This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2014.2332299, IEEE Journal of Selected Topics in Signal Processing

Optimal Foresighted Multi-User Wireless Video

Yuanzhang Xiao, *Student Member, IEEE*, and Mihaela van der Schaar, *Fellow, IEEE*
Department of Electrical Engineering, UCLA. Email: yxiao@seas.ucla.edu, mihaela@ee.ucla.edu.

***Abstract*—Recent years have seen an explosion in wireless video communication systems. Optimization in such systems is crucial – but most existing methods intended to optimize the performance of multi-user wireless video transmission are inefficient. Some works (e.g. Network Utility Maximization (NUM)) are *myopic*: they choose actions to maximize *instantaneous* video quality while ignoring the future impact of these actions. Such myopic solutions are known to be inferior to *foresighted* solutions that optimize the long-term video quality. Alternatively, foresighted solutions such as rate-distortion optimized packet scheduling focus on single-user wireless video transmission, while ignoring the resource allocation among the users.**

In this paper, we propose a general framework of *foresighted* resource allocation among *multiple* video users sharing a wireless network. Our framework allows each user to flexibly choose individual cross-layer strategies. Our proposed resource allocation is optimal in terms of the total payoff (e.g. video quality) of the users. A key challenge in developing foresighted solutions for multiple video users is that the users' decisions are coupled. To decouple the users' decisions, we adopt a novel dual decomposition approach, which differs from the conventional optimization solutions such as NUM, and determines foresighted policies. Specifically, we propose an informationally-decentralized algorithm in which the network manager updates *state- and user-dependent* resource "prices" (i.e. the dual variables associated with the resource constraints), and the users make individual packet scheduling decisions based on these prices. Because a priori knowledge of the system dynamics is almost never available at run-time, the proposed solution can learn online, concurrently with performing the foresighted optimization. Simulation results show 7 dB and 3 dB improvements in Peak Signal-to-Noise Ratio (PSNR) over myopic solutions and existing foresighted solutions, respectively.

***Keywords*—Multi-user resource allocation, multi-user wireless communication, wireless video, cross-layer optimization, foresighted optimization, network utility maximization, Markov decision processes, reinforcement learning.**

I. Introduction

Video applications, such as video streaming, video conferencing, remote teaching, surveillance etc., have become the major applications deployed over the current cellular networks and wireless Local Area Networks (LANs). The bandwidth-intensive and delay-sensitive video applications require efficient allocation of network resources (e.g. bandwidth in 4G LTE networks or temporal transmission opportunity in wireless LANs) among the users accessing the network, and efficient scheduling of each user's video packets based on its allocated resources.

Most existing solutions for multi-user wireless video transmission are *myopic* [4]–[9], meaning that these joint resource allocation and packet scheduling solutions are designed to maximize the *instantaneous*[1] video quality (i.e. the average distortion impact of the packets sent within the next transmission opportunity or time interval). These solutions can be interpreted as maximizing the instantaneous video quality by repeatedly and independently solving Network Utility Maximization (NUM) problems over time. However, current (resource allocation and packet scheduling) decisions impact the future system performance, which is not taken into consideration by the repeated NUM solutions. Optimal solutions for multi-user resource allocation and packet scheduling need to be formalized as sequential decisions given (unknown) dynamics (i.e. time-varying channel conditions and video traffic characteristics, dependencies across video packets etc.). Hence, the repeated NUM solutions are myopic and inferior to *foresighted* solutions that maximize the *long-term* video quality.

To address this limitation, several foresighted solutions have been proposed for packet scheduling; see e.g. [10]–[16]. However, these packet scheduling solutions focus on the sequential decision making of a single foresighted video user and do not consider the coupling among users. In multi-user wireless networks, the solutions developed for a single user have been shown to be highly inefficient [17][2], since they ignore the fact that users are sharing the same wireless resource, and hence, their decisions over time are coupled. A simple solution to perform multi-user resource allocation was proposed in [4]. However, this allocation is performed statically, possibly before run-time, and as shown in [17], such static resource allocation is suboptimal compared to solutions that dynamically allocate resources among multiple users given the users' video traffic and channel characteristics, which are time-varying and often unknown before run-time.

In this paper, we propose a general methodology for performing *foresighted* resource allocation among *multiple* video users sharing the common resources of a wireless network. Our framework allows each user to flexibly choose individual cross-layer strategies. Our proposed resource allocation is optimal, i.e. it maximizes the total payoff (e.g. video quality) of the users. We consider multiple users transmitting delay-sensitive video streams over a 4G cellular uplink network[3]. In such networks, the base station (BS) needs to decide how to allocate wireless resources (i.e. bandwidth) among

[1]We will define instantaneous and long-term video quality rigorously in Section III.

[2]The work [17] is the only one that develops foresighted solutions for multiple video users. We will discuss the differences between our work and [17] in detail in Section II.

[3]Although we focus on uplink video transmission, our work can be readily extended to downlink video transmission, to IEEE 802.11a/e wireless LANs (Local Area Networks) in which transmission opportunities (i.e. time slots) are allocated among users [4], and to IEEE 802.15.4e-enabled Internet-of-Things and machine-to-machine communications.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2014.2332299, IEEE Journal of Selected Topics in Signal Processing

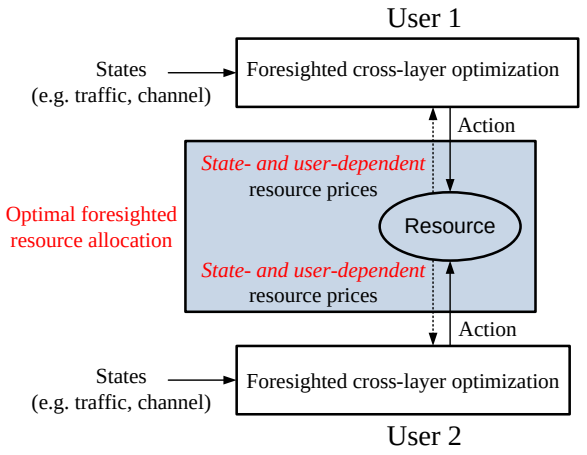

Fig. 1. Illustration of our proposed framework. Each user can flexibly choose any (foresighted) cross-layer optimization that can be formulated as a Markov decision process. Under the chosen cross-layer optimization framework, our proposed foresighted resource allocation based on state- and user-dependent prices is optimal in terms of the total long-term payoff of the users.

the multiple video users. In our proposed solution, the BS does not directly allocate the resources; instead, it mediates the resource allocation by charging each user a *state- and user-dependent* resource "price"[4]. Given the price, each user determines its own optimal packet scheduling. Hence, our approach is decentralized and enables users to make optimal decisions locally, in an informationally-decentralized manner, based on their own private information and the resource price. We propose a low-complexity algorithm in which the BS updates the resource prices using a stochastic subgradient method based on the users' resource usage while the users make foresighted decisions based on these prices. We prove that the algorithm can converge to the optimal prices, under which the users' optimal decisions maximize the long-term total payoff of the users. Moreover, our solution also allows to impose a minimum video quality guarantee for each user (i.e. the video quality for each user needs to be higher than a preset minimum video quality guarantee). Importantly, the BS and the users may not have in practice the statistic knowledge of the system dynamics (e.g. the incoming video traffic characteristics and the time-varying channel conditions). For such scenarios, we propose a post-decision state (PDS) based learning algorithm that converges to the proposed optimal solution.

Finally, at the risk of repetition, we want to emphasize that our framework is general: the users can adopt any cross-layer strategy that can be formulated as a general Markov decision process (MDP) as in Section III-A. For instance, most state-of-the-art single-user video packet scheduling schemes [10]–[16] can be formulated in this way. Given such packet scheduling schemes, our proposed foresighted resource allocation is optimal in terms of the total long-term payoff of the users.[5] We illustrate our proposed framework in Fig. 1.

[4]Note that the "price" can be a control signal instead of the price for monetary payment.

[5]Of course, different packet scheduling schemes may result in different performances at the optimum.

TABLE I. COMPARISONS WITH RELATED WORKS.

	Traffic model	Users	Foresighted	Learning	Optimal
[1]–[3]	Flow-level	Multiple	No	No	No
[4]–[9]	Packet-level	Multiple	No	No	No
[10]–[16]	Packet-level	Single	Yes	Yes	No
[17]	Packet-level	Multiple	Yes	Yes	No
Proposed	Packet-level	Multiple	Yes	Yes	Yes

The rest of the paper is organized as follows. We discuss prior work in Section II. We describe the system model in Section III and formulate the design problem in Section IV. Then we propose our solution in Section V. Simulation results in Section VI demonstrate the performance improvement of the proposed solution. Finally, Section VII concludes the paper.

II. RELATED WORKS

A. Related Works on Video Transmission

The existing works on wireless video communications can be classified based on various criteria. First, some works [1]–[3] model the video traffic with a simplified model, e.g. flow-level models using priority queues. Such a model cannot accurately capture the heterogeneous distortion impact, delay deadlines, and dependency of the video packets. Hence, the solution derived based on simplified models may not perform well in practice [16][17].

A plethora of recent works [4]–[16] adopt packet-level models for video traffic, but propose distinct solutions to optimize the video quality. Some works [4]–[9] assume that the users are *myopic*, namely they only maximize their *instantaneous* video quality over a given time interval without considering the impact of their actions on the *long-term* video quality. They cast the problem in a NUM framework to maximize the instantaneous joint video quality of all the users, and apply the NUM framework repeatedly when the channel conditions or video traffic characteristics change. However, since the users are optimizing their transmission decisions myopically, their long-term average performance is inferior to the performance achieved when the users are *foresighted* [16]. Most of the works [10]–[16] considering the foresighted decision of users focus solely on a *single* foresighted user making sequential transmission decisions (e.g. packet scheduling, retransmissions etc.). However, these single user solutions do not discuss how to allocate resources among multiple users as well as how this allocation is impacted by and impacts the foresighted scheduling decisions of individual users. Static allocations of resources, which are often assumed in the works studying the foresighted decisions of a single user, have been shown to be suboptimal compared to the solutions that dynamically allocate resources among multiple users [17].

