Joint Source Adaptation and Resource Allocation for Multi-User Wireless Video Streaming

Jianwei Huang, *Member, IEEE*, Zhu Li, *Senior Member, IEEE*, Mung Chiang, *Member, IEEE*, and Aggelos K. Katsaggelos, *Fellow, IEEE*

***Abstract*—Multi-user video streaming over wireless channels is a challenging problem, where the demand for better video quality and small transmission delays needs to be reconciled with the limited and often time-varying communication resources. This paper presents a framework for joint network optimization, source adaptation, and deadline-driven scheduling for multi-user video streaming over wireless networks. We develop a joint adaptation, resource allocation and scheduling (JARS) algorithm, which allocates the communication resource based on the video users' quality of service, adapts video sources based on smart summarization, and schedules the transmissions to meet the frame delivery deadlines. The proposed algorithm leads to near full utilization of the network resources and satisfies the delivery deadlines for all video frames. Substantial performance improvements are achieved compared with heuristic schemes that do not take the interactions between multiple users into consideration.**

***Index Terms*—Collaborative video streaming, optimization decomposition, pricing control, rate-distortion modeling, video adaptation.**

I. Introduction

A. Motivation

With the advances of mobile computing technology and deployments of 3G wireless infrastructure, video communication applications are becoming very important for service providers as a source of many new business applications. However, there are still many open problems in terms of how to efficiently provision complicated quality-of-service (QoS) requirements for mobile users. One particular challenging problem is multi-user video streaming over wireless channels, where the demand for better video quality and small transmission delays needs to be reconciled with the limited and often time-varying communication resources. The main technical difficulties are as follows:

- The video sources for most streaming applications are typically precoded stored video sequences with relative high bit rates. However, the currently deployed wireless cellular systems (e.g., [1], [2]) are designed to only support voice and lower bit rate data. In order to support video streaming over such networks, the high rate video sources need to be adapted through a variety of schemes, such as scalable video stream extraction (e.g., [3]–[5]), transcoding (e.g., [6], [7]), and summarization (e.g., [8]), before they can be accommodated by the wireless channel.
- Different video content segments have different rate-distortion characteristics, e.g., some segment may be part of an action movie and requires a lot of bits to encode, while others maybe news anchors talking that require relatively less bits to encode. In a wireless multiaccess channel, the type of multi-user content diversity in content rate-distortion characteristics should be taken into consideration while optimizing the network resource.
- The resource consumptions of video users are typically discrete, i.e., measured in frames instead of in bits. As a result, their utility functions (QoS as functions of allocated resources) are discrete as well, and typically do not have close form representations. Therefore, most of previous work on resource allocation for elastic data traffic does not directly apply here, and a new optimization framework is needed.
- The streaming applications have stringent delay requirements, which can be only satisfied under a carefully designed scheduling policy. This is a challenging task in a wireless network, since the transmissions of multiple users are typically tightly coupled either due to limited network resource (e.g., transmission power or bandwidth in downlink transmissions) or mutual interferences (e.g., in uplink transmissions).

Traditionally the mechanisms of content generation and the engineering of network resources are designed separately, and most of the above challenges are ignored in the network design. This has led to a "content-pipe divide" [40], and the need for application-aware networking. This paper presents another step towards matching content processing with network engineering so as to maximize the user perception's utility. In this paper, we propose a framework for resource allocation, source adaptation, and deadline oriented scheduling. During the resource allocation phase, the network resources are allocated to different video users by temporally treating them as "elastic data users," i.e., without considering the discrete nature of the video traffic. An optimal average resource allocation is achieved in a distributed fashion by exploiting the content diversity among users. Based

Manuscript received March 26, 2007; revised September 17, 2007. This paper was recommended by Associate Editor E. Steinbach.

J. Huang is with the Department of Information Engineering, the Chinese University of Hong Kong, Shatin, NT, Hong Kong, SAR (e-mail: jwhuang@ie.cuhk.edu.hk).

Z. Li is with the Department of Computing, the Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, SAR (e-mail: zhu.li@ieee.org).

M. Chiang is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: chiangm@princeton.edu).

A. K. Katsaggelos is with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60260 USA (e-mail: aggk@eecs.northwestern.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/TCSVT.2008.919109

on the average resource allocation, users perform source adaptations in a distributed fashion to select a set of video frames to be transmitted in order to match the allocated resource. Then two greedy deadline oriented scheduling algorithms (for uplink and downlink transmissions, respectively) are proposed to satisfy users' stringent deadline requirements by taking advantage of the variable bit rate (VBR) nature of users' traffic.

B. Background and Related Work

The problem of source adaptation has been widely explored in the video coding community, with a good review provided by [9]. Video source adaptation serves two purposes in video communication and consumption: 1) complying with the resource limitations in communication and 2) satisfying user preferences in video consumption. Resource limitations can be due to bandwidth and energy limitations in communications, disk size in storage devices, display size in hand-held devices, and battery energy and computational power in mobile devices. User preferences can be expressed in terms of the reconstructed video quality, which is a function of the frame size, peak signal-to-noise ratio (PSNR), and frame rate. There are two basic solutions to the video adaptation problem, provided by scalable coding [3]–[5] and transcoding [6], [7]. With scalable coding, a video source is encoded once in such a way that different subsets of the bit stream can be used to reconstruct the video sequence at different frame sizes, frame rates, and visual quality levels. Scalable coding offers adaptation with minimum computational burden and can be performed at routers and access points. With transcoding, a decoder and an encoder are concatenated back to back, resulting in a flexible system that is able to adapt to communication resource limitations and achieves desired video quality levels at the cost of high computational complexity. Various means for reducing the complexity of transcoding exist by taking advantage of partial decoding of the bit stream and reusing of the motion field information.

To achieve better end-to-end quality at very low bit rate over wireless networks, a more intelligent approach to video adaptation is needed. In this work, we utilize video content analysis [10] and summarization solutions [8], [11] to deliver good visual quality at low bit rates. Through content analysis and optimization, video summarization schemes select a subset of frames from the video sequence to form a concise representation of the sequence, such that the incurred loss is minimized. This extra layer of intelligence can be used to guide transcoding or scalable stream packet extraction, while achieving better quality and achieves better quality than the content-blind approaches. In a wireless network, the complex underlining channel conditions directly affect the QoS of the video applications. In order to achieve an optimal network performance, it is natural to coordinate the decisions at the application layer (e.g., source coding and adaptation) with the underlying physical layer resource allocations. There exists a rich literature in this field, and some representative work includes [12]–[19], with a good survey in [20]. One approach focuses on the design of effective protection strategies to deal with the error-prone wireless channels (e.g, [14], [15]); another one is to partition the multimedia data into various priority classes for adaptive transmission (e.g., [16]); a final approach is to take the stochastic nature of the wireless networks into consideration (e.g, [18], [19]). However, most of the previous work did not explicit consider the competition for resource among multiple users in the network. Recently, rate control for multiple video streaming over multihop wireless networks was considered [21][22], where the deadline constraints are not explicitly taken into consideration.

Cross-layer resource allocation based on optimization decomposition has also been considered in the networking community, with recent results summarized in the long survey paper [23]. Most work in this field (e.g., [24]–[27] and references therein) has considered optimization for elastic data traffic using a fluid model without delay constraints. In the video streaming applications, however, we need to further consider the discrete frame selection problem (through source adaptation) and satisfy the stringent delivery deadline constraints (through scheduling).

C. Summary of Contributions

In this paper, we start by formulating a joint optimization problem that involves both the network resource constraints at the physical layer and users' QoS service requirements at the application layer. The decomposition of the joint problem is then carried out in a systematic way such that modularity of different functional units is maintained while the interactions among them are properly designed, which leads to enhanced network efficiency without sacrificing network architectural divisions among coders, routers, and schedulers. The major contributions of this paper are as follows.

- *Framework*: This paper presents a framework for joint network optimization, source adaptation, and deadline-driven scheduling for multi-user video streaming over wireless networks. By exploiting the content diversity of the video users, the network resources can be efficiently used, and the aggregate distortion of the video users can be minimized while meeting the stringent delivery deadlines of the streaming applications.
- *Algorithm*: We develop a family of joint adaptation, resource allocation, and scheduling (JARS) algorithms, which allocate the communication resources based on the video users utility functions, adapt video sources at very low bit rate (VLBR) based on content-aware summarization and transcoding schemes, and schedule the transmission of packets to meet individual deadline constraints. The resource allocation is achieved in a distributed fashion based on dual decompositions, while the source adaptation is performed based on content summarization to minimize the delivery distortions. The scheduling is done in a centralized fashion, and requires the mobile users to send frame information to the base station. A feedback mechanism between resource allocation, source adaptation, and scheduling is established to ensure the feasibility of the solution.
- *Performance Evaluation*: The proposed algorithms lead to near full utilization of the network resources, and achieve good delivered visual quality with very low bit rate that can be supported by the existing cellular networks. Both the computational complexity and communication overhead are low. The performance improvement is significant

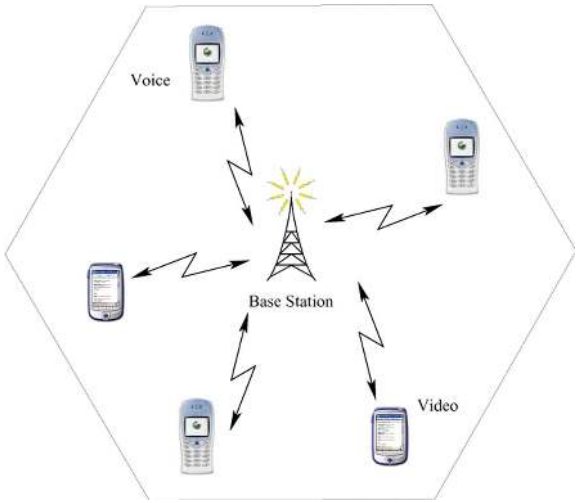


Fig. 1. Single-cell network with mixed voice and video users.

compared with the heuristic schemes that do not take the interactions between multiple users into consideration.

