An End-to-End Approach for Optimal Mode Selection in Internet Video Communication: Theory and Application

Dapeng Wu, Yiwei Thomas Hou, Bo Li, Wenwu Zhu, Ya-Qin Zhang, and H. Jonathan Chao

Abstract— **Rate-distortion (R-D) optimized mode selection is a fundamental problem for video communication over packet-switched networks. The classical R-D optimized mode selection only considers quantization distortion at the source. Such approach is unable to achieve global optimality under the error-prone environment since it does not consider the packetization behavior at the source, the transport path characteristics and receiver behavior. This paper presents an *end-to-end approach* to generalize the classical theory of R-D optimized mode selection for point-to-point video communication. We introduce a notion of *global* distortion by taking into consideration both the path characteristics (i.e., packet loss) and the receiver behavior (i.e., the error concealment scheme), in addition to the source behavior (i.e., quantization distortion and packetization). We derive, for the first time, a set of accurate global distortion metrics for any packetization scheme. Equipped with the global distortion metrics, we design an R-D optimized mode selection algorithm to provide the best trade-off between compression efficiency and error resilience. The theory developed in this paper is general and is applicable to many video coding standards, including H.261/263 and MPEG-1/2/4. As an application, we integrate our theory with point-to-point MPEG-4 video conferencing over the Internet, where a feedback mechanism is employed to convey the path characteristics (estimated at the receiver) and receiver behavior (error concealment scheme) to the source. Simulation results conclusively demonstrate that our end-to-end approach offers superior performance over the classical approach for Internet video conferencing.**

Keywords— **Internet, video conferencing, R-D optimized mode selection, global distortion metric, feedback, packetization, error concealment, MPEG-4.**

I. INTRODUCTION

VIDEO communication over the Internet is becoming an important application in recent years. A challenging problem associated with Internet video communication lies in how to cope with packet loss in the network and achieve acceptable video quality at the receiver. This is because packet loss is unavoidable in the Internet and may have significant impact on perceptual quality.

The effect of lost packets on the video presentation qual-

Manuscript received May 1, 1999; revised November 1, 1999.

D. Wu is with Carnegie Mellon University, Dept. of Electrical & Computer Engineering, 5000 Forbes Ave., Pittsburgh, PA 15213, USA.

Y. T. Hou is with Fujitsu Laboratories of America, 595 Lawrence Expressway, Sunnyvale, CA 94086, USA.

B. Li is with Hong Kong University of Science and Technology, Dept. of Computer Science, Clear Water Bay, Kowloon, Hong Kong.

W. Zhu and Y.-Q. Zhang are with Microsoft Research, China, 5F, Beijing Sigma Center, No. 49, Zhichun Road Haidian District, Beijing 100080, China.

H. J. Chao is with Polytechnic University, Dept. of Electrical Engineering, Six Metrotech Center, Brooklyn, NY 11201, USA.

ity depends on the coding scheme used at the source, the network congestion status and the error concealment scheme used at the receiver. High-compression coding algorithms usually employ inter-coding (i.e., prediction) to achieve efficiency. With these coding algorithms, loss of a packet may degrade video quality over a large number of frames, until the next intra-coded frame is received. Intra-coding can effectively stop error propagation at the cost of efficiency while inter-coding can achieve compression efficiency at the risk of error propagation. Therefore, a good mode selection between intra mode and inter mode should be in place to enhance the robustness of the video communications using intra- and inter-coding.

For video communication over a network, a coding algorithm such as H.263 or MPEG-4 [7] usually employs rate control to match the output rate to the available bandwidth. The objective of rate-controlled compression algorithms is to maximize the video quality under the constraint of a given bit budget. This can be achieved by choosing a mode that minimizes the quantization distortion between the original frame/macroblock and the reconstructed one under a given bit budget [10], [16], which is the so-called rate-distortion (R-D) optimized mode selection. We refer such R-D optimized mode selection as the classical approach. The classical approach is not able to achieve global optimality under the error-prone environment since it does not consider the network congestion status and the receiver behavior.

This paper presents an *end-to-end approach* to solve the fundamental problem of R-D optimized mode selection for peer-to-peer video communication over packet-switched networks. Under the end-to-end approach, we identify three factors that have impact on the video presentation quality at the receiver, namely, the source behavior, the path characteristics, and the receiver behavior. To put such end-to-end approach into a theoretical framework, we develop a theory for globally optimal mode selection under packet lossy environment. We begin with formulating the problem of globally optimal mode selection using the notion of global distortion metric. Then we describe the three factors in the end-to-end approach. We derive, for the first time, a set of accurate global distortion metrics for any packetization scheme. We show how to apply the global distortion metrics to specific packetization scheme. Equipped with the global distortion metrics, we design an R-D optimized mode selection algorithm to provide the best trade-off between compression efficiency and

error resilience. Our theory on R-D optimized mode selection is general and is applicable to numerous video coding standards, including H.261/263 and MPEG-1/2/4.

As an application, we integrate our theory with point-to-point MPEG-4 video conferencing over the Internet, and employ a feedback mechanism to convey the path characteristics (estimated at the receiver) and receiver behavior (error concealment scheme) to the source. Simulation results conclusively demonstrate that our end-to-end approach offers superior performance over the classical approach for Internet video conferencing.

The error resilience mechanisms in the literature include re-synchronization marking, data partitioning, data recovery (e.g., reversible variable length codes (RVLC)), and error concealment [17], [19], [24], [25], [26]. However, re-synchronization marking, data partitioning, and data recovery are targeted at error-prone environment like wireless channel and may not be applicable to Internet environment. For video transmission over the Internet, the boundary of a packet already provides a synchronization point in the variable-length coded bit-stream at the receiver side. Since a packet loss may cause the loss of all the motion data and its associated shape/texture data, mechanisms such as re-synchronization marking, data partitioning, and data recovery may not be useful for Internet video communications. On the other hand, most error concealment techniques discussed in [20] are only applicable to either ATM or wireless environment, and require substantial additional computation complexity, which may be tolerable in decoding still images but not acceptable for decoding real-time video. Therefore, we only consider simple error concealment schemes that are applicable to Internet video communication (see Section II-B.3).

Previous work on optimal mode selection that took into account of path characteristics and error concealment has been reported in Ref. [3], [9]. However, the scheme in [3] is based on simple Bernoulli loss model, which may not capture the bursty nature of the Internet. Measurements of packet loss in the Internet show that the bursty packet loss behavior can be modeled reasonably well with a 2-state Markov chain, i.e., the Gilbert model [1], [5]. In addition, the congestion status of the Internet is dynamically changing, that is, the loss probability is varying from time to time. The scheme in [3] assumes that the loss probability is fixed and known *a priori*, which may not reflect the loss behavior in the real Internet. Thus, the scheme in [3] may not achieve optimality for the dynamic Internet environment. This paper addresses these problems by employing the Gilbert path model and introducing feedback mechanism to deal with the dynamics of the Internet. In [9], only a special case is considered in the derivation of distortion metrics. Thus, the distortion metrics introduced there are not accurate. Furthermore, the distortion metrics introduced in [9] only apply to a specific packetization scheme and a specific error concealment scheme. On the contrary, this paper derives a set of accurate global distortion metrics for any packetization scheme. We also show how to extend the global distortion metrics to specific packetization scheme.

Prior work on feedback-based error resilient coding includes [15]. It has been shown in [15] that the feedback information does help the encoder select a proper coding mode to prevent error propagation and achieve better performance in terms of PSNR. However, there is no guideline on how to set the threshold in the coding mode selection [4]. This paper shows how feedback mechanism can be employed to convey transmission parameters estimated at the receiver in the real Internet environment where such parameters are not known *a priori* at the source.

Our approach aims at solving the fundamental problem of R-D optimized mode selection from the end-to-end perspective. We develop the theory of R-D optimized mode selection using the notion of global distortion. Equipped with a feedback mechanism in implementation, our approach is shown to be capable of achieving superior performance over the classical approach.

The remainder of this paper is organized as follows. Sections II and III lay the theoretical foundation of globally optimal mode selection. Specifically, Section II presents an end-to-end approach for globally optimal mode selection. In Section III, we derive a set of global distortion metrics and develop an algorithm for globally R-D optimized mode selection. As an application, Section IV presents an end-to-end implementation architecture for point-to-point MPEG-4 video conferencing over the Internet. In Section V, we use simulation results to demonstrate the performance improvement of our approach under various network configurations and varying network conditions. Section VI concludes this paper and points out future research directions.

II. AN END-TO-END APPROACH

This section describes the end-to-end approach to solving R-D optimized mode selection and set up the stage for our theoretical development in Section III. We organize this section as follows. In Section II-A, we introduce the notion of global distortion and formulate the problem of globally optimal mode selection. Section II-B examines the key factors contributing to the global distortion.

A. Problem Formulation

Figure 1 depicts an architecture for video communication across a network. On the sender side, raw bit-stream of live video is encoded by a video encoder. After this stage, the compressed video bit-stream is first packetized and then passed through the transport protocol layers before entering the network. Packets may be dropped inside the network (due to congestion) or at the destination (due to excess delay). For packets that are successfully delivered to the destination, they first pass through the transport protocol layers and depacketized before being decoded at the video decoder.

Table I lists the notations used in this paper.

Under QCIF format, N_f is 99 and N_h is 98. By using the notations, $\bar{F}_i^n \prec X_k$ means that coded macroblock $\bar{F}_i^n$ is completely contained by packet X_k; $\bar{F}_i^n \preceq X_k$ means that coded macroblock $\bar{F}_i^n$ is partially contained by packet X_k

<div align="center">TABLE I</div>

<div align="center">NOTATIONS.</div>

N_f	:	the total number of MBs in a frame
N_h	:	the highest location number of MBs in a frame ($N_h = N_f - 1$)
N_G	:	the number of MBs in a group of blocks (GOB)
F_i^n	:	the MB at location i in frame n
$\bar{F}_i^n$	:	the coded MB at location i in frame n.
$F_{\bar{i}}^n$	:	the MB (at location $\bar{i}$ in frame n) which is above F_i^n, if it exists.
$\mathcal{G}^n$	:	the set of macroblocks $\bar{F}_i^n$ ($i \in [0, N_h]$) that does not have $F_{\bar{i}}^n$.[†]
$\tilde{P}_R^{(i,n)}$	:	the probability of the event that $\bar{F}_i^n$ is received correctly.
$\tilde{P}_L^{(i,n)}$	:	the probability of the event that $\bar{F}_i^n$ is lost.
$\hat{P}_{RL}^{(i,n)}$	:	the probability of the event that $\bar{F}_{\bar{i}}^n$ is received correctly and $\bar{F}_i^n$ is lost.
$\hat{P}_{LL}^{(i,n)}$	:	the probability of the event that both $\bar{F}_{\bar{i}}^n$ and $\bar{F}_i^n$ are lost.
f_{ij}^n	:	the original value of pixel j in F_i^n (raw data).
$\tilde{f}_{ij}^n$	:	the value of reconstructed pixel j in F_i^n at the encoder.
$\hat{f}_{ij}^n$	:	the value of reconstructed pixel j in F_i^n at the receiver.
e_{ij}^n	:	the prediction error of pixel j in inter-coded F_i^n.
$\tilde{e}_{ij}^n$	:	the reconstructed prediction error of pixel j in inter-coded F_i^n.
$\tilde{f}_{uv}^{n-1}$	:	the value of reconstructed pixel v in F_u^{n-1} for prediction of f_{ij}^n.
$\hat{f}_{ml}^{n-1}$	:	the value of reconstructed pixel l in F_m^{n-1} to replace $\hat{f}_{ij}^n$ due to EC-3.[‡]
$\mathcal{I}$	:	the set of coding modes (i.e., $\mathcal{I} = \{$intra, inter$\}$).
M_i^n	:	the mode selected to code macroblock F_i^n ($M_i^n \in \mathcal{I}$).
X_k	:	the packet with sequence number k ($k \geq 0$).
η_i^n	:	the sequence number of the last packet used to packetize macroblock $\bar{F}_i^n$.
$\prec$	:	the completely containing relation between a macroblock and a packet.
$\preceq$	:	the partially containing relation between a macroblock and a packet.
$\mathcal{X}$	:	the set of packets that packetize frame 0, i.e., $\mathcal{X} = \{X_{\eta_i^0} : i \in [0, N_h]\}$.
K	:	the number of packets in set $\mathcal{X}$.

[†] For example, in QCIF, an MB in the first GOB does not have an MB above it.

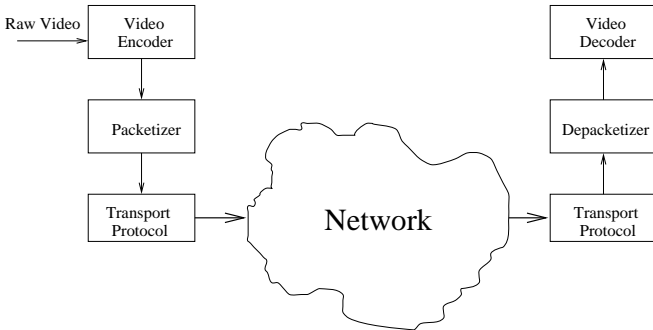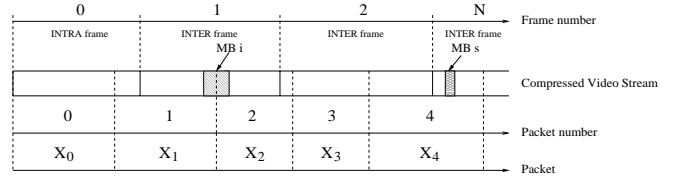[‡] EC-3 will be defined in Section II-B.3.

Fig. 1. Overview of the video communication.

Fig. 2. Illustration of packet number η_i^n.

(i.e., $\bar{F}_i^n$ is split and contained in two consecutive packets). In the example of Fig. 2 (where $N = 3$), packet 4 is the last packet used to packetize macroblock s in frame N ($\bar{F}_s^N$). So η_s^N is 4. In addition, packet 2 is the last packet used to packetize MB i in frame 1. Thus, η_i^1 is 2.

In formulating the problem of globally R-D optimized mode selection, we consider an MB at location i ($i \in [0, N_h]$) of a given frame. We assume that each MB can be coded using only one of the two modes in $\mathcal{I}$.

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 R_c on the number of bits used. This constrained problem can be formulated as

$$\min_{M_i^n} D_q(M_i^n) \quad \text{subject to} \quad R(M_i^n) \leq R_c, \quad (1)$$

where $D_q(M_i^n)$ and $R(M_i^n)$ denote the quantization distortion and the number of bits used, respectively, for macroblock F_i^n with a particular mode M_i^n.

The classical R-D optimized mode selection is optimal with respect to quantization distortion. However, under environments such as Internet, where packets may get lost either due to network congestion or due to exceeding the maximum delay threshold, the classical R-D optimized

mode selection is not optimal with respect to the distortion D_r, which measures the difference between the original image/frame/MB at the source and the reconstructed one at the receiver. This is because the classical R-D optimized mode selection does not consider the path characteristics (packet loss) and receiver behavior (error concealment), both of which affect the distortion D_r. This motivates us to propose globally R-D optimized mode selection.