The only work which proposes a solution for video resource allocation among *multiple foresighted* users is [17], which formulates the problem as a multi-user MDP (MU-MDP). Since this work [17] is most related to our proposed solution, we discuss the differences between them in detail. The challenge in foresighted multi-user video transmission is that the users' decisions are dynamic and coupled through

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2014.2332299, IEEE Journal of Selected Topics in Signal Processing

TABLE II. COMPARISONS WITH RELATED THEORETICAL FRAMEWORKS.

	Decision makers	Foresighted	Learning	Optimal
MDP [16]	Single	Yes	Yes	No
Repeated NUM [9]	Multiple	No	No	No
MU-MDP [17]	Multiple	Yes	Yes	No
Lyapunov Optimization [18]	Single	Yes	Yes	No
Proposed	Multiple	Yes	Yes	Yes

the resource (e.g. bandwidth or time) constraints. Hence, the design problem is much more complicated than in myopic multi-user video resource allocation and transmission (e.g. NUM) or in foresighted single-user video transmission given a static or pre-determined resource allocation. To overcome this challenge, a dual decomposition method has been proposed in [17], which removes the resource constraints in all the states and adds them to the objective functions in the corresponding states as penalties modulated by the *same* Lagrangian multiplier (interpreted as the price of resources).[6] However, such a solution is in general suboptimal (i.e. there is a positive duality gap) for the following reason. Since it uses the same Lagrangian multiplier (i.e. a uniform price) for the resource constraints in all the states [17], the uniform price is usually set high such that the resource constraints in all the states are satisfied. Technically, when there are some active constraints (i.e. constraints satisfied with strict inequality), the complementary slackness conditions cannot be satisfied under all the states. Hence, there will always be a duality gap due to the violation of complementary slackness conditions. In contrast, in our work, we use different prices for resource constraints under different states (e.g. under different channel conditions). In this way, we can achieve the optimal performance (i.e. there is no duality gap, unlike the MU-MDP solution with uniform price in [17]).

Table I summarizes the above discussions. Note that the optimality shown in the last column of Table I indicates whether the solution is optimal for the long-term network utility (i.e. the joint long-term video quality of all the users in the network). The existing solutions can be optimal in the corresponding models (e.g. the solutions in [10]–[16] are optimal when there is only one user).

B. Related Theoretical Frameworks

Single-user foresighted decision making in a dynamically changing environment has been studied and formulated as Markov Decision Process (MDP). As mentioned previously, the problem of multiple video users making coupled foresighted decisions (MU-MDPs) has been studied in [17]. However, as we discussed before, this MU-MDP solution is based on uniform prices and is therefore suboptimal for most multi-user wireless video scenarios, even though it has been shown to significantly outperform the myopic solutions.

It is worth noting that foresighted decision making in a dynamically changing environment can also be solved using a Lyapunov optimization framework [18]. However, the Lyapunov optimization framework is not able to take optimal decisions for wireless video streaming since it disregards the specific delay-constraints and distortion-impacts of the video traffic [19].

Table II summarizes the above discussions about existing theoretical frameworks.

In Table VI at the end of Section V, we provide rigorous technical comparisons with existing works after we describe our proposed solution in detail.

III. SYSTEM MODEL

We first present a general abstract model for multi-user wireless video transmission in 4G LTE cellular networks or IEEE 802.11a/e wireless LANs. Our technical results hold for this general model. Then we give an example of a packet-level video transmission model as in [10][16][17] to illustrate our abstract model.

A. The General Model

We consider a network with a network manager (e.g. the base station), indexed by 0, and a set $\mathcal{I}$ of I wireless video users, indexed by $i = 1, \ldots, I$. Time is slotted at $t = 0, 1, 2, \ldots$. In the rest of the paper, we will put the user index in the superscript and the time index in the subscript of variables. The multi-user wireless video transmission system is described by the following features:

1) States: Each user i has a finite state space S^i, from which a state s^i is realized and revealed to user i at the beginning of each time slot. The state s^i can consist of several components, such as the video traffic state and the channel state. An example of a simplified video traffic state can be the types of video frames (I, P, or B frame) available for transmission and the numbers of packets in each available video frame. Note that the video traffic in our model can come from video sequences that are either encoded in real-time, or offline and stored in the memory before the transmission. An example of a channel state can be the channel quality reported to the application layer by the lower layers. The network manager has a finite number of states $s^0 \in S^0$ that describe the status of the resource in the network.

2) Actions: At each state s^i, each user i chooses an action $a^i \in A^i(s^i)$. For example, an action can be how many packets within each available video frame should be transmitted. We allow the sets of actions taken under different states to be different, in order to incorporate the minimum video quality requirements that will be discussed.

3) Payoffs: Each user i has a payoff function $u^i : S^i \times A^i \to \mathbb{R}$. The payoff function u^i is concave in the action under any state. A typical payoff can be the distortion impact of the transmitted packets plus the (negative) disutility incurred by the energy consumption in transmitting the packets.

[6]There are two key differences between the dual decomposition method in [17] and the classical dual decomposition method used in myopic optimization such as in NUM. The first difference is that the users' objective functions are the long-term video quality, and hence the decomposed subproblems solved by each user are foresighted optimization problems instead of static optimization problems as in NUM. The second one is that the update of the dual variables is different from that in NUM since the subproblems are foresighted.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2014.2332299, IEEE Journal of Selected Topics in Signal Processing

4) State Transition: Each user i's state transition is Markovian, and can be denoted by $p^i(s^{i\prime}|s^i, a^i) \in \Delta(S^i)$, where $\Delta(S^i)$ is the probability distribution over the set of states.

5) Resource Constraints: Given the status of the resource (i.e. the network manager's state s^0), we can write the (linear) resource constraint as

$$f(s^0, a^1, \ldots, a^I) \triangleq f^0(s^0) + \sum_{i=1}^{I} f^i(s^0) \cdot a^i \leq 0,$$

where $f^i(s^0)$ is the coefficient under state s^0. When a^i is a vector, $f^i(s^0)$ is a vector of the appropriate length, and $f^i(s^0) \cdot a^i$ is the inner product of the two vectors.

A variety of multi-user wireless video transmission systems can be modeled as special cases of our general model. Next, we present a packet-level video transmission model as an illustration.

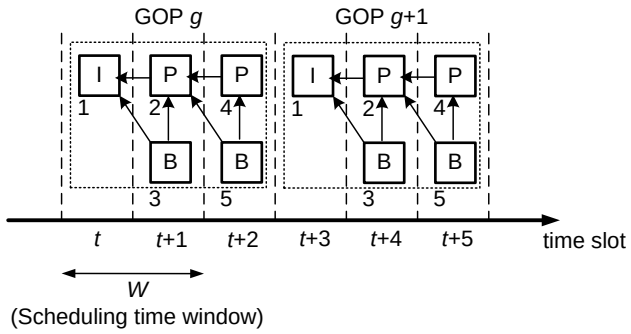

Fig. 2. Illustration of GOP (group of pictures), DU (data unit), and the context. Since the scheduling time window is $W = 2$, the contexts in different time slots are $C_t = \{DU_1^g, DU_2^g, DU_3^g\}$, $C_{t+1} = \{DU_2^g, DU_3^g, DU_4^g, DU_5^g\}$, $C_{t+2} = \{DU_4^g, DU_5^g, DU_1^{g+1}\}$, $C_{t+3} = \{DU_1^{g+1}, DU_2^{g+1}, DU_3^{g+1}\}$, and so on.

B. An Example Packet-Level Video Transmission Model

Packet-level video transmission models have been proposed in a variety of related works, including [10]–[17]. In the following, we briefly describe the model as an illustration of our general model, and refer interested readers to [10]–[17] for more details.

We first consider a specific video user i, and hence drop the superscript before we describe the resource constraints. The video source data is encoded using an H.264 or MPEG video coder under a Group of Pictures (GOP) structure: the data is encoded into a series of GOPs, indexed by $g = 1, 2, \ldots$, where one GOP consists of N data units (DUs). Each DU $n \in \{1, \ldots, N\}$ in GOP g, denoted DU_n^g, is characterized by its size $l_n^g \in \mathbb{N}_+$ (i.e. the number of packets in it), distortion impact $q_n^g \in \mathbb{R}_+$, delay deadline $d_n^g \in \mathbb{N}_+$, and dependency on the other DUs in the same GOP. The dependency among the DUs in one GOP comes from encoding techniques such as motion estimation/compensation. In general, if DU_n^g depends on DU_m^g, we have $d_n^g \geq d_m^g$ and $q_n^g \leq q_m^g$, namely DU_m^g should be decoded before DU_n^g and has a higher distortion impact than DU_n^g [16]. Note that in the case of scalable video coding, there is no dependency among the DUs, and the following representation of the model can be greatly simplified. We will keep the dependency for generality in our exposition.

Among the above characteristics, the distortion impact q_n^g, delay deadline d_n^g, and the dependency are deterministic and fixed for the same DUs across different GOPs (e.g. $q_n^g = q_n^{g+1}$) [16][17]. The size of DU_n^g is random. As in [17], we assume that the sizes of all the DUs are independent random variables, and that the sizes of the nth DUs in different GOPs have the same distribution.

