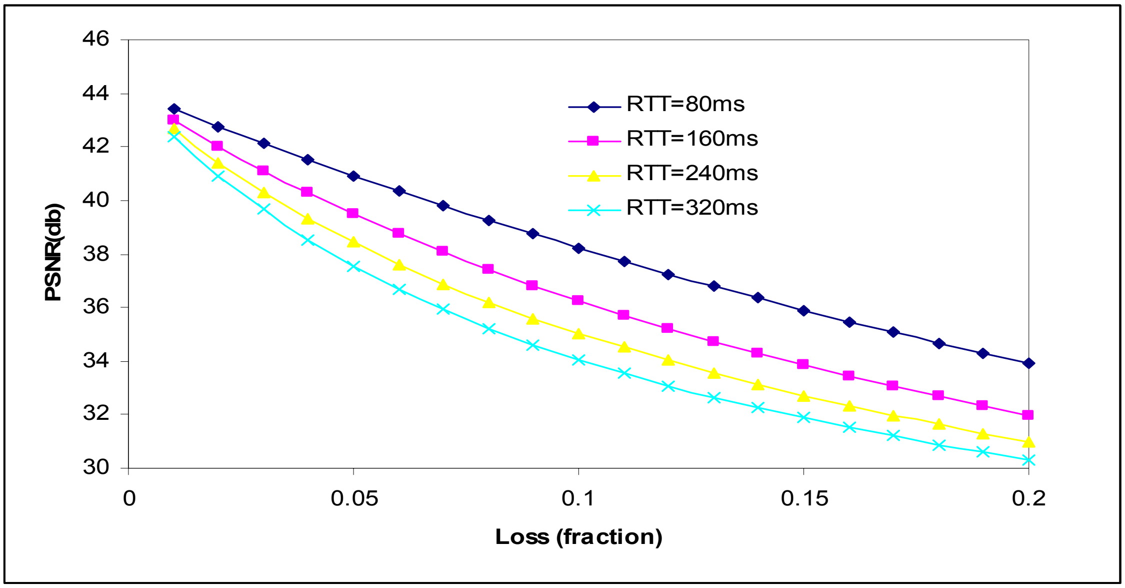

Figure 7.2.3 PSNR vs. loss with RPS NACK under different round-trip times (video News)

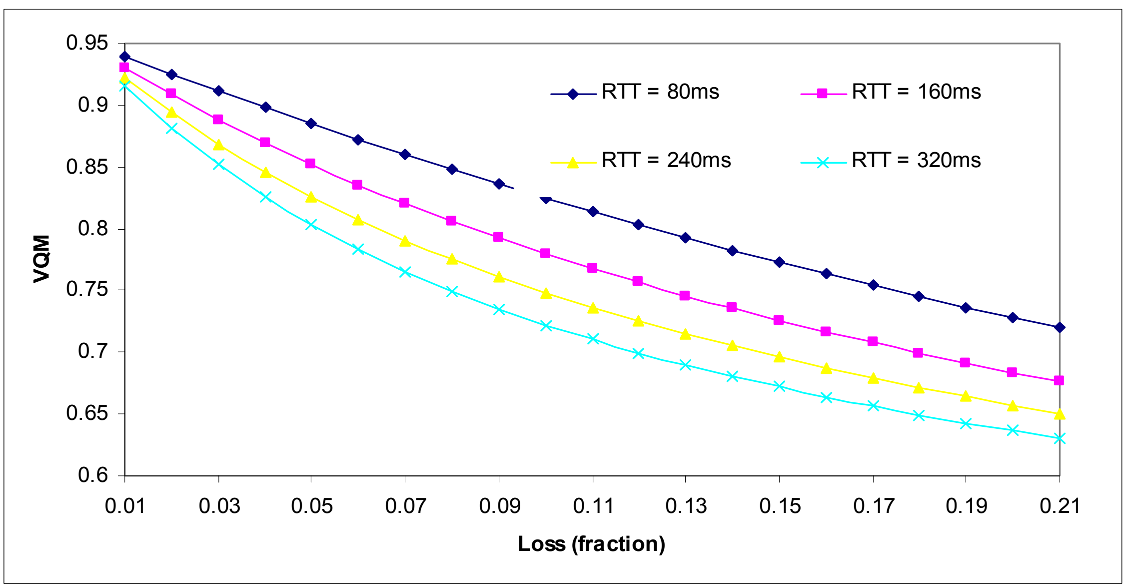

Figure 7.2.4 VQM vs. loss with RPS NACK under different round-trip times (video News)

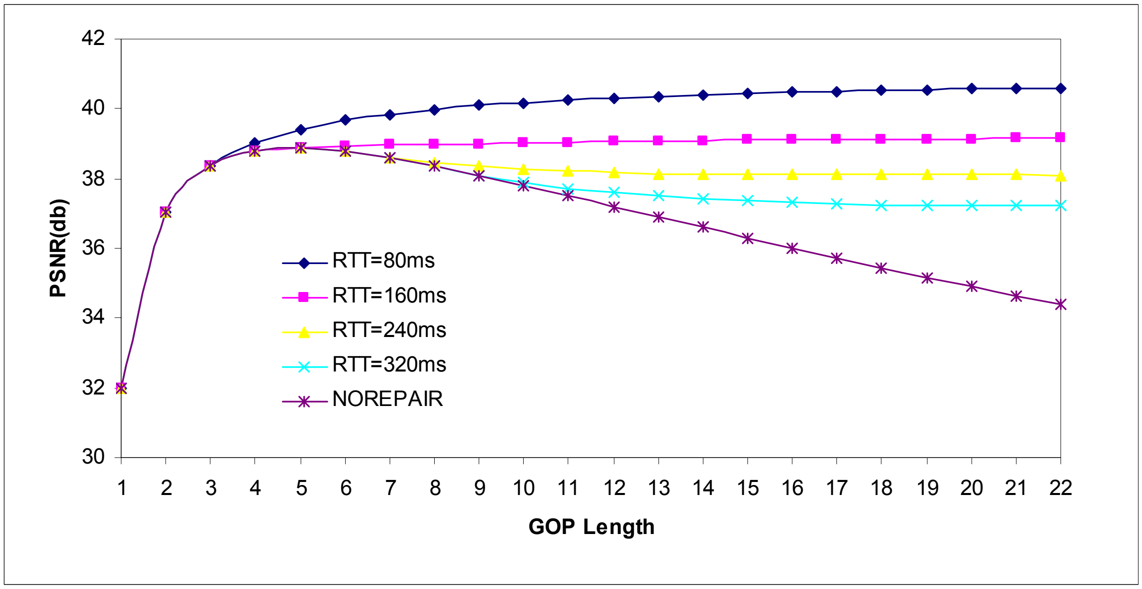

Figure 7.2.5 PSNR vs. GOP length with RPS NACK (p=0.05, video News)

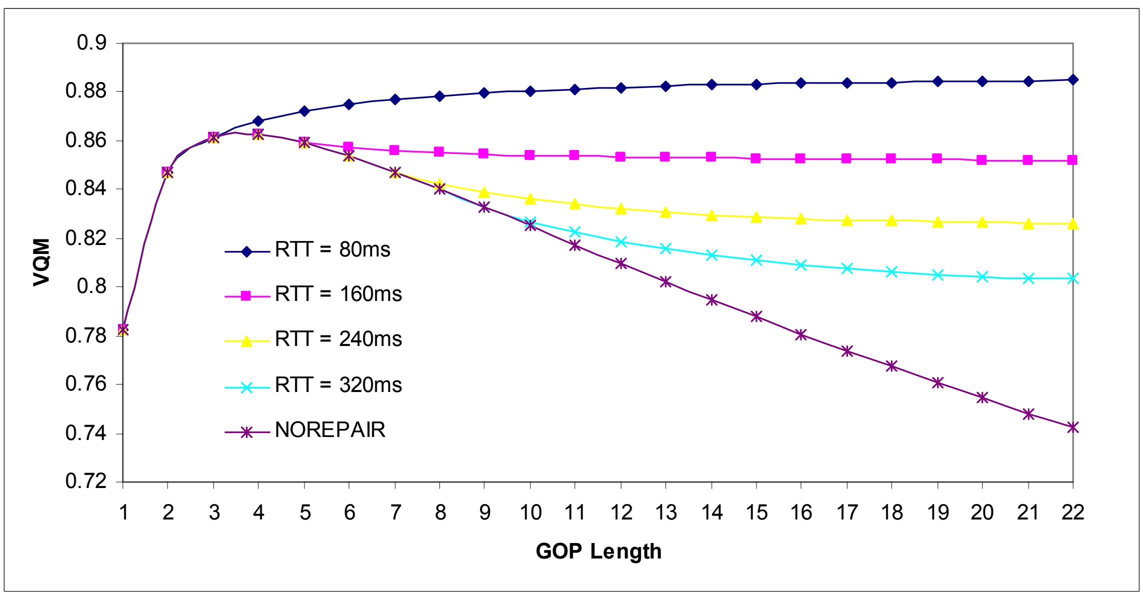

Figure 7.2.6 VQM vs. GOP length with RPS NACK (p=0.05, video News)

7.3 RPS ACK

This section continues analyzing the performance of RPS but with ACK mode. For videos encoded using RPS ACK, reference distance is major factor that affects the video quality. The longer of reference distance reduces the coding efficiency and thus lowers video quality. Again, the reference distance is primarily determined by round-trip time.

Figure 7.3.1 depicts PSNR versus round-trip time for a video encoded with RPS ACK under the same four loss rates. As the round-trip time increases the average PSNR for videos with RPS ACK degrades for all loss rates, similar to RPS NACK. However, unlike with RPS NACK, RPS ACK video quality degrades slower with an increase in round-trip time and higher packet loss rates than under lower packet loss rates. When the packet loss rate is low, the major cause of video quality degradation for RPS ACK is the increased reference distance caused by the round-trip time; whereas under higher packet loss rates, the video quality degradation for RPS ACK is attributed more to packet loss than to round-trip time. A similar trend can be observed when VQM is adopted as video quality metric as shown in Figure 7.3.2.

Figure 7.3.3 graphs PSNR versus loss probability curves for a video encoded with RPS ACK for the same four round-trip times. As loss rate increases, like RPS NACK, for RPS ACK the video quality degrades for all round-trip times. However, unlike RPS NACK, for RPS ACK, the video quality under higher round-trip times degrades slower than those under lower round-trip times. Under higher round-trip times, video quality degradation for RPS ACK is attributed more to the round-trip time than to the packet loss, whereas under lower round-trip times, packet loss is the dominant cause of video quality degradation. A similar trend can be observed when VQM is adopted as video quality metric as shown in Figure 7.3.4.

Figure 7.3.5 depicts PSNR versus GOP length for video News encoded with RPS ACK for the same four round-trip times. As GOP length increases, quality increases for videos with RPS ACK for all round-trip times shown. Since RPS ACK uses Intra coding before any frames are acknowledged, the quality for the first part of the GOP remains constant and increases only after ACKs are received by the encoder. For all GOP lengths, RPS ACK performs better under lower round-trip times than under higher round-trip times since RPS ACK under higher round-trip times has to use older frames as references for prediction. It is worth noting that below a certain GOP size, video without repair performs better than videos with RPS ACK for repair. For instance, when the round-trip time is 80 ms and GOP length is below 5, videos without RPS have better quality than videos with RPS ACK. Videos without RPS always use the previous GOB as reference and rely on Intra coding to stop error propagation. When the GOP length is small, error propagation can be quickly stopped, whereas RPS ACK always uses older frames as references. Therefore, when the loss rate is low and the GOP length is small, videos without RPS outperform videos with RPS ACK. A similar trend can be observed when VQM is adopted as a video quality metric as shown in Figure 7.3.6.

Similar trends were observed for other tested video clips representing a variety of motion characteristics, but these results are omitted here to avoid redundancy.

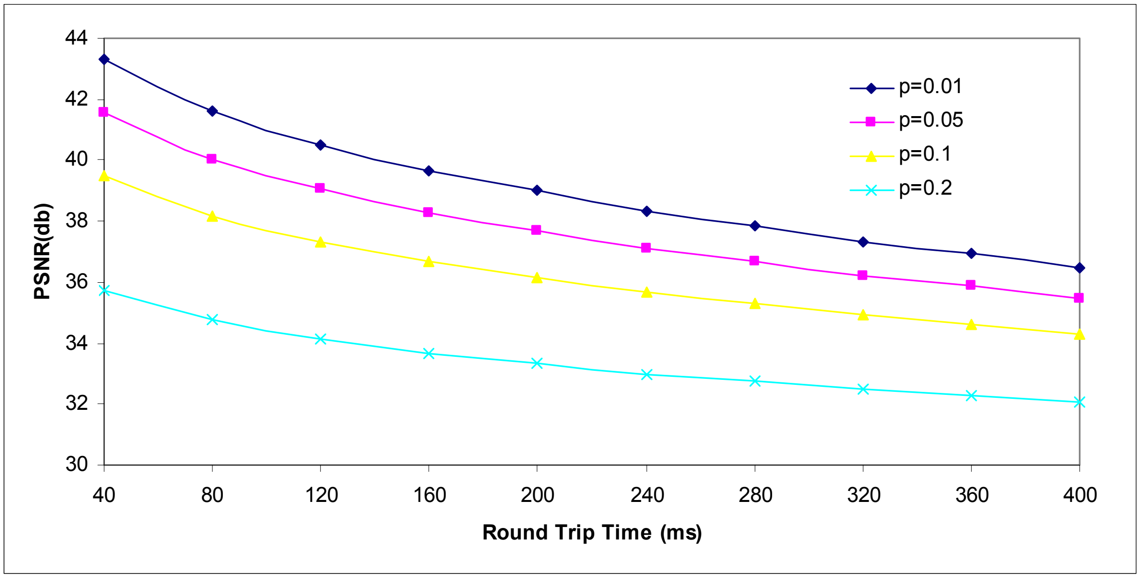

Figure 7.3.1 PSNR vs. round-trip time with RPS ACK under different loss rates (video News)

