Joint Uplink and Downlink Optimization for Real-Time Multiuser Video Streaming Over WLANs

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

***Abstract*—In this paper, a network-aware and source-aware video streaming system is proposed to support interactive multiuser communications within single-cell and multicell IEEE 802.11 networks. Unlike the traditional streaming video services, the strict delay constraints of an interactive video streaming system pose more challenges. These challenges include the heterogeneity of uplink and downlink channel conditions experienced by different users, the multiuser resource allocation of limited radio resources, the incorporation of the cross-layer design, and the diversity of content complexities exhibited by different video sequences. With the awareness of video content and network resources, the proposed system integrates cross-layer error protection mechanism and performs dynamic resource allocation across multiple users. We formulate the proposed system as to minimize the maximal end-to-end expected distortion received by all users, subject to maximal transmission power and delay constraints. To reduce the high dimensionality of the search space, fast multiuser algorithms are proposed to find the near-optimal solutions. Compared to the strategy without dynamically and jointly allocating bandwidth resource for uplinks and downlinks, the proposed framework outperforms by 2.18$\sim$7.95 dB in terms of the average received PSNR of all users and by 3.82$\sim$11.50 dB in terms of the lowest received PSNR among all users. Furthermore, the proposed scheme can provide more uniform video quality for all users and lower quality fluctuation for each received video sequence.**

***Index Terms*—Cross-layer design, joint uplink and downlink optimization, multiuser video communication, network-aware, wireless local area networks.**

I. Introduction

With the rapid advance of wireless local area network (WLAN) technology, WLAN has become ubiquitous as a broadband wireless access medium. One of the promising services supported by WLAN is interactive video streaming, whereby a pair of mobile users at different locations can exchange video streams with each other in real time. Besides the real time requirement, a wireless system providing interactive video streaming faces more challenges than the typical video-on-demand service. For instance, in each conversation session, there are two video streams being exchanged between a conversation pair; each video stream is transmitted through at least two paths, namely, an uplink to an access point and a downlink from the access point. The transmitted packets of each video stream experience different channel conditions in both links. Because the radio bandwidth resources are limited for different users' transmissions over uplink and downlink and the channel conditions change over time, dynamically allocating the limited network resources to all users can significantly improve the end-to-end quality. Moreover, various video programs exhibit different content complexities and require different amount of bandwidth to achieve similar video quality. To provide satisfactory video quality to all users, a multiuser wireless video streaming system should be aware of video source and dynamically exploit multiuser diversity through cross-layer design. In this paper, we address the above issues and propose an interactive video streaming framework to support multiple conversation pairs over WLANs.

For a wireless system with limited bandwidth resources, it is critical to determine the amount of bandwidth allocated to uplink and downlink to achieve high spectrum utilization and system service objectives. A static strategy is to allocate equal bandwidth to each link and perform optimal uplink resource allocation and optimal downlink resource allocation individually. As this simple strategy of allocating equal bandwidth is inefficient due to uneven load in both links, several works adopting unequal bandwidth assignment have been proposed. A scheme was proposed in [1] to address the unbalanced capacity and asymmetric channel bandwidth usage problem. Several call admission control schemes were presented in [2]–[4] to explore the asymmetric traffic load in both links. A scheduler to simultaneously control generic data traffic in both uplink and downlink for IEEE 802.11a networks was proposed in [5]. Bandwidth resource allocation for transmitting video over WLAN in real time is more challenging than for transmitting generic data since compressed video bitstreams exhibit different characteristics from generic data [6], [7]. For example, compressed video bitstreams have decoding dependency on the previous coded bitstreams due to the spatial and temporal prediction. Transmitting video streams in real time has a strict delay constraint (below 200 ms [8]) that belated video data is useless for its corresponding frame and will cause error propagation for the video

Manuscript received September 4, 2006; revised May 4, 2007. This work was supported in part by the U.S. National Science Foundation under Award CCR-0133704. Some preliminary results of this work were presented at the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. John Apostolopoulos.

G.-M. Su is with Marvell Semiconductor, Santa Clara, CA 95054 USA (e-mail: guanmingsu@yahoo.com).

Z. Han is with the Department of Electrical and Computer Engineering, Boise State University, Boise, ID 83725 USA (e-mail: ZhuHan@boisestate.edu).

M. Wu and K. J. R. Liu are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: minwu@umd.edu; kjrliu@umd.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2007.901518

data that are predictively encoded using that frame as reference. Besides, the encoded video rate highly fluctuates from frame to frame, which complicates the bandwidth allocation. This motivates us to investigate dynamical bandwidth allocation to all video streams in both links.

In this paper, we propose a network-aware and source-aware framework to support interactive video streaming. The proposed framework explores the diversity of content complexity exhibited by different video sequences and the heterogeneity of uplink and downlink channel conditions experienced by different users. With vertical integration of different communication layers, a cross-layer error protection mechanism is proposed for graceful quality degradations. With the awareness of current network channel conditions, in the physical layer, we employ adaptive modulation schemes and convolutional coding rates for each video stream. With the awareness of current network traffic conditions, in the medium access control (MAC) layer, the system determines the optimal time proportion for each transmitter to send video packets. In the application layer, the proposed framework performs joint optimization for video source coding and application-layer forward error coding (FEC) to achieve optimal end-to-end video quality. To provide consistent perceptual quality to all participated users, we formulate the proposed system as to minimize the maximal end-to-end expected distortion among all end users by selecting the parameters in each communication layer, subject to the transmission power and delay constraints. Since searching the optimal setting in different layers is a combinatorial problem, which is NP hard, we develop fast algorithms to find the transmission configurations for near-optimal solutions.

By considering the cross-layer parameters, such as source rate, channel coding rate, and modulation, the proposed algorithm first converts the rate and distortion (R-D) function into the expected transmission time to the expected distortion (T-D) function. By doing so, the traditional R-D function in single-user system evolves into resource-distortion function in the multiuser scenario. Subject to the limitations of radio resources, a fast near-optimal algorithm is then proposed to allocate resources to all users. Compared to the strategy without dynamically distributing bandwidth for uplinks and downlinks, the proposed framework in a single cell outperforms by about 2$\sim$8 dB for the average received peak signal-to-noise ratio (PSNR) of all users and by about 4$\sim$12 dB for the minimal PSNR among all users. In addition, the proposed scheme can provide more uniform video quality among all users and lower quality fluctuation along each received video sequence. We also extend the proposed algorithm to the multicell case, where we jointly optimize uplink and downlink transmissions. The proposed scheme outperforms the sequential uplink-then-downlink optimization scheme by about 3$\sim$7 dB for the average PSNR and about 5$\sim$11 dB for the minimal PSNR.

This paper is organized as follows. We first review the prior work in Section II. The architecture for the proposed video streaming system is described in Section III. In Section IV, we formulate the streaming system in a single cell as a min-max optimization problem under system resource constraints, and propose a fast algorithm to find the near-optimal solution. In Section V, we consider a video streaming system in a multicell scenario, which supports both intra-cell and inter-cell conversations, and extend the proposed single-cell algorithm to the multicell scenario. Simulation results are presented in Section VI and conclusions are drawn in Section VII.

II. Prior Work

In the video communication literature, systems transmitting a single video program through wireless channels have been widely studied [9], [10]. Several error resilient algorithms from source coding's perspective, such as layer coding, multiple description coding, and robust entropy coding, have been proposed to overcome the error propagation owing to the fragility of compressed bit stream. [11]. By jointly considering source/channel rate adaptation and power allocation [12]–[15], a system can provide better video quality than allocating source/channel resources separately without awareness of each other. When designing a streaming video system over WLANs, more issues need to be addressed, such as how to packetize a video stream and consider the resource limitation at various communication layers. To improve the effective throughput, a wireless video framework was proposed to utilize hybrid automatic repeat request (HARQ) with multiple descriptions in the application layer [16]. Recently, cross-layer design methodologies that jointly optimize the resource allocation in different communication layers have been shown as an effective approach to improve the overall system performance [17], [18]. By adaptively utilizing the retry limit of the MAC layer with priority queueing, a wireless video scheme exploring unequal importance of video bitstreams was proposed in [19]. With cross-layer system integration, including application layer FEC, MAC retransmission, and adaptive packet size selection, a cross-layer error protection scheme was proposed to transmit video over IEEE 802.11a network [20].

Systems providing services to multiple users, however, have more challenges than systems supporting single user. We should consider the heterogeneous channel conditions experienced by different users, admission delays, and interference from external users on MAC; and distribute system resources to each video stream with sufficient error protection. Besides the heterogeneity of channel conditions in a multiuser wireless video system, all users send/receive heterogeneous video programs simultaneously. Such a system has another dimension of diversity to explore because of different content complexity of video scenes, namely, the rates to achieve the same perceptual quality are content dependent. With the awareness of co-existence of multiple streams, joint multiuser video source coding has been proposed to leverage the diversity of video content to achieve more desired quality [21]–[25]. The common service objectives include minimizing the overall users' distortion [26] or minimizing the maximal distortion among all users [27]. In this work, we consider the fairness issue among users who subscribe the same level of quality of service. The proposed system dynamically performs rate control for each video sequence and provides consistent video quality to all users.

In general, the channel conditions along the end-to-end transmission path of a video stream are heterogeneous [28]. To achieve the same bit error rate, the required level of error protection in each path may not be the same. The FEC transcoding

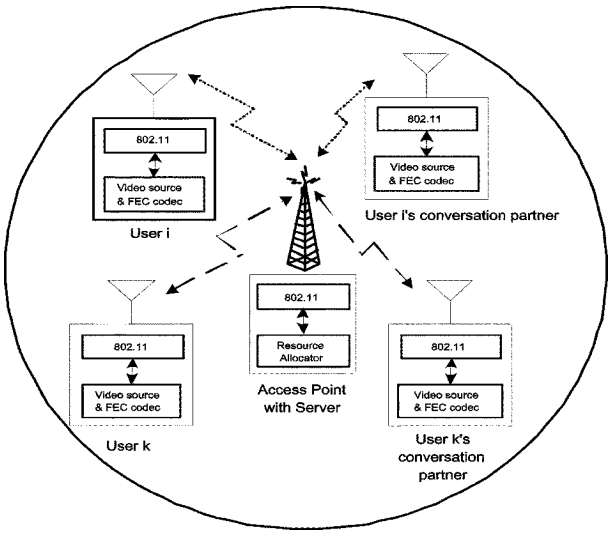


Fig. 1. System block diagram for single-cell case.

strategy, which dynamically applies optimal level of FEC in each intermediate path, can provide higher effective end-to-end throughput than traditional fixed FEC configuration throughout the end-to-end transmission path. Furthermore, adopting FEC transcoding in each intermediate transmission node can recover certain amount of corrupted packets transmitted through the preceding paths, thus preventing from further quality degradation accumulatively in the following transmission paths. With fixed allocated bandwidth and prior knowledge of the channel condition for each path, systems can be formulated as to maximize the overall throughput by determining which intermediate nodes should perform FEC transcoding for the unicast [29] and multicast scenarios [30]. In our proposed framework, the bandwidth of uplink and downlink is dynamically allocated according to the needs. The system performance can be further improved by jointly choosing the optimal channel coding rate and the bandwidth in both uplink and downlink, and performing FEC transcoding in the server located at the access point.

