

Mechanism-Based Resource Allocation for Multimedia Transmission over Spectrum Agile Wireless Networks

Ahmad Reza Fattahi, *Student Member, IEEE*, Fangwen Fu, Mihaela van der Schaar, *Senior Member, IEEE*,
Fernando Paganini, *Senior Member, IEEE*

Abstract—We propose to add a new dimension to existing wireless multimedia systems by enabling autonomous stations to dynamically compete for communication resources through adjustment of their internal strategies and sharing their private information. We focus on emerging spectrum agile wireless networks, where developing an efficient strategy for managing available communication resources is of high importance. The proposed dynamic resource management approach for wireless multimedia changes the passive way stations are currently adapting their joint source-channel coding strategies according to available wireless resources. Each wireless station can play the resource management game by adapting its multimedia transmission strategy depending on the experienced channel conditions and user requirements. The resource allocation game is coordinated by a network moderator, which deploys mechanism-based resource management to determine the amount of transmission time to be allocated to various users on different frequency bands such that certain global system metrics are optimized. Subsequently, the moderator charges the various users based on the amount of resources it has allocated to them, in order to discourage them from being dishonest about their resource requirements. We investigate and quantify both the users' and the system performance when different cross-layer strategies, and hence users' levels of smartness, are deployed by wireless stations. Our simulations show that mechanism-based resource management outperforms conventional techniques such as air-fair time and equal time resource allocation in terms of the obtained system utility. They also provide insights that can guide the design of emerging spectrum agile network protocols and applications.

Index Terms—Spectrum Agile Wireless Networks, Resource Management, Wireless Multimedia Streaming, Game Theory, Mechanism Design

I. INTRODUCTION

EMERGING wireless networks provide dynamically varying resources with only limited support for the Quality of Service (QoS) required by the delay-sensitive, bandwidth-intensive and loss-tolerant multimedia applications. This variability of resources does not significantly impact delay-insensitive applications (e.g., file transfers), but has considerable consequences for multimedia applications and often leads

Manuscript received February 1, 2006; revised October 12, 2006. This work is in part supported by NSF under grant ECS-0323153.

Ahmad Reza Fattahi, Fangwen Fu, and Mihaela van der Schaar are with the Electrical Engineering Department, University of California, Los Angeles (e-mail: {ahmad, ffw, mihaela}@ee.ucla.edu).

Fernando Paganini is with the Universidad ORT, Uruguay (e-mail: paganini@ort.edu.uy).

Digital Object Identifier 10.1109/JSAC.2007.070410.

to unsatisfactory user experience. Existing algorithms and protocols for wireless transmission do not provide adequate QoS support for multimedia applications in crowded wireless networks or when the interference is high. In particular, indoor wireless technologies are overcrowding the unlicensed spectrum while the licensed spectrum goes without much utilization [1]. To fulfill the necessary QoS requirements under such conditions, wireless stations (WSTAs) need to *dynamically and cognitively* harvest additional resources as well as optimally adapt their cross-layer strategies based on the available resources.

A. Spectrum Agility and its Challenges

A possible way of obtaining additional resources is to deploy an Opportunistic Spectrum Agile Radio (OSAR) network infrastructure [2][3][4][5][6][1], where WSTAs can benefit from the opportunistic deployment of unused spectral opportunities from various frequency bands that were initially allotted to primary users (i.e., users for which the spectrum was originally assigned such as emergency services, police, etc.). While conceptually simple, the realization of OSAR has shown to be highly challenging. While deploying these emerging OSAR networks can alleviate to some extent the need for wireless resources, the problem of efficiently dividing the additional available resources among competing, autonomous WSTAs becomes highly important and is the subject of this paper. In a recent IEEE Spectrum issue, Robert W. Lucky [8] argued for the need for new and proactive resource management schemes that are able to prevent competing users from misusing the common (shared) network resources. Importantly, he mentioned the lack of incentives for the WSTAs in current wireless networks to adhere to *fairness* or courtesy rules: "Today we worry whether Wi-Fi will exhibit the same meltdown. There is no incentive, other than the ultimate survival of the system, for users to limit their use."

Naturally, each WSTA tries to acquire as much of the network resources as possible (see e.g. [7]), unless a preemptive mechanism exists in the network [8]. Even when such preemptive mechanisms exist, the problem of determining optimal utilities and strategies for allocating the transmission opportunities among various WSTAs streaming delay-sensitive multimedia still remains unsolved. The complexity of this problem is further exacerbated by the fact that the cross-layer optimization at each WSTA involves numerous

time-varying parameters and interactions among layers. This makes the interactions among WSTAs and the resulting utility-resource tradeoffs very difficult to model. Moreover, WSTAs are considered autonomous entities that separately determine and optimize their deployed cross-layer strategies. Hence, another inherent property of the considered OSAR system is its distributed nature, both in information and actions. Last but not least, for wireless multimedia applications, the resource management is further complicated by the delay-sensitive nature of the application, i.e. multimedia data that is received after its delay deadline is useless.

B. Existing Solutions and Their Limitations

To solve the aforementioned OSAR resource management problem, several strategies can be envisioned [9], including equal-time, air-fair time, and admission-control (reservation) based schemes. In the equal-time scenario, the available time on a channel is equally divided among all present users. While this allocation strategy might seem fair, it is neither efficient nor fair, since it does not consider the content characteristics, channel conditions, and the application constraints of each user. In air-fair time allocation, each user announces a measure or request of the amount of time (or rate) it requires over the next period of service called service interval. Then, each user receives an amount of time proportional to the requested amount. This strategy represents an improvement to the equal-time method because it explicitly considers the rate requirements and, inherently, the video content characteristics of different users. Nevertheless, not only the entire performance of the network heavily depends on the truthfulness of the users, it effectively persuades users to lie! There are multiple incentives for WSTAs to lie about their requirements in addition to mere quality improvements. For instance, they might want to be able to cope with sudden variations in channel conditions or content characteristics. Moreover, WSTAs can lower their power usage via over-provisioning because it allows them to deploy less-sophisticated coding and protection schemes. Alternatively, in admission-control schemes such as the IEEE 802.11e [10], the resources are allocated on a first-come first-served basis. In this scenario, users will either be admitted at their exact resource requirement or they will be denied any service. In congested networks, this scheme is inefficient because it does not scale well with the number of users. For instance, in [7] it has been shown that a considerable number of users may be denied any service in a congested IEEE 802.11e network. Moreover, as in the case of air-fair time allocation, there are multiple incentives for the users to lie about their requirements.

C. The Proposed Novel Paradigm

To overcome the aforementioned limitations of existing allocation schemes, we propose a new paradigm for resource management in OSAR networks that allows WSTAs to dynamically *compete and pay* for available spectral opportunities. In the OSAR transmission scenario, the transmission opportunities (shown as TXOPs and defined as the smallest unit of transmission time interval) available on the multiple channels need to be allocated to the various competing WSTAs.

Our paper is based on the observation that wireless devices currently operate in such a *passive* manner that degrades the whole network's performance.

We propose to change the passive resource management in which wireless stations currently interact, by allowing them to bid for the available TXOPs. For this, we assume the presence of a central spectrum moderator (CSM) (similar to [11]) that manages the available TXOPs and divides them among the various users; Hence, in this part the algorithm is centralized. To enforce autonomous self-optimizing WSTAs to act in a *socially optimal* way, the CSM adopts a tool called *transfers* through which it charges WSTAs based on the inconvenience they cause to others by using common resources. Each WSTA transmits to the CSM a vector of private information that quantifies its expected utility as a function of potential allocated time. In this sense, the information and decision-making process is decentralized. The available TXOPs in the network are then allocated to the WSTAs by the CSM in a way that the aggregate utility gets maximized. The transfers are computed in such a way that encourages the WSTAs to declare the truth. Given the current conditions of different channels and source, each WSTA has to adopt the optimal cross-layer strategy in order to maximize its own expected utility. In this stage, our method performs in a decentralized way. Hence, the proposed algorithm partly relies on the rationality and smartness of WSTAs (e.g. how good the cross-layer strategy, compression, or protection schemes are, etc.) to play the resource management game. Therefore, the burden of optimizing the transmission parameters is shared by all WSTAs. In this paper, we will not thoroughly discuss the impact of the various cross-layer strategies on the user's performance, which is defined in terms of experienced video quality. Instead, we will use our prior cross-layer strategy design results [12], [13], [9] and focus on how to manage the resulting network resources. To recap, our approach ensures truthfulness and dynamic adaptation of users' strategies based on time-varying channel and content characteristics. In other words, it promotes collaboration in an indirect way through charging WSTAs based on the inconvenience they cause to other users rather than the used resources. In this way, WSTAs will naturally tend to distribute their requests over time in an efficient manner to avoid crowded intervals as much as possible.