1) States: Each user's state consists of the traffic state T_t and the channel state h_t. We describe the traffic state T_t first. At time slot t, as in [10][16][17], we assume that the wireless user will only consider for transmission the DUs with delay deadlines in the range of $[t, t+W)$, where W is referred to as the scheduling time window (STW). Following the model in [16][17], at time slot t, we introduce *context* to represent the set of DUs that are considered for transmission, i.e., whose delay deadlines are within the range of $[t, t+W)$. We denote the context by $C_t = \left\{DU_j^g | d_j^g \in [t, t+W)\right\}$. Since the GOP structure is fixed, the transition from context C_t to C_{t+1} is deterministic. An illustration of the context is given in Fig. 2.

Given the current context C_t, we let $x_{t,DU}$ denote the number of packets in the buffer associated with a DU in C_t. We denote the buffer state of the DUs in C_t by $x_t = \{x_{t,DU} | DU \in C_t\}$. The traffic state T_t at time slot t is then $T_t = (C_t, x_t)$, where the context C_t represents which DUs are available for transmission, and the buffer state x_t represents how many packets each available DU has left in the buffer .

Next we describe the channel state h_t. At each time slot t, the wireless user experiences a channel condition $h_t \in \mathcal{H}$, where $\mathcal{H}$ is the finite set of possible channel conditions. We assume that the wireless channel is slow-fading (i.e. remains the same in one time slot), and that the channel condition h_t can be modeled as a finite-state Markov chain [21].

In summary, the state of a user at each time slot t is $s_t = (C_t, x_t, h_t)$, which includes the current context, buffer state and channel state.

2) (Packet Scheduling) Actions: At each time slot t, the user decides how many packets should be transmitted from each DU in the current context. The decision is represented by $a_t(s_t) = \{y_{t,DU} | DU \in C_t, y_{t,DU} \in [0, x_{t,DU}]\}$, where $y_{t,DU}$ is the amount of packets transmitted from the DU.

3) Payoffs: As in [16], we consider the following *instantaneous payoff* at each time slot t:[7]

$$u(s_t, a_t) = \sum_{DU \in C_t} q_{DU} y_{t,DU} - \beta \cdot \rho\left(h_t, \|a_t\|_1\right), \qquad (1)$$

where the first term $\sum_{DU \in C_t} q_{DU} y_{t,DU}$ is the *instantaneous video quality*, namely the distortion reduction obtained by transmitting the packets from the DUs in the current context, and the second term $\beta \cdot \rho\left(h_t, \|a_t\|_1\right)$ represents the disutility

[7]The payoff function can be easily extended within our framework to include additional features in the model. For example, when there are packet loss, we can modify the first term to be the expected distortion reduction given the packet loss rate, or modify the second term to consider the additional energy consumption associated with packet retransmission.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2014.2332299, IEEE Journal of Selected Topics in Signal Processing

of the energy consumption by transmitting the packets. Since the packet scheduling action a_t is a vector with nonnegative components, we have $\|a_t\|_1 = \sum_{DU \in C_t} y_{t,DU}$, namely $\|a_t\|_1$ is the total number of transmitted packets. As in [16], the energy consumption function $\rho(h, \|a\|_1)$ is assumed to be convex in the total number of transmitted packets $\|a\|_1$ given the channel condition h. An example of such a function can be $\rho(h, \|a\|_1) = \sigma^2(e^{2\|a\|_1 b} - 1)/h$, where b is the number of bits in one packet [20]. The payoff function is a tradeoff between the distortion reduction and the energy consumption, where the relative importance of energy consumption compared to distortion reduction is characterized by the tradeoff parameter $\beta > 0$. In the simulation, we will set different values for β to illustrate the tradeoff between the distortion reduction and energy consumption.

4) The Resource Constraint: In this paper, we consider a 4G cellular network, where there are I wireless video users transmitting to the BS indexed by 0. The users access the channels in a FDMA (frequency-division multiple access) manner. The total bandwidth B is divided and shared by the users.

We assume that each user i uses adaptive modulation and coding (AMC) based on its channel condition. In other words, each user i chooses a data rate r_t^i under the channel state h_t^i. Note that the rate selection is done by the physical layer and is not a decision variable in our framework. Then as in [2][3], we have the following resource constraint:

$$\sum_{i=1}^{I} \frac{\|a_t^i\|_1 b}{r_t^i(h_t^i)} \leq B, \tag{2}$$

where $\frac{\|a_t^i\|_1 \cdot b}{r_t^i(h_t^i)}$ is the bandwidth needed for transmitting the amount $\|a_t^i\|_1 \cdot b$ of bits given the data rate $r_t^i(h_t^i)$.

In this model, the network manager's state s^0 is then the collection of channel states, namely $s^0 = (h^1, \ldots, h^I)$. The information about the channel states is fed back from the users to the BS. We can write the constraint compactly as the linear constraint $f(s_t^0, a_t^1, \ldots, a_t^N) \leq 0$ with $f^i(s_t^0) = \frac{b}{r_t^i(h_t^i)}, i = 1, \ldots, I$ and $f^0(s_t^0) = -B$.

IV. The Design Problem

Each user makes decisions based on its state s_t. Hence, each user i's strategy can be defined as a mapping $\pi^i : S^i \to \cup_{s^i} A^i(s^i)$, where $A^i(s^i)$ is the set of actions available under state s^i. We allow the set of available actions to depend on the state, in order to capture the minimum video quality guarantee. For example, we may have a minimum distortion impact reduction requirement D^i for user i at any time, which gives us

$$A^i(s_t^i) = \left\{ a_t^i : \sum_{DU \in C_t^i} q_{DU} y_{t,DU}^i \geq D^i \right\}.$$

The users aim to maximize their expected long-term payoff. Given its initial state s_0^i, each user i's strategy π^i induce a probability distribution over the sequences of states $s_1^i, s_2^i, \ldots$, and hence a probability distribution over the sequences of instantaneous payoffs $u_0^i, u_1^i, \ldots$. Taking expectation with respect to the sequences of payoffs, we have user i's *long-term payoff* given the initial state as

$$U^i(\pi^i | s_0^i) = \mathbb{E}\left\{ (1-\delta) \sum_{t=0}^{\infty} \left(\delta^t \cdot u_t^i\right) \right\}, \tag{3}$$

where $\delta \in [0, 1)$ is the discount factor.

The design problem can be formulated as

$$\begin{aligned} \max_{\pi^1, \ldots, \pi^I} \quad & \sum_{s_0^1, \ldots, s_0^I} \sum_{i=1}^{I} U^i(\pi^i | s_0^i) \\ s.t. \quad & \text{minimum video quality guarantee :} \\ & \pi^i(s^i) \in A^i(s^i), \ \forall i, s^i, \\ & \text{resource constraint :} \\ & f(s^0, \pi^1(s^1), \ldots, \pi^I(s^I)) \leq 0, \ \forall s^0. \end{aligned} \tag{4}$$

Note that the design problem (4) is a weakly-coupled MU-MDP as defined by [22]. It is a MU-MDP because there are multiple users making foresighted decisions. The MU-MDP is coupled, because the users influence each other through the resource constraints (namely the choice of one user's action depends on the choices of the other users). However, it is weakly-coupled, because the coupling is through the resource constraints only, and because one user's *instantaneous* payoff $u^i(s^i, a^i)$ is not affected by the other users' actions a^j. It is this weak coupling that enables us to decompose the multi-user problem into multiple single-user problems through prices. Such a decomposition of weakly-coupled MU-MDPs has been studied in a general setting [22] and in wireless video transmission [17], both adopting a dual decomposition approach based on *uniform* price (i.e. the same Lagrangian multiplier for the resource constraints under all the states).

Note also that we sum up the network utility $\sum_{i=1}^{I} U^i(\pi^i | s_0^i)$ under all the initial states $(s_0^1, \ldots, s_0^I)$. This can be interpreted as the expected network utility when the initial state is uniformly distributed. The optimal stationary strategy profile that maximizes this expected network utility will also maximize the network utility given any initial state.

The design problem (4) is very challenging, and has never been solved optimally. To better understand this, let us assume that a central controller would exist which knows the complete information of the system (i.e. the states, the state transitions, the payoff functions) at each time step. Then, this central controller can solve the above problem (4) as a centralized single-user MDP (e.g. using well-known Value Iteration or Policy Iteration methods) and obtain the solution to the design problem $\boldsymbol{\pi}^\star$ and the optimal value function $U^\star$. However, the multi-user wireless video system we discussed is inherently informationally-decentralized and there is no entity in the network that possesses the complete information. Moreover, the computational complexity of solving (4) by a single entity is prohibitively high. Hence, our goal is to develop an optimal decentralized algorithm that converges to the optimal solution.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2014.2332299, IEEE Journal of Selected Topics in Signal Processing

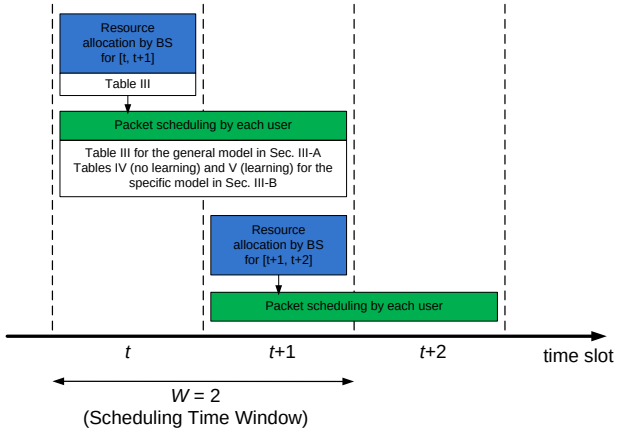

Fig. 3. Illustration of the resource allocation and packet scheduling in the proposed solution.

V. Optimal Foresighted Video Transmission

In this section, we show how to determine the optimal foresighted video transmission policies. We propose an algorithm that allows each entity to make decisions based on its local information and the limited information exchange between the BS and the users. Specifically, in each time slot, the BS sends resource prices to each user, and the users send their total numbers of packets to transmit to the BS. The BS keeps updating the resource prices based on the resource usage by the users, and obtains the optimal resource prices based on which the users' optimal individual decisions achieve the optimal network utility. An overview of the main results and the structure of the proposed solution is given in Fig. 3.

