A Pricing Mechanism for Resource Allocation in Wireless Multimedia Applications

Fangwen Fu, Tudor M. Stoenescu, and Mihaela van der Schaar, *Senior Member, IEEE*

***Abstract*—We consider the problem of multiuser resource allocation for wireless multimedia applications deployed by autonomous and noncollaborative wireless stations (WSTAs). Existing resource allocation solutions for WLANs are not network-aware and do not take into account the selfish behavior of individual WSTAs. Specifically, the selfish WSTAs can manipulate the network by untruthfully representing their private information (i.e., video characteristics, experienced channel conditions, and deployed streaming strategies). This often results in inefficient resource allocations. To overcome this obstacle, we present a pricing mechanism for message exchanges between the WSTAs and the Central Spectrum Moderator (CSM). The messages represent network-aware resource demands and corresponding prices. We prove that the message exchanges reach the Nash equilibrium and that the resulting equilibrium messages generate allocations which are efficient, budget balanced, and satisfy voluntary participation. The simulation results verify that these properties hold when the WSTAs behave strategically. Additionally, we evaluate the impact of initial prices and network congestion level on the convergence rate of message exchanges.**

***Index Terms*—Game theory, multiuser wireless multimedia streaming, pricing mechanism, resource allocation.**

I. INTRODUCTION

WIRELESS networks are envisioned to play a crucial role in the delivery of various delay-sensitive multimedia services to homes, enterprises, and campuses. A fundamental problem in enabling the large scale deployment of such networks is the absence of effective resource allocation schemes, which can arbitrate the division of the scarce wireless resource among competing (high-bandwidth and delay-sensitive) multimedia users.

There are two main challenges in designing efficient resource allocation schemes for wireless media. First, in current wireless standards, multimedia users are able to unfairly compete for resources by misrepresenting their Quality-of-Service (QoS) requirements [1]. For instance, in existing WLANs, such as IEEE 802.11a Point Coordination Function (PCF) [2] and 802.11e Hybrid Coordination Function (HCF) [3], the available resources are divided among competing stations through a polling-based mechanism. This mechanism is deployed by a Central Spectrum Moderator (CSM), e.g., the access point, and it is based on resource reservation requests that are negotiated by the autonomous wireless stations (WSTAs) when they first join the network. The CSM, often implemented at the Medium Access Control (MAC) layer, is assumed to be able to take into consideration information from other layers when determining policies to divide the available resources. In current state-of-the-art reservation (admission-control) based schemes, each wireless station tries to acquire as much of the network resource as possible by declaring a traffic specification (TSPEC) based on worst-case traffic estimates [3]. In wireless networks arbitrated by the recently standardized 802.11a/e WLAN admission control protocols, if some users misrepresent their TSPEC requirements, the performance of the entire wireless network may degrade considerably [4].

Recently, several fair resource allocation algorithms [35], [36] have been proposed for wireless multimedia applications. In [35], a max–min fairness allocation is presented using a combination of the bandwidth reservation and bandwidth borrowing to provide the network users the required QoS. In [36], a Nash bargaining solution is proposed to divide the available resources in order to achieve a utility-fair allocation. However, these proposed algorithms assume that all WSTAs truthfully reveal their resource requirements. This is not always true when the wireless users are selfish [1].

A second challenge in the design of resource management schemes for WLANs comes from the informationally decentralized nature of the wireless resource allocation [3], [4]. Each WSTA can derive different video quality benefits based on the various resources allocated by the CSM. For each WSTA, the quality benefit depends on its private information, which is represented by its video characteristics, channel conditions, as well as its deployed streaming strategies. In general, the private information of each user is not known by the CSM or other WSTAs. Also, the users are not directly aware of the other WSTAs requesting resources from the CSM. To address the informationally-decentralized nature of the network, pricing-based distributed resource allocation algorithms have been extensively investigated [6], [7], where the price reflects the congestion in the network and the network users adjust their traffic based on the resource price. However, the algorithms assume that the network users are "price-takers," i.e., the users accept the price announced by the network and do not consider the effects of their actions on the network price. If the network users anticipate these effects (we refer to such

Manuscript received November 1, 2006; revised May 7, 2007. This work was supported by NSF CAREER Award CCF-0541867 and CCF-0541453, and grants from UC Micro. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Adriana Dumitras.

F. Fu and M. van der Schaar are with the Electrical Engineering Department, University of California, Los Angeles, CA 90095-1594 USA (e-mail: fwfu@ee.ucla.edu; mihaela@ee.ucla.edu).

T. M. Stoenescu is with the NASA Jet Propulsion Laboratory, Pasadena, CA 91109 USA (e-mail: tudor@caltech.edu).

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

Digital Object Identifier 10.1109/JSTSP.2007.901519

1932-4553/$25.00 © 2007 IEEE

users as "strategic" players), the above algorithms will lead to an inefficient allocation [38]. Such resource allocations among strategic players have been extensively studied by mathematical economists in the context of mechanism design [17], [27].

In this paper, we focus on the problem of autonomous wireless stations (WSTAs) that are deployed by various noncollaborative and strategic users. From now on, we use the expressions "WSTA" and "user" interchangeably. These stations compete for wireless resources (transmission time) in order to transmit video data in real-time over a shared wireless LAN (WLAN) infrastructure. To address the abovementioned challenges, new schemes for resource allocation in multimedia environments need to be devised, which i) maximize the network utilization; ii) take into account the "self-interested" behavior of individual users that may try to selfishly influence the resource allocation process; and iii) satisfy the informational constraints imposed by the informationally decentralized nature of the investigated resource allocation problem.

In this paper, to enforce WSTAs to declare their resource requirements truthfully and to act in a socially optimal way, we adopt a game-theoretic pricing mechanism for the CSM to implement the time allocation in polling-based WLANs. In this mechanism, each WSTA communicates with the CSM. Hence, unlike existing WLAN standards, where a single message is transmitted for the resource allocation (i.e., the TSPEC), in our proposed scheme, each WSTA transmits to the CSM two messages: a resource demand and a corresponding price. The CSM announces then back to the WSTA three messages: the average price announced by the other WSTAs, the resources consumed by the other WSTAs, and the update step size of the price. Finally, this process is repeated until the message equilibrium (which represents the set of users' demands and resource prices from which no user would like to deviate) is reached. After the equilibrium is reached, the CSM allocates the negotiated transmission opportunities (TXOPs) to the WSTAs. The proposed mechanism can generate efficient TXOP allocations when the WSTAs behave "strategically."

Unlike other resource allocation schemes, which are based on Vickery–Clarke–Groves (VCG) mechanisms [17] and study the resource allocation problem from a dominant strategy perspective, the proposed mechanism implements the solutions in Nash equilibrium. The proposed approach enables us to provide solutions for the wireless multimedia resource allocation that are not only simpler in terms of communication and computational cost, but also satisfy the important properties of voluntary participation (i.e., users prefer to participate in the resource exchange rather than not participate) and are budget balanced (i.e., all the money users pay to the wireless network is allocated back to them). The budget balanced property is very important for existing WLANs since it prevents the CSM from behaving as a profit maker and trying to alter the users' allocations in order to maximize its revenue. Such a requirement holds in most of the aforementioned wireless multimedia applications.

In summary, our paper makes the following contributions.

1) It adds a new dimension to existing wireless systems by enabling them to proactively compete for the limited wireless resources based on their video characteristics and channel conditions.
2) It introduces a novel pricing-based mechanism for resource allocation and management in multiuser wireless environments, where WSTAs are allowed to compete for the available resources. The proposed mechanism, while taking into account the strategic behavior of individual users, generates allocations that i) maximize the sum of the users' expected received video quality, ii) are budget balanced, and iii) satisfy the property of voluntary participation.
3) It leads to allocations that are easier to enforce and are superior from a "fairness" standpoint [1] than the ones generated by the existing multiuser wireless multimedia resource allocation mechanisms such as IEEE 802.11e [3].

The paper is organized as follows. In Section II, we describe the wireless system for multimedia applications and formulate the centralized resource allocation problem. In Section III, we present the expected received video quality-resource function using the priority queuing model, which is required for the multimedia users to strategically maximize their own utilities. In Section IV, we propose a pricing mechanism to implement the decentralized resource allocation. In Section V, given the resource allocated by the pricing mechanism, we illustrate the real-time transmission strategies that are deployed by the WSTAs to stream their video packets over the wireless medium. In Section VI, we present the simulation results, followed by the conclusions in Section VII.

II. System Description

We consider M autonomous WSTAs $(1 \leq i \leq M)$ that are streaming video content in real-time over a shared one-hop WLAN infrastructure. These WSTAs are competing for the available wireless resource $\boldsymbol{\mathcal{R}} \in \mathbb{R}_+$, which in our system represents the amount of time (e.g., TXOPs) that can be allocated to the WSTAs. We assume that a polling-based mechanism (similar to that adopted in the QoS-enabled MAC of IEEE 802.11e [3]) is deployed by the CSM to divide the available resources among the competing WSTAs. The total available wireless resource $\boldsymbol{\mathcal{R}}$ is allocated to the M WSTAs, each of which receives $\alpha_i\boldsymbol{\mathcal{R}}$ resources with $(0 \leq \alpha_i \leq 1)$. Thus, the resource allocation vector is denoted as $\boldsymbol{\alpha} = [\alpha_1 \ \cdots \ \alpha_M]$. Note that $\sum_{i=1}^{M} \alpha_i \leq 1$, where the inequality is due to the overhead of signaling and synchronizing in the wireless network. However, for simplicity, in this paper we ignore the proportion of time used for signaling and synchronization (as this can be easily subtracted from the total available wireless resources). The multimedia users are modeled as selfish and strategic users that try to maximize their utility benefits. Given a resource allocation α_i, each multimedia user i has an expected received video quality, denoted as $Q_i^{rec}(\alpha_i)$, which depends on the video traffic characteristics, deployed transmission strategies as well as experienced channel conditions. In Section III, we will discuss how to determine $Q_i^{rec}(\alpha_i)$.

User i's utility function is summarized by a quasi linear utility function of the form $U_i(\alpha_i, \tau_i) = Q_i^{rec}(\alpha_i) + \tau_i$. The term $\tau_i \in \mathbb{R}$ is the "numeraire" commodity, which represents the tax incurred by the user when participating in the resource management game. Although the numeraire commodity generally represents money, in our problem it can be any type of tradable resource that is available to the users. In general, a constraint on the availability of the numeraire commodity needs to be imposed. In particular, we assume that each user should possess enough numeraire commodity to be able to afford to pay for the excess amount of resource needed to achieve an optimal wireless resource allocation.

The resource allocation aims to maximize the "social" welfare [1], which in our problem is characterized by the sum of the users' utilities, as in [17]. Formally, the resource allocation problem can be formulated as

$$\begin{aligned} &\max_{\boldsymbol{\alpha},\boldsymbol{\tau}} \sum_{i=1}^{M} U_i(\alpha_i,\tau_i) \\ &s.t. \sum_{i=1}^{M} \alpha_i \le 1 \\ &\quad 0 \le \alpha_i \le 1, \quad \forall i \in \{1,\ldots,M\} \\ &\quad \sum_{i=1}^{M} \tau_i \le 0 \end{aligned} \tag{1}$$

with $\boldsymbol{\tau} = [\tau_1 \cdots \tau_M]$ being the tax vector for all the users. The last constraint in the problem in (1) comes from the fact that the amount of numeraire commodity in the system after the allocation cannot exceed the amount before the allocation. The above optimization problem can potentially be solved in a centralized fashion by the use of linear and nonlinear programming techniques [28], [29]. Unfortunately, due to the informational constraints involved in the noncollaborative multiuser wireless multimedia transmission, this is not possible unless the CSM knows all the users' private information. Hence, a decentralized negotiation process is required to find the optimal solution to our problem.