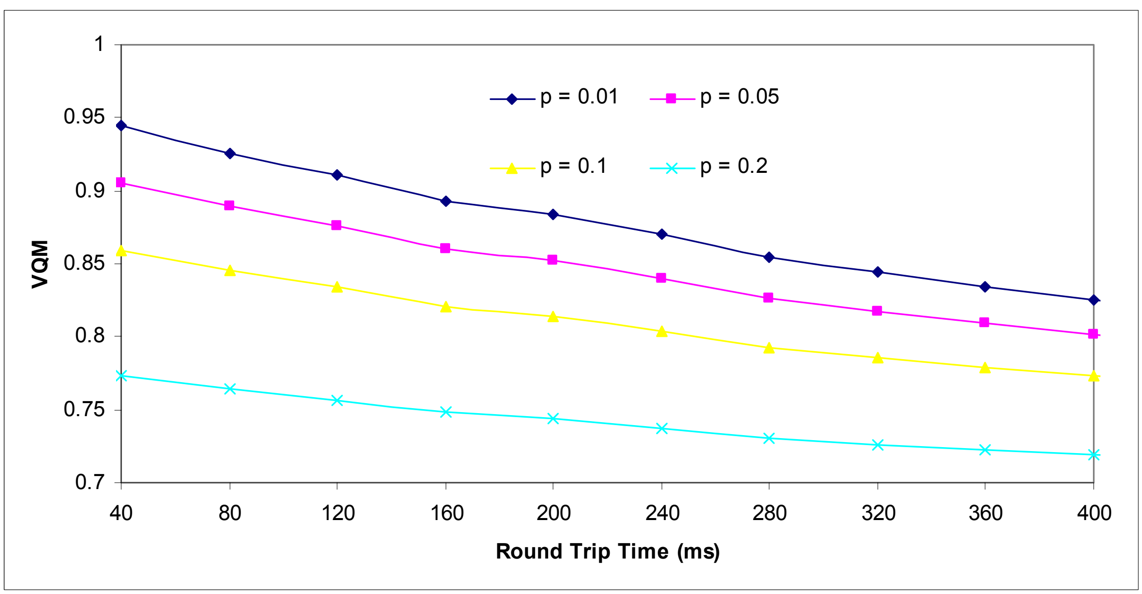

Figure 7.3.2 VQM vs. round-trip time with RPS ACK under Different Loss

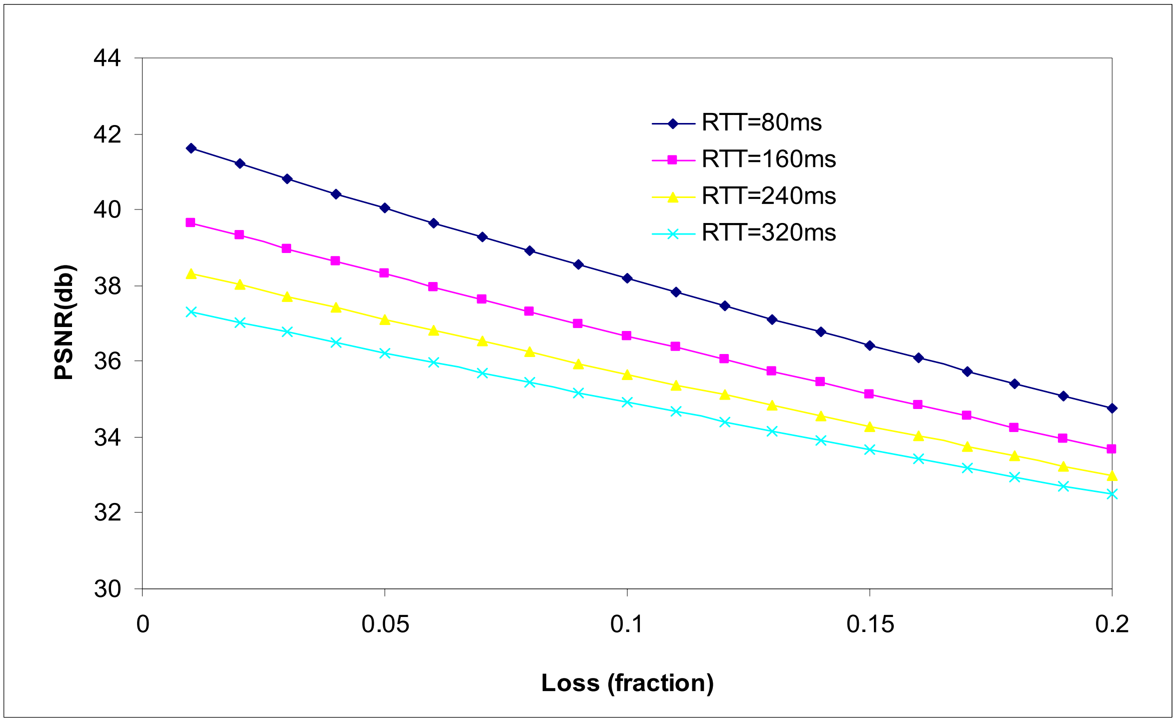

Figure 7.3.3 PSNR vs. loss with RPS ACK under different round-trip times (video News)

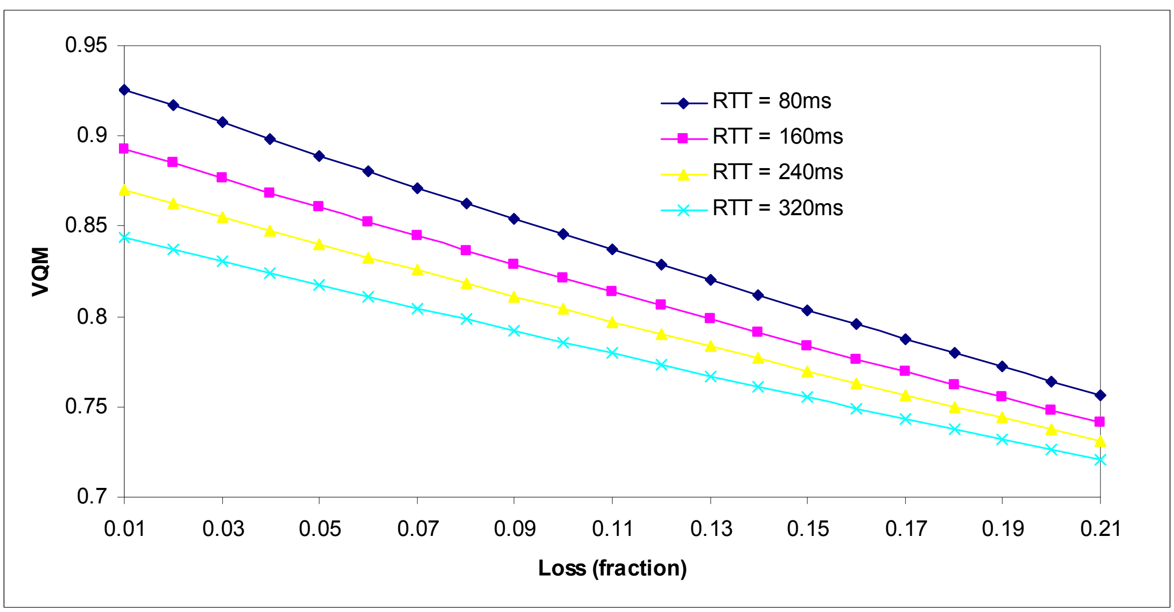

Figure 7.3.4 VQM vs. loss with RPS ACK under different round-trip times (video News)

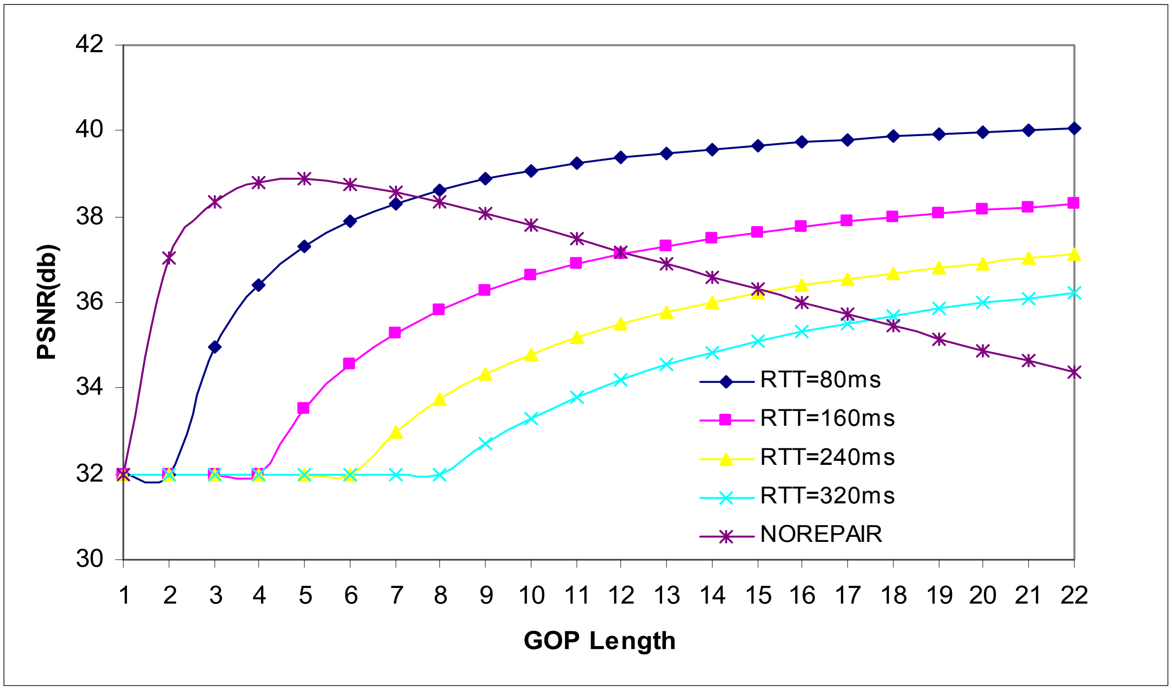

Figure 7.3.5 PSNR vs. GOP length with RPS ACK (P=0.05, video News)

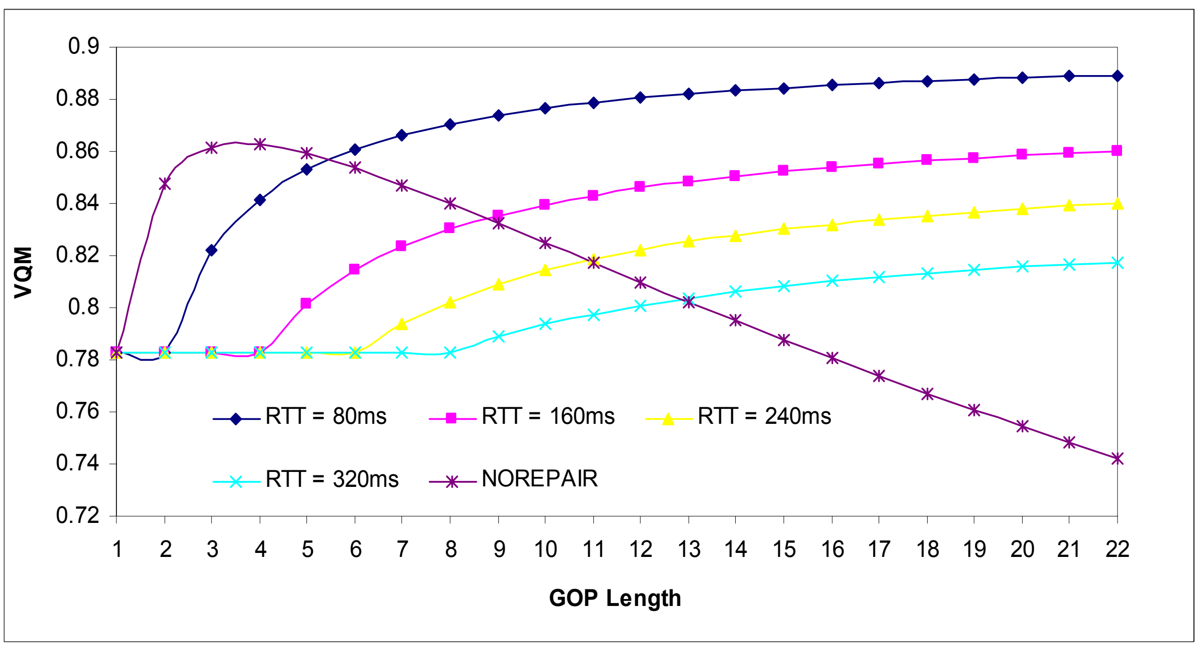

Figure 7.3.6 VQM vs. GOP length with RPS ACK (P=0.05, video News)

7.4 Intra Update

This section analyzes the performance of Intra Update under a variety of network conditions. The performance of Intra Update is mainly affected by two factors: error propagation and the coding efficiency reduction induced by Intra coding. For videos using Intra Update for repair, a transmission error propagates at least one round-trip time before the encoder can receives error message from the decoder. The encoder with Intra Update encodes a GOB in Intra mode when it is informed of the lost GOB by the decoder. The Intra coding reduces the coding efficiency. Thus to maintain a constant frame rate and bit rate, the encoder uses a coarser quantization and the overall video quality decreases. Figure 7.4.1 depicts PSNR versus fraction of intra coded GOBs for three different videos: Akiyo, News and Coastguard which represent low-, medium- and high-motion videos respectively. For all three videos, as the fraction of Intra coding increases, the video quality decreases. However, it can be observed that video quality degrades faster for low motion videos than for high motion videos. Since the intra-coded macro-blocks are independent of other GOBs, the coding efficiency for pictures containing more intra-coded macro-blocks (high motion) degrades less with an increase in Intra coding than those containing more inter-coded macro-blocks (low motion).

The impact of round-trip time on Intra Update video quality is examined next. Figure 7.4.2 depicts PSNR versus round-trip time for one video (*News*) encoded with Intra Update and GOP size 22, under four loss rates ranging from 1% to 20%. The quality for a locally concealed GOB is 50% of the best quality of a GOB. As round-trip time increases, in Figure 7.4.2, the video quality (PSNR) degrades for all loss rates since longer round-trip time causes longer duration of error propagation. However, the amount of quality degradations is not uniform. With Intra Update, video quality under the higher packet loss rates degrades faster with an increase in round-trip time than under lower packet loss rates. For Intra Update, each transmission error propagates to subsequent frames for a period of one round-trip time. Thus, higher packet loss rates induce more frequent GOB error propagation and video quality degrades more quickly with higher round-trip times. Higher packet loss rates also induce more intra-coded macro-blocks and as a result, the coding efficiency drops and video quality degrades more quickly. A similar trend can be observed when VQM is adopted as video quality metric as shown in Figure 7.4.3.

The impact of loss rate on RPS video quality is now investigated. Figure 7.4.4 provides PSNR versus loss rate curves for the *News* video encoded with Intra Update for four round-trip times ranging from 80 milliseconds to 320 milliseconds. As the loss rate increases, the video quality (PSNR) using Intra Update degrades for all round-trip times due to the increase of error propagation and the reduction of coding efficiency. However, with Intra Update, video quality under higher round-trip times degrades faster than does the video quality under lower round-trip times. With Intra Update, larger round-trip time implies longer error propagation periods that cause video quality to degrade more rapidly. A similar trend can be observed when VQM is adopted as video quality metric as shown in Figure 7.4.5.

