In Proc. of Visual Communic. and Image Proc., VCIP '97, San Jose, CA, Feb. 1997

Rate control for robust video transmission over wireless channels

Chi-Yuan Hsu, Antonio Ortega

Integrated Media Systems Center, Department of Electrical Engineering-Systems
University of Southern California, Los Angeles, California 90089-2564

Masoud Khansari

Hewlett-Packard Laboratories
Palo Alto, California 94304-1126

ABSTRACT

Video transmission over wireless links is an emerging application which involves a time-varying channel. Compared to other transmission media, wireless links suffer from limited bandwidth, and are more likely to see their performance degrade due to multipath fading. Therefore error control mechanisms, which can achieve better video quality with the available bandwidth and recover from the errors due to the channel degradation, are very important in wireless video transmission systems. Many of the proposed wireless communications systems are likely to be two-way so that a return channel can convey information to the transmitter about the channel state. Recent research has considered ways of improving the transmission reliability by making use of the feedback channel for "closed-loop" error control, including various forms of retransmission. In this paper we propose a rate control algorithm based on dynamic programming combined with Automatic Repeat reQuest (ARQ) as the error control mechanism. We formalize the constraints imposed by the real-time characteristics of video. We show how when an appropriate model for the channel is available, the overall robustness of the system can be improved through rate control at the source using the channel state information conveyed by the ARQ acknowledgement.

Keywords: wireless video, rate control, wireless links, automatic repeat request (ARQ), forward error correction (FEC), channel model, channel feedback, Viterbi algorithm.

1 INTRODUCTION

Given the bandwidth limitations, compression will be required for video transmission over wireless channels. Due to the extensive use of variable length entropy coding and prediction, compressed video is more vulnerable to transmission errors as losses can produce artifacts until synch is recovered or the prediction loop is restarted. Error protection is thus essential for a robust wireless video transmission. What makes error protection challenging in a wireless environment is the bursty nature of the errors, usually due to multipath fading. Periods of good channel behavior alternate with deep fade intervals during which getting any information across would require substantial levels of error correction overhead. A solution based on using Forward Error Correction (FEC) alone for protection would thus require that either the overhead is high throughout the transmission (thus lowering data

throughput during the "good" intervals) or that the channel be effectively unusable during the worst fade periods. In one-way communication systems, FEC is the only possible choice for error control: errors have to be detected and corrected by the receiver. In these cases the FEC overhead can be reduced by interleaving the video data to spread out clustered errors over several data packets. The Asynchronous Transfer Mode (ATM) Adaptation Layer (AAL) specification in the ITU-T recommendation I.363 [1] employs the Reed-Solomon (128,124) code combined with interleaving to spread out the bursty errors. Better error resilience with lower overhead can be achieved by interleaving the video data at the cost of a longer delay.

Many burst-error channels are such that a feedback channel is available and the receiver can send information back to the transmitter. This feedback channel can then be used to improve the error resilience by, for example, informing the transmitter of the current state of the channel [2–4]. In two-way communication systems with a feedback channel, methods for error detection and retransmission, such as Automatic Repeat reQuest (ARQ), can be used instead of, or in combination with, FEC. In the ARQ error control scheme, either acknowledgement (ACK) or negative acknowledgement (NAK) are sent back to the transmitter to indicate whether the packet was correctly received. Upon receiving a NAK the transmitter attempts to retransmit the data until the data is correctly received [5]. ARQ protocols introduce variable delays in retransmission when channel fading occurs, and such delay could be problematic for delay-critical video transmission application.

The problem of rate control in a wireless environment with variable channel rate was formalized in [6]. In this paper we address the issue of designing a rate control mechanism that can work in an ARQ environment and which can minimize the distortion produced by data losses caused by excessive delay. The rate control mechanism uses an *a priori* model of the channel behavior and relies on observations of the current channel state which are obtained at the transmitter by monitoring the incoming ACK/NAK signals. The rate control unit dynamically adjusts the source video coding rate responding to the changes in channel conditions so that in periods of poor channel performance the source rate can be reduced. Dynamic programming is used to find out the best bit allocation, in the sense of minimizing the expected value of the distortion, for each video frame inside the encoder buffer. With tightly integrated rate control and error control unit, channel feedback and *a priori* channel model, we demonstrate significant improvements in the average video quality and a corresponding reduction on the packet loss rates.

The paper is organized as follows. In Section 2 we define the timing constraints due to the constant end-to-end delay constraint between the encoder and decoder. The system considered for our experiments is described in Section 3. In particular we describe the video encoder and the transmission system, as well as the corresponding model for channel behavior. Section 4 provides a detailed description of the optimization framework. We introduce the notion of expected distortion for given channel characteristics, and show a solution can be obtained to minimize the average expected distortion. The simulation results in Section 5 demonstrate the advantage of using channel feedback in terms of the received quality and packet loss rate.