To transmit multimedia over the current WLANs, the WSTAs are required to submit their TSPECs to the CSM based on their traffic models [2], [3] before they actually start transmitting the video data. Based on the submitted TSPECs, the CSM allocates the wireless resources. After this resource allocation is finalized, the WSTAs begin to transmit their video packets. Note that the WSTAs only submit the TSPECs when they join the network. During the transmission, the WSTAs can deploy various transmission strategies to cope with the dynamics of the source characteristics and channel conditions [4]. However, in reality, in order to maximize their own utilities, the users will strategically anticipate the impact of their own actions on the network resource allocation. To prevent the users from misusing the network resource in this manner, we propose an efficient negotiation process (i.e., message exchanges between the users and the CSM) which explicitly considers the strategic behaviors of WSTAs. Similar to the TSPEC used in current WLANs, the message exchanges between the WSTAs and the CSM are composed of several scalar numbers representing the resource demand and corresponding prices based on their video quality-resource models. In Section III, we present such a video quality-resource model. However, we note that the proposed message exchange framework can be applied to other video quality-resource models. The message exchange is performed iteratively until the equilibrium is reached. In Section IV, we describe in detail the message exchange procedure. Similar to the resource negotiation performed in existing WLANs, the proposed mechanism takes place only when large variations occur in the network, e.g., new WSTAs join the wireless network. Again, similar to the current approach in existing WLANs, while the resource renegotiation is finalized, the WSTA will continue transmitting their video packets based on the newly negotiated resources.

III. Expected Received Video Quality for Multimedia Users

To analyze the interaction between the multimedia users and the CSM, a model for determining the expected received video quality $Q_i^{rec}(\alpha_i)$ for user i is required. In this section, we will derive $Q_i^{rec}(\alpha_i)$ based on a priority queuing model. Specifically, in Section III-A, the video packets from user i are divided into several priority classes based on their different contributions to the video quality. In Section III-B, a priority queuing model is used to model the video transmission. In Section III-C, $Q_i^{rec}(\alpha_i)$ is derived for user i based on the proposed priority queuing model.

A. Priority Video Classes

In [4], [8], it has been shown that partitioning the packets into different priority classes and correspondingly adjusting the transmission strategies for each class can significantly improve the overall received quality and provide graceful degradation as congestion levels and channel conditions change. Similarly, in this paper, we will divide the packets of each encoded video stream into several priority classes based on their impact on the video distortion and their delay constraints. For instance, the video data can be divided into different priority classes[1] using data partition [33], [34] for hybrid video coders (e.g., H.264) or using spatio–temporal-SNR layering for 3-D wavelet video coders [9]. The packets belonging to the same class compose one priority class and are assumed to have the same contributions to the reconstructed video quality (i.e., the same "priority"). The number of priority classes for user i equals H_i. We assume that each packet of class $h (1 \le h \le H_i)$ has the quality contribution $\lambda_{i,h}$. Note that the quality contribution $\lambda_{i,h}$ depends on the underlying content characteristics, encoding parameters, etc. and typically increases with the importance or distortion impact of the packet and can be determined as in [25]. We assume that the packets are prioritized in descending order of their quality contribution, i.e., $\lambda_{i,1} > \lambda_{i,2} > \ldots > \lambda_{i,H_i}$. For simplicity, we assume that the packet length L_i (which includes the various packet headers, etc.) is constant for a specific WSTA i. The optimal packet length can be determined as in [30].

B. Priority Queuing Model for Packet Transmission

Before starting to stream their video packets over the wireless network, multimedia users are required to submit their resource requirements to the CSM. For example, in WLAN 802.11e, the users are required to submit their TSPECs, which include mean

[1]It should be noted that the particular prioritization schemes do only affect the video quality performance of the wireless stations, and not the proposed mechanism for resource allocation. Our proposed pricing mechanism is generic and can be applied in a similar manner to different video coders.

data rates, peak data rates and delay bounds, to the CSM. However, in our resource allocation game, user i has to proactively determine its $Q_i^{rec}(\alpha_i)$ over the set of potential resource allocations α_i in order to determine the messages it should exchange with the CSM. We note that the TSPEC only captures the worst case resource requirement and does not explicitly represent the resulting quality experienced by the user, i.e., $Q_i^{rec}(\alpha_i)$. On the other hand, the received video quality is influenced by many factors, e.g., video traffic characteristics, transmission strategies, experienced channel conditions, allocated resources, etc. All these factors are actually not known a priori and can only be estimated during the actual real-time transmission. Hence, we introduce a priority queuing model for the video packet transmission based on which a WSTA can estimate the resulting $Q_i^{rec}(\alpha_i)$. The priority queuing model is especially useful for deriving $Q_i^{rec}(\alpha_i)$ for the following reasons.

i. Using priority queuing models, the uncertainty of video packets' arrivals and transmission time can be analytically described and $Q_i^{rec}(\alpha_i)$ can be explicitly determined.

ii. The transmitted packets have different contributions to the reconstructed video quality, and thus applying priority-aware transmission strategies to the different packet classes can significantly improve the received video quality.

iii. The priority queuing model captures the steady state performance, e.g., waiting time distribution, and packet loss probability, which directly impact $Q_i^{rec}(\alpha_i)$.

It is worthwhile to note that the more accurate the models are for the video traffic, transmission strategies and channel conditions, the more accurate the quality estimation $Q_i^{rec}(\alpha_i)$ becomes. On the other hand, the queuing models should also be easy to compute for each WSTA.

1) Priority Queuing Assumptions: The adopted priority queuing analysis is based on the following assumptions.

i. The packet arrivals of each priority class $h (1 \leq h \leq H_i)$ from user i is assumed to be a Poisson process as in [31], i.e., the distribution of the duration, denoted as $A_{i,h}(t)$, between two sequent packet arrivals from the same class is assumed to be exponential [10]. Thus, the average packet arrival rate can be computed as $r_{i,h} = 1/E[A_{i,h}(t)]$. We note that a more complicated packet arrival model can also be applied [11]. However, as mentioned before, the tradeoff between the model complexity and accuracy has to be taken into account.

ii. The highest priority packets present in the buffer will be the first to be transmitted whenever the transmission opportunities are available. As in [4], a packet with higher priority will be transmitted repeatedly until it reaches the destination or it expires due to its delay constraint. Moreover, while a packet is being sent, its transmission will not be interrupted by the newly generated higher priority packets arriving in the transmission buffer. This transmission policy is denominated nonpreemptive priority transmission policy [11]. Hence, the packet service (transmission) time, $X_i(\alpha_i)$, is a function of the allocated resource α_i, which can be approximated by a geometric distribution [10]. The service time can be determined based on the experienced channel conditions expressed by the SNR and the allocated transmission opportunities (e.g., TXOP). Based on these, the maximum transmission rate $R_i^{\max}$ and the packet error probability, e_i, can be easily computed, as shown in [12]. Using the geometric distribution, the first and second moments of the service time can be approximated by [10]

$$E[X_i(\alpha_i)] \approx \frac{L_i}{\alpha_i R_i^{\max}(1-e_i)}, \tag{2}$$

$$E[X_i^2(\alpha_i)] \approx \frac{L_i^2(1+e_i)}{(\alpha_i R_i^{\max})^2 (1-e_i)^2}. \tag{3}$$

iii. Different video packets have different delay deadlines. Each packet in priority class $h (1 \leq h \leq H_i)$ is assumed to have the same maximum allowable delay, $delay_{i,h}^{\max}$.

iv. We assume that the queue waiting time dominates the overall delay (i.e., the transmission time is small [3], [4]).

2) Priority Queuing Analysis: Let $E[W_{i,h}]$ be the average waiting time of the packets in priority class h before they are transmitted. For a nonpreemptive priority M/G/1 queue, the Pollaczek–Khinchin equation gives the following result [13]:

$$E[W_{i,h}(\alpha_i)] = \frac{\sum_{k=1}^{H_i} r_{i,k} E[X_i^2(\alpha_i)]}{2\left(1 - \sum_{k=1}^{h-1} r_{i,k} E[X_i(\alpha_i)]\right)\left(1 - \sum_{k=1}^{h} r_{i,k} E[X_i(\alpha_i)]\right)}. \tag{4}$$

Based on this expected average waiting time, the tail distribution of the waiting time can be calculated by:

$$\text{Prob}(W_{i,h}(\alpha_i) > t) \approx \left(\sum_{k=1}^{H_i} r_{i,k} E[X_i(\alpha_i)]\right) \times \exp\left(-\frac{t \sum_{k=1}^{H_i} r_{i,k} E[X_i(\alpha_i)]}{E[W_{i,h}(\alpha_i)]}\right). \tag{5}$$

In (5), we adopt the G/G/1 tail distribution approximation based on the work of [14], [15]. Note that we assume that

$$\sum_{k=1}^{H_i} r_{i,k} E[X_i(\alpha_i)] < 1. \tag{6}$$

In Section III-C, we will describe how the average packet arrival rate can be controlled such that the above inequality holds and, meanwhile, the expected received video quality is maximized. The probability of packet loss due to the delay deadline expiration, $P_{i,h}(\alpha_i, \mathbf{r}_i)$, for priority class h can be computed based on the tail distribution of the waiting time:

$$P_{i,h}(\alpha_i, \mathbf{r}_i) = \text{Prob}\left(W_{i,h}(\alpha_i) > delay_{i,h}^{\max}\right), \tag{7}$$

where $\mathbf{r}_i = [r_{i,1} \cdots r_{i,H_i}]$ represents the admitted average packet arrival rate and $delay_{i,h}^{\max}$ is the maximum allowable delay for priority class h.

TABLE I
GREEDY ALGORITHM THAT INCREASES THE INPUT RATE GIVEN THE RESOURCE ALLOCATION α_i FOR WSTA i

Initialization $h \Leftarrow 0, r_{i,j} \Leftarrow 0\ (1 \le j \le H_i),\ Q_i^{rec,old} \Leftarrow 0$ and set the search step Δr.

Repeat

- $r_{i,h} \Leftarrow \min\{r_{i,h} + \Delta r, r_{i,h}^{\max}\}$, if $r_{i,h} \Leftarrow r_{i,h}^{\max}$, then $h \Leftarrow h+1$.
- Compute $Q_i^{rec,new} \Leftarrow \sum_{l=1}^{h} r_{i,l}\lambda_{i,l}\prod_{k=1}^{l-1}(1 - P_{i,k}(\alpha_i, \mathbf{r}_i))$ based on Eqs. (2)~(5) and Eq. (7).
- $\Delta Q \Leftarrow Q_i^{rec,new} - Q_i^{rec,old}$.
- $Q_i^{rec,old} \Leftarrow Q_i^{rec,new}$.

Until $\Delta Q \le 0$ or $\sum_{i=1}^{H_i} r_{i,h}E[X_i(\alpha_i)] \ge 1$.

$r_{i,h} \Leftarrow r_{i,h} - \Delta r, (1 \le h \le H_i)$.

Return $r_{i,h} (1 \le h \le H_i)$.