Finally, the Intra Update model is used to investigate the impact of the GOP length on video quality. Figure 7.4.6 depicts PSNR versus GOP length for the video News encoded with Intra Update for round-trip times ranging from 80 ms to 320 ms, as well as a video with no repair. The loss rate for this experiment is 0.05. Below a GOP length of 5, video quality increases in all cases. After the GOP length reaches 5, quality for the video without RPS degrades due to error propagation. With Intra Update, when the round-trip time is 80 ms, quality increases and becomes asymptotically steady. When round-trip times are 160 ms, 240 ms and 320 ms, quality slightly decreases and becomes asymptotically steady. For all GOP lengths, videos with Intra Update perform no worse than videos without repair, and Intra Update performs better for lower round-trip times than for higher round-trip times since higher round-trip times introduce longer periods of error propagation. A similar trend can be observed when VQM is adopted as video quality metric as shown in Figure 7.4.7.

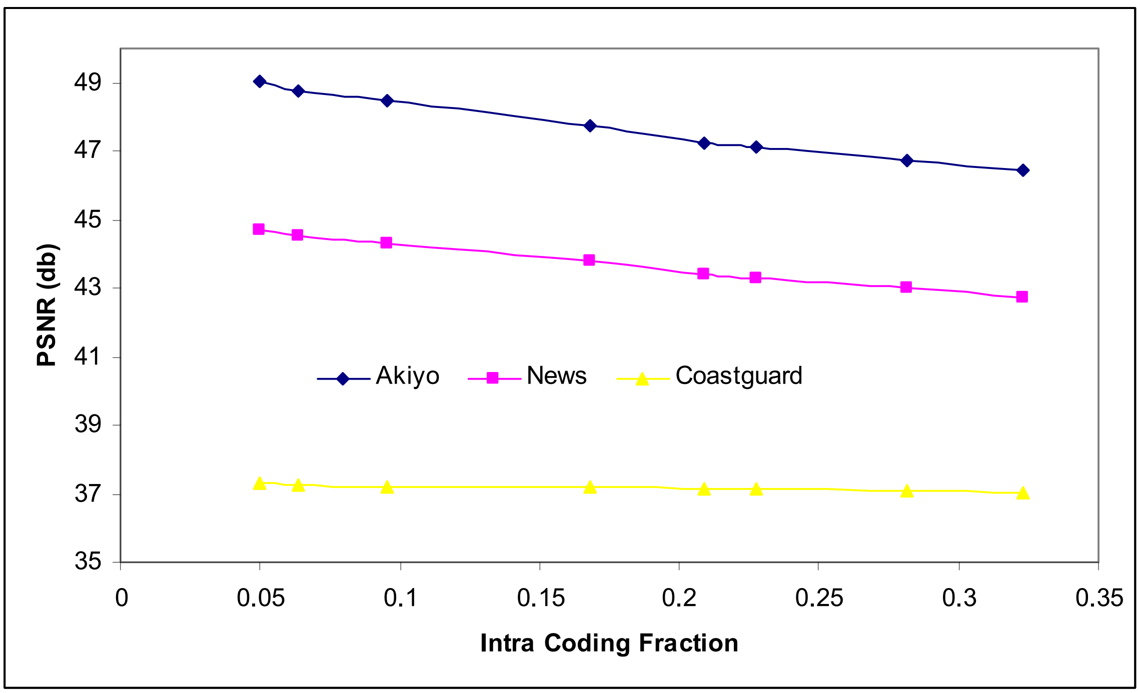

Figure 7.4.1 PSNR vs. Intra Coding Fraction for Three Videos

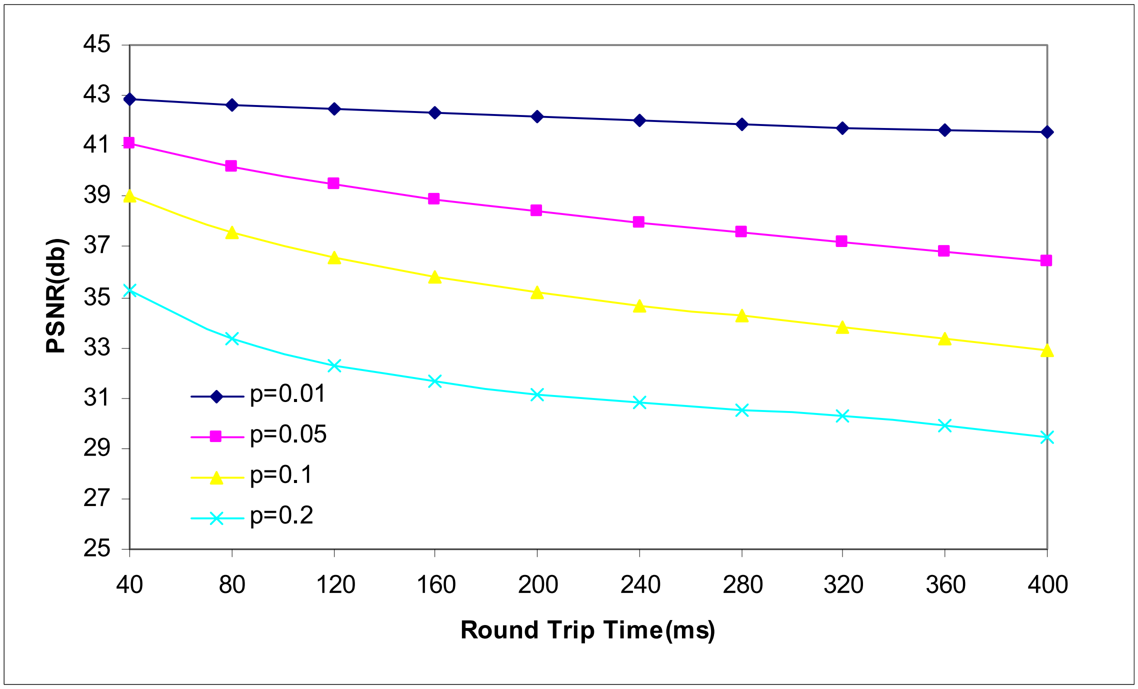

Figure 7.4.2 PSNR vs. round-trip time with Intra Update under different loss rates (video News)

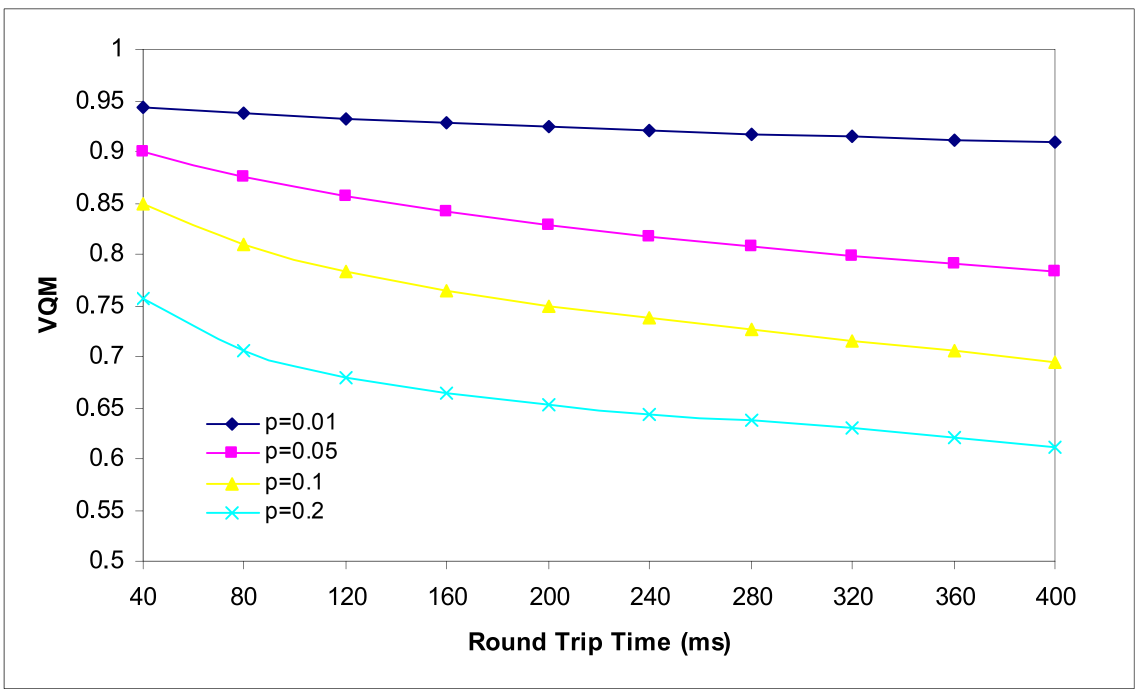

Figure 7.4.3 VQM vs. round-trip time with Intra Update under Different Loss

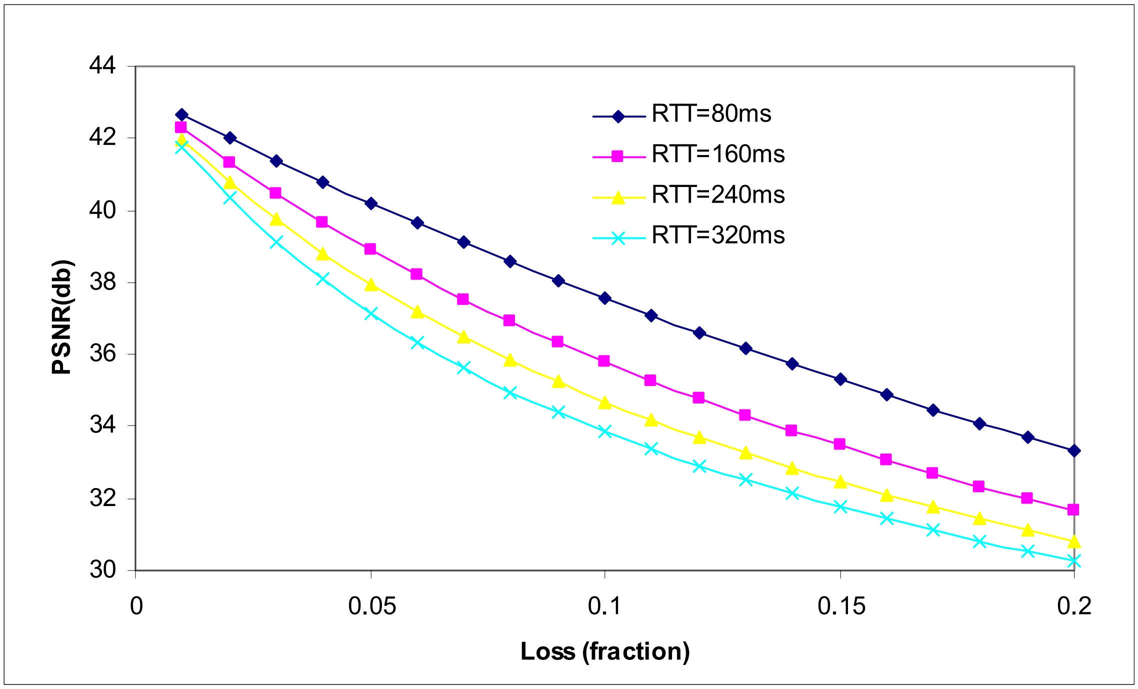

Figure 7.4.4 PSNR vs. loss with Intra Update under different round-trip times (video News)

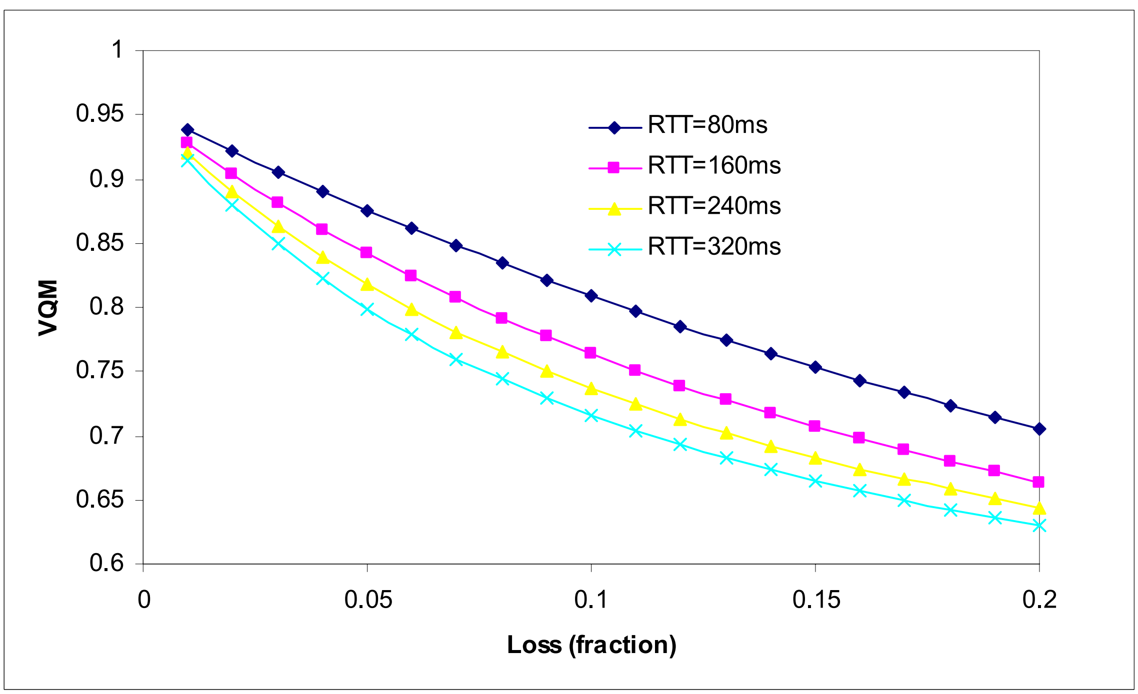

Figure 7.4.5 VQM vs. loss with Intra Update under different round-trip times (video News)