III. Proposed System Description

In this section, we present an overview of the key modules involved in our proposed video streaming system. Fig. 1 illustrates the single-cell scenario. We first introduce the video source subsystem and the corresponding R-D characteristics, and discuss the throughput and error protection mechanisms provided by the communication subsystem. We then discuss the cross-layer error protection schemes performed in resource allocator, and describe the coordination executed by the server to manage all video streams and the auxiliary control signals for resource allocation between mobile users and server.

A. Video Source Coding: MPEG-4 FGS

Unlike the traditional video codec employing hybrid motion compensation, scalable video coding provides flexibility and convenience in reaching the desired visual quality and/or the desired bit rate. MPEG-4 Fine Granularity Scalability (FGS) coding [31] is one of the scalable video codecs and encodes a video frame into a nonscalable base layer and a highly scalable FGS layer. The base layer is compressed using the nonscalable MPEG-4 codec at a low bit rate using a large quantization step, and the FGS layer is generated by encoding the residues between the original frame and the base layer. The encoded FGS bitstream is an embedded bitstream, namely, the decoder can decode any truncated segment of the bitstream of FGS layer corresponding to each frame. The more bits the decoder receives and decodes, the higher the video quality is. We use MPEG-4 FGS in this paper as an example to illustrate the proposed framework, and can extend to other codecs with similar coding structure, such as H.264 Scalable Video Coding (SVC) [32].

The R-D performance of FGS layer depends on the coding parameters used in the base layer. For simplicity, we set a fixed quantization step size in the base layer. The main task in the video coding subsystem is to determine the bit rate of FGS layer to achieve desired video quality among all users. Accurate R-D models for video bitstreams can help systems be aware of heterogeneous content complexities of co-existing streams and facilitate resource allocation. Previous studies in [22], [23], [33] and our experimental results show that a piecewise linear function is a good approximation to the R-D curve of FGS bitstreams at the frame level. We summarize this piecewise linear model as

$$D_{i,n}(r_{i,n}) = M_{i,n}^{q}\left(r_{i,n} - R_{i,n}^{q}\right) + E_{i,n}^{q}, \quad q = 0, \ldots, p-1,$$
$$\text{with } M_{i,n}^{q} = \frac{E_{i,n}^{q+1} - E_{i,n}^{q}}{R_{i,n}^{q+1} - R_{i,n}^{q}},\ R_{i,n}^{q} \le r_{i,n} \le R_{i,n}^{q+1}$$

where p is the total number of bit planes, $E_{i,n}^{q}$ denotes the i^{th} user's distortion of the n^{th} frame measured in mean-squared error (MSE) after completely decoding the first q DCT bit planes, $R_{i,n}^{q}$ represents the corresponding bit rate, and $r_{i,n}$ indicates the overall decoded bit rate. $E_{i,n}^{0}$ and $R_{i,n}^{0}$ represent the distortion and source rate of the base layer, respectively, and all $(R_{i,n}^{q}, E_{i,n}^{q})$ pairs can be obtained during the encoding process. Since resources are dynamically allocated frame by frame, we omit n to simplify the notations.

B. IEEE 802.11 MAC and PHY Layer

The IEEE 802.11 MAC protocol supports two kinds of access methods, namely, distributed coordination function (DCF) and point coordination function (PCF). The DCF is an access mechanism using carrier sense multiple access with collision avoidance (CSMA/CA). In contrast, the PCF is based on polling controlled by a point coordinator. In both mechanisms, only one user occupies all the bandwidth at each time slot. The proportion of time a user can occupy the bandwidth can be controlled by either PCF or enhanced DCF [36], [37]. In this work, we study how to determine the time proportion allocated to each user to optimize video quality. The proposed scheme can also be deployed in networks supporting similar spectrum management.

We use the IEEE 802.11a Physical (PHY) layer as an example to present the proposed framework. Other wireless LAN standards can be incorporated in a similar way. The IEEE 802.11a Physical layer provides eight PHY modes with different modulation schemes and different convolutional coding rates, and can

TABLE I
PHYSICAL LAYER MODE FOR IEEE 802.11*a*

Mode	Modulation	Channel Coding	Data Rate
1	BPSK	1/2	6 Mbps
2	BPSK	3/4	9 Mbps
3	QPSK	1/2	12 Mbps
4	QPSK	3/4	18 Mbps
5	16-QAM	1/2	24 Mbps
6	16-QAM	3/4	36 Mbps
7	64-QAM	2/3	48 Mbps
8	64-QAM	3/4	54 Mbps

offer various data rates. The configurations of these eight PHY modes are listed in Table I.

With awareness of the current channel conditions and knowledge of the available network resources, the proposed system can select the optimal PHY modes for each uplink and downlink of each user to maximize video quality. Let $P_{max}^{(U)}$ and $P_{max}^{(D)}$ be the maximal available transmission power for uplink and downlink, respectively; and $G_i^{(U)}$ and $G_i^{(D)}$ the uplink and downlink channel gain from user i to his/her conversation partner at the current time slot, respectively. Without loss of generality, we assume the same thermal noise level, σ^2, for all users. Thus, the maximal signal-to-noise ratio (SNR)[1] for uplink and downlink are

$$\Gamma_i^{(U)} = \frac{P_{max}^{(U)} G_i^{(U)}}{\sigma^2} \text{ and } \Gamma_i^{(D)} = \frac{P_{max}^{(D)} G_i^{(D)}}{\sigma^2}, \text{ respectively.} \tag{1}$$

The bit error rates (BERs) of BPSK, QPSK, 16-QAM, and 64-QAM modulation are given by the following equations as functions of the received symbol SNR denoted by Γ [34]

$$P_b^{\mathrm{BPSK}}(\Gamma) = 0.5\left(1 - \sqrt{\frac{\Gamma}{1+\Gamma}}\right) \tag{2}$$

$$P_b^{\mathrm{QPSK}}(\Gamma) = 0.5\left(1 - \sqrt{\frac{\Gamma}{2+\Gamma}}\right) \tag{3}$$

$$P_b^{\mathrm{16QAM}}(\Gamma) = 0.5\left[\left(1 - \sqrt{\frac{\Gamma}{10+\Gamma}}\right) + \left(1 - \sqrt{\frac{9\Gamma}{10+9\Gamma}}\right)\right] \tag{4}$$

and

$$P_b^{\mathrm{64QAM}}(\Gamma) = \frac{1}{24}\left(14 - 7\sqrt{\frac{\Gamma}{42+\Gamma}} - 6\sqrt{\frac{9\Gamma}{42+9\Gamma}} + \sqrt{\frac{25\Gamma}{42+25\Gamma}} - 2\sqrt{\frac{81\Gamma}{42+81\Gamma}} - \sqrt{\frac{121\Gamma}{42+121\Gamma}} + \sqrt{\frac{169\Gamma}{42+169\Gamma}}\right). \tag{5}$$

[1]We use SNR instead of SINR (signal-to-interference-ratio) in this work. If co-channel users are located far away, the interference is small and can be treated as thermal noise. If users are closely located and the hidden terminal problem has been solved, since only one user occupies all the bandwidth at each time slot, there is no interference from co-channel users.

With convolutional code, the union bound for BER [35] can be expressed as

$$P_c(\Gamma) \le \sum_{d=d_{free}}^{\infty} a_d P_d(\Gamma) \tag{6}$$

where d_{free} is the free distance of the convolutional code, a_d is the total number of error events of weight d, and $P_d(\Gamma)$ is the probability that an incorrect path at distance d from the correct path is chosen by the Viterbi decoder. When the hard decision is applied, $P_d(\Gamma)$ can be given by

$$P_d(\Gamma) = \begin{cases} \sum_{k=(d+1)/2}^{d} \binom{d}{k} (P_b)^k (1-P_b)^{d-k}, & \text{when d is odd;} \\ \frac{1}{2}\binom{d}{d/2}(P_b)^{d/2}(1-P_b)^{d/2} + \sum_{k=d/2+1}^{d} \binom{d}{k}(P_b)^k(1-P_b)^{d-k}, & \text{when d is even} \end{cases} \tag{7}$$

where P_b is the uncoded BER depending on the modulations from (2)–(5).

If user i selects the uplink and downlink PHY mode as m_i and n_i, respectively, the BER for uplink and downlink can be approximated as a function of PHY mode and SNR level

$$\mathrm{BER}_{i,m_i}^{(U)} = P_{m_i}\left(\Gamma_i^{(U)}\right) \text{ and } \mathrm{BER}_{i,n_i}^{(D)} = P_{n_i}\left(\Gamma_i^{(D)}\right), \text{ respectively} \tag{8}$$

where the function $P_{m_i}(\cdot)$ and $P_{n_i}(\cdot)$ are the union bound of BER using channel coding as defined in (6). Since different PHY modes use different modulation schemes and channel coding rates, their coded BER performances are different. At the same SNR, systems with higher PHY mode index can provide higher throughput at a cost of higher BER than ones with lower PHY mode index.

The probability that a packet is received successfully for uplink and downlink can be calculated as

$$p_{i,m_i}^{(U)} = \left(1 - \mathrm{BER}_{i,m_i}^{(U)}\right)^L \text{ and } p_{i,n_i}^{(D)} = \left(1 - \mathrm{BER}_{i,n_i}^{(D)}\right)^L, \text{ respectively} \tag{9}$$

where L is the number of bits in a packet. With a fixed packet size, $p_{i,m_i}^{(U)}$ and $p_{i,n_i}^{(D)}$ are functions of the channel gains and PHY modes.