We consider the distortion D_r, which is the difference between the original image/frame/MB at the source and the reconstructed one at the receiver. Under lossy environments such as Internet and wireless communication, the distortion D_r is a random variable, which may take the value of either (1) the quantization distortion D_q plus the distortion D_{ep} caused by error propagation, 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. That is,

$$D = E\{D_r\}, \tag{2}$$

where D_r takes the value of $(D_q + D_{ep})$ or D_c with certain probability, which is determined by path characteristics (packet loss behavior). Therefore, the global distortion is affected by three factors: sender behavior (quantization and packetization), path characteristics, and receiver behavior (error concealment).

The problem of globally R-D optimized mode selection is to find the mode that minimizes the global distortion D for a given MB, subject to a constraint R_c on the number of bits used. This constrained problem reads as follows.

$$\min_{M_i^n} D(M_i^n) \quad \text{subject to} \quad R(M_i^n) \leq R_c, \tag{3}$$

where $D(M_i^n)$ denotes the global distortion for macroblock F_i^n with a particular mode M_i^n.

The global distortion can be expressed by the sum of absolute differences (SAD), mean absolute difference (MAD), the sum of squared differences (SSD), mean squared error (MSE), or peak signal-to-noise ratio (PSNR). In this paper, we define the global distortion metrics for macroblock F_i^n in terms of MAD as follows.

$$\text{MAD}(F_i^n) = \frac{E\left\{\sum_{j=1}^{256} |f_{ij}^n - \hat{f}_{ij}^n|\right\}}{256} \tag{4}$$

In the rest of the paper, we will develop theory based on MAD. However, our underlying methodology is general and can be applied to other global distortion metrics (i.e., SAD, SSD, MSE, PSNR).

B. Key Factors in the Global Distortion Metric

This subsection discusses in detail the three factors (i.e., sender behavior, path characteristics, and receiver behavior) that contribute to the global distortion.
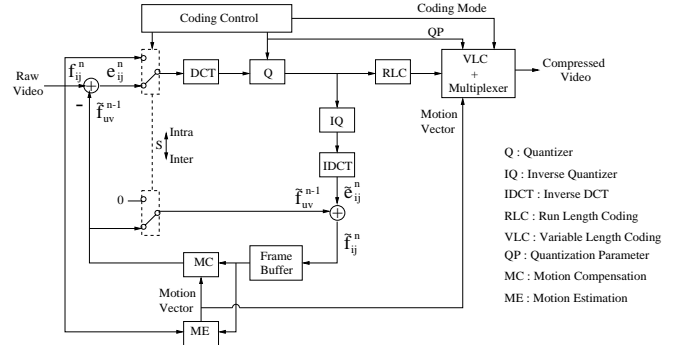

Fig. 3. Block diagram of the video encoder under the inter mode.

B.1 Source Behavior

The source behavior includes quantization and packetization, which have impact on global distortion.

A block diagram of the video encoder is depicted in Fig. 3. The switches represent the two different paths for the intra- and inter-mode.

Under the intra mode, the raw video f_{ij}^n is transformed by DCT, quantized and coded by run length coding (RLC). The resulting information, as well as coding parameters such as the coding mode and the quantization parameter (QP), is coded by variable length coding (VLC). Then the compressed video stream is formed by the multiplexer. At the same time, the pixel is reconstructed at the encoder for the prediction used by the next frame. The value of the reconstructed pixel is $\tilde{f}_{ij}^n$.

Figure 3 also illustrates the case when the encoder is operating under the inter-mode. Under such mode, the raw video f_{ij}^n is first predicted from the motion-compensated pixel in the previous frame, $\tilde{f}_{uv}^{n-1}$, resulting in the prediction error, e_{ij}^n. The prediction error, e_{ij}^n, is DCT-transformed, quantized and coded by RLC. The resulting information, as well as coding parameters such as the coding mode, the motion vector and the QP, is coded by VLC. Then the compressed video stream is formed by the multiplexer. At the same time, the pixel is reconstructed at the encoder for the prediction used by the next frame. There are two steps for the reconstruction at the encoder. First, the prediction error is reconstructed, resulting in $\tilde{e}_{ij}^n$. Second, $\tilde{e}_{ij}^n$ is added to the predicted value $\tilde{f}_{uv}^{n-1}$, resulting in the reconstructed value, $\tilde{f}_{ij}^n$. That is, the pixel is reconstructed by $\tilde{f}_{ij}^n = \tilde{e}_{ij}^n + \tilde{f}_{uv}^{n-1}$.

Before passed to the transport protocols and sent to the network, the compressed video stream will be packetized. To discuss the packetization at the source, we make the following assumption throughout the paper.

Assumption 1: The payload size of a packet is greater than the size of any MB.

For video applications over the Internet, a packet is always larger than an MB. In practice, the maximum size of an MB is 90 bytes and the default IP packet size is 576 bytes while the protocol overhead is no more than 50 bytes, resulting the payload with at least 526 bytes. Furthermore, it is also plausible to use large packet size for video applica-

tion to achieve efficiency. Therefore, Assumption 1 is valid for all practical purposes.

Under Assumption 1, since the size of a packet is always larger than the size of any MB, an MB could be split into at most two consecutive packets.

In the following, we define specific packetization schemes.

Definition 1 (Packetization Schemes)   A packetization scheme for video bit-stream is called

• PKT-1 if each generated packet has the same fixed packet size;

• PKT-2 if each generated packet solely contains a complete MB;

• PKT-3 if each generated packet solely contains a complete GOB/slice.

Under PKT-1, the packet size can be set as large as path MTU to achieve efficiency. PKT-1 is widely used due to its simplicity. If PKT-1 is used, there are only two cases for the relation between an MB and the last packet used to packetize it: (1) an MB $(\bar{F}_i^n)$ is completely contained by a single packet $(X_{\eta_i^n})$, i.e., $\bar{F}_i^n \prec X_{\eta_i^n}$; or (2) an MB $(\bar{F}_i^n)$ is split into two consecutive packets and partially contained by each packet $(X_{\eta_i^n})$,[1] i.e., $\bar{F}_i^n \preceq X_{\eta_i^n}$.

If PKT-2 is used, an MB never gets split into two packets. Thus, loss of a packet only corrupts one MB, which enhances the error resilient capability of the video. For this reason, PKT-2 was adopted by Internet Engineering Task Force (IETF) [18].

PKT-3 has similar property to that of PKT-2. That is, a GOB/slice/MB is never split into two packets.[2] Thus, loss of a packet only corrupts one GOB/slice. Therefore, PKT-3 was also adopted by IETF [27].

B.2 Path Characteristics

Path characteristics determine the probability of the event that a packet can be received correctly. A path is typically characterized by the Bernoulli model [3], [21] or the Gilbert model [1].

Bernoulli Path Model

Under the Bernoulli path model, the packet loss in the Internet is modeled as a Bernoulli process with only one parameter, packet loss probability p_l. In the end-to-end architecture for video communication over the Internet, the packet loss probability p_l can be obtained by measuring the numbers of lost packets and successfully received packets at the receiver.[3] Thus, p_l is given by

$$p_l = \frac{N_l}{N_l + N_r}, \qquad (5)$$

where N_l is the number of lost packets and N_r the number of successfully received packets.

Gilbert Path Model

Fig. 4. Gilbert path model.

R: Receive state

L: Loss state

Measurements of packet loss in the Internet have shown that the packet loss behavior can be modeled reasonably well with a 2-state Markov chain, i.e., the Gilbert model (see Fig. 4) [1]. That is, the Markov chain is in state R if a packet is received timely and correctly and in state L if a packet is lost either due to network congestion or due to exceeding the maximum delay threshold. The parameters p and q are the transition probabilities between states L and R. The durations of states L and R are exponentially distributed with respective mean lengths T_L and T_R, which are given by

$$T_L = \frac{1}{p} \qquad \text{and} \qquad T_R = \frac{1}{q}.$$

The probability of the event that the path is in state L (i.e., packet loss probability) is given by

$$P_L = \frac{T_L}{T_R + T_L} = \frac{q}{p + q}. \qquad (6)$$

The transition matrix $\mathbf{A}$ of the 2-state Markov chain is given by

$$\mathbf{A} = \begin{bmatrix} 1 - p & p \\ q & 1 - q \end{bmatrix}.$$

In the end-to-end architecture, the transition probabilities p and q are given by:

$$p = \frac{N_1}{N_1 + N_2} \qquad \text{and} \qquad q = \frac{N_3}{N_3 + N_4}, \qquad (7)$$

where N_1 is the number of successfully received packets when the previous packet is lost, N_2 is the number of lost packets when the previous packet is lost, N_3 is the number of lost packets when the previous packet is successfully received, and N_4 is the number of successfully received packets when the previous packet is successfully received.[4] N_1, N_2, N_3, and N_4 can be measured at the receiver.

Remark 1: It should be clear that when $p + q = 1$, the 2-state Markov chain becomes a Bernoulli process. That is, a Bernoulli process is a special case of the 2-state Markov chain.

B.3 Receiver Behavior

Receiver behavior, i.e., error concealment, affects the video quality. We define specific error concealment schemes as follows.

Definition 2 (Error Concealment Schemes)   An error concealment scheme is called

---

[1]To be specific, if $\bar{F}_i^n$ is split, $\bar{F}_i^n$ is partially contained by packet $X_{\eta_i^n - 1}$ and partially contained by packet $X_{\eta_i^n}$.

[2]MPEG-4 does not have the concept of GOB for video sequences with arbitrary shape. However, we can define a slice which is the part of GOB confined by two shape boundaries of the VO.

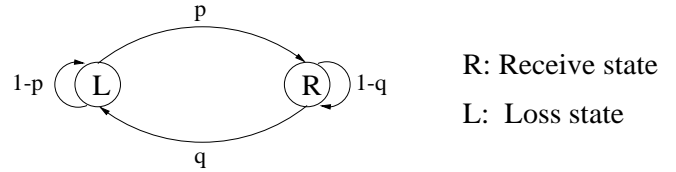[3]Packets that arrive later than the maximum expected time are regarded as lost packets.

[4]Packets that arrive later than the maximum expected time are regarded as lost.

• EC-1 if it replaces the whole frame (in which some MBs are corrupted) with the previous reconstructed frame;

• EC-2 if it replaces a corrupted MB with the MB at the same location from the previous frame;

• EC-3 if it replaces the corrupted MB with the MB from the previous frame pointed by a motion vector.[5]

We would like to stress that EC-1 and EC-2 are special cases of EC-3.

A block diagram of the video decoder is depicted in Fig. 5 where EC-3 is used. Three switches represent different scenarios as follows.

• *Switch S1:* represents the two different paths for the intra- and inter-mode.

• *Switch S2:* represents the two different paths for the two cases as follows.

— the case where the packet containing F_i^n is received correctly.

— the case where the packet containing F_i^n is lost.

• *Switch S3:* represents the two different paths for the two cases as follows.

— the case where the estimated motion vector of the corrupted F_i^n is set to zero.

— the case where the estimated motion vector of the corrupted F_i^n is copied from $F_{\bar{\imath}}^n$.

The compressed video stream is first demultiplexed by the demultiplexer and decoded by variable length decoding (VLD). The resulting coding mode information will control switch S1. The resulting QP will control the inverse quantizer (IQ). The resulting motion vector is used for motion compensation. The information containing DCT coefficient will be decoded by run length decoding (RLD), then inversely quantized, and inversely DCT-transformed.

Under EC-3, there are three cases for the reconstructed pixel at the receiver as follows.

• *Case (i):* The packet containing F_i^n is received correctly. If F_i^n is intra-coded, then we have $\hat{f}_{ij}^n = \tilde{f}_{ij}^n$, which is illustrated in Fig. 5 with state {S1: Intra, S2: No loss, S3: don't care}. If F_i^n is inter-coded, then we have

$$\hat{f}_{ij}^n = \tilde{e}_{ij}^n + \hat{f}_{uv}^{n-1}, \qquad (8)$$

which is illustrated in Fig. 5 with state {S1: Inter, S2: No loss, S3: don't care}.

• *Case (ii):* The packet containing F_i^n is lost and the packet containing the MB above ($\bar{F}_i^n$) has been received correctly. Then we have $\hat{f}_{ij}^n = \hat{f}_{ml}^{n-1}$, which is illustrated in Fig. 5 with state {S1: don't care, S2: Loss, S3: MV$\neq$0}.

• *Case (iii):* The packet containing F_i^n and the packet containing $\bar{F}_{\bar{\imath}}^n$ are all lost. Then we have $\hat{f}_{ij}^n = \hat{f}_{ij}^{n-1}$, which is illustrated in Fig. 5 with state {S1: don't care, S2: Loss, S3: MV=0}.

Thus far we have examined the three key components that contribute to the video quality in terms of global distortion. In the next section, we derive the global distortion metric based on the materials discussed in this section and design an algorithm for optimal mode selection.

[5]The motion vector of the corrupted MB is copied from one of its neighboring MB when available, otherwise the motion vector is set to zero.
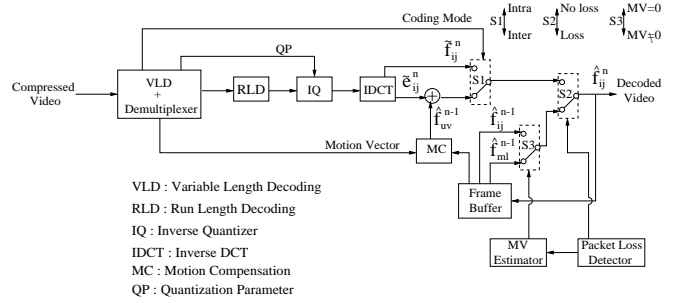

Fig. 5. Block diagram of the video decoder.

III. Optimal Mode Selection

In this section, by using an end-to-end approach, we derive, for the first time, a set of accurate global distortion metrics for any packetization scheme. We also design an algorithm for optimal mode selection based on the global distortion metrics.

Since measurements on the Internet indicate that packet loss behavior can be modeled reasonably well with a 2-state Markov chain, i.e., the Gilbert model [1], we thus characterize the path across the Internet with the Gilbert model.

We organize this section as follows. In Section III-A, we derive the global distortion metrics for an intra-coded MB and an inter-coded MB. In Section III-B, we design an algorithm for optimal mode selection based on the global distortion metrics.

A. Derivation of Global Distortion Metrics

Without loss of generality, we consider the distortion for macroblock s in frame N, where s ($s \in [0, N_h]$) is the location number and N ($N \geq 0$) is the frame number. Note that the sequence number for both frame and packet start from zero.