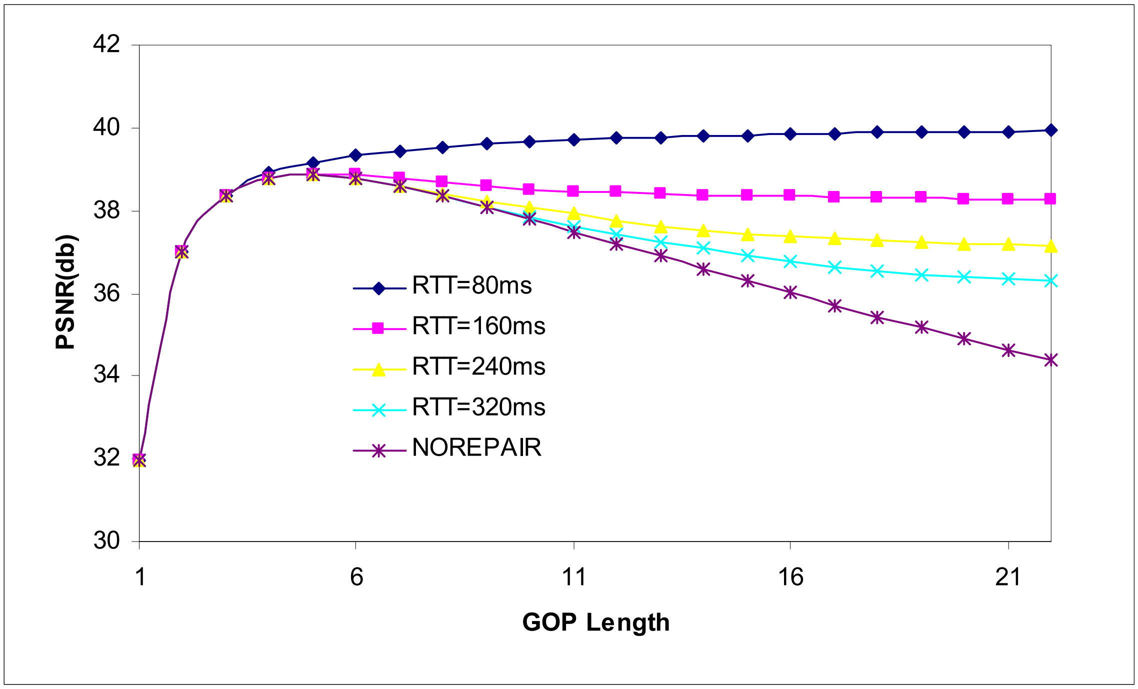

Figure 7.4.6 PSNR vs. GOP length with Intra Update (P=0.05, video News)

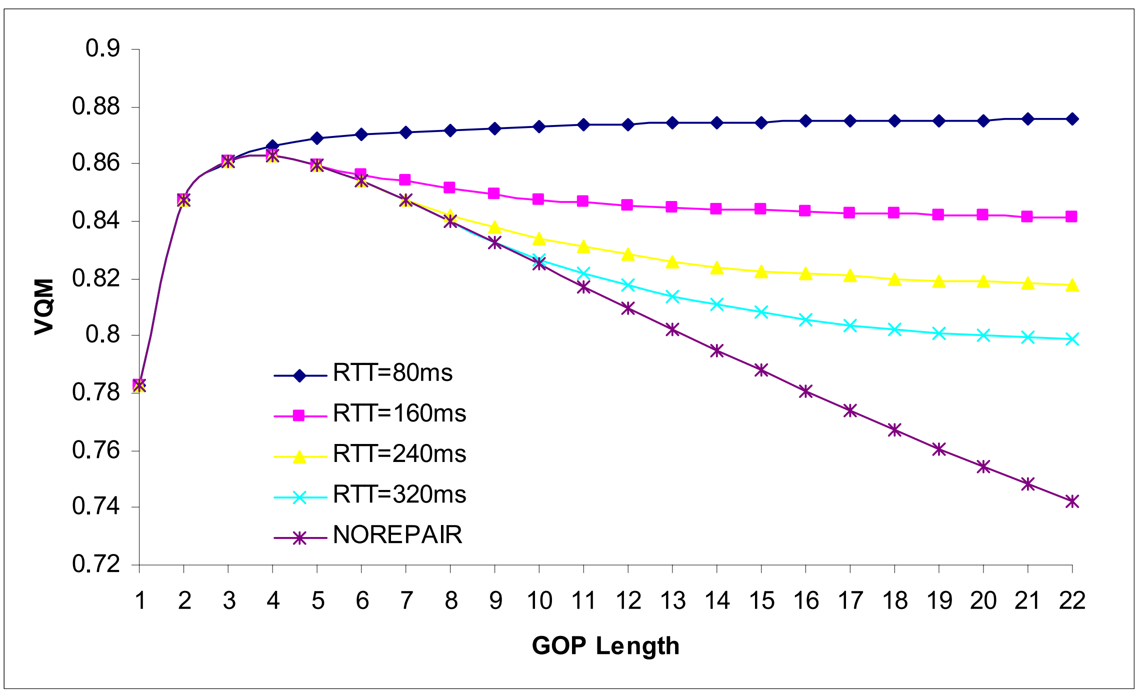

Figure 7.4.7 VQM vs. GOP length with Intra Update (P=0.05, video News)

7.5 Comparisons of Feedback-Based Error Control Schemes

In the previous sections, we used our analytical models to individually examine the performance of four feedback-based error control techniques. This section compares these four error control techniques. The objective of these comparisons is to examine the network conditions and video content that may impact the choice of technique for error control.

Both Figure 7.5.1 and 7.5.2 depict PSNR versus loss rate for the video *News* encoded using four feedback-based error control techniques but under round-trip time 80ms and 240 ms respectively. Overall, as shown in these two pictures, RPS NACK achieves the best performance when the loss rate is low while RPS ACK outperforms other repair techniques when the loss rate is high but performs the worst when the loss rate is low. Retransmission performs slightly better than Intra Update when the loss rate is low, but performs the worst when the loss rate is high.

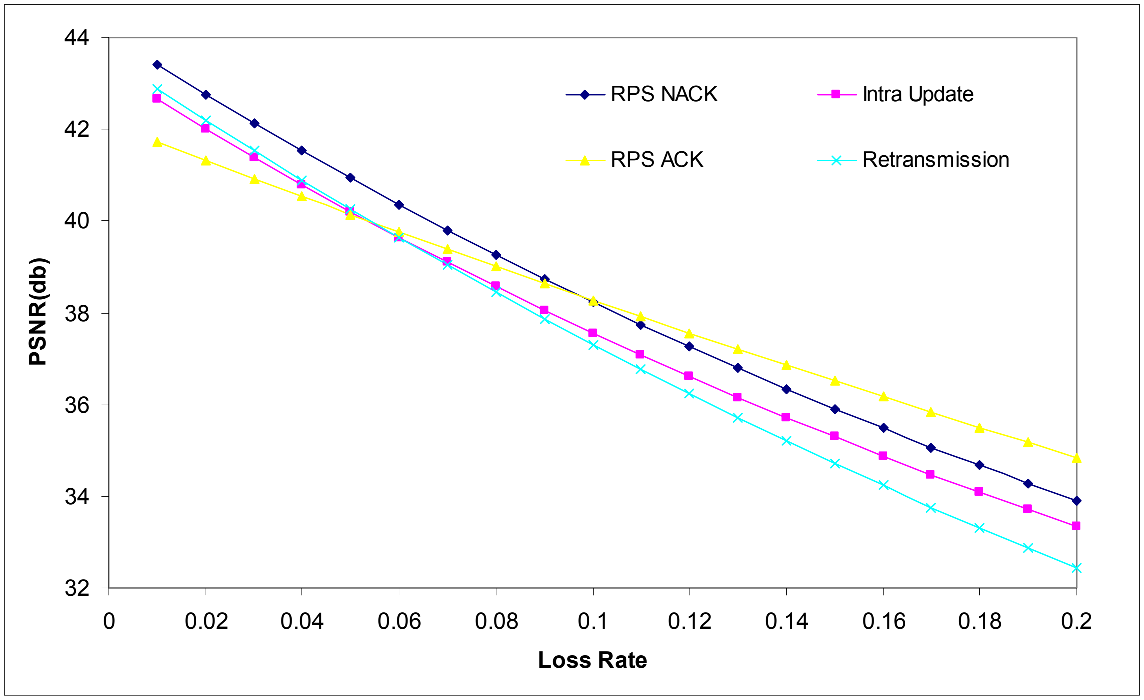

Figure 7.5.1 PSNR vs. loss for four feedback-based error control techniques (round-trip time=80ms, video News)

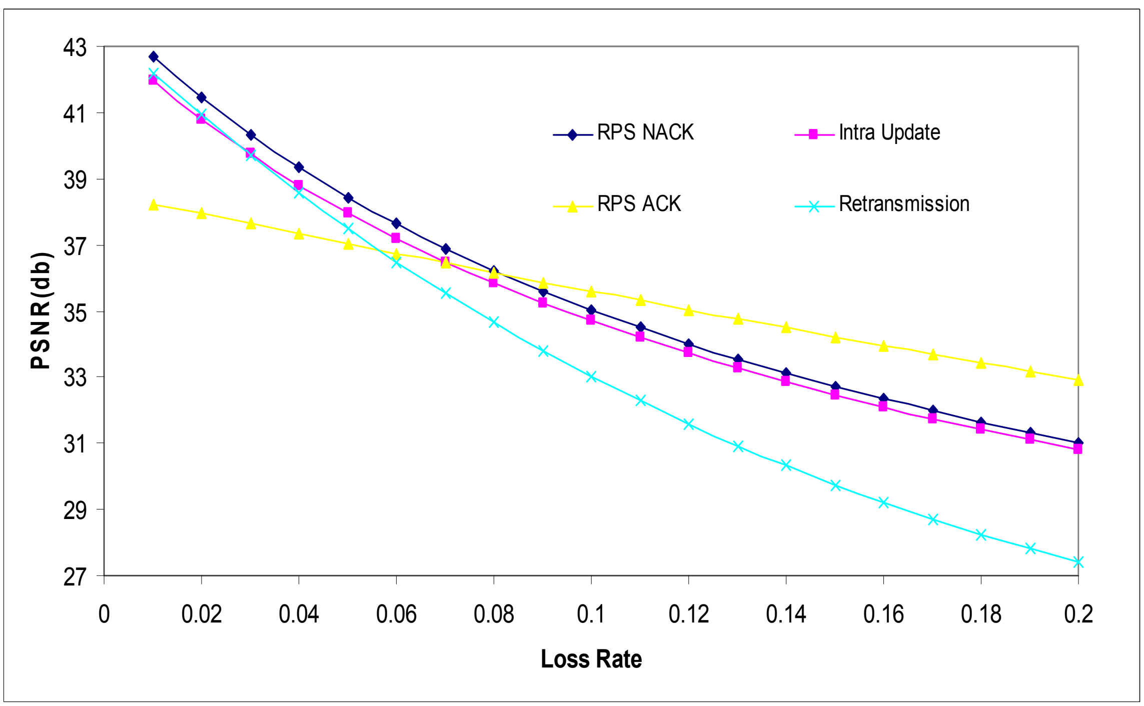

Figure 7.5.2 PSNR vs. loss for four feedback-based error control techniques (round-trip time=240ms, video News)

We next compare the performance between RPS NACK and Intra Update. During error-free transmission, both RPS NACK and Intra Update use one of the GOBs in the previous frame as a reference. However, after a transmission error, RPS NACK switches to an older but intact GOB as a reference whereas Intra Update encodes the erroneous GOB using Intra coding. Using an older GOB as a reference for prediction reduces the coding efficiency, as does using Intra coding. Therefore, the comparison of performance between RPS NACK and Intra Update can essentially be translated into comparing the impact of reference distance with the impact of Intra coding on video quality. This comparison is largely affected by video content as well as round-trip time.

Figure 7.5.3, 7.5.4 and 7.5.5 compare RPS NACK with Intra Update for three videos under round-trip time 80ms, 240ms and 400ms respectively. As shown in these three graphs, RPS NACK outperforms Intra Update when used by a low motion video (*Akiyo*) for all three round-trip times. This suggests that for low motion videos, increasing Intra coding impairs coding efficiency more than does increasing reference distance. Longer round-trip time induces longer reference distance and thus lowers coding efficiency. It can be observed that the performance gap between RPS NACK and Intra Update drops as the round-trip times increases. For videos with higher motion (*News* and *Coastguard*), the performance gap between RPS NACK and Intra Update drops and for Coastguard, there is almost no difference in performance between RPS NACK and Intra Update.

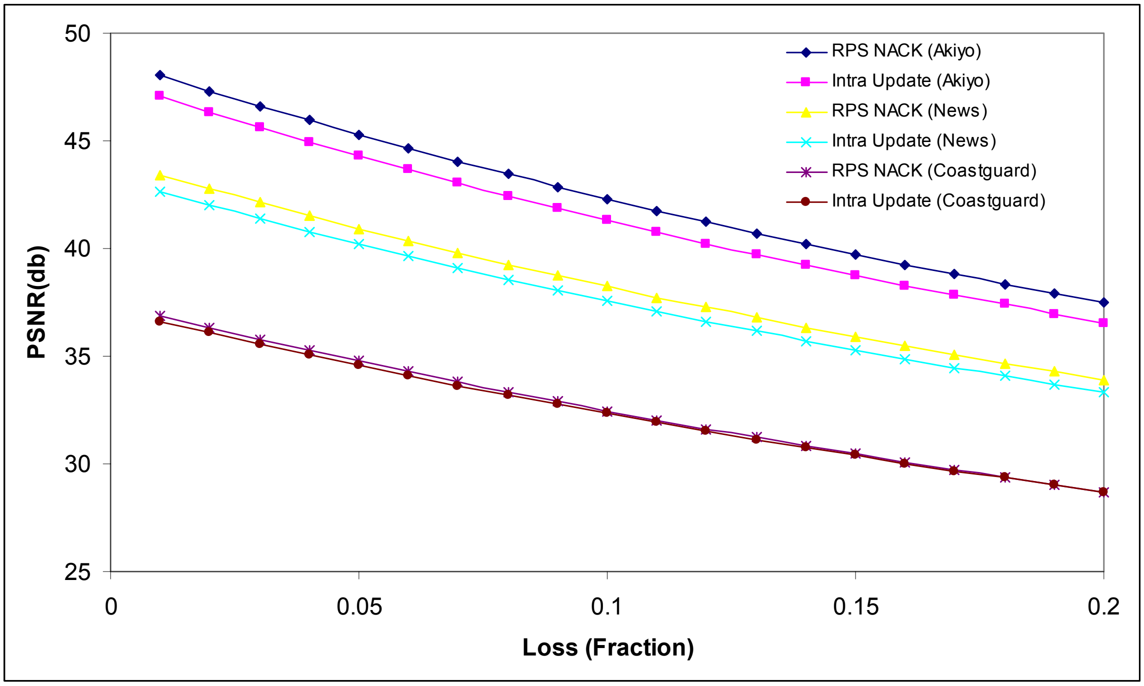

Figure 7.5.3 Comparison of RPS NACK and Intra Update with three videos (round-trip time=80ms)