2 TIMING CONSTRAINTS

Due to the Variable Bit Rate (VBR) nature of compressed video, buffers at the encoder and decoder are required to smooth out the bit rate variations so that a VBR encoder can be used over a constant rate channel. Buffering data results in delay. Larger buffers result in longer delays but also allow lower overall distortion to be achieved by proper bit-allocation between the video frames in the buffer. In a video communications system, the end-to-end delay, ΔT, between the video input and output consists of three major components:

$$\Delta T = \delta t_e(\text{Encoder buffer delay}) + \delta t_c(\text{Channel propagation delay}) + \delta t_d(\text{Decoder buffer delay}). \quad (1)$$

We assume that video encoding and decoding delays are constant and we only take into account the variable elements in the overall delay (Fig 1). In real-time video transmission, both encoder and decoder are attached to synchronous input and output devices. If no frames are dropped at the encoder, in order to preserve the

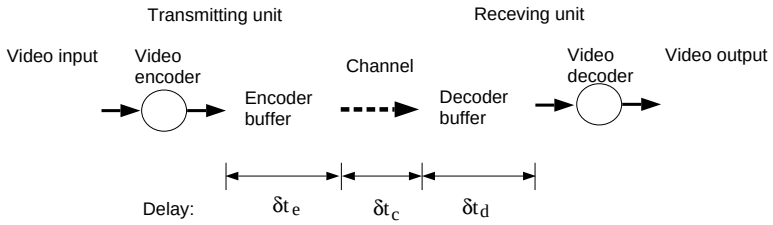


Figure 1: Delay components of a communication system.

correct displaying sequence of the video, the end-to-end delay between the synchronous devices, ΔT in (1), has to be constant. Video data that arrives later then the scheduled display time is considered lost. The channel propagation delay, which corresponds to the signal transmission time from the transmitter to the receiver, can be considered as constant in many wireless channels. If there is networking involved and the delay is variable we can assume that δt_c is the maximum rate. This will only result in overdimensioning of the encoder/decoder buffers and will not affect the overall formulation [7]. Therefore the sum of δt_e and δt_d has to be constant in (1). This also means that the number of video frames stored in the encoder and decoder buffers has to be constant. We now describe how this delay constraint translates into a constraint on the rates that can be used for each video frame.

In our discussion, the video sequence is subdivided into *video blocks*, which may consist of several macroblocks, a group of blocks (GOB) or a frame. A convenient choice of video block size is that where the average number of bits used in encoding a video block approximately equals or is smaller than the size of a transmitted packet (assumed here to be fixed). The time index is quantized and a time unit is the time interval of displaying a video block, or equivalently the time interval of transmitting a packet. Define ΔN as the constant number of video blocks in the encoder and decoder buffer. While the number of blocks inside both the transmitting and receiving buffer is constant, the number of blocks in the encoder buffer is variable and always smaller than ΔN. Denote block τ as the last block of those which are transmitted by the channel at time ν, and block ν as the one which just arrived at the encoder buffer at time ν. It is possible that not all the data of block τ is transmitted at time ν, because the required numbers of bits to encode different video blocks are different. Denote $R'(\tau)$ as the number of bits belonging to block τ that are still in the encoder buffer. Also denote $R(i)$ as the number of bits that are used by block i. The content of the encoder buffer then consists of $R'(\tau), R(\tau+1), R(\tau+2), \ldots, R(\nu)$ bits of data from blocks $\tau, \tau+1, \ldots, \nu$. The buffer content at time ν is depicted in Fig 2. A video block is considered lost if any of the packets which carry data from that block cannot be delivered to the receiving end by the scheduled display time. This delay constraint can be translated into a constraint on the number of bits that can be used for encoding those blocks which are still stored in the encoder buffer. In the example depicted in Fig 2, the end-to-end delay ΔN constrains the values of $R'(\tau), R(\tau+1), \ldots, R(\nu)$.

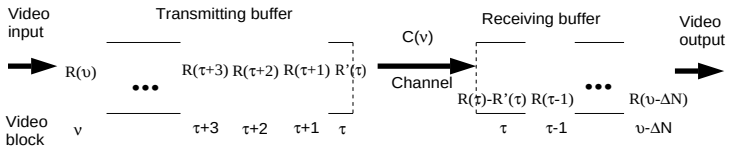


Figure 2: Transmitting buffer content

Consider block τ, which arrived at the encoder buffer at time τ. All the data which corresponds to block τ should be carried out of the decoder buffer by the channel before time $\tau + \Delta N$, otherwise the maximum delay for block τ will be exceeded. In the example depicted in Fig 2, this timing constraint on block τ means the cumulative bits of data that are transmitted by the channel in the time period $\nu+1, \nu+2, \ldots, \tau+\Delta N$ should be larger than $R'(\tau)$ to prevent any data of block τ from being over-delayed in the encoder buffer. Define $C(k)$ as the amount of data that is transmitted by channel at time k. The condition to prevent block τ from being