Assume the first I-frame of the video stream has been successfully received.[6] Given transition matrix $\mathbf{A}$ for the Gilbert path model, after transmission of n packets, the transition matrix becomes $\underbrace{\mathbf{A} \cdot \mathbf{A} \cdots \mathbf{A}}_{n}$, i.e., $\mathbf{A}^n$, where

$$\mathbf{A}^n = \left[\begin{array}{cc} 1-p & p \\ q & 1-q \end{array} \right]^n .$$

Since the resulting $\mathbf{A}^n$ is a 2×2 matrix, we can denote $\mathbf{A}^n$ as

$$\mathbf{A}^n = \left[\begin{array}{cc} P_{LL}^{(n)} & P_{LR}^{(n)} \\ P_{RL}^{(n)} & P_{RR}^{(n)} \end{array} \right],$$

where $P_{ij}^{(n)}$ ($i \in \{L, R\}$ and $j \in \{L, R\}$) denotes the transition probability from state i to state j after transmission of n packets.

Intra Mode

[6]The reason why we make this assumption is to initialize $E\{\hat{f}_{ij}^0\}$ used by the encoding process.

The following lemma shows how to compute the probability of the event that $\bar{F}_i^n$ is received correctly and the probability of the event that $\bar{F}_i^n$ is lost.

Lemma 1: The probabilities of $\tilde{P}_R^{(i,n)}$ and $\tilde{P}_L^{(i,n)}$ $(n \geq 0)$ are given respectively by

$$\tilde{P}_R^{(i,n)} = \begin{cases} 1 & \text{if } X_{\eta_i^n} \in \mathcal{X} \\[2mm] P_{RR}^{(1)} & \\ \quad \text{if } \bar{F}_i^n \preceq X_{\eta_i^n} \ (X_{\eta_i^n-1} \in \mathcal{X} \text{ and } X_{\eta_i^n} \notin \mathcal{X}) & \\[2mm] P_{RR}^{(\eta_i^n-K+1)} & \text{if } \bar{F}_i^n \prec X_{\eta_i^n} \ (X_{\eta_i^n} \notin \mathcal{X}) \\[2mm] P_{RR}^{(\eta_i^n-K)} \cdot P_{RR}^{(1)} & \text{if } \bar{F}_i^n \preceq X_{\eta_i^n} \ (X_{\eta_i^n} \notin \mathcal{X}) \end{cases}$$
$$(9)$$

and

$$\tilde{P}_L^{(i,n)} = \begin{cases} 0 & \text{if } X_{\eta_i^n} \in \mathcal{X} \\[2mm] P_{RL}^{(1)} & \\ \quad \text{if } \bar{F}_i^n \preceq X_{\eta_i^n} \ (X_{\eta_i^n-1} \in \mathcal{X} \text{ and } X_{\eta_i^n} \notin \mathcal{X}) & \\[2mm] P_{RL}^{(\eta_i^n-K+1)} & \text{if } \bar{F}_i^n \prec X_{\eta_i^n} \ (X_{\eta_i^n} \notin \mathcal{X}) \\[2mm] 1 - P_{RR}^{(\eta_i^n-K)} \cdot P_{RR}^{(1)} & \text{if } \bar{F}_i^n \preceq X_{\eta_i^n} \ (X_{\eta_i^n} \notin \mathcal{X}) \end{cases}$$
$$(10)$$

For a proof of Lemma 1, see Appendix.

The following lemma shows how to compute the probability of the event that $\bar{F}_{\bar{i}}^n$ is received correctly and $\bar{F}_i^n$ is lost and the probability of the event that both $\bar{F}_{\bar{i}}^n$ and $\bar{F}_i^n$ are lost.[7]

Lemma 2: The probabilities $\hat{P}_{RL}^{(i,n)}$ and $\hat{P}_{LL}^{(i,n)}$ $(n \geq 0)$ are given in Tables II and III.

For a proof of Lemma 2, see Appendix.

The following proposition shows how to compute MAD for the intra-coded MB under the Gilbert path model and EC-3.

Proposition 1: Under the Gilbert model and EC-3, the MAD for the intra-coded MB at location s of frame N $(N > 0)$ is given by

$$\text{MAD}(F_s^N, \text{intra}) = \frac{\sum_{j=1}^{256} |f_{sj}^N - E\{\hat{f}_{sj}^N\}|}{256}, \qquad (11)$$

where

$$E\{\hat{f}_{ij}^n\} = \begin{cases} \tilde{P}_R^{(i,n)} \cdot \tilde{f}_{ij}^n + \tilde{P}_L^{(i,n)} \cdot E\{\hat{f}_{ij}^{n-1}\} & \text{if } F_i^n \in \mathcal{G}^n \\[2mm] \tilde{P}_R^{(i,n)} \cdot \tilde{f}_{ij}^n + \hat{P}_{RL}^{(i,n)} \cdot E\{\hat{f}_{ml}^{n-1}\} & \\ \quad + \hat{P}_{LL}^{(i,n)} \cdot E\{\hat{f}_{ij}^{n-1}\} & \text{if } F_i^n \notin \mathcal{G}^n \end{cases}$$
$$(12)$$

where $\tilde{P}_R^{(i,n)}$, $\tilde{P}_L^{(i,n)}$, $\hat{P}_{RL}^{(i,n)}$ and $\hat{P}_{LL}^{(i,n)}$ are given by Eqs. (9) and (10), Tables II and III, respectively.

Proof: From Eq. (4), we have

$$\text{MAD}(F_s^N, \text{intra}) = \frac{E\left\{\sum_{j=1}^{256} |f_{sj}^N - \hat{f}_{sj}^N|\right\}}{256}. \qquad (13)$$

Due to the random nature of the path characteristics, $\hat{f}_{sj}^N$ at the receiver is a random variable while f_{sj}^N is not a random variable. Thus, Eq. (13) becomes Eq. (11).

To see that Eq. (12) holds, we consider the following two cases based on the position of F_i^n.

- *Case 1:* Suppose that F_i^n does not have $F_{\bar{i}}^n$,[8] i.e., $F_i^n \in \mathcal{G}^n$. According to the packet loss behavior, there are two cases for the random variable $\hat{f}_{ij}^n$ as follows.

  − *Subcase 1.1:* Suppose that the packet containing F_i^n is received correctly. Then we have $\hat{f}_{ij}^n = \tilde{f}_{ij}^n$, and the probability of this event is $\tilde{P}_R^{(i,n)}$.

  − *Subcase 1.2:* Suppose that the packet containing F_i^n is lost. Since the MB above F_i^n does not exist, the estimated motion vector of the corrupted F_i^n is set to zero. Thus, the corrupted pixel j in F_i^n is concealed by pixel j in F_i^{n-1} at the receiver. Then we have $\hat{f}_{ij}^n = \hat{f}_{ij}^{n-1}$, and the probability of this event is $\tilde{P}_L^{(i,n)}$. Since $\hat{f}_{ij}^{n-1}$ is also a random variable, from Fact 2, the expectation of $\hat{f}_{ij}^{n-1}$ should be chosen in computing the expectation of $\hat{f}_{ij}^n$. So we have $\hat{f}_{ij}^n = E\{\hat{f}_{ij}^{n-1}\}$.

Based on the analysis of Subcases 1.1 and 1.2, the expectation of $\hat{f}_{ij}^n$ can be given by

$$E\{\hat{f}_{ij}^n\} = \tilde{P}_R^{(i,n)} \cdot \tilde{f}_{ij}^n + \tilde{P}_L^{(i,n)} \cdot E\{\hat{f}_{ij}^{n-1}\}. \qquad (14)$$

- *Case 2:* Suppose that F_i^n has $F_{\bar{i}}^n$,[9] i.e., $F_i^n \notin \mathcal{G}^n$. According to the packet loss behavior, there are three cases for the random variable $\hat{f}_{ij}^n$ as follows.

  − *Subcase 2.1:* Suppose that the packet containing F_i^n is received correctly. Then we have $\hat{f}_{ij}^n = \tilde{f}_{ij}^n$, and the probability of this event is $\tilde{P}_R^{(i,n)}$.

  − *Subcase 2.2:* Suppose that the packet containing F_i^n is lost and the packet containing the MB above ($\bar{F}_{\bar{i}}^n$) has been received correctly. Then due to error concealment EC-3, the corrupted pixel j in F_i^n is concealed by the pixel in frame $n - 1$, which is pointed by the motion vector of $\bar{F}_{\bar{i}}^n$. Since this pixel is denoted by pixel l in macroblock m of frame $n - 1$,[10] then we have $\hat{f}_{ij}^n = \hat{f}_{ml}^{n-1}$, and the probability of this event is $\hat{P}_{RL}^{(i,n)}$. Since $\hat{f}_{ml}^{n-1}$ is also a random variable, from Fact 2, the expectation of $\hat{f}_{ml}^{n-1}$ should be chosen in computing the expectation of $\hat{f}_{ij}^n$. So we have $\hat{f}_{ij}^n = E[\hat{f}_{ml}^{n-1}]$.

  − *Subcase 2.3:* Suppose that the packet containing F_i^n and the packet containing $\bar{F}_{\bar{i}}^n$ are all lost. Since the MB above F_i^n is lost, the estimated motion vector of the corrupted F_i^n is set to zero. Then the corrupted pixel j in F_i^n is concealed by pixel j in F_i^{n-1}. Thus, we have $\hat{f}_{ij}^n = \hat{f}_{ij}^{n-1}$, and the probability of this event is $\hat{P}_{LL}^{(i,n)}$. Since $\hat{f}_{ij}^{n-1}$ is also a random variable, from Fact 2, the expectation of $\hat{f}_{ij}^{n-1}$ should be chosen in computing the expectation of $\hat{f}_{ij}^n$. So we have $\hat{f}_{ij}^n = E[\hat{f}_{ij}^{n-1}]$.

---

[7] Due to the complication of the formulae, we list them in the tables.

[8] For example, F_i^n is located at the top of the frame.

[9] For example, F_i^n is not located at the top of the frame.

[10] Refer to Table I.

TABLE II

THE PROBABILITY $\hat{P}_{RL}^{(i,n)}$.

Cases	$\hat{P}_{RL}^{(i,n)}$
$X_{\eta_i^n} \in \mathcal{X}$ or $F_i^n \prec X_{\eta_i^n}$ ($X_{\eta_i^n} \notin \mathcal{X}$ and $X_{\eta_i^n} = X_{\eta_{\bar{i}}^n}$)	0
$\bar{F}_i^n \preceq X_{\eta_i^n}$ ($X_{\eta_i^n - 1} \in \mathcal{X}$ and $X_{\eta_i^n} \notin \mathcal{X}$)	$P_{RL}^{(1)}$
$\bar{F}_i^n \preceq X_{\eta_i^n}$ and $\bar{F}_i^n \prec X_{\eta_i^n - 1}$ ($X_{\eta_i^n - 1} \notin \mathcal{X}$)	$P_{RR}^{(\eta_i^n - K)} \cdot P_{RL}^{(1)}$
$\bar{F}_i^n \preceq X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n \preceq X_{\eta_i^n - 1}$ ($X_{\eta_i^n - 1} \notin \mathcal{X}$)	$P_{RR}^{(\eta_i^n - K - 1)} \cdot P_{RR}^{(1)} \cdot P_{RL}^{(1)}$
$\bar{F}_i^n \prec X_{\eta_i^n}$ ($X_{\eta_i^n} \notin \mathcal{X}$ and $X_{\eta_i^n} \neq X_{\eta_{\bar{i}}^n}$)	$\tilde{P}_R^{(i,n)} \cdot P_{RL}^{(\eta_i^n - \eta_{\bar{i}}^n)}$
$\bar{F}_i^n \preceq X_{\eta_i^n}$ ($X_{\eta_i^n} \notin \mathcal{X}$ and $X_{\eta_i^n - 1} \neq X_{\eta_{\bar{i}}^n}$)	$\tilde{P}_R^{(\bar{i},n)} \cdot \left(1 - P_{RL}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)}\right)$

TABLE III

THE PROBABILITY $\hat{P}_{LL}^{(i,n)}$.

Cases	$\hat{P}_{LL}^{(i,n)}$
$X_{\eta_i^n} \in \mathcal{X}$ or $F_i^n \preceq X_{\eta_i^n}$ ($X_{\eta_i^n - 1} \in \mathcal{X}$ and $X_{\eta_i^n} \notin \mathcal{X}$)	0
$\bar{F}_i^n \prec X_{\eta_i^n}$ ($X_{\eta_i^n} \notin \mathcal{X}$ and $X_{\eta_i^n} = X_{\eta_{\bar{i}}^n}$)	$P_{RL}^{(\eta_i^n - K + 1)}$
$\bar{F}_i^n \preceq X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n \prec X_{\eta_i^n - 1}$ ($X_{\eta_i^n - 1} \notin \mathcal{X}$)	$P_{RL}^{(\eta_i^n - K)}$
$\bar{F}_i^n \preceq X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n \preceq X_{\eta_i^n - 1}$ ($X_{\eta_i^n - 1} \notin \mathcal{X}$)	$P_{RL}^{(\eta_i^n - K)} + P_{RL}^{(\eta_i^n - K - 1)} \cdot P_{LR}^{(1)} \cdot P_{RL}^{(1)}$
$\bar{F}_i^n \prec X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n \prec X_{\eta_i^n}$ ($X_{\eta_i^n} \notin \mathcal{X}$ and $X_{\eta_i^n} \neq X_{\eta_{\bar{i}}^n}$)	$P_{RL}^{(\eta_i^n - K + 1)} \cdot P_{LL}^{(\eta_i^n - \eta_{\bar{i}}^n)}$
$\bar{F}_i^n \prec X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n \preceq X_{\eta_i^n}$ ($X_{\eta_i^n} \notin \mathcal{X}$ and $X_{\eta_i^n} \neq X_{\eta_{\bar{i}}^n}$)	$P_{RL}^{(\eta_i^n - K)} \cdot P_{LL}^{(1)} \cdot P_{LL}^{(\eta_i^n - \eta_{\bar{i}}^n)}$ $+ P_{RL}^{(\eta_i^n - K)} \cdot P_{LR}^{(1)} \cdot P_{RL}^{(\eta_i^n - \eta_{\bar{i}}^n)}$ $+ P_{RR}^{(\eta_i^n - K)} \cdot P_{RL}^{(1)} \cdot P_{LL}^{(\eta_i^n - \eta_{\bar{i}}^n)}$
$\bar{F}_i^n \preceq X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n \prec X_{\eta_i^n}$ ($X_{\eta_i^n} \notin \mathcal{X}$ and $X_{\eta_i^n - 1} \neq X_{\eta_{\bar{i}}^n}$)	$P_{RL}^{(\eta_i^n - K + 1)} \cdot \left(1 - P_{LR}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)}\right)$
$\bar{F}_i^n \preceq X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n \preceq X_{\eta_i^n}$ ($X_{\eta_i^n} \notin \mathcal{X}$ and $X_{\eta_i^n - 1} \neq X_{\eta_{\bar{i}}^n}$)	$P_{RL}^{(\eta_i^n - K)} \cdot P_{LL}^{(1)} \cdot \left(1 - P_{LR}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)}\right)$ $+ P_{RL}^{(\eta_i^n - K)} \cdot P_{LR}^{(1)} \cdot \left(1 - P_{RR}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)}\right)$ $+ P_{RR}^{(\eta_i^n - K)} \cdot P_{RL}^{(1)} \cdot \left(1 - P_{LR}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)}\right)$

Based on the analysis of Subcases 2.1, 2.2 and 2.3, the expectation of $\hat{f}_{ij}^n$ can be given by

$$E[\hat{f}_{ij}^n] = \tilde{P}_R^{(i,n)} \cdot \tilde{f}_{ij}^n + \hat{P}_{RL}^{(i,n)} \cdot E\{\hat{f}_{ml}^{n-1}\} + \hat{P}_{LL}^{(i,n)} \cdot E\{\hat{f}_{ij}^{n-1}\} \tag{15}$$

Combining Cases 1 and 2 above, we complete the proof of Eq. (12). ■

We would like to stress that Proposition 1 holds for any packetization scheme.