C. Expected Received Video Quality-Resource Function

Given a certain resource allocation α_i, reducing the average packet arrival rate $\mathbf{r}_i = [r_{i,1} \cdots r_{i,H_i}]$ leads to a decreased number of packets for transmission, but also a lower packet loss probability $P_{i,h}(\alpha_i, \mathbf{r}_i)$. This tradeoff can be formulated as an optimization problem as follows:

$$\begin{aligned} Q_i^{rec}(\alpha_i) &= \max_{\mathbf{r}_i} \sum_{h=1}^{H_i} r_{i,h}\lambda_{i,h}\prod_{k=1}^{h-1}(1 - P_{i,k}(\alpha_i, \mathbf{r}_i)) \\ s.t.\quad & 0 \le r_{i,h} \le r_{i,h}^{\max} \\ & \sum_{h=1}^{H_i} r_{i,h}E[X_i(\alpha_i)] < 1 \end{aligned} \tag{8}$$

where $r_{i,h}^{\max}$ is the maximum average packet arrival rate of priority class h, and $\prod_{k=1}^{h-1}(1 - P_{i,k}(\alpha_i, \mathbf{r}_i))$ represents the dependency of the current class h on classes $\{1, \ldots, h-1\}$. By solving the optimization problem, we can get the expected received video quality-resource function $Q_i^{rec}(\alpha_i)$. However, this is a difficult nonlinear problem. Due to the nonpreemptive service policy, the higher priority packets arriving into the queue have to wait until the service of the lower priority packet being served is finished. Hence, reducing the arrival rate of the lower priority class will decrease the packet loss probability of the higher priority class. Based on this observation, we propose a suboptimal algorithm, which greedily increases the packet arrival rate from the highest priority class to the lowest priority class. The algorithm is illustrated in Table I. Using this algorithm, the suboptimal $Q_i^{rec}(\alpha_i)$ for various resource allocations can be determined. The expected received video quality function $Q_i^{rec}(\alpha_i)$ is a concave and increasing function with the increased resource allocation α_i (see e.g., [16]).

IV. Mechanism Design for Resource Allocation

As mentioned in the introduction, the major challenge in devising methods for allocating and managing resources in multimedia applications comes from the informationally decentralized nature of these problems and the strategic behavior of the users. In this section we introduce a pricing mechanism, which solves the problem in (1) in the decentralized way, while taking into account the strategic behavior of individual users. In Section IV-A, we provide the motivation and describe some of the desired properties of mechanisms. In Section IV-B, we develop a pricing mechanism that solves the problem in (1) using Nash equilibrium messages and satisfies the desired properties presented in Section IV-A. In Section IV-C, we discuss the convergence issue of the message exchanges.

A. Mechanism Design

1) Motivation: In the resource allocation for multimedia applications, each user generally desires to acquire as much of the network resource as possible [1]. There are several reasons for this phenomenon: i) $Q_i^{rec}(\alpha_i)$ is assumed to always be improved by increasing the amount of resources allocated; ii) if given more resources, the users might be able to cope with sudden variations in channel conditions or source characteristics; and iii) users can lower their processing power usage via over-provisioning because it allows them to deploy less complex channel coding and protection schemes. Unfortunately, in multimedia applications, the information related to $Q_i^{rec}(\alpha_i)$, such as the video source, underlying channel conditions as well as transmission strategies, is private to each user. Also, as discussed in Section I, The informationally decentralized nature of the network and the strategic behavior of the users make the design of resource allocation mechanisms a challenging research problem.

2) Mechanism Components: In solving such informationally decentralized problems one needs to devise a message exchange process between the problem's agents (i.e., the users and the CSM), at the end of which all agents agree on a resource allocation which corresponds to an optimal solution to the problem. We call such a process of communication, decisions and actions a *resource allocation mechanism* [21]. In this paper, we investigate the design of mechanisms that are able to provide solutions to the problem in (1) and take into account the strategic behavior of individual users.

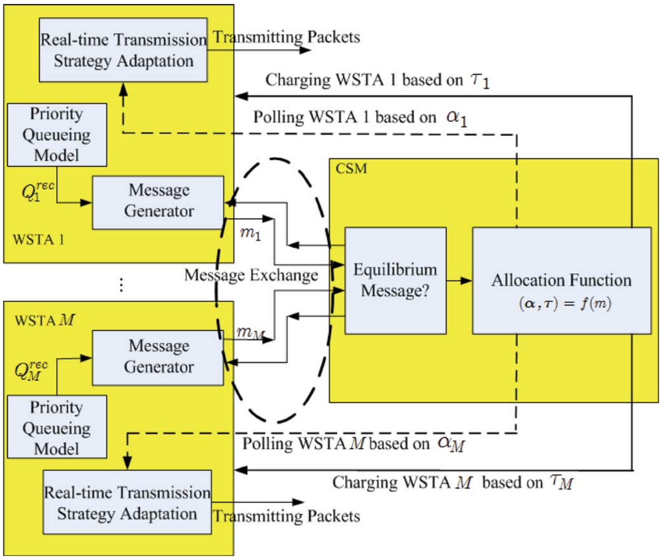

Fig. 1. Mechanism framework for the resource allocation in wireless multimedia applications.

Formally, a resource allocation mechanism is formed by a "game form" $(\boldsymbol{\mathcal{M}}; f)$ along with an "equilibrium message concept". $\boldsymbol{\mathcal{M}}$ is called the message space, and represents the set of all the allowable messages that can be transmitted in the system, while f, called the outcome function, represents a mapping from messages to allocations. In the context of the problem in (1), $\boldsymbol{\mathcal{M}} = \prod_{i=1}^{M} \boldsymbol{\mathcal{M}}_i$, $\boldsymbol{\mathcal{M}}_i$ represents the set of messages user i is allowed to transmit, and $f : \boldsymbol{\mathcal{M}} \mapsto ([0,1]^M, \mathbb{R}^M)$ maps messages to possible allocations $(\boldsymbol{\alpha}, \boldsymbol{\tau})$, with $\boldsymbol{\alpha} \in [0,1]^M$ and $\boldsymbol{\tau} \in \mathbb{R}^M$. The mechanism framework can be deployed for wireless multimedia resource allocation as shown in Fig. 1.

The equilibrium message concept plays a key role in the design of a mechanism. Given a set of specific users' utility functions (which are "private" to each user) and the designed game form $(\boldsymbol{\mathcal{M}}; f)$ (which is proposed by the mechanism designer), the equilibrium message concept will induce a map from the set of users' utility functions to that of equilibrium messages. The role of the outcome function is to map equilibrium messages to optimal resource allocations.

Before we proceed to describe some of the desired properties of mechanisms, we present two equilibrium concepts generally used in mechanism design: Nash equilibrium and dominant strategy equilibrium [17].

Nash equilibrium: A message profile $\boldsymbol{m} = (m_1, \ldots, m_M) \in \boldsymbol{\mathcal{M}}$ is said to be a *Nash equilibrium message* if for every user i,

$$U_i\left([f(m_i, \boldsymbol{m}_{-i})]_i\right) \geq U_i\left([f(m_i', \boldsymbol{m}_{-i})]_i\right), \quad \forall m_i', m_i \in \boldsymbol{\mathcal{M}}_i, \text{ and } m_i' \neq m_i, \tag{9}$$

where $\boldsymbol{m}_{-i} = (m_1, \ldots, m_{i-1}, m_{i+1}, \ldots, m_M)$ and $[f(m_i, \boldsymbol{m}_{-i})]_i$ corresponds to the allocation (α_i, τ_i) under the message profile $m = (m_i, \boldsymbol{m}_{-i})$. In words, every user i receives an allocation that maximizes its own utility when transmitting message m_i over any other possible message $m_i' \in \boldsymbol{\mathcal{M}}_i$, given that the messages of other users are fixed.

Dominant strategy equilibrium: A message profile $\boldsymbol{m} = (m_1, \ldots, m_M) \in \boldsymbol{\mathcal{M}}$ is said to be a *dominant strategy equilibrium message* if for every user i,

$$U_i\left([f(m_i, \boldsymbol{m}_{-i})]_i\right) \geq U_i\left([f(m_i', \boldsymbol{m}_{-i})]_i\right), \quad \forall m_i', m_i \in \mathcal{M}_i, m_i' \neq m_i, \text{ and } \forall \boldsymbol{m}_{-i} \in \prod_{j=1, j\neq i}^{M} \boldsymbol{\mathcal{M}}_j. \tag{10}$$

In other words, for every user i, a message m_i is a dominant strategy if given any other users profile, the user i cannot improve its utility for the allocation received by transmitting a message m_i' that is different from m_i. A dominant strategy equilibrium is a profile in which every user picks a dominant strategy. The dominant strategy equilibrium is a stronger concept than Nash equilibrium.

3) Mechanism Properties: In the resource allocation for bandwidth-intense multimedia applications, the available resource is quite limited and the involved users have the freedom to join or leave the resource allocation game. The outcomes of the mechanism for the multimedia applications should i) maximize the sum of the users' expected received video quality, ii) be budget balanced, and iii) satisfy the property of voluntary participation. Although these have been well established in the mathematical economics community [19], we would like to present them here for completeness as well as to establish the link with our multiuser wireless resource management problem. Thus, in this subsection, we provide some definitions for these properties.

Feasibility: An allocation vector $(\boldsymbol{\alpha}, \boldsymbol{\tau})$ is said to be *feasible* if it satisfies the constraints in the problem in (1), i.e.

$$(\boldsymbol{\alpha}, \boldsymbol{\tau}) \in \mathcal{S} \tag{11}$$

where $\mathcal{S} = \{(\boldsymbol{\alpha}, \boldsymbol{\tau}) | \sum_{i=1}^{M} \alpha_i \leq 1; 0 \leq \alpha_i \leq 1, \forall i \in \{1, \ldots, M\}; \sum_{i=1}^{M} \tau_i \leq 0\}$.

Voluntary participation: A feasible allocation vector $(\boldsymbol{\alpha}, \boldsymbol{\tau}) \in \mathcal{S}$ is said to satisfy the property of voluntary participation if

$$U_i(\alpha_i, \tau_i) \geq 0, \ \forall i \in \{1, \ldots, M\}. \tag{12}$$

The voluntary participation of $(\boldsymbol{\alpha}, \boldsymbol{\tau})$ can be interpreted as follows: after the resource allocation process, no user can be worse off than before in terms of the gained utility.

We now present several "efficiency" criteria which can be used as objectives for our wireless resource allocation mechanism.

Pareto efficiency: A feasible allocation vector $(\boldsymbol{\alpha}, \boldsymbol{\tau}) \in \mathcal{S}$ is said to be *Pareto efficient* if there exists no other allocation $(\boldsymbol{\alpha}', \boldsymbol{\tau}') \in \mathcal{S}$ such that

$$U_i\left(\alpha_i', \tau_i'\right) \geq U_i(\alpha_i, \tau_i), \ \forall i \in \{1, \ldots, M\} \tag{13}$$

and the inequality is strict for some i.

Utility-maximizing: A feasible allocation vector $(\boldsymbol{\alpha}, \boldsymbol{\tau}) \in \mathcal{S}$ is said to be *utility-maximizing* if it satisfies

$$\sum_{i=1}^{M} U_i(\alpha_i, \tau_i) \geq \sum_{i=1}^{M} U_i\left(\alpha_i', \tau_i'\right), \ \forall (\boldsymbol{\alpha}', \boldsymbol{\tau}') \in \mathcal{S}. \tag{14}$$

Quality-maximizing: A feasible allocation vector $(\boldsymbol{\alpha}, \boldsymbol{\tau}) \in \mathcal{S}$ is said to be *quality-maximizing* if it satisfies

$$\sum_{i=1}^{M} Q_i^{rec}(\alpha_i) \geq \sum_{i=1}^{M} Q_i^{rec}(\alpha_i'), \ \forall (\boldsymbol{\alpha}', \boldsymbol{\tau}') \in \mathcal{S}. \tag{15}$$

It is easy to show that utility-maximizing allocations imply Pareto efficient allocations. If the utility functions are quasi-linear and concave as in our problem, the Pareto efficient allocations are also utility-maximizing. To address the relationship between utility-maximizing and quality-maximizing, we first define the budget balance for the allocation.

Budget balance: A feasible allocation vector $(\boldsymbol{\alpha}, \boldsymbol{\tau}) \in \mathcal{S}$ is said to be *budget balanced* if it satisfies

$$\sum_{i=1}^{M} \tau_i = 0. \tag{16}$$

The budget balance condition has the following interpretation: the amount of numeraire commodity in the system prior to the resource allocation process equals to the amount of numeraire commodity in the system after the resource allocation process. In other words, the numeraire commodity held by the multimedia users are redistributed among them, and none of it is "thrown" away or taken away by the CSM.