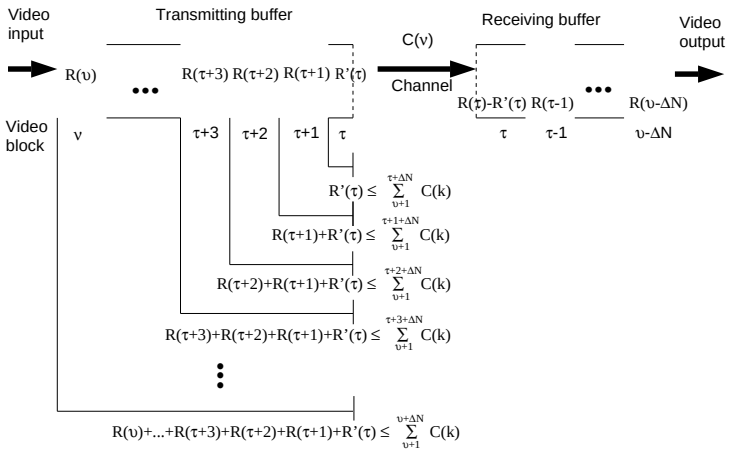


Figure 3: Constraint on the value of bit-rates $R(i)$ for each block inside the encoder buffer at time τ

over-delayed in the encoder buffer thus can be written as:

$$R'(\tau) \leq \sum_{k=\nu+1}^{\tau+\Delta N} C(k) \tag{2}$$

Likewise for block $\tau+1$ to be correctly received, its data and that of the previous blocks should be transmitted by time $\tau+1+\Delta N$. This condition can be written as:

$$R(\tau+1) + R'(\tau) \leq \sum_{k=\nu+1}^{\tau+1+\Delta N} C(k) \tag{3}$$

In general, the condition for the data of block i, $i \in \{\tau, \ldots, \nu\}$, to be delivered to the receiver without violating the delay constraint is thus:

$$\sum_{j=\tau+1}^{i} R(j) + R'(\tau) \leq \sum_{k=\nu+1}^{i+\Delta N} C(k) \tag{4}$$

From (4), the constraints on $R'(\tau), R(\tau+1), \ldots, R(\nu)$ at time ν are summarized in Fig. 3. Packet loss can be avoided by keeping, through rate control, the encoding rates within the bounds defined by (4). While this would be easy to achieve if $C(k)$ were known, in our case the channel rates depend on the channel state which varies randomly. We will show that rate control can be also formulated with a probabilistic model of the channel behavior in Section 4. We first describe the wireless system we are considering.

3 SYSTEM DESCRIPTION

3.1 Video encoder and decoder

The source video encoder used in our simulation was H.263 [8], a video compression standard targeted at low bit-rate transmission applications. A simple way to visualize our system (see Fig. 4) is to assume that the same video data coded with different quantization steps is stored in separate buffers (note that only quantization differs). The rate control unit then chooses the proper bit-stream from different buffers based on the feedback

channel condition and the R-D relations of those video blocks which are stored in these buffers. Only the data from the chosen buffer is transmitted by the channel, and the data from other buffers are discarded. Obviously, a similar result can be achieved with a layered video coder, where the transmitted video quality can be adjusted by deciding the number of layers to be transmitted. In our simulations when a video block is lost, due to a violation of the delay constraint, we replace it by its corresponding DC value at the decoder end. We assume that the DC value of a video blocks can be transmitted reliably by using FEC with high protection capability. Note that this assumption is somewhat pessimistic since it would be possible in general to replace lost blocks by blocks in the same position in the previous frame, thus achieving better error concealment.

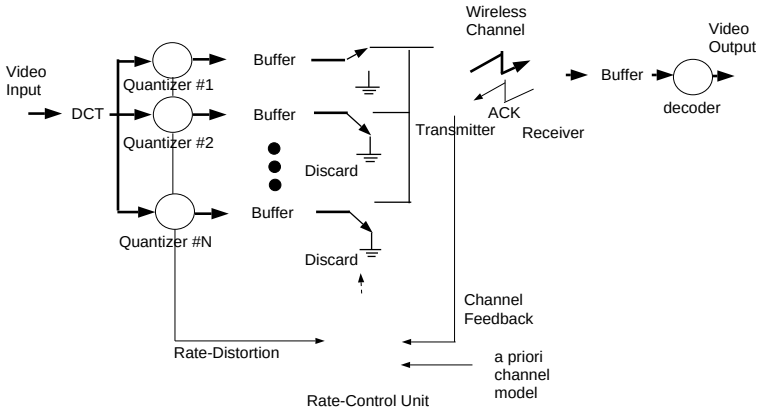


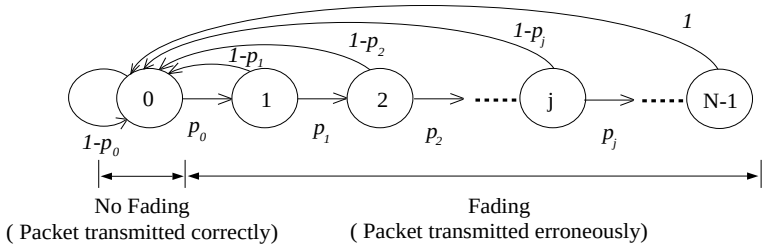
Figure 4: System block diagram.