The rest of the paper is organized as follows. We first describe the general solution framework in Section II. In Sections III and IV, we discuss in details of how the framework can be applied for uplink and downlink video streaming in wireless networks. Experimental results are given in Section V, and we conclude in Section VI.

II. Framework of Optimization, Adaptation and Scheduling

We consider a single cell model in a wireless cellular network based on code division multiple access (CDMA).[1] A *fixed* user population with both voice and video applications are considered, as shown in Fig. 1. All users communicate with the base station through one-hop transmission, thus there is no problem of multihop relay or routing. A voice transmission is successful if a target signal-to-interference-plus-noise ratio (SINR) is reached at the receiver. A video users is more flexible and can adapt to the network environment in terms of the achieved SINR and the transmission rate. However, once the video frames are transmitted, stringent delay deadlines need to be satisfied in order to guarantee the normal operation of the streaming application.

Here the network objective is to maximize the overall performance of the video users (measured in terms of video qualities), subject to the normal operations of voice users. We will achieve this by allocating various network resource (i.e., transmission power and transmission time), video source signal processing (i.e., adaptation by summarization) and scheduling (both "soft scheduling" in terms of deadline aware power allocation, and "hard scheduling" in a time-division-multiplexing fashion).

We will consider both uplink and downlink video streaming in this paper. In the uplink case, video users need to limit aggregate interference that they generate and affect the voice users. In the downlink case, the base station needs to limit the amount of transmission power allocated to the video users. In both cases, the optimal video streaming problem can be modeled in the framework of nonlinear constrained optimization. Two key questions that need to be answered are: 1) how to allocate resources among video users in an efficient manner (i.e., maximizing total user' quality or minimizing total users' distortion) and 2) how to make sure that the stringent delivery deadline requirements are met for every video frame that is chosen to for transmission.

In this section, we describe a solution framework that answers the above two questions. This framework involves three phases:

1) *Average resource allocation.* This is achieved by solving a network utility maximization (NUM) problem. The multi-user content diversity will be fully exploited to make efficient use of the network resources. A distributed pricing-based algorithm is proposed to achieve the resultant solution.
2) *Video source adaptations.* Based on the average resource allocation results in phase 1), each video user adapts the video source by solving a localized optimization problem with video summarization.
3) *Multiuser deadline oriented scheduling.* The network decides a transmission schedule based on video users' source decisions in phase 2), in order to meet the stringent deadline constraints of the streaming applications.

In some cases we may not be able to find a feasible schedule in phase 3). This implies that although the system resource is enough in an average sense [guaranteed by phase 1)], the deadline requirements might be too stringent to satisfy. In that case, we will go back to phase 1) and re-optimize the average resource allocation, but with more stringent resource constraints (e.g., less total power in downlink transmission). This will force the users to be more conservative when doing the source adaptations in phase $= 12$) (i.e., each user will transmit fewer frames), thus make it easier to achieve a feasible schedule in step 3).

This section will focus on the discussions of the essence of the above three phases. Further details for specific settings of uplink and downlink streaming will be given in Sections III and IV.

A. Average Resource Allocation

A key question of resource allocation for multimedia communication is how to deal with the VBR nature of the source. We take a decoupling approach in this paper, by first considering the resource allocation in the *average* sense without worrying about the time dependency. The time dependency will be brought back into the picture later in the source adaptation and multi-user scheduling phases.

Assume there are N video users in the cell. We characterize the QoS of a video user n by a utility function $U_n(x_n)$, which is an increasing and strictly concave function of the communication resource allocated to user n, x_n. This models various commonly used video quality measures such as the rate-PSNR function [28] and rate-summarization distortion functions [11]. It is well known from information theory [29] that the rate-distortion functions for a variety of sources are convex, and in practice, the operational rate distortion functions are usually convex as well. Thus, the utility functions (defined as negative distortion) are concave. For the average resource allocation phase, we assume that $U_n(x_n)$ is continuous in x_n. The average resource allocation is achieved by solving the following NUM problem,

[1]Although we focus on CDMA systems in this paper, the proposed framework in Section II is general and can be applied to other wireless access schemes [e.g., orthogonal frequency-division multiplexing (OFDM)] as well.

where $X_{\max}$ denotes the total limited resource available to the video users (i.e., total transmission power in the downlink case and total transmission time in the uplink case)

$$\max_{\{x_n \geq 0, 1 \leq n \leq N\}} \sum_n U_n(x_n), \text{s.t. } \sum_n x_n \leq X_{\max}. \quad (1)$$

Solving Problem (1) directly requires a centralized computation due to the coupling resource constraint. However, a distributed solution is often more desirable, since the base station typically does not know the utility functions of individual video users. Here we use the dual decomposition technique [30], where the base station sets a price on the resource, and each mobile user determines its average resource request depending on the announced price and its own source utility characteristic. This technique has been extensively used in network resource allocation for elastic data traffic (e.g., [24]–[27]). Here we will briefly review the main results, and details can be found in, for example, [31].

First, we relax the constraint in (1) with a dual variable λ and obtain the following Lagrangian

$$L(\boldsymbol{x}, \lambda) \triangleq \sum_n U_n(x_n) - \lambda \left(\sum_n x_n - X_{\max} \right) \quad (2)$$

where $\boldsymbol{x} = (x_n, 1 \leq n \leq N)$. The variable λ can be interpreted as the shadow price for the constrained resource $X_{\max}$. Then Problem (1) can be solved at two levels. At the lower level, each video user solves the following problem:

$$\max_{x_n \geq 0} \{U_n(x_n) - \lambda x_n\} \quad (3)$$

which corresponds to maximizing the surplus (i.e., utility minus payment) based on price λ. Denote the optimal solution of (3) as $x_n(\lambda)$, which is unique since the utility function is continuous, increasing and strictly concave. The video users then feedback the values of $x_n(\lambda)$ to the base station. At the higher level, the base station adjusts λ to solve the following problem:

$$\min_{\lambda \geq 0} g(\lambda) \triangleq \sum_n g_n(\lambda) + \lambda X_{\max} \quad (4)$$

where $g_n(\lambda)$ is the maximum value of (3) for a given value of λ. The dual function $g(\lambda)$ is nondifferentiable in general, and (4) can be solved using a subgradient searching method

$$\lambda^{l+1} = \max \left\{ 0, \lambda^l + \alpha^l \left(\sum_n x_n(\lambda^l) - X_{\max} \right) \right\} \quad (5)$$

where l is the search iteration index and α^l is a small step size at iteration l. The two level optimizations together solve the *dual* problem of the original NUM problem (1) (which we call the *primal* problem). This enables us to obtain a distributed solution. Base station controls the resource price according to (5), and each video user n chooses the average resource request x_n to maximize its surplus according to (3) in a distributed fashion. This avoids centralized computation and makes the solution scalable in a large network.

The difference between the optimal solutions of the primal and dual problems is known as the duality gap. Given the assumption on the utility functions, we have the property of strong duality [30] which implies zero duality gap. In other words, given the optimal dual solution λ^*, the corresponding $x_n(\lambda^*)$ for all n are the optimal solution of the primal problem (1). The complete distributed algorithm is given in Algorithm 1.

Algorithm 1 Dual-based Optimization Algorithm to solve Problem **(1)**

1: Initialization: set iteration index $l = 0$, and choose $0 < \epsilon \ll 1$ as the stopping criterion.

2: Base station announces an arbitrary initial price $\lambda^0 > 0$.

3: **repeat**

4: **for all** video user n **do**

5: Locally determine the resource consumption $x_n(\lambda^l) = \arg\max_{x_n} \{U_n(x_n) - \lambda^l x_n\}$.

6: Send the value of $x_n(\lambda^l)$ to the base station.

7: **end for**

8: Base station announces a new price $\lambda^{l+1} = \max\{0, \lambda^l + \alpha^l (\sum_n x_n(\lambda^l) - X_{\max})\}$.

9: $l = l + 1$.

10: **until** $|\lambda^l - \lambda^{l-1}| < \epsilon$.

Algorithm 1 converges under properly chosen step sizes, as stated in the following proposition (for proof, see [24]).

Proposition 1: If the step-sizes in (5) satisfy $\lim_{l \to \infty} \alpha^l = 0$ and $\sum_l \alpha^l \to \infty$ (e.g., $\alpha^l = 1/l$), then Algorithm 1 converges to the optimal solution of Problem (1).

So far we have not specified how Problem (3) is solved in Algorithm 1. Since the utility functions in video communications typically do not have closed form representations, Problem (3) needs to be solved by using various source adaptation techniques. This is different than, for example, congestion control in the Internet (e.g., [25]), where each source determines the transmission rate as a closed form function of the network congestion price.

B. Source Adaptation

The utility functions for elastic data traffic are typically defined on instantaneous allocated resources, such as the allocated bandwidth at time t. However, this is in general not suitable for video transmissions, due to the inter-dependent nature of the video frames. The video quality can be better determined by the total resource allocation during a time segment, which should be long enough for the user to perform source adaptation to determine the best set of frames to transmit.