Under PKT-2 or PKT-3, since no MB is split, Eqs. (9) and (10) can be simplified by

$$\tilde{P}_R^{(i,n)} = \begin{cases} 1 & \text{if } n = 0 \\ P_{RR}^{(\eta_i^n - K + 1)} & \text{if } n > 0 \end{cases}, \tag{16}$$

and

$$\tilde{P}_L^{(i,n)} = \begin{cases} 0 & \text{if } n = 0 \\ P_{RL}^{(\eta_i^n - K + 1)} & \text{if } n > 0 \end{cases}, \tag{17}$$

respectively. In addition, Tables II and III can be simplified by

$$\hat{P}_{RL}^{(i,n)} = \begin{cases} 0 & \text{if } n = 0 \\ P_{RR}^{(\eta_i^n - K + 1)} \cdot P_{RL}^{(\eta_i^n - \eta_{\bar{i}}^n)} & \text{if } n > 0 \end{cases}, \tag{18}$$

and

$$\hat{P}_{LL}^{(i,n)} = \begin{cases} 0 & \text{if } n = 0 \\ P_{RL}^{(\eta_i^n - K + 1)} \cdot P_{LL}^{(\eta_i^n - \eta_{\bar{i}}^n)} & \text{if } n > 0 \end{cases}, \tag{19}$$

respectively.

Under EC-1/2 and PKT-2/3, since it works just like the case where no motion vector is available under EC-3, Eq. (12) can be simplified by

$$E\{\hat{f}_{sj}^n\} = \tilde{P}_R^{(s,n)} \cdot \tilde{f}_{sj}^n + \tilde{P}_L^{(s,n)} \cdot E\{\hat{f}_{sj}^{n-1}\} \tag{20}$$

Thus, EC-1 and EC-2 are special cases of EC-3.

Remark 2: Under PKT-2 or PKT-3, from Eqs. (16) to (19), we have $\tilde{P}_R^{(i,n)} = 1$, $\tilde{P}_L^{(i,n)} = 0$, $\hat{P}_{RL}^{(i,n)} = 0$, and $\hat{P}_{LL}^{(i,n)} = 0$, if $n = 0$. Thus, from Eq. (12), we have

$$E\{\hat{f}_{ij}^0\} = \tilde{f}_{ij}^0 \qquad (i \in [0, N_h] \text{ and } j \in [1, 256]). \tag{21}$$

That is, $E\{\hat{f}_{ij}^0\}$ is initialized with $\tilde{f}_{ij}^0$, which is known at the encoder.[11] Then, $E\{\hat{f}_{ij}^n\}$ (for frame $n > 0$) can be obtained through the iteration in Eq. (12).

Inter Mode

The following proposition shows how to compute MAD for the inter-coded MB under the Gilbert path model and EC-3.

Proposition 2: Under the Gilbert model and EC-3, the MAD for the inter-coded MB at location s of frame N ($N > 0$) is given by

$$\text{MAD}(F_s^N,\text{inter}) = \frac{\sum_{j=1}^{256} |f_{sj}^N - E\{\hat{f}_{sj}^N\}|}{256}, \qquad (22)$$

where

$$E\{\hat{f}_{ij}^n\} = \begin{cases} \tilde{P}_R^{(i,n)} \cdot (\tilde{e}_{ij}^n + E\{\hat{f}_{uv}^{n-1}\}) + \tilde{P}_L^{(i,n)} \cdot E\{\hat{f}_{ij}^{n-1}\} \\ \qquad\qquad\qquad\qquad\qquad \text{if } F_i^n \in \mathcal{G}^n \\[2mm] \tilde{P}_R^{(i,n)} \cdot (\tilde{e}_{ij}^n + E\{\hat{f}_{uv}^{n-1}\}) + \hat{P}_{RL}^{(i,n)} \cdot E\{\hat{f}_{ml}^{n-1}\} \\ \quad + \hat{P}_{LL}^{(i,n)} \cdot E\{\hat{f}_{ij}^{n-1}\} \\ \qquad\qquad\qquad\qquad\qquad \text{if } F_i^n \notin \mathcal{G}^n \end{cases}$$
$$(23)$$

where $\tilde{P}_R^{(i,n)}$, $\tilde{P}_L^{(i,n)}$, $\hat{P}_{RL}^{(i,n)}$ and $\hat{P}_{LL}^{(i,n)}$ are given by Eqs. (9) and (10), Tables II and III, respectively.

Proof: Similar to the derivation of Eq. (11), we have Eq. (22). To see that Eq. (23) holds, we can also use the same way as that to derive Eq. (12). The only difference is the case where the packet containing F_i^n is received correctly. In this case, under the inter code, we have $\hat{f}_{ij}^n = \tilde{e}_{ij}^n + \hat{f}_{uv}^{n-1}$, which is Eq. (8). ∎

We would like to stress that Proposition 2 holds for any packetization scheme.

Under PKT-2 or PKT-3, since no MB is split, Eqs. (9) and (10) can be simplified by Eqs. (16) and (17), respectively. In addition, Tables II and III can be simplified by Eqs. (18) and (19), respectively.

Under EC-1/2 and PKT-2/3, since it works just like the case where no motion vector is available under EC-3, Eq. (23) can be simplified by

$$E\{\hat{f}_{sj}^n\} = \tilde{P}_R^{(i,n)} \cdot (\tilde{e}_{ij}^n + E\{\hat{f}_{uv}^{n-1}\}) + \tilde{P}_L^{(i,n)} \cdot E\{\hat{f}_{ij}^{n-1}\}. \quad (24)$$

Although the global distortion metrics derived in Section III-A only apply to the Gilbert path model, the methodology we employ (i.e., the end-to-end approach) is general and can be applied to any path model (e.g., self-similar path model).

B. *Optimal Mode Selection Based on Global Distortion Metrics*

Given the packetization scheme used by the source, the path characteristics and the error concealment scheme used by the decoder, we design a globally R-D optimized mode selection algorithm.

---

[11] This is why we assume the first I-frame of the video stream has been successfully received.

Consider a GOB denoted by $\mathcal{F}_g^n = (F_g^n, \cdots, F_{g+N_G-1}^n)$, where N_G is the number of MBs in a GOB. Assume each MB in $\mathcal{F}_g^n$ can be coded using only one of the two modes in set $\mathcal{I}$. Then for a given GOB, the modes assigned to the MBs in $\mathcal{F}_g^n$ are given by the N_G-tuple, $\mathcal{M}_g^n = (M_g^n, \cdots, M_{g+N_G-1}^n) \in \mathcal{I}^{N_G}$. The problem of globally R-D optimized mode selection is to find the combination of modes that minimizes the distortion for a given GOB, subject to a constraint R_c on the number of bits used. This constrained problem can be formulated as

$$\min_{\mathcal{M}_g^n} D(\mathcal{F}_g^n, \mathcal{M}_g^n) \quad \text{subject to} \quad R(\mathcal{F}_g^n, \mathcal{M}_g^n) \leq R_c, \quad (25)$$

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 $\mathcal{F}_g^n$ with a particular mode combination $\mathcal{M}_g^n$.

The constrained minimization problem in (25) can be converted to an unconstrained minimization problem by Lagrange multiplier technique. Under the assumption of an additive distortion measure, the Lagrangian cost function can be given by

$$\begin{aligned} J(\mathcal{F}_g^n, \mathcal{M}_g^n) &= \sum_{i=g}^{g+N_G-1} J(F_i^n, \mathcal{M}_g^n) \\ &= \sum_{i=g}^{g+N_G-1} [D(F_i^n, \mathcal{M}_g^n) + \lambda R(F_i^n, \mathcal{M}_g^n)]. \end{aligned}$$
$$(26)$$

Thus, the objective function becomes

$$\min_{\mathcal{M}_g^n} \left\{ \sum_{i=g}^{g+N_G-1} J(F_i^n, \mathcal{M}_g^n) \right\}. \qquad (27)$$

If both the rate and distortion for macroblock F_i^n are not affected by other mode that is not used by macroblock F_i^n, a simplified Lagrangian can be given by

$$J(F_i^n, \mathcal{M}_g^n) = J(F_i^n, M_i^n). \qquad (28)$$

Thus, the optimization problem of (27) reduces to

$$\sum_{i=g}^{g+N_G-1} \min_{M_i^n} J(F_i^n, M_i^n) = $$
$$\sum_{i=g}^{g+N_G-1} \min_{M_i^n} \{D(F_i^n, M_i^n) + \lambda R(F_i^n, M_i^n)\}, \quad (29)$$

where the global distortion $D(F_i^n, M_i^n)$ can be expressed by the formulae we derived in Section III-A, according to the coding mode, the packetization scheme used by the source and the error concealment scheme used by the decoder.

The problem of (29) is a standard R-D optimization problem and can be solved by the approaches described in [8], [11], [14], [22]. Different from these approaches, we use a simpler method to obtain λ.

Since a large λ in the optimization problem of (29) can reduce the bit-count of the coded frame, we employ this nature in choosing λ. To be specific, at the end of frame n, we adjust λ for frame $n + 1$ (i.e., λ_{n+1}) as follows:

$$\lambda_{n+1} = \frac{2 \cdot B_n + (\gamma - B_n)}{B_n + 2 \cdot (\gamma - B_n)} \cdot \lambda_n \qquad (30)$$
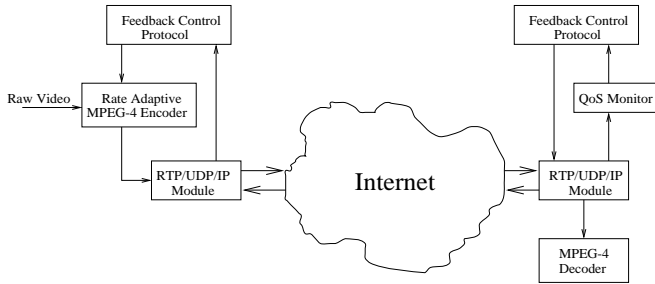
Fig. 6. An end-to-end architecture for MPEG-4 video conferencing.

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. (30) is to keep the buffer occupancy at the middle level to reduce the chance of buffer overflow or underflow. In other word, Eq. (30) also achieves the objective of rate control.

Sections II and III complete the theoretical part of our work. To evaluate the effectiveness of our approach, we integrate our theory with a specific system – an architecture for point-to-point MPEG-4 video conferencing over the Internet.

IV. AN APPLICATION FOR MPEG-4 VIDEO CONFERENCING

In this section, we present an end-to-end architecture for point-to-point MPEG-4 video conferencing over the Internet.

A. Architecture of MPEG-4 Video Conferencing

Figure 6 shows our end-to-end architecture for point-to-point MPEG-4 video conferencing over the Internet. We use the MPEG-4 rate control algorithm described in [2], [23] to control the output rate to be constant. In this paper, we set the rate fixed to investigate the error resilient capability of our architecture and algorithm.

In Fig. 6, on the sender side, raw bit-stream of live video is encoded by an MPEG-4 encoder. After this stage, the compressed video bit-stream is first packetized at the sync layer and then passed through the RTP/UDP/IP layers before entering the Internet. Packets may be dropped at a router/switch (due to congestion) or at the destination (due to excess delay). For packets that are successfully delivered to the destination, they first pass through the RTP/UDP/IP layers in reverse order before being decoded at the MPEG-4 decoder.

On the receiver side, a QoS monitor detects the packet loss through RTP sequence number and estimates the transition probabilities (e.g., p and q in the Gilbert model). When the receiver sends out a feedback RTCP packet to the source, the estimated parameters p and q are carried in the feedback packet. Once source receives such feedback, it encodes the video based on the parameters p and q through the proposed R-D optimized mode selection in Section IV-D and rate control algorithm in [2], [23].

B. Transport and Feedback Control Protocols

Since TCP retransmission introduces delays that are not acceptable for MPEG-4 video conferencing, we employ UDP as the transport protocol for MPEG-4 video streams. Because UDP does not guarantee packet delivery, the receiver needs to rely on upper layer (i.e., RTP/RTCP) to detect packet loss.

Real-Time Transport Protocol is an Internet standard protocol designed to provide end-to-end transport functions for transmitting real-time applications [12]. RTP Control Protocol (RTCP) is a companion protocol designed to provide QoS feedback to the participants of an RTP session. In order words, RTP is a data transfer protocol while RTCP is a control protocol.

RTCP provides QoS feedback through the use of *Sender Reports (SR)* and *Receiver Reports (RR)* at the source and destination, respectively. In particular, RTCP keeps the total control packets to 5% of the total session bandwidth. Among the control packets, 25% are allocated to the sender reports and 75% to the receiver reports. To prevent control packet starvation, at least one control packet is sent within five seconds at the sender or receiver. The feedback control protocol employs RTCP to convey QoS information so that QoS information can be utilized by the encoder.

RTP does not guarantee QoS or reliable delivery, but rather, provides some basic functionalities (e.g., packet sequence number). As a result, packet loss can be detected by the QoS monitor through examining the RTP packet sequence number at the receiver side. In particular, a packet that arrives after the maximum delay threshold is considered lost.

We employ the Gilbert path model since measurements of packet loss in the Internet has shown that the bursty packet loss behavior in the Internet can be modeled reasonably well with the Gilbert model [1], [5]. The path characteristics can be estimated by the QoS monitor as follows. On obtaining the packet loss information from RTP/UDP/IP module, the QoS monitor measures N_1, N_2, N_3, and N_4, and estimates the transition probabilities, p and q, through Eq. (7).

The period for estimating p and q is set to five seconds. That is, N_1, N_2, N_3, and N_4 are measured during the five-second period. At the end of each period, p and q are obtained through (7) and then transferred by an RTCP packets. Since N_1, N_2, N_3, and N_4 are reset to zero at the end of each period, the estimated p and q reflect the current network congestion status.

C. Packetization and Error Concealment

We use PKT-3 rather than PKT-2 to achieve efficiency for Internet video conferencing. In addition, when a packet is lost, we employ EC-3 to conceal the region associated with the lost packet. To be specific, each corrupted MB will be replaced with the MB in the previous frame pointed by an estimated motion vector. The estimated motion vector of the corrupted MB is copied from the MB above it when available, otherwise the motion vector is set to zero. Note

TABLE IV

SIMULATION PARAMETERS.

