# JOINT UPLINK AND DOWNLINK OPTIMIZATION FOR VIDEO CONFERENCING OVER WIRELESS LAN

*Guan-Ming Su, Zhu Han, Min Wu, and K. J. Ray Liu*

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

## ABSTRACT

In this paper, a real-time video conferencing framework is proposed for multiple conferencing pairs by joint considering the uplink and downlink conditions within IEEE802.11 networks. We formulate this system as to minimize the maximal end-to-end expected distortion received by all users by selecting the PHY modes and transmission time. Compared with the strategy of individually optimizing uplink and downlink, the proposed framework outperforms by 3.67 $\sim$ 8.65 dB for the average received PSNR among all users.

## 1. INTRODUCTION

Wireless local area network (WLAN) has become ubiquitous owing to the great demands of pervasive mobile applications. It has been shown that a cross-layer design can improve the overall performance for a wireless communication system [1, 2]. Recently, joint multiuser video source coding and communication has drawn attention since it can leverage the diversity of video contents and radio resources to achieve more desired quality [3, 4]. In a multiuser system, one important issue is how to efficiently allocate system resources to achieve fairness among all users [5].

In the scenario that a mobile user transmits a video stream to the other mobile user like video conferencing, the video stream is transmitted through at least two paths, namely, an uplink to a base station and a downlink from a base station. This causes some problems, e.g., the maximal video quality transmitted through downlink has been restricted by the packets received by the base station through uplink. Since a video stream may experience different channel conditions in the uplink and downlink, a strategy adopting optimal uplink resource allocation and then optimal downlink allocation individually is not efficient [6, 7]. In this paper, we propose a real-time video conferencing framework for multiple video conferencing pairs with joint uplink and downlink optimization. The simulation results demonstrate that the joint optimization scheme can achieve fairer quality among all users and higher overall video qualities than those of the individual uplink/downlink optimization schemes.

This paper is organized as follows. The system architecture for joint uplink and downlink video conferencing is described in Section 2. In Section 3, we formulate a minimax optimization problem under the system resource constraints. The simulation results are presented in Section 4 and conclusions are drawn in Section 5.

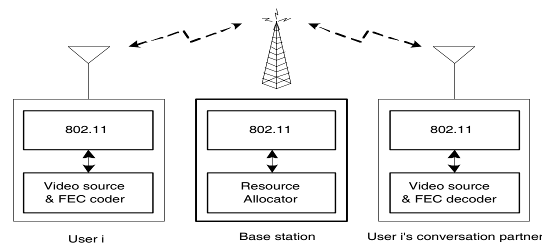


**Fig. 1**. System block diagram

## 2. SYSTEM DESCRIPTION

In this section, we first present the proposed system and then address each module used in the system.

### 2.1. System Framework

We consider the case of a single cell for simplicity. For video conferencing in multiple cells, as long as the coherent time of the channel condition is much larger than the propagation delays induced by the networks, the proposed framework can be employed in a similar way. Also, we only consider point-to-point conferencing, i.e., if there are $N$ users in this video conferencing system, then there are $N/2$ conversation pairs. Figure 1 shows the proposed framework for video conferencing system consisting of three main modules: video source and channel coding, IEEE 802.11 transmission component, and resource allocator. We divide the time line into $F$ slots per second, where $F$ is the video frame rate. At the beginning of each time slot, each user's video source coder encodes video in real time and the channel coder protects the source packets using forward error correction (FEC) coding. The mobile users send video's rate-distortion (R-D) information to the server. The resource allocator located at server gathers all users' R-D and channel information, and then performs resource allocation for all users. Subsequently, each user sends video packets to the server and the server will forward those packets to his/her conversation pair according to the allocated system resources.

