ADAPTIVE QOS CONTROL FOR MPEG-4 VIDEO COMMUNICATION OVER WIRELESS CHANNELS

Dapeng W u   Yiwei Thomas Hou[†]   Ya-Qin Zhang[‡]   W enwu Zhu   H. Jonathan Chao[§]

[*] Carnegie Mellon Univ ersity, Pittsburgh, PA, USA
[†] Fujitsu Laboratories of America, Sunn yvale, CA, USA
[‡] Microsoft Researc h, Beijing, China
[§] Polytechnic University, Brooklyn, NY, USA

ABSTRACT

This paper proposes an adaptive quality-of-service (QoS) control to increase the robustness of MPEG-4 video communication over wireless channels. More specifically, the proposed adaptive QoS control consists of optimal mode selection and delay-constrained hybrid automatic repeat request (ARQ). The optimal mode selection is employed to provide QoS support on the compression layer while delay-constrained hybrid ARQ is used to provide QoS support on the link layer. Simulation results show that the proposed adaptive QoS control achieves satisfactory quality for MPEG-4 video under dynamically changing wireless channel conditions and utilizes network resources efficiently.

1. INTRODUCTION

Robust transmission of MPEG-4 video over wireless channels is a challenging problem [3]. This is because the bit error rate (BER) in wireless channels tends to be very high and the bit errors could have devastating effect on the video presentation quality.

To address this problem, this paper introduces an adaptive QoS control approach for MPEG-4 video communication over wireless channels. We first present a transport architecture, which includes source rate adaptation, packetization, adaptive QoS control, interleaving, and modulation. Then we describe the main contributions of this paper: (1) an optimal mode selection algorithm which provides the best trade-off between compression efficiency and error resilience in rate-distortion (R-D) sense, and (2) delay-constrained hybrid ARQ which is capable of reducing the BER for the compression layer while guaranteeing delay bound and achieving high throughput.

Previous work [1] on optimal mode selection considered wireless channel characteristics and error concealment. However, the distortion metrics introduced there are not accurate since the derivation of the distortion metrics was done at the block level. In addition, the BER in the wireless channel is varying dynamically. The scheme in [1] assumes that the BER is fixed and known a priori, which may not reflect the error behavior in the wireless channel. This paper addresses these problems by deriving accurate distortion metrics at the pixel level and employing feedback mecha-

nism to deal with the time-varying nature of the wireless channel.

Previous work on hybrid ARQ scheme [6] achieves delay bound by limiting retransmission number. However, it assumes that the maximum number of retransmissions is fixed and known a priori, which may not reflect the time-varying nature of delay. We address this problem by employing delay-constrained hybrid ARQ, which is shown to achieve higher throughput and better perceptual quality.

The remainder of this paper is organized as follows. Section 2 gives an overview of our architecture for MPEG-4 video over the wireless channel. In Section 3, we describe our optimal mode selection algorithm. Section 4 presents the delay-constrained hybrid ARQ. In Section 5, we use simulation results to demonstrate the performance of our adaptive QoS control. Section 6 concludes this paper.

2. AN ARCHITECTURE FOR MPEG-4 VIDEO OVER THE WIRELESS CHANNEL

Figure 1 shows our architecture for point-to-point MPEG-4 video communication over a wireless channel. On the sender side, raw bit-stream of live video is encoded by an rate-adaptive MPEG-4 encoder, which also employs optimal mode selection. After this stage, the compressed video bit-stream is first packetized and then passed to CRC & RCPC encoder. Cyclic redundancy check (CRC) code and RCPC codes are used as the error detection and correction codes. Following the RCPC coder, an interleaver is used to randomize the error bursts in each packet. After modulation, packets are transmitted over a wireless channel. At the receiver, the video sequence is reconstructed in a manner shown in Fig. 1.