In this paper, we define the utility U_n of the n-th user as the video summarization quality of a segment of K frames within a time segment of length T, denoted by $\mathcal{V}_n = \{f_n^0, f_n^1, f_n^2, \ldots, f_n^{K-1}\}$. Let us denote the corresponding *video summary* of K' frames by $\mathcal{S}_n = \{f_n^{S,0}, f_n^{S,1}, f_n^{S,2}, \ldots, f_n^{S,K'-1}\}$, where $K' \leq K$. It is assured that $f_n^{S,0} = f_n^0$ is always included in the summary. In other words, we will only send K' out of K frames through the wireless channel, due to the limited communication resources. Assuming that all K' frames can be received error-free by the receiver, the original K frame sequence can be reconstructed

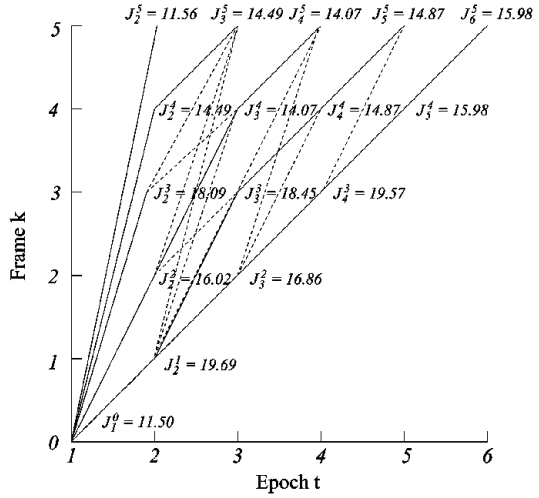


Fig. 2. Example of DP Trellis.

as $\mathcal{V}_n^S = \left\{\tilde{f}_n^0, \tilde{f}_n^1, \tilde{f}_n^2, \ldots, \tilde{f}_n^{K-1}\right\}$ by substituting the missing frames with the most recent frame that is in the summary $\mathcal{S}$. The video summary quality, which is defined as the negative of the average distortion caused by the missing frames, is given as

$$U_n(\mathcal{S}_n) = -\frac{1}{K}\sum_{k=0}^{K-1} d\left(f_n^k, \tilde{f}_n^k\right) \tag{6}$$

where $d\left(f_n^k, \tilde{f}_n^k\right)$ is the distortion between the original frame f_n^k and the reconstructed frame $\tilde{f}_n^k$. If frame f_n^k is in the summary of the K' frames, then $f_n^k = \tilde{f}_n^k$ and $d\left(f_n^k, \tilde{f}_n^k\right) = 0$. Therefore, the optimization Problem (3) can be translated into the problem of summarization with a price on the resource,

$$\max_{\mathcal{S}_n}\left\{U_n(\mathcal{S}_n) - \lambda x_n(\mathcal{S}_n)\right\}. \tag{7}$$

Remark 1: In general, the solution to (7) depends on the available adaptation schemes and the operating bit rate range of the network. Problem (7) can be solved with a *Dynamic Programming (DP)* approach. More detail can be found in [8] for the single user case. Basically, by relaxing the objective function, each candidate video summary frame in the sequence is now associated with a frame loss distortion and a bit-rate dependent on the previous video summary frame selection. Starting with the 1st frame of the sequence, a trellis is being built with edges indicating valid choices of the video summary frames. An example is shown in Fig. 2. Each node $J_{j,k}$ indicates the relaxed cost of adding frame f_k to the summary if the previous summary frame is f_j. The minimum cost choice of frame f_{j^*} is found by,

$$j^* = \min_{j<k}\left(D_{j,k} + \lambda b_{j,k}\right), \tag{8}$$

where $D_{j,k}$ is the video summary distortion for the new segment consisting of $f_j, f_{j+1}, \ldots, f_k$, and $b_{j,k}$ is the transcoding cost of predicting coding frame f_k from f_j. The minimum cost and best choice of incoming frames are computed from the trellis and then a back track program can retrieve the path to the start point and construct the optimal video summary solution for the given multiplier λ.

Remark 2: In order to utilize Algorithm 1 a, we need to find the mapping between average resource consumption x_n and summarization $\mathcal{S}_n$. For the uplink case in Section III, x_n is the total transmission time of summary frames $\mathcal{S}_n$ under a fixed transmission rate; for the downlink case in Section IV, x_n is the average transmission power needed to deliver the summary frames $\mathcal{S}_n$ within the time segment $[0, T]$.

Remark 3: By solving (3) using the summarization technique outlined here, we have moved from the continuous utility model (assumed in Section II-A) into a discrete utility model. This is because the total number of choices of the summary sequence $\mathcal{S}_n$ is finite and equals 2^K. In other words, while solving (7), user n chooses one out of 2^K possible choices of $\mathcal{S}_n$ to maximize the surplus. This also means that there exists only a finite number of choices for the corresponding $x_n(\lambda)$, and there might not be a value of λ for which $\sum_n x_n(\lambda) = X_{\max}$. That does not create a problem for the convergence of Algorithm 1, since the value of λ will still converge. However, the base station might need to announce a positive price of λ even if the total resource is not fully utilized, i.e., $\sum_n x_n(\lambda) < X_{\max}$. This is different from congestion control for elastic data traffic, where only saturated links will generate positive congestion prices.

C. Deadline Oriented Scheduling

So far we have considered average resource allocation and source adaptation, based on which each video user generates a sequence of frames to be transmitted during a given time segment. The last step is to schedule the transmissions of packets such that the delivery deadlines are met. This is essential to streaming applications. All frames have to be delivered to the receiver before their corresponding deadlines, which are determined by their positions in the original frame sequence (before summarization) and a predetermined initial delay (which allows the transmission of intra- frames that are needed at the beginning of frame sequences). The details of the scheduling algorithm will depend on the physical model of the communication networks. In the uplink case where transmissions from various users interferer each other, we propose time division multiplexing (TDM) among video users to ensure a high enough transmission rate (voice users still transmit constantly in the background). In the downlink case where transmissions are orthogonal to each other, we propose to schedule users to transmit simultaneously, with each user's transmission power determined by its current frame size, the corresponding deadline, as well as the resource consumption of other users. Further details will be given in Sections III and IV, respectively.

III. Wireless Uplink Streaming

A. Problem Formulation

In a wireless CDMA network, different users transmit using different spreading codes. These codes are mathematically orthogonal under synchronous reception. However, the orthogonality is partially destroyed when the transmissions are asynchronous, such as in the uplink transmissions. The received SINR in that case is determined by the users' transmission power, the spreading factors (defined as the ratio of the bandwidth and the achieved rate), the modulation scheme used, and the background noise. The maximum constrained resource

of the video users can be expressed as the maximum received power at the base station, derived based on a physical layer model similar as the one used in [32].

We consider the uplink transmission in a single CDMA cell with M voice users and N video streaming users. The total bandwidth W is fixed and shared by all users. Each voice user has a QoS requirement represented in bit-error rates (BER) [or frame-error rates (FER)], which can be translated into a target SINR at the base station, γ_{voice}. Each voice user also has a target rate constraint R_{voice}. Assuming perfect power control, each voice user achieves the same received power at the base station, P^r_{voice}. The total received power at the base station from all video users is denoted as $P^{r,all}_{\text{video}}$. The background noise n_0 is fixed and includes both thermal noise and inter-cell interferences.

In order to support the successful transmission of all voice users, we need to satisfy

$$\frac{W}{R_{\text{voice}}}\frac{G_{\text{voice}}P^r_{\text{voice}}}{n_0W+(M-1)P^r_{\text{voice}}+P^{r,all}_{\text{video}}}\geq\gamma_{\text{voice}}. \quad (9)$$

Here W/R_{voice} is the spreading factor, and coefficient G_{voice} reflects the fixed modulation and coding schemes used by all voice users (e.g., $G_{\text{voice}}=1$ for BPSK and $G_{\text{voice}}=2$ for QPSK). For each voice user, the received interference comes from the other $M-1$ voice users and all video users. From (9), we can solve for the maximum allowed value of $P^{r,all}_{\text{video}}$, denoted as $P^{r,\max}_{\text{video}}$

$$P^{r,\max}_{\text{video}}=\left(\frac{WG_{\text{voice}}}{R_{\text{voice}}\gamma_{\text{voice}}}-(M-1)\right)P^r_{\text{voice}}-n_0W \quad (10)$$

which is assumed to be fixed given fixed number of voice users M.

The network objective is to choose the transmission power of each video user during a time segment $[0,T]$, such that the total video's utility is maximized, i.e.,

$$\begin{aligned}\max_{\{P_n(t),1\leq n\leq N\}}\quad & \sum_{n=1}^{N}U_n\left(\int_0^T R_n\left(\boldsymbol{P}(t)\right)\right)dt\\ \text{s.t.}\quad & \sum_{n=1}^{N}h_nP_n(t)\leq P^{r,\max}_{\text{video}}\quad\forall\, t\in[0,T]\\ & 0\leq P_n(t)\leq P^{\max}_n,\quad 1\leq n\leq N \quad (11)\end{aligned}$$

where $P_n(t)$ is the transmission power of video user n at time t, $\boldsymbol{P}(t)$ is the vector of all video users' transmission power at time t, $P^{\max}_n$ is the maximum peak transmission power of user n, and h_n is the fixed channel gain from the transmitter of user n to the base station. $R_n(t)$ is the rate achieved by user n at time t, and depends on all video users' transmission power, the channel gains, the background noise, and interference from voice users. A user n's utility function is defined on the video summarization quality of its transmitted sequence during $[0,T]$, as discussed in Section II-B.

Remark 4: Problem (11) is not a special case of Problem (1), since 1) Problem (11) optimizes over N functions $(P_n(t),1\leq n\leq N)$, whereas Problem (1) optimizes over N variables $(x_n,1\leq n\leq N)$, and 2) the objective function in Problem (11) is coupled across users, whereas the objective in Problem (1) is fully decoupled. This makes (11) difficult to solve in a distributed fashion.

In order to solve Problem (11), we will resort to the framework described in Section II, where we will perform average resource allocation (in terms of average transmission power), source adaptation (to match the average resource allocation), and the deadline scheduling (to determine the exact power allocation functions by deadline aware water-filling).