The relationship between utility-maximizing and quality-maximizing is addressed in the following lemma:

Lemma 1: In the case of quasi-linear utilities, an allocation is utility-maximizing if and only if it is quality-maximizing and budget balance.

Sketch of proof: Given the quasi-linear form of the utilities, the social welfare can be rewritten as $\sum_{i=1}^{M} U_i(\alpha_i, \tau_i) = \sum_{i=1}^{M} Q_i^{rec}(\alpha_i) + \sum_{i=1}^{M} \tau_i$. Since $\sum_{i=1}^{M} \tau_i \leq 0$, we have $\sum_{i=1}^{M} U_i(\alpha_i, \tau_i) \leq \sum_{i=1}^{M} Q_i^{rec}(\alpha_i)$. If the allocation $(\boldsymbol{\alpha}^*, \boldsymbol{\tau}^*)$ is quality-maximizing and budget balanced, then we have that $\sum_{i=1}^{M} U_i(\alpha_i^*, \tau_i^*) = \sum_{i=1}^{M} Q_i^{rec}(\alpha_i^*)$ and hence, the allocation is also utility-maximizing. On the other hand, if the allocation $(\boldsymbol{\alpha}^*, \boldsymbol{\tau}^*)$ is utility-maximizing, and the budget is not balanced, i.e., $\sum_{i=1}^{M} \tau_i^* < 0$, we define a new tax $\tau_i' = \tau_i^* - (1/M)\sum_{i=1}^{M} \tau_i^*, \forall i$. Note that the new allocation $(\boldsymbol{\alpha}^*, \boldsymbol{\tau}')$ is also feasible since $\boldsymbol{\alpha}^*$ satisfies the first two constraints of the problem stated in (1) and $\boldsymbol{\tau}'$ satisfies the last constraint in the same problem. Since $\tau_i' \geq \tau_i$, we also have that $\sum_{i=1}^{M} U_i(\alpha_i^*, \tau_i^*) < \sum_{i=1}^{M} U_i(\alpha_i^*, \tau_i')$. This contradicts the optimality assumption of $(\boldsymbol{\alpha}^*, \boldsymbol{\tau}^*)$. Thus, the allocation $(\boldsymbol{\alpha}^*, \boldsymbol{\tau}^*)$ is budget balanced. Since the allocation is budget balanced, we have that $\sum_{i=1}^{M} U_i(\alpha_i^*, \tau_i^*) = \sum_{i=1}^{M} Q_i^{rec}(\alpha_i^*)$ which automatically implies that the allocation is also quality-maximizing. The reader is referred to [20] for more details about this proof.

In the context of our problem, quality-maximizing allocations are the ones which maximize the sum of expected received video quality over all the users, i.e., producing the maximum amount of video quality. However, quality-maximizing allocations are not concerned with how the numeraire commodity is being allocated, and particularly if any of the numerarie from the system has been thrown away or allocated outside of the system (i.e., given to the CSM). Utility-maximizing allocations do not only maximize the obtained video quality, but they also ensure that the numeraire commodity is not "wasted" during the allocation process.

B. Pricing Mechanism Implemented in Nash Equilibrium

In this subsection, we present a pricing mechanism that implements the problem in (1) using Nash equilibrium messages, and generates allocations that are feasible, individually rational (i.e., satisfying the property of voluntary participation) and utility-maximizing.

1) Implementation of the Pricing Mechanism: The pricing mechanism [20] is implemented in the considered multiuser resource allocation in four stages:

Stage 1: The endowment stage: At the first stage of the mechanism, each user i is endowed a certain amount of resource α_i^0, where $\sum_{i=1}^{M} \alpha_i^0 = 1$.

This initial user endowment will not affect the final resource allocation received by each user. Instead, it will affect the amount of numeraire commodity that each user will be charged in order to attain his/her optimal resource. The initial endowment is determined before playing the resource allocation game. If no information is known about the users' utilities, this initial endowment can be equal for all users (i.e., $\alpha_i^0 = 1/M, \forall i \in \{1, \ldots, M\}$).

Stage 2: Communication stage (information exchange): We note that at the mechanism's first stage, some users will have been endowed a resource larger than the optimal resource, while others will have been endowed a resource that is too small. At this stage of the mechanism all users are allowed to *trade* resources in order to achieve the desirable amount.

The users and the CSM are allowed to communicate with each other by repeating the following two steps until the message equilibrium is reached:

Step 1) Each user i submits to the CSM a message (α_i, p_i), where α_i represents the amount of resource desired, and p_i represents the user's evaluation of the "price" per unit of resource.

Step 2) After receiving the messages from all the users, the CSM conveys the following messages (p_{-i}, d_i, γ) to each user i, where

$$p_{-i} = \frac{1}{M-1} \sum_{j=1, j\neq i}^{M} p_j \tag{17}$$

represents the average of the other users price per unit of resource, and

$$d_i = \sum_{j=1, j\neq i}^{M} \alpha_j - 1 \tag{18}$$

is the excess demand when the ith user's demand is eliminated, and γ is a positive number interpreted as the update rate of the price that is enforced by the CSM when the excess demand is nonzero. This will be further discussed in Section IV-C.

To choose the demand α_i and the price p_i, the user i should be able to know the tax form which is used by the CSM to charge the users. The tax form[2] in this paper is defined as

$$t_i(\boldsymbol{\alpha},\boldsymbol{p}) = \left(\alpha_i - \alpha_i^0\right) \times p_{-i} + \left[p_i - p_{-i}\left(1 + \frac{\sum_{j=1}^{M}\alpha_j - 1}{\gamma}\right) - \chi_+(\boldsymbol{\alpha},\gamma)\right]^2 \tag{19}$$

where

$$\chi_+(\boldsymbol{\alpha},\gamma) = \max\left\{0, \frac{\sum_{j=1}^{M}\alpha_j - 1}{\gamma}\right\}. \tag{20}$$

In (19), the first term on the right hand side is the amount of numeraire commodity charged in order to purchase/sell $(\alpha_i - \alpha_i^0)$ amount of resource from/to the other users. The second term is the penalty the user i should pay due to the mismatch of the user's price to p_{-i}. $\chi_+(\cdot)$ is introduced here to prevent the following set of messages, which generate nonefficient allocations from becoming Nash equilibrium messages: i) $p_i = 0$ for all the users and ii) the total demand exceeds the available resource, i.e., $\sum_{i=1}^{M}\alpha_i - 1 > 0$.

At each iteration, given the message (p_{-i}, d_i, γ), the user i chooses (α_i, p_i) by maximizing its own utility $U_i(\alpha_i, \tau_i)$, i.e.

$$\begin{aligned}\max_{\alpha_i,p_i} U_i(\alpha_i,\tau_i) &= \max_{\alpha_i,p_i}\left\{Q_i^{rec}(\alpha_i) + \tau_i\right\}\\ &= \max_{\alpha_i,p_i}\Big\{Q_i^{rec}(\alpha_i) - \left(\alpha_i - \alpha_i^0\right) \times p_{-i}\\ &\qquad -\left[p_i - p_{-i}\left(1 + \frac{d_i+\alpha_i}{\gamma}\right) - \chi_+(d_i,\alpha_i,\gamma)\right]^2\Big\}.\end{aligned} \tag{21}$$

Note that $\tau_i = -t_i$, $\sum_{j=1}^{M}\alpha_j - 1 = d_i + \alpha_i$ and $\chi_+(d_i,\alpha_i,\gamma) = \max\{0, (d_i+\alpha_i)/\gamma\}$. The optimization in (21). can be decomposed into two subproblems by solving α_i and p_i independently:

$$\max_{\alpha_i}\left\{Q_i^{rec}(\alpha_i) - \left(\alpha_i - \alpha_i^0\right) \times p_{-i}\right\} \tag{22}$$

and

$$p_i = p_{-i}\left(1 + \frac{d_i+\alpha_i}{\gamma}\right) + \chi_+(d_i,\alpha_i,\gamma). \tag{23}$$

Note that p_{-i} is determined by the CSM based on the revealed messages in the previous iteration.

Stage 3: Allocation stage: Given an equilibrium set of messages $(\boldsymbol{\alpha}^*, \boldsymbol{p}^*)$, where $\alpha^* = [\alpha_1^*, \ldots, \alpha_M^*]$ and $\boldsymbol{p}^* = [p_1^*, \ldots, p_M^*]$, each user i is allocated the amount of resource α_i^* and is taxed as follows:

$$t_i(\boldsymbol{\alpha}^*,\boldsymbol{p}^*) = \left(\alpha_i^* - \alpha_i^0\right) \times p_{-i}^* + \left[p_i^* - p_{-i}^*\left(1 + \frac{\sum_{j=1}^{M}\alpha_j^* - 1}{\gamma}\right) - \chi_+(\boldsymbol{\alpha}^*,\gamma)\right]^2. \tag{24}$$

[2]The tax t_i computed by the CSM equals $-\tau_i$.

Stage 4: Real-time video transmission: After the resource allocation is performed, the CSM polls the multimedia users based on the allocated resource. When they are polled, the multimedia users will deploy their real-time transmission strategies to stream the delay-sensitive video packets. The real-time streaming will be discussed in Section V.

The following theorem proves that even in the case when the users behave strategically, the allocations generated by the mechanism presented above are utility-maximizing.

Theorem 1: The mechanism implemented in stages 1–4 generates the allocation in Nash equilibrium for the multiuser wireless video resource allocation problem in (1).

Proof: See Appendix I.

Lemma 2: The mechanism implemented in stages 1–4 satisfies voluntary participation.

Proof: Given the message (p_{-i}, d_i, γ), user i maximizes its own utility. To show that the allocation satisfies voluntary participation, we only need to show the maximum utility for user i is not less than 0.

$$\begin{aligned}\max_{\alpha_i,p_i} U_i(\alpha_i,\tau_i) &= \max_{\alpha_i,p_i}\Big\{Q_i^{rec}(\alpha_i) - \left(\alpha_i - \alpha_i^0\right) \times p_{-i}\\ &\qquad - \left[p_i - p_{-i}\left(1 + \frac{d_i+\alpha_i}{\gamma}\right) - \chi_+(d_i,\alpha_i,\gamma)\right]^2\Big\}\\ &\geq \max_{p_i}\Big\{Q_i^{rec}(0) - \left(-\alpha_i^0\right) \times p_{-i}\\ &\qquad - \left[p_i - p_{-i}\left(1 + \frac{d_i}{\gamma}\right) - \chi_+(d_i,0,\gamma)\right]^2\Big\}\\ &= \underbrace{Q_i^{rec}(0)}_{=0} - \underbrace{\left(-\alpha_i^0\right) \times p_{-i}}_{\leq 0}\\ &\qquad \underbrace{- \min_{p_i}\left\{\left[p_i - p_{-i}\left(1 + \frac{d_i}{\gamma}\right) - \chi_+(d_i,0,\gamma)\right]^2\right\}}_{=0}\\ &\geq 0.\end{aligned} \tag{25}$$

Since the above inequalities hold for any given messages p_{-i}, d_i, γ, the allocations generated by our mechanism satisfy the voluntary participation property.

2) Comparison Between the Pricing Mechanism and VCG Mechanism: We would like to compare now the presented pricing mechanism with another well-known game-theoretic mechanism—the VCG mechanism [17], which implements the resource allocation in dominant strategy equilibrium (rather than Nash equilibrium) as defined in Section IV-A2. The VCG mechanism is based on the "revelation principle", which in our multimedia applications has the following interpretation [18]: the CSM requests each user to "reveal" its utility function. For a profile of utility functions, the CSM computes the optimal allocations and the individual VCG tax for each user.