C. Application Layer FEC

In the IEEE 802.11 MAC protocol, a packet sent from uplink will be dropped if errors are detected and will not be forwarded to the next path or the upper communication layer. A packet loss in the base layer will cause error propagation for the video data that are predictively encoded using that frame as reference. In addition, FGS layer bitstream has strong decoding dependency owing to the intra-bitplane variable length entropy coding and the inter-bitplane DCT coefficient synchronization. The loss of a FGS layer packet containing significant bitplanes will make the following successfully received FGS layer packets containing

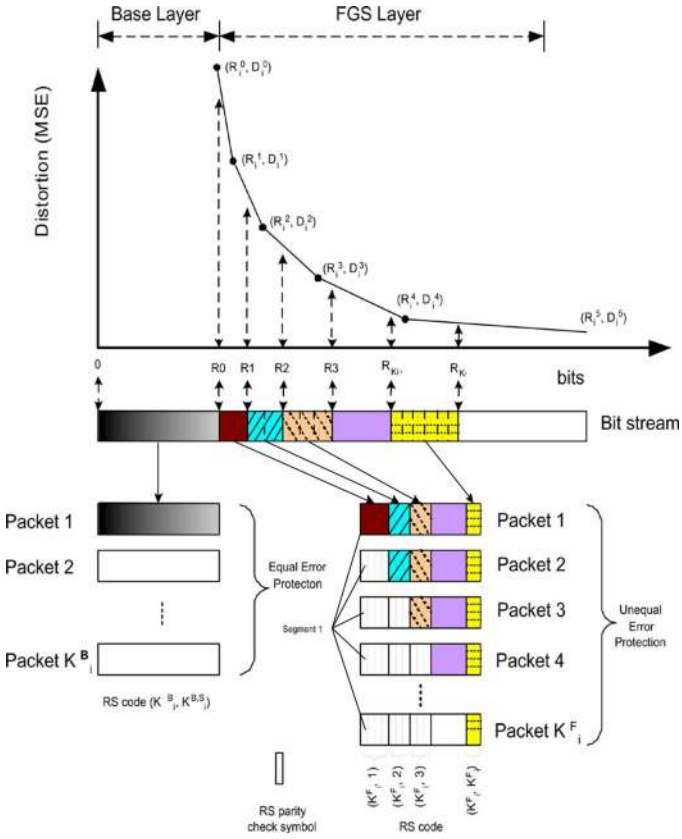


Fig. 2. Error protection scheme for application layer FEC.

lower bitplanes useless. Since we can know which packet arrives successfully at the application layer by checking the transmission index in the packet header, this wireless channel can be modelled as a packet erasure channel [16], [19], [20]. Applying FEC in the application layer across packets, such as systematic Reed–Solomon (RS) codes, has been shown as an effective way to alleviate the problem caused by packet loss [20]. An $\mathrm{RS}(K_i, k)$ encoder will generate $K_i - k$ parity symbols for k source symbols, and a corresponding RS decoder can recover the original source symbols if it receives at least k out of K_i symbols successfully when the locations of the erased symbols are known. We can apply RS codes across source packets to generate parity packets for recovering erasure packet loss.

Since MPEG-4 FGS codec is a two-layer scheme, we adopt different strategies for each layer, as shown in Fig. 2. For the nonscalable base layer, we apply a strong equal error protection strategy across packets to provide the baseline video quality. To remove the strong decoding dependency of the FGS layer bitstream and to have graceful quality fluctuation, we adopt the multiple-description forward error correction framework (MD-FEC) [38]. MD-FEC converts a prioritized bit stream into nonprioritized and packetized bit streams. Each packetized bit stream represents one description that can be independently decoded to represent the content in a coarse quality, and the final reconstructed video quality depends primarily on how many packets the receiver receives successfully, instead of depending on which packets are corrupted. The more descriptions a receiver receives successfully, the better reconstructed quality the decoder can get. The basic mechanism of MD-FEC works as follows. Let s be the number of symbols carried in a packet and K_i be the total number of packets. A segment is defined as the symbols located at the same position over the K_i packets. The FGS bit stream is converted to these K_i packets segment by segment, and an RS coding across packet is applied within each segment to provide error protection. An RS code with higher level of error protection is applied to the segment with higher priority. Fig. 2 shows the overall error protection strategy. If the receiver receives ρ packets successfully out of K_i packets, then the segments encoded with $\mathrm{RS}(K_i, k)$ codes for $k \leq \rho$ can be correctly decoded. The optimal configuration of RS code in each segment can be formulated as a constrained optimization problem and solved through the Lagrangian method [38]. There have been several works proposed to reduce the computational complexity of MD-FEC. We adopt the fast local search method [39] with complexity as $O(K_i \cdot s)$ in this work.

To decode the coded video packets generated by the MD-FEC framework, the RS decoder located at each client terminal needs to know the configuration of RS code used in each segment. The RS configuration is generated through an optimization according to the side information, namely, the R-D of video source, packet loss rate due to the selected PHY modes, and the allocated numbers of transmitted packets. With the side information, the RS configuration can be produced at both client terminals belonging to the same conversation pair. In the next subsection, we will discuss how the server located at the access point coordinates the transmission of video streams and the related side information.

D. Video Over WLAN

Fig. 3 illustrates a flowchart of the proposed system, where user i and j form a conversation pair. Let the video refreshing rate be F frames per second. We divide the time line into F slots per second, and perform distortion management to allocate system resources to every stream within one frame refreshing interval, $T = 1/F$. Note that the distortion management can be performed in a finer time scale to react to fast fading channel so that the channel gain is stable within a time slot. The distortion management consists of two phases, namely, an initialization phase and a video packet transmission phase. The tasks of initialization phase are to gather R-D information of compressed video streams and channel information, and then to perform resource allocation. The task of video packet transmission phase is to send video packets from users to their corresponding conversation partners.

There are three steps executed in the initialization phase. In the first step, each user's video source coder encodes video in real time and analyzes the R-D of the compressed video bitstream. Meanwhile, each user's communication module estimates the downlink channel condition, and then sends the R-D models, (R_i^q, E_i^q), along with the estimated channel conditions, $\Gamma_i^{(D)}$, to the resource allocator located at the access point. At the server side, the resource allocator estimates the channel conditions for the uplink, $\Gamma_i^{(U)}$, of all users. In the second step, the resource allocator gathers the R-D information with the channel information, and performs multiuser cross-layer optimization, which is the core of our proposed system and will be discussed in the next section. The resource

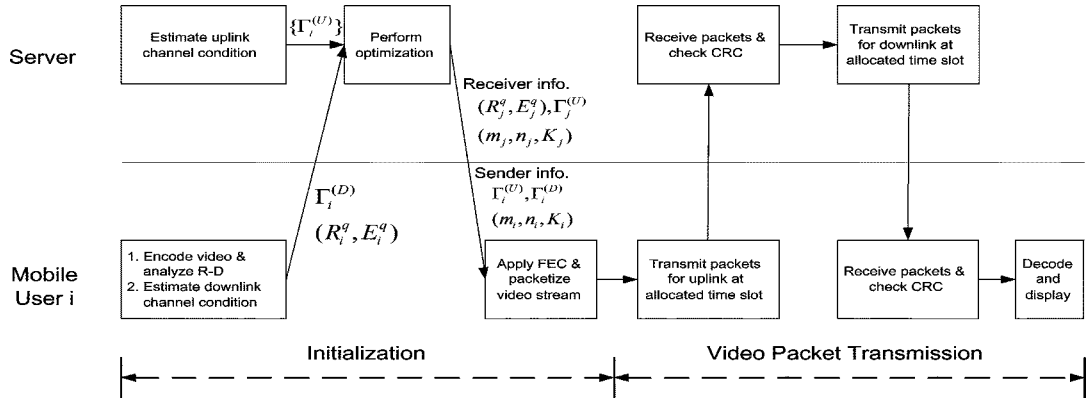


Fig. 3. Flowchart of the proposed wireless video system. User i and user j are a conversation pair.

allocator then informs each user of two sets of transmission configurations. One set is the transmitting configuration for encoding and sending video stream from each user to his/her conversation partner. The configuration information consists of the number of packets to be transmitted, K_i, the selected PHY modes for uplink, m_i, and downlink, n_i, and the channel condition of uplink, $\Gamma_i^{(U)}$ and downlink, $\Gamma_i^{(D)}$. The other set is the receiving configuration for receiving and decoding video stream from each user's conversation partner to himself/herself. This second set of configuration information consists of the expected number of packets to be received, K_j, the selected PHY modes for uplink, m_j, and downlink, n_j, the uplink channel condition, $\Gamma_j^{(U)}$, and the R-D models (R_j^q, E_j^q). In the third step, each user applies FEC and packetizes video packets according to the parameters assigned by the resource allocator. The aforementioned control signals are transmitted through control channels. We assume that the required time in this phase is negligible since the overhead rates of control signals are much smaller than the required rates for transmitting video bitstreams.

After the video is encoded, the coded video packets will be transmitted during the video packet transmission phase, which consists of two steps. In the first step, each user will transmit the FEC coded packets using the assigned PHY mode through an uplink to the access point according to the allocated time slot. In the meantime, the communication module located at the access point will check the cyclic redundancy check (CRC) of each received packet, drop the corrupted packets, and buffer the successfully received packets. In the second step, the server forwards the buffered packets to their destinations using the assigned PHY modes for the downlink path. At each mobile terminal, the communication module checks the CRC of each packet, and gathers the successfully received packets. These packets will be forwarded to the application layer for further processing so that the video frames can be reconstructed for displaying.

The critical issue in this system is how the resource allocator selects the transmission configurations for all users such that the service objective is optimized subject to the system resource constraints. We will formulate a single-cell system as an optimization problem and propose a fast algorithm in Section IV. We then extend the proposed algorithm to a multicell system in Section V.

IV. Joint Uplink-Downlink Optimization: Single-Cell Case

Based on the system described in Section III, we first study a simple case where there is only a single cell with intra-cell calls. We begin with a discussion on the video quality model when we jointly consider the channel conditions in both uplink and downlink. The interactive video streaming system is formulated as a min-max optimization problem, subject to the constraints of maximally allowed transmission time. We will present a fast algorithm to find the transmission configurations for both base and FGS layers.