### 2.2. IEEE 802.11 Wireless LAN

In this paper, we use IEEE 802.11a as an example. Other standards can be applied in a similar way. The IEEE 802.11a Physical (PHY) layer provides eight PHY modes with different modulation schemes and different convolutional coding rates and can offer data rate of 6, 9, 12, 18, 24, 36, 48, and 54 Mbps. A major task of the proposed system is to select PHY modes of both uplink and downlink for each user to transmit packets. Let $P^U_{max}$ and $P^D_{max}$ be the maximal available transmitted power for uplink and downlink, respectively; and $G^U_i$ and $G^D_i$ the uplink and downlink channel gain from user $i$ to his/her conversation partner. Thus, the maximal signal to noise ratio (SNR) for uplink and downlink are $\Gamma^U_i = P^U_{max}G^U_i/\sigma^2$ and $\Gamma^D_i = P^D_{max}G^D_i/\sigma^2$, respectively, where $\sigma^2$ is the thermal noise level that is assumed to be the same at all mobile receivers and base station. If user $i$ selects the uplink and downlink PHY mode as $m_i$ and $n_i$, respectively, the bit error rate (BER) for uplink and downlink can be approximated as a function of PHY mode and SNR level: $\text{BER}^U_{i,m_i} = P_{m_i}(\Gamma^U_i)$ and $\text{BER}^D_{i,n_i} = P_{n_i}(\Gamma^D_i)$, respectively, where the function $P(\cdot)$ can be obtained in [8, 9]. Then, the probability that a packet is received successfully for uplink and downlink can be calculated as $p^U_{i,m_i} = (1 - \text{BER}^U_{i,m_i})^L$ and $p^D_{i,n_i} = (1 - \text{BER}^D_{i,n_i})^L$, respectively, where $L$ is the number of bits in a packet. With a fixed packet size, $p^U_{i,m_i}$ and $p^D_{i,n_i}$ are functions of the channel gains and PHY modes.

IEEE 802.11 medium access control (MAC) protocol supports two kinds of access methods: distributed coordination function (DCF) and point coordination function (PCF). The DCF is the basic access mechanism using carrier sense multiple access with collision avoidance (CSMA/CA) and must be implemented in all stations. In contrast, the PCF is optional and based on polling controlled by a point coordinator. In both mechanisms, only one user occupies all the bandwidth at each time, which is similar to time division multiple access (TDMA) technology. In this paper, we use a simple TDMA scheme to allocate time slot for each user by a centralized resource allocator. Either PCF or enhanced DCF such as [10] can satisfy this requirement.

### 2.3. Error-Resilient Scalable Source Coding

The MPEG-4 Fine Granularity Scalability (FGS) coding [11] is a two-layer scheme consisting of a non-scalable base layer and a highly scalable FGS layer of which any truncated bitstream corresponding to each frame can be decoded. The more FGS bits the decoder receives and decodes, the higher the video quality is. Furthermore, the R-D curve of FGS at the frame level can be well approximated as a piecewise linear line [12, 13]. Thus, we adopt FGS codec in this work to allow convenient adjustment in video rate and quality.

In IEEE 802.11 MAC, a packet sent from uplink will be dropped and not be forwarded to the next path if errors are detected. A packet loss will cause error propagation for the base layer and make the successfully received FGS packets following a corrupted FGS packet useless due to the strong decoding dependency of the FGS layer bitstream. Applying application layer FEC, such as systematic Reed-Solomon (RS) codes, across packets has been shown as an effective

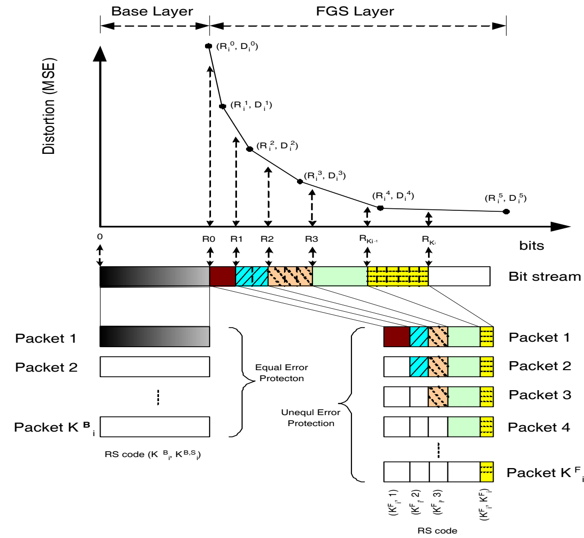


**Fig. 2**. Multiple description error protection scheme

solution to alleviate the problem caused by packet loss [1]. An RS$(K_i, k)$ encoder will generate $K_i - k$ parity symbols for $k$ source symbols, and a corresponding RS decoder can recover the original $k$ source symbols if it receives at least $k$ out of $K_i$ symbols successfully when the locations of the erased symbols are known. Since MPEG-4 FGS codec is a two-layer scheme, we adopt two different strategies for each layer as shown in Figure 2. For the non-scalable base layer, we apply a strong equal error protection strategy to provide the baseline video quality. To remove the strong decoding dependency of the FGS layer bitstream, we adopt multiple description forward error correction (MD-FEC) framework [14] with fast local search [15], since MD-FEC allocates an embedded bitstream into $K_i$ descriptions (packets) such that the more descriptions a receiver receives successfully, the better reconstructed quality the decoder can get.

