ROBUST DISTRIBUTED MULTI-POINT VIDEO CONFERENCING OVER ERROR-PRONE CHANNELS

Meng Chen, Guan-Ming Su, and Min Wu

Department of Electrical and Computer Engineering, University of Maryland, College Park.

ABSTRACT

In this paper, we propose a novel multi-point video conferencing system through error-prone channels, where the aggregation of multiple video streams and resource allocation are performed in a distributed manner. Video stream combiners, which are located in different geographical areas and serve as portals for conferees, aggregate incoming streams supplied by local users with other streams aggregated from nearby video stream combiners. A distributed multi-stream error protection scheme is performed in each video stream combiner to minimize the maximal expected video distortion among all aggregated streams. The simulation results demonstrate that our proposed scheme outperforms the traditional multicasting scheme by 1dB $\sim$ 1.4dB in terms of average PSNR.

1. INTRODUCTION

With the rapid development of video coding and communication technologies, transmitting real-time encoded video programs among multiple users has become a promising service. Multi-point video conferencing, which involves multiples conferees and realizes a virtual conference room, is one of the potential applications. Conventional multi-point video conferencing systems often consider a centralized scheme and assume error-free communication channel [1, 2]. When we consider holding a conference over a large-scale network with time-varying error-prone channels, centralized schemes require long round-trip delays for resource allocation and cannot react to fast changing conditions in both communication channel and video content.

Instead of centralized control, system designers can realize conferencing systems by utilizing receiver-driven layered multicasting algorithms [3, 4] and/or multi-hop forward error coding (FEC) transcoding [5, 6] to respond to time-varying and heterogeneous channel conditions. Since multiple streams are exchanged among multiple users, these streams may share the same transmission path. A dynamic resource allocation for each stream with awareness of other coexisting streams in the same path is more efficient than a static allocation. In this paper, we explore the multi-stream diversity to provide better video quality and study how to perform cross-layer multi-stream error protection in a distributed manner.

This paper is organized as follows. We introduce our proposed system in Section 2. In Section 3, we formulate the error protection strategy for the proposed system as an optimization problem and propose a fast algorithm to solve it. Simulation results are presented in Section 4 and conclusions are drawn in Section 5.

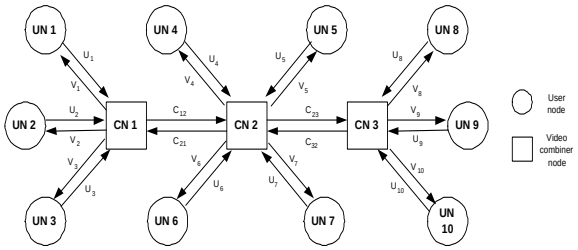

Fig. 1. Distributed multi-point video conferencing system

2. SYSTEM OVERVIEW

Figure 1 illustrates the proposed distributed multi-point video conferencing system, where there are two different types of nodes, namely, user node (UN) and video stream combiner node (CN). Each conferee j located at UN j will encode the captured video frame into compressed bitstream in real time. Then, the FEC protected bitstream and the rate-distortion (R-D) information are transmitted through uplink U_j to a neighboring CN. Every CN m will retrieve the video source bitstreams and the corresponding R-D from the successfully received FEC coded streams of each video frame. Then, the CN m performs joint multi-stream error protection to form a merged stream, and transmits the aggregated stream along with R-Ds' through a link C_{mn} to the next CN n. Each conferee will receive the merged stream containing video frames from all other conferees through downlink V_j from a neighboring CN. For transmission in each link, each node basically performs the same operation to protect and transmit streams within each video frame refreshing interval. The operation involves collecting local information including R-D embedded in all incoming streams and current channel conditions for the next hop, and performing multi-stream optimization to form a merged stream. To support multi-point video conferencing in real time, all nodes will perform optimizations simultaneously such that the aggregated video streams can reach the maximal quality subject to limited channel resources in all transmission paths. By doing so, we arrive at a distributed design for a multi-point video conferencing system with exploration of multi-stream diversity and cross-layer design.