Under our architecture, a QoS monitor is kept at the receiver side to infer channel status based on the behavior of the arriving packets, e.g., bit errors on the link layer, delay, and packet errors on the compression layer (called residual packet error). Link-layer bit errors and delay are used by delay-constrained hybrid ARQ to decide whether a retransmission should be requested. In addition, residual packet error ratio (RPER) is periodically sent back to the source; the video encoder makes the optimal mode selection based on the returned RPER.
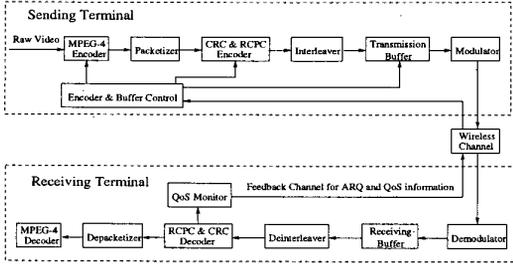
Figure 1: An architecture for MPEG-4 video over wireless channel.

Under our architecture, we consider a wireless link with fixed capacity, which is fully devoted to the connection of interest. We use rate-adaptive MPEG-4 encoder [5] so that the rate of compressed video can match the available bandwidth of the wireless link. The available bandwidth is meant to be the link capacity reduced by the forward error correction (FEC) redundancy and retransmissions. In other words, the aggregate rate of compressed video, the FEC redundancy and retransmissions has to match the link capacity in order to fully utilize the bandwidth while preventing channel overload.

Under our architecture, it is desirable to have a packetization scheme to achieve robustness against errors. To achieve error resilience, we attempt to divide the compressed video stream into packets that are as independent of each other as possible. That is, each packet should generally contain one or several complete logical entities of video information, e.g., macroblock (MB), so that the errors in a packet does not affect other packets. Furthermore, the size of each packet should be chosen to achieve a good trade-off between efficiency and robustness [5].

Our packetization scheme employs the Group of Blocks (GOB) or slice structure. If GOB structure is available, we choose a complete GOB as a packet payload. If GOB structure is not available, which is the case for a video object (VO) with arbitrary shape, we choose a complete slice as a packet payload. We define a slice as the part of GOB confined by two shape boundaries of the VO.

After packetization, the video packet is passed to the link layer (CRC & RCPC encoder). For each video packet, the CRC & RCPC encoder generates a link-layer packet by using CRC/RCPC codes.

3. OPTIMAL MODE SELECTION

In this section, we present an optimal mode selection algorithm based on the work in [4]. We organize this section as follows. Sections 3.1 describes the global distortion. In Section 3.2, we design an algorithm for optimal mode selection based on the global distortion.

3.1. Global Distortion

The problem of classical R-D optimized mode selection is to find the mode that minimizes the quantization distortion D_q for a given MB, subject to a constraint. The classical

R-D optimized mode selection is optimal with respect to quantization distortion. However, under wireless environment, where packets may get discarded due to unrecoverable errors, the classical R-D optimized mode selection is not optimal with respect to the distortion D_r, which measures the difference between the original MB at the source and the reconstructed one at the receiver [4]. This is because the classical R-D optimized mode selection does not consider the channel characteristics and receiver behavior, which also affect the distortion D_r.

Under wireless environment, the distortion D_r is a random variable, which may take the value of either (1) the quantization distortion D_q plus the error propagation distortion D_{ep}, or (2) distortion D_c caused by errors due to error concealment. We define the global distortion D as the expectation of the random variable D_r.

Now the problem of globally R-D optimized mode selection amounts to finding the mode that minimizes the global distortion D for a given MB, subject to a constraint R_c on the number of bits used. In terms of mean absolute difference (MAD), we define the global distortion metric for the MB at location i in frame n, F_i^n, as follows.

$$\text{MAD}(F_i^n) = \frac{E\left\{\sum_{j=1}^{256} |f_{ij}^n - \hat{f}_{ij}^n|\right\}}{256}, \qquad (1)$$

where f_{ij}^n is the original value of pixel j in F_i^n (raw data), and $\hat{f}_{ij}^n$ is the value of reconstructed pixel j in F_i^n at the receiver. The global distortion is affected by three factors: sender behavior, channel characteristics, and receiver behavior. The source behavior consists of quantization and packetization. Channel characteristics can be characterized by residual packet loss ratio p_d. Receiver behavior is determined by the error concealment scheme.

In our system, we employ the following error concealment scheme. The corrupted MB is replaced with the MB from the previous frame pointed by a motion vector. The motion vector of the corrupted MB is copied from one of its neighboring MB (which is above or below the corrupted MB) when available, otherwise the motion vector is set to zero.