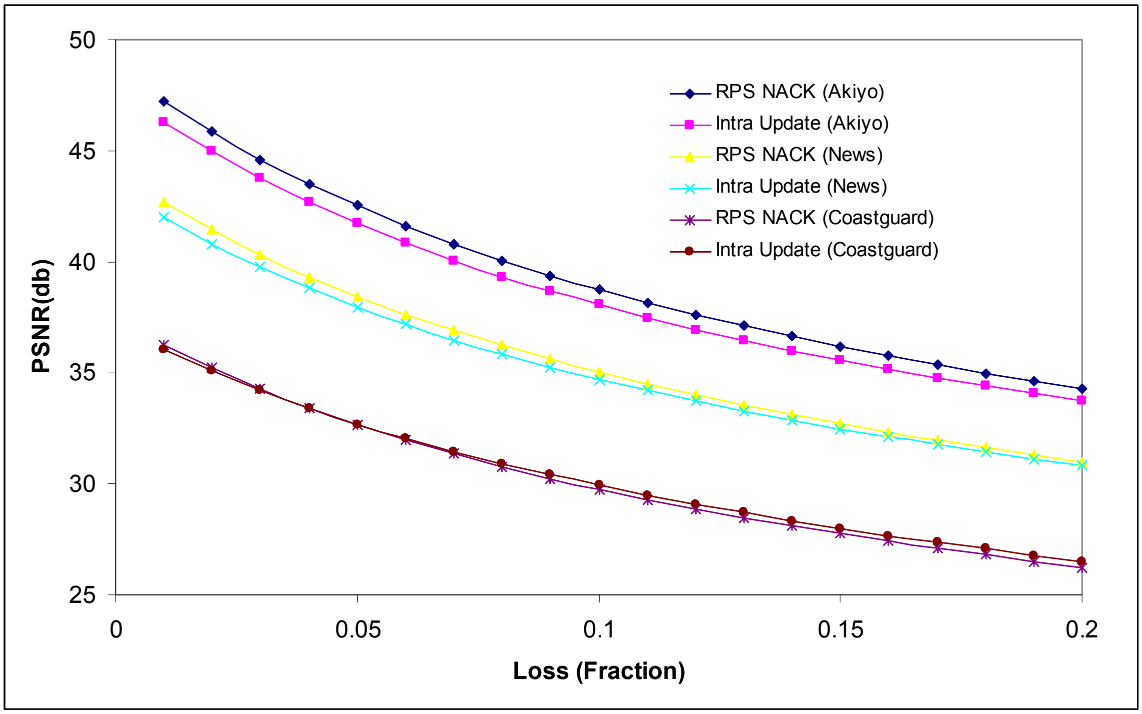

Figure 7.5.4 Comparison of RPS NACK and Intra Update with three videos (round-trip time=240ms)

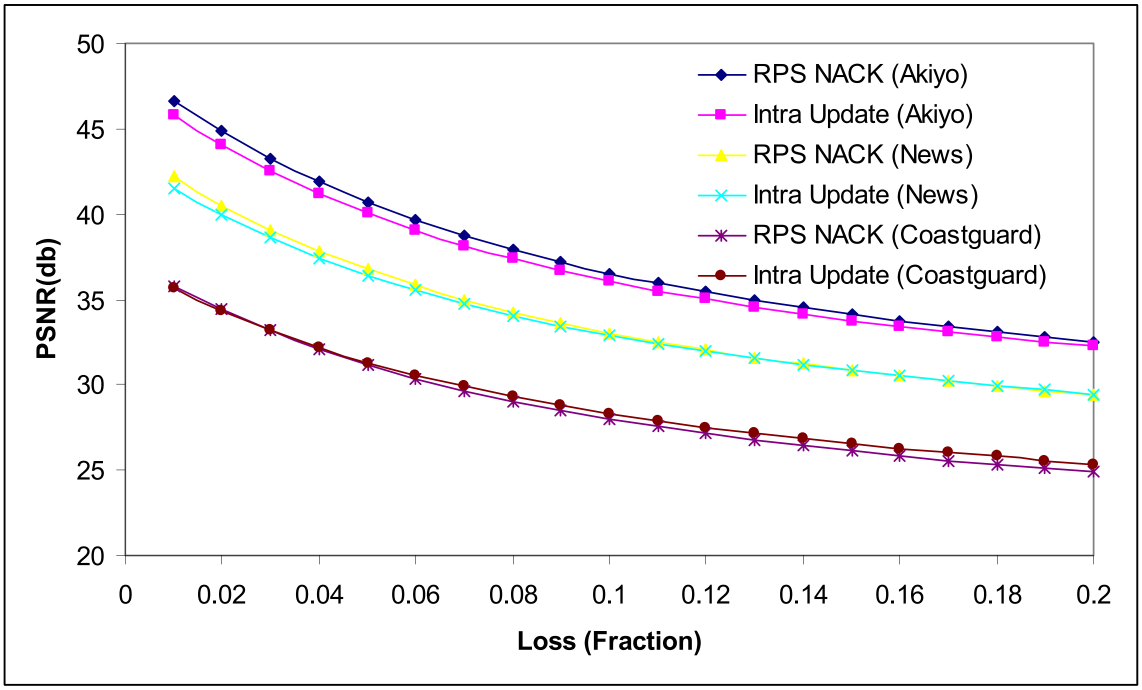

Figure 7.5.5 Comparison of RPS NACK and Intra Update with three videos (round-trip time=400ms)

We next focus on the comparison between RPS NACK and RPS ACK. Analysis thus far demonstrates that video quality for RPS ACK and RPS NACK is affected by round-trip time and packet loss rate. To make an informed choice about RPS, it is useful to know the range of packet losses within which RPS NACK performs better than RPS ACK and vise versa, and how this relationship changes with round-trip time, local concealment and video content. Figures 7.5.6-7.5.8 compares RPS ACK and RPS NACK by graphing VQM versus packet loss with each figure having a different round-trip time. All three experiments again use the *News* video clip (video content is analyzed later). As shown in Figure 7.5.6 with an 80 ms round-trip time, when the loss rate is less than 0.044, RPS NACK outperforms RPS ACK and when the loss rate is larger than 0.044, RPS ACK performs better than RPS NACK. When the round-trip time is increased from

80 ms to 160 ms in Figure 7.5.7, the same crossover point is reduced from 0.044 to 0.037. In Figure 7.5.8 with a 400 ms round-trip time, the crossover point is further reduced to 0.032. This confirms that as the round-trip time increases, the video quality with RPS NACK degrades faster than RPS ACK. For RPS NACK, increased round-trip time produces longer GOB error propagation; whereas for RPS ACK, increased round-trip time yields higher GOB reference distances. Increasing error propagation does more harm to video quality than does increasing reference distance.

We further investigate how local concealment affects the crossover point. Figure 7.5.9 shows the crossover points when the quality for a locally concealed GOB is 90%, 50% and 10% of the best quality of a GOB respectively with a round-trip time of 160ms. When the quality for a locally concealed GOB is 90% of the best quality of a GOB, the packet loss crossover point is 0.13. When the locally concealed quality is reduced to 50% of the best quality of a GOB, the crossover point is reduced to 0.037 and further reduced to 0.01 when the locally concealed quality is reduced to 10% of the best quality of a GOB. This suggests that RPS NACK outperforms RPS ACK over a wider range when there is better local concealment.

The relationship between crossover point and round-trip time for different video content is investigated next. Figure 7.5.10 shows the quality crossover point versus round-trip time for the six videos in Table 5.1. For loss rates above the trend-lines, RPS ACK performs better than RPS NACK while for loss rates below the trend-lines, RPS NACK performs better than RPS ACK. As round-trip time is increased, the crossover points are lowered for all videos. This suggests that regardless of the video content, increasing the error propagation is more harmful to video quality than increasing reference distance. For a fixed round-trip time, the crossover points for low-motion videos are higher than the crossover points for high-motion videos. This implies that RPS ACK outperform RPS NACK over a wider range of packet loss rates for high-motion videos than for low-motion videos. High-motion videos are less sensitive to changes in reference distance and thus can achieve better video quality with RPS ACK than can low-motion videos. Similar trends were observed for videos using PSNR as the quality metric as depicted in Figure 7.5.11.

Figure 7.5.12 shows crossover point versus round-trip time for two videos evaluated with both PSNR and VQM for quality metrics. Figure 7.5.12 clearly shows that for both videos the crossover points when using PSNR to measure quality are higher than when using VQM to measure quality. For instance, for the News video, when the round-trip time is 200 ms, the cross-point for PSNR is 0.085, whereas for VQM, the cross-point is 0.04. Hence, the range of loss rates where RPS NACK outperforms RPS ACK is smaller when using VQM to predict quality than when using PSNR to predict quality. This implies that VQM, a metric designed to incorporate temporal as well as spatial aspects of video, is more sensitive to loss than PSNR, a metric that captures only spatial degradations in video.

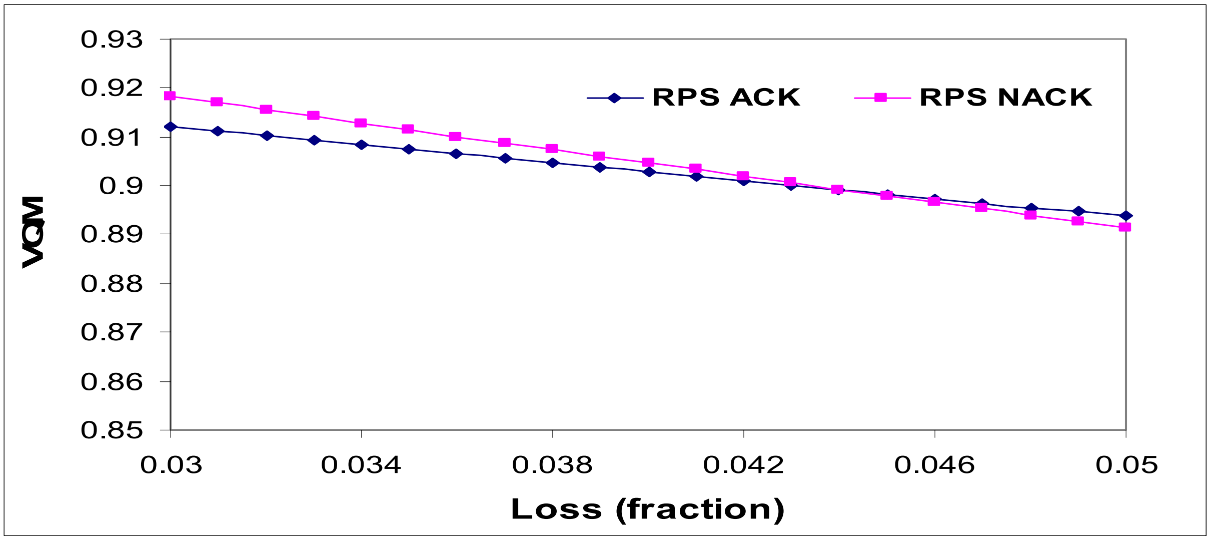

Figure 7.5.6 RPS NACK vs. RPS ACK (round-trip time = 80 ms)

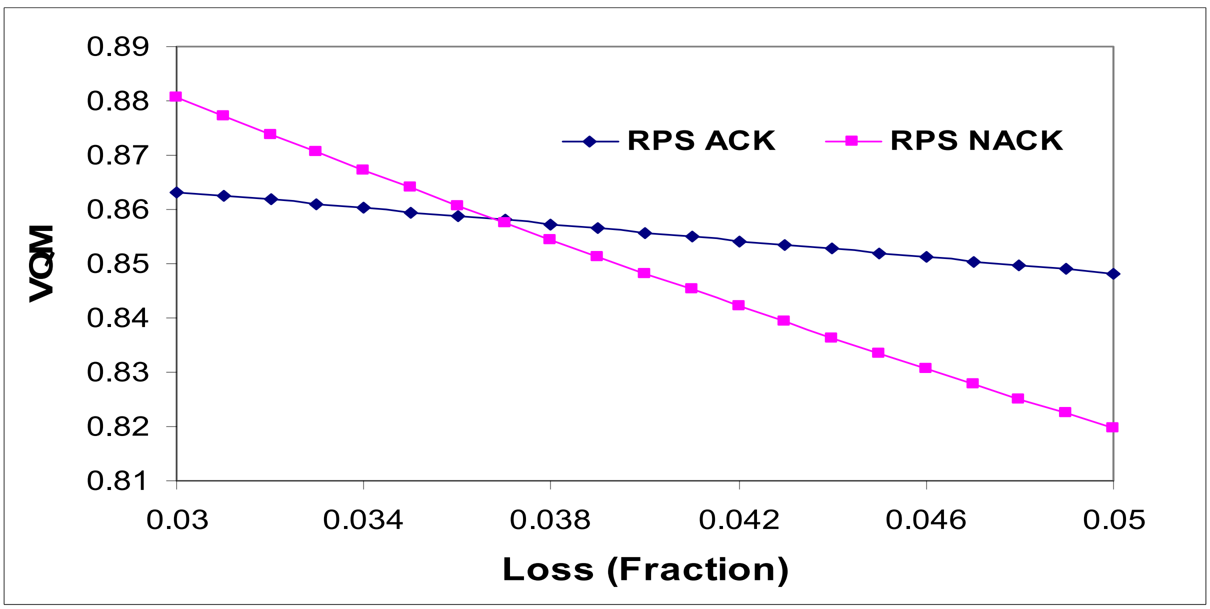

Figure 7.5.7 RPS NACK vs. RPS ACK (round-trip time = 160 ms)

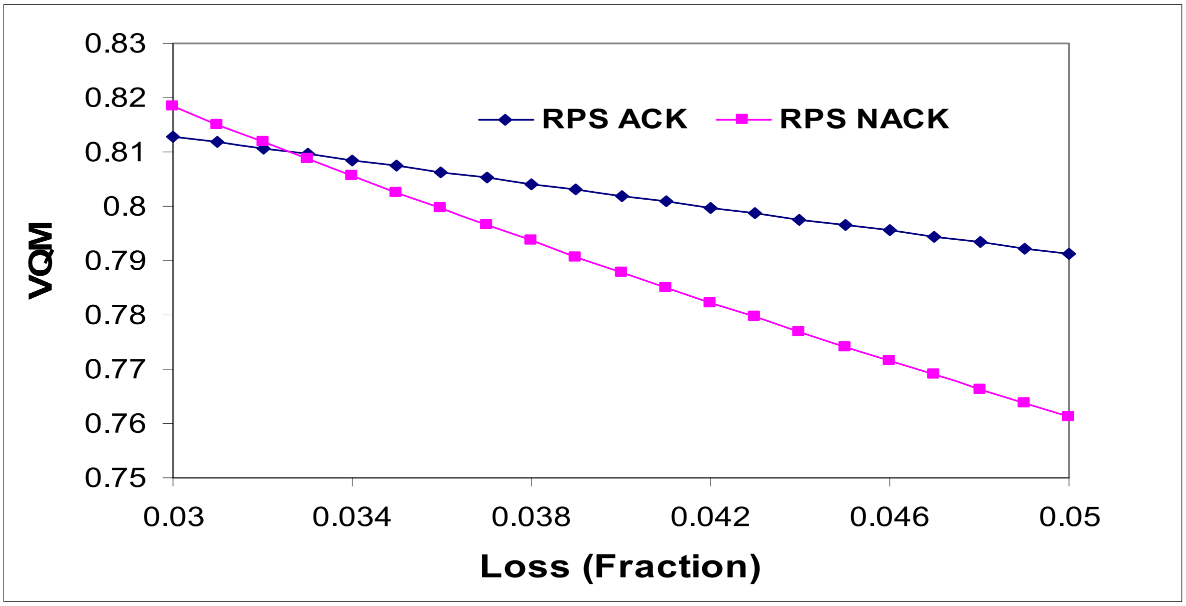

Figure 7.5.8 RPS NACK vs. RPS ACK (round-trip time = 400 ms)