The selected video codec should provide high flexibility to facilitate rate adaptation and provide accurate R-D information with low overhead. While the proposed framework can be extended to use other scalable codecs, to demonstrate the concept, we adopt MPEG-4 Fine Granularity Scalability

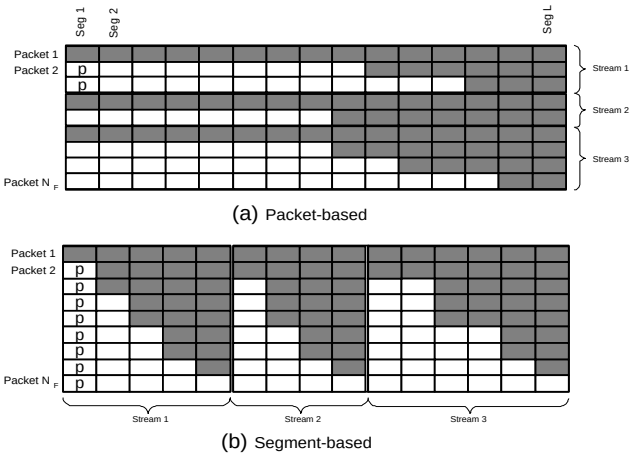

Fig. 2. Multi-stream merging strategy. The shaded and white block indicate the source and parity symbols, respectively

(FGS) coding [7] in this work. FGS is a two-layer scheme consisting of a non-scalable base layer and a highly scalable FGS enhancement layer. Any truncated FGS bitstream corresponding to each frame can be decoded. The more FGS bits the decoder receives and decodes, the higher the video quality is. In addition, the R-D function of FGS layer at the frame level can be well approximated as a piecewise linear line by interpolating the R-D pairs obtained for recovering each complete DCT bit plane [8]. Therefore, the R-D function for each video frame can be described using a small amount of bits.

We focus on the fixed-length packetization for FEC because it is a relatively matured technique to cope with erasure channel and it is widely used due to its simplicity. Because FGS enhancement layer uses bitplane based coding [7], the decoding of the symbols in its remaining bitplanes following a lost symbol may not improve the visual quality of the received video bitstream. Therefore, FGS enhancement data has a monotonically decreasing priority for error protection. We adopt an unequal error protection method of multiple descriptions through forward error correction codes (MDFEC) [9] for FGS layer, because it achieves good perceptual video quality in delivering single video stream with unequal priority for error protection. Given N packets, MDFEC performs error protection and packetization as follows: A segment is defined as a set of symbols located at the same position of each of these N packets. The FGS bitstream is filled into N packets segment by segment, and Reed-Solomon (RS) code is applied within each segment. A higher error protection level of RS code is applied for the segment with higher priority. If the receiver receives $\bar{k}$ packets successfully out of N packets, then the segments encoded with RS(N, k) codes, where $k \leq \bar{k}$, can be correctly decoded.

The traditional strategy to perform multi-stream error protection is the packet-based approach as shown in Figure 2(a). The packet-based approach assigns each stream a set of packets, N_j, and calculates the optimal MDFEC configuration within the assigned packet set. We propose a new strategy of segment-based allocation, namely, each stream can store data in certain segments in all available packets. For stream j, we need to determine the number of segments and the RS configuration of each assigned segment. Figure 2(b) illustrates the segment-based scheme. By spreading multiple video streams in a larger number of packets, the segment-based scheme can have higher effective bandwidth after FEC protection, thus provide better overall video quality.

3. MULTI-STREAM AGGREGATION

In this section, we formulate the proposed segment-based strategy as an optimization problem, and then present a fast algorithm to obtain the optimal solution.