Assuming all packets of base layer are received successfully, the end-to-end expected distortion model using MD-FEC can be represented as follows:

$$E\left[D_i(m_i, n_i, K_i)\right] = D^B_i - \sum_{k=1}^{K_i} p_i(m_i, n_i, K_i, k)\Delta D_i(K_i, k)$$

where $D^B_i$ is the distortion after receiving all base layer packets successfully, $K_i$ is the number of packets sent from user $i$, $\Delta D_i(K_i, k)$ is the distortion reduction if user $i$'s conversation partner receives one more correct packet after having $k-1$ uncorrupted packets, and $p_i(m_i, n_i, K_i, k)$ is the probability that the receiver receives at least $k$ packets successfully when transmitter sends $K_i$ packets. We have

$$p_i(m_i, n_i, K_i, k) = \sum_{\alpha=k}^{K_i} p^U_{i,m_i}(K_i, \alpha)p^D_{i,n_i}(\alpha, k), \qquad (1)$$

where $p^U_{i,m_i}(K_i, \alpha)$ is the probability that the server receives $\alpha$ packets successfully when user $i$ sends $K_i$ packets

$$p^U_{i,m_i}(K_i, \alpha) \triangleq \begin{pmatrix} K_i \\ \alpha \end{pmatrix} \left(1 - p^U_{i,m_i}\right)^{K_i - \alpha} \left(p^U_{i,m_i}\right)^{\alpha},$$

and $p^D_{i,n_i}(\alpha, k)$ is the probability that the receiver receives at least $k$ packets successfully when the server sends $\alpha$ packets

$$p^D_{i,n_i}(\alpha, k) \triangleq \sum_{\beta=k}^{\alpha} \left( \begin{array}{c} \alpha \\ \beta \end{array} \right) \left(1 - p^D_{i,n_i}\right)^{\alpha-\beta} \left(p^D_{i,n_i}\right)^{\beta}.$$

## 3. JOINT UPLINK-DOWNLINK OPTIMIZATION

In this section, we first formulate the system as a minimax optimization problem and then present the proposed algorithm to allocate resources for the base and FGS layers.

### 3.1. Problem Formulation

Since we adopt TDMA scheme, there is only one user who can send data at any moment in one cell. Let $T$ be the time duration to refresh a video frame ($T = 1/F$) and $t_i$ the whole assigned time duration for user $i$ to send a video frame to his/her conversation partner through both uplink and downlink. To facilitate our discussion, we denote $T^{max}_x$ as the time duration if user $i$ selects PHY mode $x$ to transmit a packet in a single path. Thus, if user $i$ selects PHY mode for uplink and downlink as $m_i$ and $n_i$ and sends $K_i$ packets from sender to server, the expected number of successful packets arrived at base station is $p^U_{i,m_i} K_i$, and the overall expected transmission time from the sender through a server to the receiver can be represented as:

$$t_i(m_i, n_i, K_i) = K_i(T^{max}_{m_i} + p^U_{i,m_i} T^{max}_{n_i}). \qquad (2)$$

We formulate the video conferencing system as an optimization problem to choose each user's transmission mode to minimize the maximum of all users' expected distortion, subject to the maximal available transmission time constraint. More specifically, we solve for the PHY mode for uplink $m_i$, the PHY mode for downlink $n_i$, and the number of packets to transmit $K_i$ in the minimax constrained optimization of

$$\min_i \max_{\{m_i,n_i,K_i\}} E\left[D_i(m_i, n_i, K_i)\right] \qquad (3)$$
$$\text{subject to } \sum_{i=1}^{N} t_i(m_i, n_i, K_i) \leq T.$$

Due to different properties between the base and FGS layer, we adopt different resource allocation strategies for them.

### 3.2. Base Layer

To carry all base layer rate $R^0_i$ of user $i$, we use $K^{B,S}_i = \lceil R^0_i / L \rceil$ source packets. We need to determine the uplink and downlink PHY mode $(m_i, n_i)$ and the number of parity packets, $K^{B,P}_i$, such that the resulted transmission time for the base layer is the shortest and the end-to-end BER is kept lower than a threshold (in this paper, we set the threshold $\text{BER}^B = 10^{-6}$). This can be attained in three steps: first examine the smallest required number of parity packets for each $(m_i, n_i)$ to achieve $p_i(m_i, n_i, K^{B,S}_i + K^{B,P}_i, K^{B,S}_i) \geq (1 - \text{BER}^B)$ using (1); then calculate the corresponding transmission time $t^B_i(m_i, n_i)$ using (2); and finally find the optimal setting with the shortest transmission time

$$(\hat{m}_i, \hat{n}_i) = \arg \min_{\{m_i,n_i\}} t^B_i(m_i, n_i). \qquad (4)$$

Denote $t^B_i$ as the transmission time using mode $(\hat{m}_i, \hat{n}_i)$. Thus, the overall transmission time for all users is $T^B = \sum_{i=1}^{N} t^B_i$ and the remaining transmission time for FGS layer is $T^F = T - T^B$.