D. Related Work and Paper Organization

To enable the resource exchanges among WSTAs as required by the proposed resource management, we rely on recent developments on cross-layer optimization for multimedia transmission (see [9] for a review of the topic). However, in prior work, the optimization has been performed in isolation, at each individual station, and does not consider its impact on the overall wireless system.

Game theory has been used in previous research to resolve resource allocation issues for wireless networks in a distributed and scalable manner [14][15][16]. However, previous research has not considered the benefits of dynamic and competitive resource management among WSTAs; Such a management regimen relies on users' ability to adapt their cross-layer

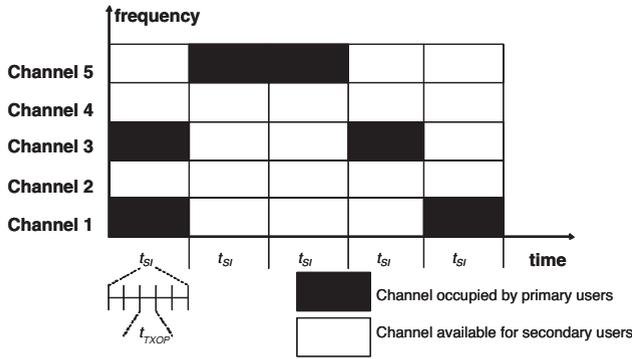


Fig. 1. Spectrum opportunities for OSAR users.

strategies to changing source properties and varying channel conditions. In [17], the authors proposed a discrete resource-utility function and maximize the aggregate utility by dynamically assigning network resources. However, this centralized allocation method passively adjusts the allocation based on the previous observations and does not take into account the dynamic user behavior. In [18], pricing schemes are introduced from the point of view of the service provider, by considering the requested quality of service and the willingness to pay. However, the relationship between the assigned resources and the gained utility is not thoroughly studied.

The proposed framework relies on related work in OSAR network development, multimedia compression, streaming, cross-layer design, and game-theory. In this paper we rely on existing research on OSAR network infrastructure [1]. The US Federal Communications Commission (FCC) has issued a Notice of Public Rulemaking and Order regarding the so-called OSAR or cognitive radio technologies [5]. The Defense Advanced Research Projects Agency (DARPA) has also started the neXt Generation (XG) Communications Program to develop new technologies that allow multiple radio systems to share the spectrum through adaptive mechanisms [2]. For more details on the OSAR infrastructure, the interested reader is referred to [2][3][4][5][6][1].

The paper is organized as follows. In Section II, we introduce the investigated OSAR system and its parameters. Section III presents the transmission strategies that are deployed by the users in playing the resource management game. The users and system utilities is introduced in section IV. Section V proposes the mechanism for time allocations and computing transfers. Numerical results are presented in section VI followed by conclusions and future work in Section VII.

II. OSAR SYSTEM DESCRIPTION

We consider an OSAR wireless communication network system, where two types of users co-exist, namely primary and secondary users. (In this paper, the expressions "user" and "WSTA" are interchangeably used). Primary users have exclusive access to designated spectral bands, while secondary users only access spectral bands when the primary users do not use that band. To realize such an opportunistic use of idle spectral resources, secondary users need to possess spectral agility [3], enabled for instance by software-defined radios. The network moderator can then locate and distribute available

resources among the various secondary users, in both spectral and temporal domains (see Figure 1).

The wireless spectrum that can be accessed by the secondary users is divided into *channels*, which represent the smallest unit of a spectral band. As in [4], we differentiate two types of spectrum agile radios, referred to as *type I* and *type II* users. The type I WSTAs use a fixed spectral bandwidth to transmit their data, but they may effectively exploit the available spectrum opportunities by dynamically hopping between the various channels. For instance, the WLANs that exist today are examples of type I agile radio WSTAs. In the type II case, the WSTAs can dynamically expand and contract their bandwidth and also adapt their physical layer and modulation strategy based on the vacant spectral opportunities present on all the available channels. In our paper, we assume that all WSTAs are type II users; In a simulation case we show the effects of sudden presence of a type I user on type II users' performances and resource allocations. Note that in our analysis, we assume that each secondary user can scan a channel, switch to a channel, and vacate a channel instantly (when claimed by the primary network) without incurring any control overhead or delay. In the investigated communication system, we assume that each WSTA is transmitting multimedia bitstreams to a single WSTA connected to the same OSAR infrastructure. However, as mentioned in [3], one of the most challenging tasks in realizing a spectral-agile network is to maintain the connectivity among the communicating WSTAs. Hence, an important role of the CSM is not only to determine the TXOP assignments per channel for the various transmit-receive WSTA pairs, but also to disseminate these assignments to both involved WSTAs, such that they can maintain their connectivity. Defining efficient protocols for communicating spectrum opportunities among WSTAs and disseminating TXOP assignments between the CSM and the users are beyond the scope of this paper, but they represent an important topic for further research as their overheads and latency may significantly impact the performance of the system.

In this paper, we assume that the various spectrum-agility functionalities are already implemented using e.g. the system described in [3][4]. Hence, we do not consider here the important problems of spectral opportunity discovery or management. We assume that based on e.g. [3], each WSTA and network moderator can maintain a spectral opportunity map, which stores the status of each channel in the considered wireless spectrum. We assume that while the network moderator has full knowledge about the available resources, the spectrum maps of the various WSTAs is in general a subset of all the available channels and/or spectral opportunities. Moreover, the available opportunities are characterized differently by the WSTAs based on their experienced channel conditions. Hence, the spectrum opportunity maps of the various WSTAs will be different. However, it is not essential for all nodes to maintain an identical spectrum map as long as the network moderator coordinates their channel assignments.

The resource allocation mechanism for OSAR has to fulfill several important properties. It needs to scale to a varying number of users having different requirements and adapt to the dynamic nature of the wireless environment and the time-

varying video source characteristics. Moreover, due to the distributed property of information, we need to adopt a distributed optimization model, in which the CSM does not need to be aware of the transmission strategies, requirements, or capabilities of the various users. Thus, a considerable portion of the complexity of the OSAR system optimization resides at the user side (i.e. they need to adapt their cross-layer strategies accordingly and determine what are the resulting utilities for different resource allocations) rather than the CSM's.

We assume that the proposed game-theoretic resource management strategy is implemented using a reservation (polling)-based MAC, where a CSM is allocating time slots to the various wireless stations every service interval (SI). The CSM has sufficient authority to allocate TXOPs to users, charge users if necessary, and deny service to WSTAs which do not comply with network regulations. The number of TXOPs in each SI equals Q ; Hence, the relation $t_{TXOP} = \frac{t_{SI}}{Q}$ holds where t_{SI} and t_{TXOP} are the durations of each SI and TXOP, respectively (See Figure 1). Each WSTA can potentially be allocated q number of TXOPs per channel per SI (where $q \in \{1, 2, \dots, Q\}$).

The considered OSAR network has N channels available whereas there are M secondary users competing for these resources. All secondary users are assumed to be of type II. In addition, we assume that the performance of each channel is characterized by each user based on the experienced Signal to Noise Ratio (SNR). Hence, each WSTA $i, i = 1, 2, \dots, M$ is assumed to be capable of measuring the SNR of channel $j, j = 1, \dots, N$, represented by $SNR_{i,j}$.

III. USERS' STRATEGIES

In this section we will analyze the actions and strategies that a WSTA can deploy. Since the network is considered a competitive adversarial environment, the way users will play the game will be of paramount importance for their own performance, as well as the overall system performance. We will first investigate and define the space of feasible strategies for a WSTA. In the following sections, based on the strategies defined in this section, we will define the utilities and show results about the game outcomes.

We assume that each user has a multimedia content to be transmitted through the wireless network. We also assume that each WSTA has information about the spectral opportunities on the various channels, as well as about their quality represented by $SNR_{i,j}$. Given the multimedia content and channel conditions, the WSTA should decide about two major set of strategies:

- *Internal Strategies*: This set of strategies, represented by \mathcal{S}_i^{int} for user i , includes the cross-layer transmission parameters and strategies used by each WSTA.
- *External Strategies*: As shown later in this paper, each user has to announce a vector of private information to the CSM at the beginning of each SI. External strategies, whose space for user i is represented by \mathcal{S}_i^{ext} , determine how each WSTA decides about the information to be transmitted to the CSM such that it results in the most available expected payoff for that user.

In the following two subsections we discuss these two sets of strategies in detail.