Next, we compare the VCG mechanism and the proposed pricing mechanism for the studied wireless video resource allocation.

i) Under the VCG mechanism, only a single message exchange between the users and the CSM is needed. Alternatively, in our pricing mechanism, an iterative message exchange between the users and the CSM is required in order to determine the allocation.

ii) In the VCG mechanism, however, the users are required to compute the expected received video quality over all the possible resource allocation and convey the entire quality-resource allocation profile to the CSM. Although the expected received video quality can be parameterized as shown in [18], the number of parameters is still large. Thus, transmitting them to the CSM will require a large overhead. Moreover, the WSTAs might not want to declare their entire quality-resource allocation profile to the CSM due to privacy issues. Alternatively, in our pricing mechanism, the users are only required to reveal their resource demands and corresponding prices, leading to a very limited overhead (as quantified in Section IV-D).

iii) In the VCG mechanism, the optimization for the resource allocation problem is computed by the CSM, thereby resulting in a high complexity cost for the wireless infrastructure provider. In our pricing mechanism, the CSM is only required to compute the average price and excess demand for each user.

iv) In the VCG mechanism, the allocation is feasible, quality-maximizing and satisfies the voluntary participation property, but it is not budget balanced. In our pricing mechanism, the resource allocation is feasible, utility-maximizing (implying quality-maximizing), budget balanced and also satisfies the voluntary participation property. Hence, the VCG mechanism can be deployed to perform the resource allocation in the quality-maximizing sense, but not in the utility-maximizing sense. In the investigated wireless resource allocation problem, quality-maximizing allocations that are not utility-maximizing are undesirable because the wireless coordinator (CSM) aims solely at assisting the resource allocation and not at making any profit.

C. Discussion of Convergence

In the previous section we have presented a pricing mechanism that generates optimal allocations to problem (1). In describing the pricing mechanism, we presented an iterative procedure that converges to a set of Nash equilibrium messages. The key assumption in this procedure is that each WSTA will communicate its best response messages at each stage (i.e., for a set of messages received from the CSM, each WSTA communicates the messages that maximize its utility function). However, the convergence of the presented pricing mechanism depends on the parameter γ, which appears in the tax function in (19). The parameter γ can be interpreted as the step size in the price update procedure. The value of γ is inversely proportional to the amount by which the prices of the users are being updated. While for large values of γ the pricing mechanism converges slowly, if γ is too small, the resource negotiation process may lead to oscillations and it will never converge. To understand this phenomenon, we need to consider the change rate of the subgradients[3] [16] of the $Q_i^{rec}(\alpha_i)$ around the optimal allocation: if the change rate is large, then a limited change in the user demand may generate a large change in its price. Since the $Q_i^{rec}(\alpha_i)$ is nondecreasing and concave, the change rate of the subgradients decreases with the amount of allocated resources. Hence, when a limited amount of resources is allocated to the users, the parameter γ should be large to ensure that the resource exchange procedure convergences.

[3]The subgradients are used because the $Q_i^{rec}(\alpha_i)$ may not be differentiable at some points.

The convergence rate of the above procedure is affected not only by the parameter γ, but also by the initial price p_{-i} announced by the CSM and the network congestion level (e.g., the number of users in the network). Even though the pricing mechanism will converge independently of the initial price p_{-i}, this will affect the number of iterations required before the equilibrium is reached. Moreover, the congestion level also affects the value of the parameter γ required to ensure convergence.

In the results section, we will assess how various settings of the parameters γ and p_{-i} impact the convergence rate under different transmission scenarios.

D. Message Exchange Overhead

In this section, we quantify the overhead associated with the message exchanges described in Section IV-B. At each iteration, each WSTA transmits to the CSM two messages: a resource demand α_i and a corresponding price p_i. Subsequently, the CSM responds to the WSTA with two messages: the average price p_{-i} announced by the other WSTAs and the resources d_i consumed by the other WSTAs. The update step size of the price γ, is kept fixed during the iteration and hence, it can be announced once by the CSM at the first iteration. Thus, in addition to the update step size γ, only four scalar numbers are exchanged between WSTA i and the CSM. This scalar message can be encapsulated into the control packets that are transmitted between WSTA i and the CSM, as in the current 802.11e WLAN standards [2], [3]. As discussed in Section II, the message exchange is only performed when large variations occur in the network, e.g., new WSTAs joining the network. In Section VI-C, we quantify the negotiation overhead under different wireless transmission scenarios by determining the number of iterations required to converge to the optimal allocation. Note that while the wireless resources are renegotiated by the WSTAs, the WSTAs that are already present in the network are continuing to transmit their video packets using the previously negotiated transmission opportunities. Hence, the reallocation will not delay the wireless video transmission already taking place. The WSTAs will only start transmitting their packets based on the newly allocated transmission opportunities *after* the new resource allocation is finalized.

V. Real-Time Transmission Strategy

As discussed in Section II, user i can strategically play the resource allocation game using message exchanges based on the pricing mechanism illustrated in Section IV. Given the resource allocation α_i, user i adapts its real-time transmission strategies to the time-varying content characteristics and channel conditions as shown in Fig. 1. Various existing real-time transmission strategies can be employed for this purpose. In [26], it was shown that optimized transmission strategies involve joint adaptation across the various layers of the protocol stack. Also, as shown in [22], the packet scheduling has a significant impact on the performance of delay-sensitive multimedia applications.

Hence, in this section, we mainly focus on the real-time packet scheduling that can be deployed at the application layer for optimized transmission. For the transmission strategies at the lower layers (MAC and physical layers), such as adaptive retransmission and modulation and coding mode selection, the interested reader is referred to our prior work in [4], [26].

The scheduling policy decides the packet transmission order (which packet should be transmitted) and time (when the packet should be transmitted) based on the estimated channel condition. In other words, the optimal scheduling policy is to choose a subset of available packets to transmit at the successive transmission opportunities such that the received video quality is maximized. It has been proved that the scheduling problem is in general NP-complete [23] and no optimal algorithms for this exist in polynomial time. Several heuristic algorithms providing suboptimal solutions have been developed [24], [25]. In [24], the authors introduced a virtual playback delay instead of the actual one to compute the subset of packets to be sent by trading-off the number of transmitted packets and the packet loss probability. In [25], the packet interdependencies are expressed as a directed acyclic graph and an iterative descent algorithm is developed to find a suboptimal scheduling by trading off the expected transmission rate and expected reconstruction distortion. However, in our proposed system, we deploy a simpler real-time scheduling policy based on the video packet priority classes.

The scheduling policy will be optimized every (group of) service interval(s) (SI) [2] to take into consideration the information (e.g., the channel conditions and the available packets in the transmission buffer) learned or measured during the previous transmission opportunities. Hence, the scheduling policy will be able to capture the time-variation of both the video traffic and the channel conditions. In the following, we refer to the real-time scheduling policy as $\pi_i \in \Pi_i$, where Π_i represents the set of possible scheduling policies.

We assume that the packets from G GOPs are present in the transmission buffer. The current time is assumed to be 0. Within GOP $g (1 \leq g \leq G)$, the number of available packets in class h equals $N_{i,g,h}$. The delay deadline[4] for all packets (after which the packets will be expired) in class h in GOP g is denoted as $d_{i,g,h}$. (Note that the delay deadline is different from the maximum allowable delay as discussed in Section III-B.) The maximum allowable delay is the maximum waiting time before the packet is available for transmission. The scheduling policy π_i assigns the transmission opportunities to the different classes of packets in the considered GOPs. Let $N_{i,g,h}^{TX} (0 \leq N_{i,g,h}^{TX} \leq N_{i,g,h})$ be the number of packets in class h in GOP g that are assigned transmission opportunities before their delay deadline expires. The scheduling policy π_i can be interpreted as choosing $N_{i,g,h}^{TX}$ for all the classes, i.e., $\pi_i = \{N_{i,g,h}^{TX}\}_{1 \leq h \leq H_i, 1 \leq g \leq G}$. Since the resource allocation is already known during the real-time scheduling, the expected received video quality is now expressed as a function of the scheduling policy π_i, i.e., $Q_i^{rec}(\pi_i)$, instead of the allocated resource α_i. Hence, $Q_i^{rec}(\pi_i)$ can be expressed as

$$Q_i^{rec}(\pi_i) = \sum_{g=1}^{G} \sum_{h=1}^{H_i} N_{i,g,h}^{TX} \lambda_{i,h}. \tag{26}$$

[4]Note that this deadline is relative to the current time.

Given the resource allocation α_i and the channel conditions, the transmitted rate $R_i^{\max}$ and the packet error probability e_i can be computed as in [12] by assuming that the channel conditions are constant during the current scheduling period (e.g., one SI). Then, the average service time $E[X_i]$ for repeatedly transmitting one packet until it is successfully received is again computed as in (2). Based on the expected transmission time for each packet, the optimal scheduling policy π_i^{opt} can be found by solving the following optimization under the delay constraints:

$$\begin{aligned} \pi_i^{opt} &= \arg\max_{\pi_i \in \Pi_i} Q_i^{rec}(\pi_i) \\ s.t. \quad & \sum_{k=1}^{g-1} \sum_{l=1}^{H_i} N_{i,k,l}^{TX} E[X_i] + \sum_{l=1}^{h-1} N_{i,g,l}^{TX} E[X_i] + N_{i,g,h}^{TX} E[X_i] \\ & \leq d_{i,g,h}, \\ & 1 \leq N_{i,g,h}^{TX} \leq N_{i,g,h}, \\ & \text{for } 1 \leq h \leq H_i,\ 1 \leq g \leq G \end{aligned} \tag{27}$$

where the first constraint comes from the fact that any packets assigned the transmission opportunities should not be expired. For the prioritized video packets available in the WSTA's transmission buffer, we can develop a greedy but optimal algorithm to the above optimization problem.

To specify the algorithm, let us consider the transmission opportunities assignment to class h in GOP g, i.e., determining $N_{i,g,h}^{TX}$. The amount of transmission opportunities which could be assigned to class h in GOP g is determined by the amount of transmission opportunities assigned to classes $\{1, \ldots, h-1\}$ of all the GOPs and the classes with priority h of GOPs $\{1, \ldots, g-1\}$ which are already computed prior to class h in GOP g, i.e., $\{N_{i,k,l}^{TX}\}_{1 \leq l < h, 1 \leq k \leq G}$ and $\{N_{i,k,h}^{TX}\}_{1 \leq k < g}$ are known.

The average starting time, $t_{i,g,h}^{start}$, to transmit the packets in class h in GOP g equals the sum of the transmission opportunities assigned to the classes transmitted prior to class h in GOP g, i.e.

$$t_{i,g,h}^{start} = \sum_{k=1}^{g-1} \sum_{l=1}^{H_i} N_{i,k,l}^{TX} E[X_i] + \sum_{l=1}^{h-1} N_{i,g,l}^{TX} E[X_i]. \tag{28}$$

The ending time, $t_{i,g,h}^{end}$ at which we stop transmitting the packets in class h in GOP g is the minimum value of the delay deadline, $d_{i,g,h}$, of class h in GOP g and the time that will not lead to the delay violation of the classes $\{1, \ldots, h-1\}$ in the GOPs $\{g+1, \ldots, G\}$, i.e.

$$\begin{aligned} t_{i,g,h}^{end} = \min\Bigg\{ d_{i,g,h}, \max\Bigg\{ t \,\Big|\, t + \sum_{j=g+1}^{k-1} \sum_{m=1}^{h-1} N_{i,j,m}^{TX} E[X_i] \\ + \sum_{m=1}^{l} N_{i,k,m}^{TX} E[X_i] \leq d_{i,k,l}, \\ \forall 1 \leq l \leq h-1, g < k \leq G \Bigg\}\Bigg\}. \end{aligned} \tag{29}$$

The "max" term above is interpreted as the maximum time to start transmitting the packets in class 1 in the GOP $g+1$ such that the packets already assigned transmission opportunities in classes $\{1, \ldots, h-1\}$ in GOPs $\{g+1, \ldots, G\}$ are not expired.