3.1. Problem Formulation

Suppose the outbound channel of a video combiner can transmit N packets to the next hop within each video frame interval. There are J streams to be merged into these N packets and there are a total number of L segments in each packet. To facilitate the discussion, we use $a_{j,l} \in \{0, 1\}$ as an indicator to represent whether segment l is allocated to stream j. The overall segment-to-stream assignment can be represented as a matrix $\mathbf{A}$ with $[\mathbf{A}]_{j,l} = a_{j,l}$. Note that each segment can be assigned to at most one video stream, i.e. $\sum_{j=1}^{J} a_{j,l} \leq 1, \forall l$. Let L_j be the number of segments assigned to stream j and the total number of segments assigned to all streams should not exceed the maximal number of segments in each packet, L. In addition, we use $f_{i,l} \in \{0, 1\}$ as an indicator to represent whether the number of source symbols assigned to segment l is equal to i. The overall source symbol-to-segment assignment can be represented as a matrix $\mathbf{F}$ with $[\mathbf{F}]_{i,l} = f_{i,l}$. For unequal error protection, we apply stronger RS codes for more important data, i.e. $\sum_{i=1}^{N} i \cdot f_{i,l} \leq \sum_{i=1}^{N} i \cdot f_{i,l+1}$, if segments l and $l + 1$ are allocated to the same video stream.

Suppose the receiver located in the next CN receives exactly n packets when CN sends N packets, the reconstructed video quality for stream j can be represented as follows:

$$D_{j,n}(\mathbf{A}, \mathbf{F}) = D_j(\sum_{l=1}^{L} \sum_{i=1}^{n} a_{j,l} \cdot i \cdot f_{i,l}), \quad (1)$$

where $D_j(\cdot)$ is stream j's R-D function for current incoming frame. The distortion reduction when receiving one more correct packet after receiving $n - 1$ packets successfully is:

$$\Delta D_{j,n}(\mathbf{A}, \mathbf{F}) = D_{j,n-1}(\mathbf{A}, \mathbf{F}) - D_{j,n}(\mathbf{A}, \mathbf{F}). \quad (2)$$

Let p_c be the packet loss rate (PLR) along the next hop and $P_c(N, n)$ be the probability that the receiver receives at least n packets successfully when the transmitter sends N packets. We have:

$$P_c(N, n) = \sum_{\alpha=n}^{N} \left(\begin{array}{c} N \\ \alpha \end{array} \right) (1 - p_c)^{\alpha} \, (p_c)^{N-\alpha}. \quad (3)$$

The expected distortion of transmitting N packets of stream j using segment assignment $\mathbf{A}$ and RS source symbols assignment $\mathbf{F}$ can be expressed as:

$$ED_j(\mathbf{A}, \mathbf{F}) = D_{j,0}(\mathbf{A}, \mathbf{F}) - \sum_{n=1}^{N} P_c(N, n) \Delta D_{j,n}(\mathbf{A}, \mathbf{F}). \quad (4)$$

We formulate the segment-based multi-stream aggregation problem as to minimize the maximal distortion among all streams by determining the assignment on segment, **A**, and source symbols **F**:

$$\min_{\mathbf{A},\mathbf{F}}(\max_{j}\ w_j \cdot ED_j(\mathbf{A},\mathbf{F})) \tag{5}$$

subject to

$$\begin{cases} \sum_{j=1}^{J} a_{j,l} \le 1, a_{j,l} \in \{0,1\}, \forall l; \\ \sum_{j=1}^{J}\sum_{l=1}^{L} a_{j,l} = L; \\ \sum_{i=1}^{N} f_{i,l} \le 1, f_{i,l} \in \{0,1\}, \forall l; \\ \sum_{i=1}^{N} i f_{il} \le \sum_{i=1}^{N} i f_{i,l+1}, \text{ if} \exists j, l\ s.t.\ a_{j,l} = a_{j,l+1} = 1; \end{cases}$$

The first two constraints restrict the segment assignment and the last two constraints enforce unequal error protection. Here, w_j is the quality weighting factor. By setting different w_j values for different video streams, we can achieve differentiated quality among the aggregated video streams.