3.2 Uplink and downlink transceivers

The system under consideration is patterned after $IS-95$ standard proposed for digital wireless communication. This system is a wide-band CDMA spread spectrum system with processing gain of 128 and chip rate of 9.83 Mchips/s, resulting in a data rate 76.8 Kbits/s. The packets are of 5 msec or 48 bytes duration of which 41 bytes is the payload. As a result the information transmission rate is 65.6 Kbits/s. In general, a wireless link consists of two radio links, namely uplink or reverse link (mobile-to-base) and downlink or forward link (base-to-mobile). Because of the asymmetry in the information available at the base and the mobile and also the different power constraints and hence processing power, the design of these two links have different requirements. On the uplink transmitter, the user data is first spread using a 1/3 convolutional encoder. The encoded bits are then further spread using 64-ary orthogonal signaling followed by symbol interleaving and QPSK PN spreading. The receiver employs both antenna and multipath diversity where a number of correlators (each correlator corresponds to a path) are assigned to each antenna. A fast closed loop power control is used to combat Rayleigh fading. On the downlink transmitter, user information is encoded using a half-rate convolutional encoder and the encoded bits are QPSK spread and transmitted using BPSK modulation. Spatial diversity through the use of different antennas (3 in our case) is used to combat fading. Also, the signal transmission is pilot assisted. Using the pilot signal received at the receiver, maximal ratio combining of different paths is achieved. The combiner is followed by de-interleave and soft decision Viterbi decoder. Because of the fast power control used for the uplink channel, the bit error bursts tend to be of shorter duration. For a given BER, there are more packets in error for the uplink channel – there are, however, fewer bits in error in each erroneously received packet. The downlink channel has fewer frames in error but, due to the bursty nature of the errors, the packet error bursts tend to be more persistent and hence of longer duration. We refer to [2] for more details.

3.3 Channel model

Using simulation results of both the uplink and the downlink transceivers, a Markov chain model was used to model the behavior of the links. We use a similar model as in [9] where it is used to evaluate the performance of different speech coders. The model has N states, and a packet error occurs in transmission when the channel is in state $i = 0, \cdots, N-1$. For all $1 \leq i \leq N-2$, the next transition is either to the next higher state (state $i+1$) or back to state 0 based on the status of the currently received data-frame. If the channel is in state $N-1$, it will always return to state 0. The effective channel rate equals to the constant packet size $\bar{C}$ when the channel is in state 0, and equals to 0 when the channel is in state $1 \leq i \leq N-1$. Note that, using this model, it is only possible to generate burst errors of at most of length $N-1$. The error statistics can then be controlled by properly assigning values to the probabilities $p_0, p_1, \cdots, p_{N-2}$. The values for the uplink and downlink channels at BER $= 10^{-3}$ are shown in the following table, where N was found to be 15 and 7 (equivalent to maximum burst error lengths of 70 msec and 30 msec) for the downlink and the uplink channel, respectively. These values are found by matching the parameters of the Markov chains to simulations of the transceivers, see Fig. 5.



	Downlink	Uplink
p_0	0.001469	0.064292
p_1	0.516068	0.100324
p_2	0.778388	0.164083
p_3	0.854118	0.149606
p_4	0.936639	0.526316
p_5	0.873529	0.000000
p_6	0.905724	
p_7	0.881041	
p_8	0.831224	
p_9	0.893401	
p_{10}	0.863636	
p_{11}	0.717105	
p_{12}	0.853211	
p_{13}	0.763441	
p_{14}	0.000000	

Figure 5: Markov model for channel behavior and transitional probability parameters for the downlink and the uplink channels

4 REAL-TIME RATE CONTROL OPTIMIZATION

A conservative approach to meet the constraints in (4) for the unknown future channel rates would be to assume worst case values for the expected channel rates. However this approach would reduce the source coding rate (and hence the quality) and so as will be seen better results can be achieved if the current channel state is fedback to the transmitter. Basically, in a channel with bursty errors, channel feedback can indicate the beginning of a channel fading period and thus allow the rate control unit to reduce the source coding rate. We will now show how to formalize the problem of selecting the encoding rates $R(\tau+1), R(\tau+2), \ldots, R(\nu)$ so as to minimize the expected distortion, including the distortion due to packet losses. The expected distortion is estimated based on the channel state feedback and the *a priori* channel model.

4.1 Expected distortion

Although the exact value of $\sum_{k=\nu+1}^{i+\Delta N} C(k)$ in (4) can not be known in advance, its probability distribution function can be estimated from the channel model and the observation of current channel conditions. Define the probability that at time ν, $\sum_{k=\nu+1}^{i+\Delta N} C(k)$ is smaller than c given that the observed channel state is s:

$$F_{\nu+1,i+\Delta N}(c|s) = Prob\left[\sum_{k=\nu+1}^{i+\Delta N} C(k) < c \quad \Big| \quad \text{channel state} = s\right]. \tag{5}$$

Therefore the probability that $\sum_{j=\tau+1}^{i} R(j) + R'(\tau)$ exceeds the upper bound is:

$$Prob\left[\sum_{k=\nu+1}^{i+\Delta N} C(k) < \sum_{j=\tau+1}^{i} R(j) + R'(\tau) \quad \Big| \quad \text{channel state} = s\right] = F_{\nu+1,i+\Delta N}(\sum_{j=\tau+1}^{i} R(j) + R'(\tau) \mid s) \tag{6}$$