A. Problem Formulation

Consider the system has a total of N users. As mentioned in Section III, for user i who encodes and sends video streams, we need to determine the PHY mode of uplink, m_i, and the PHY mode of downlink, n_i, in the physical layer, as well as the number of packets sent from user i, K_i, in the application layer. To facilitate the discussion, we use a triplet, (m_i, n_i, K_i), to represent a transmission mode. Assuming all packets of the base layer are received successfully, the end-to-end expected distortion using transmission mode (m_i, n_i, K_i) can be represented as

$$E\left[D_i(m_i, n_i, K_i)\right] = D_i^B - \sum_{k=1}^{K_i} p_i(m_i, n_i, K_i, k)\Delta D_i(K_i, k) \tag{10}$$

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

$$p_i(m_i, n_i, K_i, k) = \sum_{\alpha=k}^{K_i} p_{i,m_i}^{(U)}(K_i, \alpha) p_{i,n_i}^{(D)}(\alpha, k). \tag{11}$$

Here, $p_{i,m_i}^{(U)}(K_i,\alpha)$ is the probability that the server receives α packets successfully when user i sends K_i packets

$$p_{i,m_i}^{(U)}(K_i,\alpha) \triangleq \binom{K_i}{\alpha}\left(1-p_{i,m_i}^{(U)}\right)^{K_i-\alpha}\left(p_{i,m_i}^{(U)}\right)^{\alpha} \quad (12)$$

and $p_{i,n_i}^{(D)}(\alpha,k)$ is the probability that user i's conversation partner receives at least k packets successfully when the server sends α packets

$$p_{i,n_i}^{(D)}(\alpha,k) \triangleq \sum_{\beta=k}^{\alpha}\binom{\alpha}{\beta}\left(1-p_{i,n_i}^{(D)}\right)^{\alpha-\beta}\left(p_{i,n_i}^{(D)}\right)^{\beta}. \quad (13)$$

Note that $p_i(m_i,n_i,K_i,k)$ can be calculated off-line and stored in a lookup table for online retrieval. The complexity to calculate the end-to-end expected distortion in (10) is $O(K_i)$.

To support interactive video streaming, we set the maximum transmission delay as one video frame refreshing interval, i.e., T second. Thus, the encoded bitstream of each video frame should arrive at the end user within the refreshing interval of every video frame. As mentioned in Section III-B, we consider a system where there is only one user who can send data at any moment in one cell. Let t_i be the assigned amount of time for user i to send a video frame to his/her conversation partner through uplink and then downlink. The overall transmission time of all users, $\sum_{i=1}^{N} t_i$, should not exceed T seconds. Note that the amount of time to transmit a fixed-length packet depends on which PHY mode we apply. Denote T_x^{max} as the required transmission time if the PHY mode x is selected 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, respectively, and sends K_i packets from the sender to the server, the expected transmission time along user i's uplink is

$$t_i^{(U)}(m_i,n_i,K_i) = K_i T_{m_i}^{max}. \quad (14)$$

The expected number of packets successfully arriving at server is $p_{i,m_i}^{(U)}K_i$, and expected transmission time along user i's downlink is

$$t_i^{(D)}(m_i,n_i,K_i) = p_{i,m_i}^{(U)}K_i T_{n_i}^{max}. \quad (15)$$

The overall expected transmission time from user i through the server to his/her conversation partner is

$$t_i(m_i,n_i,K_i) = t_i^{(U)}(m_i,n_i,K_i) + t_i^{(D)}(m_i,n_i,K_i). \quad (16)$$

We formulate the overall distortion management problem in the video streaming system as an optimization problem that searches for each user's transmission mode to minimize the maximum of all users' expected distortion, subject to the maximal available transmission time. That is

$$\min_i \max_{\{m_i,n_i,K_i\}} w_i \cdot f\left(E\left[D_i(m_i,n_i,K_i)\right]\right)$$
$$\text{subject to} \sum_{i=1}^{N} t_i(m_i,n_i,K_i) \le T \quad (17)$$

where w_i is the quality weighting factor and $f(\cdot)$ the perceptual distortion function. Because of the integer valued parameters in transmission mode, the problem (17) is NP hard. The complexity of finding the optimal transmission modes for all N users through full search is $O(\kappa^N)$, where κ is the number of all feasible transmission modes bounded by the maximum transmission delay and number of PHY modes provided by WLAN. To meet the real-time requirement of the proposed system, we propose a fast algorithm in the next subsection to find a near-optimal solution to problem (17). As a proof of concept, we consider the case of providing uniform mean-squared distortion among all users, i.e., $w_i = 1, \forall i$, and $f(E[D_i(m_i,n_i,K_i)]) = E[D_i(m_i,n_i,K_i)]$.

B. Proposed Algorithm

Because the base layer and FGS layer have different properties and importance, we propose a two-stage strategy to allocate resources to the base layer first and then FGS layer. The goal of resource allocation in the base layer is to provide a strong error protection and to reduce the overall transmission time used in the base layer so that the remaining transmission time can be used for sending the FGS layer. For the FGS layer, the resource allocation strategy is to prune out inefficient transmission modes and to find the optimal solutions that gives the lowest maximal distortion among all users.

1) Base Layer: Let R_i^0 be the bit rate of the nonscalable base layer associated with user i for the current video frame. With a fixed packet size, L, user i requires $K_i^{B,S} = \lceil R_i^0/L \rceil$ source packets. The remaining rates of the last source packet, $K_i^{B,S}L - R_i^0$, is filled with the first part of the FGS layer bit stream. We need to determine the uplink and downlink PHY mode (m_i,n_i) and the number of parity packets, $K_i^{B,P}$, such that the required 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 $\mathrm{BER}^B = 10^{-6}$ as suggested in [40].

The BER requirement can be attained in three steps: we first examine the smallest number of required parity packets for each (m_i,n_i) to achieve $p_i(m_i,n_i,K_i^{B,S}+K_i^{B,P},K_i^{B,S}) \ge (1-\mathrm{BER}^B)$ using (11); then calculate the corresponding transmission time $t_i^B(m_i,n_i)$ using (16); and finally find the setting with the shortest transmission time

$$(\hat{m}_i,\hat{n}_i) = \arg\min_{\{(m_i,n_i)\}} t_i^B(m_i,n_i). \quad (18)$$

Denote t_i^B as the transmission time using mode $(\hat{m}_i,\hat{n}_i)$. Thus, the overall transmission time for all users is $T^B = \sum_{i=1}^{N} t_i^B$, and the remaining transmission time for FGS layer is $T^F = T - T^B$. An outage is reported if T^B exceeds T, which suggests that there are too many users in the system and there are not even enough resources to support base layer.

2) FGS Layer: To reduce the high dimensionality of the search space, we propose a two-step algorithm by first obtaining a one-to-one mapping function between transmission time and expected distortion (T-D) for each user and then applying bi-section search among all T-D functions to obtain the solutions. The T-D function can be

TABLE II
PROPOSED ALGORITHM TO OBTAIN TRANSMISSION TIME TO EXPECTED DISTORTION FUNCTION

a) Initialization:
1) Feasible set: $S_i = \{(m_i, n_i, K_i), \forall m_i, n_i, K_i\}$
2) Thresholds: $D_s = D_i^B$ and $T_s = 0$,
3) T-D function list: $k = 0$, $t_{i,k} = T_s$, $E[\tilde{D}_i[t_{i,k}]] = D_s$.

b) Obtain T-D function:
While $|S_i| > 0$
1) Find the next efficient mode.
For each $(m_i, n_i, K_i) \in S_i$
$NT(m_i, n_i, K_i) \triangleq t_i(m_i, n_i, K_i) - T_s$.
$(\hat{m}_i, \hat{n}_i, \hat{K}_i) = \arg\min_{\{m_i, n_i, K_i\}} \{NT(m_i, n_i, K_i)\}$.
2) Add $(\hat{m}_i, \hat{n}_i, \hat{K}_i)$ to the T-D function list.
$t_{i,k} = t_i(\hat{m}_i, \hat{n}_i, \hat{K}_i)$,
$E[\tilde{D}_i[t_{i,k}]] = E\left[D_i(\hat{m}_i, \hat{n}_i, \hat{K}_i)\right]$,
$k = k + 1$.
3) Update thresholds T_s and D_s.
$T_s = t_i(\hat{m}_i, \hat{n}_i, \hat{K}_i)$,
$D_s = E\left[D_i(\hat{m}_i, \hat{n}_i, \hat{K}_i)\right]$.
4) Remove modes whose distortion $\geq D_s$ from feasible set.
For each $(m_i, n_i, K_i) \in S_i$,
If $E[D_i(m_i, n_i, K_i)] \geq D_s$
$S = S \setminus (m_i, n_i, K_i)$.
End

obtained by first finding a set of efficient transmission modes. A transmission mode (m_i, n_i, K_i^F) is *efficient* if $E[D_i(m_i', n_i', K_i^{F'})] < E[D_i(m_i, n_i, K_i^F)]$ for all other modes $(m_i', n_i', K_i^{F'})$ with $t_i(m_i', n_i', K_i^{F'}) > t_i(m_i, n_i, K_i^F)$. We can collect all efficient transmission modes $\{(m_i, n_i, K_i^F)\}$ as set $\mathcal{S}_i$ and the corresponding transmission time $\{t_i(m_i, n_i, K_i^F)\}$ as set $\mathcal{T}_i$ obtained via an iterative algorithm as follows. The search algorithm starts from the results of receiving only base layer packets and treats it as the first efficient transmission mode. Suppose the efficient mode selected in the previous iteration can achieve expected distortion D_s and transmission time T_s. In the current iteration, the search algorithm will find the next nearest efficient mode by first pruning out all modes with distortion no less than D_s. Then, among the preserved modes, we choose the mode with the smallest increased transmission time deviated from previous selected efficient mode. 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. The algorithm to obtain a T-D function is summarized in Table II. The complexity of obtaining a T-D function for the worst case is $O(\kappa)$.

Fig. 4 shows an example how to obtain the T-D function for a user by considering only PHY mode index 1 and 2. Let $\mathrm{PHY}(a, b)$ represent the two selected PHY modes for uplink, a, and for downlink, b. For each $\mathrm{PHY}(a, b)$, we can obtain a curve for the expected transmission time and the expected distortion by using different numbers of packets. Since users choose two PHY modes for uplink and two PHY modes for downlink for a packet, there are four different curves shown in Fig. 4. As we can see, Point A is not an efficient transmission mode because we can find other transmission modes with smaller distortion and shorter transmission time (such as Point B). On the other hand, Point B is an efficient transmission mode. After finding all efficient transmission modes, we can collect them as a T-D

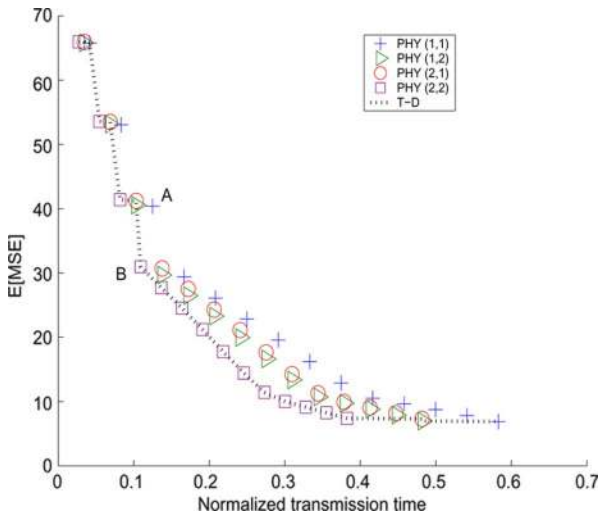


Fig. 4. Time-distortion function.

function, as shown a dotted line in Fig. 4. In general, the resulting T-D function contains points from different $\mathrm{PHY}(a, b)$ modes.