A. Internal Strategies

The benefit or *utility* that user i gains by successfully transmitting the k -th packet from transmission queue, shown here by v_i^k , is denoted as Δ_i^k ; It is defined as the *distortion reduction* at the video receiver in case the video data of packet number k is correctly decoded at the receiver. The utility Δ_i^k is expressed in our paper as the *expected mean square error* (MSE) reduction at the video decoder instead of the visual distortion reduction, since the latter is harder to quantify. Rate-distortion (R-D) models can be used for modelling the utility as a function of rate/time. These models are codec specific and such R-D models can be found for example in [19].

Let $s = [phy_i^l, mac_i^m, app_i^n] \in \mathcal{S}_i^{int}, i = 1, 2, \dots, M$ be a nominal vector of cross-layer adaptation strategies feasible to the i -th WSTA, where $\mathcal{S}_i^{int} = \mathcal{S}_i^{PHY} \times \mathcal{S}_i^{MAC} \times \mathcal{S}_i^{APP}$ and the three sets $\mathcal{S}_i^{PHY}, \mathcal{S}_i^{MAC}$, and \mathcal{S}_i^{APP} are the strategy spaces of user i in physical (PHY), medium access control (MAC), and application layers (APP), respectively. We also assume that the three strategy spaces above have a finite number of elements, with $N_i^{PHY} = |\mathcal{S}_i^{PHY}|$, $N_i^{MAC} = |\mathcal{S}_i^{MAC}|$, and $N_i^{APP} = |\mathcal{S}_i^{APP}|$. In general, the size of the strategy space is very large. However, in this paper we consider only the optimization of a limited set of parameters and strategies at various layers. For instance, at the PHY, we only allow users to adjust their modulation and coding schemes and assume that other parameters are fixed. Hence, $\mathcal{S}_i^{PHY} = \{phy_i^1, \dots, phy_i^{N_i^{PHY}}\}$ represents the nominal PHY strategy space of user $i, i = 1, 2, \dots, M$ where each element phy_i^k shows a particular vector of modulation and channel coding strategies feasible to user i on N channels. In the same manner, in the MAC layer, we only consider adaptive retry-limit adaptation per packet and hence, the strategy space can be defined as $\mathcal{S}_i^{MAC} = \{mac_i^1, \dots, mac_i^{N_i^{MAC}}\}$, where mac_i^k represents a vector of maximum retransmission numbers per packet per channel for user i . In the APP layer, users can adapt the transmission rate or scheduling strategy, $\mathcal{S}_i^{APP} = \{app_i^1, \dots, app_i^{N_i^{APP}}\}$, where app_i^k shows a specific packet scheduling in the transmission queue of user i based on the contribution of the packets in video quality, delay constraints, etc.

The SNR of channel j seen by the i -th WSTA together with its physical layer strategy phy_i determine the bit-error probability of user i on channel j which is represented by $e(SNR_{i,j}, phy_i)$ and is assumed independent and similar for all bits. Then the packet-loss probability for user i over channel j will be computed as:

$$e_{i,j}(L_i, phy_i) = 1 - (1 - e(SNR_{i,j}, phy_i))^{L_i} \quad (1)$$

where L_i is the average packet size of user i in bits. We also assume that through a single SI, the changes in channel quality are negligible and therefore the $SNR_{i,j}, i = 1, \dots, M; j = 1, \dots, N$ is constant over a certain SI. For OSAR networks that deploy similar modulation and coding schemes like IEEE 802.1e networks, it can be shown [20] that the physical-layer throughput of channel j can be approximated by:

$$R_{i,j}^{phy}(SNR_{i,j}, phy_i) = \frac{R_{max}^{phy}(phy_i)}{1 + e^{-\mu(SNR_{i,j} - \delta)}} \quad (2)$$

where $R_{max}^{phy}(\mathbf{phy}_i)$ is the maximum achievable data rate for the physical layer strategy \mathbf{phy}_i and μ, δ are constants whose values for each modulation and coding strategy \mathbf{phy}_i can be determined as in [20]. We assume that each packet is retransmitted until it is received or its deadline is expired. Given the modulation, the maximum number of retransmissions (including the initial transmission) of packet v by user i on channel j can be dynamically computed as:

$$T_{i,j}^{max}(L_i, v) = \frac{R_{phy}^{i,j} \cdot (t_i^{delay}(v) - t^{trans}(v))}{L_i} \quad (3)$$

where $t_i^{delay}(v) = \min\{\text{the deadline of packet } v, t_{SI}\}$ and $t^{trans}(v)$ is the expected time that user i begins the first transmission attempt of packet v . If a number of packets are ordered in the transmission queue of user i , then for the first packet, $t^{trans}(v)$ is the current time while for the next packets the expected transmission times of previous packets should be accounted for (based on the average number of transmissions each packet takes until successfully transmitted as in equation (5) below and channel rate computed by equation (2)). Then the probability of successfully receiving these packets can simply be computed as [21]:

$$P_{i,j}^{succ}(s) = 1 - [e_{i,j}(L_i, \mathbf{phy}_i)]^{T_{i,j}^{max}(L_i, v)} \quad (4)$$

The *average* number of transmissions until the packet is successfully transmitted, or the retransmission limit is reached, can be calculated as [21]:

$$N_{tr}^{mean}(\mathbf{phy}_i, T_{i,j}^{max}(L_i, v)) = \frac{1 - [e_{i,j}(L_i, \mathbf{phy}_i)]^{T_{i,j}^{max}(L_i, v)}}{1 - e_{i,j}(L_i, \mathbf{phy}_i)} \quad (5)$$

Hence, the average number of packets that can be correctly transmitted during the time t_{TXOP} by user i over channel j can be computed as $\bar{p}_{i,j}(t_{TXOP})$:

$$\bar{p}_{i,j}(t_{TXOP}) = \max\{p | t_{TXOP} \geq \frac{L_i}{R_{phy}(\mathbf{phy}_i, SNR_{i,j})} \sum_{k=1}^{p-1} N_{tr}^{mean}(\mathbf{phy}_i, T_{i,j}^{max}(L_i, v^k))\} \quad (6)$$

in which we simply counted the number of packets which can be successfully transmitted before the current TXOP is over [7].

Finally, the total number of packets that could on average be transmitted by user i over all channels in one TXOP is equal to:

$$\bar{p}_i^{tot} = \sum_{j=1}^N \bar{p}_{i,j}(t_{TXOP}) \quad (7)$$

The above intermediate parameters will be used later to calculate the users' utilities and the method to choose them is called *link adaptation* and is discussed in detail in [9]. The internal strategies for our delay-sensitive video transmission include modulation and coding mode selection at the PHY layer, adapting the number of retransmissions at the MAC layer and adaptive packet scheduling at the APP layer (consult [13], [1] for more details on the various cross-layer strategies that can be deployed by the WSTAs and their impact on the resulting video quality). Hence, the internal strategies

determine the expected video quality at the receiver side as a function of the allocated time. This, in turn, determines the private information characterizing the utility function for each WSTA. These vectors of private information will be revealed to the CSM as the external strategies which are described in section III-B. The form of the information space is discussed in sections IV and V. In the next subsection we introduce the external strategies of users.

B. External Strategies

After each user decides about its internal strategy and calculates its intermediate parameters, it has to announce a function of them to the CSM according to certain protocols defined in section V. These protocols determine the form or *space* of the messages and constitute a considerable portion of the contributions of this paper which are discussed rigorously in section V. On the other hand, the role of external strategies is to determine the *content* of what is to be sent to the CSM such that the expected utility gets maximized. Since the network is assumed competitive and the resources are scarce, in general, there is no guarantee that users do not lie about their private information in a way that leads to more payoff for them.

However, as we show in section V, the best external strategy for all users is to announce the true private information; Hence, we say that our mechanism is *incentive compatible*. Besides, truthfulness is the dominant strategy regardless of what strategy other users take; In more technical terms we say that announcing the truth, is implemented in *dominant strategies*. Since there is no reason for users to announce other than their true private information, we content our strategy analysis mainly to internal strategies. This fact leads to a very useful separation principle: At first, each user is interested in internal strategies and afterwards, it has to decide, based on its decision in first stage, how to play the game. The fact that announcing the truth is the equilibrium of the game implemented in dominant strategies, implies that further analysis of the external strategies is unnecessary. In the next section we will compute the expected utility of a WSTA resulting from the time allocation vector on the network channels.

IV. UTILITIES

Given the space of strategies of the WSTAs in the network, we will discuss the nature of the utility functions that WSTAs and the CSM seek to separately maximize. This paper is focused on video applications and therefore, we assume that all users are interested in transmitting video data. However, the utility and mechanism formulation in the sequel, is general enough to make the CSM capable of handling various user types. The fact that we only define the utility functions for video, is just for length limits; The only expectation from a user of any kind is that it calculates its own utility function. On the other side, the CSM even does not care about any specific application as long as each user announces some utility values.