3.2. Proposed Algorithm

Owing to different importance of base layer and FGS layer, we develop different algorithms for the two layers.

3.2.1. Base Layer

A strong error protection is applied to the base-layer source symbols to ensure baseline video quality. We aggregate all streams' base layer together and construct N_B^S source packets. It has been shown that if we can keep packet loss rate after FEC decoding lower than a threshold, PLR^B = 10^{-3}, the distortion caused by the channel error is negligible for MPEG-4 codec [10]. We can find the minimal number of packets, N_B^P, to achieve the desired PLR threshold: $P_c(N_B^S + N_B^P, N_B^S) \ge (1 - \text{PLR}^B)^{N_B^S}$. The overall number of packets used in base layer is $N_B = N_B^S + N_B^P$, and the rest of bandwidth, $N_F = N - N_B$, will be allocated for FGS layer.

3.2.2. FGS Layer

Denote $Rate_j^{max}$ be the overall FGS rate for stream j received by the video stream combiner. The maximal segments assigned to stream j will be $L_j^{max} = \lceil Rate_j^{max}/N_F \rceil$. Given a pre-determined L_j, the RS configuration for each segment and the corresponding weighted expected distortion, $S_j(L_j) = w_j \cdot ED_j(\mathbf{A},\mathbf{F})$, can be calculated using fast local search algorithm [11]. Note that $S_j(L_j)$ is a decreasing function of L_j. We propose a fast algorithm to obtain the min-max solution, consisting of three steps.

Step 1: Initialization. We start to solve this problem by considering an error-free channel. Under this condition, the function $S_j(L_j)$ become the original R-D function. We will first perform bi-section search on the original R-D functions. Then, we round the solution to the nearest feasible integer solution and use it as the initial points $\{L_j^{(0)}\}$.

Step 2: Coarse Search. At each iteration, we find the best searching direction towards the optimal solution: we exchange one segment for the stream having the largest expected distortion with one segment for the stream having smallest expected distortion. If the number of segment assigned to the stream with maximal distortion has reached L_j^{max}, we will exclude stream j at the next iteration. We repeat the above procedures until the maximal distortion can no longer be improved any more.

Step 3: Refinement. A round of refinement is performed based on the results obtained from previous coarse search. At each iteration, we perform $J - 1$ trials by exchanging one segment for the stream which has the largest expected distortion with one segment for each of the other streams. If the maximal distortion in neither trials is smaller than the one in the previous iteration, the refinement step is completed at the $\bar{k}^{th}$ iteration and $\{L_j^{(\bar{k})}\}$ is the optimal segment assignment. Otherwise, we move to the next iteration and perform coarse search again.

The optimal solution may not be unique and there exist several sets of solutions with the same maximal distortion. However, we can prove that there is no better solution than the one achieved by the proposed algorithm as follows: Let the set of optimal solution provided by the proposed solution is $\{L_j^{opt}\}$ and suppose stream $\bar{j}$ has the maximal distortion $S_{\bar{j}}(L_{\bar{j}}^{opt})$. According to Step 3 in the proposed algorithm, we have

$$S_j(L_j^{opt} - 1) \ge S_{\bar{j}}(L_{\bar{j}}^{opt}),\ \forall j. \tag{6}$$

Suppose $\{L_j^\star\}$ is a set of solution that can provide smaller maximal distortion. The overall number of segment in both sets of solutions should be equal to L. Suppose for some α, we have $L_\alpha^{opt} < L_\alpha^\star$, then there exists at least one β such that $L_\beta^{opt} > L_\beta^\star$. From (6),

$$S_\beta(L_\beta^\star) \ge S_\beta(L_\beta^{opt} - 1) \ge S_{\bar{j}}(L_{\bar{j}}^{opt}). \tag{7}$$

From (7), the maximal distortion achieved by $\{L_j^\star\}$ is not smaller than the one by $\{L_j^{opt}\}$, which contradicts the assumption that set $\{L_j^\star\}$ can provide smaller expected distortion.