End system	MPEG-4	MaxPL	526 bytes
		Rate	100 Kbps
		Frame rate	10 frames/s
		I-VOP refreshment period	50 frames
		λ_0	1
		Buffer size	1 Mbytes
	TCP	Mean packet processing delay	300 μs
		Packet processing delay variation	10 μs
		Packet size	576 bytes
		Maximum receiver window size	64K bytes
		Default timeout	500 ms
		Timer granularity	500 ms
		TCP version	Reno
	UDP	$E(T_{on})$	100 ms
		$E(T_{off})$	150 ms
		r_p	100 Kbps
		Packet size	576 bytes
Switch		Buffer size	10 Kbytes
		Packet processing delay	4 μs
Link	End system to switch	Link speed	10 Mbps
		Distance	1 km
	Switch to switch	Distance	1000 km

that a more sophisticated error concealment scheme than EC-3 may achieve better performance than that of EC-3.

D. Feedback-Based Optimal Mode Selection

For implementation purpose, our end-to-end approach also considers the impact of feedback mechanism on the video quality (in terms of global distortion). The rationale is as follows. Global optimality is not achievable without feedback since the source could not select an optimal mode without knowledge of the path characteristics and receiver behavior. In addition, the congestion status of the Internet is dynamically changing. Assigning the path characteristics (e.g., p and q) with fixed numbers may either lose compression efficiency when the network is less congested than expected, or suffer from insufficiency of error resilience when the network becomes more congested than expected. Therefore, it is not valid in the real Internet to assume that the path characteristics is known *a priori* and is fixed. From our experiments and simulations, we observe that the percentage of intra-coded macroblocks should increase as the packet loss ratio increases in order to improve the capability of error resilience. Thus, MPEG-4 video coding should adapt to the changing Internet environment, i.e., network congestion.

This motivates us to employ a feedback mechanism to convey such information to the encoder as the path characteristics (i.e., p and q) estimated at the receiver and the error concealment scheme used by the decoder. The type of error concealment scheme used by the receiver can be transmitted at the set-up period of the session.

Base on the theory in Section III, we design a globally R-D optimized mode selection algorithm for MPEG-4. Since we employ PKT-3, EC-3 and the Gilbert path model, the global distortion $D(F_i^n, intra)$ in (29) can be expressed by Eqs. (11), (12), (16), (17), (18), and (19); the global distortion $D(F_i^n, inter)$ in (29) can be expressed by Eqs. (22), (23), (16), (17), (18), and (19). Based on the feedback p and q, the proposed algorithm will choose a mode which best trades off compression efficiency and error resilience.

V. SIMULATION RESULTS

In this section, we implement the end-to-end architecture described in Section IV on our network simulator and perform a simulation study of video conferencing with MPEG-4. The purpose of this section is to demonstrate the performance improvement of our end-to-end approach over the classical approach under various simulation settings.

A. Simulation Settings

The network configurations that we use are the *peer-to-peer* (Fig. 7), and the *chain* (Fig. 9) network configurations. These network configurations have been used as standard test configurations in networking research community.

We implement the architecture depicted in Fig. 6. 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 [2], [23] to keep a constant rate at 100 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.
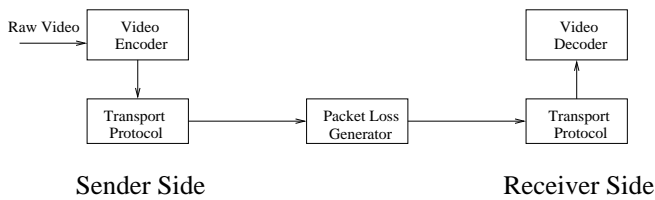
Fig. 7. A peer-to-peer network.

The encoded bit-stream is packetized with PKT-3 scheme (i.e., a packet corresponds to a GOB). Additional overhead from RTP/UDP/IP is also added to the packet before it is sent to the network. We use 576 bytes for the path MTU. Therefore, the maximum payload length, MaxPL, for MPEG-4 is 526 bytes (576 bytes minus 50 bytes of overhead) [13]. Packets may be dropped due to congestion in the network. For arriving packets, the receiver extracts the packet content to form the bit-stream for the decoder.

In addition to MPEG-4 video, we also use TCP/UDP connections to simulate the background interfering traffic. All TCP sources are assumed to be persistent during the simulation run. For UDP connections, we use an exponentially distributed on/off model with average $E(T_{on})$ and $E(T_{off})$ for on and off periods, respectively. During each on period, the packets are generated at peak rate r_p. The average bit rate for a UDP connection is, therefore, $r_p \cdot \frac{E(T_{on})}{E(T_{on})+E(T_{off})}$.

Table IV lists the parameters used in our simulation.

Under such simulation settings, 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 IV-D. However, feedback is not employed.

Encoder C: implements the globally R-D optimized mode selection described in Section IV-D. Feedback mechanism is used.

B. Performance Under the Peer-to-Peer Configuration

A peer-to-peer network configuration is shown in Fig. 7. We emphasize that such simple network configuration captures the fundamental property of a transport path within the Internet cloud (Fig. 6) since there is only one bottleneck link (i.e., the one with minimum bandwidth among all the traversing links) between the sender and the receiver. Since the network status under this configuration is stable, Encoder B and Encoder C are identical. Thus, we only examine Encoder A and Encoder B rather than Encoder C under this configuration, i.e., feedback mechanism is not necessary here.

We run our simulations under two scenarios. The first scenario is loss-free environment. It can be seen from Eq. (2) that D_r takes the value of D_q when there is no loss, thereby making the global distortion equal to quantization distortion. Thus, under the loss-free environment, Encoder B works just in the same way as Encoder A. As
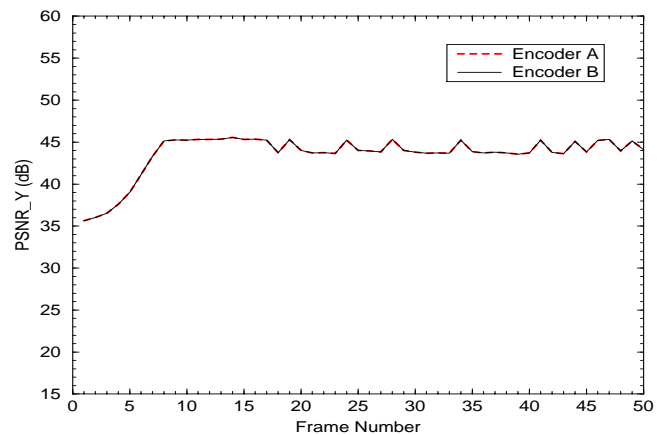
(a)
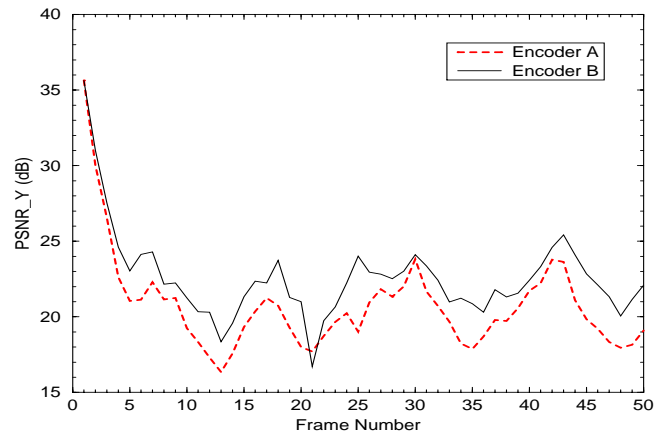
(b)

Fig. 8. PSNR at the receiver under (a) the loss-free peer-to-peer network and (b) the lossy peer-to-peer network.

a result, Encoder B has the same performance (in terms of PSNR of Y component of the video at the receiver) as that of Encoder A, which is demonstrated in Fig. 8(a). The average PSNRs are all 43.6 dB.

The second scenario of our simulation under the peer-to-peer configuration is lossy environment, where the loss pattern is characterized by the Gilbert path model with $p = 0.76$ and $q = 0.08$.[12] The resulting average packet loss ratio is 9.5%.

The performances under the second scenario is shown in Fig. 8(b). The average PSNRs correspondent to Fig. 8(b) are 20.7 dB for Encoder A and 22.6 dB for Encoder B, respectively. It can be seen that our approach (i.e., Encoder B or C) is able to improve the quality in terms of PSNR, compared with the classical approach (i.e., Encoder A).

C. Performance Under the Chain Configuration

This is a benchmark network configuration commonly used to examine traffic behavior under the impact of other traversing interfering traffic. The specific chain configura-

---

[12]We set $p = 0.76$ and $q = 0.08$ to match the real measurements collected between INRIA in France and University College London in the UK [9].
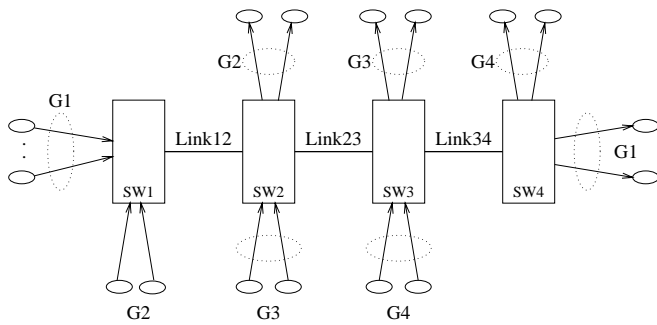
Fig. 9. A chain network.

tion that we use is shown in Fig. 9 where path G1 consisting of multiple flows and traverses from the first switch (SW1) to the last switch (SW4), while all the other paths traverse only one hop and "interfere" the flows in G1.

In our simulations, G1 consists of one MPEG-4 source, three TCP connections and three UDP connections while G2, G3 and G4 all consist of three TCP connections and three UDP connections, respectively. The link capacities on Link12, Link23, and Link34 are identical. We run simulations under two cases.

Case (a): The link capacity is 250 Kbps to simulate a network with high loss probability.

Case (b): The link capacity is 350 Kbps to simulate a network with low loss probability.

We run our simulation under the two cases for 100 seconds. Since there are only 150 continuous frames in "Miss America" sequence available, we repeated the video sequence cyclically during the simulation run.

Figure 10 shows the PSNR of the first 50 frames in the 100-second simulations. It can be observed that under both cases, 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).

Figure 11 shows the PSNR at the receiver during the 100-second simulations. The packet loss ratios corresponding to Fig. 11(a) are all 19.8%; the packet loss ratios corresponding to Fig. 11(b) are all 3.2%.

The average PSNRs corresponding to Fig. 11(a) from top to bottom are 19.0 dB, 19.7 dB, and 20.8 dB, respectively. The average PSNRs corresponding to Fig. 11(b) from top to bottom are 26.9 dB, 28.9 dB, and 30.4 dB, respectively. In Fig. 11, it can be seen that there is a period of 5 seconds starting by a peak in PSNR. This is due to intra-VOP refreshment period of 50 frames (5 second).

To examine the perceptual quality of the received video streams, we play out the decoded video sequences at the receiver. Figure 12(a) shows a sample video frame in loss-free environment; Figure 12(b) shows the same video frame under Encoder A when Case B (350 Kbps) is used; Figure 12(c) shows the same video frame under Encoder C when Case B (350 Kbps) is used. Figure 12 demonstrates
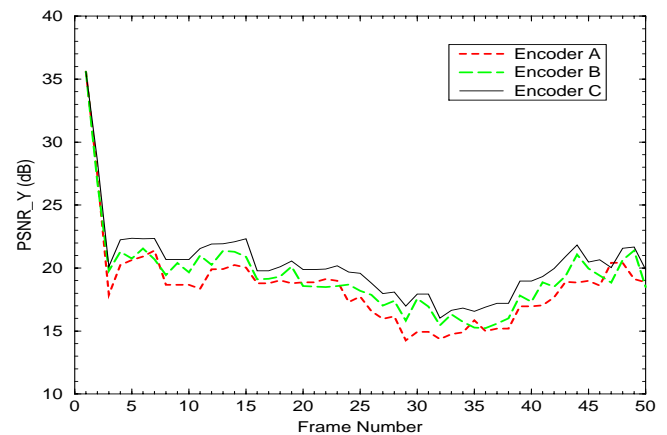
(a)
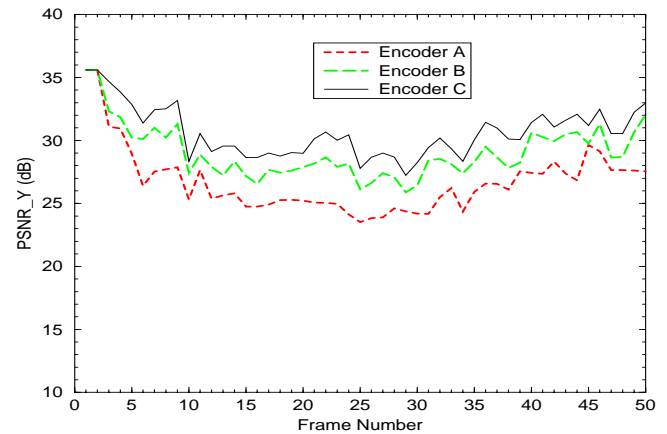
(b)

Fig. 10. PSNR at the receiver under the chain network (a) with link capacity 250 Kbps; and (b) with link capacity 350 Kbps.

that our approach (i.e., Encoder C) achieves better perceptual quality than the classical approach (i.e., Encoder A).

We summarize the packet loss ratios (PLR) and average PSNRs of all simulations in Table V. We have the following observations.

• Under loss-free peer-to-peer network, since the PLR under Encoder A and B are all zero, the PSNRs for Encoder A and B are, therefore, the same as shown in Fig. 8(a).

• Under lossy peer-to-peer network, the average PSNR for Encoder B is higher than that for Encoder A, which demonstrates the error resilience of our approach.

• Under the chain network with link capacity 250 Kbps (high loss probability), Encoder C has the highest average PSNR; Encoder A has the lowest average PSNR; the average PSNR for Encoder B is higher than that for Encoder A.

• Under the chain network with link capacity 350 Kbps (low loss probability), Encoder C has the highest average PSNR; Encoder A has the lowest average PSNR; the average PSNR for Encoder B is higher than that for Encoder A.