### 3.3. FGS Layer

We propose a two-step algorithm to determine the optimal transmission modes for FGS layer by first obtaining a one-to-one mapping function between transmission time and distortion (T-D) for each user and then using bisection search in all T-D functions to obtain the optimal solutions. The T-D function for each user can be obtained by first finding a set of efficient transmission modes. A transmission mode $(m_i,\ n_i,\ K^F_i\ )$ is *efficient* if $E[D_i(m'_i,n'_i,K^{F'}_i)] < E[D_i(m_i,n_i,\ K^F_i)]$, we have $t_i(m'_i,n'_i,K^{F'}_i) > t_i(\ m_i,\ n_i,K^F_i)$, for all other modes $(m'_i,\ n'_i,\ K^{F'}_i)$. We can collect all efficient transmission modes $\{(m_i, n_i, K^F_i)\}$ as set $\mathcal{S}_i$ and the corresponding transmission time $\{t_i(\ m_i,\ n_i,K^F_i)\}$ as set $\mathcal{T}_i$. Let $\{t_{i,k}\}$ be the transmission time sorted in an increasing order in $\mathcal{T}_i$ and the corresponding expected distortion for each transmission time $t_{i,k}$ can be obtained. Bring all $\{t_{i,k}\}$ and the corresponding expected distortion together, we have a time-distortion function $E[\tilde{D}_i[t_{i,k}]]$ for user $i$. After obtaining all T-D functions for all users, the problem (3) can be reformulated as

$$\min_i \max_{\{t_{i,k}\}} E\left[\tilde{D}_i[t_{i,k}]\right] \text{ s.t. } \sum_{i=1}^{N} t_{i,k} \leq T^F. \qquad (5)$$

Since all T-D functions are monotonically decreasing by the definition of efficient transmission mode, the optimal solution for problem (5) can be obtained using bisection search. After determining the optimal $t_{i,k}$ for all users, the corresponding transmission mode $(m_i, n_i, K^F_i)$ can be obtained from $\mathcal{S}_i$ for each user $i$.

## 4. SIMULATION RESULTS

The simulations are set up as follows. The noise power is $10^{-10}$ Watts and the maximal transmitted power for mobile user and server is 40 mW. The path loss factor is 2.5. Packet length $L$ is set to 512 bytes. The video refresh rate is 30 frames per second and thus $T = 33.33$ ms. We simulate a 4-user system in which user 1,2 and user 3,4 form two conversation pairs. User 1 to 4 receive 90-frame QCIF (176 $\times$ 144) video sequence, *Akiyo*, *carphone*, *Claire*, and *foreman*, respectively. The base layer is generated by MPEG-4 encoder with a fixed quantization step of 30 and the GOP pattern is 29 P frames after one I frame. All frames of FGS layer have up to six bit planes. For each setting, we repeat the experiments 100 times to calculate the average PSNR.

We compare our proposed scheme with a scheme that first allocates the optimal configuration based on only the uplink channel information and then optimizes the downlink configuration based on the packets received successfully by the base station. The total available transmission time budget for uplink is equal to the one of downlink for