TABLE II
RESOURCE ALLOCATIONS AND VIDEO QUALITIES OF THE ADMISSION CONTROL-BASED RESOURCE ALLOCATION APPROACH WHEN WSTAS TRUTHFULLY DECLARE THEIR RESOURCE REQUIREMENTS OR NOT

	WSTAs truthfully declaring their requirements		WSTA 1 exaggerating but other WSTAs truthfully declaring their requirements	
	Allocated resource a_i	Video quality[8] (dB)	Allocated resource a_i	Video quality (dB)
WSTA 1	0.2	30.033	0.333	33.765
WSTA 2	0.2	35.430	0.176	34.214
WSTA 3	0.2	35.448	0.176	34.210
WSTA 4	0.2	35.464	0.176	34.180
WSTA 5	0.2	35.458	0.176	34.197

[1] We use "video quality" to represent the "expected received video quality" for simplicity in the simulation results.

TABLE III
RESOURCE ALLOCATIONS, UTILITIES AND VIDEO QUALITIES OF OUR PROPOSED PRICING MECHANISM WHEN WSTAS STRATEGICALLY PLAY THE GAME OR NOT

	WSTAs strategically playing the game			WSTA 1 lying about its quality model but others strategically playing the game		
	Allocated resource a_i	Utility U_i	Video quality (dB)	Allocated resource a_i	Utility U_i	Video quality (dB)
WSTA 1	0.1740	29.5179	28.8423	0.4000	25.6559	33.1156
WSTA 2	0.2065	34.8665	35.0373	0.1500	35.1356	33.2727
WSTA 3	0.2065	34.8665	35.0373	0.1500	35.1356	33.2727
WSTA 4	0.2065	34.8665	35.0373	0.1500	35.1356	33.2727
WSTA 5	0.2065	34.8665	35.0373	0.1500	35.1356	33.2727

The number of packets, $N_{i,g,h}^{TX}$, in class h in GOP g can now be computed as

$$N_{i,g,h}^{TX} = \min\left\{N_{i,g,h}, \left\lfloor \frac{t_{i,g,h}^{end} - t_{i,g,h}^{start}}{E[X_i]} \right\rfloor\right\}. \qquad (30)$$

By determining $N_{i,g,h}^{TX}$ for the classes from the highest priority to the lowest priority as above, we can automatically obtain the optimal scheduling policy to the optimization problem in (27).

VI. SIMULATION RESULTS

To be able to efficiently stream video over wireless networks, each WSTA needs to be able to cope with instantaneous bandwidth variations due to time-varying channel conditions and network congestion due to many competing WSTAs. In order to adapt to the time-varying availability of resources, the video bitstreams need to be compressed in a prioritized scalable manner. In this paper, we used data partitioning [33] for H.264/AVC video coders [40] and spatio–temporal-SNR layering for 3-D wavelet video coders [9].

A. Assessing the Impact of Selfish Behavior of WSTAs

In the introduction section, we mention that existing WLAN resource allocation approaches assume that all WSTAs truthfully reveal their own resource requirements. This is not true when the WSTAs are selfish. In the following experiments, we will assess the impact of the selfish behavior of WSTAs on the resource allocation, the expected received video qualities and utilities of WSTAs.

We consider five WSTAs concurrently streaming video sequences over a one-hop 802.11a/e WLAN test-bed [12]. The video sequences streamed by the five WSTAs are "Foreman" at CIF resolution 30 Hz. The first video sequence is encoded using a 3-D wavelet codec [9] and the last four sequences are encoded with the H.264/AVC codec [40]. The channel conditions experienced by the five WSTAs are assumed to be similar, having an average SNR of 23 dB and a variation across the duration of the transmission of around 5 dB. We compare our proposed pricing mechanism to the resource allocation approach existing in existing 802.11e WLANs [3], [4].

We consider two cases: i) the WSTAs are assumed to truthfully submit their TSPECs [4]; ii) WSTA 1 exaggerates its requirement by 100% but other WSTAs truthfully declare their requirements. Table II shows the resource allocations and video qualities for the current 802.11e allocation in the two scenarios.

To evaluate our pricing mechanism, we also simulate two cases: a) the WSTAs strategically submit their resource demands and corresponding prices by maximizing their utilities at each iteration; b) WSTA 1 submits a fixed but higher resource demand at each iteration and other WSTAs behave as in case a). Table III shows the resource allocations, utilities and video qualities for our pricing mechanism.

From Table II, we note that, by exaggerating its requirement, WSTA 1 increases its final expected received video quality around 4 dB, while decreasing the other WSTA's performance around 1 dB. In this case, the improvement of WSTA 1's performance does not incur any penalty for WSTA 1. However, from Table III, we note that, although WSTA 1 increases its expected received video quality more than 4 dB, the final utility of WSTA 1 is actually reduced from 29.5179 to 25.6559. This is because WSTA 1 does not submit the optimal resource demand (i.e., it is unaware of the price) at each iteration and hence, it has to pay a much higher tax than when it strategically responds. In other words, our pricing mechanism enforces all the WSTAs to truthfully declare their optimal resource demands and corresponding prices at each iteration. The pricing mechanism also penalizes the selfish WSTAs that exaggerate their resource demand by imposing higher taxes. Based on these experiments,

TABLE IV
RESOURCE ALLOCATIONS AND VIDEO QUALITY OF THE CENTRALIZED OPTIMIZATION, AND THE RESOURCE ALLOCATIONS, PRICES, TAXES, UTILITIES AND VIDEO QUALITIES OF OUR PRICING MECHANISM FOR ALL THE WSTAS IN SCENARIO 1

	Centralized		Decentralized				
	Allocated resource a_i	Video quality (dB)	Allocated resource a_i	Price p_{-i}	Tax t_i	Utility U_i	Video quality (dB)
WSTA 1	0.2341	34.0236	0.2342	24.9203	0.8514	33.1726	34.0231
WSTA 2	0.2341	34.0236	0.2342	24.9203	0.8514	33.1726	34.0231
WSTA 3	0.1302	29.4093	0.1302	24.9203	-1.7403	31.1495	29.4085
WSTA 4	0.1302	29.4093	0.1302	24.9203	-1.7403	31.1495	29.4085
WSTA 5	0.2714	27.4593	0.2712	24.9203	1.7779	25.6808	27.4576
Summation	1.0000	154.3251	1.0000	/	0	154.3250	154.3207

TABLE V
RESOURCE ALLOCATIONS AND VIDEO QUALITY OF THE CENTRALIZED OPTIMIZATION, AND THE RESOURCE ALLOCATIONS, PRICES, TAXES, UTILITIES AND VIDEO QUALITIES OF OUR PRICING MECHANISM FOR ALL THE WSTAS IN SCENARIO 2

	Centralized		Decentralized				
	Allocated resource a_i	Video quality (dB)	Allocated resource a_i	Price p_{-i}	Tax t_i	Utility U_i	Video quality (dB)
WSTA 1	0.1282	30.7099	0.1269	40.1439	0.0769	30.5826	30.6588
WSTA 2	0.1270	30.6630	0.1269	40.1439	0.0769	30.5826	30.6588
WSTA 3	0.1274	30.6804	0.1269	40.1439	0.0769	30.5826	30.6588
WSTA 4	0.0560	27.1004	0.0571	40.1439	-2.7259	29.8711	27.1448
WSTA 5	0.0560	27.1005	0.0571	40.1439	-2.7259	29.8711	27.1448
WSTA 6	0.0575	27.1610	0.0571	40.1439	-2.7259	29.8711	27.1448
WSTA 7	0.2240	26.1817	0.2240	40.1439	3.9736	22.2090	26.1816
WSTA 8	0.2240	26.1815	0.2240	40.1439	3.9736	22.2090	26.1816
Summation	1.0000	225.7783	1.0000	/	0	225.7791	225.7743

we would like to emphasize that the used coder does affect the video quality performance experienced by the wireless stations, but not the proposed mechanism for the resource allocation.

B. Verifying the Properties of the Pricing Mechanism

In this subsection, we verify the properties of the pricing mechanism proposed in Section IV (i.e., the allocations produced i) are utility-maximizing, and ii) satisfy voluntary participation), and we explore the behavior of the WSTAs in different wireless transmission scenarios. To show that the allocation is utility-maximizing, we need to show that it is feasible, quality-maximizing and budget balanced. The feasibility of the resource allocation can be easily checked, i.e., $\sum_{i=1}^{M} \alpha_i \leq 1, \alpha_i \geq 0, \forall i = 1, \ldots, M$. We can compare the resource allocation produced by our pricing mechanism (denoted as "decentralized" in the following simulations) to the allocation produced by the centralized optimization (denoted as "centralized") to check if the allocations are quality maximizing. Since the expected received video quality is a concave function, the centralized optimization will generate the global optimal resource allocation. If our pricing mechanism generates the same resource allocation, then it is quality-maximizing. To verify the budget balance condition, we need to check whether the summation of the tax of all the WSTAs is zero. Finally, if all the values of the utility functions of all the WSTAs are non-negative at the equilibrium, we conclude that the WSTAs voluntarily participate our pricing mechanism. Tables IV and V corresponding to Scenarios 1~2, respectively, show the resource allocations and the expected received video quality based on the centralized optimization[5] as well as the resource allocations, the corresponding prices, tax, utilities, and expected received video quality of our pricing mechanism at the equilibrium message. In the remaining experiments, the video sequences are encoded using a 3-D wavelet codec [9].

Scenario 1: To assess the properties of the pricing mechanism, we consider five WSTAs concurrently streaming video sequences. The video sequences streamed by the five WSTAs are: "Foreman," "Foreman," "Coastguard," "Coastguard," and "Mobile," respectively, at CIF resolution 30 Hz. All the video applications are considered to tolerate a maximum delay of 533 ms [39] corresponding to the duration of one GOP. The channel conditions experienced by the five WSTAs are assumed to be similar, having an average SNR of 23 dB and a variation across the duration of the transmission of around 5 dB. The initial resource allocations are assumed to be the same, i.e., $\alpha_i^0 = 0.2, \forall i = 1, \ldots, 5$. By observing the resources allocated by the proposed pricing mechanism in Table IV, we note that they are feasible, i.e., $\sum_{i=1}^{M} \alpha_i \leq 1, \alpha_i \geq 0, \forall i = 1, \ldots, 5$. By comparing the resource allocations and the sum of expected received video qualities generated by the centralized optimization and

[5]The optimization problem in (1) is also solved in the centralized way by using the convex optimization.

our pricing mechanism, we find that the results are very similar.[6] This verifies that our pricing mechanism is quality-maximizing. The fact that the sum of the taxes of all the WSTAs equals zero shows that our allocations are budget balanced. We also note that, in equilibrium, the values of the utility functions of all the WSTAs are greater than zero. This demonstrates that our pricing mechanism satisfies the property of voluntary participation.

Scenario 2: In this scenario we verify the properties of our pricing mechanism, and we investigate how the resource allocations and prices change when the number of users in the network increases. For this, we consider eight WSTAs in the network having channel conditions similar to those in Scenario 1. The first three WSTAs stream the "Foreman" sequence, the WSTAs 4, 5, and 6 stream the "Coastguard" sequence and the last two WSTAs stream the "Mobile" sequence. The sequence formats are all at CIF resolution 30 Hz. The initial resource allocations are assumed to be the same, i.e., $\alpha_i^0 = 0.125, \forall i = 1, \ldots, 8$. It is easy to check, from Table V, that our pricing mechanism generates allocations that are feasible, utility-maximizing, and satisfy voluntary participation.

Comparing to Scenario 1, in this scenario the number of participating WSTAs is increased. This implies that the amount of resource allocated to each WSTA becomes smaller. Hence, each WSTA can only transmit the most important video packets. In particular, the results in this scenario show how our pricing mechanism is able to scale with the number of WSTAs in the network.

C. Assessing the Convergence and Convergence Rate

In Section IV-C, we discuss the convergence and corresponding rate of our pricing mechanism, which will be verified in the following experiments. By determining the convergence rate in different scenarios, we quantify the negotiation overhead of the message exchanges discussed in Section IV-D.