3.2. Optimal Mode Selection

Based on the architecture described in Section 2, the global distortion metrics can be derived. We summarize the main results as follows.

The MAD for the MB at location s of frame N $(N > 0)$ is given by

$$\text{MAD}(F_s^N) = \frac{\sum_{j=1}^{256} |f_{sj}^N - E\{\hat{f}_{sj}^N\}|}{256}, \qquad (2)$$

where

$$E\{\hat{f}_{ij}^n\} = \begin{cases} (1 - p_d) \cdot \hat{f}_{ij}^n + p_d^3 \cdot E\{\hat{f}_{ij}^{n-1}\} \\ + p_d \cdot (1 - p_d^2) \cdot E\{\hat{f}_{ml}^{n-1}\} \\ \qquad \text{if the MB is intra-coded} \\[1em] (1 - p_d) \cdot (\tilde{e}_{ij}^n + E\{\hat{f}_{uv}^{n-1}\}) + p_d^3 \cdot E\{\hat{f}_{ij}^{n-1}\} \\ + p_d \cdot (1 - p_d^2) \cdot E\{\hat{f}_{ml}^{n-1}\} \\ \qquad \text{if the MB is inter-coded} \end{cases} \qquad (3)$$

and $\tilde{f}_{ij}^n$ is the value of reconstructed pixel j in F_i^n at the encoder, $\tilde{e}_{ij}^n$ is the reconstructed prediction error of pixel j in inter-coded F_i^n, $\tilde{f}_{uv}^{n-1}$ is the value of reconstructed pixel v in F_u^{n-1} for prediction of f_{ij}^n, and $\hat{f}_{ml}^{n-1}$ is the value of reconstructed pixel l in F_m^{n-1} to replace $\hat{f}_{ij}^n$ due to the error concealment. $E\{\hat{f}_{ij}^0\}$ is initialized with $\hat{f}_{ij}^0$, which is known at the encoder. $E\{\hat{f}_{ij}^n\}$ (for frame $n > 0$) can be obtained through the iteration in (3).

The problem of globally R-D optimized mode selection is to find the combination of modes that minimizes the distortion for a given GOB/slice, subject to a constraint R_c on the number of bits used. This constrained problem can be formulated as

$$\min_{\mathcal{M}_g^n} D(\mathcal{F}_g^n, \mathcal{M}_g^n) \quad \text{subject to} \quad R(\mathcal{F}_g^n, \mathcal{M}_g^n) \leq R_c, \quad (4)$$

where $D(\mathcal{F}_g^n, \mathcal{M}_g^n)$ and $R(\mathcal{F}_g^n, \mathcal{M}_g^n)$ denote the total distortion and bit budget, respectively, for the GOB/slice $\mathcal{F}_g^n$ with a particular mode combination $\mathcal{M}_g^n$.

Problem (4) can be converted to an unconstrained minimization problem by Lagrange multiplier technique as follows.

$$\sum_{i=g}^{g+N_G-1} \min_{M_i^n}\{D(F_i^n, M_i^n) + \lambda R(F_i^n, M_i^n)\}, \quad (5)$$

where the global distortion $D(F_i^n, M_i^n)$ can be expressed by Eq. (2), N_G is the number of MBs in a GOB/slice, and g is the location number of the first MB in the GOB/slice.

We use the following method to obtain λ. At the end of frame n, we adjust λ for frame $n+1$ (i.e., λ_{n+1}) as follows:

$$\lambda_{n+1} = \frac{2 \cdot B_n + (\gamma - B_n)}{B_n + 2 \cdot (\gamma - B_n)} \cdot \lambda_n \quad (6)$$

where B_n is the current buffer occupancy at the end of frame n and γ is the buffer size. λ_n is initialized by a preset value λ_0. The adjustment in Eq. (6) is to keep the buffer occupancy at the middle level to reduce the chance of buffer overflow or underflow.

4. DELAY-CONSTRAINED HYBRID ARQ

The previous section used optimal mode selection to provide QoS support on the compression layer. This section presents QoS support mechanism for video communication on the link layer. There are two such mechanisms, namely, FEC and ARQ. However, although FEC is able to achieve bounded delay for video transmission, it lacks adaptiveness to varying wireless channel. On the other hand, ARQ is adaptive to varying channel while delay is unbounded.