4. EXPERIMENTAL RESULTS

We evaluate the performance of our proposed scheme with two alternatives. The first one is the traditional multicasting scheme with hop-by-hop FEC transcoding. In this scheme, since the resource allocation for each stream is not aware of co-transmitted streams, each stream has fixed and the same amount of bandwidth. The second alternative is the packet-based multi-stream error protection scheme shown in Figure 2(a) and the solution can be obtained via similar techniques of segment-based scheme discussed in Section 3.

The simulations are set up as follows. The network topology is shown in Figure 1, where there are 10 users and 3 video stream combiners. The video refreshing rate is 30 frames per second. The base layer is generated by MPEG-4 encoder with a fixed quantization step of 30 and the GOP pattern leading by one I frame followed by 29 P frames. All frames of FGS layer have up to six bit planes. Each user will send one video sequence with 90 frames through the uplink and receive the merged 90 frames from all other users. User 1 $\sim$ 10 send video sequence, *Akiyo*, *Carphone*, *Claire*, *Container*,

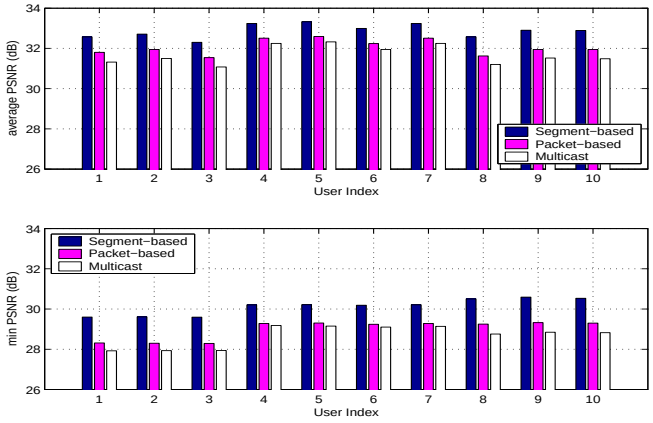

Fig. 3. Average and minimal PSNR for each user

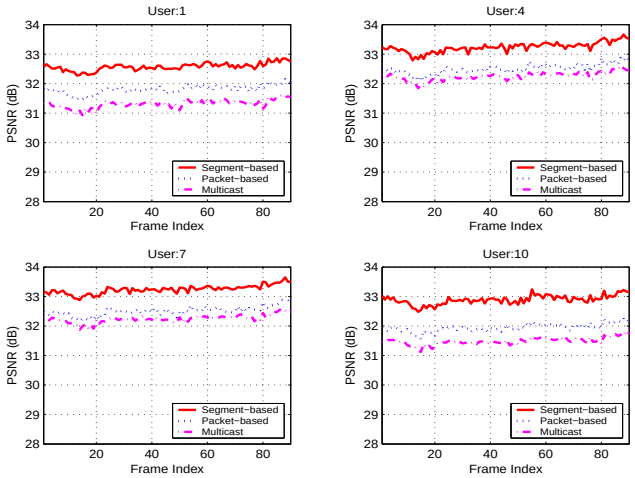

Fig. 4. Frame-by-frame average PSNR for merged stream

Foreman, *Grandmother*, *Hall objects*, *Mother and daughter*, *MPEG4 news*, and *Salesman*, respectively. The packet size is L=128 bytes. The bandwidth of uplink and downlink for all users are 930 Kbps and 5.55 Mbps, respectively. Link C_{12}, C_{23}, C_{21}, C_{32} have bandwidth 2.16 Mbps, 4.31 Mbps, 4.31 Mbps, and 2.16 Mbps, respectively. Without loss of generality, we set packet loss rate of all links as 0.1. We examine the case of consistent quality with all w_js' having the same value in problem (5). We repeat the experiments 100 times and average the received PSNR for each user.