From the probability distribution function (6), the rate control unit can decide its choice of encoding rates based on the source coding distortion and the expected distortion due to data losses. Define $D(i)$ as the distortion (measured as mean squared error) of a video block at the encoding rate $R(i)$. As mentioned above, we assume that lost video blocks are replaced by their DC value; denote $D_0(i)$ the distortion incurred in that case. The expected distortion $E[D(i)]$ of block i can be calculated from $F_{\nu+1,i+\Delta N}(c|s)$ as:

$$E[D(i)] = D(i) \cdot \left(1 - F_{\nu+1,i+\Delta N}(\sum_{j=\tau+1}^{i} R(j) + R'(\tau) \mid s)\right) + D_0(i) \cdot F_{\nu+1,i+\Delta N}(\sum_{j=\tau+1}^{i} R(j) + R'(\tau) \mid s) \tag{7}$$

given that $R(\tau+1), \ldots, R(i-1)$ are the numbers of bits used for encoding blocks $\tau+1, \ldots, i-1$, respectively, and s is the current channel state. Since the channel is modeled as a time-invariant Markov chain, the probability distribution function $F_{\nu+1,i+\Delta N}(c|s)$, which is derived from the model, is also time-invariant. $F_{\nu+1,i+\Delta N}(c|s)$ only depends on the length of the window $i + \Delta N - \nu$ in which channel rates are estimated and the observed channel state s. Therefore, $F_{\nu+1,i+\Delta N}(c|s)$ can be further simplified as:

$$F_t(c|s) = F_{\nu+1,i+\Delta N}(c|s) \quad \text{where } t = i + \Delta N - \nu. \tag{8}$$

This time-invariant function $F_t(c|s)$ can be computed from the model in advance and can be stored as a look-up table to calculate $E[D(\tau+1)], E[D(\tau+2)], \ldots, E[D(\nu)]$ from (7) for the observed channel state s.

4.2 Distortion estimation from Markovian channel model

The channel was modeled as an N state Markov chain, where p_i was defined as the transitional probability from state i to state $i+1$. Define $q_{s,r}^{(t)}$ as the probability that the channel visits state 0 (successful packet transmission) r times in t transitions. Therefore $q_{s,r}^{(t)}, s \in 0, 1, \ldots, N-1$ and $0 \le r \le t$ can be related to the probabilities in transition $t-1$ as:

$$\text{When } s \neq 0: \quad q_{s,r}^{(t)} = p_{s-1} \cdot q_{s-1,r}^{(t-1)} \quad \text{(The channel transits to nonzero state and error occurs)} \tag{9}$$

$$\text{When } s = 0: \quad q_{0,r}^{(t)} = \sum_{i=0}^{N-1} (1 - p_i) \cdot q_{i,r-1}^{(t-1)} \quad \text{(The channel transits to state 0 and no error occurs)} \tag{10}$$

$q_{s,r}^{(t)}$ can be represented by a $N \times (t+1)$ matrix, and the relations (9),(10) can be represented as:

$$Q^{(t)} = \begin{bmatrix} q_{0,0}^{(t)} & q_{0,1}^{(t)} & \cdots & q_{0,t}^{(t)} \\ q_{1,0}^{(t)} & q_{1,1}^{(t)} & \cdots & q_{1,t}^{(t)} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N-1,0}^{(t)} & q_{N-1,1}^{(t)} & \cdots & q_{N-1,t}^{(t)} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ p_0 & 0 & \dots & 0 & 0 \\ 0 & p_1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & p_{N-2} & 0 \end{bmatrix} \cdot \begin{bmatrix} q_{0,0}^{(t-1)} & q_{0,1}^{(t-1)} & \cdots & q_{0,t-1}^{(t-1)} \\ q_{1,0}^{(t-1)} & q_{1,1}^{(t-1)} & \cdots & q_{1,t-1}^{(t-1)} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N-1,0}^{(t-1)} & q_{N-1,1}^{(t-1)} & \cdots & q_{N-1,t-1}^{(t-1)} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

$$+ \begin{bmatrix} 1-p_0 & 1-p_1 & \dots & 1-p_{N-1} \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \cdot \begin{bmatrix} q_{0,0}^{(t-1)} & q_{0,1}^{(t-1)} & \cdots & q_{0,t-1}^{(t-1)} \\ q_{1,0}^{(t-1)} & q_{1,1}^{(t-1)} & \cdots & q_{1,t-1}^{(t-1)} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N-1,0}^{(t-1)} & q_{N-1,1}^{(t-1)} & \cdots & q_{N-1,t-1}^{(t-1)} \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

or equivalently in a more compact notation as:

$$\begin{aligned} Q^{(t)} &= L_1 \cdot Q^{(t-1)} \cdot R_1 \\ &+ L_2 \cdot Q^{(t-1)} \cdot R_2 \end{aligned} \tag{11}$$

where L_1,L_2 are $N \times N$ matrices, and R_1 and R_2 are $(t-1) \times t$ matrices respectively. If the channel state observed at time ν is s, then the initial probability matrix $Q^{(0)}$ can be written:

$$Q^{(0)} = \begin{bmatrix} q_{0,0}^{(0)} \\ q_{1,0}^{(0)} \\ \vdots \\ q_{N-1,0}^{(0)} \end{bmatrix}, \quad \text{where} \begin{cases} q_{s,0}^{(0)} = 1, \\ q_{i,j}^{(0)} = 0, \quad \text{when } i \neq s \text{ and } j \neq 0 \end{cases} \tag{12}$$

By recursively applying (11), $Q^{(t)}$ can be obtained given the initial probability matrix $Q^{(0)}$. Denote c as the sum of channel rates of the r correctly transmitted packets. Since we assume the packet size to be constant and equal to $\bar{C}$, $c = r \cdot \bar{C}$. The probability that less than r packets are transmitted correctly in t transitions (resulting in effective accumulated channel rates smaller than c) can be calculated from the elements of matrix $Q^{(t)}$ as:

$$F_t(c|s) = \sum_{j=0}^{r-1} \sum_{i=0}^{N-1} q_{i,j}^{(t)} \quad \text{where } c = r \cdot \bar{C} \tag{13}$$

4.3 Distortion estimation with delayed channel feedback

The expected distortion estimated from (7) and (13) is based on the channel state observed at time ν, which assumed instantaneous feedback of the channel state to the transmitter. In a more realistic scenario, the channel feedback information carried by the ACK/NAK signals could be delayed by the round-trip travel time of the signal propagation. Thus for a feedback delay of Δf time units (packet intervals) the information available at time ν corresponds to the channel state at time $\nu - \Delta f$.