B. Transmission Time Allocation and Source Adaptation

To simplify the problem and make the solution tractable, we consider the case where video users transmit in a TDM fashion. This is motivated by [33], where the authors showed that in order to achieve maximum total rate in a CDMA uplink, it is better to transmit weak power users in groups and strong power users one by one. Since video users typically need to achieve much higher rate than voice users (thus transmit at much higher power), it is reasonable to avoid simultaneous transmissions among video users, and thus avoid large mutual interference. A more important motivation for TDM transmission here is to exploit the temporal variation of the video contents, i.e., content diversity. Under such a TDM transmission scheme, the constraint resource to be allocated to the video users becomes the total transmission time of length T. The total number of bits that can be transmitted by user n is determined by the transmission time allocated to it, $t_n\in[0,T]$, and the maximum rate it can achieve while it is allowed to transmit. Let us denote this rate as R^{TDM}_n, and it can be calculated by

$$R^{\text{TDM}}_n=W\log_2\left(1+\frac{\min\left\{h_nP^{\max}_n,P^{r,\max}_{\text{video}}\right\}}{n_0W+MP^r_{\text{voice}}}\right). \quad (12)$$

Under the assumption of TDM transmission, Problem (11) can be written as follows:

$$\max_{\{t_n\geq 0,1\leq n\leq N\}}\sum_{n=1}^{N}\tilde{U}_n(t_n),\quad\text{s.t.}\ \sum_{n=1}^{N}t_n\leq T \quad (13)$$

where the new utility function $\tilde{U}_n$ is defined as

$$\tilde{U}_n(t_n)=U_n\left(R^{\text{TDM}}_nt_n\right) \quad (14)$$

i.e., a user n's total transmitted data during time $[0,T]$ is determined by the product of R^{TDM}_n and the active transmission time t_n. Now Problem (13) is a special case of Problem (1), where we replace U_n by $\tilde{U}_n$, x_n by t_n and $X_{\max}$ by T. As a result, the optimal transmission time allocation per user can be found using Algorithm 1.

Based on the discussions in Section II-B, each user locally adapts its source using summarization, which leads to the best sequence of video frames that fit into the transmission time allocation t_n. The transmission of each frame needs to meet a certain delivery deadline, after which the frame becomes useless. This requires the base station to determine a transmission schedule for all users, which will be explained next.

C. Uplink Greedy Scheduling

Our objective is to find a transmission schedule, such that all frames meet their delivery deadline, subject to a causality constraint. In a TDM based transmission, since a user n's transmission rate R^{TDM}_n is fixed, so is the transmission time of its kth summary frame. It can be calculated as $B^{S,k}_n/R^{\text{TDM}}_n$, where $B^{S,k}_n$ is the size of the frame (in bits).

In order to calculate the value of R_n^{TDM} according to (12), the user needs to know the following information: 1) the background noise plus interference $n_0W + MP_{\mathrm{voice}}^r$, and the maximum received power $P_{\mathrm{video}}^{r,\max}$. These values do not change frequently and they need to be fed back from the base station to the user only once in a while; 2) the channel gain h_n, which needs to be updated with a frequency dependent on the moving speed of the user; 3) bandwidth W, which is a fixed and publicly known parameter.

Given this information, the users calculate the transmission time for each of the summary frames, and send this information along with the absolute delivery deadline for each frame to the base station. The correspondingly signaling overhead is not significant compared with the transmission rate of the video users.[2] The base station makes the scheduling decisions based on the GREEDY approach, where the frames from all users are sorted and transmitted one by one according to their deadlines (with the earliest deadline first).

Although the GREEDY scheduling is simple, it is optimal among all TDM-based schedules.

Proposition 2: If any TDM-based scheduling algorithm can meet the deadlines of all video frames, so can the GREEDY scheduling algorithm.

Proposition 2 can be proved as follows: select any TDM-based scheduling algorithm where all deadlines are met and one or more frames are transmitted out of the deadline order. Then by rearranging the corresponding out of order frames by the deadline as in the GREEDY algorithm, all the deadline constraints are still satisfied.

If the GREEDY schedule can not meet all frames deadlines, users need to go back and solve again Problem (13) again, where the total transmission time constraint T is replaced by a value $T' < T$. In other words, the total resource constraint needs to be reduced such that the corresponding summary frames become schedulable. The complete uplink Joint Resource Allocation and Scheduling (JARS) algorithm is given in Algorithm 2.

Algorithm 2 JARS Algorithm for Video Streaming over Wireless Uplink Channels

1: Initialization: let $T' = T$ and choose $0 < \epsilon_T \ll 1$.

2: **repeat**

3: Solve Problem **(13)** using Algorithm 1 in a distributed fashion (replacing U_n by $\tilde{U}_n$ in **(14)**, x_n by t_n, and $X_{\max}$ by T').

4: **for all** user n **do**

5: Determine a summary sequence as in **Section II-B**.

6: Calculate rate R_n^{TDM} according to **(12)**.

7: Calculate transmission time for each summary frame.

8: Send the transmission time and deadline information of all summary frames to the base station.

9: **end for**

10: Base station sorts the frames in increasing order of deadlines, and determines the transmission starting and ending time of each frame accordingly.

11: If there is deadline violation, let $T' = (1 - \epsilon_T)T'$.

12: **until** no deadline violation occurs for any user.

13: Base station informs all users of the schedule, and users transmit accordingly.

D. Computational Complexity and Communication Overhead of Algorithm 2

In this section, we show that the proposed algorithm has both low computational complexity and communication overhead, and thus is very scalable with the network size. We analyze each of the three components of the algorithm as follows:

1) *Average Resource Allocation:* From the base station's point of view, this involves searching for the optimal dual price λ. Since the total resource demand (transmission time) is monotonic in the price, we can find the optimal price by simply using bisection search. For example, if we want to achieve a precision of 10^{-10} of the optimal price, we only need to have at most $\log_2(1/10^{-10}) = 34$ price iterations. Thus, the computational complexity of this step is independent of the number of users, and we denote it as C_1. Since the number of price announcements from the base station is the same as the number of iterations needed, the total communication overhead of the base station is also independent of the number of users, and we denote it as L_1.
2) *Individual Source Adaptation:* Under each fixed dual price, each mobile user performs source adaptation independently. The computational complexity of each user is independent of the number of users. The total computational complexity (over all users) of this step is linear with the number of users. We denote it as C_2N, where N is the total number of video users. Each user then need to report its resource demand back to the station, with a total communication overhead (over all users) of L_2N.
3) *Deadline Oriented Scheduling:* This involves sorting all users' frames accordingly to their deadlines. Under a fixed total system resource (total transmission time), the maximum number of frames that can supported is upper-bounded (since in practice we have a minimum frame size) and independent of the number of users. Thus, the worst-case computational complexity of this step is also independent of the number of users, and we denote it as C_3. The communication overhead involves users reporting the lengths and locations of their frames and the base station announcing the final schedule. Accordingly to the same argument (upperbound on the number of total frames), the worst-case communication overhead is also independent of the number of users, and we denote it as L_3.

Furthermore, if the scheduling phase does not lead to a feasible schedule, we need to go back and rerun the average resource allocation with a more stringent total transmission time

[2]As an example, consider the case where each video user has a maximum of 90 evenly spaced frames within in a scheduling interval of 3 s. It only takes 7 bits to specify the location of any of the 90 frames. If we further assume that each frame is no larger than 1 Mbits, it only takes no more than 20 bits to specify the size of each frame. In total, the worst case communication overhead with 4 video users and an average of 90 frames per 3 s is upperbounded by $(7 + 20) * 30 * 4/3/1000 = 1.08$ kbps, which is very small compared with the average transmission rate of video users (around 30 kbps in our numerical examples).

constraint. The total number of such iterations can be upper-bounded by Z and is again independent of the number of users.

As a conclusion, the worst-case computational complexity of the proposed algorithm is $Z(C_1C_2N+C_3)$, and the worst-case communication overhead is $Z(L_1L_2N+L_3)$, both are linear in the number of users. However, in terms of the base station and each mobile user, the computation complexity and communication overhead is independent of the total number of users. This shows that the proposed algorithm scales well with the network size.

IV. Wireless Downlink Streaming

Different from the uplink case, transmissions in the downlink are orthogonal to each other, thus it is desirable to allow simultaneous transmissions of multiple video users. The resource constraint in the downlink case is the maximum peak transmission power at the base station. The objective here is to determine the transmission power functions, $P_n(t)$, of each user n during time $t \in [0,T]$, such that the total user utility (measured in video quality) is maximized.

A. Problem Formulation and Average Power Allocation

Following the framework described in Section II, the first step is to perform average resource allocation. For the downlink case, we will allocate the transmission power to each user, subject to the total transmission power constraint (for video users) at the base station, $P_{\max}^{\text{base}}$. Since there is no mutual interference, the transmissions of the voice users need not be taken into consideration when determining the achievable rates of the video users.

At this stage, we will temporality assume that each user n will transmit at a fixed power level P_n throughout the time segment $[0,T]$. The problem we want to solve is

$$\max_{\{P_n \geq 0, 1 \leq n \leq N\}} \sum_{n=1}^{N} U_n(P_n), \quad \text{s.t.} \sum_{n=1}^{N} P_n \leq P_{\max}^{\text{base}}. \quad (15)$$

Problem (15) is a special case of Problem (1), and can be solved using Algorithm 1. Assuming that user n is allocated a constant transmission power P_n^*, its total throughput within $[0,T]$ is given by

$$TW \log_2\left(1+\frac{h_n P_n^*}{n_0 W}\right) \quad (16)$$

where h_n is the channel gain from base station to the mobile receiver, and n_0 is the background noise density at the receiver end. The user can determine its best video summary sequence based on this achieved throughput.

Due to the difference in frame sizes and locations, transmitting at constant power levels is not optimal in terms of meeting the frame delivery deadlines. Next we present an energy-efficient water-filling power allocation algorithm based on the solution of Problem (15).