A. Users' Utilities

The vector of allocated network time to each user i is a column vector of times, $\mathbf{t}_i \in \mathbb{R}^N$, which represents the time

intervals allocated to user i on the N channels available in the network (Figure 1). We also define the collection of the allocated times to all M users as an accumulated column vector of size $M \times N$ as $\mathbf{t} = [t_1^T, t_2^T, \dots, t_M^T]^T \in \mathbb{R}^{M \times N}$. In the following we will define user i 's utility as a function of the vector of allocated times to all users, \mathbf{t} . Define $size_i^{queue}$ as the size of the transmission queue of the i -th WSTA in packets. Next we define the discrete distortion reduction function of user i as a function of the *number* of transmitted packets, $V_i: \mathbb{N} \cup \{0\} \rightarrow \mathbb{R}^+$, as the following:

$$V_i(n) = \begin{cases} 0 & \text{for } n = 0 \\ \sum_{k=1}^n \Delta_i^{q_i+k-1} & \text{for } n > 0 \end{cases}$$

where q_i is the place-holder index of user i which points to the top of user i 's queue, $\Delta_i^{q_i+k}$ is the distortion reduction of the packet in place q_i+k when $q_i+k \leq size_i^{queue}$, and $\Delta_i^{q_i+k} = 0$ otherwise. Since in this paper we do not consider any temporal effects on the value of packets (i.e. how close a packet is to its deadline), the corresponding *value* index (utility) that is attached to each packet is assumed equal to the distortion reduction resulting from that packet. Therefore the packet utility, for each certain packet, is assumed fixed over time. The discrete distortion reduction function $V_i(n)$ represents user i 's distortion reduction resulting from successfully transmitting the first n packets from its transmission queue. We proceed to defining user i 's *continuous* version of the distortion reduction function, $V_i^{cont}: \mathbb{R}^+ \rightarrow \mathbb{R}^+$, as:

$$V_i^{cont}(\eta) = \begin{cases} V_i(\eta), & \text{for } \eta \text{ integer} \\ V_i(n+1)(\eta-n) - V_i(n)(\eta-n-1), & \text{o.w.} \end{cases}$$

where n is integer and $n < \eta < n+1$. The function V_i^{cont} is the linearly-interpolated version of the original V_i .

Lemma 1: In the above setup if the distortion reductions in user i 's transmission queue are sorted in descending order, i.e., $\Delta_i^k \geq \Delta_i^{k+1}, \forall i = 1, 2, \dots, M$ and $\forall k = 1, 2, \dots, size_i^{queue} - 1$, then the continuous distortion reduction function $V_i^{cont}(\eta)$ is concave in η over \mathbb{R}^+ .

Proof: Due to the fact that the distortion reductions in the queue are sorted in descending order, the function V_i^{cont} can alternatively be expressed as the point-wise minimum of $size_i^{queue} + 1$ affine functions over \mathbb{R}^+ : $V_i^{cont}(\eta) = \min\{l_i^k(\eta), k = 1, 2, \dots, size_i^{queue} + 1\}$ where the affine functions are defined as:

$$l_i^k(\eta) = (\eta - k)[V_i(k) - V_i(k-1)] + V_i(k), \quad \text{for } k \in \{1, \dots, size_i^{queue} + 1\} \quad (8)$$

Hence V_i^{cont} is concave in η over \mathbb{R}^+ [22]. ■

The next step is to state the utility of each user directly in terms of the allocated network resources, which in this case is transmission time. For an arbitrary vector w , we show its j -th element by $w(j)$. Based on the definitions of the parameters $\bar{p}_{i,j}$ and t_{SI} , the *effective* number of packets that user i can successfully transmit on average, using time vector \mathbf{t}_i , is:

$$\begin{aligned} NoP_i(\mathbf{t}_i) &= \sum_{j=1}^N \bar{p}_{i,j} \frac{t_i(j)}{t_{TXOP}} \\ &= \boldsymbol{\alpha}_i^T \cdot \mathbf{t}_i \end{aligned} \quad (9)$$

where the column vector $\boldsymbol{\alpha}_i \in \mathbb{R}^N$ is defined as $\boldsymbol{\alpha}_i = \frac{1}{t_{TXOP}} \cdot [\bar{p}_{i,1}, \dots, \bar{p}_{i,N}]^T$. Hence, user i 's utility function $U_i(\mathbf{t}) : \mathbb{R}^{M \times N} \rightarrow \mathbb{R}$ is defined as the expected distortion reduction resulting from transmitting over the corresponding N channels in the allocated times as:

$$U_i(\mathbf{t}) = V_i^{cont}(\boldsymbol{\alpha}_i^T \cdot \mathbf{t}_i) \quad (10)$$

Note that the utility function of each user seems to depend only on user i 's strategy. However, the dependency of user i 's utility on all other users' strategies is implicit through the allocated time \mathbf{t}_i which in turn is a direct function of all users' strategies. Finally, the parameters Δ_i , which shape the utility function V_i^{cont} , and the parameters $\boldsymbol{\alpha}_i$ are transmitted to the CSM in the mechanism design which will be discussed in detail in section V.

B. System Utility

Having defined the utility of every individual user, we discuss the utility function of the whole system. The goal of the CSM is in general different from that of the individual users. In this paper, we propose a CSM whose goal is to maximize the following *system utility* function:

$$U_{SYS}(\mathbf{t}) = \sum_{i=1}^M U_i(\mathbf{t}_i) \quad (11)$$

In other words, the CSM cares about the aggregate utility of all users present in the network. The following proposition paves the way for performing efficient maximization of the $U_{SYS}(\mathbf{t})$ over \mathbf{t} as will be shown in the sequel:

Proposition 2: The system utility function $U_{SYS}(\mathbf{t})$ as defined above is concave in the vector of time allocations \mathbf{t} .

Using Lemma 1 and observing that $\boldsymbol{\alpha}_i^T \cdot \mathbf{t}_i$ is a linear function of \mathbf{t}_i , proves that each user's utility function $U_i(\mathbf{t}_i)$ is concave in \mathbf{t}_i . Therefore, by definition, $U_{SYS}(\mathbf{t})$ is concave in \mathbf{t} because it is the sum of M concave functions. In the next section we will synthesize a mechanism based on which WSTAs and the CSM interact in a way that the system utility will get maximized.

V. MECHANISM DESIGN

In this section we design a mechanism to moderate the network comprised of selfish users. The key problem for mechanism-based resource management is how the CSM should allocate the time slots to users in an efficient and fair way. Assuming that WSTAs announce *correct* information, one could be optimistic that solving some sort of optimization program over users' utilities might be feasible. Unfortunately, that is a naive assumption because selfish users by definition aim at improving their own utility. Hence, they are prone to lie about their internal parameters to the CSM. The question is thus "how the penalty of a selfish user should be designed such that it refrains from requesting unnecessary transmission time?" The basic tool to prevent the lying/exaggeration problem associated with selfish users, deployed by a CSM, is the so-called *mechanism* from the game theory literature [23], [24]. Generally speaking, a mechanism is a tool for

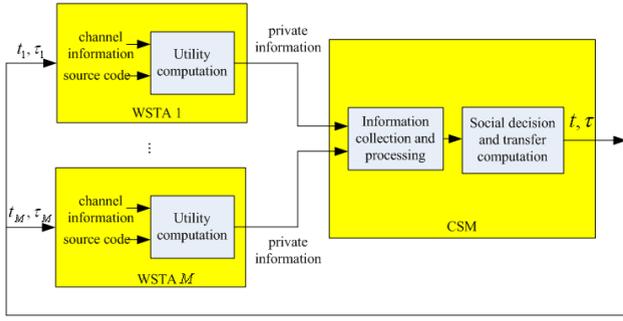


Fig. 2. The block diagram of the whole system over one SI.

efficient resource management in cases where users are non-collaborative and the information is decentralized. A mechanism has three main components: (i) The environment which in this case is the source and channel characteristics. The environment can not be affected by the users or the CSM. (ii) The message space which describes the *structure* of the private information to be exchanged by the users and the CSM. This choice plays a very important role in the resulting outcome of the mechanism and composes one of the key contributions of this paper. For the problem in hand, it is rigorously defined in section V-A.1. (iii) The outcome correspondence which determines the outcome given the messages from the users. The outcome for our problem is the vector of time allocations, \mathbf{t} , and the vector of the transfers to be charged to all users, $\boldsymbol{\tau} \in \mathbb{R}^M$. The information space together with the outcome correspondence are decisive factors as to what properties a certain mechanism possess. Determining the outcome is rendered by the CSM which does so based on the information received from WSTAs. Both \mathbf{t} and $\boldsymbol{\tau}$ are functions of the vectors of private information or types, represented by $\boldsymbol{\theta}_i, i = 1, \dots, M$, sent to the CSM from WSTAs. The transfers discussed in this paper could be monetary charges or other resources available at the WSTAs (e.g. computational resources). In the following subsections we formalize the arguments above.