Thus in (12) the initial probability $Q^{(0)}$ cannot be obtained directly from the channel condition at time ν and instead $Q^{(0)}$ has to be calculated from the channel model and the channel feedback at time $\nu - \Delta f$. Define $Q^{(-t)} = \left[q_0^{(-t)}, q_1^{(-t)}, \dots, q_{N-1}^{(-t)}\right]^T, 0 \leq t \leq \Delta f$, where $q_i^{(-t)}$ is the probability for the channel to stay at state i at time $\nu - t$. With the transitional probability defined in the Markovian channel model, the transition matrix can

be set up as:

$$P = \begin{bmatrix} 1-p_0 & 1-p_1 & \cdots & 1-p_{N-2} & 1 \\ p_0 & 0 & \cdots & 0 & 0 \\ 0 & p_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & p_{N-2} & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \tag{14}$$

Suppose the channel condition of time $\nu - \Delta f$ is s. The state probabilities at time $\nu - \Delta f$ with channel condition s can be expressed as:

$$Q^{(-\Delta f)} = \begin{bmatrix} q_0^{(-\Delta f)} \\ q_1^{(-\Delta f)} \\ \vdots \\ q_{N-1}^{(-\Delta f)} \end{bmatrix}, \quad \text{where} \begin{cases} q_0^{(-\Delta f)} = 1, \\ q_i^{(-\Delta f)} = 0, \quad \text{when } i \neq s \end{cases} \tag{15}$$

The state probabilities $Q^{(-t)}$ can be obtained from $Q^{(-t-1)}$ and P as:

$$Q^{(-t)} = P \cdot Q^{(-t-1)} \tag{16}$$

$Q^{(0)}$ can be obtained from repeatedly applying (16),

$$Q^{(0)} = \underbrace{P \cdot P \cdot \ldots \cdot P}_{\Delta f} \cdot Q^{(-\Delta f)} \tag{17}$$

Again $Q^{(0)}$ can be used to compute $F_t(c|s)$ by using (11),(12) and (13), and the expected distortion can be calculated from $F_t(c|s)$ by using (7).

4.4 Problem formulation

When the expected distortion for each video block can be estimated from the channel model and real-time channel feedback, the goal of rate control then is to find out the best bit allocation for $R(\tau+1), R(\tau+2), \ldots, R(\nu)$, with minimum sum of expected distortion $E[D(\tau+1)], E[D(\tau+2)], \ldots, E[D(\nu)]$. Assume that the video is encoded by P different quantizer steps, and $R_{x(i)}(i)$, $D_{x(i)}(i)$ are the corresponding rate and distortion in encoding block i by quantizer $x(i)$ at time i. Before the channel transmits $C(\nu+1)$ bits of data at time $\nu+1$, the rate control unit has to decide the sequence of quantizer choices $x(\tau+1), x(\tau+2), \ldots, x(\nu)$ with minimum expected distortion $E\left[\sum_{j=\tau+1}^{\nu} D_{x(j)}(j)\right]$ based on the channel state s observed at time ν. Our goal is then: *At time ν, find out the mapping $x(\tau+1, \ldots, \nu) \rightarrow (1, \ldots, P)$ that solves:*

$$\min E\left[\sum_{j=\tau+1}^{\nu} D_{x(j)}(j)\right] \tag{18}$$

given $R'(\tau)$ and channel state s

4.5 Optimal bit allocation by Viterbi algorithm

The Viterbi algorithm, a form of deterministic dynamic programming, can be used to solve the above problem. In this algorithm a trellis, which consists of nodes and branches, is formed to represent all the possible combinations

of quantizer choices for block $\tau+1$ to block ν. A branch which connects a node in stage $i-1$ to a node in stage i is associated with $R_{x(i)}(i)$ and $E[D_{x(i)}(i)]$, the rate and expected distortion when quantizer $x(i)$ is used to encode block i. Each node in stage i represents a distinct accumulated rate $R'(\tau)+\sum_{j=\tau+1}^{i} R_{x(j)}(j)$(defined as state variable for that node), and is associated with expected distortion $\sum_{j=\tau+1}^{i} E[D_{x(j)}(j)]$.

When two or more branches originated from different nodes in stage $i-1$ arrive at the same node in stage i (i.e. they have the same accumulated rate), only the branch with smallest expected distortion is kept and the remaining branches are pruned out. After the trellis is constructed, the path (consisting of connected branches from stage $\tau+1$ to stage ν) with minimum accumulated distortion can be found. The quantizer choices associated with the branches on the minimum distortion path are the optimal choice of quantizer which can achieve minimal distortion in encoding block $\tau+1$ to block ν, given the channel state s.