1) The Impact of γ on the Convergence: In this experiment we assess how γ affects the convergence of the message exchange. The simulation setup is the same as in Scenario 1 of Section VI-B. From the results in Table IV, we know that the equilibrium price is 24.9203. Fig. 2 shows the announced prices to WSTA 1 for $\gamma = 0.9$, 5 and 20, respectively, from which, we note that the small value of $\gamma(= 0.9)$ leads to the iteration oscillate and never converge, while large values of $\gamma(= 20)$ results in slow convergence. Hence, it is important for delay-sensitive multimedia applications to choose an appropriate γ.

The convergence rate for our decentralized and delay-sensitive multimedia applications is also very important. The larger the number of iterations (i.e., message exchanges) needed to determine an optimal solution, the longer it takes to perform the resource allocation. From Fig. 2, we note that a large value of γ leads to slow convergence.

2) The Impact of the Initial Price on the Convergence Rate: In this experiment we assess how the initial price affects the convergence rate of the message exchange. The simulation setup is the same as in Scenario 1 of Section VI-B. In this experiment

[6]The difference of the resource allocation between the centralized optimization and our pricing mechanism comes from the fact that we stop our mechanism after a finite number of steps, i.e., when the update step size of the price is small enough.

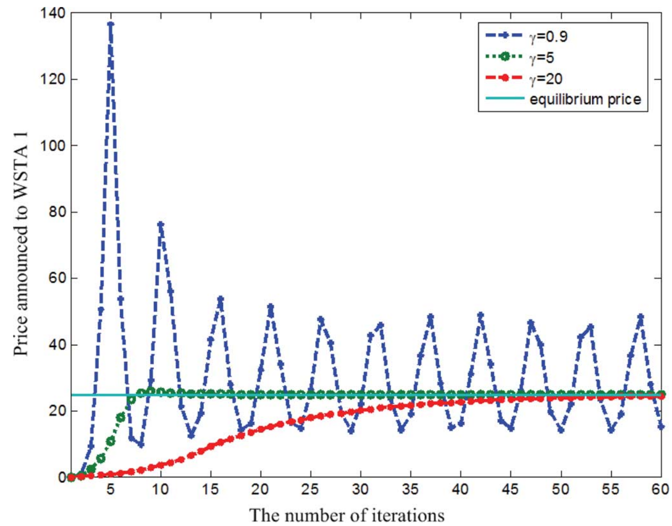

Fig. 2. Announced price to WSTA 1 for various γ.

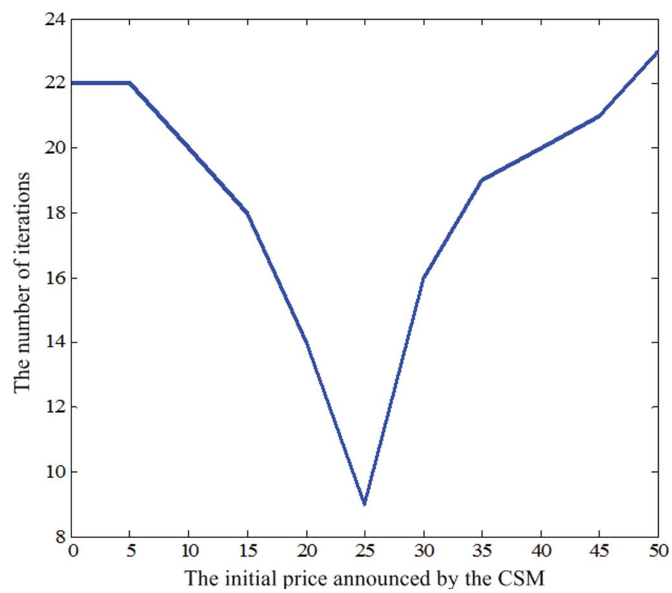

Fig. 3. Number of iterations as a function of the initial price set in the pricing mechanism.

we vary the initial price from 0 to 50 and record the number of iterations required before reaching the equilibrium message. From the results in Table IV we know that the equilibrium price is 24.9203. In Fig. 3, given various initial prices, we show how many iterations are needed in order to converge to the equilibrium price. From this figure we conclude that the number of iterations increases as the distance between the initial price and the equilibrium price becomes larger.

From the results in Scenario 2 of Section VI-B, we know that by increasing the number of the WSTAs in the network, the equilibrium price will also increase. This tells us that if the CSM has an estimate of the type of traffic desired and the number of users requesting service, it may be able to choose an initial price that is closer to the equilibrium price. By employing such strategies the speed of convergence of the mechanism may be drastically increased. A detailed investigation of this issue is part of our future research.

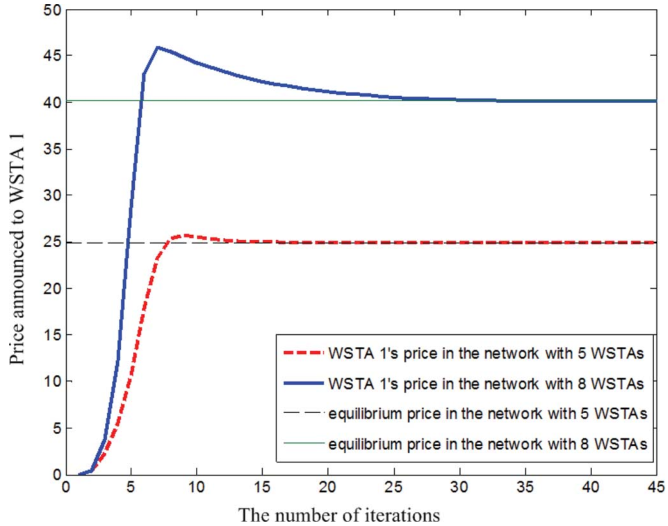

Fig. 4. Announced price to WSTA 1 for various iterations.

3) The Impact of the Number of WSTAs on the Convergence Rate: From the results in Scenario 2 of Section VI-B, we know that the increasing number of WSTAs raises the equilibrium price. In this experiment we assess how the increasing number of WSTAs affects the convergence rate of the message exchange process. We compare the network with five WSTAs and the network with eight WSTAs, as in Scenarios 1 and 2 of Section VI-B, respectively. Since we assume that the CSM has no prior information about the network we set the initial prices for both cases to zero. Fig. 4 shows the announced prices to the WSTA 1 at each iteration of the message exchange. When the announced prices converge to the equilibrium price, the message equilibriums are reached. From the figure we see that the network with eight WSTAs converges slower than the one with five WSTAs. The slow convergence of the larger network comes from the fact that the second order derivative of the quality functions of the WSTAs is larger for smaller amounts of resource allocation. A large second order derivative of the quality function can correspond to prices that are far apart even though the resource allocations are close. For this reason, the values of the prices converge slower in a more congested network.

VII. Conclusions

In this paper we model the wireless resource allocation problem as a "game" played among strategic WSTAs that are streaming video in real-time over a shared wireless network. We propose a pricing mechanism which takes into account the strategic behavior of individual WSTAs. This mechanism allows the WSTAs to exchange with the CSM a limited number of messages to reach the Nash equilibrium. The resulting Nash equilibrium messages generate the optimal resource allocations. After the allocation, each WSTA deploys a real-time transmission strategy to efficiently transmit its video bitstream. Our simulations verify that the allocations generated by the pricing mechanism i) are utility-maximizing and ii) satisfy voluntary participation. Using the proposed pricing mechanism, the WSTAs are able to appropriately "sell" or "buy" the resource based on the equilibrium price. Moreover, our results show that the equilibrium price is gracefully scales with varying channel conditions and the video traffic requirements of the WSTAs as well the number of WSTAs involved in the network (i.e., network congestion level). The convergence rate is also discussed and our simulation results show that the information about the video traffic and the number of WSTAs can accelerate the convergence rate.

Appendix I

In proving this theorem we proceed as follows. First we present the necessary and sufficient conditions for the utility-maximizing allocations of the problem in (1). Then we show that the Nash equilibria of mechanism presented in Section IV-B satisfy the utility-maximizing conditions. In order to determine an optimal solution of Problem (1) we first write the Lagrangian function:

$$\Lambda(\boldsymbol{\alpha}, \boldsymbol{\tau}, \lambda, \varphi) \triangleq \sum_{i=1}^{M} U_i(\alpha_i, \tau_i) - \lambda\left(\sum_{i=1}^{M} \alpha_i - 1\right) - \varphi \sum_{i=1}^{M} \tau_i. \tag{31}$$

At an optimal allocation $(\boldsymbol{\alpha}^*, \boldsymbol{\tau}^*) \triangleq (\alpha_i^*, \tau_i^*)_{i\in\{1,\ldots,M\}}$, the necessary and sufficient Karush–Kuhn–Tuker (KKT) conditions for optimality [32] are

$$0 \in \nabla_{\alpha_i}\Lambda(\boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \lambda, \varphi) = \nabla_{\alpha_i} Q_i^{rec}(\alpha_i^*) - \lambda, \forall i \in \{1, \ldots, M\}, \tag{32}$$

$$0 \in \nabla_{\tau_i}\Lambda(\boldsymbol{\alpha}^*, \boldsymbol{\tau}^*, \lambda, \varphi) = 1 - \varphi, \forall i \in \{1, \ldots, M\}, \tag{33}$$

$$\lambda\left(\sum_{i=1}^{M} \alpha_i^* - 1\right) = 0, \tag{34}$$

$$\varphi \sum_{i=1}^{M} \tau_i^* = 0, \tag{35}$$

where $\nabla_{x_i} f(x)$ represents the set of subgradients[7] of f at x in the ith coordinate, and λ and φ are the Lagrange multipliers for the capacity and budget constraints, respectively. Substituting (33) into (35), the KKT conditions can be reduced to

$$0 \in \nabla_{\alpha_i} Q_i^{rec}(\alpha_i^*) - \lambda, \forall i \in \{1, \ldots, M\}, \tag{36}$$

$$\lambda\left(\sum_{i=1}^{M} \alpha_i^* - 1\right) = 0, \tag{37}$$

$$\sum_{i=1}^{M} \tau_i^* = 0, \tag{38}$$

where (36) states that at optimality the marginal utility equals to the Lagrange multiplier λ, (37) is the capacity constraint, and (38) states that the allocation must be budget balanced. We are now going to investigate the Nash allocations generated by the mechanism presented in Section IV-B. In order to show that the mechanism implements in Nash equilibria the resource allocation in (1), we need to show that the Nash allocations satisfy

[7]Let S be a nonempty convex set $\mathbb{R}_+^M$, and let $f : S \mapsto R$ be concave. Then ξ is called a *subgradient* of f at $x \in S$ if $f(y) \le f(x) + \xi^T(y - x)$ for all $y \in S$.