A. The Mechanism

The goal of this subsection is to calculate the allocation of TXOPs on all N channels to M users such that the *system utility*, $U_{SYS}(\mathbf{t})$, becomes maximized. The following three steps form the resource management mechanism which take place at the beginning of every service interval.

- **Exchanging Information:** Each user i , transmits a vector of private information $\boldsymbol{\theta}_i$, to the CSM. We represent the vector of all transmitted information to the CSM by $\boldsymbol{\theta}$.
- **Allocating Times:** The CSM decides about the time allocations on all N channels in a way that maximizes the system utility U_{SYS} .
- **Computing Transfers:** The CSM calculates the transfers to be charged to users to prevent them from lying.

Figure 2 depicts the block diagram of the system. In the following we describe the three phases.

1) *Exchanging Information:* We assume that the information transmitted from user i to the CSM at the beginning of each SI is captured in the following two vectors: $\boldsymbol{\delta}_i$ and $\boldsymbol{\alpha}_i$ defined as follows:

- The vector of video distortion reductions that would result from the successful transmission of different packets in user i 's transmission queue of size $Q \cdot \bar{p}_i^{tot}$: $\boldsymbol{\delta}_i = [\Delta_{v_i^{q_1}}, \dots, \Delta_{v_i^{q_1+Q \times \bar{p}_i^{tot}-1}}]^T \in \mathbb{R}^{Q \cdot \bar{p}_i^{tot}}$. All packets are assumed to have the same delay constraint. The CSM sorts each vector of distortion reduction in descending order upon receipt;
- The vector $\boldsymbol{\alpha}_i$ of size N as defined in section IV.

The content of the information conveyed to the CSM is formed and sorted based on user i 's discretion and cross-layer strategy; e.g. how to schedule the packets and calculate corresponding distortion reductions in $\boldsymbol{\delta}_i$ or how to calculate accurate parameters $\bar{p}_{i,j}$ and $P_{i,j}^{succ}$ is fully depending on users' discretion.

2) *Allocating Times:* The optimization program of the CSM, represented by $[\text{OPT}(\{1, 2, \dots, M\})]$, can be shown by:

$$\begin{aligned} \max \quad & U_{SYS}(\mathbf{t}) \\ \text{s.t.} \quad & t_i(j) \geq 0 \end{aligned} \quad (12)$$

$$\sum_{i=1}^M t_i(j) = t_{SI} \quad \forall j = 1, \dots, N \quad (13)$$

where the optimization variables are $t_i(j), i = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$ which represent the time allocated to user i on channel j . The objective function is the system utility which shows that, by definition, the optimization is aimed at maximizing the aggregate utility. Constraint (12) simply enforces nonnegative allocated times while constraint (13) enforces that the sum of allocated times on each channel equals the length of a service interval t_{SI} . Using proposition 2, the problem $[\text{OPT}(\{1, 2, \dots, M\})]$ is a convex optimization program which could be solved very efficiently [22]. We represent the solution to the above optimization by \mathbf{t}^* . We also note that the optimization variables above are *continuous* times on each channel. In reality, after solving this optimization, we will round all allocated times to the closest integer multiple of t_{TXOP} . In other words we approach the problem by solving a convex relaxation of the original optimization which is in the number of TXOPs and hence, non-convex. This approximate approach is legitimate because of the following two reasons: First, the distortion reduction values are sorted in the decreasing order, and hence our piece-wise linear relaxation of the utility function is equivalent to their convex hull. Second, the number of packets in the queue is generally of the order of tens or hundreds. This makes the relative error incurred, in case of considering the convex relaxation of the problem, very small.

3) *Computing Transfers:* The next task of the CSM is to compute and announce the vector of transfers. The idea is that each user is charged based on the amount of net utility loss it causes other users. Formally, the CSM computes each transfer τ_i as follows:

$$\tau_i(\boldsymbol{\theta}) = \sum_{k \neq i} U_k(\mathbf{t}^*(\boldsymbol{\theta})) - \max_{\hat{\mathbf{t}} \in \mathbb{R}^{N \cdot (M-1)}} \sum_{k \neq i} U_k(\hat{\mathbf{t}}), \quad \forall i = 1, 2, \dots, M \quad (14)$$

The first term is the sum of real utilities that *other users* are making in presence of user i . The second summation however is the best aggregate utility that others would have made, had not user i been present at all. It is in fact the solution to

TABLE I
THE RESOURCE ALLOCATION ALGORITHM

For the current SI do:
1) For every WSTA $i = 1, 2, \dots, M$ do Poll WSTA i about its private information $\theta_i = [\delta_i, \alpha_i]$;
2) Compute the optimal time allocation by solving $[\text{OPT}(\{1, 2, \dots, M\})]$ and calculate t^*
3) For every WSTA $i = 1, 2, \dots, M$ do Compute WSTA's transfer, τ_i , according to (14);
4) Announce time allocations, t^* , and transfers, τ , to all WSTAs.
5) Transmission phase begins

$[\text{OPT}(\{1, \dots, i-1, i+1, \dots, M\})]$. The difference, which is always by construction non-positive, is the transfer associated with user i .

Definition: The mechanism in which the information exchange, decision on time allocation, and transfers are rendered according to the three steps described above is called *Clarke pivotal mechanism*.

There are two main reasons why we choose the Clarke mechanism. The first reason is that the transfers are computed in a very intuitive way. The intuition is that each user is being penalized for the *inconvenience* it causes to all other WSTAs. The second reason is that because $\tau_i \leq 0$ for $i = 1, 2, \dots, M$, the transfers in this mechanism are always in the form of charges and not incentives and therefore the mechanism is always feasible in terms of transfers, i.e. there will be no need for outside funds. The following proposition shows why the Clarke pivotal mechanism leaves no incentive for users to lie about their private information. First define θ_{-i} as the vector of information of all WSTAs except for user i .

Proposition 3: Assume that in our OSAR resource management problem, the Clarke pivotal mechanism is applied. Then no WSTA will have any incentive to lie about its real information regardless of what other WSTAs announce. In other words, it is dominant strategy incentive compatible.

Proof: Let us assume it is not true. Then, if WSTA i 's real information vector is represented by θ_i there should exist an information vector $\hat{\theta}_i$ such that user i receives more payoff by announcing $\hat{\theta}_i$ rather than the true information θ_i . In other words,

$$U_i(t^*(\theta_{-i}, \hat{\theta}_i), \theta_i) + \tau_i(\theta_{-i}, \hat{\theta}_i) > U_i(t^*(\theta_{-i}, \theta_i), \theta_i) + \tau_i(\theta_{-i}, \theta_i) \quad (15)$$

Replacing τ_i from (14), we get:

$$U_i(t^*(\theta_{-i}, \hat{\theta}_i), \theta_i) + \sum_{j \neq i} U_j(t^*(\theta_{-i}, \hat{\theta}_i), \theta_j) > U_i(t^*(\theta_{-i}, \theta_i), \theta_i) + \sum_{j \neq i} U_j(t^*(\theta_{-i}, \theta_i), \theta_j) \quad (16)$$

which is clearly a contradiction to the definition of t^* . Hence, such $\hat{\theta}_i$ can not exist for any i . ■

Thus, we can predict that every rational WSTA will announce the true information.

TABLE II
USERS' SPECIFICATIONS FOR SIMULATION 1

	Rate	Sequence	Resolution	Channels SNRs
User 1	384 kpbs	Foreman	352×288	[24dB 24dB]
User 2	2560 kpbs	Foreman	352×288	[24dB 24dB]

B. Implementing the Mechanism

In this subsection we will focus on more practical aspects of implementing the Clarke pivotal mechanism. The algorithm in table I recaps the implementation of the algorithm. The first step includes information exchange between each user and the CSM. Analyzing the tradeoff between more granular information and more overheads is the topic of our future research.

The next steps in the algorithm are the computation of the optimal time allocation and transfers. This amounts to solving $M+1$ convex optimization programs. In the first optimization, that finds the optimal time allocation, there are MN variables, MN linear inequalities, and N linear equalities while calculating every transfer requires solving an optimization program with $N(M-1)$ variables, $N(M-1)$ inequalities, and N equalities. Using an interior-point optimization method [22] with a logarithmic barrier function, the complexity of solving such optimization programs is of order \sqrt{MN} Newton iterations for an unconstrained convex optimization. Hence, the number of Newton iterations per SI is of order $M\sqrt{MN}$. In a deeper analysis of the complexity of the mechanism, we also find the growth of the number of basic operations per Newton iteration. Because of the very simple form of the equality and inequality constraints of the optimization program, which are linear and sparse, each Newton iteration takes basic operations of order MN .