After obtaining all T-D functions for all users, the problem (17) can be reformulated as

$$\min_i \max_{\{t_{i,k}\}} E\left[\tilde{D}_i[t_{i,k}]\right] \quad \text{subject to} \quad \sum_{i=1}^{N} t_{i,k} \leq T^F. \tag{19}$$

Based on the definition of efficient transmission mode, all T-D functions are monotonically decreasing. We solve the problem (19) using bi-section search. The search algorithm calculates the total required time to achieve a targeted distortion, and then increases the targeted distortion at the next iteration if the total required time is higher than the time constraint, T^F, and vice versa. The overall number of iterations is determined by the computation precision used in bi-section search and is typically fewer than 20 in our experiment. If T-D functions are continuous and monotonically decreasing, the solution provided by bi-section search is optimal. However, due to the discrete nature of T-D function as shown in Fig. 4, the problem (19) is NP hard [41] and the solution provided by bi-section search is suboptimal. After determining $t_{i,k}$ for all users, we can obtain the corresponding transmission mode of each user (m_i, n_i, K_i^F) from $\mathcal{S}_i$.

V. JOINT UPLINK-DOWNLINK OPTIMIZATION: MULTICELL CASE

In this section, we consider a video streaming system supporting multiple cells. With the awareness of different network traffic load in different cells, joint resource allocation among multiple cells can improve system performance. We first present the proposed system framework and discuss different types of conversation calls hold within multiple cells. We formulate this multicell system as an optimization problem to minimize the

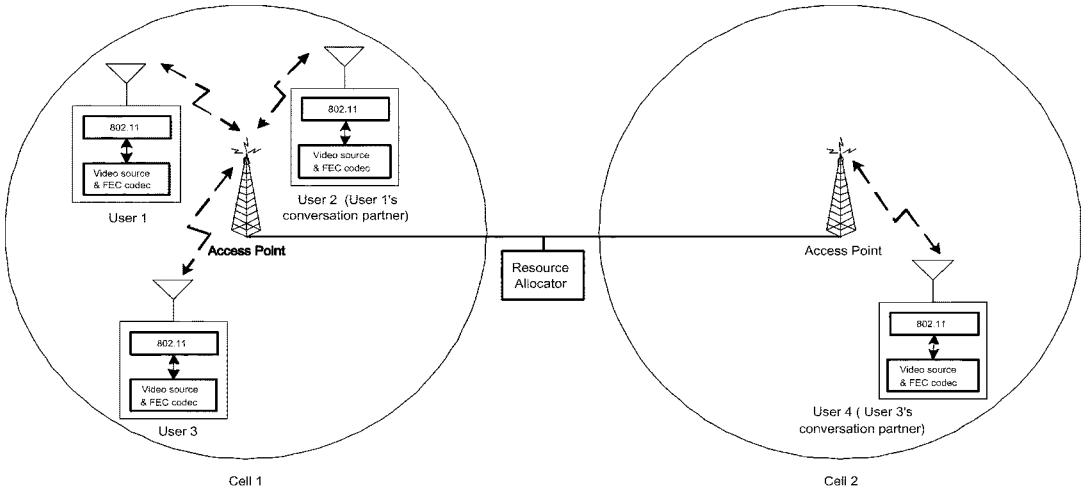


Fig. 5. System block diagram for multicell case.

maximal distortion among all users, and extend the proposed single-cell algorithm to the multicell system.

A. System Framework

Fig. 5 shows the proposed framework for multiple cells. Without loss of generality, here we use a system with two cells as an example to illustrate. For simplicity, we assume that the distance between these two cells are long enough, i.e., two sites of a company, such that they won't interfere to each other. Both cells are connected by a wired channel which is reliable without any packet loss and whose bandwidth is large enough to transmit all packets. We also assume the coherent time of the channel condition is much larger than the propagation delays induced by the wired link. A user can have either an intra-cell conversation with a user within the same cell (e.g., the conversation between user 1 and 2 in Fig. 5), or an inter-cell conversation with a user located in another cell (e.g., conversation between user 3 and 4). Similar to the distortion management used in the single-cell case, the resource allocator needs to first gather R-D information of all video streams and channel information of all links, and then performs distortion control. Note that only one transmitter is allowed to send data at any time instance in each cell. In this two-cell system, two transmitters located at different cells are allowed to transmit video packets simultaneously. The major tasks of resource allocator are how to jointly consider the traffic load in both cells and how to allocate system resources to each user in each link such that the maximal distortion among all users is minimized.

B. Problem Formulation

Suppose there are C cells in the proposed streaming system. Let $S_c^{(U)}$ and $S_c^{(D)}$ be the set of users who have requested uplink channel and downlink channel to send video streams in the c^{th} cell, respectively. As an example shown in Fig. 5, $S_1^{(U)} = \{1,2,3\}$, $S_1^{(D)} = \{1,2,4\}$, $S_2^{(U)} = \{4\}$, and $S_2^{(D)} = \{3\}$. We can formulate this video streaming system as an optimization problem that chooses each user's transmission mode to minimize the maximum of all users' expected distortion, subject to the maximal available transmission time constraint in each cell

$$\begin{aligned} &\min_{i} \max_{\{m_i,n_i,K_i\}} E\left[D_i(m_i,n_i,K_i)\right] \\ &\text{s.t.} \sum_{i\in S_c^{(U)}} t_i^{(U)}(m_i,n_i,K_i) + \sum_{i\in S_c^{(D)}} t_i^{(D)}(m_i,n_i,K_i) \le T_c, \\ &\quad \text{for } c = 1,2,\ldots C. \end{aligned} \tag{20}$$

Unlike the single-cell system containing only intra-cell calls, a multicell system needs to consider the inter-cell conversation pairs whose packets are transmitted from cells to cells. The traffic load in different cells may be different and adjusting traffic load in one cell will affect other cells' load through the inter-cell calls. We should jointly allocate time slots in all cells for the inter-cell calls and evaluate the time constraints in all cells. In fact, the problem (20) is a generalized assignment problem, which is NP hard [41]. To meet the real-time requirement, we propose a fast and suboptimal algorithm by extending the single-cell algorithm.

C. Proposed Algorithm

Similar to the single-cell case, we adopt a two-stage strategy to allocate system resources for the base layer first and then for the FGS layer.

1) Base Layer: In parallel to the single-cell case, we calculate the required number of packets, $K_i^{B,S}$, to carry all base layer's bitstream. We then find the optimal transmission mode $(\hat{m}_i, \hat{n}_i, K_i^{B,S} + K_i^{B,P})$ that has the shortest overall transmission time in both cells with end-to-end BER lower than the BER threshold, BER^B. Once the transmission modes are determined, the overall allocated transmission time for base layer in each cell can be determined as

$$\begin{aligned} T_c^B = &\sum_{i\in S_c^{(U)}} t_i^{(U)}\left(\hat{m}_i,\hat{n}_i,K_i^{B,S}+K_i^{B,P}\right) \\ &+ \sum_{i\in S_c^{(D)}} t_i^{(D)}\left(\hat{m}_i,\hat{n}_i,K_i^{B,S}+K_i^{B,P}\right), \quad \forall c. \end{aligned} \tag{21}$$

Subsequently, we can calculate the rest transmission time, $T_c^F = T - T_c^B$, to transmit FGS layer's data in each cell.

2) FGS Layer: We first obtain the T-D functions, $E[\tilde{D}_i[t_{i,k}]]$, for all users using Table II. For each valid $t_{i,k}$, we can know its corresponding transmission time in the uplink path alone, $t_{i,k}^{(U)}$, and in the downlink path alone, $t_{i,k}^{(D)}$. We reformulate the problem (20) as

$$\min_{i} \max_{\{t_{i,k}\}} E\left[\tilde{D}_i[t_{i,k}]\right]$$
$$\text{subject to} \sum_{i \in S_c^{(U)}} t_{i,k}^{(U)} + \sum_{i \in S_c^{(D)}} t_{i,k}^{(D)} \le T_c^F, \quad \text{for } c = 1, 2, \ldots C. \tag{22}$$

To solve this problem, we propose a fast algorithm performing multiple rounds of bi-section search on all T-D functions, as shown in Fig. 6. For a targeted distortion, the search algorithm calculates the total required transmission time including all uplinks and downlinks in each cell. If there is at least one cell whose overall required time is higher than the corresponding time constraint, T_c^F, the algorithm increases the targeted distortion to reduce the required amount of transmission time in the next iteration, and vice versa. Because the numbers of intra-cell calls and inter-cell calls are different in each cell, the available FGS transmission time in each cell is different. The allocated transmission time in some cells will reach the limit of time constraints first, and some cells may still have unallocated transmission time left. Thus, performing only one round of bi-section search to maintain strict fairness among all users may waste system resources in some cells. To efficiently utilize the remaining system resources, we allow further rounds of bi-section search to reduce users' distortion. A cell is defined as *inactive* if there is no more transmission time left for FGS layer. A user is *inactive* if either uplink or downlink of the corresponding video streaming path is in an inactive cell. Once a round of bi-section search is finished, the proposed multicell algorithm will remove the inactive cells and inactive users from the further assignment list. Then, another round of bi-section search is performed on the T-D functions of all active users subject to the set of time constraints in the active cells. The whole algorithm terminates when there are no more active users in this system.

VI. Simulation Results

In this section, we evaluate the performance of our proposed scheme and compare it with a traditional sequential optimization scheme. This traditional scheme assigns equal bandwidth to each uplink and downlink and allocates system resources to each link independently. More specifically, the resource allocator first allocates the optimal configuration based on only the uplink channel information and the mobile users transmit packets to access point during the first half of available transmission time. Then, based on the packets received successfully by the access point, the resource allocator optimizes the downlink configuration and the server transmits packets to each mobile user during the second half of available transmission time. We first describe the simulation setup and the performance criteria used to examine both schemes, and then present simulation results for both schemes within a single cell and multiple cells, respectively.