A. Decoupling of The Users' Decision Problems

Each user aims to maximize its own long-term payoff $U^i(\pi^i|s_0^i)$ subject to the constraints. Specifically, given the other users' strategies $\boldsymbol{\pi}^{-i} = (\pi^1, \ldots, \pi^{i-1}, \pi^{i+1}, \ldots, \pi^I)$ and states $\boldsymbol{s}^{-i} = (s^1, \ldots, s^{i-1}, s^{i+1}, \ldots, s^I)$, each user i solve the following long-term payoff maximization problem:

$$\begin{aligned} \pi^i = \quad & \arg\max_{\tilde{\pi}^i} \quad U_i(\tilde{\pi}^i|s_0^i) \\ & s.t. \qquad \tilde{\pi}^i(s^i) \in A^i(s^i),\ \forall s^i, \\ & \qquad\quad f(s^0, \tilde{\pi}^i(s^i), \boldsymbol{\pi}^{-i}(\boldsymbol{s}^{-i})) \le 0. \end{aligned} \tag{5}$$

Assuming that the user knows all the information (i.e. the other users' strategies $\boldsymbol{\pi}^{-i}$ and states $\boldsymbol{s}^{-i}$), user i's optimal value function should satisfy the following:

$$\begin{aligned} V(s^i) = \max_{a^i \in A^i(s^i)} \quad & (1-\delta)u^i(s^i, a^i) + \delta \sum_{s^{i\prime}} p^i(s^{i\prime}|s^i, a^i) V(s^{i\prime}) \\ s.t. \qquad & f(s^0, a^i, \boldsymbol{\pi}^{-i}(\boldsymbol{s}^{-i})) \le 0. \end{aligned} \tag{6}$$

Note that the above equations would be the Bellman equation, if user i knew the other users' strategies $\boldsymbol{\pi}^{-i}$ and states $\boldsymbol{s}^{-i}$ and the BS's state s^0 (i.e. the channel states of all the users). However, such information is never known to a particular user. Without such information, one user cannot solve the decision problem above because the resource constraint contains unknown variables. Hence, we need to separate the influence of the other users' decisions from each user's decision problem.

One way to decouple the interaction among the users is to remove the resource constraint and add it as a penalty to the objective function. Denote the Lagrangian multiplier (i.e. the "price") associated with the constraint under state s^0 as $\lambda^0(s^0)$. Then the penalty at state s^0 is

$$-\lambda^0(s^0) \cdot f(s^0, a^1, \ldots, a^I) = -\lambda^0(s^0) \cdot \sum_{i=1}^{I} f^i(s^0) \cdot a^i.$$

Since the term $-\lambda^0(s^0) \cdot \sum_{j \neq i} f^j(s^0) \cdot a^j$ is a constant for user i, we only need to add $-\lambda^0(s^0) \cdot f^i(s^0) \cdot a^i$ to each user i's objective function. We define $\lambda^i(s^0) \triangleq \lambda^0(s^0) \cdot f^i(s^0)$. Then we can rewrite user i's decision problem as

$$\begin{aligned} \tilde{V}^{\lambda^i(s^0)}(s^i) = \max_{a^i \in A^i(s^i)} & (1-\delta)\left[u^i(s^i, a^i) - \lambda^i(s^0) \cdot a^i\right] \\ & + \delta \cdot \sum_{s_i'} \left[p^i(s^{i\prime}|s^i, a^i) \tilde{V}^{\lambda^i(s^0)}(s^{i\prime})\right]. \end{aligned} \tag{7}$$

By contrasting (7) with (6), we can see that given the price λ^i, each user can make decisions based only on its local information since the resource constraint is eliminated. Note, importantly, that the above decision problem (7) for each user i is different from that in [17] with uniform price. This can be seen from the term $\lambda^i(s^0) \cdot a^i$ in (7), where the price $\lambda^i(s^0)$ is user-specific and depends on the state, while the uniform price in [17] is a constant λ. The decision problem (7) is also different from the subproblem resulting from dual decomposition in NUM, because it is a foresighted optimization problem that aims to maximize the long-term payoff. This requires a different method to calculate the optimal Lagrangian multiplier $\lambda^i(s^0)$ than that in NUM.

B. Optimal Decentralized Video Transmission Strategy

For the general model described in Section III-A, we propose an algorithm used by the BS to iteratively update the prices and by the users to update their optimal strategies. The algorithm will converge to the optimal prices and the optimal strategy profile that achieves the minimum total system payoff $U^\star$. The algorithm is described in Table III.

Theorem 1: The algorithm in Table III converges to the optimal strategy profile, namely

$$\lim_{t,k\to\infty} \left| \sum_{s_t^1, \ldots, s_t^I} \sum_{i=1}^{I} U_i(\pi^{i,\lambda_k^i}|s_t^i) - U^\star \right| = 0 \ .$$

Proof: See the appendix. ■

We illustrate the the BS's and users' updates and their information exchange in one time slot in Fig. 4. At the beginning of each time slot t, the BS and the users exchange information to compute the optimal resource price and the optimal actions to take. Specifically, in each iteration k, the BS updates the resource price λ_k^0. Then based on the user-specific resource price λ_k^i, each user i solves for the optimal individual strategy π^{i,λ_k^i}, and sends the BS its resource request $f^i \cdot \pi^{i,\lambda_k^i}(s_t^i)$. Then the BS updates the prices based on the users' resource

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2014.2332299, IEEE Journal of Selected Topics in Signal Processing

TABLE III. DISTRIBUTED ALGORITHM TO COMPUTE THE OPTIMAL STRATEGY AT TIME t.

Input: Performance loss tolerance ϵ

Initialization: Set $k = 0$, $\lambda_k^0 = 0$.
Each user i observes s_t^i, the BS observes s_t^0
repeat
Each user i solves the decoupled decision problem (7) to obtain π^{i,λ_k^i}
Each user i submits its resource request $f^i(s_t^0) \cdot \pi^{i,\lambda_k^i}(s_t^i)$
The BS updates the prices (stochastic subgradient update):
$\lambda_{k+1}^i(s_t^0) = \lambda_k^i(s_t^0) + \frac{1}{k+1} f(s_t^0, \pi^{1,\lambda_k^1}(s_t^1), \cdots, \pi^{I,\lambda_k^I}(s_t^I))$
until $\|\lambda_{k+1}^i(s_t^0) - \lambda_k^i(s_t^0)\| \leq \epsilon$

Output: optimal price λ_k^0, optimal strategies $\{\pi^{i,\lambda_k^i}\}_{i=1}^I$

requests using the stochastic subgradient method, which can be performed easily. The difference from the dual decomposition in NUM is that each user's decision problem in our work is a foresighted optimization problem aiming to maximize the long-term, instead of instantaneous, payoff. Our algorithm is also different from the algorithm in [17] in that we have different prices under different states.

From Fig. 4, we can clearly see what information (namely resource prices λ_k^0 and resource requests $f^i \cdot \pi^{i,\lambda_k^i}(s_t^i)$) is exchanged. The amount of information exchange is small ($O(I)$), compared to the amount of information required by each user to solve the decision problem (6) directly ($\prod_{j\neq i}|S_i|$ states plus the strategies $\boldsymbol{\pi}_{-i}$). In other words, the algorithm enables the entities to exchange a small amount ($O(I)$) of information and reach the optimal video transmission strategy that achieves the same performance as when each entity knows the complete information (i.e. the states and the strategies of all the entities) about the system.

C. Optimal Packet Scheduling

In the previous subsection, we propose an algorithm of optimal foresighted resource allocation and packet scheduling for the general video transmission model described in Section III-A. In the algorithm, each user's packet scheduling decision is obtained by solving the Bellman equation (7) (see Table III). The Bellman equation (7) can be solved by a variety of standard techniques such as value iteration. However, the computational complexity of directly applying value iteration may be high, because each user's state contains the information of all DUs and thus each user's state space can be very large. In the following, we show that for the specific model described in Section III-B, we can greatly simplify the packet scheduling decision problem. The key simplification comes from the decomposition of each user's packet scheduling problem with multiple DUs into multiple packet scheduling problems with single DU. In this way, we can greatly reduce the number of states in each single-DU packet scheduling problem, such that the total complexity of packet scheduling grows linearly, instead of exponentially without decomposition, with the number of DUs.

The decomposition closely follows the decomposition of multiple-DU packet scheduling problems proposed in [16]. The only difference is that the decision problem (7) in our work has an additional term $-\lambda^i(s^0)\cdot a^i$ due to the price, while such

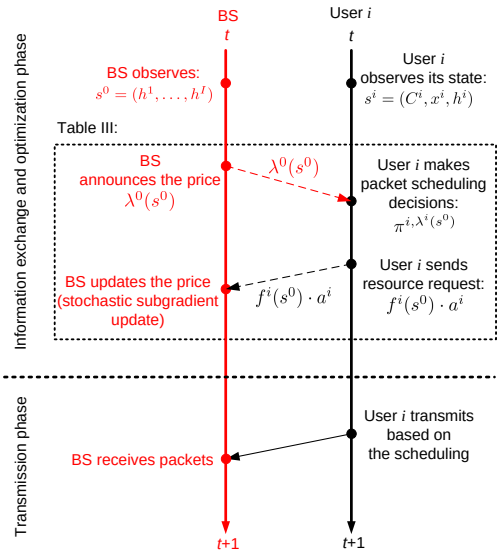

Fig. 4. Illustration of the interaction between the BS and user i (i.e. their decision making and information exchange) in one period.

a term does not exist in [16] because the *single*-user packet scheduling problem is considered in [16].