VI. SIMULATION RESULTS

In this section, we present our simulation results. In order to achieve efficient streaming of video over OSAR networks, the application layer needs to accommodate instantaneous bandwidth variations due to time-varying channel conditions, network congestion, and/or the sudden emergence of primary users. Non-scalable video coding algorithms do not provide graceful degradation and adaptability to a large range of wireless channel conditions. Hence, although the concepts proposed in this paper can potentially be deployed with state-of-the-art non-scalable coding with bitstream switching, this usually entails higher complexity and smaller granularity for real-time bandwidth adaptation and packet prioritization [25]. Consequently, in this paper we use recently-proposed scalable video coding schemes based on Motion Compensated Temporal Filtering (MCTF) using wavelets [26]. Such a 3D wavelet video compression is attractive for wireless streaming applications since it provides on-the-fly adaptation to channel conditions, support for a variety of wireless receivers with different resource capabilities and power constraints, and easy prioritization of various coding layers and video packets. More details about the deployed 3D wavelet video coder can be found from [27].

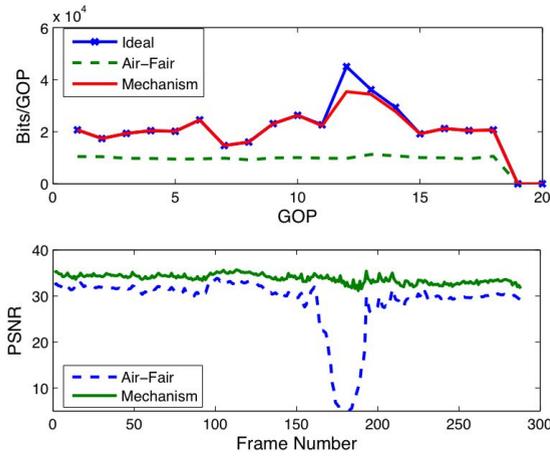


Fig. 3. Transmitted bits per GOP and PSNR for user 1 in the 2-user example.

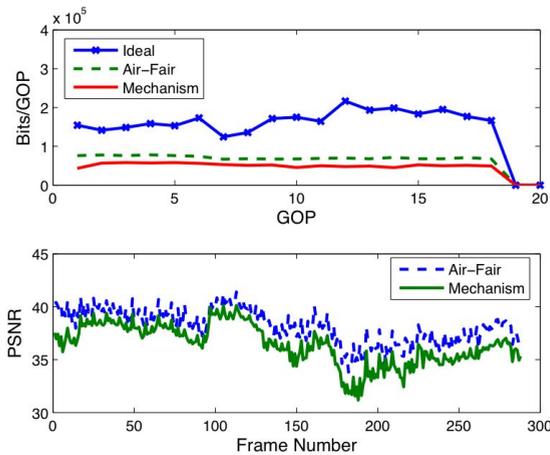


Fig. 4. Transmitted bits per GOP and PSNR for user 2 in the 2-user example.

A. A Simple Motivating Example

The first simulation, focuses on the comparison between the air-fair time paradigm and our mechanism on a specific OSAR network setup and aims at showing the incapability of the air-fair paradigm to adapt dynamically to users' needs. Table II lists the specifications of the 2 users present in the network. We assume that the network consists of 2 autonomous WSTAs transmitting real-time video over 2 OSAR transmission channels. We also assume $t_{SI} = 100\text{ms}$ and $t_{TXOP} = 10\text{ms}$. The sequences selected are CIF (352×288) with 288 frames at 30 frames per second. The packet deadlines are assumed 533ms for all packets. We use a Group Of Pictures (GOP) structure with 16 frames in each GOP, and a temporal decompositions with 4 temporal levels. We assume that no user lies about its rate requirement. Figures 3 and 4 depict the performance of the air-fair and Clarke mechanism. For reference, the graph for an ideal case is also shown, which corresponds to a case in which there is no resource limit and each WSTA can transmit at any arbitrary rate. The reported result is the instantaneous experienced PSNR (Peak Signal-to-Noise Ratio) and number of transmitted bits per GOP for each user. Since the air-fair time paradigm is a *static* time-allocation method, which does not take into account the relative importance of

TABLE III
USERS' AVERAGE EXPERIENCED PSNR FOR SIMULATION I

	User 1	User 2	Sum
Air-Fair Time	28.8dB	38.1dB	67.0dB
Mechanism	33.8dB	36.4dB	70.2dB

TABLE IV
SIMULATION SETUP

N	M	t_{SI}	t_{TXOP}
2	5	100ms	10ms

different packets, its performance is worse than our *dynamic* mechanism-based resource allocation method. Particularly, we consider the sudden drop in user 1's PSNR below 10dB in air-fair, which practically causes frozen video, at about frame number 160. The reason is that, due to sudden changes in video, user 1 temporarily needs more rate at around frame 160. The comparison shows how our mechanism handles this change dynamically and smoothly while air-fair time fails offering acceptable quality of service to user 1. The average experienced PSNRs of two users are reported in table III.

B. More Users and Spectrum Agility

In the next simulations, we consider a specific OSAR network infrastructure and quantify the performance of various resource management schemes: equal-time, air-fair time, and the proposed Clarke mechanism. We assume that the network consists of 5 autonomous WSTAs transmitting real-time video over 2 OSAR transmission channels. Other user and network specifications are the same as last simulation unless otherwise stated. We consider two congestion scenarios: In the first scenario, the network is mildly congested, i.e., the ratio of the aggregate required rate to the total available channel rate is close to 1, while in the second scenario, the network is more congested and the above ratio is much larger than 1. The experienced SNRs on the two channels for all users vary between 18dB and 29dB. Tables IV and V show the specifications of the simulation.

Case I: In the first simulation, we compare the performances of the three above resource management paradigms when user 1 announces its utility exaggerated by 30%. In other words, WSTA 1 is not a rational user and deviates from its own optimal strategy. In this case, WSTA 1 is penalizing other users by making them receive less resources; e.g. user 2 is receiving an unacceptable video quality of less than 25dB in the air-fair case. Figure 5 depicts the results. This undesirable penalty, is mitigated by the use of mechanism as compared with the two other cases: The Equal-time scenario, is clearly not efficient especially for high-demanding users; e.g. users 4 and 5, which need the largest amounts of resource, are allocated insufficient number of TXOPs and hence, they experience less than 24dB in the sense of PSNR. However, the proposed mechanism, performs much better than both aforementioned scenarios in which all users receive close to 27dB or more in terms of PSNR. This is a result of the high level of the content-awareness of the CSM.

Case II: In this case, we focus only on our mechanism and analyze the effects of changes in various assumptions that we

TABLE V
USERS' SPECIFICATIONS FOR SIMULATIONS 2,3,4,5

	Rate	Sequence	Resolution
User 1	384 kpbs	Foreman	352×288
User 2	384 kpbs	Foreman	352×288
User 3	1536 kpbs	Coastguard	352×288
User 4	2048 kpbs	Mobile	352×288
User 5	2048 kpbs	Mobile	352×288

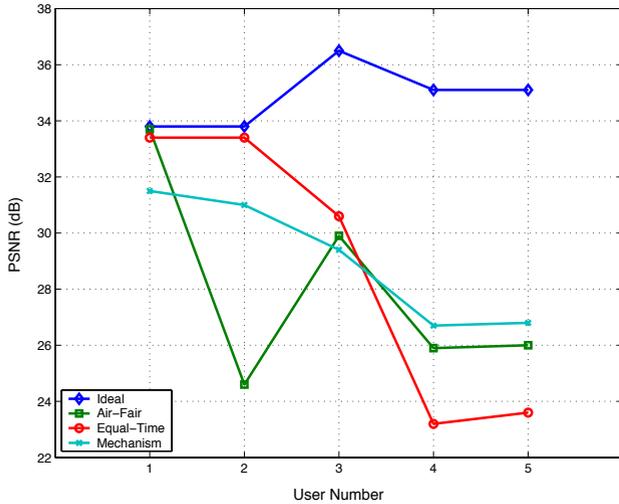


Fig. 5. High Congestion Case: Comparing experienced average PSNR for equal-time, air-fair time, and Clarke mechanism with user 1 lying.

made about users. We generally assume the following about the users: all users schedule their packets in the decreasing order of distortion reduction and they keep retransmitting a packet until it is successfully received or it is expired, and also users adjust their modulation scheme in the physical layer based on channel conditions. The above assumptions state to what extent a user is capable of doing link adaptation and taking the right strategy in playing the resource allocation game. We show through simulation results that the more advanced a user can adapt its strategy, the better video quality it can expect. This way the need for more advanced video coders and cross-layer strategy is well-justified. In the model we used, the mildly-congested network is chosen. Figure 6 depicts the experienced PSNRs of all users. In one scenario, user 5 has no packet scheduling, no retransmission of unsuccessful packets, and deploys a fixed modulation and coding schemes; i.e., this user is not smart compared to other WSTAs. From the graph it is clear that user 5 is doing worse than other users due to its lack of good strategy; it is experiencing more than 12dB loss in PSNR compared to the smarter strategy scenario. In the same figure we also show the results for another scenario where user 5 suffers from a bad estimation of channel SNRs. We simulated user 5 in a way that it always underestimates the quality of channels by about 10dB. This deficiency in its information about channels, results in flawed strategies and eventually considerable loss in the resulting PSNR which is more than 2dB compared to the true channel information.