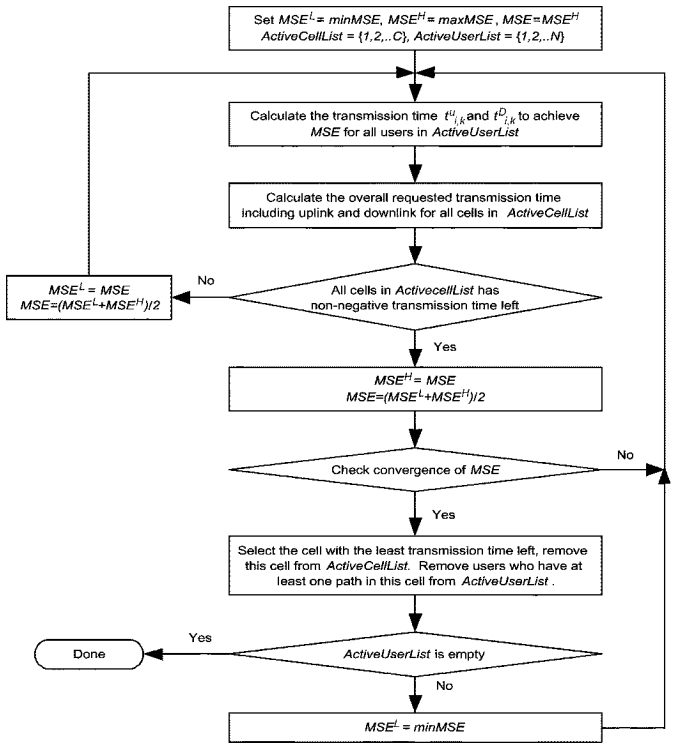


Fig. 6. Proposed algorithm for multicell case.

A. Simulation Setup

The simulations are set up as follows. The noise power is 10^{-10} Watts and the maximal transmission power for both mobile user and server is 40 mW. The path loss factor is 2.5. Packet length L is set to 512 bytes. The video format is QCIF (176×144) with refreshing rate as 30 frames per second and thus $T = 33.33$ ms. We concatenate 15 QCIF video sequences to form one testing video sequence of 5760 frames. The 15 sequences are 300-frame *Akiyo*, 360-frame *carphone*, 480-frame *Claire*, 300-frame *coastguard*, 300-frame *container*, 390-frame *foreman*, 870-frame *grandmother*, 330-frame *hall objects*, 150-frame *Miss American*, 960-frame *mother and daughter*, 300-frame *MPEG4 news*, 420-frame *salesman*, 300-frame *silent*, 150-frame *Suzie*, and 150-frame *Trevor*. 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 and the maximal available bit rate ranges from 62 K to 191 K bits/frame.

A simulation profile for an N-user system is defined as follows: the video content program for each user is 90-frame long, the first video frame starts from a randomly selected frame of the concatenated video, and the location for each user is randomly selected between 20 m to 100 m. For each simulation profile, we repeat the simulations 100 times and take the average.

B. Performance Criteria

Four performance criteria are used to evaluate the proposed scheme and the traditional scheme. Let $\mathrm{PSNR}_{i,n}$ denote the PSNR of the received video frame n for user i. Since the service

objective in the problem (17) is to minimize the maximal distortion, our first performance metric is the worst received video quality among all users. We measure the minimal PSNR among all users at frame n as $\min \mathrm{PSNR}_n = \min_i\{\mathrm{PSNR}_{i,n}\}$ and take the average of the minimal PSNRs' over M video frames

$$\mathrm{minPSNR} = \frac{1}{M}\sum_{n=1}^{M} \mathrm{minPSNR}_n. \tag{23}$$

The second metric is the average video quality received by all users, averaged over M frames

$$\mathrm{avePSNR} = \frac{1}{M}\sum_{n=1}^{M} \mathrm{PSNR}_n \tag{24}$$

where PSNR_n is the average received video quality of all users' n^{th} video frame. The higher avePSNR is, the higher system efficiency in terms of overall video quality we have.

The third metric measures the fairness through examining the deviation of video qualities received by users. If users pay the same price for certain video quality, the received qualities for these users should be similar. To quantify the fairness, we calculate the standard deviation for all users' n^{th} video frame and take the average along the whole M-frame video, i.e.,

$$\mathrm{stdPSNR} = \frac{1}{M}\sum_{n=1}^{M}\left\{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{PSNR}_{i,n} - \mathrm{PSNR}_n)^2\right\}^{\frac{1}{2}}. \tag{25}$$

The lower stdPSNR is, the fairer video quality each user receives.

The forth metric concerns the quality fluctuation. Because significant quality differences between consecutive frames can bring irritating flickering and other artifacts to viewers even when the average video quality is acceptable. To quantify the fluctuation of quality between nearby frames, we use the mean absolute difference of consecutive frames' PSNR, madPSNR, to measure the perceptual fluctuation along each video sequence and take the average over N users

$$\mathrm{madPSNR} = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{M-1}\sum_{n=2}^{M}|\mathrm{PSNR}_{i,n} - \mathrm{PSNR}_{i,n-1}|\right\}. \tag{26}$$

C. Single-Cell Case

We first use a four-user system to illustrate the proposed scheme to achieve fair video quality. User 1, 2 and User 3, 4 are teamed up to form two conversation pairs. The locations of User 1 to 4 are 91 m, 67 m, 71 m, and 20 m away from the access point, respectively. For the video content, User 1 to 4 send one frame of video sequence, *Akiyo*, *carphone*, *Claire*, and *foreman* to their corresponding conversation partner, respectively. The selected transmission modes for the FGS layer using the proposed algorithm are summarized in Table III. As we can see, User 1 to 4 selects uplink PHY modes as 4, 6, 4, and 8, respectively; and downlink PHY modes as 5, 4, 7, and 5, respectively. As expected, a link with longer transmission distance or worse channel condition requires a higher level of error protection (i.e., smaller PHY mode index) to protect video packets. We then compare the required number of packets for each user. The required number of packets for User 2 and 4 are 32 and 24, respectively, which are higher than the 17 and 8 packets for User 1 and 3, respectively. This is because User 2's sequence, *carphone*, and User 4's sequence, *foreman*, have higher content complexity than the other two sequences and require more packets to achieve similar video quality. The overall transmission time for each video stream depends on the number of packets and the selected PHY modes, and is calculated using (14)–(16). Finally, we evaluate the final reconstructed video quality. As shown in Table III, the quality of the final reconstructed video sent from User 1 to 4 are 42.97, 42.75, 42.90, and 42.49 dB, respectively, maintaining a good amount of fairness.

TABLE III
SELECTED TRANSMISSION MODES FOR FGS LAYER

	user 1	user 2	user 3	user 4
Sent video sequence	*Akiyo*	*carphone*	*Claire*	*foreman*
Uplink distance (m)	91	67	71	20
Downlink distance (m)	67	91	20	71
Uplink PHY mode, m_i	4	6	4	8
Downlink PHY mode, n_i	5	4	7	5
Number of packet, K_i	17	32	8	24
Transmission time, t_i (ms)	6.4	8.9	2.5	5.9
Received PSNR (dB)	42.97	42.75	42.90	42.49

We compare the proposed scheme with the sequential optimization scheme by keeping the same simulation setting as mentioned above, except that each user sends a 90-frame video sequence to his/her conversation partner. We repeat the experiments 100 times to calculate the average PSNR for each frame. Fig. 7 shows the frame-by-frame PSNR. As shown, the proposed scheme can provide higher minimal and average PSNR, more uniform video quality among all users, and lower quality fluctuation along each received video sequence than the sequential optimization scheme. The performance gain is attributed to the dynamic bandwidth allocation by the proposed scheme to users in uplink and downlink transmission paths. Note that the sequential optimization scheme allocates fixed $T/2$ seconds for all uplinks and another $T/2$ seconds for all downlinks. Because of the asymmetric channel conditions along uplink and downlink for each video stream and the time heterogeneity of video content, the sequential optimization scheme lacks the freedom to dynamically adjust the time budget for uplink and downlink to attain better video quality.

We evaluate the performance of both schemes with different number of users within a single cell and show the results in Fig. 8. We average the results over 100 simulation profiles as described in Section VI-A, and calculate the minPSNR, avePSNR, stdPSNR, and madPSNR as defined in Section VI-B. We can see in Fig. 8(a) that, for the minPSNR criterion, the proposed joint optimization scheme outperforms the sequential optimization scheme 3.82∼11.50 dB. In other words, the worst received quality among all users in the proposed scheme has a substantial improvement over the one in the sequential optimization scheme. Comparing the avePSNR as shown in Fig. 8(b) and the stdPSNR as shown in Fig. 8(c), the proposed scheme has higher overall quality by 2.18∼7.95 dB and lower quality deviation by 0.92∼2.95 dB among all users than the sequential optimization scheme. The proposed algorithm can provide not

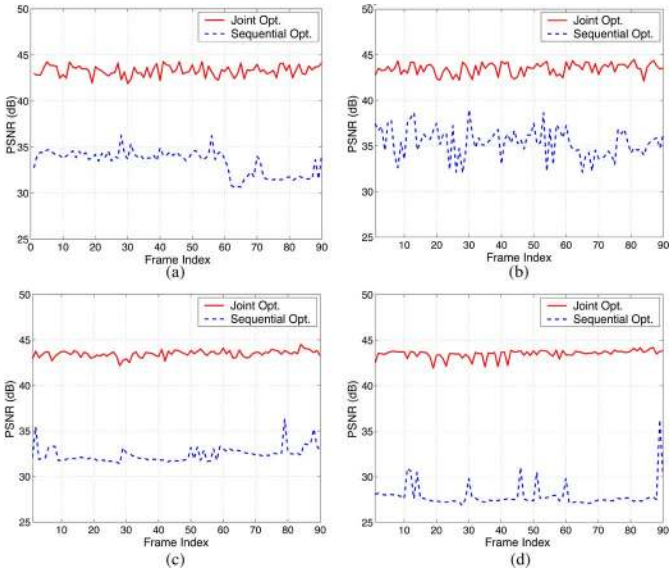


Fig. 7. Frame-by-frame PSNR for User 1 to User 4. (a) User 1; (b) user 2; (c) user 3; and (d) user 4.