(36)–(38). Note that, given a fixed (p_{-i}, d_i, γ), user i picks α_i and p_i such that its individual utility function is maximized:

$$\max_{\alpha_i, p_i} \Big\{ Q_i^{rec}(\alpha_i) - \left(\alpha_i - \alpha_i^0\right) \times p_{-i} - \left[p_i - p_{-i} \left(1 + \frac{d_i + \alpha_i}{\gamma} \right) - \chi_+(d_i, \alpha_i, \gamma) \right]^2 \Big\}. \tag{39}$$

At a Nash equilibrium message, (39) is maximized for each user $i \in \{1, \ldots, M\}$. Assume that $(\boldsymbol{\alpha}^*, \boldsymbol{p}^*)$ is a Nash equilibrium message. By the first order conditions for each $i \in \{1, \ldots, M\}$ we have that

$$\begin{aligned} 0 \in & \nabla_{\alpha_i} Q_i^{rec}(\alpha_i^*) \\ & - \left[p_{-i}^* + 2 \times \left(\frac{p_{-i}^*}{\gamma} + \frac{\partial}{\partial \alpha_i} \chi_+ (d_i^*, \alpha_i^*, \gamma) \right) \right. \\ & \times \left[p_i^* - p_{-i}^* \left(1 + \frac{d_i^* + \alpha_i^*}{\gamma} \right) \right. \\ & \left. \left. - \chi_+ (d_i^*, \alpha_i^*, \gamma) \right] \right] \Big\} \end{aligned} \tag{40}$$

$$p_i^* = p_{-i}^* \left(1 + \frac{d_i^* + \alpha_i^*}{\gamma} \right) + \chi_+ (d_i^*, \alpha_i^*, \gamma) \tag{41}$$

where

$$p_{-i}^* = \frac{1}{M-1} \sum_{\substack{j=1 \\ j \neq i}}^{M} p_j^* \tag{42}$$

and

$$d_i^* = \sum_{\substack{j=1 \\ j \neq i}}^{M} \alpha_j^* - 1 \tag{43}$$

Substituting (41) into (40), the first order conditions become:

$$p_{-i}^* \in \nabla_{\alpha_i} Q_i^{rec}(\alpha_i^*). \tag{44}$$

Since (41) has to be satisfied for all $i \in \{1, \ldots, M\}$, by summing over all i we have

$$\begin{aligned} \sum_{i=1}^{M} p_i^* &= \sum_{i=1}^{M} p_{-i}^* \left(1 + \frac{d_i^* + \alpha_i^*}{\gamma} \right) + \chi_+ (d_i^*, \alpha_i^*, \gamma) \\ &= \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} \frac{p_j^*}{N-1} \left(1 + \frac{d_i^* + \alpha_i^*}{\gamma} \right) + N \times \chi_+ (d_i^*, \alpha_i^*, \gamma) \\ &= \sum_{i=1}^{M} p_i^* \left(1 + \frac{d_i^* + \alpha_i^*}{\gamma} \right) + N \times \chi_+ (d_i^*, \alpha_i^*, \gamma) \end{aligned} \tag{45}$$

where the first line of (45) follows by (41), the second line by the definition of p_{-i}^*, and the third line by simplification. Equation (45) implies that

$$d_i^* + \alpha_i^* = \sum_{i=1}^{M} \alpha_i^* - 1 = 0. \tag{46}$$

and

$$\chi_+ (d_i^*, \alpha_i^*, \gamma) = 0. \tag{47}$$

This, along with (41) and the definition of p_{-i}^*, implies that at Nash equilibrium

$$p_i^* = p_{-i}^*, \ \forall i \in \{1, \ldots, M\} \tag{48}$$

$$p_i^* = p_j^* = p, \ \forall i, j \in \{1, \ldots, M\}. \tag{49}$$

Substituting (48) and (49) in (44) and (41), we can derive that

$$p \in \nabla_{\alpha_i} Q_i^{rec}(\alpha_i^*) \tag{50}$$

$$p \left(\sum_{i=1}^{M} \alpha_i^* - 1 \right) = 0. \tag{51}$$

By letting $p = \lambda$, (36) and (37) of the KKT first order conditions are satisfied. We now only have to show that the Nash equilibrium messages generate allocation which satisfy (38) (i.e., are budget balanced):

$$\begin{aligned} \sum_{i=1}^{M} \tau_i^* &= - \sum_{i=1}^{M} t_i^*(\boldsymbol{\alpha}^*, \boldsymbol{p}^*) \\ &= - \sum_{i=1}^{M} \left[\left(\alpha_i^* - \alpha_i^0\right) \times p_{-i}^* + \left[p_i^* - p_{-i}^* \left(1 + \frac{d_i^* + \alpha_i^*}{\gamma} \right) - \chi_+ (d_i^*, \alpha_i^*, \gamma) \right]^2 \right] \\ &= - \sum_{i=1}^{M} \left[\left(\alpha_i^* - \alpha_i^0\right) \times p + \left[p - p \left(1 + \frac{d_i^* + \alpha_i^*}{\gamma} \right) \right]^2 \right] \\ &= 0 \end{aligned} \tag{52}$$

where the second line of (52) follows from (19), the third line follows from (48), (49), and (46).

This establishes that the Nash equilibrium allocations are balanced, which proves that the mechanism presented in Section IV-B implements in Nash equilibria the resource allocation described in (1).

REFERENCES

[1] R. W. Lucky, "Tragedy of the commons," *IEEE Spectrum*, vol. 43, no. 1, p. 88, Jan. 2006.

[2] *IEEE 802.11, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, , Aug. 1999, IEEE Std..

[3] *IEEE 802.11e/D5.0, Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS), Draft Supplement*, , Jun. 2003.

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

[5] L. Badia, M. Lindstrom, J. Zander, and M. Zorzi, "Demand and pricing effects on the radio resource allocation of multimedia communication systems," in *Proc. IEEE Global Telecommunications Conf.*, 2003, vol. 7, pp. 4116–4121.

[6] F. Kelly, "Charging and rate control for elastic traffic," *Eur. Trans. Telecommun.*, vol. 8, no. 1, pp. 33–37, 1997.

[7] F. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *Oper. Res. Soc.*, vol. 49, pp. 237–252, 1998.

[8] A. Albanese and M. Luby, "PET-priority encoding transmission," in *High-Speed Networking for Multimedia Application*. Norwell, MA: Kluwer, 1996.

[9] J. R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Process.*, vol. 3, no. 5, Sep. 1994.

[10] R. G. Gallager, *Discrete Stochastic Processes*. Norwell, MA: Kluwer, 1996.

[11] D. Gross and C. M. Harris, *Fundamentals of Queuing Theory*, 3rd ed. New York: Wiley, 1998.
[12] D. Krishnaswamy, "Network-assisted link adaptation with power control and channel reassignment in wireless networks," in *3G Wireless Conference*, 2002, pp. 165–170.
[13] D. Bertsekas and R. Gallager, *Data Networks*. Upper Saddle River, NJ: Prentice-Hall, 1987.
[14] T. Jiang, C. K. Tham, and C. C. Ko, "An approximation for waiting time tail probabilities in multiclass systems," *IEEE Commun. Lett.*, vol. 5, no. 4, pp. 175–177, April 2001.
[15] J. Abate, G. L. Choudhury, and W. Whitt, "Exponential approximations for tail probabilities in queues I: Waiting times," *Oper. Res.*, vol. 43, no. 5, pp. 885–901, 1995.
[16] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.
[17] M. Jackson, "Mechanism theory," In the Encyclopedia of Life Support Systems 2003.
[18] F. Fu, A. R. Fattahi, and M. van der Schaar, "Game theoretic paradigm for resource management in spectrum agile wireless Networks," in *Proc. ICME*, 2006.
[19] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.
[20] T. M. Stoenescu and J. O. Ledyard, "A pricing mechanism which implements a network rate allocation problem in Nash equilibria," in *Proc. 45th IEEE Conf. Decision and Control*, 2006.
[21] T. M. Stoenescu and D. Teneketzis, "Informational efficiency of pricing mechanisms for communication networks," in *Proc. 44th IEEE Conf. Decision and Control*, 2005.
[22] M. Kalman and B. Girod, "Techniques for improved rate-distortion optimized video streaming," *ST J. Res.*, vol. 2, no. 1, pp. 45–54, November 2005.
[23] M. R. Garey and D. S. Johnson, *Computers and Intractability—A Guide to the Theory of NP-Completeness*, V. Klee, Ed., 23rd ed. New York: W. H. Freeman, 2002.
[24] D. Jurca and P. Frossard, "Media Streaming with conservative delay on variable rate channels," in *Proc. IEEE ICME 2006*, July 2006.
[25] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Trans. Multimedia*, vol. 8, no. 2, pp. 390–404, Apr. 2006.
[26] M. van der Schaar and N. Sai Shankar, "Cross-layer wireless multimedia transmission: Challenges, principles and new paradigms," *IEEE Wireless Commun. Mag.*, vol. 12, no. 4, pp. 50–58, Aug. 2005.
[27] T. Groves and J. Ledyard, "Incentive compatibility since 1972," Information, Incentives, and Economic Mechanisms, Essays in Honor of Leonid Hurwicz.
[28] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming, Theory and Algorithms*, ser. Wiley Interscience Series in Discrete Mathematics and Optimization. New York: , 1993.
[29] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. : Cambridge Univ. Press, 2004.
[30] E. Setton, T. Yoo, X. Zhu, A. Goldsmith, and B. Girod, "Cross-layer design of ad-hoc networks for real-time video streaming," *IEEE Wireless Commun. Mag.*, vol. 12, no. 4, pp. 59–65, Aug. 2005.
[31] T. Tian, A. Li, J. Wen, and J. Villasenor, "Priority dropping in network transmission of scalable video," in *Proc. Int. Conf. on Image Processing (ICIP)*, 2000, vol. 3, pp. 400–403.
[32] M. S. Bazaraa, J. J. Goode, and C. M. Shetty, "Optimality criteria in nonlinear programming without differentiability," *Oper. Res.*, vol. 19, no. 1, pp. 77–86, 1971.
[33] T. Stockhammer and M. Bystrom, "H.264/AVC data partitioning for mobile video communication," in *Proc. IEEE Int. Conf. Image Processing*, Singapore, Oct. 2004, pp. 545–548.
[34] J. Ye and Y. Chen, "Rate-distortion optimized data partitioning for video using backward adaptation," in *ICASSP*, 2003.
[35] A. Mala, M. El-Kadi, S. Olariu, and P. Todorova, "A fair resource allocation protocol for multimedia wireless networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 14, no. 1, 2003.
[36] H. Park and M. van der Schaar, "Bargaining strategies for networked multimedia resource management," *IEEE Trans. Signal Process.*, to appear.
[37] Y. Xue, B. Li, and K. Nahrstedt, "Price-based resource allocation in wireless ad hoc networks," in *Proc. 11th Int. Workshop on Quality of Service*, 2003.
[38] B. Hajek and S. Yang, "Stategic buyers in a sum bid game for flat networks," in *IMA Workshop 6: Control and Pricing in Communications and Power Networks*, 2004.
[39] [Online]. Available: http://www.hthoma.de/video/delay/
[40] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, Jul. 2003.

Fangwen Fu received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 2002 and 2005, respectively. He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering Department at University of California, Los Angeles (UCLA). His research interests include wireless multimedia streaming, resource management for networks and systems, applied game theory, video processing and analysis.

Tudor M. Stoenescu received the B.S degree in electrical engineering and his M.S degree in mathematics from North Dakota State University, Fargo, in 1999, and the M.S.E. and Ph.D. degrees in electrical engineering and computer science-systems from the University of Michigan, Ann Arbor, in 2001 and 2004, respectively. He was a Postdoctoral Scholar at the California Institute of Technology, Pasadena, from 2004 to 2006.

Since 2006 he has been with the Telecommunications Architecture Group at NASA—Jet Propulsion Laboratory, Pasadena. His current research activities fall in the area of resource allocation problems in network systems. Recently he has investigated and solved various problems in: Delay Tolerant Networks, Deep Space Networks, Intra and Inter aircraft wireless networks, cooperative and noncooperative network resource allocation, and peer-to-peer network formation games. He has also conducted extensive research in the areas of stochastic control, ergodic theory and wavelet theory.

Mihaela van der Schaar (SM'04) is currently an Associate Professor in the Electrical Engineering Department at the University of California, Los Angeles (UCLA). Since 1999, she was an active participant to the ISO MPEG standard to which she made more than 50 contributions and for which she received three ISO recognition awards.

Dr. van der Schaar received the NSF CAREER Award in 2004, IBM Faculty Award in 2005, Okawa Foundation Award in 2006, Best IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY Paper Award in 2005, and Most Cited Paper Award from EURASIP Journal Signal Processing: Image Communication between the years 2004–2006. She holds 28 granted U.S. patents. She is currently an associate editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE SIGNAL PROCESSING LETTERS and *IEEE Signal Processing e-Newsletter*. She is also the editor (with Phil Chou) of the book *Multimedia over IP and Wireless Networks: Compression, Networking, and Systems*.