In summary, our simulation results demonstrate that: (1) our approach can achieve better quality over the classi-

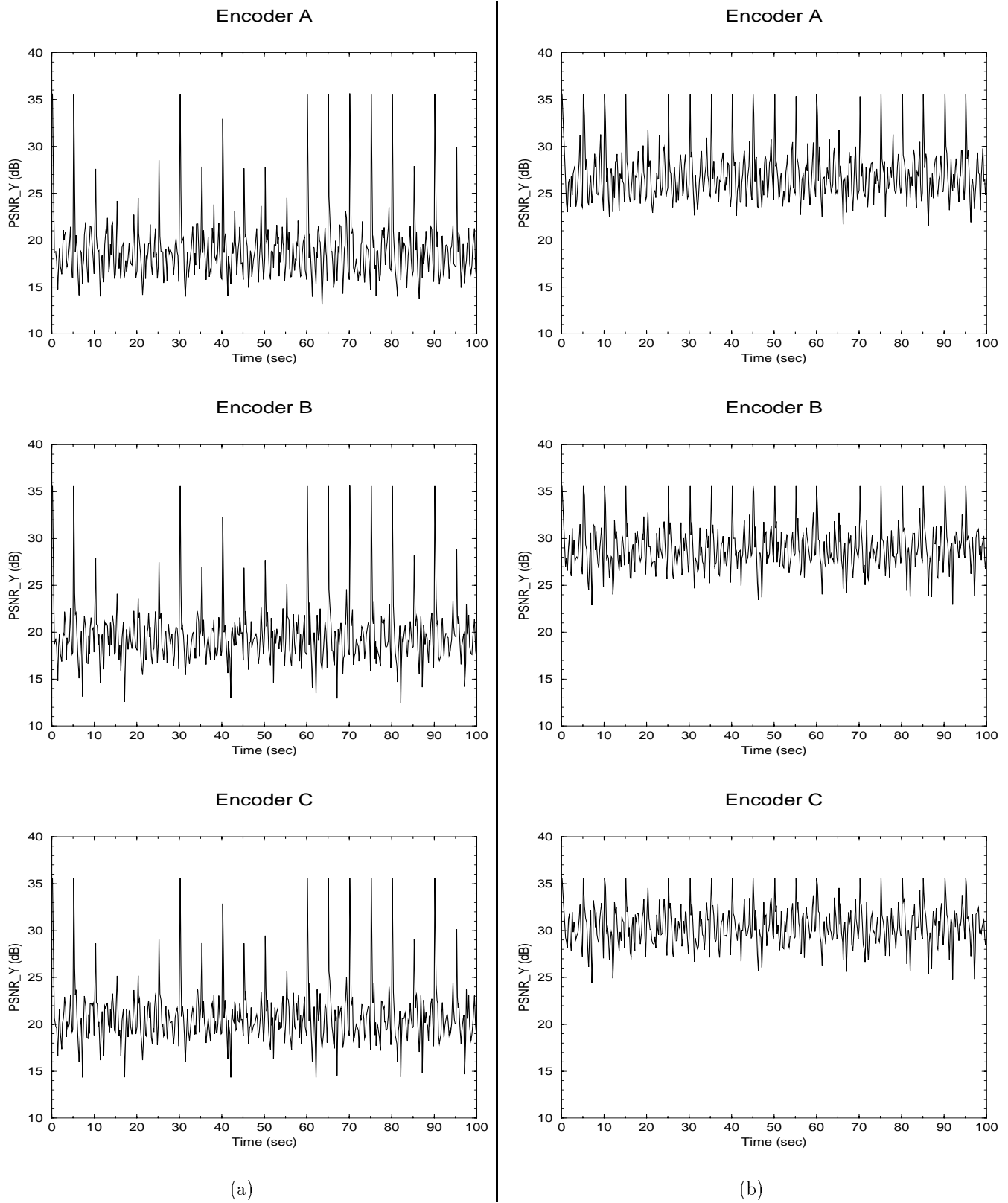(a)   (b)

Fig. 11.   PSNR at the receiver under the chain network with link capacity (a) 250Kbps; and (b) 350 Kbps.
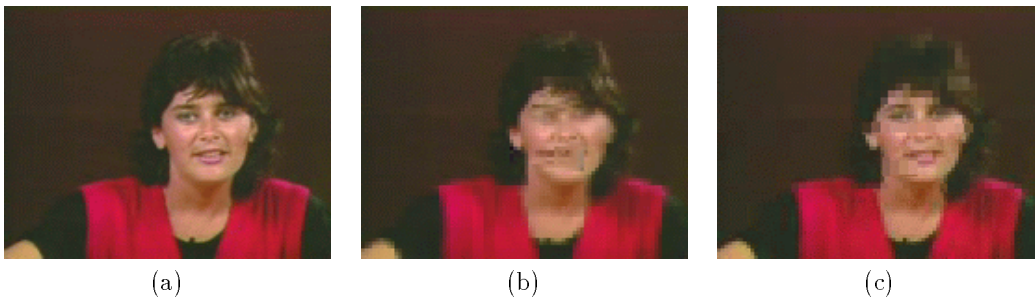
Fig. 12. (a) Sample frame under loss-free environment; (b) Sample frame for Encoder A under lossy environment; (c) Sample frame for Encoder C under lossy environment.

TABLE V

PACKET LOSS RATIO (PLR) AND AVERAGE PSNR UNDER DIFFERENT NETWORK CONFIGURATIONS.

Network configuration	PLR (%)	Average PSNR (dB)		
		Encoder A	Encoder B	Encoder C
Peer-to-peer (loss-free)	0	43.6	43.6	43.6
Peer-to-peer (lossy)	9.5	20.7	22.6	22.6
Chain (250 Kbps)	19.8	19.0	19.7	20.8
Chain (350 Kbps)	3.2	26.9	28.9	30.4

cal approach under packet loss network environment; and (2) the feedback mechanism can help to further improve presentation quality.

VI. CONCLUDING REMARKS

In this paper, we investigated the fundamental problem of R-D optimized mode selection for point-to-point Internet video communication from an end-to-end perspective, which includes source behavior, path characteristics, and receiver behavior. The main contributions of this paper are listed as follows.

• By introducing the notion of global distortion, we formulated the R-D optimized mode selection problem, which generalized the classical approach. We derived, for the first time, a set of accurate global distortion metrics for any packetization scheme. We showed how to apply the global distortion metrics to a specific packetization scheme.

• We showed how to use the global distortion metrics to design the globally R-D optimized mode selection algorithm. The proposed algorithm offers the best trade-off between compression efficiency and error resilience. Our theory is general and is applicable to many video coding standards, including H.261/263 and MPEG-1/2/4.

• We applied our theory to MPEG-4 video conferencing over the Internet. A feedback mechanism was employed to convey path characteristics and receiver behavior so that the source encoder can achieve global optimality in the mode selection. Simulation results conclusively demonstrated that our end-to-end approach offers superior performance over the classical approach for Internet video conferencing.

Globally optimal mode selection is an approach from the perspective of error-resilient source coding. We believe a hybrid scheme integrating the globally optimal mode se-

lection with forward error correction (FEC) could further enhance the robustness of video communications over error-prone networks. This will be the focus of our future research.

APPENDIX

A. Facts

From the definition of *conditional probability*, we have the following facts.

Fact 1: Denote $P(\mathcal{B}|\mathcal{A})$ the conditional probability of an event $\mathcal{B}$ assuming $\mathcal{A}$. The probability of the intersection $\mathcal{A}\mathcal{B}$ is given by

$$P(\mathcal{A}\mathcal{B}) = P(\mathcal{A}) \cdot P(\mathcal{B}|\mathcal{A}). \tag{31}$$

Fact 2: Suppose that events $\mathcal{A}_i$ form a partition of $\mathcal{A}$ and events $\mathcal{B}_{ij}$ form a partition of $\mathcal{A}_i$. Denote p_i the probability of event $\mathcal{A}_i$ and q_j the conditional probability of an event $\mathcal{B}_{ij}$ assuming $\mathcal{A}_i$. Define a random variable $\mathbf{x}$ on $\mathcal{A}$ and a random variable $\mathbf{y}_i$ on $\mathcal{A}_i$. Then we have

$$E\{\mathbf{x}\} = \sum_i (p_i \cdot E\{\mathbf{y}_i\}). \tag{32}$$

Assume that random variable $\mathbf{x}$ takes the values x_{ij}, which is correspondent to event $\mathcal{B}_{ij}$. Then Eq. (32) follows from

$$
\begin{aligned}
E\{\mathbf{x}\} &= \sum_i \sum_j [P(\mathcal{B}_{ij}) \cdot x_{ij}] = \sum_i \sum_j [P(\mathcal{A}_i \mathcal{B}_{ij}) \cdot x_{ij}] \\
&= \sum_i \sum_j [P(\mathcal{A}_i) \cdot P(\mathcal{B}_{ij}|\mathcal{A}_i) \cdot x_{ij}] \\
&= \sum_i \sum_j (p_i \cdot q_j \cdot x_{ij}) \\
&= \sum_i \left[p_i \cdot \sum_j (q_j \cdot x_{ij}) \right] = \sum_i (p_i \cdot E\{\mathbf{y}_i\}).
\end{aligned}
$$

B. Proofs of the Lemmas Under Gilbert Path Model

Proof of Lemma 1: Regarding the relationship between the macroblock and the associated packets, and the relationship between the packets and set $\mathcal{X}$, we consider four cases as follows.

• *Case (a):* Suppose that packet $X_{\eta_i^n}$ belongs to set $\mathcal{X}$, i.e., $X_{\eta_i^n} \in \mathcal{X}$.
Since we assume frame 0 has been received correctly, all the packets in set $\mathcal{X}$ must have been received correctly, that is, $X_{\eta_i^n}$ and the previous packet $X_{\eta_i^n - 1}$ have been received correctly. Since $\bar{F}_i^n$ is at most contained by two packets $X_{\eta_i^n}$ and $X_{\eta_i^n - 1}$, $\bar{F}_i^n$ is also received correctly. Thus, we have $\tilde{P}_R^{(i,n)} = 1$ and $\tilde{P}_L^{(i,n)} = 0$.

• *Case (b):* $\bar{F}_i^n$ is split and partially contained by packet $X_{\eta_i^n}$ such that packet $X_{\eta_i^n - 1}$ belongs to set $\mathcal{X}$ while packet $X_{\eta_i^n}$ does not belong to set $\mathcal{X}$, i.e., $\bar{F}_i^n \preceq X_{\eta_i^n}$ ($X_{\eta_i^n - 1} \in \mathcal{X}$ and $X_{\eta_i^n} \notin \mathcal{X}$).
In this case, $\bar{F}_i^n$ must have been split into two packets: $X_{\eta_i^n - 1}$ and $X_{\eta_i^n}$. Since we assume frame 0 has been received correctly, all the packets in set $\mathcal{X}$ must have been received correctly, that is, packet $X_{\eta_i^n - 1}$ has been received correctly. Then, if $X_{\eta_i^n}$ is correctly received, $\bar{F}_i^n$ is received correctly. Under the condition that packet $X_{\eta_i^n - 1}$ is received correctly (with probability 1), the probability of the event that the next packet $X_{\eta_i^n}$ is received correctly is $P_{RR}^{(1)}$. Thus, from Fact 1, we have $\tilde{P}_R^{(i,n)} = P_{RR}^{(1)}$.
On the other hand, if packet $X_{\eta_i^n}$ is lost, $\bar{F}_i^n$ is lost. Under the condition that packet $X_{\eta_i^n - 1}$ is received correctly (with probability 1), the probability of the event that the next packet $X_{\eta_i^n}$ is lost is $P_{RL}^{(1)}$. Thus, from Fact 1, we have $\tilde{P}_L^{(i,n)} = P_{RL}^{(1)}$.

• *Case (c):* $\bar{F}_i^n$ is completely contained by packet $X_{\eta_i^n}$ such that packet $X_{\eta_i^n}$ does not belong to set $\mathcal{X}$, i.e., $\bar{F}_i^n \prec X_{\eta_i^n}$ ($X_{\eta_i^n} \notin \mathcal{X}$).
In this case, if $X_{\eta_i^n}$ is correctly received, $\bar{F}_i^n$ is received correctly. Since we assume frame 0 has been received correctly, all the packets in set $\mathcal{X}$ must have been received correctly, that is, the last packet in set $\mathcal{X}$, X_{K-1}, has been received correctly. Under the condition that packet X_{K-1} is received correctly (with probability 1), the probability of the event that packet $X_{\eta_i^n}$ is received correctly is $P_{RR}^{(\eta_i^n - K + 1)}$. Thus, from Fact 1, we have $\tilde{P}_R^{(i,n)} = P_{RR}^{(\eta_i^n - K + 1)}$.
On the other hand, if packet $X_{\eta_i^n}$ is lost, $\bar{F}_i^n$ is lost. Under the condition that packet X_{K-1} is received correctly (with probability 1), the probability of the event that packet $X_{\eta_i^n}$ is lost is $P_{RL}^{(\eta_i^n - K + 1)}$. Thus, from Fact 1, we have $\tilde{P}_L^{(i,n)} = P_{RL}^{(\eta_i^n - K + 1)}$.

• *Case (d):* $\bar{F}_i^n$ is split and partially contained by packet $X_{\eta_i^n}$ such that packet $X_{\eta_i^n - 1}$ does not belong to set $\mathcal{X}$, i.e., $\bar{F}_i^n \preceq X_{\eta_i^n}$ ($X_{\eta_i^n - 1} \notin \mathcal{X}$).
In this case, $\bar{F}_i^n$ must have been split into two packets: $X_{\eta_i^n - 1}$ and $X_{\eta_i^n}$. Only when both $X_{\eta_i^n - 1}$ and $X_{\eta_i^n}$ are correctly received, $\bar{F}_i^n$ is received correctly. Under the condition that packet $X_{\eta_i^n - 1}$ is received correctly (with

probability $P_{RR}^{(\eta_i^n - K)}$), the probability of the event that the next packet $X_{\eta_i^n}$ is received correctly is $P_{RR}^{(1)}$. Thus, from Fact 1, the probability of the event that both $X_{\eta_i^n - 1}$ and $X_{\eta_i^n}$ are received correctly is $P_{RR}^{(\eta_i^n - K)} \cdot P_{RR}^{(1)}$. Then, we have $\tilde{P}_R^{(i,n)} = P_{RR}^{(\eta_i^n - K)} \cdot P_{RR}^{(1)}$. Otherwise (with probability $1 - P_{RR}^{(\eta_i^n - K)} \cdot P_{RR}^{(1)}$), $\bar{F}_i^n$ is lost. Thus, we have $\tilde{P}_L^{(i,n)} = (1 - P_{RR}^{(\eta_i^n - K)} \cdot P_{RR}^{(1)})$.
Combining Case (a) to (d), we have Eqs. (9) and (10). ∎

Proof of Lemma 2: Regarding the relationship between the macroblocks and the associated packets, the relationship between the packets and set $\mathcal{X}$, and the sharing packet between $\bar{F}_{\bar{i}}^n$ and $\bar{F}_i^n$, we consider nine cases as follows. The relations are listed in Table VI.

• *Case 1:* Suppose that packet $X_{\eta_i^n}$ belongs to set $\mathcal{X}$, i.e., $X_{\eta_i^n} \in \mathcal{X}$.
In this case, since we assume frame 0 has been received correctly, all the packets in set $\mathcal{X}$ must have been received correctly, that is, $X_{\eta_i^n}$ and the previous packet $X_{\eta_i^n - 1}$ have been received correctly. Since $\bar{F}_i^n$ is at most contained by two packets $X_{\eta_i^n}$ and $X_{\eta_i^n - 1}$, $\bar{F}_i^n$ is received correctly. In other words, the probability of the event that $\bar{F}_i^n$ is lost is 0. Thus, we have $\hat{P}_{RL}^{(i,n)} = 0$ and $\hat{P}_{LL}^{(i,n)} = 0$.