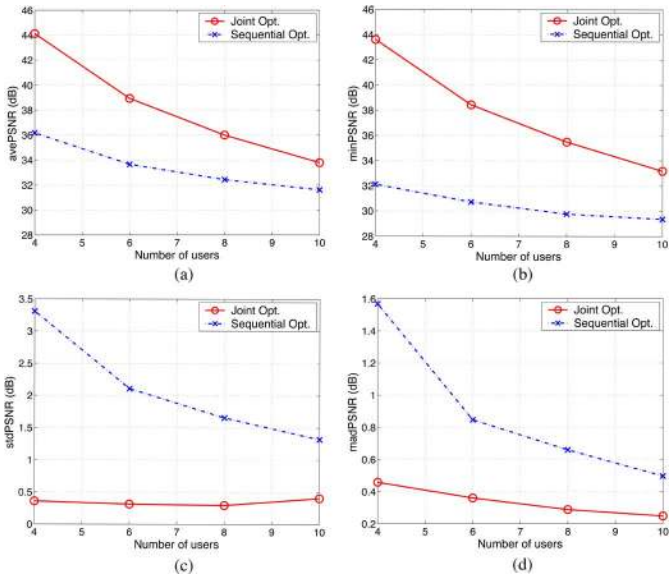


Fig. 8. PSNR performance results for different number of users for single-cell case. (a) minPSNR; (b) avePSNR; (c) stdPSNR; and (d) madPSNR.

only higher overall users' video quality but also more uniform video quality among all users. In general, a system with more users can leverage the diversity of video content complexity to provide more consistent video qualities to all users. However, we observe that the stdPSNR for the proposed system with ten users is slightly higher than the one with eight users. This is because the available FGS transmission time for the system with ten users is close to 0. In most cases, the system can allocate transmission time for the base layer only. Consequently, there are less transmission time budget left for FGS bitstreams to compensate the quality deviation among users contributed by the base layer, which results in higher stdPSNR. Fig. 8(d) shows the quality fluctuation along each received video sequence for both schemes. The proposed scheme can achieve 0.25∼1.11 dB lower madPSNR than the sequential optimization scheme.

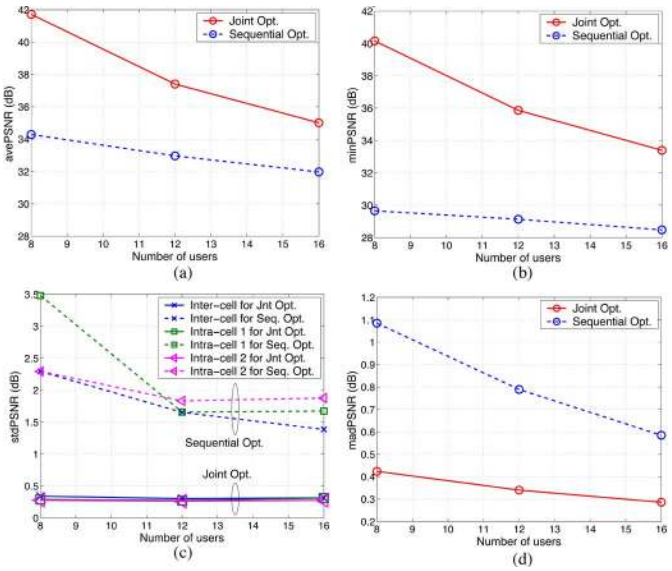


Fig. 9. PSNR results for different number of users for two-cell case. (a) minPSNR; (b) avePSNR; (c) stdPSNR; and (d) madPSNR.

By exploring multiuser diversity, the more users the proposed system has, the lower quality fluctuation each user experiences.

D. Multiple-Cell Case

For the multicell case, without loss of generality, we simulate a two-cell system in which there are 8, 12, and 16 users. For each simulation profile, each user is randomly located in either cell, the distance from each user to his/her cell's access point is randomly selected between 20 m to 100 m, and each user's first video frame is also randomly picked from the testing video sequence. We repeat the simulation using 100 different profiles and average the results to evaluate the performance. Fig. 9(a) and (b) shows the minPSNR and avePSNR using both schemes for different number of users in this system, respectively. The proposed joint uplink and downlink optimization scheme outperforms the sequential uplink and downlink optimization scheme by 4.92∼10.50 dB for the minimal PSNR and by 3.04∼7.43 dB for the average PSNR. Since there are three different types of video streaming flows in this system, namely, inter-cell call between cell one and two, intra-cell call within cell one, and intra-cell call within cell two. we shall compare the stdPSNR for each type of call separately. As revealed by Fig. 9(c), the proposed algorithm can provide lower quality deviation for all three types of calls. Fig. 9(d) shows the quality fluctuation along each received video sequence for both schemes, suggesting that the proposed scheme provides lower quality fluctuation than the sequential optimization scheme. In summary, the proposed scheme can provide higher minPSNR, higher avePSNR, lower stdPSNR, and lower madPSNR, which again demonstrates the superiority of joint uplink and downlink optimization.

To study the bottleneck effect caused by different traffic loads over different cells, we conduct another simulation in which there are 8 users and there are only two types of calls, namely, inter-cell call between cell one/two and intra-cell call within cell one. The PSNR performances with various number of intra-cell calls in cell one are shown in Fig. 10. If the system has more

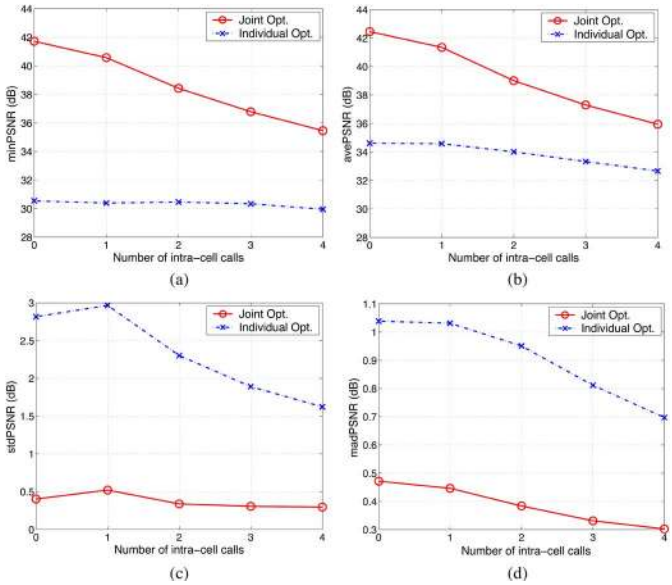


Fig. 10. PSNR results for different number of intra-cell calls for two-cell case with 8 users. (a) minPSNR; (b) avePSNR; (c) stdPSNR; and (d) madPSNR.

intra-cell calls within cell one, there are more users requesting bandwidth to deliver video streams such that cell one becomes the system's bottleneck. Consequently, the allocated bandwidth for each user is reduced, and the received video quality decreases.

Fig. 10(c) shows that the stdPSNR of a system with only one intra-cell call is slightly higher than the one without any intra-cell calls. This is because 7% of simulation profiles have all users in cell two being far away from the access point. These users adopt higher level of error protection to transmit video streams and thus require longer transmission time along the corresponding uplinks and downlinks in cell two than in cell one. Therefore, the available transmission time in cell two will saturate earlier than cell one. To utilize unassigned transmission time in cell one, our algorithm performs another round of bi-section search in cell one. It results in two different levels of video quality in the overall system and the quality deviation among all users increases.

VII. Conclusion

In summary, we have constructed a network-aware and source-aware video streaming framework for multiple conversation pairs within IEEE 802.11 networks. The proposed framework dynamically performs multidimensional resource allocation by jointly exploring the cross-layer error protection, multiuser diversity, and the heterogeneous channel conditions in all paths. We formulate the system as a min-max optimization problem to provide satisfactory video quality for all users. A fast algorithm that converts system resources into time-distortion functions is proposed to determine the transmission configuration for each user in both single-cell and multicell scenario.

We compare the proposed scheme with a traditional scheme that performs sequential optimization for uplink and downlink. Our experiments demonstrated that the proposed scheme for a single cell scenario can obtain a 2.18∼7.95 dB gain for the average received PSNR of all users and a 3.82∼11.50 dB gain for the minimal received PSNR among all users. For a two-cell case, the proposed scheme can achieve a 4.92∼10.50 dB gain for the worst received quality among all users and a 3.04∼7.43 dB gain for the average video quality. In addition, the proposed scheme can provide more uniform video quality among all users and lower quality fluctuation along each received video sequence.

References

[1] D. G. Jeong and W. S. Jeon, "CDMA/TDD system for wireless multimedia services with traffic unbalance between uplink and downlink," *IEEE J. Select. Areas Commun.*, vol. 17, no. 5, pp. 939–946, May 1999.

[2] W. S. Jeon and D. G. Jeong, "Call admission control for mobile multimedia communications with traffic asymmetry between uplink and downlink," *IEEE Trans. Veh. Technol.*, vol. 50, no. 1, pp. 59–66, Jan. 2001.

[3] W. S. Jeon and D. G. Jeong, "Call admission control for CDMA mobile communications systems supporting multimedia services," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 649–659, Oct. 2002.

[4] H. Yomo and S. Hara, "An uplink/downlink asymmetric slot allocation algorithm in CDMA/TDD-based wireless multimedia communications systems," in *Proc. IEEE Vehicular Technol. Conf.*, Fall, 2002, vol. 2, pp. 797–801.

[5] H.-Y. Wei, C.-C. Chiang, and Y.-D. Lin, "Co-DRR: An integrated uplink and downlink scheduler for bandwidth management over wireless LANs," in *IEEE Int. Symp. on Computers and Commun.*, 2003, vol. 2, pp. 1415–1420.

[6] T. V. Lakshman, A. Ortega, and A. R. Reibman, "VBR video: Tradeoffs and potentials," *Proc. IEEE*, vol. 86, no. 5, pp. 952–973, May 1998.

[7] B. Girod and N. Farber, "Wireless video," in *Compressed Video Over Networks*, M.-T. Sun and A. R. Reibman, Eds. New York: Marcel Dekker, 2001.

[8] A. Eleftheriadis, M. R. Civanlar, and O. Shapiro, "Multipoint video conferencing with scalable video coding," in *Proc. Packet Video Workshop*, Apr. 2006.

[9] P. J. Cherriman, T. Keller, and L. Hanzo, "Orthogonal frequency-division multiplex transmission of H.263 encoded video over highly frequency-selective wireless networks," *IEEE Trans. Circuits Syst. Video Techol.*, pp. 701–712, Aug. 1999.

[10] N. H. L. Chan and P. T. Mathiopoulos, "Efficient video transmission over correlated Nakagami fading channels for IS-95 CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 996–1011, Jun. 2000.