Case III: In the next set of simulations we pick the highly-congested network as opposed to the last case. In two cases, again user 5 suffers from bad strategy, and flawed channel estimation as in the previous case. In the above two

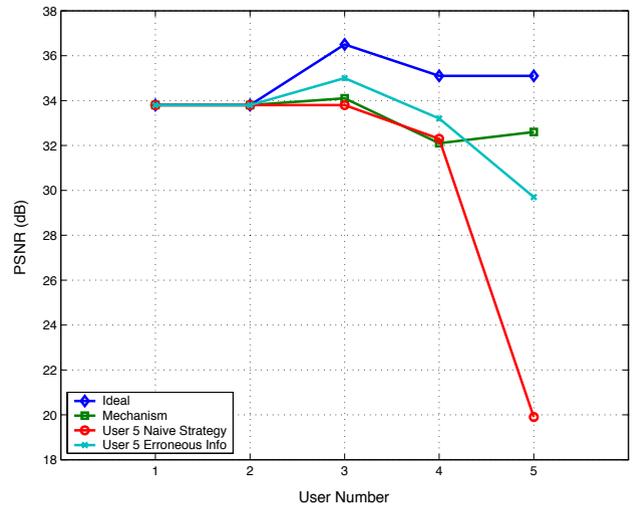


Fig. 6. Low Congestion Case: Comparing experienced average PSNR of users in cases of simple-strategy user and erroneous data user.

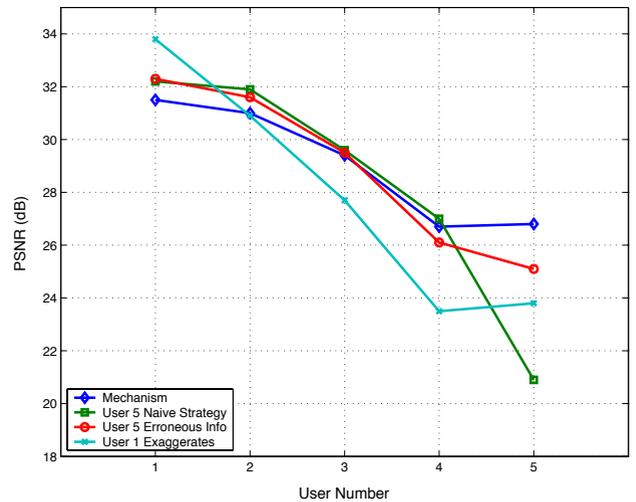


Fig. 7. High Congestion Case: Comparing experienced average PSNR of users in cases of simple-strategy user, erroneous data user, and exaggerating user.

scenarios, similar to case II, user 5 loses about 6dB and 2dB, respectively. Figure 7 shows the results. In a third scenario we assume that, even though the best external strategy for all users is to announce the real utility, user 1 announces an exaggerated version of its expected distortion reduction to the CSM; it always announces a constant number even when it has no packets to transmit. In this scenario, user 1 receives better PSNR which is a reasonable observation due to exaggerated announcements. Figure 8 shows that however user 1 is experiencing better PSNR, by paying more transfers, it is penalized for the extra resource it is claiming. According to proposition 3, the extra PSNR does not compensate the extra incurred charges.

Case IV: In this scenario, we consider a mildly-congested network where the specification of the users and the network is the same as before (as shown in tables IV and V) except for users 1 and 2, which now have higher rate requirements, i.e. their required rate equals 512kbps instead of 384kbps. A more dynamic network situation is considered here: the

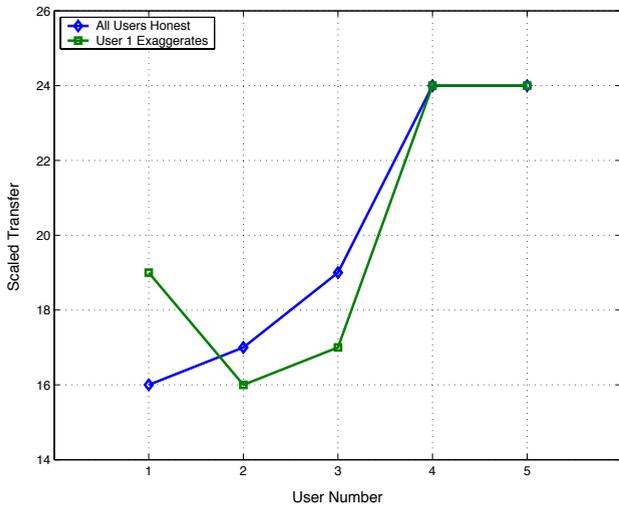


Fig. 8. High Congestion Case: Exaggerating in mechanism is not efficient; User 1 pays more.

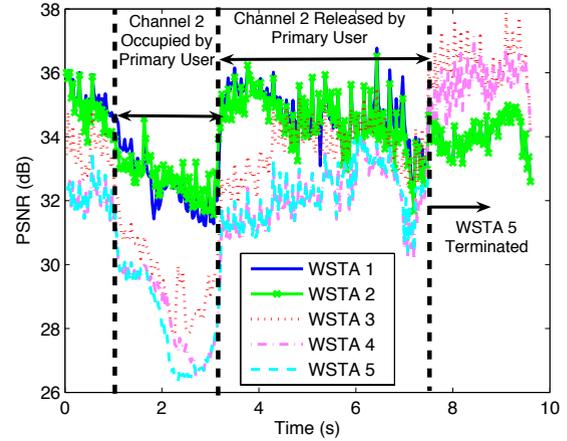


Fig. 10. OSAR Example; PSNRs of 5 users in case IV.

VII. CONCLUSION

In this paper, we considered a wireless communication network aimed at transmitting multimedia content in real-time over an OSAR infrastructure. We proposed a novel and new paradigm for resource allocation among competing WSTAs which has the potential capability of building a framework for arising OSAR protocol design. The proposed setup is a middle ground between fully centralized and fully decentralized paradigms: while there exists a special spectrum moderator called CSM, competing WSTAs adjust their cross-layer strategies dynamically based on their own discretion to play the resource allocation game as efficient as they can. We explained and also visited examples that exhibit deficiencies of the equal-time, air-fair time, and admission-control-based paradigms. In our method, each user announces a vector of private information to the CSM which consequently allocates available TXOPs in a socially-optimal way and charges transfers to users. We showed that using Clarke pivotal mechanism, unlike many proposed resource allocation schemes, announcing true private information is an equilibrium point of the resulting resource allocation game which is repeated every service interval. This result holds independently for any user regardless of what other WSTAs do; Hence, this mechanism is dominant strategy incentive compatible. We also showed by examples how taking weak strategies by WSTAs can result in a significant loss in terms of received video quality. The proposed resource management mechanism is scalable with respect to number of users, number of channels, and network load. Besides it is absolutely dynamic, making the proposed system fit in the spectrum agility framework. Another important extension of this work is the case when users are transmitting other data than video. As long as users can prioritize their applications and announce their packets and channel information to the CSM, our mechanism will successfully handle the network in exactly the same manner with no changes needed.

One of the important aspects of our future research is analyzing the effects of the granularity of the information transmitted to the CSM. The trade-off between more accurate results and less computations is the topic under scrutiny.