Two performance criteria are used to evaluate these three schemes. The first one is to measure the overall system efficiency by averaging PSNR over all 90 video frames of the aggregated stream received by each user. To illustrate results for this criteria more clearly, we also show the frame-by-frame average PSNR of the aggregated stream received by User 1, 4, 7, and 10 in Figure 4, respectively. The second one is to measure the minimal PSNR received among all incoming streams for each user, which is the objective of the formulated problem (5). As we can see from Figure 3, the two schemes which explore multiuser diversity have higher values than the traditional multicasting schemes in both criteria. The packet-based scheme outperforms the multicating scheme 0.99dB $\sim$ 1.40dB and 1.03dB $\sim$ 1.75dB for average PSNR and minimal PSNR, respectively. If we compare the segment-based scheme to the packet-based scheme, the segment-based scheme outperforms by 0.73dB $\sim$ 0.95dB and 0.91dB $\sim$ 1.31dB for average PSNR and minimal PSNR, respectively. This is because segment-based scheme can provide higher effective bandwidth after optimal FEC protection to carry more source bits compared to the packet-based scheme.

5. CONCLUSIONS

In this paper, we propose a distributed multi-point video conferencing system over error-prone channels and a novel error protection scheme for this system. By aggregating multiple streams and performing joint error protection, the proposed error protection schemes can outperform the existing multi-hop FEC multicasting scheme by more than 1 dB. With exploration of higher effective bandwidth, the segment-based error protection scheme can have up to 0.95dB performance gain compared to the packet-based error protection scheme. Thus, the proposed system and error protection scheme compose a promising framework to support multi-point video conferencing in error-prone environment.

6. REFERENCES

[1] M.-T. Sun, A. C. Loui, and T.-C. Chen, "A coded-domain video combiner for multipoint continuous presence video conferencing", *IEEE CSVT*, pp.855-863, Dec. 1997.

[2] K.-T. Fung, Y.-L. Chan, and W.-C. Siu, "Low-complexity and high-quality frame-skipping transcoder for continuous presence multipoint video conferencing", *IEEE Trans. on Multimedia*, pp.31-46, 2004.

[3] S. McCanne, M. Vetterli, and V. Jacobson, "Low-complexity video coding for receiver-driven layered multicast", *IEEE JSAC*, pp.983-1001, Jun. 1997.

[4] W.-T Tan and A. Zakhor, "Video multicast using layered FEC and scalable compression", *IEEE CSVT*, pp.373-386,Mar. 2001.

[5] Y. Shan, I. V. Bajic, S. Kalyanaraman, and J. W. Woods, "Overlay multi-hop FEC scheme for video streaming over peer-to-peer networks," *IEEE ICIP*, 2004.

[6] H. Radha and M. Wu, "Overlay and peer-to-peer multimedia multicast with network-embedded FEC," *IEEE ICIP*, 2004.

[7] H.M. Radha, M. v. d. Schaar, and Y. Chen, "The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP", *IEEE Trans. on Multimedia*, pp.53-68, Mar. 2001.

[8] X.M. Zhang, A. Vetro, Y.Q. Shi, and H. Sun, "Constant quality constrained rate allocation for FGS-coded videos", *IEEE CSVT*, pp.121-130, Feb. 2003.

[9] R. Puri, K.-W Lee, K. Ramchandran, and V. Bharghavan, "An integrated source transcoding and congestion control paragidm for video streaming in the internet", *IEEE Trans. on Multimedia*, pp.18-32, Mar. 2001.

[10] S. Gringeri, R. Egorov, K. Shuaib, A. Lewis, and B. Basch, "Robust compression and transmission of MPEG-4 video," *ACM Inter. Conf. on Multimedia*, pp.113-120, Jun. 2000.

[11] V. M. Stankovic, R. Hamzaoui, and Z. Xiong, "Real-time error protection of embedded codes for packet erasure and fading channels", *IEEE CSVT*, pp.1064-1072, Aug. 2004.