In order to guarantee optimality in the Viterbi algorithm, we must have that the state variable, $R'(\tau)+\sum_{j=\tau+1}^{i} R_{x(j)}(j)$, and the sum of expected distortion, $\sum_{j=\tau+1}^{i} E[D_{x(j)}(j)]$, of a node in stage i only depends on the preceding connected node at stage $i-1$ and the connecting branch. By using the intra-frame only video coding mode in H.263, the R-D relation for each block is independent on the quantizer choices of other blocks. From (7) the expected distortion for block i only depends on the function $F_{\nu+1,i+\Delta N}(c|s)$ (which is invariant for the same observed channel state s), state variable $R'(\tau)+\sum_{j=\tau+1}^{i} R_{x(j)}(j)$ and the choice of quantizer $x(i)$ which results in the the corresponding $R_{x(i)}(i)$ and $D_{x(i)}(i)$. Since a node in stage i is independent from all the nodes other than the one connected to it node at stage $i-1$, the Viterbi algorithm can be used to find out the minimum expected distortion solution. When the source video is encoded by dependent coding, e.g. P and B frames in H.263 or MPEG, the optimum quantizer choices can not be found by Viterbi algorithm because distortion depends on previous choices of quantizers at previous stages. However, other fast algorithms [10,7,11,12] can be used to find the suboptimal solution with the encoding rate constraints (4) or expected distortion (7) we introduced.

5 SIMULATION AND EXPERIMENTAL RESULTS

In our simulations, we consider the channel feedback delay to compute the expected distortion. The selective repeat ARQ error control scheme was used to increase the channel utilization. All the erroneous packets are back-logged for retransmission. The retransmission of erroneous packets is attempted until the transmitter receives the corresponding ACK or until all the data carried by that packet violates the delay constraint. In the latter case, the packet is discarded, and the corresponding video blocks are replaced by DC values at the decoder. We use the test video sequence "Miss America" (frames 1 to 60, luminance only, QCIF format, 176×144 pixels/frame, 11×9 macroblocks) encoded with the H.263 standard at quantization step sizes 8, 10 and 12. For simplicity, groups of 3 macroblocks constituted our basic video blocks. A single quantizer was assigned to each video block. On average one video block was transmitted by a packet with 41 bytes payload (thus 33 packets/frame and around 6 frames/sec on average).

In order to estimate the performance gain achievable by utilizing the channel feedback information, a simulation for the same communication system without the channel feedback was also performed. For comparison purposes, we also simulate the unrealistic scenario where the encoder has deterministic knowledge of future channel conditions *and* of the R-D relation of each frame. This case can serve as the bound on the achievable performance. In summary, the following three scenarios are considered in our simulation and for each of them dynamic programming is used to find out the minimal expected distortion solution:
(I) Finite delay with channel feedback: Channel feedback case where only the data in the buffer is considered.
(II) Finite delay without channel feedback: No feedback channel is available and we assume an equivalent constant channel rate equal to the average effective rate (i.e. after taking into account the channel losses).
(III) Known channel and infinite delay: The encoder has the deterministic knowledge of the channel condition and rate-distortion relation for every frame.

Six downlink and uplink channel realizations were generated from the Markov chain model. The resulting video distortions and packet loss rates are averaged over all these realizations. Figs 6 and 7 depict the resulting PSNR and packet loss rate for the four scenarios. From our simulation results, the communication systems with channel feedback outperformed the ones without channel feedback in terms of PSNR and packet loss rates in both the downlink and uplink channels. The performance of the systems with feedback channel also improved as the end-to-end delay increased. Because the transmission errors are more likely to occur in cluster in the downlink channel and are more more randomly distributed in the uplink channel, the future channel condition can be more accurately estimated from the current channel state in the downlink channel. Therefore, better video quality and lower packet loss could be achieved in the downlink channel than in the uplink channel in our experiment.

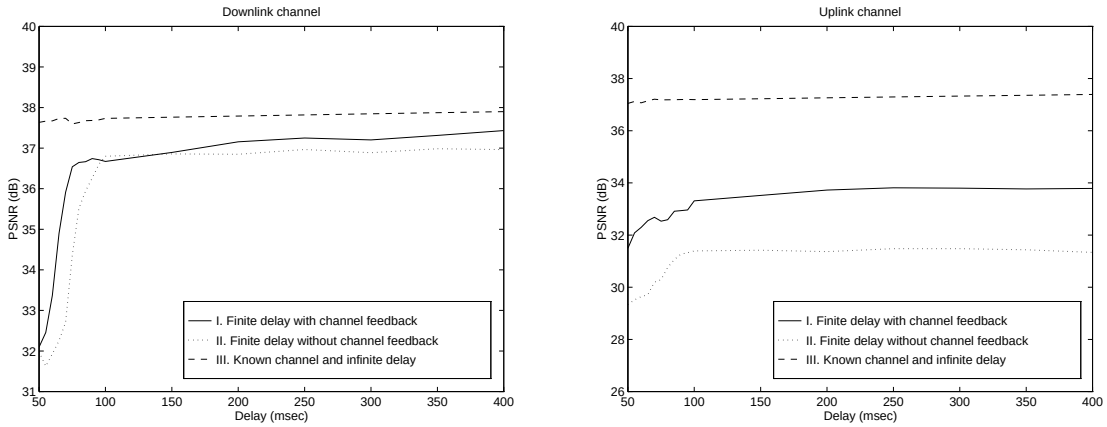


Figure 6: PSNR of the video for scenario (I), (II) and (III) with different delay constraints. The propagation delay equals 5 msec, hence the channel feedback delay equals to 10 msec in scenario (I).