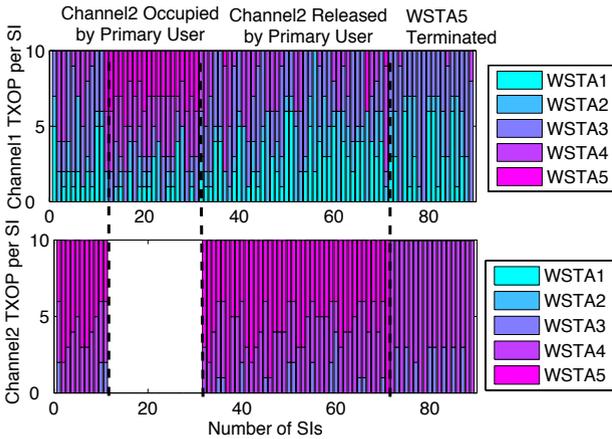


Fig. 9. OSAR Example; Time allocation to 5 users on 2 channels in case IV.

second channel is first occupied by an emerging primary user at time 1s; subsequently, the channel is released by the primary user at time 3.2s (available to the secondary users again), and finally, at time 7.5s, user 5 leaves the network because its transmission is terminated. Figures 9 and 10 show the channel allocations and the resulting PSNRs for all the secondary users, respectively. The appearance of the primary user leads to a graceful video quality drop for all the secondary users. However, as soon as the primary user leaves the network, the video qualities of all secondary users improve again. As mentioned above, the ability of the proposed mechanism to allocate resources in such a way that the secondary users experience only a graceful degradation in their video quality is essential in order to sustain real-time multimedia streaming applications. From the last simulation, we can conclude that our proposed mechanism-based resource allocation algorithm can successfully cope with the dynamic changes in channel conditions often experienced in OSAR networks.

Another issue to be further studied is the complexity of the proposed computations. In both WSTA side and the CSM's, considerable amounts of computations have to be performed at every SI. Hence, analyzing the computational burden on both sides is crucial.

REFERENCES

- [1] S. Shankar and M. van der Schaar, "Fundamentals of multimedia transmission over emerging spectrum agile wireless networks," *Tutorial IEEE ICIP Conference*, 2005.
 - [2] http://www.darpa.mil/ato/programs/XG/rfc_vision.pdf-Vision RFC.
 - [3] C. T. Chou, H. Kim, S. S. N, and K. G. Shin, "What and how much to gain from spectrum agility," *Submitted to IEEE/ACM Trans. on Networking*, June 2004.
 - [4] C. Cordeiro, K. Challapali, D. Birru, and S. Shankar, "Ieee 802.22: the first worldwide wireless standard based on cognitive radios," *DySPAN 2005. 2005 First IEEE International Symposium on*, pp. 328 – 337.
 - [5] C. Curescu and S. Nadjm-Tehrani, "Facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies," *the FCC Notice of Proposed Rulemaking and Order*, vol. ET Docket No. 03-108.
 - [6] S. S. N, C.-T. Chou, K. Challapali, and S. Mangold, "Spectrum agile radio: capacity and qos implications of dynamic spectrum assignment," *Global Telecommunications Conference*, vol. 5, no. 28, pp. 2510 – 2516, 2005.
 - [7] M. van der Schaar, Y. Andreopoulos, and Z. Hu, "Optimized scalable video streaming over ieee 802.11a/e hcca wireless networks under delay constraints," *accepted to IEEE Trans. On Mobile Computing*.
 - [8] R. W. Lucky, "Tragedy of the commons," *IEEE Spectrum*, vol. 43, p. 88, Jan 2006.
 - [9] M. van der Schaar and S. Shankar, "Cross-layer wireless multimedia transmission: Challenges, principles and new paradigms," *IEEE Wireless Communications Magazine*, Aug 2005.
 - [10] IEEE, "Ieee 802.11e/d5.0, draft supplement to part 11: Wireless medium access control (mac) and physical layer (phy) specifications: Medium access control (mac) enhancements for quality of service (qos)," 2003.
 - [11] D. Raychaudhuri, "Adaptive wireless networks using cognitive radios as a building block," *MobiCom2004 KeynoteSpeech*, 2004.
 - [12] F. Fangwen, A. Fattahi, and M. van der Schaar, "Game-theoretic paradigm for resource management in spectrum agile wireless networks," *Proceedings of the International Conference on Multimedia and Expo*, 2006.
 - [13] M. van der Schaar, S. Krishnamachari, S. Choi, and X. Xu, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 wlans," *Appl. Comput. Harmonic Anal.*, vol. 10, pp. 290–315, 2001.
 - [14] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.
 - [15] R.-F. Liao, R. Wouhaybi, and A. Campbell, "Wireless incentive engineering," *IEEE Journal of Selected Areas in Communications, Special Issue on Recent Advances in Multimedia Wireless*, 2003.
 - [16] N. Semret, R. Liao, A. Campbell, and A. Lazar, "Pricing, provisioning and peering: Dynamic markets for differentiated internet services and implications for network interconnections," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 12, pp. 2499–2513, 2000.
 - [17] C. Curescu and S. Nadjm-Tehrani, "Time-aware utility-based resource allocation in wireless networks," *IEEE Transactions on Parallel and Distributed*, vol. 16, no. 7, pp. 624 – 636, 2005.
 - [18] L. Badia, M. Lindstrom, J. Zander, and M. Zorzi, "Demand and pricing effects on the radio resource allocation of multimedia communication systems," *Proceedings IEEE Global Telecommunications Conference*, vol. 7, pp. 4116 – 4121, 2003.
 - [19] M. Wang and M. van der Schaar, "Operational rate-distortion modeling for wavelet video coders," *accepted for publication to IEEE Trans. On Signal Processing*.
 - [20] D. Krishnaswamy, "Network-assisted link adaptation with power control and channel reassignment in wireless networks," *3G Wireless Conference*, pp. 165–170, 2002.
 - [21] Q. Li and M. van der Schaar, "Providing adaptive qos to layered video over wireless local area networks through real-time retry limit adaptation," *IEEE Trans. on Multimedia*, 2004.
 - [22] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
 - [23] D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, 1991.
 - [24] M. Jackson, "Mechanism theory," *In the Encyclopedia of Life Support Systems*, 2003.
 - [25] J. Ohm, *Multimedia Communication Technology: Representation, Transmission and Identification of Multimedia Signals*. Springer, 2004.
 - [26] J. Ohm, M. van der Schaar, and J. Woods, "Interframe wavelet coding motion picture representation for universal scalability," *EURASIP Signal Processing: Image Communication, Special issue on Digital Cinema*, pp. 624 – 636, 2004.
 - [27] Y. Andreopoulos, A. Munteanu, J. Barbarien, M. van der Schaar, J. Cornelis, and P. Schelkens, "In-band motion compensated temporal filtering," *EURASIP Signal Processing: Image Communication (special issue on "Subband/Wavelet Interframe Video Coding"*, vol. 19, no. 7, pp. 653 – 673, 2004.
- Ahmad Reza Fattahi** received his BS degree with honors from Sharif University of Technology, Tehran, Iran in 1999. He received his MS degree in 2002 from UCLA where he is currently a PhD candidate in electrical engineering. His areas of research include resource management and control of distributed systems, including wireless communication networks where he is using optimization and economical methods to analyze and synthesize distributed systems. He has received several departmental and school-wide fellowships for his academic excellence. He is a student member of the IEEE and an associate member of Sigma Xi.
- Fangwen Fu** received his Bachelor and Master degrees both in electrical engineering from Tsinghua University, Beijing, China, in 2002 and 2005, respectively. He is currently in the Ph.D. program in Electrical Engineering Department, UCLA. His research interests include wireless multimedia streaming, resource management for networks and systems, applied game theory, video processing and analysis.
- Mihaela van der Schaar** received the M.S. and Ph.D. degrees from Eindhoven University of Technology, Eindhoven, The Netherlands, in 1996 and 2001, respectively. She is currently an Assistant Professor in the Electrical Engineering Department at UCLA. She has published extensively on multimedia compression, processing, communications, networking and architectures and holds 27 granted US patents. Since 1999, she was an active participant to the ISO Motion Picture Expert Group (MPEG) standard to which she made more than 50 contributions and for which she received three ISO recognition awards. She was an Associate Editor of IEEE Transactions on Multimedia and SPIE Electronic Imaging Journal and is currently an Associate Editor of IEEE Transactions on Circuits and System for Video Technology and of IEEE Signal Processing Letters. She received the NSF CAREER Award in 2004, IBM Faculty Award in 2005, Okawa Foundation Award in 2006, and the Best Paper Award for her paper published in 2005 in the IEEE Transactions on Circuits and Systems for Video Technology.
- Fernando Paganini** received his Electrical Engineering and Mathematics degrees from Universidad de la Republica, Montevideo, Uruguay, in 1990, and his M.S. (1992) and PhD (1996) degrees in Electrical Engineering from Caltech, Pasadena, CA. In 1996-1997 he was a postdoctoral associate at MIT. From 1997 to 2006 he was Assistant, later Associate Professor of Electrical Engineering at UCLA. He is currently a Professor at Universidad ORT Uruguay. His research interests are control and networks. Dr. Paganini is the recipient of the O. Hugo Schuck best paper award at the 1994 American Control Conference, the 1996 Wilts and Clouser Prizes for his PhD Thesis at Caltech, the 1999 NSF CAREER Award, the 1999 Packard Fellowship, and the 2004 George S. Axelby award from the IEEE Control Systems society.