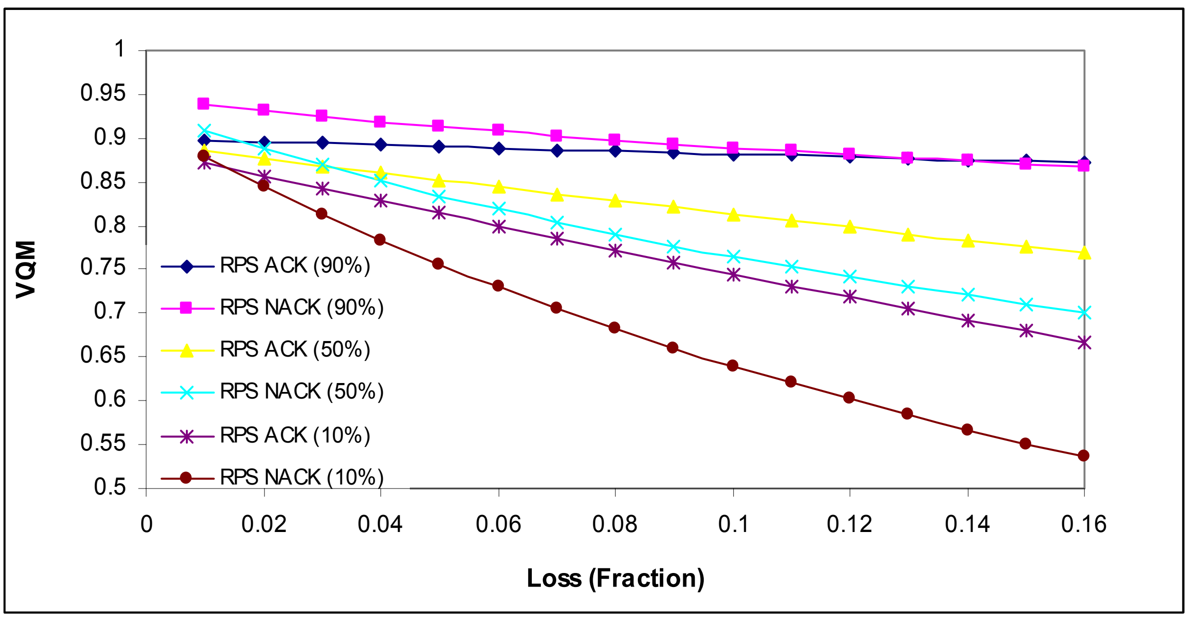

Figure 7.5.9 RPS ACK vs. RPS NACK by varying quality for locally concealed GOBs

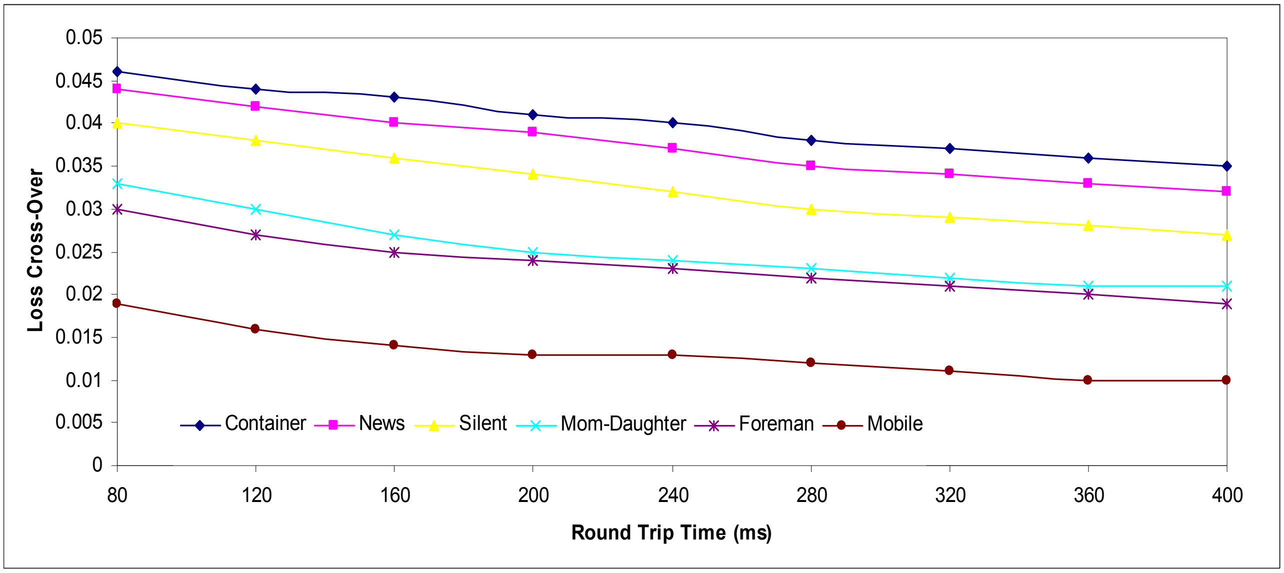

Figure 7.5.10 The loss crossover point for loss vs. round-trip time for six video clips using VQM

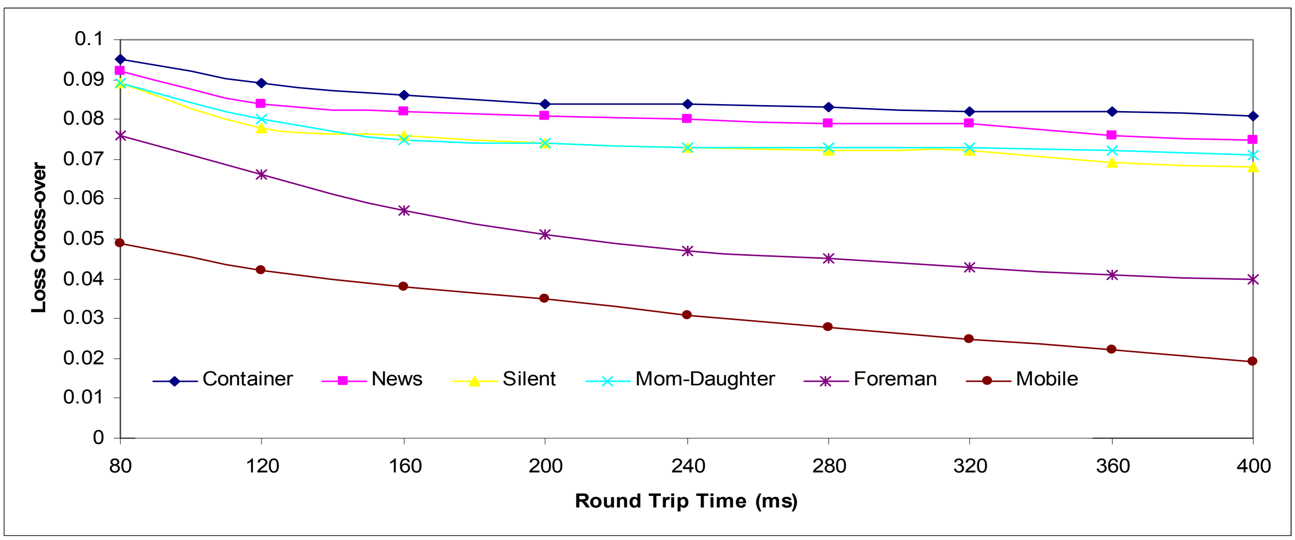

Figure 7.5.11 The loss crossover point for loss vs. round-trip time for six video clips using PSNR

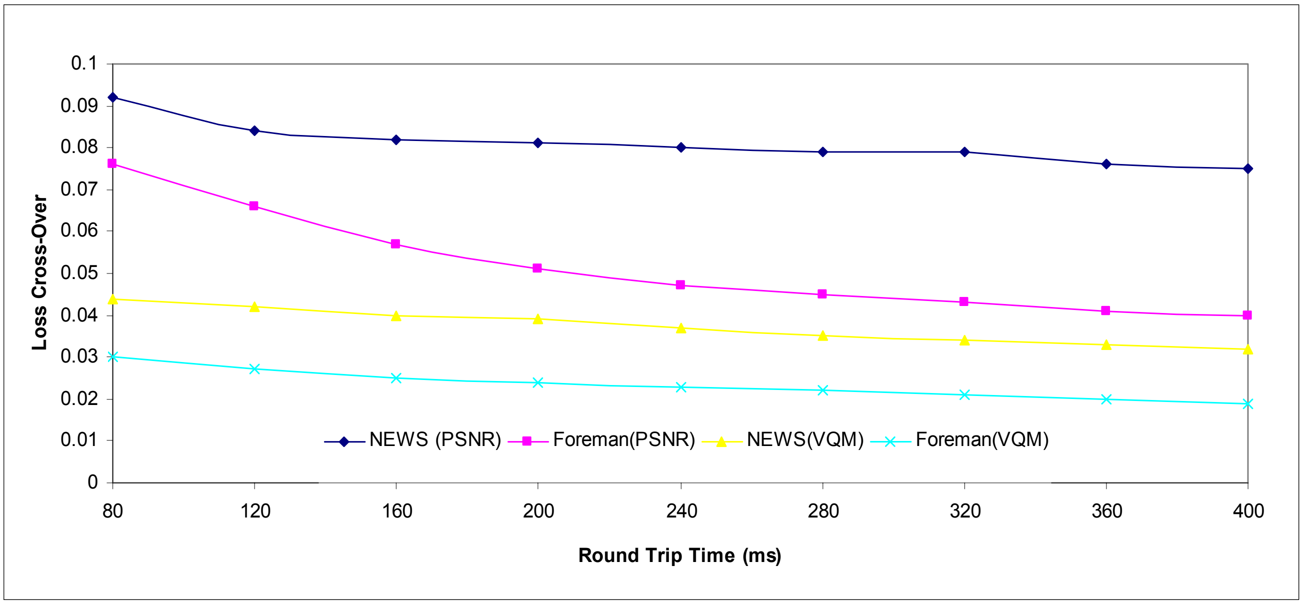

Figure 7.5.12 The loss crossover point for loss vs. round-trip time for two videos using both PSNR and VQM

Chapter 8

Conclusions

Despite many recent improvements to computer networks, streaming video quality may still be degraded by lost data packets. A single missing video packet can propagate errors to many subsequent video frames due to inter-frame encoding dependencies. Feedback-based error control techniques including Reference Picture Selection (RPS), Intra Update and Retransmission use feedback information from the decoder to adjust coding parameters to reduce error propagation due to data loss. They have been shown to be more effective than trying to conceal the error at the encoder or decoder alone since they allow the encoder and decoder to cooperate in the error control process. However, there has been no systematic exploration of the impact of video content and network conditions on the performance of feedback-based error control techniques. In particular, the impacts of packet loss, round-trip delay, network capacity constraint, video motion and reference distance on quality of videos using RPS, Intra Update and Retransmission have not been thoroughly studied.

This thesis presents the analytical models for the three major feedback-based error control techniques, including Retransmission (Full and Partial), Reference Picture Selection (NACK and ACK modes) and Intra Update. These feedback-based error control techniques have been included in H.263/H.264 and MPEG4, the state of the art in video compression standards. Given the estimated round-trip time, packet loss rate, and network capacity constraint, our models can predict the achievable video quality for a streaming video with retransmission, Intra Update and RPS (NACK and ACK mode) over a lossy network. In order to exploit our analytical models, a series of studies have been conducted to explore the effect of reference distance, capacity constraint and Intra coding on video quality. The accuracy of our analytical models in predicting the video quality under different network conditions is validated through simulations. These models are used to examine the behavior of feedback-based error control schemes under a variety of network conditions and video content through a series of analytic experiments.

8.1 Summary of Feedback-Based Error Control Technique

Reference Picture Selection (RPS) is an established video repair technique that allows a video encoder to select one of several previous frames as a reference for predictive encoding of subsequent Group of Blocks (GOBs). RPS operates in one of two modes: NACK or ACK. RPS NACK uses the previous GOB as a reference until an error is reported and then it uses an older GOB as a reference to stop error propagation. RPS NACK cannot eliminate error propagation since a packet loss results in error propagation until the NACK reaches the encoder and the newly encoded video travels back to the decoder - about a round-trip time. RPS ACK only uses acknowledged GOBs as a reference and thus eliminates error propagation entirely. However, using an older GOB as a reference reduces coding efficiency, especially for high round-trip times, and results in lower video quality. Therefore, both RPS NACK and RPS ACK have merits and drawbacks with the best choice between choosing the best RPS mode, NACK or ACK, depending upon network conditions, such as round-trip time and loss rate, and upon the video content, such as high motion or low motion.

Intra Update is similar to RPS NACK in that during error-free transmission, the encoder uses one of the GOBs in the previous frame as a reference. However, when a NACK from the decoder is received, instead of using an older, intact GOB as a reference, Intra Update simply encodes the next portion of a frame using Intra coding. Therefore, the comparison of performance between RPS NACK and Intra Update can essentially be translated into comparing the impact of reference distance with the impact of Intra coding on video quality. This comparison is largely affected by video content as well as round-trip time.

The retransmission technique considered here is different from conventional retransmission in that packets arriving after their display time are not discarded but instead used to reduce error propagation by repairing all subsequent frames. For Retransmission, a transmission error propagates at least one round-trip time, the same as RPS NACK and Intra Update. The main performance difference between Retransmission and RPS NACK and Intra Update arises because a successful repair using Retransmission requires all the GOBs in the Retransmission Range (RR)[16] to be received correctly. A large round-trip time implies a larger Retransmission Range and thus a greater probability that a transmission error may occur for a GOB within the Retransmission Range. Thus, Retransmission performs worse than RPS NACK and Intra Update when the round-trip time is high.

[16] Please refer to Chapter 4 for the definition of Retransmission Range.

When the network capacity is constrained, retransmission of every lost packet may not be feasible. We propose a Partial Retransmission scheme in which only a fraction of lost packets, those with highest priority, are retransmitted. In some cases, Partial Retransmission can achieve better performance than Full Retransmission since extra bit-rate consumption from Full Retransmission can result in greater quality degradation than Full Retransmission repairs.