B. Frame Scheduling With Greedy Water-filling Power Allocation

Next we develop an energy-efficient scheduler that tries to meet the deadlines of the frames for all users with a minimum amount of power. Compared with the uplink case, the users can transmit simultaneously in the downlink case without generating interference. The key concern is how to choose a transmit power function of each user n, $P_n(t)$ during $t \in [0,T]$, which can meet the frame delivery deadlines without violating the total power constraint, $\sum_n P_n(t) \leq P_{\max}^{\text{base}}$. This is achieved by a sequential scheduling algorithm based on a water-filling solution over *the transmission power that has been allocated*.

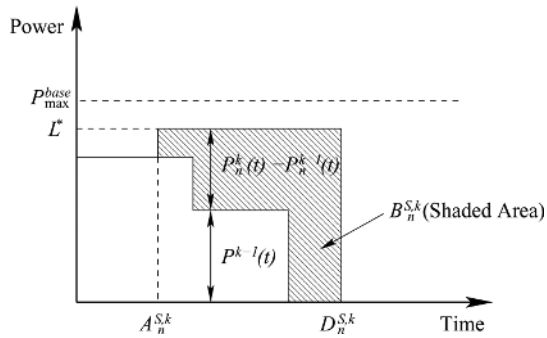


Fig. 3. Greedy water-filling transmission power allocation.

First, similarly to the uplink case, we sort the frames of all users in an increasing order of delivery deadlines. If the kth frame in the sequence belongs to user n, we will denote its frame size, frame arrival time, and delivery deadline as $\{B_n^{S,k}, A_n^{S,k}, D_n^{S,k}\}$, with the superscript S denoting summarization and $A_n^{S,k} < D_n^{S,k}$.

Then the scheduling is performed one frame at a time, starting from the frame with the earliest deadline. Assume that we have completed the scheduling up to the $(k-1)$st frame in the sequence, where the transmission power allocated to a user $j \in \{1,\ldots,N\}$ is $P_j^{k-1}(t)$ for $t \in [0,T]$. Notice that the power will be zero for any time $t > D_n^{S,k}$, since all the frames scheduled have deadlines smaller than $D_n^{S,k}$. Also let the total allocated transmission power to be $P^{k-1}(t) = \sum_j P_j^{k-1}(t)$. Assuming that the kth frame belongs to user n, the allocated transmission power to user n after the kth frame is scheduled, $P_n^k(t)$, will satisfy

$$P_n^k(t) - P_n^{k-1}(t) = \begin{cases} L - P^{k-1}(t), & t \in \left[A_n^{S,k}, D_n^{S,k}\right] \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

where L is the water-level. The extra amount of information $B(L)$ that user n can transmit during time $\left[A_n^{S,k}, D_n^{S,k}\right]$ can be computed as a function of L

$$B(L) = W \int_{A_n^{S,k}}^{D_n^{S,k}} \left(\log_2\left(1+\frac{h_n P_n^k(t)}{n_0 W}\right) - \log_2\left(1+\frac{h_n P_n^{k-1}(t)}{n_0 W}\right)\right) dt \quad (18)$$

and a fast bisection search can be performed to find the optimal value of L^* such that the kth frame can be transmitted before the deadline, i.e., $B(L^*) = B_n^{S,k}$. This is a greedy type of water filling solution and tries to satisfy the delivery deadline of the current frame with the minimum amount of total power (summed over all users). A graphical illustration of the water-filling algorithm is given in Fig. 3.

The algorithm does not stop until the power function corresponding to the last frame is computed. Each user n's complete transmission power function is then $P_n(t) = P_n^K(t)$, where

K is the total number of frames for all users. Notice that although the resulting $P_n(t)$'s may not be constant functions, the scheduler tries to spread transmission as much as possible over time such that the total power used at each time is minimum. This has the same flavor as the "lazy scheduling" in [34], which showed that the total energy consumption for transmitting a fixed amount of data decreases as the transmission time increases. Instead of focusing on data transmission in a single user environment, here we focus on the multimedia transmission in a multi-user environment.

If the water-filing algorithm leads to a peak transmission power greater than $P_{\max}^{\text{base}}$, users need to go back and solve Problem (15) again, where the maximum peak power constraint $P_{\max}^{\text{base}}$ is replaced with $P_{\max}^{\prime\text{base}} < P_{\max}^{\text{base}}$. In other words, the total resource constraint needs to be reduced such that the resultant summary frames can be schedulable. The complete downlink JARS algorithm is given in Algorithm 3.

Algorithm 3 JARS Algorithm for Video Streaming over Wireless Downlink Channels

1: Initialization: let peak power constraint be $P_{\max}^{\prime\text{base}} = P_{\max}^{\text{base}}$, and choose a resource constraint reduction factor $0 < \epsilon_P \ll 1$.

2: **repeat**

3: Video users and the base station solve Problem **(15)** using Algorithm 1 in a distributed fashion (replacing x_n by P_n and $X_{\max}$ by $P_{\max}^{\prime\text{base}}$).

4: **for all** user n **do**

5: Determine a summary video sequence as in **Section II-B**.

6: Send the frame sizes and frame deadlines to the base station.

7: **end for**

8: Base station sort the frames in increasing order of deadlines, and determines transmission power for each user using greedy water-filling scheduling.

9: If the peak power constraint is violated, let $P_{\max}^{\prime\text{base}} \Leftarrow (1-\epsilon_P)P_{\max}^{\prime\text{base}}$.

10: **until** no peak power violation at any time.

11: Base station transmits utilizing the computed transmission power functions.

C. Computational Complexity and Communication Overhead of Algorithm 3

Similar as the uplink case, we will analyze the computational complexity and communication overhead of the proposed JARS algorithm for the downlink case.

The computational complexity of the downlink algorithm is $Z(C_1C_2N+C_3)$, where Z is the maximum number of iterations that the base station need to rerun the algorithm (due to infeasible schedule), C_1 denotes the maximum number of iterations for searching the optimal dual price, C_2 denotes the maximum number of computation needed for a user to perform single user source adaptation, and C_3 represents the product of the maximum number of frames to be scheduled and the maximum iterations of bisection search needed for finding the appropriate water-filling level for each frame. Notice that in the downlink case all computation happens at the base station, where in the uplink case the computation is distributed among the base station and the mobile users.

The communication overhead for the downlink algorithm is zero, since the base station knows the information of all users' frames and performs all computation locally without any further message exchange with the users.

TABLE I
KEY SIMULATION PARAMETERS—PART 1

Symbol	Notation	Value
W	Bandwidth	1.223 MHz
n_0	Noise density	$8.3 * 10^{-7}$ mW/Hz
γ_{voice}	Voice target SINR	6 dB
P_{voice}	Voice received power	1 mW
G_{voice}	Voice modulation factor	1 (BPSK)
R_{voice}	Voice transmission rate	9.6 kbps

V. EXPERIMENTAL RESULTS

We choose four video clips with different content activity levels. Clips 1 and 2 are, respectively, frames 150–239 and frames 240–329 from the "foreman" sequence, and clips 3 and 4 are respectively frames 50–139 and 140–229 from the "mother-daughter" sequence, respectively. The codec used is H.263, and GoP structure is IPPPP. There are 90 frames within each video clip at a frame rate of 30 Hz, which corresponds to a time segment of $T = 3$ s. We will use these clips for both uplink and downlink streaming simulations.

A. Simulation Results for the Uplink Case

In the uplink case, besides the GREEDY scheduling algorithm, we also simulate the case where all four video users are allowed to transmit simultaneously with the same constant rates (SIMCONST). In other words, the received power from each of the video user is the same at the base station in SIMCONST algorithm, and no scheduling across users is needed due to simultaneous transmissions.

In Table I we list the simulation parameters that are kept constant throughout this subsection.

1) Achievable Video Rate Under Different Number of Voice Users: We first compare the video users' total achievable rate under GREEDY and SIMCONST algorithms for different voice user load. Under GREEDY, we plot the maximum rate achieved by allowing only one user transmitting. Under SIMCONST, we plot the total rate achieved by all four users. Fig. 4 shows that the video users' total achievable rate decreases with the number of voice users, and becomes zero when there are more than 26 voice users in the system. In other words, the system's ability of supporting video users depends on the current voice load in the cell. It is also clear that the GREEDY algorithm always outperforms the SIMCONST algorithm in terms of total achievable rate, due to the heavy mutual interference among users in the latter case. Later we will show that due to the ability of exploiting content diversity, the GREEDY algorithm achieves much better performance than the SIMCONST algorithm in terms of the distortion experienced by the users.

2) Transmission Time Allocation and Source Rate Adaptation: As a concrete example, let us consider a cell with 24 voice users, where the GREEDY algorithm achieves a rate of 120 kbps (for the single active user) and SIMCONST algorithm offers a rate of 29 kbps for each of the four video users at the same

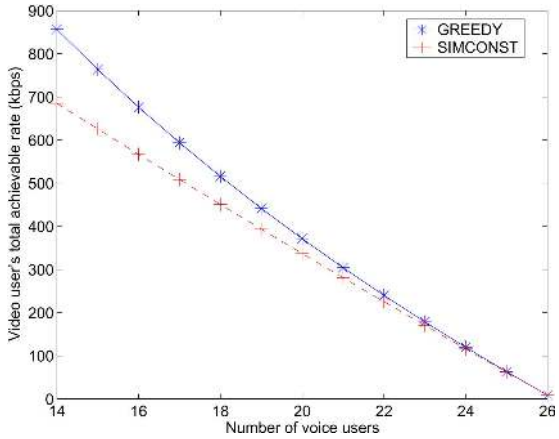


Fig. 4. Comparison of total achievable rate between GREEDY and SIMCONST algorithms.