Lemma 1 (Structural Result): Suppose $DU_1 \in C_t$ and $DU_2 \in C_t$. If DU_2 depends on DU_1, we should schedule the packets of DU_1 before scheduling the packets of DU_2.

Proof: The proof is straightforward and similar to the proof of [16, Lemma 1]. If DU_2 depends on DU_1, then DU_1 has higher distortion impact and earlier deadline, which means that it always achieves a higher distortion reduction to schedule packets of DU_1. In addition, the contributions in energy consumption and resource payment do not depend on which DU the packets come from. Hence, we should always schedule the packets of DU_1 before scheduling the packets of DU_2. ■

Although Lemma 1 is straightforward, it greatly simplifies the scheduling problem because we can now take advantage of the partial ordering of the DUs. However, this still does not solve the scheduling decision for the DUs that are not dependent on each other. Next, we provides the algorithm of optimal packet scheduling in Table IV. The algorithm decomposes the multiple-DU packet scheduling problem into a sequence of single-DU packet scheduling problems, and determines how many packets to transmit for each DU sequentially. This greatly reduces the total computational complexity (which is linear in the number of DUs) compared to solving the multiple-DU packet scheduling problem directly (in which the number of states grows exponentially withe number of DUs). The algorithm is similar to [16, Algorithm 2]. The only difference is the term $-\lambda^i(s^0)\cdot a^i$.

D. Learning Unknown Dynamics

In practice, each entity may not know the dynamics of its own states (i.e., its own state transition probabilities) or even the set of its own states. When the state dynamics are

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2014.2332299, IEEE Journal of Selected Topics in Signal Processing

TABLE IV. OPTIMAL PACKET SCHEDULING ALGORITHM FOR EACH USER.

Input: Directed acyclic graph given the current context: $DAG(C_t)$

Initialization: Set $DAG_1 = DAG(C_t)$.

For $k = 1, \ldots, |C_t|$

$$DU_k = \arg \max_{DU \in root(DAG_k)} \max_{0 \le y \le x_{t,DU}} (1-\delta)\left[q_{DU} \cdot y - \lambda_k^i(s^0) \cdot y\right] + \delta \cdot \textstyle\sum_{s^{i\prime}} \left[p^i(s^{i\prime}|s^i, a_{f,t}) \tilde{V}^{i,\lambda^{i,(k)}(s^0)}(s^{i\prime})\right]$$

$$y^*_{t,DU_k} = \arg \max_{0 \le y \le x_{t,DU_k}} (1-\delta)\left[q_{DU_k} \cdot y - \lambda_k^i(s^0) \cdot y\right] + \delta \cdot \textstyle\sum_{s^{i\prime}} \left[p^i(s^{i\prime}|s^i, a_{f,t}) \tilde{V}^{i,\lambda^{i,(k)}(s^0)}(s^{i\prime})\right]$$

$$DAG_{k+1} = DAG_k \setminus \{DU_k\}$$

End For

TABLE V. OPTIMAL DECOMPOSED PACKET SCHEDULING ALGORITHM WITH PDS-BASED LEARNING.

Input: Directed acyclic graph given the current context: $DAG(C_t)$

Initialization: Set $DAG^1 = DAG(C_t)$.

For $k = 1, \ldots, |C_t|$

$$DU_k = \arg \max_{DU \in root(DAG_k)} \max_{0 \le y \le x_{t,DU}} (1-\delta)\left[q_{DU} \cdot y - \lambda_k^i(s^0) \cdot y\right] + \delta \cdot U_{DU}(C_t, x_{t,DU} - y, h_t)$$

$$y^*_{t,DU_k} = \arg\max_{0 \le y \le x_{t,DU}} (1-\delta)\left[q_{DU} \cdot y - \lambda_k^i(s^0) \cdot y\right] + \delta \cdot U_{DU}(C_t, x_{t,DU} - y, h_t)$$

$$DAG_{k+1} = DAG_k \setminus \{DU_k\}$$

End For

not known a priori, each entity cannot solve their decision problems using the distributed algorithm in Table III. In this case, we can adapt the online learning algorithm based on post-decision state (PDS) in [16], which was originally proposed for single-user wireless video transmission, to the considered deployment scenario.

The main idea of the PDS-based online learning is to learn the post-decision value function, instead of the value function. Each user i's post-decision value function is defined as $\tilde{U}^i(\tilde{x}^i, \tilde{h}^i)$, where $(\tilde{x}^i, \tilde{h}^i)$ is the post-decision state. The difference from the normal state is that the PDS $(\tilde{x}^i, \tilde{h}^i)$ describes the status of the system *after* the scheduling action is made but before the DUs in the next period arrive. Hence, the relationship between the PDS and the normal state is

$$\tilde{x}^i = x^i - a^i, \ \tilde{h}^i = h^i.$$

Then the post-decision value function can be expressed in terms of the value function as follows:

$$\tilde{U}^i(\tilde{x}^i, \tilde{h}^i) = \sum_{x^{i\prime}, h^{i\prime}} p^i(x^{i\prime}, h^{i\prime}|\tilde{x}^i + a^i, \tilde{h}^i) \cdot \tilde{V}^i(x^{i\prime}, \tilde{h}^i).$$

In PDS-based online learning, the normal value function and the post-decision value function are updated in the following way:

$$\begin{aligned} V^i_{k+1}(x^i_k, h^i_k) &= \max_{a^i} \ (1-\delta) \cdot u^i(x^i_k, h^i_k, a^i) \\ &\quad + \ \delta \cdot U^i_k(x^i_k + (a^i - l^i_k), h^i_k), \\ U^i_{k+1}(x^i_k, h^i_k) &= (1 - \frac{1}{k}) U^i_k(x^i_k, h^i_k) \\ &\quad + \ \frac{1}{k} \cdot V^i_k(x^i_k - (a^i - l^i_k), h^i_k). \end{aligned}$$

We can see that the above updates do not require any knowledge about the state dynamics. In particular, we propose the decomposed optimal packet scheduling with PDS-based learning in Table V. Note that the difference between the learning algorithm in Table V with the algorithm assuming statistic knowledge in Table IV is that we use the post-decision state value function instead of the normal value function. It is proved in [16] that the PDS-based online learning will converge to the optimal value function. Hence, the distributed packet scheduling and resource allocation solution in Table III can be modified by letting each user perform the packet scheduling using the PDS-based learning in Table V.

E. Detailed Comparisons with Existing Frameworks

Since we have introduced our proposed framework, we can provide a detailed comparison with the existing theoretical framework. The comparison is summarized in Table VI.

First, the proposed framework reduces to the myopic optimization framework (repeated NUM) when we set the discount factor $\delta = 0$. In this case, the proposed solution reduces to the myopic solution.

Second, the Lyapunov optimization framework is closely related to the PDS-based online learning. In fact, it could be considered as a special case of the PDS-based online learning when we set the post-decision value function as $U^i(s^i) = u^i(s^i, a^i) + \|x^i - a^i + l^i\|_1^2 - \|x^i\|_1^2$, and choose the action that maximizes the post-decision value function at run-time. However, the Lyapunov drift in the above post-decision value function depends only on the total number of packets in the queue, but not on the delay deadlines, the dependency among packets, and the channel condition. In contrast, in our PDS-based online learning, we explicitly consider the impact of the video traffic (i.e. delay deadlines and dependency) and the channel condition when updating the normal and post-decision value functions.

Finally, the key difference between our proposed framework and the framework for MU-MDP [17] is how we penalize the constraints $\boldsymbol{f}(s^0, \boldsymbol{a})$. In particular, the framework in [17], if directly applied in our model, would define only one Lagrangian multiplier for all the constraints under different states s^0. This in general leads to performance loss [17]. In contrast, we define different Lagrangian multipliers to penalize the constraints under different states s^0, and enable the proposed framework to achieve the optimality (which is indeed the case as have been proved in Theorem 1).

VI. SIMULATION RESULTS

We consider a wireless network with multiple users streaming one of the following three video sequences, "Foreman" (CIF resolution, 30 Hz), "Coastguard" (CIF resolution, 30 Hz), and "Mobile" (CIF resolution, 30 Hz). We use the following system parameters by default and will explicitly specify when changes are made. The video sequences Foreman and Coastguard are encoded at a bit rate of 512 kb/s, and Mobile is encoded at 1024 kb/s due to its higher and more complicated video characteristics. Each GOP contains 16 frames and each encoded video frame can tolerate a delay of 266 ms (namely

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2014.2332299, IEEE Journal of Selected Topics in Signal Processing

TABLE VI. RELATIONSHIP BETWEEN THE PROPOSED AND EXISTING THEORETICAL FRAMEWORKS.

Framework	Relationship	Representative works
Myopic	$\delta = 0$	[4]–[9]
Lyapunov optimization	User i's post-decision value function $U^i(s^i) = u^i(s^i, a^i) + \|x^i - a^i + l^i\|_1^2 - \|x^i\|_1^2$	[18]
MU-MDP (uniform price)	Lagrangian multiplier $\boldsymbol{\lambda}(s^0) = \boldsymbol{\lambda}$ for all s^0, and $\lambda^i = \lambda^j$ for all users i, j	[17]
MU-MDP (optimal price)	Lagrangian multiplier $\boldsymbol{\lambda}(s^0)$ different for each s^0, and λ^i different for each users i	This work

TABLE VII. RESOURCE ALLOCATION AND PACKET SCHEDULING USED IN DIFFERENT SOLUTIONS.