• *Case 2:* $\bar{F}_i^n$ is split and partially contained by packet $X_{\eta_i^n}$ such that packet $X_{\eta_i^n - 1}$ belongs to set $\mathcal{X}$ while packet $X_{\eta_i^n}$ does not belong to set $\mathcal{X}$, i.e., $\bar{F}_i^n \preceq X_{\eta_i^n}$ ($X_{\eta_i^n - 1} \in \mathcal{X}$ and $X_{\eta_i^n} \notin \mathcal{X}$).
In this case, $\bar{F}_i^n$ must have been split into two packets: $X_{\eta_i^n - 1}$ and $X_{\eta_i^n}$. Since we assume frame 0 has been received correctly, all the packets in set $\mathcal{X}$ must have been received correctly, that is, packet $X_{\eta_i^n - 1}$ has been received correctly. Since $\bar{F}_{\bar{i}}^n$ is generated before $\bar{F}_i^n$, $\bar{F}_{\bar{i}}^n$ must have been received correctly. In other words, the probability of the event that $\bar{F}_{\bar{i}}^n$ is lost is 0.
With $X_{\eta_i^n - 1}$ received correctly, if packet $X_{\eta_i^n}$ is lost, $\bar{F}_i^n$ is lost. Under the condition that packet $X_{\eta_i^n - 1}$ is received correctly (with probability 1), the probability of the event that the next packet $X_{\eta_i^n}$ is lost is $P_{RL}^{(1)}$. Thus, from Fact 1, the probability of the event that packet $X_{\eta_i^n - 1}$ is received correctly and packet $X_{\eta_i^n}$ is lost is $P_{RL}^{(1)}$. That is, $\hat{P}_{RL}^{(i,n)} = P_{RL}^{(1)}$.
On the other hand, since the probability of the event that $\bar{F}_{\bar{i}}^n$ is lost is 0, from Fact 1, we have $\hat{P}_{LL}^{(i,n)} = 0$.

• *Case 3:* $\bar{F}_i^n$ is completely contained by packet $X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n$ is partially or completely contained by packet $X_{\eta_i^n}$ such that packet $X_{\eta_i^n}$ does not belong to set $\mathcal{X}$, i.e., $\bar{F}_i^n \prec X_{\eta_i^n}$ ($X_{\eta_i^n} \notin \mathcal{X}$ and $X_{\eta_i^n} = X_{\eta_i^n}$).
In this case, both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ can be lost but the event $\{\bar{F}_i^n$ is lost and $\bar{F}_{\bar{i}}^n$ is received$\}$ cannot happen. Thus, we have $\hat{P}_{RL}^{(i,n)} = 0$. On the other hand, only when packet $X_{\eta_i^n}$ is lost (with probability $P_{RL}^{(\eta_i^n - K + 1)}$), both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ are lost. Thus, we have $\hat{P}_{LL}^{(i,n)} = P_{RL}^{(\eta_i^n - K + 1)}$.

• *Case 4:* $\bar{F}_i^n$ is split and $\bar{F}_{\bar{i}}^n$ is completely contained by packet $X_{\eta_i^n - 1}$ such that packet $X_{\eta_i^n - 1}$ does not belong to set $\mathcal{X}$, i.e., $\bar{F}_i^n \preceq X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n \prec X_{\eta_i^n - 1}$ ($X_{\eta_i^n - 1} \notin \mathcal{X}$).

TABLE VI

Relationship.

	Containing F_i^n	Containing $F_{\bar{i}}^n$	$\mathcal{X}$	Sharing packet
Case 1	don't care	don't care	$X_{\eta_i^n} \in \mathcal{X}$	don't care
Case 2	$X_{\eta_i^n-1}$ and $X_{\eta_i^n}$	don't care	$X_{\eta_i^n-1} \in \mathcal{X}$ and $X_{\eta_i^n} \notin \mathcal{X}$	don't care
Case 3	$X_{\eta_i^n}$	don't care	$X_{\eta_i^n} \notin \mathcal{X}$	$X_{\eta_i^n} = X_{\eta_{\bar{i}}^n}$ (Share)
Case 4	$X_{\eta_i^n-1}$ and $X_{\eta_i^n}$	$X_{\eta_{\bar{i}}^n-1}$	$X_{\eta_i^n-1} \notin \mathcal{X}$	$X_{\eta_i^n-1} = X_{\eta_{\bar{i}}^n}$ (Share)
Case 5	$X_{\eta_i^n-1}$ and $X_{\eta_i^n}$	$X_{\eta_{\bar{i}}^n-2}$ and $X_{\eta_{\bar{i}}^n-1}$	$X_{\eta_i^n-1} \notin \mathcal{X}$	$X_{\eta_i^n-1} = X_{\eta_{\bar{i}}^n}$ (Share)
Case 6	$X_{\eta_i^n}$	$X_{\eta_{\bar{i}}^n}$	$X_{\eta_i^n} \notin \mathcal{X}$	$X_{\eta_i^n} \neq X_{\eta_{\bar{i}}^n}$ (NoShare)
Case 7	$X_{\eta_i^n}$	$X_{\eta_{\bar{i}}^n-1}$ an $X_{\eta_{\bar{i}}^n}$	$X_{\eta_i^n} \notin \mathcal{X}$	$X_{\eta_i^n} \neq X_{\eta_{\bar{i}}^n}$ (NoShare)
Case 8	$X_{\eta_i^n-1}$ and $X_{\eta_i^n}$	$X_{\eta_{\bar{i}}^n}$	$X_{\eta_i^n} \notin \mathcal{X}$	$X_{\eta_i^n-1} \neq X_{\eta_{\bar{i}}^n}$ (NoShare)
Case 9	$X_{\eta_i^n-1}$ and $X_{\eta_i^n}$	$X_{\eta_{\bar{i}}^n-1}$ and $X_{\eta_{\bar{i}}^n}$	$X_{\eta_i^n} \notin \mathcal{X}$	$X_{\eta_i^n-1} \neq X_{\eta_{\bar{i}}^n}$ (NoShare)

Under the condition that packet $X_{\eta_i^n-1}$ is received correctly (with probability $P_{RR}^{(\eta_i^n-K)}$), the probability of the event that the next packet $X_{\eta_i^n}$ is lost is $P_{RL}^{(1)}$. Then, from Fact 1, the probability of the event that packet $X_{\eta_i^n-1}$ is received correctly and packet $X_{\eta_i^n}$ is lost is $P_{RR}^{(\eta_i^n-K)} \cdot P_{RL}^{(1)}$. That is, the probability of the event that $\bar{F}_{\bar{i}}^n$ is received correctly and $\bar{F}_i^n$ is lost is $P_{RR}^{(\eta_i^n-K)} \cdot P_{RL}^{(1)}$. Thus, we have $\hat{P}_{RL}^{(i,n)} = P_{RR}^{(\eta_i^n-K)} \cdot P_{RL}^{(1)}$. On the other hand, only when packet $X_{\eta_i^n-1}$ is lost (with probability $P_{RL}^{(\eta_i^n-K)}$), both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ are lost. Thus, we have $\hat{P}_{LL}^{(i,n)} = P_{RL}^{(\eta_i^n-K)}$.

• *Case 5:* $\bar{F}_i^n$ is split and $\bar{F}_{\bar{i}}^n$ is partially contained by packet $X_{\eta_i^n-1}$ such that packet $X_{\eta_i^n-1}$ does not belong to set $\mathcal{X}$, i.e., $\bar{F}_i^n \preceq X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n \preceq X_{\eta_i^n-1}$ ($X_{\eta_i^n-1} \notin \mathcal{X}$).

In this case, only when the event {both packet $X_{\eta_i^n-2}$ and $X_{\eta_i^n-1}$ are received correctly, and $X_{\eta_i^n}$ is lost} happens, $\bar{F}_{\bar{i}}^n$ is received correctly and $\bar{F}_i^n$ is lost. Under the condition that packet $X_{\eta_i^n-2}$ is received correctly (with probability $P_{RR}^{(\eta_i^n-K-1)}$), the probability of the event that the next packet $X_{\eta_i^n-1}$ is received correctly is $P_{RR}^{(1)}$. Under the condition that packet $X_{\eta_i^n-1}$ is received correctly, the probability of the event that the next packet $X_{\eta_i^n}$ is lost is $P_{RL}^{(1)}$. Then, from Fact 1, the probability of the event that both packet $X_{\eta_i^n-2}$ and $X_{\eta_i^n-1}$ are received correctly, and $X_{\eta_i^n}$ is lost is $P_{RR}^{(\eta_i^n-K-1)} \cdot P_{RR}^{(1)} \cdot P_{RL}^{(1)}$. That is, the probability of the event that $\bar{F}_{\bar{i}}^n$ is received correctly and $\bar{F}_i^n$ is lost is $P_{RR}^{(\eta_i^n-K-1)} \cdot P_{RR}^{(1)} \cdot P_{RL}^{(1)}$. Thus, we have $\hat{P}_{RL}^{(i,n)} = P_{RR}^{(\eta_i^n-K-1)} \cdot P_{RR}^{(1)} \cdot P_{RL}^{(1)}$.

On the other hand, there are two subcases for the event {both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ are lost}.

– *Subcase 5.1:* If packet $X_{\eta_i^n-1}$ is lost, both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ will be lost. The probability of this case is $P_{RL}^{(\eta_i^n-K)}$.

– *Subcase 5.2:* If the event {both $X_{\eta_i^n-2}$ and $X_{\eta_i^n}$ are lost, and $X_{\eta_i^n-1}$ is received} happens, both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ will be lost. The probability of this case is $P_{RL}^{(\eta_i^n-K-1)} \cdot P_{LR}^{(1)} \cdot P_{RL}^{(1)}$. Combining Subcase 5.1 and 5.2, we have $\hat{P}_{LL}^{(i,n)} = P_{RL}^{(\eta_i^n-K)} + P_{RL}^{(\eta_i^n-K-1)} \cdot P_{LR}^{(1)} \cdot P_{RL}^{(1)}$.

• *Case 6:* $\bar{F}_i^n$ is completely contained by packet $X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n$ is completely contained by packet $X_{\eta_{\bar{i}}^n}$ rather than $X_{\eta_i^n}$ such that packet $X_{\eta_i^n}$ does not belong to set $\mathcal{X}$, i.e., $\bar{F}_i^n \prec X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n \prec X_{\eta_{\bar{i}}^n}$ ($X_{\eta_i^n} \notin \mathcal{X}$ and $X_{\eta_i^n} \neq X_{\eta_{\bar{i}}^n}$).

In this case, only when the event {$\bar{F}_{\bar{i}}^n$ is received correctly and $X_{\eta_i^n}$ is lost} happens (with probability $\tilde{P}_R^{(\bar{i},n)} \cdot P_{RL}^{(\eta_i^n-\eta_{\bar{i}}^n)}$), $\bar{F}_{\bar{i}}^n$ is received correctly and $\bar{F}_i^n$ is lost. Thus, we have $\hat{P}_{RL}^{(i,n)} = \tilde{P}_R^{(\bar{i},n)} \cdot P_{RL}^{(\eta_i^n-\eta_{\bar{i}}^n)}$.

On the other hand, only when the event {both $X_{\eta_i^n}$ and $X_{\eta_{\bar{i}}^n}$ are lost} happens, both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ will be lost. The probability of this event is $P_{RL}^{(\eta_{\bar{i}}^n-K+1)} \cdot P_{LL}^{(\eta_i^n-\eta_{\bar{i}}^n)}$. Thus, we have $\hat{P}_{LL}^{(i,n)} = P_{RL}^{(\eta_{\bar{i}}^n-K+1)} \cdot P_{LL}^{(\eta_i^n-\eta_{\bar{i}}^n)}$.

• *Case 7:* $\bar{F}_i^n$ is completely contained by packet $X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n$ is partially contained by packet $X_{\eta_{\bar{i}}^n}$ rather than $X_{\eta_i^n}$ such that packet $X_{\eta_i^n}$ does not belong to set $\mathcal{X}$, i.e., $\bar{F}_i^n \prec X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n \preceq X_{\eta_{\bar{i}}^n}$ ($X_{\eta_i^n} \notin \mathcal{X}$ and $X_{\eta_i^n} \neq X_{\eta_{\bar{i}}^n}$).

In this case, only when the event {$\bar{F}_{\bar{i}}^n$ is received correctly and $X_{\eta_i^n}$ is lost} happens (with probability $\tilde{P}_R^{(\bar{i},n)} \cdot P_{RL}^{(\eta_i^n-\eta_{\bar{i}}^n)}$), $\bar{F}_{\bar{i}}^n$ is received correctly and $\bar{F}_i^n$ is lost. Thus, we have $\hat{P}_{RL}^{(i,n)} = \tilde{P}_R^{(\bar{i},n)} \cdot P_{RL}^{(\eta_i^n-\eta_{\bar{i}}^n)}$.

On the other hand, there are three subcases for the event {both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ are lost}.

– *Subcase 7.1:* If the event {packets $X_{\eta_{\bar{i}}^n-1}$, $X_{\eta_{\bar{i}}^n}$ and $X_{\eta_i^n}$ are all lost} happens, both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ will be lost. The probability of this case is $P_{RL}^{(\eta_{\bar{i}}^n-K)} \cdot P_{LL}^{(1)} \cdot P_{LL}^{(\eta_i^n-\eta_{\bar{i}}^n)}$.

– *Subcase 7.2:* If the event {both $X_{\eta_{\bar{i}}^n-1}$ and $X_{\eta_{\bar{i}}^n}$ are lost, and $X_{\eta_i^n}$ is received correctly} happens, both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ will be lost. The probability of this case is $P_{RL}^{(\eta_{\bar{i}}^n-K)} \cdot P_{LR}^{(1)} \cdot P_{RL}^{(\eta_i^n-\eta_{\bar{i}}^n)}$.

– *Subcase 7.3:* If the event {both $X_{\eta_{\bar{i}}^n}$ and $X_{\eta_i^n}$ are lost, and $X_{\eta_{\bar{i}}^n-1}$ is received correctly} happens, both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ will be lost. The probability of this case is $P_{RR}^{(\eta_{\bar{i}}^n-K)} \cdot P_{RL}^{(1)} \cdot P_{LL}^{(\eta_i^n-\eta_{\bar{i}}^n)}$.

Combining Subcase 7.1 to 7.3, we have $\hat{P}_{LL}^{(i,n)} = P_{RL}^{(\eta_{\bar{i}}^n-K)} \cdot P_{LL}^{(1)} \cdot P_{LL}^{(\eta_i^n-\eta_{\bar{i}}^n)} + P_{RL}^{(\eta_{\bar{i}}^n-K)} \cdot P_{LR}^{(1)} \cdot P_{RL}^{(\eta_i^n-\eta_{\bar{i}}^n)} + P_{RR}^{(\eta_{\bar{i}}^n-K)} \cdot P_{RL}^{(1)} \cdot$

$P_{LL}^{(\eta_i^n - \eta_{\bar{i}}^n)}$.