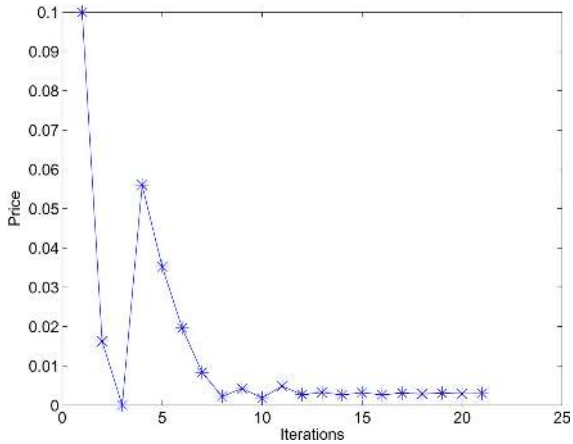


Fig. 5. Dual price iteration.

TABLE II
KEY SIMULATION PARAMETERS—PART 2

Symbol	Notation	Value
M	Number of voice users	24
$P_{\max}$	Maximum video received power	2.16 mW
R_{TDM}	Single user video rate under GREEDY	120 kbps
R_{CR}	Single user video rate under SIMCONST	29 kbps

time (Table II). First consider the pricing-based rate control algorithm. Based on the assumption of TDM scheduling, pricing on transmission time is equivalent to pricing on the achievable rate. We start from an initial price $\lambda = 0.1$, and use diminishing step-sizes $\alpha^n = 0.05/n$ that satisfy the conditions in Proposition 1. The iteration stops when the total transmission time of the four video users achieves more than 99% of the time segment length (i.e., 3 s). Fig. 5 shows the convergence of price in 21 iterations, with a final optimal price equal to $\lambda^* = 2.9{*}10^{-3}$.

Fig. 6 shows how the summary distortion per frame of each individual user decreases (or increases) as the price decreases (or increases). Depending on the video contents that determine the specific rate-distortion functions, users experience different levels of distortions under the same price. Among the four users, user 2 experiences the largest distortion due to the large temporal variations of its content. Users 3 and 4 achieve similar

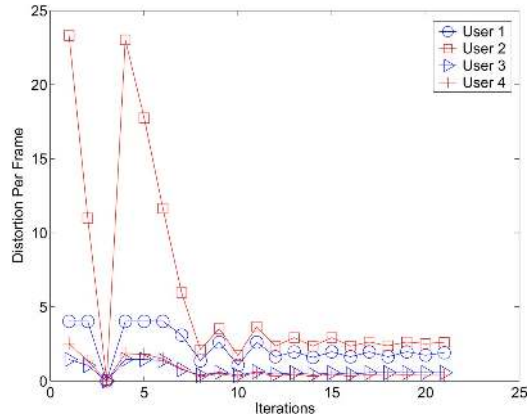


Fig. 6. Users' distortion iteration.

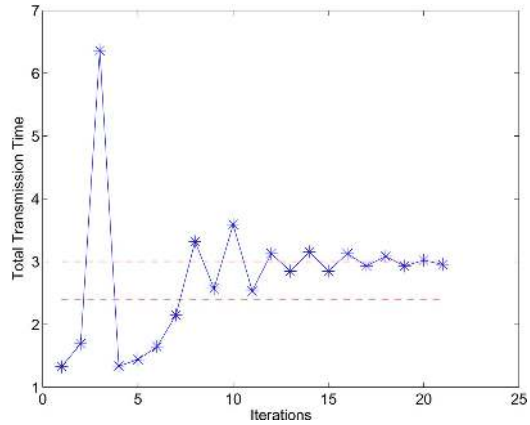


Fig. 7. Transmission time iteration.

distortions that are much smaller than that of users 1 and 2, due to the small time variations in the contents.

Fig. 7 shows how the total transmission time of the summarized frames changes during the iteration. If we relax the convergence criterion from 99% to 80% (i.e., the price converges when it first enters the region bounded by the two dashed lines in Fig. 7), then the convergence is achieved in nine iterations. This reflects a trade-off between the computational complexity and resource utilization efficiency. A system designer needs to carefully choose the iteration parameters to tradeoff the convergence speed and performance. In general, the convergence speed of the pricing algorithm depends on the video contents, the initial price, the choice of step-sizes, and the stopping criterion. Except for the video contents, which can not be adjusted by the system, all other factors can be continuously tuned based on experiences to offer the best tradeoff of convergence and performance. Typically the requirement of faster convergence inevitably leads to degraded performance since the resource (transmission time) may not be fully unutilized (e.g., by reducing the stopping criterion from 99% to 80%). This tradeoff becomes more important as the number of video users increases.

The resulting video summary distortions based on the optimal price λ^* are plotted in Fig. 8. The vertical arrows indicate video summary frame locations in the sequence. Notice that the distortion is zero at summary frame locations, since the received frames are exactly the same as the original frames before summarization. The optimal price gives a good tradeoff

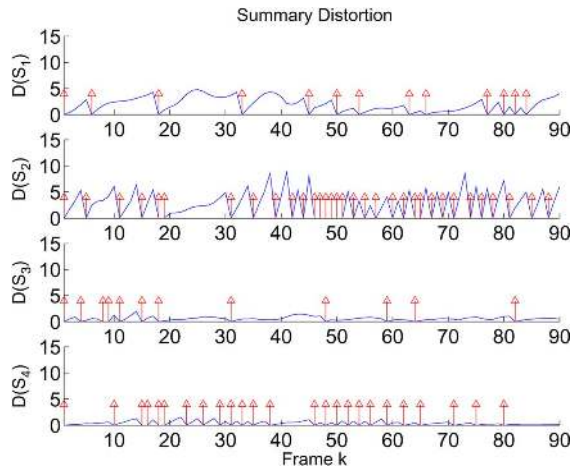


Fig. 8. Resulting video summary distortion at optimal pricing.

Fig. 10. Frame delivery deadlines under GREEDY scheduling algorithm.

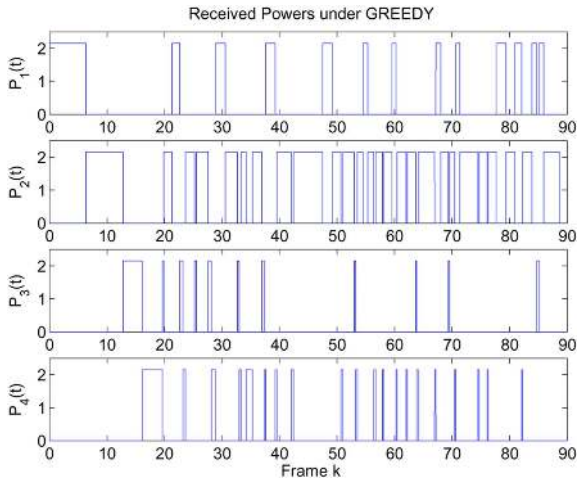


Fig. 9. Video users' received powers at the base station under GREEDY scheduling algorithm (power is measured in milliwatts).

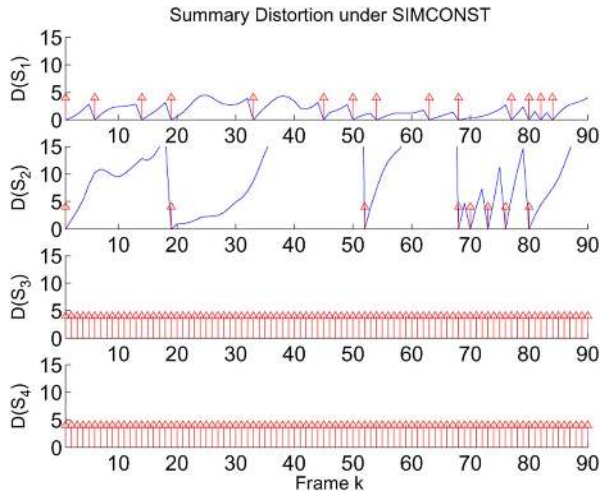


Fig. 11. Resulting summary distortion under SIMCONST scheme.

between total transmitting time and total video summary distortions. Clips 1 and 2 are coded at an average PSNR of 27.8 dB, and clips 3 and 4 at 31.0 dB. The resulting average bit rates for the four clips are 27, 68, 9.1, and 13.1 kbps, respectively.

3) Greedy Scheduling: Given the summarization results, the GREEDY algorithm performs scheduling based on sorted packet deadlines. The corresponding received power functions of users are plotted in Fig. 9 and the corresponding delivery deadlines are plotted in Fig. 10. Under an initial delay of 30 frames (1 s), the GREEDY algorithm successfully transmits all frames within 3 s and meets all deadline requirements.

Remark 5: As we mentioned in Section III, if the current summary frames can not be scheduled (i.e., deadline violation occurs), then the base station needs to increase the price and let the users recompute the summarizations. However, in all simulations that we performed, the summarization results from the pricing-based rate control are always schedulable. This is due to the fact that by taking advantage of the multi-user content diversity, the deadline requirements of the summary frames are typically spread out through the time segment, thus it is relatively easy to satisfy the scheduling constraints. This implies that as long as there are enough content differences among the video users, the two stages of the algorithm can be operated separately in practice.

4) Simconst Algorithm: In the SIMCONST algorithm, users perform summarizations based on the same guaranteed rate, so that all the summary frames can be transmitted within their individual deadline constraints. The resulting summary distortions are shown in Fig. 11. The averaged distortions per frame for all users are 1.74, 15.98, 0, and 0, respectively, with a total distortion per frame equal to 17.72. For comparison purposes, the averaged distortions per frame for all users achieved under pricing-based rate control are 1.90, 2.61, 0.56, and 0.37, respectively, with a total distortion per frame equal to 5.43. Under SIMCONST, user 2 encounters a much larger distortion due to its busy content. As a result, the total distortion per frame increases more than 200% from the pricing-based approach to SIMCONST.