	Resource allocation	Packet scheduling
Myopic [4]–[9]	myopic	earliest-deadline-first [9]
Lyapunov [18]	same as proposed (optimal)	based on Lyapunov drift
MU-MDP [17]	based on uniform price	same as proposed (optimal)
Proposed	as in Table III (optimal)	as in Table IV or V (optimal)

the duration of 8 frames, or half the duration of a GOP). The length of one time slot is 10 ms, and the scheduling time window W is 266 ms. The energy consumption function is set as $\rho(h^i, \|a^i\|_1) = \sigma^2(2^{\|a^i\|_1} - 1)/|h^i|^2, \forall i$ [16][17][20]. We quantize the continuous channel gains into 8 discrete channel states; the details of the quantization thresholds and representative values can be found in [16, Table II]. The ratio of the average channel gain to noise power is $\mathbb{E}\left\{|h^i|^2\right\}/\sigma^2 = 1.4$ ($\approx$ 1.5 dB) $\forall i$. We set the tradeoff parameter of distortion reduction and energy consumption as $\beta = 1$. The discount factor is $\delta = 0.95$.

A. Learning and Adaptation

Before comparing against the other solutions, we show that the proposed PDS-based learning algorithm converges to the optimal solution (namely the optimal value function is learned), and adapts to the nonstationary dynamics in the system (for example, tracks the optimal solution in the presence of dynamic entry and exit of users). The optimal solution is obtained by the proposed algorithm in Table III assuming the statistical knowledge of the system dynamics. We consider a scenario with 10 users streaming the video sequence Foreman. First, we compare long-term video quality (i.e. Peak Signal-to-Noise Ratio (PSNR)) achieved by the optimal solution and that achieved by the PDS-based learning algorithm. For illustrative purpose, we show the convergence of the learning algorithm in terms of long-term video quality only for two users in Fig. 5.

Next, in Fig. 6, we study how the proposed PDS-based learning algorithm adapts to the dynamic entry and exit of users. Specifically, a new user enters at time slot 5000, and then two users leave at time slot 6500. We show the PSNR of one particular user who stays in the network throughout the entire time. We can see that the proposed solution can adapt to the dynamic entry and exit of users, as long as it does not happen too frequently. Note also that under an entry or exit of users, the solutions converge faster, compared to the initial convergence starting from time 0. This is because the initial point in the new convergence process is already good (i.e. we have a "warm start").

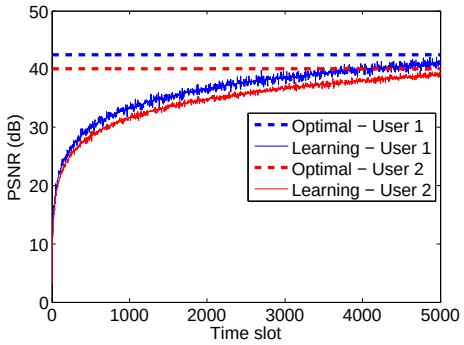

Fig. 5. Convergence of the PDS-based learning algorithm.

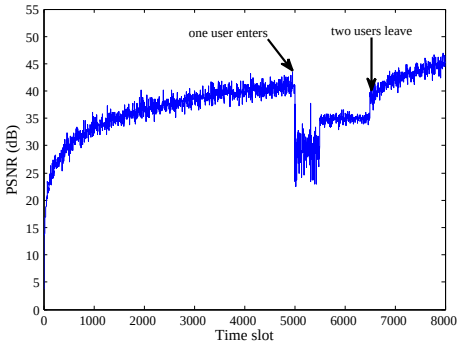

Fig. 6. Adapting to the dynamic entry and exit of users.

B. Comparison Against Existing Solutions

Now we compare against the myopic solution (i.e. repeated NUM) [4]–[9], the Lyapunov optimization solution (adapted from [18] for video transmission), and the MU-MDP solution [17]. In Table VII, we summarize the resource allocation and packet scheduling schemes adopted in the above solutions. Note that the Lyapunov solution is proposed for the single-user problem without resource allocation. To fairly compare the optimal packet scheduling with the packet scheduling in the Lyapunov solution that ignores video traffic, we adopt the proposed optimal resource allocation scheme in the Lyapunov solution. Similarly, since the proposed and MU-MDP solutions use the same optimal packet scheduling, we can fairly compare the proposed resource allocation with the suboptimal foresighted resource allocation based on uniform price.

1) Resource Allocation and Packet Scheduling Decisions Made by Different Solutions: Before comparing the performance of different solutions, we illustrate the resource allocation and packet scheduling of different solutions under one

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2014.2332299, IEEE Journal of Selected Topics in Signal Processing

realization of a sequence of traffic states and channel states. From this illustration, we can better understand the differences of the resource allocation and packet scheduling decisions made by different solutions and their impact on the video quality.

For illustrative purposes, we consider a simple but representative system with two users streaming video sequence Foreman. User 1 encodes the sequence with GOP structure "IPB" and user 2 with GOP structure "IPP". We fix the numbers of packets in the I, P, B frames to be 40, 10, 10, respectively. The scheduling time window is 2. The channel state has two values, "good" and "bad" (Note that we use binary channel states only to make our illustration easier to understand; The number of channel states in our model can be much larger than two.). When the channel state is good (bad), the two users can transmit up to 60 (40) packets in total. We illustrate the resource allocation and packet scheduling of different solutions in Tables VIII–XI.

In the myopic solution, the resource allocation is static and determined based on the total distortion impact of the users. Since the users have very similar GOP structures, the resource allocation is fixed at (0.50, 0.50) in all the time slots. The packet scheduling is EDF. From Table VIII, we can see that severe packet loss occurs due to three consecutive bad channel states. In particular, in the second bad channel state, user 1's I frame approaches its transmission deadline and needs to be transmitted, while user 2's I frame can wait to be transmitted in the next time slot. However, due to the static resource allocation, it can only transmit 20 packets and has to discard 20 packets of its I frame. The same problem occurs in the Lyapunov solution. However, since we adopt the proposed optimal resource allocation, only 7 packets of the I frame are lost in the second bad channel state. Another problem of both myopic and Lyapunov solutions is that they schedule based on deadlines without considering the distortion impact of the packets. For example, in the third bad channel state, user 1 has context (P, B, I) with the I frame of the next GOP having the latest deadline. As a consequence, both solutions schedule the packets of P and B frames, leaving many packets of I frame in the buffer. This results in the loss of 10 packets and 4 packets of the I frame in the myopic and Lyapunov solutions, respectively.

We can see from Table XI that the proposed solution achieves much better performance due to foresighted resource allocation and packet scheduling. First, notice that as in all the other solutions, the I frame experiences packet losses in the first time slot, where there are 80 packets of I frame in total and the transmission capacity is 60 packets. This packet loss is inevitable since there is no room for foresighted resource allocation and packet schedule ahead of the first time slot. The distinction of the proposed solution is evident in the bad channel states. In the second bad channel state, it allocates much more resource (72%) to user 1, although both users have a I frame to transmit. This is because user 1's I frame reaches its transmission deadline at this time slot. Hence, such an allocation meets the urgent need for resource of user 1 and avoids packet loss in its I frame. In addition, when the context is (P, B, I) with I frame of the next GOP having the latest deadline, the proposed solution will still schedule some packets of I frames at the expense of losing some packets of P and B frames. Due to foresighted resource allocation and packet scheduling, the proposed solution loses no packet of the I frames, which will result in much higher video quality than the other solutions.

The MU-MDP solution illustrated in Table X uses the same optimal packet scheduling as proposed and a suboptimal resource allocation. As we have discussed earlier, the MU-MDP solution uses a uniform price, which needs to be set high enough to avoid the violation of the resource constraints. Hence, the resource allocation is relatively conservative, as we can see from Table X. Although the initial resource allocation shown in the table will be scaled up by the BS as in [17] to ensure full resource utilization, the proportion of the resource allocation is suboptimal. This suboptimal resource allocation results in the loss of 7 packets of the I frame in the third bad channel state, because user 2 does not get enough resources.

2) Heterogeneity of Video Traffic: We compare the proposed solution with existing solutions in terms of PSNR and energy consumption. We change the tradeoff parameter β from 1 to 30 to get different PSNR and energy consumption tradeoffs. We assume that there are three users streaming the three different video sequences, respectively. The results are listed in Table XII. We can see that the proposed solution can achieve for all the users 7 dB PSNR improvement compared to the myopic solution, 5 dB PSNR improvement compared to the Lyapunov solution, and 3 dB PSNR improvement compared to the MU-MDP solution with uniform price. Moreover, the proposed solution can achieve high PSNR for all the three different video sequences, while existing solutions may result in low quality (e.g. less than 30 dB PSNR) especially for the more challenging "Coastguard" and "Mobile" videos.

3) Scaling With The Number of Users: Now we compare the proposed solution with existing solutions in terms of the average PSNR across users when the number of users increases. The energy consumption is fixed at 0.15 Joule. We assume that all the users stream the "Foreman" video sequence. Table XIII summarizes the results for different numbers of users. We can see that the performance gain of the proposed solution increases with the number of users. In particular, the proposed solution can achieve high enough PSNR for high-quality video transmission (e.g. larger than 30 dB PSNR) even when there are 20 users sharing the same resource. On the contrary, the myopic solution, the Lyapunov solution, and the MU-MDP solution may not achieve high-quality video transmission when the number of users exceeds 9, 9, and 13, respectively.

4) Different Channel Conditions and Total Rates: We compare different solutions under different channel conditions and total rates (which is the total rate of source coding and channel coding). We consider a scenario with 10 users streaming the video sequence Foreman. The channel condition (i.e. $|h|^2/\sigma^2$) varies from 1.0 dB to 2.0 dB, and the total rate varies from 256 kb/s to 1024 kb/s. We use adaptive error protection by changing the ratio of the source coding rate to the total rate. Specifically, under channel conditions of 1.0 dB, 1.5 dB, and 2.0 dB, the corresponding ratios of the source coding rate to the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2014.2332299, IEEE Journal of Selected Topics in Signal Processing

TABLE VIII. RESOURCE ALLOCATION AND PACKET SCHEDULING OF THE MYOPIC (REPEATED NUM) SOLUTION.