To overcome the problems associated with FEC and ARQ, truncated type-II hybrid ARQ has been proposed [6]. Truncated type-II hybrid ARQ imposes a limit on maximum number of retransmissions for a packet to bound the delay. That is, such scheme assumes that the maximum number of retransmissions N_m is fixed and known a priori, which may not reflect the time-varying nature of delay. If N_m is set too large, retransmitted packets may arrive too late for play-out and thereby be discarded, resulting in waste of

bandwidth; if N_m is set too small, the perceptual quality will be reduced due to unrecoverable errors that could have been corrected with more retransmissions. In the rest of this section, we present a delay-constrained hybrid ARQ to address this problem.

Our delay-constrained hybrid ARQ is aimed to achieve bounded delay, adaptiveness and efficiency. We employ CRC-16 and RCPC [6] as the FEC codes and RCPC offers variable rates to adapt to the time-varying wireless channel. Our hybrid ARQ works as follows. When errors in the received link-layer packet is detected, the receiver decides whether to send a retransmission request based on the delay bound of the packet. If the delay requirement cannot be met, the request will not be sent. Otherwise, a retransmission request will be sent to the source. Upon a retransmission request, the source only transmits the necessary incremental redundancy bits.

5. SIMULATION RESULTS

In this section, we implement the architecture described in Section 2 on our simulator and evaluate the performance of our optimal mode selection algorithm and delay-constrained hybrid ARQ. At the source side, we use the standard raw video sequence "Miss America" in QCIF format for the video encoder. The encoder employs the rate control described in [5] to keep a constant rate at 64 Kbits/s. The frame rate is 10 frames/s. The encoder is used in the rectangular mode, with intra-VOP refreshment period of 50 frames and $\lambda_0 = 1$. We consider three different encoders for MPEG-4 video as follows.

Encoder A: employs the classical approach for R-D optimized mode selection.

Encoder B: implements the globally R-D optimized mode selection described in Section 3. However, feedback of RPER is not employed.

Encoder C: implements the globally R-D optimized mode selection described in Section 3. Feedback of RPER to the source is employed.

In our simulations, we employ binary symmetric channel (BSC) as the wireless channel model. We assume that the feedback channel is error-free. Two sets of RCPC codes are used with the ARQ schemes. Both sets are generated from a rate $1/4$ code with constraint length $K = 5$ and puncturing period $V = 5$ (see Table B.186 in [2]). The rates of Set $\mathcal{A}$ are 1, 5/20, 5/18, 5/16, 5/14, 5/12, 5/10, 5/9, 5/8, 5/6, and 1/4. The rates of Set $\mathcal{B}$ are 1, 5/16, 5/10, and 1/4. At the receiver side, if an erroneous link-layer packet cannot be recovered by RCPC at its playout time, the link-layer packet and its retransmitted redundancy will be discarded. Each simulation is conducted for 100-second simulation time.

5.1. Performance of Optimal Mode Selection

In this simulation, we evaluate the performance of different optimal mode selection schemes, namely, Encoder A, B, and C. For Encoder B, we set RPER=1% for optimal mode selection. The round-trip time (RTT) is set at 0.01 second. Figure 2 shows the average peak signal-to-noise ratios (PSNR's) of Y component under BER of 10^{-5}, $5 \times$
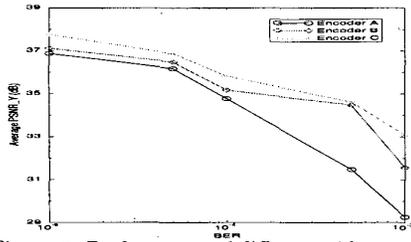
Figure 2: Performance of different video encoders.
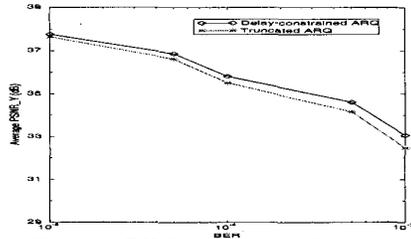
Figure 3: PSNR for hybrid ARQ schemes.