8.2 Impact of Reference Distance on Video Quality

Since RPS may have to use an older reference frame for prediction, the coding efficiency decreases as the reference distance increases since the similarity between the encoding frame and the reference frame decreases. If the network capacity is constrained, the video quality degrades as the coding efficiency drops. In order to understand how changing the reference distance affects the performance of RPS, a series of experiments are conducted to explore the relationship between video quality and reference distance. A set of video clips with a variety of visual motion are selected for study, and the video sequences are shuffled to change the reference distances. For each reshuffled video sequence, an H.264 encoder encodes the sequence and measures video quality with PSNR and VQM, two popular video quality metrics.

From analysis of the experimental results, the relationship between video quality and reference distance can be determined:

- Video quality degrades as reference distance increases.
- The degree of the video quality degradation is affected by the video content. The quality for videos with high motion tends to degrade with an increase of

reference distance slower than the quality for videos with low motion. This is largely because high-motion videos have a much larger number of inter-coded macro-blocks (P-blocks) and are thus less sensitive to the change in reference distance than are low-motion videos.

- Although these findings hold for both the PSNR and VQM measures of video quality, the characterizations of the relationship between video quality and reference distance are different. While the relationship between PSNR and reference distance can be characterized using a logarithmic function, with VQM as the video quality metric, the same relationship is best characterized using a linear function.

8.3 Analytical Models for Feedback-based Error Controls

This thesis compares feedback-based error control schemes under various network conditions and video content using a set of analytical models. Our models characterize these feedback-based error control techniques, incorporating the impact of reference distance, bandwidth constraint, and Intra coding on video quality, prediction dependency among GOBs in the reference chain and Group of Picture (GOP) length. Given a variety of network characteristics including packet-loss rate, round-trip time, capacity constraints, and measured video quality derived from empirical studies, our models predict average video quality for videos using feedback-based error controls.

The accuracy of our analytical models in predicting the video quality under different network conditions is validated through simulations. The simulations modify the input video sequences based on the given loss probability and round-trip delay to mimic the effect of packet loss as well as the change of reference distance on the video quality. Validation through simulation suggests our models accurately predict video quality.

8.4 Major Results of Analytic Experiments

Analytic experiments over a range of loss rates, round-trip times and video contents using the models show:

- RPS NACK achieves the best performance among feedback-based repair techniques when loss rate is low while RPS ACK outperforms other repair techniques when loss rate is high. However RPS ACK performs the worst when loss rate is low. Retransmission performs the worst when the loss rate is high.
- High loss rates degrade video quality for both RPS ACK and RPS NACK. However, RPS ACK performs roughly 7% better than RPS NACK when packet loss rate is high; conversely, RPS NACK yields up to 11% better video quality than RPS ACK under low packet loss conditions.
- For a given latency, the loss rate range where RPS ACK produces better video quality than RPS NACK is 2 times larger for low motion videos than it is for high motion videos.
- In general, better methods of local concealment increase the range where RPS NACK outperforms RPS ACK. For example, when the quality for a locally concealed GOB is increased from 10% of the best quality of a GOB to 90%, the range where RPS NACK outperforms RPS ACK is increased roughly from 1% to 11%.

- Videos with RPS NACK always perform the same or better than videos without repair. However, when small GOP sizes are used, videos without repair perform up to 10% better than videos with RPS ACK.
- Although the above trends hold for both VQM and PSNR, when VQM is the video quality metric the performance results are much more sensitive to network loss. For instance, the range where RPS NACK outperforms RPS ACK when using VQM to measure video quality is about half of the range when using PSNR as video quality metric.
- RPS NACK outperforms Intra Update for low-motion videos. However, the performance gap between RPS NACK and Intra Update drops from 3% to 1% when round-trip time increases. For high-motion videos, there is almost no difference in performance between RPS NACK and Intra Update.
- Retransmission is effective only when the round-trip time is low. When the round-trip time is high, Partial Retransmission achieves almost the same performance as Full Retransmission. Retransmission of 50% of loss packets achieves the best effectiveness in terms of the ratio of performance gain over bit-rate cost.
- Although the performance of feedback-based error control techniques are affected by a number of factors including packet loss, round-trip time, network capacity constraint, reference distance, Intra coding, and motion in video, the impact of these factors may vary depending upon which error control technique is chosen. RPS ACK is more sensitive to round-trip time whereas RPS NACK, Intra

Update and Retransmission are more sensitive to packet loss. Capacity constraint play very important role in Retransmission performance since retransmission of lost packets consumes extra bandwidth. Intra-coding has great impact on the quality of videos using Intra Update since Intra Update relies on Intra coding to stop error propagation; reference distance has greater impact on RPS ACK than on RPS NACK. For a given round-trip time, RPS NACK achieves better performance for low-motion videos than for high-motion videos whereas RPS ACK performs better for high-motion videos than for low-motion videos.

8.5 Major Contributions

This thesis has the following major contributions:

1. A systematic study of the effects of reference distance on video quality for a range of video coding conditions. A set of video clips with a variety of motions are selected for study, and the video sequences are shuffled to change the reference distances. For each reshuffled video sequence, an H.264 encoder encodes the sequence and measures video quality with PSNR and VQM.

2. Two utility functions that characterize the impact of reference distance on video quality based upon the study. While the relationship between PSNR and reference distance can be characterized using a logarithmic function, with VQM as the video quality metric, the same relationship can be characterized using a linear function.

3. Modeling the prediction dependency among GOBs for RPS NACK and Intra Update using a binary tree. Based on these two models, the probabilities of

correctly decoding a GOB encoded with RPS NACK or Intra Update can be calculated.

4. A study of the impact of bandwidth constraint on video quality in terms of VQM and PSNR. For both video quality metrics, the impact of bandwidth constraints on video quality can be characterized using a logarithmic function.

5. Analytical models for feedback-based error control techniques including Full Retransmission, Partial Retransmission, RPS ACK, RPS NACK and Intra Update. Our models characterize these feedback-based error control techniques, incorporating the impact of reference distance, bandwidth constraint, and Intra coding on video quality, prediction dependency among GOBs in the reference chain and Group of Picture (GOP) length.

6. Simulations that verify the accuracy of our analytical models. The simulations modify the input video sequences based on the given loss probability and round-trip delay to mimic the effect of packet loss as well as change of reference distance on video quality.

7. Analytic experiments over a range of loss rates, round-trip times and video contents using our models. The experiments explore a wide range of factors that may impact the performance of feedback-based error control techniques. The analysis based on these experiments is useful for helping select the best feedback-based repair techniques for improving video quality.

8.6 Recommendations on Selecting Feedback-based Error Control Techniques

Our analytical models and experiments shows that the performance of feedback-based error control techniques is affected by a number of factors: packet loss, round-trip time, network capacity constraint, reference distance, Intra coding, and motion in video. Therefore, the choice among RPS, Retransmission and Intra Update depends on the application requirements, network conditions, quality of service (QoS), and video content. Table 8.1 shows the suggested feedback-based error control techniques with different network conditions and amounts of video motion.

Loss Rate (p)	Round-Trip Time	Video Motion	Suggested Error Control Techniques
Low	Low	Low	RPS NACK
Low	High	Medium/High	RPS NACK, Intra Update
Medium	Low	Medium	RPS NACK Intra Update
Medium	Low/High	Low	RPS NACK
Medium	Low/High	High	RPS ACK
Medium	High	Medium	RPS ACK
High	Low/High	High/Medium/Low	RPS ACK

Table 8.1 Suggested feedback-based error control techniques; loss rate: High ($p>5\%$), Medium ($2\%<p<5\%$), Low ($<2\%$); round-trip time: Low (<160 ms), High (>400 ms)

- In a network environment where the loss rate is low (below 2%), such as ISDN or private LANs, either RPS NACK or Intra Update can be chosen for error repair. However, when the round-trip time is low and the intensity of video motion is low, RPS NACK performs significantly better than Intra Update and thus is a better choice for error repair. When the round-trip time is high and the intensity of video motion is high or medium, Intra Update performs nearly as well as RPS NACK, thus both of them can be chosen for error repair.

- In a network environment where the loss rate is medium (between 2% and 5%), such as LAN or Internet, the choice of feedback-based error control techniques depends on the round-trip time and the intensity of video motion. For high-motion videos, RPS ACK is the best choice for error repair; whereas for low-motion videos, RPS NACK is the best choice. For medium-motion videos, when the round-trip time is high, RPS ACK is the best choice; when the round-trip time is low, either RPS NACK or Intra Update can be chosen for error repair.

- In a network environment where the loss rate is high (5% and above), RPS ACK is the best choice for error repair. However, when the network capacity is constrained and the back channel uses part of the total bit-rate budget, RPS ACK may not be desirable choice since it requires frequent transmission of feedback messages. In this case, RPS NACK or Intra Update could be an alternative selection for error repair. In a lossy network environment, round-trip times and loss rates may change rapidly. In such environments, the encoder should dynamically adjust the error repair technique. For instance, when the loss rate is high, the encoder could switch to RPS ACK; whereas the loss rate is low, the encoder could switch to RPS NACK.

- In any circumstances, RPS or Intra Update is a better choice than Retransmission. However, Retransmission combined with playout buffering may be desirable for non-interactive video applications such as Internet video streaming and broadcasting due to its simplicity and wide deployment.

Chapter 9

Future Work

This chapter presents some possible future work that can be extended from this dissertation.

1. This thesis adopts both PSNR and VQM as video quality metrics. However, our analytical models make no assumption on specific video quality metrics. Future work could explore and incorporate other existing video quality metrics or develop a new quality metric that has better correlation with user perceptual quality. However, prior to introducing a new quality metric into our models, a study has to be conducted to explore the impact of changing reference distance on this new quality metric.
2. Our analytical models assume erroneously-decoded GOBs are repaired by local concealment and make no assumption on specific local concealment techniques. However, our analytic experiments show that the effectiveness of local concealment techniques affects the quality of the repaired video and thus the overall evaluation of feedback-based repair techniques. Future work could further investigate how local concealment may affect the choice of feedback-based repair techniques under various network conditions and video content.
3. Our analytical experiments assume independent packet loss with a random loss distribution. This is an assumption typically made by some analytic models and

well represents many computer networks. However, in some network situations, packet loss may be bursty, such as in wireless environments. Incorporating bursty loss requires fundamentally changing our current models. One future work could extend the analytical experiments to measure how inaccurate our models are in the presence of bursty loss.

4. In our analytical experiments, round-trip times and loss rates remained fixed for the duration of each video flow. This simplified environment allows us to clearly illustrate the effects of round-trip time and loss probability on the performance of feedback-based repair techniques. However, in practice, round-trip times and loss rates may change rapidly. Future work could explore the impact of varying round-trip times or loss rates during the lifecycle of a flow on the performance of feedback-based repair techniques.
5. Our analytical models do not impose a restriction that the GOBs in a reference chain cannot belong to one single frame. However, the experiments conducted in this thesis assume that GOBs in a reference chain reside in separate frames. Future work could explore the possibility of measuring the impact of reference distance on video quality for GOBs that are within a frame, extending the analytical experiments as appropriate.
6. Our analytical models and experiments assume reliable transmission of feedback messages. Future work could model cases where feedback messages may be lost and explore this impact on the performance of feedback-based error control techniques. Similarly, our models assume the feedback messages are transferred via a separate back channel, which is not counted into the overall bit-rate budget.

Future work could investigate the impact of the extra bandwidth consumed by feedback messages on the performance of feedback-based error control techniques, in particular on the performance of RPS ACK since RPS ACK requires more frequent transmissions of feedback messages than other feedback-based error control techniques.

7. Future work could build a videoconference system that automatically adapts to the best RPS mode (ACK or NACK) or Intra Update depending upon the network conditions and video content. For instance, when a high loss rate is detected, the encoder could switch to RPS ACK; whereas when the loss rate drops, the encoder could switch to RPS NACK.
8. Our models target H.264 videos since this standard incorporates all four feedback-based error control techniques considered in this thesis, but can generally represent any video encoding technique that uses feedback-based repairs. Future work could explore other coding standards in detail, such as MPEG4, H.263, to examine if the results derived from our models still hold.
9. The video quality function U_r in our models is obtained from our previous work [81]. It is not feasible for a real-time video system to measure U_r using the approach adopted in [81]. Future work could explore how the relationship between video quality and reference distance is affected by scene complexity and motion with a broader set of videos. The video quality functions for a variety of scene complexity and motion could then stored in a database. With this database, a real-time video system could obtain the U_r function for a specific video based on its

scene complexity and motion by matching the current data with the stored data in real time.

Bibliography

[1] Y. Wang, Q. Zhu. Error Control and Concealment for Video Communication: A Review. *Proceedings of the IEEE*, Vol. 86, No. 5, May 1998.

[2] Y. Wang, S. Wenger, J. Wen and A. K. Katsaggelos. Review of Error Resilient Coding Techniques for Real-Time Video Communications. *IEEE Signal Processing Magazine*, vol. 17, no. 4, July 2000.

[3] Y. Liu and M. Claypool. Using Redundancy to Repair Video Damaged by Network Data Loss. *In Proceedings of ISFT/SPIE/ACM Multimedia Computing and Networking (MMCN)*, Jan. 2000.

[4] J.C. Bolot, S. Fosse-Parisis, and D. Towsley. Adaptive FEC-Based Error control for Internet Telephoney. *In Proceedings of IEEE INFOCOM*, Mar, 1999.

[5] J. W. Modestino and D. G. Daut. Combined source-channel coding of images. *IEEE Trans. Commun.,* vol. COM-7, pp. 1644–1659, Nov. 1979.

[6] N. Farvardin and V. Vaishampayan. Optimal quantizer design for noisy channels: an approach to combined source-channel coding. *IEEE Trans. Inform. Theory*, vol. IT-38, pp.827–838, Nov, 1987.

[7] Y.-Q. Zhang, Y. J. Liu, and R. L. Pickholtz. Layered image transmission over cellular radio channels. *IEEE Transactions on Vehicular Technology,* vol. 43, pp. 786–796, Aug. 1994.

[8] M. Khansari and M. Vetterli. Layered transmission of signals over power-constrained wireless channels. In *Proc. International Conference on Image Processing (ICIP)*, Washington, DC, Oct. 1995, vol. III, pp.380–383.

[9] Y. Wang, Q. -F. Zhu, and L. Shaw. Maximally smooth image recovery in transform coding. *IEEE Trans .Commun.*, vol. 41, pp. 1544–1551, Oct. 1993.

[10] W. Kwok and H. Sun. Multi-directional interpolation for spatial error concealment. *IEEE Trans. Consumer Electron*, vol. 39, pp. 455–460, Aug. 1993.

[11] H. Sun and W. Kwok. Concealment of damaged block transform coded images using projections onto convex sets. *IEEE Trans. Image Processing*, vol.4, pp.470–477, Apr.1995.