• *Case 8:* $\bar{F}_i^n$ is partially contained by packet $X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n$ is completely contained by packet $X_{\eta_{\bar{i}}^n}$ rather than $X_{\eta_i^n}$ such that packet $X_{\eta_i^n}$ does not belong to set $\mathcal{X}$, i.e., $\bar{F}_i^n \preceq X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n \prec X_{\eta_{\bar{i}}^n}$ $(X_{\eta_i^n} \notin \mathcal{X}$ and $X_{\eta_i^n - 1} \neq X_{\eta_i^n})$.

In this case, the probability of the event that $\bar{F}_{\bar{i}}^n$ is received correctly and $\bar{F}_i^n$ is lost is $\tilde{P}_R^{(\bar{i},n)} \cdot (1 - P_{RL}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)})$. Thus, we have $\hat{P}_{RL}^{(i,n)} = \tilde{P}_R^{(\bar{i},n)} \cdot (1 - P_{RL}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)})$.

On the other hand, the probability of the event that both $\bar{F}_{\bar{i}}^n$ and $\bar{F}_i^n$ are lost is $P_{RL}^{(\eta_{\bar{i}}^n - K + 1)} \cdot (1 - P_{LR}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)})$. Thus, we have $\hat{P}_{LL}^{(i,n)} = P_{RL}^{(\eta_{\bar{i}}^n - K + 1)} \cdot (1 - P_{LR}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)})$.

• *Case 9:* $\bar{F}_i^n$ is partially contained by packet $X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n$ is partially contained by packet $X_{\eta_{\bar{i}}^n}$ rather than $X_{\eta_i^n}$ such that packet $X_{\eta_i^n}$ does not belong to set $\mathcal{X}$, i.e., $\bar{F}_i^n \preceq X_{\eta_i^n}$ and $\bar{F}_{\bar{i}}^n \preceq X_{\eta_{\bar{i}}^n}$ $(X_{\eta_i^n} \notin \mathcal{X}$ and $X_{\eta_i^n - 1} \neq X_{\eta_i^n})$.

In this case, the probability of the event that $\bar{F}_{\bar{i}}^n$ is received correctly and $\bar{F}_i^n$ is lost is $\tilde{P}_R^{(\bar{i},n)} \cdot (1 - P_{RL}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)})$. Thus, we have $\hat{P}_{RL}^{(i,n)} = \tilde{P}_R^{(\bar{i},n)} \cdot (1 - P_{RL}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)})$.

On the other hand, there are three subcases for the event {both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ are lost}.

– *Subcase 9.1:* If the event $\{X_{\eta_i^n - 1}, X_{\eta_{\bar{i}}^n},$ and $\bar{F}_i^n$ are all lost} happens, both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ will be lost. The probability of this case is $P_{RL}^{(\eta_{\bar{i}}^n - K)} \cdot P_{LL}^{(1)} \cdot (1 - P_{LR}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)})$.

– *Subcase 9.2:* If the event {both $X_{\eta_{\bar{i}}^n - 1}$ and $\bar{F}_{\bar{i}}^n$ are lost, and $X_{\eta_{\bar{i}}^n}$ is received correctly} happens, both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ will be lost. The probability of this case is $P_{RL}^{(\eta_{\bar{i}}^n - K)} \cdot P_{LR}^{(1)} \cdot (1 - P_{RR}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)})$.

– *Subcase 9.3:* If the event {both $X_{\eta_{\bar{i}}^n}$ and $\bar{F}_i^n$ are lost, and $X_{\eta_{\bar{i}}^n - 1}$ is received correctly} happens, both $\bar{F}_i^n$ and $\bar{F}_{\bar{i}}^n$ will be lost. The probability of this case is $P_{RR}^{(\eta_{\bar{i}}^n - K)} \cdot P_{RL}^{(1)} \cdot (1 - P_{LR}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)})$.

Combining Subcase 9.1 to 9.3, we have $\hat{P}_{LL}^{(i,n)} = P_{RL}^{(\eta_{\bar{i}}^n - K)} \cdot P_{LL}^{(1)} \cdot (1 - P_{LR}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)}) + P_{RL}^{(\eta_{\bar{i}}^n - K)} \cdot P_{LR}^{(1)} \cdot (1 - P_{RR}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)}) + P_{RR}^{(\eta_{\bar{i}}^n - K)} \cdot P_{RL}^{(1)} \cdot (1 - P_{LR}^{(\eta_i^n - \eta_{\bar{i}}^n - 1)} \cdot P_{RR}^{(1)})$.

Combining Case 1 to 9, we have Tables II and III. ∎

Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful comments.

References

[1] J-C. Bolot and T. Turletti, "Adaptive error control for packet video in the Internet," in *Proc. IEEE ICIP'96*, Sept. 1996.

[2] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 246–250, Feb. 1997.

[3] G. Cote and F. Kossentini, "Optimal intra coding of blocks for robust video communication over the Internet," to appear in *EUROSIP Image Communication Special Issue on Real-time Video over the Internet*.

[4] N. Faerber, personal communications.

[5] B. Girod, K. W. Stuhlmueller, M. Link and U. Horn, "Packet loss resilient Internet video streaming," in *SPIE Proc. Visual Communications and Image Processing (VCIP'99)*, Jan. 1999.

[6] ISO/IEC JTC 1/SC 29/WG 11, "MPEG-4 video verification model version 8.0," MPEG97/N1796, July 1997.

[7] 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.

[8] J. Lee and B. W. Dickinson, "Rate-distortion optimized frame type selection for MPEG encoding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, no. 3, pp. 501–510, June 1997.

[9] F. Le Leannec and C. M. Guillemot, "Error resilient video transmission over the Internet," in *SPIE Proc. Visual Communications and Image Processing (VCIP'99)*, Jan. 1999.

[10] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, pp. 74–90, Nov. 1998.

[11] K. Ramchandran, A. Ortega and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Trans. on Image Processing*, vol. 37, pp. 533–545, Aug. 1994.

[12] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, "RTP: a transport protocol for real-time applications," *RFC 1889*, Internet Engineering Task Force, Jan. 1996.

[13] H. Schulzrinne, D. Hoffman, M. Speer, R. Civanlar, A. Basso, V. Balabanian and C. Herpel, "RTP payload format for MPEG-4 elementary streams," *Internet Draft*, Internet Engineering Task Force, March 1998.

[14] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. on Acoust., Speech and Signal Processing*, vol. 36, no. 9, pp. 1445–1453, Sept. 1988.

[15] E. Steinbach, N. Faerber and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environments," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, no. 6, pp. 872–881, Dec. 1997.

[16] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, pp. 74–90, Nov. 1998.

[17] R. Talluri, "Error-resilience video coding in the ISO MPEG-4 standard," *IEEE Commun. Magazine*, pp. 112–119, June 1998.

[18] T. Turletti and C. Huitema, "RTP payload format for H.261 video streams," *RFC 2032*, Internet Engineering Task Force, Oct. 1996.

[19] J. Villasenor, Y.-Q. Zhang and J. Wen, "Robust video coding algorithms and systems," *Proceedings of the IEEE*, vol. 87, no. 10, pp. 1724–1733, Oct. 1999.

[20] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: a review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[21] S. Wenger and G. Cote, "Using RFC2429 and H.263+ at low to medium bit-rates for low-latency applications," *Packet Video'99 Workshop*, New York, April 1999.

[22] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell and S. K. Mitra, "Rate-distortion optimized mode selection for very low bit rate video coding and the emerging H.263 standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, no. 2, pp. 182–190, April 1996.

[23] D. Wu, Y. T. Hou, W. Zhu, H.-J. Lee, T. Chiang, Y.-Q. Zhang, 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.

[24] Y.-Q. Zhang, Y.-J. Liu and R. Pickholtz, "Layered image transmission over cellular radio channels," *IEEE Trans. on Vehicular Technology*, vol. 43, Aug. 1994.

[25] Y.-Q. Zhang and X. Lee, "Performance of MPEG codecs in the presence of errors," *Journal of Visual Communications and Image Representation*, vol. 5, no. 4, pp. 379–387, Dec. 1994.

[26] Special Issue on Wireless Video, *IEEE Trans. on Circuits and Systems for Video Technology*, April 1996.

[27] C. Zhu, "RTP payload format for H.263 video streams," *RFC 2190*, Internet Engineering Task Force, Sept. 1997.

Dapeng Wu received the B.E degree from Huazhong University of Science and Technology, and the M.E. degree from Beijing University of Posts and Telecommunications in 1990 and 1997 respectively, both in Electrical Engineering. From July 1997 to December 1999, he conducted graduate research at Polytechnic University, Brooklyn, New York. Since January 2000, he has been working towards his Ph.D. degree in Electrical Engineering at Carnegie Mellon University, Pittsburgh, PA. During the summer of 1998 and part of 1999, he conducted research at Fujitsu Laboratories of America, Sunnyvale, California, on architectures and traffic management algorithms in integrated services (IntServ) networks and differentiated services (DiffServ) Internet for multimedia applications. His current interests are in the areas of rate control and error control for video communications over the Internet and wireless networks, and next generation Internet architecture, protocols, implementations for integrated and differentiated services. He is a student member of the IEEE and the ACM.

Yiwei Thomas Hou obtained his B.E. degree (*Summa Cum Laude*) from the City College of New York in 1991, the M.S. degree from Columbia University in 1993, and the Ph.D. degree from Polytechnic University, Brooklyn, New York, in 1997, all in Electrical Engineering. He was awarded a five-year National Science Foundation Graduate Research Traineeship for pursuing Ph.D. degree in high speed networking, and was recipient of the Alexander Hessel award for outstanding Ph.D. dissertation (1997–1998 academic year) from Polytechnic University. While a graduate student, he worked at AT&T Bell Labs, Murray Hill, New Jersey, during the summers of 1994 and 1995, on design and implementations of IP/ATM internetworking; he conducted research at Bell Labs, Lucent Technologies, Holmdel, New Jersey, during the summer of 1996, on fundamental problems on network traffic management.

Since September 1997, Dr. Hou has been a Researcher at Fujitsu Laboratories of America, Sunnyvale, California. He received Intellectual Property Contribution Award from Fujitsu Laboratories of America in 1999. His current research interests are in the areas of quality of service (QoS) support for transporting multimedia (MPEG-4) over the Internet; scalable architecture, protocols, and implementations for differentiated services (DiffServ). Dr. Hou is a member of the IEEE, ACM, Sigma Xi and New York Academy of Sciences.

Bo Li received the B.S. (cum laude) and M.S. degrees in the Computer Science from Tsinghua University (Beijing) in 1987 and 1989, respectively, and the Ph.D. degree in the Computer Engineering from University of Massachusetts at Amherst in 1993. Between 1994 and 1996, he worked on high performance routers and ATM switches in IBM Networking System Division, Research Triangle Park, North Carolina. He joined the faculty of the Computer Science Department of the Hong Kong University of Science and Technology in January 1996.

Dr. Li has been on editorial board for ACM Mobile Computing and Communications Review (MC2R) and Journal of Communications and Networks (JCN). He will be serving as an editor for ACM/Baltzer Journal of Wireless Networks (WINET). He has been co-guest editing special issues for *IEEE Communications Magazine, IEEE Journal of Selected Areas in Communications* and the upcoming *SPIE/Baltzer Optical Networks Magazine.* He has been involved in organizing many conferences such as IEEE Infocom, ICDCS and ICC. He will be the international vice-chair for Infocom'2001.

Dr. Li's current research interests include wireless mobile networking supporting multimedia, voice and video (MPEG-2 and MPEG-4) transmission over the internet and all optical networks using WDM.

Wenwu Zhu received the B.E. and M.E. degrees from National University of Science and Technology, Changsha, China, in 1985 and 1988, respectively. He received the M.S. degree from Illinois Institute of Technology, Chicago, IL, and the Ph.D. degree from Polytechnic University, Brooklyn, NY, in 1993 and 1996, respectively, both in Electrical Engineering. Dr. Zhu joined Microsoft Research in China in October 1999 as a Researcher. From July 1996 to October 1999, he was with Bell Labs, Lucent technologies. From August 1988 to December 1990, he was with Graduate School, University of Science and Technology of China (USTC) , and Institute of Electronics, Academia Sinica (Chinese Academy of Sciences), Beijing. Dr. Zhu has published over 50 papers in various referred conferences and journals. His current research interests are in the areas of video over IP and wireless networks, multimedia signal processing and multimedia applications.

Ya-Qin Zhang joined Microsoft Research in China in January 1999 as the Assistant Managing Director of the company. He was previously the Director of Multimedia Technology Laboratory at Sarnoff Corporation in Princeton, New Jersey (formerly David Sarnoff Research Center and RCA Laboratories). His laboratory is a world leader in MPEG2/DTV, MPEG4/VLBR, and multimedia information technologies. He was with GTE Laboratories Inc. in Waltham, Massachusetts, and Contel Technology Center in Virginia from 1989 to 1994. He has authored and co-authored over 150 refereed papers and 30 US patents granted or pending in digital video, Internet multimedia, wireless and satellite communications. Many of the technologies he and his team developed have become the basis for start-up ventures, commercial products, and international standards.

Dr. Zhang was the Editor-In-Chief for the *IEEE Transactions on Circuits and Systems for Video Technology* from July 1997 to July 1999. He was a Guest Editor for the special issue on "Advances in Image and Video Compression" for the *Proceedings of the IEEE* (February 1995). He serves on the editorial boards of seven other professional journals and over a dozen conference committees. He has been an active contributor to the ISO/MPEG and ITU standardization efforts in digital video and multimedia.

Dr. Zhang is a Fellow of IEEE. He received numerous awards, including several industry technical achievement awards and IEEE awards. He was awarded as the "Research Engineer of the Year" in 1998 by the Central Jersey Engineering Council for his "leadership and invention in communications technology, which has enabled dramatic advances in digital video compression and manipulation for broadcast and interactive television and networking applications."

H. Jonathan Chao received the B.S.E.E. and M.S.E.E. degrees from National Chiao Tung University, Taiwan, in 1977 and 1980, respectively, and the Ph.D. degree in Electrical Engineering from The Ohio State University, Columbus, Ohio, in 1985. He is a Professor in the Department of Electrical Engineering at Polytechnic University, Brooklyn, New York, which he joined in January 1992. His research interests include large-capacity packet switches and routers, packet scheduling and buffer management, and congestion flow control in IP/ATM networks. From 1985 to 1991, he was a Member of Technical Staff at Bellcore, New Jersey, where he conducted research in the area of SONET/ATM broadband networks. He was involved in architecture designs and ASIC implementations, such as the first SONET-like Framer chip, ATM Layer chip, and Sequencer chip (the first chip handling packet scheduling). He received Bellcore Excellence Award in 1987.

He served as a Guest Editor for *IEEE Journal on Selected Areas in Communications* special issue on "Advances in ATM Switching

Systems for B-ISDN" (June 1997) and special issue on "Next Generation IP Switches and Routers" (June 1999). He is currently serving as an Editor for *IEEE/ACM Transactions on Networking.*