10^{-5}, 10^{-4}, 5×10^{-4}, and 10^{-3}. We observe that Encoder C achieves the best performance, Encoder B has the second best performance, and Encoder A performs the worst. That is, our approach achieves better performance than the classical one, even if feedback mechanism is not employed. Feedback-based scheme (i.e., Encoder C) achieves better performance than non-feedback-based scheme (i.e., Encoder B) since Encoder C has more accurate channel information than Encoder B.

5.2. Performance of Hybrid ARQ

To compare the performance of the truncated type-II hybrid ARQ [6] and our delay-constrained hybrid ARQ, we run our simulations under two scenarios where only Encoder C is employed.

In the first scenario, we set RTT=0.01 second and the maximum number of retransmissions N_m=3 for the truncated hybrid ARQ. The truncated hybrid ARQ uses the rates of Set B for RCPC codes; our delay-constrained hybrid ARQ uses the rates of Set A for RCPC codes. Figure 3 shows that our delay-constrained hybrid ARQ achieves higher PSNR than the truncated hybrid ARQ. This is because an erroneous link-layer packet can be retransmitted 10 times under our delay-constrained hybrid ARQ without violating the delay bound while the truncated hybrid ARQ only allows 3 retransmissions, resulting in lower recovery probability.

In the second scenario, we set RTT=0.03 second and the maximum number of retransmissions N_m=10 for the truncated hybrid ARQ. Both hybrid ARQ scheme use the rates of Set A for RCPC codes. Figure 4 shows that our delay-constrained hybrid ARQ achieves higher throughput than the truncated hybrid ARQ. This is because an erroneous link-layer packet may be retransmitted up to 3 times under our delay-constrained hybrid ARQ without violating the delay bound while the truncated hybrid ARQ allows 10 retransmissions, resulting in waste of bandwidth due to discard of packets that missed their playout time.
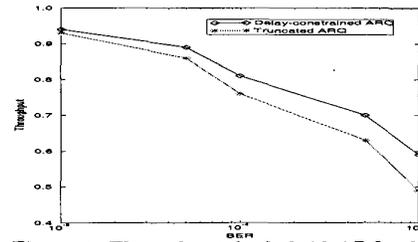
Figure 4: Throughput for hybrid ARQ schemes.

The throughput shown in Fig. 4 is defined as the ratio of average transmission rate of compression-layer packets over the channel rate.

6. CONCLUDING REMARKS

This paper proposed two adaptive QoS control mechanisms for robust transmission of MPEG-4 video over wireless channels, namely, optimal mode selection and delay-constrained hybrid ARQ. By combining the best features of error-resilient source encoding, FEC and delay-constrained retransmission, the proposed adaptive QoS control is capable of achieving bounded delay, high reliability and efficiency. Simulation results demonstrated that the proposed adaptive QoS control achieves better performance than some other proposed approaches.

7. REFERENCES

[1] G. Cote, S. Shirani and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error prone networks," submitted to IEEE J. on Selected Areas in Communications.

[2] P. Frenger, P. Orten, T. Ottosson, and A. Svensson, "Multi-rate convolutional codes," Tech. Report No. 21, Dept. of Signals and Systems, Chalmers University of Technology, April 1998.

[3] ISO/IEC JTC 1/SC 29/WG 11, "Information technology – coding of audio-visual objects, part 1: systems, part 2: visual, part 3: audio," FCD 14496, Dec. 1998.

[4] D. Wu, Y. T. Hou, B. Li, W. Zhu, Y.-Q. Zhang and H. J. Chao, "An end-to-end approach for optimal mode selection in Internet video communication: theory and application," to appear in IEEE J. on Selected Areas in Communications, 2000.

[5] D. Wu, Y. T. Hou, W. Zhu, H.-J. Lee, T. Chiang, Y.-Q. Zhang and H. J. Chao, "On end-to-end architecture for transporting MPEG-4 video over the Internet," to appear in IEEE Trans. on Circuits and Systems for Video Technology, 2000.

[6] Q. Zhang and S. A. Kassam, "Hybrid ARQ with selective combining for fading channels," IEEE J. on Selected Areas in Communications, vol. 17, no. 5, pp. 867–880, May 1999.