[11] Y. Wang and Q. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[12] H. Zheng and K. J. R. Liu, "The subband modulation: A joint power and rate allocation framework for subband image and video transmission," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 5, pp. 823–838, Aug. 1999.

[13] J. Song and K. J. R. Liu, "An integrated source and channel rate allocation scheme for robust video coding and transmission over wireless channels," *EURASIP J. Appl. Signal Process.*, vol. 2004, no. 2, pp. 304–316, Feb. 2004.

[14] C. E. Luna, Y. Eisenberg, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Joint source coding and data rate adaptation for energy efficient wireless video streaming," *IEEE J. Select. Areas Commun.*, vol. 21, no. 10, pp. 1710–1720, Dec. 2003.

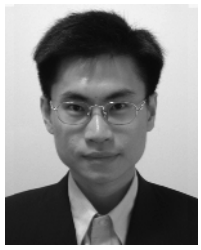
[15] Y. Li, A. Markopoulou, N. Bambos, and J. Apostolopoulos, "Joint power-playout control for media streaming over wireless links," *IEEE Trans. Multimedia*, vol. 8, no. 4, pp. 830–843, Aug. 2006.

[16] A. Majumda, D. G. Sachs, I. V. Kozintsev, K. Ramchandran, and M. M. Yeung, "Multicast and unicast real-time video streaming over wireless LANs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 524–534, Jun. 2002.

[17] Y. Shan and A. Zakhor, "Cross-layer techniques for adaptive video streaming over wireless networks," in *Proc. IEEE Int. Conf. Multimedia and Expo*, Aug. 2002, vol. 1, pp. 277–280.

[18] Y. Chen, J. C. Ye, C. R. Floriach, and K. Challapali, "Robust video streaming over wireless LAN with efficient scalable coding and prioritized adaptive transmission," in *Proc. IEEE Int. Conf. Image Processing*, Sep. 2003, vol. 3, pp. 14–17.

[19] Q. Li and M. van der Schaar, "Providing adaptive QoS to layered video over wireless local area networks through real-time retry limit adaptation," *IEEE Trans. Multimedia*, vol. 6, no. 2, pp. 278–290, Apr. 2004.
[20] 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 J. Select. Areas Commun.*, vol. 21, no. 10, pp. 1752–1763, Dec. 2003.
[21] L. Wang and A. Vincent, "Bit allocation and constraints for joint coding of multiple video programs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 6, pp. 949–959, Sep. 1999.
[22] X. M. Zhang, A. Vetro, Y. Q. Shi, and H. Sun, "Constant quality constrained rate allocation for FGS-coded videos," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 121–130, Feb. 2003.
[23] G.-M. Su and M. Wu, "Efficient bandwidth resource allocation for low-delay multiuser MPEG-4 video transmission," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 9, pp. 1124–1137, Sep. 2005.
[24] X. Lu, Y. Wang, E. Erkip, and D. Goodman, "Power optimization of source encoding and radio transmission in multiuser CDMA systems," in *Proc. IEEE Int. Conf. on Communications*, Jun. 2004, vol. 5, pp. 3106–3110.
[25] G.-M. Su, Z. Han, M. Wu, and K. J. R. Liu, "Joint uplink and downlink optimization for video conferencing over wireless LAN," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, 2005, vol. 2, pp. 1101–1104.
[26] Z. Han, G.-M. Su, A. Kwasinski, M. Wu, and K. J. R. Liu, "Multiuser distortion management of layered video over resource limited downlink MC-CDMA," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3056–3067, Nov. 2006.
[27] G.-M. Su, Z. Han, M. Wu, and K. J. R. Liu, "A scalable multiuser framework for video over OFDM networks: Fairness and efficiency," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 10, pp. 1217–1231, Oct. 2006.
[28] A. C. Begen and Y. Altunbasak, "Proxy-assisted interactive-video services over networks with large delays," *Signal Process.: Image Commun.*, vol. 20, no. 8, pp. 755–772, Sep. 2005.
[29] Y. Shan, I. V. Bajic, S. Kalyanaraman, and J. W. Woods, "Overlay multi-hop FEC scheme for video streaming over peer-to-peer networks," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 2004, vol. 5, pp. 3133–3136.
[30] M. Wu, S. S. Karande, and H. Radha, "Network-embedded FEC for optimum throughput of multicast packet video," *Signal Process.: Image Commun.*, vol. 20, no. 8, pp. 728–742, Sep. 2005.
[31] 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. Multimedia*, vol. 3, no. 1, pp. 53–68, Mar. 2001.
[32] J. Reichel, H. Schwarz, and M. Wien, "Joint Scalable Video Model JSVM-6," Joint Video Team Doc. JVT-S202, Apr. 2006.
[33] L. Zhao, J. Kim, and C.-C. J. Kuo, "MPEG-4 FGS video streaming with constant-quality rate control and differentiated forwarding," in *Proc. SPIE Conf. Visual Communications and Image Processing*, 2002, pp. 230–241.
[34] L. Hanzo, S. X. Ng, T. Keller, and W. T. Webb, *Single and Multicarrier Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space-Time Coded OFDM, CDMA and MC-CDMA Systems*. New York: Wiley, 2004.
[35] J. G. Proakis, *Digital Communication*. New York: McGraw-Hill, 1995.
[36] 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.
[37] Y. Xiao, "IEEE 802.11e: QoS provisioning at the MAC layer," *IEEE Wireless Commun.*, vol. 11, no. 3, pp. 72–79, Jun. 2004.
[38] R. Puri, K.-W. Lee, K. Ramchandran, and V. Bharghavan, "An integrated source transcoding and congestion control paradigm for video streaming in the internet," *IEEE Trans. Multimedia*, vol. 3, no. 1, pp. 18–32, Mar. 2001.
[39] V. M. Stankovic, R. Hamzaoui, and Z. Xiong, "Real-time error protection of embedded codes for packet erasure and fading channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 8, pp. 1064–1072, Aug. 2004.
[40] S. Gringeri, R. Egorov, K. Shuaib, A. Lewis, and B. Basch, "Robust compression and transmission of MPEG-4 video," in *Proc. 7th ACM Inter. Conf. on Multimedia*, Jun. 2000, pp. 113–120.
[41] S. Martello and P. Toth, *Knapsack Problems: Algorithms and Computer Implementations*. West Sussex, U.K.: Wiley, 1990.



Guan-Ming Su (S'04–M'07) received the B.S.E. degree in electrical engineering from National Taiwan University in 1996 and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 2001 and 2006, respectively.

He was with the Research and Development Department, Qualcomm, Inc., San Diego, CA, during the summer of 2005, and with ESS Technology, Fremont, CA, in 2006. He is currently with Marvell Semiconductor, Inc., Santa Clara, CA. His research interests are multimedia communications and multimedia signal processing.

Zhu Han (S'01–M'04) received the B.S. degree in electronic engineering from Tsinghua University in 1997, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1999 and 2003, respectively.

From 2000 to 2002, he is an R&D Engineer with ACTERNA, Germantown, MD. From 2002 to 2003, he was a Graduate Research Assistant at the University of Maryland. From 2003 to 2006, he was a Research Associate at the University of Maryland. Currently, he is an assistant Professor with the Electrical and Computer Engineering Department, Boise State University, Boise, ID. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, and wireless multimedia.

Dr. Han is Guest Editor for the Special Issue on Cross-layer Optimized Wireless Multimedia Communications, *Journal of Advances in Multimedia*. He is PHY/MAC Symposium vice chair of the IEEE Wireless Communications and Networking Conference, 2008. He is a member of the Technical Programming Committee for the IEEE International Conference on Communications, the IEEE Vehicular Technology Conference, the IEEE Consumer Communications and Networking Conference, the IEEE Wireless Communications and Networking Conference, and the IEEE Globe Communication Conference.

Min Wu (S'95–M'01–SM'06) received the B.E. degree in electrical engineering and the B.A. degree in economics (both with the highest honors) from Tsinghua University, Beijing, China, in 1996, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 2001.

Since 2001, she has been with the faculty of the Department of Electrical and Computer Engineering and the Institute of Advanced Computer Studies at the University of Maryland, College Park, where she is currently an Associate Professor. Previously she was with the NEC Research Institute and Panasonic Laboratories, Princeton. She co-authored two books, *Multimedia Data Hiding* (Springer-Verlag, 2003) and *Multimedia Fingerprinting Forensics for Traitor Tracing* (EURASIP/Hindawi, 2005), and holds five U.S. patents. Her research interests include information security and forensics, multimedia signal processing, and multimedia communications.

Dr. Wu is an Associate Editor of IEEE SIGNAL PROCESSING LETTERS and an Area Editor of the *IEEE Signal Processing Magazine*. She is a member of the IEEE Technical Committees on Image and Multidimensional Signal Processing, on Multimedia Signal Processing, and on Multimedia Systems and Applications. She served as Finance Chair for 2007 IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP), and Publicity Chair for 2003 IEEE International Conference on Multimedia and Expo (ICME). She received a U.S. National Science Foundation CAREER award in 2002, a University of Maryland George Corcoran Education Award in 2003, an MIT Technology Review's TR100 Young Innovator Award in 2004, and a U.S. Office of Naval Research Young Investigator Award in 2005. She is a co-recipient of the 2004 EURASIP Best Paper Award and the 2005 IEEE Signal Processing Society Best Paper Award.

K. J. Ray Liu (F'03) received the B.S. degree from the National Taiwan University and the Ph.D. degree from the University of California, Los Angeles, both in electrical engineering.

He is Professor and Associate Chair, Graduate Studies and Research, of Electrical and Computer Engineering Department, University of Maryland, College Park. His research contributions encompass broad aspects of wireless communications and networking, information forensics and security, multimedia communications and signal processing, bioinformatics and biomedical imaging, and signal processing algorithms and architectures.

Dr. Liu is the recipient of numerous honors and awards including best paper awards from IEEE Signal Processing Society (twice), IEEE Vehicular Technology Society, and EURASIP; IEEE Signal Processing Society Distinguished Lecturer, EURASIP Meritorious Service Award, and National Science Foundation Young Investigator Award. He also received various teaching and research recognitions from University of Maryland including university-level Distinguished Scholar-Teacher Award and Invention of the Year Award, and college-level Poole and Kent Company Senior Faculty Teaching Award. He is Vice President—Publications and on the Board of Governor of IEEE Signal Processing Society. He was the Editor-in-Chief of *IEEE Signal Processing Magazine* and the founding Editor-in-Chief of the *EURASIP Journal on Applied Signal Processing*.