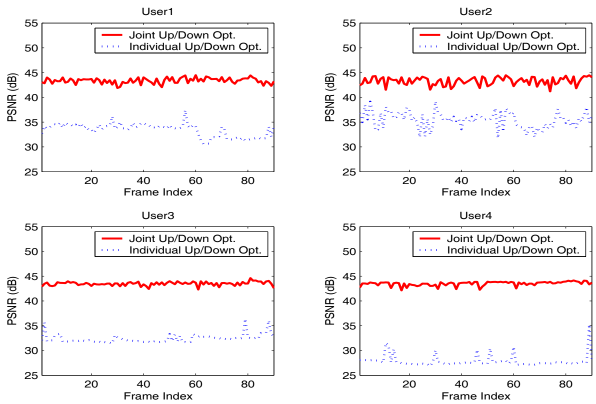


**Fig. 3**. Frame-by-frame PSNR result for User 1 to User 4.

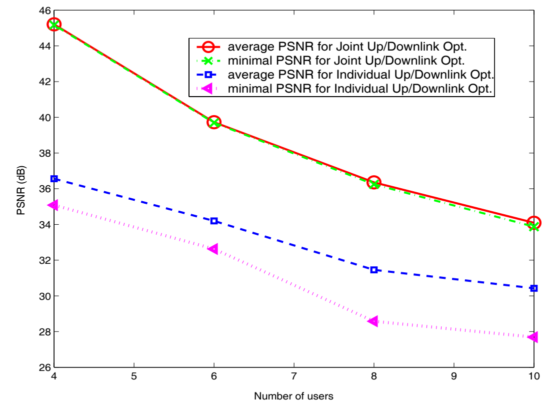


**Fig. 4**. PSNR results for different number of users.

the comparison scheme. Figure 3 shows the frame-by-frame expected PSNR results when user 1$\sim$4 are located at 91m, 67m, 71m, and 20m away from the base station, respectively. As we can see, the proposed scheme has higher average PSNR, fairer video quality among all users, and less quality fluctuation along each received video sequence than the individual uplink and downlink optimization scheme.

We repeat the simulation 20 times by randomly setting the distances between users and the base station within 20 m to 100 m. Figure 4 shows the average PSNR and minimal PSNR among all users for different number of users in the system. As we can see, the proposed joint uplink and downlink optimization scheme outperforms the individual uplink and downlink optimization scheme 3.67 $\sim$ 8.65 dB for the average PSNR and 6.19 $\sim$ 10.09 dB for the minimal PSNR. Comparing the gap between the minimal and average PSNR for these two schemes, the proposed scheme has less quality deviation among all users than the one of the individual uplink and downlink scheme, which shows the fairness achieved by the proposed scheme.

## 5. CONCLUSIONS

In this paper, we have constructed a video conferencing framework for multiple conversation pairs by joint considering the uplink and downlink within IEEE802.11 networks. We formulate the problem as a minimax problem under the system resource constraints. The simulation results demonstrate that the proposed joint uplink/downlink optimization scheme outperforms the individual optimization strategy not only in the average PSNR but also in the fairness among users. Thus, the proposed scheme is a promising framework supporting real-time multiuser video conferencing, which is the next driving demand for the wireless services.

## 6. REFERENCES

[1] M. van der Schaar, S. Krishnamachari, S. Choi, and X. Xu, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs," *IEEE JSAC*, pp.1752-1763, Dec. 2003.

[2] C. Luna, Y. Eisenberg, R. Berry, T. Pappas, A. Katsaggelos, "Joint source coding and data rate adaptation for energy efficient wireless video streaming," *IEEE JSAC,* Dec. 2003

[3] G.-M. Su, Z. Han, A. Kwasinski, M. Wu, K.J. R. Liu, and N. Farvardin, "Distortion management of real-time MPEG-4 FGS video over downlink multicode CDMA networks," *IEEE ICC,* vol.5, pp.3071-3075, Jun. 2004.

[4] X. Lu, Y. Wang, E. Erkip, and D. Goodman, "Power optimization of source encoding and radio transmission in multiuser CDMA systems," *IEEE ICC*, Jun. 2004.

[5] Z. Ji, Q. Zhang, W. Zhu, J. Lu, and Y.-Q. Zhang, "Video broadcasting over MIMO-OFDM systems," *IEEE ISCAS*, pp.844-847, May 2003.

[6] D. G. Jeong and W. S. Jeon, "CDMA/TDD system for wireless multimedia services with traffic unbalance between uplink and downlink," *IEEE JSAC,* pp.939-946, May 1999.

[7] H.-Y. Wei, C.-C. Chiang, and Y.-D. Lin, "Co-DRR: an integrated uplink and downlink scheduler for bandwidth management over WLANs," *IEEE ISCC,* pp.1415-1420, 2003.

[8] L. Hanzo, S.X. Ng, T. Keller, and W.T. Webb, *Single and multicarrier quadrature amplitude modulation*, Wiley, 2004.

[9] J. G. Proakis, *Digital communication*, McGraw-Hill, 1995.

[10] IEEE P802.11e/Draft 6.0, "Draft Amendment to IEEE Std 802.11, 1999 Edition, Medium access control enhancements for quality of service," Nov. 2003.

[11] H.M. Radha, M. van der Schaar, and Y. Chen, "The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP," *IEEE Trans. on Multimedia*, Mar. 2001.

[12] 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.

[13] L. Zhao, J. Kim and C.-C. J. Kuo, "MPEG-4 FGS video streaming with constant-quality rate control and differentiated forwarding," *SPIE VCIP, 2002.*

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

[15] 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.