Our results indicate improvements in end-to-end quality can be achieved through channel feedback. Longer end-to-end delay can also help to improve the performance. However moderately large buffering delay is required for the error control to retransmit the erroneous packet, and for the rate control to smooth out the VBR source bit-stream. It can be seen from the result of our simulation that the received video quality degrades significantly as the end-to-end delay is shorter than 100 msec, and most degradation can be attributed to the fact that the encoder buffer is too small to smooth out the VBR video traffic. While error control and rate control both need buffering data for better performance, good integration of these two mechanisms and utilization of the channel information can increase the robustness of the data transmission and the quality of the receiving video in the application of wireless video transmission.

6 ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation under grant MIP-9502227 (CAREER).

7 REFERENCES

[1] ITU-T Recommandation I.363, *B-ISDN ATM Adaptation Layer (AAL) specification.*

[2] M. Khansari, A. Jalali, E. Dubois, and P. Mermelstein, "Low bit-rate video transmission over fading channels

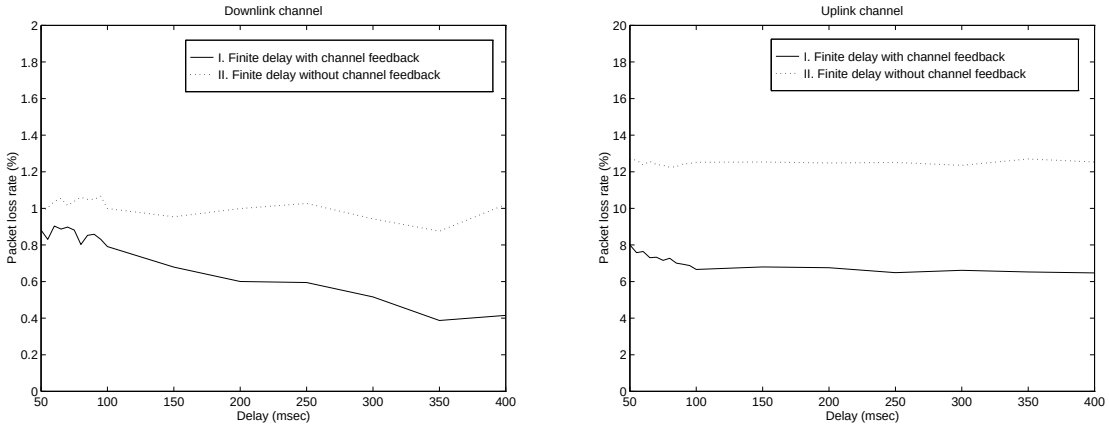


Figure 7: Packet loss rates for scenario (I) and (II) with different delay constraints. Note that no packet loss occur in scenario (III) because of the deterministic knowledge of all the channel conditions and R-D relations of source video.

for wireless microcellular systems," *IEEE Trans. on Circ. and Sys. for Video Tech.*, pp. 1–11, Feb. 1996.

[3] E. Steinbach, N. Färber, and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environments," *Submitted for publication in IEEE Trans. on CAS for video tech.*, 1996.

[4] National Semiconductor Corporation, *ITU-T/SG15/LBC-95-309 Sub-video with retransmission and intra-refreshing in mobil/wireless environments*, Oct. 1995.

[5] S. Lin, D. J. Costello, and M. Miller, "Automatic repeat request error control schemes," *IEEE Communications Magazine*, pp. 5–17, 1984.

[6] A. Ortega and M. Khansari, "Rate control for video coding over variable bit rate channels with applications to wireless transmission," in *Proc. of the 2nd. Intl. Conf. on Image Proc., ICIP'95*, (Washington, D.C.), Oct. 1995.

[7] C.-Y. Hsu, A. Ortega, and A. R. Reibman, "Joint selection of source and channel rate for VBR video transmission under ATM policing constraints," *Accepted for publication in IEEE J. on Sel. Areas in Comm., Special Issue on Real-Time Video Services in Multimedia Networks*, 1997.

[8] Telenor R & D, "TMN-Test model for H.263." `ftp://bonde.nta.no/pub/tmn/software`.

[9] V. K. Varma, "Testing speech coders for usage in wireless communications system," in *Proc. of IEEE Speech Coding Workshop*, (Montreal), 1993.

[10] A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal trellis-based buffered compression and fast approximations," *IEEE Trans. on Image Proc.*, vol. 3, pp. 26–40, Jan. 1994.

[11] J. Choi and D. Park, "A stable feedback control of the buffer state using the controlled lagrange multiplier method," *IEEE Trans. on Image Proc.*, vol. 3, pp. 546–558, Sept. 1994.

[12] J.-J. Chen and D. W. Lin, "Optimal bit allocation for video coding under multiple constraints," in *Proc. IEEE Intl. Conf. on Image Proc., ICIP'96*, (Lausanne,Switzerland), 1996.