Traffic states	I(40), P(10), B(10) I(40), P(10)	P(10), B(10), I(40) P(10), P(10)	I(40), P(10), B(10) P(0), I(40)	P(10), B(10), I(40) I(20), P(10)	I(40), P(10), B(10) P(10), P(10)
Channel state	good	bad	bad	bad	good
Resource allocation	(0.50, 0.50)	(0.50, 0.50)	(0.50, 0.50)	(0.50, 0.50)	(0.50, 0.50)
Packet scheduling	I(30), P(0), B(0) I(30), P(0)	P(10), B(10), I(0) P(10), P(10)	I(20), P(0), B(0) P(0), I(20)	P(10), B(10), I(0) I(20), P(0)	I(30), P(0), B(0) P(10), P(10)
Packet loss	I(10), both users		I(20), user 1		I(10), user 1

TABLE IX. RESOURCE ALLOCATION AND PACKET SCHEDULING OF THE LYAPUNOV SOLUTION.

Traffic states	I(40), P(10), B(10) I(40), P(10)	P(10), B(10), I(40) P(10), P(10)	I(32), P(10), B(10) P(8), I(40)	P(10), B(10), I(40) I(33), P(10)	I(40), P(10), B(10) P(8), P(10)
Channel state	good	bad	bad	bad	good
Resource allocation	(0.50, 0.50)	(0.72, 0.28)	(0.63, 0.37)	(0.35, 0.65)	(0.64, 0.36)
Packet scheduling	I(30), P(0), B(0) I(30), P(0)	P(10), B(10), I(8) P(10), P(2)	I(25), P(0), B(0) P(8), I(7)	P(10), B(3), I(0) I(27), P(0)	I(36), P(0), B(0) P(8), P(10)
Packet loss	I(10), both users		I(7), user 1	I(6), user 2	I(4), user 1

TABLE X. RESOURCE ALLOCATION AND PACKET SCHEDULING OF THE MU-MDP SOLUTION.

Traffic states	I(40), P(10), B(10) I(40), P(10)	P(10), B(10), I(40) P(10), P(10)	I(40), P(10), B(10) P(6), I(40)	P(0), B(7), I(40) I(22), P(10)	I(24), P(10), B(10) P(8), P(10)
Channel state	good	bad	bad	bad	good
Resource allocation (before scaling)	(0.50, 0.50)	(0.46, 0.22)	(0.45, 0.42)	(0.45, 0.35)	(0.54, 0.31)
Packet scheduling	I(30), P(0), B(0) I(30), P(0)	P(0), B(0), I(26) P(10), P(4)	I(12), P(10), B(0) P(0), I(18)	P(0), B(7), I(16) I(15), P(2)	I(24), P(10), B(10) P(8), P(10)
Packet loss	I(10), both users		P(6), user 2	I(7), user 2	

TABLE XI. RESOURCE ALLOCATION AND PACKET SCHEDULING OF THE PROPOSED SOLUTION.

Traffic states	I(40), P(10), B(10) I(40), P(10)	P(10), B(10), I(40) P(10), P(10)	I(40), P(10), B(10) P(8), I(40)	P(0), B(7), I(40) I(25), P(10)	I(27), P(10), B(10) P(8), P(10)
Channel state	good	bad	bad	bad	good
Resource allocation	(0.50, 0.50)	(0.72, 0.28)	(0.63, 0.37)	(0.35, 0.65)	(0.64, 0.36)
Packet scheduling	I(30), P(0), B(0) I(30), P(0)	P(0), B(0), I(28) P(10), P(2)	I(12), P(10), B(3) P(0), I(15)	P(0), B(0), I(13) I(25), P(2)	I(27), P(9), B(0) P(8), P(10)
Packet loss	I(10), both users		P(8), user 2	B(7), user 2	

total rate are 50%, 60%, and 75%, respectively. In other words, we have better error protection when the channel is worse. We summarize the PSNR achieved by different solutions in Table XIV. We can see that the proposed solution outperforms the other solutions in various channel conditions and encoding rates considered, with an average improvement in PSNR of 10 dB over the myopic solution, 7 dB over the Lyapunov solution, and 5 dB over the MU-MDP solution.

C. Comparison of Resource Allocation Under Different Packet Scheduling Algorithms

The proposed solution consists of two parts: packet scheduling and resource allocation. The proposed solution is optimal when each user adopts the optimal packet scheduling. However, the optimal packet scheduling is achieved by solving a MDP, which may have high computational complexity (even though we have decoupled the users' decision problems and thus have reduced the state space of each user's MDP). In practice, however, the users may adopt simpler but suboptimal

TABLE XII. COMPARISONS OF PSNR UNDER DIFFERENT ENERGY CONSUMPTIONS.

Energy consumption (Joule)	0.08	0.10	0.15
Myopic (repeated NUM) [4]–[9]	(30, 24, 24) dB	(40, 35, 25) dB	(45, 45, 34) dB
Lyapunov [18]	(31, 26, 26) dB	(42, 38, 27) dB	(47, 47, 37) dB
MU-MDP [17]	(34, 28, 27) dB	(44, 40, 28) dB	(50, 48, 38) dB
Proposed	(37, 32, 30) dB	(46, 43, 32) dB	(54, 51, 42) dB

TABLE XIII. COMPARISONS OF PSNR UNDER DIFFERENT NUMBERS OF USERS.

Number of users	3	5	7	9	11	13	15	20
Myopic (repeated NUM) [4]–[9]	45 dB	41 dB	36 dB	31 dB	27 dB	24 dB	20 dB	15 dB
Lyapunov [18]	47 dB	45 dB	42 dB	36 dB	29 dB	26 dB	23 dB	20 dB
MU-MDP [17]	50 dB	46 dB	44 dB	39 dB	32 dB	30 dB	26 dB	23 dB
Proposed	54 dB	52 dB	49 dB	44 dB	38 dB	36 dB	32 dB	30 dB

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2014.2332299, IEEE Journal of Selected Topics in Signal Processing

TABLE XIV. COMPARISONS OF PSNR UNDER DIFFERENT CHANNEL CONDITIONS AND TOTAL RATES.

(dB, kb/s)	(1.0, 256)	(1.0, 512)	(1.0, 1024)	(1.5, 256)	(1.5, 512)	(1.5, 1024)	(2.0, 256)	(2.0, 512)	(2.0, 1024)
Myopic [4]–[9]	27 dB	28 dB	30 dB	29 dB	30 dB	31 dB	30 dB	32 dB	34 dB
Lyapunov [18]	31 dB	33 dB	34 dB	32 dB	34 dB	35 dB	34 dB	35 dB	37 dB
MU-MDP [17]	34 dB	35 dB	36 dB	35 dB	37 dB	37 dB	38 dB	39 dB	40 dB
Proposed	38 dB	39 dB	41 dB	40 dB	41 dB	42 dB	43 dB	44 dB	46 dB

TABLE XV. COMPARISONS OF PSNR UNDER SIMPLER SCHEDULING ALGORITHMS.

Packet scheduling	EDF	FIFO	HDF	Optimal
Myopic [4]–[9]	24 dB	22 dB	25 dB	29 dB
MU-MDP [17]	34 dB	32 dB	33 dB	36 dB
Proposed	38 dB	37 dB	38 dB	41 dB

packet scheduling algorithms. Next, we demonstrate that such simpler packet scheduling algorithms can be easily combined with our proposed resource allocation scheme, and show that when using simpler scheduling algorithms, our resource allocation scheme still outperforms the resource allocation of the myopic and MU-MDP solutions with uniform price. Specifically, we consider the following three simpler scheduling algorithms:

- EDF (Earliest Deadline First) scheduling: schedule all the packets of the frame with the earliest deadline first within its deadline.
- FIFO (First-In First-Out) scheduling: schedule all the packets of the frame that comes in the buffer first within its deadline.
- HDF (Highest Distortion First) scheduling: schedule all the packets of the frame with the highest distortion impact first within its deadline.

We consider a scenario with 10 users streaming the video sequence Foreman. The channel condition (i.e. $|h|^2/\sigma^2$) is 1.5 dB, and the encoding rate is 512 kb/s. In Table XV, we compare the proposed resource allocation with myopic resource allocation and the resource allocation in the MU-MDP solution under different scheduling algorithms. We can see that our proposed resource allocation scheme outperforms the other resource allocation schemes even when we use simpler scheduling algorithms. Moreover, the performance loss induced by using simpler scheduling algorithms is smaller under our proposed solution (around 3 dB), compared to the myopic solution (around 5 dB).

VII. CONCLUSION

We propose the optimal foresighted resource allocation and packet scheduling for multi-user wireless video transmission. The proposed solution achieves the optimal long-term video quality subject to each user's minimum video quality guarantee, by dynamically allocating resources among the users and dynamically scheduling the users' packets while taking into account the dynamics of the video traffic and channel states. We develop a low-complexity algorithm that can be implemented by the BS and the users in an informationally-decentralized manner and converges to the optimal solution. Through extensive simulation, we demonstrate the performance gain of our proposed solution over existing solutions under a wide range of deployment scenarios: different number of users, different channel conditions, different video encoding rates, and different (simpler but suboptimal) packet scheduling algorithms. The simulation results show that our proposed solution can achieve significant improvements in PSNR of up to 7 dB compared to myopic solutions and of up to 3 dB compared to state-of-the-art foresighted solutions.

Our solution is versatile and can be applied to 4G LTE cellular networks or IEEE 802.11a/e wireless LANs with real-time or offline video encoding. It can be readily extended to more general multi-user resource allocation and scheduling problems with correlated traffic states and channel states, where the video traffic is generated by the video encoder at run-time based on the time-varying channel conditions.