Remark 6: We can also think the SIMCONST algorithm as a special case of the class of "reservation based" algorithms, where each user is reserved (guaranteed) a constant bit pipe during the transmission. Since users can not shared resource with each other under the reservation based algorithms, the resultant performance is typically much worse than our proposed algorithm where the content diversity among users is exploited.

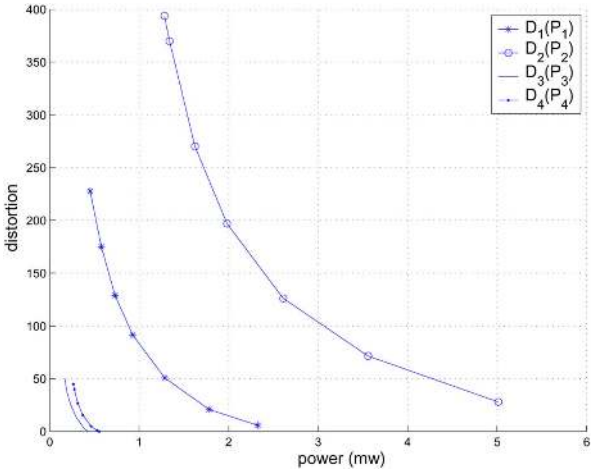


Fig. 12. Utility-average power functions for different clips.

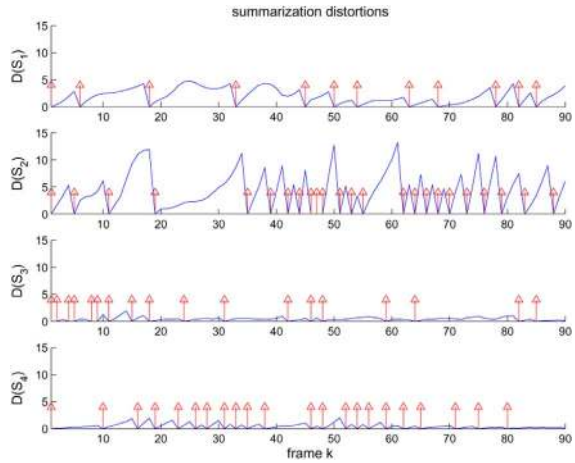


Fig. 13. Resulting video summary distortion for $P_{\max} = 2.4$.

B. Simulation for the Downlink Case

For the downlink case, channel gains are given as $H = [0.75, 1.00, 0.80, 0.65]$. This in conjunction with the clip choices, is intended to cover a range of activity levels and reflects diversity in utility as function of the transmitting power levels. Fig. 12 shows the summary distortion versus the average transmission power for these four users.

At the summarization-power allocation phase, a total transmitting power threshold of $P_{\max} = 2.4$ mW is used, and the optimal price is found to be equal to $\lambda^* = 101.45$ through the price iteration.

The resulting video summary distortions are plotted in Fig. 13. The vertical arrows indicate video summary frame locations in the sequence. Notice that the distortion is zero at these locations. The optimal price gives the best trade-off between total transmitting power and total video summary distortion. Clips 1 and 2 are coded at an average PSNR of 27.8 dB, and clips 3 and 4 at 31.0 dB. The resulting average bit rates for the four clips are 20.1, 43.3, 8.1, and 9.4 kbps, respectively. If an equal power allocation scheme is used instead, i.e., each user is allocated a power level of $P_{\max}/4 = 0.6$ mW, it is clear that clips 1 and 2 will suffer large distortions, while clips 3 and 4 will have virtually no distortions. This is not a good allocation of resources if we want to achieve best total quality.

The joint water-filing scheduler achieves a total power limit of $P_{\max} = 2.45$ mW. There is a slight loss of power efficiency through the summarization and power allocation phase that only considers the average transmitting power.

The power allocation results, $P_1(t) \sim P_4(t)$, for the video summaries generated in Fig. 13 are shown in Figs. 14 and 15. The dotted line is the total power function $P(t)$. Although each user's transmission power function is not constant over time, the total transmission power function is rather flat and achieves efficient utilization of the power resource. As a comparison, the results based on the single-user earliest deadline first serve (EDFS) scheduling are plotted in Fig. 15, which leads to a maximum power of $P_{\max} = 7.56$ mW. The efficiency of joint power scheduling is clearly demonstrated in this case.

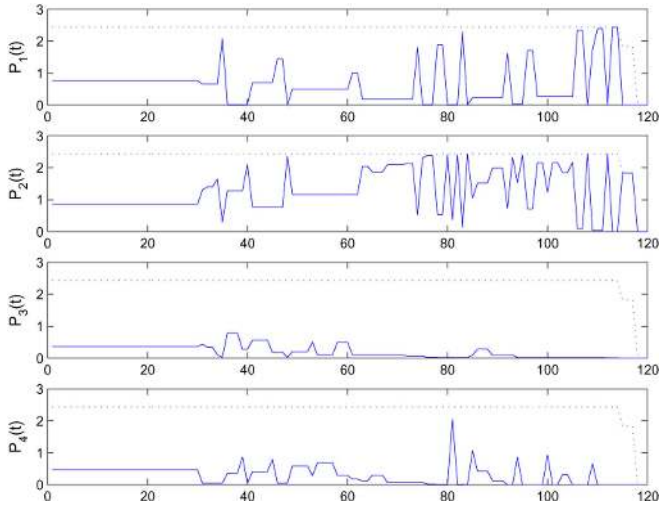


Fig. 14. Deadline-driven water-filling scheduling result. Solid lines represent the transmission power for each user, and the dotted line is the total transmission power at the base station. The horizontal axis represents the number of frames.

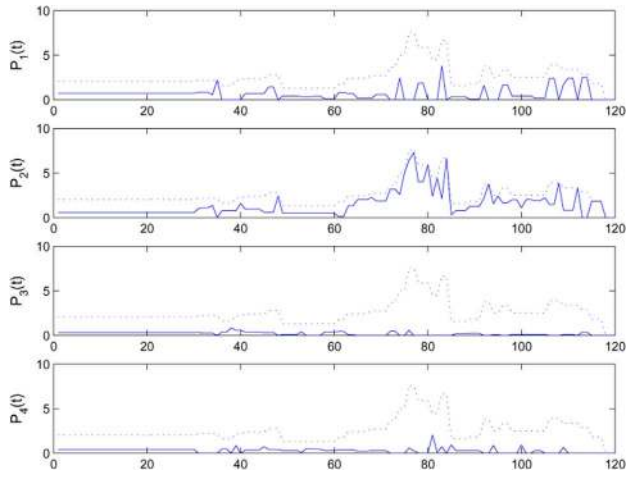


Fig. 15. Single-user based scheduling result. Solid lines represent the transmission power for each user, and the dotted line is the total transmission power at the base station. The horizontal axis represents the number of frames.

VI. Conclusion and Future Work

Traditional network engineering treats traffic as generic stream of data bits, while top down video streaming solutions view network as bit pipes of different capabilities with a loose coupling between the operations in Application layer and various layers below the application layer. As video is becoming the dominant traffic in network applications, it is essential to consider a joint optimization of video adaptation and network resource allocations to take full advantages of various features offered at different layers.

In this paper, we considered efficient multi-user video streaming over the existing wireless networks. Our objective is to maximize the total reception quality of a limited number of video users, without interrupting the service of the existing voice users. Since the video sequences in this case can only be supported at a very low bit rate, it is very important to jointly optimize both the resource allocation across users and the individual video source adaptations to achieve the best results. We formulated the problem as a NUM problem, and developed a class of joint resource allocation and scheduling algorithms for both uplink and downlink video streaming scenarios. There are three key phases in the algorithms: average resource allocation by a dual-based pricing algorithm, individual source adaptation by smart summarization, and scheduling that takes advantage of multi-user content diversity. The resulting algorithms have provable convergence, low computational complexity, and small communication overhead. They also achieve much better overall received video quality and resource utilization efficiency compared with the algorithms that are content blind. The algorithms also enjoy the benefit of distributing the computational complexity among mobile users by coordinating individual video summarization via a low overhead pricing scheme. In particular, in the four video user case that we simulated, the proposed algorithms reduce video distortion by more than 2/3 (under a fixed total network resource) in the uplink case and reduce peak resource consumption by 2/3 (with fixed total video reception quality) in the downlink case.

References

[1] *Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System, tia/eia Interim Standard 95(is-95-a)*. Washington, DC, Telecommunications Industry Association, 1995.

[2] TIA/EIA IS-856 CDMA 2000: High Rate Packet Data Air Interface Specification 2000.

[3] J. Ohm, "Advances in scalable video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 42–56, Jan. 2005.

[4] H. Radha and M. Chen, "The MPEG-4 fine-grained scalable video coding method for multimediastreaming over IP," *IEEE Trans. Multimedia*, vol. 3, no. 1, pp. 53–68, 2001.

[5] B. Kim, Z. Xiong, and W. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 8, pp. 1374–1387, Oct. 2000.

[6] J. Xin, C. Lin, and M. Sun, "Digital video transcoding," *Proc. IEEE*, vol. 93, no. 1, pp. 84–97, Jan. 2005.

[7] A. Vetro and C. Sun, "Video transcoding architectures and techniques: An overview," *IEEE Signal Process. Mag.*, vol. 20, no. 2, pp. 18–29, Apr. 2003.

[8] Z. Li, F. Zhai, and A. Katsaggelos, "Video summarization for energy efficient wireless streaming," in *Proc. SPIE Opt. Fibers: Technol.*, Rayss, Jan, Culshaw, Brian, Mignani, and G. Anna, Eds., 2005, vol. 5960, pp. 763–774.

[9] S. F. Chang and A. Vetro, "Video adaptation: Concepts, technologies, and open issues," *Proc. IEEE*, vol. 93, no. 1, pp. 148–158, Jan. 2005.

[10] B. Manjunath, J. Ohm, V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.