[12] B. Girod and N. Färber: Feedback-based error control for mobile video transmission. *Proc. IEEE, Special Issue on Video for Mobile Multimedia*, vol. 87, pp. 1707-1723, Oct. 1999.

[13] I. Rhee. Error control Techniques for interactive low-bit rate video transmission over the Internet. In *Proceedings of ACM SIGCOM*, September 1998.

[14] N. Feamster and H. Balakrishnan. Packet Loss Recovery for Streaming Video. In *Proceedings of Packet Video Workshop (PV'02)*, April 2002.

[15] S. Fukunaga, T. Nakai, and H. Inoue. Error resilient video coding by dynamic replacing of reference pictures. In *Proc. IEEE Global Telecommunications Conf. (GLOBECOM)*, London, U.K., vol.3, Nov. 1996, pp.1503–1508.

[16] Y. Tomita, T. Kimura, and T. Ichikawa. Error resilient modified inter-frame coding system for limited reference picture memories. In *Proc. Int. Picture Coding Symp. (PCS)*, Berlin, Germany, Sept. 1997, pp.743–748.

[17] N. Farber, E. Steinbach, and B. Girod. Robust H.263 compatible video transmission over wireless channels. In *Proc. Int. Picture Coding Symp. (PCS)*, Melbourne, Australia, Mar.1996, pp. 575–578.

[18] I. E. G. Richardson. H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia. *Wiley*, ISBN 0-470-84837-5, 2004.

[19] Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC). In *Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG*, *JVTG050*, 2003.

[20] T. Wiegand, G. Sullivan, and A. Luthra. Overview of the H.264/AVC Video Coding Standard. *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, No 7, pp. 560–576, July 2003.

[21] T. Stockhammer, M. M. Hannuksela, and T. Wiegand. H.264/AVC in Wireless Environments. In *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, No 7, pp. 657–673, July 2003.

[22] S. Kumar1, L. Xu, M. K. Mandal and S. Panchanathan. Error Resiliency Schemes in H.264/AVC Standard. *Elsevier J. of Visual Communication &*

Image Representation (Special issue on merging H.264/AVC Video Coding Standard), Vol. 17(2), April 2006.

[23] M. Pinson and S. Wolf. A New Standardized Method for Objectively Measuring Video Quality. *IEEE Transactions on Broadcasting*, VOL. 50, NO.3, pp. 312-322, September, 2004.

[24] S. Aign. Error concealment for MPEG-2 video. In *Signal Recovery Techniques for Image and VideoCompression and Transmission*, A. K. Katsaggelos and N. P. Galatsanos, editors, ch. 8, pp. 235-268, Kluwer Academic Publishers, 1998.

[25] H. Sun, K. Challapali, and J. Zdepski. Error concealment in digital simulcast AD-HDTV decoder. *IEEE Trans. Consumer Electronics*, vol. 38, no. 3, pp. 108-117, Aug. 1992.

[26] W. M. Lam, A. R. Reibman, and B. Liu. Recovery of lost or erroneously received motion vectors. *Proc. ICASSP'93*, Minneapolis, Apr. 1993, pp. 417-420.

[27] Q.-F. Zhu, Y. Wang, and L. Shaw. Coding and cell loss recovery for DCT-based packet video. *IEEE Trans. CAS for Video Tech.*, vol. 3, no. 3, pp. 248-258, June 1993.

[28] V. Varsa and M. M. Hannuksela. Non-normative error concealment algorithms. ITU-T SG16 Doc. VCEG-N62, 2001.

[29] M.C. Hong, L. Kondi, H. Scwab, and A. K. Katsaggelos. Video Error Concealment Techniques. *Signal Processing: Image Communications*, special issue on *Error Resilient Video*, 14(6-8), pp. 437-492, 1999.

[30] ITU-T. ITU-T Draft Recommendation H.324. Terminal for low bitrate multimedia communication. *International Telecommunications Union*, Geneva, Switzerland, Dec. 1997.

[31] M. Hemy, U. Hangartner, P. Steenkiste, and T. Gross. MPEG System Streams in Best-Effort Networks. In *Proceedings of Packet Video Workshop'99*, April 1999.

[32] P. Bocheck, A. Campbell, S.-F. Chang, and R. Lio. Utility-based Network Adaptation for MPEG-4 Systems. In Proceedings of Ninth International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV), June 1999.

[33] A. Watson and M. A. Sasse. Measuring Perceived Quality of Speech and Video in Multimedia Conferencing Applications. In *Proceedings of the ACM Multimedia Conference*, pp. 55-60, Bristol, UK, September 1998.

[34] C. J. van den Branden Lambrecht and O. Verscheure. Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System. In *Proceedings of the SPIE*, January 1996.

[35] S. Winkler. Issues in Vision Modeling for Perceptual Video Quality Assessment. *Signal Processing, 78(2)*, 1999.

[36] J.-B. Martens and L. Meesters. Image dissimilarity. *Signal Processing*, vol. 70, pp. 155–176, Nov. 1998.

[37] VQEG. Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment. Mar. 2000, [Online] Available: http://www.vqeg.org/

[38] J. C. Bolot and T. Turletti. Experience with control mechanisms for packet video in the internet. *SIGCOMM Computer Communication Review*, 28(1), January 1998.

[39] X. Su and B. W. Wah. Multi-description video streaming with optimized reconstruction-based DCT and neural-network compensations. *IEEE Transactions on Multimedia*, 3(3), May 2001.

[40] B. W. Wah, X. Su, and D. Lin. A survey of error-concealment schemes for real-time audio and video transmissions over the Internet. In *Proc. of IEEE Int'l Symposium on Multimedia Software Engineering*, pp.17–24, Taipei, Taiwan, December 2000.

[41] G. Ramamurthy and D. Raychaudhuri. Performance of packet video with combined error recovery and concealment. In *Proceedings of the IEEE INFOCOM*, pp. 753–761, April 1995.

[42] C. Papadopoulos and G. Parulkar. Retransmission-based error control for continuous media applications. In *Proceedings of the Sixth International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, May 1997, pp. 5-12.

[43] R. Marasli, P. D. Amer, and P. T. Conrad. Retransmission-based partially reliable transport service: an analytic model. In *Proc. IEEE Infocomm'96*, San Francisco, CA, May, 1996, pp. 621–628.

[44] B. C. Smith. Implementation Techniques for Continuous Media Systems and Applications. *Ph.D. dissertation*, Univ. of California, Berkeley, 1994.

[45] Q. Zhu. Device and method of signal loss recovery for real time and/or interactive communications. *U.S. Patent 5550847*, Aug. 1996.

[46] H. Yasuda, H. Kuroda, H. Kawanishi, F. Kanaya, and H. Hashimoto. Transmitting 4MHz TV signals by combinational difference coding. *IEEE Trans. Commun*., vol. COM-25, pp. 508–516, May 1977.

[47] J. G. Apostolopoulos, W. Tan, and S. J. Wee. Video streaming: concepts, algorithms and systems. *HP Laboratories*, 2002.

[48] N. Mukawa, H. Kuroda, and T. Matuoka. An inter-frame coding system for video teleconferencing signal transmission at a 1.5Mbit/s rate. *IEEE Trans. Commun.*, vol. COM-32, pp. 280–287, Mar. 1984.

[49] W. Wada. Selective recovery of video packet loss using error concealment. *IEEE J. Select. Areas Commun.*, vol. 7, pp. 807–814, June 1989.

[50] N. Farber, E. Steinbach, and B. Girod. Robust H.263 compatible video transmission over wireless channels. In *Proc. Int. Picture Coding Symp. (PCS)*, Melbourne, Australia, Mar.1996, pp. 575–578.

[51] E. Steinbach, N. Farber, and B. Girod. Standard compatible extension of H.263 for robust video transmission in mobile environments. *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp.872–881, Dec.1997.

[52] B. Girod, N. Farber, and E. Steinbach. Error-resilient coding for H.263. In *Insights into Mobile Multimedia Communication,* D. Bull, N. Canagarajah, and A. Nix, Eds. New York: Academic, 1999, pp. 445–449.

[53] ITU-T [17]. Recommendation H.263, Video coding for low bit rate communication. Version 1, Nov. 1995; Version 2, Jan. 1998; Version 3, Nov. 2000.

[54] ITU-T. Video coding for low bit rate communication annex N. ITU-T Recommendation H.263 Ver. 2, 1995.

[55] ITU-T. Video coding for low bit rate communication annex U. ITU-T Recommendation H.263 Ver. 3, 2000.

[56] ITU-T. Recommendation H.245, Control protocol for multimedia communication. MAY 2006.

[57] ITU-T. Recommendation H.271, Video back channel messages for conveyance of status information and requests from a video receiver to a video sender. May, 2006.

[58] S. Fukunaga, T. Nakai, and H. Inoue. Error resilient video coding by dynamic replacing of reference pictures. In *Proc. IEEE Global Telecommunications Conf. (GLOBECOM)*, London, U.K., Nov. 1996, vol.3, pp.1503–1508.

[17] International Telecommunication Union - Telecommunication Standardization Sector.

[59] Y. Tomita, T. Kimura, and T. Ichikawa. Error resilient modified inter-frame coding system for limited reference picture memories. In *Proc. Int. Picture Coding Symp. (PCS)*, Berlin, Germany, Sept. 1997, pp.743–748.

[60] ITU-T/SG15/LBC-96-033. An error repair method based on back channel signaling and FEC. Telenor R&D, San Jose, Jan 1996.

[61] S. Wenger. H.264/AVC over IP. *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 645–656, July 2003.

[62] T. Wiegand, X. Zhang, and B. Girod. Motion-compensating long-term memory prediction. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Santa Barbara, CA, Oct. 1997, pp.53–56.

[63] T. Wiegand, X. Zhang, and B. Girod. Long-term memory motion-compensated prediction. *IEEE Trans. Circuits Syst. Video Technol.*, vol.9, pp.70–84, Feb.1999.

[64] B. Girod, T. Wiegand, E. Steinbach, M. Flierl, and X. Zhang. High-order motion compensation for low bit-rate video. In *Proc. Europ. Signal Processing Conf.*, Rhodes, Greece, Sept. 1998, pp. 253–256.

[65] P. Mulroy, M. Nilsson. Windowed Reference Picture Selection for H.264 Transmission Error Recovery. In *Proc. Int. Picture Coding Symp. (PCS)*, Lisbon, Portugal, Nov 2007.

[66] K. Mayer-Patel, L. Le, and G. Carle. An MPEG Performance Model and Its Application to Adaptive Forward Error Correction. In *Proc. of ACM Multimedia*, Dec. 2002.

[67] H. Wu, M. Claypool, and R. Kinicki. A Model for MPEG with Forward Error Correction and TCP-Friendly Bandwidth. In *Proceedings of Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, Monterey, California, USA, June 2003.

[68] R. Marasli, P. D. Amer, and P. T. Conrad. Retransmission-based partially reliable transport service: ananalytic model. In *Proc. IEEE Infocomm'96*, San Francisco, CA, 1996, pp. 621–628.

[69] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos. Rate-distortion optimized hybrid error control for packetized video communications. *IEEE Trans. Image Processing*, 2005.

[70] K. Stuhlmüller, N. Färber, M. Link, and B. Girod. Analysis of Video Transmission over Lossy Channels. *IEEE Journal on Selected Areas in Communications,* vol. 18, no. 6, pp. 1012-1032, June 2000.

[71] Y. J. Liang, J. G. Apostolopoulos, and B. Girod. Model-Based Delay-Distortion Optimization for Video Streaming Using Packet Interleaving. *Technical Report HPL-2002-323*, HP Labs, Nov. 2002.

[72] Y. J. Liang, J.G. Apostolopoulos, and B. Girod. Analysis of packet loss for compressed video: Does burst-length matter? In *Proc. IEEE International Conference on Acoustics, Speech, and Sig nal Processin*g (*ICASSP*), Hong Kong, Apr. 2003.

[73] W. Tu and E. Steinbach. Proxy-based Reference Picture Selection for Real-time Video Transmission over Mobile Networks. *IEEE International*

Conference on Multimedia & Exp (ICME '05), Amsterdam, Netherlands, July 6-8, 2005.

[74] Y. Wang, C. Zhu and H. Li. Error resilient video coding using flexible reference frames. *Invited paper, special session on Video Coding Standards, Visual Communications and Image Processing (VCIP),* Beijing, China, Jul. 2005.

[75] X. Huang, W. Shen, Z. Zhu, X. Guo, and X. Wei. Content-Based Adaptive Intra Blocks Update for Robust Wireless Video Communications. *Wireless Communications, Networking and Mobile Computing, (WiCom),* Shanghai, *China,* Sept. 2007.

[76] C. Chen, Y. Lee, C. Lin and Y. Chen. Error Resilience Trans-coding Using Prioritized Intra-Refresh for Video Multicast over Wireless Networks. *IEEE International Conference on Multimedia and Expo (ICME),* pp. 1310-1313, Amsterdam, Netherlands., Jul. 2005.

[77] H. Chiou; Y. Lee; and C. Lin. Error-resilient Trans-coding using Adaptive Intra Refresh for Video Streaming. *Proceedings of the International Symposium on Circuits and Systems (ISCAS)*, vol. 3, pp. 777-780, Vancouver, BC, Canada, May 2004.

[78] H. Wei; Y. Tsai; and C. Lin. Prioritized Retransmission for Error Protection of Video Streaming over WLANs. *Proceedings of the International Symposium on Circuits and Systems (ISCAS)*, vol. 2, pp. 65-68, Vancouver, BC, Canada, May 2004.

[79] Y. Uchida et al. Proactive Retransmission and Buffer Management for Layered Video Transmission over Wireless Channel, *IEICE Transactions on Communications*, vol. E91-B, no. 9, pp. 2897-2906, Sept. 2008

[80] F. Hou, P.H. Ho, and X. Shen. A Novel Differentiated Retransmission Scheme for MPEG Vedio Streaming over Wireless Links, *Inernational. Journal. Wireless and Mobile Computing (IJWMC) - special issue on MAC for WLANs, WPANs, Ad hoc Networks, and Sensor Networks,* Vol. 1, Nos. 3/4, pp. 260-267, 2006.

[81] Y. Wang, M. Claypool and R. Kinicki. Impact of Reference Distance for Motion Compensation Prediction on Video Quality, In *Proceedings ACM/SPIE Multimedia Computing and Networking (MMCN) Conference*, San Jose, California, USA, Jan. 2007.

[82] Y. Wang, M. Claypool and R. Kinicki. An Analytic Comparison of RPS Video Repair, In *Proceedings ACM/SPIE Multimedia Computing and Networking (MMCN) Conference*, San Jose, California, USA, Jan. 2008.

[83] Y. Wang, M. Claypool and R. Kinicki. Modeling RPS and Evalauting Video Repair with VQM, *IEEE Transactions on Multimedia*, 2009. (to appear)