APPENDIX A
PROOF OF THEOREM 1

Due to limited space, we give a detailed proof sketch. The proof consists of three key steps. First, we prove that by penalizing the constraints $f(s^0, a^1, \ldots, a^I) \leq 0$ into the objective function, the decision problems of different entities can be decentralized. Hence, we can derive optimal decentralized strategies for different entities under given Lagrangian multipliers. Then we prove that the update of Lagrangian multipliers converges to the optimal ones under which there is no duality gap between the primal problem and the dual problem, due to the convexity assumptions made on the cost functions. Finally, we validate the calculation of the prices.

First, suppose that there is a central controller that knows everything about the system. Then the optimal strategy to the design problem (4) should result in a value function V^* that satisfies the following Bellman equation: for all $s^1, \ldots, s^I$, we have

$$V^*(s^1, \ldots, s^I) = \max_{a^1,\ldots,a^I} (1-\delta) \sum_{i=1}^{I} u^i(s^i, a^i) + \delta \cdot \mathbb{E}\left\{V^*(s^{1\prime}, \ldots, s^{I\prime})\right\} \quad s.t. \quad f(s^0, a^1, \ldots, a^I) \leq 0. \tag{8}$$

Defining a Lagrangian multiplier $\lambda(s^0) \in \mathbb{R}_+^N$ associated with the constraints $f(s^0, a^1, \ldots, a^I) \leq 0$, and penalizing the constraints on the objective function, we get the following Bellman equation:

$$V^{\boldsymbol{\lambda}}(s^1, \ldots, s^I) = \max_{a^1,\ldots,a^I} (1-\delta)\left[\sum_{i=1}^{I} u^i(s^i, a^i) + \lambda(s^0) f(s^0, a^1, \ldots, a^I)\right] + \delta \cdot \mathbb{E}\left\{V^*(s^{1\prime}, \ldots, s^{I\prime})\right\}. \tag{9}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2014.2332299, IEEE Journal of Selected Topics in Signal Processing

In the following lemma, we can prove that (9) can be decomposed.

Lemma 2: The optimal value function V^{λ} that solves (9) can be decomposed as $V^{\lambda}(s^1,\ldots,s^I)=\sum_{i=1}^{I}V^{\lambda,i}(s_i)$ for all $(s^1,\ldots,s^I)$, where $V^{\lambda,i}$ can be computed by user i locally by solving

$$V^{\lambda,i}(s^i)=\max_{a^i}\ (1-\delta)\cdot\left[u^i(s^i,a^i)+\lambda f^i(s^0,a^i)\right] \quad (10)$$
$$+\delta\cdot\sum_{s^{i\prime}}p^i(s^{i\prime}|s^i,a^i)V^{\lambda,i}(s^{i\prime}).$$

Proof: This can be proved by the independence of different entities' states and by the decomposition of the constraints $f(s^0,a^1,\ldots,a^I)$. Specifically the constraints $f(s^0,a^1,\ldots,a^I)$ are linear with respect to the actions $a_1,\ldots,a_I$. As a result, we can decompose the constraints as $f(s^0,a^1,\ldots,a^I)=\sum_{i=1}^{I}f^i(s^0,a^i)$. ■

We have proved that by penalizing the constraint $f(s^0,a^1,\ldots,a^I)$ using Lagrangian multiplier $\lambda(s_0)$, different entities can compute the optimal value function $V^{\lambda(s^0),i}$ distributively. Due to the convexity assumptions on the cost functions, we can show that the primal problem (4) is convex. Hence, there is no duality gap. In other words, at the optimal Lagrangian multipliers $\lambda^*(s^0)$, the corresponding value function $V^{\lambda^*(s^0)}(s^1,\ldots,s^I)=\sum_{i=1}^{I}V^{\lambda^*(s^0),i}(s^i)$ is equal to the optimal value function $V^*(s^1,\ldots,s^I)$ of the primal problem (8). It is left to show that the update of Lagrangian multipliers converge to the optimal ones. It is a well-known result in dynamic programming that $V^{\lambda(s^0)}$ is convex and piecewise linear in $\lambda(s^0)$, and that the subgradient is $f(s^0,a^1,\ldots,a^I)$. Note that we use the sample mean of a^i, whose expectation is the true mean value of a^i. Since $f(s^0,a^1,\ldots,a^I)$ is linear in a^i, the subgradient calculated based on the sample mean has the same mean value as the subgradient calculated based on the true mean values. In other words, the update is a stochastic subgradient descent method. It is well-known that when the stepsize $\Delta(k)=\frac{1}{k+1}$, the stochastic subgradient descent will converge to the optimal λ^* [24].

Finally, we can write the prices by taking the derivatives of the penalty terms. For user i, its penalty is $\lambda(s^0)\cdot f^i(s^0,a^i)$. Hence, its price is

$$\frac{\partial\lambda^i(s^0)\cdot f^i(s^0,a^i)}{\partial a^i}=\lambda(s_0)\cdot\frac{\partial f^i(s^0,a^i)}{\partial a^i}=\lambda(s_0)\cdot\frac{b}{r^i(h^i)}.$$

REFERENCES

[1] A. Reibman, and A. Berger, "Traffic descriptors for VBR video teleconferencing over ATM networks," *IEEE/ACM Trans. Netw.*, vol. 3, no. 3, pp. 329–339, Jun. 1995.

[2] P. Li, H. Zhang, B. Zhao, and S. Rangarajan, "Scalable video multicast with joint layer resource allocation in broadband wireless networks," in *Proc. IEEE International Conference on Network Protocols (ICNP)*, pp. 295–304, Oct. 2010.

[3] Y.-J. Yu, P.-C. Hsiu, and A.-C. Pang, "Energy-efficient video multicast in 4G wireless systems," *IEEE Trans. Mobile Comput.*, vol. 11, no. 10, pp. 1508–1522, Oct. 2012.

[4] M. van der Schaar, Y. Andreopoulos, and Z. Hu, "Optimized scalable video streaming over IEEE 802.11 a/e HCCA wireless networks under delay constraints," *IEEE Trans. Mobile Comput.*, vol. 5, no. 6, pp. 755–768, Jun. 2006.

[5] G-M. Su, Z. Han, M. Wu, and K.J.R. Liu, "Joint uplink and downlink optimization for real-time multiuser video streaming over WLANs," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 2, pp. 280–294, Aug. 2007.

[6] X. Zhang and Q. Du, "Cross-layer modeling for QoS-driven multimedia multicast/broadcast over fading channels in mobile wireless networks," *IEEE Commun. Mag.*, vol. 45, no. 8, pp. 62–70, Aug. 2007.

[7] J. Huang, Z. Li, M. Chiang, and A.K. Katsaggelos, "Joint source adaptation and resource allocation for multi-user wireless video streaming," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 5, pp. 582–595, May 2008.

[8] E. Maani, P. Pahalawatta, R. Berry, T. N. Pappas and A. K. Katsaggelos, "Resource allocation for downlink multiuser video transmission over wireless lossy networks," *IEEE Trans. on Image Process.*, vol. 17, no. 9, 1663–1671, Sep. 2008.

[9] X. Ji, J. Huang, M. Chiang, G. Lafruit, and F. Catthoor, "Scheduling and resource allocation for SVC streaming over OFDM downlink systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 10, pp. 1549–1555, Oct. 2009.

[10] P. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Trans. Multimedia*, vol. 8, no. 2, pp. 390–404, Apr. 2006.

[11] S. Khan, M. Sgroi, Y. Peng, E. Steinbach, W. Kellerer, "Application-driven cross layer optimization for video streaming over wireless networks," IEEE Commun. Mag. , pp. 122–130, Jan. 2006.

[12] C. De Vleeschouwer and P. Frossard, "Dependent packet transmission policies in rate-distortion optimized media scheduling," *IEEE Trans. Multimedia*, vol. 9, no. 6, pp. 1241–1258, Oct. 2007.

[13] Z. Li, F. Zhai and A. Katsaggelos, "Joint video summarization and transmission adaptation for energy-efficient wireless video streaming," *EURASIP J. on Adv. in Signal Processing, special issue on Wireless Video*, 2008.

[14] N. Tizon and B. Pesquet-Popescu, "Scalable and media aware adaptive video streaming over wireless networks," *EURASIP Journal on Advances in Signal Processing*, Volume 2008 (2008), Article ID 218046, 11 pages.

[15] H. Wang and A. Ortega, "Rate-distortion optimized scheduling for redundant video representations," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 225–240, Feb. 2009.

[16] F. Fu and M. van der Schaar, "Structural solutions for dynamic scheduling in wireless multimedia transmission," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 5, pp. 727–739, May 2012.

[17] F. Fu and M. van der Schaar, "A systematic framework for dynamically optimizing multi-user video transmission," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 308–320, Apr. 2010.

[18] W. Chen, M. J. Neely, and U. Mitra, "Energy-efficient transmission with individual packet delay constraints," *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 2090–2109, May 2008.

[19] F. Fu and M. van der Schaar, "Structure-aware stochasticcontrol for transmission scheduling," *IEEE Trans. Veh. Tech.* vol. 61, no. 9, pp. 3931–3945, Nov. 2012.

[20] D. Bertsekas and R. Gallager, *Data Networks*. Upper Saddle River, NJ: Prentice-Hall, 1987.

[21] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.

[22] J. Hawkins, "A Lagrangian decomposition approach to weakly coupled dynamic optimization problems and its applications," PhD Dissertation, MIT, Cambridge, MA, 2003.

[23] *IEEE Standard for Local and Metropolitan Area Networks, Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs) Amendment 1: MAC sublayer*, IEEE Computer Society, April. 16, 2012, online at: http://standards.ieee.org/getieee802/download/802.15.4e-2012.pdf

[24] H. Kushner and G. Yin, Stochastic Approximation and Recursive Algorithms and Applications. Springer, 2003.