[11] Z. Li, G. Schuster, A. Katsaggelos, and B. Gandhi, "Rate-distortion optimal video summary generation," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1550–1560, Oct. 2005.

[12] A. Katsaggelos, Y. Eisenberg, F. Zhai, R. Berry, and T. Pappas, "Advances in efficient resource allocation for packet-based real-time video transmission," *Proc. IEEE*, vol. 93, no. 1, pp. 135–147, Jan. 2005.

[13] F. Zhai, C. Luna, Y. Eisenberg, T. Pappas, R. Berry, and A. Katsaggelos, "Joint source coding and packet classification for real-time video transmission over differentiated services networks," *IEEE Trans. Multimedia*, vol. 7, no. 4, pp. 716–726, Apr. 2005.

[14] M. v. d. Schaar and S. Xu, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 wlans," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1752–1763, Oct. 2003.

[15] S. Zhao, Z. Xiong, and X. Wang, "Joint error control and power allocation for video transmissionover CDMA networks with multi-user detection," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 12, no. 6, pp. 425–437, Jun. 2002.

[16] T. Yoo, E. Setton, X. Zhu, A. Goldsmith, and B. Girod, "Cross-layer design for video streaming over wireless ad hoc networks," in *Proc. MMSP*, Sienna, Italy, Oct. 2004, pp. 99–102.

[17] J. Shin, J. Kim, and C. Kuo, "Quality-of-service mapping mechanism for packet video indifferentiated services network," *IEEE Trans. Multimedia*, vol. 3, no. 2, pp. 219–231, Feb. 2001.

[18] W. Kumwilaisak, Y. Kuo, and C. Zhang, "A cross-Layer quality-of-service mapping architecture for video delivery in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1685–1698, Oct. 2003.

[19] M. Mirhakkak, N. Schult, and D. Thomson, "Dynamic bandwidth management and adaptive applications for avariable bandwidth wireless environment," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 1984–1997, Oct. 2001.

[20] Q. Zhang, W. Zhu, and Y. Zhang, "End-to-end QoS for video delivery over wireless Internet," *Proc. IEEE*, vol. 93, no. 1, pp. 123–134, Jan. 2005.

[21] T. Nguyen and A. Zakhor, "Multiple sender distributed video streaming," *IEEE Trans. Multimedia*, vol. 6, no. 2, pp. 315–326, Feb. 2004.

[22] M. Chen and A. Zakhor, "Rate control for streaming video over wireless," in *Proc. IEEE INFOCOM*, 2004, vol. 2, pp. 1374–1387.

[23] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proc. IEEE*, vol. 95, no. 1, pp. 255–312, Jan. 2007.

[24] R. Srikant, *The Mathematics of Internet Congestion Control*. Boston, MA: Birkhauser, 2004.

[25] S. Low, "A duality model of tcp and queue management algorithms," *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 525–536, Apr. 2003.

[26] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control," *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 104–116, Jan. 2005.

[27] J. W. Lee, M. Chiang, and R. A. Calderbank, "Price-based distributed algorithm for optimal rate-reliability tradeoff in network utility maximization," *IEEE J. Sel. Areas Commu.*, vol. 24, no. 5, pp. 962–976, May 2006.

[28] Y. Wang, Y. Zhang, and J. Ostermann, *Video Processing and Communications*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[29] T. M. Cover and J. Thomas, *Elements of Information Theory*. Singapore: Wiley-Interscience, 1991.

[30] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge, MA: Cambridge Univ. Press, 2004.

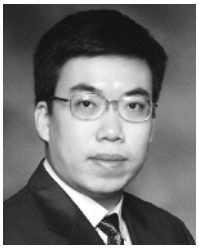
[31] D. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, 2006.

[32] A. Sampath, P. S. Kumar, and J. Holtzman, "Power control and resource management for a multimedia CDMA wireless system," in *Proc. IEEE PIMRC'95*, 1995, vol. 1, pp. 91–95.

[33] K. Kumaran and L. Qian, "Uplink scheduling in CDMA packet-data systems," *Wireless Netw.*, vol. 12, no. 1, pp. 33–43, 2006.

[34] E. Uysal-Biyikoglu, B. Prabhakar, and A. E. Gamal, "Energy-efficient packet transmission over a wireless link," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 487–499, Apr. 2002.

[35] M. Chiang, "Network distribution capacity and content-pipe gap," presented at the Conf. Inform. Sci. Syst., Princeton, NJ, Mar. 2008.



Jianwei Huang (S'01–M'06) received the B.S. degree in electrical engineering from Southeast University , Nanjing, China, in 2000, and the M.S. and Ph.D. degrees in electrical and computer engineering from Northwestern University, Evanston, IL, in 2003 and 2005, respectively.

He is currently an Assistant Professor in the Department of Information Engineering at the Chinese University of Hong Kong, Hong Kong, SAR. He was a Postdoctoral Research Associate at Princeton University, Princeton, NJ, during 2005-2007. He also had summer internships in Motorola, Arlington Heights, IL, in 2004 and 2005. He conducts research in the area of nonlinear optimization and game theoretical analysis of communication networks, with current focus on network economics, cognitive radio networks, broadband communication networks, and multimedia over wireless.

Dr. Huang has served as an Associate Editor of *Journal of Computer and Electrical Engineering*, a Lead Guest Editor of IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, a Lead Guest Editor of *Journal of Advances in Multimedia*, and a TPC Co-Chair of International Conference on Game Theory for Networks (GameNets'09). He was the recipient of a 2001 Walter P. Murphy Fellowship at Northwestern University and a 1999 Chinese National Excellent Student Award.

Zhu Li (M'01–SM'07) received the Ph.D. degree in electrical and computer engineering from Northwestern University, Evanston, IL, in 2004. He has been with the Multimedia Research Laboratory (MRL), Motorola Laboratories, Schaumburg, IL, since 2000, where he is a Principal Staff Research Engineer.

He is currently an Assistant Professor with the Department of Computing, Hong Kong Polytechnic University. He was a Principal Staff Research Engineer with the Multimedia Research Lab (MRL), Motorola Labs USA, during 2000-2008. His research interests include image/video analysis, manifold modeling and machine learning in biometrics and video search and mining, video coding and communication, optimization decomposition and distributed computing techniques in multimedia networks and systems. He has 3 issued U.S. patents, 30+ publications in book chapters, journals and conference proceedings in these areas.

Dr. Li received the Best Poster Paper Award at IEEE International Conference on Multimedia and Expo (ICME), 2006, and the DoCoMo Labs Innovative Paper Award (Best Paper) at IEEE International Conference on Image Processing (ICIP), 2007.

Mung Chiang (S'00–M'03) received the B.S. (hons.) degree in electrical engineering and mathematics, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1999, 2000, and 2003, respectively.

He is an Assistant Professor of Electrical Engineering and an affiliated faculty of Applied and Computational Mathematics and of Computer Science at Princeton University, Princeton, NJ. He conducts research in the areas of optimization of communication systems, theoretical foundation of network architectures, algorithms for wireless broadband access networks, and stochastic analysis of communications and networking.

Dr. Chiang has served as an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and for the *Journal on Optimization and Engineering*, a lead Guest Editor for IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, a Guest Editor for IEEE/ACM TRANSACTIONS ON NETWORKING and IEEE TRANSACTIONS ON INFORMATION THEORY, a Program Co-Chair of the 38th Conference on Information Sciences and Systems, and a Co-Editor of the new Springer book series on Optimization and Control of Communication Systems. He received CAREER Award from the National Science Foundation, Young Investigator Award from the Office of Naval Research, Howard B. Wentz Junior Faculty Award and Commendation List for Outstanding Teaching from Princeton University, School of Engineering Terman Award from Stanford University, and was a Hertz Foundation Fellow. For his work on broadband access networks and Internet traffic engineering, he was selected for the TR35 Young Technologist Award in 2007, a list of top 35 innovators in the world under the age of 35. His work on Geometric Programming was selected by Mathematical Programming Society as one of the top three papers by young authors in the area of continuous optimization during 2004–2007. His work on Layering As Optimization Decomposition became a Fast Breaking Paper in Computer Science by ISI citation. He also co-authored papers that were IEEE INFOCOM best paper finalist and IEEE GLOBECOM best student paper.

Aggelos K. Katsaggelos (S'80–M'85–SM'92–F'98) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1979, and the M.S. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1981 and 1985, respectively.

In 1985, he joined the Department of Electrical and Computer Engineering at Northwestern University, Evanston, IL, where he is currently a Professor. He was the holder of the Ameritech Chair of Information Technology (1997–2003). He is also the Director of the Motorola Center for Communications and a member of the Academic Affiliate Staff, Department of Medicine, Evanston Hospital. He is the editor of *Digital Image Restoration* (Springer-Verlag, 1991), coauthor of *Rate-Distortion Based Video Compression* (Kluwer, 1997), co-editor of *Recovery Techniques for Image and Video Compression and Transmission* (Kluwer, 1998), co-author of *Super-Resolution for Images and Video* (Claypool, 2007) and *Joint Source-Channel Video Transmission* (Claypool, 2007), and the co-inventor of 12 patents

Dr. Katsaggelos has served the IEEE and other Professional Societies in many capacities. He is currently a member of the Publication Board of the IEEE PROCEEDINGS and has served as Editor-in-Chief of the *IEEE Signal Processing Magazine* (1997–2002) and a member of the Board of Governors of the IEEE Signal Processing Society (1999–2001). He is the recipient of the IEEE Third Millennium Medal (2000), the IEEE Signal Processing Society Meritorious Service Award (2001), an IEEE Signal Processing Society Best Paper Award (2001), an IEEE International Conference on Multimedia and Expo Paper Award (2006), and an IEEE International Conference on Image Processing Paper Award (2007). He is a Distinguished Lecturer of the IEEE Signal Processing Society (2007–2008).